

Regression Models for Lifetime Data: An Overview

Chryst Caroni

Department of Mathematics, National Technical University of Athens, 157 80 Athens, Greece; ccar@math.ntua.gr

Abstract: Two methods dominate the regression analysis of time-to-event data: the accelerated failure time model and the proportional hazards model. Broadly speaking, these predominate in reliability modelling and biomedical applications, respectively. However, many other methods have been proposed, including proportional odds, proportional mean residual life and several other “proportional” models. This paper presents an overview of the field and the concept behind each of these ideas. Multi-parameter modelling is also discussed, in which (in contrast to, say, the proportional hazards model) more than one parameter of the lifetime distribution may depend on covariates. This includes first hitting time (or threshold) regression based on an underlying latent stochastic process. Many of the methods that have been proposed have seen little or no practical use. Lack of user-friendly software is certainly a factor in this. Diagnostic methods are also lacking for most methods.

Keywords: lifetime data; regression; proportional hazards; proportional odds; mean residual life; median residual life; proportional reversed hazards; accelerated failure time; first hitting time

1. Introduction

The purpose of the present paper is to give an overview of the several forms of regression models that have been proposed for use when the dependent variable is the time until the occurrence of an event, with simple examples being the death of a patient (survival analysis) and the failure of a machine (reliability modelling). A basic form of statistical model regresses the value y of a continuous dependent (response) random variable Y on the values of a vector of covariates (predictors or explanatory variables) \mathbf{x} recorded for the same statistical unit. The standard example of regression is of course the general linear model

$$Y = \beta' \mathbf{x} + \epsilon, \quad (1)$$

where β is a vector of regression coefficients and the random error term ϵ in the standard model follows the Normal distribution with zero mean and constant variance σ^2 , $\epsilon \sim N(0, \sigma^2)$. Taking, as usual, the values of \mathbf{x} as fixed (not random), this implies that the conditional distribution of Y given \mathbf{x} is also normal,

$$Y|\mathbf{x} \sim N(\beta' \mathbf{x}, \sigma^2). \quad (2)$$

This model is not suitable for application to lifetime data; the fact that the dependent variable—usually time or a proxy for time, such as the distance run by a vehicle—is non-negative demands a special approach. The same models also apply to other cases of non-negative dependent variables besides time, such as the load that can be applied to a sample of material before it breaks. The regression models examined here are all intended for use with dependent variables of these types. There will be no attempt to review each model in the sense of covering all the developments, as they are far too numerous to permit this. For example, it will be assumed here that the covariates \mathbf{x} are measured at the baseline (time origin of the study), whereas a necessary practical extension of any regression model is also to allow for time-varying covariates. However, this extension does not alter the



Citation: Caroni, C. Regression Models for Lifetime Data: An Overview. *Stats* **2022**, *5*, 1294–1304. <https://doi.org/10.3390/stats5040078>

Academic Editor: Wei Zhu

Received: 6 November 2022

Accepted: 3 December 2022

Published: 7 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

concept underlying the model. In this paper, it is the concept behind each model that will be presented, with comments on its strengths and weaknesses.

There will be some emphasis on parametric models. In this context, given data in the most common form of a set of n independent observations $\{(t_i, \mathbf{x}_i, \delta_i), i = 1, \dots, n\}$, where δ_i is the censoring indicator and t_i is the observed event time (when $\delta_i = 0$) or the time of right censoring (when $\delta_i = 1$), parameters can be estimated straightforwardly by maximising the likelihood

$$L = \prod_{i \in U} f(t_i) \prod_{i \in C} S(t_i) \quad (3)$$

where $f(t)$ is the density function of event times T , and $S(t) = 1 - F(t)$ is the survival function $P(T > t)$. U denotes the set of uncensored observations (observed failures) and C the set of right-censored times. As usual, uninformative right censoring is assumed. An equivalent expression for the likelihood is

$$L = \prod_{i \in U} h(t_i) \prod_{i=1}^n S(t_i) \quad (4)$$

where $h(t) = f(t)/S(t)$ is the hazard function. From these expressions, it can be seen that in order to estimate a model that incorporates dependence on covariates, it suffices to be able to express any one of the three functions $S(t|\mathbf{x})$, $f(t|\mathbf{x})$ or $h(t|\mathbf{x})$, which are equivalent in the sense that any one can be derived from any other. Almost all models are also available in a semi-parametric formulation, in which, for example, a baseline hazard function remains unspecified. Model fitting in these cases requires another approach.

One line of development of regression models for lifetime data is suggested immediately by Equation (2), where the effect of covariates on Y is to alter one of the parameters of its distribution while retaining the same functional form. This approach will appear among the models presented below. However, another type of model is more common. When the general linear model is replaced by logistic regression for a binary dependent variable (a generalised linear model), the covariates act multiplicatively on a function of the distribution of Y , namely, the odds of success $P(Y = 1|\mathbf{x})/P(Y = 0|\mathbf{x})$. Several regression models for lifetime data adopt this idea in some way.

