Properly calculating estat phtest in the presence of stratified hazards

Shawna K. Metzger
Michigan State University
East Lansing, MI
metzge69@msu.edu

Benjamin T. Jones
University of Mississippi
Oxford, MS
btjones1@olemiss.edu

Abstract. The Schoenfeld residual-based test for proportional-hazards violations in the Cox duration model is predicated on a key assumption pertaining to homoskedasticity in the residuals. This assumption is likely to be violated in the presence of stratified hazards, which is noted in `estat phtest`'s help file. We provide a wrapper command, `stratph`, that implements a straightforward modification to the residual-based diagnostic that corrects for the potential assumption violation. We use the stratified hazards example from the `stcox` help file to demonstrate the `stratph` command's workings.

Keywords: st0661, stratph, estat phtest, proportional hazards, Cox model, stratified hazards, heteroskedasticity

1   Introduction

Proportional hazards (PH) is a key assumption underlying the Cox semiparametric duration model. The assumption originates from the way in which these models' functional form is defined and states that each covariate's effect is proportional over time—if changing a covariate's value increases the risk of an event by 10% at $t = 1$, then that same change should increase the risk of an event by 10% at $t = 10$, $t = 50$, and so on. Even one covariate violating the PH assumption can yield misleading inferences regarding any of the model's covariates (Box-Steffensmeier and Zorn 2001, 974–975). Testing the assumption is therefore crucial and is typically done with residual-based diagnostics involving Schoenfeld residuals (`estat phtest`).

However, we know from the biostatistics literature that the Schoenfeld diagnostic test is predicated on a key assumption: variance homogeneity across risk sets (Therneau and Grambsch 2000, 141). Therneau and Grambsch (2000, 133–134) make this assumption for simplifying purposes (because the variance remains fairly constant across time) and also to help stabilize the Schoenfeld variance estimate across all time points because the estimate can become erratic in later periods as fewer and fewer subjects remain at risk. We also know from the biostatistics literature that, if the homogeneous variance assumption is not met, we may draw erroneous conclusions about whether a covariate's effect is proportional from standard Schoenfeld residual-based tests.

In this article, we discuss a command, `stratph`, that adjusts the Schoenfeld-based PH test in Stata in the presence of stratification to yield appropriate results because the variance homogeneity assumption is likely to be violated when multiple strata are

present.[1] This command acts as a wrapper to the `mstatecox` package's `mstphtest` command, which is written to account for the presence of strata and other potential complexities involving covariate effects. We begin by discussing the PH test adjustment and then discuss `stratph`'s general syntax. We conclude by providing a stratified hazard example from the `stcox` help file, in which the conclusions we draw about PH violations differ once we stratify the PH test.

2 Adjusting the PH test

The basic intuition to correct the PH test for a model with strata is straightforward. If researchers have a Cox model with strata (call this the overall model) but suspect that the Schoenfeld variance differs across these strata, the solution is to run the PH test on each stratum (Therneau and Grambsch 2000, 142). Doing so involves rerunning the Cox model separately on each stratum, constraining the stratum-specific Cox models' coefficient estimates to be the same as the overall Cox model's coefficient estimates, and then running the PH test using each stratum-specific model.[2] We refer to this as "stratifying the PH test". Constraining the stratum-specific estimates is the crucial part of the correction; otherwise, Stata will refit a new model using only that stratum's observations, which amounts to estimating stratum-specific covariate effects, almost certainly yielding different coefficient estimates from the overall model.

3 Syntax

`stratph` [, *estat_phtest_opts*]

`stratph` requires that a Cox model be in memory. The Cox model must also have a variable included in `strata()`; if it does not, `stratph` will print an informational message and then run `estat phtest` as if the latter had been called directly. `stratph` also requires that `strata()`'s values are discrete in nature. If they are not, you will receive an error message and be asked to recode the variable.

Any of the permissible options for `estat phtest` are permissible options for `stratph`. The sole exceptions are the plot-based options. These options produce an error message in `mstphtest`, which is the command `stratph` ultimately calls. Finally, the `detail` option is hardcoded into `mstphtest`; it does not need to be specified again but harms nothing if it is.

