Discrete Dependent Variable Models

For some weeks you have been working with regression. Two of the assumptions of regression are that the dependent variable can assume an unlimited number of values and that the dependent variable is measured at either the "interval" or "ratio" level (see page 10). However, political scientists routinely use regression when the dependent variable is restricted in the number of categories it can attain. For example, if you are in POSC 300B, you have completed Assignment #2 where the dependent variable was a percentage. Since the dependent variable could not attain any value other than those between 0 and 100, technically, this was a violation of one of the above assumptions. However, since the dependent variable could attain a great number of different values (i.e., 101 - "0" plus "1" through "100"), violating this assumption was not important. Alternatively, when the dependent variable can only assume from say 2 to 7 different values this limitation does become serious. When a variable can assume only a small number of possible values, it is termed "discrete."

Political scientists frequently encounter situations where the dependent variable is "discrete." For example, an international relations scholar may be studying the causes of war. Typically, the dependent variable in such studies has only two categories of responses (e.g., coded "1" if war occurred or "0" if war did not occur). In many studies of voting behavior the dependent variable has only two categories of responses (e.g., coded "1" if the voter voted Democratic, "0" if the voter voted Republican). In the example we will be using shortly, the dependent variable is whether a senator voted "for" (coded "1") or "against" (coded "0") the Levin Amendment to raise taxes on high income earners.

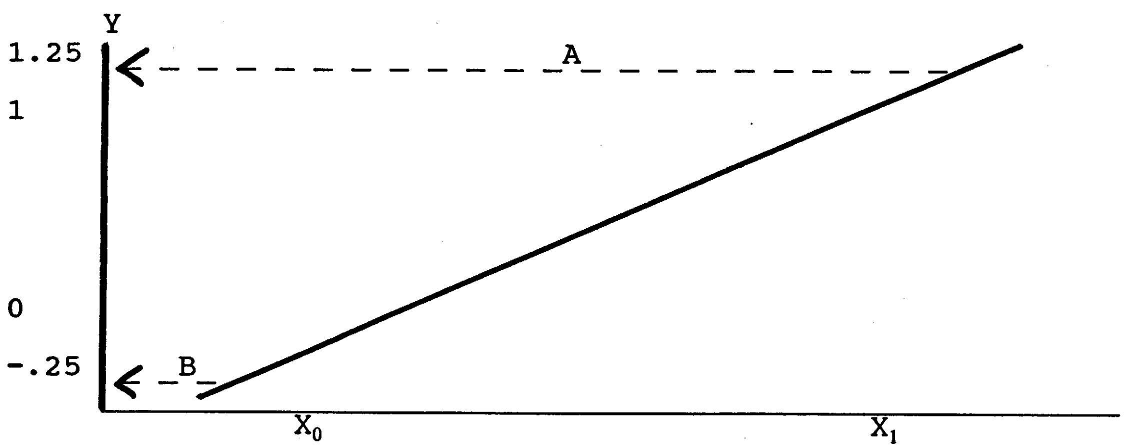
In addition to violating the regression assumption that the dependent variables has a large, or unlimited, number of categories of responses, discrete dependent variables frequently do not meet another assumption of regression: that the dependent variable is measured at either the "interval" or "ratio" level (just keep reading). For example, suppose an international relations scholar has a dependent variable with three categories of responses. The dependent variable might concern how a crisis was resolved. The options might be as follows: (1) either the challenger or the defender conceded; (2) an international body (e.g., the United Nations) tried to enforce a peaceful settlement; (3) the conflict ended in war. Is the difference between categories "1" and "2" the same as the difference between categories "2" and "3"? I doubt it. Therefore, not only is the dependent variable restricted to just three categories of responses, but we do not have an equal interval between the categories of responses. If the dependent variable does not have an equal interval of measure then it cannot be measured at either the interval or ratio level (i.e., it must be nominal or ordinal - see pages 9-10).

When the dependent variable has a small number of categories of responses (i.e., is discrete), regression is inappropriate and political scientists must use one of the techniques discussed in this reading assignment (e.g., logit or probit). The most frequent use of discrete dependent variable models in political science involve a dependent variable that has only two categories of responses. So, for right now, let us consider that situation. For reasons I will discuss shortly, when we have a dependent variable with only two categories of responses (i.e., dichotomous), regression is inappropriate. For a period of time, this situation left political scientists in an uncomfortable position. It seemed that political scientists either had to give up the many advantages of regression (see pages 35-38) or use regression and violate its assumptions. Fortunately, this is no longer the case.

Over the past two decades, the level of statistical analysis in political science has greatly improved (if you are interested, see Gary King, "On Political Methodology," Political Analysis, volume 2, 1990, pages 1-29). Two of the techniques that political scientists "imported" during this statistical upgrade were "logit" and "probit" analysis. These two techniques deliver the tremendous advantages of regression (again, see pages 35-38) in situations where the dependent variable has only two categories of responses. Additionally, both logit and probit analysis can easily be adapted to situations where the dependent variable has more than two, but still a relatively small number of possible responses (e.g., three, four, etc. categories of responses). Not surprisingly, these two techniques have become extremely popular in contemporary quantitative political science research.

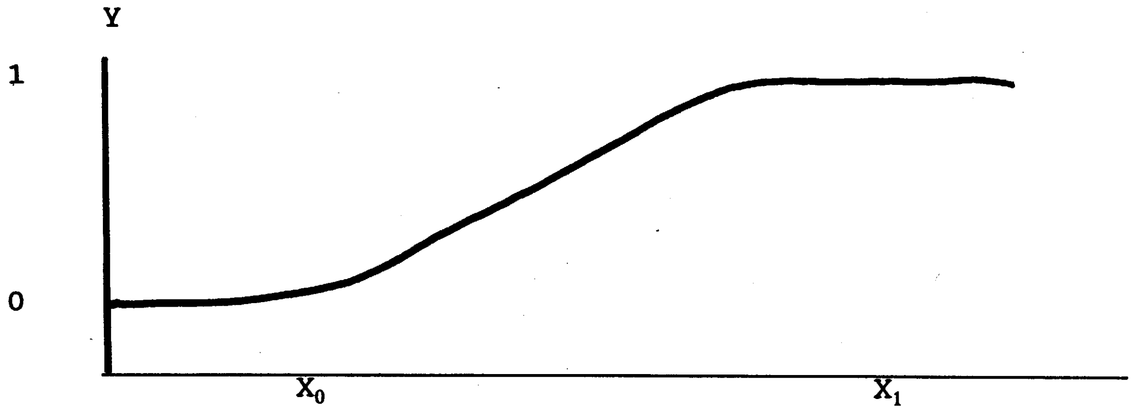
The decision of whether to use logit or probit is largely arbitrary. The researcher will obtain virtually the same results using either method. In my opinion the "logic" of logit is more readily understandable than probit. Therefore, we will examine logit in this reading assignment. Before discussing how to use logit, I will first show what happens when we use regression with a dependent variable that has only two categories of responses.

In the example below the dependent variable (Y) has only two possible categories of responses ("1" and "0").



The regression line produces nonsensical predictions. For example, values on variable X that are greater than X_1 (i.e., to the right of X_1) predict that the value of the dependent variable (Y) will be greater than 1 (e.g., 1.25, see dashed line "A" above). That is impossible; values on the dependent variable (Y) can only be 0 and 1. Similarly, values on variable X that are less the X_0 (i.e., to the left of X_0) predict that the value of the dependent variable will be less than 0 (i.e., a negative number, e.g., -.25, see dashed line "B" above), again impossible. Such predictions are nonsensical. While there are several other problems in using regression with a dependent variable having only two categories, the most important problem is the one I have just mentioned.

Consequently, when we have a dependent variable with only two categories of responses, we will use the logit "line" (i.e., model) diagrammed on the next page. Notice how this line only predicts scores on the dependent variable (Y) between 0 and 1. This eliminates the problem mentioned previously.



Logit (which comes from the word logarithm) produces coefficients (i.e., "b's") that estimate the change in the log odds ratio of the dependent variable for a one unit increase in a particular independent variable holding the level of all other independent variables constant. As this is unlikely to make much sense, let me explain. The odds ratio is simply the ratio of the probability of occurrence to the probability of non-occurrence. Thus, if an event (such as a senator voting "yes") had a probability of .7, then the probability of non-occurrence (voting "no") must be .3 (1-.7=.3). Hence the odds ratio would be .7/.3 or 2.33. Using natural logs (base = 2.71828) the log of the odds ratio would be whatever power 2.71828 must be raised to equal 2.33 (2.71828 to the .84586th power is approximately equal to 2.33 - i.e., $2.71828^{.84586}$ = approximately 2.33).

Logit coefficients tell us the change in such a "log of the odds" ratio. The higher the ratio the greater the probability of occurrence to non-occurrence. A positive sign for a coefficient would indicate that as scores on this independent variable increase (while holding the level of all other independent variables constant) the "log odds ratio" (which means that the probability of the dependent variable attaining a score of "1") increases.

Admittedly, a "log odds ratio" makes the interpretation of the coefficients difficult. However, the combination of a "log odds ratio" and a non-linear model (note that the line at the top of the this page is a curved, as opposed to a straight, line) eliminates the problem mentioned on the previous page.

To introduce logit let use an example from legislative politics. The dependent variable in this study is how the senator voted on the Levin (Democratic Senator from Michigan) Amendment to raise income taxes on those earning $200,000 or more per year. The independent variables are the degree of the senator's liberalism, the senator's political party affiliation, the percentage of the statewide vote for Michael Dukakis in 1988 (this is a measure of how liberal the state's vote are) and the median family income in the senator's state. The code names for the variables and a description of each variable appears ahead:

Levin = each senator's vote on the Levin Amendment to raise federal income tax rates 4% on incomes at, or above, $200,000 per year (0 = No, 1 = Yes). This is the dependent variable.

ADA89 = the percentage of votes each senator cast in support of positions taken by the Americans for Democratic Action. Scores range from least from 0% (least liberal) to 100% (most liberal).

Bush88 = percentage of the statewide vote for George H. W. Bush in the 1988 Presidential election. Higher scores denote greater constituency conservatism.

Medinc = state median family income in thousands of dollars (i.e., 32.1 = $32,100).

With the variables above I asked the computer to estimate a logit equation of the vote on the Levin Amendment. Through this discussion I will refer to the equation where the dependent variable is "Levin" and the independent variables are "ADA89," "Bush88" and "Medinc" as the "main equation." The results for the "main equation" appear immediately below.

