




Article

Bivariate Proportional Hazard Models: Structure and Inference

Barry C. Arnold ¹, Guillermo Martínez-Flórez ² and Héctor W. Gómez ^{3,*}

- ¹ Statistics Department, University of California Riverside, Riverside, CA 92521, USA
² Departamento de Matemáticas y Estadística, Facultad de Ciencias, Universidad de Córdoba, Córdoba 2300, Colombia
³ Departamento de Matemática, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile
* Correspondence: hector.gomez@uantof.cl

Abstract: We focus on a variety of bivariate models with proportional hazard components. Models with proportional hazard marginals are described together with a selection of models with proportional hazard conditional distributions. The bivariate distributions with marginal proportional hazards distributions are shown to be closely related to certain known bivariate exponential models. Two distinct kinds of conditional specification are investigated. Discussion is provided of cases with hazard function components that are (1) completely unknown, (2) known to belong to given parametric families and (3) completely known. Since the models are designed for use with survival data, it is inevitable that the marginal and conditional distributions will be asymmetric. However, logarithmic transformations in some cases will result in symmetric component distributions.

Keywords: bivariate exponential models; conditional specification; gumbel distribution; dependent lifetime data; functional equation



Citation: Arnold, B.C.; Martínez-Flórez, G.; Gómez, H.W. Bivariate Proportional Hazard Models: Structure and Inference. *Symmetry* **2022**, *14*, 2073. <https://doi.org/10.3390/sym14102073>

Academic Editor: Dário Ferreira

Received: 29 August 2022

Accepted: 29 September 2022

Published: 6 October 2022

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



Copyright: © 2022 by the authors. 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/>).

1. Introduction

Survival models involving families of densities with proportional hazard functions have proved to be useful for analyzing many lifetime data sets. Not infrequently bivariate survival data (involving related lifetimes) need to be analyzed. In this paper, we review several methods for generating suitable bivariate models for such situations. The key observation in the development is that proportional hazard models can be viewed as ones obtained via monotone transformations applied to exponential models. In the latter sections of the paper, related statistical inference issues are discussed. Since the models are designed for use with survival data, it is inevitable that the marginal and conditional distributions will be asymmetric. However, logarithmic transformations in some cases will result in symmetric component distributions.

2. Bivariate Distributions with Proportional Hazard Marginals

Let F_1 and F_2 be two absolutely continuous distribution functions with support $(0, \infty)$ and with corresponding densities f_1 and f_2 and hazard functions h_1 and h_2 (where $h_i = f_i / (1 - F_i)$, $i = 1, 2$).

We will say that (X_1, X_2) has a bivariate marginal proportional hazard distribution associated with F_1 and F_2 and with parameters $\alpha_1, \alpha_2 > 0$ if for $i = 1, 2$

$$f_{X_i}(x_i; F_i, \alpha_i) = \alpha_i [1 - F_i(x_i)]^{\alpha_i - 1} f_i(x_i) I(x_i > 0), \quad (1)$$

and we will write $(X_1, X_2) \sim PHM(F_1, \alpha_1; F_2, \alpha_2)$ and also $X_i \sim PH(F_i, \alpha_i)$ $i = 1, 2$ (here *PHM* is an acronym for proportional hazard marginals). Note that the name is appropriate since for $i = 1, 2$

$$h_{X_i}(x_i; F_i, \alpha_i) = \alpha_i h_i(x_i). \quad (2)$$

Observe that if $(X_1, X_2) \sim PHM(F_1, \alpha_1; F_2, \alpha_2)$ then the X_i 's admit a representation of the form

$$X_i = F_i^{-1}(1 - e^{-Y_i}), \quad i = 1, 2, \quad (3)$$

where for each i , $Y_i \sim \exp(\alpha_i)$ (i.e., $f_{Y_i}(y_i) = \alpha_i e^{-\alpha_i y_i} I(y_i > 0)$).

Note that the transformations in (3) are monotone increasing.

In fact, it is not necessary to build the model with reference to two distribution functions F_1 and F_2 . Instead we can begin with $Y_i \sim \exp(\alpha_i)$, $i = 1, 2$ and use two monotone increasing functions $g_i : (0, \infty) \rightarrow (0, \infty)$ to define $X_i = g_i(Y_i)$, $i = 1, 2$. If we denote the corresponding inverse functions by $h_i(x) = g_i^{-1}(x)$, then it is readily verified that each X_i has a proportional hazard distribution, i.e., that for $i = 1, 2$,

$$h_{X_i}(x_i; g_i, \alpha_i) = \alpha_i h_i'(x_i), \quad x_i > 0. \quad (4)$$

It is however customary to use the representation (3) involving the two distribution functions F_1 and F_2 , and we will adhere to this convention.

Of course, in the representation (3), (Y_1, Y_2) can have any bivariate exponential distribution that we wish to utilize. Popular choices of bivariate exponential distributions involving few additional parameters include:

(i) Gumbel Type I distribution, with

$$P(Y_1 > y_1, Y_2 > y_2) = \exp[-\alpha_1 y_1 - \alpha_2 y_2 - \delta \alpha_1 \alpha_2 y_1 y_2], \quad 0 \leq \delta \leq 1.$$

(ii) Gumbel Type II distribution, with

$$P(Y_1 \leq y_1, Y_2 \leq y_2) = [1 - e^{-\alpha_1 y_1}][1 - e^{-\alpha_2 y_2}][1 + \delta e^{-\alpha_1 y_1 - \alpha_2 y_2}], \quad \delta \in [-1, 1].$$

(iii) Marshall–Olkin distribution (see Marshall and Olkin [1]), with

$$P(Y_1 > y_1, Y_2 > y_2) = \exp[-\alpha_1 y_1 - \alpha_2 y_2 - \delta \alpha_1 \alpha_2 \max(y_1, y_2)], \quad \delta \in [0, \infty).$$

Many other choices are possible, see for example Kotz et al. ([2], pp. 350–385).

