

Flexible Bivariate Binary Models for Estimating the Efficacy of Phototherapy for Newborns with Jaundice

Giampiero Marra¹ & Rosalba Radice²

¹ Department of Statistical Science, University College London, UK

² Department of Economics, Mathematics and Statistics, Birkbeck, University of London, UK

Correspondence: Rosalba Radice, Department of Economics, Mathematics and Statistics, Birkbeck, University of London, Malet Street, London WC1E 7HX. Tel: 44-20-7631-6795. E-mail: r.radice@bbk.ac.uk

Received: October 31, 2014 Accepted: November 30, 2014 Online Published: December 19, 2014

doi:10.5539/ijsp.v4n1p46

URL: <http://dx.doi.org/10.5539/ijsp.v4n1p46>

Abstract

In this work we analyse the efficacy of phototherapy (treatment) on the probability of being hyperbilirubinemic (outcome) in infants. A realistic quantification of the relationship between treatment and outcome can be challenging for various reasons. First, the probability of interest might be too small. Second, confounding unmeasured variables may exist which can bias the efficacy of phototherapy at preventing significant hyperbilirubinemia. Third, relationships between covariates and the outcome variable may exhibit non-linear patterns that, if not accounted for, can bias the relationship of interest. One way of dealing with the second and third issues is to use a semiparametric recursive bivariate probit model. To address the first issue as well, we explore an extension of this model which accounts for the fact that being hyperbilirubinemic can be regarded as a rare event. The proposed approach combines the marginal distributions of treatment and outcome using copulae, and uses asymmetric link functions to deal with rare outcome events. The main features underpinning the use of asymmetric link functions within semiparametric bivariate binary models are discussed.

Keywords: asymmetric link function, recursive bivariate binary model, jaundice, phototherapy efficacy, regression spline, unmeasured confounding

1. Introduction

Although jaundice in newborns is common and generally benign, very high total serum bilirubin (TSB) levels can injure the newborn's central nervous system (Maisels et al., 2001). This is known as hyperbilirubinemia. For this reason, TSB levels in jaundiced newborns are followed and sometimes treated with phototherapy if they are at risk of rising to or have already reached potentially dangerous levels. The American Academy of Pediatrics (AAP) has published guidelines that suggest TSB levels at which phototherapy is recommended for term and late preterm newborns (Maisels et al., 2004). However, no randomized trials have quantified the efficacy of this intervention at the TSB levels at which they are currently recommended. A randomized trial of phototherapy is difficult to do because relevant outcomes, such as a TSB level exceeding the AAP's threshold for exchange transfusion, are rare, and because there are ethical obstacles to randomizing newborns not to receive a therapy recommended by the AAP (Newman et al., 1999, 2004).

Another issue with quantifying the efficacy at preventing significant hyperbilirubinemia in infants is the presence of unmeasured confounding variables. For example, because continuing exclusive breastfeeding is a risk factor for subsequent hyperbilirubinemia, phototherapy is generally more indicated and therefore might be more commonly used in infants continuing to breastfeed exclusively (Kuzniewicz et al., 2008). Thus, it is expected that the effect of not having breastfeeding data might confound the relationship between phototherapy and the outcome, by causing the effect of phototherapy to be falsely high.

To account for the problem of unmeasured confounding, the recursive bivariate probit model (RBPM) may be used (Heckman, 1978). A recent application of RBPM to estimate the efficacy of phototherapy for newborns with jaundice has appeared in Newman et al. (2012). In brief, this approach assumes that the two observable binary variables (outcome and treatment) are manifestations of a corresponding pair of unobserved, correlated, bivariate

normal latent variables, with the manifest variable taking on a value of one if the latent variable is positive and zero otherwise.

The use of RBPM is, however, subject to criticism as it makes the assumptions of bivariate normality, pre-specified form of covariate effects and symmetry of link functions. Marra & Radice (2011) introduced an extension of RBPM which relaxes the assumption of pre-specified covariate-response relationships using penalized splines. Recently, Radice et al. (2013) also relaxed the assumption of bivariate normality using copula functions. These generalizations gave rise to the semiparametric copula RBPM (SCRBPBM). A potential issue with SCRBPBM is that the assumption of symmetric link functions is maintained. This means that when the event of interest is rare (as in this case) SCRBPBM, as any other modeling approach based on symmetric link functions, will tend to underestimate the probability of interest (Czado & Santner, 1992; Chen et al., 1999). There are many ways to obtain asymmetric links; see, for instance, the recent work by Bazan et al. (2013), and references therein. One of them considers the cumulative distribution function (cdf) of an asymmetric distribution. A very popular example is the complementary log-log link which comes from the cdf of the Gumbel distribution. Alternative links can be obtained using the cdfs of the Weibull and log-normal distributions. In these cases, the cdf is completely specified and it does not depend on any unknown shape parameter. Also, no relationship between these cdfs and the usual symmetric links are established. On the contrary, links derived from cdfs which depend on an unknown shape parameter have been proposed by Prentice (1976), Aranda-Ordaz (1976) and Bazan et al. (2013), for example. These links are skewed logit and probit and include the logit and probit links as special cases. The cdf of the generalized extreme value distribution can also be employed for this purpose (Wang & Dey, 2010).

Building on the work by Radice et al. (2013), we explore a modification of SCRBPBM which uses asymmetric links for the marginal distributions of the two model equations. Specifically, we employ a skew probit link using the standard skewnormal distribution by Azzalini (1985). We opted for this link as it includes the probit link as special case and has desirable mathematical properties.

2. The Proposed Model and Its Main Features

Consider a pair of random variables (y_{1i}, y_{2i}) (treatment and outcome, respectively) for $i = 1, \dots, n$, where $y_{vi} \in \{0, 1\}$, $v = 1, 2$, and n is the sample size. The observed y_{vi} can be viewed as determined by a latent continuous variable y_{vi}^* such that $y_{vi} = \mathbf{1}(y_{vi}^* > 0)$, where $\mathbf{1}$ is the indicator function. We assume that

$$y_{vi}^* = \eta_{vi} + \epsilon_{vi}, \quad \epsilon_{vi} \sim \mathcal{SN}(\xi_v),$$

where

$$\eta_{1i} = \mathbf{u}_{1i}^T \boldsymbol{\alpha}_1 + \sum_{k_1=1}^{K_1} s_{1k_1}(z_{1k_1i}), \quad (1)$$

$$\eta_{2i} = \psi y_{1i} + \mathbf{u}_{2i}^T \boldsymbol{\alpha}_2 + \sum_{k_2=1}^{K_2} s_{2k_2}(z_{2k_2i}), \quad (2)$$

and ϵ_{vi} follows a standardised skew normal (\mathcal{SN}) distribution with shape parameter $\xi_v \in \mathbb{R}$. Note that our development also allows ϵ_{vi} to follow either a power normal (\mathcal{PN}) or a reciprocal power normal (\mathcal{RPN}), for instance; more details are given in the Appendix. Parameter ψ quantifies the effect of the treatment on the outcome on the scale of η_{2i} , \mathbf{u}_{1i}^T is defined as $(1, u_{12i}, \dots, u_{1P_1i})$ and corresponds to the i^{th} row of $\mathbf{U}_1 = (\mathbf{u}_{11}, \dots, \mathbf{u}_{1n})^T$, the $n \times P_1$ model matrix containing P_1 parametric terms (e.g., intercept, dummy and categorical variables), $\boldsymbol{\alpha}_1$ is a coefficient vector, and the s_{1k_1} are unknown smooth functions of the K_1 continuous covariates z_{1k_1i} . Similarly, $\mathbf{u}_{2i}^T = (1, u_{22i}, \dots, u_{2P_2i})$ is the i^{th} row vector of the $n \times P_2$ model matrix $\mathbf{U}_2 = (\mathbf{u}_{21}, \dots, \mathbf{u}_{2n})^T$, $\boldsymbol{\alpha}_2$ is a parameter vector, and the s_{2k_2} are unknown smooth terms of the K_2 continuous regressors z_{2k_2i} . The $s_{vk_v}(z_{vk_vi})$ are subject to the centering (identifiability) constraint $\sum_{i=1}^n s_{vk_v}(z_{vk_vi}) = 0$, $v = 1, 2$, $k_v = 1, \dots, K_v$ (Wood, 2006). η_{1i} and η_{2i} can also include smooth terms interacted with some predictor(s) and smooth functions of two or more covariates (e.g., Wood, 2006).

