

Introduction to the Virtual Issue: Recent Advances in Discrete Choice Methods in Political Science

Garrett Glasgow

Department of Political Science
University of California, Santa Barbara
Santa Barbara, CA 93106-9420
email: glasgow@polsci.ucsb.edu

A “discrete choice” situation occurs when a political agent (i.e., voter, political leader, country) faces a choice or a series of choices from among a finite set of discrete options. These types of situations are frequently studied by political scientists, but it is only in the last decade that our understanding and use of discrete choice models has really started to evolve. While even into the mid-1990s it was not unusual to see linear regression applied to binary or ordinal dependent variables (e.g., Gerber 1996), now discrete choice models are part of the standard toolkit for applied researchers, and methodological training in the use of these models is required in many graduate programs.

The leading edge of the field has also advanced rapidly. In the last decade political methodologists have moved beyond the “welcoming” of discrete choice methods from economics and other fields (Beck 2000) to the development of new discrete choice methods. In this virtual issue I have selected six articles from *Political Analysis* that I feel best exemplify this trend. These articles are:

- Logistic Regression in Rare Events Data, by Gary King and Langche Zeng.
- Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999, by Andrew D. Martin and Kevin M. Quinn.
- Back to the Future: Modeling Time Dependence in Binary Data, by David B. Carter and Curtis S. Signorino.
- An Estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions, by Anne E. Sartori.
- Structure and Uncertainty in Discrete Choice Models, by Curtis S. Signorino.
- Postestimation Uncertainty in Limited Dependent Variable Models, by Michael C. Herron.

Much of the methodological innovation in the use of discrete choice models by political scientists has occurred in areas where our discrete choice data have different properties than those found in most other social science fields. The first three articles in this virtual issue can all be grouped under this theme.

The article by **Gary King** and **Langche Zeng** (2001, Vol. 9(2), pp. 137-163) develops the “rare events logistic regression,” designed for cases where a binary dependent variable has many more non-events than events. This is an especially common problem in international relations, where among hundreds of thousands of country-dyads, there will only be a few dozen cases of war. King and Zeng show that standard binary choice models in such situations will systematically underestimate the probability of the rare event, and demonstrate some simple adjustments to the logit model that corrects for this bias. This article is the second-most frequently cited article in *Political Analysis*, and it has had a powerful effect on quantitative international relations research.

The next article in this virtual issue also addresses a common discrete choice data problem in political science. **Andrew Martin** and **Kevin Quinn** (2002, Vol. 10(2), pp. 134-153) develop a Bayesian approach for dynamic ideal point estimation, and apply it to a study of the ideal points of US Supreme Court justices over time. This article provides a lucid explanation of Bayesian methods that is accessible to those from a frequentist background, and has the potential to be influential outside of the ideal point literature where it is already well-known. As Martin and Quinn point out, the kind of roll call or voting data used in ideal point studies can be viewed as binary time-series cross-sectional (BTSCS) data with no measured covariates. The method introduced in this article could easily be extended to include covariates, giving political scientists a powerful new tool to deal with BTSCS data in other contexts, such as government transitions or the occurrence of war.

The third article in this virtual issue, by **David Carter** and **Curtis Signorino** (2010, Vol. 18(3), pp. 271-292), extends the influential work of Beck, Katz, and Tucker (1998) on time dependence in BTSCS data. Beck, Katz, and

Tucker point out that the BTSCS data we generally use in political science can also be viewed as disaggregated event history data, and show that we can deal with time dependence through the use of dummy variables to indicate elapsed time since the last event, or (better) use splines to create a smooth function of time. While these recommendations are clear, Carter and Signorino note that many applied researchers have struggled to apply them in practice. Time dummies often lead to problems with complete or quasi-complete separation, and many applied researchers appear to be unfamiliar with splines (most have simply applied the application-specific knots in Beck, Katz, and Tucker to their own data). Carter and Signorino demonstrate that simply including a cubic polynomial of elapsed time since the last event in the regression outperforms time dummy variables and generally performs as well as a spline, and is much simpler to implement and interpret for the average applied researcher. While a large variety of flexible and relatively easy to use smoothers have been developed in the last decade that might be superior to a simple cubic polynomial of time (Beck 2010), I predict that the simplicity of the approach outlined by Carter and Signorino will influence applied political science research in this area for years to come.

While one branch of innovation in discrete choice models in political science has been driven by specific data concerns, another has developed along more general theoretical or empirical lines. The next three articles in this virtual issue fall under this grouping.

Over the last decade, many political methodologists have sought to forge a closer link between our theoretical models and empirical analysis, both through the summer Empirical Implications of Theoretical Models (EITM) workshops and other venues. The article from **Curtis Signorino** (2003, Vol. 11(4), pp. 316-344) is drawn from the special *Political Analysis* issue on EITM, and extends his pioneering work on the implications of uncertainty and strategic behavior for discrete choice models. The development of strategic choice models is one of the clearest successes of the EITM approach, and this article is an important milestone. When modeling strategic interaction, the choice probabilities of our models will be a function of the preferences of multiple agents, each with their own preferences, expectations, and information – failure to account for this can lead to biased parameter estimates and incorrect inferences. In this article Signorino develops a broad class of discrete choice models for dealing with strategic behavior, further broadening the appeal of this approach to applied researchers.

