LAB NOTES

Vol. III, No. 3 (Spring 2004)

The Official Quarterly Newsletter of the Political Research Laboratory

In This Issue

A Brief History of the PRL

by Professor Herb Weisberg

The Political Research Lab (PRL or Lab) is now in its 35th year in existence. It was officially established as "Polimetrics" during the 1969-70 academic year under the leadership of Political Science Department Chair Randall Ripley (1969-91) and the first Polimetrics Director C. Richard Hofstetter.

The first incarnation of the Lab was actually initiated by Political Science Department Chair Lawrence Herson (1962-69) as the "Numerical Computation Laboratory." Those were the days of computer cards, so the original Lab equipment obtained by Professor Herson's 1963 NSF grant consisted of a keypunch machine, a counter sorter, and a Monroe/Marchant electric calculator.

The Lab has evolved in many ways since the early days of its first polling on the OSU Oval, during the campus riots of spring 1970. Polimetrics began with a data processing section (staffed by Jim Ludwig), the data archive, a computation section, and a programmer, with a simulation section added a few years later. In the early 1970s, the Lab began the OSU Poll of faculty, staff, and students.

The Lab has seen a succession of directors, with Stu Thorson taking over as head in 1976, followed by Kristi Andersen in 1981. The longest service was by Aage Clausen (1984-1993), who was responsible for such key developments as hiring Kathleen Carr as an experienced survey research professional to run our survey operation, moving to Computer Assisted Telephone Interviewing, setting up the Political Analysis Lab (PAL), and establishing LabNotes for communication with department faculty, staff, and graduate students. Since then, Herb Weisberg (1993-2000), Tom Nelson (2000-2002), and Rich Timpone have served as Lab Directors, with Professor Weisberg returning to the helm in 2002.

Graduate students who worked in the Lab have moved on to a variety of teaching and research positions, including several in the polling industry. For example, early Lab alums who have become prominent in this field include Marty Saperstein and David Krakoff in Columbus, Susan Howell at New Orleans, Cliff Zukin at Rutgers (currently a candidate for AAPOR president), Steve Yarnell in Florida, and Mark Teare and Dee Allsop in Washington, D.C. The Lab has benefited immensely from highly talented professional staff, including our long-time computer wizard Jim Ludwig, Kathleen Carr as head of the survey unit, and now James Norman and Bill Miller at the computer end. Old timers will remember when Elsbeth Connaughton, Karen Baker, and Jodi Hertvik Renshaw served as Lab office managers. More recently, Greg Strizek, Michael Young, Doug Perkins, Kevin Sweeney, and now Greg Miller have served as Assistant Directors, managing the day-to-day operations of the Lab.

Lab facilities have changed dramatically over the years. Polimetrics began when the department was still in the original University Hall, and then moved to the first floor of Derby Hall in the 1970s. We were in temporary space in the bleak surroundings of Neil Hall when Derby was being renovated in 1991-1993. With the reopening of Derby in 1993, we have our modern 2nd floor facility, plus responsibility for the electronic classroom in the basement of Derby.

The Lab has seen vast changes in technology. By the 1970s, using the computer meant using Lab terminals to the mainframe computer on the 5th floor of Baker Systems Engineering, where people went to pick up their printouts. Data were originally on punched cards, which eventually were phased out in favor of data stored on computer tapes at the Computer Center – data that Jim Ludwig subsequently burned on CD-ROMs. Computers came into the political science department twenty years ago when, in 1984, Professor Phil Stewart bought a NESTAR computer network, one of the very first computer networks on campus.

There have been several important recent transitions in the Lab. Tom Nelson started the experimental lab, with the Lab taking over the responsibility for the department's human subjects pool. The survey unit was spun off to become the basis of the University's Center for Survey Research in 1996, after which Polimetrics was renamed as the Political Research Lab. The Lab has become responsible for designing and maintaining the Department's Webpages. As it became possible to access data directly from the Web, the Lab left the data archive business. Finally, in 2003 Jan Box-Steffensmeier started up the PRISM program as a means of revitalizing the lab's emphasis on methods training.

While the Lab has seen many changes over the years, its missions remain constant: training students, research, and service to the department. We expect to continue with these missions over the coming years.

PRL Projects, Plans, and Services

As always, the PRL is involved in a number of interesting projects. Below is a sample of our Winter achievements, Spring plans, and consistently-offered services.

