On Parametric Survival Analysis

Kevin Burke
University of Limerick, Ireland
M.C. Jones
Open University, U.K.
Angela Noufaily
University of Warwick, U.K.

Abstract

We introduce a general, flexible, parametric survival modelling framework which encompasses key shapes of hazard function (constant, increasing, decreasing, up-then-down, down-then-up), various common survival distributions (log-logistic, Burr type XII, Weibull, Gompertz), and includes defective distributions (i.e., cure models). This generality is achieved using four basic distributional parameters: two scale-type parameters and two shape parameters. Generalising to covariate dependence, the scale-type regression components correspond to accelerated failure time (AFT) and proportional hazards (PH) models. Therefore, this general formulation unifies the most popular survival models which allows us to consider the practical value of possible modelling choices for survival data. Furthermore, in line with our proposed flexible baseline distribution, we advocate the use of multi-parameter regression in which more than one distributional parameter depends on covariates – rather than the usual convention of having a single covariate-dependent (scale) parameter. While many choices are available, we suggest introducing covariates through one or other of the two scale parameters, which covers AFT and PH models, in combination with a "power" shape parameter, which allows for more complex non-AFT/non-PH effects, while the other shape parameter remains covariate-independent, and handles automatic selection of the baseline distribution.

Keywords: Accelerated failure time; Flexible modelling; Multi-parameter regression; Power generalised Weibull distribution; Proportional hazards.

1 Introduction

We consider a univariate lifetime random variable, $T > 0$, whose cumulative hazard function (c.h.f.), $H(t)$, is, atypically, modelled using a flexible parametric form which we take to be

$$H(t) = \lambda\, H_0\left((\phi t)^{\gamma}\,; \kappa\right), \qquad t > 0. \tag{1}$$

Here, $H_0(\cdot;\kappa)$ is an underlying c.h.f. with shape parameter κ, and $\phi, \lambda, \gamma > 0$ are further parameters with the following distinct interpretations: ϕ controls the horizontal scaling of the hazard function, and is well known as the accelerated failure time parameter (also, $1/\phi$ is the usual distributional scale parameter); λ controls the vertical scaling of the hazard function, and is well known as the proportional hazards parameter; and γ is a second shape parameter which is explicitly defined as a power parameter (unlike κ which can enter in potentially more complicated ways, and might even represent a vector of parameters). Were $Y = \log(T)$ to be modelled as a location-scale distribution on $\mathbb{R}$, then $\mu = -\log\phi$ and $\sigma = 1/\gamma$ would be the location and scale of that distribution, respectively, these relationships driving our preference to specify γ as a power parameter rather than as a more general shape parameter.

In this article, we also propose a specific choice for $H_0(t^{\gamma};\kappa)$ based on — but adapted from — the choice

$$H_N(t^{\gamma};\kappa) = (1 + t^{\gamma})^{\kappa} - 1 \tag{2}$$

corresponding to the 'power generalised Weibull' (PGW) distribution introduced by Bagdonaviçius and Nikulin (2002). This choice has some major advantages: with just two shape parameters, the full range of simplest hazard shapes, namely, constant, increasing, decreasing, up-then-down or down-then-up (and no others), are available, the parameters γ and κ controlling this through the way they control behaviour of the hazard function near zero and at infinity. Here, we use the simple descriptive terms 'up-then-down' and 'down-then-up' to avoid the term 'bathtub-shaped', which is down-then-up but with a flat valley, the clumsy term 'upside-down-bathtub-shaped', and the terms 'unimodal/uniantimodal' which also encompass monotone hazards. Our adaptation of the PGW distribution retains the hazard behaviour outlined here, and also allows κ to control distributional choice within the family: for $\kappa > 0$, log-logistic and Burr Type XII distributions are the heaviest tailed members, Weibull distributions are 'central' within the family, and Gompertz-related distributions are the most lightly tailed. See Section 2 for details of this model, which also include links and interactions with frailty, transformation and cure models.

Any one or more of the four distributional parameters in model (1) can be made to depend, typically log-linearly, on covariates; such "multi-parameter regression" is one of the focusses of this work. Indeed, this general formulation covers the most popular survival models, e.g., the accelerated failure time (AFT) model when ϕ depends on covariates, the proportional hazards (PH) model when λ depends on covariates, and semi-parametric versions when H_0 is an unspecified function. In particular, an advantage of considering (1) is that one may evaluate the breadth of possible modelling choices. Our primary focus in this respect is to consider which distributional parameters should depend on covariates to assess, for example, whether an AFT model (ϕ regression) is, in general, likely to provide a superior fit when compared with a PH model (λ regression), the utility of a simultaneous AFT-PH model (simultaneous ϕ and λ regression components), and the merits of a shape regression component (γ or κ) in addition to the, more standard, AFT and PH components. One might also consider whether or not non-parametric components should be introduced either for functions of covariates within the regression equations, or for the baseline c.h.f., H_0, or both. However, this is beyond the scope of the current paper.

The reason for our focus on the core model structure rather than the development of non-/semi-parametric approaches is that we feel that, within the survival literature, there is a general over-emphasis placed on semi-parametric models – compared with other fields of statistics – to the extent that many useful parametric alternatives do not receive the attention they deserve. In particular, practitioners are often content with the "flexibility" afforded by a non-parametric baseline function without considering the possibly inflexible structural assumptions of the model at hand. Indeed, a structurally flexible parametric framework has the potential to outperform a less flexible semi-parametric model; for example, there might be more to be gained by contemplating the extension of a PH model (λ regression) to include a γ regression, than by extension to a non-parametric H_0. Of course, this is not to downplay the importance of a sufficiently flexible baseline function, and our proposed choice for H_0, which is based on (2), is quite general as it covers a wide variety of popular survival distributions.

After Section 2, in which we introduce and justify our choice of baseline distribution and develop its properties, we consider the extension to regression modelling in Section 3, including model interpretation and estimation. Then, the properties of estimation within this general framework, and further practical aspects, are explored using simulated and real data in Sections 4 and 5, respectively. Finally, we close with some discussion in Section 6.

2 The Specific Model for H_0

2.1 Basic Definition and Properties

Our starting point is the PGW distribution with c.h.f. given by (2) and hazard function

$$h_N(t^\gamma;\kappa) = \kappa\gamma\, t^{\gamma-1}(1+t^\gamma)^{\kappa-1}, \qquad t>0$$

(Bagdonaviçius and Nikulin, 2002; Nikulin and Haghighi, 2009; Dimitrakopoulou et al., 2007). This is a tractable distribution with readily available formulae for its (unimodal) density, survivor and quantile functions also. For fixed $\gamma, \kappa > 0$,

$$h_N(t^\gamma;\kappa) \sim \kappa\gamma\, t^{\gamma-1} \ \text{ as } \ t\to 0 \quad \text{and} \quad h_N(t^\gamma;\kappa) \sim \kappa\gamma\, t^{\kappa\gamma-1} \ \text{ as } \ t\to\infty.$$

The power parameter γ controls the behaviour of the hazard function at zero: it goes to 0 (constant) ∞ as $\gamma >$ (=) < 1. As $t\to\infty$, the hazard function goes to 0 (constant) ∞ as $\kappa\gamma <$ (=) > 1. In fact, the PGW hazard function joins these tails smoothly in such a way that its hazard shapes are readily shown to be as listed in Table 1.

Table 1: Shapes of PGW hazard functions

γ	$\kappa\gamma$	shape
1	1	constant
≤ 1	≤ 1	decreasing
≤ 1	≥ 1	down-then-up
≥ 1	≤ 1	up-then-down
≥ 1	≥ 1	increasing

Here, pairs of $\le$'s and/or $\ge$'s include the convention 'and not both equal at once'.

Since in practice H in (1) will incorporate arbitrary vertical (λ) and horizontal (ϕ) rescalings, we may adapt H_N for use as H_0 by choosing different vertical and horizontal scalings than those chosen in (2). In particular, for $\gamma, \kappa > 0$, we set

$$H_A(t^\gamma;\kappa) = \frac{\kappa+1}{\kappa}\left\{\left(1+\frac{t^\gamma}{\kappa+1}\right)^{\kappa} - 1\right\}, \qquad t>0. \tag{3}$$

Note immediately that the hazard shape properties associated with H_A remain the same as those associated with H_N, as given in Table 1. Whenever the hazard function, now given by

$$h_A(t^\gamma;\kappa) = \gamma t^{\gamma-1}\left(1+\frac{t^\gamma}{\kappa+1}\right)^{\kappa-1}, \qquad t>0, \tag{4}$$

is non-monotone, its mode/antimode is at $\{(1-\gamma)(\kappa+1)/(\kappa\gamma-1)\}^{1/\gamma}$.

Defining H_A by (3) allows us to identify an especially large number of special and limiting cases, many important and well known, some less so, as listed in Table 2. (For the 'Weibull extension' distribution, see Chen (2000) and Xie et al. (2002); the $\gamma = 1$ special case of H_N is the extended exponential distribution of Nadarajah and Haghighi (2011).) The shapes of their hazard functions, which are also given in Table 2, reflect the general shape properties of Table 1, of course.

Table 2: Special and limiting cases of adapted PGW distributions

κ	H_A	shapes of h_A	distribution	others encompassed
0	$\log(1+t^\gamma)$	decreasing, up-then-down	log-logistic	$H_A \times \lambda \Rightarrow$ Burr type XII
1	t^γ	decreasing, constant, increasing	Weibull	$\gamma = 1 \Rightarrow$ exponential
2	$t^\gamma + \frac{1}{6}t^{2\gamma}$	decreasing, down-then-up, increasing		$\gamma = 1 \Rightarrow$ linear hazard
∞	$e^{t^\gamma} - 1$	increasing, down-then-up		$H_A \times \lambda \Rightarrow$ Weibull extension; $H_A \times \lambda, \gamma = 1 \Rightarrow$ Gompertz

It can also be shown that changing distribution from (2) to (3) retains membership of the log-location-scale-log-concave family of distributions of Jones and Noufaily (2015) and therefore, inter alia, unimodality of densities. We also now note, for future reference, the attractive form of the quantile function associated with H_A, namely $Q_A(u) = \{H_A\{-\log(1-u); 1/\kappa\}^{1/\gamma} \equiv Q_0(u;\kappa)^{1/\gamma}$.

