

II. SOME MORE ADVANCED METHODS

In this section, we take up a slightly more advanced set of problems. The issues we address relate directly to the problem of causality. We said at the beginning that the goal of theory is to try to understand what is causing what. Our theories state as explicitly as possible the causal relations among our concepts. We then subject these theories to a test by examining evidence to determine whether our expectations can be verified. Thus, if we expect voters' financial condition to influence how they vote, we arrange for a test by measuring the concepts and finding out whether there is a statistical relationship between the variables. The test is interesting because there is every chance our theory is wrong. If, when we examine the evidence, we find there is no statistical relationship, we are forced to doubt our theory. That's okay. When we rethink our understanding, we may come up with a better theory.

When we find no statistical relationship, there is a kind of closure. We need to look elsewhere for a theory which will help us understand. But what if we find a relationship? We have said a relationship supports the hypothesis that x influences y , but it does not prove it. The problem in this section is to push our analysis in order to provide additional support for the hypothesis (but, alas, still not proof).

When we state a hypothesis with an independent and a dependent variable we reduce a complex problem to one involving only two variables. This is called bivariate analysis. When we complicate our analysis to approximate more closely the complexity associated with our problem, we introduce more than two variables. We conduct multivariate analysis. When our theories accommodate more complexity, they become more plausible, especially if our analysis of the evidence supports the theory.

Alternative Independent Causes

The simplest kind of multivariate analysis admits that there is more than one cause of the dependent variable. Acknowledging more than one cause may have implications for our theory, particularly for the causal argument on which the theory depends.

For example, we have looked already at the relationship between DIVOTE and V056. We found a statistical relationship between the two variables, and hence we could conclude the evidence supports the argument that personal financial condition has an impact on how people vote. We might in turn argue that this supports an important expectation about how elections promote accountability of leaders to followers.

But surely there are many other things which influence how people vote. If we want to understand the more general question of why people vote the way they do, we need to take into account these other causes of voting choice. It may also be that by asking the broader question we can learn more about the question of accountability.

Certainly one thing that is often argued to be a powerful influence on the vote is party identification. Many people develop a loyalty to one of the parties which influences how they perceive electoral politics, and, ultimately, how they vote. Some theorists understand the

influence of party identification on voting choice as contrary to political accountability because party identification can serve as a kind of screen or blinder which limits citizens' ability to act on the basis of their self-interests. Let us see how multivariate analysis might proceed to examine the alternative cause of party identification.

Our first step has already been taken. We have examined the effect of personal finance (V056) on the vote, and found a relationship. In this discussion, we change the dependent variable, presidential vote, slightly to reflect only the choice between the two major-party candidates, Bill Clinton (the Democratic incumbent) and Bob Dole (the Republican challenger). We eliminate Perot voters to simplify things for the rest of our analysis. To do this, we have created a new variable, TWOPVOTE (for two-party presidential vote). The relationship between TWOPVOTE and V056 is presented in Table 6.

Table 6. Bivariate Relationship between Personal Finances and Presidential Vote, 1996

TWOPVOTE * V056 Better Off Than Last Year? Crosstabulation

| | | V056 Better Off Than Last Year? | | | Total |
|----------|--------------|---------------------------------|--------|---------|--------|
| | | 1 Better | 2 Same | 3 Worse | |
| TWOPVOTE | 1.00 Clinton | 302 | 182 | 105 | 589 |
| | | 65.5% | 52.3% | 46.9% | 57.0% |
| | 2.00 Dole | 159 | 166 | 119 | 444 |
| | | 34.5% | 47.7% | 53.1% | 43.0% |
| Total | | 461 | 348 | 224 | 1033 |
| | | 100.0% | 100.0% | 100.0% | 100.0% |

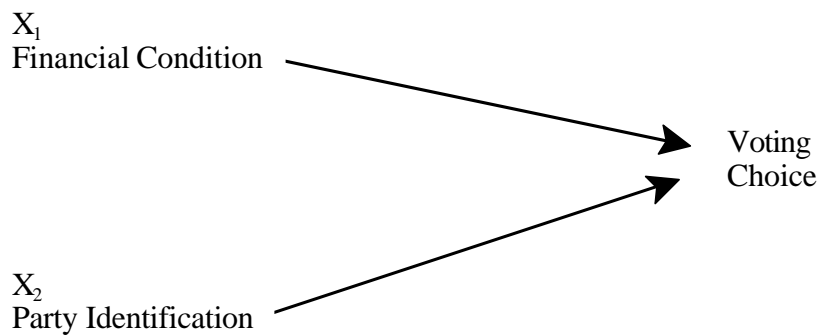
Tau-C = .167; Approximate significance = .000.