Server Updates

During the winter, James Norman made a number of upgrades to the servers. One of the most useful for students, staff, and faculty is webmail (pictured at right), which can now be accessed at http://polisci.sbs.ohio-state.edu. When you type the address into your browser, you will be asked for your name (typically the last name) and password (the same as when you use Eudora or whatever email program you use). You will then be taken to your inbox. For questions, please contact James Norman at 688-3732 or norman.67@osu.edu. **Note**: you must use Internet Explorer 5.5 or Netscape Navigator 7.1 (or higher).

Computer Upgrades

All of the computers in the experimental lab (Derby 2049A) were upgraded during the Winter quarter. The department has also purchased new computers for the PAL lab, which should see active duty sometime during the spring quarter. The computers from both labs will be used to replace machines in some of the graduate student offices.

In addition, we are still waiting for the results of our request to purchase new computers for the basement lab (Derby 0150), using technology funds from the college. If successful with this grant, we hope to install new computers in the basement sometime late spring or early summer quarter.

Wireless Network

Just a reminder that the PRL has installed a wireless network for those of you wishing to use your laptops in Derby Hall. We recommend that people interested in purchasing a wireless adapter stick to SMC (a little cheaper) or Linksys (a little more range). The 2.4 GHz/11 Mbps cards seem to work best with the department's network. If you have questions please contact Bill Miller at bmiller@polisci.sbs.ohio-state.edu or 292-1814.

Accessing the K: drive

A number of people have had problems accessing their K: drive from outside Derby Hall when using a web browser. We recommend using an ftp client, such as filezilla or netdrive rather than a web browser (both programs will be on the I: drive). To setup these programs, you must input the host, user ID, and password. The host is pol1.sbs.ohio-state.edu, the user ID is .name.grads.polisci or .name.faculty.polisci (where name is what you use to login to the network), and then use your network password. Please contact Greg Miller at 292-1061 or miller.2452@osu.edu if you have any problems.

Experiments and Subject Pool

Winter was another successful quarter for the subject pool, with over 1100 undergraduates participating in six experiments. If you are an instructor looking for ways to give students extra credit, the subject pool will run again during the Spring. Contact Zach Mears at mears.16@osu.edu or 292-0511 for more information or to sign up.

In addition, new guidelines have been developed by the PRL for those planning to use human subjects. Those planning to conduct experiments in the spring quarter or the near future should consult these new rules (which are available from Zach Mears and which will be posted on the web). These changes are designed to ease the burden on the lab staff and on the instructors who are willing to donate students to help their fellow political scientists.

Reducing Spam

Many people have expressed displeasure at the growing amount of spam – mass email selling products like viagra or inappropriate websites – that they receive on a daily basis. There is not much the PRL can do to prevent this, but there are some steps you can take to limit the level of inconvenience:

- Some email clients (e.g., Netscape, Eudora) have the ability to block emails containing certain words or from certain senders. Using this feature will reduce the amount of spam you receive over time.
- Setting your browser preferences to block all cookies will also reduce the amount of spam you receive. Cookies provide email information to those who sell such data, so blocking cookies will reduce the number of spammers who have your information (this *will* effect your ability to view those websites that require active cookies).
- The vast majority of spam we receive is sent to osu.edu rather than to the department account. You can separate these accounts (i.e., don't have one forwarded to the other), which will help separate the spam from the important email.
- Alternatively, you can create a free personal account (e.g., hotmail, yahoo, etc.) for use whenever you fill out online forms or purchase goods online. This will allow for greater privacy on your OSU and department accounts.

Spring Quarter PRISM Events

PRISM will host four meetings in the Spring as part of our ongoing *PRISM Colloquium Series*. All faculty and graduate students are invited to attend. Information regarding the colloquia is included below.

If you have any questions about these events, or if you have ideas for the future, please email Brandon Bartels at bartels.20@osu.edu or stop by his office (2049Q) to chat with him.

Saving Time with Prudent Data Management: Working with Data and Programming in Stata
Friday, April 23, 2004; 2:00 – 3:30 p.m.

Derby Hall 0150 (Basement Computer Lab)
Presenters: Brandon Bartels & Kevin Sweeney

Advanced Programming in Stata
Friday, May 7; 2:00 – 3:30 p.m.
Derby Hall 0150 (Basement Computer Lab)
Presenters: Kevin Sweeney & Brandon Bartels

Teaching PS 585
Friday, May 21, 2004; 12:00-1:30 pm
Derby Hall, Spencer Room

Dynamic Factor Analysis, with an Application to International Relations
Tuesday, May 25, 2004; 12:30-2:00 pm
Derby Hall, Spencer Room
Presenters: Kevin Sweeney & Omar Keshk