2.2 Further Links and Properties

2.2.1 Frailty Links

By way of notation, write PGW(κ, λ) for the PGW distribution with proportionality parameter $\lambda > 0$ and shape parameter $\kappa > 0$, i.e., with c.h.f. $\lambda\, H_N(t^\gamma; \kappa)$. Frailty is usually introduced into survival models by mixing over the distribution of the proportionality parameter λ. A given survival distribution can be produced from another given survival distribution by such frailty mixing if the ratio of their hazard functions is decreasing (Gupta and Gupta (1996)). Thus, when PGW(κ, λ) is mixed with a certain frailty distribution, PGW($\omega\kappa, \lambda_0$) for $0 \leq \omega \leq 1$ and appropriate $\lambda_0 > 0$ results. The following novel result identifies the mixing distribution. It is a tempered stable (TS) distribution (Tweedie, 1984; Hougaard, 1986; Fischer and Jakob, 2016). This distribution has three parameters, $0 \leq \omega \leq 1$, $\xi > 0$ and $\theta \geq 0$. However, we will take $\theta = 1$ throughout and refer to the corresponding distribution as $TS(\omega, \xi)$. It is defined through its Laplace transform given by

$$\mathcal{L}_{\omega,\xi}(s) = \exp\left[-\frac{\xi}{\omega}\left\{(1+s)^{\omega} - 1\right\}\right].$$

Result 1. Let $T|B = b \sim$ PGW(κ, b) and let $B \sim$ TS(ω, ξ). Then, $T \sim$ PGW($\omega\kappa, \xi/\omega$).

`Proof.` Denote by $g_{\omega,\xi}$ the density of TS(ω, ξ). Then,

$$\begin{aligned} P(T \geq t) &= \int_0^{\infty} \exp\{-b\, H_N(t^{\gamma}; \kappa)\} g_{\omega,\xi}(b)\, db = \mathcal{L}_{\omega,\xi}\{H_N(t^{\gamma}; \kappa)\} \\ &= \exp\left(-\frac{\xi}{\omega}\left[\{1 + (1+t^{\gamma})^{\kappa} - 1\}^{\omega} - 1\right]\right) = \exp\left\{-\frac{\xi}{\omega} H_N(t^{\gamma}; \omega\kappa)\right\}. \quad \square \end{aligned}$$

A similar result holds for the adapted PGW distribution, but is more opaque because of the more complicated constants involved; see the online Supplementary Material for details.

Some interesting special cases of Result 1 are that:

- $\omega = 1/2$: if $T|B = b \sim$ PGW(κ, b) and B follows the inverse Gaussian distribution with parameters $(1/2, 1/2)$, then $T \sim$ PGW($\kappa/2, 1$);
- $\omega = 0$: if $T|B = b \sim$ PGW(κ, b) and B follows the unit-scale gamma distribution with shape parameter ξ, then T follows the Burr Type XII distribution with power parameter γ and proportionality parameter $\xi\kappa$. (When $\kappa = 1$, this is the well known result that a Weibull distribution with gamma frailty results in the Burr Type XII distribution.)

A related frailty link between the $\kappa = \infty$ and $\kappa = 1$ adapted PGW distributions is:

- if $T|B = b$ has c.h.f. $b(e^{t^{\gamma}} - 1)$ and B follows the exponential distribution with parameter 1, then T follows the Weibull distribution.

2.2.2 As a Transformation Model

Linear transformation models (c.f. Kalbfleisch and Prentice (2002, sec. 7.5); Zeng and Lin (2007)) are concerned with c.h.f.'s of the form

$$H_T(t) = w(\theta H_0(t)) \tag{5}$$

where $\theta > 0$ depends log-linearly on covariates and both the transformation function w and baseline function H_0 are c.h.f.'s. It is easy to see that the PGW c.h.f.'s H_N and H_A are of the form (5). Either is a transformation model with H_0 the Weibull c.h.f. t^γ; in terms of the overall model, whenever the Weibull is used as baseline, $\theta^{-1/\gamma}$ is the horizontal scale parameter so that when covariates are included only in it, the transformation model is an accelerated failure time model. The transformation w is a version of the Box-Cox transformation given, in the simpler case of H_N, by $w(y) = (1+y)^\kappa - 1$ (Box and Cox, 1964; Yeo and Johnson, 2000).

If Y is the lifetime random variable following the transformation model with, for simplicity, c.h.f. H_N, then the model can also be written in the form $\gamma \log Y = -\log\theta + \log E$ where E follows the distribution with c.h.f. $(1+y)^\kappa - 1$.

2.2.3 The Associated Cure Model

The new adaptation can also be used to widen the family of PGW distributions by taking $-1 < \kappa < 0$. For clarity, define $\psi = \kappa + 1$ so that $0 < \psi < 1$. The adapted PGW c.h.f. can then be written as

$$H_A(t^\gamma; \psi) = \frac{\psi}{1-\psi}\left(1 - \frac{1}{\{1+(t^\gamma/\psi)\}^{1-\psi}}\right).$$

This corresponds to a cure model with cure probability $p_\psi \equiv \exp\{-\psi/(1-\psi)\}$. Since the (improper) survival function is in this case of the form $p_\psi^{1-S_0(t)}$, this cure model has an interpretation as the distribution of the minimum of a Poisson number of random variables (e.g. cancer cells, tumours), each following the lifetime distribution with survival function S_0 (e.g. Tsodikov et al. (2003)); here, the Poisson parameter is $\psi/(1-\psi)$ and $S_0(t) = \{1+(t^\gamma/\psi)\}^{\psi-1}$ is the survival function of a scaled Burr Type XII distribution. The hazard functions $\gamma t^{\gamma-1} h_A(t^\gamma; \psi)$ follow the shape of their $\psi \to 1$ limit — the log-logistic — being decreasing for $\gamma \leq 1$ and up-then-down otherwise.

The frailty links of Section 2.2.1 extend to a link between the adapted PGW(κ) and PGW($-L$) distributions when $\kappa > 0$ and $0 < L < 1$. The frailty distribution is the cure model version of the tempered stable distribution explored by Aalen (1992). See the online Supplementary Material for details.

2.3 Why This Particular Choice for H_0?

The PGW distribution shares the set of hazard behaviours listed in Table 1 with two other established two-shape-parameter lifetime distributions centred on the Weibull distribution, namely, the generalised gamma (GG) and exponentiated Weibull (EW) distributions; see Jones and Noufaily (2015). See Figure 1 for many examples of just how similar the hazard shapes of all three distributions are; in Figure 1, we have chosen the scale parameter such that each distribution has median one, but otherwise specified shape parameters $\gamma, \kappa > 0$ only so that all three hazard functions behave as $t^{\gamma-1}$ as $t \to 0$ and as $t^{\kappa\gamma-1}$ as $t \to \infty$.
Further effort to choose shape parameters to match hazard functions or other aspects of the distributions even more closely is possible and has been pursued for the EW and GG distributions by Cox and Matheson (2014) and extended to the PGW distribution (what they call the generalised Weibull distribution) by Matheson et al. (2017). Cox and Matheson (2014) state that the "agreement between the two distributions [GG and EW] in our various comparisons, both graphically and in terms of the K–L [Kullback-Leibler] distance, is striking"; after a similar K–L matching exercise, Matheson et al. (2017) state that "the survival and hazard functions of the [PGW] distribution and its matched GG are visually indistinguishable." It remains, therefore, to choose between PGW, GG and EW distributions on other grounds. The GG distribution includes the Weibull and gamma distributions as special cases and the lognormal as a limiting one; the EW distribution includes the Weibull and exponentiated exponential distributions. However: we have been unable to match the number of PGW's advantageous properties as in the previous subsections by similarly adapting either the GG or EW distributions; we prefer the breadth of/difference between the wide range of distributions encompassed by the PGW distribution; and we appreciate the greater tractability of the PGW distribution both mathematically and computationally (for instance, its hazard function has a simpler form compared with the GG — which involves an incomplete gamma function — and the EW).

We note that $H_N(t;\kappa)$ is of the form $\mathcal{H}^{-1}\{\kappa\mathcal{H}(t)\}$ where $\mathcal{H}(t) = \log(1+t)$ is the c.h.f. of the log-logistic distribution — and $\mathcal{H}^{-1}(t) = e^t - 1$ is the c.h.f. of the Gompertz distribution. Similarly, $H_A(t;\kappa)$ is $(\kappa+1)\mathcal{H}^{-1}\{\kappa\mathcal{H}(t/(\kappa+1))\}/\kappa$ for the same choice of $\mathcal{H}$. This explains the neatness of some of the properties given above. We have, however, failed to find a viable competitor to the PGW distribution from among this class. Interestingly, if we invert the roles of $\mathcal{H}$ and $\mathcal{H}^{-1}$ by taking $\mathcal{H}(t) = e^t - 1$ we get the well known proportional odds (PO) model which therefore has the same limiting distributions as the PGW distribution, and is very much of interest in its own right. However, when a power parameter is included, the PO model has quite different hazard behaviour from the PGW model.

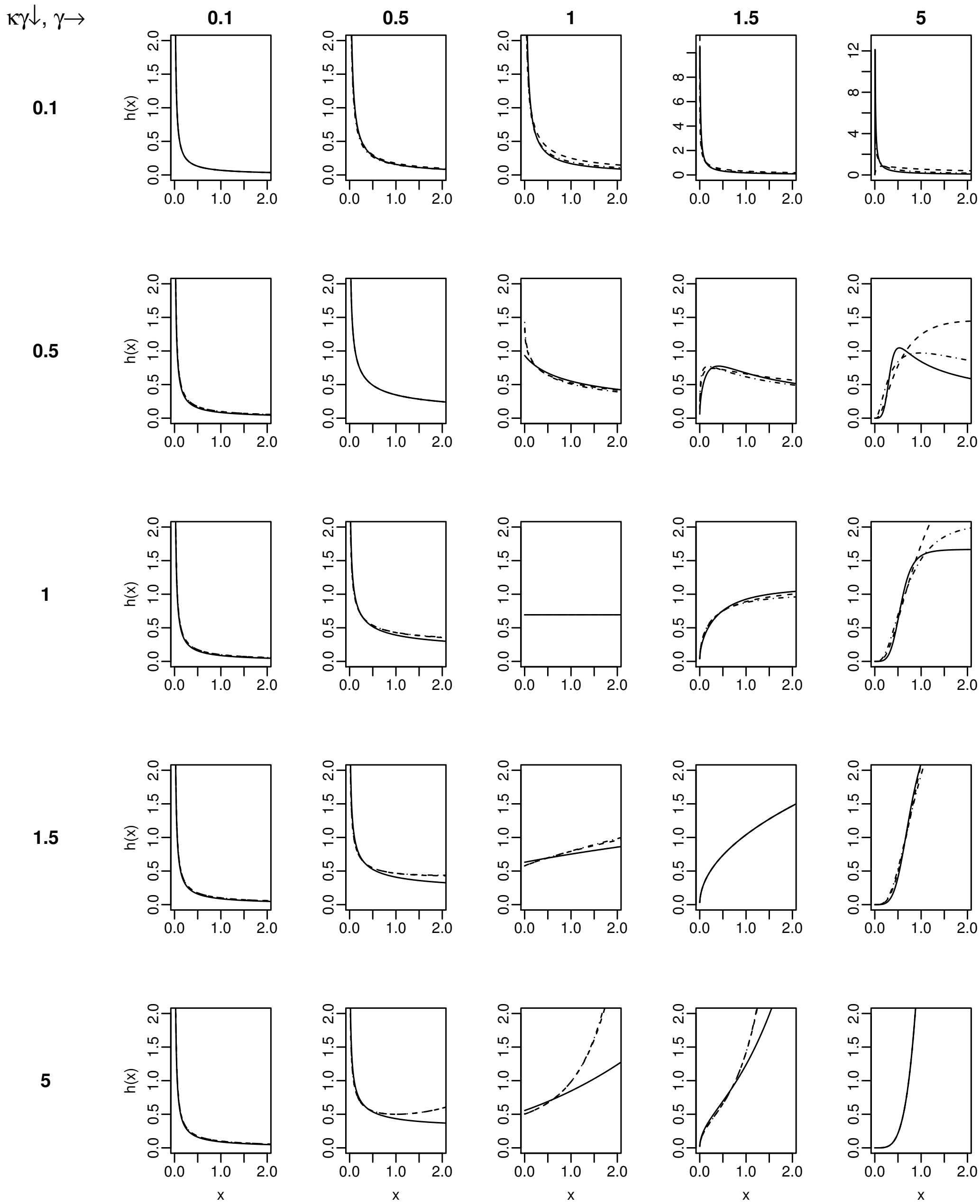

Figure 1: Hazard functions of PGW (solid), GG (dashed), and EW (dot-dashed) distributions for the values of γ and $\kappa\gamma$ specified along the top and down the left-hand side of the figure, respectively. In each case, the scale parameter is chosen such that the median of the distributions is one. Figures on the main diagonal of the matrix of figures, in each of which the PGW, GG and EW hazard functions are identical, correspond to Weibull distributions, the figure in the centre to the exponential distribution.

