Alternatives to the Cox Model

Publication details
Torben Martinussen, Limin Peng
Published online on: 22 Jul 2013

How to cite :- Torben Martinussen, Limin Peng. 22 Jul 2013, Alternatives to the Cox Model from: Handbook of Survival Analysis CRC Press
Accessed on: 07 Dec 2018

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
3

Alternatives to the Cox Model

Torben Martinussen
Department of Biostatistics, University of Copenhagen

Limin Peng
Department of Biostatistics and Bioinformatics, Emory University

CONTENTS

3.1 Additive hazards regression ........................................ 49
   3.1.1 Model specification and inferential procedures ............ 50
   3.1.2 Goodness-of-fit procedures .................................. 53
   3.1.3 Further results on additive hazard models ................. 54
       3.1.3.1 Structural properties of the additive hazard model .. 54
       3.1.3.2 Clustered survival data and additive hazard model .. 55
       3.1.3.3 Additive hazard change point model .................. 55
       3.1.3.4 Additive hazard and high-dimensional regressors ..... 56
       3.1.3.5 Combining the Cox model and the additive model ... 56
       3.1.3.6 Gastrointestinal tumour data ......................... 57

3.2 The accelerated failure time model ................................ 58
   3.2.1 Parametric models ........................................... 58
   3.2.2 Semiparametric models ....................................... 59
       3.2.2.1 Inference based on hazard specification ............. 59
       3.2.2.2 Inference using the additive mean specification ... 61

3.3 Quantile regression for survival analysis .......................... 62
   3.3.1 Introduction .................................................. 62
   3.3.2 Estimation under random right censoring with $C$ always known 63
   3.3.3 Estimation under covariate-independent random right censoring 63
   3.3.4 Estimation under standard random right censoring .......... 64
       3.3.4.1 Self-consistent approach ............................... 64
       3.3.4.2 Martingale-based approach ............................ 66
   3.3.5 Variance estimation and other inference .................... 67
   3.3.6 Extensions to other survival settings ....................... 68
   3.3.7 An illustration of quantile regression for survival analysis 68

Bibliography ........................................................... 71

3.1 Additive hazards regression

The Cox regression model is clearly the most used hazards model when analyzing survival data as it provides a convenient way of summarizing covariate effects in terms of relative risks. The proportional hazards assumption may not always hold, however. A typical vi-
olation of the assumption is time-changing covariate effects which is often encountered in bio-medical applications. A typical example is a treatment effect that varies with time such as treatment efficacy fading away over time due to, for example, tolerance developed by the patient. Under such circumstances, the Aalen additive hazard model (Aalen, 1980) may provide a useful alternative to the Cox model as it easily incorporates time-varying covariate effects. For Aalen’s model, the hazard function for an individual with a \( p \)-dimensional covariate \( X \) vector takes the form

\[
\alpha(t|X) = X^T \beta(t). \tag{3.1}
\]

The time-varying regression function \( \beta(t) \) is a vector of locally integrable functions, and usually the \( X \) will have 1 as its first component allowing for an intercept in the model. Model (3.1) is very flexible and can be seen as a first order Taylor series expansion of a general hazard function \( \alpha(t|X) \) around the zero covariate:

\[
\alpha(t|X) = \alpha(t, 0) + X^T \alpha'(t, X^*)
\]

with \( X^* \) on the line segment between 0 and \( X \). As we shall see below, it is the cumulative regression function

\[
B(t) = \int_0^t \beta(s) \, ds
\]

that is easy to estimate and its corresponding estimator converges at the usual \( n^{1/2} \)-rate. Estimation and asymptotical properties were derived by Aalen (Aalen, 1980), and further results for this model were given in Aalen (1989), Aalen (1993), and Huffer and McKeague (1991) derived efficient estimators.

The full flexibility of the model (3.1) may not always be needed and it is usually also of interest to try to simplify the model investigating whether or not a specific covariate effect really varies with time. A very useful model in this respect is the semiparametric additive hazards model of McKeague and Sasieni (1994). It assumes that the hazard function has the form

\[
\alpha(t|X, Z) = X^T \beta(t) + Z^T \gamma, \tag{3.2}
\]

where \( X \) and \( Z \) are \( p \)-dimensional and \( q \)-dimensional covariate vectors, respectively. Apart from its use for testing time-varying effects, this model is useful in its own right because it is then possible to make a sensible bias-variance trade-off, where effects that are almost constant can be summarized as such and effects that are strongly time-varying can be described as such. Note also that model (3.2) covers model (3.1) simply by leaving out the last part of model (3.2) by leaving \( Z \) empty. On the other hand, in practice, one may start out with model (3.1) and then test for a specific covariate whether its effect varies significantly with time. If this is not the case one may then simplify to model (3.2) with \( Z \) being that specific covariate, and one can then proceed in that way with some or all of the remaining covariates; the starting point, though, will now be model (3.2). Below we show how to do such inference in model (3.2). Lin and Ying (1994) considered a special case of (3.2), where only the intercept term is allowed to depend on time; this model is often referred to as the Lin and Ying model. A nice discussion of pros and cons of the additive hazards model can be found in Aalen et al. (2008).

### 3.1.1 Model specification and inferential procedures

We present estimation and inferential procedures based on model (3.2) since, as mentioned above, model (3.1) is covered by model (3.2) simply by leaving \( Z \) empty and therefore the below results also cover model (3.1).
Let $T$ be the survival time of interest with conditional hazard function $\alpha(t|X,Z)$ given the covariate vectors $X$ and $Z$. In practice $T$ may be right-censored by $U$ so that we observe $(\tilde{T} = T \wedge U, \Delta = I(T \leq U), X, Z)$. We assume that $T$ and $U$ are conditionally independent given $(X, Z)$. Let $(\tilde{T}_i, \Delta_i, X_i, Z_i)$ be $n$ iid replicates from this generic model. Under the above independent right-censoring scheme, the $i$th counting process $N_i(t) = I(\tilde{T}_i \leq t, \Delta_i = 1)$ has intensity
\[
\lambda_i(t) = Y_i(t) \left[ X_i^T \beta(t) + Z_i^T \gamma \right],
\]
where $Y_i(t) = I(t \leq \tilde{T}_i)$ is the at risk indicator. We assume that all counting processes are observed in the time-interval $[0, \tau]$, where $\tau < \infty$. Each counting process has compensator $\Lambda_i(t) = \int_0^t \lambda_i(s)ds$ such that $M_i(t) = N_i(t) - \Lambda_i(t)$ is a martingale. Define the $n$-dimensional counting process $N = (N_1, ..., N_n)^T$ and the $n$-dimensional martingale $M = (M_1, ..., M_n)^T$. Let also $X = (Y_1 X_1, ..., Y_n X_n)^T$, $Z = (Y_1 Z_1, ..., Y_n Z_n)^T$. The results presented below hold under some regularity conditions; see Martinussen and Scheike (2006). Writing the model in incremental form
\[
dN(t) = X(t)dB(t) + Z(t)\gamma dt + dM(t),
\]
and since $E\{dM(t)\} = 0$, this suggests the following (unweighted) least squares estimator of $\{B(t) = \int_0^t \beta(s) ds, \gamma\}$ McKeague and Sasieni (1994):
\[
\hat{\gamma} = \left( \int_0^\tau Z^T HZ dt \right)^{-1} \int_0^\tau Z^T HdN(t), \tag{3.3}
\]
\[
\hat{B}(t) = \int_0^t X^-dN(t) - \int_0^t X^-Z dt, \tag{3.4}
\]
where $X^-$ denotes the generalized inverse $(X^T X)^{-1}X^T$ and $H = I - XX^-$ assuming here that the required inverses exist. Considering the model with only non-parametric terms, model (3.1), means that $Z$ is empty and is readily seen from the latter display that the estimator of $B(t)$ in that case is given by
\[
\hat{B}(t) = \int_0^t X^-dN(t),
\]
which is the estimator initially proposed by Aalen (1980). Weighted estimators, and in theory efficient estimators, exist. However, they depend on the unknown parameters, and therefore, to calculate them, some kind of smoothing is needed which may compromise efficiency in practice; see McKeague and Sasieni (1994) for more on the efficient estimators. Here we will only deal with the unweighted estimators that can be calculated without any use of smoothing. It is quite obvious that these simple and direct estimators have sound properties as the following arguments show. Considering $\hat{\gamma}$, the counting process integral can be written as
\[
\int_0^\tau Z^T HdN(t) = \int_0^\tau Z^T Hd\{XdB(t) + Z\gamma dt\} + \int_0^\tau Z^T HdM(t) = \int_0^\tau Z^T HZ dt \gamma + \int_0^\tau Z^T HdM(t)
\]
since $HX = 0$, and therefore $\hat{\gamma}$ is an essentially unbiased estimator of $\gamma$ since the martingale term has zero-mean. Similar arguments concerning $\hat{B}(t)$ are readily given. The limit
distributions of the estimators are
\[ n^{1/2}\{\hat{\gamma} - \gamma\} = C_c^{-1}(\tau)n^{-1/2} \int_0^\tau Z^THdM(t), \]
\[ n^{1/2}\{\hat{B}(t) - B(t)\} = n^{1/2} \int_0^\tau X^{-}dM(t) - P(t)n^{1/2}\{\hat{\gamma} - \gamma\}, \]
where
\[ C(t) = n^{-1} \int_0^t Z^THZd\tau, \quad P(t) = \int_0^t X^+Zd\tau. \]