Main Equation

```
Logistic regression                               Number of obs   =         99
                                                  LR chi2(3)      =      47.09
                                                  Prob > chi2     =     0.0000
Log likelihood = -44.218855                       Pseudo R2       =     0.3475
```

levin	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ada89	.0568186	.0108992	5.21	0.000	.0354566	.0781805
bush88	.0629142	.0576857	1.09	0.275	-.0501477	.1759761
medinc	-.1609074	.1210805	-1.33	0.184	-.3982208	.076406
_cons	-3.320801	4.365551	-0.76	0.447	-11.87712	5.235522

As a first step in interpreting the coefficients from the "main" equation, let me suggest just looking at the sign and the level of statistical significance for each independent variable. For example, the coefficient for ada89 (i.e., the value in the "Coef." column for ada89) is .056. This coefficient value is positive (i.e., .056 not -.056). This means that *after* removing the impact of Bush88 and medinc, as the senator's score on ada89 increases (i.e., the senator becomes more liberal) the greater the probability the senator will vote "yes" on the Levin Amendment (since "1" - a "yes" vote is a higher score than "0" – a "no" vote). Since the Levin Amendment raises tax rates on high income individuals, one would logically hypothesize that the relationship between ada scores (which measure the degree of a senator's liberalism) and the probability of a particular senator supporting higher taxes on the wealthy would be positive. Thus, the more liberal the senator the more likely they should be to vote in favor of the Levin Amendment. This is exactly what a positive sign for the ada coefficient indicates. If I had used a conservative group rating the sign would have been negative and probably of similar magnitude.

The t ratio for ada89 is 5.21 (the "z" column contains the t ratios). Since 5.21 has an absolute value greater than 2.0, the result is easily statistically significant. Put another way, the null hypothesis (i.e., a senator's ada score is unrelated to the probability they will vote "yes" on the Levin Amendment) is true less than 5% of the time. Notice in the P>Z column the score for ada89 is .000. This means that the null hypothesis is true less than 1 time in 1,000 (i.e., less than .001). Since less than a 1 in 1,000 chance of committing a type 1 error (rejecting the null hypothesis when it is true) is much less than a 5% chance, we can reject the null hypothesis with great confidence. As discussed on page 89, the values in the "z" column (i.e., the t ratios) are the coefficient divided by the standard error (i.e., .056/.010 = 5.21).

Because we are using a multivariate model (we have three theoretically interesting independent variables) we are on much sounder ground in estimating the impact of ideology (liberalism) on the probability a senator will support a more progressive tax structure ("progressive" means that the tax rate an individual pays increases as their income increases). Thus, it could have been that more liberal states would elect more liberal senators but that in voting on tax progressivity senators were actually responding to their constituency and not voting their ideological preference. Our results show that constituency conservatism (support for Bush in the 1988 election) is *insignificantly* (in a statistical sense) related to the log odds of a senator supporting the Levin Amendment. Furthermore, the coefficient for ada is calculated after removing the effects of all other independent variables (which includes constituency liberalism). Hence, the possibility of a spurious result (that constituency liberalism and not the liberalism of the senator is what leads the senator to vote "yes" on the Levin Amendment) is greatly reduced. Thus, just by looking at the sign and statistical significance of each independent variable in the "main equation" we can already tell which of our hypotheses are supported (only ada89 - the liberalism of the senator - has a statistically significant effect).

Standardized Coefficients – The Relative Importance of the Independent Variables

As discussed on pages 127-129, it is often useful to assess the relative importance of the independent variables. For example, it would be useful to say that a senator's ideology (ada89) was *twice* as important as the conservatism of his constituency (Bush88) in determining his vote on the degree of progressivity of the federal tax code (Levin). Because independent variables are measure in different units and have different means and standard deviations, we *cannot* directly compare the impact of the independent variables by comparing the absolute size of their unstandardized coefficients. For example, although the coefficient in the "main equation" is larger for Bush88 than for ada89 (.062 vs. .056), it would be a mistake to concluded that Bush88 had a greater impact than ada89 on senatorial voting on the Levin Amendment. In order to make such a comparison we need to standardize the coefficients.

As discussed on pages 127-129, standardizing a coefficient (i.e., a "b") can be accomplished by taking the unstandardized coefficient and multiplying it by the ratio of the standard deviation of that independent variable to the standard deviation of the dependent variable. The absolute value of standardized coefficients can then be compared to each other. Obviously, to perform the standardization process we need both the unstandardized coefficients (i.e., b^s) and the standard deviation for each variable. The entries in the "Coef." column of the main equation provide the unstandardized coefficients for each independent variable. The results below provide the standard deviation and the mean for each variable.

Variable	Obs	Mean	Std. Dev.	Min	Max
levin	99	.4343434	.498193	0	1
ada89	99	45	33.43437	0	100
bush88	100	54.56	5.457494	44	66
medinc	99	19.46061	2.594296	14.6	28.4

From page 128 we know that the standardization formula is:

Standardized b = [(unstandardized b)] [(st. dev. of X)/(st. dev. of Y)]

Let's take ada89. We know from the main equation that the unstandardized b for ada is .056 (see the results in the coefficient column for the main equation on page 153). We know from the variable means and standard deviations output that the standard deviation of ada is 33.43 (see the above results) and the standard deviation of the dependent variable (levin) is .498. Inserting those numbers in the above formula yields the following:

(.056) (33.43/.498)

= (.056) (67.12)

= 3.758

For Bush88 the computation is:

(.062) (5.457/.498) = (.062) (10.957) = .679

For medinc the computation is:

(-.160) (2.594/.498) = (-.160)(5.208) = -.833

The above computations allow me to say that senator ideology (ada89) is approximately 5.5 times as important as constituency conservatism (3.758/.679 = 5.53) and approximately 4.5 times as important as median family income (3.758/.833 = 4.51 – remember absolute values are all that matters for standardized coefficients – i.e., we can work with .833 not -.833) in explaining senators' votes on the Levin Amendment. Finally, constituency conservatism (i.e., Bush 88) needs to be compared to median family income. Since the standardized coefficient for median family income has a greater absolute value than the standardized coefficient for constituency conservatism (.833 vs. .679), I will use compare median family income to constituency conservatism (i.e., the calculation will be .833/.679 rather than .679/.833). Median family income has approximately 1.2 times as much impact as constituency conservatism on senators' voting on the Levin Amendment. This sounds better than saying constituency conservatism is approximately 80% as important as median family income in explaining senators' votes on the Levin Amendment (.679/.833 = .81). In your term paper you will need to interpret the standardized coefficients as I did. Notice I referred to each independent variable by *concept* and **not** by an acronym. Thus, use terms such as the liberalism of the senator or senatorial liberalism instead of "ada89." This tells the reader the concept the variable is measuring.

If you have more independent variables in your term paper than I do in this example, you will have more comparisons. Thus, I compared independent variable #1 (senatorial liberalism) with independent variables #2 (constituency conservatism) and #3 (median family income). I then compared independent variable #2 with independent variable #3. You may have a much longer sequence (i.e., #1 with #2, #1 with #3, #1 with #4, #2 with #3, #2 with #4 and #3 with #4 or a still longer sequence).

It is important to grasp the significance of these findings. The results show that the ideology of the senator is much more important than the views of the senator's constituents in determining how the senator will vote. That is an important finding. This is one reason elections matter! The people you elect are not that heavily constrained by public opinion. On most votes (e.g., "Levin") the public either does not have an opinion, or does not communicate it to their congressman or senator or will not cast their vote in the next election on the basis of this vote. To an important extent, an elected official can behave as they want to.

Additionally, these results are similar to the results we have been working with on pages 195-196. Since poor states have a smaller percentage of their population earning over $200,000 per year (the people adversely affected by the Levin Amendment), senators from poorer states should be more likely to favor the Levin Amendment than senators from wealthier states (because the amendment shifts the tax burden more to wealthy states). While the sign on the coefficient is negative (-.160) as expected (i.e., all other factors being equal the *higher* the median income in the senator's state the *less* likely they will vote in favor of the Levin Amendment), the results are not that close to attaining statistical significance. This tells you a lot about how our democracy works. Answering such questions is why political scientists use the methods that you have been learning in this course. If one wanted to know what independent variables most effect how their senator votes, there is simply no viable alternative to doing the type of statistical analysis we just did.

Magnitude Assessments

As I have stressed in class, regression allows us to estimate the magnitude of the impact of each independent variable on the dependent variable *controlling* for the effects of all *other* independent variables. While magnitude assessments are somewhat more difficult in logit than in multiple regression, they are still attainable and useful.

One of the difficulties in assessing the magnitude of logit coefficients is that the impact of each independent variable on the probability of voting "yes" (or "1" in my coding scheme - see page 152) on the Levin Amendment depends on both the level of that independent variable *as well as the level of each other independent variable*. For example, replacing a Democratic senator with an ADA score of 50% (a moderate) with another Democratic senator with an ADA score of 80% (a liberal) would probably result in a relatively small increase in the probability of voting "yes" on the Levin Amendment; whereas replacing a Democratic senator with an ADA score of 20% (a conservative) with a Democratic senator with an ADA score of 50% (a moderate) would most likely result in a much greater increase in the probability of voting "yes" on the Levin Amendment. Even though in each comparison the increase in ADA scores is the same (30%, i.e., 80% - 50% = 30% and 50% - 20% = 30%) the change in the probability of voting "yes" on the Levin Amendment is different. We saw this same phenomenon, non-linearity, on page 93. For the next quiz, it would be "highly" advisable to re-read page 93 and then look at the diagram on page 152 (notice that both the diagrams on the lower half of page 93 and on page 152 are non-linear - i.e., "non-straight" lines as opposed to the "straight" line on the upper half of page 93).

The reason for non-linearity in this example is that both moderate and liberal Democratic senators were highly likely to vote for the Levin Amendment, so replacing a moderate Democratic senator with a liberal Democratic senator will probably not change the probability of voting "yes" nearly as much as changing from a conservative Democratic senator to a moderate Democratic senator. A similar line of reasoning would extend to the level of the other independent variables (e.g., degree of liberalism of the senator's constituency, etc.). Therefore, we can not use the method of interpreting "b" for a linear regression that we learned on pages 109-110 when interpreting "b" in logit. We must enter values for each independent variable (e.g., an ADA score of say 50% and then re-calculating if ADA becomes 80%) when we estimate the probabilities of voting "yes" on the Levin Amendment.

