J. Ital. Statist. Soc. (2000)
1-3, pp. 139-158

COMPARING PARAMETRIC MODELS FOR RELIABILITY DATA VIA RESIDUAL ANALYSIS

Alessandra Nardi*
Department of Systems Theory, University of Teramo

Summary

Fully parametric models, in particular the exponential and the Weibull, are widely used in reliability analysis where both the shape of the baseline hazard and the effect of a set of explanatory variables are of interest.

In order to compare the fitting of alternative parametric models we propose a graphical procedure based on log-odds and normal deviate residuals as diagnostic statistics. These residuals have been originally suggested for the purpose of outlier screening but their properties make them suitable for verifying assumptions on the distribution of the baseline hazard as well.

1. Introduction

While medical literature in survival analysis is dominated by the Cox's model, *i.e.*, a semiparametric approach to survival data, fully parametric models are widely used in reliability analysis. Here interest lies often in the shape of the hazard function, sometimes in the presence of non homogeneous observations.

Different families of distributions are available to model life-time data and discriminating among them is a central issue. Residual analysis can play an important role in this respect. We propose a graphical procedure based on the distributional properties of log-odds and normal deviate residuals (Nardi and Schemper, 1999).

We start by recalling their definitions and we investigate their sampling distributions under a correctly specified model in Section 2. In order to assess the goodness of fit of the assumed parametric model, residuals' empirical distributions can be compared with their corresponding reference distributions. In practice, even

* *Address for correspondence*: Dipartimento of Systems Theory, University of Teramo, Viale Crucioli, 122 – I-64100, Teramo. E-mail: nardi@dtso.spol.unite.it

assuming that a clear departure is observed, it may be difficult to interpret such a departure and it remains unclear in what way the model should be modified.

Therefore, in Section 3, we investigate residuals' expected behavior under a misspecified baseline hazard, but no censoring. The exponential and the Weibull regression models are considered, corresponding to the assumptions of a constant and monotone baseline hazard. The asymptotic distributions of the proposed residuals are derived, based on the convergence of maximum likelihood estimators (mles) under the alternative hypothesis. An application is described in Section 4. Section 5 deals with the effect of increasing censoring and methods for taking it into account while a final discussion is given in Section 6.

2. The residuals under a properly specified model

Let T_i denote the failure time and C_i the censoring time of the i-th unit. Suppose that there are n observations and that the data for subject i are of the form (y_i, δ_i, $\mathbf{x}_i$), where $Y_i = min(T_i, C\text{i})$, δ_i is an indicator function which equals 1 when the observed time is uncensored and $\mathbf{x}_i^t = (x_{i1},\ldots, x_{im})$, $i = 1,\ldots, n$, are observed values of m covariates. We assume that $log\ T_i$ is related to the covariates via the linear model

$$log T_i = \theta_0 + \sum_{j=1}^{m} \theta_j x_{ij} + log T_{0i} \qquad (1)$$

where θ_0 is the unknown general mean, $\theta_1,\ldots, \theta_m$ are unknown regression parameters and T_{0i} are i.i.d. random variables with common density $f(t;\ \kappa)$ which is independent of θ and completely specified up to an unknown scale or shape parameter κ. Notice that the effect of independent variables is multiplicative on the event time; for this reason this class of models is often referred to as *accelerated failure time models*. Let $\lambda_0(t;\ \kappa)$ be the hazard function of T_{0i}. It follows that both the hazard and the survival function of T_i can be written in terms of the baseline hazard $\lambda_0(t)$ as

$$\lambda(t;\ \kappa,\ \theta,\ \mathbf{x}_i) = \lambda_0(t\ e^{-\theta' \mathbf{x}_i};\ \kappa) e^{-\theta' \mathbf{x}_i}$$

and

$$S(t; \kappa, \theta, \mathbf{x}_i) = \exp\left[-\int_0^{te^{-\theta' \mathbf{x}_i}} \lambda_0(u; \kappa)\, du \right].$$

Since our attention focuses on modelling the baseline hazard function, throughout the paper we assume that the linear predictor $\theta' \mathbf{x}_i$ is correctly specified.

Log-odds and normal-deviate residuals, that we denote respectively as L_i and Ni, are defined as follows (Nardi and Schemper, 1999)

$$L_i = \log\left[\frac{S(t_i;\hat{\kappa},\hat{\theta},\mathbf{x}_i)}{1-S(t_i;\hat{\kappa},\hat{\theta},\mathbf{x}_i)}\right]$$

$$N_i = \Phi^{-1}\left[S(t_i;\hat{\kappa},\hat{\theta},\mathbf{x}_i)\right]$$

where $S(t_i; \hat{\kappa}; \hat{\theta}; \mathbf{x}_i)$ is the estimated survival function of the fitted model for individual i evaluated at his-her observed failure time and where Φ is the standard normal cdf.

At first we assume no censoring. The censored case will be dealt with in Section 5.

Notice that both the residuals are 0 if the observed failure time coincides with the estimated median failure time, which is regarded as reference time. Increasing departures from the predicted median time are reflected by increasing absolute values. Large negative and positive residuals identify too long and too short survival times. Assuming the survival function as known, L_i and N_i follow the standard logistic and the standard normal distribution, respectively. This results follows by noting that $U_i = S(T_i)$, $i = 1,\dots, n$ represent a set of n independent random variables, each having a [0, 1] uniform distribution. Being defined in terms of probability integral transform, the suggested residuals can be regarded as *generalized residuals* in the sense of Cox-Snell (Cox and Snell, 1968). Compared to the classical Cox-Snell residuals $e_i = -log(S(t_i; \hat{\kappa}, \hat{\theta}; \mathbf{x}_i))$ they offer two main advantages. They can be intuitively interpreted as a *distance* between the predicted median time and the observed failure time. Furthermore the symmetry of their reference distributions, which resembles the property of residuals in the General Linear Model, is of help in any graphical procedure. In fact, this avoids the exaggerated visual effect in the upper tail of the distribution of Cox and Snell residuals, that results from applying the logarithmic transformation to the survival function.