Methods Speakers for Spring Quarter

PRISM also will host three speakers during the spring quarter:

- *Walter Enders*, University of Alabama, "After 9/11: Is It All Different Now?" (Sponsored by the Mershon Center)
  Friday, April 9, 2004, 12:00 p.m. Please RSVP to Ann Powers, powers.108@osu.edu, if you plan to attend.
  Professor Enders is an expert in time series analysis, and his book, *Applied Econometric Time-Series*, is the leading book in the field; it is also used in the ITV Time Series course being taught now. Enders' current research focuses on the development and application of time-series models to areas in economics and finance.
- *Narayan Sastry* of the RAND Institute will give a talk in the department this quarter (date and time TBA). Sastry has specialized research interests in demography, duration models, and multi-level modeling.
- *Dave Darmofal*, University of Illinois, will also be giving a talk in the PRISM series (date and time TBA) on recent innovations in spatial modeling with applications to voting data.

Simulation and Substantive Interpretation in Statistical Modeling

By Brandon Bartels
& Kevin Sweeney

Discussion of the substantive *impact* of a variable on a dependent variable, especially for maximum likelihood models, requires more than reporting the sign and significance of coefficient. Substantive interpretation of MLE models is now practically required for publication in major journals. Post-estimation simulation of the model's parameters allows the analyst to calculate these substantive quantities of interest, and most importantly, *it allows the analyst to reflect the degree of uncertainty around those quantities*.

We present the logic of post-estimation simulation and express its importance for presenting both the estimate *and* the precision of a substantive quantity of interest. CLARIFY (hereafter ℭ) (Tomz, Wittenberg, and King 2003; see also King, Tomz, and Wittenberg 2000) is an easy-to-use post-estimation program that simulates parameters and calculates substantive quantities of interest and the degree of uncertainty around those estimates. ℭ can currently be used for linear regression, logit and probit, ordered logit and probit, multinomial logit, count models (Poisson and negative binomial), duration models (Weibull), and seemingly unrelated regressions. Importantly, we emphasize that analysts can move beyond ℭ to simulate parameters of interest from a model, and that researchers estimating models outside of the canned ℭ models should almost always execute post-estimation parameter simulation.

The Logic of Post-Estimation Parameter Simulation Using Clarify

The motivation for post-estimation parameter simulation is provided by the Monte Carlo principle: We can learn about the distributional properties of a random variable, *y*, by sampling many (*m*) times from the probability distribution that generated *y*. ℭ translates this principle to the parameters of a statistical model. Since maximum likelihood parameters have the statistical property of asymptotic normality, we can learn more about a coefficient by drawing *m* samples from a normal distribution underlying the coefficient. Post-estimation simulation used in ℭ, then, involves simulating a *distribution* of each parameter estimated by the model; think of the technique as attempting to simulate the *sampling distribution* of a parameter. ℭ uses Monte Carlo simulation to draw *m* values of each parameter from a multivariate normal distribution, where the mean of the distribution is the vector of point estimates of the coefficients from the model, and the variance is the variance-covariance matrix of those point estimates.[1]

ℭ uses these simulated parameter estimates to generate quantities of interest and most importantly, to reflect uncertainty in those quantities. For example, consider a simple logit model of intended vote choice, where the dependent variable is intention to vote for Bush or not. We want to estimate the effect of economic perceptions—measured as *very bad*, *average*, or *very good*—on the intention of a Bush vote. First, we may be interested in the probability of an intended Bush vote for someone who perceives the economy as *very bad*, holding other variables constant at a baseline value (e.g., the mean). We first estimate the model, and using ℭ, we randomly draw *m*, e.g., 1000, values of each coefficient using the procedure discussed above. To generate the point prediction of an intended Bush vote when economic perceptions is set at *very bad* and the rest of the variables are held constant at their baselines, ℭ first estimates 1000 probability estimates via the simulated parameter estimates. The estimated point prediction ℭ reports is the *mean* of these 1000 probability estimates. Let's say that the mean is 0.15, so we can report that, all else equal, the estimated probability of voting for Bush for someone who perceives the economy as very bad is 0.15.

Importantly, as we discussed above, parameter simulation allows one to report the *precision* of this quantity of interest. To report a 95% confidence interval of this point prediction, ℭ simply sorts the 1000 simulated probability estimates from lowest to highest, and reports the 25th and 975th probability estimates as the lower and upper bounds of this confidence interval. Let's say that these upper and lower bounds are 0.09 and 0.21, respectively. We can now report that with 95% confidence, the probability of voting for Bush for one who perceives the economy as very bad, all else equal, is between 0.09 and 0.21.

