Issues in Model Selection

A wide variety of models for duration data have been discussed. Each of the approaches have certain (and even uncertain!) advantages and disadvantages.

Cox vs. Parametric vs. Discrete

So many choices!! (Too many????)

Let's consider the "pros" and "cons".

…and then consider a pretty neat little hybrid model.

Advantages and Disadvantages of Modeling Strategies

Parametric Models

The principal is the ability of the model to provide parameter estimates while simultaneously producing a relatively simple and easy-to-interpret characterization of the baseline hazard rate.

The main drawback to the parametric approach is the potential for arbitrary decisions regarding the nature of the baseline hazard rate.

Cox models

A useful question to ask is, in what context is a parametric duration model naturally preferred over its alternatives, chiefly the Cox model?

The simple answer depends on the extent to which the analyst is interested in:

1) making explicit inferences regarding duration dependency.

2) the degree to which the researcher believes time dependency is substantively meaningful (as opposed to thinking of it as a nuisance).

One clear advantage of parametric models is predicting or forecasting what will happen beyond the "follow-up" period of the data—that is, making out-of-sample predictions.

If one thinks duration dependency is a nuisance and if interest centers primarily on the relationship between covariates and the hazard rate, then it is difficult to find settings where a parametric model would be preferred over a Cox model.

In practice (and we've seen this throughout), discrimination between a Cox and Weibull model may be difficult, particularly if there are a lot of failures.

Discrete-Time Models Revisited

The main advantage of discrete-time models is that they are widely used and well understood by social scientists.

BUT, there are some disadvantages of discrete-time models.

The problems parallel parametrics.

That is, the form of the duration dependency must be explicitly accounted for (or tested for).

Having to specify the nature of the duration dependency is easy to deal with.

Certainly, *ignoring* duration dependency in grouped binary data is problematic.

Cox redux?

If duration dependency is regarded as a nuisance, then the question naturally arises as to the appropriateness of discrete-time models like the logit model, when compared to the Cox alternative.

The Box-Steffensmeier—Jones strong argument: there is no reason we can think of why one would naturally prefer a standard logit model over the conditional logit Cox model *if duration dependency is assumed to be a nuisance parameter.*

. . . *but* . . .

Reconsidering the Cox Model

(Repeating) The primary advantage of the Cox model is simple: the relationship between covariates and the hazard rate can be estimated without having to make assumptions about the nature and shape of the baseline hazard rate.

But cautions apply!
In instances where inference on the baseline functions (and derivations of these functions) is of interest, the Cox model has some undesirable properties.

Chief among them: the baseline hazard function is a "high dimensional, very overfitted estimate" (Royston and Parmar, 2002).

It is easy to see why the estimate is "overfitted." The baseline hazard is closely adapted to the observed data: $\hat{h}_0(t)$ is (often) a noisy function.

Hjort (1992) takes the argument one step further: "the success of Cox regression has perhaps had the unintended side-effect that practitioners too seldomly invest efforts in studying the baseline hazard."

Yet in some settings, the form of the underlying time dependency *may* be of central interest.

Like where? Perhaps medical studies of illness progression; epidemiology; contagion effects; networks.

Are there any "middle ground" models?

Flexible Parametric Models

Royston and Parmar (2002) propose a generalized hazard model that relies on splines to characterize the duration dependency.

Their approach is conceptually similar to that advocated by Beck and Jackman (1998) and Beck, Katz, and Tucker (1998), in the political methodology literature (but there are differences).

Mainly, the Royston and Parmar approach differs from the standard parametric approach in that the distribution of the integrated hazard (from which the hazard rate can be retrieved) is not assumed to follow a specific distribution function.

Instead, the baseline function is modeled with the use of splines.

The basic models are:

$$\log H(t;\mathbf{x}) = \log H_0(t) + \beta'\mathbf{x} = s(x) + \beta'\mathbf{x} \tag{1}$$

and

$$\log O(t;\mathbf{x}) = \log O_0(t) + \beta'\mathbf{x} = s(x) + \beta'\mathbf{x}, \tag{2}$$

where $\log H_0(t)$ is the log integrated hazard and $\log O_0(t)$ is the log of the cumulative proportional odds.

The baseline functions in equations (1) and (2) (denoted as $s(x)$) are approximated by a spline function.