2.1. *Exact and asymptotic results*

When the unknown survival function is replaced by its mle, the property that the $U_i's$ are independent, each being uniformly distributed, is not valid any longer.

As first step, assume to observe a vector $\mathbf{T}' = (T_1,\dots, T_n)$ of n i.i.d. failure times, where each T_i follows an exponential distribution with common parameter δ. The mle of δ is $\hat{\delta} = \frac{n}{S}$ where $S = \sum_{i=1}^{n} T_i$. The pdf $f_{\hat{\mathbf{L}}^{(n)}}(t_1,\dots, t_n)$ of the corresponding vector of residuals can be easily derived by noting that $\mathbf{Y}^{(n)} = \frac{1}{S}\mathbf{T}$ follows a *Dirichlet* $(\mathbf{1}_n)$ distribution, where $\mathbf{1}'_n = (1,\dots, 1)$. Therefore

$$f_{\hat{\mathbf{L}}^{(n)}}(t_1,\dots,t_n) = \frac{(n-1)!}{n^n}\prod_{i=1}^{n}\frac{1}{1+\exp(t_i)}I_C(t_1,\dots,t_n)$$

where

$$C = \left\{(t_1,\dots,t_n) : t_i \in \left(\log\frac{\exp(-n)}{1-\exp(-n)}, +\infty\right)\forall i, \sum_{i=1}^{n}\log\frac{\exp(t_i)}{1+\exp(t_i)} = -n\right\}$$

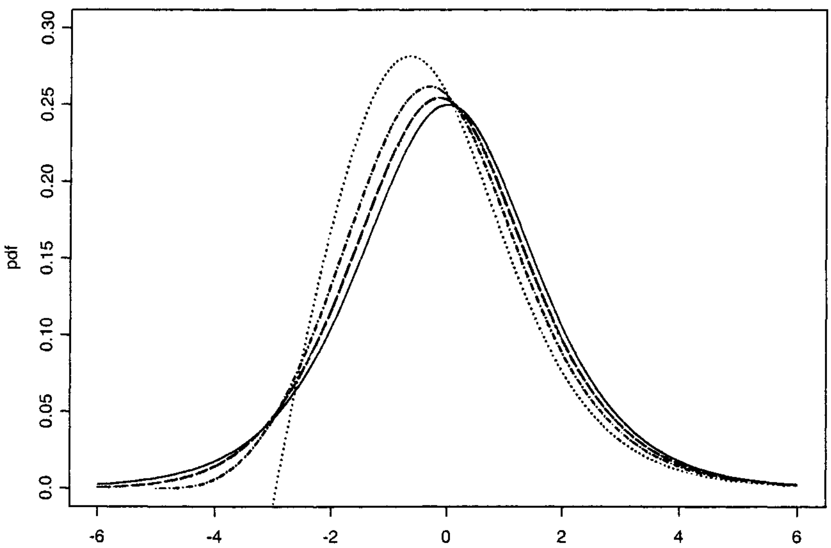

Fig. 1 – Pdf of log-odds residuals: —, Reference density; …, $n = 3$; – · –, $n = 5$; – –, $n = 10$.

Figure 1 shows the marginal density of $L_i^{(n)}$ for different choices of n. Clearly this marginal density converges to the reference logistic density as n tends to infinity, the approximation being already satisfactory for $n = 10$.

If failure times are not identically distributed, *i.e.*, a set of explanatory variables is introduced, exact results are not achievable since an explicit expression for mles is not available. However, on the basis of mles' consistency, we can prove that, assuming a correctly specified model, L_i and N_i converge in probability to their reference distributions, being asymptotically independent.

In order to assess possible violations of model assumptions, a Q-Q plot could be constructed, based on these results. In practice, even assuming that the plotted points clearly depart from linearity, indicating that the model is inappropriate, they cannot indicate in what way.

3. The residuals under a misspecified baseline hazard

In order to improve the understanding of the graphical inspection, we now investigate the expected behavior of the proposed residuals under a misspecified model for the baseline hazard function.

3.1. *Known parameters*

Throughout this sub-section the parameters are assumed to be known. Two particular cases are considered, defined by the shape of the baseline hazard $\lambda_0(t;\kappa)$:

1. an exponential regression model, H_E, where T_0 follows the standard exponential distribution.

2. a Weibull regression model, H_W, where T_0 is distributed according to the standard Weibull distribution.

We investigate the consequences on residuals' sampling distribution if an exponential model is fitted while the true model is H_W. Note that exponential regression is widely used in reliability, where a constant hazard is regarded a reasonable choice as far as there is no indication of a clear departure.

Since the discussion involves pairs of different hypotheses, different symbols are required for the set of parameters (θ_0, κ, $\theta' = (\theta_1,\ldots,\theta_m)$). This set will be denoted by (β_1, β_2, $\mathbf{b}'$) for H_W and (δ_1, $\mathbf{d}'$) for H_E.

Let a constant baseline hazard $H_E(\delta_1, \mathbf{d}')$ be falsely assumed, the true model being $H_W(\beta_1, \beta_2, \mathbf{b}')$. The random variables $U_i = S(T_i)$, $i = 1,\ldots, n$ are no more uniformly distributed. However the pdfs of log-odds and normal-deviate residuals can be still derived (Appendix A). We obtain

$$f_{L_i}(t) = \frac{\beta_2}{1+e^t}\left[\log\frac{e^t}{1+e_t}\right]^{(\beta_2-1)} \exp\left[-\left(\log\frac{1+e^t}{e^t}\right)^{\beta_2}\right] \qquad (2)$$

$$f_{N_i}(t) = \frac{\beta_2}{\sqrt{2\pi}\Phi(t)}[-\log\phi(t)]^{(\beta_2-1)}\exp[-(-\log\phi(t))^{\beta_2}]\exp\left[-\frac{t^2}{2}\right]. \qquad (3)$$

As expected the two densities depend on the shape parameter β_2. Since the regression parameters are assumed to be known, they cancel out which may not be the case when they are estimated. We will come back to this issue in the next subsection. When $\beta_2 = 1$, the exponential model is correctly assumed and the two densities coincide with the reference standard Logistic and Normal distribution, respectively.