2.1 Smooth Function Representation

An effective way of representing smooth functions of continuous variables is the regression spline approach (e.g.,

Eilers & Marx, 1996). Specifically,

$$s_{vk_v}(z_{vk_v i}) = \sum_{j=1}^{J_{vk_v}} \beta_{vk_v j} b_{vk_v j}(z_{vk_v i}) = \mathbf{B}_{vk_v}(z_{vk_v i})^T \boldsymbol{\beta}_{vk_v},$$

where the $b_{vk_v j}(z_{vk_v i})$ are known basis functions, the $\beta_{vk_v j}$ are parameters, J_{vk_v} is the number of spline bases used to represent $s_{vk_v}(\cdot)$, $\mathbf{B}_{vk_v}(z_{vk_v i})^T$ is the i^{th} vector of dimension J_{vk_v} , defined as $\{b_{vk_v 1}(z_{vk_v i}), b_{vk_v 2}(z_{vk_v i}), \dots, b_{vk_v J_{vk_v}}(z_{vk_v i})\}$, and $\boldsymbol{\beta}_{vk_v}$ is the corresponding parameter vector. The cases of smooth terms multiplied by some covariate(s) and of smooths of more than one variable follow a similar construction. Several choices for the basis functions are possible and include low rank thin plate regression splines, B-splines and cubic regression splines (e.g., Ruppert et al., 2003; Wood, 2006). Linear predictors (1) and (2) can, therefore, be written as $\eta_{1i} = \mathbf{u}_{1i}^T \boldsymbol{\alpha}_1 + \mathbf{B}_{1i}^T \boldsymbol{\beta}_1$ and $\eta_{2i} = \psi y_{1i} + \mathbf{u}_{2i}^T \boldsymbol{\alpha}_2 + \mathbf{B}_{2i}^T \boldsymbol{\beta}_2$, where $\mathbf{B}_{vi}^T = \{\mathbf{B}_{v1}(z_{v1i})^T, \dots, \mathbf{B}_{vK_v}(z_{vK_v i})^T\}$ and $\boldsymbol{\beta}_v^T = (\boldsymbol{\beta}_{v1}^T, \dots, \boldsymbol{\beta}_{vK_v}^T)$. After defining $\mathbf{X}_{1i} = (\mathbf{u}_{1i}^T, \mathbf{B}_{1i}^T)^T$ and $\mathbf{X}_{2i} = (y_{1i}, \mathbf{u}_{2i}^T, \mathbf{B}_{2i}^T)^T$, we have $\eta_{1i} = \mathbf{X}_{1i}^T \boldsymbol{\delta}_1$ and $\eta_{2i} = \mathbf{X}_{2i}^T \boldsymbol{\delta}_2$ where $\boldsymbol{\delta}_1^T = (\boldsymbol{\alpha}_1^T, \boldsymbol{\beta}_1^T)$ and $\boldsymbol{\delta}_2^T = (\psi, \boldsymbol{\alpha}_2^T, \boldsymbol{\beta}_2^T)$.

2.2 SN

The probability density function (pdf) and cdf of the standard SN distribution are given as

$$f_{SN}(\epsilon_{vi}, \xi_v) = 2\phi(\epsilon_{vi})\Phi(\xi_v \epsilon_{vi}), \quad F_{SN}(\epsilon_{vi}, \xi_v) = \Phi(\epsilon_{vi}) - 2T(\epsilon_{vi}, \xi_v),$$

where $T(\cdot, \cdot)$ denotes the Owen's T function and $\xi_v \in \mathbb{R}$. A reflection of SN can be obtained by using the result $F_{SN}(\epsilon_{vi}, \xi_v) = 1 - F_{SN}(-\epsilon_{vi}, -\xi_v)$ (Azzalini, 1985).

Note that $F_{SN}(-\epsilon_{vi}, \xi_v) \neq 1 - F_{SN}(\epsilon_{vi}, \xi_v)$ because $F_{SN}(\epsilon_{vi}, \xi_v)$ is not symmetric. Also, if $\xi_v = 0$ then $F_{SN}(\epsilon_{vi}, \xi_v) = \Phi(\epsilon_{vi})$, that is the standardised skew normal distribution reduces to a standardised normal distribution. Using the latent variable representation of the response, we have that

$$\mathbb{P}(y_{vi} = 1) = \mathbb{P}(y_{vi}^* > 0) = \mathbb{P}(\eta_{vi} + \epsilon_{vi} > 0) = \mathbb{P}(\epsilon_{vi} > -\eta_{vi}) = 1 - \mathbb{P}(\epsilon_{vi} \leq -\eta_{vi}). \quad (3)$$

If $\epsilon_{vi} \sim SN(\xi_v)$ then (8) is

$$\mathbb{P}(y_{vi} = 1) = 1 - F_{SN}(-\eta_{vi}, \xi_v) = F_{SN}(\eta_{vi}, -\xi_v),$$

whereas if $\epsilon_{vi} \sim SN(-\xi_v)$ then (8) becomes

$$\mathbb{P}(y_{vi} = 1) = 1 - F_{SN}(-\eta_{vi}, -\xi_v) = F_{SN}(\eta_{vi}, \xi_v).$$

The distribution of SN is positively or negatively asymmetric in agreement with the sign of ξ_v (Azzalini, 1985). Figure 1 shows the shape of the cdfs and pdfs of SN for different values of ξ_v . From the curves, it is evident that SN allows the probability of a binary variable to approach zero at different rates than it approaches one depending on the values of ξ and η . If $\xi = 0$ then SN yields a classic probit curve.

2.3 Copula representation and log-likelihood function

Given some marginal cdfs, it is possible to define the joint cdf of event $(y_{1i} = 1, y_{2i} = 1)$ by using the copula representation (Sklar, 1959, 1973)

$$\mathbb{P}(y_{1i} = 1, y_{2i} = 1) = C(\mathbb{P}(y_{1i} = 1), \mathbb{P}(y_{2i} = 1); \theta),$$

where C is a two-place copula function and θ is an association parameter measuring the dependence between $\mathbb{P}(y_{1i} = 1)$ and $\mathbb{P}(y_{2i} = 1)$. A nice feature of the copula approach is that the marginal distributions may come from different families. Also, choosing the marginal distributions and modeling the dependence between them can be treated as two separate but related issues. The families considered are Clayton, Frank, Gaussian, Gumbel, Joe and Student-t, as well as rotated versions (90, 180 and 270) of the Clayton, Gumbel and Joe copulas; see the table and figures in Radice et al. (2013) for these copulas' definitions and illustrations. Nelsen (2006) provides a comprehensive overview of copulas and their properties.