These limit distributions may be written as sums of essentially iid terms:
\[ n^{1/2}\{\hat{\gamma} - \gamma\} = n^{-1/2} \sum_{i=1}^n \epsilon_i^\gamma + o_p(1), \]
\[ n^{1/2}\{\hat{B}(t) - B(t)\} = n^{-1/2} \sum_{i=1}^n \epsilon_i^B(t) + o_p(1), \]
where
\[ \epsilon_i^\gamma = c^{-1}(\tau) \int_0^\tau \{Z_i - (z^Tx)(x^Tx)^{-1}X_i\}dM_i(t), \]
\[ \epsilon_i^B(t) = \int_0^\tau (x^Tx)^{-1}X_i dM_i(t) - p(t)\epsilon_i^\gamma, \]
where \( c(t) \) and \( p(t) \) denote the limits in probability of \( C(t) \) and \( P(t) \), respectively. Also, \( x^Tx \) is used as notation for the limit in probability of \( n^{-1}X^TX \), and similarly with \( z^Tx \).

The limit distributions may be simulated likewise what Lin et al. (1993) suggested in a Cox-model setting:
\[ n^{1/2}\{\hat{\gamma} - \gamma\} \sim n^{-1/2} \sum_{i=1}^n \epsilon_i^\gamma G_i, \]
\[ n^{1/2}\{\hat{B}(t) - B(t)\} \sim n^{-1/2} \sum_{i=1}^n \epsilon_i^B(t)G_i, \]
where we let \( \sim \) indicate that two quantities have the same limit distribution. In the latter display, \( G_1, \ldots, G_n \) are independent standard normals and \( \epsilon_i^\gamma \) is obtained from \( \epsilon_i^\gamma \) by replacing deterministic quantities with their empirical counterparts and by replacing \( M_i(t) \) with \( \hat{M}_i(t) \), \( i = 1, \ldots, n \), and similarly with \( \epsilon_i^B(t) \). The result is that, conditional on the data,
\[ \left( n^{-1/2} \sum_{i=1}^n \epsilon_i^\gamma G_i, n^{-1/2} \sum_{i=1}^n \epsilon_i^B(t)G_i \right) \]
will have the same limit distribution as
\[ \left( n^{1/2}\{\hat{\gamma} - \gamma\}, n^{1/2}\{\hat{B}(t) - B(t)\} \right). \]

The hypothesis of time-invariance
\[ H_0: \beta_p(t) = \beta_p, \]

focusing without loss of generality on the \( p \)th regression coefficient, may be reformulated in terms of the cumulative regression function \( B_p(t) = \int_0^t \beta_p(s)ds \) as
\[ H_0: B_p(t) = \beta_p \cdot t. \]

Martinussen and Scheike (2006) studied the test process
\[ V_n(t) = n^{1/2}(\hat{B}_p(t) - \hat{B}_p(\tau)\frac{t}{\tau}) \quad (3.5) \]
that is easy to compute and considered test statistics as 
\[ \sup_{t \leq \tau} |V_n(t)|. \]
Under \( H_0 \), we have
\[ V_n(t) = n^{1/2} \left\{ (\hat{B}_p(t) - B_p(t)) - (\hat{B}_p(\tau) - B_p(\tau)) \frac{t}{\tau} \right\}. \]
Clearly, the limit distribution of \( V_n(t) \) is not a martingale due to the term \( \hat{B}_p(\tau) \), but one may use the above resampling technique of Lin et al. (1993) to approximate its limit distribution. Specifically, the limit distribution of \( V_n(t) \) may be approximated by
\[ \hat{V}_n(t) = n^{-1/2} \sum_{i=1}^{n} \left[ \{ \hat{\epsilon}_i^B(t) \}_p - \{ \hat{\epsilon}_i^B(\tau) \}_p \right] G_i \]
where \( v_k \) is the \( k \)th element of a given vector \( v \) and where we are fixing the data. The above limit distributions and the conditional multiplier approach may also be used to test other hypothesis such as overall effect of a given covariate. Such inferential procedures and estimation are implemented in the contributed R package `timereg`; for examples of its use, see Martinussen and Scheike (2006).

### 3.1.2 Goodness-of-fit procedures

The additive hazard model is very versatile but it is still of importance to check whether the model provides an adequate fit to the data at hand. The starting point will then usually be model (3.1) and if this model is found to give an adequate fit then one may proceed to investigate the time-dynamics of the covariate effects. The goodness-of-fit procedures presented below are based on the martingale residuals
\[ M(t) = N(t) - \int_0^t X(s)d\hat{B}(s) = \int_0^t H(s)dM(s). \]
Under the model, the martingale residuals are thus themselves martingales. To validate the model fit, one possibility is to sum these residuals depending on the level of the covariates (Aalen, 1993; Grønnesby and Borgan, 1996). A \( K \)-cumulative residual process may then be defined as
\[ M_K(t) = \int_0^t K^T(s)d\hat{M}(s) = \int_0^t K^T(s)H(s)dM(s), \]
where \( K(t) = \{ K_1^T(t), \ldots, K_m^T(t) \}^T \) is an \( n \times m \) matrix, possibly depending on time. A typical choice of \( K \) is to let it be a factor with levels defined by the quartiles of the considered covariates. Based on a iid representation of \( M_K(t) \) one may calculate \( p \)-values using for instance a supremum-test, see Martinussen and Scheike (2006). With a continuous covariate one may also target a test to investigate whether the specific covariate is included in the model on the right scale, a typical alternative is to use a log-transformed version of the covariate. This construction is similar to what Lin et al. (1993) proposed under the Cox model and is based on
\[ M_c(z) = \sum_{i=1}^{n} I(X_{i1} \leq z)\hat{M}_i(\tau) \]
here assuming that the covariate under investigation is \( X_1 \). This process in \( z \) has a decomposition similarly to \( M_K \) developed above, and resampling may thus be used to validate the fit of the model (Martinussen and Scheike, 2006). These goodness-of-fit procedures are implemented in the R-package `timereg`. For other goodness-of-fit testing concerning the additive hazard model, see Gandy and Jensen (2005a) and Gandy and Jensen (2005b).
3.1.3 Further results on additive hazard models

The additive hazard is applied in some more specialized areas of survival analysis. Below we describe some of these. But first we present some structural properties of the model that underline that this model may indeed provide an appealing alternative to the proportional hazard model.

3.1.3.1 Structural properties of the additive hazard model

Consider the additive hazard model

$$\alpha(t|X, Z) = \beta_0(t) + \beta_X(t)X + \beta_Z(t)Z$$

(3.6)

and let $B_X(t) = \int_0^t \beta_X(s)\,ds$ and similarly with $B_Z(t)$. We will explore when the additive hazard structure is preserved when only conditioning on $X$ which was first investigated in Aalen (1993). The model above is formulated with $X$ and $Z$ both being scalar valued but the following can easily be generalized to cover the vector situation also. The hazard function, when only conditioning on $X$, is

$$\alpha(t|X) = \beta_0(t) + \beta_X(t)X + \beta_Z(t)Z + \psi_{Z|X}(B_Z(t)),$$

where $\psi_{Z|X}(\cdot)$ denotes minus the derivative of the log Laplace transform of $Z$ given $X$. It is therefore readily seen that additive structure is preserved if $X$ and $Z$ are independent as only the intercept term is changed. This also holds true if $(X, Z)$ has a normal distribution also allowing for dependence between the two. This is in contrast to the proportional hazard model where the proportional structure is not preserved even when $X$ and $Z$ are independent Struthers and Kalbfleisch (1986).

