LAB NOTES

Vol. IV, No. 1 (Autumn 2004)

The Official Quarterly Newsletter of the Political Research Laboratory

In This Issue

Director's Corner

by Professor Herb Weisberg

The beginning of a new academic year always brings changes in the Lab, and this year is certainly no exception. Jeff Martinson is the new Assistant Director, handling day-to-day management issues as well as editing Lab Notes. Bill Miller, James Norman, and David Sweasey are our staff information-technology specialists. Zach Mears will be working with Bill Miller on the department's web page. Jim DeLaet is taking over the Human Subjects Pool.

Additionally, on the PRISM side, David Darmofal from the University of Illinois is the Senior PRISM fellow this year and Visiting Assistant Professor, while Cory Smidt is the Junior PRISM fellow. And Yared Debebe and Matt Thompson will continue on as undergrad work-study helpers.

This is a very strong team, so be sure to come to them with your computer and methods problems. At the same time, we will miss Greg Miller, Brent Strathman, and Kevin Sweeney, who have done great service to the Lab over the last several years.

There is also exciting news about computer upgrades. The last round of distribution of new Windows XP computers to faculty, staff, and the PAL lab is now about complete. We have also upgraded the computers in the basement computer lab (150 Derby) to run Windows XP with expanded memory.

We are also currently getting a new basement computer lab (125 Derby) ready for winter 2005 classes, at which time the computers currently in 150 Derby will be moved to grad student offices, with the hope of replacing all of our Windows 98 machines.

The shift away from Windows 95 and 98 to Windows 2000 and XP allows us to handle security matters more efficiently, including automatic virus checking updates, automatic Windows critical updates, and better ability to limit downloads on lab computers (which should keep them working correctly longer).

As always, I want to encourage students and faculty to continue to make use of the University's healthy list of Site-Licensed Software, where appropriate. A listing of site-licensed software arrangements at OSU is available on the Lab web page, at:

http://psweb.sbs.ohio-state.edu/prl/faq/

Remember, SPSS is free at OSU and Stata can be purchased on the GradPlan for $129.

Faculty are expected to buy computer programs for their research on their PSA accounts; faculty who need other programs for their teaching should send the PRL a one-paragraph description of their need (with a copy to Kathleen McGraw if the software is expensive).

The K:drive for Grad Students

There are many reasons why grad students should remember to store their files on their K: drive rather than the C: drive. The K: drive is backed-up nightly, so the work you've saved there is retrievable if anything should go awry with an individual desktop unit.

Be aware also that whether in an office or in one of the common areas, "your" computer is actually being shared by a number of other individuals. Anything saved on the C: drive is potentially viewable (and corruptible) by them. The K: drive, on the other hand, is your own private storage area, accessible only by you.

In addition, in order to keep the computers operating more efficiently, we reclone them as often as once a week. When this occurs, everything on the C: drive is replaced. We do not give prior notification of this procedure.

Finally, there is the issue of remote access, which is impossible with respect to the C: drive. Conversely, those with a network account can access the K: drive from anywhere with Internet access.

As the preferred method, we recommend using an ftp client, such as FileZilla, NetDrive, or FTPCommander in order to do this. To set up 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 (where name is what you use to login to the network), and your existing network password.

Alternatively, to view your K: drive using a web browser, the procedure is as follows. In the internet address area, type:

ftp://.name.grads.polisci@pol1.sbs.ohio-state.edu

(where name is the same name you will use to login to the network); next, input the same password you use to login to the network (be sure that when you are asked for your password, the username is listed as ".name.grads.polisci"). The files in your K: drive should then appear in the browser window.

Functional Capability	***C: drive***	***K: drive***
online access	No	YES
departmental back-up	No	YES
access from all Derby computers	No	YES
personal use only	No	YES
regularly recloned (erased)	YES	No

The K:drive for Faculty

For many of the same reasons stated above, we would like to also encourage the faculty to use the K:drive instead of their C:drives.

The procedure for remote access is slightly different, however. Using an ftp client such as FileZilla, NetDrive, or FTPCommander, input your host as pol1.sbs.ohio-state.edu, the user ID is .name.faculty.polisci (where name is what you use to login to the network), and your existing network password

You also can use an internet browser to view your K: drive. The procedure is as follows. In the internet address area, type:

ftp://.name.faculty.polisci@pol1.sbs.ohio-state.edu

(where name is the same name you will use to login to the network); next, input the same password you use to login to the network (be sure that when you are asked for your password, the username is listed as ".name.faculty.polisci"). The files in your K: drive should then appear in the browser window.

Wireless Network

Many of you have utilized the department's wireless over the past several months. For those still unfamiliar with the system, it is a wireless network for anyone wishing to use a laptop to connect to the Internet in Derby Hall.

