

P R E F A C E

Researchers in the social sciences often require reference books to aid them in the computation and interpretation of statistics. These books are usually organized around a set of statistical tools and give extensive detail to the formulas used in estimation. Researchers also use computer packages to compute these statistics. A simple, one-line command will give birth to multiple regression, factor analysis, or analysis of variance.

On the bookshelves next to the statistical texts and computer manuals are the books that summarize and interpret the field. These texts of substantive theory may occasionally present statistical summaries of data, but data play a small role in such texts. To some extent, theory and data rarely touch each other in the social sciences. To bridge the gap, over the past ten or so years, an area called causal or structural analysis has developed. This area takes as its task the explication of a statistical method in terms of how the method relates to substantive theory.

This text is a general introduction to the topic of structural analysis. It is an introduction because it presumes no previous acquaintance with causal analysis. It is general because it covers all the standard, as well as a few nonstandard, statistical procedures. Since the topic is structural analysis, and not statistics, very little discussion is given to the actual mechanics of estimation. Do not expect to find computational formulas for various statistical methods. One should consult the standard sources if interested in them. Moreover, it is presumed the reader has some familiarity with the two standard multivariate methods of multiple regression and factor analysis. The emphasis is not on the mechanics of a statistical technique but rather its structural meaning.

I have attempted to present the material in an informal style. Instead of discussing the general case I have chosen small specific examples. Also, to keep the algebra simple, I have not employed

matrix algebra. This has made parts of the discussion accessible to the beginner, though it may seem somewhat awkward to the advanced student. I apologize if at times you feel as if you are being talked down to. Please try to sympathize with the concern to write a book that has something to say to both beginner and expert.

I would suggest that you try to read the book with someone else. The corrective nature of dialogue can prevent misapplication better than the addition of another 100 pages. I must warn the reader that a number of sections of the book are not easy to read. I would suggest reading the book with a pencil and calculator in hand. If you take the time to puzzle through various difficult sections, I expect you will find the result to be rewarding.

I have included many examples taken from interesting data sets—for example, the accuracy of psychics, Vietnam protest, concerns of dying cancer patients, and productivity of scientists—which I believe will facilitate the readers' own applications of the methods described in this book. In many cases, however, the models proposed for the data do not fully exploit the structure of the data. The examples then are not meant to demonstrate the successful application of the procedures, but rather the mechanics. Successful applications require full chapters, not two or three pages. Since I provide where possible the full correlation matrix, the reader is invited to fit alternative models.

The text is divided into four major sections, each with three chapters. The first three chapters are introductory: Chapter 1 discussing causal analysis and its role in the social sciences, Chapter 2 presenting a set of simple rules for manipulating covariances, and Chapter 3 introducing path analysis and defining key concepts applied in the remaining chapters.

Chapters 4 to 6 discuss classical econometric methods for estimation of structural parameters. Chapter 4 considers models for which multiple regression analysis estimates causal coefficients; Chapter 5 considers error of measurement in causes and unmeasured third variables; and Chapter 6 briefly introduces feedback models,

Chapters 5 and 6 using two-stage least squares to estimate causal parameters.

The third section considers in detail models with unmeasured variables. Chapter 7 discusses models with a single unmeasured variable and multiple indicators, whereas Chapter 8 allows for multiple unmeasured variables and Chapter 9 details causation between unmeasured variables. All the models discussed in Chapters 7 to 9 are factor analysis models, and maximum likelihood estimation is recommended.

Chapters 10 through 12 consider the application of correlational methods to experimental and quasi-experimental designs. Chapter 10 considers a causal modeling approach to experimentation, as well as additional correlational analyses that could be performed; Chapter 11 considers the analysis of the nonequivalent control group design; and Chapter 12 considers cross-lagged panel correlation analysis. The final chapter of the book, Chapter 13, ties together the loose ends.

A beginning student should read Chapters 2, 3, 4, 7, and the first section of Chapter 10. An advanced student should read the remaining chapters but should concentrate on 5, 8, and 9. I would recommend the following chapters for a course on correlational methods: Chapters 1, 2, 7, 8 (section on the multitrait-multimethod matrix), and 12. For a course on survey analysis, I would recommend Chapters 2 through 9. For a course on experimentation, I would recommend Chapters 2, 3, 4, 7, 8, 9, 10, and 11.

To some purists I may seem to be rather careless. First, I have often unnecessarily limited generalization by employing standardized variables. I did so because I did not want to confuse the beginner. I clearly state the limits of standardization in Chapters 3 and 13. Second, at some points the distinction between population value and estimate is blurred. I did so in order not to have a text filled with distracting symbols. If these practices disturb you, I apologize. But I felt that if I had to sacrifice elegance for the experts in order to obtain clarity for the beginner, I would choose clarity.

A project like this represents not the work of one, but of the community of scholars, and this is especially true of this text. Most of what is contained represents the contributions of others. I would like to thank three persons in particular: first, Don Campbell, who pointed me in this direction and redirected me when I was getting lost; whatever contributions there are in this text are his and not mine; second, Steven H. Cohen, who suffered with me, pondering many a sentence and analysis; I only wish I were beginning the project now so that he and I could collaborate in it; third, Pierce Barker, who encouraged and prodded me to finish the project; moreover, he provided detailed feedback for a number of chapters.

Countless others provided helpful feedback throughout the project. Those who occur to me now are Charles Judd, Dean Simonton, Allan Lundy, Jeffrey Berman, William McGarvey, Judith Harackiewicz, Louise Kidder, Lawrence La Voie, Mary Ellen Kenny, Michael Milburn, Reid Hastie, James Garvey, and Susan Fiske. Special thanks are due to Alice Mellian and Mary Ellen Kenny, who assisted me in the preparation of the book.

David A. Kenny

Storrs, Connecticut

January 1979

P R E F A C E

T O T H E R E V I S E D E D I T I O N

The book *Correlation and Causality* has been out of print for over a decade. Many have encouraged me to write a second edition of the book. However, with the publication of many excellent new books (e.g., Kline, 1998), such an effort seems unnecessary. The book still continues to be cited and its approach, though clearly dated, seems to be of use to many, if only as an historical document. So a revised edition may be of interest to some. I would ask people if they could continue to cite the original edition and not this revised edition.

The revised edition is essentially the same as the original 1979 book. I have corrected the errors that I knew of. I have added a few new references. However, in no way this revised edition should be considered an updated version of the manuscript. So as an example, the book does not employ the contemporary convention of denoting latent variables by having a circle around the name.

One thing that I have changed is that the original edition used the population symbol for correlation, ρ , at times when the sample symbol, r , should be used.

In preparing this preface, the saddest part was my noting of the passing of several persons who were very helpful in the preparation of the book. I want to especially note the deaths of my two parents to whom the book was dedicated and the deaths of Donald T. Campbell and Pierce Barker. Also sad is the fact that some of those who were very helpful in the preparation of the book I am no longer in contact. Despite the passage of more than a quarter of a century, I remain most grateful to all who helped.

I want to thank my two daughters, Katherine and Deirdre Kenny who have together done most of the work in preparing this revision. They were the ones who scanned the original version with Deirdre doing most of the work. It was Katherine who was mainly responsible for the design of the book, much of it done long distance in

viii CORRELATION AND CAUSALITY

Prague. Thanks are also due to Virginia Carrow, Marina Julian, and Matthew Beloin who assisted in the preparation.

If you have suggestions or find typographical errors, please send me an email at davken@rcn.com

David A. Kenny

Storrs, Connecticut
March 2004

BOOK
DEDICATION

ORIGINAL EDITION

Mary Ellen and My Parents

REVISED EDITION

Virginia Carrow

1

CORRELATIONAL
INFERENCE

Given the old saying that "correlation does not imply causation," one might wonder whether the stated project of this book—correlational inference—is at all possible. Correlational inference is indeed possible through the application of standard multivariate statistical methods to a stated structural model. Because a great amount of both confusion and controversy center around the terms "correlational" and "inference," it is first needed to have a common understanding about the meaning of each.

First, *correlational* means a statistical relationship between a set of variables, none of which have been experimentally manipulated. Although correlations and covariances can be computed from experimental data, we usually reserve the term *correlation* for a relationship between unmanipulated variables. Very often, the random assignment of units to treatment conditions, the backbone of experimental inference, is impossible and there is only correlational data. In such a case, causal inference may still be the goal.

Second, *inference* means confirmation or disconfirmation of a scientific hypothesis by the use of data. To interpret data, there must be a set of reasonable assumptions about how the data were generated and additional assumptions about how the data can be summarized. The set of assumptions about how the data were generated is usually called the *model*, and data are summarized by *statistical methods*. For at least two reasons, an inference does not imply in any way that the hypothesis is proven to be true. First, the inference is usually evaluated statistically: at best one knows only the probability the results that would have obtained by chance given the model. There is no certainty, only a

probability. Second, and more importantly, the model or assumptions on which the inference is based can always be questioned. Every inference but *cogito ergo sum* is based on a set of assumptions. One cannot ascertain simultaneously the validity of both the assumptions and the inference. Modern epistemology states that proof is a goal that is never achieved by social scientists or any scientist for that matter. As the ancient Hebrews felt about their God, the scientist should never speak the words *truth* or *proof* but always keep them in mind.

Quite clearly the strength of inference is very different for confirmation than for disconfirmation. Especially for correlational inference, disconfirmation is usually more convincing than confirmation. A disconfirmation implies that the data are incompatible with the hypothesis. A confirmation shows the opposite, that the data are compatible with the hypothesis. However, the data also normally confirm a host of alternative inferences. It shall be seen that the confirmation process in correlational inference can be greatly strengthened by having the hypothesis state not only what should happen with the data but also what should not.

Confirmatory inference is strong then, only if there are no plausible rival explanations of an effect. Campbell and Stanley (1963) in their classic text show how randomized and quasi-experimental designs may rule out a set of stated plausible rival hypotheses. Although the exact list of rival hypotheses in Campbell and Stanley does not fit very well with correlational designs, the core ideas of that text are extremely helpful in evaluating the strength of correlational inference. For instance, one important idea of Campbell and Stanley is the necessity of tests of statistical significance. Before a relationship can be interpreted, it must be demonstrated that it is not plausibly explained by chance.

The term *correlational inference* should not be taken to mean that various statistics are by themselves inferential. Regression coefficients, factor loadings, and cross-lagged correlations do not, in and of themselves, have an inferential quality. Given a plausible model, a statistic can be used for inferential purposes, but the statistic

itself is merely a passive tool. Inference goes on in the head of the researcher, not in the bowels of the computer.

CAUSATION

There is one particular type of inference that we often wish to make from correlational data: a causal inference. For many years both scientists and statisticians were reluctant to even say the word “causation.” Judea Pearl (2000), almost single-handedly, has returned the concept of causality to the parlance of scientific discourse. A causal statement, to no one's surprise, has two components: a cause and an effect. Three commonly accepted conditions must hold for a scientist to claim that X causes Y :

1. time precedence
2. relationship
3. nonspuriousness

For X to cause Y , X must precede Y in time. Such time precedence means a causal relationship is asymmetric. To see this let X cause Y with a lag in time, and we then have, X_t causes Y_{t+k} where the subscript refers to time with $k > 0$. Note that Y_{t+k} cannot cause X_t since this would violate time precedence. (It is true, however, that Y_t could cause X_{t+k} .) Causal relationships are then fundamentally asymmetric while many statistical measures of relationship are symmetric. Implicit in a causal vocabulary is an active, dynamic process that inherently must take place over time.

There is no obvious logical objection to instantaneous causation (although it would be difficult to observe), but philosophers of science have not chosen to assume it. In a sense, something like causation backwards in time happens with rational foresighted creatures. Suppose an individual named John examines the current economic conditions and sees that a depression is inevitable. John then makes a series of economic decisions to cushion himself from the

impending effects of the depression (if that is at all possible). One might argue that depression that occurs after John's decisions causes these prior decisions. We can easily see the fallacy in this line of reasoning. It is not the depression per se that causes the decision, but the perception of the impending depression. Like any perception it may not be veridical. If we allowed for the veridical perception of the future (prophecy), it would be empirically impossible to rule out backwards causation.

The second condition for causation is the presence of a functional relationship between cause and effect. Implicit in this condition is the requirement that cause and effect are variables, that is, both take on two or more values. For example, the United States Surgeon General tells us that cigarette smoking causes lung cancer. The statement that "smoking causes lung cancer" is actually shorthand for saying that, other things being equal, smoking increases the probability of lung cancer, over not smoking. Thus, to elaborate "smoking causes cancer" we must define two variables: smoking—presence or absence—and lung cancer—again presence or absence. To measure a relationship between two variables, first it is necessary to define the meaning of no relationship between variables or, as it is sometimes called, independence. Two variables are *independent* if knowing the value of one variable provides no information about the value of the other variable; more formally, X and Y are independent when the conditional distribution of Y does not vary across X . If variables are not independent, then they are said to be related.

In judging whether two variables are related, it must be determined whether the relationship could be explained by chance. Since naïve observers are very poor judges of the presence of relationships, statistical methods are used to both measure and test the existence of relationships. Statistical methods provide a commonly agreed upon procedure of testing whether a sample relationship indicates a relationship in the population.

The third and final condition for a causal relationship is nonspuriousness (Suppes, 1970). For a relationship between X and Y to be nonspurious, there must not be a Z that causes both X and Y such

that the relationship between X and Y vanishes once Z is controlled. A distinction should be made here between a spurious variable and an intervening or mediating variable. Variable Z intervenes between X and Y if X causes Z and Z in turn causes Y . Controlling for either a spurious variable or an intervening variable makes the relationship between X and Y vanish; but while a spurious variable explains *away* a causal relationship, an intervening or mediating variable *elaborates* the causal chain. Many analysts see the issue of spuriousness as the biggest stumbling block in causal analysis, and it has been variously called the third variable problem, the excluded variable, common factoredness, and cosymptomatic relationship. A commonly cited example of spuriousness is the correlation of shoe size with verbal achievement among young children. The relationship is spurious because increasing age causes increasing shoe size and verbal achievement. Spuriousness usually plays a much more subtle role. Much of the discussion of this book centers around the problem of spuriousness.

Besides the three formal requirements for causality of time precedence, relationship, and nonspuriousness, there is perhaps a fourth condition for causation that is difficult to state precisely. This fourth condition is that causality implicitly implies an active, almost vitalistic, process. It is difficult to convey this notion of causality formally just as it is to formally define space or time. The philosopher Immanuel Kant has called causality, along with space and time, a *synthetic a priori*, that is, an idea that we bring to our experience of phenomena. This distinctively human bias toward causal explanations has recently become a central topic among social psychologists who study the attribution process.

THE NEED FOR CAUSAL MODELING

Although causal analysis may pervade our everyday life, it is controversial whether causal analysis is an appropriate concern of the social sciences. It is the contention here that causal modeling should have a central position within social research, although there are other very important tasks in the social sciences such as observation,

measurement, data reduction, and theory formulation. For at least three reasons causal modeling needs to be applied to social science research:

1. Because most researchers either implicitly or explicitly construct models, a formal development of the method would assist these researchers.
2. Causal modeling can assist the development, modification, and extension of measurement and substantive theory.
3. Causal modeling can give social science a stronger basis for applying theory to solving social problems.

Beginning with the first point, since causal modeling is already being employed, an increased understanding of the process would aid researchers. Most researchers find themselves with large data sets to which they apply statistical methods. Any such method implies a statistical model. Common models with which most researchers are familiar are the analysis of variance, multiple regression, and factor analysis models. Very often a statistical model can elegantly and simply summarize the data. Although the fit of the statistical model may satisfy the curiosity of the statistician, it only whets the curiosity of the social scientist since most social scientists gather data to test substantive theory. They invest their egos and reputations in theory, not in statistical models.

Usually the skeleton of a theory is a string of causal statements, although as is later discussed, the heart of any theory is a metaphor or image. Experiments and surveys are conducted to test a set of causal statements, which are usually called hypotheses. The formation of scientific hypotheses is guided by theory and not by a statistical model. Consider an example. A theory of deviance states that deviance causes social rejection and social rejection in turn causes alienation. It then makes sense to control for social rejection if we are to measure the impact of deviance on alienation. Although it makes theoretical nonsense to control for alienation in estimating the effect of deviance on social rejection, it is perfectly permissible to control for

alienation statistically in estimating the relationship between deviance and social rejection. Many of the serious errors in data analysis are due not to lack of knowledge of statistical methods, but to a failure to apply the appropriate statistical method given the conceptual problem. Levin and Marascuilo (1972) have called such errors, Type IV errors.

The ready availability of computer packages contributes to this conceptual slippage between the idea and data analysis. Users can quickly and cheaply apply factor analysis, multiple regression, and analysis of variance to their data. Even though users may understand the statistical assumptions necessary to apply these techniques, they may not fully comprehend the conceptual assumptions. One good example is factor analysis. Many are sophisticated in the ins and outs of rotation, but few have a solid understanding of the relationship of factor analysis to substantive theory. As seen in Chapter 7, one use of factor analysis is to test for unidimensionality, that is, the set of measures taps a single construct, but few researchers understand how to apply factor analysis to answer this important question. A better understanding of the process of causal modeling would help researchers choose the appropriate statistical method, use it in the correct way, and interpret it intelligently.

The second reason why causal modeling is important is that it can be used to more exactly state theory, to more precisely test theory, and then to more intelligently modify theory. Unfortunately most of what passes for data analysis seems more like a ritual than an investigation into the underlying process of how the data are generated. A researcher who approaches data from a modeling approach is somewhat more likely to learn something new from the data. Ideally the researcher starts with a model or formulates one. Then the researcher determines if the data to be analyzed can estimate the parameters of the model and if the data can falsify the model. Such estimation and testing reveal whether the model is too general, too simple, or just plain wrong. As is seen in Chapter 3, a researcher who carefully follows all the steps of causal modeling is in the position to test theory. Although causal modeling offers an exciting possibility for

researchers, it also clearly shows the limits of data analysis in resolving theoretical issues.

A third reason for causal modeling is that it can provide a scientific basis for the application of social science theory to social problems. If one knows that X causes Y , then one knows that if X is manipulated by social policy, *ceteris paribus*, Y should then change. However, if one only knows that X predicts Y , one has no scientific assurance that when X is changed, Y will change. A predictive relationship may often be useful in social policy, but only a causal relationship can be applied scientifically.

Even when a causal law is known, one must take care in applying it. For instance, although the scientific literature is viewed by some as not very supportive of psychotherapy as a cause of "mental health" (see Smith & Glass (1977) for a contrary view), it would be premature to replace it by more "scientific" methods. Often in conventional wisdom resides an implicit knowledge that surpasses the formal knowledge of science. It may be fortunate that political and practical considerations impede the application of social science to social problems. In absence of strong theoretical causal models, the practitioners such as social workers and clinicians deserve to be listened to more than the ivory tower academic. It is a sad commentary that the social scientist is often better skilled not in designing better social programs to solve social problems but only in evaluating the success or failure of existing social programs (Weiss, 1972).

LIMITS OF CAUSAL MODELING

Although causal modeling is important in the advancement of social science, it has very definite limitations:

1. The research and data must be grounded in a solid foundation of careful observation.

2. The central ideas or driving themes of theory are not usually causal laws but are more likely images, ideas, and structure.
3. Causal modeling is open to abuse.

Regarding the first point, models are built upon qualitative examination of phenomena. Adherents of participant observation often criticize the elaborate "laws" of statistical modelers, and argue that before a tradition can be developed in an empirical domain, a wealth of qualitative lore must be built up. This careful, almost clinical sensitivity to empirical phenomena is not easily taught or appreciated. Sadly, this sensitivity seems to be primarily emphasized in the study of phenomena of behaviors not easily amenable to the quantitative methods, for example, deviance. Careful observational methods form the basis for measurement (institutionalized observation), and measurement in turn is the basis for quantitative data. Let us not forget that the validity of measures in social science is usually nothing more than face validity and face validity is only based on common sense.

Too often, even in the early stages of research on a topic, the researcher is too far separated from the observation of the empirical process. Almost as often the researcher never sees the raw data but only highly aggregated statistical summaries. True enough, interviewing and observation by themselves can lead to mistaken and facile explanations of behavior, but they nonetheless provide a wealthy source of data, albeit difficult to characterize; and from such observations models and hypotheses can be formed.

Second, although theory takes the form of causal statements, the guiding ideas of theory are not those statements but rather an image or an idea. Many of the important ideas of natural science are not causal but are pictures of a process. Although evolution, the periodic table, and the kinetic theory of gases have a mathematical form, they are fundamentally images.

Some causal modelers seem to act as if the major problem facing an area is finding a key cause of the correct variable. For instance,

if a researcher is interested in drug use, research will then be defined as measuring all the possible causes of drug use and seeing which is the best predictor. Given finite samples and the multitude of "independent variables," to no great surprise, some variables are found to be significant. Even though most of these effects are usually obvious and unsurprising, elaborate explanations are found for these variables. Then there are marginal but "interesting" variables. How easy it is to delude oneself that one's study is the study to find the key variable. Of course, no other study can then replicate the key variable. Causal modeling provides no certain path to knowledge. In fact, causal models are maximally helpful only when good ideas are tested. Good ideas do not come out of computer packages, but from people's heads.

Third, like any tool causal modeling can be misused. There is no magic in the pages that follow. The techniques to be discussed—factor analysis and multiple regression—are already in common use. This book elaborates how to more carefully and consistently relate these techniques to theory. One cannot take bad data and turn it into gold by calling it a causal model. The potential promise of the method is that it puts the theoretical assumptions up front. Ideally methodological and statistical questions become clarified and disappear, and the focus of the argument is then turned to the substantive assumptions of the researcher.

There will be a temptation to dress up statistical analyses by calling them a "causal analysis." Although the term causal modeling sounds impressive, remember it forces the researcher to make stronger assumptions in order to make stronger conclusions.

CAUSATION AND FREEDOM

There is a final criticism of causal modeling that is put forth. To some, it is repugnant to think of persons being pushed and pulled by a set of causes external to them. Causal models of human behavior give rise to a vision of persons as marionettes pulled by the strings of some set of universal laws. Ironically the same people who object to social science as an enterprise, usually also have an unrealistic view

that social science can be used to manipulate human behavior to achieve evil ends. The fear that advertising or behavior modification can change people's lives is an implicit recognition that human behavior is indeed caused. A simple illustration that our behavior is caused is that, even though we are "free" to travel when and where we want, traffic patterns are even more predictable than the weather.

Although we cannot deny that human behavior is caused, our intuitions and traditions are still correct in telling us that people are free. Even highly developed causal models do not explain behavior very well. A good rule of thumb is that one is fooling oneself if more than 50% of the variance is predicted. It might then be argued that the remaining unexplained variance is fundamentally unknowable and unexplainable. Human freedom may then rest in the error term. The hard core determinist would counter-argue the error is potentially explainable, but at the present moment science lacks an adequate specification of all the relevant causes. Einstein took this approach in his arguments against physical indeterminacy. This, however, is only a conjecture and it seems just as plausible and no less scientific that human behavior is both caused and free. There is no logical or scientific necessity to argue that behavior is totally determined.

Equating unpredictability with freedom may seem peculiar, but Ivan Steiner (1970) in his discussion of perceived freedom defines decision freedom as the condition in which the decision choice is unpredictable. Choices made with no constraints are free. Unpredictability is only one of two parts of freedom. The person must also be responsible for the behavior. The velocity of a subatomic particle may be unpredictable, but it is hardly responsible for its position. Human beings are responsible for their behavior.

There may even be a biological basis for human freedom. The fuel of evolution, both biological and cultural, is variation (Campbell, 1974). There can only be evolution if there are a set of different behaviors from which the environment selects. Perhaps nature programmed into the human brain something like a random number generator. This random number generator creates behavioral

responses some of which are by chance adaptive. The organism then begins to learn which responses are adaptive to the situation. Culture, socialization, and values continually select from this pool of variable responses. For operant conditioning to work, one needs a variable organism to shape. If human beings are fundamentally variable and, therefore, at least partially unknowable, then traditional social science has erred in presuming that human action is totally determined. Sadly, social science has taken only one track. It has emphasized control and constraint instead of emphasizing freedom and possibility.

Causal modeling can be compatible with the preceding notion of human freedom. It can also allow us to understand the causes of our behavior so that we can transcend those causes. This transcendence can be viewed in both a weak or strong sense. The weak sense is that a causal relationship can be mitigated through intervention. For instance, if high employment causes inflation, the government can take steps to reduce inflation when employment is high. If the governmental action has the desired effect of reducing inflation and little or no effect on employment, it would then be possible to reduce or even reverse the relationship between employment and inflation. The causal relationship between employment and inflation would still exist, but the statistical relationship between the two variables would vanish. By the stronger sense of transcendence, it is meant that if we understand the rules of the game (i.e., causal relationships), we can decide to change not the variables, but the rules themselves. We may, for instance, decide that rules generated by a free enterprise economy should be abolished and we should play by a set of rules determined by a socialist economic structure. Only by understanding the ways in which we are determined, can we transcend those rules.

CAUSAL LAWS

Now that we have discussed the pluses and minuses of causal modeling, let us begin to discuss causal modeling itself. A causal law is of the general form:

For all Q , X causes Y .

The term Q refers to the set of objects or persons to whom the causal law applies, X refers to the cause, and Y the effect. Neglected in many causal models is the specification of Q . In social science Q is usually some subset of persons, although Q can also be the set of words, situations, generations, or nations. If Q is the set of persons, usually additional specifications are made. For instance, it is usually assumed that the person is an adult, is awake, is not mentally retarded, and is healthy. Normally these qualifications of the population are implicitly made. Often for a causal law to operate, certain important necessary conditions must hold. For instance, the person must be motivated, must have already learned a relevant behavior, or must attend to various stimuli. In this text the "for all Q " part of a causal statement is usually omitted. One should, however, avoid the reification of causal laws. As Sartre points out (1968, p. 178), causal laws do not operate in heaven but on people:

About 1949 numerous posters covered the walls in Warsaw: "Tuberculosis slows down production." They were put there as the result of some decision on the part of the government, and this decision originated in a very good intention. But their content shows more clearly than anything else the extent to which man has been eliminated from an anthropology which wants to become pure knowledge. Tuberculosis is an object of a practical knowledge: the physician learns to know it in order to cure it; the Party determines its importance in Poland by statistics. Other mathematical calculations connecting these with production statistics (quantitative variations in production for each industrial group in proportion to the number of cases of tuberculosis) will suffice to

14 CORRELATION AND CAUSALITY

obtain a law of the type $Y = f(x)$, in which tuberculosis plays the role of independent variable. But this law, the same one which could be read on the propaganda posters, reveals a new and double alienation by totally eliminating the tubercular man, by refusing to him even the elementary role of mediator between the disease and the number of manufactured products.

Although it is all too easy to do, let us not lose sight that causal laws in the social sciences refer to people.

Because both the cause X and the effect Y are variables, the relationship between them can be put into some functional form: $Y = f(X)$ called a structural equation. The typical functional form is a linear one:

$$Y = b_0 + b_1X$$

The b_1 term is called the causal parameter and its interpretation is straightforward. If the variable X were increased by one unit, the variable Y would be increased by b_1 units. The relationship is linear because if X were plotted against Y , there would be a straight line with slope b_1 and a Y intercept of b_0 . Instead of just one X causing Y , there may be a set of X s. These X s could be combined in a number of ways, but the simplest is an additive combination:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_nX_n$$

Each causal variable, the X s, is multiplied by its causal parameter, and these parameter-variable combinations are then summed. One could concoct other mathematical forms, but the linear model is one of the simplest.

Given a set of variables in a causal model, most if not all the information can be summarized in the covariances between variables. Statisticians call the covariances the sufficient statistics. Before discussing causal modeling any further, let us take a necessary detour and develop the algebra of covariances.

2

COVARIANCE ALGEBRA

The information about the set of parameters of a linear model, causal or otherwise, is all contained in the covariances between variables. To recover the model's parameters, covariances must be manipulated. Thus, to become proficient at causal modeling the researcher must learn how to determine the covariances between variables. A covariance occurs between a pair of variables, say X and Y , and is symbolized here by $C(X,Y)$. The term $C(X,Y)$ is read as the covariance between X and Y . If it is known that $X = aZ + bU$, then it is also known that $C(X,Y) = C(aZ+bU,Y)$, and so within a covariance a variable's equation can be substituted for the variable itself. Additionally a covariance is a symmetric measure; that is, $C(X,Y) = C(Y,X)$.

DEFINITION OF COVARIANCE

So far we have only given a symbol for covariance. Mathematically it is defined as

$$E[(X - \mu_X)(Y - \mu_Y)] \quad [2.1]$$

where E is the expectation or best guess and μ_X is the population mean of variable X and μ_Y of Y . Some texts refer to covariance as σ_{XY} . The unbiased sample estimate of a covariance is given by

$$\frac{\Sigma[(X - M_X)(Y - M_Y)]}{N - 1} \quad [2.2]$$

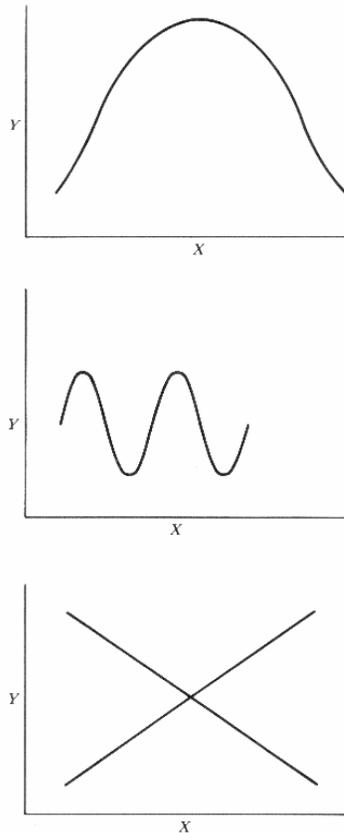
where observations are independently sampled and M_X and M_Y are sample means and N is the sample size. A common computational formula for covariance is $(\Sigma XY - \Sigma X \Sigma Y / N) / (N - 1)$. Covariance is a measure that is in the scale of both variables. If X is measured in feet and Y in yards, their covariance is in feet-yards. If the units of measurement of X or Y are changed, then the value of the covariance changes.

A covariance is a measure of association or relationship between two variables. The covariance is positive if a person who is above the mean on X is expected to be above the mean on Y . With a negative covariance the pattern is reversed: above the mean on X and below the mean on Y or below the mean on X and above the mean on Y . This relationship between X and Y is ordinarily not perfect but only a tendency. Thus if the covariance is positive usually not all the persons above the mean on X are above the mean on Y , but only the "average" person is. A zero covariance implies that being above the mean on one variable is in no way predictive of the other variable.

FACTORS INFLUENCING THE SIZE OF A COVARIANCE

Before moving on to the algebra of covariances, consider factors that may make covariance a misleading measure of association: outliers and nonlinearity. As shown by Equation 2.1, a covariance is the sum of cross-product deviations from the mean. Scores that are more deviant from the mean more heavily influence the covariance. In computing covariances researchers should check for outliers that could distort the covariances. An *outlier* is usually said to be three or more standard deviations from the mean. According to Anscombe (1968), there are three different explanations for outliers: (a) measurement errors, (b) multiple populations, and (c) random values.

Figure 2.1 *Nonlinear relations between X and Y: Curvilinear, cyclical, and crossover interaction*



Likely most outliers are due to a malfunction of the measuring instrument or an error in the transfer of data to a computer file. Some of the possible sources of measurement errors are that a respondent could misunderstand the instructions, the response could be mistakenly transposed onto coding sheets or the data analyst could make a data-entry error. There are indeed all too many sources of error. Errors are often detected by noting whether the obtained value is within the possible range of values. For instance, negative values for either

dollars earned and or for reaction time are ordinarily not allowable. For outliers that are errors, the correct value can often be recovered whereas in other cases this is impossible.

A second, and more problematic, source of outliers is multiple populations. The underlying distribution of the variable is bimodal with a second peak that is considerably different from the first. In a study on memory for phone numbers in which one person's performance is abysmally low, the researcher may discover that the person has some organic problem. The researcher should have screened the participants and excluded from the study those who had organic problems. In a study of income, if one of the sample respondents is a millionaire, this one person could considerably bias the conclusions of the study even if the sample size were large. This person's income may be a social error but it is not a measurement error. Researchers interested in unraveling causal relationships are usually well advised either to exclude from the study any highly deviant group or to oversample them for an additional separate analysis. If the interest is in estimating population parameters, it is unwise to exclude such units.

Finally, outliers are expected by chance. Given a sample size of 2000 and a normal distribution, about five persons would be three or more standard deviations from the mean and about one would be four or more. If an outlier cannot be attributed to measurement error or sampling from multiple populations, then it is reasonable to explain the outlier by chance. To reduce the effect of outliers on covariances, one might trim the distribution: Ignore the highest and lowest values or the two highest and lowest. Although trimming is not yet a common practice, statisticians are currently developing this strategy as well as other resistant methods (Tukey, 1977).

Using a covariance as a measure of association presumes that the functional relationship between variables is linear, that is, a straight-line functional relationship. For the examples in Figure 2.1, although there is marked functional association between X and Y , their covariance is zero. For all three examples, if the X s across each Y are

averaged, a vertical line is obtained, indicating no linear relationship and zero covariance. (Even though a line is obtained, a perfect vertical or horizontal line indicates zero covariance.) The first example in Figure 2.1 is a curvilinear relationship. A substantive example of a curvilinear relationship is the Yerkes-Dodson law stating that as motivation increases performance increases up to a point, then as motivation continues to increase performance decreases. We all know that moderate amounts of anxiety increase performance, but large doses are debilitating. The second example in Figure 2.1 is a cyclical relationship. A simple substantive example is one in which X is month of the year and Y is mean temperature for the month. Cyclical relationships are rather common when time is one of the variables. The final nonlinear example in Figure 2.1 is an interaction. The relationship between X and Y is linear but the slope varies for two different groups. The slope is positive, or ascending, for one group while the second group has an identical slope of opposite sign. This intersection of lines is sometimes called a crossover, or disordinal interaction. As a substantive example, it seems that among 10-month old infants fear of strangers and cognitive development are positively related (that is, the more fearful are more intelligent) whereas the relationship reverses for 12-month old infants (Jacobson, 1977). Although nonlinearity is a limitation in the generality of covariance as a measure of relationship, it need not be a limitation for the analysis of linear, structural equations. We return to the nonlinear function relationships in Chapter 13 and discuss strategies for coping with the problem.

THE FOUR RULES

So far we have defined a covariance and developed two problems in interpretation: outliers and nonlinearity. We now turn to the central topic of this chapter: covariance algebra. Since all information about causal parameters is contained in the covariances, one must find the key to unlock their precious secrets. The key is contained in four fundamental rules of covariance algebra.

The first rule states that the covariance of a variable with a constant is zero. Thus, if X is a variable and k a constant, $C(X,k) = 0$. Since k is constant across all participants, its mean (M_k) is k and so $k - M_k$ is always zero and every product with $(X - M_X)$ is zero. If we plotted X against k , we would obtain a flat line with zero slope indicating no functional relationship between X and k .

The second rule is $C(kX,Y) = kC(X,Y)$. Thus, if a variable is multiplied by a constant, the constant can be factored out of the covariance. Some simple corollaries are $C(kX,bY) = kbC(X,Y)$ and $C(X,-Y) = -C(X,Y)$.

The third rule states that a covariance of a variable with itself, or autocovariance, is simply the variance of the variable. This can be seen by taking the definitional formula for covariance and substituting X for Y , yielding

$$\frac{\Sigma[(X - M_X)(X - M_X)]}{N - 1}$$

which is simply the definition of sample variance. Using the second rule, a theorem about variance can be derived. If we multiply a variable by a constant, we multiply the variance of that variable by the constant squared: $C(kX,kX) = k^2C(X,X) = k^2V(X)$ where $V(X)$ denotes the variance of X .

The fourth and final rule is that the covariance of a variable with the sum of two variables is simply the sum of the covariances of the variable with each of the components of the sum: $C(X,Y+Z) = C(X,Y) + C(X,Z)$.

Table 2.1 Rules of Covariance Algebra

Null rule	$C(X,k) = 0$
Constant rule	$C(kX,Y) = kC(X,Y)$
Variance definition	$C(X,X) = V(X)$
Sum Rule	$C(X,Y+Z) = C(X,Y) + C(X,Z)$

So covariance algebra can be reduced to four simple rules. For convenience, they are stated in Table 2.1. To provide names for the rule let us call them the null rule ($C[X,k] = 0$), the constant rule ($C[kX,Y] = kC[X,Y]$), the variance definition ($C[X,X] = V[X]$), and the sum rule ($C[X,Y+Z] = C[X,Y] + C[X,Z]$). It should be noted here one nonrule: $C(X/Y,Z)$ does not necessarily equal $C(X,Z)$.

APPLICATIONS OF COVARIANCE ALGEBRA

Using the rules one can now expand the following covariance, applying the sum rule first

$$C(aX+bY,X+cZ) = C(aX+bY,X) + C(aX+bY,cZ)$$

again applying the sum rule

$$= C(aX,X) + C(bY,X) + C(aX,cZ) + C(bY,cZ)$$

Now applying the constant rule

$$= aC(X,X) + bC(Y,X) + acC(X,Z) + bcC(Y,Z)$$

and finally the variance definition

$$= aV(X) + bC(Y,X) + acC(X,Z) + bcC(Y,Z)$$

With some familiarity the sum rule need not be successively applied. One can compute the covariance of two sums as the sum of the covariance of all possible cross-products. For instance, find the variance of $X+Y$. The sum of covariances of cross-products is

$$C(X,X) + C(X,Y) + C(Y,X) + C(Y,Y)$$

Adding the two $C(X,Y)$ s and using the variance definition yields

$$V(X+Y) = V(X) + V(Y) + 2C(X,Y)$$

The theorem is useful and should be memorized. A more general theorem is that a variance of a sum equals the sum of variances plus twice all possible covariances.

Note that adding a constant does not influence a covariance. First, using the sum rule yields

$$C(X+a,Y) = C(X,Y) + C(a,Y)$$

and the null rule reduces the preceding to $C(X,Y)$.

Many of the familiar measures of linear relationship like regression and correlation can be defined in terms of covariances. For instance, the regression of Y on X (X the predictor and Y the criterion), or b_{YX} is simply $C(X,Y)/V(X)$. The metric of a regression coefficient is in units of Y per X . As the metric and the formula suggest, the regression coefficient is not symmetric: b_{YX} does not ordinarily equal b_{XY} . Because $b_{YX} = C(X,Y)/V(X)$, $b_{YX} = b_{XY}$ only if $V(X) = V(Y)$ or $C(X,Y) = 0$. Much more is said about regression coefficients in later chapters, but for the moment, let us think of them as the ratio of two covariances.

A scale-free measure of linear relationship is correlation:

$$\frac{C(X,Y)}{[V(X)V(Y)]^{1/2}}$$

or, as it is symbolized, r_{XY} in the sample and ρ_{XY} in the population. As is well known, a correlation varies from +1 to -1 and $r_{XY} = r_{YX}$. There is a simple relation between correlation and regression coefficients

$$r_{XY} = b_{YX}[V(X)/V(Y)]^{1/2}$$

The sign of both a correlation and a regression coefficient depends on the numerator, $C(X,Y)$. Along the same lines, a correlation and a regression coefficient can only be zero if $C(X,Y) = 0$.

A correlation can be thought of as a covariance between two standardized variables. Standardization means that the variable has been transformed to have zero mean and unit variance; so to standardize X compute $(X - M_X) / V(X)^{1/2}$. An examination of the formula for correlation shows that if X and Y are standardized variables, $C(X,Y) = r_{XY}$ because $[V(X)V(Y)]^{1/2} = 1$ given that X and Y have unit variance through standardization. The fact that the covariance of standard scores is a correlation will be of special help in the development of path analysis in the next chapter.

Using our knowledge of regression coefficients and covariances, one can determine $C(Y - b_{YX}X, X)$:

$$C(Y - b_{YX}X, X) = C(Y, X) - b_{YX}V(X)$$

by the sum and constant rule and variance definition,

$$= C(Y, X) - [C(X, Y)/V(X)]V(X)$$

by the definition of the regression coefficient,

$$= C(X, Y) - C(X, Y)$$

$$= 0$$

Some readers may recognize that $Y - b_{YX}X$ is the residual from the regression line. It has then just been shown that the residual is uncorrelated with the predictor X .

Covariance algebra is especially useful for showing that seemingly, theoretically meaningful relations between variables are, in fact, only mathematical necessities. For instance, let X be a pretest and Y the posttest. A researcher may wish to correlate "change," $Y - X$,

with initial status, X . Let us determine $r_{X,(Y-X)}$ in the special case where $V(X) = V(Y)$. Since $C(X, Y - X) = C(X, Y) - V(X)$ and $V(Y - X) = V(Y) + V(X) - 2C(X, Y)$, the definition of correlation implies that $r_{X,(Y-X)}$ equals

$$\frac{C(X, Y) - V(X)}{(V(X)[V(Y) + V(X) - 2C(X, Y)])^{1/2}}$$

and given $V(X) = V(Y)$ and some algebraic manipulation the formula becomes:

$$= -[(1 - r_{XY})/2]^{1/2}$$

Thus, the correlation of change with initial status must be negative if the variances of X and Y are equal and r_{XY} is less than one. This fact is part of what is meant by regression toward the mean. Note that the difference is negatively correlated with X . Thus, those who are high on X have on the average lower standard scores on Y and those who are low on X have higher scores on Y . The interpretation of covariances containing sums and differences must be made with great care.

Another illustration of a potentially misleading covariance is $C(X+Y, X-Y)$, which simply equals $V(X) - V(Y)$. This theorem is useful as a test of the null hypothesis that the variances of two variables measured on the same set of respondents are the same. The null hypothesis of $V(X) = V(Y)$ implies $\rho_{(X+Y)(X-Y)} = 0$, where ρ is the population correlation coefficient. One can then compute $r_{(X+Y)(X-Y)}$ and test whether it significantly differs from zero to test the null hypothesis of the equality of two variances computed from the same set of respondents.

Given a set of variables, their covariances can be summarized by a covariance matrix. For instance, the covariance matrix for X , Y , and Z

$$\begin{array}{ccc} C(X,X) & C(Y,X) & C(Z,X) \\ C(X,Y) & C(Y,Y) & C(Z,Y) \\ C(X,Z) & C(Y,Z) & C(Z,Z) \end{array}$$

The main descending diagonal of a covariance matrix is the set of autocovariances or variances and the matrix is symmetric with respect to that diagonal. A covariance matrix of standardized variables is a correlation matrix.

Covariance algebra is indeed useful for discovering statistically necessary relationships. However, the main use of it here is with structural equations. We see in the next chapter that if variables are standardized, covariance algebra can be short-circuited by path analysis. Nonetheless, on covariance algebra rests the analysis of linear equations.

3

PRINCIPLES OF
MODELING

Although most social and behavioral scientists either explicitly or implicitly use a causal language, they lack a common, systematic approach to causal analysis. Psychologists in particular have relied on a language that grew out of randomized experimental design. The key notions of independent and dependent variables from experimental design are potentially misleading in a nonexperimental analysis. What follows is an exposition of a set of terms which econometricians and sociologists have found useful in nonexperimental causal inference. Although to some the terms may be foreign and awkward at first, the reader will find them very useful in conceptualizing the problems in later chapters.

SPECIFICATION

Structural equation models require a blend of mathematics and theory. Although there are many interesting issues in the mathematics of models, the most difficult questions are those that translate theory into equations. This process of translation is called *specification*. Theory specifies the form of equations. As Blalock (1969) has pointed out, most theories in the social sciences are not strong enough to elaborate the exact form of equations, and so the causal modeler must make several theoretical assumptions or specifications. Although every specification can be stated in equation form, the specification should have some justification drawn from

theory. As stated in Chapter 1 each effect is specified as the sum of its causes, an assumption to be returned to in Chapter 13.

However, specifications need not be based only on substantive theory. There are two other sources of specification: measurement theory and experimental design. The theory of test measurement has been well developed, for example, Guilford (1954), and is useful in specifying structural equations. For instance, classical test theory posits that errors of measurement are uncorrelated with the true score. Measurement theory is often helpful in formulating structural models.

Researchers unfamiliar with the logic of experimental design are often unaware that the design of the research can yield additional specifications. For instance, factorial design and random assignment to treatment conditions yield independent causes that are uncorrelated with the unmeasured causes of the effect. Likewise, longitudinal designs may bring with them the specification that certain parameters do not change over time. Successful structural modelers exploit all three types of specifications: substantive theory, measurement theory, and experimental design. Traditionally psychologists focus on experimental design, psychometricians on measurement theory, and econometricians on substantive theory. Rather than choosing one's specification by one's discipline, specifications should be chosen to fit the problem.

What the causal modeler fears more than anything else is *specification error*: One of the assumptions of the model is incorrect. Most models contain misspecification. However, it is not sufficient to criticize a model just because it contains a specification error. One must show that it seriously biases estimates of the parameters of the model. It may be that the whole model or parts of it are very robust even with a specification error. A hypothesized specification error must be examined carefully to see exactly how it affects the model.

STRUCTURAL MODELS

Structural models have two basic elements, variables and parameters. Variables, as the name suggests, vary across the persons

for whom the causal law applies. One person has an intelligence test score of 100, another of 110, still another of 80. These variables are the raw materials of theories. In this text variables are symbolized by capital letters, for example, I for intelligence and A for achievement. Parameters do not vary across persons, but they describe the whole population. Mean and variance are examples of two parameters. With structural models, parameters are structural parameters, and each structural parameter is multiplied by a variable. The sum of these parameter-variable combinations equals the effect variable, and the resultant equation is called a *structural equation*. Structural parameters are symbolized by lower-case letters, a, b, c, \dots, z . A more standard manner of denoting parameters is to use b or β and subscript the cause and effect with the subscript for the effect written first. For instance, if X causes Y , b_{YX} or β_{YX} is the structural parameter. If all the variables are standardized, the structural parameters are called *path coefficients* and ordinarily are designated by lower case p s and subscripted in the same way as the b s and β s. For instance, p_{ji} is the path from i to j where i is the cause and j the effect. The subscripting of causal coefficients can be rather clumsy; moreover, the use of the b s implies that the causal parameters are regression coefficients. They are in fact theoretical regression coefficients, but in practice it is often the case that multiple regression analysis does not provide unbiased estimates of causal parameters.

To illustrate a structural equation, income (I) is thought to be the sum of the status of occupation (S) and the amount of education (E); so for person i :

$$I_i = aS_i + bE_i \quad [3.1]$$

All variables are subscripted by person because they vary across persons. Also throughout the text unless otherwise noted all variables are expressed in mean deviation form. For Equation 3.1, I is income minus the group mean. For all future equations the person subscript is dropped and presumed. The lower-case letters in Equation 3.1,

parameters a and b , are the causal parameters and their interpretation is straightforward. If the variable S were increased by one unit and if E were kept constant, I would be increased by a units. Of course, if a were negative and S were increased by one unit, I would be decreased.

Equation 3.1 is an example of a structural equation. By convention the effect is written on the left side of the equation and is sometimes called the *left-hand variable*; the causes are written on the right side and are sometimes called *right-hand variables*. A more common vocabulary taken from econometrics is to refer to variables that are only causes as *exogenous* variables and to the effects as *endogenous* variables. The causes are sometimes called the *independent variables*, and the effects are sometimes called the *dependent variables*. However, this terminology was developed for randomized experimental research in which the independent variable is always independent because it is experimentally manipulated. In the non-experimental case, the terms independent and dependent variable would seem inappropriate. In this text the terms cause or exogenous variable and effect or endogenous variable are used instead of independent and dependent variable.

Structural equations differ from other equations because structural equality represents not only a mathematical relationship, but also a theoretical relationship between cause and effect. For instance, the equation

$$-aS_i = bE_i - I_i$$

is mathematically equivalent to Equation 3.1, but it makes no structural sense.

In structural equation modeling, there is normally a system of equations. For instance, one might add to Equation 3.1 the following equation

$$E = cF + dV \quad [3.2]$$

where E is education, F is father's social class, and V represents all other causes of education. Again all variables are expressed in mean deviation form. Note that E is a cause in Equation 3.1 and an effect in Equation 3.2, and is called an *endogenous variable*. Variables F and S in these two equations are said to be *exogenous variables* because they are only cause other variables, while the variable I is endogenous. The set of structural equations is called the *causal* or *structural model*.

There are two types of structural models: hierarchical and nonhierarchical. A nonhierarchical model has feedback loops, either direct or indirect. An example of a direct feedback loop is the case in which X causes Y and Y causes X . An indirect loop occurs when one of the causal links of the feedback loop is indirect, for example, X causes Y and Y causes Z which causes X . As an example of a nonhierarchical model from economics, a model of supply and demand is nonhierarchical because supply causes demand and demand causes supply.

A hierarchical model has no feedback loops. Formally a model is hierarchical if the set of structural equations can be ordered in such a way that an effect in any equation does not appear as a cause in any earlier equation. If the equations cannot be so ordered the model is said to be nonhierarchical. For instance, the following set of equations is nonhierarchical:

$$\begin{aligned} X &= aY + bU \\ Y &= cZ + dV \\ Z &= eX + fW \end{aligned}$$

because Y causes X and X causes Z which causes Y . The structural model is then nonhierarchical. In this text the central concern is with hierarchical models, although nonhierarchical models are discussed in Chapter 6.

To summarize, a structural model consists of a set of equations. The effect or endogenous variable is on the left side, and on the right side is the sum of the causes each causal variable multiplied

by a causal parameter. If a variable is never an effect, it is called an exogenous variable. If there are no feedback loops, the model is said to be hierarchical.

PATH ANALYSIS

Often the variables in models have zero mean and unit variance which is called *standardization*; although this greatly simplifies the algebra, it does imply some loss of generality. First, standardization brings about the loss of the original metric, which sometimes interferes with interpretation, but in many cases the original metric is arbitrary anyway. Second, occasionally specifications or assumptions are made that involve the original metric. For instance, Wiley and Wiley (1970) suggested that unstandardized error variances may be stable over time. Such models can be handled with standardization but only very awkwardly. Third, the unstandardized metric is more valid when comparing parameters across populations. Fortunately it is generally a simple matter to transform the standardized parameters into unstandardized parameters. The costs of standardization are discussed further in Chapter 13. Unless otherwise stated the reader should presume that all variables are standardized.

Throughout this book a heuristic is used for standardized-hierarchical-linear models: path analysis. Path analysis was first developed by the biologist Sewall Wright (1921) and introduced into the social sciences by the sociologist Otis Dudley Duncan (1966). The working tools of path analysis are the path diagram, the first law, and the tracing rule.

The following set of structural equations can be summarized pictorially in the path diagram in Figure 3.1:

$$\begin{aligned} X_3 &= aX_1 + bX_2 \\ X_4 &= cX_1 + dX_2 + eX_3 \end{aligned}$$

The rules for translating equations into a path diagram are to draw an arrow from each cause to effect and between two exogenous variables draw a curved line with arrowheads at each end. The path diagram contains all the information of a system of equations, but for

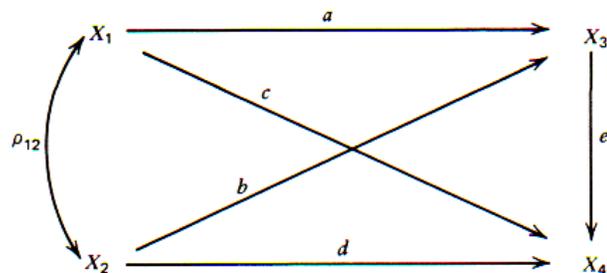


Figure 3.1. Example of a Path Model

many models the diagram is easier to comprehend. Most of us feel a little more comfortable examining the picture than the set of equations. Because the diagram contains the set of equations, it is possible to write the equations from the path diagram alone.

In Figure 3.1 the two variables X_1 and X_2 are exogenous variables and may be correlated, which is represented by a curved line with arrows at both ends. The correlation between two exogenous variables has no causal explanation, that is, the model does not specify what structural relationship brings about the correlation. The researcher may know of no causal model, or may know of many alternative causal models, or may feel the causal model is in the domain of another discipline.

It is rare, indeed, that theory specifies all the causes of a variable, although such a case is discussed in Chapter 9. Therefore, usually another cause must be added that represents all the unspecified causes of the endogenous variable. This residual term is often called the *disturbance*, the term used in this text, or error or stochastic term. The disturbance represents the effect of causes that are not specified. Some of these unknown causes may be potentially specifiable while

others may be essentially unknowable. The disturbance is an unmeasured variable and is usually denoted by the letter U , V , or W . In path analysis the disturbance has a causal coefficient. In contemporary models, the coefficient is set to one. The equations for the model in Figure 3.1 adding disturbance terms are as follows:

$$\begin{aligned} X_3 &= aX_1 + bX_2 + fU \\ X_4 &= cX_1 + dX_2 + eX_3 + gV \end{aligned}$$

The path diagram for the model is contained in Figure 3.2. Disturbances can be thought of as exogenous variables and so curved lines could have been drawn between U and V and from each to both X_1 and X_2 . It has been assumed, however, that the disturbances are uncorrelated with each other and with the exogenous variables.

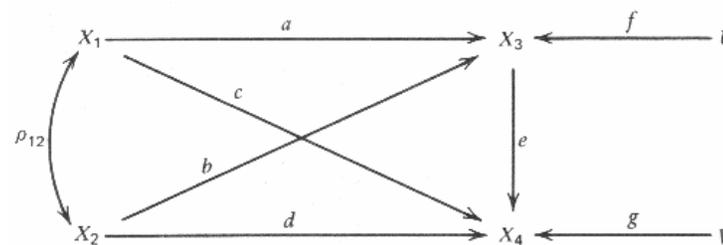


Figure 3.2. Path model with disturbance

Omitted curved lines imply that the correlation between exogenous variables is zero. The specification of uncorrelated disturbances is often made, and as with any specification it should be theoretically valid and reasonable. In path analysis, there is a path times each disturbance.

Not all the variables of a structural model are measured. For instance, the disturbance is not measured. However, even some of the other variables may not be measured. For instance, the factor analysis model discussed in Chapter 7 postulates such unmeasured variables.

THE FIRST LAW

Given a standardized structural model the correlation between any two variables can be derived by the following rule, *called the first law of path analysis*:

$$r_{YZ} = \sum p_{YX_i} r_{X_iZ}$$

where p_{YX_i} is the path or causal parameter from variable X_i to Y , p_{X_iZ} is the correlation between X_i and Z , and the set of X_i variables are all the causes of the variable Y . The first law can be shown to follow from covariance algebra. To apply the first law the variable Y must be endogenous. If both variables are exogenous, the correlation cannot be broken down and there is no need to apply the first law. To restate the first law verbally: To find the correlation between the variables Y and Z where Y is endogenous, sum the products of each structural parameter for every variable that causes Y with the correlation of each of these variables with the variable Z . A simple procedure to employ is to write all the path coefficients of the endogenous variable including the disturbance. Next to each path write the correlation of the exogenous variable of that path with the variable Z . Multiply each path and correlation and sum the products. For example, the correlation between X_1 and X_3 for the model in Figure 3.2 is

$$\rho_{31} = a\rho_{11} + b\rho_{12} + f\rho_{1U}$$

Because ρ_{11} equals 1 and because no correlation between U and X_1 is assumed (there is no curved line between the two variables), the preceding equation reduces to

$$\rho_{31} = a + b\rho_{12}$$

The reader should work through the remaining correlations for the model in Figure 3.2:

$$\rho_{32} = b + a\rho_{12}$$

$$\rho_{41} = c + d\rho_{12} + e\rho_{13}$$

$$\rho_{42} = d + c\rho_{12} + e\rho_{23}$$

$$\rho_{43} = e + c\rho_{13} + d\rho_{23}$$

$$\rho_{34} = a\rho_{14} + b\rho_{24} + f\rho_{U4}$$

Note that there are two expressions for ρ_{34} because either X_3 or X_4 can be taken as an endogenous variable. The two equations can be shown to be algebraically equivalent. The reader might be puzzled about why $f\rho_{U4}$ is included in the equation for ρ_{34} . Although $\rho_{UV} = 0$, it does not follow that $\rho_{U4} = 0$ but it does follow that $\rho_{3V} = 0$. (This can be seen more clearly once the reader learns the tracing rule.)