The article by **Anne Sartori** (2003, Vol. 11(2), pp. 111-138) is also driven in part by theoretical concerns, and addresses the often vexing problem of identification in binary-outcome selection models. It is well known that Heckman-type selection models do not perform well without an “exclusion restriction,” or a variable that appears in the selection equation that does not appear in the outcome equation. However, in many cases such an exclusion restriction is not justified theoretically, and in many cases, there is no meaningful additional variable to include in the selection equation, leaving us with the unpalatable choice of either deliberately misspecifying our models, or relying solely on the distributional assumptions of the model for identification. Sartori develops a new approach to estimating selection models for cases where the covariates that influence selection and outcome are identical, identifying the model by assuming that the unobserved factors that influence both selection and outcome are nearly perfectly correlated. This assumption is reasonable for many of the problems that political scientists care about, such as the decision to go to war (selection) and then escalate the conflict (outcome), and gives applied researchers another method by which to address selection problems.

The final article in this virtual issue is by **Michael Herron** (1999, Vol. 8(1), pp. 83-98), and appears in the inaugural issue of *Political Analysis* as a quarterly journal. Rather than a technique to address a specific political science data problem, this work improves our interpretation of discrete choice models in a wide variety of settings. In this article Herron points out that while political scientists have become increasingly aware that the substantive interpretation of discrete choice models requires additional steps such as calculating predicted probabilities, in many cases these predicted probabilities are not reported with any measure of uncertainty. Herron explains how standard errors on these predicted probabilities (and statistics based on predicted probabilities, such as percent correctly predicted) can easily be generated through simulation by taking multiple draws from the multivariate normal distribution defined by the model coefficients and covariance matrix, calculating a separate predicted probability for each draw, and reporting means and confidence intervals based on this set of predicted probabilities. This simulation approach to calculating the postestimation uncertainty of quantities such as predicted probabilities later became familiar to political scientists through Stata’s Clarify package (King, Tomz, and Wittenberg 2000), and it is now a standard method for the interpretation of discrete choice models in political science. This article is a clear example of political methodology leading rather than “welcoming” – techniques developed by political methodologists in this article and elsewhere are being

adopted in numerous other social science fields, such as economics, sociology, criminology, and public health.

What lies ahead for the development of discrete choice methods in political science? I expect that further innovation will occur in those substantive areas with discrete choice data that poses unique challenges. For instance, in parliamentary government formation situations one governing coalition of parties is selected from a set of hundreds or even thousands of potential coalitions, many of which have a near-zero probability of forming. However, since the introduction of the conditional logit model to the study of government formation a decade ago little has been done to address the unique properties of this discrete choice problem.

The rise of computing power is also changing the types of and approach to discrete choice models in political science, with maximum simulated likelihood estimation (MSLE) and Bayesian techniques opening up a new ability to model heterogeneity across observations and correlations across time and choice alternatives. As a result, models that were impossible to estimate even a few years ago are now within the grasp of the applied researcher. These simulation-based approaches are viewed by many as the second generation of discrete choice models (Train 2009), and they remain almost completely unexplored in a political science setting. We will see further leaps in the development and use of discrete choice models in political science in this area.

References

- Beck, Nathaniel L. 2000. Political Methodology: A Welcoming Discipline. *Journal of the American Statistical Association* 95:651-654.
- Beck, Nathaniel. 2010. Time is Not A Theoretical Variable. *Political Analysis* 18:293-294.
- Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable. *American Journal of Political Science* 42:1260-1288.
- Carter, David B., and Curtis S. Signorino. 2010. Back to the Future: Modeling Time Dependence in Binary Data. *Political Analysis* 18:271-292.
- Gerber, Alan. 1996. African Americans' Congressional Careers and the Democratic House Delegation. *The Journal of Politics* 58:831-845.
- Herron, Michael C. 1999. Postestimation Uncertainty in Limited Dependent Variable Models. *Political Analysis* 8:83-98.
- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science* 44:347-361.
- King, Gary, and Langche Zeng. 2001. Logistic Regression in Rare Events Data. *Political Analysis* 9:137-163.
- Martin, Andrew D., and Kevin M. Quinn. 2002. Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999. *Political Analysis* 10:134-153.
- Sartori, Anne E. 2003. An Estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions. *Political Analysis* 11:111-138.
- Signorino, Curtis S. 2003. Structure and Uncertainty in Discrete Choice Models. *Political Analysis* 11:316-344.
- Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation* (2nd Edition). New York, NY: Cambridge University Press.

About the Author: Garrett Glasgow is an Associate Professor of Political Science at the University of California, Santa Barbara. His substantive interests include individual choice behavior and parliamentary government formation. His textbook on discrete choice methods is forthcoming from Cambridge University Press.