In passing, note that if equations (1) and (2) are modeled as a linear function (that is, splines are not used to specify $s(x)$), then the Weibull model is obtained for the proportional hazards model, the log-logistic model is obtained for the proportional odds model, and the log-normal model is obtained from inverse normal model. Hence, the Royston-Parmar approach can be viewed as a generalization of some of the standard parametric models.

Note that the above holds because of the link function used by Royston-Parmar. They make use of the Aranda-Ordaz link function:

$$g(x;\theta) = \log \frac{x^{-\theta} - 1}{\theta}, \tag{3}$$

A-O showed (1981) that when $\theta = 1$, the proportional odds model is obtained, and when $\theta \rightarrow 0$, the proportional hazards model is obtained.

Pretty neat.

For analysts explicitly interested time dependency, the Royston-Parmar approach provides an attractive middle ground.

The shape of the time dependency is not determined by the distribution function (unless equations (1) and (2) are estimated linearly).

TVCs and Related Issues

Time-varying covariates are covariates having values that change within the observation plan (e.g. economic variables, campaign expenditures, etc.)

All the models discussed can "handle" TVCs, but there are some issues to be aware of.

Exogeneity Assumptions

The exogeneity condition:

$$\Pr(X(t, t+\Delta t) \mid T \geq t+\Delta t, X(t)) = \Pr(X(t, t+\Delta t) \mid X(t)), \quad (4)$$

In words: a covariate is exogenous if and only if the values it can take are independent of the duration that an observation survives.

In other words: the additional survival time shown in the left-hand term does not affect the probability that a covariate will assume some particular value.

Easy, right?

This isn't an event history issue per se.

Jump Process

If covariates only change values at discrete times, then the interval between these measurement times results in a constant value for the covariate.

This gives rise to a jump process: a TVC remains constant from t_j until t_{j+1}, at which time its value can change, or "jump" to another value.

We evaluate $S(t)$ in the duration intervals in which the TVCs remain constant.

The survivor function for a jump process looks like:

$$S(t_k) = \prod_{j=1}^{k} \Pr(T > t_j \mid T \geq t_{j-1}), \tag{5}$$

where the right-side term is

$$\Pr(T > t_j \mid T \geq t_{j-1}) = \exp\Big(- \int_{t_{j-1}}^{t_j} h(u \mid \mathbf{x}_j) du \Big) \tag{6}$$

and where $h(u \mid \mathbf{x}_j)$ is the hazard rate conditional on the covariate having the value $\mathbf{x}_j$, which is constant in the interval t_{j-1} to t_j.

Note that the integral in (6) is equivalent to

$$\int_{t_{j-1}}^{t_j} h(u \mid \mathbf{x}_j) du = H(t_k),$$

which is the integrated or cumulative hazard rate in the integrable interval, t_{j-1} to t_j.

Implications? From equations (5) and (6), it is easy see that when the survival times are a function of TVCs, the survivor function is the product of successive survivor functions defined on each interval in which the TVCs remain constant.

All this holds for $h(t)$ as well …given its connection to $S(t)$.

All this affects how we interpret our covariates too.

Inclusion of TVCs leads naturally to consideration of a counting process set-up.

TVCs and the Cox Model

Inclusion of TVCs into the Cox model entails a extension of the partial likelihood function.

Let $\mathbf{x}$ denote the covariates, some of which may be time-varying, which are indexed by (t). The likelihood function for the Cox model with TVCs is given by

$$\mathcal{L}_p = \prod_{i=1}^{K} \Big[\frac{e^{\beta' \mathbf{x_i}(t_i)}}{\sum_{j \in R(t_i)} e^{\beta' \mathbf{x_j}(t_i)}}\Big]^{\delta_i}, \tag{7}$$

and the corresponding log-likelihood function is given by

$$\log L_p = \sum_{i=1}^{K} \delta_i \Big[\beta' \mathbf{x_i}(t_i) - \log \sum_{j \in R(t_i)} e^{\beta' \mathbf{x_j}(t_i)}\Big]. \tag{8}$$

The Cox model well suited to include TVCs because the partial likelihood function is determined by the ordered failure times, but not by the actual duration times.

Calculations of the hazard ratio are only made at failure timesand so the Cox regression coefficients for the TVCs can be interpreted as the change in the log-hazard ratio for observations having a unit change in the value of the covariate at time t compared to the value of the covariate for the remaining observations in the risk set at time t.