Please be aware that PRL will be updating the password system this Fall, which will necessitate the assignment of new passwords to all wireless users. We expect to implement this change by October 18th.

Of course, further notices will precede this change. In the meantime, if you have any questions about wireless, please contact the Lab staff.

Experiments and Subject Pool

New IRB Regulations

The Office of Responsible Research Practices (ORRP), perhaps better known as "the office that handles IRB approvals for research involving human subjects," has announced that ALL personnel on research projects who are involved with human subjects in research (faculty, research assistants, students, postdocs) will need to complete online training in human subjects protection. This is not just for funded efforts, and it also includes thesis and dissertation research--anything requiring IRB approval.

If you are planning an IRB submission, you must complete the training before DECEMBER 1, 2004. This includes any renewal submissions.

Unfortunately, completion of the training requirement for individuals with NIH funding CANNOT substitute for this new training. Conversely, the ORRP's new training does satisfy NIH's requirement for training.

Finally, please make plans ahead of time for completion of this online training. The Behavioral and Social Sciences version contains 11 modules, and is estimated to take a total of 6 hours to complete. (This does not need to be at a single sitting; you can complete part of the training, save your work, and return to it at a later time). If your protocols go through the Biomedical IRB, there are even more modules to complete. For many of you, early quarter schedules are a bit more flexible, and "soon" might be a good time to get this out of the way, before end of the quarter chaos – and the December 1st deadline -- descend.

[To learn more about this and/or to take the online course, go to http://orrp.osu.edu/education3.cfm . If you have questions, email Jim DeLaet or Zach Mears]

Autumn Quarter PRISM Events

To coincide with the new academic year, PRISM is organizing some new events for the quarter as well as putting on some trusted favorites.

Methods Lunches: With a tip of our hat to our friends in judicial politics, PRISM is organizing two 'methods lunches' this quarter. On Tuesday, October 5th and Tuesday, November 9th we plan to meet at noon in the east foyer and then head on down to High Street for lunch. The lunches serve as a good opportunity for graduate students and faculty to talk about all things related to methods as well as a way to

simply get to know each other a little better. First-year graduate students who have questions about the methods program and how it might be useful to them are encouraged to come. All are welcome, of course, as we hope to see you there!

Brownbag Series: PRISM plans on continuing to hold informational brownbag presentations on methods topics. These range from presentations on specific statistical techniques to presentations on statistical software or other computer programs often used in political research. We plan on holding three presentations this quarter, kicking it off with some old favorites.

Introduction to Stata *(a PRISM brownbag classic)*
Tuesday, October 5; 3:30 – 5:00 pm
Derby Hall 0150 (Basement Computer Lab)
Presenter: Corwin Smidt

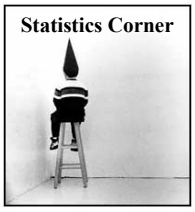
A broad overview of Stata, including information on getting started, as well as details on how to do basic data management, generate descriptive statistics commands, and run various estimation commands. The session is targeted to those who have never used Stata, but those who have a vague applied understanding of Stata may also find the session useful. This session is also geared toward first-year graduate students (since you will soon be using Stata in 685 and 686) and those enrolled in Dean Lacy's 786.

Two more brownbags are currently on taps as well; dates have yet to be set in order to determine the best time. Another brownbag classic, *Post-Estimation Techniques in Statistical Analysis: Introduction to Clarify, S-Post, and Graphing*, will go into further uses of Stata and a new presentation, *An Introduction to R*, will introduce individuals to the free version of the S-plus program.

Speaker Series: We will continue our methods speaker series this year as well. With job talks yet to be set, specific talk dates and times have not been set. All faculty and graduate students are invited to attend.

Office Hours: For those of you who may have been gone this summer, PRISM's methods fellows are now David Darmofal (darmofal.3@osu.edu) and Corwin Smidt (smidt.2@osu.edu). Faculty or graduate students having questions about methods, either in general or in specific applications, are encouraged to contact one of them or visit them during their office hours. Dave's office is 2049N and he has office hours Monday/Wednesday 11-4, Tuesday 10-5, and Friday 11-3. Cory's office is 2049Q and he has office hours Monday-Wednesday 12-2 and Thursday 10-3. Both are available by appointment as well.

If you have any questions about these events, or if you have ideas for the future, please email Corwin Smidt or stop by his office to see him.



Spatial Modeling

By David Darmofal

All political science data are spatial data: the political phenomena we seek to understand occur at specific geographic locations. And these locations are central to our understanding of these phenomena. Across a broad range of substantive concerns, from political communication to political conflict, democratization to dependency, policy diffusion to party mobilization, our theories posit that spatial location is a critical factor shaping political behavior. But our empirical methods, until recently, provided no means for us to treat these spatial effects in a rigorous, generalized, and systematic manner. Recent advances in spatial econometrics, however, now allow us to incorporate just such a treatment in our models.