You can see that there is a relationship between financial condition and the two-party vote for president, as indicated by the progressively smaller vote for Clinton as individuals' financial conditions went from "Better" to "Same" to "Worse." The relationship is summarized by the Tau C correlation coefficient of .17. This indicates a pretty weak but detectable relationship, well above 0, but a long way from a perfect relationship of 1.0.¹

¹There are two Kendall's Tau correlations of concern to us. Tau C is appropriate for tables which are asymmetric. An asymmetric table is one with an unequal number of rows and columns. Table 6 has 3 columns and 2 rows, hence we requested Tau C when we ran this crosstabulation. Tau B is appropriate when the table is symmetric, that is, when it has an equal number of rows and columns. The "significance" reported along with the correlation coefficient indicates the chances that the correlation coefficient reported in the sample could have resulted when the true value of the correlation in the population from which the sample was drawn is 0. In this case, the chances of finding a correlation of .17 in a sample this large (N = 1033) when there is no correlation in the population are extremely low (< .0005). Hence, the computer reports a "significance" of .000. We would conclude from this that the correlation between the two variables is "statistically significant."

Our theory admits that party identification may be an alternative cause; that while financial condition has an impact on voting choice, so also does party identification. People vote out of a sense of loyalty to their party, or because they have made a "standing decision" to support their party in the absence of a compelling reason to defect from their party. Assuming we accept the plausibility of both arguments, we might want to compare the impact of personal financial condition (V056) with the effect of party identification (V008) on presidential vote. We can diagram the expected relationships in Figure 1 as follows:

Figure 1. Independent Causes of Voting Choice



The first hypothesis is that personal financial condition has an effect on presidential vote choice. Table 6 presents a test of this hypothesis, and finds a statistical relationship consistent with the expectation. The next step is to specify the second hypothesis, which is that party identification also has an effect on presidential voting choice. The test for this is to cross tabulate the two variables – party identification and voting choice (see Table 7).²

²We use a version of V008, party identification, recoded into three categories.

Table 7. Crosstabulation of Two-Party Vote and Party Identification
 TWOPVOTE * TRIPTYID Crosstabulation

| | | TRIPTYID | | | Total |
|----------|--------------|---------------|------------------|-----------------|----------------|
| | | 1.00 Democrat | 2.00 Independent | 3.00 Republican | |
| TWOPVOTE | 1.00 Clinton | 400 94.3% | 149 57.3% | 41 11.7% | 590 57.1% |
| | 2.00 Dole | 24 5.7% | 111 42.7% | 309 88.3% | 444 42.9% |
| Total | | 424 100.0% | 260 100.0% | 350 100.0% | 1034 100.0% |

Tau-C = .767; significance = .000.

You can see that party identification is strongly related to presidential vote. Fully 94% of Democratic identifiers voted for Bill Clinton, while about 88% of Republicans voted for Bob Dole. Independents split more evenly, with a slight majority voting for Clinton. The frequency distributions of presidential voting show a dramatic decrease in the propensity to vote for Clinton across the three categories of party identification. This pattern is registered in a strong Tau C correlation of .77.

This analysis is "multivariate" only in the most rudimentary sense. We have actually done nothing more than compare two bivariate relationships where the dependent variable is presidential voting in both cases. We can compare them by saying, for example, that the effect of party identification on voting choice in 1996 is considerably stronger than the effect of financial condition. The "alternative cause" of party identification, based upon this analysis, seems to be quite a bit stronger than the cause originally posited by our argument that elections serve as mechanisms to promote accountability. If we judge party identification to be a concept which is not consistent with accountability (a judgment which would require a much more detailed argument than we have offered here), we would have to conclude that we have found evidence that while people may hold leaders accountable for their own financial condition to some degree, party loyalty is a much stronger influence on their voting choice.