3 Regression

3.1 Modelling Choices

Within our proposed PGW modelling framework, there are four parameters, ϕ, λ, γ, and κ, which could potentially depend on covariates. Note that most classical modelling approaches are based on having a *single* covariate-dependent distributional parameter, which we refer to as single parameter regression, where, understandably, there is a particular emphasis on scale-type parameters, e.g., the accelerated failure time (AFT) model (ϕ regression) and the proportional hazards (PH) model (λ regression). However, in line with the flexibility of the PGW distribution itself, we also consider taking a flexible multi-parameter regression (MPR) approach in which more than one parameter may depend on covariates (cf. Burke and MacKenzie (2017), and references therein, for details of multi-parameter regression). The most general linear PGW-MPR is, therefore, given by

$$\log(\phi) = x^T\tau, \qquad \log(\lambda) = x^T\beta, \qquad \log(\gamma) = x^T\alpha, \qquad \log(\kappa + 1) = x^T\nu,$$

where log-link functions are used to respect the positivity of the parameters ϕ, λ and γ, with a slightly different link function for κ to accommodate the fact that, within our adapted PGW, it can take values in the range $(-1, \infty)$ (see Section 2.2.3), $x = (1, x_1, \ldots, x_p)^T$ is a vector of covariates, and $\tau = (\tau_0, \tau_1, \ldots, \tau_p)^T$, $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$, $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_p)^T$, and $\nu = (\nu_0, \nu_1, \ldots, \nu_p)^T$ are the corresponding vectors of regression coefficients. In practice, we may not necessarily have the same set of covariates appearing in all regression components, and, in our current notation, this can be handled by setting various regression coefficients to zero.

As mentioned in Section 1, we could extend the above regression specification via non-parametric regression functions of x, but this is beyond the scope of this paper and, indeed, the MPR approach is, in itself, already flexible without this added complexity. Furthermore, although the general PGW-MPR model offers the opportunity of four regression components simultaneously, our numerical studies (Sections 4 and 5) suggest that this full flexibility is unlikely to be required in practice, and whose large parameter space can result in highly unstable estimation procedures. In particular, we find that a good practical choice is composed of the following pieces: (a) only one scale parameter (ϕ or λ) depends on covariates, where all regression coefficients for the other scale parameter (including the intercept) are set to zero, (b) the γ shape parameter may depend on covariates, and (c) the κ shape parameter is constant, i.e., only the intercept, ν_0, is non-zero in the ν vector. This choice provides a useful framework which incorporates, depending on the choice of scale regression, either an AFT (τ) or PH (β) component, allows for non-AFT/non-PH effects via the α coefficients associated with the power parameter (Section 3.2), and automatically selects the underlying baseline distribution via ν_0 from a range of popular survival distributions (Table 2) including defective distributions, i.e., cure models (Section 2.2.3).

3.1.1 Notation

Let $M(\tau, \alpha)$ and $M(\beta, \alpha)$ denote the two models discussed in the previous paragraph, e.g., the latter is the model with β and α regression components, along with the shape parameter ν_0 (but where τ is a vector of zeros). More generally, beyond these two suggested models, we will use this notation throughout the paper where the arguments of $M(\cdot)$ indicate which regression components are present in the model, the absence of either β or τ indicating that this is a vector of zeros. Irrespective of the presence of α or ν regression components, we assume that α_0 and ν_0 are always present since these are needed to characterise the baseline distribution and the shape of its hazard function (see Tables 1 and 2). Thus, for example, $M(\tau)$ and $M(\beta)$ are, respectively, AFT and PH models with two shape parameters (α_0 and ν_0), $M(\beta, \alpha, \nu)$ is a model which extends the suggested $M(\beta, \alpha)$ model so that the ν regression component is present, and $M(\beta, \tau, \alpha, \nu)$ is the most general PGW-MPR model.

3.2 Interpretation

We first consider model $M(\tau, \alpha)$ which extends the basic AFT model via the incorporation of the α regression component. Now suppose that x_j is a binary covariate and let $x_{(-j)} = (1, x_1, \ldots, x_{j-1}, 0, x_{j+1}, \ldots, x_p)^T$ be the covariate vector with x_j set to zero so that we may write $x^T\tau = x_j\tau_j + x_{(-j)}^T\tau$ and $x^T\alpha = x_j\alpha_j + x_{(-j)}^T\alpha$. As this model extends the AFT model, it is natural to consider its quantile function which is given by

$$Q(u|x) = \exp(-x^T\tau)\, Q_0(u;\kappa)^{\exp(-x^T\alpha)}$$

where $Q_0(u) = H_A\{-\log(1-u); 1/\kappa\}$ is the "baseline" quantile function defined at the end of Section 2.1. We can then inspect the quantile ratio

$$QR_j(u) = \frac{Q(u|x_j = 1)}{Q(u|x_j = 0)} = \exp(-\tau_j)\, Q_0(u;\kappa)^{\exp(-x_{(-j)}^T\alpha)\{\exp(-\alpha_j)-1\}}$$

where we see that α_j is the key parameter in determining the u-dependence. In particular, since $Q_0(u;\kappa)$ is an increasing function of u, $QR_j(u)$ increases when $\alpha_j < 0$, decreases when $\alpha_j > 0$, and is constant (i.e., the usual AFT case) when $\alpha_j = 0$. Hence, the α_j coefficient has a natural interpretation in that it characterises the nature of the effect of the binary covariate x_j, and provides a test of the AFT property for that covariate.

Now consider the model $M(\beta, \alpha)$ which extends the PH model, and whose hazard function is given by

$$h(t|x) = \exp(x^T\beta)\, h_A(t^{\exp(x^T\alpha)}; \kappa)$$

where $h_A(t^\gamma; \kappa)$ is defined in (4). The hazard ratio for the binary covariate x_j is

$$HR_j(t) = \frac{h(t|x_j = 1)}{h(t|x_j = 0)} = \exp(\beta_j + \alpha_j)\, t^{\exp(x_{(-j)}^T\alpha)\{\exp(\alpha_j)-1\}} g(t; \alpha_j, x_{(-j)}^T\alpha, \kappa)$$

where

$$g(t; \alpha_j, x_{(-j)}^T\alpha, \kappa) = \left(\frac{t^{\exp(\alpha_j + x_{(-j)}^T\alpha)} + \kappa + 1}{t^{\exp(x_{(-j)}^T\alpha)} + \kappa + 1} \right)^{\kappa-1}.$$

Clearly, α_j characterises departures from proportional hazards as $HR_j(t)$ is a constant when $\alpha_j = 0$. For $\kappa \geq 0$, we have that $\lim_{t\to 0} HR_j(t) = 0$ and $\lim_{t\to\infty} HR_j(t) = \infty$ when $\alpha_j > 0$, while $\lim_{t\to 0} HR_j(t) = \infty$ and $\lim_{t\to\infty} HR_j(t) = 0$ when $\alpha_j < 0$. Furthermore, it can be shown that $HR_j(t)$ varies monotonically in t in the following cases: (i) $\kappa \geq 1$, or (ii) $0 < \kappa < 1$ and $\alpha_j \notin (\log \kappa, -\log \kappa)$. (We do not know about monotonicity or otherwise in the remaining cases.)

From the above we see that, within the suggested $M(\tau, \alpha)$ and $M(\beta, \alpha)$ models, the parameters play clear roles: the scale coefficients (τ or β) control the overall size of the effect where negative coefficients correspond to longer lifetimes; the α shape coefficients describe how covariate effects vary over the lifetime, i.e., permitting non-AFT and non-PH effects; and the ν_0 shape parameter characterises the baseline distribution. Note that we could, alternatively, achieve non-constant effects via the ν regression component rather than the α component, for example, using $M(\beta, \nu)$ rather than $M(\beta, \alpha)$. However, in this case, the interpretation is that such non-constant effects are due to populations which arise from structurally different distributions, rather than different shapes within a given baseline distribution. The latter is arguably more natural as it creates a clear separation of parameters whereas, in the former, distribution selection and non-constant effects are intertwined. Of course, this is not to say that models with ν components instead of, or in combination with, α components will never be useful in practice. We are simply highlighting practical merits of the $M(\tau, \alpha)$ and $M(\beta, \alpha)$ models and, indeed, the general use of these models is motivated by the numerical studies of Sections 4 and 5.

