University of Illinois — Department of Economics
Fall 2001 — Roger Koenker

Economics 478
Lecture 7
Residuals for the Cox Model

The martingale formulation of counting processes leads to a natural notion of residuals for the Cox model. Recall we have

$$M_i(t) = N_i(t) - A_i(t)$$

where

$$\begin{aligned} A_i(t) &= \int_0^t Y_i(s)\exp\{x_i(s)'\beta\}\lambda_0(s)ds \\ &= \int_0^t Y_i(s)\lambda_i(s)ds \end{aligned}$$

So it is natural to define the residual process

$$\begin{aligned} \hat{M}_i(t) &= N_i(t) - \hat{A}_i(t) \\ &= N_i(t) - \int_0^t Y_i(s)\hat{\lambda}_i(s)ds \\ &= N_i(t) - \int_0^t Y_i(s)d\hat{\Lambda}_i(s) \end{aligned}$$

Here

$$d\hat{\Lambda}_i(t) = e^{x_i(t)'\beta}d\hat{\Lambda}_0(t)$$

and we may take

$$\hat{\Lambda}_0(t) = \int_0^t \frac{d\bar{N}(s)}{\sum_{j=1}^n Y_j(s)e^{x_j(s)\hat{\beta}}}$$

Note that if $\hat{\beta} = 0$, then $\hat{\Lambda}_0(t)$ is the Nelson-Aalen estimator.

In the simplest (typical) case where we don't have time varying covariates we have

$$\hat{M}_i = N_i - e^{x_i'\hat{\beta}}\hat{\Lambda}_0(t_i)$$

and is just the difference between observed and expected evaluated at the even time for each observation. Like ordinary regression residuals $EM_i = 0$ and this condition is imposed by fitting so $\sum \hat{M}_i = 0$. Due to this there is a slight negative correlation induced across residuals despite the fact that we assume $EM_iM_j = 0$.

As we have seen it is possible to consider Martingale transforms

$$R_i = \int h_i(t)dM_i(t)$$

where $h_i(\cdot)$ denotes a left continuous predictable process. This can be done for martingale residuals so we have

$$\hat{R}_i(t) = \int_0^t h_i(s) d\hat{M}_i(s)$$

and we have

$$\hat{\text{Var}}(R_i) = \int_0^t Y_i(s) h_i(s) h_i(s)' d\hat{\Lambda}_i(s)$$

The usual martingale residual is just the special case $h_i(s) \equiv 1$. Barlow and Prentice suggest three other choices: $h_i(t) = t$, corresponding to a linear time trend; $h_i(t) = x_i(t)'\hat{\beta}$ which corresponds to some misspecification of covariate effect; the third is $h_i(t) = x_i(t) - \bar{x}(t, \hat{\beta})$ where

$$\bar{x}(t, \hat{\beta}) = \frac{\sum Y_i(s) e^{x_i(s)'\hat{\beta}} x_i(s)}{\sum Y_i(s) e^{x_i(s)\hat{\beta}}}$$

This is a weighted mean up to time t of the covariate process with weights $\sum Y_i(s) e^{x_i(s)'\beta}$.

When there are time varying covariates we have multiple martingale residuals per subject but because they are defined as integrals we can sum them to obtain a martingale residual for each subject, if we wish.

The obvious hope is that M-residuals will be helpful in diagnosing misspecification. For some purposes this hope is not very realistic. For example, as noted by T&G in Section 4.2.3, the distributional properties are not readily tested, we might add that the usual regression checks are not really very good on this point either. But misspecification of covariate effects offers a bit more optimistic view point.