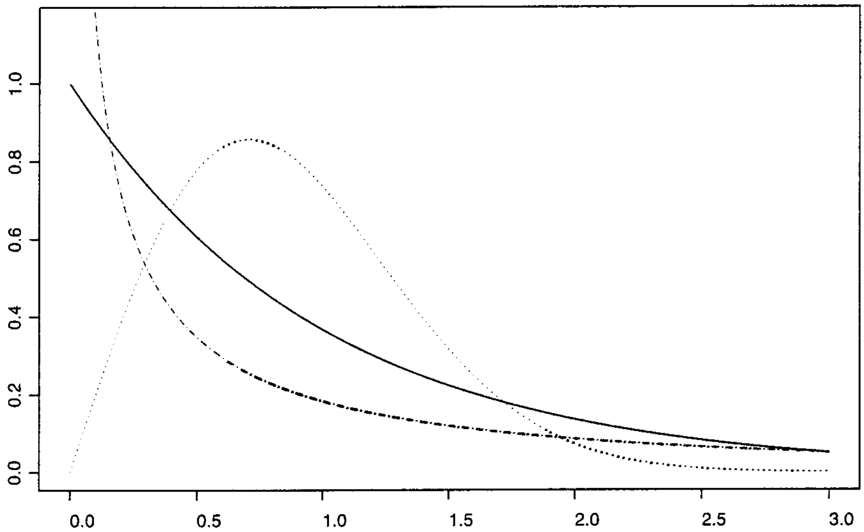

Fig. 2 – Baseline densities: —, exponential; ..., Weibull ($\beta_2 = 2$); – · –, Weibull ($\beta_2 = 0.5$).

Figures 2 and 3 show the baseline Weibull densities and the corresponding log-odds densities for β_2 equal to 0.5, 1 and 2, *i.e.*, for an increasing, a constant and a decreasing baseline hazard, respectively. Clearly a misspecified baseline hazard mainly affects the expected variability of log-odds residuals, with a negligible effect on the location of the distribution. This implies that methods based on the expected mean value of residuals will fail to detect it.

With respect to the reference density (solid line in Figure 3), log-odds residuals appear to be much more concentrated in the case of an increasing hazard, while a higher variability results from a decreasing baseline hazard. This behavior can be better understood looking at the densities in Figure 2. The Weibull density for $\beta_2 = 1.5$ (and more generally for $\beta_2 > 1$, *i.e.*, in the case of an increas-

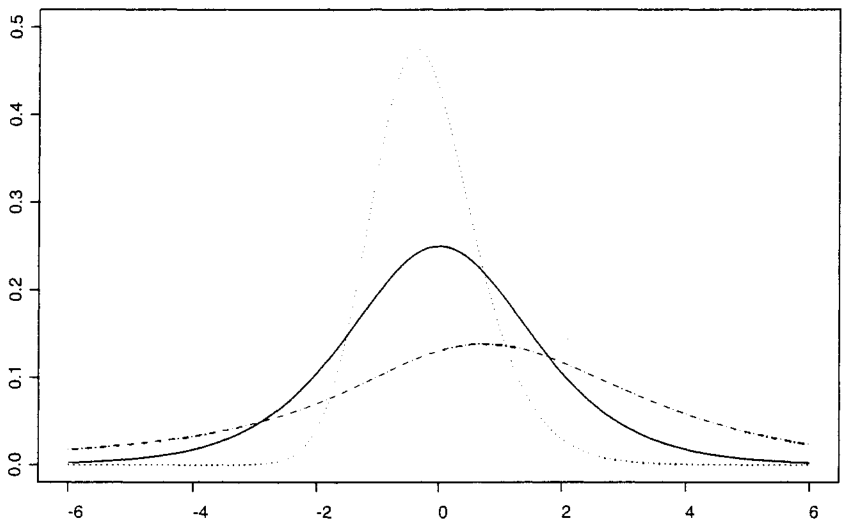

Fig. 3 – Pdf of log-odds residuals: —, assuming a correctly specified model; ..., misspecifying an exponential model, a Weibull ($\beta_2 = 2$) being the true model; – · –, misspecifying an exponential model, a Weibull ($\beta_2 = 0.5$) being the true model.

ing hazard) is more concentrated than the assumed exponential distribution in the neighborhood of 1. In other words, there is an excess of events in the central area that the exponential model cannot describe and that results in an excess of residuals. Conversely, in the case of a decreasing hazard ($\beta_2 = 0.5$), the Weibull distribution assigns greater probability mass to extreme values. Now the excess of residuals is in the tails of the distribution.

Similar arguments hold for normal-deviate residuals. It is worthwhile to remark that, with respect to the aim of this paper, log-odds and normal-deviate residuals play an identical, exchangeable role. For the sake of clarity, throughout the remainder of the paper the discussion will be carried on referring only to log-odds residuals.

3.2. *Unknown parameters: some asymptotic results*

In the previous Sub-Section we have assumed parameters to be known. We now investigate the effect of replacing their values with the corresponding mles, still focusing on residuals' distribution under the alternative hypothesis. Without loss of generality we assume that $\sum_1^n x_{ij} = 0 (j = 1,\ldots, m)$. In order to apply the asymptotic theory of mles, it is also assumed that $\lim_{n\to\infty} n^{-1} \sum_1^n \mathbf{x}_i\mathbf{x}_i'$ is a bounded positive-definite matrix.