3.3 Estimation

Consider the model formulation given in (1) with all four regression components, i.e., the $M(\tau, \beta, \alpha, \nu)$ model. Let $\phi_i = \exp(x_i^T\tau)$, $\lambda_i = \exp(x_i^T\beta)$, $\gamma_i = \exp(x_i^T\alpha)$

and $\kappa_i = \exp(x_i^T \nu) - 1$ be the covariate-dependent distributional parameters for the ith individual with covariate vector $x_i = (1, x_{i1}, \ldots, x_{ip})^T$, and τ, β, α, and ν are the associated vectors of regression coefficients. Allow independent censoring by attaching to each individual an indicator δ_i which equals one if the response is observed, and zero if it is right-censored. The log-likelihood function is then given by

$$\ell(\theta) = \sum_{i=1}^{n} \left[\delta_i \left\{ \log\left(\frac{\lambda_i \gamma_i z_i}{t_i} \right) + m_0(z_i; \kappa_i) \right\} - \lambda_i H_0(z_i; \kappa_i) \right]$$

where $\theta = (\tau^T, \beta^T, \alpha^T, \nu^T)^T$, $z_i = (\phi_i t_i)^{\gamma_i}$ and, in our proposed PGW case,

$$H_0(t) = \frac{\kappa + 1}{\kappa} \left\{ \left(1 + \frac{t}{\kappa + 1} \right)^{\kappa} - 1 \right\},$$

$$m_0(t; \kappa) = \log h_0(t; \kappa) = (\kappa - 1) \log \left(1 + \frac{t}{\kappa + 1} \right).$$

As usual, the log-likelihood function can be maximised by solving the score equations

$$(U_\tau^T X,\ U_\beta^T X,\ U_\alpha^T X,\ U_\nu^T X)^T = 0_{4p \times 1}$$

where X is an $n \times p$ matrix whose ith row is x_i, $0_{4p\times1}$ is a $4p \times 1$ vector of zeros and U_τ, U_β, U_α, and U_ν are $n \times 1$ vectors whose ith elements are as follows:

$$U_{\tau,i} = \delta_i \left\{ \gamma_i + \gamma_i z_i\, m_0'(z_i; \kappa_i) \right\} - \lambda_i \gamma_i z_i\, h_0(z_i; \kappa_i),$$

$$U_{\beta,i} = \delta_i - \lambda_i H_0(z_i; \kappa_i),$$

$$U_{\alpha,i} = \delta_i \left[1 + \log(z_i)\ \left\{ 1 + z_i\, m_0'(z_i; \kappa_i) \right\} \right] - \lambda_i z_i \log(z_i) h_0(z_i; \kappa_i),$$

$$U_{\nu,i} = \left[\delta_i \left\{ \frac{\kappa_i + 1}{\kappa_i - 1} m_0(z_i; \kappa_i) - z_i m_0'(z_i; \kappa_i) \right\} \right. \\ \left. - \lambda_i \frac{\kappa_i + 1}{\kappa_i^2 (\kappa_i - 1)} \left\{ \kappa_i - 1 + \left(1 + \frac{t}{\kappa_i + 1} \right)^{\kappa_i} a_0(z_i; \kappa_i) \right\} \right],$$

where $a_0(t; \kappa) = \kappa(\kappa + 1) m_0(t; \kappa) - \kappa^2 t\, m_0'(t; \kappa) - \kappa + 1$.

Note that the vectors $U_{\tau,i}$, $U_{\beta,i}$ and $U_{\alpha,i}$ are written generically so that they apply to any model of the form given in (1), i.e., they are not specific to the PGW case; the form of $U_{\nu,i}$, on the other hand, uses the way in which H_0 and hence m_0 and m_0' depend on κ. Thus, although the PGW is certainly a flexible choice (see Section 2), the first three score components extend immediately to other baseline distributions by replacing H_0 (and, consequently, m_0 and m_0'). Estimation then proceeds straightforwardly once the functional form of $U_{\nu,i}$ has been re-evaluated.

Furthermore, one may, alternatively, prefer to maintain an unspecified baseline distribution, whereby ν represents an infinite-dimensional (possibly covariate-independent) parameter vector. In this case, estimation equations for the regression coefficients τ, β, and α can be based on $(U_\tau^T X,\ U_\beta^T X,\ U_\alpha^T X)$ where H_0 is replaced with appropriate non-parameteric estimator (and, similarly, for m_0 and m_0'). However, while non-parametric estimation of H_0 is straightforward (say, using a Nelson-Aalen-type estimator), it is well known that terms such as m_0 and m_0', which involve h_0 and h_0', are more difficult to estimate consistently. We note that semi-parametric versions of the $M(\tau,\beta)$ and $M(\tau,\alpha)$ models have respectively been developed by Chen and Jewell (2001) and Burke et al. (2017). However, we are unaware of a semi-parametric $M(\tau,\beta,\alpha)$ model in the literature. In any case, such semi-parametric models are beyond the scope of the current paper and, indeed, a flexible parametric framework can cover a wide variety of applications as previously discussed in Section 1.

4 Simulation Studies

4.1 Without Covariates

We simulated data from a basic (no covariate) PGW distribution parameterised in terms of the following unconstrained parameters: $\tau = \log\phi$, $\beta = \log\lambda$, $\gamma = \log\alpha$ and $\nu = \log(\kappa + 1)$. The values of the first three parameters were fixed at $\tau = 0.8$, $\beta = 0.5$, $\alpha = -0.3$, respectively, while ν was varied such that $\nu \in \{0.00, 0.22, 0.41, 0.69, 1.10, 1.61, \infty\}$ (rounded to two decimal places); note that $\nu = 0$, $\nu = 0.69$ and $\nu = \infty$ correspond, respectively, to the log-logistic, Weibull and Gompertz distributions. Furthermore, the sample size was fixed at 1000 and censoring times were generated from an exponential distribution such that, for each ν value, the censoring rate was fixed at approximately 30%. Within each of the seven simulation scenarios (i.e., varying ν), we fitted four different models with the aim of understanding the stability of estimation in a reasonably large sample: (i) estimate all parameters, (ii) fix β at its true value, (iii) fix β at zero, and (iv) fix ν at its true value. Thus, τ and α are estimated in all four models. We also considered additional scenarios where $\alpha = 0.3$ but the results are similar and, therefore, are not shown here.

Each scenario was replicated 1000 times, and the results are contained in Table 3. The main findings are summarised in the following bullet points.

- Estimation of all parameters simultaneously yields unstable results (model (i)). In particular, note that the $\hat\tau$ and $\hat\beta$ estimates are highly biased with large standard deviations, and the standard deviation of $\hat\nu$ is also very large.
- When β is fixed to the truth (model (ii)), all parameter estimates are quite unbiased with reasonably small standard deviations. Note, however, that the

standard deviation of $\hat{\nu}$ can be large when ν is large. This is a consequence of the fact that the PGW distribution changes very little over a range of large ν values and, therefore, is not a concern. On the other hand, when β is fixed to zero (model (iii)), the estimates are still stable, where, now, $\hat{\tau}$ converges to a value in the range $1.4 - 1.5$ (varying smoothly with ν), while $\hat{\alpha}$ and $\hat{\nu}$ do not change dramatically from their values in model (ii). This shows that the change in β is primarily absorbed by $\hat{\tau}$. Furthermore, the fitted survivor curves for both models (not shown) are close to the truth, i.e., there is no reduction in quality of model fit as a consequence of fixing β to an incorrect value.

- Assuming that the baseline shape of the distribution, ν, is known (model (iv)), we find that $\hat{\tau}$ and $\hat{\beta}$ are still unstable as evidenced by reasonably large standard errors (although the results are not nearly as bad as in model (i)).

The results of this simulation study are clearly indicative of (near) parameter redundancy associated with the two scale-type parameters, τ and β. Indeed, in all cases where these parameters are estimated simultaneously, we have found that $\text{corr}(\hat{\tau}, \hat{\beta}) \approx 1$. Of course, $\hat{\tau}$ and $\hat{\beta}$ are perfectly co-linear in the Weibull case ($\nu = 0.69$), but it is interesting to find that this extends (approximately) beyond the Weibull distribution. This appears to be a new finding in survival modelling and implies that these parameters play similar roles across a range of popular lifetime distributions. Hence, the practitioner might simply choose whether he/she prefers an AFT parameter, τ, or a PH parameter, β. Perhaps historically too much has been made of the AFT and PH models as being competitor models; we investigate this further in Section 4.2. It is noteworthy that $\hat{\alpha}$ is very stable in all cases (and is even close to being unbiased in the worst case scenario of model (i)) suggesting that we can flexibly model this shape parameter via covariates simultaneously with one of the scale parameters (AFT, τ, or PH, β); this is in line with the findings of Burke and MacKenzie (2017) who considered only the Weibull model. Finally, it appears reasonable that the ν parameter characterises the baseline distribution and does not depend on covariates since, already without covariates, there is some instability associated with its estimation (particularly for larger ν values).

4.2 With a Covariate

We simulated survival times according to the PGW distribution with parameters $\phi = \exp(\tau_0 + \tau_1 X)$, $\lambda = \exp(\beta)$, $\gamma = \exp(\alpha_0 + \alpha_1 X)$, and $\kappa = \exp(\nu) - 1$ where $X \sim \text{Bernoulli}(0.5)$, ν was varied according to the set $\{0.00, 0.22, 0.41, 0.69, 1.10, 1.61, \infty\}$, and the remaining parameter values were fixed at $\tau_0 = 0.8$, $\tau_1 = 0.6$, $\beta = 0.0$, $\alpha_0 = 0.2$, and $\alpha_1 = -0.5$; these values were selected to yield realistic survival times. Note that, in the notation of Section 3.1.1, the true model is $M(\tau, \alpha)$. As in Section 4.1, the

Table 3: Median and standard deviation (in brackets) of estimates

Model	ν	$\hat{\tau}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\nu}$
(i) β : est, ν : est	0.00	1.87 (6.93)	-0.21 (4.98)	-0.26 (0.12)	0.18 (4.38)
	0.22	2.24 (11.52)	-0.54 (8.38)	-0.26 (0.15)	0.39 (7.24)
	0.41	2.83 (11.55)	-0.98 (8.58)	-0.26 (0.20)	0.49 (7.55)
	0.69	2.11 (8.52)	-0.48 (6.54)	-0.31 (0.22)	0.76 (7.93)
	1.10	0.83 (4.30)	0.38 (3.32)	-0.33 (0.14)	1.18 (6.82)
	1.61	0.53 (2.76)	0.57 (2.13)	-0.32 (0.11)	12.32 (6.47)
	∞	1.10 (1.53)	0.19 (1.21)	-0.32 (0.09)	13.04 (6.40)
(ii) β : true, ν : est	0.00	0.81 (0.15)	0.50 —	-0.30 (0.05)	0.00 (0.11)
	0.22	0.79 (0.15)	0.50 —	-0.30 (0.05)	0.23 (0.15)
	0.41	0.80 (0.15)	0.50 —	-0.30 (0.05)	0.40 (0.19)
	0.69	0.79 (0.15)	0.50 —	-0.30 (0.06)	0.71 (0.31)
	1.10	0.79 (0.16)	0.50 —	-0.30 (0.06)	1.12 (1.59)
	1.61	0.80 (0.14)	0.50 —	-0.30 (0.05)	1.65 (3.72)
	∞	0.84 (0.09)	0.50 —	-0.28 (0.04)	13.05 (6.23)
(iii) β : zero, ν : est	0.00	1.52 (0.12)	0.00 —	-0.29 (0.05)	0.15 (0.09)
	0.22	1.50 (0.13)	0.00 —	-0.29 (0.06)	0.33 (0.12)
	0.41	1.49 (0.13)	0.00 —	-0.30 (0.05)	0.48 (0.14)
	0.69	1.48 (0.13)	0.00 —	-0.30 (0.06)	0.68 (0.19)
	1.10	1.44 (0.13)	0.00 —	-0.31 (0.06)	0.99 (0.27)
	1.61	1.44 (0.14)	0.00 —	-0.31 (0.06)	1.27 (1.46)
	∞	1.42 (0.13)	0.00 —	-0.31 (0.06)	2.04 (4.50)
(iv) β : est, ν : true	0.00	0.78 (0.58)	0.50 (0.31)	-0.30 (0.06)	0.00 —
	0.22	0.79 (1.13)	0.50 (0.70)	-0.30 (0.06)	0.22 —
	0.41	0.91 (1.69)	0.44 (1.14)	-0.29 (0.07)	0.41 —
	0.69	0.61 (0.19)	0.64 (0.14)	-0.30 (0.03)	0.69 —
	1.10	0.91 (1.60)	0.40 (1.27)	-0.30 (0.08)	1.10 —
	1.61	0.85 (0.83)	0.46 (0.75)	-0.30 (0.06)	1.61 —
	∞	0.78 (0.44)	0.51 (0.48)	-0.30 (0.07)	∞ —