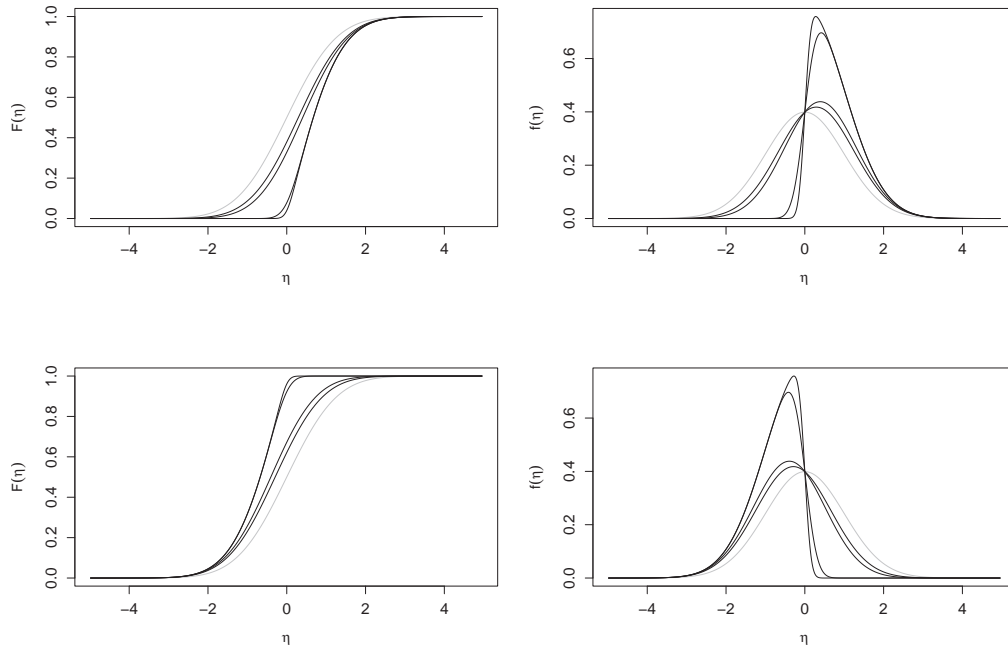


Figure 1. Cumulative distribution and probability density functions of SN Description: The first row displays the functions for $\xi = 0, 0.4, 0.6, 4, 8$, whereas the second row for $\xi = -8, -4, -0.6, -0.4, 0$, when using different values of η . The grey curves correspond to $\xi = 0$.

The log-likelihood function of the model can be written as

$$\ell = \sum_{i=1}^n \{y_{1i}y_{2i} \log \mathbb{P}(y_{1i} = 1, y_{2i} = 1) + y_{1i}(1 - y_{2i}) \log \mathbb{P}(y_{1i} = 1, y_{2i} = 0) + (1 - y_{1i})y_{2i} \log \mathbb{P}(y_{1i} = 0, y_{2i} = 1) + (1 - y_{1i})(1 - y_{2i}) \log \mathbb{P}(y_{1i} = 0, y_{2i} = 0)\},$$

where $\mathbb{P}(y_{1i} = 1, y_{2i} = 0) = \mathbb{P}(y_{1i} = 1) - \mathbb{P}(y_{1i} = 1, y_{2i} = 1)$, $\mathbb{P}(y_{1i} = 0, y_{2i} = 1) = \mathbb{P}(y_{2i} = 1) - \mathbb{P}(y_{1i} = 1, y_{2i} = 1)$ and $\mathbb{P}(y_{1i} = 0, y_{2i} = 0) = 1 - [\mathbb{P}(y_{1i} = 1) + \mathbb{P}(y_{2i} = 1) - \mathbb{P}(y_{1i} = 1, y_{2i} = 1)]$.

2.4 Treatment Effect

In our case, the aim is to investigate how the treatment changes the expected outcome. In other words, the interest is in the effect of y_{1i} on $\mathbb{P}(y_{2i} = 1|y_{1i})$, which in practice can be calculated using the commonly used average treatment effect in the specific sample at hand (SATE; Imbens, 2004). Assuming that $\mathbb{P}(y_{vi} = 1) = F_{SN}(\eta_{vi}, -\xi_v)$, SATE is defined as

$$\text{SATE}(\delta, \mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \mathbb{P}(y_{2i} = 1|y_{1i} = 1) - \mathbb{P}(y_{2i} = 1|y_{1i} = 0),$$

where

$$\mathbb{P}(y_{2i} = 1|y_{1i} = 1) = \frac{C(F_{SN}(\eta_{1i}, -\xi_1), F_{SN}(\eta_{2i}^{(y_{1i}=1)}, -\xi_2); \theta)}{F_{SN}(\eta_{1i}, -\xi_1)}, \quad (4)$$

$$\mathbb{P}(y_{2i} = 1|y_{1i} = 0) = \frac{F_{SN}(\eta_{2i}^{(y_{1i}=0)}, -\xi_2) - C(F_{SN}(\eta_{1i}, -\xi_1), F_{SN}(\eta_{2i}^{(y_{1i}=0)}, -\xi_2); \theta)}{1 - F_{SN}(\eta_{1i}, -\xi_1)}, \quad (5)$$

$\eta_{2i}^{(y_{1i}=r)}$ represents the linear predictor evaluated at $y_{1i} = r$ where r is equal to 1 or 0, δ contains all the model parameters and $\mathbf{X} = (\mathbf{x}_1 | \dots | \mathbf{x}_n)^T$ with \mathbf{x}_i defined as $(\mathbf{X}_{1i}^T, \mathbf{X}_{2i}^T)^T$. SATE(δ, \mathbf{X}) can be estimated using SATE($\hat{\delta}, \mathbf{X}$), whereas a confidence interval for it can be obtained using posterior simulation as explained in Section 4. If we use different marginal probability models then expressions (4) and (5) have to be changed accordingly.

3. Parameter Estimation

As pointed out by Azzalini & Arellano-Valle (2013), extra care must be taken when designing an algorithm to estimate the coefficients of a model that involves a link function which depends on a shape parameter. Here, the authors propose to penalize the shape parameter in estimation. We follow a similar idea and use a ridge penalty approach to penalize the shape parameters in our model.