We also need to keep in mind that the theory stated in Figure 1 indicates that there is no relationship between our two independent variables, X_1 and X_2 . As long as the theory is correct that there is no relationship between the two independent variables, the analysis in Tables 6 and 7 is correct. However, if the two independent variables are related to one another for any reason, we must take extra precautions before we can conclude we understand the independent effects of either on the dependent variable.

Control Variables and When to Use Them

A control variable is a variable which is not permitted to vary (it is held constant) while the covariation between two other variables is examined. The general rule is as follows: Use a control variable whenever there is reason to believe the original relationship of interest will change significantly when the control variable is held constant. This is a tricky idea that may take some time to grasp fully. So bear with us as we go through an example and explain more fully the meaning of control variables.

We know that financial condition has an influence on voting choice. We also know that party identification has a stronger influence on how people voted in the 1996 presidential election. With some thought, we might suppose that the effect of financial condition on voting choice depends to some degree on party identification – that, in other words, party identification and financial condition are related. Perhaps party loyalty "interferes with" the impulse to hold leaders accountable for personal financial condition. Republicans might have been more willing to blame Clinton for their ills because, after all, Clinton was a Democratic president. But were Republicans whose financial condition improved willing to give the president credit and vote for him? Presumably Democrats whose financial condition improved would be very willing to give the president from their party credit and vote for him, but would those Democrats whose circumstances had declined be likely to turn to the opposite party?

These questions suggest the need to examine the relationship between presidential vote choice and financial condition controlling for party identification. We are asking, in other words, whether the effect of financial condition on voting choice is different for Democrats than it is for Republicans, or different for Independents than it is for those who identify with one of the parties. To test this idea, we need to examine the relationship between vote choice and financial condition for each category of party identification. The results of this analysis are presented in Table 8. Table 8 is really three tables in one: it shows the crosstabulation of vote choice and financial condition for each of the three values of party identification: Democrats, Independents, and Republicans.

Table 8. Relationship between Financial Condition and Presidential Vote, Controlling for Party Identification.

| | | | | V056 Better Off Than Last Year? | | | Total |
|------------------|----------|--------------|--|---------------------------------|-----------|------------|--------|
| | | | | 1 1. Better | 2 2. Same | 3 3. Worse | |
| 1.00 Democrat | TWOPVOTE | 1.00 Clinton | | 209 | 122 | 67 | 398 |
| | | | | 96.8% | 91.7% | 91.8% | 94.3% |
| | | 2.00 Dole | | 7 | 11 | 6 | 24 |
| | | | | 3.2% | 8.3% | 8.2% | 5.7% |
| | | Total | | 216 | 133 | 73 | 422 |
| | | | | 100.0% | 100.0% | 100.0% | 100.0% |
| 2.00 Independent | TWOPVOTE | 1.00 Clinton | | 71 | 47 | 31 | 149 |
| | | | | 65.7% | 54.0% | 47.7% | 57.3% |
| | | 2.00 Dole | | 37 | 40 | 34 | 111 |
| | | | | 34.3% | 46.0% | 52.3% | 42.7% |
| | | Total | | 108 | 87 | 65 | 260 |
| | | | | 100.0% | 100.0% | 100.0% | 100.0% |
| 3.00 Republican | TWOPVOTE | 1.00 Clinton | | 21 | 12 | 8 | 41 |
| | | | | 15.4% | 9.4% | 9.2% | 11.7% |
| | | 2.00 Dole | | 115 | 115 | 79 | 309 |
| | | | | 84.6% | 90.6% | 90.8% | 88.3% |
| | | Total | | 136 | 127 | 87 | 350 |
| | | | | 100.0% | 100.0% | 100.0% | 100.0% |

Consider the Democrats first. The relationship between financial condition and voting choice is much weaker than the original bivariate relationship. This is indicated by the smaller differences in the proportion voting for Clinton across the three categories of V056: 97% of

