Interaction Terms

Thus far we have assumed that each independent variable has a separate impact on the dependent variable. For example, let us look at the following equation:

(equation 1) $Y = a + b_1X_1 + b_2X_2 + b_3X_3 + e$

You may find the next sentence confusing. You know the drill!! Just keep reading!! An important assumption that equation 1 above makes is that the impact of each independent variable (i.e., of X_1, X_2 and X_3) on the dependent variable (i.e., Y) is unaffected by the level of one of the other independent variables. In the current example, this would mean that the impact of X_1 on Y (i.e., the value of b_1) does not dependent upon the level (i.e., score) of X_2.

Let us say that we are studying the level of welfare payments in the 50 states. We will build a model that attempts to explain why some states have higher welfare payments than other states. We might have a model where the variables are as follows: Y would be the monthly dollar amount of welfare benefits that a family of four persons would receive. Our first independent variable, X_1, might be the degree of liberalism of the particular state's government. We could measure the liberalism of the state's government by several different means. One approach is to treat the legislature and the governor equally, equate the Democratic Party as being more liberal than the Republican Party and take the percentage of seats Democrats control in the state legislature and the governorship. Do not worry about the mathematics of this, but just to show you what I mean, a state with a Democratic governor and Democrats controlling 50% of the seats in the legislature might receive a liberalism score of 75%. If the governor were a Republican and the Democrats controlled 50% of the seats in the state legislature, the state might receive a liberalism score of 25%. I would assume the hypothesis would be that the more liberal the state government, the higher the welfare payment level. A second independent variable, X_2, might be the percentage of those actually voting (not just registered to vote) who are in poverty. The logic would be that the greater the percentage of the voters who would gain from higher welfare payments, the higher the welfare payments will be. A third independent variable, X_3, might be state median family income. The logic would be that the wealthier the state, the more money it could afford to pay each welfare recipient.

Equation 1 above depicts the model that I discussed in the preceding paragraph. Each independent variable is suppose to have a separate effect on the dependent variable. However, upon additional reflection we might think that the impact of the percent of the voters in poverty (X_2) on the welfare payment level (Y) might change depending upon how liberal the state's government is (X_1). If so, we will need to ask the computer to multiply the score on X_1 times the score on X_2 to capture such an effect. Such a term (i.e., adding a term such as $b_4X_1X_2$ to equation 1 above) is called an interaction term.

Let me now outline a theory as to why the impact of the percent of the voters in poverty on the welfare payment levels might be expected to change as the liberalism of the state's government changes. Remember, only use an interaction term if you have a strong theoretical reason to do so. I will try to present such reasoning on the next page.

To see why the American political system might work this way, consider the following logic. Poor people are not an important part of the Republican party voting coalition. If we elect Republicans, why are they going to try to raise welfare benefits and help what is basically a Democratic group of voters when doing so would likely alienate a large group of reliably Republican voters? However, if you were a Democratic governor, you would probably be more concerned than a Republican governor about the plight of the poor because the poor are more likely to be a significant part of the coalition that elected you. However, since raising welfare benefits means alienating part of the rest of your supporters (i.e., Democrats who want to keep taxes low) you might be reluctant to try to increase welfare payments very much. On the other hand, the larger the percentage of a Democratic governor's electoral coalition that benefits from higher welfare payments, the greater the electoral incentive for a Democratic governor to increase welfare payments.

If the above reasoning is correct (and it is certainly plausible), then the impact of both the liberalism of a state's government (X_1) and the percentage of the voting electorate who are in poverty (X_2) on the level of welfare payments (the dependent variable, i.e., "Y") would depend upon the level of the other one. Thus, the impact of the percentage of a state's voters who are in poverty (X_2) on the level of welfare payments (Y) would depend upon the liberalism of a state's government (X_1). If the liberalism of the state government is low, the percentage of those voting who are in poverty might not have much impact on the level of welfare payments. However, if the liberalism of the government is high, then the percentage of those voting who are in poverty might have an important impact on the level of welfare payments.

A model such as equation 1: $Y = a + b_1X_1 + b_2X_2 + b_3X_3 + e$ will not test the reasoning I advanced immediately above. In order to test the model I just outlined we need to estimate an equation such as the following:

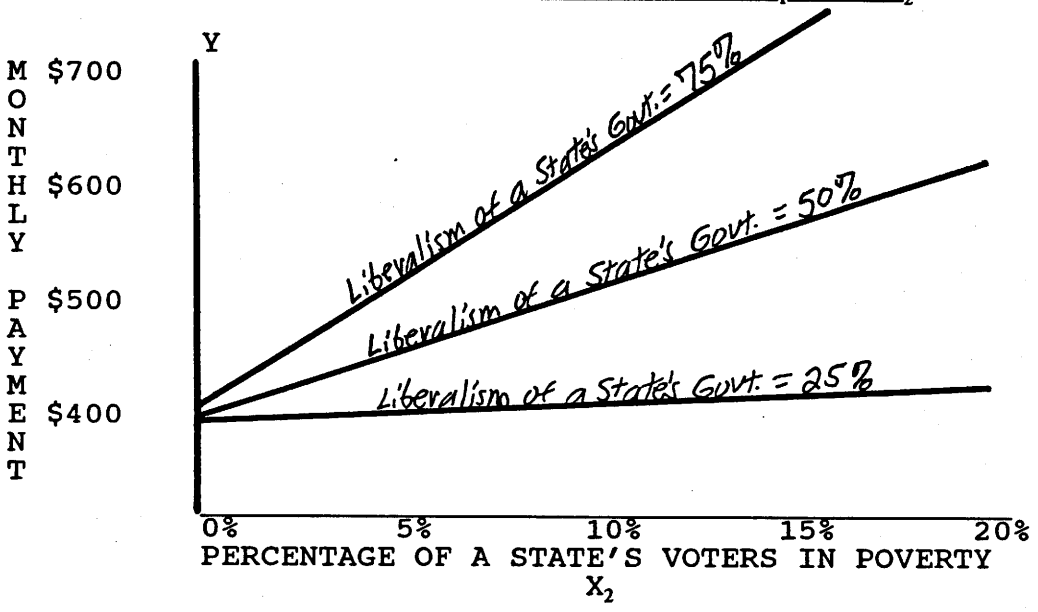
(equation 2) $Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_1X_2 + e$

The difference between equation 1 and equation 2 is the term $b_4X_1X_2$. The term $b_4X_1X_2$ says that the computer should multiply the score on X_1 times the score on X_2. For example, if a state had a liberalism score of 75% (i.e., its score on X_1 was 75) and 15% of the voting electorate in this same state were in poverty, then the computer reads the state's score on the "multiplicative" term X_1X_2 as 1125 (because 75 multiplied by 15 = 1125). The computer would then perform this same task for each other state.

Verbally, equation 2 says that the level of welfare payments in a state (Y) depends upon the liberalism of the state's government (X_1), the percentage of the state's voting electorate who are in poverty (X_2), the state's median family income (X_3) and the "interaction" (i.e., "product" - thus multiplication) of the liberalism of the state's government (X_1) multiplied by the percentage of the state's voters who are in poverty (X_2). A major study of welfare payments among the states found some support for the above pattern of reasoning. The liberalism of the state's government and the percentage of the state's voting electorate who are in poverty did effect the level of the state's welfare payments "interactively." However, the interactive effect was not nearly strong enough to attain statistical significance (i.e., b_4 in equation 2 above had a "t ratio" well below 2.0).