For a specific example, if we choose (Y_1, Y_2) to have a Gumbel Type I density, i.e.,

$$f_{Y_1, Y_2}(y_1, y_2) = \alpha_1 \alpha_2 [(1 + \delta y_1)(1 + \delta y_2) - \delta] \exp\{-\alpha_1 y_1 - \alpha_2 y_2 - \alpha_1 \alpha_2 \delta y_1 y_2\}, \quad (5)$$

and use $F_1(x_1) = x_1^{\gamma_1}$, $0 < x_1 < 1$ and $F_2(x_2) = x_2^{\gamma_2}$, $0 < x_2 < 1$, then the resulting PHM density will be of the form:

$$f_{X_1, X_2}(x_1, x_2) = \alpha_1 \alpha_2 \gamma_1 \gamma_2 [(1 - \alpha_1 \log(1 - x_1^{\gamma_1}))(1 - \alpha_2 \log(1 - x_2^{\gamma_2})) - \delta] x_1^{\gamma_1} x_2^{\gamma_2} \times \exp\{-\alpha_1 \alpha_2 \delta \log(1 - x_1^{\gamma_1}) \log(1 - x_2^{\gamma_2})\}.$$

In application of such models, it is frequently desirable to postulate that each F_i is a member of some parametric family of distributions to add flexibility to the model. The dependence structure will however be completely determined by the copula of the particular bivariate exponential distribution used in the construction.

Alternatively, one could “let the data tell us which F_i 's to use in the model”. Thus, we would seek monotone marginal transformations that will make the transformed marginal sample distributions look as much like exponential distributions as possible. This semi-parametric approach will be returned to later in the paper.

3. Bivariate Distributions with Proportional Hazard Conditionals

We will consider two types of conditioning. The first kind is quite traditional in that we consider the distribution of one variable given that a second variable takes on a particular value. The second kind involves conditioning on the event that the second variable is larger than a particular value.

3.1. The First Kind

Consider two proportional hazard families of densities as in (1). Recall that we write $X_i \sim PH(F_i, \alpha_i)$ if the corresponding densities are given by (1).

For this conditional proportional hazards paradigm, we seek to identify joint distributions for (X_1, X_2) with all conditional densities of the forms in (1). Thus, for each $x_2 > 0$ we wish to have

$$X_1|X_2 = x_2 \sim PH(F_1, \alpha_1(x_2)), \tag{6}$$

and for each $x_1 > 0$,

$$X_2|X_1 = x_1 \sim PH(F_2, \alpha_2(x_1)), \tag{7}$$

for some functions $\alpha_1(x_2)$ and $\alpha_2(x_1)$. It is not difficult to verify that this will be the case if and only if

$$X_i = F_i^{-1}(1 - e^{-Y_i}), \quad i = 1, 2, \tag{8}$$

where (Y_1, Y_2) has a joint distribution with exponential conditionals, i.e., such that for each $y_2 > 0$

$$Y_1|Y_2 = y_2 \sim \exp(\alpha_1(y_2)), \tag{9}$$

and for each $y_1 > 0$

$$Y_2|Y_1 = y_1 \sim \exp(\alpha_2(y_1)). \tag{10}$$

The class of all densities with such exponential conditionals is identified in Arnold and Strauss [3] and is of the form:

$$f_{Y_1, Y_2}(y_1, y_2) = k(\delta)\alpha_1\alpha_2 \exp[-\alpha_1y_1 - \alpha_2y_2 - \alpha_1\alpha_2\delta y_1y_2]I(y_1 > 0, y_2 > 0), \tag{11}$$

where $\alpha_1 > 0$, $\alpha_2 > 0$ and $\delta \geq 0$. The normalizing constant, $k(\delta)$, in (10) can be expressed in terms of the exponential integral function. Thus

$$k(\delta) = \frac{\delta \exp\{-1/\delta\}}{\int_{1/\delta}^{\infty} \frac{\exp\{-w\}}{w} dw}. \tag{12}$$

Consequently, bivariate densities with conditionals of the proportional hazard form will be given by

$$f_{X_1, X_2}(x_1, x_2) = \alpha_1\alpha_2k(\delta)f_1(x_1)[1 - F_1(x_1)]^{\alpha_1-1}f_2(x_2)[1 - F_2(x_2)]^{\alpha_2-1} \times \exp\{-\alpha_1\alpha_2\delta \log[1 - F_1(x_1)] \log[1 - F_2(x_2)]\}. \tag{13}$$

This model is discussed in Arnold and Kim [4] and we will follow their nomenclature and call it a proportional hazard conditionals model of the first kind. If (X_1, X_2) has a density of the form (13), we will write

$$(X_1, X_2) \sim PHC(I)(F_1, \alpha_1; F_2, \alpha_2; \delta).$$

3.2. The Second Kind

Again consider two proportional hazard families of densities as in (1). Recall that we write $X_i \sim PH(F_i, \alpha_i)$ if the corresponding densities are given by (1). The second kind of conditional model (also introduced in Arnold and Kim [4]) involves conditioning on events of the form $\{X_1 > x_1\}$ and $\{X_2 > x_2\}$. Thus, we seek to identify joint survival functions for (X_1, X_2) such that for each $x_2 > 0$,

$$X_1|\{X_2 > x_2\} \sim PH(F_1, \alpha_1(x_2)), \tag{14}$$

and for each $x_1 > 0$,

$$X_2|\{X_1 > x_1\} \sim PH(F_2, \alpha_2(x_1)), \tag{15}$$

for some functions $\alpha_1(x_2)$ and $\alpha_2(x_1)$.

To analyze this situation (since F_1 and F_2 are known) it is again convenient to write

$$X_i = F_i^{-1}(1 - e^{-Y_i}), \quad i = 1, 2, \quad (16)$$

where the Y_i 's are exponential random variables.