We may also want to report an estimate and associated precision of a first difference, i.e., the change in the probability of voting for Bush given a change in economic perceptions. Say we wanted to estimate the change in the probability of an intended Bush vote as economic perceptions change from *average* to *very good*. ℭ estimates *m* of these first differences, and then reports the mean and the 95% confidence interval of the estimate. Assume that the estimated first difference is 0.35. This information alone would lead us to conclude that a change in economic perceptions from *average* to *very good* increases the probability of an intended Bush vote by 0.35. Now, if the lower and upper bounds of the confidence interval are 0.25 and 0.45, respectively, we can conclude with 95% confidence that this first difference is 0.35, plus or minus about 0.10. Reporting this confidence interval is important because it conveys the degree of dispersion around the first difference and also allows one to conclude whether the first difference is statistically different from zero.

Simulation for Substance without Using Clarify

ℭ is useful because it allows analysts to implement the simulation technique in a simple manner. However, the more important point behind the program is that simulation is a powerful technique to gain substantive leverage over our statistical results for *any* type of

[1] Note that the Clarify procedure differs from bootstrapping in that it is parametric, while bootstrapping is nonparametric. For a further discussion, see King et al. (2000, 352).

statistical model. The coefficients and their attendant directionality and statistical significance contain relatively little interesting information (especially in maximum likelihood models), and "substantive" results reported with the coefficients alone will hide the uncertainty that surrounds our point estimates. This holds for *all types* of statistical models, not just those canned in ℭ. We might call this critical underlying point "simulation for substance."

A review of the recent (i.e., post King, Tomz, and Wittenberg 2000) literature in the major journals in our field shows that scholars have overwhelmingly failed to extend the notion of "simulation for substance" to statistical models that are not canned in ℭ. This is a major problem not only because their results are not presented as informatively as they might have been and therefore the import of the piece is not capitalized on, but also because taking the additional step is relatively easy and *does not* require the ℭ program. To show this here we will consider a member of the increasingly popular sample selection model family: censored probit.[2] Simulation for this model is not canned in ℭ, but can easily be programmed by the analyst. The left column (on the next page) walks through the model and relevant simulation, the right column displays Stata code for achieving "simulation for substance" in the censored probit model.

In conclusion, since post-estimation parameter simulation maximizes the quality of the presentation of results from a statistical model, analysts should be encouraged to use this technique whenever possible, even for models outside of ℭ. Building on last quarter's session on post-estimation, PRISM will offer a session this quarter on advanced programming, which will cover the topic of post-estimation parameter simulation for models not canned in ℭ. See below for more information on this session.

[2] Space limitations preclude us from explaining sample selection models in more detail, but we chose this example because, as a two stage model, it may appear very complicated to simulate quantities of interest. It is not. You will likely need to be somewhat familiar in order to follow what is below. To bone up on your own: Heckman (1979) is the most cited reference for sample selection models, Dubin and Rivers (1989) extend the Heckman model to cases where the dependent variable in the outcome equation is binary (i.e. censored probit), and Timpone (2002) and Sweeney and Fritz (2004) provide examples of "simulation for substance" applied to the censored probit model.

References

Dubin, Jeffery, and Douglas Rivers. 1989. "Selection Bias in Linear Regression, Logit, and Probit Models." *Sociological Methods and Research* 18:360-90.

Greene, William H. 2000. *Econometric Analysis*. 4th Edition. Upper Saddle River, New Jersey: Prentice Hall.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153-61.

King, Gary; Michael Tomz; and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44: 341-355.

Sweeney, Kevin, and Paul Fritz. 2004. "Jumping on the Bandwagon: An Interest Based Explanation for Great Power Alliances." *Journal of Politics* 66(2):428-49.

Timpone, Richard J. 2002. "Estimating Aggregate Policy Reform Effects: New Baselines for Registration, Participation, and Representation." *Political Analysis* 10:154-77.

Tomz, Michael, Jason Wittenberg, and Gary King. 2003. *CLARIFY: Software for Interpreting and Presenting Statistical Results*. Version 2.1. Cambridge, MA: Harvard University. http://gking.harvard.edu.

1. Selection Equation:

$$y1_j = z_j \gamma + u_{2j}$$

Outcome Equation:

$$y2_j = x_j \beta + u_{1j}$$

Where:

$$u_1 \sim N(0,1)$$
$$u_2 \sim N(0,1)$$
$$corr(u_1, u_2) = \rho$$

/*Step One: Estimate Model*/ [3]

```
>heckprob y2 x1 x2 x3 x4, sel(y1 = z1 z2 z3 z4) robust
```

2. Simulate the model parameters by drawing from the multivariate normal distribution. Note: there are 11 – 4 Xs, 4 Zs, 2 constants, and ρ (the correlation between the errors).