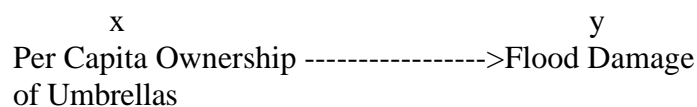
Democrats who were better off voted for the president, and the support for Clinton dropped only slightly to about 92% for the other two categories. There is a very weak Tau-C correlation (.05; significance = .039), which indicates that the correlation is very close to zero (although it is significantly different from zero). Among independents, the relationship is stronger, which is suggested by the monotonic decline in support for Clinton as financial condition varies from better to same to worse (Tau-C = .161; significance = .014), while among Republicans the relationship is not significantly different from zero (Tau-C = .059; significance = .120). The fact that the relationship is strongest among independents makes sense since independents do not have partisan predispositions which might distract them from voting on the basis of their financial interests. In any event, we have found that the relationship between our original two variables has changed rather significantly when we control for party identification.³

One final point from this table is worth noting. You can see that Democrats, whatever their financial status, were most likely to vote for Clinton, and Republicans were least supportive of the president, regardless of their financial status. This pattern reflects the fact that party identification has a powerful effect on voting choice. It is an effect that can easily be seen because each of the three groups in Table 8 is made up exclusively of respondents with a common value on party identification. These respondents are the same ones we analyzed in Tables 6 and 7; Table 2 just incorporates all three variables explicitly into the analysis.

Spurious Relationships

Let us be more specific about the kinds of theoretical expectations which lead the analyst to control for a third variable. We have said the general rule is that one expects the relationship to change in some way when controls are imposed. Perhaps the most fundamental reason to control for a third variable is to test for a spurious relationship. The possibility of a spurious variable poses a threat to the causal interpretation of any bivariate relationship. This is why a correlation between two variables is consistent with a causal hypothesis, but the correlation in no way proves that a causal relationship exists.

Suppose we are interested in studying what causes flood damage in cities across the nation. We assign the project to our Legislative Research Bureau, and they develop the following hypothesis:



Before we can stop them, eager researchers from the Bureau collect the data necessary to test the hypothesis. They measure the ownership of umbrellas in a sample of cities and they measure the amount of flood damage each city has experienced. These data are dutifully entered into the computer, and a crosstabulation of the dependent variable, flood damage with the independent

³The control for party identification is imperfect because we have collapsed the original seven-point variable (V008) into three categories. This means that some of the relationship we observe between financial condition and voting choice could still be due to variation in the strength of party identification.

variable, per capita ownership of umbrellas, yields the following impressive statistical relationship:

Table 9. Relationship between Umbrella Ownership and Flood Damage (Hypothetical data)

| | | Umbrella Ownership | |
|--------------|------|--------------------|------|
| | | Low | High |
| Flood Damage | Low | 82% | 14% |
| | High | 18% | 86% |
| | | 100% | 100% |

Tau B = .72

Obviously there is a substantial correlation between umbrella ownership and flood damage, but is there a causal relationship? Of course not. The reason there is a correlation is because the theory leading to the hypothesis is in error. There really is a correlation between the two variables, but it is spurious. That is, both umbrella ownership and flood damage are caused by the same third variable which is not included in the analysis. This variable, of course, is the amount of rainfall:

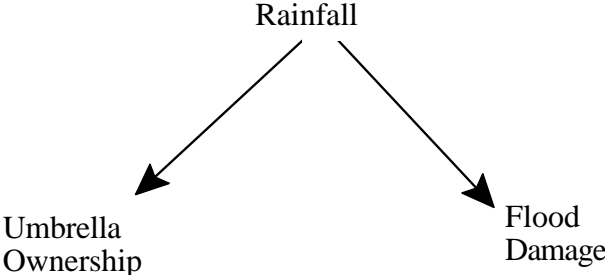
[Figure 2 here]

This new theory is much more plausible than the original, and it shows that the alleged independent variable, umbrella ownership, is in fact a dependent variable. Umbrella ownership and flood damage are correlated with one another because they are both caused by the same independent variable, the amount of rainfall.

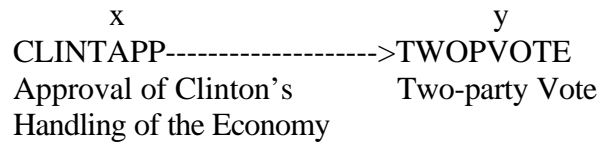
What to do? One thing we should do is be careful not to state stupid theories. But that advice will only take us so far. Not every theory that is wrong is stupid, and many times we must worry about possible spurious effects without necessarily believing the relationship is completely spurious. Or, we may need to convince ourselves and others that a relationship of interest is not tainted by some hidden variable that is producing the effect we observe.