2. Accelerated Failure Time

The most direct approach to the construction of a regression model for a non-negative dependent variable (denoted T from here on) is to transform it to the whole real line, which can be done most simply by taking its logarithm. Model (1) above applies, with Y replaced by $\ln T$; thus, T follows a log-normal distribution. This distribution is widely used in the analysis of lifetime data. It appears less commonly in biomedical applications but has been recommended there, for example, for analysing certain data on cancer [1,2].

More generally, Equation (1) can be replaced by

$$\ln T_{\mathbf{x}} = \mu + \beta' \mathbf{x} + \sigma \epsilon, \quad (5)$$

where ϵ is an error term that follows a distribution with location parameter zero and scale parameter one, and σ is a scale parameter. Taking $\epsilon \sim N(0, 1)$ gives the log-normal distribution for T . Another very important case is obtained when ϵ follows the Gumbel distribution (with location zero and scale one), implying that T follows the Weibull distribution, which is the most widely used distribution in engineering applications of lifetime data analysis.

From the above Equation (5) for $\ln T_{\mathbf{x}}$, the survival function of T given \mathbf{x} is

$$S(t|\mathbf{x}) = P(T_{\mathbf{x}} > t) = P(\mu + \beta' \mathbf{x} + \sigma \epsilon > \ln t). \quad (6)$$

Hence,

$$S(t|\mathbf{x}) = S_0 \left(t e^{-\beta' \mathbf{x}} \right), \quad (7)$$

where S_0 is a baseline survival function. This shows that the covariates' effect is to alter the time scale, and hence the name accelerated life or accelerated failure time (AFT) for such a model. More generally, the model can be formulated as $S(t|\mathbf{x}) = S_0(tg(\mathbf{x}))$ for a non-negative function g .

The AFT model predominates in lifetime data analysis in the fields of engineering and technology. The concept of changing the time scale corresponds to the way that many experimental studies of lifetimes are carried out, such as applying heavier loads than normal to components in order to compress the period over which failures occur so that data can be collected within a reasonable length of time. However, it is interesting to note that the heavily used Weibull regression can be regarded not only as an AFT model but also as belonging to the other major class of models, proportional hazards (see Section 4), which is much less used in the field of reliability modelling. In the parameterisation that will be employed here, the Weibull distribution has the hazard function $h(t) = \eta t^{\eta-1} / \alpha^\eta$ for positive parameters α (scale) and η (shape). Thus, the hazard is monotonic, increasing if $\eta > 1$, decreasing if $\eta < 1$ or constant if $\eta = 1$. In contrast, the log-normal distribution's hazard function always increases to a peak and then declines as t increases. This makes the obvious point that models even within the same framework may have quite different properties that need to be checked against the actual behaviour of the data.

3. Proportional Odds

The proportional odds (PO) regression model [3,4] follows the idea indicated at the end of Section 1. In fact, effectively, it is logistic regression. The odds of the occurrence of the event by time t

$$\theta(t) = \frac{F(t)}{1 - F(t)} = \frac{1 - S(t)}{S(t)} \quad (8)$$

is related to covariates in the PO model by

$$\theta(t|\mathbf{x}) = \theta_0(t)g(\mathbf{x}), \quad (9)$$

where $\theta_0(t)$ is a baseline odds function and the non-negative function $g(\mathbf{x})$ is usually $g(\mathbf{x}) = e^{\beta'\mathbf{x}}$. This gives the model

$$\ln \theta(t|\mathbf{x}) = \ln \theta_0(t) + \beta'\mathbf{x}. \quad (10)$$

Because this form is so similar to logistic regression, which is the familiar regression model for binary data, and the regression coefficients can be interpreted in the same way, it might be expected that the PO model would find extensive use in the biomedical fields and elsewhere. In fact, however, PO models are used infrequently, with the exception of the one based on the log-logistic distribution of event times, which also fits into the accelerated failure time framework [5]. This lack of use of the PO model must to a large extent be explained by its general absence from computer packages for statistical analysis. Detailed discussions of PO models can be found in [6].

The equation for $\theta(t|\mathbf{x})$ implies that the survival function in a PO model is given by

$$S(t|\mathbf{x}) = \{1 + \theta_0(t)g(\mathbf{x})\}^{-1}. \quad (11)$$

Hence, the ratio of the hazard functions for two units with different covariate values is

$$\frac{h(t|\mathbf{x}_1)}{h(t|\mathbf{x}_2)} = \frac{g(\mathbf{x}_1)}{1 + \theta_0(t)g(\mathbf{x}_1)} / \frac{g(\mathbf{x}_2)}{1 + \theta_0(t)g(\mathbf{x}_2)} \quad (12)$$

This ratio is monotonic in t and it tends to one as $t \rightarrow \infty$, because $\theta_0(t) \rightarrow \infty$ (since $S_0(t)$ tends to zero) while $g(\cdot)$ is independent of t . Therefore, it is a feature of the PO model that initial differences in hazards between units with different values of \mathbf{x} ultimately disappear.