When dealing with data from a non-randomized study it may be of interest to calculate the causal hazard function as it corresponds to the hazard function under a randomized study. To do so we briefly introduce Pearl’s (Pearl, 2000) do operator, $do(X = x)$ (or $\hat{X} = x$), which is an intervention setting variable $X$ to $x$. Therefore, the conditional distribution of $T$ given $X$ is set to $x$, $P(T|\hat{X} = x)$, is different from $P(T|X = x)$ where the latter is the conditional distribution of $T$ given that we find $X$ equal to $x$. Taking again model (3.6) as the starting point it may be seen, using the so-called “G-computation formula” (Robins, 1986; Pearl, 2000), that

$$\alpha(t|\hat{X} = x) = \tilde{\beta}_0(t) + \beta_X(t)x,$$

(3.7)

$$\tilde{\beta}_0(t) = \beta_0(t) + \beta_Z(t)\psi_Z(B_Z(t)),$$

where $\psi_Z(\cdot)$ denotes minus the derivative of the log Laplace transform of $Z$. It is seen from display (3.7) that the additive hazard structure is again preserved and therefore the causal effect of $X$ can be estimated from model (3.6) formulated for the observed data. Such a property does not hold true for the proportional hazard model; see Martinussen and Vansteelandt (2012) for more details on these issues. The additive hazard model also forms the basis for investigation of so-called “direct effects” in Fosen et al. (2006); Martinussen (2010); Lange and Hansen (2011); and Martinussen, Vansteelandt, Gerster and Hjelmborg (2011). In a situation as depicted in Figure 3.1, the direct effect of $X$, formulated via the hazard function on absolute scale, is

$$\alpha(t|\hat{X} = x + 1, \hat{K}) - \alpha(t|\hat{X} = x, \hat{K})$$
3.1.3.2 Clustered survival data and additive hazard model

Clustered failure time data arise in many research areas, for example, in studies where time to disease onset is recorded for members of a sample of families. One approach to analyze such data is to use a frailty model, and recently Martinussen, Scheike and Zucker (2011) proposed the use of the additive hazard model in such a setting. The Aalen additive gamma frailty hazard model specifies that individual \( i \) in the \( k \)th cluster has conditional hazard function

\[
\alpha_{ik}(t|Z_k, X_{ik}) = Z_k X_{ik}^T \beta(t),
\]

(3.8)

where \( Z_k \) is the unobserved frailty variable, taken to be gamma distributed with mean 1 and variance \( \theta \), \( X_{ik} \) is the covariate vector and \( \beta(t) \) is the vector of time-dependent regression coefficients. The counting process associated with individual \( i \) in the \( k \)th cluster has the following intensity with respect to the marginal history:

\[
\lambda_{ik}^m(t) = \{1 + \theta X_{ik}^T B(t-)^{-1}\} Y_{ik}(t) X_{ik}^T \beta(t),
\]

(3.9)

where \( Y_{ik}(t) \) is the at-risk indicator. Thus the marginal intensity still has an additive structure, but with the design depending on \( B(t-) \). Based on this, an estimator was developed that has a recursive structure, and large sample properties were presented; see Martinussen, Scheike and Zucker (2011) for more details. Another approach was taken by Cai and Zeng (2011) where the frailty effect also enters at the additive scale

\[
\alpha_{ik}(t|Z_k, X_{ik}) = \beta_0(t) + X_{ik}^T \beta + Z_k,
\]

(3.10)

using the Lin and Ying version of the additive model and where \( Z_k \) follows a one-parameter distribution with zero-mean. Estimators and large sample are provided in Cai and Zeng (2011).

As alternative to the frailty approach one may instead use a marginal regression model approach if interest centers on estimating regression effects at the population level and estimation of correlation is of less interest. For use of the fully nonparametric version of the additive hazards model in this context, see Martinussen and Scheike (2006). Instead one may also apply the Lin and Ying version of the additive model; see Pipper and Martinussen (2004) and Yin and Cai (2004).

3.1.3.3 Additive hazard change point model

As described in Section 3.1 the Aalen additive hazard model is well suited to investigate the time dynamics of the effect of a given covariate. Hence one can test the constant effects model against a model where the structure of the effect is left fully unspecified. An appealing model in between the constant effects model and the full Aalen model is the change-point
model in which the effect of a covariate is constant up to an unknown point in time and changes thereafter to a new value. This model can be written as

$$\alpha(t|X, Z) = X'\beta(t) + Z\{\gamma_1 + \gamma_2 I(t > \theta)\},$$

(3.11)

where $\alpha(t|X, Z)$ is the hazard rate, $X$ is a $p$-dimensional covariate and $Z$ is a scalar covariate, $\beta(t)$ is a vector of unknown locally integrable time-dependent regression functions, and $\gamma_1$, $\gamma_2$ and $\theta$ are unknown parameters with the last being the change-point parameter that is also taken as an unknown parameter. To test whether or not there is evidence of a change-point effect corresponds to testing the hypothesis $H_0 : \gamma_2 = 0$, which is a nontrivial problem as the change-point parameter $\theta$ is only present under the alternative. Martinussen and Scheike (2007) derived a test tailored to investigate this hypothesis, and also gave estimators for all the unknown parameters defining the model. They also proposed a test that can be used to compare the change point model to full Aalen model with no prespecified form of the covariate effects.

### 3.1.3.4 Additive hazard and high-dimensional regressors

It has become of increasing interest to develop statistical methods to explore a potential relationship between a high-dimensional regressor, such as gene-expression data, to the timing of an event such as death. It has been demonstrated that the additive hazard model may also be useful in this respect. The starting point for these developments has so far been the additive hazard model with all effects being constant with time, that is, the Lin and Ying model,

$$\alpha(t|Z) = \beta(t) + Z^T \gamma,$$

where $Z$ is $p$-dimensional covariate vector; in this case with $p$ large. With gene expression data, $p$ may be very large and often much larger than the sample size $n$. This is referred to as the $p >> n$ situation. Based on the Lin and Ying model Martinussen and Scheike (2009a) derived a partial least squares estimator via the so-called “Krylov sequence,” and Ma et al. (2006) derived a principal components analysis in this context. Other popular approaches in a high-dimensional regressor setting is the Lasso and Ridge regression to mention some; see Bovelstad et al. (2009) for a comparison of methods using the Cox model as the base model. One complication with the additive hazard model compared with the proportional hazard model is, however, that there is no simple likelihood to work with for the additive model. Leng and Ma (2007) suggested to use a least squares criterion, similarly to the partial likelihood for Cox’s regression model, for the Lin and Ying model:

$$L(\beta) = \beta^T \left\{ \int Z^T(t)H(t)H(t)dN(t) \right\} \beta - 2\beta^T \left\{ \int Z^T(t)H(t)dN(t) \right\}.$$  

(3.12)

using the notation introduced in Section 3.1. This criterion was also suggested independently by Martinussen and Scheike (2009b) that provided further motivation. This criterion opens the route to many popular estimation procedures applied in the high-dimensional regressor setting when the underlying model is the Lin and Ying model. Recently Gorst-Rasmussen and Scheike (2012) took the same model to propose a sure independence screening method that shows good performance. For estimation in the high-dimensional case one may use the contributed R package ahaz that can also do other types of estimation and inference within the additive hazard model context.

### 3.1.3.5 Combining the Cox model and the additive model

The Cox model and the Aalen model may be combined in various ways to improve model fit or to enhance interpretation of parameters. Scheike and Zhang (2002) considered the model

$$\alpha(t|X, Z) = \{X^T \beta(t)\} \exp\{Z^T \gamma\},$$

where $\alpha(t|X, Z)$ is the hazard rate, $X$ is a $p$-dimensional covariate and $Z$ is a scalar covariate, $\beta(t)$ is a vector of unknown locally integrable time-dependent regression functions, and $\gamma$ is an unknown parameter.
Alternatives to the Cox Model

3.1.3.6 Gastrointestinal tumour data

A typical instance where the Cox model provides a poor fit is when crossing survival curves is observed as shown in Figure 3.2 (left display) that displays the Kaplan-Meier estimates of survival probabilities of the chemotherapy and chemotherapy plus radiotherapy groups of gastric cancer patients in a randomized clinical trial (Stablein and Koutrouvelis, 1985); these data were also considered in Zeng and Lin (2007). The crossing of the survival curves can be explained by the patients receiving the more aggressive treatment (radiotherapy) who are at elevated risk of death initially but may benefit from the treatment in the long run if they are able to tolerate the treatment initially. That the Cox model provides a poor fit to the data is confirmed by the supremum test based on the score process Lin et al. (1993) with $p < 0.001$. We fitted Aalen's additive hazards model to the gastrointestinal tumour data:

$$\alpha(t|X) = \beta_0(t) + \beta(t)X,$$

where $X$ is the indicator of chemotherapy plus radiotherapy treatment. The cumulative regression coefficient corresponding to the combined therapy group is depicted in Figure 3.2 (right display), showing nicely that a time varying effect is indeed the case for these
data with a negative effect of the combined treatment in the first 300 days or so, and then apparently with an adverse effect thereafter. For these data one may be interested in applying the change-point model

$$\alpha(t|X) = \beta_0(t) + X\{\gamma_1 + \gamma_2 I(t > \theta)\},$$

with $\theta$ being the change-point parameter. For the gastrointestinal tumour data we obtain the estimates $\hat{\theta} = 315$ and $(\hat{\gamma}_1, \hat{\gamma}_2) = (0.00175, -0.00255)$ with the estimated cumulative regression function superimposed on the Aalen estimator in Figure 3.2 indicating that this model gives a good fit to these data.