All numbers are rounded to two decimal places. For the models with fixed parameters, the "estimated" value shown is the value at which the parameter is fixed, and its standard error is then indicated by "—". While τ and β are not simultaneously estimable when $\nu = 0.69$ (Weibull case), the estimation procedure still yields values (which depend completely on initial values) such that the constant $\lambda\phi^{\gamma}$ is preserved in the sense that its value is the same as the case where one of β or τ were held constant.

sample size and censored proportion were, respectively, set at 1000 and 30% (with censoring times generated from an exponential distribution). Within each of the seven scenarios (i.e., varying ν), we fitted the following three regression models: the more general $M(\tau, \beta, \alpha)$, the true $M(\tau, \alpha)$, and the misspecified $M(\beta, \alpha)$, respectively. The results, based on 1000 simulation replicates, are given in Table 4.

Table 4: Median and standard deviation (in brackets) of estimates

				Model(τ, β, α)			
ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$
0.00	0.98 (1.21)	0.55 (1.28)	-0.20 (1.37)	0.02 (0.70)	0.23 (0.13)	-0.50 (0.18)	0.05 (0.76)
0.22	0.98 (1.40)	0.51 (2.08)	-0.20 (1.63)	0.04 (1.61)	0.23 (0.17)	-0.50 (0.24)	0.27 (0.44)
0.41	1.00 (1.65)	0.57 (2.95)	-0.24 (1.99)	0.10 (2.48)	0.24 (0.18)	-0.51 (0.26)	0.40 (0.39)
0.69	1.18 (2.13)	0.34 (4.39)	-0.43 (2.63)	0.19 (3.85)	0.22 (0.20)	-0.51 (0.22)	0.61 (3.56)
1.10	0.39 (1.50)	0.08 (2.88)	0.35 (1.89)	0.21 (2.66)	0.17 (0.12)	-0.49 (0.18)	1.31 (6.88)
1.61	0.55 (1.00)	0.47 (1.62)	0.26 (1.34)	-0.04 (1.61)	0.19 (0.12)	-0.51 (0.18)	2.60 (7.04)
∞	0.88 (0.53)	0.69 (0.88)	-0.14 (0.81)	-0.06 (0.99)	0.21 (0.12)	-0.51 (0.17)	15.12 (6.46)
				Model(τ, α)			
ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$
0.00	0.80 (0.09)	0.60 (0.13)	0.00 —	0.00 —	0.21 (0.06)	-0.50 (0.06)	0.00 (0.06)
0.22	0.81 (0.09)	0.60 (0.12)	0.00 —	0.00 —	0.21 (0.06)	-0.50 (0.06)	0.22 (0.09)
0.41	0.80 (0.08)	0.59 (0.10)	0.00 —	0.00 —	0.20 (0.06)	-0.50 (0.06)	0.40 (0.11)
0.69	0.79 (0.08)	0.60 (0.10)	0.00 —	0.00 —	0.20 (0.06)	-0.50 (0.06)	0.70 (0.17)
1.10	0.80 (0.09)	0.60 (0.09)	0.00 —	0.00 —	0.20 (0.06)	-0.50 (0.06)	1.12 (0.84)
1.61	0.81 (0.08)	0.60 (0.08)	0.00 —	0.00 —	0.20 (0.07)	-0.50 (0.06)	1.59 (2.44)
∞	0.82 (0.05)	0.62 (0.06)	0.00 —	0.00 —	0.22 (0.05)	-0.50 (0.06)	13.16 (6.78)
				Model(β, α)			
ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$
0.00	0.00 —	0.00 —	0.88 (0.11)	0.03 (0.08)	0.18 (0.05)	-0.52 (0.05)	-0.36 (0.12)
0.22	0.00 —	0.00 —	0.91 (0.11)	0.04 (0.08)	0.18 (0.06)	-0.51 (0.06)	-0.06 (0.15)
0.41	0.00 —	0.00 —	0.93 (0.13)	0.05 (0.09)	0.19 (0.06)	-0.50 (0.06)	0.21 (0.21)
0.69	0.00 —	0.00 —	0.98 (0.13)	0.06 (0.09)	0.20 (0.06)	-0.50 (0.06)	0.72 (0.35)
1.10	0.00 —	0.00 —	1.03 (0.14)	0.07 (0.11)	0.22 (0.06)	-0.50 (0.07)	1.83 (4.33)
1.61	0.00 —	0.00 —	1.18 (0.10)	0.08 (0.12)	0.27 (0.05)	-0.50 (0.07)	15.16 (6.52)
∞	0.00 —	0.00 —	1.54 (0.10)	0.10 (0.14)	0.37 (0.05)	-0.48 (0.07)	16.88 (1.28)

All numbers are rounded to two decimal places. For the models with fixed parameters, the "estimated" value shown is the value at which the parameter is fixed, and its standard error is then indicated by "—".

Mirroring the case with no covariates (Section 4.1), we find that estimation is highly unstable when we attempt to estimate τ and β coefficients simultaneously in $M(\tau, \beta, \alpha)$, whereas all parameters can be estimated reliably in the true $M(\tau, \alpha)$. Furthermore, within the misspecified $M(\beta, \alpha)$, the parameter estimates are stable in the sense that the standard deviations are small and the β coefficients converge to values which vary smoothly with ν. Compared with fitting the true model, the $\hat{\nu}$ values change somewhat, while the $\hat{\alpha}$ coefficients are reasonably unchanged (apart from $\hat{\alpha}_0$ at larger ν values). Note that the results are broadly similar for smaller sample sizes of $n = 500$ and $n = 100$ although the standard deviations are quite large for most estimated parameters in the latter case (see the online Supplementary Material for details).

While the above discussion is concerned with the stability in estimating model parameters, we now turn to goodness-of-fit by inspecting the fitted baseline survivor curves, i.e., the survivor curve for an individual with $X = 0$ which we denote by $S_0(t)$. In particular, we focus on this estimated baseline survivor function evaluated at three true quantiles, namely, $Q_0(u)$, $u = 0.1, 0.5, 0.9$, since $\hat{S}_0(Q_0(u))$ is an estimate of the probability $1 - u$. Boxplots of these estimates over simulation replicates arising from the true model, Model(τ, α), and the misspecified model, Model(β, α), are shown in Figure 2 (we ignore Model(τ, β, α) due to the instability in estimated parameters). We also display the estimates from two simpler (misspecified) models, Model(τ) and Model(β), wherein X has been dropped from the α component (specifically, α_1 is set to zero but α_0 is a free parameter in these simpler models).

In the first two models (i.e., those with α regression components), the very good fit is not altered much by the choice of τ or β regression (apart from a little bias in $M(\beta, \alpha)$ when $\nu = \infty$). This is in line with the earlier findings of Section 4.1 and reiterates the fact that the difference between the AFT (τ) and PH (β) specifications is minimal. However, when the α regression is dropped, the quality of the model fit decreases considerably (but again note that the fit is similar between these two non-α models). Thus, it is clear that the inclusion of the α regression is a bigger model alteration than the move from AFT to PH specifications – and this is true over a considerable range of distributions (ν values).

5 Lung Cancer Analysis

We now consider our modelling framework in the context of a lung cancer study which was the subject of a 1995 Queen's University Belfast PhD thesis by P. Wilkinson (previously analysed in Burke and MacKenzie (2017)). This study concerns 855 individuals who were diagnosed with lung cancer between 1st October 1991 and 30th September 30 1992, who were then followed up until 30th May 1993 (approximately 20% of survival times were right-censored). The main aim of this study was to investigate differences between the following treatment groups: palliative care, surgery,