It is worthwhile to remark that, for the models considered in this paper, the estimators of the regression coefficients $\theta_1,\ldots, \theta_m$ are asymptotically consistent, independently of distributional assumptions. However, when using a false model, their asymptotic efficiency (evaluated by the ratio of the determinants of the covariance matrices) can be substantially reduced (Pereira, 1978).

Assume that the Weibull regression model holds but the exponential model is falsely assumed. Let $(\beta_1, \beta_2, \mathbf{b}')$ be the true parameters, which refer to the Weibull regression model, and $(\delta_1^{(n)}, \mathbf{d}^{(n)\prime})$ the mles corresponding to the fitted exponential regression. Then, the following probability limits

$$\delta_1^{(n)} \to_P \beta_1 + \log \Gamma\left(\frac{\beta_2 + 1}{\beta_2}\right)$$

$$\mathbf{d}^n \to_P \mathbf{b}$$

can be derived by minimizing the Kullback-Leibler distance between the null and the alternative models (Appendix B).

Starting from these probability limits, the results outlined in Appendix B show that log-odds residuals converge to the random variables L_i^{W-E}, the density of which follows

$$f_{L_i^{W-E}}(t) = \frac{\beta_2 c^{\beta_2}}{1+e^t}\left[\log\frac{1+e^t}{e^t}\right]^{(\beta_2-1)} \exp\left[-\left(\left(\log\frac{1+e^t}{e^t}\right)c\right)^{\beta_2}\right] \tag{4}$$

where

$$c = \Gamma\left(\frac{\beta_2+1}{\beta_2}\right).$$

As expected, the asymptotic distribution depends on the shape parameter of the true baseline Weibull density. Notice that (4) coincides with the corresponding density (2), but for the presence of c. Its effect can be evaluated by comparing, in Figure 4, the asymptotic distribution of log-odds residuals with the corresponding density for known parameters, assuming $\beta_2 = 2$.

In the case of unknown parameters, the misspecification of the baseline hazard is still reflected by an anomalous concentration of residuals. An additional shift of the asymptotic density towards the left side can be explained by the presence of the additional term c in the probability limit of $\delta_1^{(n)}$. The histogram in the background is based on 5000 simulated trials, assuming a sample size of 20. Values of a single covariate x_1 were assumed binomially distributed with Prob(x_1 = 1) = Prob(x_1 = 2) = 0.5. The corresponding survival times were generated from a Weibull distribution, setting $\beta_2 = 2$ $\beta_1 = 1.39$ and $b_1 = -0.29$. An exponential model was fitted to the generated data and log-odds and normal deviate residuals of the last subject of each trial were recorded. The approximation to the asymptotic density appears satisfactory even for such a limited sample size.

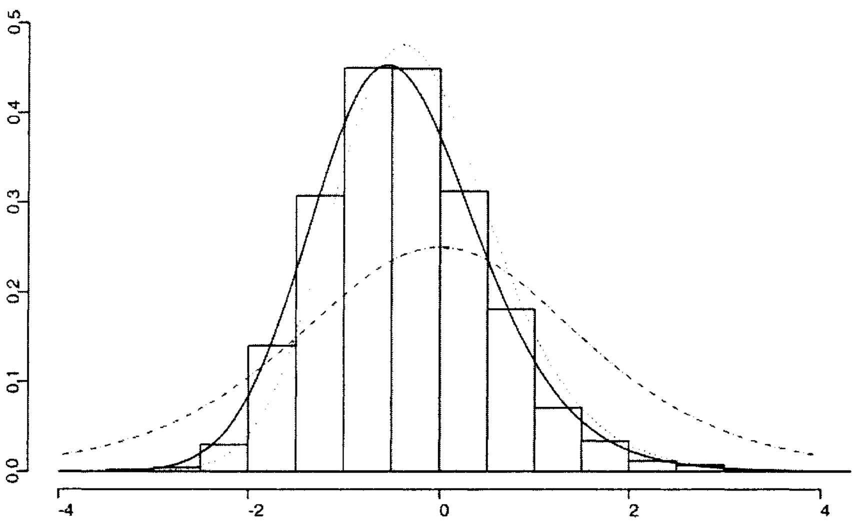

Fig. 4 – Distribution of log-odds residuals assuming the exponential regression model while theWeibull model holds: —, asymptotic pdf; …, pdf assuming known parameters; – · –, reference logistic pdf.

4. An application

The following example illustrates the use of the results above. It is a study of the lifetimes of Klevar 49/Epoxy spherical vessels that are subjected to a constant sustained pressure until vessel failure have been made. The NASA space shuttle uses Klevar 49/Epoxy spherical pressure vessels in a sustained pressure mode throughout the usage life of the vessels and several commercial application are also subjected to this service condition. The study was done to generate baseline data and to predict vessel life under different levels of pressure. Four data sets are considered here, at stress levels decreasing from 90% to 60%. An exponential model was fitted to each data set and log-odds residuals from the fitted models are shown in Figure 5.

Both at 90% and 80% stress level the residuals' empirical distributions are very close to the logistic reference density, providing evidence in favor of the null hypothesis of constant hazard. Conversely, at low stress levels, the residuals' distributions clearly departs from the reference density, suggesting the presence of a monotone increasing hazard; the lower the stress level, the stronger the evidence.