Let us denote the copula log-likelihood as $\ell(\delta)$, where $\delta^T = (\delta_1^T, \delta_2^T, \theta, \xi_1, \xi_2)$. Because θ is bounded in most copula cases, we replace θ with θ_* (see the table in Radice et al. (2013)). The use of $\delta_*^T = (\delta_1^T, \delta_2^T, \theta_*, \xi_1, \xi_2)$ will ensure that in optimization $\delta_* \in \mathbb{R}^p$, where p is the total number of parameters. To avoid overfitting, which is likely to happen when a model includes smooth components, we employ in estimation the (second-order roughness) penalty term $\sum_{v=1}^2 \sum_{k_v=1}^{K_v} \lambda_{vk_v} \int \{s''_{vk_v}(z_{vk_v})\}^2 dz_{vk_v}$, where the λ_{vk_v} are smoothing parameters controlling the trade-off between fit and smoothness (e.g., Ruppert et al., 2003; Wood, 2006). Because regression splines are linear in their model parameters, the overall penalty can be written as $\beta^T \mathbf{S}_\lambda \beta$ where $\beta^T = (\beta_1^T, \beta_2^T)$, $\mathbf{S}_\lambda = \sum_{v=1}^2 \sum_{k_v=1}^{K_v} \lambda_{vk_v} \mathbf{S}_{vk_v}$ and the \mathbf{S}_{vk_v} are positive semi-definite known square matrices. Expressions for the $b_{vk_v,j}(z_{vk_v,i})$ needed to represent the smooth terms and the respective \mathbf{S}_{vk_v} can be found in Ruppert et al. (2003) and Wood (2006). The ridge penalties for the shape parameters are simply given by ξ_1^2 and ξ_2^2 . The function to maximize is

$$\ell_p(\delta_*) = \ell(\delta_*) - \frac{1}{2} (\beta^T \mathbf{S}_\lambda \beta + \xi_1^2 + \xi_2^2), \quad (6)$$

where $\beta^T \mathbf{S}_\lambda \beta + \xi_1^2 + \xi_2^2 = \delta_*^T \tilde{\mathbf{S}}_\lambda \delta_*$ with $\tilde{\mathbf{S}}_\lambda = \text{diag}(\mathbf{0}_{p_1}^T, \lambda_{1k_1} \mathbf{S}_{1k_1}, \dots, \lambda_{1K_1} \mathbf{S}_{1K_1}, \mathbf{0}_{p_2}^T, \lambda_{2k_2} \mathbf{S}_{2k_2}, \dots, \lambda_{2K_2} \mathbf{S}_{2K_2}, 0, 1, 1)$ and $\mathbf{0}_{p_v}^T = (0_{v1}, \dots, 0_{vp_v})$. Note that in (6) we need to assume that $\lambda^T = (\lambda_{1k_1}, \dots, \lambda_{1K_1}, \lambda_{2k_2}, \dots, \lambda_{2K_2})$ is known; joint estimation of δ and λ via maximization of (6) would clearly lead to $\hat{\lambda} = \mathbf{0}$, which is not practically useful (e.g., Wood, 2006). Building on the approach described in Radice et al. (2013), estimation of δ and λ is achieved in two steps which are iterated until convergence:

step 1 For a given parameter vector value $\delta_*^{[a]}$ and holding the smoothing parameter vector fixed at $\lambda^{[a]}$, find an estimate of δ_* using the trust region Newton approach:

$$\min_{\mathbf{p}} \tilde{\ell}_p(\delta_*^{[a]}) \stackrel{\text{def}}{=} - \left\{ \ell_p(\delta_*^{[a]}) + \mathbf{p}^T \mathbf{g}_p^{[a]} + \frac{1}{2} \mathbf{p}^T \mathcal{H}_p^{[a]} \mathbf{p} \right\} \quad \text{so that} \quad \|\mathbf{p}\| \leq r^{[a]},$$

$$\delta_*^{[a+1]} = \arg \min_{\mathbf{p}} \tilde{\ell}_p(\delta_*^{[a]}) + \delta_*^{[a]},$$

where, at iteration a , $\mathbf{g}_p^{[a]} = \mathbf{g}^{[a]} - \tilde{\mathbf{S}}_{\lambda^{[a]}} \hat{\delta}_*^{[a]}$, $\mathcal{H}_p^{[a]} = \mathcal{H}^{[a]} - \tilde{\mathbf{S}}_{\lambda^{[a]}}$, $\mathbf{g}^{[a]}$ is made up of $\mathbf{g}_1^{[a]} = \partial \ell(\delta_*) / \partial \delta_1|_{\delta_1=\delta_1^{[a]}}$, $\mathbf{g}_2^{[a]} = \partial \ell(\delta_*) / \partial \delta_2|_{\delta_2=\delta_2^{[a]}}$, $\mathbf{g}_3^{[a]} = \partial \ell(\delta_*) / \partial \theta_*|_{\theta_*=\theta_*^{[a]}}$, $\mathbf{g}_4^{[a]} = \partial \ell(\delta_*) / \partial \xi_1|_{\xi_1=\xi_1^{[a]}}$ and $\mathbf{g}_5^{[a]} = \partial \ell(\delta_*) / \partial \xi_2|_{\xi_2=\xi_2^{[a]}}$, and the Hessian matrix has a 5×5 matrix block structure with $(r, h)^{th}$ element $\mathcal{H}_{r,h}^{[a]} = \partial^2 \ell(\delta_*) / \partial \delta_r \partial \delta_h|_{\delta_r=\delta_r^{[a]}, \delta_h=\delta_h^{[a]}}$, $r, h = 1, \dots, 5$, where $\delta_3 = \theta_*$, $\delta_4 = \xi_1$ and $\delta_5 = \xi_2$. $\|\cdot\|$ denotes the Euclidean norm and $r^{[a]}$ represents the radius of the trust region. At each iteration, $\tilde{\ell}_p(\delta_*^{[a]})$ is minimized subject to the constraint that the solution falls within a trust region with radius $r^{[a]}$. The proposed solution is then accepted or rejected and the trust region expanded or shrunk based on the ratio between the improvement in the objective function when going from $\delta_*^{[a]}$ to $\delta_*^{[a+1]}$ and that predicted by the quadratic approximation; see Nocedal & Wright (2006) for full details. Note that, near the solution, the trust region Newton algorithm typically behaves as a Newton algorithm. As pointed out by Nocedal & Wright (2006), this approach usually works better than classic optimization methods.

step 2 For a given smoothing parameter vector value $\lambda^{[a]}$ and holding the main parameter vector value fixed at $\delta_*^{[a+1]}$, find an estimate of λ :

$$\lambda^{[a+1]} = \arg \min_{\lambda} \mathcal{V}_u(\lambda) \stackrel{\text{def}}{=} \frac{1}{\check{n}} \|\mathbf{z}^{+, [a+1]} - \mathbf{A}_\lambda^{[a+1]} \mathbf{z}^{+, [a+1]}\|^2 - 1 + \frac{2}{\check{n}} \text{tr}(\mathbf{A}_\lambda^{[a+1]}), \quad (7)$$

where $\check{n} = 4n$, $\mathbf{z}^{+, [a+1]} = \sqrt{\mathbf{W}^{[a+1]}} \mathbf{z}^{[a+1]}$, $\sqrt{\mathbf{W}^{[a+1]}}$ is the square root of a block diagonal matrix made up of $\mathbf{W}_i^{[a+1]}$ with $(r, h)^{th}$ element $-\partial^2 \ell(\delta_*)_i / \partial \eta_{ri} \partial \eta_{hi}|_{\eta_{ri}=\eta_{ri}^{[a+1]}, \eta_{hi}=\eta_{hi}^{[a+1]}}$, $r, h = 1, \dots, 4$, $\eta_{3i} = \xi_1$, $\eta_{4i} = \xi_2$, $\mathbf{z}_i^{[a+1]}$ is $\tilde{\mathbf{X}}_i \delta_*^{[a+1]} + \mathbf{W}_i^{-1, [a+1]} \mathbf{d}_i^{[a+1]}$, $\mathbf{d}_i^{[a+1]} = \left\{ \partial \ell(\delta_*)_i / \partial \eta_{1i}|_{\eta_{1i}=\eta_{1i}^{[a+1]}}, \dots, \partial \ell(\delta_*)_i / \partial \eta_{4i}|_{\eta_{4i}=\eta_{4i}^{[a+1]}} \right\}^T$, $\tilde{\mathbf{X}}_i = \text{diag}\{\mathbf{X}_{1i}^T, \mathbf{X}_{2i}^T, 1, 1\}$,

