4

Parametric regression models of survival analysis

In some sense, regression modeling is the hallmark of the modern survival analysis, especially in behavioral science. As indicated in Chapter 2, descriptive approaches do not possess the capability to test causal effects because some 'lurking' covariates can significantly confound the true association between two factors under examination. Regression modeling provides a powerful methodology to test theoretical hypotheses about causality and compare effects across population groups, given its capacity to adjust for the influences of potential confounders. In survival analysis, regression models are used for analyzing the causal linkage between an outcome lifetime variable (such as the hazard rate, the event time, or the survival function) and one or more independent variables, with one or more variables serving as controls. As the specification of independent and control variables involves prior knowledge and theoretical hypotheses, construction of a lifetime regression model needs to be guided by existing theories, previous findings, and specific research questions. Therefore, regression modeling, both generally and with specific regard to survival analysis, is a theory-based methodology, not just an amalgamation of mathematical rules, statistical procedures, and algebraic manipulations.

In this chapter, I describe parametric regression modeling on survival data, which generally points to statistical techniques combining a known parametric distribution of survival times with multivariate regression modeling procedures. In particular, I first portray general specifications and statistical inferences of parametric regression modeling. Then, several widely used parametric regression models, such as the exponential, the Weibull, and the log-logistic regression models, are illustrated with empirical examples. Some other parametric regression models are also briefly described. Special attention is paid to the Weibull regression model, given its widespread applicability in survival analysis and high flexibility in describing empirical lifetime data. Lastly, I summarize the chapter with a comparison of several key parametric regression models.

Survival Analysis: Models and Applications, First Edition. Xian Liu.

4.1 General specifications and inferences of parametric regression models

Like constructing other types of regression models, parametric regression modeling in survival analysis starts with the specification of parameters. First, for illustrative convenience, I specify a causal model involving a number of factors, including one or more independent variables and one or more control variables, and a dependent lifetime variable. Statistically, both the independent and the control variables are regarded as covariates in a multivariate regression model. According to a specific theoretical framework, the covariates, denoted by $\boldsymbol{x}$, a $1 \times M$ vector containing elements $(x_1, x_2, \ldots)'$, are those causally associated with the dependent lifetime outcomes. The dependent variable is the event time T, which may be censored. In parametric regression perspectives, this dependent variable is assumed to follow a known probability distribution whose parameters may depend on $\boldsymbol{x}$. The effect of covariates on the parameters can be modeled either using the survival time or the hazard rate at time t as a function of a parameter vector $\boldsymbol{\theta}$. Empirically, $\boldsymbol{x}$ may contain explanatory variables (independent and control) and some theoretically relevant interactions predictive of the survival time or of the hazard rate, such as medical treatment, severity of a given medical condition, and the interaction between those two variables in a typical biomedical study. The elements in $\boldsymbol{x}$ are sometimes called the *exogenous variables* as their causes are not generated within a regression model.

There are two popular perspectives in terms of parametric regression modeling. The first perspective is the parametric hazard rate model in which covariates impact the hazard function. In this class of parametric regression modeling, the proportional hazard rate model, in which the effects of covariates are assumed to be multiplicative, is most common. The second class uses the log transformation of event times as the outcome measure, often assumed to be linearly associated with covariates. In this section, I describe general specifications of these two approaches and the associated likelihood functions in the presence of right censoring.

4.1.1 Specifications of parametric regression models on the hazard function

In terms of parametric regression models on the hazard function, each observation under investigation is assumed to be subject to an instantaneous hazard rate, $h(t)$, of experiencing a particular event, where $t = 0, 1, \ldots, \infty$. Because the hazard rate is always nonnegative, the effects of covariates are generally specified as a multiplicative term $\exp(\boldsymbol{x}'\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ represents a vector of regression coefficients to be estimated. Consequently, the hazard rate model on the causal relationships between covariates and lifetime processes is written by

$$h(t|\boldsymbol{x}) = h_0(t)\exp(\boldsymbol{x}'\boldsymbol{\beta}) \tag{4.1}$$

where $h_0(t)$ denotes a known baseline hazard function for continuous time T and $\boldsymbol{\beta}$ provides a set of the effects of covariates on the hazard rate, with the same length as $\boldsymbol{x}$. As can be seen from Equation (4.1), the effects of $\boldsymbol{x}$ on the hazard rate are multiplicative, or proportional, so that the predicted value of $h(t)$, given $\boldsymbol{x}$, denoted by $\hat{h}(t|\boldsymbol{x})$, can be restricted in the range $(0, \infty)$. Nonmultiplicativity of the effects of given covariates, if it exists, can be captured by specification of one or more interaction terms, contained in $\boldsymbol{x}$. Given this specification, the class of parametric regression models represented by Equation (4.1) is generally referred to as the *proportional hazard rate model*. In this survival perspective, all individuals are assumed to follow a univariate hazard function, with individual heterogeneity primarily reflected in the proportional scale change for a stratification of distinct population subgroups; that is, any two individuals have hazard rates that are constant multiples of one another (Lawless, 2003). The concept of *proportionality* in this context, nevertheless, points more to a statistical perspective than to a substantive linkage. With specification of interaction terms for some covariates, for example, the effects of those variables are no longer proportional or multiplicative. The issue of nonproportionality will be discussed further in later chapters.

The term $h_0(t)$ in Equation (4.1) represents a parametric baseline hazard function within the context of parametric regression modeling, attaching to a particular probability distribution of event time T. For example, if the baseline hazard rate is constant throughout an observation interval, we have $h_0(t) \equiv \lambda$, which leads to an *exponential regression model*, given by

$$h(t, \boldsymbol{x}; T \sim Exp) = \lambda \exp(\boldsymbol{x}'\boldsymbol{\beta}). \tag{4.2}$$

Equation (4.2) shows that the exponential regression model on the hazard rate is simply the product of a constant baseline hazard rate and a multiplicative term representing the effect of the covariate vector $\boldsymbol{x}$.

Similarly, when the observed hazard function varies monotonically over time, the Weibull regression model may be applied, given by

$$h(t, \boldsymbol{x}; T \sim \text{Weib}) = \lambda \tilde{p}(\lambda \boldsymbol{t})^{\tilde{p}-1} \exp(\boldsymbol{x}'\boldsymbol{\beta}), \tag{4.3}$$

where the definitions of λ and $\tilde{p}$ are described in Chapter 3 (Section 3.2). Equation (4.3) displays the Weibull regression mode as the Weibull distributional function, represented by Equation (3.17), multiplied by the effect term $\exp(\boldsymbol{x}'\boldsymbol{\beta})$.

Given its intimate association with the hazard function, the survival function for T given $\boldsymbol{x}$ can be readily derived from Equation (4.1):

$$\begin{aligned} S(t; \boldsymbol{x}) &= \exp\left[-\int_0^t h_0(u)\,\mathrm{d}u \exp(\boldsymbol{x}'\boldsymbol{\beta})\right] \\ &= \exp\left[-\exp(\boldsymbol{x}'\boldsymbol{\beta}) \int_0^t h_0(u)\,\mathrm{d}u\right] \\ &= \exp[-\exp(\boldsymbol{x}'\boldsymbol{\beta}) H_0(t)], \end{aligned} \tag{4.4}$$

where $H_0(t)$ is defined as the continuous cumulative baseline hazard function at time t.

After some algebra, Equation (4.4) can be converted to the following formation:

$$\begin{aligned} S(t; \boldsymbol{x}) &= \{\exp[-H_0(t)]\}^{\exp(\boldsymbol{x}'\boldsymbol{\beta})} \\ &= [S_0(t)]^{\exp(\boldsymbol{x}'\boldsymbol{\beta})}, \end{aligned} \tag{4.5}$$

where $S_0(t)$ is the baseline survival function, mathematically defined as

$$S_0(t) = \exp\left[-\int_0^t h_0(u)\,\mathrm{d}u\right]$$
$$= \exp[-H_0(t)].$$

Equation (4.5) demonstrates that the survival function at t given $\boldsymbol{x}$ can be expressed as the baseline survival function $S_0(t)$ raised to the power of the multiplicative effects of covariates on the hazard function.

Given Equations (4.1) and (4.5), the probability density function of T given $\boldsymbol{x}$ is

$$f(t;\boldsymbol{x}) = h_0(t)\exp(\boldsymbol{x}'\boldsymbol{\beta})\exp\left[-\exp(\boldsymbol{x}'\boldsymbol{\beta})\int_0^t h_0(u)\,\mathrm{d}u\right]. \tag{4.6}$$

Parametric regression models on the hazard rate can be conveniently expressed in terms of a generalized linear regression function. Taking log values on both sides of Equation (4.1) yields

$$\begin{aligned}\log[h(t|\boldsymbol{x})] &= \log[h_0(t)] + \boldsymbol{x}'\boldsymbol{\beta}\\ &= \alpha^* + \boldsymbol{x}'\boldsymbol{\beta},\end{aligned} \tag{4.7}$$

where α^* is the logarithm of the baseline hazard function, serving as the intercept in this transformed linear regression function.

Equation (4.7) reflects some distinct similarities between the hazard rate model and other types of generalized linear regression models, such as the logistic regression. Without specification of an explicit baseline hazard function, however, Equation (4.7) does not display a direct link between $\boldsymbol{x}$ and event time T. Additionally, the model does not conveniently include a term representing random disturbances as do other generalized linear models. Empirically, specification of an error term in Equation (4.7) may raise some prediction problems because the hazard rate is not observable. These limitations can be well addressed by using another family of parametric regression models that regresses the logarithm of the event time over covariates, as described below.

4.1.2 Specifications of accelerated failure time regression models

As indicated earlier, parametric regression models in survival analysis can be created on log time over covariates, from which a different set of parameters needs to be specified. Suppose that $Y = \log T$ is linearly associated with the covariate vector $\boldsymbol{x}$. Then

$$Y = \mu^* + \boldsymbol{x}'\boldsymbol{\beta}^* + \tilde{\sigma}\varepsilon, \tag{4.8}$$

where $\boldsymbol{\beta}^*$ is a vector of regression coefficients on Y or $\log T$, μ^* is referred to as the intercept parameter, and the parameter $\tilde{\sigma}$ is an unknown scale parameter. The term ε represents random errors that follow a particular parametric distribution of T with survival function S, cumulative distribution function F, and probability density function f. Compared to Equation (4.7), Equation (4.8) looks more like a standard generalized linear regression model, thereby producing tremendous appeal for statisticians. With the location term ($\boldsymbol{x}'\boldsymbol{\beta}^*$) and scale parameter

$\tilde{\sigma}$, the baseline parametric distribution of survival times can be conveniently modeled by a term of additive random disturbances ($\log T_0$).

Using the equation $\log T \geq \log t$, the survival function, conditional on $\boldsymbol{x}$, can be expressed in the form of an extreme value distribution, as discussed in Chapter 3. Given Equation (4.8), the survival function for the ith individual is

$$S_i(t) = \mathrm{P}\left[\left(\mu^* + \boldsymbol{x}_i'\boldsymbol{\beta}^* + \tilde{\sigma}\varepsilon_i\right) \geq \log t\right]$$
$$= \mathrm{P}\left(\varepsilon_i \geq \frac{\log t - \mu^* - \boldsymbol{x}_i'\boldsymbol{\beta}^*}{\tilde{\sigma}}\right). \tag{4.9}$$

Let ε_i be a component of the error vector $\boldsymbol{\varepsilon}$, $S(t) = \Pr(\varepsilon_i \geq t)$, $F(t) = \Pr(\varepsilon_i < t)$, and $f(t) = \mathrm{d}F(t)/\mathrm{d}t$. Because the term $\boldsymbol{x}'\boldsymbol{\beta}^*$ is independent of the disturbance parameter ε_i, the survival function $S(t)$ with respect to $\log T$ can be modeled by specifying a random component and a fixed component, given by

$$S(t|\boldsymbol{x}) = S_0\left(\frac{\log t - \mu^* - \boldsymbol{x}'\boldsymbol{\beta}^*}{\tilde{\sigma}}\right), \quad -\infty < \log t < \infty, \tag{4.10}$$

where S_0, independent of $\boldsymbol{x}$, is the survival function of the distribution of ε and $\boldsymbol{x}'\boldsymbol{\beta}^*$ defines the location of T, referred to as an *accelerated factor*.

As $H(t) = -\log S(t)$, the cumulative hazard function can be expressed in terms of Equation (4.10):

$$H(t|\boldsymbol{x}) = -\log S_0\left(\frac{\log t - \mu^* - \boldsymbol{x}'\boldsymbol{\beta}^*}{\tilde{\sigma}}\right)$$
$$= H_0\left(\frac{\log t - \mu^* - \boldsymbol{x}'\boldsymbol{\beta}^*}{\tilde{\sigma}}\right), \quad -\infty < \log t < \infty. \tag{4.11}$$

where H_0 is the cumulative hazard function of $\boldsymbol{\varepsilon}$.

Similarly, by differentiating Equation (4.11), the corresponding hazard function is

$$h(t|\boldsymbol{x}) = \frac{1}{\tilde{\sigma}t} h_0\left(\frac{\log t - \mu^* - \boldsymbol{x}'\boldsymbol{\beta}^*}{\tilde{\sigma}}\right), \quad -\infty < \log t < \infty. \tag{4.12}$$

where h_0 is the baseline hazard function for the distribution of ε_i, also independent of $\boldsymbol{x}$.

Given its simplicity and flexibility, the above log-linear model, as represented by Equation (4.10), can be applied to formulate a large number of families of parametric distributions in survival analysis, including the exponential, the Weibull, the extreme value, the normal, the logistic, the lognormal, and the gamma distributions. In contrast, the parametric proportional hazard perspective only applies to few parametric distributions. A selected distribution of ε_i defines the formation of the intercept, scale, and shape parameters, in turn deriving the survival, the hazard, and the density functions, as will be described extensively in the following sections.

As defined, if $Y = \log T$ follows an extreme value distribution, T has an *accelerated failure time distribution* (Lawless, 2003; Meeker and Escobar, 1998). As the effect of covariate vector $\boldsymbol{x}$ on T changes the location, but not the shape, of a distribution of T, the parametric regression models that can be formulated by Equation (4.10) are generally referred to as *accelerated failure time (AFT) regression models* or *log-location-scale regression models* (in much of the following text, I call them simply the AFT regression models). One added advantage of using the specification of the AFT regression is that it covers a wide range of the survival time distribution, given the range of $\log T$ $(-\infty, \infty)$.

Using the above description, the AFT regression perspective can also be formulated with respect to the random variable T, rather than to $\log T$, for convenience of specifying a likelihood function. As Equation (4.8) is a log-linear model, exponentiation of both sides of this equation leads to

$$\begin{aligned} T &= \exp\left(\mu^* + \boldsymbol{x}'\boldsymbol{\beta}^*\right)\exp\left(\sigma\varepsilon\right) \\ &= \exp\left(\mu^* + \boldsymbol{x}'\boldsymbol{\beta}^*\right)\tilde{E}, \end{aligned} \tag{4.13}$$

where $\tilde{E} = \exp\left(\boldsymbol{\sigma\varepsilon}\right) > 0$ has the hazard function $h_0\left(\tilde{e}\right)$ and is independent of $\boldsymbol{\beta}^*$. Because $\exp(\boldsymbol{x}'\boldsymbol{\beta}^*)$ is positive valued, T is restricted in the range $(0, \infty)$. Given the equation $T \geq t$, the survival function for the ith individual can be mathematically expressed by

$$S_i\left(t\right) = \mathrm{P}\left(T_i \geq t\right) = \mathrm{P}\left[\exp\left(\mu^* + \boldsymbol{x}'\boldsymbol{\beta}^* + \tilde{\sigma}\varepsilon_i\right) \geq t\right]. \tag{4.14}$$

As the term $\boldsymbol{x}'\boldsymbol{\beta}^*$ is independent of the disturbance parameter, the survival function $S(t)$ can be written by two separate components:

$$S_i\left(t\right) = S_0\left[t\exp\left(\boldsymbol{x}'\boldsymbol{\beta}^*\right)\right], \tag{4.15}$$

where S_0 is a fully specified survival function, defined as

$$S_0\left(t\right) = \mathrm{P}\left[\exp\left(\mu^* + \tilde{\sigma}\varepsilon\right) \geq t\right].$$

The hazard function for T can be readily derived from Equation (4.15), given its intimate relationship with the survival function:

$$h\left(t|\boldsymbol{x}\right) = h_0\left[t\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\right]\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right). \tag{4.16}$$

From Equation (4.16), the survival function can be respecified in terms of the AFT hazard function:

$$\begin{aligned} S\left(t|\boldsymbol{x}\right) &= \exp\left\{-\int_0^t h_0\left[u\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\right]\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\mathrm{d}u\right\} \\ &= \exp\left\{-H_0\left[t\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\right]\right\}. \end{aligned} \tag{4.17}$$

Given Equations (4.16) and (4.17), the density function in the formation of such an AFT perspective is defined as

$$f\left(t;\boldsymbol{x}\right) = h_0\left[t\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\right]\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\exp\left\{-H_0\left[t\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\right]\right\}. \tag{4.18}$$

In the AFT regression models, the effect of covariates determines the time scale in such a way that if $\exp(\boldsymbol{x}'\boldsymbol{\beta}^*) > 1$, the effect of the covariate vector $\boldsymbol{x}$ is to decelerate the survival process, and if $\exp(\boldsymbol{x}'\boldsymbol{\beta}^*) < 1$, the effect is to accelerate it (Lawless, 2003). Obviously, such multiplicative effects of covariates on survival times take the opposite direction to those on the hazard function – $\exp(\boldsymbol{x}'\boldsymbol{\beta})$ – unless two corresponding coefficients are both 0. Statistically, the two sets of regression coefficients are intimately associated and mutually convertible if a parametric distribution of T can be formulated by both the proportional hazard and the AFT perspectives.

While the parametric hazard rate and the AFT models represent two different parametric families, two parametric functions can be expressed in terms of both perspectives – the exponential and the Weibull distributions. In the Weibull regression model, for example, the two sets of regression coefficients, $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$, reflect two parameter sets under the same parametric distribution of T; therefore, one set of parameters can be readily converted to the other. As a special case of the Weibull regression model, the exponential regression model has the same feature. These issues will be extensively discussed in the following two sections.

