Residuals

Schoenfeld Residuals

Assume p covariates and n independent observations of time, covariates, and censoring, which are represented as (t_i, x_i, c_i), where $i = 1, 2, ..., n$, and $c_i = 1$ for uncensored observations and zero otherwise. Schoenfeld residuals "are based on the individual contributions to the derivative of the log partial likelihood" (Hosmer and Lemeshow 1999, 198). To derive the Schoenfeld residuals, one takes the derivative for the kth covariate,

$$\begin{aligned} \frac{\partial L_p(\beta)}{\partial \beta_k} &= \sum_{i=1}^{n} c_i \left\{ x_{ik} - \frac{\sum_{j \in R(t_i)} x_{jk} e^{x'_j \beta}}{\sum_{j \in R(t_i)} e^{x'_j \beta}} \right\} \\ &= \sum_{i=1}^{n} c_i \left\{ x_{ik} - \overline{x}_{w_i k} \right\}, \end{aligned} \tag{1}$$

where

$$\overline{x}_{w_i k} = \frac{\sum_{j \in R(t_i)} x_{jk} e^{x'_j \beta}}{\sum_{j \in R(t_i)} e^{x'_j \beta}}. \tag{2}$$

Hosmer and Lemeshow (1999, 198) show, the estimator of the Schoenfeld residual for the ith subject on the kth covariate are then obtained by substituting the partial likelihood estimator of the coefficient, $\hat{\beta}$:

$$\hat{r}_{S_i k} = c_i(x_{ik} - \hat{\overline{x}}_{w_i k}), \tag{3}$$

where

$$\hat{\overline{x}}_{w_i k} = \frac{\sum_{j \in R(t_i)} x_{jk} e^{x'_j \beta}}{\sum_{j \in R(t_i)} e^{x'_j \beta}}. \tag{4}$$

Highly nonobvious, but there *is* some intuition here!

• Schoenfeld residual can be thought of as the observed minus the expected values of the covariates at each failure time.

• If the residual exhibits a random (i.e. unsystematic) pattern at each failure time, then this gives evidence the covariate effect is not changing with respect to time—precisely the PH assumption. If it is systematic, it suggests that as time passes, the covariate effect is changing.

• Why? Because if the PH property holds, then we would expect the difference between covariate values at failure times versus a weighted average of the covariate values to display no temporal trends. In residual plots, we might expect the slope of the (re-scaled) Schoenfeld residuals with respect to time should be zero. This is the basic idea behind graphical methods.

Cox-Snell Residuals

Basic issue involving the use of these kinds of residuals is goodness-of-fit of the Cox model. If a Cox model, given by

$$h_i(t) = h_0(t)e^{\beta' \mathbf{x}} \tag{5}$$

holds, then estimates of survival times from the posited model (i.e $\hat{S}_i(t)$) should be similar to the true value of $S_i(t)$.

Enter Cox-Snell residuals. The Cox-Snell residual is given by

$$r_{CS_i} = \exp(\hat{\beta}' \mathbf{x_i}) \hat{H}_o(t_i), \tag{6}$$

where $\hat{H}_o(t_i)$ is the estimated integrated baseline hazard (or cumulative hazard). This is the Nelson-Aalen estimator shown in your first handout.

From some results shown in the book (not here), the Cox-Snell residuals can also be written as:

$$-\log \hat{S}(t_i)$$

that is, the log of the estimated survival time.

Issue: if Cox model "fits" then the residuals should be distributed as unit exponential (i.e. "behave" as if they are from a unit exponential distribution).

Procedure: Compute the Kaplan-Meier estimator *on the Cox-Snell residuals*. From these estimates, compute the integrated hazard. Plot the integrated hazard based on the residuals against the hazard rate estimates backed out of the Cox model. If model "holds", plot should have a 45-degree slope.

Counting Process

This concept gets us to martingale residuals (and also helps us see how TVCs can be incorporated and how discrete-models can be applied).

The Triple: i enters the risk period at time t_0 and is observed to time t at which point the individual is observed as either failing or surviving (or being right-censored). As usual, we denote this as δ.

The observed survival time is fully described by the triple:

(t_0, t, δ).

That is, i is observed from $(t_0, t]$ and either fails ($\delta = 1$) or survives ($\delta = 0$).

To construct the triple, we need to know this:

• First, the analyst must know when the observation entered the risk period (i.e. t_0).

• Second, the analyst must know at what time the TVC changes values.

• Third, the analyst must know when (or if) the observation failed (i.e. δ).