### 3.2 The accelerated failure time model

Another useful model in survival analysis is the accelerated failure time (AFT) model that can be formulated using survival functions. Let $T$ be a survival time and $Z$ a covariate vector that does not depend on time and let $S(t|Z)$ be the conditional survival function. The AFT model assumes that

$$S(t|Z) = S_0(te^{Z^T\beta}), \quad (3.13)$$

where $S_0$ is a baseline survival function. From (3.13) it is seen that covariates act multiplicatively on time so that their effect is to accelerate (or decelerate) time-to-failure relative to $S_0$. An equivalent and popular formulation of AFT-model is the following linear regression for the log-transformed event time, $\log(T)$, given $Z$ such that

$$\log(T) = -Z^T\beta + \varepsilon, \quad (3.14)$$

where the distribution of $e^\varepsilon$ is given by $S_0$ and $\varepsilon$ is assumed to be independent of $Z$. The AFT model has received considerable attention in the statistical literature in the last two decades or so. Although, in its semiparametric version (see below) it is computationally intensive, it is now considered as an alternative to the Cox model. It is appealing due to its direct relationship between the failure time and covariates, which was also noted by David Cox in Reid (1994). One may extend the model to cover the case where covariates may change with time by considering

$$e^\varepsilon = \int_0^T e^{Z^T(t)\beta} dt$$

see for example Zeng and Lin (2007a), but we will only consider the simpler case here. The AFT model can also be formulated in terms of the hazard function for $T$ given $Z$:

$$\lambda(t|Z) = \lambda_0(te^{Z^T\beta})e^{Z^T\beta}, \quad (3.15)$$

where $\lambda_0(t)$ is the hazard associated with the distribution of $e^\varepsilon$.

#### 3.2.1 Parametric models

If we are willing to specify a specific form of $S_0$ then the AFT model is fully parametric. For instance, the Weibull model is given by

$$S_0(t) = e^{-kt^\alpha}, \quad k, \alpha > 0.$$
In such a case, estimation and statistical inference is straightforward using standard likelihood methods for survival data. For the Weibull model we have the hazard function \( \lambda_0(t) = k \alpha t^{\alpha-1} \) and hence, by (3.15),

\[
\lambda(t|Z) = k \alpha (te^{Z^T \beta})^{\alpha-1} e^{Z^T \beta} = \lambda_0(t) e^{Z^T \beta}
\]

being also a proportional hazard function with a Weibull baseline hazard function, so in this specific example the two models coincide.

### 3.2.2 Semiparametric models

We will now look into the case where \( S_0(t) \) is left unspecified, or, equivalently, \( \varepsilon \) in (3.14) is not specified, which is referred to as the “semiparametric AFT model.” There exist two approaches to tackle this situation. One develops estimation based on the hazard specification (3.15), while the other one uses the additive mean model specification (3.14) as a starting point.

#### 3.2.2.1 Inference based on hazard specification

Let \( \hat{C} \) be the censoring time for \( T \), and put \( \hat{T} = T \land C \) and \( \Delta = I(T \leq C) \). Hence the intensity of the counting process \( N(t) = I(\hat{T} \leq t) \Delta \) is thus given by \( Y(t) \lambda(t) \) with \( Y(t) = I(t \leq \hat{T}) \) being the at-risk indicator and \( \lambda(t) \) given in (3.15). Assume that \( n \) i.i.d. counting processes are being observed subject to this generic hazard model. We thus consider \( N(t) = \{N_1(t), \ldots, N_n(t)\} \) the \( n \)-dimensional counting process of all subjects. Define also the time-transformed counting process

\[
N^*(t) = (N_1(te^{-Z^T_1 \beta}), \ldots, N_n(te^{-Z^T_n \beta}))
\]

with associated at-risk process \( Y^*_i(t, \beta) = Y_i(te^{-Z^T_i \beta}), \ i = 1, \ldots, n \). This time-transformation is made because then the intensity of \( N_i(te^{-Z^T_i \beta}) \) is

\[
\lambda_i^*(t) = Y_i^*(t) \lambda_0(t),
\]

which immediately suggests that \( \Lambda_0(t) = \int_0^t \lambda_0(s) ds \) should be estimated by the Breslow-type estimator

\[
\hat{\Lambda}_0(t, \beta) = \int_0^t \frac{1}{S_0^*(s, \beta)} dN_i^*(s)
\]

(3.16)

where \( dN_i^*(t) = \sum_{i=1}^n dN_i^*(t) \), and

\[
S_0^*(t, \beta) = \sum_{i=1}^n Y_i^*(t, \beta)
\]

if \( \beta \) were known. The efficient score function for \( \beta \) is

\[
\sum_{i=1}^n \int_0^\infty \frac{\partial}{\partial \beta} (\lambda_i(t, \beta)) \lambda_i^{-1}(t, \beta) (dN_i(t) - Y_i(t) \lambda_i(t) dt)
\]

(3.17)

\[
= \sum_{i=1}^n \int_0^\infty \left( \frac{\lambda'_0(te^{Z^T_i \beta})te^{Z^T_i \beta}}{\lambda_0(te^{Z^T_i \beta})} + 1 \right) Z_i(dN_i(t) - Y_i(t) \lambda_i(t) dt)
\]

\[
= \sum_{i=1}^n \int_0^\infty \left( \frac{\lambda_0(u)u}{\lambda_0(u)} + 1 \right) Z_i(dN_i^*(u) - Y_i^*(u, \beta) d\Lambda_0(u)),
\]
and inserting \( d\hat{\Lambda}_0(u, \beta) \) for \( d\Lambda_0(u) \) gives

\[
U_W(\beta) = \sum_{i=1}^{n} \int_{0}^{\infty} W(u)Z_i \left( dN^*_i(u) - \frac{Y^*_i(u, \beta)}{S_0^*(u, \beta)} dN^*_i(u) \right)
\]

\[
= \sum_{i=1}^{n} \int_{0}^{\infty} W(u) \left( Z_i - E^*(u, \beta) \right) dN^*_i(u),
\]

(3.18)

where

\[
S^*_1(u, \beta) = \sum_{i=1}^{n} Y^*_i(u, \beta) Z_i, \quad E^*(u, \beta) = \frac{S^*_1(u, \beta)}{S_0^*(u, \beta)},
\]

and

\[
W(u) = \left( \frac{\lambda_0(u)u}{\lambda_0(u)} + 1 \right)
\]

(3.19)

is the efficient weight function. We cannot use (3.18) directly for estimation purposes since the weight function \( W(u) \) involves the unknown baseline hazard function \( \lambda_0(u) \) and its derivative \( \lambda'_0(u) \). These can be estimated and inserted into (3.18) but it is not recommendable since it is hard to get reliable estimates of especially \( \lambda'_0(t) \). A way around this is to take (3.18) and replace the weight function \( W(u) \) with one that can be computed as for example \( W(u) = 1 \) or \( W(u) = n^{-1}S_0^*(u, \beta) \) referred to as the “log-rank and Gehan weight functions,” respectively. A practical complication is that the score function \( U_W(\beta) \) is a step function of \( \beta \) so \( U_W(\beta) = 0 \) may not have a solution. The score function may furthermore not be component-wise monotone in \( \beta \). It is actually monotone in each component of \( \beta \) if the Gehan weight is chosen (Fygenson and Ritov, 1994). The estimator \( \hat{\beta} \) is usually chosen as the one which minimizes \( \| U_W(\beta) \| \). The contributed R package lss calculates the Gehan rank estimator. It has been established, under regularity conditions, that \( n^{1/2}(\hat{\beta} - \beta) \) is asymptotically zero-mean normal with covariance matrix \( A_W^{-1}B_WA_W^{-1} \), where \( A_W \) and \( B_W \) are the limits in probability of

\[
\frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} W(u) (Z_i - E^*(u, \beta)) \otimes (\frac{\lambda'_0(u)u}{\lambda_0(u)} + 1) dN^*_i(u),
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} W(u)^2 (Z_i - E^*(u, \beta))^2 dN^*_i(u),
\]

respectively; see Tsiatis (1990) and Ying (1993). It is seen that \( A_W \) and \( B_W \) coincide in the case where \( W(u) \) is taken as the efficient weight function (3.19). The asymptotic covariance matrix depends on \( \lambda'_0 \), which is difficult to estimate. One may, however, apply a resampling technique avoiding estimation of \( \lambda'_0 \); see Lin et al. (1998) and Jin et al. (2003).

The score function can also be written on the (log)-transformed time scale

\[
U_W(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \tilde{W}(t) \left( Z_i - E^*(e^t, \beta) \right) dN^*_i(e^t)
\]

\[
= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \tilde{W}(t) \left( Z_i - \tilde{E}(t - Z_i^T \beta, \beta) \right) d\tilde{N}_i(t - Z_i^T \beta),
\]

(3.20)

where \( \tilde{N}_i(t) = I(\log(\tilde{T}_i) \leq t)\Delta_i, \tilde{Y}_i(t) = I(t \leq \log(\tilde{T}_i)) \).

\[
\tilde{E}(t, \beta) = \sum_{i=1}^{n} Z_i \tilde{Y}_i(t) / \sum_{i=1}^{n} \tilde{Y}_i(t), \quad \tilde{W}(t) = \frac{\lambda'_0(t)}{\lambda_0(t)}
\]
Alternatives to the Cox Model

with \( \lambda_0(t) \) the hazard function for \( \varepsilon \). Another way of motivating (3.20) is by classical non-parametric testing theory; see Tsiatis (1990) and Kalbfleisch and Prentice (2002) for much more details on that approach.