While the values you select for each independent variable can be arbitrary, I would suggest that you experiment (alter the values) of one of the statistically significant independent variables and enter the mean value for each remaining independent variable (unless the mean value is not possible - e.g. a dummy variable where the senator could only score "1" or "0" – a mean of say .54 would not be a possible score in the "real world"). In my opinion an interesting comparison is between Democratic senators scoring approximately 80% on the ADA scale and Republican senators scoring approximately 20% on the ADA scale. My reasoning is that in large urbanized states (e.g., California, New York, etc.) Democratic senators typically score around 80% on liberalism scales (such as ADA) and Republican senators typically score around 20%. Thus, this comparison provides a relatively accurate depiction of how the probability of voting for a progressive tax shift (e.g., the Levin Amendment) changes by replacing a Republican senator in a state such as California with their likely Democratic opponent. Such a comparison helps appraise what is "at stake" in a typical senatorial election in a large urbanized state.

Now, I'll perform a magnitude comparison for a senator whose is relatively liberal (scores 80 on ADA) versus a senator who is relatively conservative (i.e., scores 20 on ADA). Policy-wise, this will give you an idea of what a Democratic vs. Republican Senate contest in California represents. For the liberal senator, take the y intercept value (i.e., _cons in the "main equation" on page 153 - cons doesn't stand for conservatism but rather the "constant" or "y intercept" – it is the lowest entry in the "Coef." column on page 153) and the value in the "Coef." for each independent variable (i.e., the "b" value for that variable). Go three places to the right of the decimal point. When you undertake this type of analysis for your term paper only go further to the right of the decimal point if you need to in order to get a non-zero entry (e.g., if the "b" value was - 0.00001035 record it as -.00001).

I will start with the y intercept (i.e., _cons) and then go to the top of the "Coef." column. Leave space between the y intercept value and each of the "b" values. From the "Coef." column of the "main equation" on page 153 we have:

-3.320 + .056 + .062 -.160

Each of the above numbers comes from the coefficient column of my main equation. You may have more independent variables in your term paper (and hence more coefficient values) than I have. Although none of my "b" values are in scientific notation, you may encounter a coefficient values such as 5.92e-06 in your statistical results. If the value was listed as 5.92e-06 it means to go 6 decimal places to the *left*. Thus, the actual value is .00000592. As hopefully you remember from several discussions in the textbook, in order to get a predicted score on the dependent variable (i.e., the probability the senator will vote "yes" on the Levin Amendment) we need to set the level of the *each* of the independent variables. Thus, we will need to multiply the ada89 coefficient of .056 by a score on ADA. As I mentioned previously, a Democratic senate candidate in California will typically score about 80 on ADA while a Republican will score about 20. Let's work the Democrat first. If we set the score on ada89 at 80, our computation now would read as follows:

-3.320 + (80)(.056) + .062 -.160

Now we need values for Bush88 and Medinc. From the data on the bottom of page 154 we know that the mean score on Bush88 is 54.5. Therefore, in the typical state Bush received 54.5% of the vote in the 1988 presidential election. This mean is different than his percentage of the *total* vote in the U.S. because the computer arrived at this 54.5% figure by adding up all 50 state scores and dividing by 50. Thus, all states counted equally. Since states differ in population size a "mean of the state vote" would *not* likely equal the percentage of the national vote Bush received. The mean score on medinc is 19.460 (i.e., in the average state the median household income was $19,460 in 1980). Plugging these two mean scores into the above data yields the following:

-3.320 + (80)(.056) + (54.5)(.062) + (19.46)(-.160)

Which becomes:

-3.320 + 4.48 + 3.379 - 3.11 = 1.429

Be careful with negative numbers! Both the y intercept (-3.320) and the median family income computation (-3.11) are negative. Thus, we'll subtract them. I would add the positives and subtract the negatives from them. Thus, 4.48 + 3.379 = 7.859 and -3.320 - 3.11 = -6.43. Subtracting the total value of the "negatives" from the total value of the "positives" yields 1.429 (i.e., 7.859- 6.43 = 1.429). Now we need to make this number (i.e., 1.429) the exponent of base "e" (i.e., $2.71828^{1.429}$). Taking a scientific calculator I did the following: (1) entered 1.429; (2) pushed the "INV" key (your calculator may have a "second function" or "shift" key – probably toward the upper left side of the calculator); (3) pushed the "ln" key and received an answer of 4.17. Test yourself. Thus, if you enter 1.429 and push the right keys you should get an answer of 4.17. If so, you know that you are using your calculator correctly. This tells me that 2.71828 to the 1.429 power (i.e., $2.71828^{1.429}$) is 4.17. The final part of the computation is take the number I just calculated (4.17) and divide it by this same number plus 1 (i.e., 4.17 divided by 1 + 4.17 = 4.17/5.17 = 80.6). This tells me that a senator who scored 80 on ADA, came from a state where Bush received 54.5% of the popular vote in 1988 and whose state median family income was $19,460 had approximately an 80.6% probability of voting in favor of the Levin Amendment.

The computations are identical for the conservative senator who scored "20" on ADA except that the number you multiply the ADA coefficient by is now 20 instead of 80. Thus, it would be as follows:

-3.320 + (20)(.056) + (54.5)(.062) + (19.46)(-.160)

-3.320 + 1.12 + 3.379 -3.11

= -1.93 (i.e., a negative exponent – notice it is -1.93, not 1.93)

At this point we have $e^{-1.93}$ (or approximately $2.71828^{-1.93}$). Entering -1.93 in my calculator, pressing "INV" and then pressing "ln" yields a value of .145. As I did above, the computation then becomes this number (i.e., .145) divided by the same number plus "1." Thus, .145/(.145 + 1) = .145/1.145 = .126. Therefore, a senator who has a liberalism score of 20% from a state in which Bush received 54.5% of the 1988 presidential vote and had a state median family income of $19,460 in 1980 had a 12.6% probability of voting "yes" on the Levin Amendment. That's quite a gap (i.e., 80.6% is much higher than 12.6%). Elections matter!

Goodness of Fit: Pseudo R^2 or the Likelihood Ratio Index

It may be important to assess how well our logit model performs. In regression we use R^2 to assess model performance (pp. 82-84). Although R^2 is *not* appropriate to use with logit, Psuedo R^2 (also referred to as the likelihood ratio index) is a somewhat analogous statistic to R^2 (William H. Greene, Econometric Analysis, second edition, pp. 651-653). Pseudo R^2 is calculated as follows:

$$\text{Pseudo } R^2 \text{ (i.e., the Likelihood Ratio Index)} = 1 - (L_1 / L_0)$$

Where: L_1 = Log likelihood from the "main equation" (-44.218 – see page 153) and L_0 = Log likelihood from a logit equation using just the constant (i.e., no independent variables – for my data this is -67.765 – printout not shown).

For my data the Pseudo R^2 would be calculated as follows:

$$\text{Pseudo } R^2 = 1 - (-44.218/-67.765)$$

$$\text{Pseudo } R^2 = 1 - (.652)$$

$$\text{Pseudo } R^2 = .35$$

Now go back to page 153 and look at the Pseudo R^2 value for the "main equation." Isn't it (within rounding of) .35? Yes! That is how it was calculated. The logic of this is similar to R^2 (pp. 82-84). We compare how well we did using our independent variables to predict the dependent variable (i.e., our "main equation") versus how well we would do *without* any independent variables (i.e., predicting the mean score on the dependent variable for all observations – see pp. 82-84). Let me offer the following standard for interpreting Pseudo R^2 values:

0 to .19: the model has a "low" degree of explanatory capability

.20 to .60: the model has a "moderate" degree of explanatory capability

.61 and higher: the model has a "high" degree of explanatory capability

A very famous econometrician I like to consult, William H. Greene, thought the above standard was reasonable, but cautioned *against* making any assessment of explanatory "power" from a Pseudo R^2 statistic. Notice I used the term "capability," not "power" in the table above. I only offer the above standard to provide some assessment of how well a logit model performs. Unlike R^2, Pseudo R^2 does *not* indicate the percentage of variation in the dependent variable explained by all of the independent variables together.

Probability and Odds

The last several pages have shown you how to convert odds in logit into probabilities. It is important to understand the difference between probability and odds. Probability is the likelihood (or chance) that a particular event will occur (e.g., a particular team wins). Odds are the probability that a particular event will occur (e.g., the team wins) divided (i.e., a ratio) by the probability that the particular event will not occur (e.g., the team does not win). I mention this because people often confuse these two concepts. An op-ed piece in the Los Angeles Times (June or July 1998) made the following statement: the odds of African-Americans convicted of murder receiving a death sentence are four times the odds faced by other defendants similarly convicted.

Most people probably would have interpreted the preceding statement to mean that if 99% of African-Americans in such circumstances received a death sentence the probability among non-African-Americans would be around 25% (i.e., 99% is almost four times 25%). The actual comparison is African-Americans 99%, non-African-Americans approximately 96%. If 99% of African-Americans convicted of murder receive a death sentence the odds are 99 to 1 (i.e., the probabilities must sum to 100%; 99% + 1% = 100% and the odds are the probability of occurrence to non-occurrence: thus 99% to 1%). If 96% of non-African-Americans convicted of murder receive a death sentence the odds are 96 to 4 (obviously 96% + 4% = 100%), or 24 to 1. Since 99 to 1 is slightly over 4 times 24 to 1 (i.e., 99 is over 4 times 24) the odds are over 4 times as great. However, the difference in the probability of execution is much closer: 99% vs. 96%. There is still discrimination, but not nearly as much as one might think if they did not know the difference between odds and probability.

Probit

On page 151 I mentioned that probit (from the term "probability unit") is a similar statistical technique to logit. Like logit, probit is often used by political scientists when the dependent variable has only two categories of responses. In order to show the similarity between the results of probit and logit I have re-estimated the "main equation" from page 153 in probit (see below). Notice that the Pseudo R^2 value (.3476) in the probit equation below is virtually identical to the logit estimate on page 153. Additionally, note that the signs (i.e., positive or negative) on all the coefficients are the same as on page 153. Furthermore, notice that the level of statistical significance of each of the independent variables in the probit equation is very similar to the results from the logit equation on page 153.