Perhaps a diagram will help convey the central point. In the diagram on page 132, I graph the relationship between the

percentage of a state's voters who are in poverty (X_2) and the level of welfare payments (Y - the amount of money a family of four receives monthly on public assistance) for three different levels of state government liberalism (X_1): a state government that is 25% liberal, a state government that is 50% liberal and a state government that is 75% liberal (just keep reading). Notice below how the line representing the relationship between the percentage of a state's voters who are in poverty (X_2) and the monthly welfare payment for a family of four (Y) is steeper the greater the liberalism of the state's government (X_1). If the steepness of the line representing the relationship between X_2 and Y changes (i.e., is either more steep or less steep) as the level (i.e., score) on X_1 changes (as in the diagram below) then we need an interactive term between X_1 and X_2 (e.g., the $b_4X_1X_2$ term in equation 2 on page 131). If the three lines below were of the same steepness, we would not need the interaction term between X_1 and X_2.



Make sure that you can distinguish between non-linear models (e.g., the logarithmic and polynomial models on pages 93-105) and an interactive model. The next several sentences may, at first, be confusing. Just keep reading!!! If the model is non-linear (as on pages 93-105), the impact of the percentage of a state's voters who are in poverty (i.e., X_2) on the level of welfare payments (i.e., Y) changes as the level of X_2 itself changes. Thus, in the models on pages 93-105, if the percentage of a state's voters who are in poverty (X_2) increases from 5% to 6%, the impact on the state's welfare monthly welfare payments will be different (i.e., either more or less) than if the percentage of the state's voters who are in poverty increases from 19% to 20%. Each increase is 1% but if the average amount of change in the dependent variable is different if the percentage of the state's voters who are in poverty increases from 5% to 6% than if the percentage of the state's voters who are in poverty increases from 19% to 20%, the model is non-linear. Such models are why you read pages 93-105.

Alternatively, an interactive model says that the impact of the percentage of the state's voters who are in poverty (X_2) on the level of welfare payments (Y) depends not upon what percentage of the state's voters are in poverty (i.e., the level of X_2 itself), but rather, upon the level of some other independent variable (e.g., X_1, the liberalism of the state's government).

Heteroscedasticity

Thus far we have examined the consequences of violating some of the important assumptions of the regression model (e.g., non-linearity and multicollinearity). In this section we investigate violating another important assumption of the regression model: homoscedasticity.

Homoscedasticity means that the variance (i.e., width) of the points around the regression line does not change (i.e., is constant) as the value of X increases. Conversely, if the variance (i.e., width) of the points around the regression line changes (i.e., either increases or decreases) as the value of X increases, we have a case of heteroscedasticity. Figure 1 below shows a case of homoscedasticity. The points are approximately equally distant from the regression line regardless of the level of variable X. Both Figures 2 and 3 show cases of heteroscedasticity. In Figure 2 the distance of the points from the regression line increases as the scores on variable X increase (i.e., as we move from left to right on the X axis). Conversely, in Figure 3 the distance of the points from the regression line decreases as the scores on variable X increase (as we move from left to right on the X axis). While the forms are different, the distance of the points from the regression line is unequal in both Figures 2 and 3. Therefore, both Figures 2 and 3 are heteroscedastic.

Figure 1.

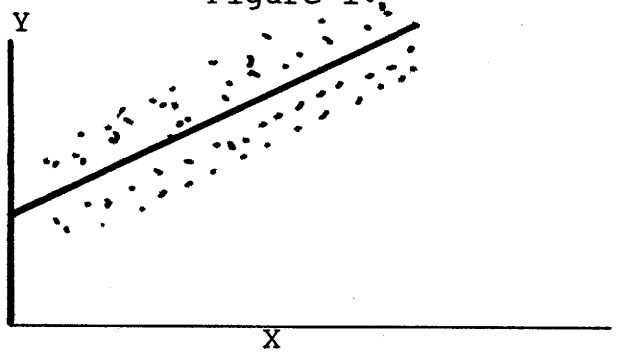


Figure 2

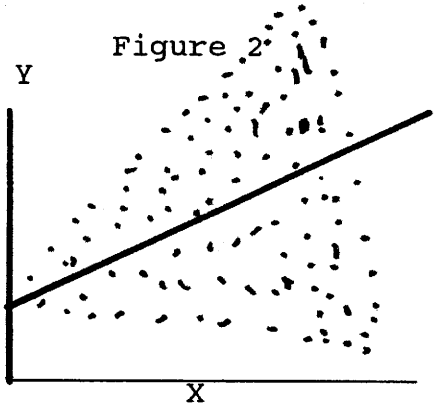
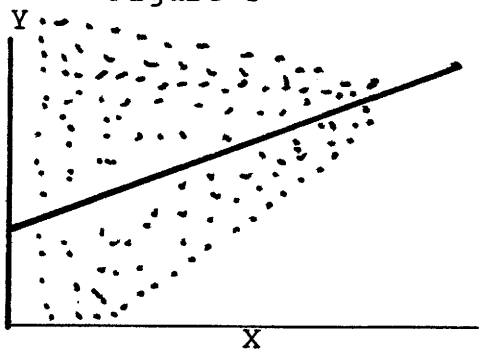


Figure 3



After visualizing heteroscedasticity in Figures 2 and 3 it seems reasonable to inquire about its consequences. After all, why should we be concerned about heteroscedasticity? As a first approach in answering this question, let us re-examine the process by which the regression line is calculated. As we have discussed previously, for the equation $Y = a + bX + e$, we use the least squared errors approach to estimate "a," "b," and "e" (pages 85-

87). As the name implies, the "least squared errors" approach minimizes the sum of squared errors (i.e., the sum of e^2 or "ESS" - see pages 85-87). This means that it is desirable to obtain the "best" fit in the worst fitting portion of the data. Put another way, if we could reduce a prediction error from "10" to "8" the squared errors would be reduced from "100" [$10^2 = (10)(10) = 100$] to "64" [$8^2 = (8)(8) = 64$]. This would reduce our squared errors by 36 units (from 100 to 64). If we reduce a prediction error from "4" to "2" the squared errors decline from "16" ($4^2 = 16$) to "4" ($2^2 = 4$). This only reduces the squared error by 12 units (from 16 to 4). In each instance we reduced the absolute size of the prediction error by 2 (10 - 8 = 2 and 4 - 2 = 2). However, the squared errors were reduced much more by reducing the prediction error from 10 to 8 rather than from 4 to 2. Therefore, in calculating the "best" fitting regression line according to the lowest sum of squared errors principle, the computer will minimize large prediction errors. This gives extra "weight" (or emphasis) to those levels of X where the errors in predicting Y are greatest. Therefore, the dots that are furthest from the regression line (i.e., the observations with the largest absolute values for "e") have the most impact on the estimated values for "a" and "b." The assumption of homoscedasticity is that each dot has approximately equal influence on the computation of "a" and "b."

The aforementioned mathematics of heteroscedasticity produces a highly undesirable outcome. The standard error of each coefficient (i.e., each "b") is usually increased but the least squared errors method typically underestimates the "true" standard errors (see A. H. Studenmund, Using Econometrics, second edition, pp. 374-375). If the estimated standard error of "b" is too small then the "t ratio" will be too large (remember that the standard error of "b" is the denominator of the "t ratio" - see page 89). If the "t ratio" is too large we will be more likely to reject the null hypothesis when the null hypothesis is actually true (i.e., commit a type I error - see page 89). For example, if our estimate of "b" is 5 and the "true" standard error of "b" is 4 then the "t ratio" is 1.25 (5/4 = 1.25). However, if due to heteroscedasticity, the computer reports a standard error of "b" of only 2, then the reported "t ratio" would be 2.5 (5/2 = 2.5). We would then mistakenly conclude that "b" was statistically significant (because the reported "t ratio" would be 2.5 - hence greater than the "2.0" threshold, see the third paragraph on page 90) when it is not (i.e., the "true" t ratio is only 1.25).