4.1.3 Inferences of parametric regression models and likelihood functions

Statistical inference of parametric regression models is performed to specify likelihood functions for the estimation of the parameter vector $\boldsymbol{\theta}$ in the presence of censoring. Specifically, a likelihood function with survival data describes the probability of a set of parameter values given observed lifetime outcomes. As generally applied in generalized linear regression, it is more convenient to generate parameter estimates by maximizing a log-likelihood function than by maximizing a likelihood function itself, given its relative simplicity. The unique aspect of statistical inference in survival analysis is the way of handling censoring. For analytic convenience, censoring is usually assumed to be random; that is, conditional on model parameters, censored times are assumed to be independent of each other and of actual survival times. If this assumption holds, censoring is noninformative, so that unbiased parameter estimates can be derived.

For considering right censoring in a likelihood function of parametric regression modeling, I first review some general conditions and random processes described in Chapter 1. Specifically, for each individual in a random sample and with the definition of a parameter vector $\boldsymbol{\theta}$, survival processes can be described by three random variables (t_i, δ_i, $\boldsymbol{x}_i$). The random variable of time t_i is defined as

$$t_i = \min(T_i, C_i) \tag{4.19}$$

and δ_i is given by

$$\begin{cases} \delta_{\mathrm{i}} = 0 \text{ if } T_i > t_i, \\ \delta_i = 1 \text{ if } T_i = t_i. \end{cases} \tag{4.20}$$

The random variables contained in the covariate vector $\boldsymbol{x}$ are specified earlier.

Given these random variables, a likelihood function with respect to the parameter vector $\boldsymbol{\theta}$ for a random sample of n individuals can be readily specified:

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} L_i(\theta) \\
&= \prod_{i=1}^{n} f(t_i; \boldsymbol{\theta}, \boldsymbol{x}_i)^{\delta_i} S(t_\mathrm{i}; \boldsymbol{\theta}, \boldsymbol{x}_i)^{1-\delta_i}. \qquad (4.21)
\end{aligned}
$$

As defined, the likelihood function $L(\boldsymbol{\theta})$ is the probability of a set of parameter values given n observed lifetime outcomes. When $\delta_i = 1$, $L_i(\boldsymbol{\theta})$ is the probability density function for the occurrence of a particular event as the second term in the second equation is 1. Likewise, when $\delta_i = 0$, $L_i(\boldsymbol{\theta})$ is the survival function for a censored time as the first term is 1. In either case, individual i's survival time, actual or censored, is accounted for in the likelihood.

The above likelihood function can be expanded in terms of a parametric regression model on the hazard rate characterized with a baseline hazard function and a vector of regression coefficients $\boldsymbol{\beta}$ including the intercept parameter. Suppose that the survival data with random variables $(t_i, \delta_i, \boldsymbol{x}_i)$ come from a parametric hazard rate regression with parameter vector $\boldsymbol{\theta}$. Then the likelihood function can be written as

$$
L(\boldsymbol{\theta}) = \prod_{i=1}^{n} [h_0(t)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})]^{\delta_i} \exp\left[-\int_0^t h_0(u)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{d}u\right]. \qquad (4.22)
$$

With the hazard and the survival functions explicitly expressed in Equation (4.22), a third function, the p.d.f. of T, is implicitly reflected therein because, when $\delta_i = 1$, the product of the first and second terms on the right of the equation gives rise to the density function.

Taking log values on both sides of Equation (4.22), a log-likelihood function is

$$
\begin{aligned}
\log L(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \left\{\delta_i \log[h_0(t)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})] - \int_0^t [h_0(u)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})]\mathrm{d}u\right\} \\
&= \sum_{i=1}^{n} \left\{\delta_i [\log h_0(t) + (\boldsymbol{x}_i'\boldsymbol{\beta})] - \int_0^t h_0(u)\mathrm{d}u \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\right\}. \qquad (4.23)
\end{aligned}
$$

Clearly, the log-likelihood function is computationally simpler than the likelihood function itself because products become sums and exponents transform to coefficients. Thus, it is easier to maximize this log-likelihood function with respect to the unknown parameter vector $\boldsymbol{\theta}$, which may contain some specific functional factors as well as regression coefficients, given a specific baseline distributional function. The generalization of a maximum likelihood estimator for $\boldsymbol{\theta}$ in the hazard rate regression model will be described in next subsection.

The likelihood function can be also constructed in terms of the AFT regression perspective. In the presence of right censoring, the likelihood function of an AFT regression model, given Equations (4.17) and (4.18), is

$$
L(\theta) = \prod_{i=1}^{n} \left\{h_0(t)\left[t\exp\left(-\boldsymbol{x}_i'\boldsymbol{\beta}^*\right)\right]\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\right\}^{\delta_i} \exp\left\{-H_0\left[\mathrm{t}\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\right]\right\}, \qquad (4.24)
$$

where the parameter vector $\boldsymbol{\theta}$ now contains different components from those on the hazard function.

The log-likelihood function of an AFT regression model can be derived by taking log values of both sides of Equation (4.24), given by

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\langle \delta_i \log\left\{h_0(t)\left[t\exp\left(-\boldsymbol{x}_i'\boldsymbol{\beta}^*\right)\right]\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\right\} - H_0\left[t\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\right]\right\rangle$$
$$= \sum_{i=1}^{n} \left\{\delta_i \left[\log h_0(t) + \log t - (\boldsymbol{x}_i'\boldsymbol{\beta})^2\right] - H_0\left[t\exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right)\right]\right\}. \tag{4.25}$$

Maximization of Equation (4.25) with respect to $\boldsymbol{\theta}$ yields the maximum likelihood (ML) estimates of the parameters contained in $\boldsymbol{\theta}$ in the AFT perspective.

4.1.4 Procedures of maximization and hypothesis testing on ML estimates

There are standard procedures to maximize a likelihood function with respect to the parameter vector $\boldsymbol{\theta}$, thereby generating robust, efficient, and consistent parameter estimates. Specifically, the process starts with the construction of a score statistic vector, denoted by $\tilde{\boldsymbol{U}}_i(\boldsymbol{\theta})$ for individual i and mathematically defined as the first partial derivatives of the log-likelihood function with respect to $\boldsymbol{\theta}$, given by

$$\tilde{\boldsymbol{U}}_i(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log L_i(\boldsymbol{\theta}) = \left[\frac{\partial}{\partial \theta_m} \log L_i(\boldsymbol{\theta})\right]_{M\times 1}, \tag{4.26}$$

where M is the dimension of $\boldsymbol{\theta}$ and θ_m is the mth component in $\boldsymbol{\theta}$ ($m = 1, \ldots, M$). This generalized likelihood function can apply to either Equation (4.23) or Equation (4.25) given different components in $\boldsymbol{\theta}$. The minus second derivatives of the log-likelihood function yields the estimator of variances and covariances for $\tilde{\boldsymbol{U}}_i(\boldsymbol{\theta})$:

$$\tilde{\boldsymbol{V}}_i(\boldsymbol{\theta}) = -\left(\mathrm{E}\frac{\partial^2 \log L_i}{\partial \theta_m \theta_{m'}}\right)_{M\times M}. \tag{4.27}$$

With right censoring assumed to be noninformative, the central limit theorem applies to the generation of the following total score statistic:

$$\tilde{\boldsymbol{U}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \tilde{\boldsymbol{U}}_i(\boldsymbol{\theta}). \tag{4.28}$$

Statistically, $\tilde{\boldsymbol{U}}(\boldsymbol{\theta})$ is asymptotically normal, given the large-sample approximation to the joint distribution of parameters, with mean 0 and variance–covariance matrix

$$\tilde{\boldsymbol{V}}(\boldsymbol{\theta}) = \sum\nolimits_{i=1}^{n} \tilde{\boldsymbol{V}}_i(\boldsymbol{\theta}).$$

Consequently, the parameter vector $\boldsymbol{\theta}$ can be estimated efficiently by solving the equation

$$\tilde{\boldsymbol{U}}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta}) = \mathbf{0}. \tag{4.29}$$

The above procedure is the formalization of a typical maximum likelihood estimator (MLE). The vector containing MLE parameter estimates are generally referred to as $\hat{\boldsymbol{\theta}}$. For a large sample, $\hat{\boldsymbol{\theta}}$ is the unique solution of $\tilde{\boldsymbol{U}}(\boldsymbol{\theta}) = \mathbf{0}$, so that $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}$ and distributed as multivariate normal, given by

$$\hat{\boldsymbol{\theta}} \sim N\left[\mathbf{0}, \tilde{\boldsymbol{V}}(\boldsymbol{\theta})^{-1}\right]. \tag{4.30}$$

This asymptotic distribution facilitates testing of hypotheses on $\boldsymbol{\theta}$. Specifically, the above maximum likelihood estimator is based on the observed Fisher information matrix, denoted by $\boldsymbol{I}\left(\hat{\boldsymbol{\theta}}\right)$, given by

$$\boldsymbol{I}\left(\hat{\boldsymbol{\theta}}\right) = -\left(\frac{\partial^2 \log L\left(\hat{\boldsymbol{\theta}}\right)}{\partial \hat{\theta}_m \hat{\theta}_{m'}}\right)_{M \times M}. \tag{4.31}$$

As can be easily recognized, $\tilde{\boldsymbol{V}}(\boldsymbol{\theta})$, formulated in Equation (4.27), contains the expected values of $\boldsymbol{I}\left(\hat{\boldsymbol{\theta}}\right)$.

As the replacement of $\tilde{\boldsymbol{V}}(\boldsymbol{\theta})$ with $\boldsymbol{I}\left(\hat{\boldsymbol{\theta}}\right)$ does not affect the asymptotic distributions of $\tilde{\boldsymbol{U}}(\boldsymbol{\theta})$, the term

$$\tilde{\boldsymbol{U}}'\left(\hat{\boldsymbol{\theta}}\right)\boldsymbol{I}\left(\hat{\boldsymbol{\theta}}\right)^{-1}\tilde{\boldsymbol{U}}\left(\hat{\boldsymbol{\theta}}\right) \tag{4.32}$$

is distributed asymptotically as $\chi^2_{(M)}$, so that the hypotheses about $\boldsymbol{\theta}$ can be statistically tested given a value of α, the degree of freedom, and the null hypothesis H_0: $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$ or simply $\hat{\boldsymbol{\theta}} = \mathbf{0}$. This test statistic on $\hat{\boldsymbol{\theta}}$ is the well-known score test.

The inverse of the observed information matrix yields the estimators of the variances and covariances for parameter estimates. For the parametric hazard regression model, for example, the variance–covariance matrix for the estimates of $\boldsymbol{\beta}$ can be written as

$$\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{I}\left(\hat{\boldsymbol{\beta}}\right)^{-1}. \tag{4.33}$$

The standard errors of $\boldsymbol{\beta}$ can be estimated by taking the square roots of the diagonal elements contained in $\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\beta}}\right)$.

As $\hat{\boldsymbol{\theta}}$ is asymptotically $N_M[\boldsymbol{\theta}, \boldsymbol{I}(\theta)^{-1}]$, $\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)$ is asymptotically $N_M[\mathbf{0}, n\boldsymbol{I}(\theta)^{-1}]$. Consequently, hypothesis testing on $\hat{\boldsymbol{\theta}}$ can also be performed by another test statistic for H_0: $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$:

$$\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)'\boldsymbol{I}\left(\hat{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right), \tag{4.34}$$

which is also distributed as $\chi^2_{(M)}$. Specifically, $\boldsymbol{I}\left(\hat{\boldsymbol{\theta}}\right)/n$ is an asymptotically consistent estimator for $\boldsymbol{I}(\boldsymbol{\theta})/n$. This test statistic, referred to as the Wald statistic, can be used to test a subtest of parameters, as will be introduced in Chapter 5.

The third main test for null hypotheses of parameter estimates in $\hat{\boldsymbol{\theta}}$ uses the likelihood ratio test. The likelihood ratio with respect to $\boldsymbol{\theta}$ is

$$LR(\boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta})}{L(\hat{\boldsymbol{\theta}})}, \tag{4.35}$$

where $L(\boldsymbol{\theta})$ is the likelihood function for the model without one or more parameters, whereas $L(\hat{\boldsymbol{\theta}})$ is the likelihood function containing all parameters.

In the likelihood ratio test, the null hypothesis about $\boldsymbol{\theta}$ can be set either as H_0: $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ for all components in $\hat{\boldsymbol{\theta}}$ or as H_0: $\theta_m = \hat{\theta}_m$ for a single component in $\boldsymbol{\theta}$. Under such a null hypothesis, the likelihood ratio statistic, written by

$$\Lambda = 2\log L(\hat{\boldsymbol{\theta}}) - 2\log L(\boldsymbol{\theta}), \tag{4.36}$$

is asymptotically distributed as $\chi^2_{(M)}$. Consequently, statistical testing can be performed with larger values of Λ associated with smaller p-values, thereby providing more evidence against H_0. More specifically, if Λ is associated with a p-value smaller than α, the null hypothesis about $\hat{\boldsymbol{\theta}}$ should be rejected. When the likelihood ratio test is performed on all components in $\hat{\boldsymbol{\theta}}$, the likelihood ratio test displays the model fit statistic with empirical data. This test can also be conducted for one or more components in $\boldsymbol{\theta}$, but the parameters need to be partitioned for reparametrization (Kalbfleisch and Prentice, 2002; Lawless, 2003).

There are several modifications of the −2 log-likelihood statistic, such as Akaike's information criterion (*AIC*), the Bayesian information criterion (*BIC*), and the corrected version of the *AIC* (*AICC*). As they usually generate very close fit statistics as the likelihood ratio test, especially in large samples, I do not describe these modified statistics further in this text.

4.2 Exponential regression models

If survival data arise from an exponential distribution, the hazard rate is constant throughout all values of t, as specified by Equation (4.2). At the same time, the exponential regression can be approached by another family of parametric regression modeling that regresses the log survival time over covariates. An important feature of the exponential regression is that the estimated regression coefficients derived from the hazard rate and the AFT models have identical absolute values but with opposite signs, and also they share exactly the same standard errors. This section describes statistical inferences of the exponential regression model in terms of both the proportional hazard and the accelerated failure time functions. An empirical example is provided for practical illustration.

4.2.1 Exponential regression model on the hazard function

Equation (4.2) specifies the hazard rate as the product of a constant baseline hazard rate λ and the multiplicative effect term $\exp(\boldsymbol{x}'\boldsymbol{\beta})$. To further simplify this exponential regression equation, statisticians often view $\log\lambda$ as a special coefficient and place it into the regression coefficient vector $\boldsymbol{\beta}$. As a result, the first element of $\boldsymbol{x}$ contains value 1 and the first element of $\boldsymbol{\beta}$ contains value $\log\lambda$. Then Equation (4.2) can be simplified by

$$h(t, \boldsymbol{x}; T \sim Exp) = \exp(\boldsymbol{x}'\boldsymbol{\beta}). \tag{4.37}$$

When all components of $\boldsymbol{x}$ except the first element are 0, $h(t, \boldsymbol{x}; T \sim Exp) = \lambda$, given $\exp(\log\lambda) = \lambda$, thereby giving rise to the baseline hazard rate. Likewise, if all covariates are centered at sample means, the constant baseline hazard rate is the expected hazard for the population a given sample represents.

Equation (4.37) derives the following survival function given the exponential distribution of event time T:

$$S(t, \boldsymbol{x}; T \sim Exp) = \exp[-t\exp(\boldsymbol{x}'\boldsymbol{\beta})]. \tag{4.38}$$

Given specifications of the hazard and the survival functions, the density function of T, conditional on $\boldsymbol{x}$, is

$$f(t, \boldsymbol{x}; T \sim Exp) = \exp(\boldsymbol{x}'\boldsymbol{\beta})\exp[-t\exp(\boldsymbol{x}'\boldsymbol{\beta})]. \tag{4.39}$$

With the assumption of noninformative censoring, the likelihood function of the exponential regression model with respect to $\boldsymbol{\beta}$ is written as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \{\exp(\boldsymbol{x}'\boldsymbol{\beta})\exp[-t\exp(\boldsymbol{x}'\boldsymbol{\beta})]\}^{\delta_i}\{\exp[-t\exp(\boldsymbol{x}'\boldsymbol{\beta})]\}^{1-\delta_i}. \tag{4.40}$$

After some algebra, Equation (4.40) reduces to

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \exp(\delta_i \boldsymbol{x}_i'\boldsymbol{\beta})\exp[-t_i\exp(\boldsymbol{x}_i'\boldsymbol{\beta})]. \tag{4.41}$$

Then, the log-likelihood function with respect to $\boldsymbol{\beta}$ is

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \boldsymbol{x}_i \boldsymbol{\beta} - \sum_{i=1}^{n} t_i \exp(\boldsymbol{x}_i'\boldsymbol{\beta}). \tag{4.42}$$

Maximization of Equation (4.42) with respect to $\boldsymbol{\beta}$ yields the MLE estimates in the exponential regression model on the hazard rate. In particular, the score statistic vector $\tilde{\boldsymbol{U}}_i(\boldsymbol{\beta})$ within the construct of exponential regression modeling contains components

$$\tilde{\boldsymbol{U}}_m(\boldsymbol{\beta}) = \frac{\partial \log L(\boldsymbol{\beta})}{\partial \beta_m} = \sum x_{mi}[\delta_i - t_i\exp(\boldsymbol{x}_i'\boldsymbol{\beta})], \quad m = 1, \ldots, M. \tag{4.43}$$

The Fisher information matrix of the second partial derivatives of the log likelihood function with respect to $\boldsymbol{\beta}$ is

$$\boldsymbol{I}\left(\hat{\boldsymbol{\beta}}\right) = -\left[\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial\beta_m \partial\beta_{m'}}\right]_{M\times M} = \sum x_{mi} x_{m'i} t_i \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right). \tag{4.44}$$

In empirical research, the observed information matrix is generally used to yield asymptotic results given the large-sample approximation theory and the central limit theorem. As indicated earlier, its inverse is regularly applied as the approximated variance–covariance matrix of parameter estimates. The standard error estimates are the square roots of the diagonal elements in the inverse of the observed information matrix. As commonly known in the estimation of a regression model, high correlation between two covariates may yield ill effects on the quality of ML parameter estimates. Therefore, if the covariate vector $\boldsymbol{x}$ contains one or more interaction terms, the covariates used for constructing those interactions need to be rescaled to be centered at their sample means or at some specific values for reducing numeric instability and collinearity.