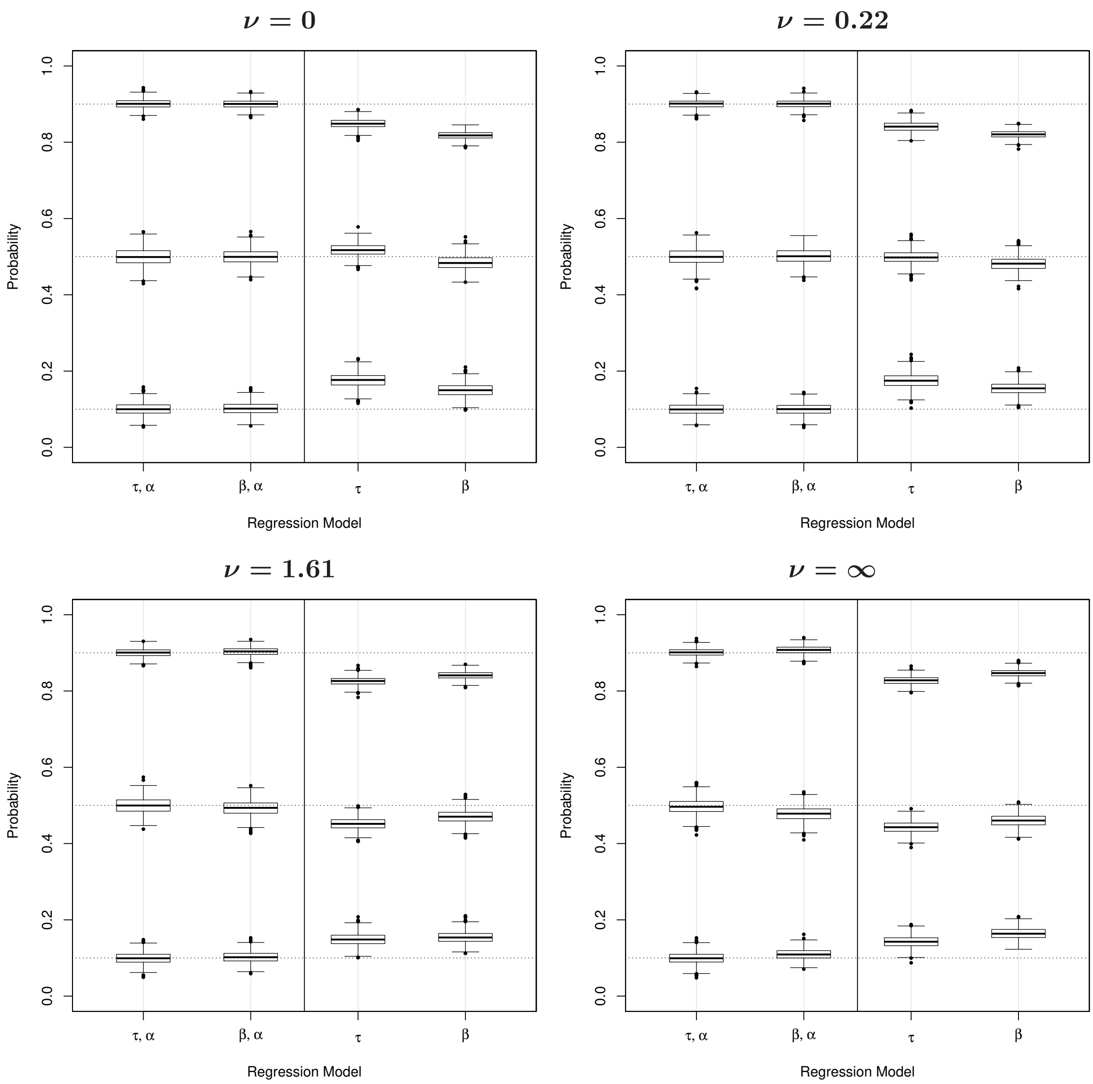

Figure 2: Boxplots of estimated baseline survivor probabilities evaluated at the true 90th, 50th, and 10th percentile times, respectively (such that the true probabilities are 0.1, 0.5, and 0.9), vertically stacked for each of four fitted models indicated by the x-axis labels.

chemotherapy, radiotherapy, and a combined treatment of chemotherapy and radiotherapy. In our analysis we take palliative care (which is a non-curative treatment providing pain relief) as the reference category. Note that, while various other covariates were captured for each individual, the purpose of this section is to briefly illustrate our general modelling framework in the context of the treatment covariate. However, a more detailed covariate analysis (using a Weibull MPR model) is given in Burke and MacKenzie (2017) where strategies for variable selection in MPR models are also discussed.

Within our framework, many choices can be made in terms of the inclusion of covariates, i.e., we may choose to include any combination of the τ, β, α, or ν regression components. Our simulation work (Section 4) suggests the following choice: one scale regression component, τ or β, along with the α shape regression component, while ν_0 characterises the overall distribution without any dependence on covariates. We now investigate this further in the context of the lung cancer data. We consider the 12 models which arise from including treatment in at least one scale regression component, τ and β (note, however, that our simulation work suggests that models with both τ and β components will be highly unstable). These models are summarised in Table 5 and denoted using the notation of Section 3.1.1.

We immediately see that the largest AIC values are associated with the simpler single component (i.e., τ and β only) models which suggests that the models are not sufficiently flexible to capture the more complex non-AFT/non-PH effects observed here. Although, in this particular application, the AFT (τ only) model has a much lower AIC (and BIC) than that of the PH (β only) model, the fit can be improved significantly by modelling shape (α or ν) in addition to scale. However, modelling more than one scale component or more than one shape component does not lead yield further reductions in AIC or BIC, i.e., the additional model complexity is not warranted. Furthermore, these more complex models are quite unstable as evidenced by non-positive definite observed information matrices (specifically, we find negative variance estimates), and very large τ and β coefficients in models with simultaneous τ and β components. It is worth highlighting that two of these more complex models, Model(τ, α, ν) and Model(τ, β, α), actually do have positive definite observed information matrices but we can confirm that the standard errors are large to the extent that nearly all of the estimated effects are not statistically significant within these models.

We now turn to the models which contain both one scale and one shape regression component: $M(\tau, \alpha)$, $M(\beta, \alpha)$, $M(\tau, \nu)$, and $M(\beta, \nu)$. Interestingly, the models with α components have very close AIC values, indicating that the choice of τ or β component is not so important within these already flexible models (in line with the findings of Section 4.2). Note that, while $M(\beta, \nu)$ has the smallest AIC value, it is

Table 5: Summary of models fitted to lung cancer data

Model	Dim.	$\|\tau\|_1$	$\|\beta\|_1$	$\|\alpha\|_1$	$\|\nu\|_1$	Pos.Def.	Δ_{AIC}	Δ_{BIC}
τ	7	5.2	—	—	—	yes	27.2	8.2
β	7	—	4.0	—	—	yes	53.1	34.1
τ, α	11	4.8	—	1.6	—	yes	1.5	1.5
β, α	11	—	11.9	3.2	—	yes	2.2	2.2
τ, ν	11	6.3	—	—	32.6	yes	8.9	8.9
β, ν	11	—	9.6	—	1.8	yes	0.0	0.0
τ, β	12	20.5	22.3	—	—	no	2.1	6.8
τ, α, ν	15	5.2	—	1.7	1.2	yes	7.2	26.2
β, α, ν	15	—	13.4	4.0	0.5	no	4.6	23.6
τ, β, α	16	6.3	7.9	2.6	—	yes	3.1	26.9
τ, β, ν	16	30.5	58.7	—	8.9	no	5.8	29.6
τ, β, α, ν	20	14.9	17.0	2.1	13.6	no	10.1	52.9

Dim., the dimension of the model, i.e., number of parameters; $\|\tau\|_1$, the ℓ_1 norm of the τ regression coefficients ignoring intercept, i.e., $\sum_{i=1}^{4} |\tau_i|$, with similar definitions for $\|\beta\|_1$, $\|\alpha\|_1$, and $\|\nu\|_1$; Pos.Def., statement of whether or not the observed information matrix is positive semidefinite; Δ_{AIC}, the AIC values for each model minus $\mathrm{AIC}_{M(\beta,\nu)} = 3873.8$ (the lowest AIC in the set); Δ_{BIC}, analogous to Δ_{AIC} where the lowest BIC is $\mathrm{BIC}_{M(\beta,\nu)} = 3926.1$.

not much lower than those of the former two models. Finally, $M(\tau, \nu)$ highlights the phenomenon that ν coefficients can grow quite large which can, in more complex multi-factor analyses, cause instability in estimation. On the other hand, within $M(\tau, \alpha)$ and $M(\beta, \alpha)$ there is only one ν parameter, ν_0, which can potentially exhibit this behaviour but this is much more straightforward to handle within estimation routines, and simply indicates the nature of the baseline distribution (i.e., something close to a sub-log-logistic cure distribution, as $\nu_0 \to -\infty$, or a Gompertz distribution, as $\nu_0 \to \infty$).

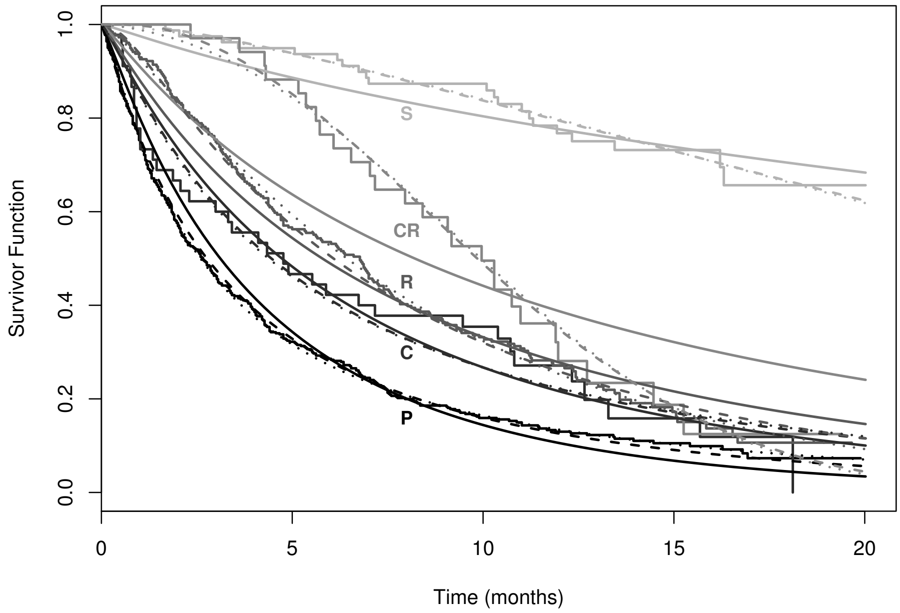

Figure 3: Kaplan-Meier survivor curves (step, solid) for each of the five treatment groups (P = palliative, C = chemotherapy, R = radiotherapy, CR = chemotherapy and radiotherapy, and S = surgery) with fitted curves overlayed for $M(\beta)$ (solid), $M(\beta, \alpha)$ (dashed), and $M(\beta, \nu)$ (dotted).

We now briefly explore model fit and interpretation in the context of the PH-PGW model, $M(\beta)$, and the two associated shape-regression extensions $M(\beta, \alpha)$ and $M(\beta, \nu)$. The advantage, in terms of model fit, of shape regression components is clear from Figure 3, while the $M(\beta, \alpha)$ and $M(\beta, \nu)$ models themselves are virtually indistinguishable. Table 6 displays the estimated regression coefficients. We can see that both $M(\beta)$ and $M(\beta, \alpha)$ suggest a baseline distribution which is between a log-logistic ($\nu = 0$) and a Weibull ($\nu = 0.69$), while $M(\beta, \nu)$ assumes a separate baseline distribution for each treatment group. Interestingly, in all three models, all shape parameters (ν and α) are positive which indicates that the hazards are increasing with time in each treatment group (Table 1). While all three models are in agreement when it comes to the overall effectiveness of each treatment as viewed in terms of the scale coefficients (albeit the chemotherapy effect is only statistically significant in $M(\beta)$), the positive shape coefficients in $M(\beta, \alpha)$ suggest that the effectiveness of each treatment reduces to some extent over time (see Section 3.2) – especially in the case of the combined treatment of chemotherapy and radiotherapy.

Table 6: Fitted model coefficients

	Model(β)		Model(β, α)				Model(β, ν)			
	Scale		Scale		Shape		Scale		Shape	
Intercept	-1.40	(0.08)	-1.13	(0.09)	0.12	(0.07)	-1.04	(0.10)	0.21	(0.06)
Palliative	0.00	—	0.00	—	0.00	—	0.00	—	0.00	—
Surgery	-2.18	(0.23)	-4.77	(0.97)	1.06	(0.28)	-3.96	(0.66)	0.55	(0.15)
Chemo	-0.38	(0.17)	-0.55	(0.33)	0.13	(0.18)	-0.60	(0.36)	0.11	(0.13)
Radio	-0.56	(0.09)	-1.46	(0.21)	0.52	(0.11)	-1.48	(0.19)	0.36	(0.06)
C+R	-0.86	(0.20)	-5.13	(0.96)	1.50	(0.22)	-3.57	(0.60)	0.82	(0.13)
$\hat{\alpha}_0$	0.15	(0.08)	$*$				0.27	(0.07)		
$\hat{\nu}_0$	0.46	(0.06)	0.35	(0.05)			$*$			

The $*$ symbol indicates that the shape parameter already appears as the intercept in the shape regression component.