The first law can also be applied to the correlation of a variable with its disturbance:

$$\rho_{3U} = a\rho_{1U} + b\rho_{2U} + f\rho_{UU}$$

which reduces to

$$\rho_{3U} = f$$

and similarly

$$\rho_{4U} = g$$

If the disturbance is uncorrelated with the exogenous variables, the path from the disturbance to its endogenous variable equals the correlation of the disturbance with that variable. This equivalence will be useful in later sections.

THE TRACING RULE

Path diagrams and the first law hold for nonhierarchical models as well as hierarchical models. But the tracing rule does not

apply, at least in any simple fashion, to nonhierarchical models. The tracing rule is a simple, nonmathematical rule: The correlation between X_i and X_j equals the sum of the product of all the paths obtained from each of the possible tracings between i and j . The set of tracings includes all the possible routes from X_i to X_j given that (a) the same variable is not entered twice and (b) a variable is not entered through an arrowhead and left through an arrowhead. Again using Figure 3.2 one obtains

$$\rho_{31} = a + b\rho_{12}$$

There are two possible tracings: first, a direct path from X_1 to X_3 and a tracing from X_1 to X_2 to X_3 . Note that the correlations between exogenous variables are treated no differently. A tracing is not allowable from X_1 to X_4 to X_3 because X_4 is entered through an arrowhead and left through an arrowhead. The solution for the correlation of X_1 with X_3 yields an identical expression for both the tracing rule and the first law. They yield identical results only when all the causes of the endogenous variables are exogenous. The remaining correlations for Figure 3.2 are

$$\begin{aligned}\rho_{32} &= b + a\rho_{12} \\ \rho_{41} &= c + d\rho_{12} + ea + eb\rho_{12} \\ \rho_{42} &= d + ae\rho_{12} + c\rho_{12} + eb \\ \rho_{43} &= e + ac + ad\rho_{12} + bc\rho_{12} + bd \\ \rho_{3U} &= f \\ \rho_{4V} &= g\end{aligned}$$

There seem to be different solutions from the tracing rule and the first law for correlations involving X_4 . To show the results are the same, though apparently different, examine the solution for ρ_{41} obtained from the first law:

$$\rho_{41} = c + d\rho_{12} + e\rho_{13}$$

Substitute the value obtained from the first law for ρ_{13}

$$\rho_{41} = c + d\rho_{12} + ea + eb\rho_{12}$$

which is identical to the value obtained by the tracing rule. The results obtained by the tracing rule can always be reproduced by using the results from the first law. In general, the first law should be preferred because it ordinarily yields a result that is simpler than the tracing rule. Moreover, there is more chance of making an error with the tracing rule because one may not be sure if all possible tracings have been made. The tracing rule is nonetheless useful, particularly with models with unobserved variables.

It may be helpful to work through another example. A researcher is interested in the causal model for verbal (X_4) and mathematical achievement (X_5). Theory tells her that parental socioeconomic status (X_1), intelligence (X_2), and achievement motivation (X_3) cause the two achievement variables. She also assumes that there are other common causes of the two achievement variables. The structural equations are:

$$\begin{aligned}X_4 &= aX_1 + bX_2 + cX_3 + dU \\ X_5 &= eX_1 + fX_2 + gX_3 + hV\end{aligned}$$

with the side conditions

$$\begin{aligned}\rho_{1U} = \rho_{2U} = \rho_{3U} = \rho_{1V} = \rho_{2V} = \rho_{3V} &= 0 \\ \rho_{UV} &\neq 0\end{aligned}$$

The assumption of omitted causes of X_4 and X_5 is embodied by the correlation of the disturbances. The path diagram is in Figure 3.3, and using either the first law or tracing rule the correlations of the variables are:

$$\rho_{41} = a + b\rho_{12} + c\rho_{13}$$

$$\begin{aligned} \rho_{42} &= a\rho_{12} + b + c\rho_{23} \\ \rho_{43} &= a\rho_{13} + b\rho_{23} + c \\ \rho_{51} &= e + f\rho_{12} + g\rho_{13} \\ \rho_{52} &= e\rho_{12} + f + g\rho_{23} \\ \rho_{53} &= e\rho_{13} + f\rho_{23} + g \\ \rho_{4U} &= d \\ \rho_{5V} &= h \\ \rho_{4V} &= d\rho_{UV} \\ \rho_{5U} &= h\rho_{UV} \end{aligned}$$

Using the first law either

$$\rho_{54} = e\rho_{14} + f\rho_{24} + g\rho_{34} + h\rho_{V4}$$

or equivalently

$$\rho_{45} = a\rho_{15} + b\rho_{25} + c\rho_{35} + d\rho_{U5}$$

and the tracing rule

$$\begin{aligned} \rho_{45} &= ae + bf + cg + af\rho_{12} + be\rho_{12} + ag\rho_{13} + ec\rho_{13} \\ &\quad + bg\rho_{23} + fc\rho_{23} + dh\rho_{UV} \end{aligned}$$

As can be seen for ρ_{45} , the expressions can be complicated, but if the simple rules are followed it all becomes routine.

It should be noted that the laws of path analysis hold for all correlations in a causal model, including the correlation of a variable with itself. Consider, for instance, the following equation:

$$X_3 = aX_1 + bX_2 + cU$$

where

$$\rho_{1U} = \rho_{2U} = 0$$

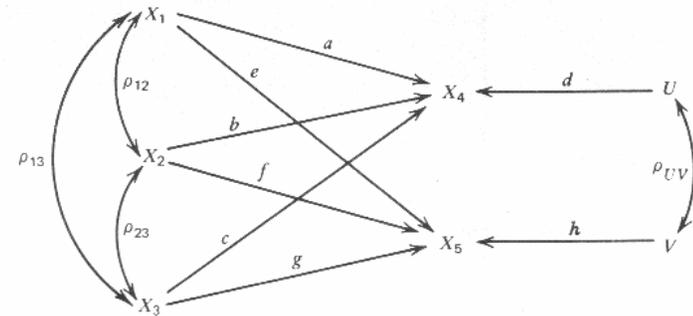


Figure 3.3 Model with correlated disturbances

Using the first law

$$\rho_{33} = a\rho_{13} + b\rho_{23} + c\rho_{3U}$$

Because it is known that ρ_{33} must equal 1, it follows that the preceding equation equals 1.

To use the tracing rule to find ρ_{33} simply draw another X_3 as was done in Figure 3.4. Now find all the possible tracings yielding

$$\rho_{33} = a^2 + b^2 + 2ab\rho_{12} + c^2 \quad [3.3]$$

A third solution for ρ_{33} is to solve for the variance of $aX_1 + bX_2 + cU$ which is known to be one because of standardization. This would be an expression identical to Equation 3.3, a fact that is later useful in estimating the path from the disturbance to the endogenous variable.

The reader now knows how to solve for correlations of a given standardized, hierarchical, linear model. A major goal of structural modeling is not to find such correlations, but rather to find the values of the causal parameters. The path coefficients of a given model are solved for by a four-stage process:

1. Measure as many variables of the model as possible.
2. Compute the correlations between the measured variables.
3. Using the first law or the tracing rule, derive the formulas for correlations.
4. Substituting in the correlations computed from data, solve for the path coefficients.

The first two steps are simple enough for anyone who has gathered data previously. The reader has just learned how to do the third step, and with some practice the reader will find that it is simple to do. The trouble comes with the last step; as is seen in the next section it may not even be possible to solve for the parameters from the correlations. In practice, this four-step procedure can be bypassed through the use of a structural equation modeling program. Still these programs in essence use this four-step procedure.

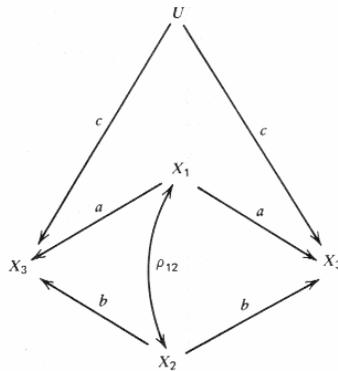


Figure 3.4 Path diagram with the effect written twice

IDENTIFICATION

Determining whether estimation of parameters of the model is possible brings one to the issue of identification. Although

identification may seem to be a strange term to use when speaking of a set of equations, it is the term that econometricians have chosen. Identification has nothing to do with the number of observations (there must be at least as many observations as parameters plus two), but rather with the number of correlations between the measured variables. A necessary but not sufficient condition to be able to identify and, therefore, estimate the causal parameters of a set of structural equations is that the number of correlations between the measured variables be greater than or equal to the number of causal parameters. This necessary condition is called the minimum condition of identification. Given that the two are equal, it may be possible that the model is just-identified or saturated; that is, there is one and only one estimate for each causal parameter. If there are more correlations than parameters, the structural model is said to be overidentified; that is, there is more than one way of estimating a causal parameter in the system. For instance, the logic of the F test in one-way analysis of variance is that there are two estimates of error variance under the null hypothesis: variability within and between groups. The error variance is then overidentified. A set of equations is said to be *underidentified* if there are more parameters than correlations. If a parameter were not identified, an infinite number of values would satisfy the equations.

As was stated earlier, knowing that the correlation of a variable with itself equals unity yields a solution for the path coefficient of the disturbance if the values of the other parameters are known. For instance, in Equation 3.3 the path coefficient for the disturbance equals

$$c^2 = 1 - a^2 - b^2 - 2ab\rho_{12} \quad [3.4]$$

Given that the metric of the disturbance is arbitrary, by convention one takes the positive square root of Equation 3.4. Parameter c is a constrained parameter; that is, its value is determined by the other parameters of the model. The implication of this for the identification issue is that one need not worry about the identification of the paths

from the disturbances to endogenous variables, but one need consider only the identification of the other parameters of the model. When the disturbance is uncorrelated with the exogenous variables and given that

$$Y = \sum p_{YX_i} X_i + p_{YU} U$$

the solution for path from U to Y is

$$\rho_{YU} = (1 - \sum p_{Y_i} \rho_{iY})^{1/2}$$

Procedurally, how does one assess the status of identification? First one determines the number of correlations between the observed or measured variables. If n variables are measured, the number of correlations is $n(n - 1)/2$. Then count the number of parameters making sure to include (a) all the path coefficients, (b) all correlations between exogenous variables, (c) all correlations between disturbances but not to include the path coefficients of the disturbances. For instance, for the model in Figure 3.3, if X_1 through X_5 are measured, there are 10 correlations ($5 \times 4/2 = 10$) and 10 causal parameters: six path coefficients, $a, b, c, g, e,$ and f ; three correlations between exogenous variables, $\rho_{12}, \rho_{13},$ and ρ_{23} ; and one correlation between disturbances, ρ_{UV} . The model may be just-identified (it is, in fact). I say “may” because one parameter might be overidentified and all the remaining parameters might be underidentified. Remember that having the number of correlations equal to the number of parameters is only a minimum condition for identification.

The model in Figure 3.3 would be underidentified if X_3 were not measured because there still are ten parameters but only six correlations. The model would be overidentified if it were assumed that the disturbances were uncorrelated because there would be nine parameters and ten correlations.

If a model is overidentified, there are two estimates of a causal parameter or some function of causal parameters; that is, one can express

a causal parameter as two or more different functions of the correlations. If the functions of correlations are set equal to each other, there is an equation that says two sets of correlations equal each other. The resultant equality is called an overidentifying restriction on the model. Often a subset of overidentifying restrictions implies one or more of the other overidentifying restrictions. The smallest set of restrictions that implies the remaining ones yields the number of independent restrictions on the model. The number of this smallest set is referred to as the degrees of freedom of the model. If a model is truly overidentified, the number of free parameters plus the number of independent overidentifying restrictions is equal to the number of correlations. Overidentifying restrictions can be viewed as a constraint on the structure of a correlation matrix. Thus, the number of free correlations is effectively reduced by the number of overidentifying restrictions. One must then subtract the number of independent restrictions from the number of correlations to determine if the minimum condition is met. These restrictions play an important role in structural modeling because they can be used to test the validity of the model.

Any overidentifying restriction can be stated as a null hypothesis and then tested. Rarely does an overidentifying restriction exactly hold in the sample, but if the model is valid it should hold within the limits of sampling error. Overidentifying restrictions also increase the efficiency of parameter estimation (Goldberger, 1973). If there are two estimates of the same parameter, those estimates can be pooled to obtain a new estimate whose variance is less than or equal to the variance of either original parameter estimate. One of the key themes of this book is the importance of these over-identifying restrictions.

A JUST-IDENTIFIED MODEL

Before proceeding any further with the concept of identification, consider an example. The equation for the path diagram in Figure 3.5 is

$$X_3 = aX_1 + bX_2 + cU$$

The correlations are

$$\rho_{31} = a + b\rho_{12} \quad [3.5]$$

$$\rho_{32} = b + a\rho_{12} \quad [3.6]$$

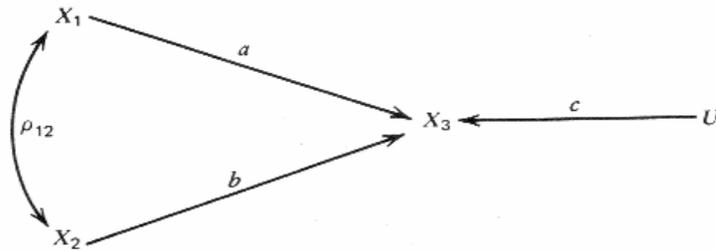


Figure 3.5 A just-identified model

Adding ρ_{12} there are three correlations ($3 \times 2/2$) and three parameters, a , b , and ρ_{12} , and the model is possibly just-identified or saturated. Solutions for a and b exist because there are two linear equations with two unknowns. Multiplying ρ_{12} times Equation 3.6 yields

$$\rho_{23}\rho_{12} = b\rho_{12} + a\rho_{12}^2 \quad [3.7]$$

Subtracting Equation 3.7 from Equation 3.5 yields

$$\rho_{13} - \rho_{12}\rho_{23} = a - a\rho_{12}^2$$

We now solve for parameter a yielding

$$a = \frac{\rho_{13} - \rho_{12}\rho_{23}}{1 - \rho_{12}^2} \quad [3.8]$$

$$1 - \rho_{12}^2$$

And by analogy, the solution for b is

$$b = \frac{\rho_{23} - \rho_{12}\rho_{13}}{1 - \rho_{12}^2} \quad [3.9]$$

As was seen in Equation 3.4, $c^2 = 1 - a^2 - b^2 - 2ab\rho_{12}$. By knowing the model was just-identified one knows, at least, that it may be possible to find solutions for a , b , and c . The reader may have noticed something interesting about Equations 3.8 and 3.9. The solution for parameter a is simply the beta weight or standardized regression coefficient of X_3 on X_1 controlling for X_2 , whereas b is the beta weight of X_3 on X_2 controlling for X_1 . Compare Equations 3.8 and 3.9 with the formulas on page 192 in McNemar (1969). The reader may breathe a sigh of relief because one will not need to solve a different set of simultaneous equations for every causal model. As shall be seen, regression coefficients, partial correlations, factor loadings, and canonical coefficients estimate causal parameters of different models. After progressing through the book, the reader should be able to tell what statistical technique estimates the parameters for a given path model, which should give the reader more insight to the causal interpretation of these statistical techniques.

AN OVERIDENTIFIED MODEL

For the model in Figure 3.6 variables X_1 through X_4 are measured while the disturbances and F are unmeasured. The disturbances of X_1 are correlated with X_2 and those of X_3 with X_4 . (As an exercise write out the structural equations for the model in Figure 3.6.) There are six correlations ($4 \times 3/2$) and six free parameters: four path coefficients, a , b , c , and d ; and two correlations between

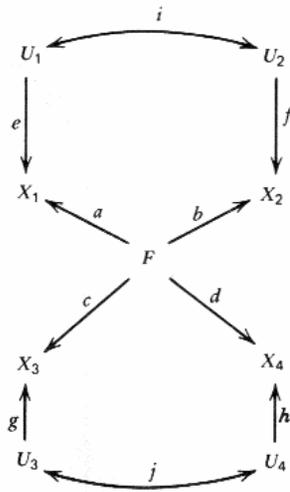


Figure 3.6 Model with an overidentifying restriction

disturbances, i and j . The model is possibly just-identified. The correlations obtained by the tracing rule are

$$\rho_{12} = ab + eif$$

$$\rho_{13} = ac$$

$$\rho_{14} = ad$$

$$\rho_{23} = bc$$

$$\rho_{24} = bd$$

$$\rho_{34} = cd + ghj$$

Try as you can, you will never be able to express the parameters in terms of the correlations. Even though there are six correlations and six free parameters satisfying the minimum condition of identifiability, none of the parameters are identified. This is due to the overidentifying restriction of

$$\rho_{13}\rho_{24} = \rho_{14}\rho_{23}$$

because both $\rho_{13}\rho_{24}$ and $\rho_{14}\rho_{23}$ equal $abcd$. Recall that an overidentifying restriction is a constraint on the correlation (covariance) matrix of the measured variables. The overidentifying restriction, even if true, should only perfectly hold in the population. Given the overidentifying restriction plus the six parameters there is not enough information to estimate the parameters of the model, because the overidentifying restriction reduces the number of free correlations to five, which is less than the number of free parameters. Even though none of the parameters of the model are estimated, the model can still be tested by taking the overidentifying restriction as a null hypothesis. Such is also the case for cross-lagged panel correlation, which is discussed in Chapter 12.

UNDERIDENTIFICATION

Should one give up if the model is underidentified? As has been already stated, even though as a whole the model may be underidentified (a) various parameters of the model may still be just-identified, (b) some parameter may even be overidentified, or (c) as in Figure 3.6 no parameters are identified, but there may be an overidentifying restriction on the model. But, alas, in many cases none of the preceding three conditions hold. To achieve identification or even overidentification various strategies can be employed. One strategy is to measure more variables. This is useful when an exogenous variable is correlated with the disturbance or when an exogenous variable is measured with error. Unfortunately, adding more variables and more equations can at times only make matters worse. Another strategy to bring about identification is to reduce the number of parameters. There are three common means to do this. First, one can assume that certain parameters are zero. As has been seen and as shall continue to be seen, one often assumes that the disturbance is uncorrelated with the exogenous variables. One may assume that certain paths are zero as in Chapter 5 with the method that econometricians call "instrumental variable" estimation. The second

method to reduce the number of parameters is to set two of them equal to each other, thereby decreasing the number of free parameters by one. For instance path coefficients are set equal as in the case of cross-lagged panel correlation in Chapter 12, or in the case of disturbances set equal across equations, or correlations between disturbances are set equal for the multitrait-multimethod matrix in Chapter 8. The third and last strategy is to make a proportionality constraint.

Given four parameters, a through d , one might assume $a/b = c/d$. Thus only three parameters are free because if three are known the fourth can be solved for. For instance, if a , b , and c are known, d equals bc/a . There are other procedures available for reducing the number of parameters but the three preceding are most common.

All these strategies are viable only if they can be justified by substantive theory. Making new specifications just to be able to identify the parameters of a causal model is perhaps the worst sin of causal modelers. Obviously, identification is necessary for causal modeling, but one must not sloppily add constraints just to achieve identification.

Duncan (1976) has suggested yet another approach to under identification. He suggests fixing one parameter, or as many as needed, to a reasonable value and solving for the remaining parameters. Then by changing the original parameter again, another set of parameter estimates are obtained. This is done repeatedly, until a range of possible values are obtained. Often it will happen that the obtained range of values is, under this procedure, rather limited. Occasionally, there are no reasonable solutions; for instance, a correlation between disturbances is larger than one or a parameter's solution is imaginary (the square root of a negative value). The lack of a reasonable solution would indicate some sort of specification error.

EMPIRICAL UNDERIDENTIFICATION

Even though a system of equations may be identified in principle, in practice there may be no solution for a parameter. It may

happen that when the correlations or covariances are substituted into the expression that should estimate the parameter, the denominator of that expression is zero or very near zero. Because division by zero is algebraically undefined, there is no solution. The case in which the denominator of an expression that estimates a parameter is equal or nearly equal to zero, is called in this text empirical underidentification. As is seen in the next chapter multicollinearity is an example of empirical underidentification.

After one has determined that a set of equations is identified in principle and obtained estimates of each parameter, one should carefully examine the denominator of each estimate and note the condition under which it equals zero; this condition defines empirical under identification.

Although the estimate is defined when the denominator is nonzero, if the denominator is very close to zero the estimate is practically useless because its standard error may be huge. *Empirical underidentification* is defined by zero or near-zero denominators in the estimates of structural parameters. There is the problem of how to define "near zero." When working from a correlation matrix with denominators less than .1, the estimates are often so unstable as to be worthless. Empirical underidentification results in very unstable parameter estimates and very large standard errors. For some models with empirical underidentification, the computer program never arrives at a satisfactory solution (i.e., fails to converge). Kenny, Kashy, and Bolger (2000) compare empirical underidentification to a black hole, which is defined in Einstein's relativity theory by an equation with a denominator of zero.

A SINGLE UNMEASURED VARIABLE

One need not measure all the variables in a model besides disturbances to achieve identification. Consider the following model

$$X_1 = aF + eU_1$$

$$\begin{aligned}X_2 &= bF + fU_2 \\X_3 &= cF + gU_3\end{aligned}$$

with the following side conditions

$$\begin{aligned}\rho_{U_j F} &= 0 \quad (j = 1, 2, 3) \\ \rho_{U_i U_j} &= 0 \quad (i \neq j)\end{aligned}$$

(As an exercise draw a path diagram for the preceding set of equations.) The model is just-identified with X_1 , X_2 , and X_3 measured and F being unmeasured. There are three correlations and three free parameters, a , b , and c . The correlations are

$$\begin{aligned}\rho_{12} &= ab \\ \rho_{13} &= ac \\ \rho_{23} &= bc\end{aligned}$$

The parameters are then

$$\begin{aligned}a^2 &= \frac{\rho_{12}\rho_{13}}{\rho_{23}} \\ b^2 &= \frac{\rho_{12}\rho_{23}}{\rho_{13}} \\ c^2 &= \frac{\rho_{13}\rho_{23}}{\rho_{12}}\end{aligned}$$

This is a single-factor model where a , b , and c are factor loadings (cf. Duncan, 1972; Harmon, 1967). This model is more extensively discussed in Chapter 7.

STATISTICAL ESTIMATION, INFERENCE AND TESTING

The discussion has been mainly restricted to the population by assuming that the population values of correlation coefficients are known. Rarely if ever are these correlations known and in practice the researcher must use sample estimates. Given only sample data, statistical methods must be used to estimate parameters. Ideally, the estimates should be sufficient, unbiased, consistent, and efficient. [See Hays (1963) for definitions.] Fortunately, most multivariate statistical methods can be adapted to estimation of the parameters of structural models.

If the equations are simple enough or the researcher ingenious enough, it may be possible to solve for the structural parameters from the set of equations that are obtained from the tracing rule or the first law. Such a solution is commonly called a path analytic solution. Often one can just substitute the sample estimates of correlations into the path analytic solution values. This ordinarily gives good estimates if the model is just-identified. However, if the parameter is overidentified there will be multiple path analytic solutions. For such cases traditional statistical methods are to be preferred.

Statistical methods are also needed to find standard errors of parameter estimates, and to test whether a given parameter is zero or whether two parameters are equal. Statistical methods are, in addition, used in testing overidentifying restrictions. To make these significance tests, ordinarily additional specifications must be made which usually concern the distribution of a variable or the disturbance.

The fit of the model can be determined. In essence, the discrepancy between the observed covariances and the model predicted covariances are compared. In this way, the overall fit of the model can be determined.

REDUCED FORM

With some models it is simpler not to deal with the parameters themselves but to work with the reduced form coefficients. To solve

for the reduced form coefficients the structural equations are expressed solely in terms of the exogenous variables. As an example, given the following set of equations

$$X_2 = aX_1 + bU \quad [3.10]$$

$$X_3 = cX_1 + dX_2 + eV \quad [3.11]$$

within Equation 3.11 one can substitute for X_2 what it equals in Equation 3.10 or

$$X_3 = cX_1 + d(aX_1 + bU) + eV$$

$$X_3 = (c + da)X_1 + dbU + eV$$

The reduced form coefficient for X_1 is not a single path coefficient but a function of coefficients ($c + da$). At times it is easier to solve for the reduced form coefficients and then solve for the path coefficients. Furthermore, as is seen in the next chapter, reduced form is useful for solving for indirect paths from one variable to an endogenous variable (Alwin & Hauser, 1975; Baron & Kenny, 1985; Finney, 1972), i.e., mediation.

TINKERING WITH EQUATIONS

Sometimes for social-policy considerations the researcher wants to know the effect of altering structural equations. For instance, this was a central concern of Jencks et al. (1972) in the *Inequality* volume. They continually ask questions such as, "If the father's occupational status no longer affected the amount of the child's education, by how much would the variance of child's education be reduced?" or "Would reducing inequality in one domain significantly reduce in equality in another domain?" For a structural equation many parameters affect the variance of the endogenous variable. Given the unstandardized structural equation of

$$X_3 = aX_1 + bX_2$$

its variance is defined as

$$V(X_3) = a^2V(X_1) + b^2V(X_2) + 2abC(X_1, X_2) \quad [3.12]$$

using the variance definition and sum rule of the previous chapter. Examining Equation 3.12 note that there are three ways to affect variance of the endogenous variable X_3 :

1. Change the causal parameters a and b .
2. Change the variance of the exogenous variables.
3. Change the covariance between casual variables.

Mathematically, one is not free to alter these parameters in any way because a number of side conditions must be met. For instance, the absolute value of any correlation must be less than or equal to one, as must the absolute values of the partial correlations of $\rho_{12.3}$, $\rho_{13.2}$, and $\rho_{23.1}$.

If the variance of X_1 is altered, the covariance between X_1 and X_2 may as well be affected. If it is assumed that the change in variance equally affects the component correlated with X_2 and the component uncorrelated, then the altered covariance is equal to the original covariance times the square root of the factor by which the variance was changed.

If there is a set of structural equations, special care must be taken when a variable is altered that is an effect in one equation and is a cause in another equation. If its variance is changed one must specify exactly how the structural equation of the endogenous variable is altered. Different changes in its equation, even if the resultant variances are the same, have different consequences for the variances of the other endogenous variables.

Finally, one should not assume that, by reducing the variance of an exogenous variable, one automatically decreases the variance of the endogenous variable. Though rather uncommon, it is possible for

the variance to increase. Similarly an increase in the variance of an exogenous variable may result in a decrease in the variance of the endogenous variable.

CONCLUSION

A great amount of material has been covered. The reader should know how to express correlations or covariances in terms of causal parameters and how to draw a path diagram from a set of equations and vice versa. The reader should also know what is meant by the following terms: endogenous and exogenous variable, standardized-hierarchical-linear structural model, specification, identification, specification error, overidentifying restriction, and reduced-form coefficients. The remainder of the book is the application of these concepts to different structural models. The steps of structural modeling are:

1. From theory elaborate a model.
2. Choose a measurement and design model and respecify the structural model to conform to design and measurement specifications.
3. Determine the status of identification of the model.
4. If identified, determine the correlations (covariances) between the measured variables and from the correlations (covariances) estimate the parameters of the model and test the fit of the model.

In the next chapter it is shown that multiple regression can be used to estimate the causal parameters for certain models when measured variables cause measured variables.

4

MODELS WITH OBSERVED VARIABLES AS CAUSES

To many readers multiple regression is synonymous with path analysis. Although it is true that most empirical applications of path analysis use multiple regression analysis to estimate causal parameters, many models require the use of more sophisticated techniques to estimate their parameters. This chapter elaborates the assumptions, or specifications, that must be made to interpret ordinary multiple regression coefficients as path coefficients. More complicated techniques are discussed in later chapters. The chapter is divided into three sections. The first section briefly explains the elements of multiple regression analysis. The second section details how regression coefficients can be interpreted as causal parameters. The third section considers a number of technical issues.

MULTIPLE REGRESSION

Regression analysis begins with a set of predictor variables, say father's occupation (F) and intelligence (I), and a criterion, say school grades (G). What the researcher is interested in is the "best" linear combination of the predictor variables that predicts the criterion. For the grade example such a linear prediction equation is

$$b_1F + b_2I$$

Given errors of prediction,

$$G = b_1F + b_2I + E \quad [4.1]$$

where E is lack of fit, or residual term. All variables but E are standardized. At issue in predicting G is the choice of the coefficients b_1 and b_2 , or, as they are commonly called, regression coefficients. The coefficients are chosen to maximize the correlation of the quantity $b_1F + b_2I$ with G . It can be shown that maximizing this correlation is the same as minimizing the sum of squared residuals. Minimizing sums of squares is accomplished by the method of least squares, from which it follows that the correlation of any predictor variable with the residual term must be zero. Using the first law to correlate each of the predictors with grades, the resulting equations are:

$$r_{GF} = b_1 + b_2r_{FI}$$

and

$$r_{GI} = b_1r_{FI} + b_2$$

Since there are two linear equations in two unknowns, the unknowns b_1 and b_2 can be solved for:

$$b_1 = \frac{r_{GF} - r_{FI}r_{GI}}{1 - r_{FI}^2} \quad [4.2]$$

$$b_2 = \frac{r_{GI} - r_{FI}r_{GF}}{1 - r_{FI}^2} \quad [4.3]$$

(Compare with Equations 3.8 and 3.9 of the previous chapter.) Note that each regression coefficient is simply a function of both the correlations among the predictor variables and the correlations of the predictors with the criterion. As the number of predictor variables increases, the equations can be solved in the same way, although not surprisingly the regression coefficients are equal to a more complicated

function of the correlations. Fortunately, virtually every statistical package can estimate regression equations.

Kerchoff (1974, p. 46) reports the correlations between a number of variables for 767 twelfth-grade males. In Table 4.1 some of these correlations are reproduced. Of particular interest now are the correlations between F , I , and G :

$$\begin{aligned} r_{FI} &= .250 \\ r_{GF} &= .248 \\ r_{GI} &= .572 \end{aligned}$$

Table 4.1. Correlations Taken from Kerchoff^{a,b}

I	1.000						
S	-.100	1.000					
E	.277	-.152	1.000				
F	.250	-.108	.611	1.000			
G	.572	-.105	.294	.248	1.000		
X	.489	-.213	.446	.410	.597	1.000	
A	.335	-.153	.303	.331	.478	.651	1.000
	I	S	E	F	G	X	A

^aI, intelligence; S, number of siblings; E, father's education; F, father's occupation; G, grades; X, educational expectation; A, occupational aspiration.
^bN = 767.

Solving for b_1 and b_2 by Equations 4.2 and 4.3, respectively, yields

$$\begin{aligned} b_1 &= \frac{.248 - (.250)(.572)}{1 - .250^2} = .112 \\ b_2 &= \frac{.572 - (.250)(.248)}{1 - .250^2} = .544 \end{aligned}$$

The interpretation of the regression weights in the standardized case or, as they are more commonly called, *beta weights*, is straightforward. If someone is one standard deviation above the mean in father's occupation and at the mean in intelligence, he is on the

average .112 of a standard deviation above the mean in school grades. Similarly, someone at the mean in father's occupation and one standard deviation above in intelligence is .544 of a standard deviation above the mean in school grades.

Often a researcher is interested in how well the regression equation predicts the criterion. One measure is to obtain the variance of the predicted scores. The variance of the predicted score is defined as

$$\begin{aligned} &= C(b_1F + b_2I, b_1F + b_2I) \\ &= C(b_1F + b_2I, G - E) \end{aligned}$$

Because $C(F, E) = C(I, E) = 0$ given least squares, it follows that

$$= b_1^2 + b_2^2 + 2b_1b_2r_{FI} \tag{4.4}$$

which holds for the standardized case. Another measure of fit is the multiple correlation which is the correlation of the predicted score of G , or $b_1F + b_2I$, with G . The multiple correlation, or $R_{G(FI)}^2$ can be shown to equal

$$b_1r_{FG} + b_2r_{IG}$$

because $C(G, b_1F + b_2I) = b_1r_{GF} + b_2r_{GI}$, $V(G) = 1$, and $V(b_1F + b_2I)$ equals Equation 4.4. Note that the variance of the predicted scores, Equation 4.4, is simply the multiple correlation squared. Solving for the multiple correlation of our example yields

$$[(.112)(.248) + (.544)(.572)]^{1/2} = .582$$

The variance of the residuals is simply $1 - R^2$ because G is standardized and the predicted G is uncorrelated with the residuals.

In general, the multiple correlation is the square root of the sum of the products of each regression coefficient and the corresponding correlation of the predictor with the criterion. The

variance of the errors is simply one minus the multiple correlation squared.

Very often researchers are interested in more than predicting the criterion: They want the "best" statistical estimates of the betas and they want to test hypotheses about the regression equation. For estimates of the betas to be unbiased, consistent, and efficient, the following must be assumed:

1. Independence: Each observation should be sampled independently from a population. Independence means that the errors, the E s, are statistically independent; that is, the covariances of errors are zero. The sampling unit should be the same as the unit in the analysis; otherwise this assumption will be violated. Independence is also violated if the data are repeated measurements on a single unit since scores tend to be proximally autocorrelated, that is, data closer together in time and space are more highly correlated than data further apart.
2. Homoscedasticity: The variance of the errors should not be a function of any of the predictor variables. If the residuals are heteroscedastic, occasionally the dependent variable can be transformed to preserve homoscedasticity.

To test hypotheses concerning the regression equation, the assumption must be added that the errors are normally distributed. These assumptions of independence, homoscedasticity, and normality can be empirically tested (Bock, 1975). While violations of the assumptions of homoscedasticity and normality do not seriously distort the level of statistical significance, violation of the independence assumption does.

To test whether a given beta weight is not zero, compute the multiple correlation without the variable (R_1) and with the variable (R_2) in the equation. Then under the null hypothesis that the coefficient is zero the following is distributed as

$$t_{(N-k-1)} = \frac{[(R_2^2 - R_1^2)(N - k - 1)]^{1/2}}{(1 - R_2^2)^{1/2}} \quad [4.5]$$

where N is the total number of sampling units, k is the number of predictors in the full equation, and t is Student's t distribution. The number of degrees of freedom is $N - k - 1$. To test whether there is an effect of father's occupation, first omit it from the equation and R_1 is simply the correlation of I with G , or .572. Since $N = 767$, there are 764 degrees of freedom and earlier R_2 was found to be .582. Substituting these values into Equation 4.5 one obtains

$$t(764) = \frac{[(.582^2 - .572^2)764]^{1/2}}{(1 - .582^2)^{1/2}} = 3.65$$

which is statistically significant at the .05 level. Similarly, the test that the effect of intelligence is nil is

$$t(764) = \frac{[(.582^2 - .248^2)764]^{1/2}}{(1 - .582^2)^{1/2}} = 17.90$$

which is also statistically significant.

If two or more predictors are removed from the regression equation, a similar procedure is followed. Compute R_1 omitting the variables that one wishes to exclude and R_2 for the full equation. It can then be shown that under the null hypothesis of zero coefficients, the following is distributed as

$$F_{(m, N-k-1)} = \frac{[(R_2^2 - R_1^2)(N - k - 1)]}{m(1 - R_2^2)} \quad [4.6]$$

where m equals the number of excluded variables and F is Fisher's F distribution. The F of Equation 4.6 has m degrees of freedom in the

numerator and $N - k - 1$ in the denominator. To test the effect of removing both father's occupation and intelligence, the R_1 value is zero since there are no predictor variables. Substituting into Equation 4.6 one obtains

$$F(2,764) = \frac{[(.582^2 - .000^2)(764)]}{2(1 - .582^2)} = 195.67$$

which is statistically significant. Testing whether all the beta weights are zero is identical to testing whether the multiple correlation is zero.

A beta weight with only one predictor is identical to a correlation coefficient. A beta weight with two or more predictors is not a correlation and can even be larger than one. For instance,

$$X_3 = .75X_1 + 1.25X_2$$

where $r_{12} = -.6$ is a perfectly valid regression equation. (As an exercise show $r_{12} = 0$ and $r_{23} = .8$.) Although beta weights larger than one are not impossible, they are empirically improbable. One should be suspicious when one obtains weights larger than one.

Beta weights or path coefficients can be easily restated in an unstandardized metric. For instance, given

$$X_3' = aX_1' + bX_2' + cU'$$

where the prime notation expresses standardization, one can re-express a and b in an unstandardized metric: The unstandardized coefficient of X_1 is $a[V(X_3)/V(X_1)]^{1/2}$, and of X_2 , $b[V(X_3)/V(X_2)]^{1/2}$. Thus, to destandardize a coefficient, multiply it by the standard deviation of the effect variable and divide it by the standard deviation of the causal variable. To standardize an unstandardized coefficient, reverse the operation: Multiply it by the standard deviation of the causal variable and divide it by the standard deviation of the endogenous variable.

CAUSAL INTERPRETATION OF REGRESSION COEFFICIENTS

Researchers are often interested not only in predicting grades but also in explaining what causes grades. Is it permissible to interpret the regression coefficients obtained from the regression analysis as causal parameters? The answer is, as it is in the case of most important questions, it depends. Regression coefficients can be interpreted as causal coefficients if certain assumptions are met. These assumptions are the same as those of multiple regression: In this case the justification is not to maximize prediction or to perform significance tests, but because the assumptions are specified by theory. The assumption of independence can usually be assured by the sampling design of the research. Homoscedasticity and normality of errors are usually not discussed in most social science theories (although they are in some biological theories), but as stated earlier, these are robust assumptions which may only be "approximately" met. Often a transformation of the measure into a more meaningful metric will aid in meeting these two assumptions.

As a result of least squares, the residuals are uncorrelated with the predictor variables. In structural modeling, this result requires that the disturbance be uncorrelated with the causes of the endogenous variables. As shall be seen, the assumption of uncorrelated errors implies that:

1. The endogenous variable must not cause any of the variables that cause it; that is, there is no reverse causation.
2. The causal variables must be measured without error and with perfect validity.
3. None of the unmeasured causes must cause any of the causal variables; that is, there are no common causes, or third variables.

Obviously, lack of reverse causation, perfect measurement, and lack of common causes are rather stringent assumptions. Reverse causation can be ruled out by theory or logic; or example, variables measured at one point in time do not cause variables measured earlier in time. If reverse causation cannot be ruled out, however, a nonhierarchical model must be specified the parameters of which cannot be estimated by an ordinary regression analysis (see Chapter 6). Perfect measurement is assured for some variables such as race and sex, close approximations exist for some variables such as intelligence, but most motivational, attitudinal, and behavioral measures are not close to being measured perfectly. The common cause problem can be solved by measuring the third variables (although these variables must be perfectly measured), but it is still logically impossible to demonstrate that all third variables have been excluded if the multiple correlation is less than one. The next chapter presents partial solutions to the problems of unreliability and unmeasured third variables.

A final assumption for the application of regression analysis to structural models when the variables are measured cross-sectionally is that a state of equilibrium has been reached. Thus if X_1 is assumed to cause X_2 with a lag of k units, and if X_1 and X_2 are contemporaneously measured at time t , equilibrium exists if $X_1 = X_2$; that is, X_1 did not change between times $t - k$ and t .

The path coefficients are, then, estimated by the beta weights, and the causal coefficient for the disturbance of the endogenous variable simply equals $(1 - R^2)^{1/2}$, where R^2 is the squared multiple correlation of the regression equation. In some treatments of path analysis, the disturbance has no path coefficient. For this case the variance of the disturbance is set at $1 - R^2$. Here the variance of the disturbance is fixed at one and its path coefficient is estimated, thereby making all of the variables of the model standardized. Actually, $(1 - R^2)^{1/2}$ is a biased estimate of the disturbance path since R^2 is also biased. A less biased estimate of the multiple correlation is $R^2(N - k - 1)/(N - 1)$. One could then use this corrected R^2 to estimate the disturbance path. For ease of computation here the uncorrected

estimate is used. However, if sample size is small, the corrected R^2 should be used.

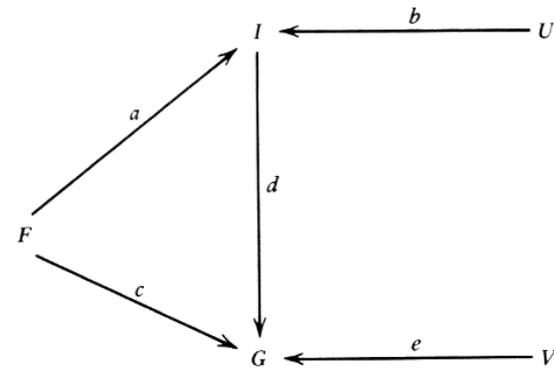


Figure 4.1 Example of a structural model taken from Kerchoff.

To illustrate the use of multiple regression to estimate causal parameters, return to the three variables. A simple model is

$$I = aF + bU \quad [4.7]$$

$$G = cF + dI + eV \quad [4.8]$$

That is, father's occupation (F) causes intelligence (I), and father's occupation and intelligence both cause school grades (G). It is assumed $\rho_{UF} = \rho_{VF} = \rho_{VI} = 0$; that is, the disturbances are uncorrelated with the causal variables, which in turn implies $\rho_{UV} = 0$. (As an exercise, prove why $\rho_{UV} = 0$.) The path diagram for Equations 4.7 and 4.8 is in Figure 4.1.

The path from F to I can be estimated by the regression of I on F because the disturbance of I is uncorrelated with F . The univariate beta weight is simply the correlation coefficient. To see this, solve for r_{FI} , by either the first law or the tracing rule, to obtain a , the path from I to F . Parameter a then is estimated by .250. To test whether such a value could be explained by sampling error, use Equation 4.5

$$t(765) = \frac{[.250^2 765]^{1/2}}{(.9375)^{1/2}} = 7.14$$

Figure 4.1 Example of a structural model taken from Kerchoff. (Note $R_1 = 0$ because no variables remain in the equation.) The path for the disturbance equals $(1 - R^2)^{1/2}$ where R is the multiple correlation. Since there is only one predictor, and $R = .250$, $b = (1 - .250^2)^2 = .968$. To solve for c and d , regress G on F and I which was done earlier in this chapter. Thus, the estimate of c equals .112 and of d equals .544. Also computed was the multiple correlation for that equation and obtained was the value of .582. The estimate of e is then $(1 - .582^2)^{1/2} = .813$. Earlier these beta weights were shown to be statistically significant.

It is just that simple. If the disturbances are uncorrelated with causal variables, then beta weights estimate path coefficients and $(1 - R^2)^{1/2}$ estimates the path from the disturbance to the endogenous variable. Tests of hypotheses about the path coefficients can be obtained from ordinary least squares (standard multiple regression analysis).

Note that in applying regression analysis to estimate structural parameters, one estimates a single regression equation. The use of stepwise regression has no place in structural modeling. With stepwise regression one ordinarily does not know which variables are important but in structural modeling, theory has already specified the important variables.

CAUSAL LINKAGES

For the model in Figure 4.1, F causes G both directly and indirectly. This can be seen by examining r_{GF} using the tracing rule:

$$r_{GF} = c + ad$$

There is a direct linkage, path c , and an indirect linkage, F causes I , and, in turn, I causes G . The variable I mediates a link or intervenes between F and G . The mathematics of indirect paths are discussed later in this chapter.

The distinction between direct and indirect causes is an important one in social science. When a causal link has been demonstrated either experimentally or nonexperimentally, usually some inquiring person will investigate the variables that mediate that causal linkage; that is, an attempt will be made to show that X causes Y through Z and that once Z has been controlled, the relationship between X and Y vanishes. There are usually a host of mediating variables, and through such a process variables in social science become refined and theory expands. One could argue that all causal linkages are indirect, or mediated, thereby making a true direct link an impossibility. A causal linkage, therefore, is said to be direct only with respect to other variables specified in the system under consideration. Certain classes of variables are not entered into structural equations generally because they are not topics of study in that field. For instance, it is likely that physiological variables mediate the psychological, and physical laws mediate certain chemical laws. Researchers do not strive to find absolute linkages, but rather look for linkages relative to their interests and discipline.

Returning to Figure 4.1, I is related to G in two ways: by a direct causal linkage and by a spurious linkage through F since F causes I and G . A spurious relationship means that the association between two variables is a result of both being caused by a third variable. Although all indirect causes need not be entered into the structural equation, spurious variables must be. For instance, since the linkage of I and G through F in Figure 4.1 does not reflect the causal effect of I on G , the variable F must be measured and controlled. Care must be taken to specify from theory these spurious causes.

Although it makes no substantive sense in this example to say that I causes F , even if it were so, the path coefficients of the effect of I and F on G (c and d) would not change. Similarly, if no model for the relationship between the exogenous variables is posited and a curved line is drawn between them, the estimates of c and d again remain the same. This result follows from the earlier rule that multiple regression correctly estimates structural coefficients if the disturbances are

uncorrelated with the specified causes. Nothing need be said about the causal relationships among the causal variables. Even if the structural model among the causal variables contains specification error, estimates of the paths to the endogenous variables will not be affected by this error.

PARTIAL CORRELATION

Consider a different causal model for our three variables:

$$\begin{aligned} I &= aF + bU \\ G &= cF + dV \end{aligned}$$

where $e = r_{UV} \neq 0$ and $r_{UF} = r_{VF} = 0$. As with the model in Figure 4.1, this model assumes F causes I and G . But in this case the disturbances U and V are correlated because various common causes have been omitted. Because there are three free parameters, a , c , and e , and three correlations, the parameters may be just-identified or saturated. From the path diagram in Figure 4.2, the three correlations obtained by the tracing rule are:

$$\begin{aligned} r_{FI} &= a \\ r_{FG} &= c \\ r_{IG} &= ac + bde \end{aligned} \tag{4.9}$$

Path coefficients a and c are directly estimated by the correlations. Parameters b and d can be solved from the multiple correlations

$$\begin{aligned} b &= (1 - r_{FI}^2)^{1/2} = .968 \\ d &= (1 - r_{FG}^2)^{1/2} = .969 \end{aligned}$$

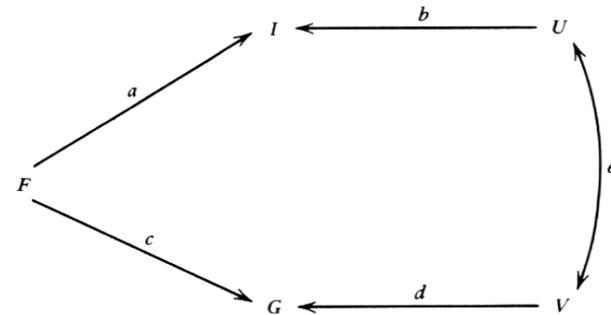


Figure 4.2 Correlated disturbances.

Because a , b , c , and d are known, one can solve for e from Equation 4.9:

$$\begin{aligned} &= \frac{r_{IG} - ac}{bd} \\ &= \frac{r_{IG} - r_{FI}r_{FG}}{[(1 - r_{FI}^2)(1 - r_{FG}^2)]^{1/2}} \end{aligned} \tag{4.10}$$

Solving through obtains .544 as an estimate of e . Some readers may recognize Equation 4.10 as the partial correlation of I and G with F partialled out, or as it is symbolized $r_{IG.F}$. A partial correlation can be tested against the null hypothesis that it equals zero by the following formula:

$$t_{(N-k-2)} = \frac{r_p(N-k-2)^{1/2}}{(1 - r_p^2)^{1/2}} \tag{4.11}$$

where r_p is the partial correlation and k is the number of variables partialled. (The preceding formula is no different from the standard test of a correlation.) Testing the estimate of e yields a $t(764) = 17.92$. If one goes back to compare the test of significance with the test of d in Figure 4.1, one notes that both yield identical values within rounding error. This is no accident since they are equivalent. Blalock (1964) had

originally suggested using partial correlations to test assumptions about path coefficients. A better approach is to use multiple regression which besides providing significance tests also estimates the structural coefficients. Although partial correlations and regression coefficients may be related, the choice of which to compute depends on the substantive meaning the researcher wishes to attach to a coefficient. A partial correlation is the correlation between the unspecified causes of two endogenous variables, while a regression coefficient is the path from a causal variable to an endogenous variable.

At present, for the correlation between disturbances to be identified, it must be assumed that neither endogenous variable causes the other. For instance, a path drawn from I to G in Figure 4.2 would create an underidentified model since there are four free parameters and only three correlations. Thus, to estimate the correlation between disturbances, the assumption must be made that neither endogenous variable causes the other, that is, neither variable appears in the equation of the other.

Many texts give the formula for higher-order partials and computer programs will compute them. If you are caught without a computer, remember that the partial is the correlation between residuals from two regression equations using the same predictor variables. One need only compute the covariance between residuals and their variances. The actual residuals need not be computed, but only their theoretical variances and covariance must be.

SEMIPARTIAL CORRELATION

There is yet another model for the three variables. For this model the disturbance is a cause of two endogenous variables:

$$\begin{aligned} I &= aF + bU \\ G &= cF + dU + eV \end{aligned}$$

where $\rho_{UF} = \rho_{FV} = \rho_{UV} = 0$. The correlations can be obtained by applying the tracing rule to the path diagram in Figure 4.3. They are

$$\begin{aligned} r_{FI} &= a \\ r_{FG} &= c \\ r_{GI} &= ac + bd \end{aligned}$$

The solution for both a and c is directly given again. The estimate b is $(1 - r_{FI}^2)^{1/2}$, or .968. From the correlation between G and I one can solve for d :

$$\begin{aligned} d &= \frac{r_{GI} - ac}{b} \\ &= \frac{r_{GI} - r_{FI}r_{FG}}{[(1 - r_{FI}^2)^2]^{1/2}} \end{aligned}$$

which is the semipartial correlation, or part correlation, between G and I controlling for F , or as it is symbolized $r_{G(I.F)}$. A test of the semipartial is given by

$$t(N-m-1) = \frac{r_s(N-m-1)^{1/2}}{(1-R^2)^{1/2}} \quad [4.12]$$

where r_s is the semipartial, m is the number of causes of the variable that is caused by the disturbance (e.g., G in Figure 4.3), and R^2 is the multiple correlation of that variable. This R^2 can be simply computed by first computing R^2 with the causal disturbance omitted and then adding to it the squared semipartial. For instance, for Figure 4.3 d is estimated by the value of .527, R^2 is then $.248^2 + .527^2 = .339$, and the test of statistical significance is

$$t(764) = \frac{.527(764)^{1/2}}{(1-.339)^{1/2}}$$

Again this is the same as was obtained for the test of significance of d in Figure 4.1 and of e in Figure 4.2. However, the interpretation is

different. The semipartial is the effect of a disturbance of one endogenous variable on another endogenous variable. At present, to identify this path it must be assumed that neither of the endogenous variables causes the other. Note that the semipartial is a nonsymmetric measure of association. Ordinarily $r_{1(2,3)}$ will not equal $r_{2(1,3)}$. The semipartial is useful in the computation of regression coefficients in stepwise regression. We should not confuse this usefulness with its causal interpretation. The semipartial has not and probably should not have much role in structural analysis.

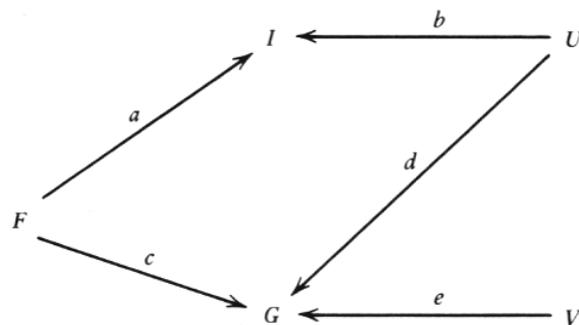


Figure 4.3 Path from a disturbance to two endogenous variables.

EXAMPLE

It has been shown that beta coefficients estimate path coefficients if the unspecified causes are uncorrelated with the causal variables, that partial correlations estimate the correlation of the disturbances of two endogenous variables if neither variable appears in the equation of the other, and that the semipartial estimates a path of a disturbance to another endogenous variable, again if neither endogenous variable appears in the equation of the other. Here these principles are applied to determine the statistical methods for estimating causal parameters of the four models in Figure 4.4. All four models have in common the characteristic that they are just-identified. Each model has four observed variables, and therefore six correlations;

and since each model has six free parameters, each model may be just-identified. Those six parameters can be identified as follows:

Model I: Parameters a , b , and c are simply estimated by the correlations between the exogenous variables. To estimate d , e , and f regress X_4 on X_1 , X_2 , and X_3 , and the resulting beta weights are estimates of path coefficients. Parameter g is estimated by $(1 - R^2)^{1/2}$, where R is the multiple correlation of X_4 with $dX_1 + eX_2 + fX_3$.

Model II: Parameter a is estimated by regressing X_2 on X_1 . Parameters b and c are estimated by regressing X_3 on X_1 and X_2 , and d , e , and f by regressing X_4 on X_1 , X_2 , and X_3 . The estimates of d , e , and f are the same for Models I and II. The difference between the two models is that Model I makes no statement about the causal relationships between X_1 , X_2 , and X_3 , while Model II does. The disturbances of g , h , and i can be estimated by $(1 - R^2)^{1/2}$, where R is the multiple correlation for each equation.

Model III: To estimate parameter a regress X_2 on X_1 ; b , regress X_3 on X_1 ; and c , regress X_4 on X_1 . Since there is only one predictor variable in each equation, the estimated causal coefficient equals the correlation of cause and effect. Parameters h , i , and j can be estimated by partial correlations since X_2 , X_3 , and X_4 do not cause each other: $h = r_{23.1}$, $i = r_{24.1}$, and $j = r_{34.1}$. The disturbances can be estimated by $(1 - R^2)^{1/2}$.

Model IV: Parameter a is simply estimated by r_{12} . Parameters c and d can be estimated by regressing X_3 on X_1 and X_2 , and b and e by regressing X_4 on X_1 and X_2 . Since X_3 does not cause X_4 or vice versa, h is simply the semipartial correlation of X_3 with X_4 controlling for X_1 and X_2 or as it is symbolized $r_{3(4,12)}$. The paths for the residuals are estimated in the usual way. Remember to include U as a cause of X_4 in the multiple correlation for X_4 .

The beta weights of each of the preceding four models can be tested against the null hypothesis that they are zero by applying the t test given by Equation 4.5, the partial correlations in Model III by the t test given by Equation 4.11, and the semipartial in Model IV by Equation 4.12.

