

Methodology Minor Field Exam

Fall 2014

For the minor field exam, you must answer two questions, one in the morning session and one in the afternoon session. In the afternoon session, you may use the software of your choice and will have access to the internet—which you may use to help you analyze data but NOT to communicate with anyone. You are free to use whatever word processing software you like to write your answers. The questions must be answered in the allotted time.

Morning Session: Statistical Theory and Modeling Decisions

Answer one of the following two questions:

1. *Bayesian Statistics:* Bayesian approaches to statistics have become increasingly popular in recent years. The Bayesian framework, while it is often used to answer the same questions to which frequentist methods have been applied, rests on fundamentally different philosophical foundations and uses different methods for estimation. In your view, which of these two things (the foundations or the estimation methods) is the most attractive feature of the Bayesian approach? Be sure to discuss the main differences in both of these things, commenting on both the advantages and disadvantages of the Bayesian approach relative to the standard frequentist approach in terms of fundamental differences as well as estimation methods and applicability to modern problems in political methodology.

Next, consider a linear model in the Bayesian context where income is the dependent variable and gender, race, and education are the independent variables. Describe, in detail, how a Gibbs sampler would be used to estimate the coefficients in this model. Additionally, write out all necessary steps to estimate this model. Once estimated, describe the process necessary to make valid inferences from these results. Specifically, what are the main differences in interpretation and model checking between the frequentist and Bayesian approaches? If there are additional steps required in the Bayesian estimation, carefully describe what these are.

2. *Causal Inference:* Causal inference in the presence of a reciprocal relationship can be difficult. For instance, many scholars believe that nations' economic development leads to the development of democratic institutions, **and** nations' expansion of democratic institutions leads to economic development. Suppose, then, that you worked for a think tank that wanted to make a projection of how much per capita GDP would rise if a country expanded its democratic institutions enough to rise one point on the POLITY scale. It could be difficult to isolate the one-way effect of democracy on economic development if you believe there is a feedback loop between the two.

Why it is so problematic to isolate a causal effect whenever there is a reciprocal relationship between variables? How prevalent of a problem is this in Political Science?

There are a variety of methods designed to deal with the problem of reciprocal causation, but each works best with different varieties of data. Name **three** methods for conducting causal inference when addressing this problem of reciprocal causation. For each method please answer four subquestions: For what kind of data is the method most suitable? How is the method implemented in practice? What are the strengths and weaknesses of the method? What is an example of a real problem to which you would consider applying this technique?

Consider the opening example of the reciprocal relationship between economic development and democratic institutions. As a policy analyst, how would you go about making your forecast to report back to policymakers? Be sure to describe both the kind of data and the estimation method you would use. Why will you stake your reputation on this research design? Do you have any reason for doubt in your forecast? Why or why not?

Afternoon Session: Analyzing Data

Answer one of the following two questions:

3. Please analyze the data set *incumbent.dta* using a logistic regression model. The data are available here:

http://spia.uga.edu/faculty_pages/monogan/teaching/incumbent_logit.dta

The data set contains information on members of the House of Representatives in 1990. The outcome of interest is the variable *returned*—whether the member was returned to the House in 1992, and the input variables are (you must use them all):

- age: The incumbent's age
- surplus: Surplus the incumbent could take home if they retired
- marginal: Dummy—whether the race was close last time
- resistm: Dummy—whether the incumbent's district was redistricted

Present the results of this model in a table including the coefficients, the standard errors, the proportional reduction in error, and any additional information you would like. What can you conclude from the *z*-ratios associated with each coefficient? What can you conclude from the proportional reduction in error?

Please test the conditional hypothesis that the effect of redistricting varies according to an incumbent's age. Please illustrate the nature of this conditioned relationship using predicted probabilities with confidence intervals. Please assess the substantive effect of the other input variables as well, reporting odds ratios and predicted probabilities. What are the tradeoffs of these two interpretation techniques?

Next, estimate the same model but *without including the interaction term* and use a likelihood ratio test to determine whether the more complex model offers a significantly better fit. Lastly, if you wanted to do some kind of residual analysis, how would you go about that in principle?

4. Please analyze the data set *unrest.dta* using a count model. The data are available here:

http://spia.uga.edu/faculty_pages/monogan/teaching/unrest_count.dta

The outcome of interest is the variable *unrest*—a count of protest events in a given country, and the input variables (you must use them all) are:

- CL: Freedom House civil liberties index (1 – 7 scale, with higher values indicating lower levels of civil liberties).
- soviet: Dummy—whether a country is a former Soviet block country
- polity: an index that ranges from –10 to 10 measuring level of democracy (higher values = more democratic)
- politysq: polity squared
- urbanpop: Percentage of a country’s population that lives in an urban setting

Start by fitting a Poisson model and reporting these results. Please test for overdispersion in these data and describe what overdispersion is and why it is potentially a problem. What conclusions can you draw from these tests? What is the best choice of count model for these data and how did you make this choice?

For every set of results you report, present the results in a table (separate or combined, across models) including the coefficients, the standard errors, at least one fit statistic, and any additional information you would like. For the one model you determine to be best for these data, please tell us: What can you conclude from the z -ratios associated with each coefficient? For all models, what can you determine from the fit statistic?

Now test the hypothesis that the effect of Civil Liberties on unrest events is different in former Soviet countries than in the rest of the world. For the one model you determine to be the best for these data, Please illustrate the nature of this conditioned relationship using predicted counts with confidence intervals. For this one model, please assess the substantive effect of all the other input variables as well. When interpreting the effects of other predictors, you may choose among the methods of: partial changes in the conditional mean, factor change in the conditional mean, discrete change in the conditional mean (e.g., predicted counts), or predicted probabilities of counts.