To make this discussion more useful, let me suggest a potential example of heteroscedasticity from political science. A famous theory from comparative politics concerns the relationship between overlapping divisions and the level of strife in a society. Christopher Achen summarizes this theory well by stating:

> Religious, political, sectional, or other divisions within a society are said to "overlap" when each of them divides the society in approximately the same way. For example, if most people in Quebec speak French and few Canadians elsewhere do, and if most Quebecers are Catholic while most other Canadians are Protestant, then linguistic, religious and sectional divisions overlap in Canada. The division rule asserts that, all else being equal, societies whose divisions overlap will experience more strife than those with nonoverlapping divisions (Christopher H. Achen, Interpreting and Using Regression, p. 13).

Thus, we would expect that the degree to which divisions overlap would be "positively" associated with the level of domestic strife.

In order to test such a hypothesis we might survey citizens in all nations and calculate the mean (average) level of overlapping linguistic, religious and sectional division within each society. Thus we would be aggregating individual scores to achieve a figure for each nation. Therefore the "unit of analysis" would be the nation, as opposed to the individual.

In addition to the level of overlapping division, a second potentially important independent variable might be the age of the nation's population. As the young are more likely to commit crimes than the old, it might be reasonable to hypothesize that the older the population of a nation, the lower the level of domestic strife. Hence, age of the population and level of domestic strife should be "negatively" associated. I also realize that a potentially important factor might be the dispersion in ages in a nation as opposed to the average age level. For example, it might be that a bi-polar society (many young, many old and a few middle-aged) might have a different level of strife than a society with the same "mean" age but composed of a more even distribution of ages. Thus, we could desire the "standard deviation" of ages as well as the "mean" age. However, for the sake of simplicity, we will ignore this possibility for now. Let us then regress the level of strife (variable Y) on the level of overlapping division (variable X_1) and the mean age of the population (variable X_2) as follows:

(Equation #1) $Y = a + b_1X_1 + b_2X_2 + e$

where Y = level of domestic strife, X_1 = degree of overlapping division and X_2 = mean age of the population.

Should we expect heteroscedasticity in equation #1? To answer this question remember that the error term ("e") includes the impact of omitted independent variables (see page 80). Furthermore, let us keep in mind that heteroscedasticity means that the squared value of the error term (i.e., "e^2") changes as either the level (i.e., amount) of some factor changes. Therefore, if the squared value of the error term is likely to change as either an included, or excluded, independent variable changes, heteroscedasticity should be expected.

For example, the squared value of the prediction errors for Y (i.e., "e^2") may be related to either X_1 or an omitted independent variable (e.g., a potential X_3 that we have left out of our model/equation). Additionally, the squared value of the error term (again, "e^2") may be related to the interaction of our independent variables (e.g., a term such as $b_3X_1X_2$). Thus, e^2 may be related to X_1 multiplied times X_2 and not to either X_1 or X_2 alone. Remember that an interaction term means that the impact of one independent variable on the dependent variable changes as the level of another independent variable changes (see pages 130-132). Finally, perhaps e^2 is related to the square of one of the independent variables (e.g., X_1^2).

Rather than just present a mathematical test and correction procedure for heteroscedasticity, I suggest we first try to think through the situation. In any quantitative analysis we should be trying to discover what process(es) lead to the set of data we have. In our situation this means what forces lead different nations to have different levels of strife? Obviously, if there are one, or more, previously omitted independent variables that theory suggests should be in our model, we should try to incorporate them. However, this does not mean that we should include every conceivable independent variable. Remember, one of the purposes of a model is to isolate those independent variables

which theory suggests should be important. If we just include every conceivable independent variable we are likely to cause some of the theoretically important independent variables to become statistically insignificant. For example, if the theoretically "unimportant" independent variables are related to the theoretically "important" independent variables, high multicollinearity may occur. If so, we may incorrectly conclude that a theoretically important independent variable is statistically insignificant (i.e., commit a "type II error"). Such errors cause us to build poor theories/models.

Of the mathematical possibilities that I have previously mentioned, the most likely to cause heteroscedasticity are either an omitted interaction term or an omitted independent variable. The impact of political variables often depends upon the level of other independent variables. As William D. Berry notes, "anytime a high value for an independent variable seems to be a necessary but not sufficient condition for an observation (in our case a nation) having a high value on the dependent variable, heteroscedasticity should be suspected" (William D. Berry, Understanding Regression Assumptions, page 75). Applied to our situation this could mean that although a high degree of overlapping division might be necessary for a high level of domestic strife, it is not sufficient. Thus, a nation could have a high degree of overlapping division and not have a high degree of domestic strife. However, if the nation had both a high degree of overlapping division and a low mean age, then it might be very likely to have a high degree of domestic strife. If so, heteroscedasticity is likely because equation #1 does not contain an interaction term between a nation's degree of overlapping division and its mean age (i.e., there was not a multiplicative term between X_1 and X_2 - note the omission of $b_3X_1X_2$ in equation #1).

If a reasoning process such as the one I just went through indicates that heteroscedasticity is likely to be a problem, a formal test for it should be conducted. Although there are a variety of different tests for heteroscedasticity, one which is increasingly popular is the White test (named for econometrician Hal White).

As you read the description of the White test below, try to follow the logic of the test. In other words, why do we undertake the various steps in the test? On a quiz I will not expect you to memorize either the order of the steps or what each step asks us to do. Rather, I would ask you either the purpose of the White test or what is the logic of the test. Just keep reading!!

The first step in executing the White test is to estimate equation #1. Specifically, we are interested in the values for the error term ("e") in equation #1. To review, remember that each value for "e" (also called the "residual") is the difference between the actual score on variable Y and the predicted score on variable Y (see page 79). For example, suppose we estimate equation #1 and the results are as follows: $a = 7$, $b_1 = .600$, and $b_2 = -.300$. Let us further stipulate that the value of X_1 for nation #1 is 9 and the value of X_2 for nation #1 is 34.3. If so, the predicted value of Y (or $\hat{Y}$) for nation #1 would be 2.11 {7 + [(.600)(9)] + [(-.300)(34.3)] = 2.11}. If the actual value of Y for nation #1 is 3, the value for "e" for nation #1 would be .89 [3 (the actual value of Y for nation #1) - 2.11 (the predicted value of Y for nation #1) = .89]. Next, we square each nation's value for "e" (i.e., "e^2"). For nation #1 the value of e^2 is .79 (.89 times .89 = .79). The final stage of step #1 of the White test is to ask the computer to store the values of e^2 in its "memory."

The second step in the White test is to build a model to explain the values of e^2 we generated in step #1. Therefore, we will use e^2 as the dependent variable in a second equation. Since heteroscedasticity means that the size of e^2 varies, we want to find independent variables that can "explain" (or are related to) the values of e^2. If we can "explain" a high percentage of the variation in the values of e^2, we probably have heteroscedasticity. Equally important, we should have a good idea of what is causing the heteroscedasticity (i.e., whatever is important in explaining variation in e^2). This is the "logic" of the White test.