As demonstrated by Equation (4.37), given the covariate effect term $\exp(\boldsymbol{x}'\boldsymbol{\beta})$, any two individuals are subject to hazard rates that are multiplicative or proportional to one another throughout all t values. This feature can be extended to the condition that the ratio of two hazard rates associated with two successive integer values of a covariate can be expressed in terms of $\exp(b_m)$. Let b_0 be the intercept ($\log\lambda$) and b_m be the regression coefficient of covariate x_m, both contained in the coefficient vector $\boldsymbol{\beta}$. Also let two successive integer values of x_m be x_{m0} and $x_{m0} + 1$. If all other covariates are scaled to 0, the ratio of the hazard rates for $x_m = x_{m0}$ and $x_m = x_{m0} + 1$, denoted by $HR(x_{m0}, x_{m0} + 1)$, is given by

$$\begin{aligned} HR\left(x_{m0}, x_{m0}+1\right) &= \frac{\lambda \exp\left[b_0 + b_m\left(x_{m0}+1\right)\right]}{\lambda \exp\left(b_0 + b_m x_{m0}\right)} \\ &= \exp\left[\left(x_{m0}+1-x_{m0}\right)b_m\right] \\ &= \exp\left(b_m\right). \end{aligned} \tag{4.45}$$

Given this simple inference, the term $\exp(b_m)$ is referred to as *the hazard ratio of covariate m*, with the abbreviated term HR. This definition holds for all proportional hazard regression models, including the Weibull regression model and the semi-parametric regression model described in later chapters. Exponentiation of elements in the coefficient vector $\boldsymbol{\beta}$ thus yields a hazard ratio vector. Such a relative risk holds when other covariates take nonzero values because the additional terms would appear in both the numerator and the denominator in Equation (4.45) and thereby can cancel out.

For proportional hazard rate models, an individual hazard ratio displays the multiplicative effect on the hazard rate with a 1-unit change in a given covariate. If the covariate is a dichotomous variable, the hazard ratio has an intuitive meaning of relative risk. For example, the hazard ratio for gender (1 = female, 0 = male) shows the proportion of the hazard rate for women as relative to the hazard rate for men, other variables being equal. In the case of a continuous covariate, the hazard ratio presents the increased risk ($HR > 1$) or the decreased risk ($HR < 1$) of experiencing a particular event with a 1-unit increase in the value of a given covariate.

Computing $100 \times [\exp(b_m) - 1]$ yields the percentage change in the hazard rate with a 1-unit change in covariate x_m. Consider the gender example: if the hazard ratio is 0.75, the value of $100 \times (0.75 - 1) = -25$, indicating a 25 % reduction in the hazard rate for women as compared to the hazard rate for men, other variables remaining constant.

4.2.2 Exponential accelerated failure time regression model

As indicated in Chapter 2, the exponential distribution can be modeled as an accelerated failure time function. Let $Y = \log T$ and $\tilde{\gamma} = \lambda^{-1}$. Then Equation (4.2) can be written in a linear form as

$$Y = \boldsymbol{x}'\boldsymbol{\beta}^* + \varepsilon_{\exp}, \tag{4.46}$$

where $\varepsilon_{\exp}$ can be viewed as a special case of the extreme value distribution with the scale parameter equaling 1 (the extreme value distribution is briefly described in Chapter 3 and will be discussed further in Section 4.3). The term $\log\tilde{\gamma}$ is included in $\boldsymbol{\beta}^*$ as an intercept parameter. Exponentiation of both sides of this equation leads to the following AFT model:

$$\begin{aligned} T &= \exp(\boldsymbol{x}'\boldsymbol{\beta}^*)\exp(\varepsilon_{\exp}) \\ &= \exp(\boldsymbol{x}'\boldsymbol{\beta}^*)\tilde{E}_{\exp}, \end{aligned} \tag{4.47}$$

where $\tilde{E}$ follows an exponential distribution, as defined. Given the above specifications, the survival function is

$$S(t; \boldsymbol{x}, \boldsymbol{\beta}^*) = \exp[-\exp(y - \boldsymbol{x}'\boldsymbol{\beta}^*)], \tag{4.48}$$

where $y = \log(t)$. Accordingly, the density function can be expressed in terms of the extreme value distribution, given by

$$f(t; \boldsymbol{x}, \boldsymbol{\beta}^*) = \exp[(y - \boldsymbol{x}'\boldsymbol{\beta}^*) - \exp(y - \boldsymbol{x}'\boldsymbol{\beta}^*)]. \tag{4.49}$$

Given the density and the survival functions, the likelihood function with respect to $\boldsymbol{\beta}^*$ is

$$L(\boldsymbol{\beta}^*) = \prod_{i=1}^{n} \{\exp[(y - \boldsymbol{x}'\boldsymbol{\beta}^*) - \exp(y - \boldsymbol{x}'\boldsymbol{\beta}^*)]\}^{\delta_i} \{\exp[-\exp(y - \boldsymbol{x}'\boldsymbol{\beta}^*)]\}^{1-\delta_i}. \tag{4.50}$$

The log-likelihood function is

$$\log L(\boldsymbol{\beta}^*) = \sum_{i=1}^{n} \delta_i (y - \boldsymbol{x}'\boldsymbol{\beta}^*) - \exp(y - \boldsymbol{x}'\boldsymbol{\beta}^*). \tag{4.51}$$

According to the procedures described earlier, maximizing the above log-likelihood function with respect to $\boldsymbol{\beta}^*$ yields the MLE estimates of $\tilde{\gamma}$ and the regression coefficients.

The second partial derivative of this function, for a single component, is defined as

$$\frac{\partial^2 L(\boldsymbol{\beta}^*)}{\partial \beta_m \beta_{m'}} = -\sum_{i=1}^{n} x_{mi} x_{m'i} \exp(y - \boldsymbol{x}'\boldsymbol{\beta}^*). \tag{4.52}$$

The variance–covariance matrix of $\boldsymbol{\beta}^*$ can be derived from the inverse of the observed information matrix $\boldsymbol{I}\left(\hat{\boldsymbol{\beta}}^*\right)$, from which the null hypothesis can be statistically tested, as described in Subsection 4.1.3.

In the exponential accelerated failure time model, the hazard rate given $\boldsymbol{x}$ is regarded as the inverse function of Equation (4.48), given by

$$h\left(t; \boldsymbol{x}, \boldsymbol{\beta}^*\right) = \exp\left(-\boldsymbol{x}'\boldsymbol{\beta}^*\right), \tag{4.53}$$

where $\log\tilde{\gamma}$ is the first element in $\boldsymbol{\beta}^*$. A comparison between Equations (4.53) and (4.37) demonstrates that as $[\exp(-\boldsymbol{x}'\boldsymbol{\beta}^*)]$ is equivalent to $\exp(\boldsymbol{x}'\boldsymbol{\beta})$, $\boldsymbol{\beta}^* = -\boldsymbol{\beta}$. Additionally, the standard errors of the hazard and AFT parameter estimates are equal. Therefore, the estimation of either the hazard rate or the AFT perspective yields both sets of parameter estimates through some straightforward transformation.

As $\exp(\boldsymbol{x}'\boldsymbol{\beta}^*)$ represents a vector of multiplicative effects on T, the individual term $\exp\left(b_m^*\right)$ indicates the ratio of survival times for $x_m = x_{m0}$ and $x_m = x_{m0} + 1$, referred to as the *time ratio of covariate m* or TR. Consequently, exponentiation of the elements in the coefficient vector $\boldsymbol{\beta}^*$ yields a time ratio vector. An individual time ratio displays the multiplicative effect on T with a 1-unit change in a given covariate. Suppose that covariate x_m is a dichotomous variable representing gender with 1 = female and 0 = male. Then the time ratio displays the proportion of survival time for women as relative to the survival time for men, other variables being equal. In the case of a continuous covariate, the time ratio presents whether the survival time for the occurrence of a particular event increases ($TR > 1$) or decreases ($TR < 1$) with a 1-unit change in the value of a covariate of interest. Like the hazard ratio, interpreting a time ratio is much like interpreting an odds ratio for a logistic regression model. The term $100\left[\exp\left(b_j^*\right) - 1\right]$ represents the percentage change in the event time with a 1-unit change in covariate x_m, other covariates being fixed. As $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}$ take opposite signs, the time ratio and the corresponding hazard ratio bear opposite directions as deviating from unity.

Given the estimates of $\boldsymbol{\beta}^*$, the exponential survival function can be approximated given values of $\boldsymbol{x}$. Some other functions, intimately related to the distribution of T, are also obtainable, either from $\boldsymbol{\beta}^*$ or from $\boldsymbol{\beta}$. For example, according to Equation (3.14), the estimated median time to the occurrence of a particular event is

$$t_m(\boldsymbol{x}, \boldsymbol{\beta}) = \frac{\log 2}{\exp(\boldsymbol{x}'\boldsymbol{\beta})} = \exp\left(\boldsymbol{x}'\boldsymbol{\beta}^*\right)\log 2. \tag{4.54}$$

Likewise, the estimated pth percentile of the survival time distribution, given $\boldsymbol{x}$, can be estimated by replacing the value of $\log 2$ with $\log 1/(1 - p)$:

$$t_p(\boldsymbol{x}, \boldsymbol{\beta}) = \frac{\log\left(\dfrac{1}{1-p}\right)}{\exp(\boldsymbol{x}'\boldsymbol{\beta})} = \exp\left(\boldsymbol{x}'\boldsymbol{\beta}^*\right)\log\left(\frac{1}{1-p}\right). \tag{4.55}$$

Therefore, with the estimates of $\boldsymbol{\beta}$ available, the median and various percentiles of survival time can be estimated by placing values of $\boldsymbol{x}$ into either of the above two equations.

4.2.3 Illustration: Exponential regression model on marital status and survival among older Americans

In Section 2.3.4, I provide an empirical example about marital status and the probability of survival among older Americans. Although the survival curve for currently married persons is found to differ significantly from the curve among the currently not married, this relationship can be confounded by some 'lurking' covariates. In the present illustration, I reexamine this association by considering two potential confounders – age and educational attainment. I consider these two covariates because each of them is often observed to be causally associated with both marital status and mortality, and thus may potentially yield some confounding effects.

In the analysis, marital status is a dichotomous variable with 1 = 'currently married' and 0 = else, named 'Married.' An individual's age is the actual years of age and educational attainment, an approximate proxy for socioeconomic status, is measured as the total number of years in school, assuming the influence of education on mortality to be a continuous process (Liu, Hermalin and Chuang, 1998). Both control variables are measured at the AHEAD baseline survey and are then rescaled to be centered at the sample means for convenience of analysis. The two centered control variables are termed, respectively, 'Age_mean' and 'Educ_mean.' My goal is to assess whether the significant difference in the probability of survival between currently married and currently not married persons holds after adjusting for the confounding effects of age and educational attainment. For illustrating how to apply the exponential regression model, I estimate a one-year survival function, assuming the baseline hazard rate to be constant within this relatively short observation interval.

In the SAS system, the PROC LIFEREG procedure is used to fit parametric regression models. This procedure is entirely based on the AFT perspective; as a result, it does not directly model the hazard rate and does not produce the corresponding hazard ratio. In the exponential regression perspective, however, parameter estimates on log(survival time) can be readily converted to the parameter estimates on the log(hazard) and the hazard ratios by using the SAS ODS procedure.

Below is the SAS program for the estimation of the exponential regression model on marital status and survival.

SAS Program 4.1:

```
options ps=56;
libname pclib 'C:\<the location that contains the dataset "chapter4_data>;;
options _last_=pclib.chapter4_data;

proc format;
  value Rx 1 = 'Married' 0 = 'Not married';

data new;
  set pclib.chapter4_data;
  format Married Rx.;

.......

Censor = 1;
  if death = 0 then Censor = 1;
```

```
proc SQL;
  create table new as
  select *, age-mean(age) as age_mean,
  educ-mean(educ) as educ_mean
  from new;
quit;

* Compute estimates of the exponential regression model *;
ods graphics on ;
ods output Parameterestimates = Exponential_AFT;
proc lifereg data=new outest=Exponential_outest;
  class married;
  model duration*Censor(1) = married age_mean educ_mean / dist = Exponential;
  output out = new1 cdf = prob;
run;
ods graphics off;
ods output close;

* Compute and print correspoing PH estimates and HR *;
data Exponential_AFT;
  if _N_ = 1 then set Exponential_outest;
  set Exponential_AFT;
  PH_beta = -1*estimate;
  AFT_beta = estimate;
  HR = exp(PH_beta);
  option nolabel;

proc print data = Exponential_AFT;
  id Parameter;
  var AFT_beta PH_beta HR;
  format PH_beta 8.5 HR 5.3;
  title1 "LIFEREG/Exponential on duration";
  title2 "ParameterEstimate dataset e/PH_beta & HR computed manually";
  title4 "PH_beta = -1*AFT_beta";
  title6 "HR = exp(PH_beta)";
run;

proc sgplot data = new1;
  scatter x = duration y = prob / group = married;
  discretelegend;
  title1 "Figure 4.1. Cumulative distribution function: Exponential";
run;
```

In SAS Program 4.1, a temporary SAS data file 'New' is created from the dataset chapter4_data. The dichotomous variable Censor is constructed, with 1 = censored and 0 = not censored. The SAS PROC SQL procedure creates two centered covariates as controls – Age_mean and Educ_mean – which are then saved into the temporary SAS data file for further analysis.

In the SAS LIFEREG procedure, I first ask SAS to estimate the regression coefficients of covariates on log(duration), with parameter estimates saved in the temporary SAS output file Exponential_AFT. In this step, the CLASS statement specifies the variable 'Married' as a classification variable. The MODEL statement specifies three covariates, with the distribution specified as exponential. The OUTPUT statement creates the output dataset New1. Given the option 'cdf = Prob,' the temporary dataset New1 contains the variable Prob created to yield the cumulative distribution function at the observed responses.

The following part converts the saved log time coefficients (AFT_beta) into log hazard coefficients (PH_beta) and hazard ratios (HR). Then I request SAS to print three sets of parameter estimates: regression coefficients on log(duration), regression coefficients on log(hazards), and hazard ratios. For exponential regression models, the conversion from AFT parameter estimates to the HR estimates is conducted simply by changing the sign of each parameter estimate. Because both sets of parameter estimates, AFT and PH, share the same standard errors, conversion of standard errors is unnecessary.

Lastly, the PROC SGPLOT procedure produces a graph of the cumulative distribution values versus the variable Duration. Some of the analytic results from SAS Program 4.1 are presented below.

SAS Program Output 4.1:

```
                              The LIFEREG Procedure
                                Model Information

                    Data Set                            WORK.NEW
                    Dependent Variable             Log(duration)
                    Censoring Variable                    censor
                    Censoring Value(s)                         1
                    Number of Observations                  2000
                    Noncensored Values                        56
                    Right Censored Values                   1944
                    Left Censored Values                       0
                    Interval Censored Values                   0
                    Number of Parameters                       4
                    Name of Distribution             Exponential
                    Log Likelihood                  -301.3722192


                     Number of Observations Read        2000
                     Number of Observations Used        2000

                                 Fit Statistics

                -2 Log Likelihood                          602.744
                AIC (smaller is better)                    610.744
                AICC (smaller is better)                   610.764
                BIC (smaller is better)                    633.148

              Analysis of Maximum Likelihood Parameter Estimates

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

Intercept        1   5.8817    0.1768   5.5351   6.2283 1106.15     <.0001
married       0  1   0.4064    0.2912  -0.1644   0.9772    1.95     0.1629
married       1  0   0.0000     .        .        .        .          .
age_mean         1  -0.0605    0.0215  -0.1027  -0.0184    7.93     0.0049
educ_mean        1   0.0268    0.0371  -0.0458   0.0994    0.52     0.4698
Scale            0   1.0000    0.0000   1.0000   1.0000
Weibull Shape    0   1.0000    0.0000   1.0000   1.0000

                      LIFEREG/Exponential on duration
            ParameterEstimate dataset e/PH_beta & HR computed manually

                            PH_beta = -1*AFT_beta

                              HR = exp(PH_beta)

                Parameter        AFT_beta      PH_beta        HR

                Intercept         5.88173     -5.88173     0.003
                married           0.40642     -0.40642     0.666
                married           0.00000      0.00000     1.000
                age_mean         -0.06052      0.06052     1.062
                educ_mean         0.02679     -0.02679     0.974
                Scale             1.00000     -1.00000     0.368
                Weibull Shape     1.00000     -1.00000     0.368
```

In SAS Program Output 4.1, the model information and a statistic of model fitness are presented first. There are 56 uncensored (deaths) and 1944 right-censored observations, given a fairly short observation interval. As all covariates are measured at baseline and the vital status for each individual is known, all censored observations are of Type I right censoring. The log-likelihood for this exponential regression model is −301.3722, and this statistic is often used for comparing the goodness of fit among multiple models, as will be described in Chapter 8. The section of fit statistics displays four indicators of model fitness, with the −2 log-likelihood statistic described earlier. In summary, values of these four statistics are close, generating the same conclusion about model fitting for this analysis.

The table of ML parameter estimates demonstrates the LIFEREG estimates on log(duration) derived from the maximum likelihood procedure. The intercept, 5.8817 ($SE = 0.1768$), is statistically significant at $\alpha = 0.05$. Within a fairly short observation period of 12 months, however, currently married persons do not display a significantly different survival function from the curve among currently not married after adjusting for the effects of age and educational attainment. Because the variable 'Married' is dichotomous with one degree of freedom, the parameter estimate is given only for currently not married persons (in PROC LIFEREG, the group with the greatest value is treated as the reference). Consequently, the intercept provides an estimate for a currently married older person with an average age and an average educational attainment. For the two control variables, age is negatively associated with survival time, statistically significant ($\chi^2 = 7.93$; $p = 0.0049$). Educational attainment, on the other hand, is positively linked to survival time, which is not statistically significant ($\chi^2 = 0.52$; $p = 0.4698$).