4. Proportional Hazards

In the linear model (2) and also in the class of generalised linear models that includes it, the covariates' effect on a distribution is modelled as affecting one of its parameters (see also Section 11 below). Taking up this idea for the Weibull distribution, the survival function

$$S(t|\mathbf{x}) = \exp\{-(t/\alpha)^\eta\} \quad (13)$$

becomes

$$S(t|\mathbf{x}) = \exp\{-(t/\alpha(\mathbf{x}))^\eta\} \quad (14)$$

when dependence on covariates is introduced into the scale parameter α . Let $\alpha(\mathbf{x}) = \alpha e^{\beta'\mathbf{x}}$ or simply $\alpha(\mathbf{x}) = e^{\beta'\mathbf{x}}$ since the constant α can be absorbed into the exponent. It follows that the hazard function $h(t) = -\frac{d}{dt} \ln S(t)$ is

$$h(t|\mathbf{x}) = \eta t^{\eta-1} e^{\beta'\mathbf{x}}. \quad (15)$$

This implies that the ratio of the hazard functions $h(t|\mathbf{x}_1)$ and $h(t|\mathbf{x}_2)$ of two units with different values of the covariates is independent of time; one is always the same multiple of the other for all time. This is the proportional hazards (PH) property. In general, the PH regression model is defined by

$$h(t|\mathbf{x}) = h_0(t)g(\mathbf{x}), \quad (16)$$

for non-negative $g(\mathbf{x})$ where $h_0(t)$ is a baseline hazard function. The defining feature of the PH model—that the difference between the hazards of two units remains the same for ever—is in sharp contrast to the PO model, in which such differences disappear over time. It follows that it is impossible for any distribution to possess both the PO and PH properties.

Although to some extent this review emphasises parametric models, semi-parametric versions of most of the classes of models mentioned here can be found in the literature. In particular, Cox's famous PH regression model [7] is and always has been a semi-parametric model in which the baseline hazard $h_0(t)$ need not be specified. Along with its apparently easy interpretability and the ready availability of software, this is undoubtedly part of its appeal to non-specialists and helps to account for its widespread use in the biomedical sciences.

However, it has often been pointed out that there is no convincing rationale behind the PH model. For example, Oakes [8] states that Cox himself said that there is usually no simple physical or biological motivation for the assumption of proportional hazards. Cox concluded the original paper with the claim that the model seemed flexible and satisfactory as an empirical method of data reduction [7]. However, it can be asked, if the model is wrong, why are these parameter estimates a good summary of the data [9]?

As the AFT and PH models predominate, studies have been carried out comparing and contrasting them. One important finding is that the AFT model is generally more robust to misspecification.

Many basic statistical models have been extended in one or more ways. In the case of PH, a generalised model has been proposed in the following form [10]:

$$h(t|\mathbf{x}) = r\{\mathbf{x}(t)\}q\{H(t|\mathbf{x})\}h_0(t), \quad (17)$$

where r and q are positive functions. Thus, the hazard rate at time t depends not only on the current values of the covariates (as in PH) but also on their history, as expressed by the cumulative hazard $H(t)$. One special case is the generalised linear PH model [11], in which $r(\mathbf{x}) = e^{\beta'\mathbf{x}}$ as usual and

$$q\{H(t|\mathbf{x})\} = e^{\gamma H(t|\mathbf{x})}, \quad (18)$$

so that

$$h(t|\mathbf{x}) = e^{\beta'\mathbf{x} + \gamma H(t|\mathbf{x})} h_0(t). \quad (19)$$

Thus, the cumulative hazard up to this moment in time is treated as an additional, unknown covariate.

5. Proportional Reversed Hazards

The PH and PO models share the structure

$$\psi(t|\mathbf{x}) = \psi_0(t)g(\mathbf{x}), \quad (20)$$

where ψ represents the hazard function or the odds, respectively, and $g(\mathbf{x})$ is a non-negative function which is usually, but not necessarily, $e^{\beta'\mathbf{x}}$. This suggests further “proportional” models in which ψ stands for another function that is related to the distribution of lifetimes of the event under study. One of these is the proportional reversed hazards (PRH) model [12]. Whereas the hazard is related to the occurrence of the event in the interval of length δt after time t , the reversed hazard $r(t)$ is related to the conditional probability that an event occurred in the interval of length δt before time t . The definition is

$$r(t)\delta t = P(t - \delta t < T \leq t | T \leq t) = f(t)\delta t / F(t). \quad (21)$$

Thus, $r(t) = (d/dt) \ln F(t)$, whereas $h(t) = -(d/dt) \ln S(t)$. In the PRH model, covariates act multiplicatively on the baseline function $r_0(t)$,

$$r(t|\mathbf{x}) = g(\mathbf{x})r_0(t). \quad (22)$$

An equivalent expression is $F(t|\mathbf{x}) = [F_0(t)]^{g(\mathbf{x})}$, similar to $S(t|\mathbf{x}) = [S_0(t)]^{g(\mathbf{x})}$, which describes the PH model. The PRH property is possessed by, for example, the exponentiated Weibull distribution, which has baseline distribution function $F(t|\alpha) = [1 - \exp(-t^\alpha)]^\theta$, and is one special case of a wider family of distributions that has this property [13].