The conditions (14) and (15) are then equivalent to the statements

$$Y_1 | \{Y_2 > y_2\} \sim \exp(\alpha_1(y_2)), \quad (17)$$

and

$$Y_2 | \{Y_1 > y_1\} \sim \exp(\alpha_2(y_1)). \quad (18)$$

Denote the survival functions of Y_1 and Y_2 by $\psi_1(y_1) = P(Y_1 > y_1)$ and $\psi_2(y_2) = P(Y_2 > y_2)$. It then follows that

$$\psi_2(y_2)e^{-\alpha_1(y_2)y_1} = P(Y_1 > y_1, Y_2 > y_2) = \psi_1(y_1)e^{-\alpha_2(y_1)y_2}. \quad (19)$$

Taking logarithms we have:

$$\log \psi_2(y_2) - \alpha_1(y_2)y_1 = \log \psi_1(y_1) - \alpha_2(y_1)y_2. \quad (20)$$

This is a Stephanos–Levi–Civita–Suto functional Equation (see Arnold et al. [5], p. 13) which is readily solved to yield the following expression for the joint survival function of (Y_1, Y_2) :

$$P(Y_1 > y_1, Y_2 > y_2) = \exp[-\alpha_1 y_1 - \alpha_2 y_2 - \alpha_1 \alpha_2 \delta y_1 y_2] \quad (21)$$

for $y_1 > 0, y_2 > 0$, where $\alpha_1 > 0, \alpha_2 > 0$ and $0 \leq \delta \leq 1$. This is recognizable as Gumbel's Type I bivariate exponential distribution (with exponential marginals). From Equation (21) we obtain the joint survival function of (X_1, X_2) in the form:

$$P(X_1 > x_1, X_2 > x_2) = [1 - F_1(x_1)]^{\alpha_1} [1 - F_2(x_2)]^{\alpha_2} \times \exp\{-\alpha_1 \alpha_2 \delta \log[1 - F_1(x_1)] \log[1 - F_2(x_2)]\}. \quad (22)$$

Then, the joint cumulative distribution function is

$$F(x_1, x_2) = [1 - F_1(x_1)]^{\alpha_1} [1 - F_2(x_2)]^{\alpha_2} \exp\{-\alpha_1 \alpha_2 \delta \log[1 - F_1(x_1)] \log[1 - F_2(x_2)]\} + [1 - (1 - F_1(x_1))^{\alpha_1}] + [1 - (1 - F_2(x_2))^{\alpha_2}] - 1. \quad (23)$$

and the joint density function is

$$f_{X_1, X_2}(x_1, x_2) = \alpha_1 \alpha_2 f_1(x_1) [1 - F_1(x_1)]^{\alpha_1 - 1} f_2(x_2) [1 - F_2(x_2)]^{\alpha_2 - 1} \times \exp\{-\alpha_1 \alpha_2 \delta \log[1 - F_1(x_1)] \log[1 - F_2(x_2)]\} \times [(1 - \alpha_1 \delta \log[1 - F_1(x_1)])(1 - \alpha_2 \delta \log[1 - F_2(x_2)]) - \delta]. \quad (24)$$

The vector (Y_1, Y_2) with density (21) has exponential marginals, i.e., $Y_i \sim \exp(\alpha_i)$, for $i = 1, 2$, and thus $X_i \sim PH(F_i, \alpha_i)$ for $i = 1, 2$. Consequently, for $j, j' = 1, 2$ the conditional densities are given by

$$f_{X_j | X_{j'}}(x_j) = \alpha_j f_j(x_j) [1 - F_j(x_j)]^{\alpha_j - 1} \exp\{-\alpha_j \alpha_{j'} \delta \log[1 - F_j(x_j)] \log[1 - F_{j'}(x_{j'})]\} \times [(1 - \alpha_j \delta \log[1 - F_j(x_j)])(1 - \alpha_{j'} \delta \log[1 - F_{j'}(x_{j'})]) - \delta]. \quad (25)$$

If (X_1, X_2) has a survival function of the form (22), we write

$$(X_1, X_2) \sim PHC(II)(F_1, \alpha_1; F_2, \alpha_2; \delta),$$

note that X_1 and X_2 in (22) will be independent if and only if $\delta = 0$.

4. If F_1 and F_2 Are Known

Suppose that we have available a sample of size n , $(X_{1,j}, X_{2,j})$, $j = 1, 2, \dots, n$ from one of the bivariate proportional hazard models discussed in this paper. Since F_1 and F_2 are known, it is appropriate to transform the data to obtain

$$Y_{i,j} = -\log(1 - F_i(X_{i,j})), \quad i = 1, 2, \quad j = 1, 2, \dots, n,$$

and thus to have a sample $(Y_{1,j}, Y_{2,j})$, $j = 1, 2, \dots, n$ from the corresponding well-known bivariate exponential distribution. See the following references for appropriate estimation strategies for these bivariate exponential data sets:

- Gumbel [6],
- Besag [7],
- Arnold and Strauss ([3,8]),
- Castillo and Hadi [9],
- Arnold et al. [5].

5. If F_1 and F_2 are Known to Belong to Some Given Parametric Families

We will illustrate this with a particular example. Other examples may be treated in analogous fashion. Suppose that in the PHC(II) model, (23), we replace $F_1(x_1)$ by $F_1(x_1; \theta)$ and $F_2(x_2)$ by $F_2(x_2; \tau)$, where the parameters θ and τ are unknown. In this case, the model becomes more complicated, but we can still envision success in estimating all the parameters in the model. As a specific example, consider the following distributions of the Weibull form:

$$F_1(x_1; \theta) = 1 - e^{-x_1^\theta}, \quad x_1 > 0, \quad (26)$$

and

$$F_2(x_2; \tau) = 1 - e^{-x_2^\tau}, \quad x_2 > 0. \quad (27)$$

The corresponding log-likelihood function is of the form

$$\begin{aligned} \ell(\boldsymbol{\theta}; X_1, X_2) &= n \log(\alpha_1 \alpha_2 \theta \tau) + (\theta - 1) \sum_{i=1}^n \log(x_{1i}) + (\tau - 1) \sum_{i=1}^n \log(x_{2i}) - \alpha_1 \sum_{i=1}^n x_{1i}^\theta \\ &\quad - \alpha_1 \alpha_2 \delta \sum_{i=1}^n x_{1i}^\theta x_{2i}^\tau - \alpha_2 \sum_{i=1}^n x_{2i}^\tau + \sum_{i=1}^n \log \left[\left((1 + \alpha_1 \delta x_{1i}^\theta) (1 + \alpha_2 \delta x_{2i}^\tau) - \delta \right) \right] \end{aligned}$$