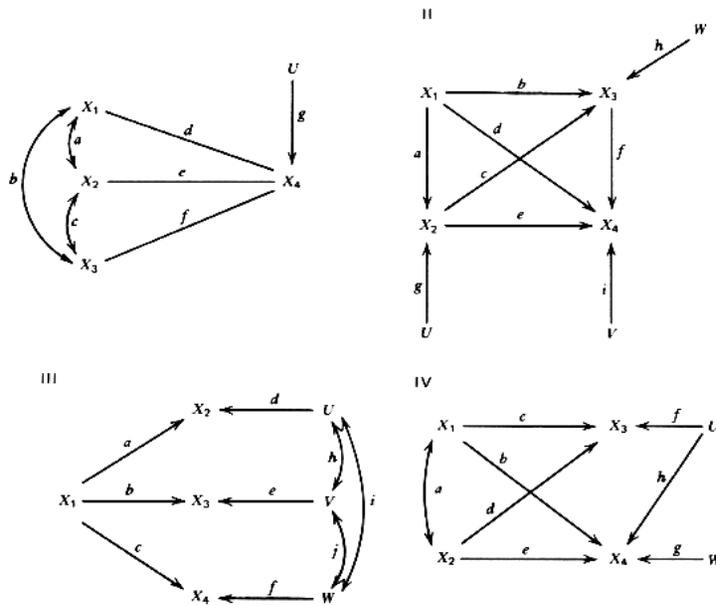


Figure 4.4 Just-identified regression models ($X_1, X_2, X_3,$ and X_4 are measured).

After estimating the coefficients and testing them against zero, one begins what is perhaps the most interesting part of causal analysis. If certain beta coefficients are not significantly different from zero, the researcher should question the utility of a path there. But as in any significance test there may be errors of inference. There may actually be a path but either the sample size is too small or the causal variable is too highly correlated with the other causal variables. (This second problem is called multicollinearity, a problem examined later in this chapter.) But given high power, a zero path should make the researcher question the validity of a causal linkage, at least for the population at hand.

The researcher should also examine the size (high, medium, or low) and the direction (positive or negative) of the paths. If they are anomalous, then the researcher should question the specification of the

model. For instance, if one were to find that the path from actual similarity to attraction is negative, one should suspect that something is wrong with the model. Unfortunately, it may be necessary to estimate many different causal models and find the model that makes the best sense. Obviously, inference is stronger if only one model is stated a priori (without looking at the data) and data confirm that model than if a model is fitted to the data. Causal modeling is meant to be a confirmatory method (Tukey, 1969) but we often find ourselves using it in an exploratory mode.

Even when causal modeling is perfectly applied, it can only confirm or disconfirm that a given model fits the correlations; it never proves the model. An infinite number of models can fit any set of correlations. Ideally the researcher designs the study in such a way that two or more theoretically opposing causal models have divergent implications for the correlations. To increase the validity of a causal model, it is useful to add to the structural equations a variable that one theory says is a cause and another theory says is not. If the first theory is correct, then the relevant path should be nonzero, while for the other theory it would be zero. Unfortunately, researchers are reluctant to spend their time measuring variables that they do not believe to be important. But a model can be strengthened if certain variables show the desired zero paths and other variables the desired nonzero paths. This gives the model both convergent and discriminant validity (see Chapter 13); that is, theory is confirmed in terms of both what should and what should not happen.

To illustrate this point, assume that a researcher did not believe that there was a path from X_2 to X_3 for Model II in Figure 4.4. The researcher should still estimate the path, and test whether it is zero. Along the same lines, in Model IV, if the researcher did not believe that there was a path from U to X_4 , the researcher should still estimate the path by $r_{3(4.12)}$ and test whether it significantly differs from zero.

TECHNICAL ISSUES

OVERIDENTIFICATION

Returning to Figure 4.4, for Model II, if the path labeled c is deleted, the model is overidentified. There is one less parameter than correlation. The early path analysts originally suggested a seemingly sensible strategy for overidentification. Examine the model in Figure 4.5. The model is overidentified since there are three correlations and only two free parameters. As with many overidentified models, it is possible to obtain two estimates of all the parameters of the model: $a = r_{12}$ and $a = r_{13}r_{23}$, and $b = r_{23}$ and $b = r_{13}/r_{12}$. Setting both the estimates equal and restating in terms of population parameters yields the overidentifying restriction of

$$\rho_{13} - \rho_{12}\rho_{23} = 0 \quad [4.13]$$

Because there are two estimates of a parameter, how should the two estimates be pooled? A simple solution would be to add the two together and divide by two. Goldberger (1970) has shown that the statistically best solution is to use the least-squares solution and ignore any other solution. The least-squares solution is more efficient than a pooled solution. Thus, if the theory is correct, one should estimate the path coefficients by regression analysis yielding r_{12} as an estimate of a and r_{23} as an estimate of b . I would not argue for exactly this solution. It is possible that there is a specification error of a path from X_1 to X_3 . If there is such a path, the estimate of the path from X_2 to X_3 will be distorted through this specification error. It might be wiser to regress X_3 on both X_1 and X_2 and to expect the path from X_1 to X_3 to be nonsignificant. The estimate of that path would be

$$\frac{r_{13} - r_{12}r_{23}}{1 - r_{12}^2}$$

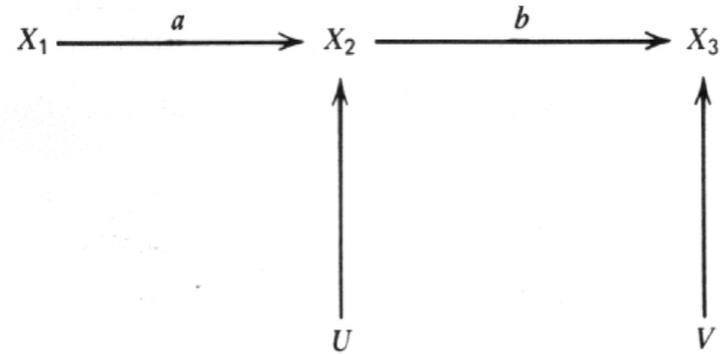


Figure 4.5 Causal chain.

Note that for the preceding to equal zero, Equation 4.13 must hold. Thus, the overidentifying restriction is tested by the prediction of the zero path. A price is paid for this strategy. The variance of the estimate of b is increased, but at least the researcher knows that b is not biased if X_1 causes X_3 . We pay for less bias by decreased efficiency.

It is my opinion that for models for which regression is appropriate, the model should always be just-identified; that is, as many parameters as correlations should be estimated. Theory will specify which parameters will be zero and nonzero. The researcher also has a choice of what connection to have among the endogenous variables. For instance, for Model IV in Figure 4.4, if there were no path from U to X specified, there would be a number of alternatives for adding another parameter. One could have (a) path from X_3 to X_4 , (b) a path from X_4 to X_3 , (c) a path from U to X_4 , (d) a path from V to X_3 , or (e) correlated disturbances between U and V . Whatever additional parameter one chooses to estimate, the null hypothesis of a zero value tests the same overidentifying restriction and the same t statistic is obtained.

SPECIFICATION ERROR

Too often researchers examine the simple, or raw, correlation coefficient as an indication of causal effects. The naive logic is that if X causes Y , then X and Y should be correlated, and if X does not cause Y , they should be uncorrelated. Neither statement is true. After controlling for other causal variables, a strong relationship can vanish and a zero relationship can become strong. To illustrate both of these effects, imagine a researcher who believes that three variables, X_1 , X_2 , and X_3 , cause X_4 . The researcher's implicit causal model is

$$X_4 = aX_1 + bX_2 + cX_3 + dU$$

with the assumption that U is uncorrelated with X_1 , X_2 , and X_3 . Suppose the correlations of three causes with X_4 are

$$\begin{aligned} r_{14} &= .00 \\ r_{24} &= .45 \\ r_{34} &= .45 \end{aligned}$$

Given these correlations, the researcher immediately jumps to the conclusion that X_2 and X_3 are relatively strong causes of X_4 while X_1 has no effect. We, however, would know better. A correlation between a hypothetical cause and effect is a function of not only the causal effect, but also the correlation of the cause with the other causes. This can be seen by applying the first law:

$$\begin{aligned} r_{14} &= a + br_{12} + cr_{13} \\ r_{24} &= ar_{12} + b + cr_{13} \\ r_{34} &= ar_{13} + br_{23} + c \end{aligned}$$

Knowing that $r_{12} = -.3$, $r_{13} = .5$, and $r_{23} = .6$ allows one to solve for a , b , and c by multiple regression. The resultant values are $c = -.3$, $b = .0$, and $a = .6$. So, although X_1 is uncorrelated with X_4 , X_1 still has an effect

on X_4 ! (The variable X_1 is sometimes called a *suppressor variable* [Lord & Novick, 1968, pp. 271-272].) Also, although X_2 is correlated with X_4 , its causal effect is zero! Clearly one must go beyond the simple raw correlation to make causal interpretations.

Table 4.2. Path Coefficients with One Omitted Variable

Path Coefficient	Omitted Variable			
	X_1	X_2	X_3	None
a	—	-.300	.148	-.300
b	.281	—	.495	.000
c	.281	.600	—	.600

What happens to the estimates of causal parameters if the researcher omits one of the causal variables? If either X_1 or X_3 are omitted, the other causal variables' path coefficients are biased since the variables are correlated with disturbances, and thus the model contains specification error. In Table 4.2 are the estimates of path coefficients with one causal variable omitted. If X_2 is omitted, the estimates of the path coefficients are unchanged since X_1 and X_3 remain uncorrelated with the disturbances. As was stated earlier, including variables that have no causal effects in no way biases the estimates of path coefficients. In fact, it adds increased validity to the model by making an explicit prediction of a zero path. Returning to Table 4.2, if the model contains specification error by omitting either X_1 or X_3 , the causal coefficients become misleading. For example, when X_3 is omitted, the path from X_1 to X_4 is .148, when it should be $-.3$. This bias is due to X_1 being correlated with the disturbance since X_3 , which causes X_4 , is contained in the disturbance and is correlated with X_1 .

To see more formally the effect of an omitted variable, examine the path diagram in Figure 4.6. If variable X_3 is omitted then the estimate of d is

$$\begin{aligned} \frac{r_{14} - r_{12}r_{24}}{1 - r_{12}^2} &= \frac{d + er_{12} + fr_{13} - r_{12}(e + dr_{12} + fr_{23})}{1 - r_{12}^2} \\ &= \frac{d(1 - r_{12}^2) + fr_{13} - r_{12}r_{23}}{1 - r_{12}^2} \\ &= d + f\beta_{31.2} \end{aligned}$$

Thus there is no bias in estimating the effect of a causal variable if either the omitted variable does not cause the endogenous variable or the regression of the omitted variable on the causal variable controlling for the remaining causal variables is zero.

An example of the necessity of controlling for variables can be taken from Fine (1976). He compared the ability of psychics and nonpsychics to predict the future. He took 59 predictions of psychics from the National Enquirer and 61 predictions from nonpsychics. He then had observers rate how likely the predictions were to occur. A year later he had a different set of observers judge whether the events had occurred during the previous year. There are then three variables: psychic-nonpsychic (P), likelihood of outcome (L), and actual outcome (A). The psychic-nonpsychic variable is a dummy variable in which one is assigned for psychic and zero for nonpsychic. The sample size is 120, the total number of predictions.

The correlations among the three variables are

$$\begin{aligned} r_{PL} &= -.440 \\ r_{PA} &= -.329 \\ r_{LA} &= .572 \end{aligned}$$

The simple correlation of being a psychic with actual outcome is highly statistically significant, $t(118) = -3.78$, indicating psychics did poorer than nonpsychics. Estimating the following structural equation:

$$A = aP + bL + cU$$

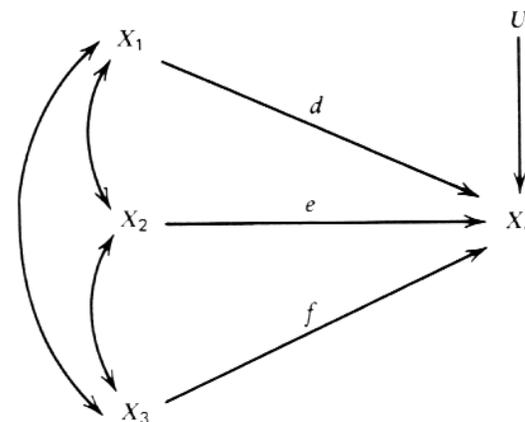


Figure 4.6 The omitted variable problem.

yields an estimate a to be $-.095$, b to be $.530$, and c to be $.816$. Thus, once one has controlled for the likelihood of the outcome, the difference between psychics and nonpsychics is reduced substantially. The test of a is

$$\begin{aligned} t(117) &= \frac{[.335 - .327](117)]^{1/2}}{.665^{1/2}} \\ &= 1.14 \end{aligned}$$

and of b

$$\begin{aligned} t(117) &= \frac{[.335 - .108](117)]^{1/2}}{.665^{1/2}} \\ &= 6.31 \end{aligned}$$

and thus the effect of being a psychic is not significant. Even though being a psychic correlates negatively with the outcome, it correlates about as negatively as it should, given that psychics make predictions that are unlikely and unlikely predictions do not usually come true.

MULTICOLLINEARITY

Multiple regression estimates the effect of a predictor variable on a criterion taking into account both the correlations between the predictor variable and the other predictor variables and the effects of the other predictor variables. As is explained in Chapter 10, for classical experimental design the predictors, or independent variables, are uncorrelated when equal numbers of subjects are assigned to each cell. If the independent variables are uncorrelated, a simpler, more elegant computational strategy can be employed: the analysis of variance. However, the strategy of partitioning of variance makes little or no sense in the case when the independent variables are correlated (Duncan, 1970, 1975). As the causal variables become more and more intercorrelated, the standard errors of path coefficients increase. This increased uncertainty in regression coefficients due to intercorrelated predictor variables is called multicollinearity. Increased standard errors means lower power. Gordon (1968) provides an excellent review of some of the obvious and not so obvious effects of multicollinearity. Much has been made of the "problem of multicollinearity," but I prefer to see it not so much a problem as a cost that is paid by not employing an experimental factorial design with equal cell size (see Chapter 10). It may be a cost that is well worth paying to increase external validity (Campbell & Stanley, 1963).

Before discussing the effects of multicollinearity, its measurement should be first reviewed. Most measures of multicollinearity center around the multiple correlation of a predictor variable with the other predictors. A helpful fact not usually mentioned in most multivariate texts is that the diagonal elements of the inverse of the correlation matrix are $1/(1 - R_i^2)$, where R_i is the multiple correlation of variable i with the other predictor variables. Some computer programs output *tolerance* which is usually defined as $1 - R_i^2$, where R_i is defined as previously (Harris, 1975, p. 283). As R_i increases, multicollinearity becomes more of an issue.

Multicollinearity is literally built into a set of structural equations. If X_1 causes X_2 and X_2 causes X_3 , it is all but inevitable that X_1 and X_2 are correlated. If X_1 is a strong cause of X_2 , it may be difficult, if not impossible, to disentangle the causal effects of X_1 and X_2 on X_3 with a small sample. For example, in Figure 4.5, if a equaled .87 and b equaled .20, one would need close to 783 units to have a 80% chance of finding the estimate of a statistically significant at the .05 level. However, if a were zero and b were .20, one would need only 193 units to have 80% power (cf. Cohen, 1969, p. 99). Thus, causal effects among the causal variables tend to reduce power through multicollinearity.

Multicollinearity also enters structural analysis when two or more measures of the same cause are employed. Imagine that

$$X_2 = .3X_1 + .954U$$

where $r_{1U} = 0$. Suppose the researcher does not measure X_1 but has two errorful measures of X_1 :

$$\begin{aligned} X_3 &= .9X_1 + .436E \\ X_4 &= .9X_1 + .436F \end{aligned}$$

where $\rho_{1E} = \rho_{1F} = \rho_{EU} = \rho_{FU} = 0$, and $\rho_{EF} = 0$. The path model for the three equations is in Figure 4.7. The correlations between the measured variables are $\rho_{34} = .81$, $\rho_{23} = .27$, and $\rho_{14} = .27$. Now suppose the researcher estimates the following equation:

$$X_2 = aX_3 + bX_4 + cV$$

The estimates for a and b are both .149 and the multiple correlation is .284 which approaches .3. The surprising fact is that whereas the theoretical .27 correlation of either X_3 and X_4 with X_2 needs only 54 units to be significant about 50% of the time, the theoretical beta weight of .149 needs 474 units to be significant about 50% of the time,

again using the .05 level of significance. This is a case in which two is not better than one. Adding another measure of a cause increases the multiple correlation but it tends both to make beta weights smaller and to make the standard errors of the beta weights larger.

As this example illustrates, special care should be taken not to include in the regression equation more than one measure of the same construct. I do not mean to imply that a construct should not have more than one measure. Quite the opposite, the multiple indicator strategy is often useful, as is demonstrated in Chapter 8. But the researcher should be suspicious of highly correlated causal variables in a regression equation, since it may well be that they are multiple indicators of the same cause. It is not sufficient for the researcher simply to define two variables as measures of different constructs. If the two variables correlate as highly as the average of their reliabilities, then the measures have no discriminant validity (Campbell & Fiske, 1959). It is not enough to say that measures tap different constructs; it should be demonstrated empirically.

Multicollinearity becomes an insuperable problem when the correlation between two causal variables is unity. To see this, examine the following structural equation:

$$X_3 = aX_1 + bX_2 + cU$$

where $r_{1U} = r_{2U} = 0$. By the first law the correlations are

$$\begin{aligned} r_{13} &= a + br_{12} \\ r_{23} &= ar_{12} + b \end{aligned}$$

Ordinarily the researcher would solve for a and b , but if $r_{12} = 1$, then $r_{13} = r_{23} = a + b$. There is a single equation left in two unknowns for which there is no solution. This is an example of empirical underidentification.

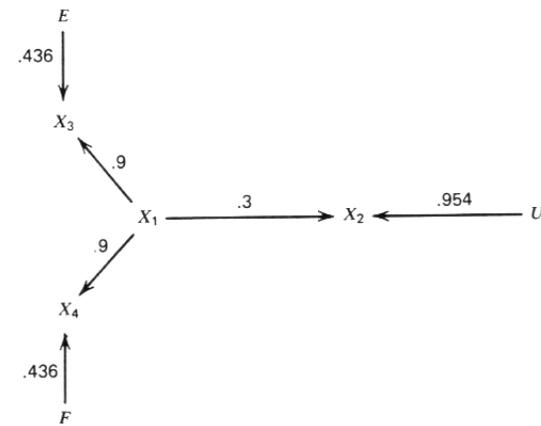


Figure 4.7 Two indicators of an exogenous variable.

Much to our despair and happiness, we rarely obtain correlations or multiple correlations of one; but we can bring about perfect multicollinearity by not being careful. First, as the number of predictors approaches the number of units, multicollinearity increases. Having more predictors than subjects ensures perfect multicollinearity. Second, careless creation of linear composites can result in perfect multicollinearity. Say, for instance, a researcher is interested in the effects of power in heterosexual relationships. The researcher obtains a measure of the resources of the male and the resources of the female. Power in the relationship is then operationalized as the difference between the male's and female's resources. The researcher regresses the involvement of one partner on the two resource variables and power. To the researcher's surprise the computer program blows up and no estimates are possible. This is due to a linear dependency in the regression analysis. Knowing any two of the measures perfectly predicts the third.

Probably the best procedure for the researcher to follow is to regress involvement on the two resource variables. The test of the effect of power is that the regression coefficients for the resource measures are significantly different.

EQUALITY OF REGRESSION COEFFICIENTS

In the previous section it was mentioned that to test the effect of power in heterosexual relationships it should be tested whether the coefficient for male's resources equaled the coefficient for female's resources. It has not, however, been shown how to test the equality of two coefficients within the same equation. The test is relatively simple. Imagine for Equation 4.8 that it is assumed that c equals d . This would imply

$$G = c(F + I) + eV$$

Thus, to test whether two coefficients are equal the researcher compares the R^2 of an equation with both causal variables in the equation with the R^2 of the sum of the two in the equation. If the parameters are equal, nothing should be gained by estimating separate parameters.

For the particular example

$$\begin{aligned} V(F+I) &= V(F) + V(I) + 2C(F,I) \\ &= 1 + 1 + 2(.25) \\ &= 2.5 \end{aligned}$$

and

$$\begin{aligned} C(F+I,G) &= C(F,G) + C(I,G) \\ &= .572 + .248 \\ &= .820 \end{aligned}$$

This makes $r_{F+I,G} = .820/(2.5)^{1/2} = .519$. Because the quantity $(F + I)$ is the only predictor of G , the path from $F + I$ to G is also .519 as is the multiple correlation. Recall that the multiple correlation with both F and I in the equation is .582. The test of whether computing both c and d separately significantly increases R^2 by Equation 4.5 is

$$\begin{aligned} t(764) &= \frac{[(.582^2 - .519^2)764]^{1/2}}{(1 - .582^2)^{1/2}} \\ &= 8.95 \end{aligned}$$

which is highly significant. A researcher should carefully consider whether the assumption of the equality of regression coefficients refers to the standardized or unstandardized coefficient. Ordinarily it refers to the unstandardized coefficient (Cudeck, 1989). The question of standardization is revisited in the last chapter of this book.

Tests of the equality of the same coefficients in the same equation computed from two samples is given in Snedecor and Cochran (1967). More complicated tests of the equality of coefficients is given in Werts, Rock, Linn, and Jöreskog (1976).

MEASURING INDIRECT EFFECTS

Earlier it was stated that a causal variable can have a direct and an indirect effect. For instance, for the path model in Figure 4.8 of the Kerchoff data in Table 4.1, intelligence is a direct and an indirect cause of occupational aspiration. The total effect equals the direct effect plus the sum of the indirect effects. The direct effect is given by the path coefficient. Some early path analysts stated that the total effect is simply the correlation between the effect and the causal variable. However, a little reflection shows that the correlation between variables reflects more than direct and indirect effects.

If the causal variable is exogenous, then part of its correlation with the endogenous variable may be due to its correlation with the other exogenous variables and the effects of these exogenous variables. To see this let us examine the correlation of intelligence (I) with grades (G):

$$r_{IG} = p_{GI} + p_{GS}r_{IS} + p_{GE}r_{IE} + r_{GF}r_{IF}$$

where “*p*” symbolizes a path coefficient. Note that there is a direct effect of intelligence on grades but there are other equivocal effects due to the correlation of intelligence with the other exogenous variables. Since there is no causal model for the exogenous variables, one does not know whether the remaining part of the correlation of *I* with *G* is due to spuriousness, *I* and *G* having common causes, or indirect effects. However, for this particular example, indirect effects are rather implausible; for example, it is implausible to argue that son's IQ causes father's education.

If the causal variable is not exogenous, then part of the correlation between it and the variable that it causes is due to spuriousness. For instance, one reason why grades and educational expectation are correlated is the fact that they share the common causes of *I*, *S*, *E*, and *F*. Spuriousness often explains a considerable portion of the correlation between two variables.

It is then totally illegitimate to find the indirect effects by subtracting the direct effect from the correlation between cause and effect. A highly general method is given by Alwin and Hauser (1975) and it involves the use of reduced form. Let us first translate Figure 4.8 into a set of equations:

$$G = .526I - .030S + .119E + .040F + .807U \quad [4.14]$$

$$X = .160I - .112S + .173E + .151F + .406G + .718V \quad [4.15]$$

$$A = -.047I - .014S - .048E + .094F + .143G + .588X + .731W \quad [4.16]$$

First note that there are no indirect effects of *G*; there are only direct effects. To find the indirect effects of the four exogenous variables on *X* take Equation 4.15 and substitute Equation 4.14 for *G*:

$$\begin{aligned} X &= .160I - .112S + .173E + .151F + .406(.526I - .030S \\ &\quad + .119E + .040F + .807U) + .718V \\ &= (.160 + (.406)(.526))I + (-.112 + (.406)(-.030))S \\ &\quad + (.173 + (.406)(.119))E + (.151 + (.406)(.040))F \\ &\quad + (.406)(.807)U + .718V \end{aligned}$$

$$= .374I - .124S + .221E + .167F + .328U + .718V$$

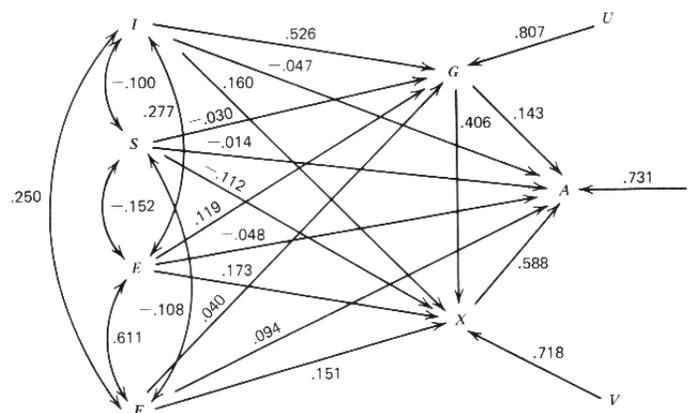


Figure 4.8 Model taken from Kerchoff.

(One could combine the two disturbance terms to form a single disturbance.) To find the indirect effect take the coefficient from the reduced form and subtract off the direct effect to find the indirect effect. So the indirect effects are .214 for intelligence, $-.012$ for number of siblings, .048 for father's education, and .016 for father's occupation.

One need not compute the reduced form coefficients by substituting equations; one could have simply regressed *X* on *I*, *S*, *E*, and *F* to obtain the reduced-form coefficients. When the number of variables is large, this regression procedure is advisable to minimize both computational and rounding errors. However, reduced form is not necessarily needed to find indirect effects. The tracing rule can be used to find the indirect effects. For instance, there is a single indirect effect from *I* to *X* through *G* and it equals $(.526)(.406) = .214$. To find the indirect effect of *G* on *A*, note that it can only go through *X*. The indirect effect is then $(.406)(.588) = .239$.

To find the indirect effect of *I*, *S*, *E*, and *F* on *A* solve for the reduced form. First, substitute Equation 4.14 into *G* of Equation 4.16:

$$\begin{aligned}
A &= -.047I - .014S - .048E + .094F + .143(.526I - .030S \\
&\quad + .119E + .040F + .807U) + .588X + .731W \\
&= (-.047 + (.143)(.526))I + (-.014 + (.143)(-.030))S \\
&\quad + (-.048E + (.143)(.119))E + (.094 + (.143)(.040))F \\
&\quad + (.143)(.807)U + .588X + .731W \\
&= .028I - .018S - .031E + .100F + .115U + .588X + .731W
\end{aligned}$$

Now we substitute the reduced form equation for X into the preceding equation:

$$\begin{aligned}
&= .028I - .018S - .031E + .100F + .115U + .588(.374I \\
&\quad - .124S + .221E + .167F + .328U + .718V) + .731W \\
&= (.023 + (.588)(.374))I + (-.018 + (.588)(-.124))S \\
&\quad + (-.031 + (.588)(.221))E + (.100 + (.588)(.167))F \\
&\quad + (.110) + (.588)(.328))U + .588(.718)V + .731W \\
&= .248I - .091S + .099E + .198F + .308U + .422V \\
&\quad + .731W \qquad [4.17]
\end{aligned}$$

To compute the indirect effect, subtract the direct effects given in Equation 4.16 from the total effects of Equation 4.17. They are then .295 for I , $-.077$ for S , .147 for E , and .104 for F . As a check compute the indirect effects of I on A by the tracing rule. There are three indirect effects: (a) through G , (b) through X and (c) through G then X . These three different values are $(.526)(.143) = .075$, $(.160)(.588) = .094$, and $(.526)(.406)(.588) = .126$, respectively. These three values sum to .295, the same value that was obtained from the reduced form method.

If the reduced form coefficients are computed by hand, there is a straightforward way to check the computations. One checks to see if the reduced form equation has a variance of one. This is done here for the reduced form equation of A . Since the disturbances are uncorrelated with each other, the variance of the sum of the disturbances is

$$\begin{aligned}
V(.308V + .422U + .731W) &= .308^2 + .422^2 + .731^2 \\
&= .807
\end{aligned}$$

To find the variance of the remainder of the equation, simply compute the multiple correlation squared of the reduced form equation:

$$\begin{aligned}
p_{AI}r_{AI} + p_{AS}r_{AS} + p_{AE}r_{AE} + p_{AF}r_{AF} &= (.248)(.335) + (-.091)(-.153) \\
&\quad + (.099)(.303) + (.198)(.331) \\
&= .193
\end{aligned}$$

Because the causal variables are uncorrelated with the disturbance, the variance of variable A equals $.807 + .193 = 1$; thus the reduced form coefficients check out.

CONCLUSION

Multiple regression analysis is the backbone of structural modeling. Beta weights estimate path coefficients if the causes and the effect variables are measured and the disturbance is uncorrelated with the causal variables. Partial correlation estimates the correlation between disturbances when neither endogenous variable causes the other. In the next chapter the problem of the disturbance correlated with an causal variable is considered.

5

MEASUREMENT ERROR IN THE CAUSAL VARIABLES AND THIRD VARIABLES

In the previous chapter it is assumed that the causal variables in a structural equation are measured without error and are uncorrelated with the disturbance. In this chapter violations of these two assumptions are considered. It is demonstrated that an estimation procedure called *two-stage least squares* can be used to estimate such models in some very special circumstances.

The first part of this chapter carefully explores the assumptions of classical measurement theory and recasts them into a path analytic formulation. The second section analytically demonstrates the effect of measurement error in structural equations. The third section illustrates how by adding instrumental variables, measurement error in a causal variable can be allowed. The fourth section considers a general estimation procedure for models with instrumental variables called two-stage least squares. The final section outlines the instrumental variable approach to correcting for bias due to unmeasured third variables that cause both the endogenous and causal variables.

MEASUREMENT ERROR

What sets aside psychometrics from other statistical approaches is the almost excessive concern about measurement error. Traditionally, measurement error had received only cursory coverage

in econometrics. This has since changed and the errors in variable problem, measurement error, has been recently given much more attention. However, psychometrics from its inception has highlighted the problem of measurement error. What follows is a discussion of psychometric models of measurement error using path analysis.

The classical model states that an individual's measured score equals the true score plus error of measurement

$$X = T + E \quad [5.1]$$

where X is measured score, T is true score, and E is error of measurement. It is usually also assumed that the mean of E is zero and the $C(T, E)$ equals zero. The reliability of measure X is defined as

$$\frac{V(T)}{V(X)} \quad [5.2]$$

that is, the ratio of true variance in the measure to the total variance. Reliability then can be thought of as the percent of variance of the measure that is due to the true score. In this text the usual symbol for reliability of measure X is r_{XX} because as is later shown, reliability can be viewed as an a correlation of a variable with itself. If the reliability is known, it is often useful to compute the error variance. It equals

$$\begin{aligned} V(E) &= V(X) - V(T) \\ &= V(X) \frac{(1 - V(T))}{(1 - V(X))} \\ &= V(X)(1 - r_{YY}) \end{aligned} \quad [5.3]$$

The correlation of X with T simply equals

$$r_{XT} = \frac{C(X, T)}{[V(X)V(T)]^{1/2}}$$

$$\begin{aligned}
 &= \frac{V(T)}{[V(X)V(T)]^{1/2}} \\
 &= \frac{[V(T)]^{1/2}}{[V(X)]^{1/2}} \\
 &= r_{XX}^{1/2} \qquad [5.4]
 \end{aligned}$$

Thus the square root of the reliability equals the correlation of the measured score with the true score. The square root of a measure's reliability is the upper limit for the correlation of the measure with any other measure when the errors of measurement of a variable are uncorrelated with all other variables.

Returning to Equation 5.1 it is clear that the variables are not standardized. Note that if $V(X) = 1$, then neither T nor E can be standardized since $V(T) + V(E) = 1$. Thus the classical model cannot be directly cast into a path analysis framework. To do so the measurement model must be reformulated as

$$X = aT + bE \qquad [5.5]$$

where X , T , and E are all now standardized variables. In Figure 5.1, there is a path diagram for Equation 5.5. From the path diagram it follows that $a^2 + b^2 = 1$ and, therefore, $b = (1 - a^2)^{1/2}$. It also follows that $r_{TX} = a$ and $r_{XX} = a^2$. Thus with standardized variables, the reliability is simply the square of the path from true score to the measure. The path for the error equals the square root of the quantity one minus the reliability.

There are then two different ways to formulate the measurement model of a single variable: the classical approach of Equation 5.1 and the path analytic approach of Equation 5.5. The text focuses on the path-analytic approach, but it is often compared to the classical approach.

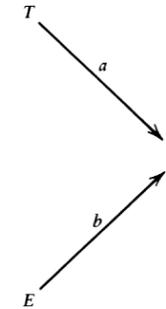


Figure 5.1 Path model for the classical measurement model.

In Figure 5.2, two variables, T_X and T_Y , are assumed to be correlated. However, neither variable is measured directly but each with error. The measure X is the errorful indicator of T_X and Y of T_Y . The errors of measurement, E_X and E_Y , are assumed to be uncorrelated with the true scores and with each other. By the tracing rule it follows that

$$r_{XY} = abr_{TXTY}$$

Therefore the correlation between true X and true Y equals

$$r_{TXTY} = \frac{r_{XY}}{(r_{XX}r_{YY})^{1/2}}$$

since $a^2 = r_{XX}$ and $b^2 = r_{YY}$. The preceding formula is the classical correction for attenuation formula. It states that the correlation between true scores equals the correlation between the measured variables divided by the geometric mean of the reliabilities of the measures. The correlation between the measured variables is said to be attenuated, and to be equal to the true correlation times some number that is less than or equal to one. The effect, then, of measurement error is to make

correlations between measures underestimate the absolute value of the true correlation.

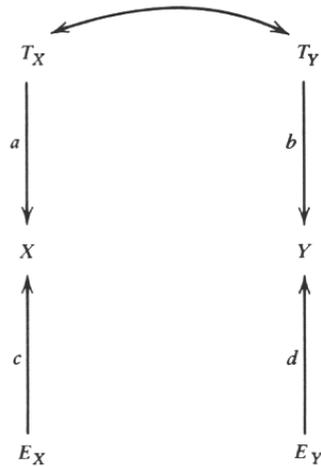


Figure 5.2 Correction for attenuation.

Because correlations are biased by measurement error, it should be obvious that ignoring measurement error in variables may seriously distort estimates of structural parameters. Later in this chapter, bias in structural parameters is discussed. Various classical approaches to the estimation of measurement error are considered now.

Typically, estimation of reliability proceeds through the use of other measures. We are all probably familiar with the use of alternative forms or test-retest to estimate reliability. In some sense the logic of these procedures can be seen by examining the path diagram in Figure 5.3. True score T_X causes both X_1 and X_2 . Their errors of measurement are mutually uncorrelated and uncorrelated with the true score. It is also specified that X_1 and X_2 are equally reliable ($r_{X_1X_1} = r_{X_2X_2} = a^2$). Note that it then follows that given uncorrelated errors of measurement the correlation between two equally reliable indicators of the same construct estimates the reliability of the measures. In fact such a correlation is sometimes given as the very definition of reliability.

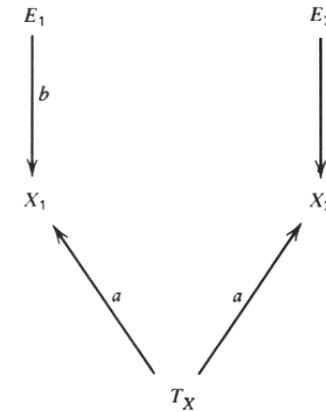


Figure 5.3 Multiple measures model.

Classical test theory distinguishes between three different types of measures: parallel, tau equivalent, and congeneric. In this text variables are ordinarily considered in mean deviation form and so only the assumptions made about the equivalence of variances need be considered. Let two measures, X_1 and X_2 , have true scores whose correlation is one such that

$$\begin{aligned} X_1 &= T + E_1 \\ X_2 &= kT + E_2 \end{aligned}$$

where $C(E_1, E_2) = C(E_1, T) = C(E_2, T) = 0$. Measures X_1 and X_2 are said to be parallel if $k = 1$ and $V(E_1) = V(E_2)$. It then follows that parallel measures have equal variances and equal covariances; that is, if X_1 , X_2 , and X_3 are parallel measures, it follows that $V(X_1) = V(X_2) = V(X_3)$ and $C(X_1, X_2) = C(X_1, X_3) = C(X_2, X_3) = V(T)$.

Tests X_1 and X_2 are said to be tau equivalent if $k = 1$. Unlike parallel measures, tau equivalent measures may have unequal error variances. However, the covariance between tau-equivalent measures is constant; that is, if X_1 , X_2 , and X_3 are tau equivalent then $C(X_1, X_2) = C(X_1, X_3) = C(X_2, X_3) = V(T)$.

Congeneric measures are the most general. Error variances may be unequal and the parameter k need not equal one. The only constraint on the covariances is that they are single-factored, which is extensively discussed in Chapter 7. Different methods of estimating reliability make different assumptions about the measures. For instance, the simple correlation between two measures (assuming uncorrelated errors of measurement) estimates the reliability of only parallel measures. A test-retest correlation also estimates the reliability of parallel measures given no true change. Cronbach's alpha estimates the reliability of the sum of the measures only if the measures are tau-equivalent or parallel. If the measures are congeneric then it underestimates the reliability of the sum.

Standardizing measures may alter their character. Parallel measures remain parallel after standardization since their variances are equal. However, standardization turns tau-equivalent measures into congeneric measures. Ordinarily congeneric measures remain congeneric after standardization with the interesting exception that they become parallel if the original measures were all equally reliable.

Both parallel and tau-equivalent specifications imply that the true score is in the units of measurement of the measured variable. For both cases the true score has a structural coefficient of one. These assumptions then are harmonious with the previously designated classical model of Equation 5.1. Congeneric measures are consistent with the path analytic assumptions of Equation 5.5 since with congeneric measures the true score need not be in the metric of the measure.

MEASUREMENT ERROR IN STRUCTURAL EQUATIONS

The effect of measurement error on the path coefficients is different for the endogenous and causal variable. First error in the endogenous variable is considered and then error in the causal variable is examined. Assume two causal variables, Z and X , cause Y_T :

$$Y_T = aZ + bX + cU$$

where the disturbance U is assumed to be uncorrelated with Z and X . Both Z and X are measured without error, but Y_T is fallibly measured by Y :

$$Y = Y_T + E$$

where the covariances of E with Y_T , Z , X , and U are zero. If Y , Y_T , and E are standardized then the equation for Y is

$$Y = dY_T + eE$$

where d^2 is the reliability of Y . The unstandardized regression coefficients b_{YZX} and b_{YXZ} yield a and b , respectively. They are unbiased because measurement error in Y is absorbed into its disturbance. However, if the variables are standardized the beta weights equal ad and bd . Because d is less than or equal to one, these beta weights will be less than or equal to the absolute value of the true structural coefficient. To summarize, under the classical specification, error of measurement in the endogenous variable does not bias the unstandardized coefficient but it does attenuate the standardized coefficient. To correct the beta weight for attenuation one should divide it by the square root of the reliability of the endogenous variable. Note that if the true beta weight is zero no bias results.

Unfortunately measurement error in the causal variable is not nearly so simple. Consider first the following structural equation

$$Y = aX_1 + bU$$

where $r_{UXT} = 0$. The variable Y is measured without error and the structural equation for X is

$$X = cX_T + dE$$

where E is uncorrelated with both X_T and U . It is assumed that all variables are standardized which makes c^2 the reliability of X . The path from X to Y is then ac . Thus there is attenuation since the true path is a . Even if one assumes the classical specification, the path from X_1 to Y is attenuated by c , the square root of the reliability of X . Consider now two causal variables, X_T and Z and an endogenous variable Y . As in Figure 5.4, X_1 is measured with error by X . Assuming all the variables in the figure are standardized, the correlations are

$$\begin{aligned} r_{XY} &= d(a + fb) \\ r_{XZ} &= df \\ r_{YZ} &= b + af \end{aligned}$$

the "path coefficient" from X to Y is then

$$\begin{aligned} \beta_{YX.Z} &= \frac{r_{XY} - r_{YZ}r_{XZ}}{1 - r_{XZ}^2} \\ &= \frac{da + dbf - (df)(b + af)}{1 - d^2f^2} \\ &= \frac{da(1 - f^2)}{1 - d^2f^2} \end{aligned} \quad [5.6]$$

It can be shown that $d(1 - f^2)/(1 - d^2f^2)$ equals $\beta_{(X_T)X.Z}$ or d_Z^2/d where d_Z^2 equals the reliability of X after Z has been partialled out. If f^2 is greater than zero, then d_Z^2/d is less than d . Thus the bias in Equation

5.6 is an attenuation bias. Its size depends mainly on the reliability of X . The beta coefficient of Y on Z controlling for X equals

$$\begin{aligned} \beta_{YZ.X} &= b + \frac{af - (df)[d(a + bf)]}{1 - d^2f^2} \\ &= b + \frac{af(1 - d^2)}{1 - d^2f^2} \end{aligned}$$

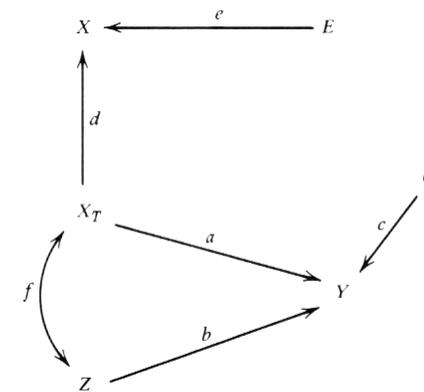


Figure 5.4 Errorful exogenous variable.

Thus, the beta weight estimates the true path plus a biasing term. That term equals zero if any of the following three conditions hold:

1. The path from X_T to Y is zero ($a = 0$),
2. the reliability of X is one ($d^2 = 1$), or
3. the correlation of X_T with Z is zero ($f = 0$).

Because X_T is measured with error, one cannot successfully partial out its effects. This can be more clearly seen if X is perfectly unreliable. It then makes no sense to say that we control for X_T by X . If X_T and Z are correlated, then Z will be correlated with the disturbance of Y that contains Y_T .

The direction of the bias is not known in practice. If, however, both a and f are positive, the direction of bias is positive. Thus the bias tends to overestimate the path coefficients of the perfectly measured variables, given that the effect of X_T is positive and its correlation with Z is also positive.

If two or more causal variables are measured with error, the effects of measurement error are difficult to predict. There are two sources of bias: First, there is attenuation bias due to error in the causal variable. Second, added to the attenuated estimate of the coefficient is a term whose size is influenced by the unreliability of the other causal variables. It can occur that these two sources of bias cancel each other, but it is unwise to hope for such an outcome.

One must always carefully consider the problem of measurement error in the causal variable. It should not be assumed that its effect will only attenuate estimates of coefficients; it can make the estimate of a truly zero coefficient nonzero or it can yield a coefficient with the opposite sign as the true coefficient. However, a careful analysis may reveal that bias is negligible if

1. Reliability is high.
2. The paths from the true scores to the endogenous variable are small.
3. The causal variables have low intercorrelations with each other.

Besides ignoring the problem there are three analytic strategies to the unreliable causal variable issue. The first is to correct the correlations for attenuation due to measurement error and then to use these disattenuated correlations to estimate causal parameters. Although this method provides consistent estimates of parameters, it has a shortcoming. It divorces reliability estimation from causal modeling. It has been shown that the measurement process can be expressed in terms of a causal model and therefore all the parameters including reliabilities should be considered as part of one model. This

is not the case when one simply disattenuates the correlations. Moreover the resulting parameter estimates cannot be tested for statistical significance in the usual way.

The second strategy for handling error in the causal variables is to have parallel forms or multiple indicators of the same construct, which is pursued in Chapters 7 through 9. This strategy introduces unmeasured variables to the model. The third strategy is the method of instrumental variables, which is considered in the next section of this chapter.

INSTRUMENTAL VARIABLES

Consider the path diagram in Figure 5.5. If there were a path from Y to Z with X , Y , and Z measured, the model would be underidentified since there are four free parameters and only three correlations. The instrumental variable strategy is to buy the identification of the reliability coefficient by assuming a path from an causal variable is zero. In Figure 5.5 there are only three free parameters since there is no path from Y to Z . Note that given that $r_{YZ} = ar_{XTY}$ and $r_{XTY} = r_{XY}/b$, it follows that

$$\frac{a}{b} = \frac{r_{YZ}}{r_{XZ}}$$

Moreover, by the tracing rule $r_{XZ} = ab$. It then follows that

$$b^2 = \frac{r_{XZ} r_{XY}}{r_{YZ}}$$

and

$$a^2 = \frac{r_{XZ} r_{YZ}}{r_{XY}}$$

Because b is usually taken to be positive, the sign of a is determined by r_{XZ} or r_{YZ}/r_{XY} , both of which should have the same sign.

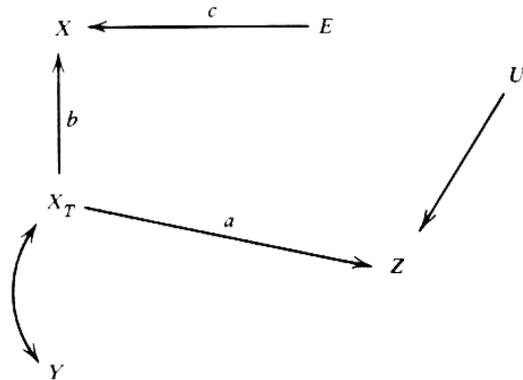


Figure 5.5 Instrumental variable illustration.

If one takes the classical specification of test score theory ($X = X_T + E$), then

$$a = \frac{C(Y,Z)}{C(X,Y)}$$

and

$$r_{XX} = \frac{r_{XZ} r_{XZ}}{r_{YZ}}$$

Clearly to be able to estimate a both r_{XY} and b^2 should not be small; otherwise, empirical underidentification results.

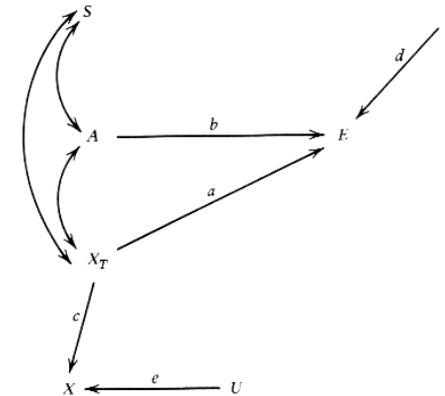


Figure 5.6 Breakdown of instrumental variable estimation.

Instrumental variable estimation can disastrously break down if the model is misspecified. Consider the set of correlations taken from unpublished data of L. Brush in Table 5.1. The model specified in Figure 5.6 is given for illustrative purposes only and was not specified

S	1.00			
A	.21	1.00		
X	.14	.70	1.00	
E	.21	.62	.53	1.00
	S	A	X	E

Table 5.1 Correlations of Sex, Ability, and Expectations for Success in Mathematics.

by Brush. Sex S does not affect expectations E , but rather its effect is mediated through anxiety X and ability A . The anxiety variable is assumed to be measured with error. Using the first law of path analysis

$$r_{ES} = ar_{SXT} + br_{SA}$$

$$r_{EA} = ar_{AXT} + b$$

Because

$$r_{AXT} = r_{AX}/c \text{ and } r_{SXT} = r_{SX}/c,$$

$$r_{ES} = ar_{SX}/c + br_{SA}$$

$$r_{EA} = ar_{AX}/c + b$$

The solution for a/c is

$$\frac{a}{c} = \frac{r_{ES} - r_{SA}r_{EA}}{r_{SX} - r_{AX}r_{SA}}$$

and

$$b = \frac{r_{SX}r_{EA} - r_{AX}r_{ES}}{r_{SX} - r_{AX}r_{SA}}$$

The estimate for a/c is -11.40 and for b is 8.60 using the estimated correlations in Table 5.1. Such results, though not impossible, are highly implausible. If one proceeds to solve for a , r_{AXT} , and d further implausible values arise. These highly anomalous values are possible if instrumental variable estimation is carelessly applied.

Instrumental variable estimation can yield informative estimates as in the next example taken from Jöreskog (1973). Children were measured at fifth (X_1 and X_2), seventh (X_3 and X_4), and ninth grade (X_5 , and X_6). Of interest are tests of mathematical ability (X_1 , X_3 , and X_5) and science knowledge (X_2 , X_4 , and X_6). The correlation matrix is contained in Table 5.2 and the model in Figure 5.7. The two seventh grade measures (X_3 and X_4) are assumed to be measured with error. The wave 1 measures are taken to be instruments and the wave 3 measures as the endogenous variables. Clearly there is measurement error in these variables also, but errors in the endogenous variables only attenuate estimates and error in the instruments only reduces the efficiency of estimation. However, it must be assumed that such errors are mutually uncorrelated.

It can be shown that

$$r_{15} = \frac{ar_{13}}{e} + \frac{br_{14}}{f}$$

$$r_{25} = \frac{ar_{23}}{e} + \frac{br_{24}}{f}$$

$$r_{16} = \frac{cr_{13}}{e} + \frac{dr_{14}}{f}$$

$$r_{26} = \frac{cr_{23}}{e} + \frac{dr_{24}}{f}$$

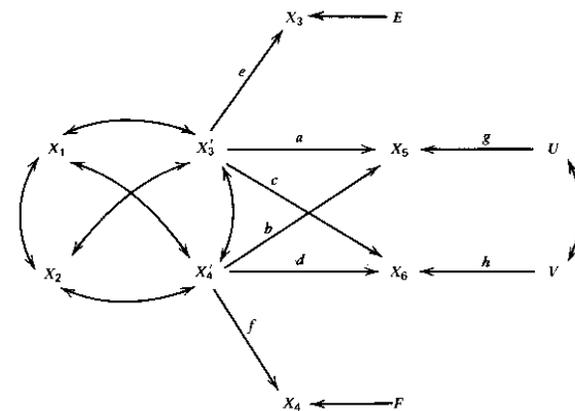


Figure 5.7 Mathematics and science ability example.

The solutions are then

$$\frac{a}{e} = \frac{r_{15}r_{24} - r_{14}r_{25}}{r_{13}r_{24} - r_{23}r_{14}}$$

$$\frac{b}{f} = \frac{r_{13}r_{25} - r_{23}r_{15}}{r_{13}r_{24} - r_{23}r_{14}}$$

$$\frac{c}{e} = \frac{r_{16}r_{24} - r_{14}r_{26}}{r_{13}r_{24} - r_{23}r_{14}}$$

$$e = \frac{r_{13}r_{24} - r_{23}r_{14}}{r_{13}r_{26} - r_{23}r_{16}}$$

$$\frac{d}{f} = \frac{r_{13}r_{26} - r_{23}r_{16}}{r_{13}r_{24} - r_{23}r_{14}}$$

$$f = \frac{r_{13}r_{24} - r_{23}r_{14}}{r_{13}r_{26} - r_{23}r_{16}}$$

Substituting in the correlations from Table 5.2 yields

$$a/e: 1.071$$

$$b/f: 108$$

$$c/e: -.144$$

$$d/f: 1.112$$

It also follows that

$$ae = r_{35} - r_{34}b/f$$

$$bf = r_{45} - r_{34}a/e$$

$$ce = r_{36} - r_{34}d/f$$

$$df = r_{46} - r_{34}c/e$$

yielding

$$ae: .840$$

$$bf: -.088$$

$$ce: -.119$$

$$df: .842$$

It then follows that

$$a: .948$$

$$b: -.097$$

$$c: -.131$$

$$d: .968$$

$$e^2: .784, .824$$

$$f^2: .814, .757$$

Table 5.2 Mathematics and Science Ability Measured at Fifth, Seventh, and Ninth Grades^{a,b}

X ₁	1.000					
X ₂	.755	1.000				
X ₃	.730	.664	1.000			
X ₄	.651	.700	.710	1.000		
X ₅	.712	.636	.763	.673	1.000	
X ₆	.619	.683	.671	.740	.736	1.000
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆

^aData taken from Jöreskog (1973).

^bN = 730.

The two negative estimates seem somewhat anomalous but perhaps proficiency in one area interferes with ability in other areas. Compare the ordinary multiple regression estimates with the preceding and note how markedly different they are:

$$a: .575$$

$$b: .265$$

$$c: .294$$

$$d: .532$$

Since there are two estimates of e^2 and f^2 , a simple pooling strategy is taken by using an arithmetic average. The estimate of $p_{3'4'}$ is then $.71/((.804)(.786))^{1/2} = .893$. The estimates of g and h are

$$g = (1 - a^2 - b^2 - 2abr_{3'4'})^{1/2} = .508$$

$$h = (1 - c^2 - d^2 - 2cdr_{3'4'})^{1/2} = .522$$

and r_{UV} is

$$\frac{r_{56} - adr_{3'4'} - bcr_{3'4'} - ac - bd}{gh} = .463$$

The instrumental variable estimates yield a set of generally plausible and informative values.

TWO-STAGE LEAST-SQUARES ESTIMATION

The classical method of estimating parameters from models with instrumental variables is called *two-stage least squares* and is commonly referred to as *2SLS*. This section first explains the estimation procedure and then discusses the assumptions necessary for such an estimation procedure.

Denote the endogenous variable as Z and its disturbance as U . The set of n variables which are assumed to be measured with error are called X ; the set of p instruments are called I and the set of q causal variables measured without error are called Y . The structural equation then is

$$Z = \sum_i a_i X_i' + \sum_k b_k Y_k + U$$

where the prime designates X_i true. Considered will be two alternative specifications of the measurement model. First the classical model where

$$X_i = X_i' + E_i$$

with $V(X_i')/V(X_i) = c_i^2$. The path analytic specification is

$$X_i = c_i X_i' + (1 - c_i^2)^{1/2} E_i$$

where $V(X_i) = V(X_i') = V(E_i)$. Note this second specification does not assume that the X variables are standardized but only that measured variance equals true variance. For both specifications the reliability of X_i is c_i^2 .

As its name suggests two-stage least squares involves two different multiple regression runs. The first stage involves regressing

each X_i on I and Y . This will involve n separate regression equations. From each regression, one can estimate the predicted X_i or X_i^* given I and Y . Note that if X_i is standardized, the X_i^* will not ordinarily have unit variance. The results of this first stage of estimation can be thought of as an attempt to estimate the true scores of the X_i . Since it will be seen that the measurement error in X_i is uncorrelated with I and Y , the X_i^* contain only true score variance.

Now that the X variables have been purified the second stage can begin. The endogenous variable Z is now regressed on X_i^* and Y . The interpretation of the coefficients for the Y variables is straightforward. They are the estimates of b_k , the structural coefficients for Y_k . The interpretation of the coefficients for the X^* depends on the test theory model that is chosen. If the classical model is chosen, then they are simply the structural coefficients for the true scores. If the path analytic model is chosen another stage of estimation remains. It can be shown that second-stage estimates equal a_i/c_i . It also can be shown that

$$a_i c_i = \frac{C(X_i, Z) - \sum_i C(X_i, X_m) a_m / c_m - \sum_k C(X_i, X_k) b_k}{V(X_i)}$$

Because all the preceding terms on the left-hand side can be estimated from the data (recall that a_m/c_m are the second-stage estimates for X_m), one can solve for an estimate $a_i c_i$. Then with estimates of a_i/c_i and $a_i c_i$, one can solve for a_i and c_i^2 . Note that the estimates of c_i^2 should, in principle, be less than one. Also for a_i not to be imaginary, the solution for $a_i c_i$ and a_i/c_i should both have the same sign.

To estimate the reliability of X_i given the classical test theory specification compute

$$c_i^2 = \frac{C(X_i, Z) - \sum_i C(X_i, X_m) a_m - \sum_k C(X_i, X_k) b_k}{a_i V(X_i)}$$

where a_i , a_m , and b_i are the second-stage coefficients.

The second-stage coefficients cannot be tested for significance in the usual way. Currently, 2SLS is not used very often, but rather a

structural equation modeling program is used to estimate these models. Additionally, estimated X and Y variables may be highly colinear since the Y variables were used to estimate X . The power of the tests of significance may then be rather low.