The third step in executing the White test is to use each independent variable in equation #1 taken to both the first and second powers (i.e., for example, X_1 and X_1^2 - this latter term is X_1 "squared") and all possible interaction terms (since we only have two independent variables this amounts to including an X_1X_2 term) to explain the variation in e^2. Accordingly, we estimate equation #2 below. The subscripts differentiate the terms in equation #2 from those used in equation #1. Please note that e^2 (the squared error term from equation #1) is not the square of e_1 (the error term in equation #2):

(Equation #2) $e^2 = a_1 + b_3X_1 + b_4X_2 + b_5X_1^2 + b_6X_2^2 + b_7X_1X_2 + e_1$

The fourth step in performing the White test is to take the value for R^2 from equation #2, multiply it by "N" and compare it against the appropriate chi-square value. This will make more sense when I explain it. Suppose that the value for R^2 for equation #2 is .10. This would mean that the various independent variables in equation #2 "explained" 10% of the variation in the squared prediction errors for Y from equation #1. If we have data for 90 nations then "N" = 90. So, multiplying R^2 by "N" yields a product of 9 [i.e., (.10)(90) = 9]. We then compare "9" against the appropriate value for chi-square.

Rather than discuss the chi-square test in detail, I will tell you that we need a chi-square value of approximately 11.07 (or greater) to conclude that we have a high enough degree of heteroscedasticity to warrant further action. Put another way, our White test score of "9" is sufficiently "low" (i.e., less than 11.07) that we can conclude the independent variables in equation #2 are not sufficiently related to the squared prediction errors from equation #1 to indicate heteroscedasticity. The threshold chi-square value (11.07) depends upon the number of independent variables in equation #2 (we have 5: X_1, X_2, X_1^2, X_2^2 and X_1X_2) and the level of significance the researcher desires. I used the generally accepted .05 level of significance.

Remember from past readings/discussion that the level of significance is equal to the probability of committing a "type 1" error (i.e., rejecting the null hypothesis when the null hypothesis is actually "true"). For our purposes the null hypothesis is that the squared errors in equation #1 are homoscedastic (i.e., of equal - or approximately equal - value). In applying the White test, we are trying to see if we can reject the null hypothesis (i.e., homoscedasticity) in favor of the alternative hypothesis that the squared errors in equation #1 are not of equal size or value (i.e., heteroscedastic).

As noted above, the results from the White test were statistically insignificant at the .05 level (remember to achieve statistical significance at the .05 level we needed a score of 11.07, but we scored only 9 - i.e., less than the threshold level).

Therefore, if we reject the null hypothesis of homoscedasticity in favor of the alternative hypothesis of heteroscedasticity we would be mistaken more than 5% of the time. Thus, we should not reject the null hypothesis of homoscedasticity. Accordingly, we would report the results from equation #1 and not worry about heteroscedasticity.

What if the results from the White test had shown significant heteroscedasticity? As I mentioned previously, my first response would have been to seriously consider re-specifying equation #1. Perhaps I should have concluded from prior theory and research that an interaction term between X_1 and X_2 belonged in equation #1. Additionally, I would have looked at the coefficient for the interaction term in equation #2 (i.e., "b_7"). If this term had a "t ratio" with an absolute value of 2.0 (or greater) this would mean that it was a significant predictor of the value of the squared errors in equation #1. This would be a good reason to include such a term in a "revised" equation #1. Thus, I could add an interaction term (such as $b_3X_1X_2$) to equation #1. On the other hand, perhaps I omitted a theoretically important independent variable (other than the interaction term) in my original version of equation #1. If so, I should include it.

It is important to note that had the White test revealed significant heteroscedasticity in my "original" version of equation #1, I would need to re-apply the White test to my "revised" version of equation #1. Remember that the dependent variable in equation #2 (the equation for the White test) is e^2 from equation #1. If I "revise" equation #1 (e.g., adding the interaction term $b_3X_1X_2$ to the "original" version of equation #1), the values for the error term in the "revised" version of equation #1 will likely differ from those in the "original" version. For example, the prediction error for nation #1 in the "original" version of equation #1 is .89. Thus, the squared value is .79 (.89 times .89 = .79). If, for a "revised" version of equation #1, the prediction error for nation #1 is .50 the "squared" error becomes .25 (.50 times .50 = .25). Obviously, .25 is not equal to .79. Therefore, the values of the dependent variable in the White test (the values for e^2) are likely to be different for a "revised" version of equation #1 than for the "original" version of equation #1. Consequently, if we "revise" equation #1 we must re-apply the White test to see if heteroscedasticity is present after the "revision" of equation #1.

What if we have heteroscedasticity and either we can not "revise" the equation or a "revision" does not eliminate heteroscedasticity? In either event we can use a mathematical adjustment. Let us explore this from an intuitive perspective. Suppose that someone reported the following 3 test scores: 5, 20, and 100. Now suppose you find out that each test score was reported to be 5 times greater than it actually was. Wouldn't the logical response be to divide each test score by 5? Sure! Thus, instead of 5, 20, and 100 the test scores (after "adjustment") would be 1 (5/5 = 1), 4 (20/5 = 4) and 20 (100/5 = 20). In this example, "5" is called the "proportionality factor."

Similarly, if the residuals of equation #1 (i.e., the values for the error term, "e") are heteroscedastic and we then divide them by the "proportionality factor," the residuals should become homoscedastic. The value of the "proportionality factor" is derived from the most statistically significant independent variable in the White test. The "proportionality factor" will be symbolized by the letter "Z." As all terms of an equation must be treated the same, we will need to divide all terms in equation #1 by the "proportionality factor."

Equation #3 (below) is equivalent to equation #1 with all terms divided by the "proportionality factor" ("Z"):

(Equation #3) $Y/Z = a/Z + b_1(X_1/Z) + b_2(X_2/Z) + e/Z$

The above procedure is referred to as "weighted" least squares. It is called "weighted" least squares because each term is "weighted" (here divided) by the "proportionality factor" (in this case "Z"). Weighted least squares is a form of "generalized" least squares. The idea of "generalized" least squares is that we incorporate the information (in this case "Z") that will mathematically correct for a violation of a regression assumption (in this case the assumption of homoscedasticity). We estimate equation #3 by the least squared errors method.

Potential quiz questions include: (1) What is an interaction term and under what circumstances would a political scientist use one? (Be prepared to answer in words, not with a diagram.) (2) What is heteroscedasticity and why are political scientists concerned about it? (Again, be prepared to answer in words, not with a diagram.) (3) If we have heteroscedasticity, what should be our first response (see pages 135-136)? (4) If our first response does not eliminate the heteroscedasticity, what should be our second response (see the next to the last paragraph on page 138)? Pages 140-149 immediately ahead discuss autocorrelation. A good way to prepare for a quiz on autocorrelation is to take questions 2-4 above and simply substitute autocorrelation for heteroscedasticity. (Obviously, the page numbers cited in questions 3-4 above do not deal with autocorrelation). Other potential quiz questions include: (5) What is the logic of the White test for heteroscedasticity? (6) What is the logic of the mathematical adjustment for autocorrelation on pages 144-145 ahead? (7) Why would a political scientist use the distributed lag model discussed on pages 148-149 ahead?

Autocorrelation

Previously we discussed the consequences and adjustments for violating the regression assumption of homoscedasticity. In this section we will make a similar examination of the regression assumption of no autocorrelation (autocorrelation is defined ahead - just keep reading).