The origin of the PRH model was a suggestion in [14] and received its first development in [15], where it was pointed out that if a set of lifetimes satisfies this model, then their reciprocals satisfy a PH model. However, right-censored observations are transformed into left-censored ones, so the usual PH estimation procedures do not apply.

6. Proportional Lifetimes

6.1. Proportional Mean Residual Life

The mean residual life (MRL) or life expectancy of a unit with current age t is defined as its expected additional lifetime beyond this point,

$$\mu(t) = E[T - t | T \geq t], \quad (23)$$

which, if it exists, is equal to

$$\mu(t) = \frac{\int_t^\infty S(u)du}{S(t)}. \quad (24)$$

The proportional MRL model [16–18] proposes the relationship

$$\mu(t|\mathbf{x}) = \mu_0(t)g(\mathbf{x}), \quad (25)$$

where $\mu_0(t)$ is the baseline MRL function. An example of its application outside the usual reliability and biomedical fields—specifically, in finance—is discussed in [19]. Another proposed model closely related to the proportional MRL model is the proportional vitalities model [20]. Vitality at time t is defined as the conditional lifetime $E(T|T > t)$, which is equal to the expected remaining lifetime from t onwards (i.e., the MRL) plus the time already lived—that is, $t + \mu(t)$. This new model does not seem to have been used anywhere at the time of writing.

There is extensive literature on MRL [21], which especially in the field of reliability is a quantity of great interest. One possible advantage of a proportional MRL model versus,

say, PH, is a more communicable interpretation. It seems clearer to say that a new drug offers 10% longer average life than the old treatment, compared to saying that it reduces the hazard by 5%. Oakes and Dasu [16,17] suggest that the MRL function is preferable to the hazard function for modelling lifetime data, because it is a summary of the entire remaining distribution rather than just giving the immediate risk of failure. However, Hougaard [22] presents the view that mean lifetime is an inappropriate measure in biostatistics. The main reasons are the difficulty of estimating means in the presence of a skew distribution and right censoring, but also the possibility that a proportion of the population will never experience the event under study (see Section 10 below).

In [23], a proportional MRL model was assumed for times until death following diagnosis of dementia. This facilitated model-fitting in the part of the sample that consisted of people who already had a diagnosis, but at an unknown date, because in combination with the length-biased sampling, it transformed into a PH model. However, model-checking for the proportional MRL assumption could only be carried out indirectly by checking the PH assumption in this part of the sample, and not directly by examining the larger prospective part of the sample because of the lack of diagnostic methods.

6.2. Proportional Mean Past Lifetimes

The mean past lifetime at time t is the expected time since the occurrence of the event under study

$$k(t) = E[t - T | T \leq t], \quad (26)$$

and similarly to (24), this equals

$$k(t) = \frac{\int_0^t F(u) du}{F(t)}. \quad (27)$$

The proportional mean past lifetime regression model

$$k(t|\mathbf{x}) = k_0(t)g(\mathbf{x}), \quad (28)$$

was proposed in [24].

6.3. Proportional Median Residual Lifetime

The median residual lifetime at time t is the additional time required until the survival function falls to half of its current value $S(t)$. This is given by

$$m(t) = S^{-1}\left(\frac{1}{2}S(t)\right) - t. \quad (29)$$

The median is a more natural measure than the mean for use with the skewed distributions of lifetime data. A proportional median residual lifetime model has been proposed along the usual lines [25]. (There appears to still be a gap in the market for a proportional median past lifetime model.) A technical difficulty that has to be overcome in its estimation is that a given median residual life function is not uniquely related to one survival function.

7. Accelerated Hazards

The accelerated hazards model combines ideas from both the PH and AFT models. The covariates change the hazard function, but in the same way that they enter the AFT model. As seen in Section 2, the covariates have the effect of shifting the time scale in the AFT model, with t changing to $te^{\beta'x}$. The accelerated hazards model [26] puts the AFT shift in the hazard function

$$h(t|\mathbf{x}) = h_0\left(te^{\beta'x}\right). \quad (30)$$

A feature of the model is that it allows the survival functions of different subgroups to cross over (neither PH nor AFT permits this) as well as the hazard functions (allowed by AFT but not by PH).