What do we mean by spatial effects? There are two broad classes of spatial effects. The first is spatial heterogeneity. Spatial heterogeneity may take several forms. Independent variables, for example, may have differing effects in different spatial locations. Or the variance in the errors in our estimates may not be constant across all spatial locations. Spatial heterogeneity can be handled via standard econometric techniques. The spatial dimension of this heterogeneity, however, is rarely considered in standard econometric texts, and can be quite helpful in diagnosing heterogeneity.

The second form of spatial effect is spatial autocorrelation, which exists when similarity in attributes of interest is related to the locations of observed units. Enduring rivalries between neighboring states, policy diffusion between neighboring governments, and shared partisanship between neighboring citizens are all political science examples of spatial autocorrelation. Spatial autocorrelation, unlike spatial heterogeneity, cannot be handled in a general manner via standard approaches to statistical inference such as OLS models, or probit and logit models for limited dependent variables. Instead, a generalized treatment of spatial autocorrelation requires the use of spatial econometrics.

The first step in spatial modeling is to conduct a spatial analysis to determine whether one's data are, in fact, spatially autocorrelated. A critical initial decision is the definition of neighbors. This definition should be based on substantive theory. Neighbors can be defined literally, as contiguous observations. Alternatively, neighbors can be defined according to a distance criterion, with observations beyond a certain distance defined as non-neighbors. In fact, the definition of neighbors is completely generalizable, and need not have a spatial component at all. For example, observations within geographically dispersed social networks could be considered neighbors. The basic logic of spatial econometrics is thus applicable to a variety of research questions.

Once the researcher has defined neighboring units, tests of global and local spatial autocorrelation can be conducted. The most widely used measure of global spatial autocorrelation is Moran's *I*, which is defined as:

$$I = \frac{N}{S_O} \frac{\sum_i \sum_j w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_i \left(x_i - \bar{x}\right)^2}$$

where w_{ij} is the (i,j)th element in a spatial weights matrix, S_o is the sum of the weights, N is the number of observations, x_i and x_j are the values at locations i and j on the variable of interest, and $\bar{x}$ is the mean of the variable of interest. A significant positive Moran's *I* indicates positive spatial autocorrelation (neighboring units share similar values on the variable of interest), while a significant negative Moran's *I* indicates negative spatial autocorrelation (neighboring units have dissimilar values on the variable of interest). If the global Moran's *I* indicates that the data are spatially autocorrelated, the next step is to determine which observations are autocorrelated with their neighbors. This is determined through a local indicator of spatial association (a LISA statistic). The most common LISA statistic is the local Moran's *I*, which is defined (using the same notation as for the global Moran) as:

$$I_i = \frac{\sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(x_i - \bar{x})^2}$$

The local Moran has two important characteristics. First, it provides an individual measure of local spatial autocorrelation for each observation *i* in the data (in relation to *i*'s neighbors *j*), an indication of whether this autocorrelation is statistically significant, and the direction of this autocorrelation (i.e., positive vs. negative). Second, the sum of the local Moran's *I*'s is proportional to the global Moran's *I*. This allows the global Moran to be

decomposed to determine which observations are producing the global result, which observations are insignificant, and which observations run counter to the global trend (see Anselin 1995).

Once the researcher has identified the spatially autocorrelated observations, the next step is to model this autocorrelation via a spatial econometric model. The critical distinguishing feature of spatial econometric models is that unlike more familiar statistical models, they directly incorporate spatial dependence between observations. These models take two principal forms, depending upon the hypothesized form of the autocorrelation.

Consider the example of shared partisanship between neighbors. We may believe that Democrats live near Democrats and Republicans live near Republicans because of political communication between neighbors. In other words, citizens communicate their partisanship and influence the partisanship of their neighbors. This suggests a direct spatial effect of neighbors' partisanship on an individual's partisanship.

We would model such an effect through a spatial lag model. Such a model takes the form $y = \rho Wy + X\beta + \varepsilon$, where Wy is a spatially lagged dependent variable, ρ is a spatial autoregressive parameter, such that $\rho > 0$ indicates positive spatial autocorrelation, $\rho < 0$ indicates negative spatial autocorrelation, and $\rho = 0$ indicates no spatial autocorrelation, and y, X, β, and ε are the same as in a standard model. Within the framework of the partisanship example, y, the dependent variable of interest, is the partisanship we seek to explain, Wy is neighbors' partisanship, and ρ captures the direct effect of neighbors' partisanship.