with \mathbf{X}_{1i} and \mathbf{X}_{2i} defined in Section 2.1, $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1 | \dots | \tilde{\mathbf{X}}_n)^T$, $\mathbf{A}_\lambda^{[a+1]} = \sqrt{\mathbf{W}^{[a+1]}} \tilde{\mathbf{X}} (\tilde{\mathbf{X}}^T \mathbf{W}^{[a+1]} \tilde{\mathbf{X}} + \check{\mathbf{S}}_\lambda)^{-1} \tilde{\mathbf{X}}^T \sqrt{\mathbf{W}^{[a+1]}}$ is the hat matrix, $\check{\mathbf{S}}_\lambda = \text{diag}(\mathbf{0}_{p_1}^T, \lambda_{1k_1} \mathbf{S}_{1k_1}, \dots, \lambda_{1K_1} \mathbf{S}_{1K_1}, \mathbf{0}_{p_2}^T, \lambda_{2k_2} \mathbf{S}_{2k_2}, \dots, \lambda_{2K_2} \mathbf{S}_{2K_2}, 1, 1)$, $\text{tr}(\mathbf{A}_\lambda^{[a+1]})$ represents the effective degrees of freedom (*edf*) of the penalized model (e.g., Wood, 2006), and γ is a tuning parameter which can be increased from its usual value of 1 to obtain smoother models (Kim & Gu, 2004). Function $\mathcal{V}_u(\lambda)$ can be regarded as an approximate Akaike information criterion (AIC) (Wood, 2006). A similar approach has been employed in a related context by Radice et al. (2013) and Marra & Radice (2011) to which we refer the reader for more details. Note that quantities needed to construct the working linear model are efficiently set up by using sparse algebra. In addition, the working linear model quantities in (7) are constructed for a fixed value of θ_* , hence keeping the dimensionality of the problem to the lowest possible. This is sensible because this parameter is not either penalized or specified as function of a linear predictor whose components may need to be penalized.

4. Model Selection and Confidence Intervals

Model selection can be achieved using the AIC which, in our case, is $\text{AIC} = -2\ell(\hat{\delta}_*) + 2\text{edf}$, where the log-likelihood is evaluated at the penalized parameter estimates and $\text{edf} = \text{tr}(\hat{\mathbf{A}}_\lambda)$.

For the construction of confidence intervals, we use the Bayesian covariance matrix $\mathbf{V}_{\delta_*} = -\mathcal{H}_p^{-1}$ which has been shown to produce intervals with close to nominal coverage probabilities as opposed to its frequentist counterpart (Marra & Wood, 2012). As explained, for instance, in Wood (2006) and Radice et al. (2013), another advantage of using a Bayesian result for interval construction is that intervals for non-linear functions of the model coefficients (e.g., SATE) can be conveniently obtained by simulation from the posterior distribution of δ_* . This can be achieved as follows:

step 1 Draw n_{sim} random vectors from $\mathcal{N}(\hat{\delta}_*, \hat{\mathbf{V}}_{\delta_*})$.

step 2 Calculate n_{sim} simulated realizations of the function of interest. Let us assume that a Gaussian copula is employed and that the function of interest is $\text{SATE}(\delta, \mathbf{X})$ where $\delta^T = (\delta_1^T, \delta_2^T, \theta = \tanh(\theta_*), \xi_1, \xi_2)$. We then have n_{sim} simulated realizations $\delta^{sim,1}, \dots, \delta^{sim,n_{sim}}$ and hence SATEs.

step 3 Using the n_{sim} realizations of SATE calculate the lower, $(\varsigma/2)$, and upper, $1 - \varsigma/2$, quantiles.

n_{sim} can be set to 1000 as the 3 steps are relatively inexpensive to run, whereas the significance level is usually set to 0.05. Note that an approximate confidence interval for SATE could also be obtained using the delta method in Radice et al. (2013).

5. Analysis of Phototherapy Data

5.1 Data

The aim is to estimate the efficacy of phototherapy in newborns with jaundice. Following previous studies, we considered 20731 infants born in 12 Northern California Kaiser Permanente Medical Care Program hospitals from 1 January 1995 to 31 December 2004 whose birth weight was at least 2000 grams and whose gestational age at least 35 weeks, with TSB levels within 3 mg/dL of the AAP's phototherapy threshold.

The data includes the following variables:

- Binary outcome variable (**threshold**) indicating the cross of the AAP exchange transfusion threshold within 48 hours of the qualifying TSB.
- Binary treatment variable (**phototherapy**) representing the receipt of hospital phototherapy within 8 hours of the qualifying TSB.
- Measured confounders:
 - Infant age:
 - * **age1**=1 if infant age < 24 hrs and 0 otherwise;
 - * **age2**=1 if infant age 24 < infant age < 48 hrs and 0 otherwise;
 - * **age3**=1 if infant age 48 < infant age < 72 hrs and 0 otherwise;
 - * **age4**=1 if infant age 72 < infant age < 96 hrs and 0 otherwise;

- * $\text{age5}=1$ if infant age > 96 hrs and 0 otherwise.
- Birth weight in grams (weight , continuous variable)
- Gestational age in weeks (gest , continuous variable).
- Gender (gender), binary variable ($\text{gender}=1$ if male, 0 otherwise).
- Qualifying TSB level:
 - * $\text{TSB1}=1$ if $-3 < \text{TSB} < -2$ and 0 otherwise;
 - * $\text{TSB2}=1$ if $-2 < \text{TSB} < -1$ and 0 otherwise;
 - * $\text{TSB3}=1$ if $-1 < \text{TSB} < 0$ and 0 otherwise;
 - * $\text{TSB4}=1$ if $0 < \text{TSB} < 1$ and 0 otherwise;
 - * $\text{TSB5}=1$ if $1 < \text{TSB} < 2$ and 0 otherwise;
 - * $\text{TSB6}=1$ if $2 < \text{TSB} < 3$ and 0 otherwise.

5.2 Model Specification

The linear predictors for the phototherapy and threshold equations were specified as

$$\begin{aligned}\eta_{1i} &= \alpha_{10} + \alpha_{11}\text{age2} + \alpha_{12}\text{age3} + \alpha_{13}\text{age4} + \alpha_{14}\text{age5} + \alpha_{15}\text{gest} + \alpha_{16}\text{gender} + \alpha_{17}\text{TSB2} + \alpha_{18}\text{TSB3} \\ &\quad + \alpha_{19}\text{TSB4} + \alpha_{20}\text{TSB5} + \alpha_{21}\text{TSB6} + s_{11}(\text{weight}), \\ \eta_{2i} &= \alpha_{20} + \psi\text{phototherapy} + \alpha_{21}\text{age2} + \alpha_{22}\text{age3} + \alpha_{23}\text{age4} + \alpha_{24}\text{age5} + \alpha_{25}\text{gest} + \alpha_{26}\text{gender} \\ &\quad + \alpha_{27}\text{TSB2} + \alpha_{28}\text{TSB3} + \alpha_{29}\text{TSB4} + \alpha_{30}\text{TSB5} + \alpha_{31}\text{TSB6} + s_{21}(\text{weight}),\end{aligned}$$

where s_{11} and s_{21} are the unknown smooth functions described in Section 2.1. The smooth components were represented using penalized thin plate regression splines with basis dimensions equal to 10 and penalties based on second order derivatives (Wood, 2006). The non-linear specification for weight arises from the fact that this covariate is likely to affect $\mathbb{P}(\text{phototherapy} = 1)$ and $\mathbb{P}(\text{threshold} = 1)$ non-linearly. gest was included as a parametric component because it did not have enough unique covariate values to justify the use of a smooth function. Regarding the use of link functions, we considered the probit and skew normal links for each copula model mentioned in Section 2.3.