The score function $U(\boldsymbol{\theta}) = (U(\theta), U(\tau), U(\delta), U(\alpha_1), U(\alpha_2))$, has elements which are derivative of the log-likelihood function with respect to the parameters and thus are given by

$$U(\alpha_1) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha_1} = \frac{n}{\alpha_1} - \sum_{i=1}^n x_{1i}^\theta - \alpha_2 \delta \sum_{i=1}^n x_{1i}^\theta x_{2i}^\tau + \delta \sum_{i=1}^n \frac{(1 + \alpha_2 \delta x_{2i}^\tau) x_{1i}^\theta}{(1 + \alpha_1 \delta x_{1i}^\theta) (1 + \alpha_2 \delta x_{2i}^\tau) - \delta},$$

$$U(\alpha_2) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha_2} = \frac{n}{\alpha_2} - \sum_{i=1}^n x_{2i}^\tau - \alpha_1 \delta \sum_{i=1}^n x_{1i}^\theta x_{2i}^\tau + \delta \sum_{i=1}^n \frac{(1 + \alpha_1 \delta x_{1i}^\theta) x_{2i}^\tau}{(1 + \alpha_1 \delta x_{1i}^\theta) (1 + \alpha_2 \delta x_{2i}^\tau) - \delta},$$

$$U(\delta) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \delta} = -\alpha_1 \alpha_2 \sum_{i=1}^n x_{1i}^\theta x_{2i}^\tau + \sum_{i=1}^n \frac{(\alpha_1 x_{1i}^\theta + \alpha_2 x_{2i}^\tau - 1) + 2\alpha_1 \alpha_2 \delta x_{1i}^\theta x_{2i}^\tau}{(1 + \alpha_1 \delta x_{1i}^\theta) (1 + \alpha_2 \delta x_{2i}^\tau) - \delta},$$

$$U(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^n \log(x_{1i}) - \alpha_1 \sum_{i=1}^n x_{1i}^{\theta} \log(x_{1i}) - \alpha_1 \alpha_2 \delta \sum_{i=1}^n x_{1i}^{\theta} x_{2i}^{\tau} \log(x_{1i}) \\ + \alpha_1 \delta \sum_{i=1}^n \frac{(1 + \alpha_2 \delta x_{2i}^{\tau}) x_{1i}^{\theta} \log(x_{1i})}{(1 + \alpha_1 \delta x_{1i}^{\theta})(1 + \alpha_2 \delta x_{2i}^{\tau}) - \delta'}$$

$$U(\tau) = \frac{\partial \ell(\theta)}{\partial \tau} = \frac{n}{\tau} + \sum_{i=1}^n \log(x_{2i}) - \alpha_2 \sum_{i=1}^n x_{2i}^{\tau} \log(x_{2i}) - \alpha_1 \alpha_2 \delta \sum_{i=1}^n x_{1i}^{\theta} x_{2i}^{\tau} \log(x_{2i}) \\ + \alpha_2 \delta \sum_{i=1}^n \frac{(1 + \alpha_1 \delta x_{1i}^{\theta}) x_{2i}^{\tau} \log(x_{2i})}{(1 + \alpha_1 \delta x_{1i}^{\theta})(1 + \alpha_2 \delta x_{2i}^{\tau}) - \delta'}$$

By equating the scores to zero we obtain the score equations, i.e., $(U(\theta) = \mathbf{0})$. These equations are typically solved by means of Newton–Raphson or quasi-Newton numerical methods to obtain the maximum likelihood estimators $\hat{\theta} = (\hat{\theta}, \hat{\tau}, \hat{\delta}, \hat{\alpha}_1, \hat{\alpha}_2)$ of the parameter vector $\theta = (\theta, \tau, \delta, \alpha_1, \alpha_2)$. The observed information matrix of θ is given by $K(\theta) = -\frac{d}{d\theta} U(\theta) = (K_{\theta_j \theta_{j'}})$, i.e., with elements of the form of minus the second derivative of the log-likelihood function with respect to the parameters. The Fisher information matrix of vector θ , $I(\theta)$ is given by $I(\theta) = \mathbf{E}(K(\theta))$ and should be calculated numerically.

When, we use the base distributions (26) and (27) in the $PHC(I)$ model, a technique known as pseudo-likelihood estimation (see Arnold and Strauss [10]) will provide estimates of all five parameters in the model. Besag [7], defined the pseudo-likelihood estimator of θ as the value θ_0 of θ that maximizes the pseudo likelihood function, which in the present bivariate situation is based on the conditional PH densities and is given by

$$L_P(\beta; X_1, X_2) = \prod_{i=1}^n f_{X_1|X_2}(x_{1i}|x_{2i}) f_{X_2|X_1}(x_{2i}|x_{1i}). \quad (28)$$

Thus for the example with distributions:

$$F_1(x_1; \gamma_1) = x_1^{\gamma_1}, \quad 0 < x_1 < 1,$$

and

$$F_2(x_2; \gamma_2) = x_2^{\gamma_2}, \quad 0 < x_2 < 1,$$

we have the following log pseudo likelihood.

$$\ell_P(\beta; X_1, X_2) = \gamma_1 \sum_{i=1}^n \log(x_{1i}) + 2 \sum_{i=1}^n \log[(1 - \alpha_1 \delta \log(1 - x_{1i}^{\gamma_1}))(1 - \alpha_2 \delta \log(1 - x_{2i}^{\gamma_2})) - \delta] \\ + n \log(\alpha_1 \alpha_2 \gamma_1 \gamma_2) + \gamma_2 \sum_{i=1}^n \log(x_{2i}) - 2\alpha_1 \alpha_2 \delta \sum_{i=1}^n \log(1 - x_{1i}^{\gamma_1}) \log(1 - x_{2i}^{\gamma_2})$$

Parallel to the definition of the score function, the pseudo–score function is defined to be the vector whose coordinates are partial derivatives of the log-pseudo-likelihood function with respect to each of the parameters in the model. It is denoted by $U_p(\beta) = (U_p(\gamma_1), U_p(\gamma_2), U_p(\delta), U_p(\alpha_1), U_p(\alpha_2))'$.

The estimating equations are constructed by setting the elements of the pseudo–score vector equal to zero. Solutions of these equations correspond to the pseudo-likelihood estimates of the parameters of the model. Typically, these solutions are obtained numerically using iterative methods such as Newton–Raphson or quasi-Newton.