A significant ρ greater than zero indicates that neighbors' partisanship has a direct positive effect on individuals' partisanship (i.e., Democrats are Democrats, in part, because their neighbors' Democratic partisanship makes them more likely to be Democrats, and similarly for Republicans). A significant ρ less than zero means that neighbors' partisanship has a direct negative effect on individuals' partisanship (i.e., Democrats are Democrats, in part, because their neighbors' Republican partisanship makes them more likely to be Democrats, and vice versa for Republicans). An insignificant ρ that is not distinguishable from zero means that neighbors' partisanship has no direct effect on individuals' partisanship.

The consequences of ignoring this type of spatial autocorrelation are quite significant for statistical inference. If a spatial lag model is the proper specification, and this direct spatial dependence is ignored, we risk biased coefficient estimates and inaccurate inferences. Given how many of our theories suggest a direct contextual influence on political behavior, the significant consequences of ignoring these spatial effects suggest a broader use of the spatial lag model in political science.

The second form of spatial econometric model is a spatial error model. In this model, the spatial autocorrelation between neighboring observations is not the product of a direct spatial effect. Instead, it results from the fact that similar observations are located near each other.

Consider again, the example of shared partisanship between neighbors. Perhaps Democrats live near Democrats and Republicans near Republicans not because of a direct effect of neighbors' partisanship, but instead, because co-partisans tend to live near each other because of a shared external factor. For example, partisanship is correlated with socioeconomic status in the United States. And place of residence is correlated with socioeconomic status in the United States. Thus, Democrats may live near other Democrats because of a shared socioeconomic status, and likewise for Republicans. Assume that we lack a measure of individuals'

socioeconomic status. If so, we would model this effect of omitted socioeconomic status through an autocorrelated error term in a spatial error model.

The spatial error model takes the form $y = X\beta + \varepsilon$, where $\varepsilon = \lambda W\varepsilon + \xi$, $W\varepsilon$ is a spatially lagged error term, λ is a spatial autoregressive parameter, such that $\lambda > 0$ indicates positive spatial autocorrelation, $\lambda < 0$ indicates negative spatial autocorrelation, and $\lambda = 0$ indicates no spatial autocorrelation, ξ is an independent and identically distributed (i.i.d.) error term and y, X, and β are the same as in a standard model.

A significant λ greater than zero indicates positive spatial autocorrelation in variables omitted from the model (e.g., similar socioeconomic status between neighbors). A significant λ less than zero indicates negative spatial autocorrelation in variables omitted from the model (e.g., dissimilar socioeconomic status between neighbors). An insignificant λ that is not distinguishable from zero indicates no significant spatial autocorrelation in omitted variables.

The consequences of ignoring spatial error dependence are less severe than those of ignoring spatial lag dependence. In the presence of ignored spatial error dependence, estimates will remain unbiased. However, they will be inefficient, and standard error estimates will be biased downward. This has the consequence of producing Type I errors in statistical inference (rejecting true null hypotheses).

In choosing between the spatial lag and spatial error specifications, substantive theory should be one's guide. However, often times we lack an a priori theory that would argue for lag or error dependence. When this is the case, we can examine diagnostics from OLS models that can point to the correct specification. There are several varieties of such diagnostics, including Lagrange multiplier tests, robust Lagrange multiplier tests (which are robust against the alternative specification), and the Kelejian-Robinson error test (which relaxes the assumption of normality common to the LM tests). By examining the significance levels of various diagnostics, we can determine whether a spatial lag or a spatial error model is the proper specification.

Both spatial data analysis and spatial econometrics will be covered in a course on Spatial Modeling that I will be teaching in the Winter quarter. The course runs from January 21st through March 11th. Additional information on the course can be found at:

http://psweb.sbs.ohio-state.edu/faculty/jbox/ITV/ITVCourses.html#future

.

As the Senior PRISM Fellow, I also am happy to help with any difficult stats questions or data issues you may have. My office is 2049N Derby Hall.

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	Jeff Martinson	2049R	292-1061	martinson.3@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
Systems Manager	David Sweasey	2049R	292-1061	sweasey.1@osu.edu
Graduate Assistant	Jim DeLaet	2049F	292-0511	delaet@polisci.sbs.ohio-state.edu
Graduate Assistant	Zach Mears	2049F	292-0511	mears@polisci.sbs.ohio-state.edu
Undergraduate Assistant	Yared Debebe	2049H	292-0511	debebe.1@osu.edu
Undergraduate Assistant	Matt Thompson	2049H	292-0511	thompson.1070@osu.edu
PRISM Senior Fellow	David Darmofal	2049Q	292-9661	darmofal@polisci.sbs.ohio-state.edu
PRISM Junior Fellow	Corwin Smidt	2049Q	292-9661	smidt@polisci.sbs.ohio-state.edu

At least one staff member is on-call between 8:30 AM and 6:00 PM, Monday through Friday.

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

To contact the HELP DESK, go to http://inet.sbs.ohio-state.edu/ and click on "HELP DESK."

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