The last table in SAS Program Output 4.1 presents the three sets of parameter estimates – AFT_beta, PH_beta, and HR. In the exponential regression model, as mentioned above and can be seen from the table, the PH_beta of a given covariate is simply the AFT_beta value times −1. The hazard ratio of each covariate (HR) is presented to highlight the multiplicative effect on the mortality of older Americans within a one-year time interval. For example, a one-year increase in age would increase the hazard rate by 6 % ($HR = 1.06$), other variables being equal.

Figure 4.1 displays the cumulative distribution function contained in the output dataset New1 for currently married and currently not married older persons. Two c.d.f.'s are shown without demonstrating a clear pattern of mortality differences. As 'month' is used as the time scale, there are substantial tied observations, especially in later months. This lack of association between marital status and survival might be due to the short observation interval, during which the impact of married life on the probability of survival cannot be captured. For a longer observation interval, on the other hand, the selection of the exponential regression model is obviously not appropriate because the hazard rate, especially among older persons, increases over time.

Another possibility for this lack of association is misspecification of the baseline distributional function of T. Although hazard rates within a one-year period are often considered to be relatively stable, this assumption might not be entirely appropriate for older persons. A graphical check on the distribution of the hazard rate can provide some implications concerning whether or not the exponential distribution is a good fit for the present analysis. Given this reason, I plot a graph of negative log survival versus time for the entire sample, using the Kaplan–Meier survival estimates. The plot should display a straight line if the survival data within a one-year interval arise from an exponential distribution, as indicated in Chapter 3. The SAS program for plotting this curve is given below.

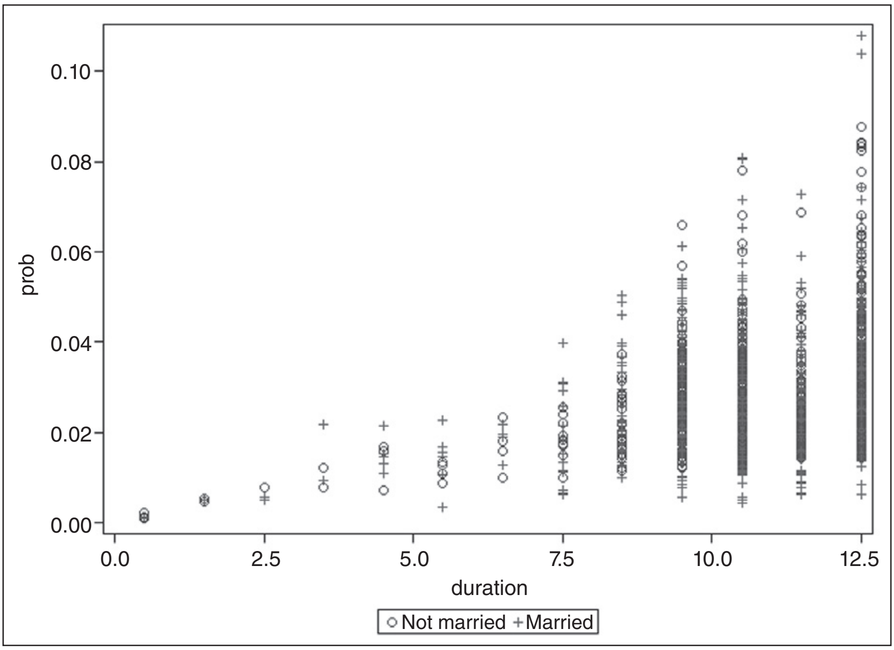

Figure 4.1 Cumulative distribution function: exponential.

SAS Program 4.2:

```
……
Status = 1;
if death=0 or time > 12 then Status = 0;

ods html;
ods graphics on;

ods select NegLogSurvivalPlot;
lifetest data = new OUTSURV = out1 plots = logsurv;
  time duration * Status(0);
run;

proc print data=out1;
    title1 "Figure 4.2.3-2. Log S(t) versus t";
run;

ods graphics off;
ods trace off;
```

In SAS Program 4.2, I replace the variable Censor with the variable Status (1 = event, 0 = censored) for executing the PROC LIFETEST procedure, constructing a negative $\log \hat{S}(t)$ versus t plot using the Kaplan–Meier survival estimates. The ODS Graphics is enabled by specifying the ODS GRAPHICS ON statement with the ODS select name as NegLogSur-

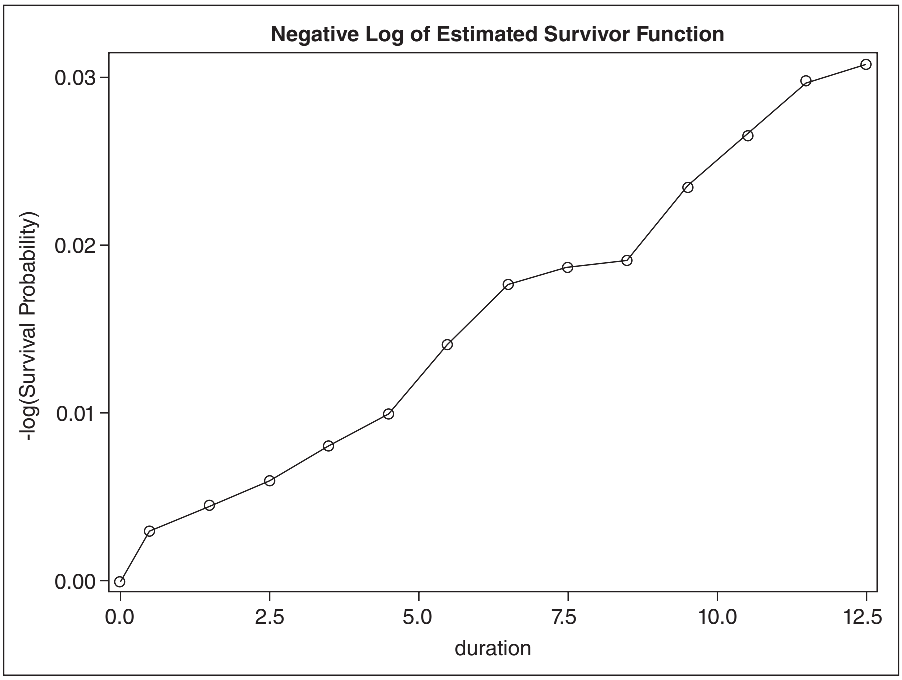

Figure 4.2 Plot of negative log survival versus time.

vivalPlot. In the PROC LIFETEST statement, the option PLOTS = LOGSURV asks SAS to construct the graph of negative $\log \hat{S}(t)$ versus t. SAS Program 4.2 yields the plot in Figure 4.2, where the negative log $S(t)$ versus t displays a roughly straight line, though with substantial variations, thus suggesting that the use of the exponential distributional function for a one-year observation interval is probably not unreasonable among older Americans. It must be emphasized that this graphical check cannot be used to make a final decision on an underlying distribution of survival data. More plausible checking methods will be discussed in Chapters 5 and 6.

4.3 Weibull regression models

Like exponential regression models, the Weibull regression can be approached in terms of two lifetime functions. As a result, the Weibull parameters can be estimated either by the Weibull proportional hazard or by the accelerated failure time model. Given the underlying distributional function of T, one set of parameter estimates can be easily converted to the other through some transformational procedures.

This section first describes the basic specifications and inferences of both the proportional hazard and the AFT regression models. Then I specify the intimate associations between the Weibull proportional hazard and AFT parameters, with a detailed description of the transformational procedures. Lastly, I provide an empirical illustration on how to apply the Weibull regression model using SAS.

4.3.1 Weibull hazard regression model

For a sample of n individuals with right censoring, the Weibull proportional hazard regression model is expressed in terms of a scale parameter λ, a shape parameter $\tilde{p}$, and a multiplicative term $\exp(\boldsymbol{x}'\boldsymbol{\beta})$. Specifically, the hazard rate at time t for individual i, within the construct of the Weibull hazard model, can be written as

$$h_i(t, \boldsymbol{x}; T \sim Weib) = \lambda\tilde{p}(\lambda t)^{\tilde{p}-1}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}), \tag{4.56}$$

where the covariate vector $\boldsymbol{x}$ and the coefficient vector $\boldsymbol{\beta}$ are specified previously. Equation (4.56) suggests that all individuals follow a univariate Weibull baseline hazard function, defined by $\lambda\tilde{p}(\lambda t)^{\tilde{p}-1}$. Population heterogeneity is reflected by the multiplicative effects of an individual's observed socioeconomic, biological, or demographic characteristics, represented by the covariate vector $\boldsymbol{x}$. In other words, the covariate effect $\exp(\boldsymbol{x}'\boldsymbol{\beta})$ changes the scale but not the shape of the distribution in T. Given this specification, the Weibull hazard regression model is a proportional perspective.

The Weibull survival rate for individual i can be readily derived from differentiating Equation (4.56). After some simplification, it becomes

$$S_i(t, \boldsymbol{x}; T \sim Weib) = \exp\left[-\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\lambda t^{\tilde{p}}\right]. \tag{4.57}$$

By definition, the density function of T in the Weibull distribution can be expressed by the product of the Weibull hazard function and the survival function. Consequently, the likelihood function of n observations is

$$L(\lambda, \tilde{p}, \boldsymbol{x}) = \prod_{i=1}^{n}\left[\lambda\tilde{p}(\lambda t)^{\tilde{p}-1}\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\right]^{\delta_i}\exp\left[-\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\lambda t^{\tilde{p}}\right]. \tag{4.58}$$

In Equation (4.58), if $\delta_i = 1$, $L_i(\lambda, \tilde{p}, \boldsymbol{x})$ is the density function for a particular event, given the product of the two terms on the right of equation; when $\delta_i = 0$, $L_i(\lambda, \tilde{p}, \boldsymbol{x})$ is the survival function for a censored survival time as the first term on the right of the equation is 1.

Taking log values on both sides of Equation (4.58), the log-likelihood function is

$$\log L(\lambda, \tilde{p}, \boldsymbol{x}) = \sum_{i=1}^{n}\Big\{\delta_i\left[\log\lambda + \log\tilde{p} + (\tilde{p}-1)\log(\lambda t) + \boldsymbol{x}_i'\boldsymbol{\beta}\right] - \left[\exp(\boldsymbol{x}_i'\boldsymbol{\beta})(\lambda t)^{\tilde{p}}\right]\Big\}. \tag{4.59}$$

After some algebra, Equation (4.59) can be simplified, removing a term that does not involve unknown parameters, given by

$$\log L(\lambda, \tilde{p}, \boldsymbol{x}) = \sum_{i=1}^{n} \left\{ \delta_i \left[\log(\lambda \tilde{p}) + \tilde{p} \log t_i + \boldsymbol{x}_i' \boldsymbol{\beta} \right] - \lambda \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) t^{\tilde{p}} \right\}. \tag{4.60}$$

Maximizing Equation (4.60) requires that this log-likelihood function be differentiated with respect to model parameters λ, $\tilde{p}$, and $\boldsymbol{\beta}$. I do not, however, elaborate on the concrete estimating procedures for the proportional hazard parameters because in most statistical software packages the more flexible AFT function is used to generate the standard parametric parameters from which the proportional hazard parameter estimates can be obtained by some simple transformations.

Once the maximum likelihood estimates of parameters λ, $\tilde{p}$, and $\boldsymbol{\beta}$ are derived, the hazard, survival, and other related functions can be approximated, given values of selected values of $\boldsymbol{x}$. Whereas the hazard and survival functions can be estimated by applying Equations (4.56) and (4.57), the estimated median survival time, given $\boldsymbol{x}$, is

$$\hat{t}_m\left(\hat{\lambda}, \hat{\tilde{p}}, \boldsymbol{x}\right) = \left[\frac{\log 2}{\hat{\lambda} \exp\left(\boldsymbol{x}' \hat{\boldsymbol{\beta}}\right)} \right]^{1/\hat{\tilde{p}}}. \tag{4.61}$$

Equation (4.61) is simply the extension of Equation (3.25), with the addition of an effect term. Similarly, the estimated pth percentile of the survival time distribution, conditional on $\boldsymbol{x}$, can be estimated by replacing the value of $\log 2$ with $\log 1/(1-p)$:

$$\hat{t}_p\left(\hat{\lambda}, \hat{\tilde{p}}, \boldsymbol{x}\right) = \left[\frac{\log\left(\frac{1}{1-p}\right)}{\hat{\lambda} \exp\left(\boldsymbol{x}' \hat{\boldsymbol{\beta}}\right)} \right]^{1/\hat{\tilde{p}}}. \tag{4.62}$$

In the above equation, the reader might want to be aware of the difference between $\tilde{p}$, the Weibull shape parameter, and p, the probability value in a time distribution.

As the Weibull hazard regression model is a proportional function, the term $\exp(b_m)$ reflects the hazard ratio of covariate x_m, highlighting the proportional change in the hazard rate with a 1-unit increase in x_m, other covariates being unchanged.

4.3.2 Weibull accelerated failure time regression model

The Weibull function is often formulated in the form of the extreme value distribution because $\log(T)$, which can be expressed as a function of the Weibull parameters, follows the extreme value distribution. More important, if $\log T$ has an extreme value distribution, the survival time T follows an accelerated failure time distribution that can be applied to many families of parametric distributions, as described in Section 4.1. Given this flexibility, the Weibull regression model can be more conveniently specified and estimated within the structure of the AFT regression perspective.

Let $Y = \log(T)$ and $p^* = 1/\tilde{p}$ be the scale parameter. Replacing $\log t$ and $\tilde{\sigma}$ in Equation (4.9) with y and p^*, the Weibull survival function of T can be written as an AFT regression model:

$$S\left(t; \boldsymbol{x}, \boldsymbol{\beta}^*, p^*\right) = \exp\left[-\exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^*}{p^*}\right)\right], \quad -\infty < y < \infty. \tag{4.63}$$

Notice that Equation (4.63) is a simplified expression. As indicated in Subsection 4.1.2, the coefficient vector $\boldsymbol{\beta}^*$ includes an intercept parameter μ^*. While intimately related to the Weibull hazard rate parameter λ, the parameter μ^* will be further specified and discussed in Subsection 4.3.3.

In view of the fact that $S(t)$ can be expressed in terms of an accelerated failure time function, the Weibull hazard function can be formulated as an inverse in terms of the AFT regression model, given by

$$h\left(t; \boldsymbol{x}, \boldsymbol{\beta}^*, p^*\right) = \left(p^*\right)^{-1} \exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^*}{p^*}\right), \quad -\infty < y < \infty. \tag{4.64}$$

The value of p^* determines the shape of the Weibull hazard function. If $p^* = 1$, the Weibull AFT model reduces to the exponential AFT regression model and, consequently, the hazard function is constant (or we can say the exponential AFT regression model is a special case of the Weibull AFT regression model with $p^* = 1$). When $p^* > 1$, the hazard rate decreases over time. When $0.5 < p^* < 1$, the hazard increases at a decreasing rate, with its shape being concave. When $0 < p^* < 0.5$, the hazard rate increases with an increasing rate, and its shape is convex. Lastly, when $p^* = 0.5$, the hazard rate is constantly increasing.

With the Weibull survival and hazard functions defined as the AFT inverse functions, the AFT density function in the Weibull regression model can be easily written within the formation of an AFT regression:

$$\begin{aligned} f\left(t; \boldsymbol{x}, \boldsymbol{\beta}^*, p^*\right) &= \exp\left[-\exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^*}{p^*}\right)\right]\left(p^*\right)^{-1} \exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^*}{p^*}\right) \\ &= \left(p^*\right)^{-1} \exp\left[\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^*}{p^*} - \exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^*}{p^*}\right)\right], \quad -\infty < y < \infty. \end{aligned} \tag{4.65}$$

As can be seen from Equation (4.65), the Weibull AFT regression model is simply an extension of the exponential AFT regression model with the addition of a scale parameter p^*. The reader might want to compare Equations (4.63) and (4.65) with Equations (4.48) and (4.49) for such associations.

After some simplification, the likelihood function with respect to $\boldsymbol{\beta}^*$ for the Weibull AFT regression model is

$$L\left(\boldsymbol{\beta}^*, p^*\right) = \prod_{i=1}^{n} \left[\left(p^*\right)^{-1} \exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^*}{p^*}\right)\right]^{\delta_i} \exp\left[-\exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^*}{p^*}\right)\right]. \tag{4.66}$$

According to Equation (4.66), the likelihood refers to either a density function when $\delta_i = 1$, or the probability of survival when $\delta_i = 0$, so survival times for both events and censored cases are accounted for.

The log-likelihood function with respect to Equation (4.66) is

$$\log L\left(\boldsymbol{\beta}^*, p^*\right) = \sum_{i=1}^{n} \left\{ \delta_i \left[\left(-\log p^*\right) + \left(\frac{\mathrm{y} - \boldsymbol{x}'\boldsymbol{\beta}^*}{p^*} \right) \right] \right\} - \exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^*}{p^*} \right). \tag{4.67}$$

Maximizing this log-likelihood function with respect to $\boldsymbol{\beta}^*$ and p^* yields the MLE estimates of the Weibull AFT model parameters. Here, as the estimating procedures are standard, described extensively in Section 4.1, I do not elaborate on the detailed inference. Given the above specifications in the AFT formulation, used by most major statistical software packages like SAS, the survival, density, and hazard functions in the Weibull proportional hazard model are regularly considered to be the inverse functions of the AFT parameters. As a result, given the Weibull AFT parameter estimates, the parameter estimates of the Weibull proportional hazard models can be obtained by transforming the AFT parameters, as presented in Subsection 4.3.3.

4.3.3 Conversion of Weibull proportional hazard and AFT parameters

As indicated in Chapter 3, the Weibull AFT parameters can be expressed as transforms of the Weibull hazard model parameters, and it is also true for vice versa. A comparison between Equations (4.56) and (4.64) or between (4.57) and (4.63) can help the reader better to understand the mathematical associations between the two sets of Weibull parameters.