A further extension includes AFT, PH and accelerated hazards in one model. It has the hazard function

$$h(t|\mathbf{x}) = h_0\left(te^{\beta_1' \mathbf{x}}\right)e^{\beta_2' \mathbf{x}}. \quad (31)$$

This is the extended hazard regression model [27]. Setting $\beta_1 = \mathbf{0}$ gives the PH model, $\beta_1 = \beta_2$ gives AFT and $\beta_2 = \mathbf{0}$ corresponds to accelerated hazards. It has been applied to the study of airline safety [28]. Furthermore, the extended model (31) has itself been extended [29] in a way that allows the spread parameter of the lifetime distribution to also depend on covariates, thus linking to the multi-parameter models discussed below.

8. Additive Hazards

In all the models with a proportionality definition, the covariates act in some way multiplicatively. In contrast, in the additive hazards model [30], they have an additive effect on the baseline hazard

$$h(t|\mathbf{x}) = h_0(t) + \beta' \mathbf{x}. \quad (32)$$

One reason for considering additive effects is thinking of a machine where parts operate in series, so that if one fails, the system fails. In that case, the hazard function of the machine's failure is the sum of the separate components' hazards. In a biostatistical application, Xie et al. [31] compared additive hazard models with the Cox PH model and discussed the different information that they would provide for public health planning and intervention.

9. Some Relations between Models

As noted above, there is a unique distribution—namely, the Weibull distribution—that possesses both the PH and AFT properties, and another—the log-logistic distribution—that is unique in possessing both the PO and AFT properties. However, no distribution can be both PO and PH, because of the incompatible behaviour of the ratios of hazard functions. Furthermore, the PRH property can be shown to be distinct from PH and PO as follows. By definition,

$$h = \frac{f}{S} = \left(\frac{f}{F}\right)\left(\frac{F}{S}\right) = r\theta \quad (33)$$

where h , r and θ are, respectively, the hazard function, reversed hazard function and odds of a given lifetime distribution, which has distribution function F , survival function S and probability density function f . Use subscripts 1 and 2 to denote these functions evaluated for two different covariate vectors. Then,

$$\frac{h_2}{h_1} = \left(\frac{r_2}{r_1}\right)\left(\frac{\theta_2}{\theta_1}\right) \quad (34)$$

shows how the ratios of the hazard, reversed hazard and odds functions are related. It is apparent that if any two of these ratios are independent of time, then all three must be. However, we know that no lifetime distribution possesses both the PO and PH properties, as noted in Section 4. Consequently, it is also impossible for any distribution to be both PH and PRH, or PO and PRH; therefore, the PH, PRH and PO properties are mutually exclusive.

If the PMRL and PH properties both hold, then the lifetime distribution must be the generalised Pareto [16]. Similarly, there is only one distribution that possesses both the proportional mean past lifetimes and PRH properties [24], namely, the power distribution

$$F(t) = t^\alpha, \quad 0 < t < 1, \quad \alpha > 0. \quad (35)$$

10. First Hitting Time Regression

As described in Section 4, the Weibull PH model is a single-parameter regression model: the covariates act on only one of its parameters while the other remains unchanged. However, if a distribution has more than one parameter, it is possible in principle to allow each one to depend on the covariates. A useful model in which this is done is the first hitting time (FHT) regression model [32,33]. It is also known as threshold regression, but this name is widely used in econometrics for a different model. Reviews and applications of the FHT regression model can be found in [34,35].

An FHT regression model arises if a lifetime is thought of as the realisation of an underlying Wiener process with drift μ and variance σ^2 . It could be thought, for example, that a patient's survival depends on a latent health status $Y(t)$ with $Y(0) = y_0 > 0$, and their death occurs when this falls to the threshold zero for the first time. The first passage time from y_0 to zero follows the inverse Gaussian distribution

$$f(t|y_0, \mu, \sigma^2) = \frac{y_0}{(2\pi\sigma^2 t^3)^{1/2}} \exp\left[-\frac{(y_0 + \mu t)^2}{2\sigma^2 t}\right], \quad (36)$$

and dependence on covariates is modelled by

$$\mu = \beta' \mathbf{u}, \ln y_0 = \gamma' \mathbf{v}, \quad (37)$$

where the parameters β and γ are vectors of regression coefficients and \mathbf{u} and \mathbf{v} are vectors of covariates (possibly but not necessarily the same vector). Because the above density can be rewritten in terms of μ/σ and y_0/σ , the variance σ^2 is usually taken to be equal to one without loss of generality. In the special case where only the drift depends on covariates while the starting level is constant, this single-parameter inverse Gaussian regression is a generalised linear model. If $\mu > 0$, the Wiener process has a non-zero probability of never reaching the boundary, given by $1 - e^{-2\mu y_0}$. (The above density still applies, conditional on reaching the boundary). This potentially represents a model for a cure fraction (also known as long-term survivors or non-susceptibles)—a proportion of the population that will never experience the event under study, no matter how long the study continues. The most common method of representing that phenomenon is by a mixture model

$$S_{pop}(t) = \pi S(t) + 1 - \pi \quad (38)$$

where S_{pop} is the population survival function, S is the survival function of the susceptibles and π is the proportion of susceptibles in the population. For some applications, the FHT model seems to present an attractive alternative to the mixture model, because the long-term survivors arise as a result of the process that applies to all the population, instead of there being a subset that is marked out from the beginning as not susceptible to the event. However, it appears that in practice it is often necessary to use a mixture model with the FHT in order to achieve good fit to data [36]. Furthermore, it has been found recently that the FHT mixture model with $\mu > 0$ and the FHT model with drift $-|\mu|$ provide identical fits to the data [37]. This identifiability problem is in addition to the possible multicollinearity mentioned in the following section.