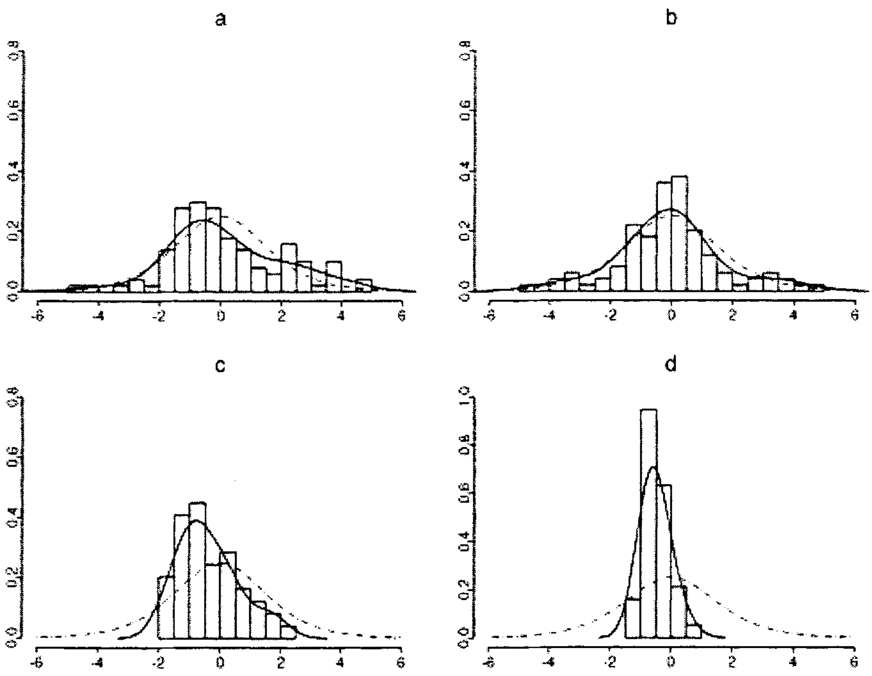

Fig. 5 – Log-odds residuals for the Klevar 49/Epoxy study: (a) 90% stress level; (b) 80% stress level; (c) 70% stress level; (d) 60% stress level.

In order to improve the graphical understanding, in Figure 6 the kernel estimates of the residuals' empirical distributions have been over-imposed for different stress levels. Compare Figure 5 with Figure 3 where the residuals' pdfs corresponding to misspecified exponential models are shown. The empirical distribution at 60% stress level clearly resembles the expected distribution in case of monotone increasing hazard: it is much more concentrated than the reference logistic density, showing a slight shift towards the left side. This finding is consistent with the results obtained by Barlow, Toland and Freeman (1984) in a Bayesian framework.

Notice that the shape parameter of the Weibull distribution could be indirectly estimated by maximizing the likelihood function of observed residuals.

5. Censoring

Survival data are typically censored and it is important to determine how the residuals' distribution is affected by the presence of censoring. We assume a cor-

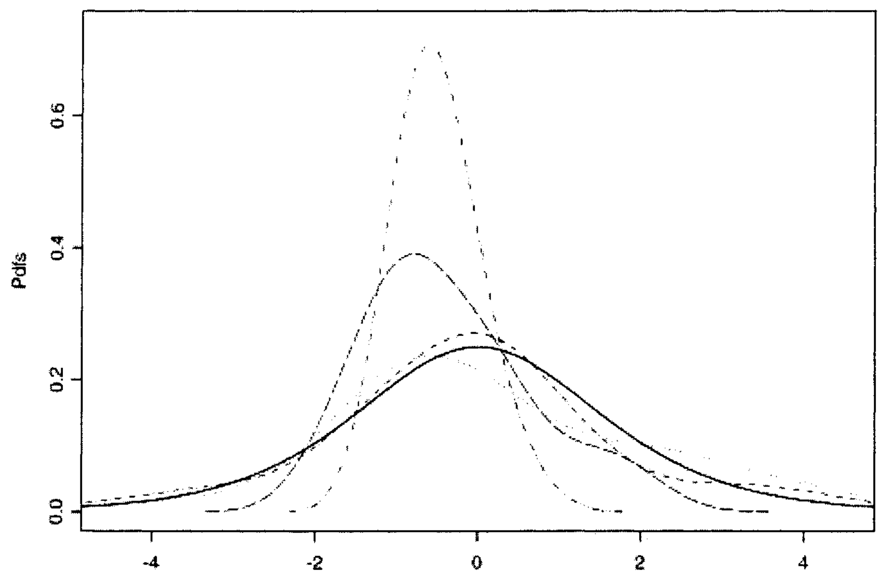

Fig. 6 – Empirical pdfs of log-odds residuals fot the Klevar 49/Epoxy study: —, reference logistic pdf; …, 90% stress level; – · –, 80% stress level; – –, 70% stress level; – ··· –, 60% stress level.

rectly specified model and an independent non-informative censoring mechanism. Note at first that, since the survival function is decreasing in time, right censored data result in left censored residuals.

Figure 7 shows the empirical distribution of log-odds residuals from 5000 simulated trials for $n = 100$. In each trial values of a single covariate x_1 were assumed binomially distributed with $Prob(x_1 = 1) = Prob(x_1 = 2) = 0.5$. The corresponding survival times were generated from a λ-exponential distribution, setting $\lambda = 1$ and $\lambda = 2$ for $x_1 = 1$ and $x_1 = 2$, respectively. An exponential model was fitted to the generated data and log-odds and normal deviate residuals of the last subject of each trial were recorded to guarantee independence of observations. The generated samples were censored using the procedure by Gehan and Thomas (1969) to model a clinical trial. Subjects were assumed to enter the study in a constant rate in an interval $(0, \tau)$ and then to fail according to the described survival time distributions. The value of τ, the time of analysis, was determined as in Green *et al.* (1979) to achieve an expected 70% of censored survival times.