Let $Y = \log(T)$, $p^* = 1/\tilde{p}$. Then the parameters in the Weibull proportional hazard regression model can be expressed in terms of the Weibull AFT parameters, given by

$$\lambda = \exp\left(-\frac{\mu^*}{p^*} \right), \tag{4.68}$$

$$\beta_m = -\frac{\beta_m^*}{p^*}, \tag{4.69}$$

where μ^* is the intercept parameter, defined in Subsection 4.1.2, with its value being the first element in the coefficient vector $\boldsymbol{\beta}^*$.

Similarly, the Weibull AFT parameters can be formulated with respect to the Weibull proportional hazard model parameters:

$$\mu^* = -\frac{\log \lambda}{\tilde{p}}, \tag{4.70}$$

$$\beta_m^* = -\frac{\beta_m}{\tilde{p}}. \tag{4.71}$$

While the conversion of the Weibull parameters is straightforward, transformation of the standard errors from one perspective to the other is more complex. Because the Weibull parameters, either the AFT or the proportional hazard, follow an asymptotic multivariate normal distribution, the transformation of a covariance matrix is regularly considered within an integrated multivariate framework. The delta method, described in Appendix A, is generally applied to approximate the variance and covariance transforms, viewing the original and the transformed variances and covariances as two random matrices.

I start with the Weibull AFT regression model. Suppose that $\hat{\boldsymbol{\theta}}_1$ is a random vector of the Weibull AFT parameters $\left(\hat{\boldsymbol{\theta}}_1 = \hat{\theta}_{11}, \ldots, \hat{\theta}_{1M}\right)'$ with mean μ and variance matrix $\hat{\boldsymbol{V}}\left(\hat{\boldsymbol{\theta}}_1\right)$, and $\hat{\boldsymbol{\theta}}_2 = g\left(\hat{\boldsymbol{\theta}}_1\right)$ is a transform of $\hat{\boldsymbol{\theta}}_1$ where g is the link function specified by Equations (4.68) and (4.69). From the Taylor series expansion,

$$\hat{\boldsymbol{\theta}}_2 = g\left(\hat{\boldsymbol{\theta}}_1\right) = g(\boldsymbol{\mu}) + \left[\frac{\partial g(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} | \boldsymbol{\theta}_1 = \boldsymbol{\mu}\right](\boldsymbol{\theta}_1 - \boldsymbol{\mu}) + \mathrm{O}\left(\|\boldsymbol{\theta}_1 - \boldsymbol{\mu}\|^2\right), \tag{4.72}$$

where $\left(\|\boldsymbol{\theta}_1 - \boldsymbol{\mu}\|^2\right)$ is of a higher order. A first-order Taylor series expansion of $g\left(\hat{\boldsymbol{\theta}}_1\right)$ gives rise to an approximation of mean

$$\mathrm{E}\left[g\left(\hat{\boldsymbol{\theta}}_1\right)\right] \approx g(\boldsymbol{\mu}) \tag{4.73}$$

and the variance–covariance matrix $\hat{\boldsymbol{V}}\left(\hat{\boldsymbol{\theta}}_2\right)$:

$$\boldsymbol{V}\left[g\left(\hat{\boldsymbol{\theta}}_1\right)\right] \approx \left[\frac{\partial g(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1}\bigg| \boldsymbol{\theta}_1 = \boldsymbol{\mu}\right]' \boldsymbol{V}_{\hat{\boldsymbol{\theta}}_1} \left[\frac{\partial g(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1}\bigg| \boldsymbol{\theta}_1 = \boldsymbol{\mu}\right]. \tag{4.74}$$

Because $g\left(\hat{\boldsymbol{\theta}}_1\right)$ in the Weibull regression is a smooth nonlinear function of the $\hat{\boldsymbol{\theta}}_1$ values, $g\left(\hat{\boldsymbol{\theta}}_1\right)$ can be approximated by a linear function of $\hat{\boldsymbol{\theta}}_1$, in the region with a nonnegligible likelihood. Consequently, the variance–covariance matrix $\hat{\boldsymbol{V}}\left(\hat{\boldsymbol{\theta}}_2\right)$ can be formulated in the following linear form (Meeker and Escobar, 1998):

$$\begin{aligned}\boldsymbol{V}\left[g\left(\hat{\boldsymbol{\theta}}_1\right)\right] \approx & \sum_{m=1}^{M}\left[\frac{\partial g(\boldsymbol{\theta}_1)}{\partial \theta_{1m}}\right]^2 \boldsymbol{V}\left(\hat{\boldsymbol{\theta}}_1\right) \\ & + \sum_{m=1}^{M}\sum_{\substack{m'=1 \\ m'\neq \mathrm{m}}}^{M}\left[\frac{\partial g(\boldsymbol{\theta}_1)}{\partial \theta_{1m}}\right]\left[\frac{\partial g(\boldsymbol{\theta}_1)}{\partial \theta_{1m'}}\right]\mathrm{cov}(\theta_{1m}, \theta_{1m'}).\end{aligned} \tag{4.75}$$

The second component on the right of Equation (4.75) can be omitted if the elements in $\hat{\boldsymbol{\theta}}_1$ are uncorrelated.

Klein and Moeschberger (2003) expand Equation (4.75) by creating an example for times to death from laryngeal cancer. For convenience of further analysis and simplification, some of the transformational equations are displayed below, with certain notational modifications:

$$\begin{aligned}\mathrm{cov}\left(\hat{\beta}_m, \hat{\beta}_{m'}\right) \approx & \frac{\mathrm{cov}\left(\hat{\beta}^*_m, \hat{\beta}^*_{m'}\right)}{\left(\hat{p}^*\right)^2} - \frac{\hat{\beta}^*_m \,\mathrm{cov}\left(\hat{\beta}^*_m, \hat{p}^*\right)}{\left(\hat{p}^*\right)^3} - \frac{\hat{\beta}^*_{m'} \,\mathrm{cov}\left(\hat{\beta}^*_{m'}, \hat{p}^*\right)}{\left(\hat{p}^*\right)^3} \\ & + \frac{\hat{\beta}^*_m, \hat{\beta}^*_{m'} V\left(\hat{p}^*\right)}{\left(\hat{p}^*\right)^4}, \quad m, m' = 1, \ldots, M;\end{aligned} \tag{4.76}$$

$$V(\hat{\lambda}) \approx \exp\left(-2\frac{\hat{\mu}^*}{\hat{p}^*}\right)\left[\frac{V(\hat{\mu}^*)}{(\hat{p}^*)^2} - 2\frac{\hat{\mu}^* \operatorname{cov}(\hat{\mu}^*, \hat{p}^*)}{(\hat{p}^*)^3} + \frac{(\hat{\mu}^*)^2 V(\hat{p}^*)}{(\hat{p}^*)^4}\right]; \tag{4.77}$$

$$V(\tilde{p}) \approx \frac{V(\hat{p}^*)}{(\hat{p}^*)^4}; \tag{4.78}$$

$$\operatorname{cov}(\hat{\beta}_m, \hat{\lambda}) \approx \exp\left(-\frac{\hat{\mu}^*}{\hat{p}^*}\right)\left[\frac{\operatorname{cov}(\hat{\beta}^*_{\mathrm{m}}, \hat{\mu}^*)}{(\hat{p}^*)^2} - \frac{\hat{\beta}^*_j \operatorname{cov}(\hat{\beta}^*_m, \hat{p}^*)}{(\hat{p}^*)^3} - \frac{\hat{\mu}^* \operatorname{cov}(\hat{\mu}^*, \hat{p}^*)}{(\hat{p}^*)^3} + \frac{\hat{\beta}^*_m \hat{\mu}^* V(\hat{p}^*)}{(\hat{p}^*)^4}\right], \quad m = 1, \ldots, M; \tag{4.79}$$

$$\operatorname{cov}(\hat{\beta}_m, \hat{\tilde{p}}) \approx \frac{\operatorname{cov}(\hat{\beta}^*_m, \hat{p}^*)}{(\hat{p}^*)^3} - \frac{\hat{\beta}^*_m V(\hat{p}^*)}{(\hat{p}^*)^4}, \quad m = 1, \ldots, \mathrm{M}; \tag{4.80}$$

$$\operatorname{cov}(\hat{\lambda}, \hat{\tilde{p}}) \approx \exp\left(-\frac{\hat{\mu}^*}{\hat{p}^*}\right)\left[\frac{\operatorname{cov}(\hat{\mu}^*, \hat{p}^*)}{(\hat{p}^*)^3} - \frac{\hat{\mu}^* V(\hat{p}^*)}{(\hat{p}^*)^4}\right]. \tag{4.81}$$

The square roots of the transformed variances for the Weibull proportional hazard parameters yield the estimated standard errors, in turn deriving corresponding confidence intervals for the transforms. Based on the above expanded equations, Klein and Moeschberger (2003) transform the standard errors of the Weibull AFT coefficients to the standard errors of the proportional hazard coefficients using the data of patients with cancer of the larynx. The results are presented in Tables 12.1 and 12.2 of their book.

With more parameters specified, the expansion of Equation (4.75) to transform variances and covariances becomes much more tedious and cumbersome. Allison (2010) discusses the characteristics of the Weibull hazard function for different values of the scale p^* and presents related graphs (Figure 2.3, p. 21). He notes that test results of the null hypothesis on $\log T$ coefficients also serve as the results of the null hypothesis on log hazard coefficients. This is a reasonable proposition because the baseline distribution of $\log T$ and $\log h(t)$ are actually two transformed profiles of the Weibull baseline distribution of T. As a result, corresponding estimates in $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ should be subject to the same p-values with identical Wald statistics. This association suggests that test results from the Weibull AFT regression model can be borrowed to perform the significance test on the estimated regression coefficients in the Weibull hazard rate regression models, without performing the transformation of the AFT variance–covariance matrix for the proportional hazard parameter estimates.

Given this intimate association, standard errors of the estimated regression coefficients and the scale parameter contained in $\hat{\boldsymbol{\theta}}_2$ can be approximated by using a much simpler approach. Here, I first demonstrate the transformation of the standard error for a single estimated regression coefficient $\hat{\beta}^*_m$. Suppose that the two regression coefficients, $\hat{\beta}^*_m$ and $\hat{\beta}_m$, have exactly the same p-value associated with a common Wald statistic distributed as χ^2.

Given the large-sample approximate normal distribution of the ML estimator, it is reasonable to establish the following equation of approximation:

$$\frac{\beta_m^*}{SE\left(\beta_m^*\right)} \approx \frac{-\beta_m}{SE\left(\beta_m\right)}. \tag{4.82}$$

Multiplying both the denominator and numerator on the right side of Equation (4.82) with the scale parameter p^* yields

$$\frac{\beta_m^*}{SE\left(\beta_m^*\right)} \approx \frac{-\beta_m\left(p^*\right)}{SE\left(\beta_m\right)\left(p^*\right)}. \tag{4.83}$$

Because $-\beta_j p^* = \beta_j^*$, we have

$$SE\left(\beta_m\right) \approx \frac{SE\left(\beta_m^*\right)}{p^*}. \tag{4.84}$$

Therefore, the standard error of $\hat{\beta}_m$ can be approximated from the standard error of $\hat{\beta}_m^*$ over the scale parameter p^*. Standard errors of other hazard regression coefficients can be estimated using the same approach. Consequently, the tedious computation of the delta method can be avoided.

Using the same logic, the standard error of $\tilde{p}$, the shape parameter for the Weibull proportional hazard model, is

$$SE\left(\tilde{p}\right) \approx \frac{SE\left(p^*\right)}{\left(p^*\right)^2}. \tag{4.85}$$

The interested reader might want to practice how Equation (4.85) is derived. Notice that squaring Equation (4.85) gives rise to Equation (4.78) about $V\left(\tilde{p}\right)$ derived from the delta method.

To highlight the applicability of the above simplified calculations, I transform the standard errors of the Weibull AFT coefficients presented in Table 12.1 of Klein and Moeschberger's (2003) book to the standard errors of the Weibull proportional hazard coefficients, using Equations (4.84) and (4.85). Table 4.1 displays the original AFT parameter estimates and their standard errors, the transformed proportional hazard coefficients, the transformed standard errors of the proportional hazard parameters from Klein and Moeschberger's calculation (KM SE), and the transformed standard errors from Equations (4.84) and (4.85).

The transformed standard errors of the PH regression coefficients and the scale parameter from the simple approach are almost identical to those calculated from the delta method. I do not report the transformed standard error estimate of the intercept parameter because the application of Equation (4.84) yields the transformed standard error of $\log \hat{\lambda}$ (β_0) rather than the standard error for the scale parameter $\hat{\lambda}$ itself. Therefore, while the standard error of the

Table 4.1 Parameter estimates of Weibull AFT regression and the PH transforms.

Variable	Parameter estimate	Standard error	PH parameter estimate	KM SE for transforms	SE from new equations
Scale	0.88	0.11	1.13	0.14	0.14
Z_1	–0.15	0.41	0.17	0.46	0.47
Z_2	–0.59	0.32	0.66	0.36	0.36
Z_3	–1.54	0.36	1.75	0.42	0.41
Z_4	–0.02	0.01	0.02	0.01	0.01

model intercept is less important, the standard error of $\hat{\lambda}$ can be obtained by using the delta method.

Such a transformation, from the Weibull AFT variance–covariance matrix to the corresponding matrix for the proportional hazard parameter estimates, can also be performed by applying the bootstrap approximation method. Specifically, bootstrap sampling is used to simulate the repeated sampling process in a large number of bootstrap samples for computing the needed standard errors or confidence intervals of certain estimators, thereby reducing the dependence on large-sample approximation. The number of bootstrap samples is generally recommended as between 2000 and 5000. Theoretically, the bootstrapping method derives a range of an estimator, rather than a one-on-one transform, because each procedure can derive a unique set of parameter estimates. Additionally, the bootstrap approximation does not account for covariance between parameters and thus may introduce some specification problems. Therefore, this approach is a second-choice alternative when the large-sample approximation is valid and the likelihood-based methods are not too demanding computationally. This approximation method is standard, not specifically designed for survival analysis. Therefore, I do not describe the detailed procedures of bootstrap further in this text. The interested reader can read Efron and Tibshirani (1993) for more details.

I would like to make a recommendation that if the estimation of confidence intervals is not needed, the researcher might want to use the *p*-values of the AFT parameter estimates directly for the significant test on the Weibull proportional hazard parameters, thus saving the energy of performing further computations.

4.3.4 Illustration: A Weibull regression model on marital status and survival among older Americans

In Subsection 4.3.3, I examined the association between marital status and the probability of survival among older Americans for a twelve-month observation, adjusting for the confounding effects of age and educational attainment. The results did not display a significant effect of current marriage on the one-year mortality of older Americans. This lack of association, as indicated earlier, might be due to a relatively short observation interval, during which the impact of married life on the probability of survival cannot be captured.

In the present illustration, I revisit this topic by extending the observation period from 12 to 48 months (four years). The underlying null hypothesis is that currently married older persons do not have a better chance of survival than do those not currently married within a much lengthened observation period. With an extended observation interval, I specify a Weibull distribution of survival time to model the baseline survival function among older Americans. As previously specified, marital status is a dichotomous variable: 1 = 'currently married' and 0 = else, with the variable name given as 'Married.' Age and educational attainment are used as the control variables, and their centered measures, 'Age_mean' and 'Educ_mean,' are applied in the regression analysis.

As the hazard rate, especially among older persons, tends to increase over time, the use of the Weibull regression model is theoretically reasonable because of its monotonic property. As mentioned in Chapter 3, if the survival data arise from a Weibull distribution, the plot of $\log\left[-\log\hat{S}(t)\right]$ versus $\log(t)$ should display a roughly linear line. Given this feature, I first plot such a graph before starting the formal regression analysis, using the Kaplan–Meier survival estimates. The SAS program for plotting this curve is given below.

SAS Program 4.3:

```
......
ods graphics on;

ods select LogNegLogSurvivalPlot;
proc lifetest data = new OUTSURV = out1 plots = loglogs;
  time Months * Status(0);
run;
```

In SAS Program 4.3, I use the variable Status (1 = event, 0 = censored) again for executing the PROC LIFETEST procedure to construct a log negative $\log\hat{S}(t)$ versus log t plot using the Kaplan–Meier survival estimates. The ODS Graphics is enabled by specifying the ODS GRAPHICS ON statement with the ODS select name switching to LogNegLogSurvivalPlot. In the PROC LIFETEST statement, the option PLOTS = LOGLOGS tells SAS to construct a graph of the log negative $\log\hat{S}(t)$ function versus $\log t$ from the survival data of older Americans. SAS Program 4.3 yields the plot in Figure 4.3.

The plot of the $\log\left[-\log\hat{S}(t)\right]$ versus log (months) displays an approximately straight curve, thereby indicating the use of the Weibull regression in the present illustration to be reasonable. Nevertheless, plotting the graph of a linear relationship between a transformed survival function and a transformed time function is a tentative, rough approach for a functional check; it cannot be used to make a final decision regarding whether or not a given parametric regression fits the survival data, as also indicated in Section 4.2.

Next, the PROC LIFEREG procedure is used to fit the Weibull regression model on the survival data of older Americans. In the SAS system, the PROC LIFEREG procedure is based on the AFT perspective, so that it does not directly estimate hazard rate parameters. The Weibull parameter estimates on log(hazards), however, can be obtained from the estimates on the log(survival time) by using the SAS ODS procedure. Below is the SAS program for the estimation of the Weibull regression parameters on marital status and survival for a four-year period.