Main Equation Estimated in Probit

```
Probit regression                                 Number of obs   =         99
                                                  LR chi2(3)      =      47.11
                                                  Prob > chi2     =     0.0000
Log likelihood =  -44.21223                       Pseudo R2       =     0.3476
```

levin	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ada89	.0330582	.0056853	5.81	0.000	.0219152	.0442012
bush88	.0328563	.0330742	0.99	0.321	-.031968	.0976806
medinc	-.0882789	.0680363	-1.30	0.194	-.2216276	.0450698
_cons	-1.836464	2.507755	-0.73	0.464	-6.751573	3.078644

Logit coefficients are typically about 1.6-1.7 times the size of probit coefficients. Notice that the coefficient value for ada89 in the logit equation on page 153 is .056 whereas the corresponding probit value is .033 (.056/.033 = 1.70). As with logit, we can undertake a magnitude assessment in probit. Using the same values for the independent variables as were used on pp. 156-159 (i.e., a senator who scored 80% on ada89, etc.) and multiplying them by the appropriate coefficient values from the probit equation yields the following:

-1.836 + (80)(.033) + (54.5)(.032) + (19.46)(-.088)

= -1.836 + 2.64 + 1.74 - 1.712

= .832

Instead of being the exponent of base "e," .832 is a z score. Since the z score is positive we know given these variable values (e.g., 80 on ada89, etc.) the senator has greater than a 50% probability of voting "yes" on the Levin Amendment (if the z score value had been 0 this would have corresponded to the mid-point on a normal curve – see page 24 – which would indicate a 50% probability of voting "yes" with negative z scores indicating less than a 50% probability and positive z scores indicating greater than a 50% of voting "yes" – see pages 25-27).

Consulting a z score table (i.e., area under a normal curve), I find that a z score of .83 (close to our score of .832) means that .2967 of the area under a normal curve (i.e., 29.6%) lies between a z score (i.e., probit score) of 0 and 1 standard deviation above the mean. Since our probit score is positive, in order to obtain the probability of a senator voting "yes" on the Levin Amendment I need to add .50 (the area at and below a z score of 0) to the area above 0 (i.e., .2967). Doing so yields .7967 (.50 + .2967 = .7967). This means that under these conditions (i.e., scores on the independent variables) a senator has approximately an 80% probability (.796) of voting "yes" on the Levin Amendment. With logit the corresponding result was 80.6% (see page 159). That's pretty close! Incidently, the reason you will be using logit instead of probit in your term paper is that cost of this packet would have been higher with probit because you would have had to pay a royalty fee for access to a z score table.

Ordered and Multinomial Logit/Probit Models

Both logit and probit are extremely flexible statistical techniques. As we have just seen, both logit and probit are useful when the dependent variable has only two categories of responses (e.g., the Levin Amendment where a senator could either vote "yes" or "no"). However, there are many instances when a political scientist will encounter a nominal or ordinal (see pages 9-10) dependent variable with three or more categories of responses. For example, a scholar in comparative politics might be studying health care coverage in Western European nations. In Western European nations (e.g., Great Britain, France, etc.), unlike the United States, the government typically provides medical coverage to all citizens as a matter of right. However, the degree of coverage may vary over time within the same nation as well as between nations.

A comparative politics scholar may use a group of political and economic variables (e.g., degree of democracy, average annual percentage of growth in the economy over the past five years, etc.) as independent variables to explain whether a nation increased the scope of its medical coverage (i.e., increased the number of different diseases/procedures the government would cover), kept the scope the same (i.e., did not change which diseases/procedures were and were not covered), or reduced the scope (i.e., eliminated coverage of some diseases/procedures that were previously covered). In this case, the dependent variable is "ordinal" because there is a continuum (most to least coverage) and there is not an equal distance between the categories (i.e., the difference between increasing coverage and keeping it the same is probably not equal to the difference between keeping it the same and reducing it – see pages 9-10).

If we have an ordinal level dependent variable with more than two categories of responses the choice between ordinal logit and probit is arbitrary: either is fine. However, if the dependent variable is nominal (i.e., does *not* have a natural continuum – hence lacks the ability to be "ordered" – see page 9) and has *more* than two categories of responses, the choice between what is termed multinomial logit and multinomial probit is *not* arbitrary. Each technique is appropriate in some circumstances but not others. For example, if one is trying to explain voter support for Clinton, Dole and Perot in the 1996 presidential election, the dependent variable (who the individual voted for) can probably not be ordered. Thus, should a continuum from most to least liberal be Clinton, Dole, Perot or Clinton, Perot, Dole? We can't be sure.

The choice between multinomial logit and probit boils down to this: if the removal of one of the categories of responses on the dependent variable does ***not*** effect the ratio of the probability of each of the *other* options occurring, use multinomial logit, if this ***does not*** hold (i.e., removing one option ***does*** effect the ratio of the probabilities of selecting the other options - which is typically the case), use multinomial probit (just keep reading). For example, in explaining individuals' votes in the 1996 presidential election, removing Perot would probably change the ratio of the probability that someone would vote for Clinton as opposed to Dole. Let us say that with Perot in the race the ratio of the probability a particular voter will vote for Clinton instead of Dole is 1.5 (e.g., a 45% probability of voting for Clinton and a 30% probability of voting for Dole is a ratio of 1.5 – thus 45% / 30% = 1.5). However, if Perot drops out of the race, let us say that the ratio for this same voter increases to 2.0 (e.g., a 66% probability of voting for Clinton vs. a 33% probability of voting for Dole - 66% / 33% = 2.0). As in this example, if the ratio would likely change (which is typically the case) use multinomial probit.

Hazard Models

Occasionally a political scientist encounters a dependent variable which can only assume two different values but the data are collected over time (just keep reading). For example, in recent years there have been a number of studies by scholars in comparative politics that test theories about independent variables that influence how long a government in a parliamentary system lasts. Remember that many foreign countries (e.g., Great Britain) have an election if the ruling political party or ruling coalition (if no party has a majority of the seats in the legislature) does not prevail on a vote in the national legislature. The data for such a study are likely to be taken monthly and be organized as follows: the dependent variable is coded "0" if in that month no election was called and "1" if an election was called (probably because the ruling government or coalition did not prevail on a vote in the national legislature); some of the independent variables (e.g., the number of political parties) are likely to be the same for all months of the study (e.g., no new parties may have emerged or none dissolved in a particular nation over the ten years of the study) while some of the independent variables might change every month (e.g., the unemployment rate). In such circumstances a political scientist uses what is termed a "hazard" model (just keep reading).

The logic of a "hazard" model in this example is that every month a government in a parliamentary system has not "failed" on a vote in the national legislature, it is "at risk" to "fail" (e.g., if a government has not "failed" by the 10th month of its existence, it is then "at risk" of "failing" in the 11th month). Thus, the question is: what explains why some governments in parliamentary systems take more months before they "fail" on a vote in the national legislature than other governments? Independent variables in such studies often include the following: the number of political parties (the hypothesis being that the greater the number of political parties in the system the shorter the time before the ruling government fails on a vote in the national legislature); the unemployment rate (the hypothesis being that the higher the unemployment rate the shorter the time before the ruling government fails on a vote in the national legislature). There are a number of different methods of estimating the coefficients (i.e., the "b^s") in a "hazard" model.

Event Count Models

Another type of discrete dependent variable model that has become prominent in political science over the decade is the event count model. The dependent variable in an event count model is simply the number of events that occur in a given time period (just keep reading). International relations scholars often test theories with independent variables (e.g., level of democracy, popularity of the ruling regime, percentage of growth in the economy) to explain how many disputes a particular nation will be involved in with other nations each year. Scores on the dependent variable might range from "0" (no international disputes involving the particular nation in a given year) to as many as perhaps "6." The dependent variable in an event count model is a ratio level measure (on ratio level measures see page 11) and hence should be amenable to regression. However, since the dependent variable in an event count model typically has a narrow range of scores (often between "0" and "6"), it is not "continuous" (see the first paragraph on page 150) and hence, regression is *not* appropriate. Like a "hazard" model, the results from an event count model can estimate the score on the dependent variable given scores on each of the independent variables (as we did on pp. 156-159) and the impact of each particular independent variable on the dependent variable (as we did on pp. 109-110). Other recent examples of event count models include presidential scholars building models to explain the annual number of Supreme Court appointments presidents' make.

Potential quiz questions include: (1) Why would a political scientist use logit or probit instead of regression (don't use a diagram)? (2) Why can't logit coefficients be interpreted like regression coefficients (i.e., like "b" on pages 109-110)? (3) What is the difference between odds and probability? (4) With a dichotomous dependent variable how much difference is there between logit and probit results? (5) If the dependent variable has more than two categories of responses, how does this impact the choice of either logit or probit?

Causal Models

Suppose you are a scholar in comparative politics. A topic you might be interested in is the effect of the degree of democracy in less well-developed nations on the growth rate of the economies of less well-developed nations. You might be interested in such a topic because economic growth rates in developing nations have important consequences both for the "developing" as well as the "developed" world. For example, in the "developing" nations, economic growth rates effect the standard of living of the people and such non-economic factors as the level of political violence. For the "developed" nations, the growth rate of the "developing" nations can impact the cost of credit and perhaps even the need to use military force abroad. Therefore, scholars in comparative politics have often been interested in this topic. Let us say that we formulate a model where Y (the dependent variable) is the rate of growth of a nation's gross national product. Let us further say that previous literature and theory suggest three independent variables that we should use: the degree of democracy in the nation (X_1); the domestic economic priorities of the government (X_2); and the level of investment and credit from foreign nations (X_3).

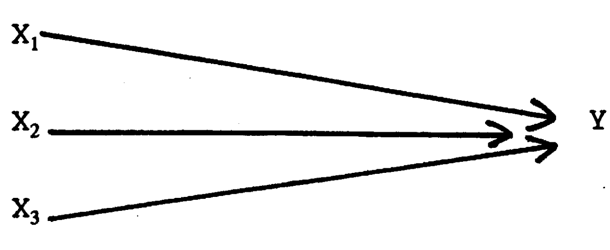
In the comparative political science literature there are well-developed measures of all the variables we will need. Rather than to needlessly detain us with a discussion of how each variable is measured, let us just say that we will use the commonly accepted measure of each variable. How to measure a variable is not the main purpose of this section. The main purpose of this section is to introduce a new type of modeling (which you will see the need for shortly). Let us say we have data for 60 developing nations. Since the dependent variable (Y - the growth rate of the economy) is measured at the interval level (see pages 10-11) and has many possible categories of responses, we will use least squares regression (see page 150). The model that we might estimate could be as follows:

(equation 1) $Y = a_1 + b_1X_1 + b_2X_2 + b_3X_3 + e_1$

Note: I put subscripts on "a" and "e" to differentiate them from other "a's" and "e's" that will be introduced shortly.