The hazard ratios for these models are shown in Figure 4 where those of $M(\beta, \alpha)$ and $M(\beta, \nu)$ are clearly similar. They suggest that while the various treatments reduce the hazard in the first few months, their effect is weakened over time and, perhaps, even become inferior to palliative care in the longer term (however, note that very few individuals remain in the sample beyond 15 months). Clearly SPR models, such as $M(\beta)$, cannot account for covariate effects of this sort.

It is worth highlighting the fact that the basic findings here are qualitatively similar to those of Burke and MacKenzie (2017) who analysed this lung cancer dataset using $M_{\kappa=1}(\beta, \alpha)$, i.e., a Weibull MPR model. However, the framework of the current paper permits us to consider a much wider range of model structures and distributions in which $M_{\kappa=1}(\beta, \alpha)$ appears as a special case. In particular, $M(\beta, \alpha)$ from Table 6 yields a 95% confidence interval for κ, $[0.28, 0.57]$, which does not support the Weibull ($\kappa = 1$) baseline distribution. Furthermore, $\text{AIC}_{M_{\kappa=1}(\beta,\alpha)} - \text{AIC}_{M(\beta,\alpha)} = 57.6$, and we can confirm that the improvement in quality of fit is most evident in the palliative care group (which $M_{\kappa=1}(\beta, \alpha)$ does not capture so well). Thus, although the basic findings are unaltered in this particular application, the PGW MPR approach yields a better solution in which uncertainty in selecting the baseline distribution is accounted for. Of course, the PGW MPR model will readily adapt to other applications which might differ significantly (both qualitatively and quantitatively) from $M_{\kappa=1}(\beta, \alpha)$.

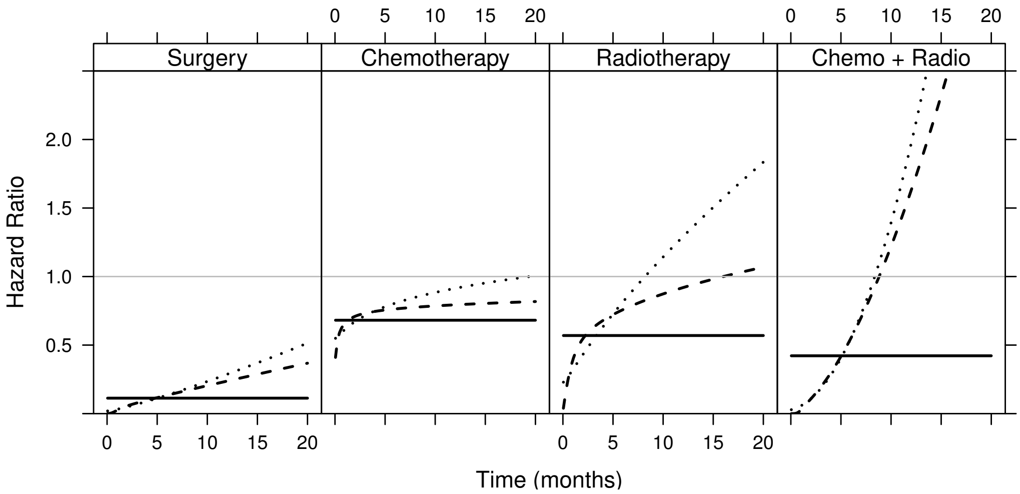

Figure 4: Hazard ratios for each treatment relative to palliative care for $M(\beta)$ (solid), $M(\beta, \alpha)$ (dashed), and $M(\beta, \nu)$ (dotted). Line of equality (grey) also shown.

6 Discussion

Our proposed PGW-MPR modelling framework appears to be highly flexible and should adapt readily to a wide variety of applications in survival analysis and reliability. In particular, this framework includes the practically important AFT and PH models, and generalises them through shape regression components. Furthermore, the adapted PGW baseline model covers the primary shapes of hazard function (constant, increasing, decreasing, up-then-down, down-then-up) within some of the most popular survival distributions (log-logistic, Burr type XII, Weibull, Gompertz) using only two shape parameters.

In practice, the full four-component PGW-MPR model is likely to be more flexible than is required for most purposes. However, the work of this paper suggests that covariates should appear via just one scale-type component (τ or β), along with the α shape component which permits survivor functions with differing shapes and indicates departures from more basic AFT or PH effects, while ν is a covariate-independent parameter which allows us to choose among distributions within one unified framework. Interestingly, we have found that the scale-type parameters (τ and β) are highly intertwined in the sense that they cannot be estimated simultaneously within the same model reliably, and are highly correlated. Furthermore, we find that the AFT model can approximate the PH model reasonably well (and vice versa).

This is true across the full range of distributions (varying ν), going well beyond the well-known Weibull case in which the two model structures are equivalent. The implication of this is that, in terms of performance gain, the movement from AFT to PH modelling (or vice versa) might not be very large, whereas we have found that modelling the shape is a more fruitful alteration to the regression specification.

Finally, the perspective of this paper has been to investigate survival modelling generally, to cover some of the most popular models, and to discover some of the better modelling choices that can be made within this framework. Although we have developed these ideas in a fully parametric context, non-parametric equivalents, while possible, are beyond the scope of the present paper (but are investigated in a separate line of work (Burke et al., 2017)). However, it is worth highlighting that perhaps too much emphasis is placed on non-/semi-parametric approaches in survival analysis whereby undue weight is attached to the flexibility of the baseline distribution in comparison to that of the flexibility of its regression structure. Indeed, it can even be that the manifestation of complex baseline distributions is the result of a structurally misspecified model. Our general approach to survival modelling provides a framework within which one can consider the most important components of survival modelling (including which might potentially be modelled non-parametrically), and we believe that this kind of insight can lead to better modelling practice in general.

Acknowledgements

This research was supported by the Irish Research Council (New Foundations Award).

References

Aalen, O. (1992). Modelling heterogeneity in survival analysis by the compound poisson distribution. *Ann. Appl. Probab.*, 2:951–972.

Bagdonaviçius, V. and Nikulin, M. (2002). *Accelerated Life Model; Modeling and Statistical Analysis*. Chapman & Hall/CRC, Boca Raton, FL.

Box, G. and Cox, D. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B*, 26:211–252.

Burke, K., Eriksson, F., and Pipper, C. B. (2017). Semiparametric multi-parameter regression survival modelling. Submitted.

Burke, K. and MacKenzie, G. (2017). Multi-parameter regression survival modelling: An alternative to proportional hazards. *Biometrics*, 73:678–686.

Chen, Y. and Jewell, N. (2001). On a general class of semiparametric hazards regression models. *Biometrika*, 88:687–702.

Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statist. Probab. Lett.*, 49:155–161.

Cox, C. and Matheson, M. (2014). A comparison of the generalized gamma and exponentiated weibull distributions. *Statist. Med.*, 33:3772–3780.

Dimitrakopoulou, T., Adamidis, K., and Loukas, S. (2007). A lifetime distribution with an upside-down bathtub-shaped hazard function. *IEEE Trans. Reliab.*, 56:308–311.

Fischer, M. and Jakob, K. (2016). ptas distributions with application to risk management. *J. Statist. Distributions Applic.*, 3:1–18.

Gupta, P. and Gupta, R. (1996). Ageing characteristics of the weibull mixtures. *Prob. Eng. Info. Sci.*, 10:591–600.

Hougaard, P. (1986). Survival models for hetergeneous populations derived from stable distributions. *Biometrika*, 73:387–396.

Jones, M. and Noufaily, A. (2015). Log-location-scale-log-concave distributions for survival and reliability analysis. *Elec. J. Statist.*, 9:2732–2750.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, Hoboken, NJ., 2 edition.

Matheson, M., Muñoz, A., and Cox, C. (2017). Describing the flexibility of the generalized gamma and related distributions. *J. Statist. Distributions Appl.*, 4:Article 15.

Nadarajah, S. and Haghighi, F. (2011). An extension of the exponential distribution. *Statist.*, 45:543–558.

Nikulin, M. and Haghighi, F. (2009). On the power generalized weibull family: model for cancer censored data. *Metron*, 67:75–86.

Tsodikov, A., Ibrahim, J., and Yakovlev, A. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *J. Amer. Statist. Assoc.*, 98:1063–1078.

Tweedie, M. (1984). An index which distinguishes between some important exponential families. In Ghosh, J. and Roy, J., editors, *Statistics: Applications and New directions*, pages 579–604.

Xie, M., Tang, Y., and Goh, T. (2002). A modified weibull extension with bathtub-shaped failure rate function. *Reliab. Eng. Sys. Safety*, 73:279–285.

Yeo, I. and Johnson, R. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87:954–959.

Zeng, Z. and Lin, D. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *J. Roy. Statist. Soc. Ser. B*, 69:507–564.

Frailty Links for Adapted PGW Distribution

Write APGW(κ, λ) for the adapted PGW distribution with proportionality parameter $\lambda > 0$ and shape parameter κ i.e. with c.h.f. $\lambda H_A(t^\gamma; \kappa)$. Also, write the three-parameter version of the TS distribution as $TS(\omega, \xi, \theta)$, $0 \leq \omega \leq 1$, $\xi > 0$, $\theta \geq 0$, having Laplace transform

$$\mathcal{L}^H_{\omega,\xi,\theta}(s) = \exp\left[-\frac{\xi}{\omega}\left\{(\theta + s)^\omega - \theta^\omega\right\}\right]$$

and density $g^H_{\omega,\xi,\theta}$.

Result A1. Let $T|B = b \sim$ APGW(κ, b), $\kappa > 0$, and let $B \sim$ TS $\left(\omega, \frac{\kappa^{\omega-1}(\omega\kappa+1)}{(\kappa+1)^\omega}, \frac{\kappa+1}{\kappa}\right)$. Then, $aT \sim$ APGW$(\omega\kappa, 1)$ where $a = \{(\kappa + 1)/(\omega\kappa + 1)\}^{1/\gamma}$.