6. Results

For each fitted model the AIC, estimated θ and estimated SATE are reported in Tables 1, 2 and 3. The results for the model cases that did not achieve convergence are not reported. Specifically, for the models with probit margins and 90 and 180 degrees rotated Clayton copulas and 270 degrees rotated Gumbel copula, the algorithm failed to find a solution. For the models with skew normal margins and Frank, Student-t, 180 degrees rotated Clayton copula, 270 degrees rotated Gumbel copula, not-rotated Joe and 270 degrees rotated Joe copula, the algorithm did not converge. Numerical problems sometimes arise when the model employed is not appropriate to fit the data at hand.

The model with lowest AIC is the 180 degrees rotated Gumbel copula with skew normal margins, although other copula models with very similar AIC values are good candidate as well. The findings are consistent with the interpretation that babies who are less likely to receive phototherapy are at lower risk of being hyperbilirubinemic. In other words, unmeasured confounding (e.g., breast feeding) is negatively associated with both the probability of receiving phototherapy and the probability of being hyperbilirubinemic. The estimated θ for Gumbel180, 3.03(2.06, 5.06), suggests that unmeasured confounding is present. The estimated SATE is -0.78 ($-1.23, -0.59$). The result is consistent with the interpretation that the probability of being hyperbilirubinemic decreases by 0.78% when a baby receives phototherapy as compared to a baby who does not receive it. When comparing the estimated results among all the models two points are worth noting. First, with some exceptions (e.g., the 270 degrees rotated Clayton copula, 90 degrees rotated Gumbel and Joe copulas), the estimated SATE appears to be robust to different margins and copula specifications. Second, the estimated θ are similar for the models whose AIC values are reasonably close to each other, whereas for the models which are not supported by the data the estimates vary considerably.

Table 1. AIC values for the copula models with probit and skew normal margins

		Margins	
		Probit	Skew normal
Copula	Gaussian	20219.64	20213.52
	Frank	20220.76	-
	Student-t	20215.53	-
	Clayton0	20214.33	20214.34
	Clayton90	-	20224.10
	Clayton180	-	-
	Clayton270	20223.51	20221.34
	Gumbel0	20222.07	20219.84
	Gumbel90	20223.07	-
	Gumbel180	20216.39	20213.19
	Gumbel270	-	-
	Joe0	20222.07	-
	Joe90	20223.45	20222.63
	Joe180	20214.21	20214.22
	Joe270	20222.07	-

Table 2. Estimated θ (and confidence interval) obtained from the fitted copula models with probit and skew normal margins

		Margins	
		Probit	Skew normal
Copula	Gaussian	0.87 (0.69,0.95)	0.88 (0.73,0.95)
	Frank	-2.44 (-5.29,0.53)	-
	Student-t	-0.46 (-0.83,0.18)	-
	Clayton0	3.04 (1.63,5.62)	2.34 (0.77,8.15)
	Clayton90	-	-
	Clayton180	-	-
	Clayton270	-0.31 (-8.47,-0.01)	-0.35 (-8.52,-0.02)
	Gumbel0	1 (1,Inf)	0.75 (0.61,0.85)
	Gumbel90	-1.2 (-4.07,-1.02)	-
	Gumbel180	3.03 (2.06,5.06)	3.05 (2.09,5.08)
	Gumbel270	-	-
	Joe0	1 (1,Inf)	-
	Joe90	-1.33 (-9.74,-1.02)	-1.31 (-9.70,-1.00)
	Joe180	3.93 (2.49,6.61)	3.90 (2.46,6.57)
	Joe270	-1 (-Inf,-1)	-

Table 3. Estimated SATE (and confidence interval) obtained from the fitted copula models with probit and skew normal margins

	Copula	Margins	
		Probit	Skew normal
	Gaussian	-0.77 (-1.24,-0.64)	-0.81 (-1.55,-0.53)
	Frank	-0.78 (-0.99,-0.58)	-
	Student-t	-0.78 (-0.99,-0.53)	-
	Clayton0	-0.78 (-1.23, -0.64)	-0.78 (-1.24,-0.63)
	Clayton90	-	-0.76 (-0.97,135.80)
	Clayton180	-	-
	Clayton270	-1.06 (-1.45, 0.83)	-1.08 (-1.40,-0.84)
	Gumbel0	-0.76 (-0.99, -0.59)	-0.81 (-1.65,-0.52)
	Gumbel90	-0.96 (-1.28,-0.77)	-
	Gumbel180	-0.77 (-1.18,-0.62)	-0.78 (-1.23,-0.59)
	Gumbel270	-	-
	Joe0	-0.76 (-0.99, -0.59)	-
	Joe90	-1.13 (-1.48,-0.89)	-1.16 (-25.11,-0.06)
	Joe180	-0.79 (-1.20,-0.64)	-0.79 (-1.15,-0.64)
	Joe270	-0.76 (-0.99,-0.59)	-

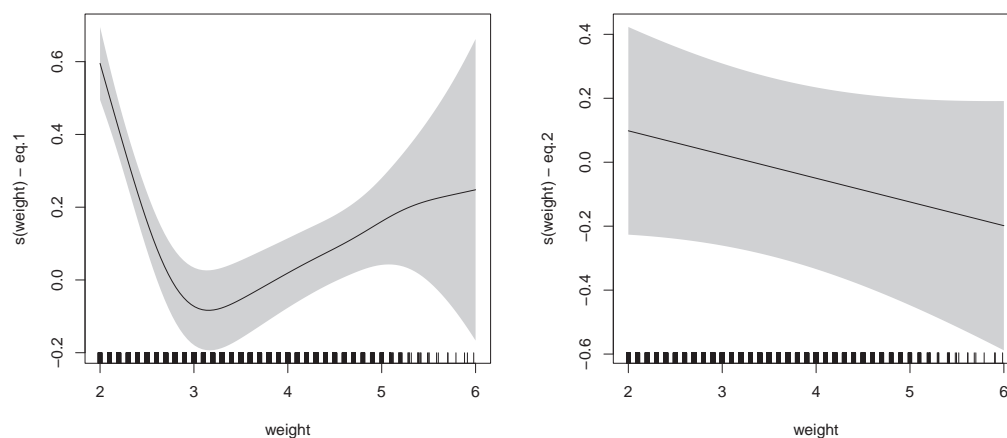


Figure 2. Estimated smooth components for the continuous variable **weight** in the **phototherapy** equation (eq. 1) and **threshold** equation (eq. 2)

Description: These were obtained by using the rotated 180 degrees Gumbel copula with skew normal margins. Results are on the scale of the respective linear predictors. Grey shades represent 95% confidence intervals and the 'rug plot', at the bottom of each graph, is used to show the covariate values.

