

Review of Event History Analysis with Stata by Blossfeld, Golsch, and Rohwer

Frank Kalter
Institute of Sociology
University of Leipzig
Leipzig, Germany
fkalter@sozio.uni-leipzig.de

Abstract. This article reviews *Event History Analysis with Stata* by Blossfeld, Golsch, and Rohwer.

Keywords: gn0039, causal modeling, event history analysis, TDA, teaching

1 Introduction

This is a book for which I have been waiting many years! I was introduced to event history analysis as a young research assistant in the early '90s, and I was lucky enough to attend a workshop by Hans-Peter Blossfeld and Götz Rohwer once in Cologne, Germany. Here we received a week-long introduction to event history analysis, and I have seldom again profited from such a short time investment. First, we got used to the statistical package Transition Data Analysis (TDA), written by Götz Rohwer. TDA is a freeware program that, at the time, by far outperformed alternative (expensive) software packages for event history analysis in terms of features and speed. Second, we learned to apply event history analysis. The workshop was designed in an extremely motivating, hands-on style, using data and examples from actual research. We moved step by step, always linking the formal content to precise TDA command files. You could improve your understanding by simply rerunning the commands, and later on, you could easily modify the original files to satisfy your own research needs. This step-by-step process is how methods should be taught.

Fortunately, *Techniques of Event History Modeling* appeared in 1995, with a second edition in 2002 (Blossfeld and Rohwer 1995, 2002), presenting an update and extension of the workshop in book form and again delivering all the necessary material. I still use *Techniques of Event History Modeling* to refresh or extend my knowledge, and I know many self-learners who have learned successfully from it. Having found that this textbook is perfect for teaching students, I have used it for this purpose very effectively.

In spite of its unquestionable advantages, TDA has always produced some discomfort. As just one example, the production of graphics for survivor and hazard curves as well as their ability to be imported into text files has never been easy or convincing. Most importantly, however, TDA alone was never enough. The major hurdle of event history analysis does not lie in running the models but in constructing the necessary data files.

To do that, everybody—including me—referred to her or his usual software package. The export and import involved with this process did cost a lot of time when modifying analyses in daily research, but teaching fluency in both software packages was an even more severe difficulty.

Then, fortunately, Stata began to implement more and more event history features and soon became a serious alternative for event history analysis. People began to *translate* the TDA command files into Stata do-files, thus saving the advantages of the book while avoiding the transaction costs of switching between statistical packages. With the increased use of Stata and all the advantages Stata provides, this translation was an obvious thing to do. And this translation is exactly what Blossfeld, Golsch, and Rohwer have addressed in the book reviewed here.

2 Contents

Chapter 1 starts with demonstrating the usefulness of event history analysis both in general and specifically for the social sciences. The chapter explains the relative advantages of event history data as compared to cross-sectional or panel data, justifying the reasoning with references to recent research. Blossfeld, Golsch, and Rohwer put specific emphasis on the need for investigating causal relationships in the social sciences and on the role of event history analysis within this attempt. Building on these discussions, the authors introduce basic terminologies and statistical concepts of event history analysis.

Chapter 2 continues with an explanation of the basic structure of event history data files and how to use them with Stata. The reader becomes familiar with the training data file (a subsample of 600 episodes from the German Life History Study), which provides the basis for all following analyses. This chapter provides some rudimentary information on how to handle Stata to allow even Stata beginners to rerun all the do-files used as examples throughout the book. Chapter 3 discusses the two basic nonparametric descriptive methods, i.e., life table and product-limit (Kaplan–Meier) estimation, along with methods to compare survivor functions. The authors helpfully demonstrate how the output resulting from Stata do-files could also be calculated by hand according to the formulas—this way the reader acquires an especially deep understanding of both the output and the underlying algorithms.

The next five chapters focus on parametric models. Chapter 4 starts with the simplest one, the exponential transition rate model, which assumes that the baseline transition rate is constant over time. Step by step, the reader learns about models without covariates, models with time-constant covariates, models with multiple events, and models with repeated events. Building on this information, chapter 5 introduces the piecewise-constant exponential model. This model allows the baseline transition rate to vary between different time intervals while being constant within. This often provides a very flexible fit to the data, and piecewise-constant exponential models turn out to be very useful tools in practical research. Chapter 6 then adds another important step by introducing time-dependent covariates and the technique of episode splitting. The authors' general ideas and precise Stata handling are embedded into a broader theoretical

discussion of the causal approach to interdependent processes. The chapter examines four examples from actual research in order to illustrate the utility of time-dependent covariates and the flexibility of the episode-splitting method. Chapter 7 further enriches the toolbox by adding four parametric models that can be fitted by Stata: Gompertz models, Weibull models, loglogistic models, and lognormal models. In each of these models, the general shape of the baseline transition rate over time is determined by two basic parameters. The sections give a short discussion of the underlying formulas, followed by estimations of models without covariates, covariates linked to one of the model's parameters, and covariates linked to both of the parameters. Chapter 8 rounds off the treatment of parametric models, suggesting methods to check parametric assumptions either by direct graphical methods or by checking the pseudoresiduals.

Chapter 9 treats semiparametric models, specifically the Cox model. It gives an idea of the partial-likelihood estimation, shows how to introduce time-dependent covariates, discusses the proportionality assumption together with consequences of its violation, and shows how to get graphical insights into the underlying baseline transition rates. The final chapter is devoted to problems of model specification. The discussion begins with a look at the consequences of unobserved heterogeneity on the time-dependent shape of the transition rate by showing examples and segues to models with a gamma mixture as possible strategies to this problem.