The pseudo-likelihood estimator $\hat{\beta}$ of β obtained in the above fashion can be verified to be consistent and asymptotically normally distributed with covariance matrix given by

$$\Sigma_p = J^{-1}(\beta) K(\beta) J^{-1}(\beta)$$

(see Arnold and Strauss [10]), where for $l, m = 1, 2$

$$K_{lm}(\boldsymbol{\beta}) = \mathbf{E} \left[\left\{ \frac{\partial \ell_p(\boldsymbol{\beta})}{\partial \beta_l} \right\} \left\{ \frac{\partial \ell_p(\boldsymbol{\beta})}{\partial \beta_m} \right\}' \right], \quad J_{lm}(\boldsymbol{\beta}) = -\mathbf{E} \left[\frac{\partial^2 \ell_p(\boldsymbol{\beta})}{\partial \beta_l \partial \beta_m} \right].$$

As a consistent estimate of the asymptotic variance-covariance matrix of the pseudo-likelihood estimator, we will use the sandwich estimator proposed by Cheng and Riu [11]. This estimator is developed as follows.

Let $U_{pi}(\boldsymbol{\beta}) = \frac{\partial \ell_{pi} \boldsymbol{\beta}}{\partial \boldsymbol{\beta}}$, be the vector of pseudo-scores for the i -th observation. Then define

$$\hat{J}_n(\boldsymbol{\beta}) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial U_{pi}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\tilde{\boldsymbol{\beta}}},$$

which is the sum over all the observations of the matrices of second derivatives of $\ell_p(\boldsymbol{\beta})$ evaluated at the pseudo-likelihood estimator $\tilde{\boldsymbol{\beta}}$. In addition, define

$$\hat{K}_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n U_{pi}(\boldsymbol{\beta}) U_{pi}(\boldsymbol{\beta})' \Big|_{\tilde{\boldsymbol{\beta}}}.$$

Using this, we construct a consistent sandwich estimator of the asymptotic variance-covariance matrix in the form

$$\hat{\Sigma}(\tilde{\boldsymbol{\beta}}) = \frac{1}{n} \hat{J}_n^{-1}(\tilde{\boldsymbol{\beta}}) \hat{K}_n(\tilde{\boldsymbol{\beta}}) \hat{J}_n^{-1'}(\tilde{\boldsymbol{\beta}}).$$

A detailed discussion and analysis of such a model, but with power Lindley base distributions, utilizing pseudo-likelihood estimation, may be found in Martínez-Flórez et al. [12].

6. If F_1 and F_2 are Unknown

All but one of the bivariate proportional hazard models described in this paper have marginals of the proportional hazard form. The exception is the PHC(I) model which, for unknown F_1 and F_2 , we will discuss in Section 7. For the other models, we know that if F_1 and F_2 were known, we could transform the data to obtain a sample from a well-known bivariate exponential model. Consequently, if we consider an estimate $\tilde{F}_{1,n}(x)$ of F_1 based on $X_{1,1}, X_{1,2}, \dots, X_{1,n}$ and an estimate $\tilde{F}_{2,n}(x)$ of F_2 based on $X_{2,1}, X_{2,2}, \dots, X_{2,n}$, we can transform the data using

$$Z_{1,j} = -\log \tilde{F}_{1,n}(X_{1,j})$$

and

$$Z_{2,j} = -\log \tilde{F}_{2,n}(X_{2,j}),$$

and then we will have approximately a sample from a bivariate distribution with standard exponential marginals and we can then estimate the parameters in this exponential model. Note that for identifiability in unknown F_1 and F_2 models we have to fix α_1 and α_2 to be equal to 1.

For the PHC(II) model with Weibull component distributions, given in (26) and (27), a small simulation study of the performance of the maximum likelihood parameter estimates has been implemented for a variety of sample sizes and for several parametric configurations. With minimal loss of generality we set $\alpha_1 = \alpha_2 = 1$ throughout the simulation study. Three values of the dependence parameter δ were used, namely 0.15, 0.30 and 0.45, together with four sample sizes $n = 30, 50, 70, 90$. The table presents results for three representative choices of values for θ and τ . As for measures of performance, the relative bias (RB) and the square root of the mean squared error (MSE) are given.

The results in Table 1 confirm that both the relative bias and the root mean-squared error of the estimates decrease as sample size increases.

Table 1. RB and \sqrt{MSE} for the PHC(II)-Weibull model.