In order to understand the assumption of no autocorrelation it is first necessary to review the meaning of the error term ("e"). Consider the following equation: $Y = a + bX + e$. Remember that the value for "e" for each observation is the difference between the actual value for Y (i.e., Y) and the predicted value for Y (i.e., $\hat{Y}$). The predicted value for Y (i.e., $\hat{Y}$) is equal to: $a + bX$ (see pages 78-79). Therefore, $e = [Y - (a + bX)]$. Equivalently, $e = Y - \hat{Y}$. Thus, if for observation #1 the actual score on Y were 9 and the predicted score on Y were 7, the value of "e" for observation #1 would be 2 (i.e., $e = Y - \hat{Y}$ and $9 - 7 = 2$). Since "2" is a "positive" number this would be an example of a "positive" prediction error. Thus, if our model predicts a value for Y which is less (i.e., lower) than what actually occurs this is called a "positive" error. Conversely, had our model predicted that Y would be 11 when it was actually 9, this would result in a prediction error of "-2" (i.e., $e = Y - \hat{Y}$ and $9 - 11 = -2$). Therefore, if our model predicts a value for Y that is greater (i.e., higher) than what actually occurs, we have a "negative" error.

The assumption of no autocorrelation means that the direction of the errors (i.e., positive or negative) from consecutive observations (i.e., the prediction errors for observation #1, observation #2, observation #3, etc.) are uncorrelated (unrelated). Therefore, without autocorrelation the direction of the prediction error for observation #1 is unrelated to the direction of the prediction error for observation #2. For example, if we have a positive value for "e" for observation #1 this will not increase the probability that the value of "e" for observation #2 will also be positive. Thus, the assumption of no autocorrelation means that knowing the direction of the error for observation #1 is of no help in predicting the direction of the error for observation #2. Hence, with no autocorrelation the direction of consecutive errors is random (i.e., no discernable pattern).

However, if we have a positive value for "e" for observation #1 and this increases the probability that we will have a positive value for "e" for observation #2, we have "positive" autocorrelation. Similarly, if we have a negative value for "e" for observation #1 and this increases the probability that we will also have a negative value for "e" for observation #2, we again have "positive" autocorrelation. Remember from mathematics courses that if you multiply two numbers of the same sign the answer is positive. Thus, 9 times 9 equals 81. Similarly, -9 times -9 also equals 81. For this reason a series of errors of the same sign (i.e., positive, positive, positive, positive; or, negative, negative, negative, negative) is called "positive" autocorrelation.

Conversely, if you multiply numbers of opposite signs, the answer is negative. Thus, both $[(9)(-9)]$ and $[(-9)(9)]$ equal -81. For this reason, a series of errors of opposite signs (i.e., positive, negative, positive, negative; or, negative, positive, negative, positive) is called "negative" autocorrelation.

The following diagrams should make the preceding points clearer. Figure 1 below shows a pattern of no autocorrelation. As the score on variable X increases, there is no discernable pattern to the errors (remember that the error is the difference between the point, that is, the actual value of Y and the value of Y predicted by the regression line, thus, the value of Y).

Figure 1 - No Autocorrelation

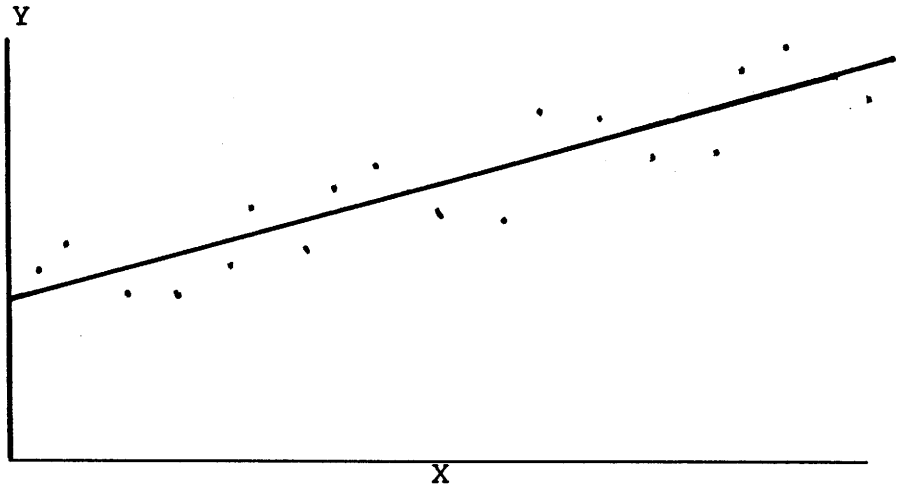


Figure 2 below shows a pattern of "positive" autocorrelation. Notice that there are six consecutive dots above the regression line [i.e., positive errors, because the actual value for Y is greater than the predicted value for Y (the value that appears on the regression line)] followed by seven consecutive dots below the regression (i.e., negative errors, because the actual value for Y is less than the predicted value for Y) followed by six consecutive dots above the regression line.

Figure 3 (page 142) shows a pattern of "negative" autocorrelation. Notice the alternating pattern of the dots. One dot is above the regression line, the next dot below the regression line, and so forth.

Figure 2 - Positive Autocorrelation

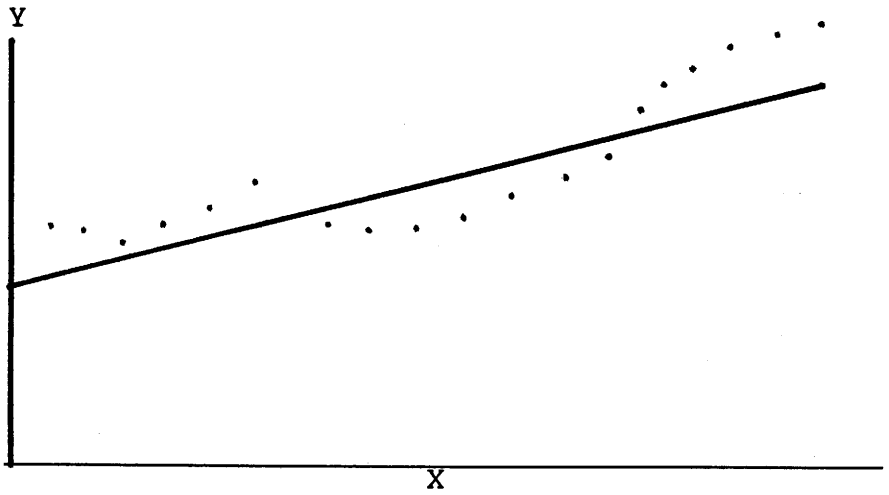
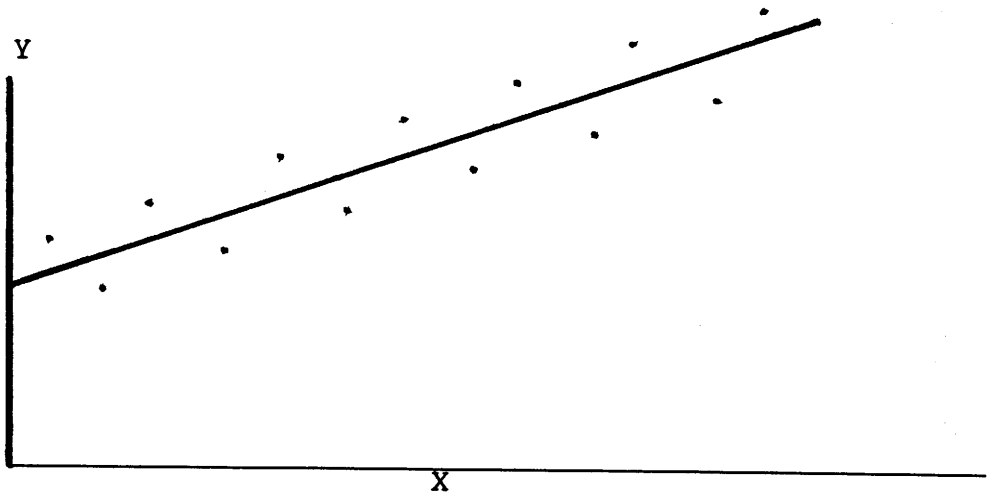