3 Assessment

As *Event History Analysis with Stata* (Blossfeld, Golsch, and Rohwer 2007) is basically a Stata “translation” of the TDA-based *Techniques of Event History Modeling* (Blossfeld and Rohwer 1995, 2002), it automatically inherits all the strengths of the latter. Above all, it is the book's general didactical concept that makes it a convincing introduction and distinguishes it from rival books. The most basic characteristics in this respect have already been implicitly mentioned in this review: in this book you learn by doing and you learn step-by-step, from simple things to more complicated ones and in a well-designed structure both within chapters and from chapter to chapter. Each necessary step is documented in a do-file, which you can download from <http://web.uni-bamberg.de/sowi/soziologie-i/eha/stata/>, together with the basic dataset. Running and modifying the do-files gives you immediate and easy access to your own research problems, especially because the discussion of outputs is closely linked to the underlying statistical concepts. Moreover, the authors wrote the book in a very clear language, kept formulas and statistical theory within necessary limits, and cared about general methodological issues and the link to sociological theory. You learn not only to apply event history analysis but also about its general and practical drawbacks and opportunities in the social sciences.

Event History Analysis with Stata takes over not only many strengths but also some shortcomings from the former TDA-based version. For example, what has often been criticized is Blossfeld and Rohwer's (1995, 2002) tendency to concentrate only on continuous-time models while not treating discrete-time models and their tendency to

put a bit too much emphasis on parametric models—these tendencies have not changed in *Event History Analysis with Stata*. One may also miss some recent, more elaborate developments in event history modeling.

In addition to these inherited limitations, the update itself seems to contain one basic shortcoming. Certainly, marrying *Techniques of Event History Modeling* to Stata is worth doing, as argued in the introduction. But what we find here is only the mere intersection of the former contents with the features and possibilities of Stata. This could have been much more! One may have wished, for example, that the authors would have chanced to enrich Stata by delivering ado-files for procedures that are still unique for TDA, e.g., the parametric sickle model or the generalized loglogistic model. But the respective pages of the former book were simply cut. Conversely, specific features and possibilities of Stata are not picked up to enrich the content of the former book and to enhance some of its procedures; some are minor but nice Stata options (for example, built-in smoothers to get graphical representations of hazard rates in the product-limit estimator), while others are helpful Stata commands that lead to more fundamental changes in the logic of certain techniques (like using the elegant combination of `stsplit` and `stjoin` to generate time-dependent dummy variables instead of the clumsy solution suggested in section 6.4). As in spoken languages, software translating is often not preferable in a straight one-to-one way—sometimes it is not even possible.

It also seems that, now and then, the Stata version would have required a slight change in the organization of the book: while, for example, the piecewise-constant exponential model is specified by changing only a model selection parameter in TDA, it here uses the `stsplit` command in section 5.2. Episode splitting, however, is not introduced until chapter 6 of the book.

Given the strength of the book, many of these critiques are minor or could be regarded as a matter of taste in the end. Nevertheless, the authors missed a chance to improve the splendid course even further. And it surely is an unquestionable shortcoming that—apart from not using respective features itself—the book does not refer to additional Stata options or alternative Stata solutions.

4 Conclusion

As was the case with its TDA-based predecessors, *Event History Analysis with Stata* offers a wonderful introduction to survival analysis for practicing social scientists who want to learn to apply the techniques successfully in their fields of interest. The systematic and practical approach makes it an ideal textbook for students or a perfect course for self-learners. It is also well-suited as a standard reference book for active researchers. But in all three cases you will—occasionally, at least—have to use an additional reference in order to learn more about the event history-specific features of Stata or to fully employ all the possibilities it offers. For example, *An Introduction to Survival Analysis using Stata* (Cleves et al. 2008) will surely be a helpful complement. In contrast to what the authors state in the preface, however, *Techniques of Event History Analysis* (Blossfeld and Rohwer 1995, 2002) is not a necessary complement—it is

simply replaced. Basically, *Event History Analysis with Stata* is a third edition—now, finally, in Stata. Therefore, it would have been more appropriate to stay with the old title. If you already own one of the former editions, you do not need to buy the new one—just download the do-files; you will only miss some updates to the references to research examples. If you do not yet own one of the former editions, you should definitely buy the new Stata-based book simply because it is more convenient to have the information all together. Certainly, you will recommend this book to your students or other event-history beginners and advanced learners as the most important reference.

5 References

- Blossfeld, H., K. Golsch, and G. Rohwer. 2007. *Event History Analysis with Stata*. Mahwah, NJ: Lawrence Erlbaum.
- Blossfeld, H., and G. Rohwer. 1995. *Techniques of Event History Modeling: New Approaches to Causal Analysis*. Mahwah, NJ: Lawrence Erlbaum.
- . 2002. *Techniques of Event History Modeling: New Approaches to Causal Analysis*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum.
- Cleves, M., W. Gould, R. Gutierrez, and Y. Marchenko. 2008. *An Introduction to Survival Analysis Using Stata*. 2nd ed. College Station, TX: Stata Press, in press.

About the author

Frank Kalter is professor of sociology at the University of Leipzig, Germany. His major teaching and research interests include methods and statistics, rational choice theory, and the sociology of migration and integration. His work has appeared in journals like *European Sociological Review*, *Journal of Mathematical Sociology*, *Rationality and Society*, and *Research in Social Stratification and Mobility*. He is currently conducting several comparative research projects on the causes of migration and on mechanisms of immigrants' structural assimilation, collecting primary data and applying longitudinal data analysis in each.