If there are more instruments than variables needing instruments ($p > n$), the model is overidentified. One advantage of two-stage least squares is that overidentified models present no special problem. For complicated models, however, it is not a simple matter to test the overidentifying restriction. One strategy is as follows: Regress each I on X^* and Y ; compute for each I_j the residual; and denote them as Q . Although there were originally n instrumental variables, only p minus n of the Q variables are independent. Now compute the first $p - n$ principal components of Q . Now include these $p - n$ components in the second-stage regression equation. The test of the overidentifying restrictions is simply whether these principal components' coefficients are zero.

To apply instrumental variable estimation the following assumptions must hold:

1. The errors of measurement in the X variables must be uncorrelated with each other and the other variables in the model.
2. The instruments must not directly cause the endogenous variable.
3. The instruments must be uncorrelated with the disturbance in the endogenous variable.
4. The number of variables needing instruments must be less than or equal to the number of instruments.

The first assumption is a classical test theory specification. The second assumption creates the zero paths which bring about identification. In a sense this is the key assumption. The instrument can cause Z but its effect must be mediated by X or Y . The third assumption is the usual

uncorrelated disturbance assumption. The fourth assumption is self-explanatory.

To avoid empirical underidentification one must assume that the I and X variables share variance independent of the Y variables. For instance, X may cause I or vice versa, or some third variable besides Y may cause them. If there were little or no independent association between X and I , the X^* variables would be virtually entirely explained by the Y variables. Then on the second stage the X and Y variables will be sufficiently colinear to preclude reliable estimation.

Secondly, the estimates of instrumental variable regression will be empirically underidentified if weights for the I variables on two or more X variables are proportional. For instance, if the first-stage equations for X_1 and X_2 were

$$\begin{aligned} X_1^* &= .2I_1 + .3I_2 \\ X_2^* &= .4I_1 + .6I_2 \end{aligned}$$

then the correlation between the two would be 1.0, resulting in perfect multicollinearity between X_1^* and X_2^* in the second stage. If the coefficients of the I variables for X_j^* are linearly or are nearly linearly dependent with those of any other set of other X variables, empirical underidentification results.

One advantage of a two-stage least-squares solution is that it can estimate the parameters of an overidentified model. For some simple models the test of the overidentifying restrictions is rather straightforward. Imagine the case in which X_1 causes X_4 and X_1 is measured with error and both X_2 and X_3 can be used as instruments of X_1 . Because both r_{24}/r_{12} , and r_{34}/r_{13} estimate the same parameter, there is the following overidentifying restriction:

$$\rho_{24}\rho_{13} - \rho_{34}\rho_{12} = 0$$

which is called a vanishing tetrad; its significance test is discussed in Chapter 7.

Although Duncan (1975, p. 98) has argued that the test of the over-identifying restriction in this instance is of little practical value, such a test of the overidentifying restriction should be performed. Take first the case in which the overidentifying restriction is met; that is, the correlations satisfy the overidentifying restriction within the limits of sampling error. Duncan claims that this is of little value since it is a test of the null hypothesis. One should not, according to Duncan, accept that the overidentifying restriction holds, but rather one cannot accept that it does not hold. I do not accept this conclusion. We often find ourselves testing the null hypothesis. In this case, the test of the null hypothesis tells us whether we can safely conclude that either the instrument does not cause the causal variable or is correlated with the disturbance. Since many of us have difficulty of thinking of good instrumental variables, the overidentifying restriction should be helpful in that it allows us to test our assumptions. As with any significance test, sample size should be large and power high. If one stringently objects to testing the null hypothesis, one could set up a reasonable alternative hypothesis and treat it as the null hypothesis.

Duncan is just as forceful in arguing against the usefulness of an overidentifying restriction not being met. For instance, in the case preceding, if the overidentification is not met one does not know whether X_2 or X_3 , or both, are bad instruments. If an overidentifying restriction is not met, it rarely localizes the source of specification error.

However, not meeting the overidentifying restriction clearly tells us that something is amiss. It forces us to rethink and respecify the model. If the model is just identified, we are always lulled into the feeling that the model is correctly specified. Overidentification affords us with many opportunities to be humbled. It does not bring truth or certainty, but like any good empirical test, it offers the possibility of confirmation or rejection of theory. If the model is overidentified the

researcher should make every effort to test the overidentifying restrictions.

UNMEASURED THIRD VARIABLE

Instrumental variable estimation can also be applied to the case in which an causal variable is correlated with the disturbance of the endogenous variable. This would happen if there were an unmeasured third variable that caused both the causal and endogenous variables. There must be as many instrumental variables as there are causal variables that are correlated with the disturbance. Again the instrumental variables by definition cannot cause the endogenous variable or be correlated with the disturbance. There are two important differences for this case from the case in which the causal variable is measured with error. First, the instrument should not be caused by the variable that needs an instrument. If it is caused, then the instrument would most probably be correlated with the disturbance of the endogenous variable. Second, when the path-analytic specification is assumed there is no need to go through a further step to solve for the structural parameters.

Consider the model in Figure 5.8. The variables are taken from Table 4.1 of the previous chapter. The variable X is educational aspirations, E is father's education, G is grades, F is father's occupation, and I is intelligence. Both E and G are assumed to be correlated with the disturbance in X and F and I are instruments. To solve for a and b one could employ two-stage least squares. However, one can also solve for the relevant parameters by path analysis. The two relevant equations are

$$\begin{aligned} r_{XF} &= ar_{FE} + br_{FG} \\ r_{XI} &= ar_{IE} + br_{IG} \end{aligned}$$

The solution then is

$$a = \frac{r_{FX}r_{IG} - r_{IX}r_{FG}}{r_{FE}r_{IG} - r_{IE}r_{FG}}$$

$$b = \frac{r_{IX}r_{FE} - r_{FX}r_{IE}}{r_{FE}r_{IG} - r_{IE}r_{FG}}$$

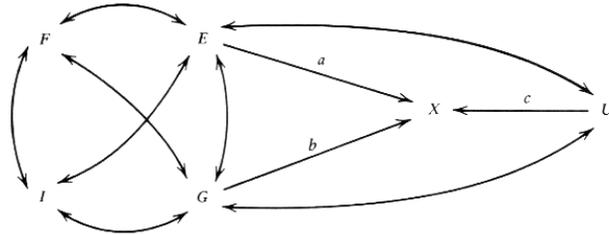


Figure 5.8 Kerchoff example with two exogenous variables correlated with the disturbance.

Substituting in the correlations in Table 4.1 a value of .403 is obtained for a and .660 for b . The solution for c is

$$V(cU) = V(X - aE - bG)$$

$$c^2 = 1 + a^2 + b^2 + 2abr_{EG} - 2ar_{EX} - 2br_{GX}$$

which yields .607 for c^2 and .779 for c . The correlations of the causal variable with the disturbance can be solved from the correlation of the causal variable with the endogenous variable

$$r_{EX} = a + br_{EG} + cr_{EU}$$

$$r_{GX} = ar_{EG} + b + cr_{GU}$$

which yields a value of $-.194$ for r_{EU} and $-.233$ for r_{GU} . These negative values seem implausible.

It is instructive to derive the overidentifying restriction that tests whether $\rho_{EU} = 0$. If ρ_{EU} were zero, then both F and I could be used as instruments for G . There is then an excess of one instrument. The estimate of b , if only F is used as an instrument, is

$$\frac{\rho_{FX} - \rho_{FE}\rho_{XE}}{\rho_{FG} - \rho_{EG}\rho_{FE}}$$

and the estimate of b , if only I is used as an instrument, is

$$\frac{\rho_{IX} - \rho_{IE}\rho_{XE}}{\rho_{IG} - \rho_{EG}\rho_{IE}}$$

Because both should estimate b , the following overidentifying restriction holds:

$$(\rho_{FX} - \rho_{FE}\rho_{XE})(\rho_{IG} - \rho_{EG}\rho_{IE}) = (\rho_{IX} - \rho_{IE}\rho_{XE})(\rho_{FG} - \rho_{EG}\rho_{FE})$$

which implies that

$$\rho_{FX.E}\rho_{IG.E} - \rho_{IX.E}\rho_{FG.E} = 0$$

which is a vanishing tetrad (see Chapter 7) of partial correlations.

This chapter has discussed a number of topics: The theory of measurement error, error in causal variables, and causal variable with disturbance correlations. Chapters 7, 8, and 9 return to the problem of measurement error. In the next chapter it is demonstrated that instrumental variable estimation can be applied to nonhierarchical models.

6

RECIPROCAL CAUSATION

All the models discussed in the previous two chapters have not allowed feedback. Those models did not allow for a variable to both cause and be caused by another variable. In this chapter attention is directed toward nonhierarchical or feedback models, which the econometricians call simultaneity. The chapter is divided into three sections. The first section outlines general considerations of nonhierarchical models. The second section discusses estimation. The third and final section presents a nonhierarchical model for halo effect.

GENERAL CONSIDERATIONS

The simplest nonhierarchical model is

$$X_1 = aX_2 + bU \quad [6.1]$$

$$X_2 = cX_1 + dV \quad [6.2]$$

where $r_{UV} = 0$, as in the path diagram in Figure 6.1. It is later shown that the disturbances of each equation are correlated with the causal variable in the equation, that is, $r_{2U} \neq 0$ and $r_{1V} \neq 0$. Because each variable causes the other, it is not possible to use multiple regression to estimate the path coefficients. There is another obvious problem with the model in Figure 6.1. There are two free parameters, a and c , and only one correlation if only X_1 and X_2 are measured. The model then is underidentified.

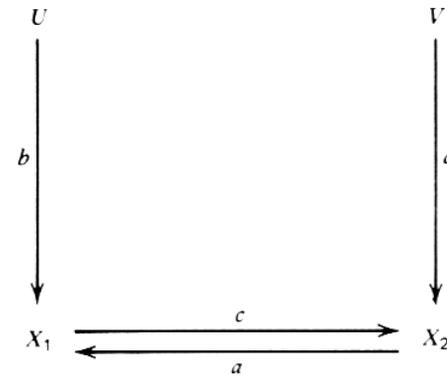


Figure 6.1 A simple feedback model.

To derive the correlation between X_1 and X_2 , take the covariance between Equations 6.1 and 6.2 to obtain

$$r_{12} = acr_{12} + bcr_{1U} + adr_{2V} \quad [6.3]$$

To solve for r_{1U} and r_{2V} take the covariance of X_1 with Equation 6.1 and X_2 with 6.2 to obtain

$$1 = ar_{12} + br_{1U}$$

and

$$1 = cr_{12} + dr_{2V}$$

Rearranging yields

$$br_{1U} = 1 - ar_{12}$$

and

$$dr_{2V} = 1 - cr_{12}$$

Substituting these values for br_{1U} and dr_{2V} into Equation 6.3 yields

$$r_{12} = acr_{12} + c(1 - ar_{12}) + a(1 - cr_{12})$$

Simplifying gives the result of

$$r_{12} = \frac{a + c}{1 + ac}$$

To achieve identification one might assume that $a = c$ and then solve for a :

$$r_{12} = \frac{2a}{1 + a^2}$$

Rearranging yields

$$r_{12}a^2 - 2a + r_{12} = 0$$

Applying the quadratic theorem results in

$$a = \frac{2 \pm (4 - 4r_{12}^2)^{1/2}}{2r_{12}}$$

$$a = \frac{1 \pm (1 - r_{12}^2)^{1/2}}{r_{12}}$$

As an example, if $r_{12} = .5$ then a and c would both equal .286 or 3.732. (Recall from Chapter 4 that standardized regression coefficients can be greater than one.)

EQUILIBRIUM ASSUMPTION

For the model in Figure 6.1, assume that $a = .4$ and $c = .6$. It then follows that $r_{12} = .806$. If the score of a person on X_1 is increased from zero to one, the person's score on X_2 is changed by .6. But this in turn affects the X_1 score, since X_2 causes X_1 . The score on X_1 is now increased by $(.6)(.4) = .24$. This in turn affects X_2 ad infinitum. The total change in X_2 is

$$.6 + .6(.6)(.4) + .6(.6)^2(.4)^2 + \dots + .6(.6)^i(.4)^i + \dots$$

which equals .790.

Since the change in X_2 echoed back to change X_1 , the X_1 score changes by more than the initial 1. The total change in the person's score on X_1 is

$$1 + (.6)(.4) + (.6)^2(.4)^2 + \dots + (.6)^i(.4)^i + \dots$$

or

$$\frac{.6}{1 - (.6)(.4)}$$

which equals 1.32. In general, with Equations 6.1 and 6.2 if a person's score on X_1 is changed by g units, the total change in X_2 is $cg/(1 - ac)$ and in X_1 is $g/(1 - ac)$. Since it was stated in Chapter 1 that causation is ordinarily assumed to occur with a time lag, a model with feedback would create effects that would last forever. Of course the effect may very well dampen out very quickly, but in theory the effect never stops.

Because of this "eternal" effect of feedback it is critical to examine systems of variables that are in equilibrium, that is, where a causal effect is not just beginning but has generally dampened out. Heise (1975, p. 227) nicely illustrates how mistaken conclusions can be made when the system of variables is not in equilibrium.

MODIFIED TRACING RULE

Feedback models create very special issues for the analyst. For instance, although the first law still holds, with feedback models the tracing rule does not hold in a simple way. The tracing rule can be modified so that it may hold for feedback models. The discussion must be divided into two parts. First discussed is the case of determining r_{12} where X_1 causes X_2 and X_2 and X_1 are involved in a feedback relationship. The key here is that X_1 and X_2 are not in a feedback relationship.

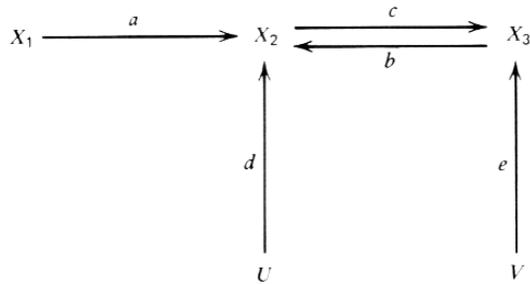


Figure 6.2 An exogenous variable in a nonhierarchical system.

Examining the path diagram in Figure 6.2, one might be tempted to apply the tracing rule and say r_{12} equals a . Examine, however, the tracings from X_1 to X_2 . There is first a path directly from X_1 to X_2 but there is also a path from X_1 to X_2 to X_3 and back again to X_2 . This would seem to violate the tracing rule since that rule does not allow tracing through the same variable twice. One can, however, argue that because due to feedback the values of X_2 change, the "same" variable is not traced through. Whatever the rationale there are the following paths: $a + abc + a(bc)^2 + a(bc)^3$ and so on. It then follows that

$$r_{12} = a \sum (bc)^i = \frac{a}{1 - bc}$$

Thus in a sense the tracing rule still applies but there are an infinite number of tracings.

The correlation r_{13} in Figure 6.2 can be computed in a similar fashion:

$$r_{13} = ac \sum (bc)^i = \frac{ac}{1 - bc}$$

If a path is added from X_1 to X_3 and is labeled f , it would then follow that

$$r_{12} = \frac{a + fb}{1 - bc}$$

$$r_{13} = \frac{f + ac}{1 - bc}$$

Thus, to apply the modified tracing rule for the correlation between a variable and a second variable that is involved in a feedback relationship with a third variable, simply apply the tracing rule learned in Chapter 3. This quantity is then divided by $1 - gh$, where g is the path from one of the feedback variables to the other and h is the second feedback path.

The second part of the modified rule concerns a correlation between X_1 and X_2 , where X_1 is in a feedback relationship and X_2 is directly or indirectly caused by X_1 . The feedback relationship that X_1 is in may or may not involve X_2 but X_1 does cause X_2 . First find the reduced form of the variable in the feedback relationship, that is, the equation without feedback incorporated into it. Then redraw the path diagram for the reduced form and apply the tracing rule. (This procedure also works for the previous part.) As an example, to determine r_{23} in Figure 6.2, write the equation for X_2 , $aX_1 + bX_3 + dU$, and for X_3 , $cX_2 + eV$. Substituting the equation for X_3 into the X_2 equation yields

$$X_2 = aX_1 + b(cX_2 + eV) + dU$$

which reduces to

$$X_2 = \frac{a}{1 - bc} X_1 + \frac{d}{1 - bc} U + \frac{be}{1 - bc} V$$

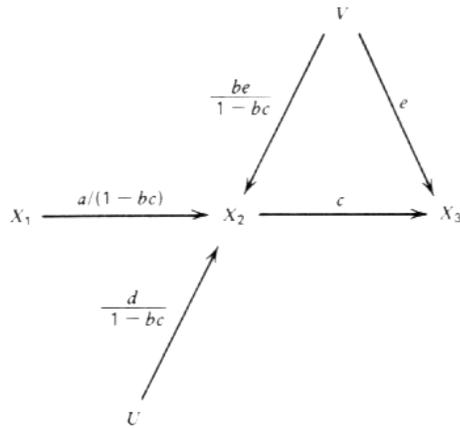


Figure 6.3 Reduced from path diagram.

Now to compute r_{23} redraw the path diagram as in Figure 6.3. The correlation between X_2 and X_3 then is $c + be^2/(1 - bc)$. Note that a different but equivalent answer would have been obtained had the reduced form of X_3 been solved for. Although the tracing rule is valid when so modified, generally it is much simpler to apply the first law. I state the rule only because it may occasionally be helpful.

Returning to the model in Figure 6.2, it happens to be identified. As has been shown

$$r_{12} = \frac{a}{1 - bc}$$

$$r_{13} = \frac{ac}{1 - bc}$$

and with some algebraic work it can be shown that

$$r_{23} = \frac{b + c}{1 + bc}$$

The solution for c is straightforward:

$$c = \frac{r_{13}}{r_{12}}$$

Because c is known, one can solve for b from the p_{23} equation to find

$$b = \frac{r_{13} - r_{23}r_{12}}{r_{23}r_{13} - r_{12}}$$

and a in terms of b and c

$$a = r_{12}(1 - bc)$$

To test whether a is zero, one tests whether $\rho_{12} = 0$; to test if b is zero, $\rho_{13,2} = 0$; and to test c we see if $\rho_{13} = 0$. It should be noted that the model in Figure 6.2 is not identified if $r_{UV} \neq 0$.

To estimate d take the covariance of X_2 with the equation for X_2 to obtain $1 = ar_{12} + br_{23} + dr_{21}$. Since $r_{21} = d/(1 - bc)$ by the use of the modified tracing rule, it follows that

$$1 = ar_{12} + br_{12} + \frac{d}{1 - bc}$$

Now solve for d to obtain

$$d = ((1 - ar_{12} - br_{23})(1 - bc))^{1/2}$$

A simpler and more direct solution for d is to take the variance of

$$dU = X_2 - aX_1 - bX_3$$

which is

$$b^2 + 2abr_{13} - 2ar_{12} - 2br_{23}$$

The solution for d is then the square root of the preceding expression. By a similar procedure one can obtain e as

$$(1 + c^2 - 2cr_{23})^{1/2}$$

or one can derive e by taking the covariance of X_3 with its structural equation

$$e = ((1 - cp_{23})(1 - bc))^{1/2}$$

Needless to say the two solutions for both d and e are algebraically equivalent.

INDIRECT FEEDBACK

Consider the model in Figure 6.4. There is an indirect feedback loop: X_1 causes X_3 which in turn causes X_2 , and X_2 causes X_1 . The equations are

$$\begin{aligned} X_1 &= eX_2 + fU \\ X_2 &= gX_3 + hW \\ X_3 &= iX_1 + jS \end{aligned}$$

Substituting the equation of X_3 into the equation of X_2 yields

$$X_2 = giX_1 + gjS + hW \tag{6.5}$$

Note that there now is a formal equivalence between the two equations 6.4 and 6.5 and the two equations 6.1 and 6.2; that is, let $e = a, f = b, gi = c$, and $gjS + hW = dV$. The correlation then between X_1 and X_2 is

$$\frac{e + gi}{1 + egi}$$

and similarly r_{13} equals

$$\frac{i + eg}{1 + egi}$$

and r_{23} equals

$$\frac{g + ie}{1 + egi}$$

Although the model is just-identified with three free parameters and three correlations, there is no simple solution to the equations.

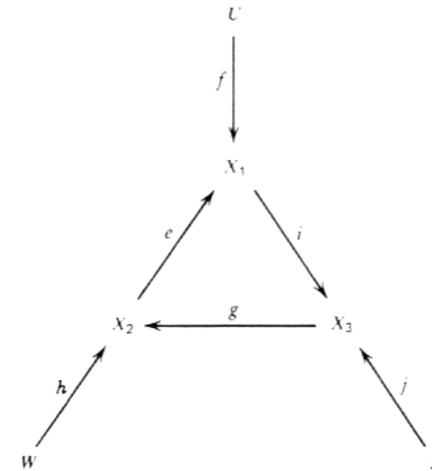


Figure 6.4 Indirect feedback.

ESTIMATION

Most of the preceding discussion would not be found in the consideration of feedback models in other texts. Those texts concentrate on the topic in this section: the use of two-stage least squares to estimate parameters for models with feedback.

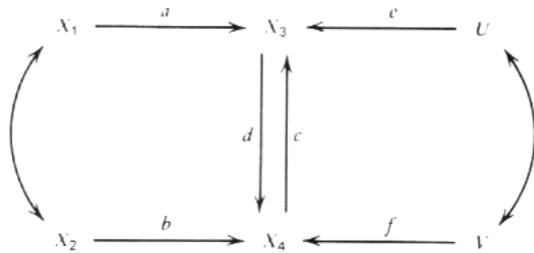


Figure 6.5 Two-stage least-squares example.

Consider the model in Figure 6.5. Variables X_3 and X_4 are involved in a feedback relationship. Variables X_1 and X_2 are purely exogenous with X_1 causing X_3 and X_2 causing X_4 . The disturbances of X_3 and X_4 have been allowed to be correlated. Naively, one might think that one could simply regress X_3 on both X_1 and X_4 and regress X_4 on X_2 and X_3 to obtain estimates of the path coefficients. They do not ordinarily provide unbiased estimates of the parameters since X_4 is correlated with the disturbance of X_3 and X_3 is correlated with the disturbance of X_4 . Using the modified tracing rule these correlations are

$$r_{4U} = \frac{ed + fr_{UV}}{1 - cd}$$

and

$$r_{3V} = \frac{fc + er_{UV}}{1 - cd}$$

Note the correlation still exists even if $r_{UV} = 0$. Since the disturbance is correlated with a cause, simple multiple regression breaks down. However, as was seen in the last chapter, in the case of an exogenous variable correlated with a disturbance, identification is still possible given certain zero paths. In Figure 6.5 estimation may be possible since there is no path from X_2 to X_3 and from X_1 to X_4 . Using the terminology

of the previous chapter, X_2 can be used as an instrument for the X_3 equation and X_1 for the X_4 equation. For the X_3 equation the first law yields

$$\begin{aligned} r_{13} &= a + cr_{14} \\ r_{23} &= ar_{12} + cr_{24} \end{aligned}$$

Solve for a and c to obtain

$$a = \frac{r_{12}r_{24} - r_{23}r_{14}}{r_{24} - r_{12}r_{14}}$$

and

$$c = \frac{r_{23} - r_{12}r_{13}}{r_{24} - r_{12}r_{14}}$$

To estimate b and d for the X_4 equation a parallel procedure is followed. Using the first law one obtains

$$r_{14} = br_{12} + dr_{13}$$

The solutions for b and d are then

$$b = \frac{r_{24}r_{13} - r_{23}r_{14}}{r_{13} - r_{12}r_{23}}$$

and

$$d = \frac{r_{14} - r_{23}r_{12}}{r_{13} - r_{12}r_{23}}$$

To find e first note that

$$eU = X_3 - aX_1 - cX_4$$

Taking the variance of eU yields

$$e^2 = 1 + a^2 + c^2 - 2ap_{13} - 2cp_{34} + 2acp_{14}$$

To solve for e simply take the square root. By a similar set of steps it can be shown that

$$f^2 = 1 + b^2 + d^2 - 2bp_{24} - 2dp_{34} + 2bdp_{23}$$

To find r_{UV} simply take the covariance of eU with fV to obtain

$$r_{34}(1 + cd) + bcr_{24} + abr_{12} + adr_{13} - c - d - ar_{14} - br_{23}$$

To solve for r_{1V} simply divide by ef .

Ordinarily two-stage least squares is computationally more efficient than the preceding procedure. Also, as was noted in the previous chapter, overidentified models present no difficulty for two-stage least square estimation.

To apply two-stage least squares to an equation for X_i with a set of m exogenous variables and k variables which are directly or indirectly caused by X_i , at least k of the exogenous variables must

1. Not directly cause X_i .
2. Not be correlated with the disturbance of X_i .

These k exogenous variables are called the set of instruments. To avoid empirical underidentification, the regression of each of the k endogenous variables on the instruments must first yield moderate to high multiple correlations and the predicted variables of the second stage must not be highly correlated among each other. The reader should refer to Chapter 5 for a more detailed discussion of these conditions.

To apply two-stage least squares, one exogenous variable must be omitted from the equation of each variable in a feedback loop.

This variable can be called the instrument. However, the same variable cannot be omitted from both equations. To understand this, examine the path diagram in Figure 6.6 which is taken from Duncan (1975, p. 84). It would seem that X_2 could be used as an instrument for both the X_3 and X_4 equation. However, it cannot be since the model in Figure 6.6 is underidentified. Although there are six free parameters and six correlations, there is a hidden pair of overidentifying restrictions. To find them use the modified tracing rule for feedback models:

$$r_{13} = \frac{a + bc}{1 - cd}$$

$$r_{14} = \frac{b + ad}{1 - cd}$$

$$r_{23} = r_{12} \frac{a + bc}{1 - cd}$$

$$r_{24} = r_{12} \frac{b + ad}{1 - cd}$$

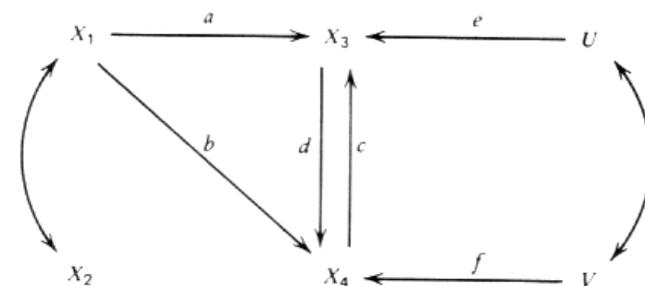


Figure 6.6 Duncan example.

The following overidentifying restrictions then hold:

$$\rho_{23} - \rho_{12}\rho_{13} = 0$$

and

$$\rho_{24} - \rho_{12}\rho_{14} = 0$$

These overidentifying restrictions simply state that X_2 shares no variance with X_3 or X_4 except through X_1 . Given these overidentifying restrictions, none of the parameters of the model can be estimated.

Duncan has noted that the "estimate" of c is

$$\frac{r_{23} - r_{12}r_{13}}{r_{24} - r_{12}r_{14}}$$

and of d is

$$\frac{r_{24} - r_{12}r_{14}}{r_{23} - r_{12}r_{13}}$$

He then points out the curious fact these two estimates are reciprocals of each other. However, neither of these two values are mathematical estimates. Note that, given the overidentifying restriction, the denominators of each expression are zero which makes the formulas mathematical nonsense. Thus neither of the "estimates" are really estimates. The "estimates" of a and b are

$$a = \frac{r_{13}r_{24} - r_{12}r_{14}}{r_{24} - r_{12}r_{14}}$$

$$b = \frac{r_{23}r_{14} - r_{13}r_{24}}{r_{23} - r_{12}r_{13}}$$

Again note that both of the denominators are zero.

A NONHIERARCHICAL MODEL FOR HALO EFFECT

Instead of an empirical example of nonhierarchical models, a conceptual example is presented in this section. Within the topic of person perception in social psychology, it is well known that the ratings of traits

and behaviors tend to be biased by the halo effect (Berman & Kenny, 1976). Observers who believe that two traits go together will in fact see the traits as going together whether they in fact do or not.

There are many possible types of models that one could posit for halo. What follows is one such model. Imagine that two traits, friendly and intelligent, have been rated by two raters for a number of rates.

Designate F as the true friendliness of the ratees and I as the true intelligence. Then let F_1 and I_1 be the ratings of the two traits by rater 1 and F_2 and I_2 by rater 2. In Figure 6.7 these six variables are theoretically related in a path diagram. It is first assumed that the ratings of F are indeed caused by F and similarly I causes I_1 and I_2 . What brings about halo is that F_1 and I_1 are assumed to be involved in a feedback relationship. So also are F_2 and I_2 . Thus the ratings of the two traits mutually influence each other. To simplify the model, F and I are uncorrelated and various equalities of parameters have been assumed in Figure 6.7.

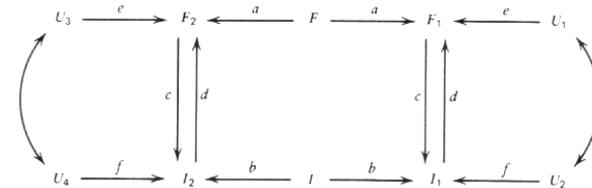


Figure 6.7 Halo effect model.

Various authors (Campbell & Fiske, 1959; Stanley, 1961) have stated or implied that the different rater-different trait correlation is free from halo. Thus if the two traits are uncorrelated ($r_{FI} = 0$ as in Figure 6.7) then the different rater-different trait correlations should be zero, $r_{F1I2} = r_{F2I1} = 0$. This might be true for some models of halo but it is not true for the model in Figure 6.7. Note that both r_{F1I2} and r_{F2I1} equal

$$\frac{a^2c + b^2d}{1 - cd}$$

7

SINGLE UNMEASURED
VARIABLES

In this chapter unmeasured exogenous variables are considered and a factor analysis vocabulary is used as well as a regression analysis vocabulary. Many of the important variables in causal models cannot be directly observed but can be measured only with error. Measured variables are often imperfect glimpses of an underlying reality, and unmeasured variables must be added to the structural model. There are four sections to this chapter. The first section introduces the factor analysis model and discusses the problem of identification. The second section takes up the question of estimation, and the third considers specification error. The final section discusses the measurement of factors.

FACTOR MODELS

IDENTIFICATION

Our initial concern is with the factor model and not with the estimation of parameters of the model or, as it is usually called, factor analysis. In a factor model the measured variables are assumed to be a linear function of a set of unmeasured variables. These unmeasured, latent variables are called factors. Traditionally, one distinguishes three different types of factors: common, specific, and error. A common factor is an unmeasured variable that appears in the equation of two or more measured variables. A specific factor appears in a single equation and represents true variance that is not common to the other variables. An error factor is simply the errors of measurement of a single

variable. Procedurally it is difficult to distinguish between specific and error factors, and so they are usually lumped together to form the unique factor.

There are two different types of parameters in factor models: factor loadings and factor correlations. A factor loading is the regression coefficient of a measured variable on a factor. If both the factor and measured variable are standardized, the loading is a beta weight. The factor loadings are multiplied by the factors and summed to equal the measured variable. The variable is said to load on a factor. In special cases the factor loading can equal the correlation of the variable with the factor.

The factors themselves may be intercorrelated. The correlations are called factor correlations. Traditionally, the unique factors are uncorrelated with the common factors and with each other. As is demonstrated in this chapter, the unique factors may be correlated. If the common factors are uncorrelated with each other, the solution is said to be orthogonal, and if correlated, it is said to be oblique.

The variance of each measured variable can be partitioned into three components: common, specific, and error. The common variance, called the communality, is the squared multiple correlation of the common factors with the measured variables, and its usual symbol is h^2 . If the common factors are orthogonal (that is, uncorrelated), then the communality simply equals the sum of squared common factor loadings. The communality plus the specific variance equals the reliability of a variable and the remaining variance is due to error. The specific and error variance may be lumped together to form the unique variance.

In this chapter, we see that with causal models the usual bugaboos of both choosing communalities and choosing a rotation scheme are avoided. Additionally, it is shown that the common factors take the role of exogenous variables, the unique factors of disturbances, the factor loadings of path coefficients, and the factor correlations of correlations between exogenous variables.

ONE FACTOR, TWO INDICATORS

Single-factor models presume that the measured variables are all indicators of a latent construct. Using the vocabulary of test theory introduced in Chapter 5, the measures are a set of congeneric tests. Measures that correlate more highly with the latent construct have proportionately less unique variance and can be used to name the construct. In the limiting case a correlation of one implies the measure and the construct are one in the same.

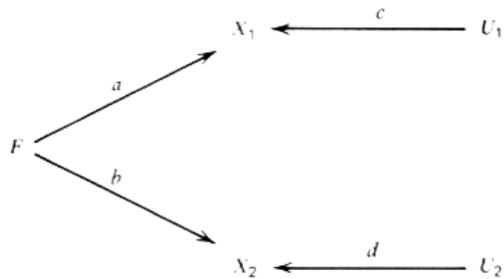


Figure 7.1 Single-factor model, two indicators.

In Figure 7.1 there is a causal model in which X_1 and X_2 are caused by F ; U causes X_1 , and V causes X_2 , where both U and V are disturbances which are called errors of measurement in latent variable models. If X_1 and X_2 are measured, all the measures are endogenous variables, all the causes are unmeasured, and the relationship between X_1 and X_2 is spurious. Disturbances in latent variable models are referred to *measurement errors* and are often symbolized by E . What is the status of identification of the model in Figure 7.1? There is only a single correlation, r_{12} , but there are two free causal parameters, a and b . As with models from the previous chapters, the paths from the disturbances are constrained since $c = (1 - a^2)^{1/2}$ and $d = (1 - b^2)^{1/2}$. (Alternatively, the disturbance paths are fixed to 1.0 and $V(U) = 1 - a^2$ and $V(V) = 1 - b^2$. Since there are more parameters than correlations, the model is underidentified. The parameters of the model cannot be estimated, but both a^2 and b^2 must be at least r_{12}^2 . To achieve

identification it might be assumed that $a = b$. It would then follow that $a = b = r_{12}^{1/2}$. However, if the value of r_{12} were negative, it would contradict any assumption that $a = b$ since a parameter would have to be imaginary, the square root of minus one.

The units of measurement of the factor are arbitrary. The approach taken here, which is consistent with the factor analysis tradition, is to standardize the factor. An alternative approach is to use the units of one of the measures. In this case a loading of one measure is set to one and the variance of the factor is free.

ONE FACTOR, THREE INDICATORS

The status of identification changes if there are three measured variables that are caused by the latent variable, F . The measures X_1 , X_2 , and X_3 are sometimes called the *indicators* of the latent construct. Also the correlation of an indicator with a factor is called an *epistemic correlation* (Costner, 1979) and is the factor loading. As shown in Chapter 3, given

$$\begin{aligned} X_1 &= aF + dU_1 \\ X_2 &= bF + eU_2 \\ X_3 &= cF + fU_3 \\ \rho_{FU_i} &= 0 \quad (i = 1, 2, 3) \\ \rho_{U_i U_j} &= 0 \quad (i \neq j) \end{aligned}$$

it follows from the tracing rule that $r_{12} = ab$, $r_{13} = ac$, and $r_{23} = bc$. Because there are three parameters and three correlations, the model is just-identified. The estimates are

$$\begin{aligned} a^2 &= \frac{r_{12} r_{13}}{r_{23}} \\ b^2 &= \frac{r_{12} r_{23}}{r_{13}} \end{aligned}$$

$$c^2 = \frac{r_{13} r_{23}}{r_{12}}$$

The disturbance paths are simply

$$\begin{aligned} d &= (1 - a^2)^{1/2} \\ e &= (1 - b^2)^{1/2} \\ f &= (1 - c^2)^{1/2} \end{aligned}$$

The direction of the scale of the factor is arbitrary. One can then take either the positive or negative root of a^2 or any of the other parameters. As a rule one normally tries to make as many loadings as possible positive. Once the sign of one parameter has been set, the sign of the others have been determined. For instance, if a is taken to be positive and if r_{12} is negative and if r_{13} is positive, b must be negative and c positive. However, if a is taken to be negative, then b is now positive and c is negative. In either case, r_{23} must be negative, and if it is not, some specification error is indicated. In general, it must hold that $r_{12}r_{13}r_{23} > 0$ for there to be no contradiction about the direction of the scale of the unmeasured variable. With three variables for there to be no contradiction all or only one of the correlations should be positive. If $r_{12}r_{13}r_{23} < 0$, one of the factor loadings must be imaginary.

Thus with three indicators the single factor model is identified. However, if any of the loadings (a , b , or c) are zero, or near zero, the model is empirically underidentified. For instance, let a equal zero and note that the estimates for b^2 and c^2 are undefined since $r_{13} = r_{12} = 0$. Thus to identify empirically the loadings on a single factor, the loadings must be of moderate size.

It is also mathematically possible that the value obtained for a^2 or any loading to be larger than one. For a^2 to be larger than one, it follows that $|r_{12}r_{13}| > |r_{23}|$. Although a value larger than one is mathematically possible, in the population it indicates a specification error. Because a is the correlation of F with X_1 , it cannot be greater

than one. Although correlations computed from data cannot be larger than one, correlations constructed from data can be out of bounds. If a^2 is larger than one, then the solution for d is imaginary. Estimates of parameters larger than one in factor analysis are examples of the Heywood case (Harmon, 1967, p. 117).

Using the sample data, it is possible to obtain an estimate of a loading larger than one or imaginary, even if the model is correctly specified. A method then is needed to test whether the anomalous estimate is due either to sampling error or to specification error. With a little bit of algebraic manipulation it can be shown that both imaginary and loadings larger than one occur only in the sample if $r_{12,3}$, $r_{13,2}$, or $r_{23,1}$ is negative. Thus a test that the anomaly is plausibly explained by sampling error is that a negative partial correlation does not significantly differ from zero. If it is significantly negative, then specification error is indicated.

ONE FACTOR, FOUR INDICATORS

In Figure 7.2 there is one common factor, F , and four endogenous variables, X_1 through X_4 . The disturbances or errors of measurement, U_1 through U_4 , are assumed to be uncorrelated with F and with each other. The model is over-identified in some way because there are four free parameters and six correlations. Using the tracing rule, the six correlations are

$$\begin{aligned} r_{12} &= ab \\ r_{13} &= ac \\ r_{14} &= ad \\ r_{23} &= bc \\ r_{24} &= bd \\ r_{34} &= cd \end{aligned}$$

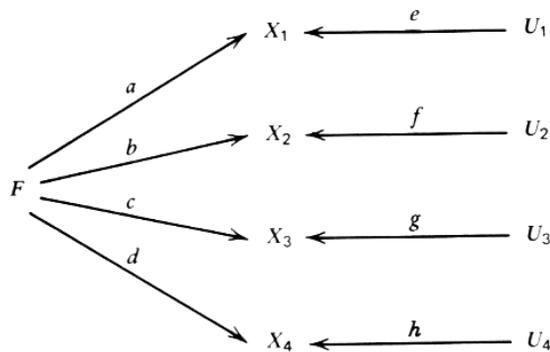


Figure 7.2 Single-factor model, four indicators.

One way to solve for the parameters is to view this four-variable model as a set of four different three-variable models; that is, separately analyze variables $X_1, X_2,$ and $X_3,$ then analyze $X_1, X_2,$ and $X_4,$ then, $X_1, X_3,$ and $X_4,$ and finally, $X_2, X_3,$ and $X_4.$ Then separately solve for the parameters from each of these three variable cases as was done in the previous section. Working through this strategy yields the following solutions for each parameter:

$$a^2 = \frac{r_{12}r_{13}}{r_{23}} = \frac{r_{12}r_{14}}{r_{24}} = \frac{r_{13}r_{14}}{r_{34}}$$

$$b^2 = \frac{r_{12}r_{23}}{r_{13}} = \frac{r_{12}r_{24}}{r_{14}} = \frac{r_{23}r_{24}}{r_{34}}$$

$$c^2 = \frac{r_{13}r_{23}}{r_{12}} = \frac{r_{13}r_{34}}{r_{14}} = \frac{r_{23}r_{34}}{r_{24}}$$

$$d^2 = \frac{r_{14}r_{24}}{r_{12}} = \frac{r_{14}r_{34}}{r_{13}} = \frac{r_{24}r_{34}}{r_{23}}$$

There is an interesting pattern in the subscripts of the preceding formulas. Although it makes no sense algebraically, cancel the subscripts of the correlations. For instance, for the first expression for $a^2,$ the 2 and 3 cancel leaving two 1's. In every case two subscripts

of the endogenous variable remain. Using "subscript algebra," all that remains is the subscript squared which provides a perfect analog to the parameter estimate which is also squared. I do not recommend subscript algebra as a replacement for careful analysis, but it can serve as a useful mnemonic and check.

The disturbance or measurement error paths can be simply estimated by

$$e = (1 - a^2)^{1/2}$$

$$f = (1 - b^2)^{1/2}$$

$$g = (1 - c^2)^{1/2}$$

$$h = (1 - d^2)^{1/2}$$

Because there are multiple estimates for each parameter, overidentifying restrictions can be solved for. All the preceding equalities of the parameter solutions imply only the three following overidentifying restrictions:

$$\rho_{12}\rho_{34} - \rho_{13}\rho_{24} = 0$$

$$\rho_{12}\rho_{34} - \rho_{14}\rho_{23} = 0$$

$$\rho_{13}\rho_{24} - \rho_{14}\rho_{23} = 0$$

because $\rho_{12}\rho_{34} = \rho_{13}\rho_{24} = \rho_{14}\rho_{23} = abcd.$ It can be quickly seen that if any of the two preceding restrictions holds, the third immediately follows. There are then only two independent overidentifying restrictions, which was already known because there are four free parameters and six correlations. This excess of correlations yields two independent over-identifying restrictions, that is, two constraints on the correlation matrix.

These overidentifying restrictions are called *vanishing tetrads.* They are called vanishing tetrads because four variables are involved and the difference between pairs of products equals zero; that is, they vanish. Note that the tetrads satisfy subscript algebra. For instance, rearranging Equation 7.1, the following holds

$$\frac{\rho_{12}\rho_{34}}{\rho_{13}\rho_{24}} = 1$$

and the subscripts all cancel to yield one.

ONE FACTOR, N INDICATORS

In the general case, given n measures and a single factor model, there are $(n-1)(n-2)/2$ estimates of each parameter. Because with every four variables there are three tetrads, the total number of tetrads are

$$\frac{n(n-1)(n-2)(n-3)}{8}$$

The total number of independent tetrads is the number of correlations, $n(n-1)/2$, minus the number of free parameters, n , which yields $n(n-3)/2$ restrictions on the correlations, which is sometimes referred to as the degrees of freedom of the model.

Vanishing tetrads are very common overidentifying restrictions in models with unobserved variables. More often than not if there is an overidentifying restriction, it is a vanishing tetrad.

TEST OF A VANISHING TETRAD

Deferred until this chapter is the significance test of a vanishing tetrad. Duncan (1972) resurrected a test for the vanishing tetrad which was originally suggested by Spearman and Holzinger (1924). The test is not well known and yields a different value depending on the order of the products of the correlations. I have suggested a test for the vanishing tetrad (Kenny, 1974) that can be made by using a computer program available at most installations: canonical correlation. The use of canonical correlation analysis for models with unmeasured variables has been suggested by other authors

(e.g., Hauser & Goldberger, 1971), but none had explicitly suggested the canonical correlation as a test for the vanishing tetrad.

For those who may be unfamiliar with canonical correlation, there are two sets of variables, say X_1, X_2, \dots, X_p , and a second set Y_1, Y_2, \dots, Y_q . There are then p variables in set X and q in set Y . The canonical correlation is the correlation between a linear combination of set X with a linear combination of set Y where the weights for both X and Y are set to maximize that correlation. If both p and q are greater than one, a second canonical correlation can be obtained such that the new linear combinations are orthogonal to the prior combinations. The total number of canonical correlations is equal to p or q , whichever is smaller.

Canonical correlation can be used to test a vanishing tetrad of $\rho_{13}\rho_{24} - \rho_{14}\rho_{23} = 0$ (Equation 7.3). Let X_1 and X_2 be set "X" and X_3 and X_4 be set "Y." To determine how the four variables divide into the sets note that the tetrad does not involve r_{12} and r_{34} . The simple rule is that X_1 and X_2 form one set (r_{12}) and X_3 and X_4 (r_{34}) the second set. For the tetrad of Equation 7.1 the missing correlations are r_{14} and r_{23} . Thus X_1 and X_4 form one set and X_2 and X_3 the second set.

It can be shown that if $r_{13}r_{24} - r_{14}r_{23} = 0$, then the second canonical correlation will be zero. Thus the null hypothesis, Equation 7.3, can be tested by the hypothesis that the second canonical correlation is zero. In the same way the remaining tetrads can be tested. A significance test for this special case of $p = q = 2$ is given by Bartlett's χ^2 approximation (Tatsuoka, 1971, p. 189):

$$\chi^2 = -(N-3.5) \log_e(1-r_2^2) \quad [7.4]$$

where N is the sample size and r_2 is the estimated second canonical correlation and the χ^2 has one degree of freedom.

In general the computation of canonical correlations is rather involved, but a handy solution is relatively simple when both p and q are two. Solve the following quadratic equation for ρ^2

$$a\rho^4 + b\rho^2 + c = 0 \quad [7.5]$$

where

$$a = (1 - r_{12}^2)(1 - r_{34}^2)$$

$$b = (r_{14} - r_{12}r_{24})(r_{13}r_{34} - r_{14}) + (r_{23} - r_{13}r_{12})(r_{24}r_{34} - r_{23}) \\ + (r_{13} - r_{12}r_{23})(r_{14}r_{34} - r_{13}) + (r_{24} - r_{14}r_{12})(r_{23}r_{34} - r_{24})$$

$$c = (r_{13}r_{24} - r_{14}r_{23})^2$$

Because Equation 7.5 is quadratic, there are two solutions for ρ^2 . The second canonical correlation squared is set equal to the smaller solution of Equation 7.5. If c equals zero, then the smaller solution of ρ^2 of Equation 7.5 is zero. Note that c exactly equals the square of a tetrad.

As an illustration, consider the correlations from Duncan (1972, p. 53): $r_{12} = .6508$, $r_{13} = .7304$, $r_{14} = .7548$, $r_{23} = .7100$, $r_{24} = .6999$, and $r_{34} = .6642$, where the sample size is 730. A test of the hypothesis that $r_{13}r_{24} - r_{14}r_{23} = 0$ yields a second canonical correlation of .050 and $\chi^2(1) = 1.80$, which is not statistically significant at the .05 level. The solution of Equation 7.5 yields $a = 0.322$, $b = -0.247$, and $c = 0.0006$.

ESTIMATION

The preceding sections present only the path analytic estimates of the path from the factor to this measure. As the reader is no doubt aware, there are a myriad of possible estimates of factor loadings from different methods of factor analysis. Before discussing the mechanics of estimating path coefficients, consider a data set to be used as an illustration. The data are taken from unpublished data of B. Malt. One

hundred eighteen terminal cancer patients were asked to state the strength of their concerns about a number of issues. The four concerns are concerns about friends (X_1), concerns about existence (X_2), concerns about family (X_3), and concerns about self (X_4). The correlations between these four variables are contained in Table 7.1. A path model for the four variables, contained in Figure 7.2, has the four specific concerns caused by some unmeasured general concern.

A single factor model adequately fits the data because the three tetrads appear to vanish

$$(.242)(.719) - (.551)(.416) = -.055$$

$$(.242)(.719) - (.577)(.311) = -.005$$

$$(.551)(.416) - (.577)(.311) = .050$$

The fact that the three preceding tetrads do not perfectly vanish can be explained by sampling error because the χ^2 for the tetrads are 1.00, 0.01, and 1.42, respectively. None of the χ^2 values even approaches statistical significance. Remember if the χ^2 is not significant, the tetrad vanishes and a single-factor model is indicated.

Because a single-factor model fits the data, the next question is, what are the path coefficients from the factor to each measure? Earlier it was stated that these path coefficients are factor loadings. A comparison of four different methods of estimating the loadings follows:

1. Pooled path analytic.
2. Principal components.
3. Principal factors.
4. Maximum likelihood.

Table 7.1 Four Concerns of Terminal Cancer Patients: Correlations below Diagonal and Residuals from the Maximum Likelihood Solution above the Diagonal^a

	X_1	X_2	X_3	X_4
Concern about friends (X_1)	1.000	-.040	.027	-.008
Existential concern (X_2)	.242	1.000	-.037	.028
Concern about family (X_3)	.551	.311	1.000	-.002
Concern about self (X_4)	.577	.416	.719	1.000

^aN = 118.

Earlier it was shown that there are three different formulas for path analytic solutions when there are four measured variables. If one substitutes in the sample estimates of the correlations, the estimates of the loadings for the correlations in Table 7.1 are

$$a: .655, .579, .665$$

$$b: .370, .418, .424$$

$$c: .841, .829, .733$$

$$d: .996, .868, .981$$

These different estimates of the same parameter are fairly homogeneous the largest difference being only .128. This is not surprising because the overidentifying restrictions were met. When the tetrads vanish, the different estimates are exactly the same. Now the problem arises of how to combine these parameter estimates. One solution might be a simple arithmetic average. This is a flawed strategy because it fails to weight the estimates by the size of the correlations. An alternative strategy is recommended by Harmon (1967): Examining the estimates of a^2 note that they all equal the product of two correlations divided by a third. Why not simply pool the numerators and the denominators separately? Then to estimate a^2 divide the pooled numerator by the pooled denominator to obtain

$$\frac{r_{12}r_{13} + r_{12}r_{14} + r_{13}r_{14}}{r_{23} + r_{24} + r_{34}}$$

The numerator involves all possible pairs of correlations with variable X_1 , and the denominator is the sum of all correlations not involving X_1 . This procedure assumes that all the correlations are positive in sign. If some of the correlations are negative, it is advisable to sum absolute values. In a similar fashion we can pool the estimates for the other path coefficients. These pooled path analytic estimates for the correlations in Table 7.1 are in Table 7.2. They fall between the three path analytic values. It is interesting to note that the loading for concerns about self, or X_4 , is almost equal to one. Thus, the unmeasured underlying concern is very close to a concern about the self.

Table 7.2 Factor Loadings for the Correlations in Table 7.1

Variable	Solution			
	Pooled Path Analytic	Principal Components	Iterated Communalities	Maximum Likelihood
X_1	.639	.774	.650	.652
X_2	.407	.568	.415	.432
X_3	.800	.857	.805	.804
X_4	.930	.893	.906	.897

The second solution discussed is perhaps the best known: the method of principal components. The nonmathematical logic of this solution is as follows: Let us create a linear composite of the measured variables, choosing the weights of this composite so that we maximize the sum of squared correlations of the variables with the composite. Because a squared correlation can be thought of as variance explained, we choose the composite to maximize its ability to explain the variance of the measured variables. Mathematically this sort of problem turns out to be an eigenvector problem, which is computationally a bit messy but tractable. The resultant correlation of the composite with each measure is the factor loading. There is clearly something backward about a principal component solution, at least when one is doing causal modeling. With principal components the factor is directly measurable

because it is defined as a weighted sum of the measured variables. In causal modeling, each measured variable or any finite sum of them is only an imperfect indicator of the factor.

The factor loadings from principal components are in Table 7.2. Although their relative ranking compares favorably to the pooled path analytic solution, they are hardly the same solution. The most striking fact about the principal component solution is that in three of the four cases the factor loadings are larger than the pooled path analytic solution. If we use the principal components solution to estimate the correlations between the measured variables, we overestimate all six of the correlations. The components solution does not attempt to fit the correlations, but rather it attempts to maximize the sum of squared factor loadings. Because this is the goal, a principal components solution does not provide a very good estimate of path coefficients.

To obtain the true factor loadings the communalities should be placed in the main diagonal of the correlation matrix before the eigenvectors are computed. But because the factor loadings are needed to compute the communalities, there is a chicken-egg problem. We need the communalities to compute the factor loadings and we need the loadings to compute the communalities. One way around this problem is to find the communalities by iteration. Start with some communality estimates and then solve for the factor loadings by the method of principal factors. (A principal factors solution attempts to maximize the explanation of only the common variance, not all the variance as does principal components.) Taking these factor loadings we solve for communalities, and with these new communalities we solve for the factor loadings. We continue until the factor loadings stabilize. To hasten the convergence, the initial communalities are the squared multiple correlations of each variable with the other variables taken as the predictors.

The iterated communality solution for the factor loadings is contained in Table 7.2. The values more closely resemble the pooled path analytic solution than the principal components solution. The

solution does result in somewhat less variable loadings than the path analytic solution.

The final solution discussed is maximum likelihood factor analysis. Of the four solutions only maximum likelihood provides a statistical estimate of factor loadings or path coefficients and only maximum likelihood factor analysis provides significance tests of goodness of fit of the model. A likelihood is the analog of probability for continuous distributions. One can compute the likelihood of a set of estimates of factor loadings given the sample correlation matrix and the assumption of multivariate normality. The method of maximum likelihood chooses the most likely estimates. Like the previous method, it is also an iterative solution. Although the mathematics of maximum likelihood factor analysis may remain a "black box" for most of us, suffice it to say the computations have proved to be tractable only by computer. The pioneering work was done by Lawley and Maxwell (1963), and the major computational breakthrough was made by Jöreskog. There is little doubt that maximum likelihood factor analysis will revolutionize factor analysis because it puts factor analysis estimation on a firm statistical ground.

Unlike most traditional methods of factor analysis, maximum likelihood factor analysis can be employed in either a confirmatory or an exploratory manner. Most methods of factor analysis can only be employed in exploratory fashion; that is, little or no structure can be imposed on the solution. In the case of a model with multiple factors the solution is not unique and can be rotated to an infinite number of alternative solutions. If the maximum likelihood solution is combined with a confirmatory approach, the solution is unique and no rotation of the solution is possible.

If a confirmatory approach is chosen, the model must be identified. The researcher must impose enough constraints to make the parameter estimates unique. The two common types of constraints are fixing the parameters to a given value and forcing two parameter estimates to be equal. With maximum likelihood factor analysis one can fix factor loadings, factor variances, factor correlations, factor

covariances, or communalities to a given value. Usually the fixed value is zero or one. For instance, a loading or a factor correlation is set to zero or a loading or factor variance is set to one. Equality constraints are also possible: Two loadings or factor correlations may be set equal to each other. In this text maximum likelihood factor analysis is always employed in a confirmatory manner. Thus maximum likelihood factor analysis is at times referred to as confirmatory factor analysis.

In Table 7.2 is the maximum likelihood solution. It is closest to the iterated communality solution. Its estimated factor loadings are less variable than any of the other solutions. Maximum likelihood judges the .242 correlation between X_1 and X_2 to be too small, which raises the estimate of the loadings for those two variables compared to the pooled path analytic and iterated communality solutions. In general, the solution with iterated communalities provides a solution very close to the maximum likelihood solution.