Parameters	n	$\hat{\theta}$		$\hat{\tau}$		$\hat{\delta}$		$\hat{\alpha}_1$		$\hat{\alpha}_2$	
		RB	\sqrt{MSE}	RB	\sqrt{MSE}	RB	\sqrt{MSE}	RB	\sqrt{MSE}	RB	\sqrt{MSE}
(1.25, 0.75, 0.15, 1, 1)	30	0.1964	0.2630	0.4915	0.4029	1.6753	0.3424	0.2011	0.2591	0.5361	0.5931
	50	0.1911	0.2583	0.4864	0.3853	1.4262	0.2828	0.1983	0.2342	0.5339	0.5705
	70	0.1842	0.2590	0.4819	0.3782	1.3480	0.2549	0.1974	0.2264	0.5330	0.5633
	90	0.1717	0.2597	0.4751	0.3686	1.3117	0.2376	0.1933	0.2165	0.5307	0.5541
(1.25, 0.75, 0.30, 1, 1)	30	0.1948	0.2675	0.4966	0.4078	1.6798	0.3393	0.1957	0.2542	0.5358	0.596
	50	0.1934	0.2620	0.4830	0.3823	1.4519	0.2841	0.1957	0.2310	0.5356	0.5731
	70	0.1887	0.2611	0.4827	0.3771	1.3260	0.2514	0.1935	0.2238	0.5332	0.5653
	90	0.1756	0.2579	0.4762	0.3688	1.2861	0.2351	0.1916	0.2178	0.5424	0.5629
(1.25, 0.75, 0.45, 1, 1)	30	0.1894	0.2610	0.5316	0.4336	0.4177	0.3002	0.1918	0.2463	0.6127	0.6808
	50	0.1827	0.2512	0.5149	0.4090	0.4022	0.2688	0.1904	0.2297	0.6018	0.6428
	70	0.1799	0.2483	0.4966	0.3884	0.4009	0.2604	0.1900	0.2210	0.5904	0.6193
	90	0.1679	0.2512	0.4905	0.3799	0.3958	0.2376	0.1779	0.2126	0.5878	0.6136
(1.75, 0.5, 0.15, 1, 1)	30	0.4228	0.7457	1.2543	0.6503	1.6778	0.3378	0.4287	0.4473	1.4193	1.4930
	50	0.4189	0.7427	1.2332	0.6305	1.5481	0.2941	0.4270	0.4353	1.4191	14.435
	70	0.4193	0.7422	1.2311	0.6262	1.3103	0.2455	0.4245	0.4347	1.4048	1.4378
	90	0.4093	0.7326	1.2276	0.6215	1.2531	0.2294	0.4227	0.4346	1.3939	1.4363
(1.75, 0.5, 0.30, 1, 1)	30	0.4218	0.7455	1.2862	0.6652	0.8103	0.3439	0.4204	0.4352	1.4968	1.5718
	50	0.4207	0.7428	1.2554	0.6408	0.7520	0.2984	0.4152	0.4288	1.4907	1.5178
	70	0.4171	0.7395	1.2544	0.6373	0.6981	0.2686	0.4154	0.4281	1.4804	1.5167
	90	0.4098	0.7334	1.2421	0.6300	0.6730	0.2539	0.4143	0.4250	1.4733	1.5102
(1.75, 0.5, 0.45, 1, 1)	30	0.4179	0.7376	1.2780	0.6602	0.4190	0.2985	0.4149	0.4270	1.5069	1.5796
	50	0.4169	0.7361	1.2606	0.6456	0.4115	0.2725	0.4148	0.4250	1.4762	1.5240
	70	0.4126	0.7326	1.2354	0.6268	0.4056	0.2597	0.4122	0.4239	1.4655	1.4988
	90	0.4034	0.7225	1.2303	0.6238	0.3916	0.2455	0.4062	0.4227	1.4550	1.4831
(0.75, 1.5, 0.15, 1, 1)	30	0.3903	0.3270	0.2506	0.3899	1.3177	0.3091	0.3590	0.4395	0.2778	0.3212
	50	0.3661	0.2980	0.2434	0.3887	1.1349	0.2508	0.3479	0.4006	0.2776	0.3044
	70	0.3516	0.2811	0.2493	0.3877	1.1461	0.2343	0.3471	0.3816	0.2761	0.2962
	90	0.3452	0.2721	0.2305	0.3871	1.0392	0.2106	0.3449	0.3755	0.2748	0.2929
(0.75, 1.5, 0.30, 1, 1)	30	0.3877	0.3280	0.2439	0.3831	0.6461	0.3136	0.3794	0.4639	0.2677	0.3001
	50	0.3699	0.3006	0.2424	0.3817	0.6403	0.2784	0.3539	0.4050	0.2660	0.2871
	70	0.3568	0.2861	0.2365	0.3785	0.6328	0.2611	0.3527	0.3896	0.2600	0.2884
	90	0.3532	0.2784	0.2273	0.3767	0.6066	0.2373	0.3513	0.3819	0.2492	0.2823
(0.75, 1.5, 0.45, 1, 1)	30	0.3933	0.3324	0.2529	0.3914	0.3934	0.2906	0.3774	0.4611	0.2664	0.3023
	50	0.3685	0.2996	0.2439	0.3900	0.3893	0.2662	0.3653	0.4186	0.2639	0.2872
	70	0.3557	0.2834	0.2495	0.3883	0.3649	0.2523	0.3595	0.3955	0.2594	0.2856
	90	0.3482	0.2747	0.2286	0.3861	0.3486	0.2379	0.3551	0.3833	0.2526	0.2829

7. If F_1 and F_2 Are Unknown in the PHC(I) Model

Our model is of the $PHC(I)(1, F_1; 1, F_2; \delta)$ form, i.e.,

$$f_{X_1, X_2}(x_1, x_2) = k(\delta)[(1 - \log(1 - F_1(x_1)))(1 - \log(1 - F_2(x_2))) - \delta] \times f_1(x_1)f_2(x_2) \exp\{-\delta \log[1 - F_1(x_1)] \log[1 - F_2(x_2)]\}. \tag{29}$$

where $\delta, F_1,$ and F_2 are unknown. Although it would be easy to estimate $\delta,$ via pseudolikelihood, if F_1 and F_2 were known, it is not apparent how to estimate F_1 and F_2 assuming that δ is known. So it is not clear how to implement an iterative strategy for estimating F_1, F_2 and δ simultaneously. Perhaps our only choice is to assume that the F_1 's belong to some parametric families of distributions, with once more utilizing pseudo likelihood to avoid dealing with $k(\delta).$

8. Application

The data analyzed in this example consist of the maximum water levels registered at two stations on the Fox river in Wisconsin during the period 1918–1950. Measurements were made at an upstream location (Berlin, X_1) and a downstream location (Wrightstown, X_2). This data set was previously analyzed by Gumbel and Mustafi [13] using a bivariate extreme model.

In our analysis of this data set we will fit four models namely:

- The Arnold and Strauss [3] bivariate exponential conditionals distribution. denoted by BEC.
- Gumbel's [6] first bivariate exponential distribution, denoted by BG(I).
- The proportional hazard conditionals Weibull extension of the BEC distribution, denoted by PHC(I)-W.
- The proportional hazard conditionals Weibull extension of the BG(I) distribution, denoted by PHC(II)-W.

In both of the Weibull proportional hazard conditionals extensions mentioned above, i.e., PHC(I)-W and PHC(II)-W, as described in Section 3, we use the following choices for the component distributions F_1 and F_2 :

$$F_1(x_1; \theta) = 1 - e^{-x_1^\theta}, \quad x_1 > 0 \quad \text{and} \quad F_2(x_2; \tau) = 1 - e^{-x_2^\tau}, \quad x_2 > 0.$$

Using the Arnold and Strauss [3] density given in Equation (11), the density of the PHC(I)-W is given by

$$f_{X_1, X_2}(x_1, x_2) = \alpha_1 \alpha_2 \theta \tau k(\delta) x_1^{\theta-1} x_2^{\tau-1} \exp\left(-\alpha_1 x_1^\theta - \alpha_2 x_2^\tau - \alpha_1 \alpha_2 \delta x_1^\theta x_2^\tau\right).$$

The corresponding log-pseudo-likelihood function for a sample of size n takes the form

$$\begin{aligned} \ell(\theta; X_{(1)}, X_{(2)}) &= n \log(\alpha_1 \alpha_2 \theta \tau) + (\theta - 1) \sum_{i=1}^n \log(x_{1i}) + \sum_{i=1}^n \log\left[(1 + \alpha_1 \delta x_1^\theta)(1 + \alpha_2 \delta x_2^\tau)\right] \\ &+ (\tau - 1) \sum_{i=1}^n \log(x_{2i}) - \alpha_1 \sum_{i=1}^n (1 + \alpha_2 \delta x_2^\tau) x_1^\theta - \alpha_2 \sum_{i=1}^n (1 + \alpha_1 \delta x_1^\theta) x_2^\tau. \end{aligned}$$

The log-pseudo-likelihood for the BEC model is obtained from the expression for the PHC(I)-W by setting $\theta = 1$ and $\tau = 1$.