Figure 2 shows the estimated smooth components for the **phototherapy** and **threshold** equations. The effect of **weight** is non-linear for the treatment equation suggesting that the propensity of receiving phototherapy first decreases and then increases as **weight** increases. As for the outcome equation the pointwise confidence intervals for the smooth function of **weight** suggest that **weight** does not have a linear or non-linear effect.

Table 4. Estimated coefficients of the measured confounders obtained from the fitted 180 degrees rotated Gumbel copula model with skew normal margins

	eq. 1	eq. 2
TSB2	0.04*	0.16
TSB3	0.14***	0.54***
TSB4	0.27***	0.81***
TSB5	0.37***	1.21***
TSB6	0.54***	1.56***
gender	0.04***	0.12**
gest	-0.04***	-0.15***
age2	-0.19***	-0.17
age3	-0.42***	-0.74***
age4	-0.72***	-1.19***
age5	-0.67***	-1.12***

Decription: (eq. 1) refers to the phototherapy equation and (eq. 2) to the threshold equation. ***, **, * and indicate significance at 0.001, 0.01, 0.05 and 1.

For completeness, Table 4 reports the estimated parametric coefficients for Gumbel180 with skew normal margins. As the qualifying TSB level increases the propensity of receiving phototherapy and that of being hyperbilirubinemic increase. Male babies are more likely to receive phototherapy and to be at higher risk of being hyperbilirubinemic. By increasing gestational age, both the propensity of receiving phototherapy and that of being hyperbilirubinemic decrease. Finally as age increases babies are less likely to receive phototherapy and less likely to be hyperbilirubinemic.

7. Discussion

The use of asymmetric link functions in the context of recursive bivariate models estimating the efficacy of phototherapy for newborns with jaundice has been explored. The skew probit link approach by Azzalini (1985) has been employed. We also explored the use of power probit and reciprocal power probit links. The choice of these link functions has been driven by the fact that they include the probit link as special case and have desirable properties. Our approach can also be employed for modeling jointly two imbalanced binary responses with a focus on predictive performance, for instance.

The approach has been applied to a study where the aim was to quantify the efficacy at preventing significant hyperbilirubinemia in infants. Results showed that phototherapy significantly decreases the probability of being hyperbilirubinemic. Also, our empirical findings suggest that the estimated quantity of interest (SATE) is robust to relaxing the assumption of symmetric links.

Appendix

The pdf and cdf of the standard power normal (\mathcal{PN}) distribution, denoted in our context by $\epsilon_{vi} \sim \mathcal{PN}(\xi_v)$, are given as

$$f_{\mathcal{PN}}(\epsilon_{vi}, \xi_v) = \xi_v [\Phi(\epsilon_{vi})]^{\xi_v-1} \phi(\epsilon_{vi}), \quad F_{\mathcal{PN}}(\epsilon_{vi}, \xi_v) = [\Phi(\epsilon_{vi})]^{\xi_v},$$

respectively, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, and $\xi_v > 0$. A new random variable can be defined such that

$$F_{\mathcal{RPN}}(\epsilon_{vi}, \xi_v) = 1 - [\Phi(-\epsilon_{vi})]^{\xi_v} \text{ and } f_{\mathcal{RPN}}(\epsilon_{vi}, \xi_v) = \xi_v [\Phi(-\epsilon_{vi})]^{\xi_v-1} \phi(\epsilon_{vi}).$$

In this case $\epsilon_{vi} \sim \mathcal{RPN}(\xi_v)$ denotes the standard reciprocal \mathcal{PN} distribution. The name comes from that fact that $F_{\mathcal{RPN}}(\epsilon_{vi}, \xi_v) = 1 - F_{\mathcal{PN}}(-\epsilon_{vi}, \xi_v)$. Therefore, the standard \mathcal{PN} and standard \mathcal{RPN} distributions are distinct although closely related because one is the reflection of the other.

Note that $F_{\mathcal{PN}}(-\epsilon_{vi}, \xi_v) \neq 1 - F_{\mathcal{PN}}(\epsilon_{vi}, \xi_v)$ or $F_{\mathcal{RPN}}(-\epsilon_{vi}, \xi_v) \neq 1 - F_{\mathcal{RPN}}(\epsilon_{vi}, \xi_v)$ because $F_{\mathcal{PN}}(\epsilon_{vi}, \xi_v)$ and $F_{\mathcal{RPN}}(\epsilon_{vi}, \xi_v)$ are not symmetric. Also, if $\xi_v = 1$ then $F_{\mathcal{PN}}(\epsilon_{vi}, \xi_v) = F_{\mathcal{RPN}}(\epsilon_{vi}) = \Phi(\epsilon_{vi})$. See Bazan et al. (2010) for full details.

Using the latent variable representation of the response, we have that

$$\mathbb{P}(y_{vi} = 1) = \mathbb{P}(y_{vi}^* > 0) = \mathbb{P}(\eta_{vi} + \epsilon_{vi} > 0) = \mathbb{P}(\epsilon_{vi} > -\eta_{vi}) = 1 - \mathbb{P}(\epsilon_{vi} \leq -\eta_{vi}). \quad (8)$$

If $\epsilon_{vi} \sim \mathcal{PN}(\xi_v)$, then (8) becomes

$$\mathbb{P}(y_{vi} = 1) = 1 - [\Phi(-\eta_{vi})]^{\xi_v} = 1 - F_{\mathcal{PN}}(-\eta_{vi}, \xi_v) = F_{\mathcal{RPN}}(\eta_{vi}, \xi_v),$$

whereas if $\epsilon_{vi} \sim \mathcal{RPN}(\xi_v)$ then (8) is

$$\mathbb{P}(y_{vi} = 1) = 1 - \{1 - [\Phi(\eta_{vi})]^{\xi_v}\} = \{1 - F_{\mathcal{RPN}}(-\eta_{vi}, \xi_v)\} = F_{\mathcal{PN}}(\eta_{vi}, \xi_v).$$

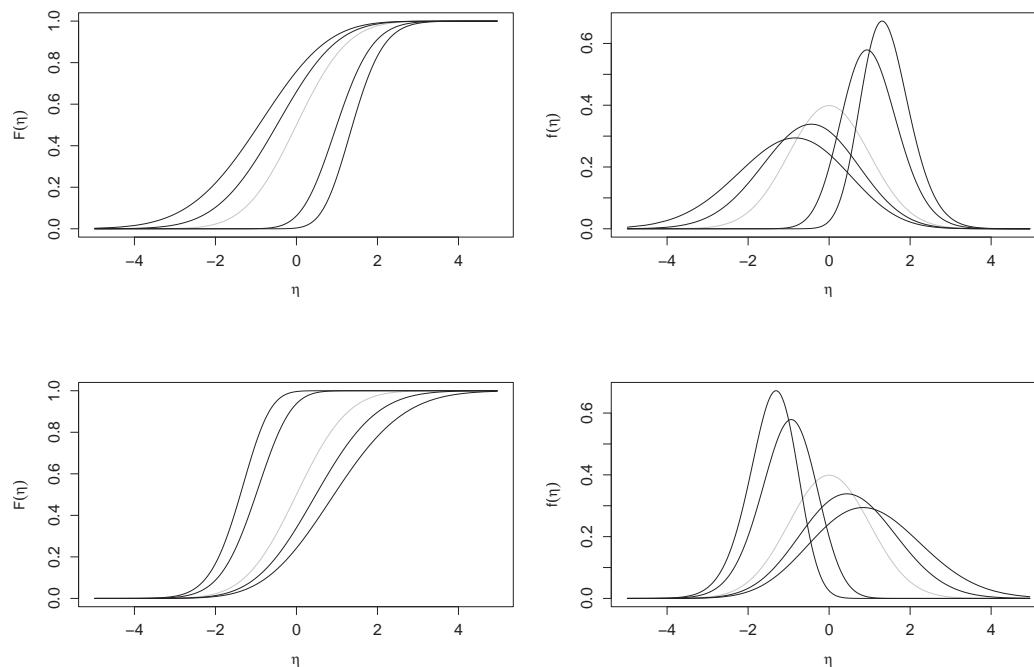


Figure 3. Shape of the cdfs and pdfs of \mathcal{PN} (top row) and \mathcal{RPN} (bottom row) for different values of ξ_v