In confirmatory factor analysis for overidentified models one can also test the fit of the model to the observed correlations. The test is a χ^2 goodness of fit test with as many degrees of freedom as the number of independent overidentifying restrictions. If the χ^2 is not significant then the model fits the data. Thus one is testing the null hypothesis. The χ^2 test is an approximation that is valid only for large samples. For the correlations in Table 7.1 the χ^2 is 1.94 and is not significant, which indicates that a single-factor model satisfactorily fits the data. One can view the χ^2 as the combined test of the vanishing tetrads. In Table 7.1 above the main diagonal, we have the residuals from the maximum likelihood solution that is, the observed correlations minus the predicted correlations. For a single-factor model the predicted correlations between two variables is simply the product of the standardized factor loadings. If one intends to compare loadings across populations or over time, one ordinarily should estimate the unstandardized loadings. One would fix the loadings of one of the measures to some arbitrary nonzero value, usually one, and the remaining loadings and factor variance would then be free. Because an

unstandardized solution is desired, a variance-covariance should be the input to the factor analysis. If one has obtained the standardized solution, one can destandardize in the following way. For the above example assume for illustrative purposes $V(X_1) = 2$, $V(X_2) = 3$, $V(X_3) = 2$, and $V(X_4) = .5$. If we fix the unstandardized loading of X_1 on F as one, it follows from the standardized solution that

$$r_{1F} = \frac{C(X_1, F)}{(V(X_1)V(F))^{1/2}} = .652$$

and from the unstandardized solution that

$$\frac{C(X_1, F)}{(V(F))} = 1.000$$

Solving through for $V(F)$ yields

$$V(F) = 2(.652^2) = .850$$

the loadings for the other variables can now be destandardized to yield

$$r_{2F} \frac{V(X_2)^{1/2}}{V(F)^{1/2}} = 0.811$$

$$r_{3F} \frac{V(X_3)^{1/2}}{V(F)^{1/2}} = 1.233$$

$$r_{4F} \frac{V(X_4)^{1/2}}{V(F)^{1/2}} = 0.688$$

Standardized loadings are often more interpretable, but the unstandardized loadings should, in principle, replicate better across populations and occasions.

A second illustration is taken from the doctoral dissertation of Hamilton (1975). One hundred ninety persons viewed an audio-visual presentation of a military court martial. The case, adapted from a real case (the names were changed to protect the guilty), concerned a soldier charged with the crime of killing four Korean civilians. After viewing the case, 190 observers were asked to state a verdict (X_1), choose a sentence (X_2), rate the soldier's innocence (X_3), and rate his responsibility (X_4). The intercorrelation matrix of the four variables is in Table 7.3. As might be expected, innocence correlates negatively with the three other variables. If possible, it is useful to reflect or reverse the scale of certain variables to obtain all positive correlations. This is not done here to show that negative factor loadings in no way interfere with the analysis.

Table 7.3 Judgements of Person in Mock Trial: Correlations below the Main Diagonal and Residuals above Diagonal^{a,b}

	X_1	X_2	X_3	X_4
Verdict (X_1)	1.000	-.001	-.008	-.012
Sentence (X_2)	.412	1.000	.042	.055
Responsibility (X_3)	.629	.403	1.000	.001
Innocence (X_4)	-.585	-.270	-.500	1.000

^aN = 190.

^bData taken from Hamilton (1975).

In Table 7.4 are the four different sets of estimates of the path coefficients for the single-factor model. Again, only the principal components solution markedly differs from the three other solutions. Again it tends to overestimate the loadings. The other three sets of estimates are very similar, and again the iterated solution is closer to the maximum likelihood solution than the pooled path analytic solution. Also the maximum likelihood solution is the least variable. In Table 7.3 above the main diagonal we have the residuals. Except for the sentence variable (X_3), they are small.

The test of tetrads 7.1, 7.2, and 7.3 yields χ^2 values of 0.99, 0.64, and 2.58, respectively, none of which are statistically significant. The goodness of fit χ^2 for the maximum likelihood solution is 2.62 which also is not significant.

Table 7.4. Factor Loadings for the Correlations in Table 7.3

Variable	Solution			
	Pooled Path Analytic	Principal Components	Iterated Communalities	Maximum Likelihood
X_1	.860	.865	.854	.854
X_2	.475	.631	.481	.484
X_3	.779	.830	.762	.746
X_4	-.637	-.767	-.656	-.672

It should be repeated that in this chapter the specification has been made to give the factor unit variance. Such a specification is consistent with traditional factor analysis models, but it is inconsistent with the common practice in the structural modeling literature. There the usual practice is to designate one of the indicator variables as a marker. Its loading is then fixed to one and the remaining loadings are then free as well as the variance of the factor. Such a strategy is much more valid if the researcher either intends to compare loadings over time, compare the loading of different populations, or desires to fix error variances equal, that is, assumes the measures are tau equivalent.

CONVERGENT AND DISCRIMINANT VALIDATION

It is unfortunate that researchers rarely check for single factoredness. To some degree this is a legacy due to Spearman's two-factor theory of intelligence. Spearman argued that all cognitive measures can be characterized by two factors. One is the g factor of general intelligence, and the second factor is unique. Although called the "two-factor theory of intelligence," a consequence of the theory is

that measures of cognitive skills should be single factored. Subsequent research showed that Spearman's theory did not hit the mark, and because then there has been an overabundance of emphasis on multiple-factor solutions.

Table 7.5 Schwartz and Tessler Transplant Data^a

	χ^2
Kidney	
Relative	1.01 ^b
Stranger	6.47 ^c
Heart	
Relative	.97
Stranger	3.13
Marrow	
Relative	2.49
Stranger	15.86 ^c

^aN = 195.

^bDegrees of freedom equals two.

^cp < .05.

Tests of single factoredness can demonstrate discriminant validity. Brewer, Campbell, and Crano (1970) argue that, before causal models of the type discussed in Chapter 4 are entertained, it should be demonstrated that the measures are not all indicators of the same construct. Their argument is especially strong for the case in which all the measures were obtained by the same method. There is a disappointing reluctance to follow the Brewer et al. suggestion. It seems that researchers are not willing to test a model that is inconsistent with their hypotheses. One should, however, remember that by its very nature a null hypothesis is inconsistent with the experimental hypothesis.

Consider the study by Schwartz and Tessler (1972) on the relationship between attitudes and behavior. The dependent variable of their study is behavioral intention to donate an organ for transplant. The independent variables are attitude, personal norms, and social

norms. Using multiple regression Schwartz and Tessler conclude that personal norms are the major cause of behavioral intention followed by attitude and, finally, social norms. Because all the measures are self-report, it is plausible that all the measures tap the same latent variable. It is not inconceivable that one's stated attitude, behavioral intention, and report of personal and social norms are all indicators of some underlying disposition. One might posit that the four measures are caused by a single factor. In Table 7.5 are the χ^2 tests of goodness of fit of such a single-factor model for six transplant operations. Because four parameters are estimated (the factor loadings of the four variables) and because there are a total of six correlations between the four variables, the χ^2 goodness of fit tests has two degrees of freedom. For four of the six transplant operations a single-factor model is consistent with the data, that is, the χ^2 is not significant. In the giving of kidney and marrow to a stranger cases, the single-factor model is inconsistent with the correlations. In all six instances the estimated factor loadings are greatest for behavioral intention variable (.833 to .916) which suggests that the underlying factor reflects behavioral intentions more closely than it does attitude or norms. There is some support for Schwartz and Tessler's use of multiple regression but no overwhelming support. Discriminant validation is not clearly indicated.

SPECIFICATION ERROR

Although all the models that have been elaborated have nicely fit a single-factor model, unfortunately this does not always happen. When the model does not fit the data, some sort of specification error is indicated. (Moreover even if the model fits, a simpler model may fit better.) An additional specification of correlated disturbances (that is, correlated unique factors) can be made to salvage a single-factor model. Correlated disturbances imply that there is an additional source of covariance between two

measures besides the factor. A researcher can specify as many correlated disturbances as there are overidentifying restrictions, but the researcher must be careful about which are made because certain patterns of correlated disturbances may yield an underidentified model. For instance, with four variables there are two independent overidentifying restrictions, and so two sets of disturbances can be correlated. If the disturbances of X_1 and X_2 and also of X_3 and X_4 are correlated, the model is underidentified. As shown in Chapter 3, an overidentifying restriction holds:

$$\rho_{13}\rho_{24} - \rho_{14}\rho_{23} = 0$$

The overidentifying restriction plus the six free parameters is one more than the six correlations. Thus, if correlated disturbances are added to the model, one must be careful about the status of identification. In general, to identify an n -variable single-factor model with correlated disturbances, there must be first at least three variables whose disturbances are not intercorrelated and, second, for every other variable, its disturbance must be uncorrelated with at least one other variable whose factor loading can be estimated. Some patterns of correlated disturbances are also illogical. For instance, if the disturbance of X_1 is correlated with X_2 and X_3 , more often than not one should have the disturbances of X_2 and X_3 be correlated. It makes almost no sense to have them uncorrelated.

Ideally the correlated disturbances can be specified by theory. More often they are unanticipated and may be indicated by large residuals. Of course, adding correlated disturbances after examining residuals makes the data analysis more exploratory than confirmatory. Moreover, examination of the residuals may be misleading (Costner & Schoenberg, 1973).

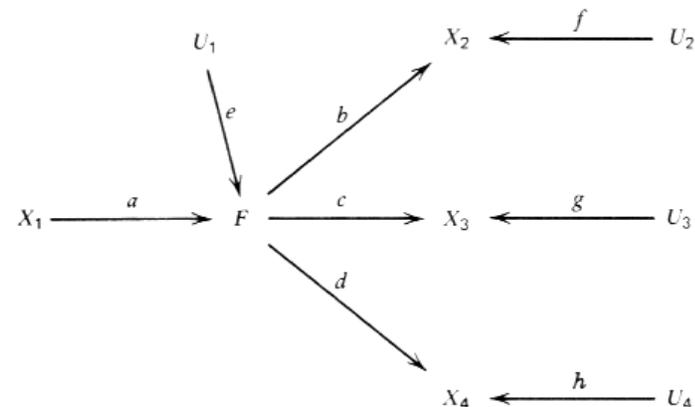


Figure 7.3 One indicator as a cause.

Recall that the overidentifying restriction of a single-factor model is a vanishing tetrad. If the disturbances of any two variables are allowed to be correlated, then the correlation between those two variables cannot be used in any tetrad. Thus, if the disturbances of X_1 and X_2 are correlated, tetrads 7.1 and 7.2 do not necessarily hold. However, the fact that only tetrad 7.3 holds does not necessarily imply that the disturbances of X_1 and X_2 are necessarily correlated. If the disturbances of X_3 and X_4 are correlated, again only tetrad 7.3 holds. Thus, one cannot necessarily use the tetrads to determine exactly which disturbances are correlated. The data always are consistent with a multitude of structural models.

If correlated disturbances are specified, the model does not necessarily always recover the sample variances. Thus, there are residuals not only for the correlations (covariances) but also for the variances.

A final type of specification error cannot be tested. In Figure 7.3 is a model somewhat different than the model in Figure 7.2. The variable X_1 causes F whereas in Figure 7.2 F causes X_1 . There is, however, no way to distinguish the two models empirically. They both have the same overidentifying restrictions. One might then reinterpret

the correlations in Table 7.1 as showing that concern about one's self causes some latent variable which in turn causes concerns about family, friends, and life.

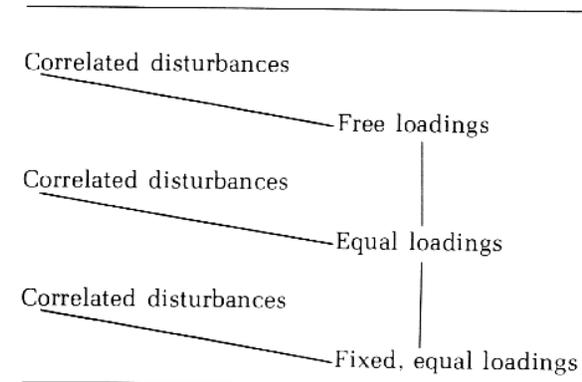
HIERARCHICAL MODEL TESTING

It is possible to compare models in confirmatory factor analysis to judge whether the fit improves when more parameters are added to the model or, alternatively, to judge whether the fit worsens when parameters are deleted from the model. This procedure of comparing models is similar to model comparison methods in analysis of variance (Appelbaum & Cramer, 1974), multiple regression, and log-linear models for contingency tables (Fienberg, 1977). In Table 7.6 are a series of single-factor models. Near the top of the tree note the entry, free loadings. This represents a model in which the loadings for each variable are free to vary. Below that entry is a model with equal loadings. This model is more restrictive than the model above it, because a model with equal loadings is a special case of a model with free loadings. Still more restrictive is a model in which the loadings are all fixed equal to some value, say zero. A model with fixed equal loadings is a special case of a model with free equal loadings.

If there are two models, say A and B, and model A is more restrictive than B, the test of the restrictions added to model A can be obtained by subtracting the χ^2 for model B from the χ^2 for model A. The difference between the χ^2 values is itself a χ^2 with degrees of freedom equal to the difference in the degrees of freedom of the two χ^2 values. For example, to test the additional constraints of equal loadings over free loadings, one takes the smaller χ^2 and subtracts it from the larger. The degrees of freedom of this χ^2 are the corresponding difference in degrees of freedom. If the difference χ^2 is not significant, then one cannot reject the null hypothesis that the factor loadings are equal. Similarly one can compare a model with equal loadings to a model in which the

loadings are all fixed at zero. The comparison of the model with free loadings versus a model with loadings equaling zero simultaneously tests equal loadings and zero loadings.

Table 7.6 A Hierarchy of Models



Each of the three models in Table 7.6 has a branch. For each model a pair of disturbances has been correlated, the same pair for each model. One can then test for the existence of correlated errors in three different ways: with free loadings, with equal loadings, and zero loadings. The resultant χ^2 values will each have one degree of freedom but will not necessarily be equal. The same parameter is being tested but it is added to different models. One cannot then speak of a unique test that a certain parameter equals zero. It can be discussed only in the context of model comparison. This is also true of other statistical methods. For example one can test whether a beta weight is zero in multiple regression only in the context of the other predictor variables. The beta weight may be significant with one set and not with another.

Clearly then it is not sufficient to estimate a single model and be satisfied if it fits. Somewhat simpler models may fit the data better, and possibly even more complex models may indicate important specification errors

FACTOR MEASUREMENT

Actually the goal of most users of factor analysis is not to test a factor model but to measure factors. A measured factor for a participant is usually called a factor score. Very often a model can be estimated and tested without actually measuring the factors. Unfortunately researchers suffer from the delusion that, unless one can assign "real scores" to participants, the variable is not real. Just because latent variables are not observable does not mean that they are not real. Much in the same way, atoms are not observable, but they are very real, nonetheless.

One approach to factor measurement is multiple regression. Because the correlation between each of the indicators and the factor is known (the factor loadings in the single-factor case), multiple regression can be used to predict factor scores. Thus, the multiple regression prediction equation is used to estimate factor scores since the true factor scores are not available. For a single-factor model the regression weights are proportional to

$$\frac{b_i}{1 - b_i^2} \quad [7.6]$$

where b_i is the loading of X_i on the factor. If any loading is near one then Formula 7.6 will be very large. A simple rule is that if a loading is .975 or better, then simply take that variable as the factor score. If the loading is less than .975 then all the weights will be less than 20.

Instead of using multiple regression weights to create the factor scores, one could simply add the variables, weighting them equally. The use of equal weights is called *unit weighting*. Regardless of whether regression or unit weights are used, it is useful to derive the correlation of any composite with the factor since the correlation squared measures the reliability of the factor score. Let a_i stand for the weight for variable X_i ; then the composite is defined as $\sum_i a_i X_i$. For a

composite of measured *standardized* variables, $\sum_i a_i X_i$, its covariance with a standardized factor is

$$\sum_i a_i b_i \quad [7.7]$$

where b_i is the correlation of the variable with the factor. Thus with unit weights the covariance is $\sum_i b_i$ because the a_i are all one. In the single-factor case b_i can be estimated by the factor loading of X_i . To find the correlation of the factor with the composite, one must divide Formula 7.7 by the standard deviation of the sum or composite. The variance of the composite is $\sum_i \sum_j a_i a_j r_{ij}$ where a_j is the weight for X_j , and so the correlation is

$$\frac{\sum_i a_i b_i}{(\sum_i \sum_j a_i a_j r_{ij})^{1/2}}$$

The squared correlation between the factor and composite is the reliability of the composite. One can then compare the reliability estimates by using different methods of forming the composites. If the weights are chosen by multiple regression, the reliability of the composite is at a maximum. Any other composite theoretically has lower reliability. However, the regression weighted composite does capitalize on chance, and most certainly reliability estimates are inflated. Moreover, a unit weighting of variables often correlates .95 or better with the regression weighted composite.

Cronbach's alpha is a commonly suggested measure of reliability of linear composites. It tends to underestimate the reliability of a composite as the factor loadings become more variable. It does have the advantage of computational ease. The formula for alpha using unit weights of standardized scores is

$$\alpha = \frac{nM_r}{1 + (n-1)M_r}$$

where n is the number of measures and M_r is average correlation between measures. The formula is sometimes called the generalized Spearman-Brown formula.

As an illustration, the reliability of factor scores is estimated for the study by Hamilton in Table 7.3. First one regresses F , the factor, on X_1 through X_4 . The correlations among the X variables are given in Table 7.3 and the correlations with the factor are the maximum likelihood factor loadings. Obtained is the following regression estimate of the factor score using Formula 7.6:

$$3.155X_1 - .632X_2 + 1.682X_3 - 1.225X_4$$

The covariance of the preceding with F is 5.078 and its variance is 30.859. Thus the squared correlation of the composite with F is .836 which is the reliability of the regression weighted composite. If we simply sum the measures by $X_1 + X_2 + X_3 - X_4$, the variance is 9.598. The correlation of the simple sum with regression composite is .973. We have then gained little or nothing by doing the regression analysis. The reliability of the sum is .791, which is only slightly less than .836, and so little was sacrificed by unit weights. Finally Cronbach's alpha of the sum is

$$\begin{aligned}\alpha &= \frac{(4)(.4665)}{1 + (3)(.4665)} \\ &= .778\end{aligned}$$

which closely approximates the .791 value.

Before moving on to models with multiple unmeasured variables, which are discussed in the next chapter, some discussion about the meaningfulness of unmeasured variables is in order. Some researchers do not believe that factors have any meaning because the researcher does not know the score of each participant on the factor. Since the researcher can only derive factor scores, which do not

perfectly correlate with factor, the factor remains a mysterious platonic ideal. Such a view suffers from an over-reliance on single operationalism; that is, the explanatory concepts in science are measured variables. A multiple operational approach would argue that explanatory concepts are rarely directly measured in the social sciences. Our measures of motivation, attitudes, intelligence, group cohesion, and ideology only imperfectly tap the constructs that have been posited. From this point of view measurement and unreliability go hand in hand. Thus to measure one must triangulate by using multiple operations. However, even the sum of the multiple operations does not define the construct. As is seen in the next two chapters very strong tests can be made of models with multiple operationalizations of multiple constructs. In none of the models is the construct measured, but rather the construct is simply posited to be imperfectly exemplified by its multiple operationalizations. Latent constructs are no less real than are measured variables unreal, since the factors more closely correspond to scientific concepts while measures are inherently noisy and provide only the shadows of social reality.

8

MULTIPLE UNMEASURED VARIABLES

Most models with unmeasured variables usually have more than one measured variable, each with multiple indicators. The presence of multiple unmeasured variables adds many new complexities beyond a single unmeasured variable. The concern of this chapter is with what has been called the *measurement model*. It is the model that maps the measures or indicators onto the factors or unmeasured variables. The unmeasured variables may well be correlated, but in this chapter such correlations and covariance's remain unanalyzed. The following chapter concerns itself with a structural analysis of these correlations.

This chapter is organized in three sections. The first section extensively examines two factor models. Most of the important material in this chapter can be found in this section. The second section considers models with three or more factors. In the third and last section the multitrait-multimethod matrix is analyzed.

TWO-CONSTRUCT MULTIPLE INDICATOR MODEL

IDENTIFICATION

In Figure 8.1 is a structural model with two unmeasured variables, *F* and *G*, and five measured variables, *X*₁ through *X*₅. The unmeasured variables or factors are assumed to be uncorrelated both

with each other and with the disturbances. There are a total of 10 free parameters, the loadings on the factors, and a total of 10 correlations (5 x 4/2). It would seem that the model is just-identified since there are as many parameters as correlations. Try as you can, you will never be able to solve for the parameters. It happens that a rather complex overidentifying restriction holds:

$$[\rho_{24}\rho_{35} - \rho_{25}\rho_{34}] [\rho_{12}(\rho_{13}\rho_{45} - \rho_{15}\rho_{34}) - \rho_{14}(\rho_{13}\rho_{25} - \rho_{15}\rho_{23})] - [\rho_{23}\rho_{45} - \rho_{25}\rho_{34}] [\rho_{12}(\rho_{14}\rho_{35} - \rho_{15}\rho_{34}) - \rho_{13}(\rho_{14}\rho_{25} - \rho_{15}\rho_{24})] = 0$$

The reader will be happy to be spared a proof of the preceding condition, first shown by Kelley (1935). Given the preceding overidentifying restriction, one correlation is lost and there is no solution for the parameters of the model. Thus, even though the model contains an overidentifying restriction, not a single parameter is estimable. The problem cannot be solved by adding more variables. Even if there were 100 variables each loading on the two factors, it would only increase the number of overidentifying restrictions and still the parameters would remain underidentified.

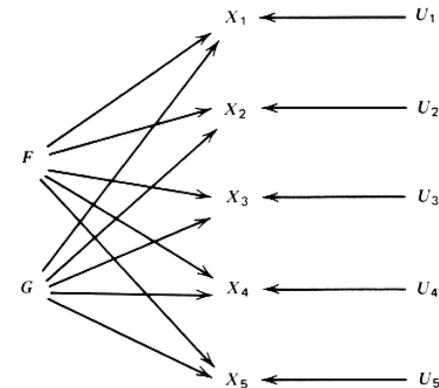


Figure 8.1 Two-factor model with uncorrelated factors.

To be able to identify and then estimate factor loadings for models with two or more factors, we must have constraints on the factor solution. In general, to estimate the factor loadings and factor covariance matrix, first the minimum condition of identifiability must be satisfied (that is, as many correlations as free parameters) and second p^2 parameters must be fixed where p is the number of factors. These p^2 fixed parameters may be either factor loadings, correlations (covariances) between factors, or the variances of the factors. These parameters are usually fixed by setting their values to zero. For instance, a loading may be set to zero implying that the factor does not cause the particular measure, or a correlation (covariance) between two factors may be set to zero. A second common method of fixing a parameter is to set it to the value of one. For factor variances this implies that the factor is standardized, or for factor loadings a value of one puts the factor in the same units of measurement as the variable. Other possible methods of fixing parameters are equality and proportionality constraints.

Even when p^2 parameters are fixed the model may still not be identified since it is only a necessary but not sufficient condition for identification. Recall that the condition also holds for single factor models. Since $p^2 = 1$, there must be one constraint. It is usually either that the variance of the factor is one or one factor loading is one. Forcing the equality of two loadings is not by itself a sufficient condition to identify the model.

With two factor models there must be $2^2 = 4$ constraints. If four constraints are not made or if they are not the right four, the model is not identified. Since there is no unique solution, in factor analysis parlance the solution can be rotated. Rotation and all its attendant problems are brought about because most factor analysis methods do not specify an identified model. The square root method of factor analysis does presume an identified model (Harmon, 1967, pp. 101-103). It constrains the p factor variances to be one (p constraints), the factor correlations to be zero ($p(p - 1)/2$ constraints), and for the loading matrix to be triangular, that is, all variables load on the first

factor, all but the first variable load on the second factor, and so on until only the last measure loads on the last factor ($p(p - 1)/2$ constraints). Such a model may be identified with a total of p^2 constraints. However, to satisfy the minimum condition of identifiability, a loading of each factor must be fixed. Although it may be identified it rarely makes any structural sense.

The approach taken in this chapter is to allow each measure to load on only one factor. In most cases this will lead to identification, as well as allow the factors to be correlated.

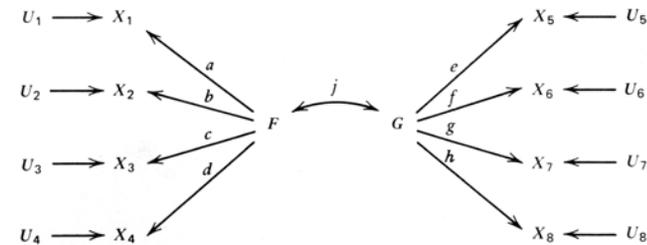


Figure 8.2 Two-factor multiple indicator model.

THE MODEL

Consider a two-factor model with each factor having four indicators as in Figure 8.2. Four measures, X_1 through X_4 , are indicators of F and the other four, X_5 through X_8 , are indicators of G . The two factors are correlated, or oblique. (The modern convention is draw circles around the latent variables and boxes around the measures.) In Figure 8.2 it is assumed that indicators of one construct are not caused by the other construct (so-called simple structure) and that the unique factors (disturbances or errors of measurement), are uncorrelated with each other and with each of the two constructs. An indicator then is defined as a measure that is caused by a single construct. Thus, in this case the indicators of G do not load on F and

the indicators of F do not load on G . The factor loading matrix is as follows:

	<u>F</u>	<u>G</u>
X_1	a	0
X_2	b	0
X_3	c	0
X_4	d	0
X_5	0	e
X_6	0	f
X_7	0	g
X_8	0	h

As in the previous chapter, the correlation of the indicator with the construct is called the *epistemic correlation* or more commonly the *factor loading*. In Table 8.1 the tracing rule has been used to derive the theoretical correlations for the model in Figure 8.2. The correlations in the box of Table 8.1 are called the cross-correlations. These correlations involve the correlations between indicators of different constructs. The correlations above the box are correlations of the indicators of F , and beside the box are the correlations of the indicators of G .

Table 8.1. Theoretical Correlations for Figure 8.2

X_1									
X_2	ab								
X_3	ac	bc							
X_4	ad	bd	cd						
X_5	aej	bej	cej	dej					
X_6	afj	bfj	cfj	dfj	ef				
X_7	agj	bgj	cgj	dgj	eg	fg			
X_8	ahj	bhj	chj	dhj	eh	fh	gh		
	X_1	X_2	X_3	X_4	X_5	X_6	X_7		X_8

The minimum condition of identifiability has been established for the model in Figure 8.2 since there are 28 correlations between measured variables, and only 9 free parameters, 8 factor loadings, and

1 factor correlation. Because 8 factor loadings have been constrained to be zero and the variance of the factors to be 1, enough constraints have been made to obtain a solution. Not only is the model identified, there is also a plethora of overidentifying restrictions, 19 in all (28 - 9). If the researcher chooses to set the units of the factor to that of a measure by fixing the measure's loading to one, then there are still 19 overidentifying restrictions. In such a case the loading is no longer free, but the factor variance is free.

ESTIMATION

The traditional factor analytic solution to the model in Figure 8.2 would be a principal components solution with some form of oblique rotation. However, unless all the overidentifying restrictions are perfectly met, no rotation will force the loadings to be zero where they should be. Some readers may have been faced with the difficult decision of how oblique should an oblimax rotation be. A Procrustes rotation (Gorsuch, 1974, pp. 166-168) might be tried, but it requires not only the specification of the zero loadings but also exact values of the nonzero loadings. Another factor analytic solution to the model in Figure 8.2 is the multiple group solution (Harmon, 1967; Gorsuch, 1974) which is rarely used. Like most factor analytic solutions, it too suffers from the communality problem; that is, communalities must be specified in advance. A variant of the multiple group solution to the model in Figure 8.2 is cluster analysis (Tyron & Bailey, 1970).

The path analytic estimates which were originally proposed by Costner (1969) are straightforward. To estimate the epistemic correlations simply pool all the estimates of each. An estimate of the epistemic correlation for X_i can be shown to be $(r_{ij}r_{ik}/r_{jk})^{1/2}$ where X_j is an indicator of the same construct as X_i whereas X_k need not be. Then to pool the estimates of the squared epistemic correlation, sum the absolute values of the numerators of all estimates and divide it by the sum of the absolute values of the denominators. If there are a indicators of construct F and b of G , there are a total $(a - 1)(b + a/2 - 1)$ estimates

of the epistemic correlation or loading for each indicator of G and $(b - 1)(a + b/2 - 1)$ for each indication of F . To estimate r_{FG} pool all of its estimates. They are of the form

$$r_{FG}^2 = \frac{r_{ik}r_{jl}}{r_{ij}r_{lk}}$$

where i and j are indicators of F and l and k indicators of G . There are a total of $ab(a - 1)(b - 1)/2$ possible estimates to be pooled. Again the numerators, and then the denominators, of each estimate are summed.

The estimate is the ratio of the pooled numerator to the pooled denominator. The only estimation procedure based on a statistical analysis is maximum likelihood factor analysis (Jöreskog, Gruvaevs, & van Thillo, 1970). A two-factor model is specified, with each variable loading on a single factor. The two factors can be correlated. A structural equation modeling program can allow correlated errors, equal standardized or unstandardized loadings, zero-factor correlations, or equal error variances. The overall fit of the model can be evaluated and restricted models can be compared by χ^2 goodness of fit tests. All the examples are estimated by employing maximum likelihood factor analysis.

SPECIFICATION ERROR

Instead of testing whether all the overidentifying restrictions hold for the model in Figure 8.2 simultaneously, it is useful to partition the restrictions into tests of different constraints on the model.

One obvious set of restrictions to examine is the correlations within a construct. The correlations within each construct should evidence a single-factor structure. For factor F , since there are four measures, there are two overidentifying restrictions. These are referred to as the test of homogeneity within indicators of a construct. More compactly, the restrictions are called *homogeneity within*. As shown in the previous chapter, homogeneity within implies tetrads of the form

$\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$, where i, j, k , and l are all indicators of the same construct.

Homogeneity between constructs can also be defined. An examination of the correlations in Table 8.1 between indicators of F and G (the cross-correlations), evidences a number of overidentifying restrictions. The restrictions follow from the fact that the cross-correlations can be used to estimate the ratio of epistemic correlations. For instance, to estimate a/b one could examine

$$\frac{a}{b} = \frac{r_{15}}{r_{25}} = \frac{r_{16}}{r_{26}}$$

which implies a vanishing tetrad of

$$\rho_{15}\rho_{26} - \rho_{16}\rho_{25} = 0$$

In general, denoting i and j as indicators of F , and k and l as indicators of G , the following vanishing tetrad holds:

$$\rho_{ik}\rho_{jl} - \rho_{il}\rho_{jk} = 0$$

The final set of restrictions is called the test of consistency of the epistemic correlations. From the correlations within a construct, the epistemic correlations can be estimated. From the cross-correlations the ratio of these epistemic correlations can also be estimated. These two sets of estimates should be consistent. A test of consistency is again a vanishing tetrad because

$$\frac{a}{b} = \frac{r_{13}}{r_{23}} = \frac{r_{15}}{r_{25}}$$

which implies

$$\rho_{13}\rho_{25} - \rho_{23}\rho_{15} = 0$$

Because from the ratios a/b , b/c , and c/d all the remaining ratios can be obtained, there are then a total of three restrictions on the consistency of the estimates of each construct. In general consistency of F implies a tetrad of the form

$$\rho_{ij}\rho_{kl} - \rho_{ik}\rho_{jl} = 0$$

where i, j , and k are indicators of F and l an indicator of G .

If it is assumed that there are a indicators of construct F and b of G , the total number of correlations is $(a + b)(a + b - 1)/2$ and the total number of free parameters in the model is $a + b + 1$. Subtracting the latter from the former yields $(a + b)(a + b - 3)/2 - 1$, which is the total number of overidentifying restrictions on the model. These restriction can be divided into homogeneity and consistency restrictions. There are a total of $(a^2 - a)(a - 2)(a - 3)/8$ tetrads testing homogeneity within construct F of which only $a(a - 3)/2$ are independent. Similarly for construct G , there are $(b^2 - b)(b - 2)(b - 3)/8$ tetrads with $b(b - 3)/2$ independent restrictions. For homogeneity between constructs, there are a total of $ab(a - 1)(b - 1)/4$ tetrads of which $(a - 1)(b - 1)$ are independent. For the consistency of F , there are a total of $(a^2 - a)(a - 2)b/2$ tetrads, but only $a - 1$ are independent given homogeneity within construct F and homogeneity between constructs F and G . Similarly for the consistency of construct G , there are $(b^2 - b)(b - 2)a/2$ tetrads of which only $b - 1$ are independent. If one adds up the total number of independent restrictions they equal $(a + b)(a + b - 3)/2 - 1$, which is the number of overidentifying restrictions on the entire model.

To assess homogeneity within a construct one needs at least four indicators of the construct. For homogeneity between, one needs only two indicators for both constructs. To test for consistency, one needs three indicators for the construct whose consistency is being tested and only one indicator of another construct.

To statistically evaluate the different constraints on the model, the following procedure can be used. Certainly there are other

procedures that are better, but this method is fairly straightforward and simple. To test for homogeneity within a construct, test for a single factor by using maximum likelihood factor analysis. If the indicators of the factor are not single-factored, test each of the tetrads using canonical correlation. Hopefully, an examination of the tetrads that fail to vanish may point to the troublesome indicators. To test for homogeneity between, perform a canonical correlation analysis treating the indicators of one construct as one set and the indicators of the other construct as the second set. The test that the second canonical correlation is zero will test for homogeneity between constructs. If the test fails then the individual tetrads can be tested to point out the troublesome indicators.

There is no simple test for consistency. A rather tedious strategy is to test all the tetrads. This should be done only after both the homogeneity between and within tests have been passed.

If a given test fails, what sort of specification error is indicated? If homogeneity within is not satisfied, it indicates that the disturbances of the measures are correlated or there may be a second factor. If homogeneity between is not satisfied, most likely disturbances are correlated across constructs. For instance, two indicators may share method variance.

The specification error for consistency is rather interesting. For Figure 8.2 imagine there is a path from the construct G to X_1 and thus X_1 would load on both F and G . In such a case homogeneity within and between would be satisfied as well as the consistency of the indicators of G , but the indicators of F would not evidence consistency. In particular, tetrads involving the correlations of X_1 with an indicator of F and an indicator of G would not vanish. So in general, violations of homogeneity within indicate correlated errors within indicators of a construct, violations of homogeneity between indicate disturbances correlated across different constructs, and violations of consistency indicate that certain measures load on both factors. Of course, if the model is grossly misspecified such tests may not be particularly

diagnostic. However, a model with a few specification errors can be improved by such an analysis.

IDENTIFICATION REVISITED

The reader is probably wondering what is the minimum number of indicators necessary for multiple indicator models. Since obtaining indicators of constructs may be difficult, it would be helpful to know the smallest number that are needed. Consider the model in Figure 8.3. There are only two indicators of each of the two constructs, four variables in all. The model is, however, identified. For instance, the parameter a^2 equals:

$$a^2 = \frac{\rho_{12}\rho_{23}}{\rho_{23}} = \frac{\rho_{12}\rho_{14}}{\rho_{24}}$$

Note, however, that the denominators of both the preceding formulas depend on r_{FG} : $r_{23} = bcr_{FG}$ and $r_{24} = bdr_{FG}$. Thus, if r_{FG} equals zero or nearly equals zero, the model is empirically underidentified since the denominators are zero or near zero. Thus, assuming no correlated disturbances, it is possible to estimate loadings given that the construct correlates with other constructs in the model. Of course, as in single-factor models the standardized loadings must not be small or else the model is again empirically underidentified.

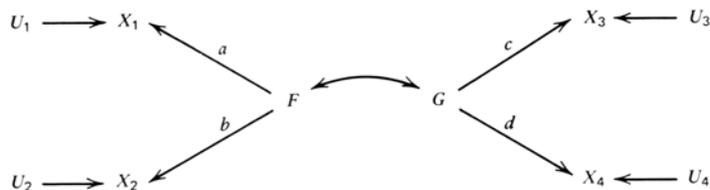


Figure 8.3 Two-factor model, each factor with two indicators.

Although two indicators are sufficient to allow for identification, it is desirable to have at least three indicators per

construct. One reason is fairly obvious. The model may be identified in principle, but underidentified empirically, that is, the constructs may correlate low or the loadings may be too small. A second problem is specification error. The investigator may be unable to allow for correlated disturbances if there are only two indicators per construct. Consider the model in Figure 8.5, which is discussed later in this chapter. Note that the errors of X_1 and X_4 , X_2 and X_5 , and X_3 and X_6 are correlated. Such a model may be very plausible, if the X_1 and X_4 are the same variable measured at times 1 and 2, respectively, and X_2 and X_5 are again the same variable measured at two points in time as well as X_3 and X_6 . The model in Figure 8.5 is identified with three indicators at each time point; it would not be if there were only two indicators.

Just as three indicators are better than two, so are four better than three. With only three indicators it is not possible to test for homogeneity within a construct. With four such a test can be made and possible correlated errors within indicators of a construct can be spotted. Having five indicators is only slightly better than having four. I suggest the following rule of thumb: Two might be fine, three is better, four is best, and anything more is gravy. To complicate matters slightly in some very special circumstances, it is seen in the next chapter that a construct need have only a single indicator.

ILLUSTRATIONS

In Table 8.2 there are a set of correlations taken from data gathered by Brookover and reanalyzed by Calsyn and Kenny (1977). The measures were taken from 556 white eighth-grade students and are:

- X_1 —self-concept of ability
- X_2 —perceived parental evaluation
- X_3 —perceived teacher evaluation
- X_4 —perceived friends' evaluation
- X_5 —educational aspiration
- X_6 —college plans

Table 8.2. Correlations between Ability and Plans below Diagonal Residuals above Diagonal^{a,b}

	1	2	3	4	5	6
X_1	1.000	-.003	.005	-.020	.014	.026
X_2	.730	1.000	-.004	.020	-.009	-.006
X_3	.700	.680	1.000	.010	-.016	-.018
X_4	.580	.610	.570	1.000	.011	-.020
X_5	.460	.430	.400	.370	1.000	.000
X_6	.560	.520	.480	.410	.720	1.000

^aData taken from Calsyn and Kenny (1977) and correlations are computed to only two decimal places.

^b $N = 556$.

The three perceived measures were all the perceptions of the student. In Figure 8.4 there is a model for the six variables. The four measures of evaluation of the child are assumed to be indicators of self-perception of ability, A , and the two aspiration measures are called plans, or P . Also in Figure 8.4 there are the parameter estimates of the model using maximum likelihood factor analysis. The epistemic correlations are all high, as is the correlation between A and P . It is instructive to note that none of the cross-correlations are as large as the estimate of the true correlation between constructs because all the cross-correlations are attenuated because of measurement error in the indicators. It is instructive to compare alternative strategies of estimating the correlation between A and P . If one randomly picked a single indicator of A and correlated it with a randomly chosen indicator of P , the correlation between A and P would be estimated as .454. If one unit weighted (added up the standard scores) both sets of indicators, the correlation would be .571. If one computed a canonical correlation between the two sets of indicators, it would be .597. Even though the reliabilities of the indicators are moderate to large, all the estimates by standard procedures are appreciably lower than the .666 value estimated by confirmatory factor analysis.

The degrees of freedom for this model are $(6)(3/2) - 1 = 8$. The overall $\chi^2(8)$ is 9.26, well within the limits of sampling error. In fact, as can be seen by looking above the diagonal in Table 8.4, the largest residual is only .026. The power of the goodness of fit test should be high since the sample size and the correlations are large. To obtain a χ^2 test for the homogeneity of the indicators of A , a test of single-factoredness of the four indicators of A is conducted. A χ^2 of 3.13 is obtained with $(4)(1/2) = 2$ degrees of freedom. This χ^2 is not significant and so single factoredness of the A indicators is indicated. Since there are only two indicators of the P factor, one cannot test for homogeneity within it. The test of homogeneity between constructs A and P can be performed by testing whether the second canonical correlation between indicators of A and of P is zero. The χ^2 equals 1.95 with 3 degrees of freedom which is clearly not significant. To test for consistency, tests of the 24 tetrads were performed. Since there are multiple tests alpha (.05) was divided by 24, the number of tests. All the tetrads vanished using this significance level.

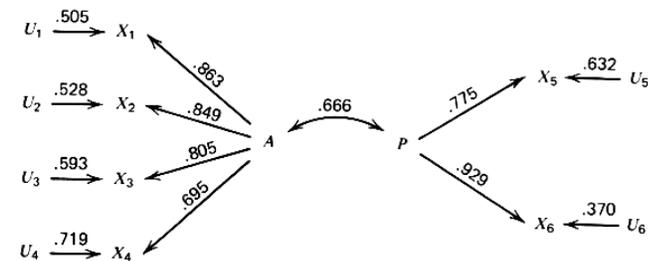


Figure 8.4 Multiple indicator model for the correlations in Table 8.2.

The second illustration is taken from a secondary analysis conducted by Milburn (1978). One hundred and forty adult smokers were asked questions on a four-point scale about smoking:

X_1 : unbearable to run out of cigarettes

X_2 : impossible to cut down

X_3 : feel more relaxed after smoking

The X_4 to X_6 measures are the same with the participants being measured one year later. All participants were smokers from a town in northern California. In Table 8.3 are the correlations between items and the standard deviation of each item. All the solutions reported below are unstandardized since the loadings of X_1 and X_4 are forced to one, and the variance-covariance matrix serves as input. The variance of the time 1 and time 2 factor is then free to vary.

Table 8.3. Correlations of Cigarette Smoking Items at Two Time Points, X_1 to X_3 - Time 1 and X_4 to X_6 Time 2. Residuals on and above Diagonal^{a,b}

X_1	.012	.038	-.010	.009	-.070	.019
X_2	.445	.016	-.017	-.018	-.040	.022
X_3	.274	.302	-.012	.122	-.041	.020
X_4	.601	.268	.340	-.002	-.040	.026
X_5	.201	.340	.162	.270	-.031	.042
X_6	.222	.249	.484	.268	.293	.022
Standard deviation	1.053	1.097	.909	.937	1.055	.800

^aN = 140.

^bData taken from Milburn (1978).

The first solution attempted forced the loadings of first, X_2 and X_5 and second, X_3 and X_6 , to be equal. Thus the unstandardized loadings were constrained to be stable over time. The χ^2 for the model is 47.19 with 10 degrees of freedom. Clearly some part of the model is misspecified. Because the same items were used at both points in time it is likely that the unique factors are correlated over time. To test this a canonical correlation analysis was set up treating $X_1, X_2,$ and X_3 as one group of variables and the remaining variables as the second set. The test of the $\chi^2(4)$ equals 29.72, which is highly significant. Thus correlated errors across constructs is indicated.

The model was respecified allowing the disturbances of the same measure to be correlated over time. The variance of the disturbances was set at one. The resultant parameter estimates are contained in Figure 8.5. The fit of the model is quite good: $\chi^2(7) =$

10.30. Residuals from the covariance matrix are reproduced in Table 8.3. Interestingly, the factor variance decreases over time being 0.406 at time 1 and 0.262 at time 2. Respondents seem to be becoming more homogeneous over time. The autocorrelation of the factor is $.251/((.406)(.262))^{1/2} = .770$.

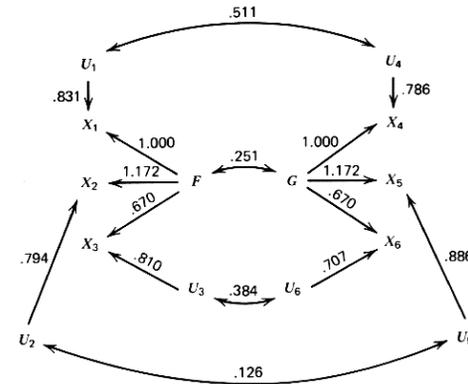


Figure 8.5 Unstandardized model for the covariance matrix in Table 8.3

To test the equality of the loadings over time, a model was set up that allowed the loadings to vary over time. However, to identify the model, the loadings for X_1 and X_4 were kept at one. The χ^2 for this model is 9.44 with 5 degrees of freedom. The test of equality constraints is the difference of the general model from the restricted model: $\chi^2(7)$ of 10.30 minus $\chi^2(5)$ of 9.44 or $\chi^2(2)$ of 0.86, which is not statistically significant, indicating that the fit does not suffer by imposing the equality constraints.

THREE-OR-MORE-CONSTRUCTS MULTIPLE INDICATOR MODEL

COMPLICATIONS

The addition of a third construct to multiple indicator models adds only a few complications. As the number of constructs grows the

number of correlations (covariances) between the constructs increases geometrically. The researcher may desire to specify a structural model for those relationships. Such models are discussed in the next chapter.

Multiple indicator models with many constructs usually tend to be highly overidentified. For instance, a model with four constructs, each with four indicators, with each indicator loading on only one factor, and no correlated errors has 98 overidentifying restrictions! For such models a small number of specification errors can greatly distort the parameter estimates and lack of fit can be virtually undetectable if one solely relies on the overall χ^2 test. The problem is analogous to an overall F test in the analysis of variance. Since there are so many degrees of freedom on the numerator, the overall test has low power to detect a few but important differences. Clearly one must think through the model and explicitly test for specific sorts of specification error. Failure to do so may often lead to parameter estimates that are very misleading.

It is not unusual for the structural model to contain a mix of constructs. Some of the constructs require multiple indicators while others may not. For instance, the model may contain demographic variables such as sex, ethnicity, or age. It seems reasonable to argue that these variables have perfect or near perfect reliabilities: They are perfect indicators with epistemic correlations of one. How can such variables be incorporated into multiple indicator models?

The simplest way to do so is to create a factor for each such measure. Then force the loading of the measure on the construct to be one and do not give the measure a disturbance. The factor will then become the measure, and its variance is set to the variance of the measure. Thus, if the variable is already standardized, the factor will also be standardized.

Before turning to an example of a multiple construct model, it should be noted that specification error can be evaluated in a different manner with three or more constructs. Earlier, consistency of estimates was defined. It involved comparing the ratio estimates of loadings determined by the within and between construct correlations. With

three or more constructs consistency can be evaluated by comparing two sets of between construct correlations. For instance, note that in Figure 8.6 the ratio of the path from L to X_1 to the path from L to X_2 equals

$$\frac{\rho_{13}}{\rho_{23}} = \frac{\rho_{14}}{\rho_{24}} = \frac{\rho_{15}}{\rho_{25}} = \frac{\rho_{16}}{\rho_{26}}$$

The preceding yields six vanishing tetrads only three of which are independent:

$$\rho_{13}\rho_{24} - \rho_{23}\rho_{14} = 0 \quad [8.1]$$

$$\rho_{13}\rho_{25} - \rho_{23}\rho_{15} = 0$$

$$\rho_{13}\rho_{26} - \rho_{23}\rho_{16} = 0$$

$$\rho_{14}\rho_{25} - \rho_{24}\rho_{15} = 0$$

$$\rho_{14}\rho_{26} - \rho_{24}\rho_{16} = 0$$

and

$$\rho_{15}\rho_{26} - \rho_{25}\rho_{16} = 0 \quad [8.2]$$

Two of the overidentifying restrictions are constraints of homogeneity first between L and A (Equation 8.1) and second between L and W (Equation 8.2). The final independent restriction can be called the consistency of L with W and A . Only when a construct has just two indicators does it make sense to evaluate consistency with two other constructs. If there are three or more indicators this form of consistency follows from both simple consistency and homogeneity between constructs. In a similar fashion the consistency of A with L and W and the consistency of W with L and A can be defined.

In Table 8.4 are the correlations between six subtests of the Iowa Test of Basic Skills for 1501 inner-city children (Crano, Kenny, & Campbell, 1972) The six tests can be thought of as two indicators of language usage (capitalization, X_1 , and punctuation, X_2), work skills (map reading, X_3 , and use of graphs, X_4), and arithmetic (concepts, X_5 , and problems, X_6).

Table 8.4. Correlation between Six Achievement Tests below the Diagonal and Residuals above^{a,b}

X ₁	1.0000	-.0000	-.0060	.0030	-.0130	.0200
X ₂	.6108	1.0000	.0130	-.0090	.0080	-.0150
X ₃	.4591	.4373	1.0000	.0000	.0040	-.0090
X ₄	.4679	.4164	.4384	1.0000	.0030	-.0000
X ₅	.4816	.4597	.4701	.4694	1.0000	-.0000
X ₆	.4782	.4030	.4225	.4316	.5736	1.0000
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆

^aN = 1501.

^bData taken from Crano, Kenny, & Campbell (1972).

In Figure 8.6 is a path model and the parameter estimates from the Jöreskog program. The work skill measures have the lowest loadings. Not surprisingly the correlations among the factors are very high. The $\chi^2(6)$ goodness of fit test is 9.98, which is not significant at conventional levels of significance. There are six degrees of freedom since the number of correlations is 15 (6 x 5/2) and there are 9 parameters: 6 epistemic correlations, and 3 correlations between factors. The residuals from the model are above the diagonal in Table 8.4.

It is impossible to test for homogeneity within a construct since there are only two indicators per construct. However, a test for homogeneity between each pair of constructs is possible. The test of

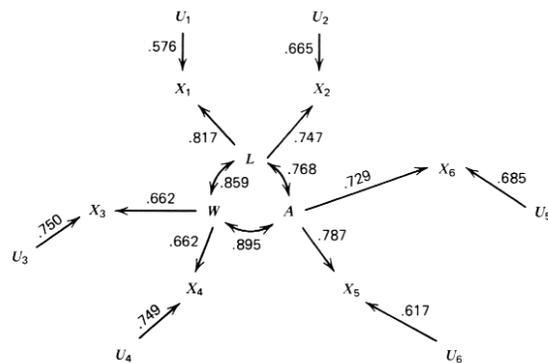


Figure 8.6 Multiple indicator model for the correlations in Table 8.4

homogeneity between language usage and work skills is

$$\rho_{13}\rho_{24} - \rho_{14}\rho_{23} = 0$$

between language usage and arithmetic

$$\rho_{15}\rho_{26} - \rho_{16}\rho_{25} = 0$$

and between work skills and arithmetic

$$\rho_{35}\rho_{46} - \rho_{36}\rho_{45} = 0$$

The χ^2 tests of the second canonical correlation for each tetrad are 1.58, 7.06, and .17, respectively. Obviously there is not homogeneity between language usage and arithmetic whereas there is homogeneity between the other two pairs. To test for consistency we compute all the tetrads of the form

$$\rho_{ij}\rho_{kl} - \rho_{ik}\rho_{jl} = 0$$

where *i* and *l* are the indicators of one construct, *j* an indicator of a second construct, and *k* an indicator of a third construct. There are a total of 12 such tetrads. Using .05/12 as the alpha level, all the 12 such tetrads vanish.

MULTITRAIT-MULTIMETHOD MATRIX

Campbell and Fiske (1959) suggested the multitrait multimethod matrix as a means for assessing validity. This matrix is a correlation matrix between the traits or constructs, each of which is measured by the same set of methods. In a sense, the measures are formed by factorially combining traits and methods. The resulting correlations can be characterized as same-trait, different-method correlations, different-trait, same-method correlations, and different-

trait, different-method correlations. According to Campbell and Fiske (1959) convergent validity is indicated if the same-trait, different-method correlations are large. Discriminant validity is indicated if those same correlations are larger than the different-trait, different-method correlations. Method variance is indicated if the different-trait, same-method correlations are larger than the different-trait, different-method correlations.

These rules of thumb are useful as far as they go, but ideally the matrix should be analyzed more formally. Numerous procedures have been put forward to factor analyze the matrix, two of which are considered below. There is no single way to specify a factor model.

The first is called the classical test theory formalization. The traits are factors whereas the disturbances or unique factors are allowed to be correlated across measures using the same method. Such a model is identified if there are at least two traits and three methods. Assuming the model fits the data, then convergent validation is assessed by high loadings on the trait factors, discriminant validation by low to moderate correlation between the trait factors, and method variance by highly correlated disturbances.

The second formalization was originally suggested by Jöreskog (1971) and has been extensively reviewed by Alwin (1974). Here each method and trait are taken to be factors and the factors may be intercorrelated. Consequently, each variable loads on both a trait and method factor. Discriminant and convergent validation are assessed as with the previous formalization, but method variance is now indicated by the loadings on the method factors. For such a model to be identified, at least three traits and three methods must be measured.

This second formalization should, in general, be preferred over the first. The first presumes that method variance is uncorrelated with both traits and the other methods. The second procedure does, however, assume that the effect of a method is single-factored whereas the first procedure does allow for multidimensional effects.

ILLUSTRATIONS

Jaccard, Weber, and Lundmark (1975) measured two constructs, attitudes toward cigarette smoking (C) and capital punishment (P), by four different methods: semantic differential (method 1), Likert (method 2), Thurstone (method 3), and Guilford (method 4) methods.

Table 8.5. Correlations of Four Attitude Measures with Residuals on and above the Diagonal^{a,b}

C_1	.001	.019	-.006	-.008	.059	.068	.057	.056
C_2	.780	-.010	-.009	-.023	.046	.047	.116	.036
C_3	.810	.770	.005	.024	-.007	-.037	-.002	.011
C_4	.760	.710	.810	.004	-.085	-.114	-.116	-.101
P_1	.290	.230	.190	.100	.021	.011	.016	.004
P_2	.280	.290	.180	.090	.840	.016	.014	-.012
P_3	.260	.310	.240	.080	.810	.890	-.006	-.033
P_4	.270	.240	.230	.150	.840	.910	.850	-.037
	C_1	C_2	C_3	C_4	P_1	P_2	P_3	P_4

^a $N = 35$.

^bData taken from Jaccard et al. (1975).

The sample size was only 35. In Table 8.5 are the correlations between the eight variables. Convergent validation is indicated because the measures of the same construct correlate highly. Discriminant validation is also indicated because the cross-construct correlations are all smaller than the within-construct correlations. There is some indication of systematic method variance since the correlations between different constructs using the same method are larger than the correlations between measures of different constructs and different methods.

Consider the path diagram in Figure 8.7 for the Jaccard et al. study. In this figure there are two types of variables, measured and unmeasured. The two traits, C = cigarette smoking and P = capital punishment, are the common factors and their four measures load on each. None of the cigarette smoking measures load on the capital punishment factor or vice versa. The factors designated E_1 through E_8

are the unique factors (measurement errors) and one measure loads on each. These unique factors are assumed to be uncorrelated with the trait factors and with the other unique factors with the important exception that unique factors may be correlated across measures using the same method. This exception allows for the method effects. The model in Figure 8.7 is sometimes called the *correlated uniqueness model*.

The minimum condition of identifiability is met because the number of correlations, 28, exceeds the number of parameters. There are 13 parameters, 8 factor loadings, and 5 factor correlations. The difference is 15, which is the total number of independent overidentifying restrictions on the model in Figure 8.7.

The values in Figure 8.7 are the parameter estimates. Note that the loadings on each trait factor are high and fairly homogeneous. There is a small positive correlation between the cigarette smoking and advocacy of capital punishment factors. There are also only small, positive correlations between disturbances.

The $\chi^2(15)$ test of goodness of fit yields a value of 10.35. Because the χ^2 is not statistically significant, the parameters of the model fit the correlations. Given the low sample size, most certainly other models fit the data.

Discriminant validation implies that the correlation between traits is low. If both traits were identical, the correlation between trait factors would be near one. One can force a structural equation modeling program to make the correlation between traits to be one. (Making the correlation between factors one is equivalent to a single-factor model.) The resultant $\chi^2(16)$ for that model equals 120.75. To test for discriminant validation simply subtract this χ^2 with the earlier one and subtract their degrees of freedom to obtain a $\chi^2(1)$ of 110.40. Since this value is highly significant, the assumption that the trait correlation is unity is not confirmed and discriminant validation is indicated.

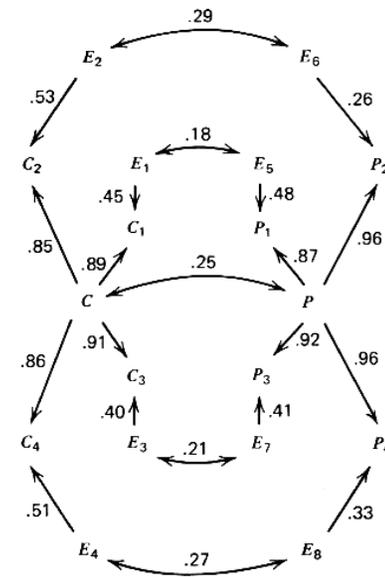


Figure 8.7 Multiple indicator model for correlations in Table 8.5.