The estimated covariate tells us by how much the risk of an event "jumps" due to a change in the value of the TVC. We now turn to an illustration.

TVCs and Parametric Models

Inclusion of TVCs into parametric models presents no special problems. The likelihood function with TVCs is evaluated in terms of k successive intervals.

The likelihood function for the Weibull with exogenous TVCs is given by

$$h[t \mid \mathbf{x}(t^{-})] = e^{-\beta'\mathbf{x}(t^{-})}p \ \ (e^{-\beta'\mathbf{x}(t^{-})} \ t)^{p-1}.$$

Here t^{-} denotes that the change in the covariate is observed prior to t. Setting $\lambda = \exp(-\beta'\mathbf{x}(t^{-}))$, the survivor function for the Weibull model with exogenous TVCs is

$$S[t \mid \mathbf{x}(\mathbf{t}^{-})] = e^{-(\lambda t)^{p}}$$

and the density function for the Weibull is

$$f[t \mid \mathbf{x}(\mathbf{t}^{-})] = \lambda p(\lambda t)^{p-1}e^{-(\lambda t)^{p}}.$$

From $f(t)$ and $S(t)$, we can construct a likelihood function for the t duration times:

$$\mathcal{L} = \prod_{j=}^{k} \{\lambda p(\lambda t)^{p-1}e^{-(\lambda t)^{p}}\}^{\delta_i}\{e^{-(\lambda t)^{p}}\}^{1-\delta_i}. \tag{9}$$

Aside from the modification that the likelihood is defined for the k intervals on which the covariates are measured, equation (9) identical to previous likelihood results.

Main difference: For TVCs, the coefficients provide information on how the hazard rate changes for a unit change in the TVC.

As with the Cox model, this incremental change results in a jump process.

Table 1: Weibull Model of Challenger Deterrence

Variable	Estimate (s.e.)
South	$-.53$ (.42)
Party	.23 (.32)
Prior Vote	-6.44 (1.66)
War Chest	-2.58 (1.38)
Shape Parameter	2.43 (.35)
N	1376
Log-Likelihood	-118.84