The model pictured in Figure 1 below is a diagram of the relationships in equation 1.

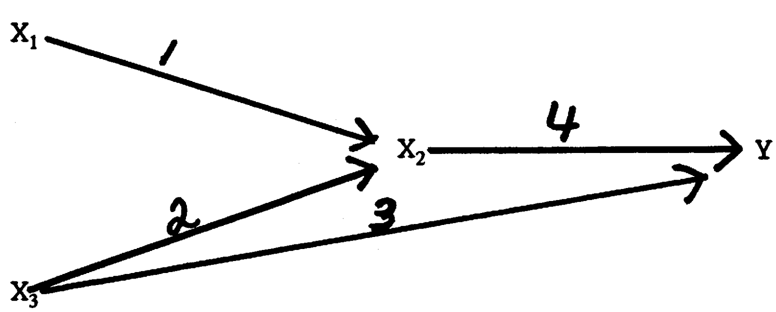
Figure 1



Suppose the results from this model indicate that the independent variable we are most interested in, the degree of democracy in the nation (i.e., X_1), is statistically insignificant. That is, suppose the "t ratio" for "b_1" in equation 1 has an absolute value of less than 2.0. If so, we are pushed to the conclusion that the form of government has no effect on the growth rate of the economy.

The model in equation 1, which is visualized in Figure 1 (page 165), is a model of what are called "direct" effects. That is, each independent variable effects the dependent variable directly. For example, in Figure 1 (page 165) there is an arrow from each independent variable "directly" to the dependent variable. One reason why the level of democracy in less well-developed nations might not be statistically significant in equation 1 is that a model of "direct" effects may not be appropriate. For example, it might be that the more democratic the nation, the more willing the nation is to let its citizens own land. After all, if you think the citizenry is capable of selecting the leaders of the country, you probably also think they are capable of managing property. If so, the degree of democracy impacts the type of domestic economic policies pursued (i.e., landholding policies, perhaps also the supply of money, tax rates, etc.). The type of domestic economic policies pursued effect the growth rate of the economy. In such a situation the impact of the degree of democracy on the growth rate of the economy is through the effect that the degree of democracy has on the type of domestic economic policies pursued. If independent variable X_1 (the degree of democracy) effects independent variable X_2 (domestic economic policies) and X_2 effects the dependent variable (Y - the growth rate of the economy), political scientists say that the effect of X_1 on Y is "indirect" (i.e., through the effect X_1 has on X_2). Since the level of foreign investments (X_3) may directly effect the growth rate of the economy (Y) and also the type of economic policies the government pursues (X_2), foreign investments (X_3) may have both "direct" and "indirect" effects on the rate of economic growth (Y). Figure 2 below shows a model where the degree of democracy (X_1) has only an "indirect" effect on the growth rate of the economy (Y) and where foreign investment and credit (X_3) have both "direct" and "indirect" effects on the growth rate of the economy (Y). Just keep reading! Figure 2 below will become clear on page 167. In political science literature, a model where independent variables have both "direct" and "indirect" effects on the dependent variable is commonly referred to as either a "causal model" or a "recursive model."

Figure 2 - Causal Model



X_1 = degree of democracy of the regime

X_2 = domestic economic policies

X_3 = investments/credits from foreign nations

Y = percent of growth in the gross national product

Figure 2 (page 166) is noticeably different from Figure 1 (page 165). In Figure 1 each independent variable has only "direct" effects on the dependent variable. Put another way, no independent variable in Figure 1 effects the dependent through one of the other independent variables. Figure 2 is different. Look at "paths" 1 and 2 in Figure 2. Paths 1 and 2 indicate that the degree of democracy (X_1) and the level of foreign investment and credits (X_3) each effect the domestic economic policies of the government (X_2). The domestic economic policies of the government, in turn, effect the growth rate of the economy (Y) through path 4.

In diagrams like Figures 1 and 2, the variables with "arrows" (i.e., "paths") pointing toward them are what we have referred to as dependent variables. Since we have "two" dependent variables in Figure 2 (i.e., two variables with "arrows" pointing toward them - X_2 and Y) we will need two equations to estimate the values for the "arrows" in Figure 2. As we can not have more than one dependent variable in an equation, we must have a number of equations equal to the number of dependent variables in the model. In our case, since there are two dependent variables (i.e., two variables with "arrows" pointing toward them, X_2 and Y) we will need two equations to estimate the model posited by Figure 2. Each "arrow" or "path" in Figure 2 is the graphical representation of one of the "b^s" in the two equations that follow.

The results from equation 2 below produce "paths" (i.e., lines) "1" and "2." in Figure 2. Specifically, using least squares regression (on least squares see pages 85-87) whatever the value the computer estimates for b_4 in equation 2 becomes the value for "path 1" in Figure 2. The value of "path 2" in Figure 2 is the estimate of "b_5" in equation 2.

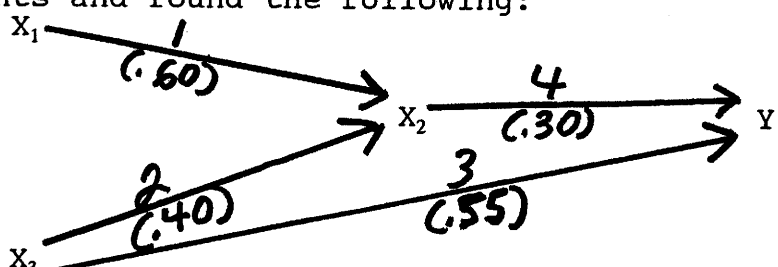
(equation 2) $X_2 = a_2 + b_4X_1 + b_5X_3 + e_2$

Paths 3 and 4 in Figure 2 come from equation 3 below. Specifically, "b_7" in equation 3 is our estimate of path 3 in Figure 2. Path 4 in Figure 2 is the graphical representation of b_6 in equation 3.

(equation 3) $Y = a_3 + b_6X_2 + b_7X_3 + e_3$

When reporting the results from causal or recursive models (such as equations 2 and 3), political scientists often use standardized coefficients (i.e., see pages 127-129)

Suppose that we estimated equations 2 and 3, standardized the coefficients and found the following:



We could say that investments/credits from foreign nations (X_3) have approximately twice the direct effect on growth as domestic economic policies (because .55 is approximately twice the size of .30). Additionally, we could say that the "total" effect (direct plus indirect) of foreign investment/credits is approximately .67 [to estimate indirect effects we multiply paths: (.40)(.30) = .12 and .12 + .55 = .67) which is over three times the effect of democracy (X_1) on growth [(.60)(.30) = .18 and .67 is over three times the size of .18].

A critic might say that since growth (Y) might influence the economic policies the government pursues (X_2) we should have an arrow from Y to X_2. Estimating such a relationship is the purpose of pages 168-178.

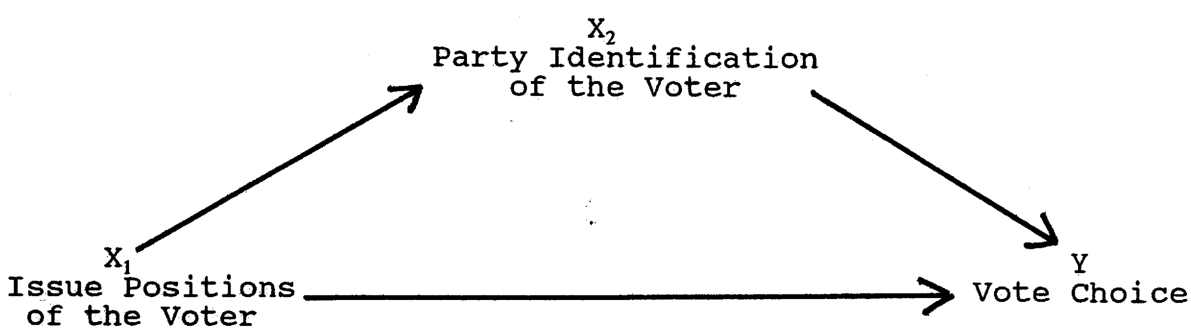
Simultaneous Equation Models

The major purpose of this reading is to show you how to estimate simultaneous equation models. Additionally, I hope the mental process we just went through demonstrates some of the advantages of model building in political science. Such a process forces a political scientist to think through not only which variables should be related to which other variables, but how they are likely to be related. For example, does X_1 effect Y directly, and/or indirectly? Should Y effect X_1? What signs should the various relationships have (i.e., positive or negative)?

One of the prime goals of this course is to heighten your skill in abstracting hypotheses from written materials, conversations, or just your thoughts, and then mentally constructing the hypotheses in the form of arrow diagrams and/or equations. Such processes force the analyst to think through complex realities. This is extremely useful, even in situations where we may not have data on the various variables, and therefore can not rigorously "test" the model. Consciously (or unconsciously) we use models every day. One may choose to attend college and major in a particular subject based upon the likely ramifications (e.g., career path, foregone earnings, etc.). Obviously the accuracy of our model could have important personal repercussions. It would be advantageous to carefully model such a process and critically think through the various linkages.

In discussing how to estimate simultaneous equation models, I will draw on an example from the vast political science literature on electoral behavior. This example was inspired from both Herbert Asher's Causal Modeling, second edition, and William Berry's Nonrecursive Causal Models. Let us say that we wanted to estimate the importance of various factors on a voter's ultimate decision concerning which candidate to support in the general election. As data would be collected on individual voters, the "unit of analysis" would be the individual. An initial model might be formulated as follows:

Figure 3 - Initial Voting Model



Most scholars of electoral behavior would not be satisfied with the causal model posited in Figure 3. Many would argue that two-way relationships (e.g., X_1 influences Y, and Y in turn influences X_1) probably exist between many combinations of the variables in Figure 3. For example, a voter's party identification

(X_2) might influence their position on various issues. Hence we should have an arrow from X_2 to X_1. Likewise, as Asher notes, "one could even envision vote choice (Y) affecting the explanatory variables (X_1 and X_2); for example a vote cast for a Democratic candidate might lead one to conclude that he or she was a Democratic loyalist" (Herbert Asher, Causal Modeling, second edition, p. 68). Thus, a change in Y could produce a change in X_2. Unfortunately, estimating a simultaneous equations model (e.g., where X_1 influences Y and Y in turn influences X_1) is much more difficult than estimating either a single equation model (e.g., Figure 1 - page 165) or a causal (or recursive) model (Figure 2, page 166 and Figure 3, page 168).

