Replication file for: "Beyond the Hazard Ratio: Generating Expected Durations from the Cox Proportional Hazards Model"

Jonathan Kropko
University of Virginia
jkropko@virginia.edu

Jeffrey J. Harden
University of Notre Dame
jeff.harden@nd.edu

Journal article meta analysis desciption
Last update: 9/5/16

Overview

This metaanalysis examines use of the Cox proportional hazards model and methods for interpreting results in five political science journals from 1990-2015: *APSR*, *AJPS*, *BJPS*, *JOP*, and *IO*. We first searched for articles using Google Scholar. Then we coded the articles based on (1) the language used to frame hypotheses and (2) the methods used to interpret Cox model results. We describe the details of these procedures below.

Our central objectives were to assess (1) the type of language researchers typically use to frame their hypotheses when employing the Cox model and (2) the methods they typically use to interpret Cox model results. On the first objective we considered two possible framing styles: a "risk" frame and a "duration" frame. A risk frame discusses hypotheses with respect to the risk of event occurrence. For example: "as X increases, the risk of event Y occurring also increases." In such a case the researcher is not primarily concerned with duration, but rather focuses on how the covariates make the event more or less likely to occur. In contrast, a duration frame discusses the hypothesis in terms of event time, as in "as X increases, the number of days until event Y occurs decreases." In this case the length of time that an event takes is of central importance.

Searching for articles

We searched scholar.google.com for "cox proportional hazards" OR "cox model" OR "cox regression" in each journal listed above, one journal at a time. We set the date range to 1990-2015. We downloaded all of the articles returned by these searches, checked each one to make sure it included analysis with the Cox model, then saved it for coding in the next step. We included any paper that reported the estimation of a Cox model in the main text. This produced 80 total papers, ranging in publication date from 1996-2015 (available in the file "meta_analysis.csv").

Coding hypothesis text

First, we identified the number of hypotheses in each article and copied the text of those hypotheses. If multiple analyses were presented, we only included hypotheses pertaining to the Cox model(s) reported in the main text. If no hypotheses were presented with the Cox model (i.e., in a descriptive analysis), we used the authors' descriptions of the model specification (i.e., variables used and purpose of the estimation).

Next, we placed all of the hypotheses' text into one single string, omitted common English stop words, then counted word frequencies of the remaining words. This produced a list of 1,571 unique words, from which we identified words as either predominantly part of a risk frame and words predominantly used in a duration frame. Our general rule was to code any word that related to probability, likelihood, or chance in the risk frame category and any word relating to time in the duration frame category.[1] In all, we coded eight unique words as risk frame words and 47 unique words as duration frame words. We put these into a separate file ("frame_words.csv") for further analysis.

Hypothesis frame results

The fact that we coded many more words as duration frame words gives some preliminary evidence that authors tend to frame their hypotheses with respect to the time until event occurrence more often than the risk of event occurrence. However, this may be skewed by the possibility that authors simply have more choices when it comes to duration frame words. Looking at word counts of all eight risk frame words and the top eight duration frame words reveals a larger count of risk frame words: 119 instances of risk frame words and 106 duration frame words. However, the full count of all the duration frame words is 215—substantially larger than the total count of risk frame words.

We also coded each article as either predominantly using a risk frame, duration frame, or equal use of both frames. We accomplished this in two ways: a count of unique words and a count of total words. First, we counted how many unique words from each frame appeared in the text of the hypotheses. This approach did not give additional weight to the same word appearing more than once. We then coded an article's frame as the type with the most instances of unique words from its list. Second, for each article we counted the total number of instances of words in its hypotheses from each frame (i.e., allowing for repeats of the same word). In both cases if an equal number of risk and duration frame words appeared, we coded the article as equal.

Using the unique word count, we coded 33 articles as using a predominantly duration frame, 16 with a risk frame, and 14 with equal use of both frames. With the total word count these numbers are 33, 20, and 10, respectively. Almost half (29) of the articles use words from both frames, 20 contain no risk frame words, and 14 contain no duration frame words.

Coding methods used

Our second objective was to code which method(s) each article used to interpret the results of the estimated Cox model. This was accomplished by reading the results section and identifying each unique method used. We created a total of four categories based on what we found in the text, which we define below. Note that all of the articles discussed the sign and significance of the Cox model coefficient estimates. The categories reflect any interpretation beyond sign and significance. The articles employed an average of 1.3 of these interpretation methods. 45 articles used one method, 17 articles used two methods, and 1 article employed three different methods.

[1] This was a subjective assessment, and we encourage interested readers to replicate our analysis. The full list of words is contained in the file "word_frq.csv".

- *Hazard ratios* (24 articles). This category includes any article that reported the exponentiation of one or more Cox model coefficients, as well as a discussion about the resulting multiplicative effect of a one-unit change in the covariate of interest.
- *Changes to the hazard rate* (33 articles). This category includes any article that reported a marginal change in the hazard rate (usually expressed as a percentage) corresponding to a substantively interesting change in the values of a covariate.
- *Empirical estimates of the hazard and/or survival functions* (17 articles). This category includes any article that graphically displayed an estimate of the baseline hazard from the model and/or computed the survivor function. Typically this was done for different covariate values to show the effect of changes to the covariate.
- *Only sign and significance of the coefficient estimates* (8 articles). This category includes any article that did not report any interpretation of the Cox model other than the sign and significance of the relevant coefficient estimates.

The most important finding from this analysis is the fact that all of the articles that go beyond sign and significance in their interpretation of the Cox model focus on the hazard rate, whether through hazard ratios, changes to the hazard rate, or estimation and graphing of the baseline hazard rate. The closest that any article comes to interpreting results with respect to duration is the few (3 articles) that construct estimates of the survivor function. However, even in those cases the focus is on the relative order of event occurrence, not estimates of the expected duration. To further emphasize this point, one article in our data did report expected durations after estimating a Cox model, but those estimates came from re-estimating the model using the Weibull parameterization (Senese and Quackenbush 2003, 714).[2]

Conclusions

This analysis yields two important insights. First, we find that political scientists employing the Cox model over the last 25 years tend to discuss their theoretical expectations in the language of time until event occurrence. Language related to the risk of event occurrence also appears, but it is less common than duration-based framing. Approximately 70% of the articles in our sample contain more duration words or an equal amount of duration and risk words. It is clear that researchers' substantive interests usually center on the duration of some political phenomenon, not just its likelihood of occurring.

This first finding contrasts sharply with the second finding, which is that researchers nearly exclusively rely on interpretation of the hazard rate after estimating the Cox model. We found no instances where researchers generated expected durations from their Cox model estimates. Thus, researchers who employ the Cox model are typically forced to switch the manner in which they discuss their research when moving from hypotheses to results. This provides motivation for our research, which provides a method for generating expected durations from the Cox model. The Cox ED approach allows researchers to maintain consistency between the language they use to describe their theoretical framework and the language they use to communicate their empirical findings.

[2] Senese, Paul D. and Stephen L. Quackenbush. 2003. "Sowing the Seeds of Conflict: The Effect of Dispute Settlements on Durations of Peace." *Journal of Politics* 65 (3): 696-717.