In a similar fashion one can test for the equality of the factor loadings and factor correlations. To test whether the measures of C have significantly different loadings, constrain them to be equal. The χ^2 test of such a model is 10.71 with 18 degrees of freedom. Three degrees of freedom are added over 15 since only one factor loading need be estimated, not four. Again subtracting χ^2 values, the χ^2 test of equal loadings on C is .36 with 3 degrees of freedom. A similar result is obtained for the measures of P: $\chi^2(3) = 1.17$. Thus, there is no significant difference between methods in tapping the underlying traits. A similar test of the equality of correlated errors yields another nonsignificant χ^2 of .16 with 3 degrees of freedom.

Table 8.6. Television Style Scales, Correlations below Diagonal and Residuals above^{a,b}

P_1	1.000	.007	-.017	.025	.023	-.006	-.005	-.031	.001
C_1	.640	1.000	.007	.005	-.012	-.007	.003	-.012	-.009
I_1	.490	.550	1.000	-.012	.020	-.044	-.005	.030	.016
P_2	.540	.440	.250	1.000	-.003	.002	-.013	.004	.007
C_2	.330	.350	.280	.640	1.000	.001	.008	.024	.012
I_2	.240	.280	.260	.540	.700	1.000	-.036	-.032	-.023
P_3	.680	.570	.430	.520	.330	.220	1.000	.008	.001
C_3	.520	.610	.510	.440	.430	.300	.670	1.000	-.001
I_3	.430	.460	.660	.250	.310	.350	.530	.690	1.000
	P_1	C_1	I_1	P_2	C_2	I_2	P_3	C_3	I_3

^a $N = 382$.

^bData taken from Rosengren et al. (1976).

The second example of a structural analysis of the matrix is taken from Rosengren, Windahl, Hakansson, and Johnsson-Smargdi (1976). They investigated three different scales designed to measure style of viewing television: parasocial viewing (P), capture (C), and identification (I). Three hundred and eighty-six Swedish adolescents completed the three scales for television in general (method 1), a parasocial show (method 2), and an identification show (method 3). The resultant nine scales were intercorrelated, and are presented in Table 8.6. Since there are three traits and three methods, each measure can load on each trait and each method factor. Such a solution was attempted but it yielded an anomalous result. The $\chi^2(3)$ is less than one and the largest residual is only .009. The fit is excellent but the parameter estimates are odd to say the least. The loadings on the identification factor are less than .2 for two of the scales and one is even negative. The correlation of this factor with two other factors is larger than one. A loading on another factor is larger than one, an implausible if not impossible value. It was judged that the loadings on this factor are empirically underidentified. Moreover, it seems the data are overfit, that is, that too many parameters are estimated. Evidence for overfit is that χ^2 is smaller than its degrees of freedom and the anomalous estimates.

Because the identification scale has only one variable with a large loading, the factor cannot be distinguished from the disturbance and the loadings are empirically underidentified. It was decided the

Table 8.7. Loading and Correlation Matrix for the Correlation Matrix in Table 8.6^a

	P	C	1	2	3
P_1	.574		.593		
C_1		.518	.844		
I_1			.803		
P_2	.531			.630	
C_2		.241		.869	
I_2				.822	
P_3	.683				.601
C_3		.627			.970
I_3					.972
P	1.000				
C	.673	1.000			
1	.066	-.324	1.000		
2	.046	-.079	.461	1.000	
3	-.083	-.412	.825	.467	1.000
	P	C	1	2	3

^aNote the I factor is omitted because of empirical underidentification.

identification factor should be dropped. This left two trait factors and three method factors. The resultant solution did not fit nearly as well: $\chi^2(11) = 19.28$. The residuals presented above the diagonal in Table 8.6 are not large, however. The loadings and factor correlations in Table 8.7 indicate more method variance than trait variance and a rather high correlation between the two traits. (No path diagram is presented since it would be too cluttered to be interpretable.) Employing the criteria offered by Campbell and Fiske, Rosengren et al. arrived at very different conclusions.

Kenny and Kashy (1992) discuss estimation difficulties for the multitrait-multimethod matrix. In fact, they recommend estimating the correlated uniqueness model.

In the next chapter the correlations among the factors are analyzed. It is seen that one can impose on unmeasured variables the same type of models that are imposed on measured variables in Chapters 4, 5, and 6.

9

CAUSAL MODELS
WITH UNMEASURED
VARIABLES

The preceding two chapters have exclusively discussed the question of multiple indicators of latent constructs. What has been developed is called the *measurement model*. This chapter is devoted to causal models in which the latent variables are endogenous and exogenous; that is, there is a causal model for the unmeasured variables. In the previous chapter it was shown how correlations between unmeasured variables can be estimated. Here these correlations are analyzed and this combined model is called a hybrid model (Kline, 1998). The chapter is divided into two major sections. The first section elaborates the issues for structural models with latent endogenous and exogenous variables. The second and larger section works through five examples that were chosen to illustrate the broad range of models amenable to this type of analysis.

GENERAL CONSIDERATIONS

MEASUREMENT MODEL

It is helpful to employ matrices to summarize the measurement and causal models. Unfortunately the term matrix strikes fear in the heart of too many researchers. The purpose of matrices here is just the same as that of a path diagram. A matrix conveniently summarizes a set of structural equations. For the measurement model there is a

matrix of measures by constructs. The row variables are all the measured variables while the column variables are the unmeasured variables. A nonzero entry in the matrix denotes that the measured variable is caused by the construct or, to put it differently, the measure is an indicator of the construct. Normally there is the specification that certain entries are zero. If this matrix seems to be similar to a factor loading matrix, it is no accident since the constructs are factors.

It is helpful to make a distinction between a measure that loads on a single factor and one that loads on two or more factors. Strictly speaking a measure indicates a construct only if the measure has only one nonzero entry, that is, loads on only one factor. In contrast, an example of a measure with multiple entries can be taken from the Jöreskog (1971) model for the multitrait-multimethod matrix discussed in the previous chapter. There each measure is assumed to be caused by both a method and a trait factor.

Normally each of the measures is not totally determined by the latent constructs or common factors. A measure has a disturbance or a unique factor. In some special cases a latent factor is defined by a measure which means the construct has only a single indicator and that measure has no disturbance. For instance, if sex were one of the measured variables, it would have no disturbance and the construct on which sex loads is then defined as sex. In the main, however, a disturbance is needed for each measure.

These disturbances are usually assumed to be uncorrelated with the constructs. They need not, however, be uncorrelated with each other. It may be that measurement errors are correlated across measures, and a covariance matrix for the disturbances is then needed. The diagonal values of that matrix would be the variances and the off-diagonal values would be the covariances.

The disturbances can be viewed as measurement errors given a special definition of measurement error. Classical definitions of error usually center on the notion of unreproducibility. If a component of a score cannot be replicated by an alternative measure or by the same measure with a time delay, then the component is considered

error. But as Cronbach et al. (1972) state, the notion of replicability depends on the context or, as they would say, the type of generalization to be made. Reliability and measurement error then do not reside in the measure, but in the goals of the research. It may be that part of a measure's disturbance is systematic, replicable, and stable. Cook and Campbell (1976) have argued that such components should be called irrelevancies and need to be considered as errorful even if they are replicable.

Perhaps the simplest way to refer to the disturbances is that they are unique factors. A unique factor contains both specific and error variance. The important point to remember is that the specific factor is specific relative to the measures included in the study. As an example consider a measure that loads on two factors, A and B , which are uncorrelated. If the study includes only measures that are indicators of A , then the B component of the measure will be considered specific. If only indicators of B are included, then the A component will be specific.

STRUCTURAL MODEL

The latent constructs are ordinarily correlated. Perhaps the most serious limitation of factor analysis, as it is customarily practiced, is that the solution is usually constrained to be orthogonal. One may, however, be primarily interested in the correlations between the factors. The traditional factor analysis solution to the analysis of the correlations between factors is to perform a factor analysis on the correlation matrix between factors. The resulting factors are called second-order factors. However, since the factors are constructs, albeit unmeasured, any type of structural analysis can be performed on the correlation (covariance) matrix. Models discussed in all the previous chapters are then possible, as well as models discussed in the subsequent chapters. For instance, an example discussed later in this chapter has two latent variables involved in a feedback relationship.

Table 9.1. Structural Model: The F Variables Are Endogenous and the G Variables Are Exogenous

	G_1	G_2	...	G_m	F_1	F_2	F_3	...	F_n
F_1									
F_2									
F_3									
⋮									
F_n									

In general, the matrix that summarizes the structural model has the following form. As in Table 9.1, the rows of the matrix are the n constructs that are endogenous and the column includes the complete set of constructs of the measurement model, that is, m exogenous variables and n endogenous variables. Entries into the matrix denote whether a particular construct causes the endogenous variable, and the columns are the causes and the rows the effects. Each row of the matrix then summarizes the structural equation for each unmeasured endogenous construct.

The columns of the matrix are ordered such that the first set of variables includes only the constructs that are exogenous. The whole matrix can be divided into two submatrices. The first is a rectangular matrix of endogenous variables by the exogenous variables. The second matrix is square, and it is the endogenous variables by the endogenous variables. This partitioning of the matrix is illustrated in Table 9.1. The $n \times m$ matrix of endogenous by exogenous variables represents the causal effects of the exogenous variables.

Now consider $n \times n$ matrix as in Table 9.1. The diagonal values are always set to one and simply represent that a variable is in its own equation. One special type of $n \times n$ matrix is called lower-triangular. A lower-triangular matrix contains zero entries above the descending diagonal. Below that diagonal the entries may

be nonzero. If the $n \times n$ matrix can be arranged by shifting rows and columns to be lower-triangular, then the structural model is hierarchical; that is, it involves no feedback. If the $n \times n$ matrix cannot be arranged to make it lower-triangular, then the model is nonhierarchical.

Also to be considered are the two final parts of this very general formulation. First, the exogenous variables may be correlated and so one must consider the correlation (covariance) matrix for these variables. Second, the disturbances of the n endogenous variables may be correlated with each other or with the exogenous variables.

The preceding general formulation can represent every previously discussed model in this text. For instance, the models discussed in Chapters 4 and 6 have a very elementary measurement model. Each measure is a perfect indicator for each construct. Put into more familiar language, each measure is the operational definition for the construct. The matrix for the measurement model would then be a square matrix with ones in the diagonal, that is, an identity matrix. Since each construct is defined by the measure, there are no disturbances for the measurement model. The matrix for the structural model would be lower-triangular for the hierarchical structural models in Chapters 4 and 5 and would not be lower-triangular for the models discussed in Chapter 6.

For the model in Chapter 5 in which an exogenous variable correlates with the disturbance, the measurement model is an identity matrix, and the structural model implies a lower-triangular matrix. For the exogenous variables measured with error in Chapter 5, the errorful measures need a disturbance. The matrix for the measurement model remains an identity matrix if the classical model of measurement error is assumed.

In Chapter 7 there is only a measurement model and its matrix is a column vector (a string of numbers). For the models in Chapter 8 there is a multicolumn measurement model. Generally each row of the matrix has the following restriction: One value is nonzero and the others are zero. Again there is no structural model.

IDENTIFICATION

For the structural model to be identified the correlations (covariances) among the factors must be identified. Assuming that such correlations are identified, the researcher then proceeds to use the rules established in Chapters 4, 5, and 6. For instance, if the researcher can establish that the model among the latent variables is hierarchical and that none of the exogenous variables are correlated with disturbances, multiple regression can be used to estimate the model, assuming the correlations among the latent variables can be estimated. Although estimation is more extensively discussed in the next section, it is important to note here that any of the estimation procedures discussed in previous chapters can be applied to virtually any correlation (covariance) matrix. The correlations can be estimated directly from sample data, estimated indirectly as by factor analysis, or simply postulated by theory. However, if the correlations are indirectly estimated or postulated, one should not employ the standard tests of significance. For instance, it is totally illegitimate to test the correlation between two unmeasured variables by the standard t -test of a correlation.

Assuming the correlations among the factors are identified, the status of identification of the structural model is straightforward. Overidentification of the measurement model can in no way aid in the identification of the structural model. Thus, adding more indicators cannot facilitate the estimation of the structural model except in the sense that it may identify the factor correlations. Recall that the identification of the structural model presumes that the factor correlations are identified.

The identification of the measurement model was discussed in the previous chapter. Assuming that loadings and factor correlations are moderate, only two indicators per construct are needed. Three indicators are safer, and four safer still. In some very special cases when the structural model is overidentified, only one indicator of a construct is necessary. Instrumental variable estimation discussed in

Chapter 5 is one such an example. When one exogenous variable is known to have a zero path, then such a restriction can buy the identification of a path from another exogenous variable to its indicator. Thus, an over-identified structural model can in some very special cases aid in the identification of the measurement model. More typically the over-identification of the structural model does not help in the identification of the measurement model. If the model is overidentified overall, it is usually helpful to be able to test for specification error in the various components of the model. One reasonable strategy is to fit a just-identified structural model. Any lack of fit of the model to the correlations can be attributed to specification error in the measurement model. The measurement model can be adjusted to achieve more satisfactory fit, and then the overidentified structural model can be fitted. Any significant increase in lack of fit can then be attributed to specification error in the structural model.

ESTIMATION

Quite clearly the estimation task is quite complex. The measurement model requires a factor analysis type model and the structural model is typically a multiple regression type model. None of the previously discussed methods can handle such a general model. Some special models can be estimated by a two-step procedure. If the structural model is just-identified, the first step is to estimate the measurement model by confirmatory factor analysis. From such an analysis one would obtain the correlations (covariances) among the factors. The second step is to input these correlations to a multiple regression or two-stage least squares program for the estimation of the structural model. As mentioned earlier, one cannot use the standard test of significance for these estimates.

Instead of the two-step procedure, a structural equation modeling program (e.g., LISREL; Jöreskog & Sörbom, 1976) can be

used to estimate the parameters of such models by maximum likelihood. The fit of the model can be evaluated by a goodness of fit test. With the computer program LISREL eight different matrices are set up:

1. The loading matrix for the indicators of the endogenous variables.
2. The loading matrix for the indicators of the exogenous variables.
3. The paths from latent endogenous variables to latent endogenous variables.
4. The paths from latent exogenous variables to latent endogenous variables.
5. The covariance matrix for the latent exogenous variables.
6. The covariance matrix for the disturbances of the latent endogenous variables.
7. The covariance matrix for the disturbances of indicators of the latent endogenous variables.
8. The covariance matrix for the disturbances of indicators of the latent exogenous variables.

In most structural equation modeling programs the variances of the latent endogenous variables cannot be fixed and, therefore, cannot be standardized. Thus, to identify the model, either one loading on each latent endogenous variable must be fixed to some nonzero value, or the variance of the disturbance of the latent variable must be fixed to some nonzero value. The former strategy is to be preferred if results are to be compared across different populations or occasions. Programs do take the solution and then standardize the latent variables to produce a standardized.

Structural equation modeling programs such as LISREL are not an easy programs to learn how to use. These programs do, however, provide the most complete solution to the estimation problem of structural models.

EXAMPLES

THERE CONSTRUCTS, EACH WITH TWO INDICATORS

In the previous chapter a model was discussed from Crano et al. (1972). In Figure 9.1 there is added a structural model in which language usage causes work skills which causes arithmetic, each factor having two indicators. Language usage is denoted as L , work skills as W , and arithmetic as A .

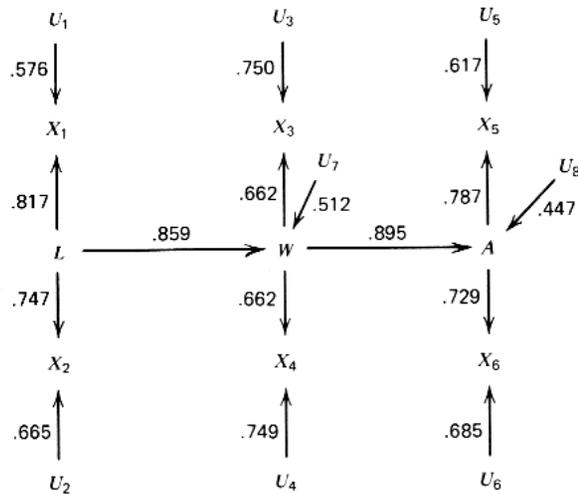


Figure 9.1 Structural model for Crano, Kenny, and Campbell example.

In Table 9.2 there are the two matrices that summarize both the measurement and structural models. The first matrix denotes which construct causes each measure. An "X" indicates the measure loads on the factor and "O" denotes it does not. Table 9.2 should not be confused with any variant of the game tic-tac-toe.

Table 9.2. Measurement and Structural Models for Figure 9.1

Measurement Model			
	L	W	A
X_1	X	O	O
X_2	X	O	O
X_3	O	X	O
X_4	O	X	O
X_5	O	O	X
X_6	O	O	X

Structural Model			
	L	W	A
L	O	O	O
W	X	O	O
A	O	X	O

The rows of the top matrix in Table 9.2 are the six measures and the columns are the three latent constructs for the model in Figure 9.1. Note that each latent construct causes two different measures. The structural model indicated by the bottom matrix in Table 9.2 is hierarchical since the matrix for it is lower-triangular; that is, all the X's are below the main descending diagonal. The measurement model can be estimated as in the previous chapter. It is identified since each construct has two indicators. In Chapter 8, it was noted that the overall fit of the model is good, but there is some problem in the homogeneity between language usage and arithmetic. The measurement model also yields the following estimates of correlations between factors: $r_{LW} = .859$, $r_{LA} = .768$, and $r_{WA} = .895$ (see Figure 8.6). These correlations are the major ingredients in the estimation of the structural model.

Because the structural model is hierarchical, one can use multiple regression to estimate the path coefficients. The path from L to W is simply p_{LW} or .859. If paths from both L and W to A are specified, one can solve for them by using the multiple regression formulas 4.5 and 4.6:

$$p_{AL} = \frac{r_{AL} - r_{AW}r_{LW}}{1 - r_{LW}^2} = -0.003$$

$$p_{AW} = \frac{r_{AW} - r_{AL}r_{LW}}{1 - r_{LW}^2} = 0.898$$

To find the coefficient for the disturbances solve for both

$$(1 - r_{LW}^2)^{1/2} = .512$$

$$(1 - r_{AW}p_{AW} - r_{AL}p_{AL})^{1/2} = .446$$

which are the disturbance paths for W and A , respectively. No special problems arise as a result of having a structural model with unobserved variables, once the correlations (covariances) between the unobserved variables are known. Special considerations do arise if the researcher wishes to estimate a structural model that is overidentified, as in Figure 9.1. Such a model makes the correlations among the latent variables overidentified by the structural model. The structural equation modeling programs were developed to estimate these models.

Consider the model in Figure 9.1. It is like the model in Figure 8.6, but in this case there is an overidentified structural model for the latent variables. The estimates are in Figure 9.1. The $\chi^2(7) = 9.98$ for the model. If a path is added from L to A , $\chi^2(6) = 9.98$. Thus the test of the measurement model is $\chi^2(6) = 9.98$ and the structural model is $\chi^2(1) = .00$. Remember the test of the measurement model is to fit a just-identified structural model. Lack of fit then is due only to specification error in the measurement model. To test an overidentified structural model, one fits that model. Then any increased lack of fit over the just-identified structural model is due to specification error in the structural model.

What is unusual about the structural model in Figure 9.1 is that the path from L to A is virtually zero. Thus the fit of the overidentified model and the just-identified model do not differ and the other

parameters are virtually unchanged. Such a result is highly unusual. However, the zero path can be compared against other more standard estimation procedures. If one simply regresses one indicator of A on one indicator of L and one of W , one would obtain on the average

$$p_{AL} = .320 \quad p_{AW} = .305$$

a result that is radically different from the multiple indicator results. If one unit weights the two indicators of each construct and then regresses the sum of the two indicators of A on the sums for L and W , one would obtain

$$p_{AL} = .298 \quad p_{AW} = .469$$

At least for this example an unmeasured variable approach yields a conclusion that is very different from multiple regression.

EXAMPLE WITH FOUR CONSTRUCTS, TWO MEASURED AND TWO UNMEASURED

This example is taken from the Westinghouse Learning Corporation evaluation of Head Start. As some readers no doubt know, there is a great deal of controversy that centers around this work (see Campbell & Erlebacher, 1970). Magidson (1977) has undertaken a reanalysis of these data and the example is taken from his analysis, though the analysis following differs from his in many respects. Magidson's sample consists of 303 white 6-year-old first graders who attended summer Head Start. He chose this group because the original analysis obtained the strongest negative effect for this group. The group consists of 148 children who received Head Start and 155 control children.

For the following analysis six measured variables whose correlations are in Table 9.3 are considered. The variables are mother's education (X_1), father's education (X_2), father's occupation (X_3), income

(X_4), and a dummy for Head Start (H), and the outcome variable (Y). The Head Start dummy variable is coded one for those who attended Head Start and zero for those who did not. Note that the correlation between H and Y is negative, indicating that those who completed Head Start had lower test scores.

Table 9.3. Head Start Correlations with Residuals above the Main Diagonal^{a,b}

X_1	1.000	.000	-.045	.041	.009	.012
X_2	.468	1.000	.032	-.017	.012	-.017
X_3	.241	.285	1.000	.000	-.008	.017
X_4	.297	.209	.407	1.000	.010	-.022
H	-.118	-.084	-.220	-.179	1.000	-.000
Y	.275	.215	.255	.190	-.094	1.000
	X_1	X_2	X_3	X_4	H	Y

^aData taken from Magidson (1977).

^b $N = 303$.

The proposed causal model is contained in Figure 9.2. For the moment ignore the parameter estimates. The model postulates two latent variables. The first is labeled E in Figure 9.2 and it can be viewed as an educational background factor. The second factor relates to socioeconomic outcomes and is labeled S . Obviously the model of latent factors in Figure 9.2 is unrealistic since the measured variables have themselves a structural relationship. For instance, no doubt father's education causes father's occupation. However, it may be plausible to view the child's cognitive skills as a function of the two proposed factors. The educational background factor would directly cause cognitive gain through imitation and the socioeconomic outcome factor would allow the parents to purchase goods that would increase the child's cognitive skills.

The measurement model for the Head Start example is straightforward. There are four constructs, E , S , H , and Y . The constructs E and S each have two indicators, and H and Y have only a single indicator that defines the construct. There is only a single equation

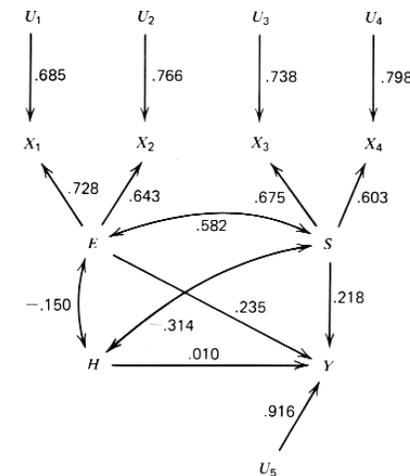


Figure 9.2 Head Start model.

for the structural model. The equation is that E , S , and H cause Y . Note that the model in Figure 9.2 can be viewed as a simple three-factor model: E , S , and H . The variable Y simply loads on all three factors. The analysis then is rather straightforward: Simply estimate this three-factor model. The first factor, E , has three "indicators" X_1 , X_2 , and Y ; the second, S , also has three "indicators" X_3 , X_4 , and Y ; and the third, H , has two "indicators" Y and H . The variable H 's loading on factor H is fixed to one. The solution is presented in Figure 9.2. Not surprisingly there is a high correlation between E and S . Both E and S correlate negatively with H but the socioeconomic outcomes factor correlates more negatively than the educational factor. Just such a difference is to be expected since eligibility into Head Start depended on socioeconomic outcomes. The pattern is reversed for the effects of the factors on cognitive skills. The educational factor has a slightly larger effect than the socioeconomic factor. Head Start has a slightly positive effect which is clearly not statistically significant. However, the effect is not negative as would be obtained had a simple regression analysis been performed with only measured X_1 , X_2 , X_3 , and X_4 partialled out. Thus, the structural analysis wipes out the negative effect.

There are a total of 10 free parameters for the Head Start model. Because there are a total of 15 correlations, the number of overidentifying restrictions is five. The $\chi^2(5)$ equals 6.13, which is clearly not statistically significant. In Table 9.3 are the residuals for the model and they are not large. The largest residual occurs in a place where a specification error surely occurs. It is between mother's education (X_1) and father's occupation (X_3) and is negative.

The five overidentifying restrictions of the model are totally due to the measurement model. Recall that the structural model is just-identified. Using the vocabulary of the previous chapter the five overidentifying restrictions can be partitioned as follows:

1. The homogeneity between E and S .
2. The consistency of E for H .
3. The consistency of S for H .
4. The consistency of E for Y .
5. The consistency of S for Y .

As an exercise show that the following 11 vanishing tetrads hold for the model in Figure 9.2 and show that only five are independent:

$$\begin{aligned} \rho_{13}\rho_{24} - \rho_{14}\rho_{23} &= 0 \\ \rho_{13}\rho_{2H} - \rho_{1H}\rho_{23} &= 0 \\ \rho_{14}\rho_{2H} - \rho_{1H}\rho_{24} &= 0 \\ \rho_{13}\rho_{2Y} - \rho_{1Y}\rho_{23} &= 0 \\ \rho_{14}\rho_{2Y} - \rho_{1Y}\rho_{24} &= 0 \\ \rho_{1Y}\rho_{2H} - \rho_{1H}\rho_{2Y} &= 0 \\ \rho_{23}\rho_{4H} - \rho_{2H}\rho_{34} &= 0 \\ \rho_{13}\rho_{4H} - \rho_{1H}\rho_{34} &= 0 \\ \rho_{13}\rho_{4Y} - \rho_{1Y}\rho_{34} &= 0 \\ \rho_{23}\rho_{4Y} - \rho_{2Y}\rho_{34} &= 0 \\ \rho_{3Y}\rho_{4H} - \rho_{3H}\rho_{4Y} &= 0 \end{aligned}$$

NONHIERARCHICAL RELATIONSHIP BETWEEN LATENT VARIABLES

The preceding models that have been considered have had relatively simple structural models. The model to be considered now contains almost every imaginable complexity. It contains, among other things:

1. Unmeasured endogenous variables with multiple indicators.
2. A nonhierarchical (i.e., feedback) relationship between latent variables.
3. An overidentified structural model.
4. Equality constraints for causal parameters.

The data are taken from Duncan, Haller, and Portes (1971), which has become a classic in the causal modeling literature. The substantive question of their paper concerns itself with the effect of peers on aspirations. The measured variables are parental encouragement of the child (X_1), child's intelligence (X_2), socioeconomic status of the parents (X_3) and the educational and occupational aspirations of the self (Y_1 and Y_2). The same variables were measured for the person named by the child as the best friend (ordered the same as for self: X_4 , X_5 , X_6 , Y_3 , and Y_4). There are then a total of 10 variables, 5 for the self and 5 for the friend. The sample consists of 329 adolescent males and the correlations are presented in Table 9.4. In Figure 9.3 there is a model for the data that Duncan, Haller, and Portes postulated with some modifications. First, Duncan et al. postulated that the disturbances in the aspiration measures of the self are correlated. For the sake of simplicity, it is assumed here that such correlations are zero. Second, the model is constrained to be the same for both the self and the friend. Duncan et al. give no compelling reason not to force such equality. They argue that the self named the other child as a friend while the friend did not necessarily make a

and for the structural model

$$G = aX_1 + bX_2 + cX_3 + dX_6 + gH + hU_1$$

$$H = aX_4 + bX_5 + cX_6 + dX_3 + gG + hU_2$$

In Table 9.5 is the measurement model in matrix form. Note that since the exogenous variables are measured, they each define their own factor and have loadings of one. Both G and H have two indicators. In Table 9.6 there is the structural model. The exogenous variables are X_1 through X_6 , and G and H are endogenous. The model is not hierarchical because G causes H and vice versa.

Table 9.5. Measurement for the Duncan, Haller, and Portes Example.

Measures	Constructs							
	X_1	X_2	X_3	X_4	X_5	X_6	G	H
X_1	1	0	0	0	0	0	0	0
X_2	0	1	0	0	0	0	0	0
X_3	0	0	1	0	0	0	0	0
X_4	0	0	0	1	0	0	0	0
X_5	0	0	0	0	1	0	0	0
X_6	0	0	0	0	0	1	0	0
Y_1	0	0	0	0	0	0	e	0
Y_2	0	0	0	0	0	0	0	f
Y_3	0	0	0	0	0	0	0	e
Y_4	0	0	0	0	0	0	0	f

A structural equation modeling program like LISREL can be used to estimate the parameters for the model in Figure 9.3. Equality constraints in most structural equation programs can only be placed on the unstandardized model. Recall that in most structural equation modeling programs the latent endogenous variables are not standardized. Any equality constraint does not necessarily hold for the standardized solution. It was, therefore, decided to fix the disturbance path h to one, and the solution is presented in Figure 9.3 without the correlations among the exogenous variables.

Table 9.6. Structural Model for the Duncan, Haller, and Portes Example

Endogenous Variables	Exogenous Variables							
	X_1	X_2	X_3	X_4	X_5	X_6	G	H
G	a	b	c	0	0	d	0	g
H	0	0	d	a	b	c	g	0

There are a total of 17 free parameters and 45 correlations. The 17 free parameters are

1. Nine correlations among the exogenous variables.
2. Four paths from the exogenous variables to the latent endogenous variables.
3. One path from the endogenous variable to the other.
4. The correlation between the disturbances of G and H .
5. Two paths from G and H to their two indicators.

There are then a total of 28 restrictions. The test of these restrictions yields a $\chi^2(28) = 35.00$ which is nonsignificant. These 28 restrictions can be divided in the following way. First, 13 of the restrictions are due to the assumption that the causal process is the same for friend and self.

There are 13 equality constraints because $a, b, c, d, g, e,$ and f are duplicated and the following correlations among the exogenous variables are assumed to be equal: $\rho_{15} = \rho_{24}, \rho_{16} = \rho_{34}, \rho_{26} = \rho_{35}, \rho_{12} = \rho_{45}, \rho_{13} = \rho_{46},$ and $\rho_{23} = \rho_{56}$. Thus the equality constraints account for 13 of the 28 overidentifying restrictions. There are 15 remaining constraints on the measurement and structural model. There are two constraints on the structural model since there are two zero paths from the exogenous variables to each endogenous variable. There is then one extra instrument for the equations for both G and H , creating two overidentifying restrictions. The remaining 13 restrictions are due to the measurement model. One of the 13 constraints is homogeneity

between G and H , and the remaining constraints are the consistency of the epistemic correlations of both G and H for X_1 through X_6 .

Table 9.7. Partition of χ^2 into Three Components of Specification Error

	χ^2	df
Equality constraints	8.17	13
Structural model	1.62	2
Measurement model	25.21	13
Total	35.00	28

In Table 9.7 the overall χ^2 has been partitioned into the three different types of restrictions. The three possible types of specification error are equality of parameters, the structural model, and the measurement model. To obtain the χ^2 to test the equality constraints, first a model is estimated in which the equality constraints are not forced to hold. The resulting $\chi^2(15)$ for this model equals 26.83 and subtracting this from the overall χ^2 yields a test of the equality constraints. To test the constraint of the structural model, paths are added from friend's intelligence to self's ambition and from self's intelligence to friend's ambition with no equality constraints. The χ^2 equals 25.21. Subtracting this χ^2 from the previous χ^2 yields $\chi^2(2) = 1.62$ which tests whether the path from X_2 to H and X_5 to G are zero given that the paths from X_1 to H and from X_4 to G are zero. (This turns out to be equivalent to testing whether the paths from X_4 to G and X_1 to H are zero given zero paths from X_2 to H and X_5 to G .) The χ^2 that remains then tests the measurement model. There does appear to be specification error in the measurement model. An examination of the residuals shows a large value for $r_{Y_1Y_3}$. The model should be respecified to allow for correlated disturbances.

The reader can now appreciate the wide range of model that can be estimated by a structural equation modeling program. If a model of the complexity of Figure 9.3 can be estimated, then no model should pose any problem. The researcher still must carefully study the model

and determine the status of identification of the measurement and structural model.

CAUSAL CHAIN, SINGLE-INDICATOR

Perhaps one of the most interesting models is the causal chain with each construct having only a single indicator. As shall be seen, some of the parameters are overidentified while others are underidentified.

In Figure 9.4 there is a model of a simple chain. As is typical of such models the chain is composed of a single variable measured at multiple points in time. In this case the variable is the number of citations of 240 chemists. The citations were counted by Hargens, Reskin, and Allison (1976) for four years, 1965 through 1968. In Table 9.8 are the correlations for the four waves of data. In Figure 9.4, X denotes the measured number of citations and T the "true" number of citations. Note T is caused only by the prior value of T . The technical name for such a process is first-order autoregressive. The correlational structure generated by a first-order autoregressive process is called a simplex (Humphreys, 1960). A simplex tends to have lower and lower correlations as one moves from the main descending diagonal. The mathematical definition of a simplex matrix is that the variables can be ordered such that

$$\rho_{ik} = \rho_{ij}\rho_{jk}$$

$$i < j < k$$

Stated in words, if one partials any variable that comes between two variables in the causal chain, the partial correlation is zero.

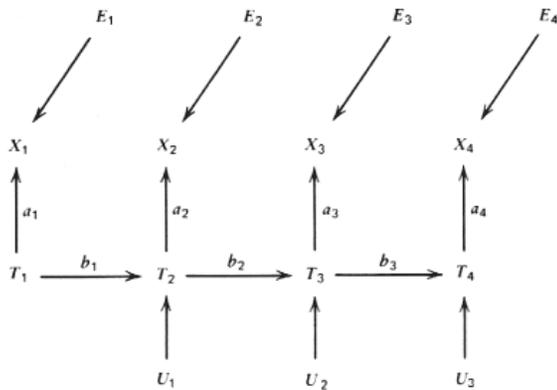


Figure 9.4 First-order autoregressive model with measurement error.

Only the true score is governed by a first-order autoregressive process and only its correlational structure is a simplex. The measured score cannot be described so simply. Actually the measured score is assumed to be the sum of two autoregressive processes: the true score and the errors of measurement. The errors are a degenerate case of a first-order autoregressive process since the regression of E on a prior value of E is zero. The correlational structure of the measured scores is called a quasi-simplex; that is, it would be a simplex if the scores were corrected for attenuation due to measurement error. A quasi-simplex differs from a pure simplex in that $\rho_{ik,j}$ does not equal zero and is usually positive.

Estimation in the quasi-simplex concerns two types of parameters. First, there are the paths from each construct to its measure that constitutes the measurement model. Second there is the estimation of the causal chain or the structural model. It has long been known that variants of principal components analysis do not even come close to estimating the parameters of either a simplex or its cousin the quasi-simplex. Diagonal factoring does estimate parameters of a simplex model, but requires prior knowledge of communalities to estimate quasi-simplex parameters. The approach

here is first to discuss path analytic solutions following Heise (1969) and Wiley and Wiley (1970) and, second, to elaborate a confirmatory factor analysis approach to estimation Jöreskog (1970; Werts, Jöreskog, & Linn, 1971).

Table 9.8. Correlation Matrix for the Chemist Example; Correlations below the Main Diagonal, Variances on the Diagonal, and Residuals to the Model with Equal Reliabilities above the Diagonal^{a,b}

X_1	2.590	.002	-.014	.018
X_2	.515	2.150	.005	-.006
X_3	.469	.613	2.730	-.002
X_4	.436	.520	.552	2.510
	X_1	X_2	X_3	X_4

^aData taken from Hargens, Reskin, and Allison (1976).

^b $N = 240$.

For the model in Figure 9.4 there are six correlations

$$\begin{aligned}
 r_{12} &= a_1 a_2 b_1 & r_{23} &= a_2 a_3 b_2 \\
 r_{13} &= a_1 a_3 b_1 b_2 & r_{24} &= a_2 a_4 b_2 b_3 \\
 r_{14} &= a_1 a_4 b_1 b_2 b_3 & r_{34} &= a_3 a_4 b_3
 \end{aligned}$$

Note that there are seven parameters and six correlations, and thus the model must be underidentified.

Even though the model is underidentified, the parameter estimates for a_2^2 , a_3^2 , and b_2^2 are overidentified:

$$a_2^2 = \frac{\rho_{12}\rho_{23}}{\rho_{13}} = \frac{\rho_{12}}{\rho_{14}}$$

$$a_3^2 = \frac{\rho_{13}\rho_{34}}{\rho_{14}} = \frac{\rho_{23}\rho_{34}}{\rho_{24}}$$

$$b_2^2 = \frac{\rho_{14}\rho_{23}}{\rho_{12}\rho_{34}} = \frac{\rho_{13}\rho_{24}}{\rho_{12}\rho_{34}}$$

Both a_2^2 and a_3^2 are reliabilities and so the positive root is taken. The sign of b_2 is the same as the sign of ρ_{23} . Because the preceding parameters are overidentified, one can solve for an overidentifying restriction. It is

$$\rho_{14}\rho_{23} - \rho_{13}\rho_{24} = 0$$

This is a vanishing tetrad which can be tested by checking whether the second canonical correlation is zero. The result of such a test yields $\chi^2(1) = .508$ which is not significant, indicating that the overidentifying restriction is met.

The estimates of a_2 , a_3 , and b_2 for the chemists' data are

$$\begin{aligned} a_2: & .820, .784 \\ a_3: & .807, .771 \\ b_2: & .970, .926 \end{aligned}$$

The two estimates of each parameter are roughly similar, which is not surprising since the overidentifying restriction holds. The "low" reliabilities are discussed later.

One is not able to estimate a_1 , a_4 , b_1 , and b_3 . One can, however, solve for the product terms, a_1b_1 and a_4b_3 after a_2 , a_3 , and b_2 have been estimated. They can be estimated by

$$a_1b_1 = \frac{r_{12}}{a_2} = \frac{r_{13}}{a_3b_2}$$

$$a_4b_3 = \frac{r_{34}}{a_3} = \frac{r_{24}}{a_2b_2}$$

Because all the parameters of the model are also correlations, they should be within +1 and -1. However, given sampling error, an

estimate may fall out of this range. Note that if the reliabilities or autoregressive parameters are zero or near zero the model is empirically underidentified.

In general given k waves of measurement, the reliabilities of the first and last wave are not identified. Also not identified is the path from the true score at the first wave to the true score at the second wave and the path from the true score at the next to the last wave to the true score at the last wave. All the other parameters are identified if k equals three, and overidentified if k is greater than three. All the overidentifying restrictions are of vanishing tetrad form:

$$\rho_{im}\rho_{jk} - \rho_{ik}\rho_{jm} = 0$$

$$i < j < k < m$$

Because the model is partially underidentified, various strategies have been put forward to bring about identification. Heise (1969) suggested that one might assume that $a_1 = a_2 = a_3 = a_4$. This implies that reliability does not change over time. This yields a just-identified model for the three-wave case and an additional overidentifying restriction for the four-wave case. Alternatively one might force what seem to be very ad hoc specifications of $a_1 = a_2$ and $a_3 = a_4$ or even more implausibly $a_1 = a_3$ and $a_2 = a_4$.

Wiley and Wiley (1970) have argued that it may be more plausible to assume that the error variance, $V(E)$, and not the reliability, is constant over time. Unlike the previous models this requires the use of the unstandardized metric and refers to the covariance matrix, not the correlation matrix. The measurement model is

$$X_i = T_i + E_i$$

where $V(E_1) = V(E_2) = V(E_3) = V(E_4)$. The measures X are said to be tau equivalent since their error variances are assumed to be equal. The structural model is

$$T_{i+1} = b_i T_i + V_i$$

The covariance matrix for the four-wave case is in Table 9.9. One can then estimate $V(E)$ by both

$$V(X_2) - \frac{C(X_1, X_2)C(X_2, X_3)}{C(X_1, X_3)}$$

$$V(X_3) - \frac{C(X_2, X_3)C(X_3, X_4)}{C(X_2, X_4)}$$

The estimates of $V(T_i)$ can now be computed by $V(X_i) - V(E)$. One can then proceed to estimate $b_1, b_2, b_3, b_1b_2, b_2b_3,$ and $b_1b_2b_3$. Note that what is analyzed is the covariance matrix not the correlation matrix.

The third solution to this underidentification problem has not been previously suggested in this context. It is, however, very popular in the time series literature. It assumes that the autoregressive process is stationary, that is, that $b_1 = b_2 = b_3$. Such an assumption brings about identification for the four-wave case but adds no further overidentifying restrictions. It does not bring about identification for the three-wave case as do assumptions about constant reliabilities or error variances.

Table 9.9. Theoretical Covariance Matrix with Tau Equivalent Tests and a First-Order Autoregressive Process

X_1	$V(T_1) + V(E)$			
X_2	$b_1V(T_1)$	$V(T_2) + V(E)$		
X_3	$b_1b_2V(T_1)$	$b_2V(T_2)$	$V(T_3) + V(E)$	
X_4	$b_1b_2b_3V(T_1)$	$b_2b_3V(T_2)$	$b_3V(T_3)$	$V(T_4) + V(E)$
	X_1	X_2	X_3	X_4

Table 9.10. Model Parameter Estimates for Autoregressive Models; All Parameters Are Standardized Except Error Variances and the Equal Error Variance Model

	Free	Equal Reliabilities	Equal Stabilities	Equal Error Variance
a_1	—	.802	.675	1.000
a_2	.807	.803	.807	1.000
a_3	.797	.800	.797	1.000
a_4	—	.802	.732	1.000
b_1	—	.796	.945	.680
b_2	.945	.946	.945	1.120
b_3	—	.864	.945	.803
$V(E_1)$	—	.925	1.410	.857
$V(E_2)$.748	.768	.748	.857
$V(E_3)$.996	.975	.996	.857
$V(E_4)$	—	.896	1.165	.857
χ^2	.549	.587	.549	1.553
df	1	2	1	2

All the preceding strategies have the particular problem that there is no clear way to pool the overidentified information. LISREL or any structural equation modeling program can estimate all the preceding models. Inputting the correlation matrix into a structural equation modeling program and forcing the equality of the paths from T_i to T_{i+1} does not result in equal reliabilities because in most structural equation modeling programs one cannot standardize in the latent endogenous variables of $T_2, T_3,$ or T_4 . To estimate a model with equal reliabilities then one forces the disturbance paths to the measures to be equal. In Table 9.10 are the estimates for the four models. All the estimates are standardized except the error variances and the estimates for the equal error variance model. For the first model there are no constraints; this makes $a_1, a_4, b_1, b_3, V(E_1),$ and $V(E_4)$ underidentified. The fit of the model is quite good. Assuming equal reliabilities does not result in a significant increase in lack of

fit. Note that either due to rounding error or small specification error, the a parameters are not exactly equal.

The assumption of equal stabilities, $b_1 = b_2 = b_3$, yields the result that the reliabilities of the first and last waves are lower than the two middle waves. Finally, the model with equal error variances also fits, although not quite as well as the model with equal reliabilities. Table 9.10 gives the unstandardized parameter estimates for this last model.

The reader is still probably puzzled by the low autocorrelation for true citations. Understanding the implicit assumptions of the autoregressive model can explain the low correlation. A test-retest correlation can be less than one for three very distinct reasons: (a) true change in causes, (b) measurement error, and (c) change in structure. The first two reasons are most certainly already known by the reader, while the third may not be. Assume that X is measured at two distinct points in time, say 1 and 2. The variable X_1 is defined as the sum of its causes, $\sum a_i Z_{i1}$ and X_2 similarly, $\sum b_i Z_{i2}$. Assume also that none of the causes change and they are all orthogonal. The test-retest correlation will only be one if $a_1/b_1 = a_2/b_2 = a_3/b_3 = \dots = a_n/b_n$. Thus if the causal structure changes over time, the test-retest correlation is not unity. There is then the implicit assumption that the structural equations do not change over time for the data to be first-order autoregressive. In Chapter 12 this assumption is called quasi-stationarity and is discussed in much more detail.

A second implicit assumption of an autoregressive process is homogeneous stability: All true causes change at the same rate; that is, they all have the same autocorrelation. If a measure is multidimensional and the different dimensions change at different rates, the process cannot ordinarily be described as first-order autoregressive. Moreover, all the causes that are totally unstable over time are defined as errors of measurement. Thus the reliability estimates that were obtained for the citations of chemists are not low. It is true that if someone else did the counting this new count should correlate very highly with the old count. Thus it is not the reliability of

the counts per se but the reliability of their ability to indicate some latent trait. The actual number of citations is not a perfectly valid indicator of "citability." See Hargens, Reskin, and Allison (1976) for an alternative explanation.

COMPOSITE CAUSES

In Figure 9.5 is a model that is very different from all others that have been considered in this text. The variable S is unmeasured and has no disturbance. It is assumed to be a composite or linear function of X_1, X_2, X_3 , and X_4 . The variable S is called a composite cause because a composite of X_1 through X_4 causes Y_1, Y_2 , and Y_3 . Note that S mediates the causal effect of the X variables on the Y set. Composite variables are more commonly called indices. For instance, one could index a child's aggressiveness by combining physical and verbal aggression. Population change is indexed by number of births minus the number of deaths plus the number of immigrants minus the number of emigrants.

The reader should compare the role of S in Figure 9.5 with the role of say L in Figure 9.1. They are both unmeasured, but once the coefficients leading into S are determined. The construct S is exactly known. However, the construct L can be estimated only with imperfect reliability. The X variables cause S in Figure 9.5, whereas

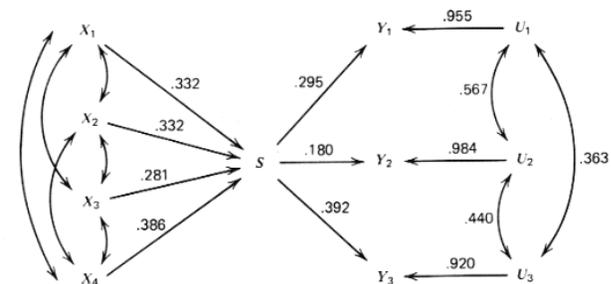


Figure 9.5 Composite cause model.

in Figure 9.1 they are caused by L . Finally it is assumed that S 's X variables are free of measurement error, whereas the indicators of L are unreliable. Clearly there is only a superficial similarity between S and L . It is shown later that there are differences in estimating the two models also.

The four components of S are father's education, mother's education, father's occupation, and average parental income making S a composite cause. The Y variables are as follows: mental ability, high school grades, and college plans. Data are taken from Hauser (1973), and the correlations are reproduced in Table 9.11.

To estimate the parameters of the model, first set up the following canonical correlation analysis. The X variables form one set and the Y variables form the second set. The paths from the X variables to S are given by the canonical variate coefficients for the X variables. The estimate of S obtained from such a canonical analysis is

$$.332X_1 + .332X_2 + .281X_3 + .386X_4 \quad [9.1]$$

To determine the path from S to Y_1 note that it equals r_{SY_1} . Thus for the path from S to Y_1 one computes the covariance between Y_1 and Equation 9.1 which is

$$\begin{aligned} p_{Y_1S} &= r_{Y_1S} \\ &= p_{X_1S}r_{X_1Y_1} + p_{X_2S}r_{X_2Y_1} + p_{X_3S}r_{X_3Y_1} + p_{X_4S}r_{X_4Y_1} \\ &= (.322)(.244) + (.322)(.230) + (.281)(.212) + (.386)(.203) \\ &= .295 \end{aligned}$$

The paths from S to Y_2 and Y_3 are similarly defined:

$$\begin{aligned} p_{Y_2S} &= (.332)(.151) + (.332)(.149) + (.281)(.127) + (.386)(.116) \\ &= .180 \\ p_{Y_3S} &= (.332)(.306) + (.332)(.269) + (.281)(.299) + (.386)(.304) \\ &= .392 \end{aligned}$$

The correlations among disturbances are simply partial correlations: $r_{Y_1Y_2.S}$, $r_{Y_1Y_3.S}$, and $r_{Y_2Y_3.S}$. The model is overidentified and the set of overidentifying restrictions are of the form

$$\rho_{X_iY_k}\rho_{X_jY_m} - \rho_{X_iY_m}\rho_{X_jY_k} = 0$$

Note that this set of restrictions is equivalent to assuming homogeneity between the X and Y set (see the previous chapter), which is why the second canonical correlation can be used to test for homogeneity between constructs. The degrees of freedom are $(n - 1)(m - 1)$ where n is the number of variables in set X and m the number in set Y . The $\chi^2(6) = 11.31$, which is not statistically significant at the .05 level. A structural equation modeling program produced the same parameter estimates and $\chi^2(6) = 11.33$.

Normally models for composite causes are much more complicated. The reader should consult Hauser (1973) and Hauser and Goldberger (1971) for more details. The reader should note that while canonical correlation analysis is useful in estimating causal models, the canonical correlation itself does not estimate a meaningful structural parameter.

Table 9.11. Correlations for Socioeconomic Variables and Outcomes^{a,b}

X_1	1.000						
X_2	.505	1.000					
X_3	.494	.318	1.000				
X_4	.389	.291	.523	1.000			
Y_1	.244	.230	.212	.203	1.000		
Y_2	.151	.149	.127	.116	.586	1.000	
Y_3	.306	.269	.299	.304	.435	.469	1.000
	X_1	X_2	X_3	X_4	Y_1	Y_2	Y_3

^aData taken from Hauser (1973).

^b $N = 3427$.

CONCLUSION

Five specific examples have been analyzed. Quite clearly, structural modeling with unobserved variables is possible. As with any model one must take care in the specification and identification of such models. Always check the estimates and residuals to determine if the results are meaningful. To some extent the measurement model can be separately considered. First, the measurement model can be specified and, if it is identified, one can proceed to estimate the covariances among the latent variables. Then one can specify the structural model and, if it is identified, go ahead and estimate it. Occasionally an overidentified structural model can bring about the identification of a previously underidentified measurement model.

The fit of the model can be tested as follows. First the measurement model can be estimated with a just-identified structural model. Any significant lack of fit can be attributed to the measurement model. Then the full model can be estimated and the increased lack of fit can be attributed to the structural model.

A structural equation modeling program is a powerful tool for estimating models with latent variables. When working with the program, one should take care not to fall prey to the many common traps:

1. Check to make sure one has not obtained a local minimum by altering the starting values.
2. Check the degrees of freedom to see if it agrees with a simple counting.
3. Do not rely on the program to determine if the model is identified. Try to independently establish the status of identification.
4. Carefully study the parameter estimates and look for anomalous results.

When I run a structural equation modeling program I presume I have made an error. I check and recheck my results. As George Harrison put it, "with every mistake, we must surely be learning."

10

RANDOMIZED EXPERIMENTS

Although the focus of this text is the analysis of nonexperimental data, in this chapter the topic is randomized experimentation. Structural modeling can add new insights to randomized experiments as well as encourage researchers to apply nonexperimental methods even within experimental settings. The first section of the chapter discusses the logic of experimentation. The second section considers alternative dummy coding schemes. The third section presents an example that uses path-analytic techniques to estimate structural parameters. The fourth section discusses the advantages of a multiple regression approach to experimental data. The fifth section considers alternative structural formulations of repeated measures designs.

LOGIC OF EXPERIMENTATION

The key feature of the randomized experiment is *random assignment of experimental units to conditions*. To randomly assign units the experimenter must be able to control or manipulate the causal variable. Occasionally natural experiments like the draft lottery occur (Staw, 1974), but in the main randomization requires that the experimenter control the causal variable. This control has brought about the term that is used to refer to the exogenous variable in an experiment: the *independent variable*. This variable is called independent since it is under experimental control and is not related to any other important exogenous variables. The endogenous variable is called the dependent variable since any covariation between it and the

independent variable is brought about through the independent variable.

The requirement of random assignment forces social scientists to choose as independent variables those variables that are under experimental control. To some degree experimentation has locked social science and particularly psychology into an epistemology that outside forces determine the behavior of the individual. Since stimulus variables can ordinarily be manipulated whereas response variables can only be manipulated with difficulty, it is hardly surprising that experimentalists are predisposed to a stimulus-response approach to human behavior (see Chapter 12).

The experimental researcher possesses a vast arsenal of experimental designs from which to choose. For the moment consider experimental designs commonly called between-subject designs. Persons are randomly assigned to receive one particular treatment. In within-subject designs persons receive multiple treatments.

Imagine a two-group experiment in which Np participants are in the experimental group and $N(1 - p)$ or Nq are in the control group, N being the total number of persons and p the probability of being assigned to the experimental group. Let X be a dummy variable such that when the subject is in the experimental group he or she receives one and if in the control group, the person receives a zero. Calling the dependent variable Y , it follows that $V(X) = pq$ and $C(X, Y) = (Y_E - Y_C)pq$ where Y_E is the experimental group mean and Y_C the control group mean. The regression of Y on X is the mean difference between the experimental and control group. Not surprisingly, going from 0 to 1 in X , the dependent variable increases by $(Y_E - Y_C)$.

In Figure 10.1 there is a path diagram. In this case X is the dummy variable, 1 for treatment and 0 for control, and Y is the dependent variable and U is the residual from the regression of Y on X or $Y - b_{YX}X$. The important point about experimentation is that one knows that in the population the disturbance is uncorrelated with the exogenous variable. Given of random assignment, the independent variable must be uncorrelated with all other causes of the dependent

variables. However, in any sample the disturbance is correlation with a causal variable due to sampling error, making a guarantee of a zero correlation good only in the long run. Clearly, experimentation provides the researcher a specification of great importance. In Chapter 4 it was shown that the key assumption for the causal interpretation of regression coefficients is that the disturbance is uncorrelated with the causal variables; however, in experimentation this is guaranteed!

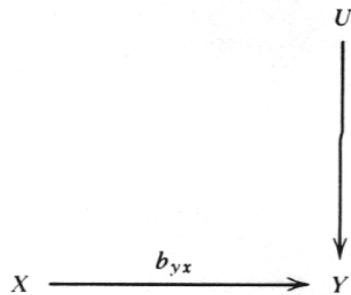


Figure 10.1 Path model for the experiment.

Of course, demonstrating a causal effect is only part of the game. One must attach to that effect a verbal interpretation. The result must be put within some theoretical context. Moreover, often one must rule out the hypothesis that the independent variable is confounded with some irrelevant variable. Our labeling the independent variable in particular does not ensure that it is in fact that variable; Cook and Campbell (1976) call this problem construct validity. Experimentation in and of itself rules out a number of classic confounding variables such as history, maturation, and selection, but numerous other artifacts like experimenter bias and demand characteristics remain. However, the strength of causal inference within randomized experiments gives it the highest degree of interval validity (Campbell & Stanley, 1963) of any method discussed in this text.

Another virtue of randomized experiments is the possibility of uncorrelated causal variables. This strength of experiments rest again on the fact that the experimenter controls the causal variables. Consider a 2 x 2 design. One independent variable, called A , has two levels, A_1 and A_2 , and the other called B , also has two levels, B_1 and B_2 . There are then four cells in the design, A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . With multiple regression three dummy variables are created, that is, as many dummies as there are cells in the design less one. In Table 10.1 there is what is called the model matrix or design matrix (Bock, 1975). The columns are the cells of the design and the rows are the dummy variables. The first dummy, X_1 , assigns a 1 to participants in A_1 and a -1 to those in A_2 . Obviously this dummy tests the A main effect. The second dummy, X_2 , assigns a 1 to those in the B_1 cells and a -1 to those in the two B_2 cells and is a test of the B main effect. The last dummy, X_3 , is formed by taking a product of the main effect dummies, X_1X_2 , and tests interaction.