Examples: MIDs

Table 1: Example of Counting Process Data with a Yearly TVC

Year	Dyad Id.	Interval (Start, Stop]	Censoring Indicator	Economic Growth	Contiguity Status
1951	2020	(0, 1]	0	0.01	1
1952	2020	(1, 2]	0	0.03	1
1953	2020	(2, 3]	0	0.02	1
1954	2020	(3, 4]	0	0.01	1
1955	2020	(4, 5]	0	0.01	1
1956	2020	(5, 6]	0	0.01	1
1957	2020	(6, 7]	0	0.02	1
1958	2020	(7, 8]	0	-0.01	1
1959	2020	(8, 9]	0	0.00	1
1960	2020	(9, 10]	0	0.00	1
1961	2020	(10, 11]	0	0.00	1
1962	2020	(11, 12]	0	0.01	1
1963	2020	(12, 13]	0	0.02	1
1964	2020	(13, 14]	0	0.04	1
1965	2020	(14, 15]	0	0.04	1
1966	2020	(15, 16]	0	0.04	1
1967	2020	(16, 17]	0	0.04	1
1968	2020	(17, 18]	0	0.03	1
1969	2020	(18, 19]	0	0.02	1
1970	2020	(19, 20]	0	0.01	1
1971	2020	(20, 21]	0	0.01	1
1972	2020	(21, 22]	0	0.01	1
1973	2020	(22, 23]	1	0.03	1
1974	2020	(23, 24]	0	0.01	1
1976	2020	(25, 26]	0	0.00	1
1977	2020	(26, 27]	0	0.02	1
1978	2020	(27, 28]	0	0.04	1
1979	2020	(28, 29]	1	0.03	1
1980	2020	(29, 30]	0	0.00	1
1981	2020	(30, 31]	0	0.00	1
1982	2020	(31, 32]	0	-0.01	1
1983	2020	(32, 33]	0	0.00	1
1984	2020	(33, 34]	0	0.02	1
1985	2020	(34, 35]	0	0.04	1
1961	2041	(0, 1]	0	-0.05	0
1962	2041	(1, 2]	0	-0.01	0
1963	2041	(2, 3]	1	-0.01	0
1964	2041	(3, 4]	0	0.00	0

Data are from Oneal and Russet (1997).

Each period has its own record of data (because the TVC changed values at each observation time). In other instances, analysts will be interested in TVCs that change intermittently across time.

Warchests:

This data set records the duration (in weeks) that passes from the start of a campaign cycle until the emergence of a high quality challenger against the incumbent.

The TVC in this example does not change with each "click of

Table 2: Example of Event History Data Set with Time-Varying Covariates

Case I.D.	Weeks to Event	Southern District	Incumbent's Party	1988 Vote	"War Chest" (in Millions)	Cesoring Indicator
100	26	0	0	.62	.003442	0
100	50	0	0	.62	.010986	1
201	26	1	0	.59	.142588	0
201	53	1	0	.59	.15857	0
201	65	1	0	.59	.202871	0
201	75	1	0	.59	.217207	0
516	26	0	1	.79	.167969	0
516	53	0	1	.79	.147037	0
516	65	0	1	.79	.164970	0
516	72	0	1	.79	.198608	1
706	26	0	0	.66	.139028	0
706	53	0	0	.66	.225633	0
706	65	0	0	.66	.225817	0
706	78	0	0	.66	.342801	0
706	83	0	0	.66	.262563	1
905	26	1	0	1.00	.211122	0
905	53	1	0	1.00	.270816	0
905	65	1	0	1.00	.262153	0

These data are taken from Box-Steffensmeier's (1996) data set on challenger entry and House war chests. The "war chest" variable is expressed in millions of dollars. The status indicator is coded 1 if a quality challenger emerged and 0 if not (denoting a right-censored case).

the clock," as in the case of the economic growth in the MIDs data. The other variables in the table are time-independent—their values do not change within the observation plan.

Martingale Residuals

Consider the linear function given by

$$\delta_i(t) = H_i(t) + M_i(t). \tag{7}$$

$\delta_i(t)$ denotes the number of observed events (i.e. it's a count) that occur at each failure time t.

Result shows the observed counts is a function of two components: a systematic component given by $H_i(t)$ and a random component given by $M_i(t)$.

Rewriting the integrated hazard to separate out the baseline integrated hazard from the covariates. This gives

$$H_i(t) = \exp(\hat{\beta}'\mathbf{x_i})\hat{H}_o(t_i). \tag{8}$$

This function is kind of like our summary of the total number of expected events. (Why expected? Remember the hazard is a function of covariates: i.e. a function of factors used to "predict" events.) Substituting equation (8) into (7), we get

$$\delta_i(t) = \exp(\hat{\beta}'\mathbf{x_i})\hat{H}_o(t_i) + M_i(t). \tag{9}$$

The martingale is the stochastic component and in residual form gives

$$M_i(t) = \delta_i(t) - \exp(\hat{\beta}'\mathbf{x_i})\hat{H}_o(t_i). \tag{10}$$

Implication: In "OLS terms" this residual is similar to the difference between the observed number and expected number of events.

In passing, note that the Cox-Snell residuals (repeating from an earlier slide) are given by

$$r_{CS} = \exp(\hat{\beta}'\mathbf{x_i})\hat{H}_o(t_i). \tag{11}$$

Verify this result is equivalent to equation (8). Bottom line: the expected count *is* the Cox-Snell residual. Put differently, the Cox-Snell residual is equivalently thought of as an expected count.

Martingale residuals are useful in evaluating functional form of a covariate.