In continuing our research on the effects of the relevance of the incumbent's economic policy on voting choice we might well argue that what is critical is approval or disapproval of the way the president has handled the national economy. Thus, those who approve of the way

Figure 2. Spurious Relationship between Umbrella Ownership and Flood Damage



the president handled the economy vote for him, those who disapprove voted against him. There is a variable in the data set that measures respondents' approval of Clinton's handling of the economy (a dichotomized version of V046). So our hypothesis is:



When we crosstabulate these variables, we find a strong relationship:

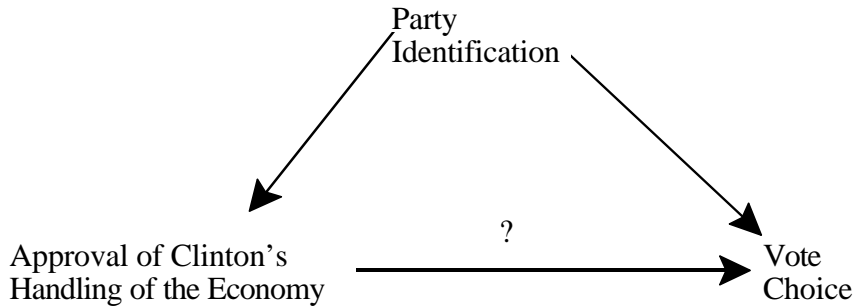
Table 10. Relationship between Approval of Clinton's Handling of the Economy and Vote Choice
TWOPVOTE * CLINTAPP Crosstabulation

| | | CLINTAPP | | Total |
|----------|--------------|---------------|-----------------|----------------|
| | | 1.00 approve | 2.00 disapprove | |
| TWOPVOTE | 1.00 Clinton | 541 78.6% | 40 12.5% | 581 57.6% |
| | 2.00 Dole | 147 21.4% | 281 87.5% | 428 42.4% |
| Total | | 688 100.0% | 321 100.0% | 1009 100.0% |

Tau-B = .624; Approximate significance = .000.

You may be impressed with the strength of the relationship in Table 10, but unconvinced by our reasoning. You may have reason to believe the relationship is either completely or partially spurious. You know that party identification has an influence on the vote. And you may think it also has an influence on whether people approve or disapprove of Reagan's handling of the economy. Democrats may have given Clinton the benefit of the doubt and said that they approved his handling of the economy, whereas Republicans may have concluded that Clinton was not responsible for the healthy economy (perhaps they gave credit to the Republican-controlled Congress). Your theory, then, is very similar to the rainfall example above:

Figure 3. Possibly Spurious Effects of Party Identification on Presidential Approval and Vote Choice



In order to test your theory, you would first show that party identification has an effect on vote for president (see Table 7) You would also need to show that party identification has an effect on respondents' rating of how well Clinton had handled the economy. This relationship is presented in Table 11 below:

Table 11. Relationship between Party Identification and Approval of Clinton's Handling of the Economy

CLINTAPP * TRIPTYID Crosstabulation

| | | TRIPTYID | | | Total |
|----------|-----------------|---------------|------------------|-----------------|----------------|
| | | 1.00 Democrat | 2.00 Independent | 3.00 Republican | |
| CLINTAPP | 1.00 approve | 507 88.5% | 291 62.2% | 179 42.0% | 977 66.6% |
| | 2.00 disapprove | 66 11.5% | 177 37.8% | 247 58.0% | 490 33.4% |
| Total | | 573 100.0% | 468 100.0% | 426 100.0% | 1467 100.0% |

Tau-C = .417; Approximate significance = .000.

Indeed, just as your theory expects, party identification has an effect on voting choice and on whether people approved of Clinton's handling of the economy. Table 11 shows that Democratic identifiers were much more likely to approve of Clinton's handling of the economy than were Republican. As usual, independents fell in between.