The log-likelihood function for a sample of size n from the PHC(II)-W is of the form given in Equation (28), with simple change of notation. In this case the corresponding log-likelihood function for the BG(I) model is again obtained by setting $\theta = 1$ and $\tau = 1$.

Using the Fox river data, maximizing the log-likelihood for the models BG(I) and PHC(II)-W and the log-pseudo-likelihood for the models BEC and PHC(I)-W, we obtain the estimates of the parameters of the four models given in Table 2 (with standard errors in parentheses).

To compare model fitting, we use the AIC (Akaike [14]) criterion, namely $AIC = -2\hat{\ell}(\cdot) + 2p$. We also consider the BIC (Schwarz [15]) criterion, namely $BIC = -2\hat{\ell}(\cdot) + \log(n)p$, criterion where p is the number of parameters for the model being considered. The best model is the one with the smallest AIC or BIC.

According to the values of the AIC and BIC criteria for the Fox river data, the best model is the PHC(I)-W followed by the PHC(II)-W model.

Table 2. Estimates (standard errors) for the fitted models.

Estimates	BG(I)	PHC(II)-W	BEC	PHC(I)-W
$\hat{\alpha}_1$	0.2198 (0.0185)	0.0775 (0.0389)	0.0940 (0.0081)	0.0277 (0.0077)
$\hat{\alpha}_2$	0.0656 (0.0056)	0.0055 (0.0029)	0.0283 (0.0089)	0.0046 (0.0008)
$\hat{\delta}$	0.8344 (0.3041)	0.9053 (0.3341)	3.8624 (0.3254)	0.8753 (0.0006)
$\hat{\theta}$		1.6396 (0.2979)		2.0337 (0.0007)
$\hat{\tau}$		1.8895 (0.1995)		1.8384 (0.0006)
AIC	400.6863	364.4660	415.2929	353.4773
BIC	405.1758	371.9485	419.7824	360.9598

Since the BEC and BG(I) models are special cases of the PHC(I)-W and PHC(II)-W models, respectively, obtained by setting $\theta = \tau = 1$, we may test the hypotheses

$$H_0 : (\theta, \tau) = (1, 1) \text{ versus } H_1 : (\theta, \tau) \neq (1, 1)$$

for comparing the PHC(II)-W and PHC(I)-W models with the BG(I) and BEC models, respectively.

Using the likelihood ratio statistic,

$$\Lambda = \frac{\ell_f(\hat{\beta})}{\ell_{f-W}(\hat{\beta})} \tag{30}$$

we obtain

$$-2 \log(\Lambda) \sim \chi_2^2. \tag{31}$$

The corresponding values of $-2 \log(\Lambda)$ in each case are provided in Table 3 (note in the BEC-PHC(I)-W comparison the log-pseudo-likelihoods have been utilized instead of log-likelihoods) which are greater than the value of the $\chi_{2,99\%}^2 = 9.210$ indicating that the PHC(II)-W and PHC(I)-W models are significantly better at the 1% level. Thus, the PHC(II)-W and PHC(I)-W models appear to be good alternative for fitting the set data. The choice between the PHC(I)-W and PHC(II)-W is not so clear-cut, but perhaps the PHC(I)-W might be considered to be marginally better.

The graphs in Figures 1a,b and 2a,b show the contours of the densities BG(I) and BEC and of the fitted models for PHC(II)-W and PHC(I)-W, respectively.

Table 3. Comparison of likelihood ratio statistics.

	PHC(II)-W vs. BG(I)	PHC(I)-W vs. BEC
$-2 \log(\Lambda)$	40.2203	65.8156

Under the assumption that the forms of the F_i 's are unknown, we use the transformations

$$Z_{1,j} = -\log \tilde{F}_{1,n}(X_{1,j}) \text{ and } Z_{2,j} = -\log \tilde{F}_{2,n}(X_{2,j}),$$

to arrive at a BG(I) model with joint survival function

$$S(z_1, z_2) = \exp(-z_1 - z_2 - \delta z_1 z_2).$$

Then, using the expression for the maximum likelihood estimate of δ provided by Kotz et al. ([2], p. 352), we obtain $\hat{\delta} = 0.2986$, a much smaller value than the estimated

value of the parameter δ obtained assuming a known form for the F_i 's. Perhaps this indicates that the Weibull choices for the F_i 's are not optimal.

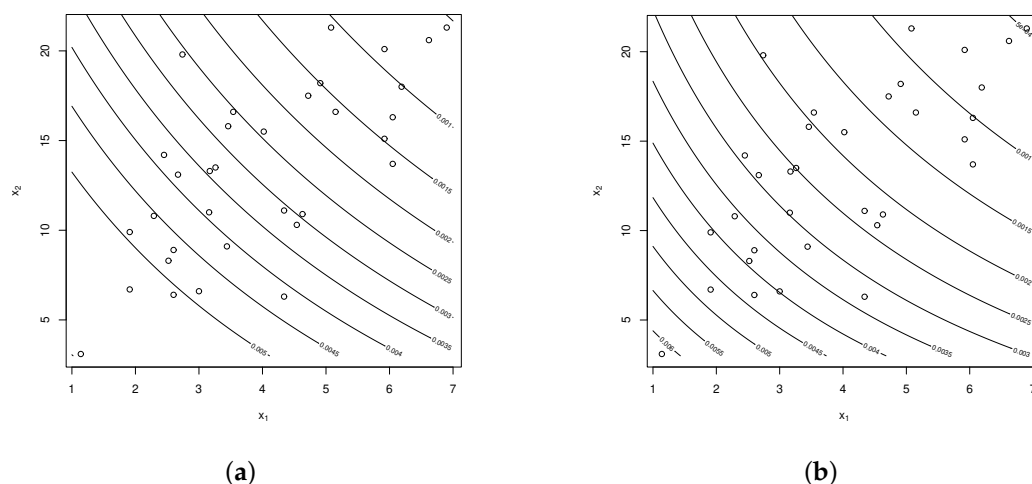


Figure 1. Contours for (a) BG(I) model and (b) BEC model.