Consider the decomposition

$$\begin{aligned} \hat{M}_i(t) &= N_i(t) - A_i(t) + A_i(t) - \hat{A}_i(t) \\ &= M_i(t) + (A_i(t) - \hat{A}_i(t)) \end{aligned} \quad (1)$$

where the second term represents the error in estimating the compensator term of the process. Following the discussion in T&G, let's consider the case of a single covariate: we specify the effect of the covariate as linear, but really the model should be

$$\lambda_i(t) = \lambda_0(t) e^{f(x_i)}$$

In settings like this we have a pseudo-true value of the parameter, β, say β^*, that minimizes the KL divergence between the parametric model and the true process. (In classical least squares regression this minimizes

$$\beta^* = \arg\min_\beta E_x (x_i'\beta - f(x_i))^2$$

where E_x denotes expectation with respect to the marginal distribution of x.)

Let

$$\rho(t) = \frac{EY(t)e^{f(x)}}{EY(t)e^{x\beta^*}} = \frac{E(e^{f(x)}|Y(t)=1)}{E(e^{x\beta^*}|Y(t)=1)}$$

denote the ratio of the mean risk scores at time t, for the true and pseudo true models. Then

$$\begin{aligned} \hat{A}_i(t) &= \int_0^t Y_i(s)e^{x_i\hat{\beta}}d\hat{\Lambda}_0(s) \\ &= \int_0^t Y_i(s)e^{x_i\beta^*}\rho(t)\lambda_0(s)ds + o_p(1) \end{aligned} \tag{2}$$

To "justify" this approximation, note that

$$\begin{aligned} d\hat{\Lambda}_0(s) &= \frac{\Sigma_j dN_j(s)}{\Sigma_i Y_j(s)e^{x_j\hat{\beta}}} \\ &= \frac{n^{-1}\Sigma_j dM_j(s) + dA_j(s)}{n^{-1}\Sigma_j Y_j(s)e^{x_j'\hat{\beta}}} \\ &\to \frac{\Sigma Y_j(s)e^{f(x_j)}\lambda_0(s)ds}{\Sigma Y_j(s)e^{x_j'\beta^*}} \\ &\approx \rho(s)\lambda_0(s)ds \end{aligned}$$

Now write our initial decomposition (1) as

$$\begin{aligned} \hat{M}_i(t) = M_i(t) \quad &+ \quad e^{f(x_i)}\int_0^t Y_i(s)\lambda_0(s)ds \\ &- \quad e^{x_i'\beta^*}\int_0^t Y_i(s)\rho(s)\lambda_0(s)ds + o_p(1) \end{aligned} \tag{3}$$

since $\hat{\beta} \to \beta^*$. Combining (2) and (3) we have, evaluating all processes at a common point, say, the termination date t of the experiment,

$$E(\hat{M}|x) = e^{f(x)}\int_0^t \pi(x,s)\lambda_0 ds - e^{x'\beta^*}\int_0^t \pi(x,s)\rho(s)\lambda_0(s)ds,$$

where $\pi(x,s) = E(Y_i(s)|X=x)$ is the probability of being at risk at time s given that you have covariate value x.

From (2)

$$E\hat{A}(t|x) = e^{x\beta^*}\int_0^t \pi(x,s)\rho(s)\lambda_0(s)ds$$

so the ratio

$$\frac{E\hat{M}(t|x)}{E\hat{A}(t|x)} = \frac{e^{f(x)}}{e^{x'\beta}}\cdot\gamma(x,t) - 1$$

where

$$\gamma(x,t) = \frac{\int_0^t \rho(s)\pi(x,s)ds}{\int_0^t \pi(x,s)ds}$$

and should be "nearly" $\perp\!\!\!\perp$ of x. Why?

Thus, a diagnostic plot of

$$G(x) = \log\left(\frac{E\hat{M}(t|x)}{E\hat{A}(t|x)} + 1\right) = f(x) - x'\beta^* + \log\gamma$$

e.g. $G(X) + x'\hat{\beta}$ vs x should reveal some missed nonlinearity.

To respond to the "why?" above. T&G suggest that this holds when the null model holds and censoring doesn't depend on x.

All of this can be done easily in R. The default version of the `resid(·)` function gives the Martingale residuals and they can then be plotted versus covariates. To explore nonlinearity lowess or other smoothed fits to the resulting scatter plots may be useful.