But to show the relationship is spurious, you must take one additional step. You must control for the potentially spurious variable. If the relationship between approval of Clinton and voting choice is spurious because the positions respondents take on both variables are caused by party identification, the bivariate correlation should disappear when party identification is held constant. Thus, in the rainfall example, if we examined the relationship between flood damage and umbrella ownership among cities with equal amounts of rainfall, no correlation would exist. This is because the covariation between umbrella ownership and flood damage depends upon both variables' covarying with rainfall. When rainfall is not permitted to vary (when it is held constant), the relationship disappears.

Table 12. Relationship between Clinton Approval and Vote Choice Controlling for Party Identification

TWOPVOTE * CLINTAPP * TRIPTYID Crosstabulation

| TRIPTYID | | CLINTAPP | | Total |
|------------------|--------------|---------------|-----------------|---------------|
| | | 1.00 approve | 2.00 disapprove | |
| 1.00 Democrat | 1.00 Clinton | 375 96.6% | 19 63.3% | 394 94.3% |
| | 2.00 Dole | 13 3.4% | 11 36.7% | 24 5.7% |
| | Total | 388 100.0% | 30 100.0% | 418 100.0% |
| 2.00 Independent | 1.00 Clinton | 125 75.3% | 22 25.3% | 147 58.1% |
| | 2.00 Dole | 41 24.7% | 65 74.7% | 106 41.9% |
| | Total | 166 100.0% | 87 100.0% | 253 100.0% |
| 3.00 Republican | 1.00 Clinton | 41 30.8% | | 41 12.1% |
| | 2.00 Dole | 92 69.2% | 205 100.0% | 297 87.9% |
| | Total | 133 100.0% | 205 100.0% | 338 100.0% |

As Table 12 shows, the relationship between vote choice and approval of Clinton's handling of the economy does not disappear when party identification is controlled. The relationship is perhaps a bit weaker among the three partisan groups, which lends some credence to the argument that party identification has something of a spurious effect. But it is clear that the effect of approval remains. Among Democrats, majorities of both those who approved and disapproved of Clinton voted for the president, but the difference was about 33% (Tau-B = .37; Significance = .001). Among independents, the effect remains very strong (a 50% difference in voting for Clinton, Tau-B = .48; significance = .000). Among Republicans, 31% of those who approved of Clinton voted for him, while none of those who disapproved voted for the president (Tau-B = .46; significance = .000). We can also see that party identification had an effect on voting since Democrats, whatever their opinion about Clinton's handling of the economy, were more

likely to vote for him than were Independents or Republicans.

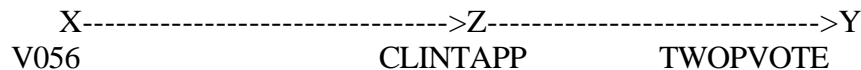
Since our original hypothesis survived the argument that it was spurious, our contention that the relationship is a causal one has gained credibility. We still have not proved it. Nor can we prove it, because there could always be some spurious variable – some equivalent of rainfall – that we have not controlled for or even thought of. It is important for us to be as careful as we can in thinking about other factors that might undermine our original causal understanding, and to include them in our analysis as further tests of the logic of our causal analysis.

Intervening Variables

Spurious variables present a threat to a causal interpretation of any statistical relationship. We control for potentially spurious variables because if these variables truly have spurious effects, they would destroy or undermine our theory. When a relationship survives the threat of a plausible spurious variable – when, in other words, the relationship continues to be strong – the original theory is supported. But it is support of a negative sort. The theory has survived an alternative explanation.

A theory which can posit an intervening variable as part of its explanation and subjects that explanation to a test provides additional evidence in support of the underlying argument. If the evidence supports the argument for an intervening variable, this can be rather powerful backing for the original, causal argument. Using our continuing example, let us see how this works.

Our original hypothesis was that voters' financial condition has an impact on how they vote. We found a weak but significant relationship between vote choice and financial condition, which we took as evidence in support of the argument. We can make our original argument more complete by specifying an intervening variable between our independent and dependent variables which is part of a causal chain. In other words, we can indicate more precisely *why* we believe that presidential voting choice reflects personal financial condition. The reasoning might be as follows: "Presidential voting choice reflects personal financial condition because people whose financial condition has deteriorated blame the president by disapproving of his handling of the economy. Those whose financial condition has improved approve of the president's handling of the economy." We can diagram this theory as follows:



This theory states that presidential voting choice is influenced by personal financial condition because of the intervening effect of the rating of Clinton's handling of the economy. That is, for financial condition to cause voting choice, it must first cause people to approve or disapprove of the president's handling of the economy. Approval of the incumbent's economic policy, in turn, is what causes people to vote for or against him.