Table 10.1. Model Matrix a 2 X 2 (A X B) Factorial Design

Dummy Variable	Cell			
	A_1B_1	A_1B_2	A_2B_1	A_2B_2
$X_1:A$	1	1	-1	-1
$X_2:B$	1	-1	1	-1
$X_3:A \times B$	1	-1	-1	1

The textbook recommendation for factorial design is to have equal number of participants in each cell, say n . Examine now the correlation between X_1 and X_2 . First, all three means are zero because half the participants have a score of 1 and the other half have -1 . Figuring $\sum X_1X_2$, it is n for A_1B_1 , $-n$ for A_1B_2 , $-n$ for A_2B_1 , and n for A_2B_2 all of which sums to zero. Working through $\sum X_1X_3$ and $\sum X_2X_3$ again zero is obtained. Thus, in this case and in general, given factorial design and equal number of participants, the main effects and their

interactions are all unconfounded. Multicollinearity evaporates as a problem. Actually the whole strategy of analysis of variance is based on this fact. Given factorial design, the multiple correlation squared (the between group sums of squares) can be partitioned unambiguously into separate effects. The analysis of variance can be viewed as a computational shortcut for balanced (equal n) factorial designs.

DUMMY VARIABLES

Consider again the dummy variables and model matrix. The values of the model matrix may be designated λ_{ij} where the first subscript i refers to the dummy variable X_i and the second subscript j refers to the column which in turn refers to a cell of the factorial design. The mean of a dummy variable X_i is

$$M_i = \frac{\sum n_j \lambda_{ij}}{N} \tag{10.1}$$

where n_j is the number of persons in the j^{th} cell and N is the total number of participants. If $\sum_j \lambda_{ij}$ equals zero, then M_j is also zero when n_j is constant across all conditions. The covariance between two dummies is given by

$$C(X_i, X_k) = \frac{\sum_j n_j \lambda_{ij} \lambda_{kj}}{N} - \frac{(\sum_j \lambda_{ij})(\sum_j \lambda_{kj})}{N^2} \tag{10.2}$$

If the covariance is zero, the two dummies are said to be orthogonal. If $\sum_j \lambda_{ij} = \sum_j \lambda_{kj} = \sum_j \lambda_{ij} \lambda_{kj} = 0$, then Equation 10.2 is also be zero if the cell sizes are equal. It is noted here that the variance of a dummy is a special case of Equation 10.2:

$$C(X_i, X_i) = \frac{\sum_j n_j \lambda_{ij}^2}{N} - \frac{(\sum_j \lambda_{ij})^2}{N^2}$$

A researcher is presented a choice in the coding of multilevel independent variables. As an example consider a four-level independent variable and three different types of coding: Helmert, deviation, and effects coding.

Table 10.2. Three Different Methods of Coding a Multilevel Nominal Variable

		Level			
		I	II	III	IV
Helmert:	X_1	1	-1/3	-1/3	-1/3
	X_2	0	1	-1/2	-1/2
	X_3	0	0	1	-1
Deviation:	X_1	1	-1/3	-1/3	-1/3
	X_2	-1/3	1	-1/3	-1/3
	X_3	-1/3	-1/3	1	-1/3
Effects (simple):	X_1	1	0	0	-1
	X_2	0	1	0	-1
	X_3	0	0	1	-1

In Table 10.2 there are examples of these three types of codes. One use for Helmert codes would be in the case of hierarchically ordered groups. For instance, imagine an educational evaluation that involves a considerable amount of training of students. An experiment was designed with four groups:

1. Pure control.
2. Training only.
3. Training plus one month's use.
4. Training plus six months' use.

Helmert contrasts could be used to contrast group 1 against groups 2, 3, and 4; group 2 against groups 3 and 4; and group 3 against group 4. The first contrast would test the combined effect of the

innovation, the second would test whether use was of value, and the last contrast would test whether 6 months' use is better than one month's. Note that if there is equal n in the groups, all the Helmert contrasts have zero mean and are orthogonal.

Deviation coding, the second method in Table 10.2, seems to compare a group with the other groups. In the univariate case if level I is coded with a 1 and the other three levels with $-1/3$, the univariate regression of the dependent variable on X_i yields mean for the level minus the grand mean. However, if the other deviation codes are created as in Table 10.2, the partial regression of the dependent variable on X_1 partialling out X_2 and X_3 yields the mean for the level minus the grand mean all of which is divided by two.

Although a deviation coding seems to compare one group with the other groups, when all the codes are entered in one equation, it actually compares one group with one other group. To see this, note that X_1 , X_2 , and X_3 all compare groups 1, 2, and 3. The only part that is unique about each contrast is the comparison of each group with the fourth group. Thus, when the other variables are partialled out, only the comparison of each group with group four remains. Deviation codings are clearly not orthogonal, but are negatively correlated.

The final sets of codes are called effects coding or simple coding by Bock (1975). These codes work in just the opposite way as the deviation codes. When only one effect is in the equation, the regression coefficient X_1 yields the difference between the mean for the level minus the grand mean all of which is divided by 2; when all the terms are in the equation, the effect of X_1 equals the mean of the level minus the grand mean.

They are called effects codes because they yield regression coefficients identical to analysis of variance effect estimates. Effects coding is not orthogonal even in the equal n case and usually yields positive correlations. For example, with equal n and any number of categories, the correlation between two dummies is .5.

Clearly the creation of dummies must be done with care. They do not always estimate or test what they seem to be. There are two

helpful rules that can aid one in setting up dummy variables. However, the rules should not be slavishly followed if they conflict with theoretical considerations. First try to keep the dummies as orthogonal as possible. A second helpful rule is to express dummies in mean deviation form, which yields three advantages:

1. It can be quickly determined if two dummies are orthogonal by $\sum_j n_j \lambda_{ij} \lambda_{kj}$.
2. The intercept in the regression equation estimates the grand mean.
3. If interaction contrasts are formed by the multiplication of dummy variables, the interaction term is usually not correlated very much with each of the main effects (Althausser, 1971).

This last point is an important one. Consider a 2 x 2 factorial design with equal n . If one uses 0 and 1 coding for the main effect, the resulting coding for interaction is 0, 0, 0, and 1. However, using 1 and -1 codes for the main effects, there is a second set of codes for the interaction, 1, -1 , -1 , and 1. This second set of codes is uncorrelated with the main effects, whereas the first set has a .289 correlation with each main effect.

EXAMPLE

Sibley (1976) investigated a possible source of bias in the perception of grammaticality of sentences drawn from articles by linguists. She speculated that linguists "set up" the reader to perceive the sentence in the way desired by the linguist. Sibley tested this hypothesis experimentally. The experiment involved 16 linguists and 16 non-linguists. Half of each received 120 sentences from two articles by prominent linguists in the same order as given by the original author while the other half received the sentences in a random order. The final independent variable was the instructions given to the participants.

Half were told to make an intuitive judgment while the other half were told to use grammatical rules if necessary. The experiment is then a 2 x 2 x 2 factorial design. Four participants were in each condition and participants were randomly assigned to order by instruction condition. The dependent variable is the number of "errors" made over the 120 sentences, an error being defined as a disagreement with the author.

Table 10.4. Cell Means for Linguistics Study

	C_1		C_2	
	B_1	B_2	B_1	B_2
A_1	18.00 ^a	24.50	16.75	22.75
A_2	24.25	26.25	23.50	32.00

^aCell size equals four.

In Table 10.3 there is the model matrix for the design. The first dummy, X_1 , compares the linguists with the non-linguists. The second, X_2 , compares the author's order with the random order, and X_3 compares the intuition instruction with the rule instruction. Dummy X_4 is formed by multiplying X_1 by X_2 ; X_5 by X_1 and X_3 ; X_6 by X_2 and X_3 ; and X_7 by X_1 , X_2 , and X_3 . These four variables are the interaction effects. All dummies have zero mean and are orthogonal. In Table 10.4 are the cell means. First note that each dummy variable is a dichotomy and so one can use the formula given in the next chapter to compute the point biserial correlation between each of the dichotomies with the dependent variable Y . For instance, for r_{X_1Y} the formula is $(20.5 - 26.5)(0.5)/5.847$ yielding -0.513 . The standard deviation of the dependent variable is 5.847. In Table 10.5 are the other correlations for

Table 10.3. Model Matrix for Linguistics Study^a

	$A_1B_1C_1$	$A_1B_1C_2$	$A_1B_2C_1$	$A_1B_2C_2$	$A_2B_1C_1$	$A_2B_1C_2$	$A_2B_2C_1$	$A_2B_2C_2$
X_1 : A	1	1	1	1	-1	-1	-1	-1
X_2 : B	1	1	-1	-1	1	1	-1	-1
X_3 : C	1	-1	1	-1	-1	-1	1	-1
X_4 : A x B	1	1	-1	-1	-1	-1	1	1
X_5 : A x C	1	-1	1	-1	-1	-1	-1	1
X_6 : B x C	1	-1	-1	1	1	-1	-1	1
X_7 : A x B x C	1	-1	-1	1	-1	1	1	-1

^a A_1 , linguist; A_2 , nonlinguist; B_1 , author's order; B_2 , random order; C_1 , intuition set; C_2 , rule set.

each of by prominent linguists in the same order as given by the original author while the other half received the sentences in a random order. The final independent variable was the instructions given to the participants. Half were told to make an intuitive judgment while the other half were told to use grammatical rules if necessary. The experiment is then a $2 \times 2 \times 2$ factorial design. Four participants were in each condition and participants were randomly assigned to order by instruction condition. The dependent variable is the number of "errors" made over the 120 sentences, an error being defined as a disagreement with the author.

Table 10.4. Cell Means for Linguistics Study

	C_1		C_2	
	B_1	B_2	B_1	B_2
A_1	18.00 ^a	24.50	16.75	22.75
A_2	24.25	26.25	23.50	32.00

^aCell size equals four.

In Table 10.3 there is the model matrix for the design. The first dummy, X_1 , compares the linguists with the non-linguists. The second, X_2 , compares the author's order with the random order, and X_3 compares the intuition instruction with the rule instruction. Dummy X_4 is formed by multiplying X_1 by X_2 ; X_5 by X_1 and X_3 ; X_6 by X_2 and X_3 ; and X_7 by X_1 , X_2 , and X_3 . These four variables are the interaction effects. All dummies have zero mean and are orthogonal. In Table 10.4 are the cell means. First note that each dummy variable is a dichotomy and so one can use the formula given in the next chapter to compute the point biserial correlation between each of the dichotomies with the dependent variable Y . For instance, for r_{X_1Y} the formula is $(20.5 - 26.5)(0.5)/5.847$ yielding -0.513 . The standard deviation of the dependent variable is 5.847. In Table 10.5 are the other correlations for

each of the dummies. Because the dummies are uncorrelated, the correlation is the standardized regression coefficient which in turn is the path coefficient. To find R^2 simply compute the sum of all the correlations squared or .577. In Table 10.5 is the t -test for each coefficient using Equation 4.1. Table 10.5 also contains the unstandardized coefficients, which are identical to the ANOVA effect estimates. Finally, in Figure 10.2 is a path diagram.

Table 10.5. Analysis of Variance Table^a

Effect	r and β	Unstandardized	
		Coefficient	t
A	-.513	-3.000	-3.48
B	-.492	-2.875	-3.34
C	-.043	-.250	-.29
$A \times B$	-.043	-.250	-.29
$A \times C$.171	1.000	1.16
$B \times C$.128	.750	.87
$A \times B \times C$	-.150	-.875	-1.02

^a $R^2 = .577$.

The results of Sibley's experiment show that linguists make fewer errors than non-linguists and that more errors are made when the sentences are in a random order than in the author's order which confirmed Sibley's hypothesis. No interactions are indicated.

DUMMY VARIABLE REGRESSION

There has been recent recognition that the analysis of variance (ANOVA) is only a special case of multiple regression. As has just been seen by judicious choice of dummy variables, multiple regression can reproduce the results of ANOVA. A relatively straightforward presentation of this process is given in Cohen and Cohen (1975), and a more complicated but highly general presentation is given in Bock (1975).

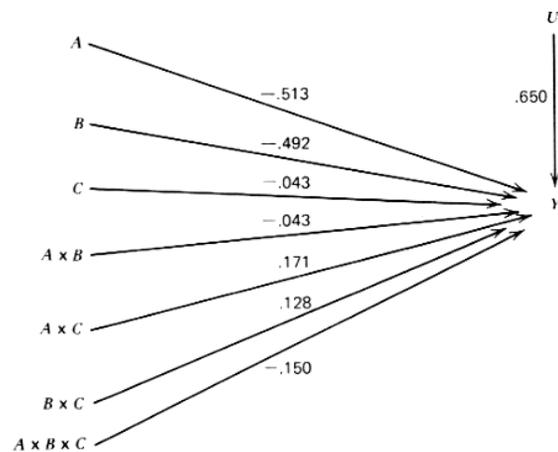


Figure 10.2 Linguistics example.

The virtues of multiple regression over ANOVA are manifold. First, the use of dummy variables encourages researchers to scale their independent variables. Since a scientific law is often best stated as a functional relationship between the independent variable and the dependent variable, the independent variable should be measured on a scale. Often variables in experiments are only operationalized with two levels, high and low, but with multilevel independent variables, the researcher using multiple regression is forced to choose dummy variable weights for the independent variable. Many times there is a natural metric for the independent variable, for example, percent similarity, status, amount of information, or level of motivation. These variables can be easily scaled and the scale weights can be used as a variable and entered into the regression analysis. It can then be tested whether these scaled values explain the variation among the cell means and, more importantly, whether there is no systematic dependent variable variation after the scaled independent variable has been controlled. For instance, Byrne and his associates have consistently shown that attraction is a direct function of percent similarity and that there is no evidence of nonlinearity. Byrne can then state not only a

relationship between the independent variable and the dependent variable, but he can also state the functional form of the relationship which allows both interpolation and extrapolation of results.

Even if there is no natural scale for the independent variable, one could be derived by the use of scaling methods. One candidate for scaling independent variables might be multidimensional scaling. Levels of an independent variable could be first judged regarding their similarity. Then the similarity judgments could be multi-dimensionally scaled. The dimensions obtained from the scaling could be used in the dummy coding of independent variables, and then it could be judged which independent variable explained the variation in the cell means.

Another strategy might be to have expert judges scale the independent variable. The cells in a factorial experiment could be judged as to how effective they are in producing change according to a theoretical formulation. "Judged effectiveness" would then be an independent variable in the regression analysis and it would be tested whether it adequately explains the variation among the cell means. Ideally all the systematic variation of the dependent variable should be explained by the scaled independent variable or some simple transformation of the scale (e.g., logarithmic), and the residual effect should be nonsignificant.

Scaling the independent variable provides a dividend of increased power. Because the scaled independent variable is captured by a single variable, it is tested by a single degree of freedom, and F s with one degree of freedom on the numerator have potentially more power than those with multiple degrees of freedom.

A second advantage of multiple regression over ANOVA is the generality of design. ANOVA is best suited to the case of equal number of participants in each treatment condition. In fact, ANOVA can be viewed as a simplification in the computations of multiple regression in the equal n case. In the all too common unequal n situation, researchers often employ highly questionable correction procedures. Some researchers randomly discard precious data; others gather data from additional participants, ignoring the chance that either

the subject population or the experimental setting have changed. Equal or unequal sample size presents no special conceptual problems within regression analysis. With unequal n it is randomized that effects become partially confounded, and the efficiency of effect estimates are reduced. This situation, known as multicollinearity, is not so much a problem but rather the price we pay for unequal n , and of course it is present in either variance or regression analyses.

Table 10.6. Unweighted versus Weighted Means Example

Means			
B			
A	1	2	3
1	7	7	13
2	7	7	13
First set of sample sizes			
B			
A	1	2	3
1	10	5	10
2	10	5	10
Second set of sample sizes			
B			
A	1	2	3
1	10	10	5
2	10	10	5

Given multicollinearity, the between-cells sum of squares cannot be orthogonally partitioned. Unfortunately some researchers have advocated methods that attempt to orthogonally partition the sum of squares: unweighted means analysis and hierarchical model testing. Although the former is an ANOVA strategy and the latter is a multiple

regression strategy, both seek equal n ANOVA-like partitioning of variance. The following shows how both unweighted means analysis and hierarchical model testing can both yield anomalous results with unequal n .

Table 10.7. Illustration of Hierarchical Model Testing

First Set			Second Set		
Means					
B			B		
A	1	2	A	1	2
1	10	20	1	10	18
2	20	10	2	12	20
Sample Sizes					
B			B		
A	1	2	A	1	2
1	5	45	1	9	15
2	5	45	2	15	9

Unweighted means analysis assumes that all the cell means are based on the harmonic mean. Unweighted means is meant to be only an approximation and is not, as is commonly thought, meant to handle accidental unequal sampling. To illustrate the arbitrariness of the unweighted means analysis, consider the means in Table 10.6. These artificial data indicate that there is a main effect of independent variable B . The values in column 3 have a mean of 13.0 and those in columns 1 and 2 have a mean of 7.0. Confidence in the existence of the main effect should be a function of the distribution of the two sets of sample sizes. The first set of sample sizes in Table 10.6 would inspire greater confidence than the second set of sample sizes because, in the first set of sample sizes, the deviant means in column 3 are associated with large sample sizes whereas in the second set the deviant means are associated with small sample sizes. Since both have harmonic cell means of 7.5, an unweighted means analysis yields a sum of squares of

360 for the column main effect in both cases. A weighting of means by sample sizes yields a sum of squares for B more in accord with intuition: 432 for the first set of values and 288 for the second set. Unweighted means analysis fails to weight the deviancy of a mean from other means by its sample size.

The second approach to be criticized is called a hierarchical approach, or method III by Overall and Spiegel (1969). The effects are first ordered, say A , B , and $A \times B$. First, the effects of A are fitted ignoring the effects of B and $A \times B$. The effect of B is then fitted controlling for A and ignoring $A \times B$. Finally, the $A \times B$ effect is fitted controlling for both A and B . In this way the between-cells sum of squares can be orthogonally partitioned. Note, however, that the effect of A is confounded with B and $A \times B$ and the effect of B with $A \times B$ since the latter effects are not controlled. This fact can create anomalous results. Examine, for instance, the first set of means in Table 10.7. They clearly illustrate only an $A \times B$ effect. The between-group sums of squares is 2500. Using the hierarchical approach, the weighted marginal means for A are 19 and 11. Because of unequal n they are not 15 and 15. Computing sums of squares for A yields 1600. None of the sum of squares can be attributed to B after fitting A . Fitting the interaction controlling for the main effects yields a sum of squares of 900. Thus, the total between-cells sum of squares has been orthogonally partitioned between A and $A \times B$. Counter-intuitively, more of the sum of squares is attributed to A than to $A \times B$. This is due to the strong confounding of the A and $A \times B$ effect. Using the dummy codes for A and $A \times B$, the correlation of A with $A \times B$ is .8. Since A explains 64% of the variance of $A \times B$, 64% of the interaction effect is attributed to A (i.e., $.64 \times 2500 = 1600$). Clearly, anomalous conclusions are very possible with the hierarchical mode of model testing. The logic of the hierarchical method implies that if the interaction is statistically significant the main effect should not be tested since the test of the main effect has ignored the significant interaction. In practice, however, most users of the hierarchical approach do not examine the confounding and do not stop testing.

It should be noted that ignoring the effects of the other independent variables (i.e., not controlling for them) does not always inflate the sum of squares attributed to an effect. It is possible, although it is empirically unusual, that effects could be grossly underestimated. Given the second set of means in Table 10.7, the marginal means of A are both 15 ignoring B , but clearly there is a main effect of A . This effect is formally identical to the suppressor variable phenomenon in multiple regression.

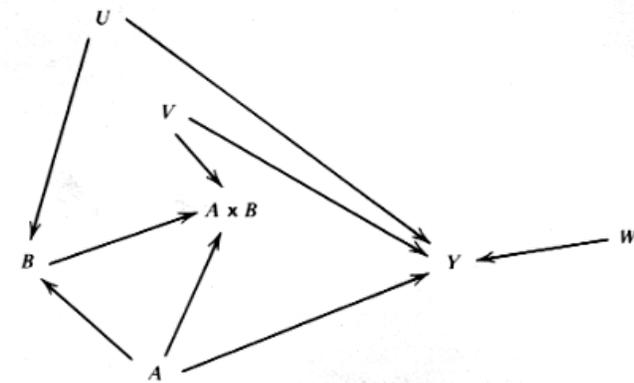


Figure 10.3 Hierarchical ANOVA.

The hierarchical method implicitly assumes that the independent variables can be causally ordered. For instance, as in Figure 10.3, A causes B and $A \times B$, B causes $A \times B$ and not A , and $A \times B$ does not cause A or B . Then the dependent variable Y is regressed on A , U , and V . The effect for A includes the direct effect of A plus its indirect effect through both B and $A \times B$. The effect of U includes the direct effect of B plus the indirect effect of B through $A \times B$. The effect of V includes only the direct effect of $A \times B$. Thus the direct effect of A could be zero, but A could be judged to have a significant effect if A and B are correlated and B has a direct effect. The procedure yields an orthogonal partitioning of variance since A , U , and V are uncorrelated; however, it does not test effects in a sensible manner. Moreover, the

independent variables in experiments have no causal orderings and the notion of indirect effects makes no sense.

Given unequal n , confounding or multicollinearity is inevitable in factorial designs. The solution to the problem is not to find some way to achieve orthogonal partitioning as in the equal n case. Multicollinearity is not a problem in experimental design; rather, it is a cost that must be paid given the failure to have equal n .

With regression analysis there is a more reasonable approach to the testing of causal effects. The significance of each effect can be judged by comparing the sum of squares with the effect in the model with the sum of squares with the effect removed. All other effects are included in both models. This strategy is called method I by Overall and Spiegel (1969) and has been recommended by a chorus of authors (cf. Overall, Spiegel, & Cohen, 1975). For example, for the first set of means in Table 10.7 no variance would be attributed to the main effects and 900 of the sum of squares would be attributed to interaction. Given confounding, 1600 of the sum of squares cannot be attributed. The reader should also consult an alternative approach given by Appelbaum and Cramer (1974).

A third advantage of multiple regression is the ease in pooling interactions into the error term. Anticipated nonsignificant interactions can be dropped from the regression equation, increasing the number of degrees of freedom in the error term and thereby increasing power. This procedure allows for a large number of between-subject independent variables even with a small number of participants in each cell.

A fourth advantage of the use of multiple regression is that it is flexible as to the type of independent variables. Nominal, ordinal, and interval independent variables are possible in multiple regression. ANOVA is limited to discrete independent variables and analysis of covariance is limited to a set (usually one) of independent variables measured at the interval level of measurement. Some researchers are reluctant to apply multiple regression because of the assumption of linearity. However, because multiple regression reproduces the results

of ANOVA, the same set of assumptions applies to both. As with ANOVA, multiple regression assumes an additive model which is a linear equation. Within regression analysis nonlinearity and interaction can be handled by creating additional dummy variables.

ANOVA is in no danger of becoming an endangered statistical species. One major virtue of ANOVA is that it deals with means and the discussion of means often makes greater intuitive sense than regression coefficients. Means can be only indirectly computed from regression coefficients. Multiple regression is often cumbersome to apply: The possible number of terms in the regression equation is the number of cells in the design, and an equation with as many as 64 terms is rarely very comprehensible. The cumbersomeness becomes almost intolerable when the complications of repeated measures and mixed models are added. Many of the features of multiple regression can be applied to ANOVA by the use of special contrasts that are sometimes called planned comparisons. Obviously the researcher should be flexible about when to use ANOVA or multiple regression, but quite clearly multiple regression has some important advantages.

MULTIPLE DEPENDENT MEASURES

Often a researcher is at a loss how to analyze data when there is more than one dependent variable. At the very least, six different strategies are available:

1. Perform repeated measures ANOVA.
2. Perform an ANOVA on the sum of the measures.
3. Perform a MANOVA.
4. Sum the measures weighted by a factor analysis.
5. Covary on one of the measures.
6. Perform a univariate analysis on each of the measures.

The choice among the six strategies can be facilitated by examining three causal models. Imagine two independent variables, X_1

and X_2 , and two measured dependent variables, Y_1 and Y_2 . In Figure 10.4 path diagrams represent three different causal models for the four variables.

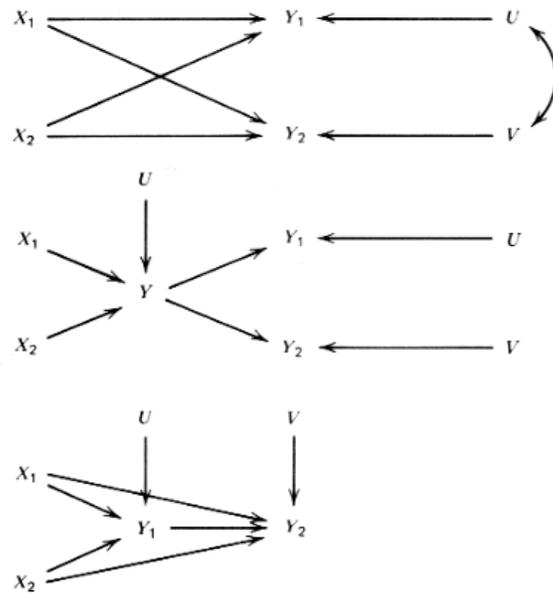


Figure 10.4 Models for repeated measures.

The top causal model in Figure 10.4 assumes that the independent variables cause each of the dependent variables. There are two different approaches to the analysis of this type of model: univariate and multivariate. The univariate strategy would be to perform separate ANOVA's on each of the dependent variables. However, as the number of dependent variables increases, such a strategy would result in a large increase in the probability of both Type I and Type II errors. To reduce the probability of these errors, a multivariate approach should be used. A multivariate F tests the effect of an independent variable on a weighted sum of the dependent variables. The weights for linear combination of the dependent variable

are chosen to maximize the univariate F of that linear combination. Significance levels are adjusted because the F is spuriously high. The multivariate F is then "hypothesisless" since the particular nature of the effect is specified by a statistical, not conceptual, criterion.

The middle path diagram in Figure 10.4 contains an unmeasured variable, Y , that the independent variables cause, and Y , in turn, causes the measured dependent variables. Unlike the top path diagram, the independent variables cause a construct of which the dependent variables are only indicators. An example of this model might be when some manipulation causes arousal which is unmeasured; possible indicators of arousal are fear, tension, and heart rate increase. To test this model, consider two sub-models of the middle figure: correlated errors and uncorrelated errors. Correlated errors are represented by the curved line between U and V which is simply a theoretical partial correlation between Y_1 and Y_2 with Y partialled out. If Y totally explains the correlation between Y_1 and Y_2 , there are no correlated errors. Given correlated errors, a multivariate analysis is the appropriate method. There should, however, be only one statistically significant solution. As in Figure 9.5, only the first root of the canonical analysis should be statistically significant. One should not perform the analysis separately for each independent variable since this would not capture the specification that Y mediate the effects of both independent variables on both dependent variables. One can perform a canonical correlational analysis and then test each independent variable with a stepdown F (cf. Bock, 1975); that is, each independent variable is removed from the analysis and the reduction of the canonical correlation is then judged.

If it is assumed that there are no correlated errors, then a factor analytic approach is appropriate. Given no correlated errors, the correlation between the dependent variables is totally explained by the unmeasured variable, that is, by the factor Y . Scores on the factor can then be estimated and treated as dependent variables. To estimate the factor loadings at least three dependent variables are needed. Ideally the raw correlation matrix is not factored but the correlation matrix

with independent variables partialled out. If the model is well specified, this should yield a single-factor solution and factor scores can then be estimated. Finally, a univariate ANOVA can be performed on the factor scores. This F is more powerful than the multivariate F since it allows for specification of the construct. More efficient and more general estimation procedures can be obtained by using confirmatory factor analysis, as discussed in Chapter 7, 8, and 9. Alwin and Tessler (1974) used confirmatory factor analysis in an experimental investigation of clients' reactions to initial interviews. They had multiple measures for both independent and dependent variables.

A repeated measures ANOVA is sometimes appropriate for the middle path diagram in Figure 10.4. Since factor loadings are often roughly equal, a simple adding of the variables, as in a repeated measures ANOVA, yields a composite highly correlated with factor scores. Because adding the variables implies an equal weighting, this assumption can be tested by no significant independent variable by repeated measures interaction. To perform a repeated measures ANOVA, all the dependent variables should be measured on the same scale.

The bottom diagram in Figure 10.4 postulates a causal relationship between the dependent variables. Dependent variable Y_1 is assumed to cause Y_2 . Variable Y_1 can be a manipulation check, an intervening variable, or a prior measure of Y_2 . This model is like the middle diagram but in this case the intervening variable is measured. The appropriate analysis for this model is to compute two univariate ANOVAs. In the first Y_1 is the dependent variable, and in the second Y_1 is a covariate and Y_2 the dependent variable. If one prefers regression analysis, one first regresses Y_1 on X_1 and X_2 and then Y_2 on X_1 , X_2 , and Y_1 . The suggestion of covarying on Y_1 breaks the rule that the covariate should not be caused by the independent variables, but such a rule is important for the main thrust of covariance analysis: increasing power. Covariance analysis can also be used to control for mediating variables like Y_1 . One must assume that the reliability of Y_1 is high, say at least

.90. (If not, this model becomes the middle diagram in Figure 10.4.) If the reliability of Y_1 is not unity, the effects of X_1 and X_2 on Y_2 are ordinarily though not always overestimated (see Chapter 5). Often for this model it is postulated that Y_1 mediates the causal effect of X_1 and X_2 on Y_2 . For instance, Insko et al. (1973) argued that implied evaluation mediates the effects of similarity on attraction. They showed that by covarying implied evaluation out of attraction there was no effect of similarity. This does not demonstrate that the relationship between similarity and attraction is spurious but only that the causal relationship is mediated by an intervening variable, implied evaluation. Serious multicollinearity problems can develop if the relationship between X_1 or X_2 with Y_1 is strong. This would especially be true if Y_1 is a manipulation check. It should be made clear that covarying on a dependent variable is permissible only if it can be assumed that the same variable mediates a causal relationship between the independent variable and another dependent variable.

Thus, a structural equation modeling approach is useful in determining the mode of statistical analysis. Note that it is not the design per se that determines the statistical analysis, but the design plus the causal model. The difficult question is often not the choice of the statistical analysis but the choice of the causal model.

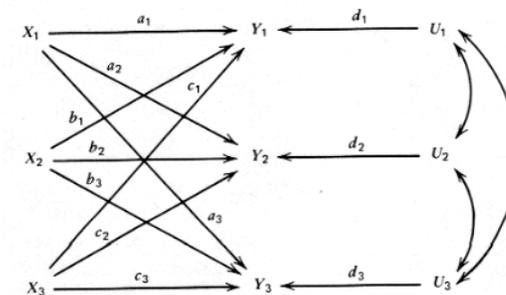


Figure 10.5 A model with between- and within-subject variables.

The combination of both between- and within-subjects (repeated measurements) variables can prove to be rather awkward when one sets up a structural model. Nonetheless structural models for such designs can provide richer information than either a repeated measures analysis of variance or multiple regression analysis. For instance, consider a 2 x 2 design with three repeated measurements. Call the within-subjects variable trials and let three dummy variables represent the 2 x 2 between-subjects variables: X_1 for the main effect of A , X_2 for the main effect of B , and X_3 for the $A \times B$ interaction. In Figure 10.5 is a path diagram for the six variables. The X and Y variables are unstandardized, but the U variables are standardized. There are no curved lines between the X variables because it is assumed that there are an equal number of participants in each of the four cells of the design; however, if cell sizes were unequal no special problems would arise. The minimum condition of identifiability is met since there are 15 correlations and 15 free parameters, 9 paths, 3 covariances among the X variables, and 3 correlations among the disturbances. Even though the researcher knows that the covariances between X_1 , X_2 , and X_3 are zero, they still need to be estimated. The model is a hierarchical model and can be estimated by multiple regression and partial correlation analysis.

One can employ a structural equation modeling program to test the usual analysis of variance hypotheses. For instance, a model in which $c_1 = c_2 = c_3$ tests the $A \times B$ by trials interaction. That is, if there were no interaction the coefficients would be equal. Recall that an interaction specifies that a causal law varies as a function of some other variable. The test of the A by trials interaction sets $a_1 = a_2 = a_3$ and the B by trials interaction sets $b_1 = b_2 = b_3$.

If none of the between-subjects variables interact with trials, then one can proceed to test the A and B main effects and their interaction. They are: the A main effect — $a_1 = a_2 = a_3 = 0$; the B main effect — $b_1 = b_2 = b_3 = 0$; the $A \times B$ interaction — $c_1 = c_2 = c_3 = 0$. Such an analysis does not make the restrictive assumptions of repeated measures analysis of variance: homogeneity of variance, $d_1 = d_2 = d_3$,

and homogeneity of correlation, $r_{U_1U_2} = r_{U_1U_3} = r_{U_2U_3}$. In fact, these assumptions can be tested by a structural equation modeling program (e.g., LISREL). It is true that the program does assume multivariate normality and the χ^2 test is only approximate; however, it does provide more natural parameter estimates.

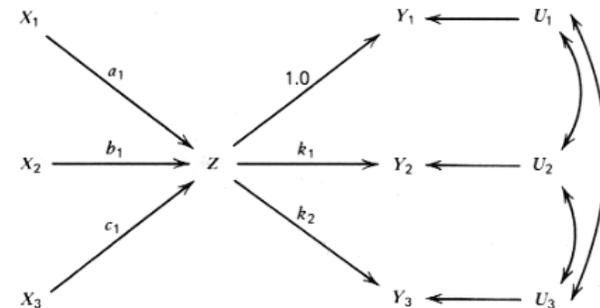


Figure 10.6 Interactions as an artifact of scale.

As an example of the flexibility of the structural approach, consider the researcher who suspects that the interactions of trials with A , B , and $A \times B$ are an artifact of the scaling of the Y variables. Thus if Y_2 were rescaled by Y_2/k_1 and Y_3 by Y_3/k_2 , the interactions would disappear. Such a model presumes that $a_2/a_1 = b_2/b_1 = c_2/c_1 = k_1$ and $a_3/a_1 = b_3/b_1 = c_3/c_1 = k_2$. To test such a model, one estimates the model in Figure 10.6. Note that variable Z has no disturbance. If the model fits the data, then a rescaling can explain the interactions. Such a model can be tested by a multivariate analysis of variance or canonical correlation analysis, but neither provides direct estimates of parameters nor can they allow respecifications such as the Y variables being unmeasured.

CONCLUSION

It has been seen that logic of experimental inference can be sharpened by an understanding of the issues of correlational inference.

Moreover, even in an experimental context the researcher can employ correlational methods. A structural modeling approach forces researchers to elaborate more complicated causal chains whereas the simple-minded analysis of variance method focuses on a single dependent variable. Social science does not need fewer randomized experiments but rather experiments that also analyze the data correlationally.

11

THE NONEQUIVALENT
CONTROL GROUP
DESIGN

A growing concern of social scientists is the evaluation of social programs (Campbell, 1969). Increasingly social scientists are called upon to evaluate educational innovations, behavior modification programs, social work strategies, and job training methods. This chapter shows that structural modeling is equally important in evaluation of social programs as it is in discipline research.

Weiss (1972) has stated that the most common design in evaluation research is a pretest-posttest design in which participants are not randomly assigned to groups or, as Campbell and Stanley (1963) refer to it, the nonequivalent control group design. Thus with this design there is a pretest and posttest, but the control and experimental participants may differ at the pretest because of the failure to randomize. This initial pretest difference may be because:

1. A treatment is administered to classroom, school, or school system and another classroom, school, or school system is taken as a control group.
2. A randomized experiment is planned but because of mortality, contamination of the control persons by the experimentals, or variation in the experimental

treatment, the true experiment has become a quasi-experiment.

3. Because of scarce resources the treatment is only given to a select group.
4. Participants select their own treatment level.

Anyone familiar with the problems encountered in natural settings must realize that although randomization and true experimentation are ideal goals in research, they are not always possible. A pretreatment measure does allow the researcher to assess how confounded the experimental treatment is with the participants' predisposition to respond to the dependent variable.

What follows is a structural analysis of the nonequivalent control group design. The chapter is divided into three sections. The first section specifies a general model. The second section outlines a set of analysis strategies for the design. The third section discusses various special cases of the general model.

GENERAL MODEL

For the nonequivalent control group design there are at least three measured variables: X_1 , a pretest measure of the variable that the treatment is to change; X_2 , a posttest measure; and T , the treatment variable coded 1 for the experimentals and 0 for the controls. The causal function of X is divided into three unmeasured, latent components:

- G — group difference such as sex, race, classroom, and others,
- Z — individual differences within groups,
- E — totally unstable causes of X (errors of measurement).

Variable G must be included in the specification of causes of X because in field settings it can rarely be assumed that one samples from

a single homogeneous population. Usually multiple populations are sampled with each population having different mean level of X .

For most of what follows all variables, both measured and unmeasured, are standardized because standardization decreases algebraic complexity and because in this case it means little loss of generality. Occasionally, however, the unstandardized measure is employed, as in measures of treatment effect.

The equations for X_1 and X_2 can be expressed in terms of their causes G , Z , and E and any treatment effect:

$$X_1 = a_1G + b_1Z_1 + e_1E_1 \quad [11.1]$$

$$X_2 = a_2G + b_2Z_2 + tT + e_2E_2 \quad [11.2]$$

where the subscripts 1 and 2 refer to time. It is assumed that the group differences variable, G , is perfectly stable making its autocorrelation unity which explains why G needs no time subscript. Relative position within groups, Z , may not be perfectly stable making its autocorrelation less than one, while errors of measurement, E , are perfectly unstable making its autocorrelation zero. (Uncorrelated measurement errors are assumed by only one analysis strategy discussed here: analysis of covariance with reliability correction.) It is also assumed that all unmeasured variables are uncorrelated with each other with the previously stated exception that Z_1 and Z_2 may be correlated ($r_{Z_1Z_2} = j$).

If the treatment is correlated with the pretest, it must then be confounded with the causes of the pretest. Thus, the treatment must be correlated with group differences, relative position within groups, errors of measurement, or any combination of the three. So in writing the structural equation for the treatment variable, the variables G , Z , and E must be included. It is assumed that the occasion of selection of the persons into treatment occurs at the pretest, thus making T confounded with Z and E at time one. The causal function of the treatment variable is then

$$T = qG + mZ_1 + sE_1 + fU \quad [11.3]$$

where U is a residual term that is uncorrelated with all the other variables in the model and is simply a variable that represents all other causes of selection besides G , Z , and E .

Equations 11.1, 11.2, and 11.3 have been expressed in the path diagram in Figure 11.1. The correlations between the measured variables are:

$$r_{X_1T} = qa_1 + mb_1 + se_1 \quad [11.4]$$

$$r_{X_2T} = qa_2 + mjb_2 + t \quad [11.5]$$

$$r_{X_1X_2} = a_1a_2 + b_1b_2j + tr_{X_1T} \quad [11.6]$$

The treatment variable has been dummy coded, 1 for treatment and 0 for control. A correlation of a dichotomous variable with another variable is sometimes called a point biserial correlation and is itself a product moment correlation. Campbell (1971) has suggested that these treatment-effect correlations are very useful in assessing treatment effects. A simple correlation may seem to be an unusual measure of difference between treated groups, but it can be shown that both mean difference and the t ratio can be easily transformed into r . The formula for t to r is

$$r = \frac{t}{(t^2 + N - 2)^{1/2}}$$

where N is the total number of participants. The formula for the difference between treated mean (X_T) and control mean (X_C) to r is

$$r = \frac{(X_T - X_C)(P_T P_C)^{1/2}}{s_X}$$

where s_Y is the variability of the observations, P_T the proportion of the total sample in the treated group, and P_C the control group ($P_T + P_C = 1$). The treatment-effect correlation is then closely related to t and the difference between means, but like all measures of correlation it is a standardized measure of the strength of a relationship. The use of dummy codes for categorical independent variables is discussed in the previous chapter.

ANALYSIS STRATEGIES

Designate μ_{T1} as the population mean of the treatment group at the pretest, and μ_{C1} as the control group mean at the pretest. Both μ_{T2} and μ_{C2} refer to the population means at the posttest. Those means now refer to the unstandardized pretest and posttest. Ordinarily, in the nonequivalent control group design there is an initial difference such that $\mu_{T1} - \mu_{C1}$ is nonzero. These pretreatment differences then make the interpretation of any difference between groups on the posttest problematic. At issue is whether pretreatment differences should increase, decrease, or remain stable given no treatment effect.

The literature suggests four modes of statistical analysis to measure treatment effects in the nonequivalent control group design:

1. Analysis of covariance.
2. Analysis of covariance with reliability correction.
3. Raw change score analysis.
4. Standardized change score analysis.

In Table 11.1 the differences between these four techniques can be illustrated by considering what each technique takes as the dependent variable. In covariance analysis the pretest is a covariate and the residual from the posttest-pretest regression is the dependent variable.

Table 11.1. Dependent Variables and Null Hypotheses for Four Modes of Analysis Where X_1 is the Pretest, X_2 the Posttest, and T the Treatment Variable^a

Statistical Technique	Dependent Variable	Measure of Treatment Effect	Correlational Null Hypothesis
Analysis of covariance	$X_2 - b_{21.T}X_1$	$(\bar{X}_{2T} - \bar{X}_{2C}) - b_{21.T}(\bar{X}_{1T} - \bar{X}_{1C})$	$\rho_{X_2T} - \rho_{X_1T}\rho_{X_1X_2} = 0$
Covariance with reliability correction	$X_2 - b_{21.T}^rX_1$	$(\bar{X}_{2T} - \bar{X}_{2C}) - b_{21.T}^r(\bar{X}_{1T} - \bar{X}_{1C})$	$\rho_{X_2T} - \rho_{X_1T}\rho_{X_1X_2}/\rho_{X_1X_1} = 0$
Raw change score analysis	$X_2 - X_1$	$(\bar{X}_{2T} - \bar{X}_{2C}) - (\bar{X}_{1T} - \bar{X}_{1C})$	$\rho_{X_2T} - \rho_{X_1T}\sigma_{X_1}/\sigma_{X_2} = 0$
Standardized change score analysis	$X_2 - (s_{X_2.T}/s_{X_1.T})X_1$	$(\bar{X}_{2T} - \bar{X}_{2C}) - s_{X_2.T}/s_{X_1.T}(\bar{X}_{1T} - \bar{X}_{1C})$	$\rho_{X_2T} - \rho_{X_1T} = 0$

^a $b_{21.T}$ is the unstandardized partial regression of X_2 on X_1 , controlling for T , and * indicates corrected for attenuation.

The results of analysis of covariance are virtually equivalent to a multiple regression analysis in which the pretest and treatment variable are the predictors, and the posttest the criterion. Both Lord (1960) and Porter and Chibucos (1974) have suggested that the estimated regression coefficient is attenuated by measurement error in the pretest, and so the regression coefficient must be corrected for attenuation. For raw change score analysis the pretest is simply subtracted from the posttest.

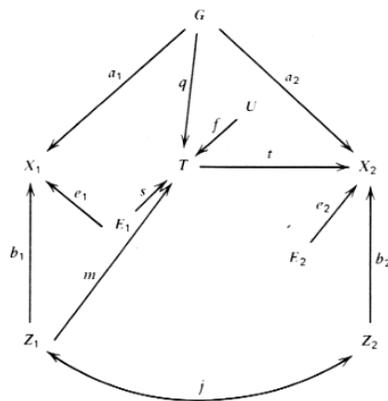


Figure 11.1 Model of selection for the nonequivalent control group design.

Although change score analysis is often condemned (e.g., Cronbach & Furby, 1970; Werts & Linn, 1970), it is perhaps the most common way of analyzing this design. The popularity of this mode of analysis is no doubt due to its seeming ease of interpretation and the fact that it can be viewed as part of a repeated measures analysis of variance. The main effect for treatment in raw change score analysis is equivalent to the Time by Treatment interaction of a repeated-measures analysis of variance (Huck & McLean, 1975). It should be noted that using the change score as the dependent variable and the pretest as a covariate yields significance test results that are identical to the analysis of covariance with a metric of change (Werts & Linn, 1970).

Another mode of analysis is standardized change score analysis. As is seen later, this is actually a correlational method, but it is possible to view the dependent variable of the analysis as the posttest minus the pretest multiplied by the ratio of time 2 pooled within treatment groups standard deviation to the time 1 pooled within treatment groups standard deviation. Kenny (1975a) presents the dependent variable as the difference between the standardized (unit variance, zero mean) pretest and posttest. For reasons to be discussed later, the dependent variable in Table 11.1 is preferable.

There is for each mode of analysis a statistical expression that equals zero if there were no treatment effects, that is, a null hypothesis for each method. Discussed are two different ways of expressing this null hypothesis: means and treatment-effect correlations. Although both ways are algebraically equivalent, each has its own distinct advantage. The use of means gives a measure of treatment effect in the metric of the posttest. Treatment-effect correlations are useful in determining the relative difference between the experimental and control group, that is, relative to the standard deviation.

For each of the four modes of analysis the measure of treatment effect can be viewed as the difference between means at the posttest ($\mu_{T2} - \mu_{C2}$) minus the difference at the pretest ($\mu_{T1} - \mu_{C1}$) times some correction factor. Porter and Chibucos (1974) call this correction factor the index of response. In Table 11.1 are the measures of treatment effect for each of the four methods. For analysis of covariance the correction factor is $b_{21.T}$ the unstandardized partial regression coefficient of X_2 on X_1 controlling for T . For analysis of covariance with reliability correction the correction factor is $b_{21.T}$ corrected for attenuation due to measurement error. The appropriate reliability is the pooled within-treatment groups reliability. Raw change score analysis has no correction factor or, more precisely, a correction factor of one. For standardized change score analysis the correction factor is $s_{X2.T}/s_{X1.T}$ where $s_{X2.T}$ is the pooled within-treatment standard deviation of the posttest.

All four null hypotheses can also be viewed as setting the posttest-treatment correlation equal to the pretest-treatment correlation times a correction factor. For analysis of covariance the correction factor is the pretest-posttest correlation. Since the pretest-posttest correlation is ordinarily less than one, analysis of covariance implies that in the absence of treatment effects, the posttest-treatment correlation is less in absolute value than the pretest-treatment correlation. The correction factor of the analysis of covariance with reliability correction is the pretest-posttest correlation divided by the reliability of the pretest.

This correction factor is ordinarily less than one, making posttest-treatment ordinarily less in absolute value than the pretest-treatment correlation. The correction factor for raw change score analysis is the standard deviation of the pretest divided by the standard deviation of the posttest. The correction factor for standardized change score analysis is equal to one.

To demonstrate the treatment-effect correlation measures of treatment effects in Table 11.1 take the covariance between the treatment with the dependent variable and simplify the expression. For analysis of covariance the covariance of the treatment with the dependent variable is

$$C(T, X_2 - b_{X_2X_1.T}X_1) = C(T, X_2) - b_{X_2X_1.T}C(T, X_1)$$

With some algebraic manipulation, the above reduces to

$$r_{TX_2} - r_{TX_1}r_{X_1X_2}$$

The preceding formula is the numerator of the formula for $r_{TX_2.X_1}$ and $\beta_{TX_2.X_1}$, that is, the partial correlation and standardized partial regression coefficient of X_2 and T controlling for X_1 .

The logic for analysis of covariance with reliability correction is the same as previously with the inclusion of the reliability of the pretest in the formula.

For raw change score analysis the covariance is

$$C(T, X_2 - X_1) = C(T, X_2) - C(T, X_1)$$

Dividing through by $s_{X_2} s_T$ and with some algebraic manipulation, the following equation is obtained:

$$C(T, X_2) - C(T, X_1) \frac{s_{X_2}(1 - r_{TX_2}^2)^{1/2}}{s_{X_1}(1 - r_{TX_1}^2)^{1/2}}$$

For standardized change score analysis, it can be shown that

$$r_{TX_2} - r_{TX_1} \frac{(1 - r_{TX_2}^2)^{1/2}}{(1 - r_{TX_1}^2)^{1/2}}$$

The preceding equals zero except in trivial cases only if $r_{TX_2} = r_{TX_1}$ equals zero. Campbell and Erlebacher (1970) have suggested a fifth method: covariance analysis with common factor coefficient correction. This method closely resembles analysis of covariance with reliability correction, the difference being that the regression coefficient is divided by the pretest-posttest correlation of X instead of the reliability of X_1 .

This correction yields the same correlational null hypothesis as that for standardized change score analysis since the pretest-posttest correlations cancel each other.

It should be clear from Table 11.1 that the measures of treatment effect of the four modes of analysis are different hypotheses except in highly trivial cases. Each of the four modes of analysis has been advocated as the method of analysis for the nonequivalent control group design by various authors. Other authors (e.g., Lord, 1967) have pointed out that it is paradoxical that different methods yield different conclusions. Cronbach and Furby (1970) state that treatments cannot be compared for this design. The literature on this design is, therefore, very confusing and not at all instructive to the practitioner.

SPECIAL CASES OF THE GENERAL MODEL

The validity of any mode of analysis depends on its match with the process of selection into groups. In the remaining part of this chapter, various special cases of the general model of selection are considered.

Each mode of analysis is appropriate for a given model of selection. These special cases of the general model yield overidentifying restrictions given no treatment effects. It will then be shown that these overidentifying restrictions derived from the model of selection match the null hypothesis of each of the four modes of analysis given in Table 11.1.

Three different types of selection processes are considered. The first type is selection based on the measured pretest. This is the only type of selection considered for which selection is controlled by the researcher. In this case either by design or accident the treatment is correlated with the entire pretest and, therefore, correlated with all the causes of the pretest: G , Z , and E . For this type of selection process the analysis of covariance is appropriate. The second type of selection is selection based on the true pretest. For this case the treatment is related to G and Z and not E , making the analysis of covariance with reliability correction appropriate. The third type of selection is based on group differences. Participants are assigned to the treatment because of demographic and social variables, G . Given that the effect of these demographic variables is stationary over time, standardized change score analysis is the appropriate mode of analysis. Finally discussed is the case in which the occasion of selection into treatment groups is not at the pretest.

SELECTION BASED ON THE PRETEST

At times it may be possible to control selection into treatment groups, but it is decided not to randomly assign participants to treatment conditions. One reason for not randomly assigning is that it

may not be fair for all types of participants to have the same probability of receiving the treatment, that is, certain persons (e.g., the injured or disadvantaged) are considered more deserving of the treatment. One strategy, called the regression discontinuity design by Thistlewaite and Campbell (1960), is to assign participants to the treatment on the basis of pretest.

Persons scoring above (below) a certain point would be given the treatment and those scoring below (above) or equal to that point would be the controls. (Since the pretest is measured on an interval scale and treatment is dichotomous, r_{X_1T} does not equal one but it is high.) Actually a pretest itself need not be used, but rather any measure of "deservingness." The treatment, then, is a function of the pretest (X_1) and random error (U):

$$T = kX_1 + fU \quad [11.7]$$

In this case assignment to the treatment groups is deliberately confounded with the pretest and, therefore, with the causes of the pretest while in the true experimental case, the treatment is deliberately uncorrelated with the pretest through randomization. But like the true experimental case, a specification has been gained by controlling the assignment to groups. In the population the treatment should correlate with the unmeasured causes of pretest (G , Z_1 , and E_1) to the degree to which they cause the pretest. Substituting Equation 11.1 for X_1 with Equation 11.7 yields

$$T = k(a_1G + b_1Z_1 + e_1E_1) + fU$$

Taking the covariance of G , Z_1 , and E_1 with the preceding equation yields a_1k , b_1k , and e_1k . Because one can also solve for the same covariances from Equation 11.3, it follows that

$$\begin{aligned} a_1k &= q \\ b_1k &= m \end{aligned}$$

$$e_1k = s$$

or alternatively $q/a_1 = m/b_1 = s/e_1 = k$. Given Equations 11.4 and 11.8 it follows that

$$\begin{aligned} r_{X1T} &= a_1(a_1k) + b_1(b_1k) + e_1(e_1k) \\ &= k(a_1^2 + b_1^2 + e_1^2) \end{aligned}$$

Since $a_1^2 + b_1^2 + e_1^2 = 1$ through standardization, it follows that

$$r_{X1T} = k \quad [11.9]$$

If no treatment effects are assumed ($t = 0$), given Equations 11.5 and 11.8 it follows that

$$\begin{aligned} r_{X2T} &= a_2(a_1k) + b_2j(b_1k) \\ &= k(a_1a_2 + b_1b_2j) \end{aligned}$$

Given no treatment effects and Equation 11.6, it follows that

$$r_{X2T} = kr_{X1X2}$$

Solving both Equations 11.9 and 11.10 for k , it then follows in the population that

$$\rho_{TX2} - \rho_{TX1}\rho_{X1X2} = 0$$

which is the null hypothesis for analysis of covariance in Table 11.1.

Thus, when participants are assigned to the treatment on the basis of the pretest, analysis of covariance is the appropriate mode of analysis. Rubin (1977) and Overall and Woodward (1977) have shown that much more complicated decision rules than regression discontinuity can be used to determine selection and covariance analysis can still be valid. However, the researcher must still control

selection into treatment groups. Special attention should be paid to the validity of covariance's assumptions of homogeneity and linearity of regression.

Lack of linearity can usually be remedied by including higher-order terms such as the pretest squared. Given the little or no overlap in the pretest distributions of the treated and untreated groups, violations of linearity could lead to erroneous conclusions and may be indistinguishable from a treatment by pretest interaction.

It seems to be an all too common occurrence that randomization of participants into treatment groups produces treatment differences even before the treatment is administered. Although one realizes that the probability of such pretest differences is 1 out of 20, given the conventional .05 level of significance, there is nonetheless the feeling of being persecuted by fate. It seems that the experiment has been doomed and there is no way to achieve valid inference.

Unhappy randomization does not mean cases of failure to randomize or cases in which there is randomization but a selected group of control and experimental participants fail to provide posttest data. It is a randomized experiment with a pretest difference between the experimentals and controls. Randomization has not failed, as it is sometimes mistakenly thought; only an unlikely type of event has occurred.

Valid inference is possible in a way similar to our discussion of the regression discontinuity design. If there is a pretest difference, the treatment is confounded with the causes of the pretest. The expected degree of this confounding with each cause is proportional to its causal effect on the pretest. For a randomized experiment with the pretest correlated with the treatment, the analysis of covariance is not only appropriate but necessary.

SELECTION BASED ON THE PRETEST TRUE
SCORE

The main concern here is not for when the researcher controls selection into treatment groups but for when this is not the case. Two types of subject selection are discussed: first, selection based on the pretest true score (G and Z) and, second, selection based only on group differences (G).

The model previously discussed assumes that the treatment is correlated with the errors of measurement in the pretest. This is generally implausible when selection is uncontrolled. If persons select themselves into programs, the variables that determine selection into the treatment are likely to be correlated with true ability and not chance performance on the pretest. The participants, not the experimenter, control selection making the treatment correlated with only the true causes of the pretest. These true scores are not actually measured but the causes of selection like motivation, expectation, and encouragement are correlated with true ability. As earlier, it is assumed that the ratio of the effect of group differences on the treatment to its effect on the pretest equals the ratio of the effect of individual differences on the treatment to its effect on the pretest; that is,

$$\frac{q}{a_1} = \frac{m}{b_1} = k \quad [11.11]$$

This assumption presumes that selection into treatment groups is on the basis of the true pretest, $a_1G + b_1Z_1$, and not on errors of measurement or any other function of G and Z . In the case of selection based on the measured pretest, a similar hypothesis was made but that hypothesis was justified by the design of the research. In this case Equation 11.11 is an assumption that must be justified by evidence from the selection process itself.