/*Step Two: Draw $\tilde{\beta}$ from multivariate normal, mean $\hat{\beta}$ and Covariance Matrix $\hat{\Sigma}$.*/

```
>matrix params = e(b)
>matrix P = e(V)
>drawnorm b1-b11, means(params) cov(P) double
```

3. Stata estimates the hyperbolic arctangent of ρ, so we must simulate to get the actual ρ.

/*Step Three: Generate Simulated Rho*/

```
>gen simrho = (exp(2*b11)-1)/(exp(2*b11)+1)
```

4. Initiate a looping structure to generate *m* (in this case 1,000) simulated first differences for the effect of x1 on y2 comparing when x1 is at its mean (the base model) to when x1 is at a value two standard deviations denoted (_m2sd) below its mean.

/*Step Four: The Loop*/

```
>local i =1
```

/*A. Generate variables that will be used to fill in a cell of Substantive Table*/

```
>generate base_y2=.
>generate x1_m2sd=.
>while `i' <=1000 {
```

/*B. Generate zγ for the selection equation.*/

```
>quietly generate select = b6[`i'] + (b7[`i']*z1)
+ (b8[`i']*z2) + (b9[`i']*z3) + (b10[`i']*z4)
```

/*C. Generate xβ for the outcome equation.*/

```
>quietly generate outcome = b1[`i'] +
(b2[`i']*x1) + (b3[`i']*x2) + (b4[`i']*x3) +
(b5[`i']*x4)
```

D. This is the meat of the simulation. The first three commands generate the probability of being selected and experiencing the outcome (p_11) for the base model. For the censored probit, this probability is (Greene 2000, 857):

$$\Phi_2[\beta' x, \gamma' z, \rho]$$

Then, we do the same for the case where X1 is two standard deviations below its mean (x1_m2sd).

This gives us 1 simulated first difference.

/*D. Generate the relevant first difference*/

```
>quietly generate p_11 =
binorm(outcome,select,simrho)
>quietly summarize p_11, meanonly
>quietly replace base_y2=r(mean) in `i'
>quietly generate x1_m2sd=outcome -
(b1[`i']*x1) + (b1[`i']*-0.2)
>quietly generate p11_x1_m2sd
=binorm(x1_m2sd,select,simrho)
>quietly summarize p_x1_m2sd, meanonly
>quietly replace x1_m2sd=r(mean) in `i'
```

5. To get the other 999 we drop the three variables we just generated and repeat the loop until `i' = 1,000.

/*Step Five: Do the Loop Again*/

```
>drop select outcome p_11 x1_m2sd  p11_x1_m2sd
>disp `i'
>local i=`i'+1
>}
```

6. When we're done with the 1,000 simulations, we can use the centile command to get the relevant distributions. To do this for each variable in the model, we would embed this loop within a larger looping structure.

/*Step Six: Get the Median and Confidence Intervals for your first difference*/

```
>centile base_y2 x1_m2sd, centile(2.5 50 97.5)
```

[3] Stata commands are preceded by >.

PRL Staff Information

Title	Name	Office	Phone	Email
Director	Herb Weisberg	2049D	292-6572	weisberg.1@osu.edu
Associate Director	Jan Box-Steffensmeier	2049S	292-9642	jboxstef+@osu.edu
Assistant Director	Greg Miller	2049R	292-1061	miller.2452@osu.edu
Systems Manager	James Norman	2049K	688-3732	norman.67@osu.edu
Systems Manager	Bill Miller	2049E	292-1814	bmiller@polisci.sbs.ohio-state.edu
Graduate Assistant	Brent Strathman	2049F	292-0511	strathman.5@osu.edu
Graduate Assistant	Zach Mears	2049F	292-0511	mears.16@osu.edu
Undergraduate Assistant	Yared Debebe	2049H	292-0511	debebe.1@osu.edu
PRISM Senior Fellow	Kevin Sweeney	2049Q	292-9661	sweeney.101@osu.edu
PRISM Junior Fellow	Brandon Bartels	2049Q	292-9661	bartels.20@osu.edu

To contact all Lab Staff simultaneously, email us at prl@polisci.sbs.ohio-state.edu

Previous issues of Lab Notes and other valuable information can be found at the Lab's website:
http://psweb.sbs.ohio-state.edu/prl/index.htm