`Proof`

$$\begin{aligned}
S_T(t) &= \int_0^\infty \exp\{-bH_A(t^\gamma; \kappa)\} g^H_{\omega, \frac{\kappa^{\omega-1}(\omega\kappa+1)}{(\kappa+1)^\omega}, \frac{\kappa+1}{\kappa}}(b)\, db \\
&= \mathcal{L}^H_{\omega, \frac{\kappa^{\omega-1}(\omega\kappa+1)}{(\kappa+1)^\omega}, \frac{\kappa+1}{\kappa}}\{H_A(t^\gamma; \kappa)\} \\
&= \exp\left[-\frac{\kappa^{\omega-1}(\omega\kappa + 1)}{\omega(\kappa + 1)^\omega}\right. \\
&\qquad \left.\times\left\{\left(\frac{\kappa+1}{\kappa} + \frac{\kappa+1}{\kappa}\left\{\left(1 + \frac{t^\gamma}{\kappa+1}\right)^\kappa - 1\right\}\right)^\omega - \left(\frac{\kappa+1}{\kappa}\right)^\omega\right\}\right] \\
&= \exp\left[-\frac{(\omega\kappa + 1)}{\omega\kappa}\left\{\left(1 + \frac{t^\gamma}{\kappa+1}\right)^{\omega\kappa} - 1\right\}\right] = \exp\left[-H_A\left\{\left(\frac{t}{a}\right)^\gamma; \omega\kappa\right\}\right]. \quad \square
\end{aligned}$$

Now write ATS(ρ, ω, ξ), $\rho > 1$, $\omega > 0$, $\xi \geq 0$, for Aalen's (1992) cure model adaptation of the TS distribution having Laplace transform

$$\mathcal{L}^A_{\rho,\omega,\xi}(s) = \exp\left[-\frac{\rho}{(1-\rho)\xi}\left\{1 - \left(1 + \frac{\xi\omega}{\rho}s\right)^{1-\rho}\right\}\right]$$

and density $g^A_{\rho,\omega,\xi}$.

Result A2. Let $T|B = b \sim$ APGW(κ, b), $\kappa > 0$, and, for $0 < L < 1$, let

$$B \sim \text{ATS}\left(1 + \tfrac{L}{\kappa}, \tfrac{\kappa+L}{1-L}, \tfrac{1-L}{\kappa+1}\right).$$

Then, $cT \sim$ APGW$(-L, 1)$ where $c = \{(\kappa + 1)/(1 - L)\}^{1/\gamma}$.

`Proof`

$$
\begin{aligned}
S_T(t) &= \int_0^\infty \exp\{-bH_A(t^\gamma;\kappa)\} g^A_{1+\frac{L}{\kappa},\frac{\kappa+L}{1-L},\frac{1-L}{\kappa+1}}(b)\,db \\
&= \mathcal{L}^A_{1+\frac{L}{\kappa},\frac{\kappa+L}{1-L},\frac{1-L}{\kappa+1}}\{H_A(t^\gamma;\kappa)\} \\
&= \exp\left(-\frac{L-1}{L}\left[1-\left\{1+\left(1+\frac{t^\gamma}{\kappa+1}\right)^\kappa-1\right\}^{-L/\kappa}\right]\right) \\
&= \exp\left[-\frac{1-L}{L}\left\{\left(1+\frac{t^\gamma}{\kappa+1}\right)^{-L}-1\right\}\right] = \exp\left[-H_A\left\{\left(\frac{t}{c}\right)^\gamma;-L\right\}\right]. \quad \square
\end{aligned}
$$

Simulation Studies with Smaller Samples

This section displays analogous results to those of Table 4 but for $n = 500$ and $n = 100$ respectively.

Table 7: Median and standard deviation (in brackets) of estimates when $n = 500$

			Model(τ, β, α)				
ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$
0.00	1.06 (1.57)	0.57 (1.63)	-0.25 (1.73)	0.01 (1.13)	0.23 (0.21)	-0.49 (0.32)	0.07 (1.02)
0.22	1.30 (1.62)	0.45 (2.66)	-0.52 (1.95)	0.21 (2.13)	0.27 (0.24)	-0.53 (0.37)	0.27 (0.73)
0.41	1.05 (1.73)	0.81 (3.17)	-0.19 (2.11)	-0.03 (2.72)	0.27 (0.24)	-0.50 (0.34)	0.35 (0.78)
0.69	0.88 (2.12)	0.45 (4.19)	-0.15 (2.60)	0.19 (3.70)	0.21 (0.26)	-0.51 (0.28)	0.65 (4.26)
1.10	0.43 (1.51)	0.26 (2.93)	0.33 (1.91)	0.04 (2.72)	0.15 (0.16)	-0.49 (0.23)	1.53 (7.11)
1.61	0.53 (1.09)	0.40 (1.98)	0.25 (1.47)	-0.01 (2.03)	0.19 (0.16)	-0.50 (0.24)	14.75 (7.06)
∞	0.91 (0.72)	0.68 (1.26)	-0.15 (1.07)	-0.02 (1.41)	0.19 (0.16)	-0.51 (0.24)	15.43 (6.48)
			Model(τ, α)				
ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$
0.00	0.79 (0.13)	0.60 (0.19)	0.00 —	0.00 —	0.20 (0.09)	-0.50 (0.09)	0.00 (0.09)
0.22	0.80 (0.12)	0.59 (0.16)	0.00 —	0.00 —	0.20 (0.09)	-0.50 (0.09)	0.22 (0.13)
0.41	0.79 (0.12)	0.58 (0.15)	0.00 —	0.00 —	0.20 (0.09)	-0.50 (0.08)	0.42 (0.18)
0.69	0.80 (0.13)	0.58 (0.14)	0.00 —	0.00 —	0.20 (0.09)	-0.50 (0.08)	0.72 (0.30)
1.10	0.79 (0.13)	0.60 (0.12)	0.00 —	0.00 —	0.20 (0.09)	-0.51 (0.09)	1.13 (1.73)
1.61	0.80 (0.11)	0.60 (0.11)	0.00 —	0.00 —	0.20 (0.09)	-0.50 (0.09)	1.62 (4.26)
∞	0.84 (0.07)	0.62 (0.08)	0.00 —	0.00 —	0.24 (0.08)	-0.51 (0.09)	13.84 (6.97)
			Model(β, α)				
ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$
0.00	0.00 —	0.00 —	0.88 (0.16)	0.03 (0.12)	0.17 (0.08)	-0.51 (0.08)	-0.35 (0.18)
0.22	0.00 —	0.00 —	0.91 (0.17)	0.04 (0.11)	0.19 (0.08)	-0.51 (0.08)	-0.07 (0.23)
0.41	0.00 —	0.00 —	0.94 (0.18)	0.04 (0.12)	0.19 (0.08)	-0.51 (0.08)	0.22 (0.30)
0.69	0.00 —	0.00 —	0.98 (0.20)	0.07 (0.14)	0.20 (0.09)	-0.50 (0.09)	0.72 (0.94)
1.10	0.00 —	0.00 —	1.04 (0.17)	0.08 (0.15)	0.22 (0.08)	-0.50 (0.09)	1.74 (5.10)
1.61	0.00 —	0.00 —	1.20 (0.15)	0.09 (0.17)	0.27 (0.07)	-0.49 (0.10)	14.00 (7.03)
∞	0.00 —	0.00 —	1.54 (0.15)	0.12 (0.20)	0.37 (0.08)	-0.49 (0.10)	16.85 (1.76)

All numbers are rounded to two decimal places. For the models with fixed parameters, the "estimated" value shown is the value at which the parameter is fixed, and its standard error is then indicated by "—".

Table 8: Median and standard deviation (in brackets) of estimates when $n = 100$

ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$
				Model(τ, β, α)			
0.00	1.32 (2.72)	0.88 (2.91)	-0.49 (3.18)	-0.01 (1.97)	0.39 (0.43)	-0.52 (0.63)	0.08 (2.53)
0.22	1.29 (2.21)	0.85 (3.20)	-0.39 (2.74)	0.01 (2.39)	0.38 (0.42)	-0.49 (0.56)	0.17 (3.18)
0.41	1.10 (1.85)	0.83 (3.21)	-0.22 (2.24)	0.02 (2.73)	0.35 (0.44)	-0.51 (0.52)	0.24 (3.41)
0.69	0.87 (1.94)	0.67 (3.47)	-0.14 (2.48)	-0.03 (3.12)	0.21 (0.45)	-0.49 (0.46)	0.87 (6.34)
1.10	0.61 (1.52)	0.50 (2.84)	0.05 (1.96)	-0.02 (2.69)	0.17 (0.33)	-0.52 (0.37)	12.66 (7.62)
1.61	0.70 (1.22)	0.57 (2.46)	0.05 (1.72)	0.05 (2.52)	0.17 (0.28)	-0.52 (0.37)	15.31 (7.02)
∞	0.89 (0.89)	0.75 (1.81)	-0.16 (1.44)	-0.14 (2.05)	0.23 (0.26)	-0.53 (0.34)	15.95 (6.12)
				Model(τ, α)			
ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$
0.00	0.79 (0.29)	0.59 (0.44)	0.00 —	0.00 —	0.24 (0.20)	-0.50 (0.20)	-0.02 (0.27)
0.22	0.79 (0.28)	0.58 (0.39)	0.00 —	0.00 —	0.21 (0.20)	-0.50 (0.21)	0.22 (1.47)
0.41	0.76 (0.27)	0.58 (0.35)	0.00 —	0.00 —	0.21 (0.20)	-0.51 (0.20)	0.45 (1.93)
0.69	0.79 (0.26)	0.58 (0.31)	0.00 —	0.00 —	0.22 (0.20)	-0.51 (0.20)	0.72 (3.68)
1.10	0.82 (0.24)	0.59 (0.25)	0.00 —	0.00 —	0.24 (0.20)	-0.51 (0.20)	1.07 (5.51)
1.61	0.83 (0.20)	0.63 (0.24)	0.00 —	0.00 —	0.26 (0.19)	-0.52 (0.19)	1.54 (6.86)
∞	0.89 (0.16)	0.66 (0.19)	0.00 —	0.00 —	0.30 (0.18)	-0.50 (0.21)	4.00 (7.52)
				Model(β, α)			
ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$
0.00	0.00 —	0.00 —	0.89 (0.36)	0.04 (0.25)	0.19 (0.17)	-0.51 (0.17)	-0.34 (0.44)
0.22	0.00 —	0.00 —	0.89 (0.38)	0.07 (0.28)	0.19 (0.18)	-0.50 (0.18)	0.01 (0.79)
0.41	0.00 —	0.00 —	0.92 (0.38)	0.03 (0.30)	0.19 (0.19)	-0.51 (0.19)	0.24 (1.99)
0.69	0.00 —	0.00 —	1.04 (0.38)	0.06 (0.33)	0.23 (0.17)	-0.51 (0.19)	0.70 (3.81)
1.10	0.00 —	0.00 —	1.21 (0.35)	0.07 (0.33)	0.29 (0.16)	-0.50 (0.20)	1.27 (5.92)
1.61	0.00 —	0.00 —	1.35 (0.36)	0.09 (0.38)	0.34 (0.17)	-0.50 (0.21)	2.32 (7.23)
∞	0.00 —	0.00 —	1.66 (0.36)	0.13 (0.43)	0.42 (0.17)	-0.50 (0.22)	15.98 (6.32)

All numbers are rounded to two decimal places. For the models with fixed parameters, the "estimated" value shown is the value at which the parameter is fixed, and its standard error is then indicated by "—".