Two of the steps necessary to test this theory have already been taken. These are, (1) Test for a relationship between x and y. (This has been done, and we found a weak relationship in

Table 6. The Tau-C correlation was .17.) and, (2) Test for a relationship between z and y. Part of this theory, in effect, treats CLINTAPP as the independent variable. (This has been done in Table 10. We found a strong relationship between z and y. Tau-B was .62).

Two additional steps are required to test for an intervening effect: (3) We must treat z as the dependent variable, and examine the data for a relationship between it and the independent variable. We cannot argue that $x \rightarrow z \rightarrow y$ unless there is evidence that x causes z. Table 13 presents the relevant crosstabulation, and finds a relationship. Those whose financial condition improved were noticeably more likely to approve of Clinton's handling of the economy than were those whose financial condition had deteriorated. As our theory expects, then, there is a tendency for people to lay the credit/blame for their financial status at the feet of the incumbent president.

Table 13. Relationship between Financial Condition and Approval of Clinton's Handling of the Economy.

| CLINTAPP * V056 Better Off Than Last Year? Crosstabulation | | V056 Better Off Than Last Year? | | | Total |
|--|-----------------|---------------------------------|-----------|------------|--------|
| | | 1 1. Better | 2 2. Same | 3 3. Worse | |
| CLINTAPP | 1.00 approve | 501 | 307 | 176 | 984 |
| | | 75.2% | 67.3% | 49.7% | 66.7% |
| | 2.00 disapprove | 165 | 149 | 178 | 492 |
| | | 24.8% | 32.7% | 50.3% | 33.3% |
| Total | | 666 | 456 | 354 | 1476 |
| | | 100.0% | 100.0% | 100.0% | 100.0% |

Tau-C = .207; Approximate significance = .000.

The final step to test for an intervening effect involves multivariate analysis: (4) Control for the intervening variable. If z intervenes between x and y, the relationship between x and y will disappear when z is held constant. Understand that a strict intervening variable (as depicted by our theory above) means that the effect of x on y depends on x first causing z which in turn causes y. Thus, if z is not permitted to vary (if it is held constant, or controlled) there is no way x and y can covary. That is why the correlation between x and y disappears when z is controlled.

We present the relationship between vote choice and financial condition controlling for the hypothesized intervening variable, approval of Clinton, in Table 14. Among those who approved of Clinton's handling of the economy, there is no relationship between financial condition and the vote, just as the theory predicts. Those who were better off were no more likely to vote for Clinton (82% than those who were worse off (82%)) (Tau-C = .06; Approximate significance = .094). Among those who disapproved of Clinton's performance, there is actually an apparent (but not significant) tendency for people who were worse off to vote for Clinton (17%) more than those who were better off (10%). In other words, the tendency of those who were better off to vote for the incumbent president does not hold among people whose approval of Clinton's handling of the economy is constant. As the theory expects, the reason financial condition affects vote choice is because it first affects approval, which in turn affect vote choice.

Table 14. Relationship between Financial Condition and Vote Choice Controlling for Approval of Clinton

TWOPVOTE * V056 Better Off Than Last Year? * CLINTAPP Crosstabulation

| CLINTAPP | | V056 Better Off Than Last Year? | | | Total |
|-----------------|-----------------------|---------------------------------|-----------|------------|--------|
| | | 1 1. Better | 2 2. Same | 3 3. Worse | |
| 1.00 approve | TWOPVOTE 1.00 Clinton | 291 | 165 | 85 | 541 |
| | | 82.2% | 71.7% | 81.7% | 78.6% |
| | 2.00 Dole | 63 | 65 | 19 | 147 |
| | | 17.8% | 28.3% | 18.3% | 21.4% |
| | Total | 354 | 230 | 104 | 688 |
| | | 100.0% | 100.0% | 100.0% | 100.0% |
| 2.00 disapprove | TWOPVOTE 1.00 Clinton | 10 | 10 | 20 | 40 |
| | | 9.8% | 9.6% | 17.4% | 12.5% |
| | 2.00 Dole | 92 | 94 | 95 | 281 |
| | | 90.2% | 90.4% | 82.6% | 87.5% |
| | Total | 102 | 104 | 115 | 321 |
| | | 100.0% | 100.0% | 100.0% | 100.0% |