An interesting variant of (3.15), considered by Chen and Jewell (2001), is

\[
\lambda(t) = \lambda_0(t e^{Z^T \beta_1}) e^{Z^T \beta_2},
\]

(3.21)

which contains the proportional hazard model (\( \beta_1 = 0 \)) and the accelerated failure time model (\( \beta_2 = 0 \)), and for \( \beta_2 = 0 \) what is called “the accelerated hazard model” (Chen and Wang, 2000). Chen and Jewell (2001) suggested estimating equations for estimation of \( \beta = (\beta_1, \beta_2) \) and showed for the resulting estimator, \( \hat{\beta} \), that \( n^{1/2}(\hat{\beta} - \beta) \) is asymptotically zero-mean normal with a covariance that also involves the unknown baseline hazard (and its derivative). They also suggested an alternative resampling approach for estimating the covariance matrix without having to estimate the baseline hazard function or its derivative.

With these tools at hand, one may then investigate whether it is appropriate to simplify (3.21) to either the Cox model or the AFT model.

3.2.2.2 Inference using the additive mean specification

There exist other ways of estimating the regression parameters \( \beta \) of the semiparametric AFT-model, which build more on classical linear regression models estimation. Starting with (3.14), let \( V = \log(T) \), \( U = \log(C) \) and \( \tilde{V} = V \wedge U \), and write model (3.14) as

\[
V = -\beta_0 - Z^T \beta + \varepsilon
\]

assuming that \( \varepsilon \) is independent of \( Z = (Z_1, \ldots, Z_p)^T \) and has zero mean. If \( V \) was not right-censored, then it is of course an easy task to estimate the regression parameters using least squares regression. A number of authors (Miller, 1976; Buckley and James, 1979; and Koul et al., 1981) have extended the least-squares principle to accommodate censoring; we describe the approach of Buckley and James. The idea is to replace \( V \) with a quantity that has the same mean as \( V \), and which can be computed based on the right-censored sample. With

\[
V^* = V \Delta + (1 - \Delta) E(V | V > U, Z),
\]

and \( \Delta = I(V \leq U) \), then \( E(V^* | Z) = E(V | Z) \). Still, \( V^* \) is not observable but it can be estimated as follows. Since

\[
E(V | V > U, Z) = -Z^T \beta + \frac{\int_{U+Z^T \beta} F(v) \, dv}{1 - F(U + Z^T \beta)}
\]

with \( F \) the distribution of \( V + Z^T \beta \), one can construct the so-called “synthetic data points:

\[
\hat{V}_i^*(\beta) = V_i \Delta_i + (1 - \Delta_i) \left( -Z_i^T \beta + \frac{\int_{U_i+Z_i^T \beta} \hat{F}(v) \, dv}{1 - \hat{F}(U_i + Z_i^T \beta)} \right),
\]

where \( \hat{F} \) is the Kaplan-Meier estimator based on \( (\hat{V}_i + Z_i^T \beta, \Delta_i), \ i = 1, \ldots, n \). One may then estimate the parameters from the normal equations leading to the following estimating equation for the regression parameter vector \( \beta \):

\[
U(\beta) = \sum_{i=1}^n (\hat{V}_i^*(\beta) + Z_i^T \beta)(Z_i - \bar{Z}) = 0,
\]

(3.22)
where $\bar{Z} = n^{-1} \sum_i Z_i$. Equation (3.22) needs to be solved iteratively if it has a solution. This estimator is referred to as the “Buckley-James estimator.” The large sample properties of the resulting estimator were studied by Ritov (1990). Equation (3.22) can also be written as, with $S(v) = v$,

$$
\sum_{i=1}^{n} \left( \Delta_i S(U_i + Z_i^T \beta) + \left(1 - \Delta_i \right) \frac{\int_{U_i + Z_i^T \beta}^{\infty} S(v) d\hat{F}(v)}{1 - \hat{F}(U_i + Z_i^T \beta)} \right) (Z_i - \bar{Z}) = 0, \tag{3.23}
$$

that may be derived from a likelihood principle; the efficient choice of $S(v)$ being $S(v) = f'(v)/f(v)$ with $f(v) = F'(v)$ the density function. Ritov (1990) also established an asymptotic equivalence between the two classes of estimators given by (3.20) and (3.23).

The estimating function $U(\beta)$ in (3.22) is neither continuous nor monotone in $\beta$, which makes it difficult to calculate the estimator in practice. Jin et al. (2006) proposed an iterative solution to (3.22) with a preliminary consistent estimator as starting value; this estimator is also available in the R package lss.

### 3.3 Quantile regression for survival analysis

#### 3.3.1 Introduction

In survival analysis, quantile regression (Koenker and Bassett, 1978) offers a significant extension of the accelerated failure time (AFT) model. Let $T$ denote the failure time of interest and $Z \equiv (Z_1, \ldots, Z_p)^T$ denote a $p \times 1$ vector of covariates. Define $\bar{Z} = (1, Z^T)^T$ and $V = \log(T)$. For the response $V$, the $\tau$-th conditional quantiles of $V$ given $\bar{Z}$ is defined as $Q_V(\tau|\bar{Z}) \equiv \inf \{ t : \Pr(V \leq t|\bar{Z}) \geq \tau \}$, where $\tau \in (0,1)$. We adopt the same quantile definition for other response random variables. With survival data, a common quantile regression modeling strategy is to link the conditional quantile $Q_V(\tau|\bar{Z})$ to $\bar{Z}$ through a linear model:

$$
Q_V(\tau|\bar{Z}) = \bar{Z}^T \beta_0(\tau), \quad \tau \in [\tau_L, \tau_U], \tag{3.24}
$$

where $\beta_0(\tau)$ is a $(p + 1) \times 1$ vector of unknown regression coefficients possibly depending on $\tau$, and $0 \leq \tau_L \leq \tau_U \leq 1$. Like the AFT model, the quantile regression model (3.24) offers an easy coefficient interpretation as a covariate effect on event time. Specifically, a non-intercept coefficient in $\beta_0(\tau)$ represents the change in the $\tau$th quantile of $\log(T)$ given one unit change in the corresponding covariate.

It is important to note that the AFT model,

$$
\log(T) = Z^T b + \varepsilon, \quad \varepsilon \perp \perp Z, \tag{3.25}
$$

is a special case of model (3.24) with $\beta_0(\tau) = (Q_\tau(\tau), b^T)^T$. Here $\perp \perp$ stands for independence, and $Q_\tau(\tau)$ represents the $\tau$th quantile of $\varepsilon$. From this view, we see that the AFT model (3.25) requires the effects of covariates on $Q_V(\tau|\bar{Z})$ be constant for all $\tau \in [\tau_L, \tau_U]$ (i.e., location shift effects). In contrast, the quantile regression model (3.24) flexibly formulates coefficients as functions of $\tau$, thereby permitting covariate effects to vary across different quantile levels. As a result, it can accommodate a more general relationship between $T$ and $Z$ compared to the AFT model (3.25).

Note, when $\tau_L = \tau_U$, model (3.24) would only assert “local” linearity between the quantile of $\log(T)$ and $\bar{Z}$ at a single quantile level, and therefore imposes weaker restrictions than a “global” quantile regression model corresponding to a case with $\tau_L < \tau_U$. A practical
advantage of conducting “global” quantile regression is the capability of investigating the dynamic pattern of covariate effects.

In this section, we shall primarily focus on quantile regression with randomly right-censored data. Let $C$ denote time to censoring. Define $\hat{T} = T \wedge C$ and $\Delta = I(T \leq C)$. The observed data under right censoring consists of $n$ i.i.d. replicates of $(\hat{T}, \Delta, \tilde{Z})$, denoted by $(\hat{T}_i, \Delta_i, \tilde{Z}_i)$, $i = 1, \ldots, n$. In addition, we define $\hat{V} = \log(\hat{T})$, $\tilde{V}_i = \log(\hat{T}_i)$, $U = \log(C)$, and $U_i = \log(C_i)$. In Sections 3.3.2, 3.3.3 and 3.3.4, we consider three different random censoring scenarios, with general inference procedures discussed in Section 3.3.5. Extensions to more complex survival settings are briefly described in Section 3.3.6.