As explained by Gupta & Gupta (2008), the \mathcal{PN} density is unimodal and is skewed to the right if $\xi_v > 1$ and to the left if $0 < \xi_v < 1$. On the other hand, by considering the reciprocal formulation, the \mathcal{RPN} density is also unimodal and is skewed to the left if $\xi_v > 1$ and to the right if $0 < \xi_v < 1$. Figure 3 shows the shape of the cdfs and pdfs of \mathcal{PN} and \mathcal{RPN} for different values of ξ_v . From the curves, it is evident that \mathcal{PN} and \mathcal{RPN} allow the probability of a binary variable to approach zero at different rates than it approaches one depending on the values of ξ and η . When $\xi = 1$, \mathcal{PN} and \mathcal{RPN} yield a classic probit curve.

References

- Aranda-Ordaz, F. J. (1976). On two families of transformations to additivity for binary response data. *Biometrika*, 68, 357–363. <http://dx.doi.org/10.1093/biomet/68.2.357>
- Azzalini, A. (1985). A class of distributions which includes the normal one. *Scandinavian Journal of Statistics*, 12, 171–178. <http://>
- Azzalini, A. & Arellano-Valle, R. B. (2013). Maximum penalized likelihood estimation for skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference*, 143, 419–433. <http://dx.doi.org/10.1016/j.jspi.2012.06.022>
- Bazan, J. L., H. Bolfarine, & Branco, M. B. (2010). A framework for skew-probit links in binary regression. *Communications in Statistics & Theory and Methods*, 39, 678–697. <http://dx.doi.org/10.1080/03610920902783849>
- Bazan, J. L., Romeo, J. S., & Rodrigues, J. (2013). Bayesian skew-probit regression for binary response data. *Brazilian Journal of Probability and Statistics*. <http://dx.doi.org/10.1214/13-BJPS218>

- Chen, M. H., Dey, D. K., & Shao, Q. (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, 94, 1172–1186. <http://dx.doi.org/10.2307/2669933>
- Czado, C. & Santner, T. J. (1992). The effect of link misspecification on binary regression inference. *Journal of Statistical Planning and Inference*, 33, 213–223. [http://dx.doi.org/10.1016/0378-3758\(92\)90069-5](http://dx.doi.org/10.1016/0378-3758(92)90069-5)
- Eilers, P. H. C. & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89–121. <http://dx.doi.org/10.1002/wics.125>
- Gupta, R. D. & Gupta, R. C. (2008). Analyzing skewed data by power normal model. *Test*, 17, 197–210. <http://dx.doi.org/10.1007/s11749-006-0030-x>
- Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46, 931–959. <http://dx.doi.org/10.2307/1909757>
- Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86, 4–29. <http://dx.doi.org/10.1162/003465304323023651>
- Kim, Y. J. & Gu, C. (2004). Smoothing spline gaussian regression: More scalable computation via efficient approximation. *Journal of the Royal Statistical Society Series B*, 66, 337–356. <http://dx.doi.org/10.1046/j.1369-7412.2003.05316.x>
- Kuzniewicz, M. W., Escobar, G. J., Wi, S., Liljestrand, P., McCulloch, C., & Newman, T. (2008). Risk factors for severe hyperbilirubinemia among infants with borderline bilirubin levels: a nested case-control study. *Journal of Pediatrics*, 153, 234–240. <http://dx.doi.org/10.1016/j.jpeds.2008.01.028>
- Maisels, M. J., Baltz, R. D., Bhutani, V. K., & et al. (2001). Neonatal jaundice and kernicterus. *Pediatrics*, 108, 763–765. <http://dx.doi.org/10.1542/peds.108.3.763>
- Maisels, M. J., Baltz, R. D., Bhutani, V. K., & et al. (2004). Management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. *Pediatrics*, 114, 297–316. <http://dx.doi.org/10.1542/peds.114.1.297>
- Marra, G. & Radice, R. (2011). Estimation of a semiparametric recursive bivariate probit model in the presence of endogeneity. *Canadian Journal of Statistics*, 39, 259–279. <http://dx.doi.org/10.1002/cjs.10100>
- Marra, G. & Radice, R. (2013). A penalized likelihood estimation approach to semiparametric sample selection binary response modeling. *Electronic Journal of Statistics*, 7, 1432–1455. <http://dx.doi.org/10.1214/13-EJS814>
- Marra, G. & Wood, S. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39, 53–74. <http://dx.doi.org/10.1111/j.1467-9469.2011.00760.x>
- Nelsen, R. (2006). *An Introduction to Copulas*. New York: Springer.
- Newman, T. B., Escobar, G. J., Gonzales, V., Armstrong, M., Gardner, M., & Folck, B. (1999). Frequency of neonatal bilirubin testing and hyperbilirubinemia in a large health maintenance organization. *Pediatrics*, 104, 1198–1203.
- Newman, T. B., Liljestrand, P., & Escobar, G. J. (2004). Infants with bilirubin levels of 30 mg/dl or more in a large managed care organization. *Pediatrics*, 111, 1303–1311.
- Newman, T. B., Vittinghoff, E., & McCulloch, C. (2012). Efficacy of phototherapy for newborns with hyperbilirubinemia: A cautionary example of an instrumental variable analysis. *Medical Decision Making*, 32, 83–92. <http://dx.doi.org/10.1177/0272989X11416512>
- Nocedal, J. & Wright, S. J. (2006). *Numerical Optimization*. New York: Springer-Verlag.
- Prentice, R. L. (1976). A generalization of the probit and logit methods for dose-response curves. *Biometrika*, 32, 761–768. <http://dx.doi.org/10.2307/2529262>
- R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Radice, R., Marra, G., & Wojtys, M. (2013). *Copula Regression Spline Models for Binary Outcomes*. Working Papers 321, University College London.
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, New

York.

Sklar, A. (1959). Fonctions de répartition é n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

Sklar, A. (1973). Random variables, joint distributions, and copulas. *Kybernetika*, 9, 449–460.

Wang, W. & Dey, D. (2010). Generalized extreme value regression for binary response data: An application to b2b electronic payments system adoption. *The Annals of Applied Statistics*, 4, 2000–2023. <http://dx.doi.org/10.1214/10-AOAS354>

Wood, S. N. (2006). *Generalized Additive Models: An Introduction With R*. Chapman & Hall/CRC, London.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).