Summary

When we introduce more complexity into our theories, we must also learn to introduce more complexity into our analysis. We must confront the evidence at a level of complexity that approximates our thinking about the world. Multivariate methods of analysis allow us to do this. Multivariate analysis permits us to state more precise and more complete conditions which support our theory. When the evidence is consistent with our conditions, the theory has more credibility. Multivariate analysis also serves the crucial purpose of allowing us to examine competing explanations for the relationships we observe. A statistical relationship does not prove a causal relationship between the independent and dependent variables. The relationship could be a spurious one. We cannot test for possible spurious effects without the multivariate technique of controlling for a third variable.

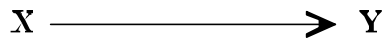
The rules we have discussed in this part are admittedly difficult to grasp when they are first encountered. You should review the examples in this handout and be sure you understand why the steps were taken, and how the conclusions from the evidence were drawn. To help you out, we provide a summary of the relationships we have examined, and the rules or steps you should follow in testing the hypotheses implied. These rules are summarized on the following page.

Key Concepts

- Bivariate analysis
- Multivariate analysis
- Alternative cause
- Control variable
- Spurious variable
- Intervening variable

SUMMARY RULES FOR QUANTITATIVE ANALYSIS OF CAUSAL HYPOTHESES

Bivariate Relationship

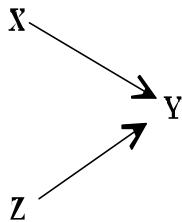


Hypothesis: **X** causes **Y**.

Test: Crosstabulate **X** with **Y**.

Expectations: Significant statistical relationship between **X** and **Y**.

Independent Cause

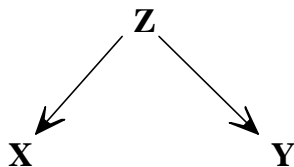


Hypotheses: **X** causes **Y**; **Z** causes **Y**. **X** does not cause **Z**; **Z** does not cause **X**.

Test: Crosstabulate **X** with **Y** and **Z** with **Y**.

Expectation: Both **X** and **Z** are significantly related to **Y** (but **X** and **Z** are *not* related to each other).

Spurious Effect



Spurious Effect (continued)

Hypotheses: **X** is correlated with **Y**, but does not cause **Y**. **X** and **Y** are correlated because both are caused by **Z**.

Test: Crosstabulate **Z** with **X**, and **Z** with **Y**. Crosstabulate **X** with **Y** controlling for **Z**.

Expectations: **Z** and **X** and **Z** and **Y** are significantly related. The relationship between **X** and **Y** disappears with **Z** is controlled.

Intervening Variable



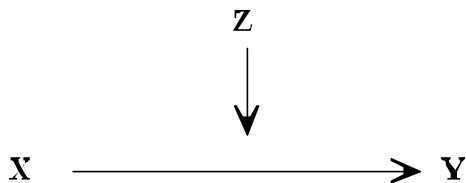
Hypotheses: **X** causes **Y**

because **X** causes **Z**, which causes **Y**.

Test: Crosstabulate **X** with **Y**, **X** with **Z**, and **Z** with **Y**. Crosstabulate **X** with **Y** controlling for **Z**.

Expectations: **X** correlates with **Y**; **X** correlates with **Z**; **Z** correlates with **Y**. The correlation between **X** and **Y** disappears when **Z** is controlled.

Interaction Effect



Hypothesis: The causal effect of **X** on **Y** depends on the value of **Z**.

Test: Crosstabulate **X** and **Y**. Crosstabulate **X** and **Y** controlling for **Z**. (Note: Requires multiplicative term in regression analysis.)

Expectation: The correlation between **X** and **Y** varies in strength depending on the value of **Z**.