### 3.3.2 Estimation under random right censoring with $C$ always known

Early work on quantile regression with censored data by Powell (1984, 1986) was targeted at type I censoring cases where the censoring time $C$ is fixed and prespecified. Motivated by the fact that $Q_V(\tau|\tilde{Z}) = \{\tilde{Z}^T\beta_0(\tau)\} \wedge U$, an estimator of $\beta_0(\tau)$ is defined as the minimizer of

$$r(b, \tau) = \sum_{i=1}^{n} \rho_\tau(\tilde{V}_i - (\tilde{Z}_i^T b) \wedge U_i)$$

with respect to $b$, where $\rho_\tau(x) = x(\tau - I(x < 0))$. This estimation method can also be applied to a more general case where $C$ is independent of $T$ given $\tilde{Z}$, and is always known but not necessarily fixed. In the absence of right censoring (i.e., $C_i = \infty$), $r(b, \tau)$ reduces to the check function of Koenker and Bassett (1978), the objective function for defining sample regression quantiles with complete data without censoring. Note that $r(b, \tau)$ is not convex in $b$ and thus it may have multiple local minima. Further efforts have been made to improve the numerical performance of this approach by several authors, for example: Fitzenberger (1997); Buchinsky and Hahn (1998); and Chernozhukov and Hong (2001). An implementation of Powell’s method is available in the crq function in the contributed R package quantreg by Koenker (2012).

### 3.3.3 Estimation under covariate-independent random right censoring

Under the assumption that $T$ and $C$ are independent and $C$ is independent of $\tilde{Z}$ (i.e., covariate-independent random censoring), a natural estimating equation for $\beta_0(\tau)$, derived from Ying et al. (1995)’s work, is given by

$$n^{-1/2} \sum_{i=1}^{n} \tilde{Z}_i \left[ I\{\hat{V}_i - \tilde{Z}_i^T \beta(\tau) > 0\} \right] = 0,$$

where $\hat{G}(\cdot)$ is the Kaplan-Meier estimate for $G(\cdot)$, the survival function of $\hat{C}$. The estimating function, the left-hand side (LHS) of (3.26), is not continuous in $\beta(\tau)$, and thus an exact zero-crossing may not exist. The solution to Equation (3.26) may be alternatively defined as a minimizer of the $L_2$ norm of the estimating function. Such an objective function is discontinuous and may have multiple minima. To solve Equation (3.26), as suggested by Ying et al. (1995), the grid search method may be used for cases with low-dimensional $\tilde{Z}$ and simulated annealing algorithm (Lin and Geyer, 1992) may be used for high-dimensional cases.

For the same right censoring scenario, an alternative estimating equation for $\beta_0(\tau)$ is suggested by the work of Peng and Fine (2009). Specifically, the inverse probability of censoring weighting (IPCW) technique (Robins and Rotnitzky, 1992) can be used to
estimate $\beta_0(\tau)$ based on the fact that

$$E \left\{ \frac{I(\tilde{V} \leq t, \Delta = 1)}{G(\tilde{V})} \right\} = \Pr(\tilde{V} \leq t | \tilde{Z}).$$

This leads to the following estimating equation for $\beta_0(\tau)$:

$$n^{-1/2} \sum_{i=1}^{n} \tilde{Z}_i \left[ \frac{I(\tilde{V}_i \leq \tilde{Z}_i^T \beta(\tau), \Delta_i = 1)}{G(\tilde{V}_i)} - \tau \right] = 0. \quad (3.27)$$

Note, the estimating function in (3.27) is not continuous but monotone (Fygenson and Ritov, 1994). By the monotonicity of (3.27), the solution to Equation (3.27) can be reformulated as the minimizer of the $L_1$-type convex function (of $b$),

$$\sum_{i=1}^{n} \left\{ I(\Delta_i = 1) \left| \frac{\tilde{V}_i}{G(\tilde{V}_i)} - b^T \frac{\tilde{Z}_i}{G(\tilde{V}_i)} \right| \right\} + M - b^T \sum_{i=1}^{n} \left( -\frac{\tilde{Z}_i I(\Delta_i = 1)}{G(\tilde{V}_i)} + 2\tilde{Z}_i \tau \right),$$

where $M$ is an extremely large positive number selected to bound $b^T \sum_{i=1}^{n} \left( -\frac{\tilde{Z}_i I(\Delta_i = 1)}{G(\tilde{V}_i)} + 2\tilde{Z}_i \tau \right)$ from the above for all $b$'s in the compact parameter space for $\beta_0(\tau)$. This minimization problem can be readily solved by the $l1fit$ function in S-PLUS or the $rq$() function in the contributed R package quantreg (Koenker, 2012).

### 3.3.4 Estimation under standard random right censoring

Standard random right censoring refers to a censoring mechanism where $C$ is only assumed to be independent of $T$ given $\tilde{Z}$. It is less restrictive than the censorship considered in Sections 3.3.2 and 3.3.3.

Two major types of approaches have been developed for quantile regression with survival data subject to standard random right censoring; one type employs the principle of self-consistency (Efron, 1967) and the other type utilizes the martingale structure associated with randomly right-censored data. Hereafter, we shall refer to them as a self-consistent approach and martingale-based approach, respectively. Here we focus on approaches that are oriented to the following global linear quantile regression model,

$$Q_V(\tau | \tilde{Z}) = \tilde{Z}^T \beta_0(\tau), \quad \tau \in [0, \tau_U], \quad (3.28)$$

which is a special case of model (3.24) with $\tau_L = 0$. Of note, model (3.28) entails $\exp \{ \tilde{Z}^T \beta_0(0) \} = 0$, which is useful boundary information to be employed in the estimation of $\beta_0(\tau)$ with $\tau > 0$. Some efforts, for example, by Wang and Wang (2009), have been made to address a local linear quantile regression model (i.e., $\tau_L = \tau_U$) by incorporating nonparametric estimation of the distribution function of $C$ given $\tilde{Z}$.

#### 3.3.4.1 Self-consistent approach

Portnoy (2003) made the first attempt to tackle the estimation of model (3.28) under the standard random right censoring assumption by employing the principle of self-consistency (Efron, 1967). The critical idea behind his algorithm is about splitting a censored event time, $\tilde{V}_i$ with $\Delta_i = 0$, between $U_i$ and $\infty$ with appropriately designed weights, sharing the same spirit as that adopted for Efron’s self-consistent estimator of survival function (Efron, 1967). The initial iterative self-consistent algorithm (Portnoy, 2003) was simplified into a grid-based sequential estimation procedure (Neocleous et al., 2006), which is implemented by
the \textit{crq} function in the contributed R package \textit{quantreg} (Koenker, 2012). The corresponding asymptotic studies were established by Portnoy and Lin (2010).

The self-consistent approach can be formulated through stochastic integral equations (Peng, 2012). Define $N_i(t) = I(\check{V}_i \leq t, \Delta_i = 1)$, $R_i(t) = I(\check{V}_i \leq t, \Delta_i = 0)$, and $F_V(t) \equiv \Pr(V \leq t)$. First, consider Efron’s (Efron, 1967) self-consistent estimating equation for $F_V(t)$ in the one-sample case:

$$F_V(t) = n^{-1} \sum_{i=1}^{n} \left\{ N_i(t) + R_i(t) \frac{F_V(t) - F_V(\check{V}_i)}{1 - F_V(\check{V}_i)} \right\}.$$  \hfill (3.29)

Expressing the right-hand side (RHS) of (3.29) by a stochastic integral and further applying stochastic integral by parts, one can rewrite Equation (3.29) as

$$F_V(t) = n^{-1} \sum_{i=1}^{n} \left[ N_i(t) + R_i(t) \{1 - F_V(t)\} \int_{0}^{t} \frac{R_i(u)}{(1 - F_V(u))^2} dF_V(u) \right].$$  \hfill (3.30)

With $t$ replaced by $\check{Z}_i^T \beta(\tau)$, Equation (3.30) evolves into an estimating equation for $\beta_0(\tau)$:

$$n^{1/2} S_n^{(SC)}(\beta, \tau) = 0,$$  \hfill (3.31)

where $S_n^{(SC)}(\beta, \tau)$ equals

$$n^{-1} \sum_{i=1}^{n} \check{Z}_i \left[ N_i(\check{Z}_i^T \beta(\tau)) + R_i(\check{Z}_i^T \beta(\tau))(1 - \tau) \int_{0}^{\tau} \frac{R_i(\check{Z}_i^T \beta(u))}{(1 - u)^2} du - \tau \right].$$

Define a grid of $\tau$-values, $\mathcal{G}$, as $0 < \tau_1 < \tau_2 < \ldots < \tau_M = \tau_U$. Let $\|\mathcal{G}\|$ denote the size of $\mathcal{G}$, max$_{k=0,...,M} (\tau_{k+1} - \tau_k)$. Without further mentioning, $\mathcal{G}$ will be adopted throughout Section 3.3.4. A self-consistent estimator, $\check{\beta}_{SC}(\cdot)$, is defined as a cadlag step function that only jumps at the grid points of $\mathcal{G}$ and approximates the solution to Equation (3.31). The algorithm taken to obtain $\check{\beta}_{SC}(\cdot)$ is outlined as follows.