In order to estimate a simultaneous equations model, it is first necessary to understand a process called equation "identification." The basic idea of equation "identification" is to see if we can "identify" a unique (i.e., single) value for each of the coefficients (i.e., the "b^s"). In the examples that follow, we will try to find a "unique" value for b_1 (the coefficient of X_1) and b_2 (the coefficient of X_2). Previously, we have written such an equation as: $Y = a + b_1X_1 + b_2X_2 + e$. In order to simplify the discussion, we will omit both "a" (the Y intercept) and "e" (the error term). Hence our "basic" equation will be: $Y = b_1X_1 + b_2X_2$. This form will be used in equations 5-10.

In the discussion that follows, do not be concerned with how we solve for a value for b_1 or b_2. This would needlessly detain us. The critical factor is whether, or not, we can find a unique (i.e., single) value for b_1 and a unique value for b_2. If we can not obtain unique values for b_1 and b_2, we will be unsure of the impact of each of our independent variables on the dependent variable. For example, if X_1 increases by one unit, is the impact on Y (i.e., b_1) .220, .911, -.700, or what? Obviously, we want a unique (i.e., single) value for each of the "b^s." If we do obtain unique values for b_1 and b_2, our equation is said to be "identified." If we can not obtain unique values for b_1 and b_2 our equation is "underidentified." As previously mentioned, our "basic" equation for this discussion (equations 5-10) will be: $Y = b_1X_1 + b_2X_2$. Let us suppose that: $Y = 7$, $X_1 = 2$ and $X_2 = 3$. Substituting these values into our "basic" equation yields equation 5 below:

(equation 5) $7 = b_12 + b_23$

There is not a "unique" solution to equation 5. An infinite number of values for b_1 and b_2 could be consistent with the constraint that the total value of the right-side of the equation equals 7. Thus, b_1 could equal 2 and b_2 could equal 1 {e.g., [7 = (2)(2) + (1)(3)]}. Alternatively, b_1 could equal 5 and b_2 could equal -1 {e.g., [7 = (5)(2) + (-1)(3)]}. This example illustrates a basic principle of equations: in order to obtain unique values for each "unknown" we can not have more "unknowns" than equations. Since we have two unknowns (b_1 and b_2), we will need a two equation system in order to produce unique values for each unknown while solving both equations.

For example, in addition to the information provided by equation 5, suppose we have the following equation:

(equation 6) $-2 = b_11 - b_24$

We can now produce a unique solution for the values of b_1 and b_2 that will solve both equations. These values are $b_1 = 2$ and $b_2 = 1$. Try inserting these values in equations 5 and 6 and solving both equations. As it turns out, $b_1 = 2$ and $b_2 = 1$ are the only values that will solve both equations 5 and 6.

Both equations 5 and 6 are "linear" because no number or letter is taken to a power other than "1" (i.e., no term such as b_1^2 - see pages 93-100). Furthermore, the terms in equation 5 are not multiples of the corresponding terms in equation 6. For example, the first term in equation 5 (i.e., 7) is -3.5 times the first term in equation 6 (i.e., -2). That is, 7/-2 = -3.5. However, the second term in equation 5 ($b_1 2$) is only 2 times the second term in equation 6 ($b_1 1$). (Remember that $b_1 2$ is equal to $2b_1$. Obviously, $2b_1$ is 2 times $1b_1$.) Since -3.5 is not equal to 2, equation 6 is not a multiple of equation 5. As a result, both equations 5 and 6 are said to be "independent" of each other. Furthermore, as no letter or number in either equations 5 or 6 is taken to a power other than "1," both equations are "linear." Therefore, equations 5 and 6 are "linearly independent" of each other.

However, suppose that instead of equation 6 our other equation in this system was as follows:

(equation 7) $14 = b_1 4 + b_2 6$

Equation 7 is of no help in determining unique values for b_1 and b_2. This is because equation 7 is a multiple of equation 5 (each term in equation 7 is twice the corresponding term in equation 5 - e.g., 14 is twice 7 and $b_1 4$ is twice $b_1 2$). Equations 5 and 7 are said to be linearly "dependent." Hence, equation 7 provides no additional "information" to equation 5. Thus, if equation 7 were our only information to supplement equation 5, we would still be left with one equation in two unknowns and hence no unique solution for the values of b_1 and b_2.

To be "identified" the number of "unknowns" in a system of equations must be equal to the number of "linearly independent" equations. Hence, if only equations 5 and 6 constituted our system they would both be "identified." However, if equations 5 and 7 constituted our system neither equation would be "identified." Equivalently, we could say that both equations 5 and 7 were "underidentified."

While "identification" and "underidentification" are two levels of equation identification, there is also a third possibility - "overidentification." Overidentification results when there are more linearly independent equations than unknowns. Suppose we have the following three equations:

(equation 8) $7 = b_1 2 - b_2 1$

(equation 9) $0 = b_1 1 + b_2 3$

(equation 10) $2 = b_1 3 - b_2 2$

As Asher notes,

> If we used equations 8 and 9 to solve, we would get $b_1 = 3$ and $b_2 = -1$. Equations 8 and 10 generate solutions of $b_1 = 12$ and $b_2 = 17$, while equations 9 and 10 yield $b_1 = 6/11$ and $b_2 = -2/11$. While there is only a finite set of solutions here, it still is the case that different pairs of equations give very dissimilar results (i.e., values for b_1 and b_2), ... an unsatisfactory situation for making ... estimates (Herbert Asher, Causal Modeling, second edition, p. 54).

The combination of linearly independent equations and multiple values for the unknowns is called "overidentification." As will be shown later, we can obtain unique estimates of the "b^s" if an equation is either "identified" or "overidentified." If an equation is "underidentified" we either have to modify it in order to achieve identification, or not estimate the results.

Previously in this course we have dealt with single equation (Figure 1, page 165) and causal models (Figure 2, page 166). In each situation we were able to estimate the models without considering the notion of "identification." We were able to avoid the question of "identification" because the assumptions we made insured that the equations would be "identified."

The fewer unknowns we have to estimate, the easier it is to achieve "identification." As we have seen, if all the information we have is that: $7 = b_1 2 + b_2 3$ we can not obtain a unique value for b_1. However, if we could eliminate the term $b_2 3$ (i.e., reduce the number of unknowns) we could then obtain a unique value for b_1 [i.e., 3.5: $7 = (b_1)(2) = (3.5)(2)$]. Consequently, each method of equation identification will attempt to reduce the number of unknowns in the model.

There are two primary methods of achieving equation "identification." The first method is by imposing coefficient restrictions. Remember that in causal models we assume that roughly half of the possible coefficients (i.e., "b^s") are equal to zero. For example, in Figure 3 while partisan attachment (X_2) is hypothesized to influence vote choice (Y), we explicitly set the reverse relationship (coefficient) equal to zero. Thus, there was not an arrow from Y to X_2 because we assumed that such an arrow would represent a coefficient (a "b" representing the path from Y to X_2) with a value of zero (hence no relationship). Such a path is not an unknown, since it is known to have a value of .00. Therefore, if we can exclude a possible path (i.e., an arrow or a "b") we will reduce the number of unknowns. There is simply no reason to include paths (arrows) between unrelated variables.

In causal models (Figure 2, page 166 and Figure 3, page 168) we also utilize the second method of identification: covariance restrictions. Remember that the error term (the "residual" - see page 79) includes the impact of independent variables that are omitted from the equation. To estimate the values for the arrows (the coefficients, i.e., the "b^s") in Figure 3 (page 168) we need to estimate two equations. This is because we have two variables with arrows pointing toward them (X_2 and Y). In both Figure 2 (page 166) and Figure 3 (page 168) we have assumed that the error terms in each equation are uncorrelated (i.e., have a correlation of .00; on correlation see page 75) with the error terms in all other equations. This is called the covariance restriction. The covariances of the error terms are thus known to have a value of .00 (i.e., "restricted" to .00). Thus, if we can make the assumption of uncorrelated errors we will have reduced the number of unknowns.

Now let us write the two equations we need to estimate in order to obtain the paths (the coefficients, i.e., the "b^s") depicted in Figure 3 on page 168.

(equation 11) $X_2 = a_1 + b_1 X_1 + e_1$

(equation 12) $Y = a_2 + b_2 X_1 + b_3 X_2 + e_2$

Equation 11 says that the partisan identification of a voter (X_2) is entirely a function of that voter's position on issues (X_1) and an error term (e_1). Equation 12 says that an individual's vote choice (Y) is entirely the function of the voter's position on issues (X_1), the voter's partisan attachment (X_2) and an error term (e_2).

The covariance (or correlation) restriction of zero between values for e_1 and e_2 from equations 11 and 12 is unlikely to be met. The reason for this is that we are assuming that the variables omitted from one equation are unrelated to the variables omitted from the other equation. Think of this in the context of equations 11 and 12. Could we plausibly say that all variables omitted from our explanation of X_2 (the voter's party identification) are entirely unrelated to the variables we have omitted from our explanation of the vote choice (Y)? I doubt it. For example, the partisan direction of previous votes an individual cast are likely to influence both their current party identification (X_2) and their vote choice in the current election (Y). If so (and we have omitted partisan voting history from the model) partisan voting history is now part of both e_1 and e_2. If a variable is part of both error terms then the error terms must be correlated. Thus, the covariance restriction of zero would be violated.

Now it should be clear why causal models (e.g., Figure 3, page 168) are sometimes implausible. The assumptions they make in order to achieve identification can be unrealistic. The covariance restriction, as we just discussed, may be untenable for some models. The only other method of achieving identification is the coefficient restriction (restricting some possible b^s to zero). While this may also be untenable in some instances, typically, it gives us our best chance of achieving equation identification. If equation identification is either impossible or implausible for a multi-equation system, we would switch to a single equation model.

The key to equation "identification" lies in the plausibility of using the coefficient restriction. If we exclude a coefficient from an equation, we have "restricted" its value to .00. Thus, the coefficient is "known" to have a value of .00. Therefore, by excluding a coefficient (i.e., a "b") we will have reduced the number of "unknowns." The "order condition" for equation identification says that "... if we have a model consisting of "k" linear equations, then for any equation in that model to be "identified," it must exclude at least "k-1" of the variables that appear in the model" (Herbert Asher, Causal Modeling, second edition, p. 56).