Figure 3 - Negative Autocorrelation



As both autocorrelation and heteroscedasticity pertain to the behavior of the error term (the residual) it is helpful to differentiate between them. If you compare Figures 2 and 3 above with Figures 2 and 3 on page 133 the difference will be clear. Whereas heteroscedasticity is concerned with how far the points are from the regression line (i.e., the width of the points around the regression line), autocorrelation is concerned with the direction of the points from the regression line (i.e., whether the points are above or below the regression line).

Why should we be concerned about autocorrelation? Because autocorrelation has the same important consequence as heteroscedasticity: the reported standard errors of the coefficients (i.e., the b^s) are highly likely to be too low (if you are confused, turn to page 134, read the second paragraph and substitute "autocorrelation" for "heteroscedasticity"). Thus, with autocorrelation, we will be more likely to commit a "type I" error (reject the null hypothesis when the null hypothesis is actually true). Additionally, in the overwhelming majority of circumstances, autocorrelation leads to a reported value of R^2 that is too high (Robert S. Pindyck and Daniel L. Rubinfeld, Econometric Models and Economic Forecasts, third edition, page 138). Thus, we will think that our model has explained a greater percentage of the variation in the dependent variable than it actually has.

When is autocorrelation likely to be a problem? Autocorrelation is likely to be a problem with time-series data. Time-series data mean that we have repeated observations on the same variables over time. For example, suppose we want to study what factors effect U.S. Defense expenditures. Since the federal budget is adopted annually (i.e., each year), it would be reasonable to gather U.S. Defense expenditures on an annual basis (e.g., the amount the U.S. spent on defense expenditures for each year since 1945). This would be time-series data.

Time-series data have an important advantage over cross-sectional data (all data collected at one point in time, e.g., the same year). Recall from our earlier discussion of research design (see pages 5-9) that political scientists almost invariably use nonexperimental data (the researcher can not alter the levels of the independent variables). If X is a cause of Y then X must change before Y changes. We can be more certain of this if we have time-series (as opposed to cross-sectional) data. If we know that both the independent and dependent variables changed values, but also that X (the independent variable) changed prior (at an earlier

time period) to Y (the dependent variable), we are on much sounder ground in concluding that X is a cause of Y. Cross-sectional data are less certain on this critical "time-of-change" question. Therefore, we would prefer time-series data. Unfortunately, for reasons we soon examine, time-series data are likely to be plagued by autocorrelation.

To make this discussion more useful, let us examine how we might model U.S. Defense expenditures annually from the end of World War II until the present. Suppose we hypothesize that the amount the U.S. spends on defense (the dependent variable) is a function of the amount of our gross national product (the independent variable). Such a model might be plausible because how much you spend (variable Y) is likely to be related to how you have to spend (variable X). If we regressed annual U.S. Defense expenditures on annual U.S. GNP we would probably produce a graph similar to Figure 2. This would likely occur because our model is probably mis-specified. Among other independent variables that we have mistakenly omitted is U.S. military involvement. All other factors being the same (such as the dollar amount of GNP), the U.S. will probably spend more money on the military during periods of military involvement. Remember that the error term ("e") includes the effects of omitted independent variables (see page 80). If we omit an independent variable measuring military involvement, the effect of military involvement on defense spending will be included in the error term. Since military involvement likely increases defense spending, the regression line will probably underestimate U.S. Defense expenditures during war periods and overestimate defense expenditures during peace time. As war (or peace) often continues for many years, the prediction errors will likely be in the same direction for many consecutive years. Hence, we will have positive autocorrelation. For this reason, positive autocorrelation is much more likely than negative autocorrelation. More generally, as most variables either increase or decrease continually over time, an omitted independent variable is likely to result in positive autocorrelation. For example, doesn't GNP almost always increase each successive year? Yes! Therefore, omitting GNP (if theory indicates it should be included as an independent variable) is highly likely to result in positive autocorrelation.

As with any violation of regression assumptions, it is wise to re-examine the model. Remember that the goal in any quantitative analysis is to uncover the process that produced the scores on the dependent variable. As the above discussion indicates, we should include omitted independent variables suggested by theory. In some cases "time" (i.e., creating a variable with a value of "one" in time period one, "two" in time period two, etc.) can serve as a "proxy" for an omitted variable that either increases or decreases by approximately the same amount in each time period (e.g., each year if the data are collected annually). Revising the model might remove the autocorrelation, or more likely, make the degree of autocorrelation statistically insignificant.

What if we have autocorrelation and either we can not "revise" our model or the "revision" of the model does not eliminate autocorrelation? For example, suppose we have autocorrelation and any of the following three possibilities occurs: (1) theory does not suggest an additional independent variable; (2) we can not obtain a measure for an additional independent variable that theory suggests (e.g., we can not validly represent the omitted independent variable by "time" - or any other available measure); (3) while theory suggests an additional independent variable autocorrelation remains after we include it. At this point we will need to make a purely mathematical adjustment for autocorrelation.

When we discussed heteroscedasticity, we utilized a mathematical adjustment procedure called "generalized" least squares (i.e., "GLS," see pages 138-139). Generalized least squares means that we use the least squared errors criterion for drawing a regression to fit data which have been adjusted to incorporate the information that is violating the regression assumption. With autocorrelation the regression assumption that is violated is that the error terms are uncorrelated. Rather than the direction of the errors being random, we have a systematic pattern. For example, as we have seen, with positive autocorrelation prediction errors for successive observations on Y tend to be in the same direction (i.e., of the same "sign" - either consecutively positive or negative). Thus, with positive autocorrelation there is a positive correlation of the error terms from consecutive observations (on correlation see page 75 of this packet). Again, the regression assumption is that the error terms are uncorrelated, hence the correlation between the error terms should be .00 (or, in a statistical sense, "insignificantly" different from .00). Positive autocorrelation would mean that we will have some statistically significant non-zero positive correlation between .00 and 1.0 (the highest positive value a correlation coefficient can attain - again, see page 75).

If we knew the strength of the correlation between consecutive error terms, we could transform (alter) the scores on our variables and the error term by the strength of the correlation. We could then perform the least squared errors estimation procedure for "a" and "b" on the transformed scores. The results should be "free" from autocorrelation. For example, suppose we estimated the following equation by the least squared errors method:

(Equation #1) $Y_t = a + b_1X_t + e_t$

> Note: the subscripts indicate that we are working with time-series data. For example, if subscript "t" was the year 1996, and we had annual (yearly) data, then subscript "t-1" would be for the year 1995.