Note the substantial departure of the empirical distribution of log-odds residuals from the reference standard logistic density: the left tail of the distribution, which corresponds to long survival times, is truncated and residuals are concentrated in the interval (0, 2). Actually, the empirical density can be re-

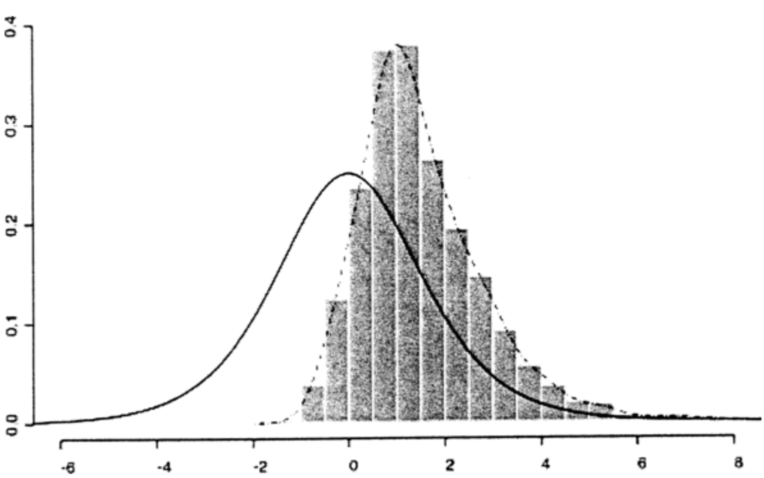

Fig. 7 – Empirical distribution of log-odds residuals assuming 70% of censoring: —, reference logistic density; – · –, kernel density estimate.

garded as a mixture of two distributions: while uncensored residuals can still be assumed to follow a logistic density, this is not the case for the censored ones. Note that even uncensored residuals cannot be regarded as representative of a logistic population since censoring is not acting uniformly on the positive real line.

Assume now that the baseline hazard has been wrongly specified: the bias due to censoring and the bias due to the misspecified baseline family overlap and it may be difficult to distinguish one from the other. The extent to which we can still recognize the effect of a wrong baseline family depends on both the percentage of censoring and the families of distributions considered.

The first proposal to accommodate residuals to censoring dates back to 1977 (Crowley and Hu, 1977). The main idea behind it is that the distribution of the unknown *true* residual, given $T_i > c_i$, is related to the uniform distribution of $S(T_i)$ in $[0, S(c_i)]$. Thus a censored residuals can be replaced by its conditional mean or median value. By denoting with l_i^c the observed censored residual, we have (Nardi and Schemper, 1999)

$$E[L_i \mid L_i \leq l_i^c] = l_i^c - \frac{1+e^{l_i^c}}{e^{l_i^c}} \log(1+e^{l_i^c}).$$

Figure 8 shows the empirical distribution of log-odds residuals, being censored residuals replaced by their conditional mean values. Due to the averaging process, the adjusted residuals tend to be less extreme then the corresponding unobservable L_i. It results in an anomalous concentration of the empirical distribution which still substantially departs from the reference density.

In order to avoid the reduced variability in adjusted residuals, we propose to randomly sample from the residuals' conditional distributions and to proceed in the spirit of Rubin's multiple imputation (Rubin, 1987). Assume that individual i is censored at c_i. Then, instead of replacing the censored residual with its conditional expectation, we randomly generate r normal deviate residuals from the distribution of $(L_i | T_i > c_i)$, r being the number of imputations. Note that this procedure can be applied to any diagnostic tool, provided that the sampling distribution is known. Each of the imputed residuals for individual i is weighted $1/r$. Figure 9 shows the empirical distribution of log-odds residuals for $r = 3$, assuming 70% of censoring. The left tail of the distribution is now reconstructed avoiding the concentration effect and the approximation to the logistic density is satisfactory, despite the high percentage of censoring. It is worth to remark that the imputation is done under the null hypothesis of a correctly specified model. This may lead to a conservative behaviour in assessing departures from model assumptions when the percentage of censoring is high.

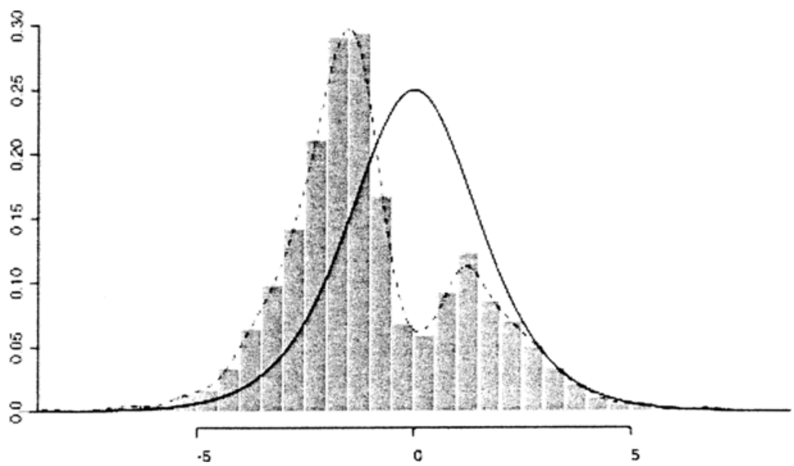

Fig. 8 – Empirical distribution of log-odds residuals assuming 70% of censoring (censored residuals have been replaced with their expected values): —, reference logistic density; – · –, kernel density estimate.

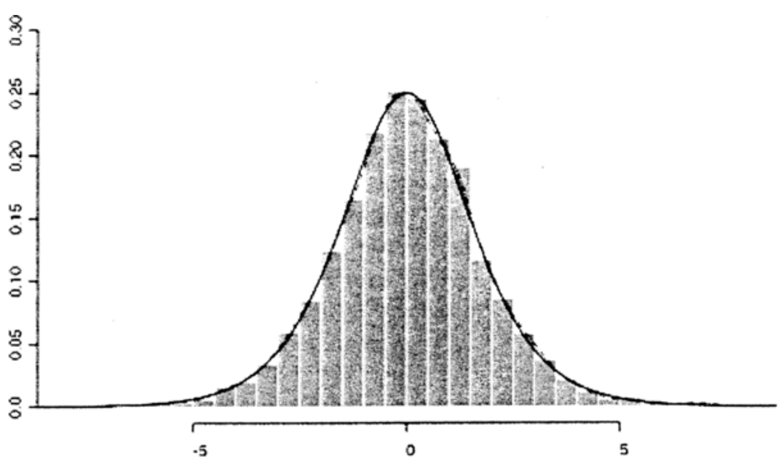

Fig. 9 – Empirical distribution of log-odds residuals assuming 70% of censoring (censored residuals have been replaced with imputed values): —, reference logistic density; – · –, kernel density estimate.