1. Set exp$\{\check{Z}_i \check{\beta}_{SC}(0)\} = 0$ for all $i$. Set $k = 0$.

2. Given exp$\{\check{Z}_i \check{\beta}_{SC}(\tau_l)\}$ for $l \leq k$, obtain $\check{\beta}_{SC}(\tau_{k+1})$ as the minimizer of the following weighted check function:

$$\left[ \sum_{\Delta_i = 1} \rho_{\tau}(\check{V}_i - \check{Z}_i^T b) + \sum_{\Delta_i = 0} \left\{ \hat{w}_{k+1,i} \rho_{\tau}(\check{V}_i - \check{Z}_i^T b) \right\} \right],$$  \hfill (3.32)

where $\hat{w}_{k+1,i} = \sum_{l=0}^{k} R_i(\check{Z}_i^T \check{\beta}_{SC}(\tau_l)) \left( \frac{1 - \tau_{k+1}}{1 - \tau_{l+1}} - \frac{1 - \tau_{k+1}}{1 - \tau_l} \right)$, and $Y^*$ is an extremely large value.

3. Replace $k$ by $k + 1$ and repeat step 2 until $k = M$ or only censored observations remain above exp$\{\check{Z}_i \check{\beta}_{SC}(\tau_{k-1})\}$.

Large sample studies for $\check{\beta}_{SC}(\cdot)$ are facilitated by the stochastic integral equation representation of (3.31). Specifically, under certain regularity conditions and given $\lim_{n \to \infty} \|\mathcal{G}\| = 0$, $\sup_{\tau \in [\nu, \tau_U]} \|\check{\beta}_{SC}(\tau) - \beta_0(\tau)\| \to 0$, where $0 < \nu < \tau_U$. If $n^{1/2} \lim_{n \to \infty} \|\mathcal{G}\| = 0$ is further satisfied, then $n^{1/2} \{\check{\beta}_{SC}(\tau) - \beta_0(\tau)\}$ converges weakly to a Gaussian process for $\tau \in [\nu, \tau_U]$. Peng (2012) also investigated several variants of $\check{\beta}_{SC}(\cdot)(\cdot)$, showing their asymptotic equivalence to $\check{\beta}_{SC}$ as well as their connection with the self-consistent estimator proposed by Portnoy (Portnoy, 2003; Neocleous et al., 2006).
3.3.4.2 Martingale-based approach

Model (3.28) can also be estimated based on the martingale structure of randomly right-censored data (Peng and Huang, 2008). Define \( \Lambda_V(t|\bar{Z}) = -\log\{1 - \Pr(V \leq t|\bar{Z})\} \), \( N(t) = I(\bar{V} \leq t, \Delta = 1) \), and \( M(t) = N(t) - \Lambda_V(t \wedge Y|\bar{Z}) \). Let \( N_i(t) \) and \( M_i(t) \) be sample analogs of \( N(t) \) and \( M(t) \), respectively, \( i = 1, \ldots, n \). Since \( M_i(t) \) is the martingale process associated with the counting process \( N_i(t) \), \( E\{M_i(t)|\bar{Z}_i\} = 0 \) for all \( t > 0 \). This implies

\[
E\left\{ \sum_{i=1}^{n} \bar{Z}_i \left[ N_i(\bar{Z}_i^\tau \beta_0(\tau)) - \Lambda_V(\bar{Z}_i^\tau \beta_0(\tau) \wedge \bar{V}_i|\bar{Z}_i) \right] \right\} = 0. \tag{3.33}
\]

By the monotonicity of \( \bar{Z}_i^\tau \beta_0(\tau) \) in \( \tau \in [0, \tau_U] \) under model (3.28), we have

\[
\Lambda_V(\bar{Z}_i^\tau \beta_0(\tau) \wedge \bar{V}_i|\bar{Z}_i) = \int_{0}^{\tau} I\{\bar{V}_i \geq \bar{Z}_i^\tau \beta_0(u)\}dH(u), \tag{3.34}
\]

where \( H(x) = -\log(1 - x) \). Coupling the equalities (3.33) and (3.34) suggests the estimating equation,

\[ n^{1/2} S_n^{(PH)}(\beta, \tau) = 0, \tag{3.35} \]

where

\[ S_n^{(PH)}(\beta, \tau) = n^{-1} \sum_{i=1}^{n} \bar{Z}_i \left[ N_i(\bar{Z}_i^\tau \beta(\tau)) - \int_{0}^{\tau} I\{\bar{V}_i \geq \bar{Z}_i^\tau \beta(u)\}dH(u) \right]. \]

An estimator of \( \beta_0(\tau) \), denoted by \( \hat{\beta}_{PH}(\tau) \), can be obtained through approximating the stochastic solution to Equation (3.35) by a cadlag step function. The sequential algorithm for obtaining \( \hat{\beta}_{PH}(\tau) \) is outlined as follows:

1. Set \( \exp\{\bar{Z}_i^\tau \hat{\beta}_{PH}(\tau_0)\} = 0 \) for all \( i \). Set \( k = 0 \).

2. Given \( \exp\{\bar{Z}_i^\tau \hat{\beta}_{PH}(\tau_l)\} \) for \( l \leq k \), obtain \( \hat{\beta}_{PH}(\tau_{k+1}) \) as the minimizer of the following \( L_1 \)-type convex objective function:

\[
l_{k+1}(h) = \sum_{i=1}^{n} \Delta_i \bar{V}_i - \delta_i \bar{Z}_i^\tau h \left[ Y^* - \sum_{l=1}^{n} (-\Delta_l \bar{Z}_l^\tau h) \right] \]

\[
+ \left| Y^* - \sum_{r=1}^{n} \left( 2 \bar{Z}_r^\tau h \right) I\{\bar{V}_r \geq \bar{Z}_r^\tau \hat{\beta}_{PH}(\tau_l)\} \{H(\tau_{l+1}) - H(\tau_l)\} \right|, \]

where \( Y^* \) is an extremely large value.

3. Replace \( k \) by \( k + 1 \) and repeat step 2 until \( k = M \) or no feasible solution can be found for minimizing \( l_k(h) \).

The \texttt{crq} function in the contributed R package \texttt{quantreg} (Koenker, 2012) provides an implementation of \( \hat{\beta}_{PH}(\tau) \) based on an algorithm slightly different from the one presented above. More recently, Huang (2010) derived a grid-free estimation procedure for model (3.28) by using the concept of quantile calculus.

Peng and Huang (2008) established the uniform consistency and weak convergence of \( \hat{\beta}_{PH}(\cdot) \). Moreover, \( \hat{\beta}_{PH}(\cdot) \) was shown to be asymptotically equivalent to the self-consistent estimator \( \hat{\beta}_{SC}(\cdot) \) (Peng, 2012). The numerical results reported in Koenker (2008) and Peng (2012) confirm this theoretical result and show comparable computational performance between these two approaches.
3.3.5 Variance estimation and other inference

The estimators of $\beta_0(\tau)$ discussed in Sections 3.3.2–3.3.4 generally have asymptotic variances that involve unknown density functions. Under random right censoring with known censoring time or covariate-independent censoring, we can adapt Huang (2002)’s technique to avoid density estimation. Specifically, let $\hat{\beta}(\tau)$ be general notation for an estimator of $\beta_0(\tau)$, and $S_n(\beta(\tau), \tau)$ denote the estimating function associated with $\hat{\beta}(\tau)$, for example, the LHS of (3.26) and (3.27). Asymptotic theory may show that $S_n(\beta_0(\tau), \tau)$ converges to a multivariate normal distribution with variance matrix $\Sigma(\tau)$. Suppose one can obtain a consistent estimator of $\Sigma(\tau)$, denoted by $\hat{\Sigma}(\tau)$. The following are the main steps to estimate the asymptotic variance of $\hat{\beta}(\tau)$:

1. Find a symmetric and nonsingular $(p+1) \times (p+1)$ matrix $E_n(\tau) \equiv \{e_{n,1}(\tau), \ldots, e_{n,p+1}(\tau)\}$ such that $\hat{\Sigma}(\tau) = \{E_n(\tau)\}^2$.

2. Calculate $D_n(\tau) = \left( S_n^{-1}\{e_{n,1}(\tau), \tau\} - \hat{\beta}(\tau), \ldots, S_n^{-1}\{e_{n,p+1}(\tau), \tau\} - \hat{\beta}(\tau) \right)$, where $S_n^{-1}(e, \tau)$ is defined as the solution to $S_n(b, \tau) - e = 0$.

3. Estimate the asymptotic variance matrix of $n^{1/2}\{\hat{\beta}(\tau) - \beta_0(\tau)\}$ by $n\{D_n(\tau)\} \otimes 2$

Bootstrapping procedures are also frequently used for variance estimation under quantile regression. Resampling methods that follow the idea of Parzen and Ying (1994) were shown to yield valid variance estimates and other inference (Peng and Huang, 2008). Simple bootstrapping procedures based on resampling with replacement also seem to have satisfactory performance (Portnoy, 2003; Peng, 2012).