As described above, the FHT regression model differs conceptually from all the regression models presented earlier, and from the PH model in particular. However, as it is a parametric model for survival time, its hazard function can be recovered without difficulty from the fitted model, if desired. This is facilitated by a simple command in the Stata implementation [33]. FHT regression is an important alternative in many circumstances where the PH assumption is untenable. Thus, many developments continue to be made in the methodology of FHT regression, and it is notable that these are driven by practical applications. Recent examples include its adaptation for high-dimensional data [38], application to group sequential designs in randomised trials [39], a Bayesian model allowing for

latent heterogeneity [40], and application to observational studies and clinical trials with delayed entry [41]. Earlier examples can be found in [34].

11. Other Multi-Parameter Regression Models

As remarked above, a regression model can be formulated based on any parametric distribution by making some or all of its parameters depend on covariates (which may or may not be the same covariates for each parameter). Unlike the regression models described above, these general multi-parameter regression models usually lack a basis for their interpretation, unlike, for example, the PH and FHT models.

The GAMLSS framework [42] allows up to four parameters to depend on covariates via generalised additive models, for an extensive selection of distributions. The methods are implemented in a set of R packages. As censoring is permitted, these distributions can be fitted to typical lifetime data. Although the available distributions include the inverse Gaussian, it employs a parameterisation different from the one used in Equation (36) for the FHT model.

On a much smaller scale than GAMLSS, the R package “mpr” [43] allows up to three parameters of a few common lifetime distributions (not including the inverse Gaussian) to depend on covariates [44]. Great flexibility in multi-parameter modelling is a feature of a new approach [45] in which the model is based on the cumulative hazard function of the lifetime,

$$H(t) = \lambda H_0\{(\phi t)^\gamma; \kappa\} \quad (39)$$

where λ is a PH scale parameter, ϕ is an AFT scale parameter, κ is a shape parameter and γ is a power parameter. The baseline distribution is taken as a form of the power generalised Weibull distribution, which encompasses various common lifetime distributions, and in some cases allows a cure fraction.

One issue that may arise in multi-parameter modelling is that there exists intrinsically a sort of multicollinearity. This can be illustrated most clearly in the FHT regression model of Section 10. A covariate could produce the same effect in either or both of two ways, for example, shortening lifetimes by either lowering the starting value y_0 or increasing the absolute value of the drift μ . It might be difficult to identify the significant predictors in the presence of this effect. Furthermore, evidence from studies discussed by Caroni [34] suggests that it is fairly common in applications for there to be conflict between the signs of the estimated regression coefficient. Opposite signs for the estimates $\hat{\beta}_i$ and $\hat{\gamma}_i$ corresponding to the covariate x_i are discordant in the sense that one indicates longer lifetimes and the other shorter lifetimes. Interpretation may be difficult. In the context of the latent process underlying the FHT regression model, it will sometimes be clear which covariates should be placed in which parameter, but in the more general multi-parameter models, there may be no guide.

12. Discussion

Several of the models presented in this paper seldom appear in applied work. Some have been around for about 30 years without really having entered into use, even though the introductory section of papers on their statistical properties often claim that they have attracted interest and have been found useful. There are at least three reasons for this. First, there is no convincing rationale for their defining feature (why should proportionality of the relevant function hold?). Second, they are not easy to use. Not only are they absent from the major user-friendly packages, there may also not be anything else available such as an R package. Third, the overwhelming predominance of an established method, such as Cox’s PH regression in biomedical applications, means that most users will take the safe course of using it, a choice that is unlikely to be challenged by referees and is of course readily available in the packages.

In parallel with the lack of practical applications of most of the methods described here, is an absence of diagnostic methods for model checking. This was noted earlier, specifically for the proportional MRL model, but applies generally. Work is required on

this practical aspect of modelling in order to turn some of these mathematical ideas into real statistical tools.