6. Discussion

We have proposed a graphical procedure to discriminate among alternative parametric models on the basis of residuals' properties. The idea of comparing the empirical distribution of residuals with their reference distribution, assuming a properly specified model, is not new. In 1997 Crowley and Hu (Crowley and Hu, 1977) and Kay (Kay, 1977) suggested to compare the empirical distribution of Cox and Snell residuals to the expected unit-exponential density. The proposal have had little success and some serious issues have arisen.

The comparison between residuals' empirical and reference distribution was intended as a global goodness of fit test for the Cox's model. Now, because Λ_0 is completely unspecified, this global test gives no information about the real fit of the model. If no covariates are in the model, $\hat{\Lambda}_0$ can be made to fit the data exactly; the order statistics of the residuals are then exactly the order statistics of the (possibly censored) unit exponential distribution, even if important covariates have been omitted. For details see Lagakos (1981) and Baltazar and Peña (1995).

The skewness of the reference exponential distribution makes a visual assessment of possible departures from the reference distribution difficult to be determined so that plots of residuals are not easy to interpret. Furthermore, even when a clear departure from the reference distribution is observed, there is no indication in what way the fitted model should be improved.

In our proposal Cox and Snell residuals have been replaced with log-odds and normal deviate residuals, both of them having a symmetric, unimodal reference distribution. Our goal is to verify the appropriateness of a fully parametric models with respect to the assumptions on the shape of the baseline hazards. In other words, our attention focuses on the shape parameter κ, the regression parameters playing the role of nuisance parameters. We have addressed the question how to proceed if the graphical inspection indicates that the model does not fit, by investigating the distribution of residuals under the alternative hypothesis.

Our study has been limited to the comparison between the exponential (the null hypothesis) and the Weibull model (the alternative hypothesis) since they are widely used in reliability analysis. However the idea can be applied to any couple of alternative parametric models.

It is worth to remark that the residuals' density under the alternative hypothesis depends only on the shape parameter, *i.e.*, nuisance parameters have be eliminated by moving on the scale of residuals. Since residuals' distributions are known both under the null and the alternative hypothesis, a formal test could be derived. When the competing models are nested, as in our case, the classic likelihood ratio test guarantees already high efficiency. Conversely in the case of separate families of hypothesis a test based on residuals properties could be a valid alternative to the modified likelihood ratio test proposed by Cox in the late 1961 (Cox, 1962).

Appendix

Appendix A: Proofs of residuals' densities assuming known parameters

Constant versus monotone hazard

Suppose the Weibull regression model

$$\log T_i = \beta_1 + \sum_1^m b_j x_{ij} + \log T_{0i}$$

$$f_{T_i}^W(t) = \beta_2\, t^{\beta_2 - 1} \exp[-t^{\beta_2}]$$

holds. Then

$$f_{T_i}^W(t) = \beta_2 (B_i\, t)^{\beta_2 - 1} \exp[-(B_i\, t)^{\beta_2}] B_i$$

where

$$B_i = \exp\left[-\left(\beta_1 + \sum_{j=1}^m b_j x_{ij}\right)\right]$$

If we incorrectly assume an exponential density as baseline distribution we have

$$S_i = S(T_i) = \exp\left[-\left(\delta_1 + \sum_1^m d_j x_{ij}\right) T_i\right]$$

Since parameters are assumed to be known and the linear predictor correctly specified $\delta_1 = \beta_1$, $b_j = d_j$, $j = 1,\ldots, m$. Then, being $S(T_i)$ a monotone transformation, the density of S_i can be obtained by applying the change of variable formula

$$f_{S_i}(t) = \frac{\beta_2}{t} (-\log t)^{\beta_2 - 1} \exp[-(\log t)^{\beta_2}]$$

The densities (2) and (3) of Section 3 can be derived by applying the logit and the probit transformation to S_i and taking into account that both the transformations are monotone.

Appendix B: Proofs of residuals' asymptotic distributions

Throughout this appendix we assume that $\lim_{n\to\infty} n^{-1}\sum_1^n \mathbf{x}_i\mathbf{x}_i'$ is a bounded positive-definite matrix. Under this assumption, the models considered here fulfill the regularity conditions required in the asymptotic theory of mles.

Constant versus monotone hazard

Suppose theWeibull regression model (see Appendix A) holds, while an exponential baseline density is incorrectly assumed. The corresponding pdfs are

$$\begin{aligned} f_{T_i}^W(t) &= \beta_2 (B_i t)^{\beta_2 - 1} \exp[-(B_i t)^{\beta_2}] B_i \\ &= f_{T_{0i}}^W (B_i t) B_i \end{aligned}$$

where

$$B_i = \exp\left[-\left(\beta_1 + \sum_{j=1}^{m} b_j x_{ij}\right)\right]$$

and

$$f_{T_i}^E(t) = \exp[-(D_i^{(n)} t)] D_i^{(n)} = f_{T_{0i}}^E (D_i^{(n)} t) D_i^{(n)} \tag{6}$$

where

$$D_i^{(n)} = \exp\left[-\left(\delta_1^{(n)} + \sum_{j=1}^{m} d_j^{(n)} x_{ij}\right)\right]$$