Equation "identification" is done on the basis of a particular equation. Therefore, in a system of equations we may find that some equations are "identified" while other equations are "underidentified." For our purposes the difference between "identification" and "overidentification" is not critical. What is important is simply whether the equation is "identified" or not. If the equation is either "identified" or "overidentified" we have sufficient information to estimate the "b^s." If the equation is "underidentified" then we must either modify it or not try to estimate it.

Technically, the "order condition" is a necessary, but not sufficient condition for identification. Thus, it is possible that an equation could meet the "order condition" but not be identified. For example, if according to the "order condition," two equations are "identified," but all of the variables in one equation are contained in the other equation, then the "order condition" would lead us to the incorrect conclusion that both equations were identified, when in fact they are not (see William D. Berry, Nonrecursive Causal Models, p. 56). The "rank condition," which is both necessary and sufficient in all circumstances, requires a knowledge of matrix algebra and is therefore beyond the scope of the present effort. However, as the above limitation of the "order condition" is not generally encountered in most practical research situations, the "order condition" is usually both necessary and sufficient for identification (Herbert Asher, Causal Modeling, second edition, p. 56).

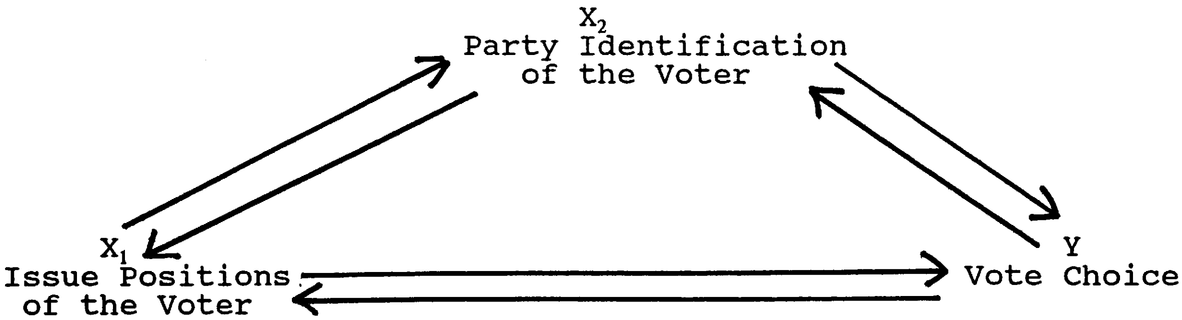
Our ability to utilize the above discussion on identification will be enhanced if we can distinguish between two types of variables. Each variable that has an arrow pointing toward it (thus an arrow going from some other variable toward the variable in question) is called an "endogenous" variable. In Figure 3 (page 168), for example, X_2 and Y are both "endogenous" variables. Endogenous variables are completely "determined" within the system of equations. The values (scores) on the endogenous variables are "determined" by (or "depend" upon) the values (scores) on the independent variables. An endogenous variable is what we have previously referred to as the dependent variable. However, when you have more than one equation (such as a "system" i.e., multi-equation model) there is more than one dependent variable. There are several equations, each with a different dependent variable. Keep in mind, however, that there is only one dependent variable in each equation.

What is potentially confusing with multi-equation models is that the role of a particular variable changes by equation. Thus, the same variable may be either an explanatory (i.e., independent) variable in one equation in the system and then the dependent variable in another equation. Look again at equations 11 and 12 (page 171) and notice how the role of X_2 varies with the particular equation. In equation 11, X_2 is the dependent variable whereas in equation 12, X_2 is one of the independent variables. The number of endogenous variables equals the number of equations in the system (one equation for each endogenous variable). Variables that only appear as independent variables (e.g., X_1 in equations 11 and 12, page 171) are called "exogenous" variables. The values for exogenous variables are completely determined outside the system of equations. Thus, exogenous variables never appear as the dependent variable in any of the equations in the system.

Now let us apply what we have discussed to the electoral behavior model in Figure 3 (page 168). As currently composed, Figure 3 is a multi-equation model because data collected on the same observations (i.e., the same people since the "unit of analysis" is an individual of voting age) is used to construct scores on a group of variables that appear in a system of equations in which at least one of the variables is used as an independent variable in one equation and the dependent variable in another equation.

The multi-equation system depicted in Figure 3 is recursive (or causal) because their is no simultaneous causation between any pair of variables. Thus, for example, if X_1 is a cause of Y then Y can not be a cause of X_1. As discussed previously, this last requirement is probably untenable in the overwhelming majority of research situations. Hence, it is generally more realistic to think that many pairs of variables are mutually causal (i.e., X_1 may influence Y and Y may in turn influence X_1). So, in order to make the model in Figure 3 more realistic, let us add a reciprocal (two-way) linkage between each pair of variables. Figure 4 (next page) shows the revised voting model. Following the model, I show the equations that we must estimate in order to ascertain coefficient values for the linkages in Figure 4.

Figure 4 - Revised Voting Model



(equation 13) $X_1 = a_1 + b_1X_2 + b_2Y + e_1$

(equation 14) $X_2 = a_2 + b_3X_1 + b_4Y + e_2$

(equation 15) $Y = a_3 + b_5X_1 + b_6X_2 + e_3$

Since each endogenous variable appears in each equation, all equations in the system are simultaneous. Remember that according to the order condition ".... for any equation in the model to be identified, it must exclude at least k-1 of the variables that appear in the model" (Herbert Asher, Causal Modeling, second edition, p. 56). As there are three equations, "k" equals three. Therefore, to be "identified" according to the order condition an equation must exclude at least two variables (3 - 1 = 2). A quick glance at the above equations indicates that none of them are identified. Equations 13, 14 and 15 each contain all three of the variables in the model and hence exclude no variables. Therefore, as Figure 4 now stands, we can not estimate any of the equations. Now you can see how greater realism also makes our task more difficult.

So, what do we do? To remedy the situation, we need to find variables that would logically be included in some equations but excluded from other equations. For example, if we can find at least two variables that would logically be independent variables in equations 14 and/or 15, but would not logically be independent variables in equation 13, equation 13 would become "identified."

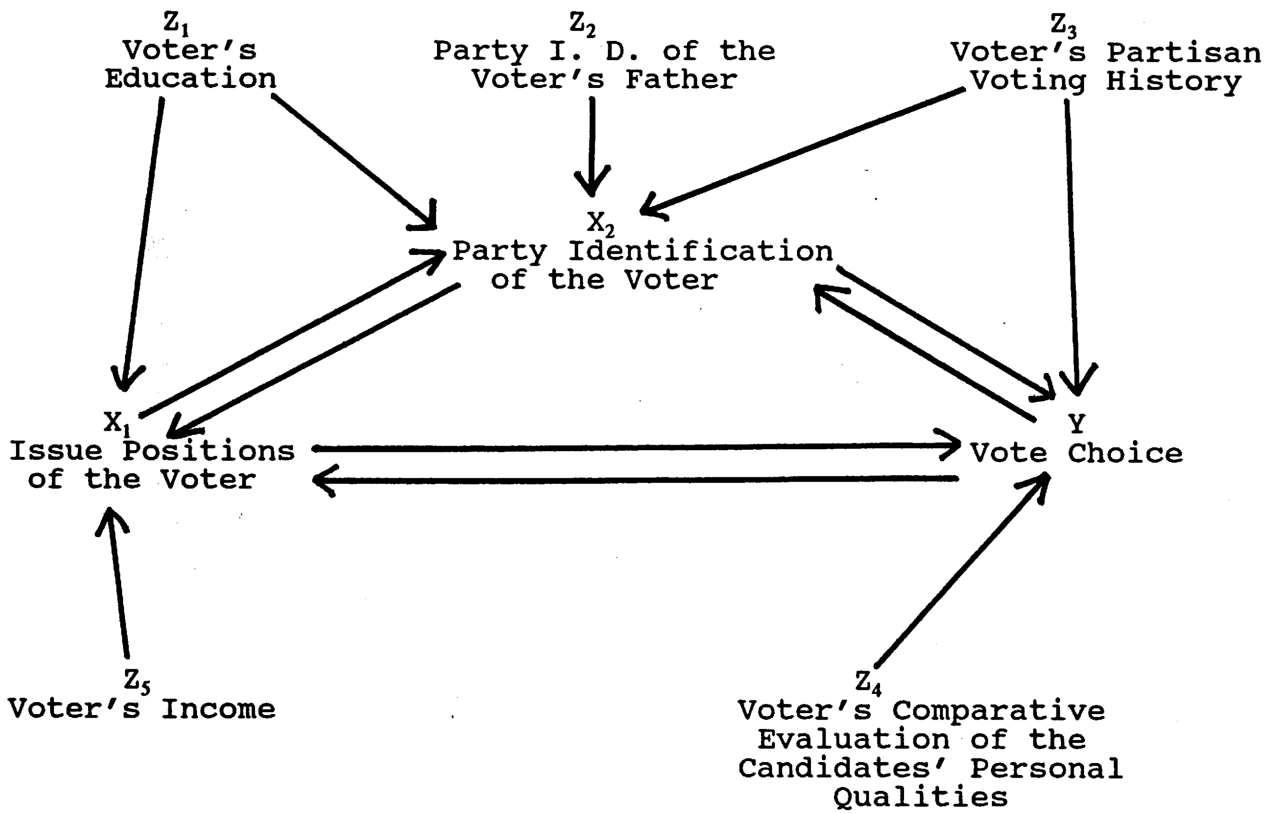
Something that should be clear is that the analyst must carefully think through all the equations in the model. Furthermore, this process must precede data collection. It is of little value to think of a variable that could aid identification after the data have been collected. For example, in the situation we are confronting, it would be impossible to re-survey the same individuals in order to collect data on additional variables. We must think through the identification of specific equations prior to collecting the data.

Unfortunately, one of the most famous studies of the impact of constituency opinions on the behavior of congressmen is a classic example of not thinking through what exogenous variables would be needed in advance of collecting the data. Among other relationships, the authors wanted to see how the representative's perception of constituency attitudes affected the representative's opinion and vice versa (i.e., a simultaneous relationship).