One practical appeal of quantile regression with survival data is its capability of accommodating and exploring varying covariate effects. Second-stage inference can be employed to serve this need. Given $\hat{\beta}(\tau)$ for a range of $\tau$, it is often of interest to investigate: (1) how to summarize the information provided by these estimators to help understand the underlying effect mechanism; and (2) how to determine whether some covariates have constant effects so that a simpler model may be considered.

A general formulation of problem (1) may correspond to the estimation of some functional of $\beta_0(\cdot)$, denoted by $\Psi(\beta_0)$. A natural estimator for $\Psi(\hat{\beta})$ is given by $\Psi(\hat{\beta})$. Such an estimator may be justified by the functional delta method provided that $\Psi$ is compactly differentiable at $\beta_0$ (Andersen et al., 1998).

Addressing question (2) can be formulated as a testing problem for the null hypothesis $H_{0,j}: \beta_0^{(j)}(\tau) = \rho_0$, $\tau \in [\tau_L, \tau_U]$, where the superscript $(j)$ indicates the $j$th component of a vector, and $\rho_0$ is an unspecified constant, $j = 2, \ldots, p+1$. An example test procedure is presented in Peng and Huang (2008). Of note, accepting $H_{0,j}$ for all $j \in \{2, \ldots, p+1\}$ may indicate the adequacy of an AFT model. This naturally renders a procedure for testing the goodness-of-fit of an AFT model.

Model checking is often of practical importance. When the interest only lies in checking the linearity between covariates and conditional quantile at a single quantile level, one can adapt the approaches developed for uncensored data, for example, the work by Zheng (2000), Horowitz and Spokoiny (2002), and He and Zhu (2003). When the focus is to test a global linear relationship between conditional quantiles and covariates, a natural approach is to use a stochastic process which has mean zero under the assumed model. For example, a diagnostic process for model (3.28) (Peng and Huang, 2008) may take the form,

$$K_n(\tau) = n^{-1/2} \sum_{i=1}^{n} q(\tilde{Z}_i) M_i(\tau; \hat{\beta}),$$
where $q(\cdot)$ is a known bounded function, and

$$M_i(\tau; \beta) = N_i\left(\exp\{\tilde{Z}_i^{T} \beta(\tau)\}\right) - \int_{0}^{\tau} I\{\tilde{V}_i \geq \tilde{Z}_i^{T} \beta(u)\}dH(u).$$

It can be shown that $K(\tau)$ converges weakly to a zero-mean Gaussian process, whose distribution can be approximated by that of

$$K^*(\tau) = n^{-1/2} \sum_{i=1}^{n} q(\tilde{Z}_i)M_i(\tau; \hat{\beta})(1 - \zeta_i) + n^{-1/2} \sum_{i=1}^{n} q(\tilde{Z}_i)\{M_i(\tau; \beta^*) - M_i(\tau; \hat{\beta})\}.$$

Here $\beta^*(\tau)$ denotes the resampling estimator obtained by perturbing the estimating equation (3.35) by $\{\zeta_i\}_{i=1}^{n}$, which are independent variates from a nonnegative known distribution with mean 1 and variance 1. An unusual pattern of $K(\cdot)$ compared to that of $K^*(\cdot)$ would suggest a lack-of-fit of model (3.28). A formal lack of fit test is given by the supremum statistic, $\sup_{\tau \in [l, u]} |K(\tau)|$, where $0 < l < u < \tau_U$. The $p$-value may be approximated by the empirical proportion of $\sup_{\tau \in [l, u]} |K^*(\tau)|$ exceeding $\sup_{\tau \in [l, u]} |K(\tau)|$.

### 3.3.6 Extensions to other survival settings

In practice, survival data may involve more complex censoring mechanism than random right censoring. Truncation can also present, for example, in many observational studies. There are recent method developments for quantile regression in more complex survival scenarios. For example, Ji et al. (2012) proposed a modification of Peng and Huang (2008)’s martingale-based approach for survival data subject to known random left censoring and/or random left truncation in addition to random right censoring. Quantile regression with competing risks or semicompeting risks data was addressed by the work of Peng and Fine (2009) and Li and Peng (2011).

There are continued research efforts to address quantile regression for other important survival scenarios, such as interval censored data and dependent censoring data. As a promising regression tool for survival analysis, quantile regression may be recommended in a greater extent in real applications to provide complementary and yet useful secondary analysis.

### 3.3.7 An illustration of quantile regression for survival analysis

To illustrate quantile regression for survival analysis, we use a dataset from a dialysis study that investigated predictors of mortality risk in a cohort of 191 incident dialysis patients (Kutner et al., 2002). Analysis covariates included patient’s age (AGE), the indicator of fish consumption over the first year of dialysis (FISHH), the indicator of baseline HD dialysis modality (BHDPD), the indicator of moderate to severe symptoms of restless leg symptoms (BLEGS), the indicator of education equal or higher than college (HIEDU), and the indicator of being black (BLACK). We first fit the data with AFT model (3.25), where $T$ stands for time-to-death. In Table 3.1, we present the estimation results based on the log-rank estimator, Gehan’s estimator, and the least-squares estimator. All covariates except for BLEGs are consistently shown to have significant effects on survival by all three different estimators. For the coefficient of BLEGs, Gehan’s estimator and the least-squares estimator yield significant $p$ values while the log-rank estimator does not.

We next conduct quantile regression based on model (3.28) for the same dataset. Figure 3.3 displays Peng and Huang (2008)’s estimator of $\beta_0(\tau)$ along with 95% pointwise confidence intervals. In Figure 3.3, we observe that the coefficient for BLEGs diminishes
TABLE 3.1
Results from fitting AFT model to the dialysis dataset. Coef: coefficient estimate; SE: standard error.

<table>
<thead>
<tr>
<th></th>
<th>Gehan’s estimator</th>
<th>Log-rank estimator</th>
<th>Least-squares estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef</td>
<td>SE</td>
<td>p value</td>
</tr>
<tr>
<td>AGE</td>
<td>−0.031</td>
<td>0.005</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>FISHH</td>
<td>0.402</td>
<td>0.139</td>
<td>0.004</td>
</tr>
<tr>
<td>BHDPD</td>
<td>−0.505</td>
<td>0.156</td>
<td>0.001</td>
</tr>
<tr>
<td>BLEGs</td>
<td>−0.340</td>
<td>0.166</td>
<td>0.040</td>
</tr>
<tr>
<td>HIEDU</td>
<td>−0.352</td>
<td>0.133</td>
<td>0.008</td>
</tr>
<tr>
<td>BLACK</td>
<td>0.640</td>
<td>0.144</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

gradually with $\tau$ whereas estimates for the other coefficients seem to be fairly constant. We apply the second-stage inference to formally investigate the constancy of each coefficient. The results confirm our observation from Figure 3.3, suggesting a varying effect of BLEGs and constant effects of the other covariates. This may lead to an interesting scientific implication that BLEGs may affect the survival experience of dialysis patients with short survival times, but may have little impact on that of long-term survivors. In addition, the evidence for the nonconstancy of BLEGs coefficient may indicate some degree of lack-of-fit of the AFT model for the dialysis data.

We further estimate the average quantile effects defined as $\int_0^u \bar{\beta}^{(i)}(u) du$ ($i = 2, \ldots, 7$). The results are given in Table 3.2. We observe that the estimated average effects are similar to the AFT coefficients obtained by Gehan’s estimator and the least-squares estimator, but have relatively larger discrepancies with the log-rank estimates. This may indirectly reflect the presence of some nonconstant covariate effect. When quantile regression model (3.28) holds without any varying covariate effects, we would expect to see more consistent results between Gehan’s estimator, which emphasizes early survival, and the log-rank estimator, which treats early and late survival information equally.

TABLE 3.2
Estimation of average covariate effects based on quantile regression. AveEff: Estimated average effect; SE: standard error.

<table>
<thead>
<tr>
<th></th>
<th>AveEff</th>
<th>SE</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>−0.030</td>
<td>0.003</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>FISHH</td>
<td>0.327</td>
<td>0.116</td>
<td>0.005</td>
</tr>
<tr>
<td>BHDPD</td>
<td>−0.489</td>
<td>0.162</td>
<td>0.003</td>
</tr>
<tr>
<td>BLEGs</td>
<td>−0.369</td>
<td>0.161</td>
<td>0.022</td>
</tr>
<tr>
<td>HIEDU</td>
<td>−0.350</td>
<td>0.137</td>
<td>0.011</td>
</tr>
<tr>
<td>BLACK</td>
<td>0.654</td>
<td>0.144</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
FIGURE 3.3
Peng and Huang’s estimator (solid lines) and 95% pointwise confidence intervals (dotted lines) of regression quantiles in the dialysis example.
Bibliography


URL: http://dx.doi.org/10.1007/978-0-387-68560-1


Alternatives to the Cox Model


Martinussen, T. and Vansteelandt, S. (2012), A note on collapsibility and confounding bias in Cox and Aalen regression models, Technical report, Department of Biostatistics, University of Copenhagen.