Let $\delta_1^{(n)}$ and $\mathbf{d}^{(n)}$ denote the mles of the corresponding unknown parameters. As $n \to \infty$, $\delta_1^{(n)}$ and $\mathbf{d}^{(n)}$ converge to those values that minimize the Kullback-Leibler divergence between the true and the falsely assumed model. This is equivalent to maximize the expected value of the log-likelihood function for the exponential model, the expectation being under the true Weibull model (see also Pereira, 1978). We have

$$
\begin{aligned}
E_W[l^E] &= \sum_{i=1}^{n} \int_0^{+\infty} \log f_{T_i}^{E}(t)\, f_{T_i}^{W}(t)\, dt \\
&= \sum_{i=1}^{n} \int_0^{+\infty} [\log D_i^{(n)} + \log f_{T_{0i}}^{E}(D_i^{(n)}\, t)]\; B_i\, f_{T_{0i}}^{W}(B_i\, t)\, dt \\
&= \sum_{i=1}^{n} \left[\log D_i^{(n)} - D_i^{(n)} \int_0^{+\infty} t\, B_i\, f_{T_{0i}}^{W}(B_i\, t)\, dt \right] \\
&= \sum_{i=1}^{n} \left[\log D_i^{(n)} - \frac{D_i^{(n)}}{B_i} \Gamma\left(\frac{\beta_2 - 1}{\beta_2} \right) \right]
\end{aligned}
\tag{7}
$$

By computing the partial derivatives with respect to $\delta_1^{(n)}$ and $\mathbf{d}^{(n)}$, we get the following system of equations

$$
\frac{\partial}{\partial d_1^{(n)}} E_W[l^E] = \sum_{i=1}^{n} \left[\frac{D_i^{(n)}}{B_i} \Gamma\left(\frac{\beta_2 - 1}{\beta_2} \right) - 1 \right] = 0
$$

$$
\frac{\partial}{\partial d_{ij}^{(n)}} E_W[l^E] = \sum_{i=1}^{n} \left[\left(\frac{D_i^{(n)}}{B_i} \Gamma\left(\frac{\beta_2 - 1}{\beta_2} \right) - 1 \right) x_{ij} \right] = 0 \quad j = 1, \ldots, m
$$

leading to the probability limits

$$
\delta_1^{(n)} \rightarrow_P \quad \beta_1 + \log \Gamma\left(\frac{\beta_2 + 1}{\beta_2} \right) \tag{8}
$$

$$
\mathbf{d}^n \quad \rightarrow_P \quad \mathbf{b} \tag{9}
$$

Starting from these results the following probability limits hold

$$
\begin{aligned}
D_i^{(n)} = \exp\left[-\left(\delta_1^{(n)} + \sum_{j=1}^{m} d_j^{(n)}\, x_{ij} \right) \right] &\rightarrow_P \quad \exp\left[-\left(\beta_1 + \log \Gamma\left(\frac{\beta_2 + 1}{\beta_2} \right) + \sum_{1}^{m} b_j\, x_{ij} \right) \right] \\
&= D_i^{W-E} \\
(T_i, D_i^{(n)}) &\rightarrow_P \quad (T_i, D_i^{W-E}) \\
S_i^{(n)} = \exp[-(T_i, D_i^{(n)}) &\rightarrow_P \quad \exp[-(T_i, D_i^{W-E})] = S_i
\end{aligned}
$$

The results above are justified by the continuity of the applied mappings and by the convergence of double random variables given the convergence of the marginal components. Notice that T_i follows here the true Weibull model, while $S_i^{(n)}$ is defined according to the falsely assumed exponential model. By transforming T_i according to S_i we get the limiting distribution of $S_i^{(n)}$ that follows

$$f_{S_i}(t) = \frac{\beta_2 c}{t}(-c\log t)^{\beta_2 - 1}\exp[-(-c\log t)^{\beta_2}] \tag{10}$$

where

$$c = \Gamma\left(\frac{\beta_2 + 1}{\beta_2}\right)$$

Finally, applying the logit transformation to $S_i^{(n)}$, we have

$$L_i^{(n)} = \log\frac{S_i^{(n)}}{1 - S_i^{(n)}} \to_P \log\frac{S_i}{1 - S_i} = L_i^{W-E}$$

where

$$f_{L_i^{W-E}}(t) = \frac{\beta_2 c^{\beta_2}}{1+e^t}\left[\log\frac{1+e^t}{e^t}\right]^{(\beta_2 - 1)} \exp\left[-\left(\left(\log\frac{1+e^t}{e^t}\right)c\right)^{\beta_2}\right] \tag{11}$$

REFERENCES

BALTAZAR-ABAN, I. and PEÑA, E. A (1995), Properties of hazard-based residuals and implications in model diagnostics. *Journal of the American Statistical Association*, 90, 185-197.

COX, D. R. (1961), Test of separate families of hypotheses. *Proc. 4th Berkeley Symp.*, 1, 105-123.

COX, D. R. and SNELL, E. J. (1968), A general definition of residuals (with discussion). *Journal of the Royal Statistical Society*, B, 30, 248-275.

CROWLEY, J. and HU, M. (1977), Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association*, 72, 27-36.

GEHAN, E. A. and THOMAS, D. G. (1969), The performance of some two-sample tests in small samples with and without censoring. *Biometrika*, 56, 127-132.

GREEN, S. B., LININGER, L., GAIL, M. H. and BYAR, D. P. (1979), Comparison of four tests foe equality of survival curves in the presence of stratification and censoring. *Biometrika*, 66, 417-428.

KAY, R. (1977), Proportional hazards regression models and the analysis of censored survival data. *Applied Statistics*, 26 (3), 227-237.

LAGAKOS, S. W. (1981), The graphical evaluation of explanatory variablesin proportinal hazard regression models. *Biometrika*, 68, 93-98.

NARDI, A. and SCHEMPER, M. (1999), New residuals for cox regression and their application to outlier screening. *Biometrics*, 55, 523-529.

PEREIRA, B. (1978), Test and efficiencies of separate regression models. *Biometrika*, 65 (2), 319-327.

RUBIN, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.