---

1. In R, the `survival` package's `cox.zph` function automatically makes this adjustment for strata starting with `survival` 3.0–10. The same function also calculates the actual score test proposed by Grambsch and Therneau (1994), instead of the quick approximation of it that pre-`survival` 3.0–10 and Stata do.
2. In Stata, this is done by setting the likelihood optimizer's starting values for the stratum-specific models equal to the overall Cox model's coefficient estimates and then preventing the optimizer from iterating by using the `iter(0)` option, consistent with the approach presented by Therneau and Grambsch (2000, 141).

stratph requires the mstatecox package (Metzger and Jones 2018), which pertains to multistate Cox models. Basic Cox models with stratified hazards are a special case of multistate Cox models, allowing us to exploit mstatecox's functionality. Stratified hazards are ubiquitous for multistate models, as are more complex covariate patterns that may vary by strata, both of which have implications for PH testing. For these reasons, mstatecox includes the mstphtest command, which tests the PH assumption by using the Schoenfeld-based test for every unique value of the Cox model's strata() variable. It also automatically constrains the coefficients to be the same as the original Cox model if needed. When stratph is finished executing, it purges all of mstatecox's intermediate calculations and variables. Like mstphtest, stratph returns the individual coefficient table and global test matrix for each stratum in r-class memory.

4   Example

We use the stratified hazard example from stcox's help file, involving stan3.dta, a dataset of patients admitted to the Stanford Heart Transplantation Program (Crowley and Hu 1977).[3] This example is particularly attractive, aside from it using one of Stata's stock datasets, because the conclusions we draw about PH-violating covariates are different once we properly stratify the PH test.[4]

We begin by executing the code from the stcox help file, which posits that the underlying rate at which heart transplant candidates die differs depending on when they were accepted into the transplant program: 1967–1969, 1970–1972, or 1973–1974. If such differences exist, it suggests we should stratify the baseline hazard by these groups:

```
. // Load data
. webuse stan3
(Heart transplant data)
. // Modify data to reflect changes in treatment 1970 and 1973
. generate pgroup = year
. recode pgroup min/69=1 70/72=2 73/max=3
(172 changes made to pgroup)
. // Fit Cox model
. stcox age posttran surg, strata(pgroup)
  (output omitted)
```

---

3. We deviate slightly from the Stata example in this application by omitting year as a covariate in the model because its effect is already reflected in the separate strata.
4. We assume that no model misspecification exists (Keele 2010) for the purposes of this demonstration.

If we run `estat phtest` at this juncture using the default time-scale transform, t, we obtain

```
. estat phtest, detail
Test of proportional-hazards assumption
Time function: Analysis time
```

	rho	chi2	df	Prob>chi2
age	0.10608	1.11	1	0.2916
posttran	−0.00280	0.00	1	0.9809
surgery	0.16004	1.90	1	0.1680
Global test		3.24	3	0.3565

None of the three coefficients' p-values are statistically significant at conventional levels. Neither is the global test. Based on this output alone, we would conclude that no evidence of a PH violation exists.

However, as we have discussed, `estat phtest` does not properly adjust for the presence of strata. Once we stratify the PH test, a very different picture emerges:

```
. stratph
*** NOTE: transition ID value = pgroup's value
> Transition 1
Test of proportional-hazards assumption
Time function: Analysis time
```

	rho	chi2	df	Prob>chi2
age	0.15985	1.13	1	0.2887
posttran	0.09217	0.23	1	0.6300
Global test		1.85	2	0.3967

```
> Transition 2
Test of proportional-hazards assumption
Time function: Analysis time
```

	rho	chi2	df	Prob>chi2
age	0.12325	0.46	1	0.4988
posttran	−0.06857	0.16	1	0.6925
surgery	0.37641	2.19	1	0.1392
Global test		2.87	3	0.4116

```
> Transition 3
Test of proportional-hazards assumption
Time function: Analysis time
```

	rho	chi2	df	Prob>chi2
age	−0.47396	3.14	1	0.0765
posttran	−0.37257	1.31	1	0.2527
surgery	0.12466	0.93	1	0.3351
Global test		5.22	3	0.1563

```
NOTE: the following variables are collinear within the listed stratum
pgroup = 1
  surgery
```