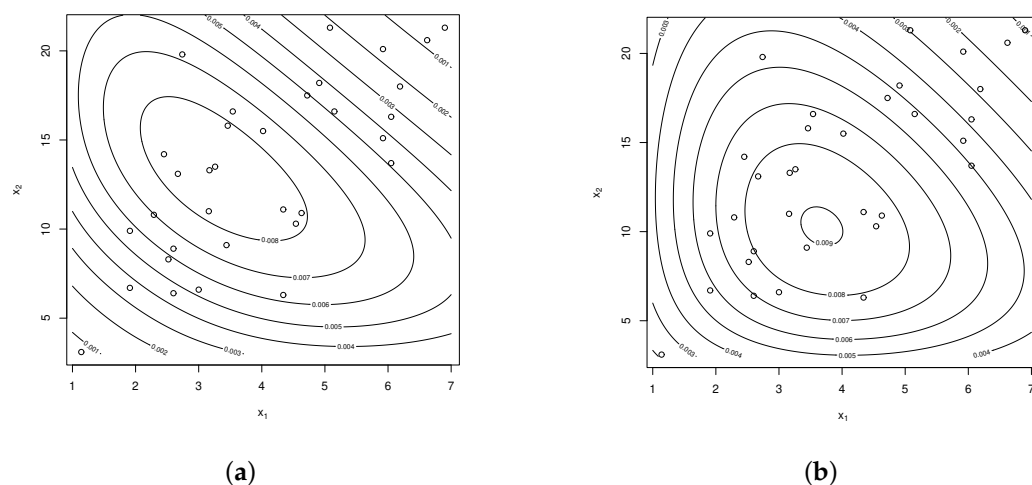


Figure 2. Contours for (a) PHC(II)-W model and (b) PHC(I)-W model.

9. Discussion

The bivariate models discussed in this paper utilize quite different approaches to their construction and thus can be expected to exhibit significantly different distributional properties, especially with regard to dependence. Future research on such models should put some focus on the problem of selecting the appropriate one of these models for a particular data set. Of course, the old stand-by of fitting via maximum likelihood and comparing models via AIC and BIC is always available.

Author Contributions: Conceptualization, B.C.A., G.M.-F. and H.W.G.; Formal analysis, B.C.A., G.M.-F. and H.W.G.; Investigation, B.C.A., G.M.-F. and H.W.G.; Methodology, B.C.A. and G.M.-F.; Software, G.M.-F.; Supervision, B.C.A. and H.W.G.; Validation, B.C.A., G.M.-F. and H.W.G. Writing—original draft preparation, B.C.A., G.M.-F. and H.W.G.; Funding acquisition, H.W.G. All of the authors contributed significantly to this research article. All authors have read and agreed to the published version of the manuscript.

Funding: The research of H.W. Gómez was supported by SEMILLERO UA-2022 project, Chile. The research of G. Martínez-Flórez was supported by Universidad de Córdoba, Montería, Colombia.

Data Availability Statement: The data can be found in Gumbel and Mustafi (1967).

Acknowledgments: The authors thank the Editor and three anonymous referees for their constructive comments and suggestions, which have greatly helped them to improve the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Marshall, A.W.; Olkin, I. A Multivariate Exponential Distribution. *J. Am. Statist. Assoc.* **1967**, *62*, 30–44. [[CrossRef](#)]
2. Kotz, S.; Balakrishnan, N.; Johnson, N.L. *Continuous Multivariate Distributions*; John Wiley and Sons: Hoboken, NJ, USA, 2000.
3. Arnold, B.C.; Strauss, D.J. Bivariate distributions with exponential conditionals. *J. Am. Statist. Assoc.* **1988**, *83*, 522–527. [[CrossRef](#)]
4. Arnold, B.C.; Kim, Y.H. Conditional proportional hazards models. In *Lifetime Data: Models in Reliability and Survival Analysis*; Jewell, N.P., Kimber, A.C., Lee, M.L.T., Whitmore, G.A., Eds.; Springer: Boston, MA, USA, 1996; pp. 21–28.
5. Arnold, B.C.; Castillo, E.; Sarabia, J.M. *Conditional Specification of Statistical Models*; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 1999.
6. Gumbel, E.J. Bivariate Exponential Distributions. *J. Am. Statist. Assoc.* **1960**, *55*, 698–707. [[CrossRef](#)]
7. Besag, J. Statistical Analysis of Non-Lattice Data. *J. R. Stat. Soc. Ser. D* **1975**, *24*, 179–195. [[CrossRef](#)]
8. Arnold, B.C.; Strauss, D.J. Bivariate distributions with conditionals in prescribed exponential families. *J. R. Stat. Soc. Ser. B* **1991**, *53*, 365–375. [[CrossRef](#)]
9. Castillo, E.; Hadi, A.S. Modeling Lifetime Data with Application to Fatigue Models. *J. Am. Statist. Assoc.* **1995**, *90*, 1041–1054. [[CrossRef](#)]
10. Arnold, B.C.; Strauss, D.J. Pseudolikelihood estimation: Some examples. *Sankhya Ser. B* **1991**, *53*, 233–243.
11. Cheng, C.; Riu, J. On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. *Technometrics* **2006**, *48*, 511–519. [[CrossRef](#)]
12. Martínez-Flórez, G.; Arnold, B.C.; Gómez, H.W. A bivariate power Lindley survival distribution. 2022. *Unpublished Work*.
13. Gumbel, E.; Mustafi, C.K. Some Analytical Properties of Bivariate Extremal Distributions. *J. Am. Statist. Assoc.* **1967**, *62*, 569–588. [[CrossRef](#)]
14. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
15. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]