Many areas of statistical analysis have had to respond in recent years to the demands of new types of data, such as high-dimensional data, in which the number of variables may greatly exceed the number of observations. Methods of machine learning now present alternatives to more traditional statistical methods and may offer greater flexibility. Thus, straightforward regression models may be replaced or supplemented by decision trees and random forests for survival data [46]. The Cox model has also been combined with neural networks [47,48]. A recent review of the Cox model on the occasion of its 50th anniversary suggests that, as far as prediction is concerned, machine learning techniques tend to outperform Cox regression [49]. This area of research is very active.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AFT	Accelerated Failure Time
FHT	First Hitting Time
MRL	Mean Residual Life
PH	Proportional Hazards
PO	Proportional Odds
PRH	Proportional Reversed Hazards

References

1. Chapman, J.W.; O'Callaghan, C.J.; Hu, N.; Ding, K.; Yothers, G.A.; Catalano, P.J.; Shi, Q.; Gray, R.C.; O'Connell, M.J.; Sargent, D.J. Innovative estimation of survival using log-normal survival modelling on ACCENT database. *Br. J. Cancer* **2013**, *108*, 784–790. [[CrossRef](#)] [[PubMed](#)]
2. Royston, P. The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. *Stat. Neerl.* **2001**, *55*, 89–104. [[CrossRef](#)]
3. Bennett, S. Analysis of survival data by the proportional odds model. *Stat. Med.* **1983**, *2*, 273–277. [[CrossRef](#)] [[PubMed](#)]
4. Bennett, S. Log-logistic regression models for survival data. *Appl. Stat.* **1983**, *32*, 165–171. [[CrossRef](#)]
5. Lawless, J.F. A note on lifetime regression models. *Biometrika* **1986**, *73*, 509–512. [[CrossRef](#)]
6. Collett, D. *Modelling Survival Data in Medical Research*, 3rd ed.; Chapman & Hall: Boca Raton, FL, USA, 2014.
7. Cox, D.R. Regression models and life tables. *J. R. Statist. Soc. B* **1972**, *34*, 187–220. [[CrossRef](#)]
8. Oakes, D. An introduction to survival models: In honor of Ross Prentice. *Lifetime Data Anal.* **2013**, *19*, 442–462. [[CrossRef](#)]
9. Freedman, D.A. Survival analysis: A primer. *Am. Stat.* **2008**, *62*, 110–119. [[CrossRef](#)]
10. Bagdanovicius, V.B.; Nikulin, M.S. Generalized proportional hazards model based on modified partial likelihood. *Lifetime Data Anal.* **1999**, *5*, 329–350. [[CrossRef](#)]
11. Bagdanovicius, V.B.; Hafdi, M.A.; Himdi, K.E.; Nikulin, M. Statistical analysis of the generalized linear proportional hazards model. *J. Math. Sci.* **2005**, *127*, 1673–1681. [[CrossRef](#)]
12. Gupta, R.C.; Gupta, R.D. Proportional reversed hazard rate model and its applications. *J. Stat. Plan. Inf.* **2007**, *137*, 3525–3536. [[CrossRef](#)]
13. Popovic, B.V.; Genc, A.I.; Domma, F. Generalized proportional reversed hazard rate distributions with application in medicine. *Stat. Methods Appl.* **2022**, *31*, 459–480. [[CrossRef](#)]
14. Kalbfleisch, J.D.; Lawless, J.F. Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS. *J. Am. Stat. Assoc.* **1989**, *84*, 360–372. [[CrossRef](#)]
15. O'Neill, T.J. The inverse proportional hazards model. *Statist. Probab. Lett.* **1991**, *12*, 125–129. [[CrossRef](#)]
16. Oakes, D.; Dasu, T. A note on residual life. *Biometrika* **1990**, *77*, 89–104. [[CrossRef](#)]
17. Oakes, D.; Dasu, T. Inference for the proportional mean residual life model. *Inst. Math. Stat. Lect. Notes* **2003**, *43*, 105–116.
18. Zahedi, H. Proportional mean remaining life model. *J. Stat. Plan. Inf.* **1991**, *29*, 221–228. [[CrossRef](#)]
19. Cai, Z.; Su, J.; Gufianti. A regression analysis of expected shortfall. *Stat. Interface* **2015**, *8*, 295–303. [[CrossRef](#)]
20. Shrahili, M.; Albabtain, A.A.; Kayid, M.; Kaabi, Z. Stochastic aspects of proportional vitalities model. *Mathematics* **2020**, *8*, 1823. [[CrossRef](#)]
21. Jeong, J.-H. *Statistical Inference on Residual Life*; Springer: New York, NY, USA, 2014.