`stratph`'s output begins by disambiguating what `Transition #` means: it is the `strata()` variable's value. The command then proceeds to calculate the PH test for each unique value of `strata()` and output the result.[5]

From the stratum-specific tests, we see no evidence of PH violations in the `pgroup = 1` stratum. We also see no evidence of violations in `pgroup = 2`. The story changes when we examine `pgroup = 3`. In this stratum, age is weakly significant ($p = 0.08$), which offers suggestive evidence of a PH violation. Although the global test fails to achieve statistical significance in this case, we should not necessarily dismiss the individual covariate tests indicating violations of the PH assumption (Box-Steffensmeier, Reiter, and Zorn 2003, 45). As a result, there is evidence it may be appropriate to correct for a PH violation for age in the third stratum. Doing so would likely involve stratum-specific covariates of some sort, which the multistate model literature discusses at length (Metzger and Jones 2016; Putter, Fiocco, and Geskus 2007).

5   Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 21-4
. net install st0661    (to install program files, if available)
. net get st0661        (to install ancillary files, if available)
```

---

5. In this specific case, `surgery` is dropped from the strata-specific PH test for `pgroup = 1` because `surgery`'s value is constant within this stratum. Stratifying the PH test involves refitting the model separately on each stratum, leading to the collinearity issue. In such cases, `stratph` will print a note to inform the user which variables were omitted because of collinearity.

6   References

Box-Steffensmeier, J. M., D. Reiter, and C. Zorn. 2003. Nonproportional hazards and event history analysis in international relations. *Journal of Conflict Resolution* 47: 33–53. https://doi.org/10.1177/0022002702239510.

Box-Steffensmeier, J. M., and C. J. W. Zorn. 2001. Duration models and proportional hazards in political science. *American Journal of Political Science* 45: 972–988. https://doi.org/10.2307/2669335.

Crowley, J., and M. Hu. 1977. Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* 72: 27–36. https://doi.org/10.2307/2286902.

Grambsch, P. M., and T. M. Therneau. 1994. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 81: 515–526. https://doi.org/10.1093/biomet/81.3.515.

Keele, L. 2010. Proportionally difficult: Testing for nonproportional hazards in Cox models. *Political Analysis* 18: 189–205. https://doi.org/10.1093/pan/mpp044.

Metzger, S. K., and B. T. Jones. 2016. Surviving phases: Introducing multistate survival models. *Political Analysis* 24: 457–477. https://doi.org/10.1093/pan/mpw025.

———. 2018. mstatecox: A package for simulating transition probabilities from semi-parametric multistate survival models. *Stata Journal* 18: 533–563. https://doi.org/10.1177/1536867X1801800304.

Putter, H., M. Fiocco, and R. B. Geskus. 2007. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine* 26: 2389–2430. https://doi.org/10.1002/sim.2712.

Therneau, T. M., and P. M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

About the authors

Shawna K. Metzger is an assistant professor (fixed term) in James Madison College at Michigan State University. Her substantive research examines the interplay between disputed issues and interstate conflict, and her methodological research focuses on event history analysis. She has published in *Conflict Management and Peace Science*, *International Interactions*, *International Studies Quarterly*, the *Journal of Conflict Resolution*, the *Journal of Legislative Studies*, and *Political Analysis*.

Benjamin T. Jones is an associate professor of political science at the University of Mississippi. His research explores the links between civil conflict combatants and foreign powers and the role of foreign powers in civil conflicts, as well as empirically modeling the role of time in political phenomena more broadly. His work has appeared in journals such as *Political Analysis*, the *British Journal of Political Science*, and *International Studies Quarterly*.