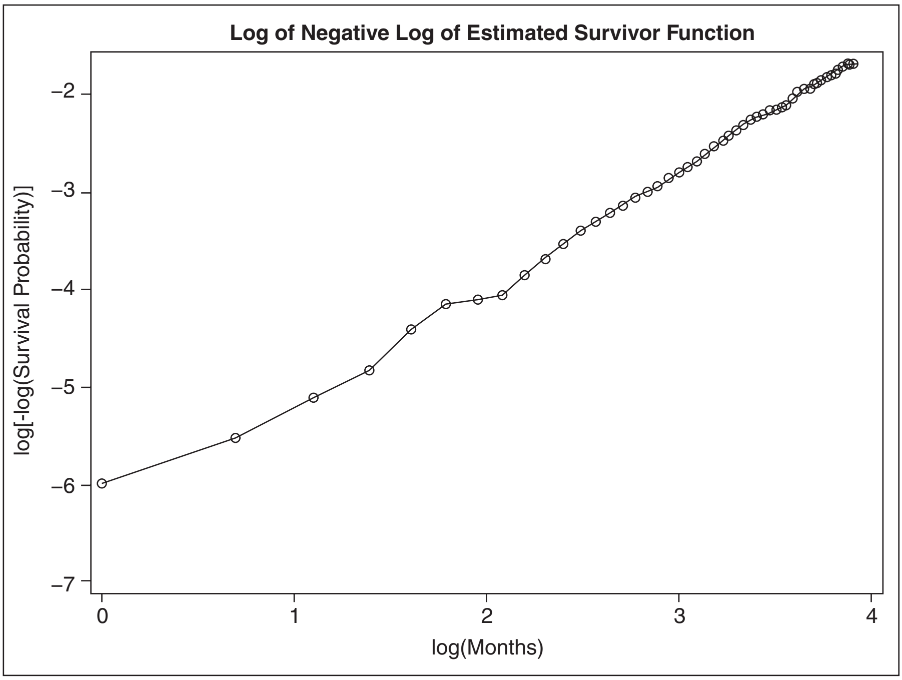

Figure 4.3 Plot of log negative log survival versus log time.

SAS Program 4.4:

```
......

* Compute xbeta estimates of the Weibull accelerated failure time model *;
ods graphics on ;
ods output Parameterestimates = Weibull_AFT;
proc lifereg data=new outest=Weibull_outest;
  class married;
  model duration*Censor(1) = married age_mean educ_mean / dist = Weibull;
  output out = new1 cdf = prob;
run;
ods graphics off;
ods output close;

* Compute and print corresponding PH estimates and HR *;
data Weibull_AFT;
  if _N_=1 then set Weibull_outest(keep=_SCALE_);
  set Weibull_AFT;
  string = upcase(trim(Parameter));
  if (string ne 'SCALE') & (string ne 'WEIBULL SHAPE')
  then PH_beta = -1*estimate/_SCALE_;
  else if PH_beta = .;
  AFT_beta = estimate;
  HR = exp(PH_beta);
  drop _SCALE_;
options nolabel;
```

```
proc print data=Weibull_AFT;
  id Parameter;
  var AFT_beta PH_beta HR;
  format PH_beta 8.5 HR 5.3;
  title1 "LIFEREG/Weibull on duration";
  title2 "ParameterEstimate dataset w/PH_beta & HR computed manually";
  title4 "PH_beta = -1*AFT_beta/Scale";
  title6 "HR = exp(PH_beta)";
run;

proc sgplot data = new1;
  scatter x = duration y = prob / group = married;
  discretelegend;
  title1 "Figure 4.4. Cumulative distribution function: Weibull";
run;
```

SAS Program 4.4 is a revised version of SAS Program 4.1, given the adjustment of the code for fitting a Weibull regression model. Therefore, the earlier part of the program is not presented. Briefly, in the omitted part a temporary SAS data file 'New' is created from the dataset chapter4_data. The survival time variable 'duration' is now specified in terms of a 48-month observation period. Accordingly, the dichotomous variable 'Censor,' 1 = censored and 0 = not censored, is redefined to conform to a 48-month observation window, with the Type I right censoring fixed at the end of the 48-month interval. The construction of the two centered control variables, Age_mean and Educ_mean, is saved into the temporary SAS data file 'New' for the Weibull regression analysis.

In SAS Program 4.4, I first ask SAS to estimate the regression coefficients of covariates on log(duration), with parameter estimates saved in the temporary SAS output file Weibull_AFT. The MODEL statement specifies three covariates, with the time distribution specified as Weibull. In SAS, the DIST = WEIBULL option can be omitted because the LIFEREG procedure fits the default Type 1 extreme-value distribution by using log(duration) as the response, equivalent to fitting the Weibull AFT regression model. The OUTPUT statement creates the output dataset New1.

The next section converts the saved Weibull log time coefficients (AFT_beta) into the Weibull log hazard coefficients (PH_beta) and the Weibull hazard ratios (HR), given the formulas in Subsection 4.3.3. Compared to SAS Program 4.1, an additional parameter SCALE is now displayed in the transformation procedure. Next, I request SAS to print three sets of parameter estimates – the regression coefficients on log(duration), the regression coefficients on log(hazards), and the hazard ratios. The test results of the null hypothesis on log time coefficients also serve as the results of the null hypothesis on log hazard coefficients because the two sets of the Weibull parameter estimates share the same *p*-values. As a result, in the present illustration I borrow test results from the Weibull AFT regression models to perform the significance tests on the regression coefficients in the Weibull hazard rate regression model. If needed, calculation of the standard errors for proportional hazard coefficients can be easily programmed in SAS.

Lastly, the PROC SGPLOT procedure is used to produce a plot of the cumulative distribution values versus the variable Duration. Some of the analytic results from SAS Program 4.4 are presented below.

SAS Program Output 4.2:

The LIFEREG Procedure

Model Information

Data Set	WORK.NEW
Dependent Variable	Log(duration)
Censoring Variable	censor
Censoring Value(s)	1
Number of Observations	2000
Noncensored Values	332
Right Censored Values	1668
Left Censored Values	0
Interval Censored Values	0
Number of Parameters	5
Name of Distribution	Weibull
Log Likelihood	-1132.63769

Fit Statistics

-2 Log Likelihood	2265.275
AIC (smaller is better)	2275.275
AICC (smaller is better)	2275.305
BIC (smaller is better)	2303.280

Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	5.3314	0.1071	5.1215	5.5414	2477.89	<.0001
married	Married	1	-0.0547	0.0946	-0.2401	0.1307	0.33	0.5631
married	Not married	0	0.0000	.	.	.	.	.
age_mean		1	-0.0588	0.0076	-0.0737	-0.0439	60.13	<.0001
educ_mean		1	0.0163	0.0121	-0.0074	0.0400	1.82	0.1773
Scale		1	0.7944	0.0425	0.7154	0.8822		
Weibull Shape		1	1.2588	0.0673	1.1336	1.3978		

PH_beta = -1*AFT_beta/Scale

HR = exp(PH_beta)

Parameter	AFT_beta	PH_beta	HR
Intercept	5.33143	-6.71104	0.001
married	-0.05470	0.06886	1.071
married	0.00000	0.00000	1.000
age_mean	-0.05879	0.07401	1.077
educ_mean	0.01631	-0.02054	0.980
Scale	0.79443	.	.
Weibull Shape	1.25877	.	.

In SAS Program Output 4.2, the class level information and a statistic of model fitness are presented first. There are 332 uncensored observations (deaths) and 1668 right-censored observations, giving a much extended observation interval. As mentioned in the previous example, all censored observations are of Type I right censoring. The log likelihood for this Weibull regression model is −1132.6377. The section of fit statistics displays four indicators of model fitness, with the −2 log-likelihood statistic probably being the mostly widely used. In summary, values of these four statistics are close, generating the same conclusion about model fitting for this analysis.

The table Analysis of Maximum Likelihood Parameter Estimates displays the LIFEREG estimates on log(duration) given the Weibull distribution, derived from the maximum likelihood procedure. The intercept, 5.3314 ($SE = 0.1071$), is very strongly statistically significant. With a much extended observation period, however, currently married older persons are still not shown to have a significantly higher survival probability than the currently not married, after adjusting for the effects of age and education ($\beta_1 = 0.0547$, $p = 0.5631$). Therefore, it is likely that the bivariate association between marital status and survival, shown in Chapter 2 and statistically significant, is spurious. For the two control variables, age is negatively and statistically significantly associated with survival time ($\beta_2 = -0.0588$, $p < 0.0001$). Educational attainment is positively linked to survival time, but this effect is not statistically significant ($\beta_3 = 0.0163$, $p = 0.1773$). Here, the AFT coefficients estimated for the Weibull regression model are similar to those for the exponential regression within a shortened observation interval. The Weibull shape parameter $\tilde{p}$ is simply the reciprocal of the AFT parameter estimate p^* (1/0.7944 = 1.2588). The value of $\tilde{p}$, greater than 1, exhibits a monotonically increasing hazard function within the 48-month interval, as expected. This shape parameter estimate can be easily tested by $(\tilde{p} - 1.0)/SE$, which is statistically significant (0.2588/0.0673 = 3.8455, $p < 0.01$).

The last section in SAS Program Output 4.2 presents the three sets of the Weibull parameter estimates – AFT_beta, PH_beta, and HR. The hazard ratio of each covariate (HR) is presented to highlight the multiplicative effects of covariates on the mortality of older Americans within a four-year time interval. For example, a one-year increase in age would increase the hazard rate by about 8 % ($HR = 1.077$), other variables being equal. This effect, again, displays tremendous similarity to the hazard ratio in the exponential regression model.

The plot in Figure 4.4 displays the cumulative distribution function, contained in the output dataset New1, for currently married and currently not married older persons. Two cumulative distribution functions from the Weibull regression model are displayed, one for currently married older persons (o) and one for those currently not married (+). Even with an extended observation period, the plot does not demonstrate a clear pattern of the relationship between marital status and the cumulative probability of death. This lack of association, after controlling for the two 'lurking' variables, might be due to the fact that the probability of survival and marital status are both affected by age and education, so that the bivariate association between current marriage and survival, statistically significant, is spurious without a true causal relationship. A competing explanation is that the extension of the observation period to four years is still not long enough to capture the impact of current marriage on the mortality of older Americans. In any case, this illustration of multivariate regression analysis highlights the importance of performing regression modeling for testing a hypothesis on the causal linkage between two factors in a survival analysis.

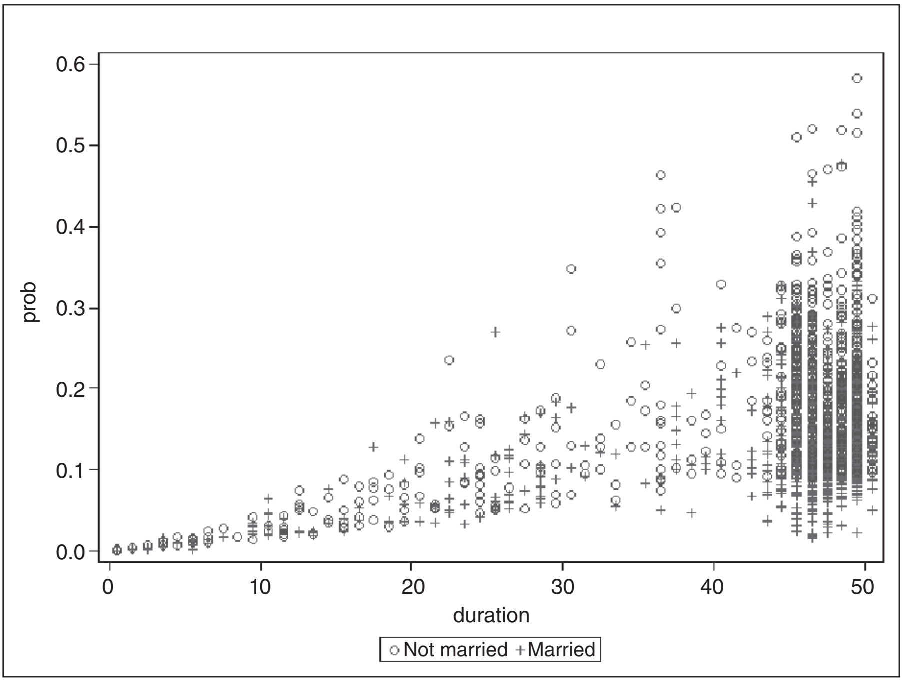

Figure 4.4 Cumulative distribution function: Weibull.

4.4 Log-logistic regression models

The Weibull proportional hazard rate model is popular in biomedical and demographic studies because it is easy to understand and the distribution agrees with most schedules of human mortality and morbidity. This regression model, however, has limitations under certain circumstances. The Weibull distribution is monotonic, so cannot be used to describe lifetime events in which the hazard rate changes direction. For example, human mortality decreases over age before reaching age 10, and then it increases steadily until all cohort members die out. Obviously, it is inappropriate to use the Weibull function to describe this long-standing process. As indicated in Chapter 3, there are several parametric lifetime functions that describe events whose rate changes direction over the life course. Lognormal and log-logistic functions are perhaps the most popular perspectives in this regard. When there are considerable censored observations in empirical data, the log-logistic distribution is believed to provide more accurate parameter estimates than the lognormal model, thereby serving as a preferable distributional function to model survival data with heavy censoring.

This section describes the log-logistic regression model and its statistical inference. In line with the previous two sections, an empirical illustration is provided to show how to apply this parametric regression model in empirical research.

4.4.1 Specifications of the log-logistic AFT regression model

The log-logistic regression model is formally specified as an accelerated failure time function because it can be conveniently expressed in the formation of Equations (4.8) and (4.9).

Accordingly, most statistical software packages generate parameter estimates of the log-logistic regression model using the AFT function.

Let $Y = \log T$ and $\hat{b}$ be the scale parameter for the log-logistic distribution. Replacing $\log t$ and $\tilde{\sigma}$ in Equation (4.9) with y and $\hat{b}$, the log-logistic survival function of T with covariate vector $\boldsymbol{x}$ is

$$S\left(t; \boldsymbol{x}, \boldsymbol{\beta}^{*}, \hat{b}\right) = \left[1 + \exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}}\right)\right]^{-1}, \quad -\infty < y < \infty, \tag{4.86}$$

where the vector $\boldsymbol{\beta}^*$ contains parameters of the logistic function of Y because $\log T$ follows a logistic distribution if T is distributed as log-logistic. As usually practiced in specifying a parametric survival model, the AFT parameter vector includes an intercept parameter, referred to as μ^* (the first element in $\boldsymbol{x}'\boldsymbol{\beta}^{*}$).

From Equation (4.86), the hazard function in the log-logistic AFT regression can be readily defined, given their intimate relationship:

$$h\left(t; \boldsymbol{x}, \boldsymbol{\beta}^{*}, \hat{b}\right) = \frac{\exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}}\right)}{\hat{b}\left(1 + \frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}}\right)}, \quad -\infty < y < \infty. \tag{4.87}$$

With the AFT survival and the hazard functions specified, the density function of T in the log-logistic accelerated failure time regression model is

$$f\left(t; \boldsymbol{x}, \boldsymbol{\beta}^{*}, \hat{b}\right) = \frac{\exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}}\right)}{\hat{b}\left[1 + \frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}}\right]\left[1 + \exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}}\right)\right]}, \quad -\infty < y < \infty. \tag{4.88}$$

Given the above three basic log-logistic AFT regression functions, the likelihood function for a sample of n individuals is given by

$$L\left(\boldsymbol{x}, \boldsymbol{\beta}^{*}, \hat{b}\right) = \prod_{i=1}^{n} \left\{\frac{\exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}}\right)}{\hat{b}\left(1 + \frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}}\right)}\right\}^{\delta_i} \left[1 + \exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}}\right)\right]^{-1}, \quad -\infty < y < \infty. \tag{4.89}$$

Clearly, the contribution of an event time to the likelihood is the density because when $\delta_i = 1$, the product of the hazard rate and the probability of survival defines a density function. If the observation is censored, $\delta_i = 0$, so the likelihood is the probability of survival.

The corresponding log-likelihood function can then be readily specified:

$$\log L\left(\boldsymbol{x}, \boldsymbol{\beta}^{*}, \hat{b}\right) = \sum_{i=1}^{n} \delta_i \left\{ -\log \hat{b} + \left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}} \right) - 2\log\left[1 + \exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}} \right) \right] \right\} \\ - (1 - \delta_i) \log\left[1 + \exp\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^{*}}{\hat{b}} \right) \right]. \tag{4.90}$$

Analogous to inferences for other parametric regression models, maximizing the above log-likelihood function with respect to the model parameters yields the MLE estimates of the log-logistic AFT regression model. The inverse of the observed information matrix generates the variance–covariance matrix of the log-logistic parameter estimates. As the estimating procedures follow the standard algorithms, as described extensively in Section 4.1, I do not elaborate the concrete equations for the derivation of parameter estimates in the log-logistic AFT regression model.

4.4.2 Retransformation of AFT parameters to untransformed log-logistic parameters

The AFT regression modeling is widely applied for the derivation of parameter estimates due to its flexibility and convenience. The log-logistic regression model is no exception. Some statisticians, however, consider it more convenient to interpret analytic results using the original, or untransformed, log-logistic parameters on survival time per se. This stance highlights the importance of converting the log-logistic AFT function into the untransformed log-logistic distribution, given parameter estimates obtained from the AFT perspective.

For retransformation of the log-logistic AFT parameters to the original log-logistic parameters on survival times, first I specify the log-logistic parameters on T in terms of the AFT parameters on $\log T$, given by

$$\lambda = \exp\left(-\frac{\mu^{*}}{\hat{b}} \right), \tag{4.91}$$

$$\beta_j = -\frac{\beta_j^{*}}{\hat{b}}, \tag{4.92}$$

$$\hat{p} = \frac{1}{\hat{b}}, \tag{4.93}$$

where μ^{*} is the intercept parameter, the first element in the coefficient vector $\boldsymbol{\beta}^{*}$. In form, these log-logistic transforms are much like those specified for the Weibull proportional hazard transforms although the log-logistic function cannot be expressed as a proportional hazard function. As a result, the transformation procedures described in Subsection 4.3.3 can be well applied for the log-logistic regression.

Given the specification of the above untransformed log-logistic regression parameters, the survival function given $\boldsymbol{x}$ can be written in the form of the untransformed log-logistic perspective, given by

$$S(t;\boldsymbol{x},\boldsymbol{\beta},\lambda,\tilde{p}) = \left[1+\lambda\exp(\boldsymbol{x}'\boldsymbol{\beta})t^{\tilde{p}}\right]^{-1}. \tag{4.94}$$

The reader might want to compare Equation (4.94) with Equation (4.86) for a better understanding of the implications of such a transformation. Given a univariate log-logistic distribution of T, the two sets of parameters just reflect two profiles of the same distributional function, thus being intimately associated. In other words, the parameters on $\log T$ can be easily transformed into the parameters on T.