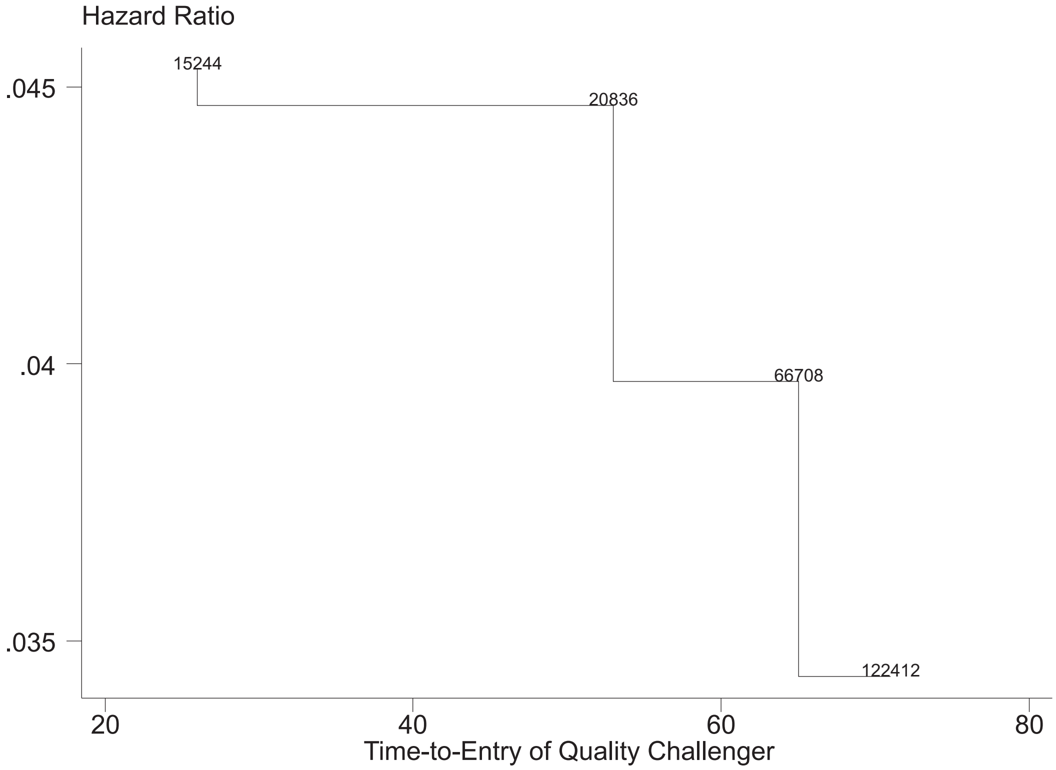

Figure 1: *This figure plots the estimated hazard ratios of quality challenger entry for a selected incumbent. The change in the hazard ratios are a function of changes in the value of the TVC.*

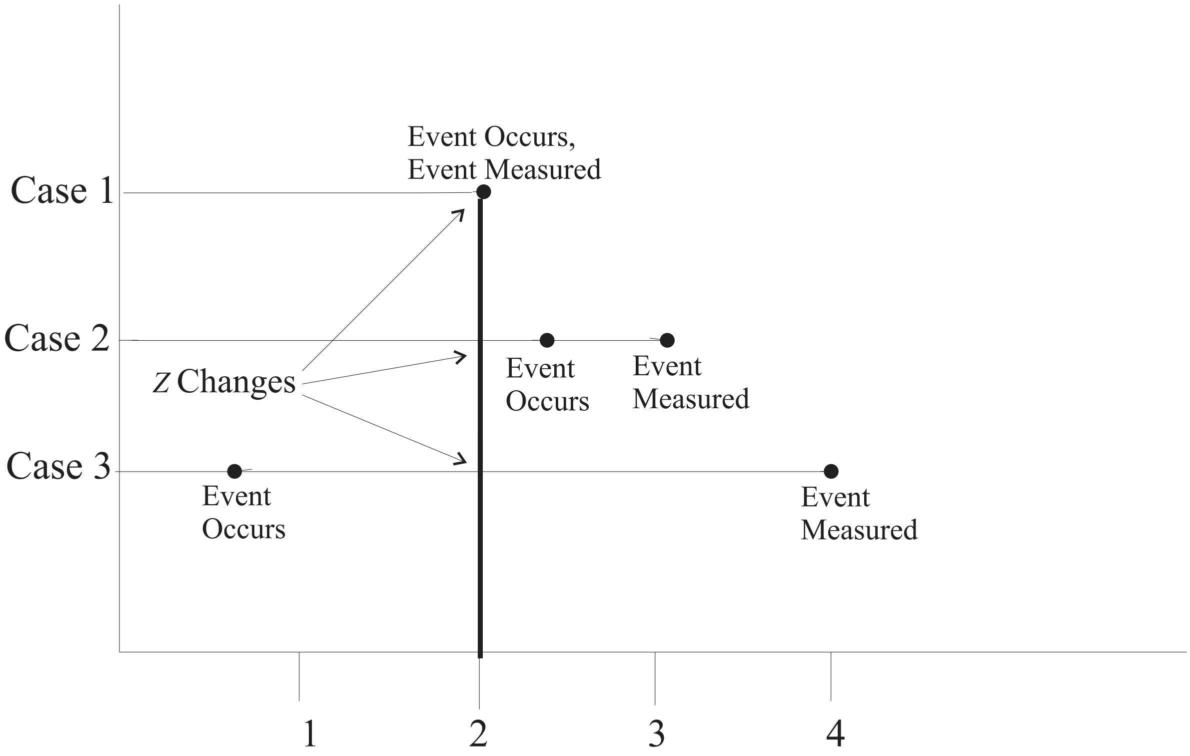

Figure 2: *This figure represents some problems that can occur when event times are imprecisely measured and TVCs are included in the analysis. In Case 1, the event occurs and the TVC changes at the same point; in Case 2, the event, although imprecisely measured, occurs after the change to the TVC; in Case 3, the event actually occurs before the TVC changes value.*

Temporal Dependence

Standard assumption:

$$I^{-1} = \frac{-\partial^2 \log(L)(\beta)}{\partial\beta\partial\beta'}.$$

Usually won't hold with TVCs (why?).

"Sandwich" estimators:

$$V = I^{-1}G'GI^{-1},$$

For survival analysis, Lin and Wei derived a robust variance estimator.

Ignore at your own peril!