22. Hougaard, P. Fundamentals of survival data. *Biometrics* **1999**, *55*, 13–22. [[CrossRef](#)]
23. Lee, C.H.; Ning, J.; Kryscio, R.J.; Shen, Y. Analysis of combined incident and prevalent cohort data under a proportional mean residual life model. *Stat. Med.* **2019**, *38*, 2103–2114. [[CrossRef](#)] [[PubMed](#)]
24. Asadi, M.; Berred, A. Properties and estimation of the mean past lifetime. *Statistics* **2012**, *46*, 405–417. [[CrossRef](#)]
25. Bandos, H. Regression on Median Residual Life Function for Censored Survival Data. Ph.D. Thesis, University of Pittsburgh, Pittsburgh, PA, USA, 2007.
26. Chen, Y.Q. Accelerated hazards regression model and its adequacy for censored survival data. *Biometrics* **2001**, *57*, 853–860. [[CrossRef](#)] [[PubMed](#)]
27. Ciampi, A.; Etezadi-Amoli, J. A general model for testing the proportional hazards and the accelerated failure time hypotheses in the analysis of censored survival data with covariates. *Comm. Statist. Theory Methods* **1985**, *14*, 651–667. [[CrossRef](#)]
28. Shyur, H.-J.; Keng, H.; I, I.-K.; Huang, C.-L. Using extended hazard regression model to assess the probability of aviation event. *Appl. Math. Comput.* **2012**, *218*, 10647–10655. [[CrossRef](#)]
29. Louzada-Neto, F. Extended hazard regression model for reliability and survival analysis. *Lifetime Data Anal.* **1997**, *3*, 367–381. [[CrossRef](#)]
30. Aalen, O.O. A linear regression model for the analysis of life times. *Stat. Med.* **1989**, *8*, 907–925. [[CrossRef](#)]
31. Xie, X.; Strickler, H.D.; Xue, X. Additive hazard regression models: An application to the natural history of human papillomavirus. *Comput. Math. Models Med.* **2013**, *2013*, 796270. [[CrossRef](#)]
32. Lee, M.-L.T.; Whitmore, G.A. Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary. *Stat. Sci.* **2006**, *21*, 501–513. [[CrossRef](#)]
33. Xiao, T.; Whitmore, G.A.; He, X.; Lee, M.-L.T. Threshold regression for time-to-event analysis: The stthreg package. *Stata J.* **2012**, *12*, 257–283. [[CrossRef](#)]
34. Caroni, C. *First Hitting Time Regression Models*; ISTE & John Wiley: London, UK; Hoboken, NJ, USA, 2017.
35. Lee, M.-L.T. A survey of threshold regression for time-to-event analysis and applications. *Taiwan J. Math.* **2019**, *23*, 293–305.
36. Balka, J.; Desmond, A.F.; McNicholas, P.D. Review and implementation of cure models based on first hitting times for Wiener processes. *Lifetime Data Anal.* **2009**, *15*, 147–176. [[CrossRef](#)]
37. Lee, M.-L.T.; Whitmore, G.A. Multivariate threshold regression models with cure rates: Identification and estimation in the presence of the Esscher property. *Stats* **2022**, *5*, 172–189. [[CrossRef](#)]
38. De Bin, R.; Stikbakke, V.G. A boosting first-hitting-time model for survival analysis in high-dimensional settings. *Lifetime Data Anal.* **2022**, *in press*. [[CrossRef](#)]
39. Chen, Y.; Lawrence, J.; Lee, M.-L.T. Group sequential designs for randomized trials using “first hitting time” model. *Stat. Med.* **2022**, *41*, 2375–2402. [[CrossRef](#)]
40. Race, J.A.; Pennell, M.L. Semi-parametric survival analysis via Dirichlet process mixtures of the First Hitting Time model. *Lifetime Data Anal.* **2021**, *27*, 177–194. [[CrossRef](#)]
41. Lee, M.-L.T.; Lawrence, J.; Chen, Y.; Whitmore, G.A. Accounting for delayed entry into observational studies and clinical trials: Length-biased sampling and restricted mean survival time. *Lifetime Data Anal.* **2022**, *28*, 637–658. [[CrossRef](#)]
42. Rigby, R.A.; Stasinopoulos, M.D.; Heller, G.Z.; De Bastiani, F. *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2021.
43. R Package ‘mpr’. Available online: <https://cran.r-project.org/web/packages/mpr/mpr.pdf> (accessed on 5 November 2022).
44. Burke, K.; MacKenzie, G. Multi-parameter regression survival modeling: An alternative to proportional hazards. *Biometrics* **2017**, *73*, 678–686. [[CrossRef](#)]
45. Burke, K.; Jones, M.C.; Noufaily, A. A flexible parametric modelling framework for survival analysis. *Appl. Stat.* **2020**, *69*, 429–457. [[CrossRef](#)]
46. Ishwaran, H.; Kogalur, U.B.; Blackstone, E.H.; Lauer, M.S. Random survival forests. *Ann. Appl. Stat.* **2008**, *2*, 841–860. [[CrossRef](#)]
47. Katzman, J.L.; Shaham, U.; Cloninger, A.; Bates, J.; Jiang, T.; Kluger, Y. DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* **2018**, *18*, 24. [[CrossRef](#)] [[PubMed](#)]
48. Kvamme, H.; Borgan, O.; Scheel, I. Time-to-event prediction with neural networks and Cox regression. *J. Mach. Learn. Res.* **2019**, *20*, 1–30.
49. Andersen, P.K. Fifty years with the Cox proportional hazards regression model. *J. Indian Inst. Sci.* **2022**, *in press*. [[CrossRef](#)]