Given such an intimate relationship, the log-logistic hazard rate can be written as a function of the untransformed parameters on event time T:

$$h(t;\boldsymbol{x},\boldsymbol{\beta},\lambda,\tilde{p}) = \frac{\lambda\tilde{p}\exp(\boldsymbol{x}'\boldsymbol{\beta})t^{\tilde{p}-1}}{1+\lambda\exp(\boldsymbol{x}'\boldsymbol{\beta})t^{\tilde{p}}}. \tag{4.95}$$

Equation (4.95) derives exactly the same hazard rate as that from Equation (4.84), using a different expression for the formation with a new set of parameters.

As can be expressed by the product of the survival and the hazard functions, the p.d.f. of the original log-logistic regression on T is

$$f(t;\boldsymbol{x},\boldsymbol{\beta},\lambda,\tilde{p}) = \frac{\lambda\tilde{p}\exp(\boldsymbol{x}'\boldsymbol{\beta})t^{\tilde{p}-1}}{\left[1+\lambda\exp(\boldsymbol{x}'\boldsymbol{\beta})t^{\tilde{p}}\right]^{2}}. \tag{4.96}$$

Given the above specifications, the log-logistic regression model on T can be articulated by retransforming the AFT parameter estimates.

The variance–covariance matrix of parameter estimates λ, $\boldsymbol{\beta}$, and $\tilde{p}$ can be approximated by applying the delta method, with the procedure described in Subsection 4.3.3. As displayed in 4.3.3, however, standard errors of the untransformed parameter estimates, except for μ^*, can be approximated by using a much simpler approach based on the large-sample approximate normal distribution of the ML estimators (see Equations (4.84) and (4.85)). If the value of the standard error is not of particular concern, the significance test on $\boldsymbol{\beta}$ and $\tilde{p}$ can be performed by borrowing the test results on $\boldsymbol{\beta}^*$ and $\hat{b}$, particularly since the two sets of regression coefficients share the same p-values.

Unlike the case in the Weibull regression model, the term $\exp(\boldsymbol{x}'\boldsymbol{\beta})$ in the log-logistic regression cannot be viewed as the proportional effects on the lifetime outcome. This effect vector, however, can be understood as representing odds ratios on the probability of survival at time t. According to the specification of the logistic regression model, the odds of the survival rate at t, given $\boldsymbol{x}$, is

$$\text{Odds}\left[S(t;\boldsymbol{x},\boldsymbol{\beta},\lambda,\tilde{p})\right] = \frac{S(t;\boldsymbol{x},\boldsymbol{\beta},\lambda,\tilde{p})}{1-S(t;\boldsymbol{x},\boldsymbol{\beta},\lambda,\tilde{p})} = \frac{\left[1+\lambda\exp(\boldsymbol{x}'\boldsymbol{\beta})t^{\tilde{p}}\right]^{-1}}{\left[\lambda\exp(\boldsymbol{x}'\boldsymbol{\beta})t^{\tilde{p}}\right]^{-1}}. \tag{4.97}$$

Given $[\exp(\boldsymbol{x}'\boldsymbol{\beta})]^{-1} = \exp(-\boldsymbol{x}'\boldsymbol{\beta})$, Equation (4.97) can be rewritten as

$$\text{Odds}\left[S(t;\boldsymbol{x},\boldsymbol{\beta},\lambda,\tilde{p})\right] = \exp(-\boldsymbol{x}'\boldsymbol{\beta})\frac{\left[1+\lambda t^{\tilde{p}}\right]^{-1}}{\left[\lambda t^{\tilde{p}}\right]^{-1}} = \exp(-\boldsymbol{x}'\boldsymbol{\beta})\frac{S(t|\boldsymbol{x}=0)}{1-S(t|\boldsymbol{x}=0)}. \tag{4.98}$$

Conceivably, a single effect term $\exp(-b_m)$ indicates the change in the individual odds of surviving with a 1-unit increase in covariate x_m, other variables being equal. This interpretation highlights the fact that the term $\exp(-\boldsymbol{x}'\boldsymbol{\beta})$ is odds proportional to the probability of survival. Given a special link to the logistic distribution, the log-logistic survival function is the only parametric survival model that possesses both the proportional odds and the AFT representations (Klein and Moeschberger, 2003).

4.4.3 Illustration: The log-logistic regression model on marital status and survival among the oldest old Americans

In Subsection 4.3.4, I reexamined the association between marital status and the probability of survival among older Americans for a 48-month observation, adjusting for the confounding effects of age and educational attainment. The results did not confirm current marriage to have a statistically significant impact on the four-year survival rate of older persons. The use of the Weibull regression model is considered to be theoretically sound because the death rate generally monotonically increases among older persons. Some scientists contend, however, that the population hazard rate is the average risk among surviving individuals who are physically selected from the 'survival of the fittest' process (Hougaard, 1995; Liu *et al.*, 2010; Vaupel, Manton, and Stallard, 1979). Empirically, the observed hazard function for a population or a population subgroup is often found to rise exponentially at younger and middle age, and then the rate of increase will be leveled-off or even decline at some oldest ages. This phenomenon is thought to be because selection eliminates less fit individuals from a given population as its age increases. Its mortality rate increases less rapidly with age than it would otherwise be because the surviving members of the population are the ones who are most fit.

In the present illustration, I analyze the association between marital status and survival among older persons aged 85 or over, generally termed the oldest old. Given such an old population, I assume a log-logistic distribution of survival time, seeking to capture the possible nonmonotonic process among those oldest old. Using this parametric regression model, the shape parameter of the log-logistic regression model displays whether the hazard rate is monotonic over time. The Weibull regression model, though highly flexible in general situations, does not have the capability to capture the possible directional change over time in the hazard rate.

As previously specified, marital status is a dichotomous variable: 1 = 'currently married' and 0 = else, with the variable name given as 'Married.' Specification of age and educational attainment remains the same, and their centered measures, 'Age_mean' and 'Educ_mean,' are used in the regression analysis as control variables. The following SAS program estimates the log-logistic regression model on marital status and survival among the oldest old Americans within a 4-year observation period.

SAS Program 4.5:

```
......

* Compute xbeta estimates of the log-logistic accelerated failure time model *;
ods graphics on ;
ods output Parameterestimates = Llogistic_AFT;
```

```
proc lifereg data=new outest=Llogistic_outest; where age ge 85;
  class married;
  model duration*Censor(1) = married age_mean educ_mean / dist = Llogistic;
  output out = new1 cdf = prob;
run;
ods graphics off;
ods output close;

* Compute and print corresponding untransformed estimates and OR *;
data Llogistic_AFT;
  if _N_=1 then set Llogistic_outest(keep=_SCALE_);
  set Llogistic_AFT;
  string = upcase(trim(Parameter));
  if (string ne 'SCALE') & (string ne 'Llogistic SHAPE')
  then UT_beta = -1*estimate/_SCALE_;
  else if UT_beta = .;
  AFT_beta = estimate;
  OR = exp(-UT_beta);
  P = 1 / _SCALE_;
  drop _SCALE_;
options nolabel;

proc print data=Llogistic_AFT;
  id Parameter;
  var AFT_beta UT_beta OR P;
  format UT_beta 8.5 OR 5.3;
  title1 "LIFEREG/Llogistic on duration";
  title2 "ParameterEstimate dataset ll/UT_beta & OR computed manually";
  title4 "UT_beta = -1*AFT_beta/Scale";
  title6 "OR = exp(-UT_beta)";
run;
```

SAS Program 4.5 bears tremendous resemblance to SAS Program 4.4, with some code adjustments for fitting the log-logistic regression. In the early part of the program, not presented, the procedures of constructing temporary SAS data files ('NEW,' 'NEW1') and specifying various variables are exactly the same as in SAS Program 4.4.

In the SAS LIFEREG procedure, I first ask SAS to estimate the parameters on log(duration), with parameter estimates saved in the temporary SAS output file Llogistic_AFT. The 'WHERE AGE GE 85' option informs SAS that only the oldest old persons are analyzed. The CLASS and MODEL statements also remain the same except for the specification of the DIST = LLOGISTIC option. The OUTPUT statement is identical to SAS Program 4.4, which creates the output dataset New1 containing the variable Prob to display the cumulative distribution function at observed responses.

The next section converts the saved log-logistic log time coefficients (AFT_beta) to the log-logistic untransformed time coefficients (UT_beta) and odds ratios (OR), using the formulas described in Subsection 4.4.2. Notice, compared to SAS Program 4.4, that the parameter estimation of OR takes the opposite direction, as deviated from 1, to the corresponding hazard ratios in the Weibull model because the proportional odds in the log-logistic regression reflect multiplicative changes in the probability of survival, rather than in the hazard rate. Next, I request SAS to print three sets of parameter estimates – the regression coefficients on log(duration), the regression coefficients on time duration itself, and the odds ratios on the probability of survival. As mentioned previously, test results of the null hypothesis on log time coefficients also serve as the results of the null hypothesis testing on time coefficients.

The main body of the analytic results from SAS Program 4.5 is presented in the following output table.

SAS Program Output 4.3:

```
                         The LIFEREG Procedure

                           Model Information

              Data Set                              WORK.NEW
              Dependent Variable               Log(duration)
              Censoring Variable                      censor
              Censoring Value(s)                           1
              Number of Observations                     193
              Noncensored Values                          66
              Right Censored Values                      127
              Left Censored Values                         0
              Interval Censored Values                     0
              Number of Parameters                         5
              Name of Distribution                 LLogistic
              Log Likelihood                    -161.9482104


                            Fit Statistics

           -2 Log Likelihood                           323.896
           AIC (smaller is better)                     333.896
           AICC (smaller is better)                    334.217
           BIC (smaller is better)                     350.210

           Analysis of Maximum Likelihood Parameter Estimates

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

Intercept                 1   4.5877    0.3415   3.9183   5.2571  180.46     <.0001
married      Married      1  -0.2498    0.1828  -0.6081   0.1086    1.87     0.1719
married      Not married  0   0.0000     .        .        .         .         .
age_mean                  1  -0.0242    0.0256  -0.0744   0.0259    0.90     0.3440
educ_mean                 1   0.0022    0.0214  -0.0398   0.0441    0.01     0.9192
Scale                     1   0.5344    0.0611   0.4271   0.6687

                    UT_beta = -1*AFT_beta/Scale

                        OR = exp(-UT_beta)

            Parameter    AFT_beta     UT_beta       OR        P

            Intercept     4.58770    -8.58444     5348    1.87118
            married      -0.24977     0.46736    0.627    1.87118
            married       0.00000     0.00000    1.000    1.87118
            age_mean     -0.02422     0.04533    0.956    1.87118
            educ_mean     0.00217    -0.00406    1.004    1.87118
            Scale         0.53442      .          .       1.87118
```

In SAS Program Output 4.3, there are 193 oldest old persons in the survival data, among whom are 66 uncensored (deaths) and 127 right censored. As indicated earlier, all censored observations are of Type I right censoring. The log-likelihood for this log-logistic AFT regression model is −161.9482, yielding the–2 log-likelihood statistic at 323.896.

The table Analysis of Maximum Likelihood Parameter Estimates displays the LIFEREG log-logistic parameter estimates on log(duration), derived from the maximum likelihood procedure. The intercept, 4.5877 ($SE = 0.3414$), is statistically significant. None of the three regression coefficients is statistically significant, perhaps because, at the oldest ages, all differentials in the probability of survival have vanished thanks to the 'selection of the fittest' effect. The log-logistic untransformed shape parameter $\hat{p}$ is simply the reciprocal of the AFT parameter estimate $\hat{b}$ (1/0.5344 = 1.8712), presented at the last column of the last section. The value of $\hat{p}$, greater than 1, indicates a nonmonotonic hazard function; namely, among the oldest old Americans, the hazard rate tends to increase initially and then decrease as time progresses. This untransformed shape parameter estimate on survival time can be easily tested by $(\hat{p}-1.0)/SE$, analogous to the formula for the Weibull proportional hazard shape parameter. Obviously, this shape parameter estimate is statistically significant, reflecting the existence of a selection effect among the oldest old Americans (the selection issues will be discussed extensively in Chapter 9).

The last section also presents the three sets of the log-logistic parameter estimates – AFT_beta, UT_beta, and OR. The odds ratio (OR), or proportional odds (PO), of each covariate is displayed to highlight the multiplicative effect on the odds of survival among the oldest old Americans. For example, a one-year increase in age would lower the individual odds of survival by 4.4 % ($HR = 0.956$), other covariates being equal.

Although the value of $\hat{p}$, greater than 1, suggests a nonmonotonic trend in the hazard rate, it may be interesting to examine whether the hazard schedules of the two marital status groups really change direction within a four-year observation period. Accordingly, below I use the SAS PROC GPLOT procedure to generate two hazard rate curves for the oldest old, one for currently married and one for currently not married. Equation (4.95) is applied, using the parameter estimates on survival time because they are more interpretable. In estimating these hazard curves, the values of control variables are fixed at sample means, so that the functions for those currently married and those currently not married can be compared effectively without being confounded. Because age and educational attainment are both centered at their sample means, their XBETA values need not be considered in calculating the predicted hazard rates. The SAS program for this plot is provided as follows.

SAS Program 4.6:

```
* Calculating hazard rates for log-logistic regression model *;

options ls=80 ps=58 nodate;
ods select all;
ods trace off;
ods listing;
title1;
run;

data Llogistic;
  do group = 1 to 2;
  do t = 1 to 50 by 0.1;
```

```
  if group = 1 then hazard = (0.00019 * 1.87118 * exp(0.46736) * (t**0.87118))
        / (1 + 0.00019 * exp(0.46736) * (t**1.87118));
  if group = 2 then hazard =  (0.00019 * 1.87118 * exp(0) * (t**0.87118))
        / (1 + 0.00019 * exp(0) * (t**1.87118));
  output;
  end;
  end;
run;

proc format;
  value group_fmt 1 = 'Not currently married'
                2 = 'Currently married'
run;

goptions targetdevice=winprtc reset=global gunit=pct border cback=white
         colors=(black red blue)
         ftitle=swissxb ftext=swissxb htitle=2 htext=1.5;

proc gplot data = Llogistic;
  format group group_fmt.;
  plot hazard * t = group;
     symbol1 c=black  i=splines  l=1  v=plus;
     symbol2 c=red    i=splines  l=2  v=diamond;
     title1 "Figure 4.6. Hazard functions for currently married and currently not
married persons";
     run;
quit;
```

SAS Program 4.6 produces two hazard functions for currently married and currently not married older persons, respectively. The value of λ, exp(−4.587 70/0.534 42) = 0.000 19, is derived from Equation (4.91). The resulting plot is displayed in Figure 4.5, which demonstrates that within the four-year observation period, the hazard rates for currently married and currently not married oldest old increase consistently, but the rate of increase is decreasing (the second derivative of the hazard function is negative). Therefore, the value of $\widehat{p}$ greater than 1 simply represents a long-term nonmonotonic trend in the mortality of the oldest old; it does not describe an actual nonmonotonic process within a limited observation period. Therefore, these two log-logistic curves can also be described by using the Weibull regression model regardless of the value of $\widehat{p}$.

It is also interesting to note that the oldest old Americans who are currently married have higher mortality rates than do their counterparts who are currently not married, although this effect is not statistically significant. The emergence of this pattern among the oldest old is perhaps related to changes in the distribution of an individual 'frailty' due to the selection effect. This issue will be further discussed in Chapter 9.

A competing possibility for the absence of a significant association between marital status and the mortality of oldest old Americans comes from the fact that a four-year observation period is still not stretched out enough to capture the dynamics of the effect. In Chapter 5, I will analyze this issue by specifying a much longer observation interval using the Cox proportional model.

4.5 Other parametric regression models

As can be summarized from previous sections, each of the exponential, the Weibull, and the log-logistic regression models has its unique characteristics and practicability in survival

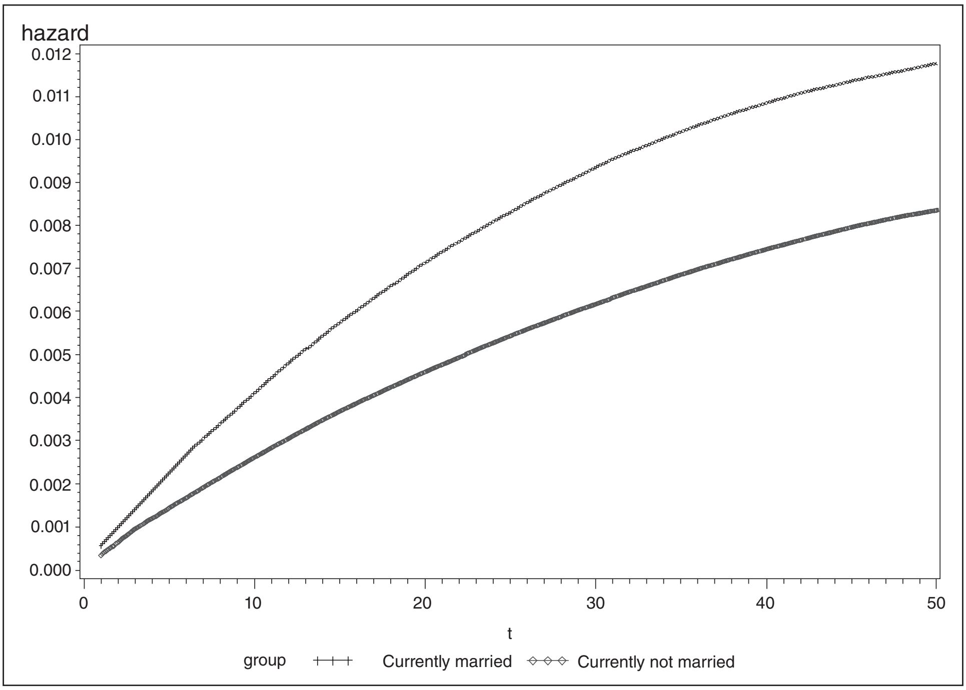

Figure 4.5 Hazard functions for currently married and currently not married persons.

analysis. The exponential regression model describes the survival data with a constant baseline hazard rate, whereas the Weibull regression model provides a flexible perspective to portray the hazard function, which is monotonically increasing or decreasing over time. Unlike the above two parametric models, the log-logistic regression model simulates nonmonotonic processes of the relative risk on the hazard rate that increase initially and then consistently decrease in the later stage of a life course.