Now suppose we discovered that our results contained positive autocorrelation. If we knew the strength of the positive correlation between the error terms we could remove the autocorrelation by transforming the data as they are below and using the least squared errors method on the transformed data. The logic of the ensuing transformation will be explained shortly. Just keep reading! Thus, we would estimate the following equation:

(Equation #2) $Y_t - pY_{t-1} = a(1-p) + b_1 (X_t - pX_{t-1}) + (e_t - pe_{t-1})$

> Note: "p" (rho, pronounced "row") is the size of the correlation coefficient between consecutive error terms (e.g., the correlation between the error term for predicting Y at time "t" with the error in predicting Y at time "t-1").
> Further note: "p" performs the same role in adjusting for autocorrelation that "Z" performs in adjusting for heteroscedasticity (see pages 138-139).

In order to understand both autocorrelation and equation #2, one must understand the error term. With time-series data, it is useful to think of the current value of the error term as both a function of the error term from the immediately preceding time period and a random component:

(Equation #3) $e_t = pe_{t-1} + u_t$

where:

e_t = the prediction error for Y at time "t" (e.g., 1996).

p = the correlation of the error terms for consecutive time periods (e.g., the error in predicting Y at time "t" with the error in predicting Y at time "t-1", the error in predicting Y at time "t-1" with the error in predicting Y at time "t-2", etc.).

e_{t-1} = the prediction error for Y at time "t-1" (e.g., 1995).

u_t = the random portion of the prediction error of Y at time "t." That is, the portion of the error in predicting Y at time "t" which is independent, or unrelated, to the prediction error of Y at time "t-1". Thus, u_t should meet the assumptions that e_t was suppose to meet. The random component (u_t) can be thought of as representing random measurement error in the dependent variable. Thus, we may randomly mis-measure scores on Y.

The practical importance of the previous decomposition of the error term can be appreciated if we work through the following algebraic manipulation:

(Equation #3) $e_t = pe_{t-1} + u_t$ (now subtract pe_{t-1} from each side)

$$e_t - pe_{t-1} = pe_{t-1} - pe_{t-1} + u_t$$

$$e_t - pe_{t-1} = u_t$$

As note previously, u_t is the random component of e_t and thus u_t conforms to the assumptions e_t was suppose to meet. If we change the error term in equation #1 on page 144 from e_t to $e_t - pe_{t-1}$ we have an error term which is equivalent to u_t (note above that $e_t - pe_{t-1} = u_t$). This new error term is random (as e_t was suppose to be). Now, look again at the error term in equation #2 on page 144. Hopefully, it is now clear why we use $e_t - pe_{t-1}$ instead of e_t. In order to maintain the proportionality an equation requires, we have to similarly transform all other terms in equation #2 on page 144. For example, note that in equation #2 we replace Y_t with $Y_t - pY_{t-1}$.

Now we need a value for "p." In order to obtain an approximation of "p" we could estimate equation #3. If "p" were close to 1.0 (e.g., .87) this would indicate strong positive autocorrelation. If "p" were close to -1.0 (e.g., -.87) this would indicate strong negative autocorrelation. If "p" is greater than 1.0 in absolute value (e.g., 1.1 or -1.1) it would mean that the error term would continually be increasing in absolute value over time (i.e., "explode"). This is unreasonable. Thus, like the correlation coefficient, "p" should be between 1.0 and -1.0. The greater the absolute value of "p" the more e_t is a function of (i.e., related to) e_{t-1}.

In order to estimate equation #2, we could insert our estimate of "p" (from equation #3) into equation #2 and then estimate equation #2 by the least squared errors method. The results from our estimation of equation #2 should be analogous to the results we would have obtained from equation #1 if equation #1 had not been plagued by autocorrelation.

Several additional points are worth noting. First, the previous discussion assumed that autocorrelation was a "first-order" process. A "first-order" process means that prediction errors from immediately adjacent time periods are related. While this is generally true, it does not have to be. For example, the error term at time "t" (e.g., 1996) could be related to the error term three time periods in the past (e.g., 1993). This would be third-order autocorrelation (the number of "orders" being equal to the number of time periods apart). Such a process is highly unlikely. What theory would indicate such a process? Second, in practical application virtually no one uses the method I just described for estimating "p." The conventional approaches for estimating "p" are merely more sophisticated methods of estimating equation #3.

Now that you have an understanding of "p," we can discuss the approach political scientists use to determine if they have a severe enough case of autocorrelation to warrant switching from equation #1 on page 144 to equation #2 on page 144. While we can look at a scatter plot of the data to suggest whether, or not, we might have a severe case of autocorrelation (see Figures 2-3), it would be useful to have a more formal measure. Intuitively, one method might be to assess the probability of a series of prediction errors in the same direction. Thus, if the probability of a positive error is .5, then the probability of two successive positive errors is .25 [(.5)(.5) = .25; i.e., one chance in four]. Remember that in the absence of autocorrelation, the direction of consecutive errors is an independent event. Hence, obtaining a positive error at time "t-1" should have no effect on the probability of obtaining a positive error at time "t." Thus, as a test for autocorrelation, we could compute the probabilities of obtaining various patterns of consecutive errors (e.g., five consecutive positive errors, etc.) This is the logic of a famous test for autocorrelation (called the "Geary," "Sign," or "Runs" test).

The preceding approach to testing for autocorrelation is only useful for large samples (e.g., 50, or more, observations). Unfortunately, we often have relatively short time-series (e.g., data for 30 consecutive years). This is an important reason why the most widely used test for autocorrelation is the Durbin-Watson test. The formula for the Durbin-Watson test is as follows:

$$DW = \frac{\sum_{t=2}(e_t - e_{t-1})^2}{\sum e_t^2}$$

As a first step in understanding the Durbin-Watson test, look again at page 65. Think of the data this way: instead of ten different counties we have ten annual (yearly) observations on one county. Let us choose Los Angeles county. Thus, "county #1" would now be the first annual observation for Los Angeles county (e.g., 1987). Similarly, "county #10" would be the tenth annual observation for Los Angeles county (e.g., 1996). The computations for the numerator begin by subtracting the prediction error for observation #1 from the prediction error for observation #2 and squaring the difference. We will "lose" one computation because we can not subtract the error at time "zero" from observation (time) #1. Hence, with ten observations we will have nine figures to sum in the numerator, but ten figures to sum in the denominator. Page

65, column #9 shows the value of the error term for each observation. Using this data, our first computation would be $[(-1.05) - (-1.07)]^2 = [-1.05 + 1.07]^2 = [.02]^2 = .0004$. Our second computation would be $[(-.05) - (-1.05)]^2 = [-.05 + 1.05]^2 = [1]^2 = 1$. We would repeat this process for the seven additional possibilities and then add the results. This would yield the numerator of the Durbin-Watson formula (shown on page 146). Due to "sign" changes resulting from subtraction (minus a negative number becomes the same as adding a positive number), consecutive prediction errors of the same "sign" (i.e., positive autocorrelation) produce a smaller value for the numerator than consecutive errors of opposite signs (i.e., negative autocorrelation). To obtain the denominator, we square each of the ten errors and then add (sum) them (see page 65, column 10).

We can gain greater insight into the Durbin-Watson test by noting that with a little algebraic manipulation the Durbin-Watson statistic equals the following:

$$DW = 2(1 - p)$$

Note: As previously, "p" = the correlation coefficient for the error terms (see pages 144-145).

Looking at the above formula we can see that as "p" approaches 0 (hence no autocorrelation) the value of the Durbin-Watson statistic approaches 2 (substitute 0 for "p" in the above formula). Additionally, a high positive value for "p" (e.g., .90) would drive the Durbin-Watson statistic near 0 (substitute .9 for "p" in the above formula). Finally, a high negative value for "p" (e.g., -.90) would raise the Durbin-Watson statistic near 4.0 (substitute -.90 for "p" in the above formula).