While the study is a seminal work in the field, the failure to collect data on the exogenous variables necessary to "identify," and, hence, estimate the necessary simultaneous equations model has limited the value of the study (see Herbert Asher, Casual Modeling, second edition, page 60). Simultaneous equation models were neither widely known, nor widely used, in political science until the mid-1970s. Relative to many other social science disciplines, political science has been in the forefront of designing new and innovative data sets. Unfortunately, until the mid-to-late 1970s, the level of statistical analysis in political science was noticeably less sophisticated than in some other social science disciplines (see Gary King, Unifying Political Methodology).

In order to show how equations 13-15 could be identified, let me present a "final" version of the voting model. This model is adapted from William D. Berry, Nonrecursive Causal Models, pp. 48; 58-59.

Figure 5 - Final Voting Model



(equation 16) $X_1 = a_1 + b_1X_2 + b_2Y + b_3Z_1 + b_4Z_5 + e_1$

(equation 17) $X_2 = a_2 + b_5X_1 + b_6Y + b_7Z_1 + b_8Z_2 + b_9Z_3 + e_2$

(equation 18) $Y = a_3 + b_{10}X_1 + b_{11}X_2 + b_{12}Z_3 + b_{13}Z_4 + e_3$

The first step in estimating equations 16-18 is to see if each equation is either "identified," or "overidentified." Remember that to be "identified" according to the order condition, each equation must exclude k -1 variables. As "k" equals three (there are three equations), each equation must exclude two variables (3 - 1 = 2). By adding five new exogenous variables (Z_1 through Z_5) to our previous voting model each equation is now either "identified" or "overidentified." For example, equations 16 and 18 both exclude three variables (equation 16 excludes Z_2, Z_3 and Z_4 while equation 18 excludes Z_1, Z_2 and Z_5). Thus equations 16 and 18 are "overidentified." As equation 17 excludes two variables (Z_4 and Z_5), it is "identified." As all three equations are either "identified" or "overidentified" we can estimate the entire three equation model shown in Figure 5.

While all three equations represented by the model in Figure 5 are "identified," it is more important to stress that the model seems plausible from a theoretical perspective. The variables labeled as "exogenous" (i.e., the Z') would not seem to be "caused" by the "endogenous" variables (i.e., X_1, X_2 or Y). For example, it is difficult to believe that because a voter changed their position on abortion that this would encourage them to seek additional education. Hence, it makes sense that there is not an arrow from X_1 to Z_1. Additionally, it seems plausible that not all exogenous variables are related to all endogenous variables. For example, an increase in a voter's income may make them more likely to vote Republican, but this probably occurs because the increased income leads them to change their position on some issues (e.g., the need for government social welfare programs or progressive taxation). If so, the impact of income on the vote would be through issue positions (i.e., indirect, from Z_5 to Y through X_1 rather than from Z_5 to Y directly).

Unfortunately, we can not simply execute regression on equations 16-18. The reason for this has to do with the error term. Remember that we previously discussed the possible correlation of error terms from different equations (e.g., e_2 possibly being correlated with e_3). We overcame this problem by no longer assuming that the error terms from pairs of equations in the system were uncorrelated. As we have just seen, dropping the assumption of uncorrelated error terms made it much more difficult for our equations to be identified.

Now our problem is that an independent variable could be correlated with the error term in the same equation. For example, in equation 16, X_1 is in part a function of Y (i.e., Y is one of the independent variables in equation 16). Additionally, in equation 18, Y is in part a function of X_1 (i.e., X_1 is one of the independent variables in equation 18). The problem is that since e_1 in equation 16 influences X_1 (remember X_1 is the dependent variable in equation 16) and X_1 influences Y (X_1 is an independent variable in equation 18), then Y is correlated with e_1 in equation 16. Think of it this way, if e_1 increases, then X_1 increases. If X_1 increases, then Y increases. Therefore, Y must be correlated with e_1 in equation 16. This is a violation of a basic assumption of regression.

The problem posed by the correlation of Y and e_1 is that the estimate of the coefficient of Y in equation 16 (b_2) is likely to be either too high or too low (i.e., "biased"). The reason for this is that e_1 (the error term in equation 16) contains the impact of omitted independent variables (i.e., any independent variables that should have been in equation 16 but were not included - see page 79). If Y is correlated with e_1, then Y is also correlated with all those omitted independent variables that makeup e_1. Both the direction and strength of the correlation between Y and e_1 will determine the direction and the amount of "bias" in the estimate of b_2.

Fortunately this problem can be solved in a rather straightforward manner. What we need is a variable that has the same magnitude of relationship with X_1 that Y has and is uncorrelated with e_1. Political scientists solve this problem by finding an "instrumental" variable to replace Y in equation 16. In most instances they create a new variable that fulfills the aforementioned criteria. Obvious candidates to serve as "instruments" (i.e., replace "Y" in equation 16) are the exogenous variables (the Z^s). The exogenous variables are attractive "instruments" to use for three reasons: first, according to the structure of the model, they are related to the endogenous variables (in this case variable "Y"); second, as they are entirely determined "outside" the model (i.e., they never appear as the dependent variable in any equation) they should not be correlated with the error terms in any of the equations; finally, we have data (scores) on the exogenous variables.

By far the most common process of creating the instrumental variable(s) and estimating either "identified" or "overidentified" equations is through the use of two-stage least squares regression. Stage one is to obtain the necessary instrumental variables. This is accomplished by regressing each endogenous variable on all exogenous variables in the system of equations. In our situation this would mean estimating the following three equations:

(equation 19) $\hat{X}_1 = c_1 + d_1Z_1 + d_2Z_2 + d_3Z_3 + d_4Z_4 + d_5Z_5$

(equation 20) $\hat{X}_2 = c_2 + f_1Z_1 + f_2Z_2 + f_3Z_3 + f_4Z_4 + f_5Z_5$

(equation 21) $\hat{Y} = c_3 + g_1Z_1 + g_2Z_2 + g_3Z_3 + g_4Z_4 + g_5Z_5$

Note: To avoid confusing equations 19-21 with equations 16-18 I have replaced "a" with "c" and "b" with "d," "f," and "g." Moreover, note there are not error terms on the right-hand side of equations 19-21 because we are only interested in obtaining predicted values for X_1, X_2 and Y. Notice the "hats" (the symbol for predicted values) over each variable to the left of the equal sign. Remember that the value of the error term (for each observation is the difference between the actual value and the predicted value (see "e" on page 80). Thus, you estimate the predicted value of a variable without an error term (see page 78).

Equations 19-21 will produce a predicted value for each observation on each endogenous variable (X_1, X_2 and Y are the endogenous variables). This is the first stage in two-stage least squares regression.

The second stage of two-stage least squares regression is to insert the predicted values of each of the endogenous variables in place of the actual values of each of the endogenous variables when that particular endogenous variable is used as an independent variable. For example, to estimate equation 16, we would substitute the predicted values for X_2 obtained from equation 20 for the actual values of X_2. Furthermore, we would substitute the predicted values of Y from equation 21 for the actual values of Y. Thus, the second stage of two-stage least squares regression for equation 16 would be as follows:

(equation 16) $X_1 = a_1 + b_1\hat{X}_2 + b_2\hat{Y} + b_3Z_1 + b_4Z_5 + e_1$

Similarly, the following two equations would yield estimates of all the remaining linkages in our "final" voting model:

(equation 17) $X_2 = a_2 + b_5\hat{X}_1 + b_6\hat{Y} + b_7Z_1 + b_8Z_2 + b_9Z_3 + e_2$

(equation 18) $Y = a_3 + b_{10}\hat{X}_1 + b_{11}\hat{X}_2 + b_{12}Z_3 + b_{13}Z_4 + e_3$

Two conclusions are warranted from looking at the immediately preceding version of equations 16-18. First, when an endogenous variable appears as an independent variable we use the predicted values for that variable obtained from equations 19-21. For example, notice the "hats" (symbol for "predicted" values) above X_2 and Y in equation 16 on page 177. Secondly, when an endogenous variable appears as the dependent variable, we use the actual values for that variable. Hence, X_1 in equation 16 on page 177 does not have a "hat" above it.

Simultaneous equation systems are a very significant advance in modern political science. Scholars in electoral behavior have found that the political party affiliation of the voter is less important than previously thought (see for example, Morris P. Fiorina, Retrospective Voting in American National Elections, pp. 155-211). For many years electoral scholars assumed that a voter's sense of party identification was an "independent" variable that affected their issue positions, but not vice versa. Simultaneous equation models revealed that party identification both influenced, and was influenced by, the issues positions of the voter.

While simultaneous equations are an important advance in modern quantitative political science, sometimes the method of achieving equation identification by using exogenous variables (as I did by using the Z' on page 175) is not viable. Put another way: one could argue that in certain models, all variables are endogenous (just keep reading). Applying this to Figure 5 on page 175, each of the Z' would have arrows pointing toward them (e.g., a path from X_1 to Z_1 as well as from Z_1 to X_1).

Recently such models have been introduced into political science. The most frequent use of such models is in international relations. For example, Kinsella (American Journal of Political Science, August, 1994, pp. 557-581) explains the level of conflict between the Arab states and Israel in 1985 (the dependent variable) by three independent variables: the level of conflict between the Arab states and Israel over the immediately preceding three years (i.e., 1982-84), the level of Russian arms shipments to the Arab states and Israel over the immediately preceding three years (again, 1982-84) and the level of U.S. arms shipments to the Arab states and Israel over the immediately preceding three years (again, 1982-84). In a second equation, he explains the level of Russian arms shipments to the Arab states and Israel in 1985 with the same three independent variables. In a third equation, he explains U.S. arms shipments to the Arab states and Israel in 1985 with the same three independent variables. Thus, we have three equations with exactly the same three independent variables. Furthermore, each equation contains all three variables in the system. While the mathematics are complicated, it boils down to this: because we are using prior values of the independent variables, we can estimate each of the three equations mentioned above. When a past value of the dependent variable is used as an independent variable, it is referred to as an "autoregressive" variable. Because it uses autoregressive variables, the above model is called a "vector autoregressive model" (or "VAR").

Potential quiz questions include: (1) Why would a political scientist use a causal model (answer in words, not diagrams)? (2) Write the equations necessary to estimate the causal model in Figure 2, page 166 (equations are on page 167 - notice which "b" in equations 2 and 3 on page 167 corresponds to which arrow/path in Figure 2 on page 166). (3) What limitation of causal models may require us to use a simultaneous equations model (answer in words)? (4) What does it mean if an equation is "identified"? (5) Why would a political scientist use two-stage least squares regression?