There are some other regression models that have also seen tremendous applications in survival analysis, each corresponding to a specific pattern of survival processes. In this section, I briefly introduce two additional parametric models: the lognormal and the gamma regression models.

4.5.1 The lognormal regression model

As mentioned in Chapter 3, the lognormal distribution is another popular parametric function for the description of a nonmonotonic process in the hazard function. Its wide applicability is based on its simplicity because the logarithm of a lognormal distribution is normally distributed with mean μ and variance σ^2. Given the lognormal distribution, the AFT survival function, given covariate vector $\boldsymbol{x}$ and parameter vector $\boldsymbol{\beta}^*$, can be modeled by

$$S\left(t; \boldsymbol{x}, \boldsymbol{\beta}^*, \breve{b}\right) = 1 - \Phi\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^*}{\breve{b}}\right), \quad -\infty < y < \infty, \tag{4.99}$$

where $\Phi(\cdot)$ indicates the standard normal cumulative distribution function (the probit function), $\breve{b}$ is the scale parameter for the lognormal distribution, and the coefficient vector $\boldsymbol{\beta}^*$ includes an intercept parameter μ^*. The probability density function of the lognormal regression model is

$$f\left(t; \boldsymbol{x}, \boldsymbol{\beta}^*, \breve{b}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y - \boldsymbol{x}'\boldsymbol{\beta}^*}{\breve{b}}\right)^2\right]^{-}, \quad -\infty < y < \infty. \tag{4.100}$$

With the survival and density functions specified, the hazard function can be readily formulated as the ratio of $f\left(t; \boldsymbol{x}, \boldsymbol{\beta}^*, \breve{b}\right)$ over $S\left(t; \boldsymbol{x}, \boldsymbol{\beta}^*, \breve{b}\right)$. It then follows that the log-likelihood function can be written for the derivation of lognormal parameter estimates by using the standard estimating procedures described in Section 4.1.

Though simpler in formulation than the log-logistic regression model, the lognormal regression model does not behave well in the presence of heavy censoring. Accordingly, this model can be applied to describe nonmonotonic hazards only when the survival data do not contain many censored observations.

4.5.2 Gamma distributed regression models

The gamma distribution is often used as a part of a regression model in survival analysis. In particular, the gamma model has been applied for reflecting a skewed distribution of random errors when the assumption of normality on individual disturbances cannot be satisfied. As described in Chapter 3, the density function of a gamma distribution is

$$f(t) = \frac{\lambda(\lambda t)^{\tilde{k}-1} e^{-\lambda t}}{\Gamma\left(\tilde{k}\right)}, \quad t > 0. \tag{4.101}$$

where $\tilde{k}$ is the shape parameter and λ is the scale parameter. The mean and variance of survival time T for the gamma distribution are

$$\mathrm{E}[T; T \sim \Gamma(\lambda, \tilde{k})] = \frac{\tilde{k}}{\lambda}, \tag{4.102}$$

$$V[T; T \sim \Gamma(\lambda, \tilde{k})] = \frac{\tilde{k}}{\lambda^2}. \tag{4.103}$$

When the gamma distribution is used to address skewed disturbances, the hazard rate model with an arbitrarily assigned error term can be written as

$$\mathrm{E}\left[h\left(t, \boldsymbol{x}, \tilde{k}, \lambda\right)\right] = \frac{h_0(t)\exp(\boldsymbol{x}'\boldsymbol{\beta})\tilde{k}}{\lambda} = h_0(t)\exp\left[\boldsymbol{x}'\boldsymbol{\beta} + \log\left(\frac{\tilde{k}}{\lambda}\right)\right]. \tag{4.104}$$

Given the flexibility of the gamma distribution, the above specification has been applied in modeling bivariate survival times and individual frailty (Vaupel, Manton, and Stallard, 1979; Wild, 1983; Yashin and Iachine, 1995). For convenience of analysis, researchers often impose the conditions that $\lambda = \tilde{k}$ and $\sigma^2 = 1/\tilde{k} = 1/\lambda$. Without these conditions, the estimation of a gamma distributed regression model becomes tedious and cumbersome due to the addition of more model parameters. In Chapter 9, I will describe the use of the gamma model as a statistical means to account for selection bias in survival data with dependence.

With the specification of one more parameter, the gamma distribution is extended to the *generalized gamma model*. This extended gamma distribution is sometimes used to test the efficiency of the exponential, the Weibull, and lognormal regression models because this distribution contains those parametric functions as special cases. Given three parameters, the hazard function can take a variety of shapes, thus possessing the capability to model some complicated lifetime events. When a parametric regression model involves a large number of covariates, however, the use of the generalized gamma function can cause overparameterization and overfitting of survival data, thereby making the model statistically inefficient.

4.6 Parametric regression models with interval censoring

The above descriptions of parametric regression models are based on the assumption that all censored observations are right censored. In reality, the exact timing for the occurrence of a particular event is often not observed, and instead the information available to the researcher is only a time interval (t_{i-1}, t_i) during which an event occurs. In clinical trials and most observational surveys of a panel design, all event times are recorded in terms of specific time intervals, rather than of exact time points, thus suggesting interval censoring. When a time interval is small, T_i can be approximated by taking the midpoint of a particular interval, as mentioned in Chapter 1. When the interval is large, however, regular inference of survival processes may overlook some important information, in turn leading to some bias in parameter estimates.

This section describes valid inferences to tackle the issue of interval censoring in the regression analysis. An empirical illustration is then provided to demonstrate how to perform such analysis using SAS. The inference for left censoring bears some resemblance to that of interval censoring, but it is not particularly displayed in this text.

4.6.1 Inference of parametric regression models with interval censoring

Inference of parametric regression models with interval censoring is generally based on the assumption that the time span of a censored interval is independent of the underlying exact failure time. Given this assumption, the estimating approach for interval censoring is simply the combination of the standard procedures on right-censored survival data and the basic likelihood function for interval censoring, provided in Chapter 1.

For a random sample of n individuals with independent interval censoring, the event time for the ith individual is denoted by $T_i \in (\bar{l}_i, \bar{r}_i]$, where $i = 1, 2, \ldots n$, and $\bar{l}$ and $\bar{r}$ indicate left and right endpoints of a particular time interval in which the event occurs. Suppose $\bar{l}_i < \bar{r}_i$ for all $i = 1, \ldots, n$ ($\bar{l}_i$ may be 0 and $\bar{r}_i$ may be ∞), and $\bar{L}_i < \bar{R}_i$ are the observed left and right endpoints for the interval censored T_i. It follows that the likelihood of experiencing a particular event for the ith individual is

$$L_i(\boldsymbol{\theta}) = F\left(\bar{R}_i; \boldsymbol{\theta}\right) - F\left(\bar{L}_i; \boldsymbol{\theta}\right), \tag{4.105}$$

where $F(t; \boldsymbol{\theta})$ is the cumulative distribution function for the event time given $\boldsymbol{\theta}$.

Assuming a continuous parametric function for the distribution of F, the likelihood function with respect to $\boldsymbol{\theta}$ for a sample of n individuals can be conveniently written as an AFT function:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} [F(\tilde{u}_i) - F(\tilde{v}_i)] \\ &= \prod_{i=1}^{n} [S(\tilde{v}_i) - S(\tilde{u}_i)], \end{aligned} \tag{4.106}$$

where $S(t; \boldsymbol{\theta})$ is the parametric survival function and $\tilde{u}_i$ and $\tilde{v}_i$ are defined by

$$\tilde{u}_i = \frac{\log \bar{R}_i - \mu^* - \boldsymbol{x}_i'\boldsymbol{\beta}}{\tilde{\sigma}},$$
$$\tilde{v}_i = \frac{\log \bar{L}_i - \mu^* - \boldsymbol{x}_i'\boldsymbol{\beta}}{\tilde{\sigma}}.$$

Equation (4.106) shows that each observation contributes two pieces of information to the likelihood, $S\left(\bar{L}_i, \boldsymbol{\theta}_i\right)$ and $S\left(\bar{R}_i, \boldsymbol{\theta}_i\right)$, which follow the same distributional function. As defined, when both $\bar{L}_i$ and $\bar{R}_i$ are observed and $\bar{L}_i < \bar{R}_i$, the observation is interval censored; when both $\bar{L}_i$ and $\bar{R}_i$ are observed but $\bar{L}_i = \bar{R}_i$, $L_i(\boldsymbol{\theta})$ is the density function of T_i. Likewise, if $\bar{L}_i$ is observed but $\bar{R}_i$ is not, the observation is right censored.

Taking log values on both sides of Equation (4.106) yields the log likelihood function of independent interval censored data:

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left[S\left(\tilde{v}_j\right) - S\left(\tilde{u}_j\right)\right]. \tag{4.107}$$

With interval censoring assumed to be random, the total score statistic is

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\partial\left[S\left(\bar{L}_i, \boldsymbol{\theta}\right) - S\left(\bar{R}_i, \boldsymbol{\theta}\right)\right] / \partial \boldsymbol{\theta}}{S\left(\bar{L}_i, \boldsymbol{\theta}\right) - S\left(\bar{R}_i, \boldsymbol{\theta}\right)}. \tag{4.108}$$

As a standard procedure, the maximum likelihood estimates of $\boldsymbol{\theta}$ can be obtained by the solution of $\partial \log L(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} = \mathbf{0}$. The information matrix for interval-censored survival data can be obtained by applying the term $-\partial^2 \log L(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$, with its inverse yielding the approximates of the variances and covariances for parameter estimates with interval censoring. In the SAS system, interval censoring can be handled by applying the PROC LIFEREG procedure, as shown in the following illustration.

4.6.2 Illustration: A parametric survival model with independent interval censoring

In Subsection 4.3.4, I provided an empirical example of the Weibull regression model on marital status and the probability of survival among older Americans, using 'month' as the time scale. Strictly speaking, this specification of event time can cover up some information of variability in survival times because the measure ‘month’ in fact records the midpoint of a time interval. For example, a duration of 10.5 months actually represents all event times ranging from 10.00 months to 10.99 months.

In the present illustration, I revise SAS Program 4.4 by fitting a Weibull regression model that particularly specifies interval censored survival times. My intention is to examine whether the previous model of right censoring is subject to significant bias on the Weibull parameter estimates. Other parts of this revised Weibull regression model, including specification of covariates, remain the same as described in Section 4.3.

Specifically, I first create two time variables, time1 and time2, specifying the lower and upper endpoints of a given censoring interval, respectively. As mentioned earlier, if both values of time1 and time2 for an individual are observed and the lower value (time1) is less than the upper value (time2), the observation is interval censored. If the upper value is missing, the lower endpoint serves as the right-censored value. If the lower value is greater than the upper value or both values are missing, the observation is not used in the estimation. (Left-censored observations, not covered by this illustration, can be indicated by the condition that if the lower value is missing, the upper value is used as the left-censored value.) The following SAS program constructs the two time variables. Using these two endpoints, the Weibull regression model is respecified with interval censored data.

SAS Program 4.7:

```
......
if death = 0 then time1 = duration + 1;
  else time1 = duration;

if death = 0 then time2 = .;
  else time2 = duration + 1;
......

proc lifereg data=new outest=Weibull_outest;
  class married;
  model (time1, time2) = married age_mean educ_mean / dist = Weibull;
  output out = new1;
run;

......
```

Other parts of the SAS code, designated by dots, are exactly the same as those in SAS Program 4.4.

In SAS Program 4.7, the variables time1 and time2 are created (the variable ‘duration’ is now an integer, rather than a midpoint), with time1 specified as the lower and time2 as the upper endpoints of a censoring interval. Nonmissing values of both time1 and time2 indicate the occurrence of a death within the interval (time1, time2); therefore, SAS treats the observation as interval censored. If time1 is not missing but time2 is missing, the observation is viewed as right censored, with time1 moving forward by one unit (duration + 1) as the right-censored survival time. Compared to SAS Program 4.4, the only difference in the PROC LIFEREG procedure is the specification of two endpoints in the MODEL statement. Specifically, replacing the ‘MODEL DURATION*CENSOR(1) =’ statement in SAS Program 4.4, two endpoint variables are now specified in a parenthesis after the word MODEL, with the first being the lower value and the second being the upper value.

SAS Program 4.7 yields the following SAS output.

SAS Program Output 4.4:

```
                         The LIFEREG Procedure

                           Model Information

              Data Set                             WORK.NEW
              Dependent Variable                Log(time1)
              Dependent Variable                Log(time2)
              Number of Observations                  2000
              Noncensored Values                         0
              Right Censored Values                   1668
              Left Censored Values                       0
              Interval Censored Values                 332
              Number of Parameters                       5
              Name of Distribution                 Weibull
              Log Likelihood                  -2139.390037


                            Fit Statistics

          -2 Log Likelihood                         4278.780
          AIC (smaller is better)                   4288.780
          AICC (smaller is better)                  4288.810
          BIC (smaller is better)                   4316.785

             Analysis of Maximum Likelihood Parameter Estimates

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

Intercept                   1   5.3371    0.1069   5.1276   5.5466 2493.00     <.0001
married       Married       1  -0.0547    0.0943  -0.2397   0.1302    0.34     0.5618
married       Not married   0   0.0000     .        .        .        .         .
age_mean                    1  -0.0587    0.0076  -0.0736  -0.0439   60.29     <.0001
educ_mean                   1   0.0163    0.0121  -0.0074   0.0399    1.82     0.1771
Scale                       1   0.7923    0.0424   0.7134   0.8800
Weibull Shape               1   1.2621    0.0676   1.1364   1.4017

                      PH_beta = -1*AFT_beta/Scale

                           HR = exp(PH_beta)

           Parameter        AFT_beta      PH_beta        HR

           Intercept         5.33707     -6.73602     0.001
           married          -0.05473      0.06908     1.072
           married           0.00000      0.00000     1.000
           age_mean         -0.05874      0.07414     1.077
           educ_mean         0.01628     -0.02054     0.980
           Scale             0.79232       .           .
           Weibull Shape     1.26212       .           .
```

The log-likelihood value for the Weibull regression model with interval censoring is −2139.39, compared to −1132.638 estimated for the Weibull model of right censoring (see SAS Program Output 4.2). Accordingly, the −2 log-likelihood statistic, following a chi-square distribution, is 4278.78 for the interval-censored survival model, compared to 2265.275 in the right-censored model. The increase in the value of this statistic is not surprising because, in estimating the interval-censored parameters, each observation contributes two pieces of information to the total likelihood. Similar to the previous model, values of all four model fit statistics are very close, thereby generating the same conclusion on model fitness.

With respect to the maximum likelihood parameter estimates, the PROC LIFEREG procedure on the interval-censored data yields almost identical results to those obtained from the Weibull regression model with right censoring. The intercept, 5.3371 ($SE = 0.1069$), differs very slightly from 5.3314 ($SE = 0.1071$) estimated for the previous Weibull model. Additionally, the two Weibull regression models, right or interval censored, generate an equal regression coefficient of marital status ($\beta_1 = -0.0547$), with only a tiny difference in the standard error (0.0943 versus 0.0946). The regression coefficients of age and educational attainment are almost the same, with identical values of standard errors. Given such similarities, the corresponding proportional hazard coefficients and the hazard ratios are very close to those reported in SAS Program Output 4.2. The Weibull scale and shape parameters are also similar between the two Weibull regression models.

The above comparison provides very strong evidence that when a time interval for event occurrences is not uncommonly wide, the use of the midpoint as the substitute for an exact event time is appropriate and statistically efficient.

4.7 Summary

This chapter describes general inferences of parametric regression modeling and a number of specific parametric regression models that are widely used in survival analysis. Parametric regression models can be approached by two perspectives, the proportional hazard rate model and the accelerated failure time regression model. Each of those two approaches has its own strengths and limitations. In such scientific disciplines as medicine, biology, epidemiology, and sociology, researchers are inclined to analyze variability in the hazard rate as a function of model covariates because the hazard model describes the relative risk of an event occurrence in a direct, straightforward fashion. Therefore, the proportional hazard regression model becomes the most applied approach in those fields. On the other hand, statisticians prefer to apply the accelerated failure time perspective for specifying and estimating various parametric survival models, given its functional flexibility and statistical applicability for a large number of families of parametric distributions. Fortunately, for the parametric regression models that can be expressed in terms of both perspectives, one set of parameter estimates can be readily converted to the other set through mathematical transformations, including conversion of standard errors and confidence intervals.

Of various parametric regression models, the exponential, the Weibull, and the log-logistic functions are perhaps the most widely applied regression models in survival analysis. The exponential regression model is used to describe survival processes characterized by a constant baseline hazard rate. The Weibull regression model, on the other hand, portrays a hazard rate function that monotonically increases or decreases over time. Because many survival processes follow a monotone pattern, the flexible Weibull regression perspective boasts widespread applications in a variety of disciplines, ranging from medicine to sociology, epidemiology, engineering, and so forth. Unlike the above two parametric perspectives, the log-logistic regression model delineates nonmonotonic risk processes, with the hazard rate increasing initially and then consistently decreasing in the late stage of a life course.

If a particular probability distribution of survival data can be identified and validated, statistical inference based on a parametric regression perspective is considered more efficient and more precise than those derived from survival models in the absence of an explicit distributional function (Collett, 2003), particularly for survival data with a small sample size (Holt and Prentice, 1974). Given such statistical strengths, parametric regression modeling is popular in some disciplines where a true distributional function of survival times can be easily identified (e.g., industrial reliability analysis, mathematical biological studies). In other disciplines, however, ascertaining a parametric distribution is not a simple undertaking. The major concern is the selection effect inherent in survival processes. Because frailer individuals tend to die sooner, the estimation of individual mortality risks, aging progression, and mortality differentials among various population subgroups can be biased by some unobservable characteristics. Consequently, a particular observed parametric pattern of survival can be just an artifact shaped by successive exits of frailer subjects. Such uncertainty has gradually made parametric regression modeling a much less applied technique than the semi-parametric regression model.