The Durbin-Watson statistic is bounded by a range of 0 - 4.0 with 2.0 indicating no autocorrelation, values below 2.0 indicating positive autocorrelation and values above 2.0 indicating negative autocorrelation. Although standards vary and the Durbin-Watson statistic has outcomes which are inconclusive, a safe rule of thumb would be that with four, or five, independent variables and a time-series of roughly 35 observations, a Durbin-Watson statistic of approximately 1.2, or lower, indicates a degree of positive autocorrelation sufficient to require the remedial action of equation #2 (page 130). The corresponding Durbin-Watson value for negative autocorrelation is approximately 2.8, or higher. Instead of the Durbin-Watson test, some argue that if the absolute value of "p" is greater than .30 we should use equation #2 (page 144).

As autocorrelation is almost exclusively confined to time-series data, I have included in this discussion some of the advantages of using time-series data. Let me continue this section by briefly discussing an important application of time-series data, distributed lag models. It is quite common that the impact of an independent variable on the dependent variable does not entirely occur within one time period. Over the past two decades a burgeoning literature to which political scientists, as well as both economists and sociologists, have actively contributed has been quantitative political economy. Scholars in this developing field have been concerned with how changes in economic variables (e.g., unemployment and inflation) effect the electoral prospects of political parties and, in turn, how political parties alter important economic outcomes (e.g., unemployment, inflation, the distribution of income, etc.). To do this, political scientists have had to model processes where causality is likely to span several time periods. For example, the effect of a tax cut may

extend over many years. As many topics of interest to political scientists fit this same pattern (e.g., a hostile act by a foreign power may effect its diplomatic relations with other nations for many years), we need to be able to model such processes.

Distributed lags (where the impact of an independent variable is "distributed" over time) are crucial in developing realistic models. For example, suppose we have a Democratic president and s/he encourages the Federal Reserve Board to lower interest rates as a way of increasing economic activity and hence reducing unemployment. Such a policy would be expected (and past research by political scientists indicates rather conclusively that this has occurred) because middle and lower income voters, who are the backbone of the Democratic party, are also the most likely workers to be unemployed during a recession. To assess the impact of the interest rate on the unemployment rate we could estimate the following equation:

$$Y_t = a + b_1X_t + e_t$$

Where: Y = the annual (yearly) unemployment rate in percentage points. Note: Y_t = 1996, Y_{t-1} = 1995

X = the annual (yearly) interest rate in percentage points. Note: X_t = 1996, X_{t-1} = 1995

However, as current year unemployment is caused, in part, by business decisions from previous years (such as the time needed to build a new plant) our model is probably mis-specified. Most likely we should incorporate the interest rate for several years prior to the current year. Unfortunately, if we use a series of prior year scores on variable X as independent variables, they are likely to be highly inter-correlated. Hence, multicollinearity is likely to be a critical problem (on multicollinearity see pages 115-123).

Fortunately, a transformation called the Koyck distributed lag will greatly help our situation. Instead of including a series of prior values of X as independent variables, the Koyck distributed lag allows us to estimate the impact of previous time periods of X on the current time period of Y by using Y_{t-1} as an independent variable. Thus, we estimate the following equation:

$$Y_t = c + b_1X_t + b_2Y_{t-1} + e_t$$

Note: I used "c" as the y intercept (usually designated as "a") because the intercept must be calculated from the value for "c" in the above equation. As the value of the intercept is not of great consequence I will not discuss its derivation.

Using the results from the above equation, we can estimate the impact of prior years' values of X on the current year value of Y. To do this, we multiply the value of b_1 times the value of b_2. The formula is b_1 (the coefficient of X_t) multiplied times b_2 (the coefficient of Y_{t-1}) with b_2 taken to a power equal to the number of time periods we want to go back. Do not panic! Just keep reading!

For example, suppose that $b_1 = .200$ and $b_2 = .800$. This indicates that the impact of a one percent increase in the current year interest rate (we do not use any prior year interest rates, i.e., we go back zero time periods) on the current year unemployment rate is $(b_1)(b_2^0) = (.200)(.800^0) = (.200)(1) = .2$ (i.e., a two-tenths of one percent increase in the current year

unemployment rate). Remember that anything to the zero power equals "1" (i.e., $.800^0 = 1$). The impact of a one percent increase in last year's interest rate on this year's unemployment rate is $(b_1)(b_2^1) = (b_1)(b_2) = (.200)(.800) = .16$ percent. The impact of a one percent increase in the interest rate of two years ago on this year's unemployment rate is $(b_1)(b_2^2) = (b_1)(b_2)(b_2) = (.200)(.800)(.800) = .128$ percent. Notice that the exponent of b_2 is equal to the number of time periods we go back on X. Thus, if we go back two time periods (i.e., two years) on X (the impact of X_{t-2} on Y_t), the exponent of b_2 is "2" (i.e., b_2^2).

The results indicate that the effect of prior years' interest rates on the current year unemployment rate decays rather slowly (from .2 to .16 to .128). This occurs because b_2 (.800) is close to 1.0. The cumulative (i.e., total) impact of a one percent increase in the interest rate of two years ago (i.e., X_{t-2}, 1994) on annual unemployment rates over the next two years (i.e., 1995-96, or the impact of X_{t-2} on Y_{t-2}, Y_{t-1} and Y_t) is .2 + .16 + .128 = .488 percent. This is almost one-half of one percent.

The Koyck lag assumes that the impact of X on Y declines for prior values of X. Thus, in the previous example, b_2 had to have an absolute value of less than 1.0. Some distributed lag schemes (e.g., the Almon or polynomial lag) permit earlier time periods of X to have a greater impact on the current time period of Y than more recent time periods of X. However, theory in political science is rarely strong enough to suggest which previous time periods of X should have the greatest impact on the current level of Y. Consequently, such approaches are rarely used.

Over the past several decades political scientists have increasing used data that are both time-series and cross-sectional. For example, many recent studies in comparative political economy use data from approximately 15 nations (e.g., the U.S., Canada, Australia, Japan and a large group of European nations) annually over the past 25 years. The time-series aspect is that we have repeated data every year (i.e., "annual"). For example, we might use the unemployment rate for 15 nations for each of the past 25 consecutive years. Thus, for each nation this means 25 consecutive annual scores (e.g., one score for each nation for each year from 1972 to 1996). The cross-sectional aspect of such data is the number of nations. Thus, we have data in any one year, say 1996, on 15 different nations. The number of observations (i.e., the "N") in a study of 15 nations annually for each of the past 25 years is 375 (15 times 25 = 375). The great advantage in such a study is that we have far more observations than if we used either one nation (time-series) or one year (cross-sectional). For example, if we used just one nation we would have only 25 observations (e.g., one score for that nation for each year from 1972 to 1996). If we did a study by year, for each year we would only have data for 15 observations (e.g., one score for each of 15 nations in the year 1996). As we know from statistical significance (page 63), more observations are preferable to fewer observations. So, in the current example, by "pooling" (i.e., merging) time-series data with cross-sectional data, we can increase the number of observations from either 15 or 25 to 375. Such a design is called either a "Pooled Time-Series" or a "Time-Series Cross-Sectional" (TSCS) design. Pooled Time-Series or Time-Series Cross-Sectional studies often have both autocorrelation (from the time-series component) and heteroscedasticity (from the cross-sectional component). Remember, some potential quiz questions are contained in the last paragraph on page 139.