If it is assumed that $s = 0$ and Equation 11.11 holds, correlations 11.4 and 11.5 become

$$\begin{aligned} r_{X1T} &= k(a_1^2 + b_1^2) \\ r_{X2T} &= k(a_1a_2 + b_1b_2j) + t \end{aligned} \quad [11.12]$$

Because the reliability of the pretest ρ_{X1X1} is defined as $a_1^2 + b_1^2$, it follows that $r_{X1T} = k\rho_{X1X1}$. Also given no treatment effects and Equation 11.6, it follows that $r_{X2T} = k\rho_{X1X2}$. The analysis of covariance null hypothesis will not equal zero because

$$\begin{aligned} r_{TX2} - r_{TX1}r_{X1X2} &= kr_{X1X2} - kr_{X1X1}r_{X1X2} \\ &= kr_{X1X2}(1 - r_{X1X1}) \end{aligned}$$

Except in trivial cases, the preceding will only equal zero if the reliability of the pretest is perfect.

The reason for this bias is that the within groups regression coefficient is attenuated because the pretest is measured with error. The posttest should be regressed not on the measured pretest as in covariance analysis, but on the true pretest (see Chapter 5). It is because of this bias that both Lord (1960) and Porter and Chibucos (1974) have suggested a correction for the analysis of covariance. One can view this correction as correcting the regression coefficient for unreliability in the pretest. To make the reliability correction we must have an estimate of the reliability of the pretest. Assuming that there is an estimate of reliability and $s = 0$ and Equation 11.11 holds, the analysis of covariance with reliability correction yield unbiased estimates of treatment effects.

Porter and Chibucos (1974) discuss in detail computational methods for true score correction. Regress the pretest on the treatment variable to obtain b_{X1T} . Now compute

$$X^d = b_{X1T}T + \rho_1(X_1 - b_{X1T}T)$$

where ρ_1 is the within-treatment group reliability of the pretest and X^d is the adjusted pretest. The next step is to regress X_2 on X^d and T , and

the regression coefficient for T is unbiased if s equals zero and Equation 11.11 holds.

A difficulty with the reliability correction procedure is the necessity of having a reliability estimate. The inclusion of any ad hoc estimate, for example, internal consistency, into the correction formula almost certainly increases the standard error of the estimate. Ideally the reliability estimation procedure should be part of a general model. Following Lord (1960) consider, for example, a parallel measure of X_1 , say Y . Let

$$Y = a_1G + b_1Z_1 + e_3E_3$$

where E_3 is uncorrelated with all other unmeasured variables. If it is assumed Equation 11.11 holds and that $t = s = 0$, then it follows that

$$r_{X_1Y} = a_1a_3 + b_1b_3 \quad [11.13]$$

$$r_{X_2Y} = a_2a_3 + b_2b_3j \quad [11.14]$$

Given Equations 11.12, 11.13, and 11.14, it follows that

$$\rho_{X_1X_1} = \frac{\rho_{TX_1}\rho_{X_1Y}}{\rho_{TY}}$$

Substituting the preceding formula for reliability into the reliability correction formula in Table 11.1 yields the following null hypothesis:

$$\rho_{X_2T} = \frac{\rho_{TY}\rho_{X_1X_2}}{\rho_{YX_1}}$$

or equivalently in vanishing tetrad form:

$$\rho_{X_2T}\rho_{X_1Y} - \rho_{TY}\rho_{X_1X_2} = 0$$

The vanishing tetrad can be tested by the null hypothesis of a zero second canonical correlation between variables X_1 and T and variables X_2 and Y as in Chapter 7. Note that it has been assumed that the true score of X_1 was used to select persons into treatment groups, not the true score of Y .

However, if both true scores were identical, it follows that $a_1/a_3 = b_1/b_3$ and that

$$\rho_{X_1X_2}\rho_{YT} - \rho_{X_1T}\rho_{YX_2} = 0$$

SELECTION NOT AT THE PRETEST

It has been assumed that the occasion for selection into treatment is at the pretest. For most programs the occasion for selection is not so well defined. Treated and control participants may drop out of the program or move out of the area, some controls may enter the program, unsuccessful or unhappy treated participants may not show up at the posttest, and so on. Sometimes the "pretest" takes place well before the treatment begins; for example, test scores of the previous year are used as a pretest for a remedial program for the current year. For many real world programs the occasion of selection into the program is not identical with the pretest.

For assignment based on the true score, the correlation of the treatment with the effect variable is not the highest at the pretest. This would substantially bias analysis of covariance with reliability correction if selection does not occur at the pretest. Imagine that the researcher does not measure X_1 but uses X_0 as a "pretest." For such a case the analysis of covariance with reliability correction would ordinarily be biased. For instance, if parameters a , b , and j remained stationary over time, then the treatment-effect correlations are equal to each other, given no treatment effects. Thus, if selection occurs midway between the pretest and posttest, or "averages out" midway, and if the researcher can assume stationarity, then standardized change

score analysis is the appropriate form of analysis. If selection is based on group differences, the occasion of selection is irrelevant since group differences are perfectly stable.

SELECTION ON THE BASIS OF GROUP DIFFERENCES

For most social programs, assignment to the treatment is not based on some psychological individual difference, that is, true score, but on some sociological, demographic, or social psychological characteristic. This "sociological" selection is brought about in a variety of ways:

1. It may be a matter of policy or legislation that treatment is available to a particular social group, for example, persons living in particular census tracts. It may be virtually impossible to find some members of that social group who did not receive the treatment.
2. Some treatments are administered to members of an entire organization, for example, school system, and members of the treated organization are compared with another organization.
3. To receive a selective treatment a person or their sponsor must be highly motivated, or have political connections and organizational "savvy." These volunteers differ systematically from nonvolunteers on a number of characteristics (Rosenthal & Rosnow, 1975).
4. Sometimes the treatment either is a sociological or demographic variable or hopelessly confounded with one. Examples of this are a study on the effects of dropping out of high school and testing for differences in socialization between males and females.

The reader can probably also conceive other patterns of sociological selection. Suffice it to say, it is a rather common form of selection into social programs.

In terms of the general model, selection based on group differences implies that $s = m = 0$. However, treatment effects are still not identified if they exist, and consequently there is no overidentifying restriction given the absence of treatment effects. To be able to estimate treatment effects one must take additional assumptions or add on additional measures. Following Kenny (1975a), it is shown in this chapter that the fan spread hypothesis can allow for identification.

One way to gain an overidentifying restriction is to assume some form of stationarity, that is, assume that the effect of group differences is the same at the pre- and posttest. Campbell (1967) has argued for just such a model with what he called the "fan spread hypothesis." The hypothesis is that associated with the mean differences between groups is a difference in maturation: Those with the higher mean mature at a greater rate than those with the lower mean. Campbell calls this the "interaction of selection and maturation" and has used this interaction as an argument against raw change score analysis. Since the mean difference between groups is widening over time, change score analysis indicates only the more rapid rate of maturation of the initially higher group. The fan spread hypothesis is that accompanying increasing mean differences is increasing variability within groups. In its strictest form the fan spread hypothesis is that the difference between group means relative to pooled standard deviation within groups is constant over time.

The rationale for the fan spread hypothesis is that the different groups are members of different populations living in different environments. The different environments create and maintain different levels of performance and different rates of growth. Given that growth is a cumulative process, variability increases over time. The groups would eventually asymptote at different levels.

Formally, the model for selection on the basis of group differences is

$$T = qG + fU$$

Fan spread and no treatment effects implies that $a_1 = a_2$, making $r_{X1T} = r_{X2T}$, that is, equality of the treatment-effect correlations.

Recall that fan spread is only a hypothesis. It may not hold in any particular data set. Ideally fan spread should be tested internally by the data itself (Kenny & Cohen, 1980). Alternative specifications other than fan spread are also possible (Bryk & Weisberg, 1977; Linn & Werts, 1977).

If the variance within treatment groups of the pretest and posttest is stationary over time, then the results of standardized change score analysis and raw change score analysis converge. Thus, for cases in which variance is stationary, raw change score analysis is to be preferred since standardization is needed only to stabilize variance over time. Power transformations may also be applied to stabilize within treatment variance, in a way akin to meeting the homogeneity of variance assumptions of analysis of variance.

It is not advisable to actually standardize the pretest and posttest and perform a standardized change score analysis as was implied in Kenny (1975a), since the standard deviation is itself affected by any treatment effect. Rather, one should adjust the pretest by $s_{X2.T}/s_{X1.T}$ and then perform a change score analysis. As defined earlier, $s_{X2.T}$ is the standard deviation of the posttest with the variance of the treatment partialled out, that is, in the case of discrete treatment groups, the pooled within group treatment standard deviations. A difficulty with using this adjustment is that sampling variation is surely introduced because $s_{X2.T}/s_{X1.T}$ is a sample statistic, not a population value. Nonetheless, using $X_2 - (s_{X2.T}/s_{X1.T})X_1$ as the dependent variable yields an interpretable, simple metric in the posttest.

One might also simply compare treatment-effect correlations over time. This has the disadvantage that an interpretable measure of

treatment effects must then be derived. It does, however, have the advantage that the summary inferential statistic does not vary according to which score is taken as the pretest and which is the posttest. Note that the statistical test results differ if $X_2 - (s_{X2.T}/s_{X1.T})X_1$ is the dependent variable or if $X_1 - (s_{Y1.T}/s_{X2.T})X_2$ is the dependent variable.

The test of statistical significance between two treatment-effect correlations is the Hotelling-Williams test (Williams, 1959). It is

$$t(N-3) = \frac{(r_{X2T} - r_{X1T})[(N-1)(1 + r_{X1X2})]^{1/2}}{[2d(N-1)/(N-3) + (r_{X2T} + r_{X1T})^2 + (1 - r_{X1X2})^3/4]^{1/2}}$$

where d equals

$$1 - r_{X1T}^2 - r_{X2T}^2 - r_{X1X2}^2 + 2r_{X1T}r_{X2T}r_{X1X2}$$

and N is the sample size.

It is more likely that mean differences are proportional to the true standard deviations as opposed to the measured standard deviations. Thus the index of response should be

$$\frac{s_{X2}\rho_{X2X2}^{1/2}}{s_{X1}\rho_{X1X1}^{1/2}}$$

where ρ_{X2X2} is the reliability of the posttest and ρ_{X1X1} the reliability of the pretest. All standard deviations and reliabilities are defined within treatments. Note that if s_{X2}/s_{X1} is used as the index of response it must then be assumed that $\rho_{X1X1} = \rho_{X2X2}$.

Before moving on to consider an example it should be pointed out that all the statistical techniques can be applied in a multivariate fashion. That is, if there are background variables like age, sex, and ethnicity, they can be entered into the analysis since all four of analysis techniques can be viewed as application of multiple regression. For analysis of covariance the posttest is the criterion and the pretest, treatment, and background variables are the predictors. For

the analysis of covariance with reliability correction the pretest must be adjusted before it is entered into the prediction equation. The pretest is regressed on the treatment and background variables to compute X_1' or the predicted pretest. Then $\rho(X_1 - X_1')$ is entered into the regression equation (where ρ is the reliability of $X_1 - X_1'$) along with the treatment and background variables. For raw change score analysis the criterion is simply raw change and the treatment and background variables are the predictor variables. For standardized change the criterion is the posttest minus the pretest times the standard deviation of the posttest with the treatment and background variables partialled out divided by the pretest standard deviation again with the treatment and background variables partialled out. The predictor variables are again the treatment and background variables.

The treatment need not be conceived of as a simple dichotomy. It may be a multilevel variable, and multiple dummy variables are then formed. Moreover, the treatment variable may be measured at interval level of measurement, as in the number of treatment hours received and treatment-background variable interactions may be entered.

To compute treatment-effect correlations, one needs to compute partial treatment-effect correlations. That is the background variables, and other treatment variables should be partialled out.

EXAMPLE

Steven M. Director (1974) reanalyzed an evaluation of a manpower training program. The interesting aspect of the design is that there are two pretreatment measures of yearly income, the measures being separated by one year. The analyses in this example use the two pretest measures as the pretest and posttest. Since the treatment had not been administered, the analysis should indicate no effect.

In Table 11.2 are the basic summary statistics. Note that those who received the manpower training started lower than the control

group and remained lower. In fact the gap actually widened by \$134.00. The pretest-posttest correlation between income is .808.

In Table 11.3 are the indices of response for four methods of analysis. There is no reliability measure for the pretest, but for illustrative purposes it has been set to a .90 value. Note that all four estimates of the nonexistent treatment are negative, indicating that the phantom program reduced income. Since the program is compensatory, the larger the index of response, the less negative the estimate of treatment effect.

Standardized change gives the least negative estimate of effect and covariance the largest. Only the covariance estimate reaches statistical significance. Although this example makes covariance look bad, I could have just as easily chosen an example that makes standardized change analysis look bad.

Table 11.2. Means and Standard Deviations in Dollars of the Manpower Training Evaluation

Group	Year	
	1966	1967
Experimental (N = 181)	2024 (2294) ^a	2744 (2719)
Control (N = 1827)	3140 (2574)	3994 (2767)

^aStandard deviation.

A critical issue has been the direction of bias in analysis of covariance when it is misapplied. Campbell and Erlebacher (1970) have stated that covariance tends to underadjust, that is, it insufficiently controls for initial differences. They argued that this was especially damaging for compensatory programs. For a compensatory program, the difference on the covariate favors the advantaged group.

If covariance analysis underadjusts, then it would tend to make compensatory programs look bad as it did in the previous example. A program that had no effect would be made to look harmful and the effects of a slightly beneficial program might be wiped out. Cronbach and Furby (1970), however, state that the effect of misapplying covariance analysis is not so certain. In some cases it will underadjust and in other cases it will overadjust. Cronbach, Rogosa, Floden, and Price (1978) give an explicit mathematical formalization of the cases in which analysis of covariance will under- and overadjust, but the mathematics of this very important paper go well beyond this text.

to various issues: unreliability in the covariate, growth rates over time, stationarity, and most importantly the process of selection.

Table 11.3. Estimates in Dollars of the Effects of the Manpower Training Program

Test	Index of Response	"Treatment Effect"
Analysis of covariance	.872	-\$276
Covariance with reliability correction	.969	-\$169
Raw change score analysis	1.000	-\$134
Standardized change score analysis	1.084	-\$40

Treatment effects from the nonequivalent control group design does not imply a single method of data analysis. Campbell and Stanley (1963) tend to give the mistaken impression that there is an ideal statistical analysis for each experimental and quasi-experimental design. It is true that the design can introduce certain specifications, but I can think of no design that can always be analyzed by a single statistical method.

The validity of each of the statistical methods discussed in this chapter depends on the causal model it presumes. One has no guarantee that any particular model is true. It is almost certainly the case that none of the models is exactly true. Special care must be given

12

CROSS-LAGGED
PANEL CORRELATION

The models discussed in this chapter were developed exclusively for longitudinal data. Whereas the models discussed in the previous chapters that used either multiple regression or factor analysis to estimate structural parameters can be used in a longitudinal context, they were primarily developed to analyze cross-sectional data. Cross-lagged panel correlation (CLPC) was first suggested by Campbell (1963) and has been extensively reviewed by Kenny (1973, 1975b). In its simplest form CLPC involves two constructs, say X and Y , measured at two points in time, say 1 and 2. There are then four variables X_1 , X_2 , Y_1 , and Y_2 and six correlations: two synchronous correlations, $r_{X_1Y_1}$ and $r_{X_2Y_2}$, two autocorrelations, $r_{X_1X_2}$ and $r_{Y_1Y_2}$, and two cross-lagged correlations, $r_{X_1Y_2}$ and $r_{X_2Y_1}$. As the name suggests, CLPC is the comparison of the cross-lagged correlations, which can be expressed as a cross-lagged differential: $r_{X_1Y_2}$ minus $r_{X_2Y_1}$. Campbell's original suggestion is that if X caused Y , then the cross-lagged differential would be positive and if Y caused X the differential would be negative. (Unless otherwise stated, all correlations in this chapter are assumed to be positive.) Campbell and his students (Kenny, 1973, 1975b; Rickard, 1972; Rozelle & Campbell, 1969) have elaborated the method. Another tradition (Bohrnstedt, 1969; Duncan, 1969; Goldberger, 1971; Heise, 1970; Pelz & Andrews, 1964) has suggested replacing CLPC with multiple regression or partial correlation analysis. Still another approach to panel data has been the application of factor

analysis (Jöreskog & Sörbom, 1976). This approach usually involves oblique solutions in which factors are time specific.

The fourth and oldest formal approach to panel data is the sixteenfold table described by Lazarsfeld (1972; Yee & Gage, 1968). This approach involves nominal data, usually dichotomous. Goodman (1973) has applied log-linear analysis to the sixteen-fold table.

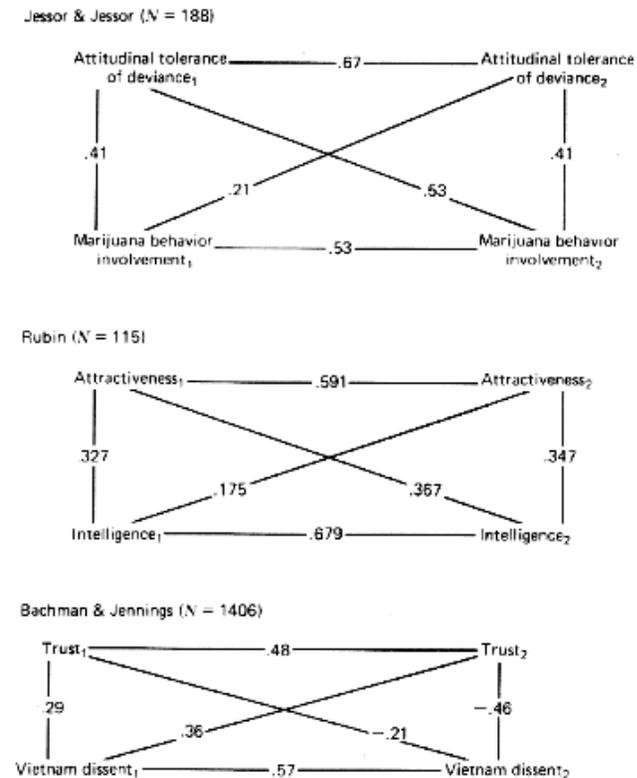


Figure 12.1 Crosslag examples.

In Figure 12.1 there are three highly atypical but very interesting examples of CLPC. As someone who attended college in the late 1960s, I chose as examples, marijuana use, women's consciousness, and Vietnam protest. The first example is taken from

Jessor and Jessor (1977) and the two variables are marijuana behavior involvement and attitudinal tolerance of deviance. The marijuana variable is a four-item scale of the amount of marijuana usage and the attitudinal tolerance of deviance is a 26-item measure of the tolerance of “conventional” deviant behaviors like lying, stealing, and aggression. The sample consists of 188 males of high school age and the lag between waves is one year. The crosslags in Figure 12.1 suggest that tolerance for deviance causes an increase in marijuana use.

The second example is taken from unpublished data of Z. Rubin. The subjects are 115 college-age female members of dating couples. The two measures are the woman’s perception of her physical attractiveness and her perception of her intelligence. The lag between measurements is 15 months. The crosslags indicate that the opposite of the “dumb blond” hypothesis occurs. Instead they point to perceived attractiveness causing an increase in perceived intelligence.

The final example is taken from Bachman and Jennings (1975). The two variables are trust in the government and Vietnam dissent. The lag is one year and the sample consists of 1406 male late-adolescents. The crosslags indicate that dissent against the war in Vietnam caused a decrease in trust. Perhaps the lack of trust brought about by Vietnam laid the groundwork for public acceptance of the crimes of Watergate.

LOGIC OF CLPC

CLPC is a quasi-experimental design (Campbell & Stanley, 1963). At the heart of quasi-experimental inference is the attempt to rule out plausible alternative explanations of a causal effect, that is, biases or artifacts. In correlational analysis the chief alternative explanation of any causal effect is spuriousness. Any statistical relationship, be it simple correlation, partial correlation, or regression coefficient, can be attributed not to causality but to spuriousness. As discussed in Chapter 1, Suppes (1970) has even defined a causal relationship negatively as a nonspurious relationship. Ideally these

spurious causes should be measured and controlled in the nonexperimental case.

Randomized experiments control for spuriousness by random assignment to treatment conditions. As reviewed in Chapter 10, random assignment guarantees that there is no systematic relationship in the population between the treatment and the dependent variable given the null hypothesis of no treatment effect. Thus, any relationship between the treatment and the dependent variable that cannot be plausibly explained by chance is attributed to the causal effects of the treatment. Although random assignment permits researchers to make strong causal inferences, it brings with it some potentially burdensome methodological limitations. True experimentation rules out of consideration as independent variables any variable that cannot be manipulated and then randomly assigned. Many important variables, usually individual differences, are not subject to experimental manipulation as simply as the intensity of light. Returning to the examples in Figure 12.1, it is clearly difficult to manipulate variables like attitudinal tolerance of deviance, perceived intelligence, or Vietnam dissent. Researchers spend considerable time theorizing about intelligence, attitude change, extroversion-introversion, and evoked potential, but since these variables are attached to rather than assigned to the organism, they are studied more often as dependent rather than independent variables. It might be argued that the stimulus-response or input-output orientation within psychology may reflect the limitation. that experimental treatments be manipulatable variables. The requirement of manipulating the independent variable also prevents researchers from examining certain variables because of ethical considerations. For instance, malnutrition has been proposed as an important cause of children's cognitive ability, but it would be highly unethical to randomly assign children to levels of malnutrition. Thus, for practical and ethical reasons it is not always possible to use random assignment to control for spuriousness.

The null hypothesis of CLPC tested by equality of the crosslags is that the relationship between X and Y is due to an

unmeasured third variable and not causation. Before causal models are entertained, the third variable explanation should be ruled out. The logic of true experimentation is similar. Before accepting that the treatment has an effect, the null hypothesis of sampling error must be ruled out. Given the inapplicability of true experimentation in numerous areas, CLPC can be used to test for spuriousness.

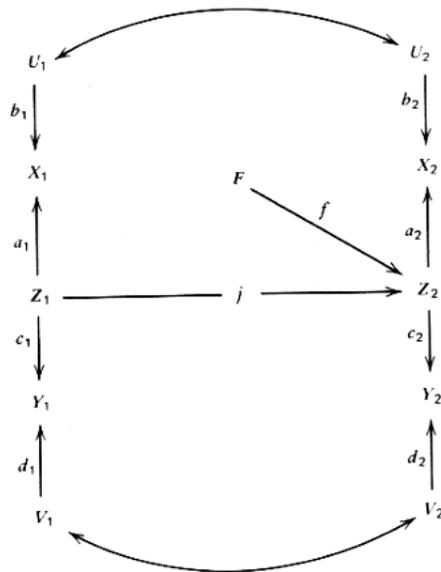


Figure 12.2 The null model for CLPC.

The null model for CLPC is illustrated in Figure 12.2. A third variable, Z_1 , causes X_1 and Y_1 simultaneously. (Actually Z may cause X and Y with a lag, and the lag would be the same for both X and Y .) Over time Z changes and at time 2, Z_2 causes X_2 and Y_2 . Given a model of spuriousness in Figure 12.2 the structural equations for X and Y are as follows:

$$X_1 = a_1Z_1 + b_1U_1 \quad [12.1]$$

$$X_2 = a_2Z_2 + b_2U_2 \quad [12.2]$$

$$Y_1 = c_1Z_1 + d_1V_1 \quad [12.3]$$

$$Y_2 = c_2Z_2 + d_2V_2 \quad [12.4]$$

where U , V , and Z are all uncorrelated with each other but are autocorrelated. Each unmeasured variable takes on a different role. The variable Z is the unmeasured variable that brings about the relationship between X and Y and is called the *third variable*. The variable U includes all causes of X besides Z . It includes true causes as well as errors of measurement. The variable V plays the same role for Y . It is assumed that U and V are correlated over time. Although it is not necessary for much of what is to follow, it is assumed that Z is first-order autoregressive (see Chapter 9). The model in Figure 12.2 is clearly underidentified. There are seven free parameters and only six correlations. Interestingly enough, the autocorrelation of Z or $r_{Z_1Z_2}$ is identified by the square root of the following:

$$\frac{r_{X_1Y_2}r_{X_2Y_1}}{r_{X_1Y_1}r_{X_2Y_2}}$$

When the square root is taken, the sign should be taken from the cross-lagged correlations. Given that it is a correlation, its value should be between plus or minus one. Because none of the other parameters are identified, how can spuriousness be tested? The key assumption of CLPC is stationarity. This assumption capitalizes on the fact that the same variables are measured at each point in time. By stationarity it is meant that a variable's causal structure does not change over time; that is, its structural equation is the same at both points in time. It is important to distinguish stationarity from stability. Stability refers to unchanging levels of a variable over time, whereas stationarity refers to an unchanging causal structure. For the two-wave, two-variable case stationarity implies

$$a_1 = a_2$$

$$c_1 = c_2$$

Given stationarity there are two overidentifying restrictions: equality of the synchronous correlations, $\rho_{X_1Y_1} = \rho_{X_2Y_2}$, and equality of the crosslagged correlations, $\rho_{X_1Y_2} = \rho_{X_2Y_1}$. The strategy of CLPC is to examine the synchronous correlations to test for stationarity, and if stationarity is satisfied the cross-lagged correlations can be used to test for spuriousness.

Unfortunately, even if the synchronous correlations are equal, stationarity may not be satisfied. For example, if a_1/c_2 were to equal a_2/c_1 , then synchronous correlations would be equal whereas the crosslags would be unequal.

The opposite can also happen: if $a_1/c_1 = a_2/c_2$, the crosslags are equal whereas the synchronous correlations are unequal. Thus given the model in Figure 12.2, equal synchronous correlations is neither a necessary nor sufficient condition for equal crosslags. Quite obviously, it is highly questionable whether crosslags can be used to test for spuriousness in the simple two-wave, two-variable case: First, equal synchronous correlations are needed to demonstrate stationarity. Although it is rather common for the synchronous correlations to be equal, a fair percentage of the time they are not. Second, even if the synchronous correlations are equal, stationarity may still not hold. In such a case unequal crosslags do not indicate a violation of spuriousness but unstationarity.

Besides stationarity, the second key assumption of CLPC is synchronicity: Both variables must be measured at the same point in time. To see why this assumption is important, examine the model:

$$\begin{aligned} X_t &= aZ_t + bU_t \\ Y_t &= cZ_t + dU_t \\ Z_t &= jZ_{t-1} + fF_t \end{aligned}$$

where U , V , and Z are autocorrelated but not cross-correlated. The synchronous correlation is then ac and the crosslag of $r_{X_t, Y_{t+k}}$ is $acj^{|k|}$. Note that the synchronous correlation is a special case of the crosslag formula where $k = 0$. Now if X is measured at times 1 and 3 and Y at

times 2 and 4, but if X_1 and Y_2 are considered wave 1 and X_3 and Y_4 are considered wave 2, then the "crosslags" would not be equal because

$$\begin{aligned} r_{X_1Y_4} &= acj^3 \\ r_{X_3Y_2} &= acj \end{aligned}$$

These unequal crosslags would not be due to a violation of spuriousness or even stationarity but due to a violation of synchronicity. Not surprisingly, the variables measured closer together in time correlate higher than those measured further apart in time. Synchronicity is then an important assumption of CLPC.

It is instructive to examine the formula for the crosslag in more detail: $acj^{|k|}$. Note that if j is positive, as it would be expected to be, the correlation between the two variables reaches an absolute maximum when the lag is zero. As the lag (i.e., k) increases, the correlation decreases in absolute value. Campbell has called this phenomenon temporal attenuation (Rozelle & Campbell, 1969) or temporal erosion (Campbell, 1971). Temporal erosion is an important empirical fact to remember in examining longitudinal data. The intuition behind temporal erosion is that two variables measured closer together in time should, *ceteris paribus*, be more highly correlated. This "fact" may not hold when either spuriousness or stationarity do not obtain. However, this is the very reason that temporal erosion has been emphasized by Campbell. Relationships that do not erode over time may indicate nonspurious effects. Note that given spuriousness and stationarity the crosslags must be less than or equal to in absolute value the synchronous correlations (Cook & Campbell, 1976). What governs this decrease in correlation is the rate of change of the third variable, Z .

At first glance synchronicity would seem to be an easy assumption to satisfy. Panel studies are defined as the replication at two different points in time of a cross-sectional survey on the same set of persons. However, because of the problems of retrospection and aggregation, synchronicity may not be so easily satisfied. Some variables in panel studies ask subjects to recall behaviors, attitudes, or

experiences of the past. These questions either directly or indirectly ask subjects to retrospect. In some sense the data may not be generated at the time of measurement but at some time prior to measurement.

Another problem for synchronicity is aggregation. Many variables are aggregated or averaged over time. A good example of a measure of this type is grade point average. If grade point average is to be taken as a measure of ability, at what time point does it measure ability? It is actually an aggregation of performances evaluated by a teacher. The aggregation problem is well known in the econometrics literature where many of the important economic variables are aggregated across time, for example, gross national product and unemployment rates.

Thus both a lack of synchronicity and a lack of stationarity are potential explanations of a difference between cross-lagged correlations. If the model in Figure 12.2 is correct, then both stationarity and synchronicity together would imply equal crosslags. The null hypothesis that the crosslagged differential is zero is then a test of spuriousness.

What if the crosslagged differential is nonzero? Asymmetrical crosslags may indicate a causal effect; more generally they indicate that there is a factor that causes one of the measured variables and then causes the other measured variable at a later point in time. This factor, called the causal factor, is one of many factors that make up the causal variable. Saying "X causes Y" is shorthand for "something in X later causes Y." It need not be the case that the measure of X is valid or that the causal factor is the same as the true score of X. Although this problem of verbally explaining a causal effect is also present in randomized experiments, it is not as severe. One knows from an experiment that X causes Y but the experiment does not necessarily tell what in X causes Y. Different theoretical perspectives focus on different aspects of an experimental treatment to explain the same causal effect. The problem of interpreting cross-lagged differences centers on the construct validity of measures just as it does in

experimental research. The more valid, reliable, and unidimensional the measure, the more straightforward is the interpretation.

To illustrate the difficulties in interpreting cross-lagged differences, consider the marijuana example in Figure 12.1. A naive interpretation would say that tolerance of deviance causes marijuana use. A more precise interpretation would be that some variable that causes tolerance for deviance later causes marijuana use. Perhaps there is a developmental sequence in which there is first tolerance for deviance and then marijuana use. Although lightning may precede thunder with a time lag, one would not argue that lightning causes thunder.

An alternative model to that in Figure 12.2 has S cause X simultaneously and Y with a lag of one unit of time. (Alternatively, S causes X with a lag of k units and Y with a lag of k + 1 units.) The equations then are

$$\begin{aligned} X_t &= a_t Z_t + g_t S_t + b_t U_t \\ Y_t &= c_t Z_t + h_t S_t + d_t V_t \end{aligned}$$

In general the correlation between X and Y is

$$r_{X_t, Y_{t+k}} = a_t c_{t+k} r_{Z_t, Z_{t+k}} + g_t h_{t+k} r_{S_t, S_{t+k-1}}$$

For the special case in which there are two waves at times 1 and 2, the synchronous correlations are

$$\begin{aligned} r_{X_1 Y_1} &= a_1 c_1 + g_1 h_1 r_{S_0 S_1} \\ r_{X_2 Y_2} &= a_2 c_2 + g_2 h_2 r_{S_1 S_2} \end{aligned}$$

and the crosslags are

$$\begin{aligned} r_{X_1 Y_2} &= a_1 c_2 r_{Z_1 Z_2} + g_1 h_2 \\ r_{X_2 Y_1} &= a_2 c_1 r_{Z_1 Z_2} + g_2 h_1 r_{S_0 S_2} \end{aligned}$$

If stationarity were assumed, $a_1 = a_2$, $c_1 = c_2$, $g_1 = g_2$, and $h_1 = h_2$, then except for trivial cases the crosslags would be unequal. Note that this more general model diverges from the result of the model in Figure 12.2. The difference between crosslags assuming stationarity is

$$(1 - r_{S0S2})gh \quad [12.5]$$

If it is now assumed that S behaves with a stationary first-order autoregressive process, it then follows that the synchronous correlations are equal and that it is possible for a crosslag to be larger than a synchronous correlation in absolute value. Such a result is not possible for the model in Figure 12.2.

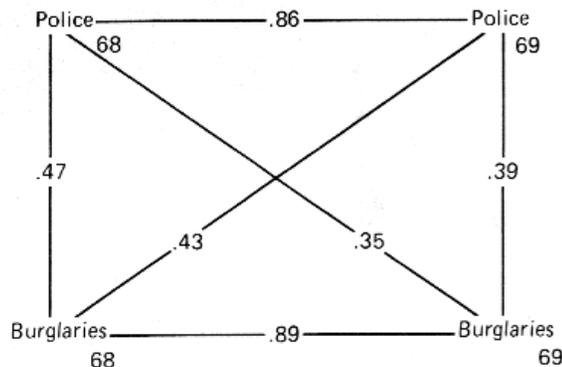


Figure 12.3 *Confounded hypotheses: correlation of number of police and number of burglaries per capita measure in 1968 and 1969 in 724 cities.*

Rozelle and Campbell (1969) and Yee and Gage (1968) have pointed out a difficulty in interpreting cross-lagged differences—competing, confounded pairs of hypotheses. There are two sources of a causal effect, X and Y , and two directions of that effect, positive and negative, making a total of four possible hypotheses. In terms of the previously elaborated causal model the source refers to which variable S causes first and the direction refers to whether g and h have the same

or different sign. Finding $r_{X1Y2} > r_{X2Y1}$ is consistent with both X causing an increase in Y and Y causing a decrease in X . Finding $r_{X1Y2} < r_{X2Y1}$ is consistent with both Y causing an increase in X and X causing a decrease in Y . To illustrate the difficulties of confounded hypotheses consider the example taken from Kidder, Kidder, and Snyderman (1974) in Figure 12.3. The variables are the number of burglaries and the number of police for 724 United States cities. At first glance it appears that burglaries cause an increase in the number of police. An alternative "law and order" explanation is that the number of police causes a decrease in the number of burglaries. Both hypotheses are equally plausible. The data are not consistent with two other hypotheses: police increase burglaries or burglaries decrease the number of police. These later hypotheses are not ruled out but their effects if they exist, are overwhelmed by the effects of one or both of the two former hypotheses.

Rozelle and Campbell suggested a no-cause baseline be computed to test both of the confounded hypotheses. Their procedure is as follows:

1. Compute the test-retest correlations of both variables and correct them for attenuation.
2. Average these two correlations to obtain a measure of the stability.
3. Multiply the estimate by the average of the two synchronous correlations.
4. The resulting value is a no-cause baseline to which both crosslags can be compared.

The implicit logic of the no-cause baseline is that given temporal erosion, the crosslags should be less than the synchronous correlations by some factor. Given stationarity and spuriousness, the crosslags should equal the synchronous correlation times the autocorrelation of Z . Rozelle and Campbell assume that the autocorrelation can be estimated from the autocorrelations of X and Y . Unfortunately there are

two difficulties with the Rozelle and Campbell baseline. First, it requires that the researcher have estimates of each variable's reliability because each autocorrelation must first be corrected for attenuation. Second, and more problematic, is the hidden assumption that all the nonerrorful causes of X and Y change at the same rate over time, that is, have the same autocorrelation. This assumption is called homogeneous stability. Evidence consistent with this assumption is that the two unattenuated autocorrelations are equal. Given the necessity of reliability estimates and homogeneous stability assumption, it would seem that the Rozelle and Campbell baseline is of limited practical use for longitudinal studies.

Although the sign of the synchronous correlations is neither a necessary nor sufficient condition for the direction of the causal effect, it is, nonetheless, suggestive of its direction. If the synchronous correlations are positive, they are supportive of X causing increases in Y or Y causing increases in X . Negative synchronous correlations indicate decreases. Moreover, occasionally the researcher may know the source of causation and the only empirical issue is the direction, or the direction is known and the only empirical issue is the source. In this way some of the confounded hypotheses can be ruled out a priori.

Given homogeneous stability the crosslags should always be smaller in absolute value than the synchronous correlations given spuriousness, stationarity, and synchronicity. Thus a crosslag larger than the synchronous correlations (assumed to be equal given stationarity) is indicative of a causal effect. It should be made clear that if X causes increases in Y and homogeneous stability is the case, then the crosslag from X to Y need not necessarily be larger than the synchronous correlations since both instability of spurious causes and misspecified causal lag would tend to make the crosslag smaller than the synchronous correlations.

NONSIGNIFICANT DIFFERENCES

What does an insignificant difference between the cross-lagged correlations indicate? Strictly speaking one should not accept the null hypothesis of spuriousness, that is, the hypothesis that the variables do not cause each other but are cosymptoms of some set of common causes. There are several alternative explanations. First, it may be that both X and Y equally cause each other in a positive feedback loop making the crosslags equal. Without a no-cause baseline such a model cannot be distinguished from spuriousness. Second, it may be that X causes Y or vice versa, but the magnitude of the effect is too small to be detected. In my experience, it is very difficult to obtain statistically significant differences between cross-lagged correlations even when the sample size is moderate, 75 to 300. The cross-lagged differential depends on the stability of the causal factor, r_{SOS2} from Equation 12.5. The more stable this factor is the smaller the differential. In the limiting case in which the factor does not change at all, the differential is zero. Cross-lagged analysis is, therefore, inappropriate for examining the causal effect of variables that do not change over time. For these variables their effects might best be diagnosed using other quasi-experimental models (see Chapter 11). (These models actually identify causal effects through unstationarity, that is, as increases in synchronous correlations.) Large cross-lagged differences are also difficult to obtain because the measured lag may not correspond to the causal lag. Normally the lag between measurements is chosen because of convenience, not theory, since theory rarely specifies the exact length of the causal lag. Moreover, for most social science theories it is doubtful that there is a single lag interval. Rather, the lag is variable across time or, as the econometricians call it, a distributed lag. Finally CLPC is most appropriate for the analysis of variables that have moderate or large correlations. My own experience is that the analysis of variables with low correlations (less than .3), yields disappointing and confusing results. In sum, CLPC requires at least moderate sample sizes, variables that change, lagged effects, and at least moderate

synchronous correlations. To obtain such moderate correlations the instruments must be reliable. Never attempt a crosslagged analysis on measures with doubtful reliability.

Given the low power of CLPC the researcher should design the longitudinal study to include many replications. Ideally a cross-lagged difference should replicate across

1. Different time lags.
2. Different groups of subjects.
3. Different operationalizations of the same construct.

For instance, most of the causal effects in Crano, Kenny, and Campbell's (1972) study of intelligence and achievement can be summarized as abstract skills causing concrete skills. In one of the best empirical applications of cross-lagged analysis Calsyn (1973) demonstrates all three of the preceding types of replications to show that academic achievement causes academic self-concept.

SIGNIFICANCE TESTS

The hypotheses tested in a cross-lagged analysis are, first, the equality of synchronous correlations to test for stationarity and, second, the equality of crosslags to test for spuriousness. One cannot use Fisher's z transformation (McNemar, 1969, pp. 157-158) to test for the significance of the differences between these correlations since the correlations are correlated. One can, however, use a rather bulky but easily programmable test cited by Peters and Van Voorhis (1940) and attributed to Pearson and Filon.

Because the formula is not easily accessible it is reproduced here. Let 1, 2, 3, and 4 be variables, N be sample size, and

$$k = (r_{12} - r_{24}r_{14})(r_{34} - r_{24}r_{23}) + (r_{13} - r_{12}r_{23})(r_{24} - r_{12}r_{14}) \\ + (r_{12} - r_{13}r_{23})(r_{34} - r_{13}r_{14}) + (r_{13} - r_{14}r_{34})(r_{24} - r_{34}r_{23})$$

The following then has approximately a standard normal distribution:

$$Z = \frac{N^{1/2}(r_{14} - r_{23})}{[(1 - r_{14})^2 + (1 - r_{23})^2 - k]^{1/2}}$$

STATIONARITY REVISITED

The pivotal assumption on which cross-lagged analysis rests is stationarity. As demonstrated later, the assumption can be somewhat relaxed in the multivariate case; nonetheless it is still critical. Put in words, stationarity implies that a variable has the same proportions of "ingredients" at both points in time. More formally, it implies that the structural equations are the same. The exogenous variables may be changing, but they are weighted the same way at both times. Others have called this assumption a steady-state process.

Stationarity cannot be guaranteed by using the same operationalization at both points in time. It may well be that the same test measures different constructs at different times. Unfortunately stationarity may be least appropriate in the situation in which it is most needed: rapid developmental growth. Developmentalists today are emphasizing the discontinuities of growth. Implicit in such stage theories of growth is that the causal model shifts. Variables that are important at one stage of development are irrelevant at other stages of development. Instead of assuming stationarity, it may be more reasonable to assume that a variable becomes more causally effective over time. This effect would be indicated by not only asymmetrical crosslags, but also changing (usually increasing) synchronous correlations. Thus in many nontrivial instances, changing synchronous correlations may be indicative of causal effects.

One good rule of thumb is that stationarity is more plausible in cases in which the lag between measurements is short. For instance, I found in the analysis of four-wave data sets that the adjacent waves were relatively more stationary, than were the more distant waves.

BACKGROUND VARIABLES

Very often panel studies contain measures like sex, ethnicity, social class, and other background variables which are potential sources of spuriousness. There are two different strategies for handling these background or social grouping variables. The first is to perform separate analyses on each sex, race, or social group. The second is to subtract out the effects of the background variable. The first strategy is preferred if different causal patterns are expected for different social groups. For instance, Crano, Kenny, and Campbell (1972) found contrasting causal relationships for lower and middle class children in the relationship between intelligence and achievement. However, sample size often prohibits this strategy.

The second strategy—subtracting out the effects of background variables—can be done by computing partial correlations between the relevant variables controlling for the background variables. If the background variables are nominally coded then dummy variables can be created for them. This procedure assumes that the causal processes are the same within social groups although the groups may differ in mean level. Controlling for background variables often increases the stationarity of the data. After the background variables are partialled, the synchronous and cross-lagged partial correlations can be examined.

There are two helpful rules in choosing variables to partial out. First, the variable to partial out should independently explain, at least, a moderate amount of variance. Otherwise nothing is changed by the partialling. Second, D. T. Campbell has suggested that any control variable should in principle be able to explain as much variance of the time 1 variables as the time 2 variables. For instance, imagine a study of cognitive skills that had only a time 2 measure of intelligence. Given temporal erosion, the intelligence measure will correlate more highly with the time 2 measures than the time 1 measures, and, therefore, it would be inappropriate as a variable to partial on.

THREE-WAVE, TWO-VARIABLE MODEL

The model in Figure 12.2 can be extended to three waves. The general equations are

$$\begin{aligned} X_t &= a_t Z_t + b_t U_t \\ Y_t &= c_t Z_t + d_t V_t \end{aligned}$$

The usual assumptions are made that Z , U , and V are autocorrelated but not cross-correlated. If Z is first-order autoregressive, then $\rho_{13} = \rho_{12}\rho_{23}$ for Z . It then follows that the following two vanishing tetrads hold:

$$\begin{aligned} \rho_{X1Y2}\rho_{X2Y3} - \rho_{X1Y3}\rho_{X2Y2} &= 0 \\ \rho_{X2Y1}\rho_{X3Y2} - \rho_{X3Y1}\rho_{X2Y2} &= 0 \end{aligned}$$

Note that both of these tetrads hold even if there is not stationarity: $a_1 \neq a_2 \neq a_3$ and $c_1 \neq c_2 \neq c_3$. Even if Z were not autoregressive ($\rho_{13} \neq \rho_{12}\rho_{23}$), the following overidentifying restriction would hold:

$$\rho_{X1Y2}\rho_{X2Y3}\rho_{X3Y1} - \rho_{X2Y1}\rho_{X3Y2}\rho_{X1Y3} = 0$$

Unfortunately these overidentifying restrictions are of little practical use. In the first place, three-wave data are difficult to obtain. Second, it must be presumed that the causal lag is or is very near to the measured lag between the two waves. If the lag is much different, the tetrads still vanish. Finally the tetrads presume a first-order autoregressive process, and such an assumption may often be unrealistic.

For example, examine the crosslags in Figure 12.4. An example of three waves from Jessor and Jessor (1977) is contained. The synchronous correlations are stationary, and the crosslags clearly indicate a causal effect from attitudinal tolerance of deviance (ATD) to marijuana behavior involvement (MBI). However, both the tetrads vanish: $\chi^2(1)$ equals 1.23 and .03.

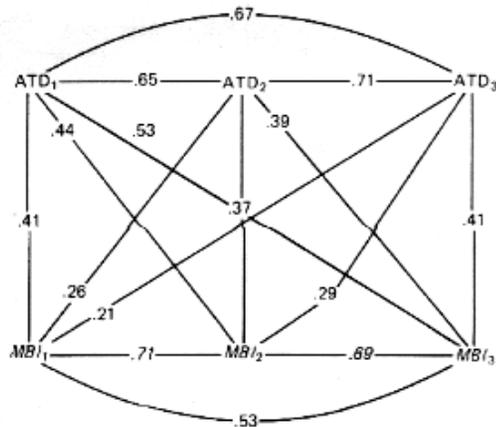


Figure 12.4 Three-wave example.

Just as a two-wave study does not begin to solve all the inference problems of the cross-sectional study, the three-wave study is no panacea. Quite bluntly, the delivery of longitudinal studies falls far short of its inflated promises. The view that "one more wave" is all that is needed to make the conclusions of the study airtight is clearly mistaken. Longitudinal studies are very expensive and usually have severe attrition problems. They should only be undertaken with the sober realization of their limitations.

TWO-WAVE, N-VARIABLE MODEL

For the previous models of CLPC, it was assumed that only a single unmeasured exogenous variable caused both endogenous variables. Now it will be assumed that all the endogenous variables are caused by a set of p exogenous variables, $F_1, F_2, F_3, \dots, F_p$. These p variables are ordinarily all unmeasured and may be considered factors. The equation for endogenous variable X_i at time 1 is

$$X_{i1} = \sum_j a_{ij} F_{1j} + c_{i1} U_{i1}$$

where a_{ij} are the common factor loadings and U_{i1} is a unique factor for X_{i1} uncorrelated with the F s. The U s can be considered errors of measurement and may be correlated over time. All the factors and measured variables have unit variance. Similarly the equation for X_i at time 2 is

$$X_{i2} = \sum_j b_{ij} F_{2j} + c_{i2} U_{i2}$$

All the unmeasured variables are standardized at both time points. It is also assumed that all the common factors, the F s, are orthogonal to each other. Although the orthogonality assumption is totally unrealistic, it greatly simplifies the algebra. Later in this chapter, the model is modified to allow the factors to be correlated.

The stationarity assumption is

$$k = b_{i1}/a_{i1} = b_{i2}/a_{i2} = \dots = b_{ip}/a_{ip} \tag{12.6}$$

This assumption has been called quasi-stationarity (Kenny, 1975b). It simply states that the path coefficients for each variable increase proportionately over time. The term k_i^2 has a simple interpretation. It is the ratio of the X_{i2} communality to X_{i1} communality. This is so because the communality of X_j is $\sum_j b_{ij}^2$ and the communality of X_{j1} is $\sum_j a_{ij}^2$. Given Equation 12.6, it follows that $\sum b_{ij}^2 = k_i^2 \sum a_{ij}^2$. It then follows that the ratio of communalities is k_i^2 . Because for a variable with no specific factor the communality and reliability are identical, one can view k_i^2 as reliability ratio assuming no specific variance. Quasi-stationarity also implies that the standardized structural equation of each variable is invariant with respect to time once the unique factor has been omitted from the equation and the coefficients are accordingly changed to standardize the equation.

As has been shown in Kenny (1973) these communality ratios can be estimated in a straightforward manner. Given the above model, the correlation of X_i with X_j at time 1 is a $\sum_k a_{ik} a_{jk}$ and at time 2 is

$\sum_k b_{ik} b_{jk}$. Given quasi-stationarity, it then follows that $\sum_k b_{ik} b_{jk} = k_i k_j \sum_k a_{ik} a_{jk}$. If q_{ij} stands for the ratio of the time 2 correlation of X_i with X_j to the time 1 correlation of the same variables, then given the above model, q_{ij} simply equals $k_i k_j$. Now given three variables X_i , X_j , and X_m , it follows that

$$k_i^2 = \frac{q_{ij} q_{im}}{q_{jm}} \quad [12.7]$$

Both k_j^2 and k_m^2 can be obtained in a similar fashion:

$$k_j^2 = \frac{q_{ij} q_{jm}}{q_{im}}$$

$$k_m^2 = \frac{q_{im} q_{jm}}{q_{ij}}$$

If the communality ratio k_i^2 is larger than one, the communality is increasing over time; if less than one, the communality is decreasing; and if one, the communality is unchanged. Although it is theoretically possible for k_i to be negative, it is unlikely. It would be indicated by changes in sign of the synchronous correlations and probably negative autocorrelations. Because k_i is not likely to be negative, one can routinely take the positive root of k_i^2 .

It follows from the earlier equations for the synchronous correlations that if the time 1 synchronous correlation of X_i and X_j is multiplied by $(k_i k_j)^{1/2}$ and the time 2 correlation is divided by the same value, the two synchronous correlations should be equal. With this correction, both correlations should now equal $(k_i k_j)^{1/2} \sum_k a_{ik} a_{jk}$.

In a similar fashion the cross-lagged correlations can be corrected to be made equal. The cross-lagged correlations between X_i and X_j are $\sum_k a_{ik} b_{jk} r_k$ where X_i is measured before X_j and $\sum_k b_{ik} a_{jk} r_k$ where X_j is measured before X_i and r_k is the autocorrelation of F_k at times 1 and 2. Given quasi-stationarity, one crosslag is $k_j \sum_k a_{ik} a_{jk} r_k$ and the other is $k_i \sum_k a_{ik} a_{jk} r_k$. If the first crosslag is multiplied by $(k_i/k_j)^{1/2}$ and the

second by $(k_j/k_i)^{1/2}$, both corrected crosslags should be equal. Unequal corrected crosslags would then indicate a violation of spuriousness.

The very same conclusions can be obtained for an oblique set of factors if stationarity assumptions are imposed. It must be assumed that the synchronous correlation of F_i with F_j does not change over time alld that the crosslags of F_i and F_j are equal; that is, $r_{i1j2} = r_{i2j1}$.

SIGNIFICANCE TESTING AND ESTIMATION

An example is perhaps the best way to illustrate the rather detailed process of estimation. The example is taken from the Educational Testing Service (ETS) evaluation of Sesame Street by Ball and Bogatz (1970). The data to be discussed here were obtained from the Cook, Appleton, Conner, Shaffer, Tamkin, and Weber (1975) reanalysis of the data. The sample to be discussed includes 348 preschoolers from five cities. The children are predominately disadvantaged. All the children were encouraged to view Sesame Street by the ETS staff. The children were measured at two points in time with a six-month lag. During six months the children watched, to differing degrees, Sesame Street.

The four dependent variables to be considered are Body Parts, a 32-item test to measure knowledge of body parts; Forms, an 8-item test to measure recognition of geometric figures; Letters, a 65-item test to measure recognition of letters of the alphabet; and Numbers, a 54-item test to measure number recognition. All tests were devised by ETS and were tailored to the objectives of Sesame Street.

In Table 12.1 above the diagonal is the partial correlation matrix for the four measures at the two points in time. For this analysis 10 variables were controlled or partialled out making the effective sample size 338. Some of these are dummy variables to control for sex, city, and race, and others are variables measured at the interval level of measurement, such as age and social class.

To employ the methods elaborated in Kenny (1975b), first the elements of the earlier described q_{ij} matrix is created by dividing the

Table 12.1. Sesame Street Correlations with Background Variables Partialled Out^a

	Time 1				Time 2			
	Body Parts	Forms	Letters	Numbers	Body Parts	Forms	Letters	Numbers
Time 1								
Body Parts	1.000	.469	.248	.491	.434	.287	.241	.343
Forms		1.000	.337	.515	.304	.292	.319	.380
Letters			1.000	.570	.167	.203	.350	.343
Numbers				1.000	.290	.344	.448	.571
Time 2								
Body Parts					1.000	.497	.425	.506
Forms						1.000	.566	.595
Letters							1.000	.745
Numbers								1.000

^aN = 336.

time 2 synchronous correlation by the time 1 synchronous correlation between the same variables. For instance, the q for letters and forms is $.566/.377 = 1.680$.

Table 12.2. The Time 2 Synchronous Correlation Divided by the Time 1 Synchronous Correlation

	Body Parts	Forms	Letters
Forms	1.060		
Letters	1.714	1.680	
Numbers	1.031	1.155	1.307

Table 12.2 presents the q matrix. One can solve for the communality ratios by substituting the sample-based estimates into Equation 12.7. There are three solutions for each of the four communality ratios. In general given n variables there are $(n - 2)(n - 1)/2$ solutions. The solutions are

- Body Parts: 1.081, 0.945, 1.351
- Forms: 1.039, 1.188, 1.485
- Letters: 2.716, 2.173, 1.900
- Numbers: 1.124, 0.786, 0.899

In theory all the preceding estimates should be the same, but they differ because of sampling error.

The problem now is how to pool the three estimates of the communality ratios. Kenny (1975b) suggests the following procedure. The three estimates for Body Parts are

$$\frac{(.497)(.425)(.337)}{(.469)(.248)(.566)} \quad \frac{(.497)(.506)(.515)}{(.469)(.491)(.595)} \quad \frac{(.425)(.506)(.570)}{(.248)(.491)(.745)}$$

Each estimate is the product of three correlations divided by three correlations. What can then be done is simply to sum the numerators of each estimate and sum the denominators. Then divide the two in order to obtain an estimate of the communality ratio. For instance, for Body Parts the pooled estimate is

$$\frac{(497)(.425)(.337) + (497)(.506)(.515) + (425)(.506)(.570)}{(.469)(.248)(.566) + (.469)(.491)(.595) + (.248)(.491)(.745)} = 1.101$$

The remaining pooled communality ratios are 1.277 for Forms, 2.135 for Letters, and 0.935 for Numbers. Note that all the estimates are compromises of the three unpooled estimates. A problem does arise if some of the correlations are negative. In such cases the absolute values should be summed.

Table 12.3. Pearson-Filon Test of Equality of the Synchronous Correlations

	Uncorrected for Shifts in Communality	Corrected for Shifts in Communality
Body Parts-Forms	.501 ^a	-.972
Body Parts-Letters	2.811	.566
Body Parts-Numbers	.292	.150
Forms-Letters	3.924	.127
Forms-Numbers	1.595	.614
Letters-Numbers	4.243	-1.231

^aThe value is positive if the time 2 synchronous correlation is larger than the time 1 correlation and negative if the time 1 correlation is larger.

Next it must be determined whether the synchronous correlations are equal after they have been corrected for shifts in communality. The hypothesis to be tested is $\rho_{1j1}(k_i k_j)^{1/2} = \rho_{1j2}(k_i k_j)^{1/2}$. The Pearson-Filon test can be used to compare the two correlated correlations. There are, however, two problems with this method. It ignores, first, that the *ks* are sample estimates and, second, that the *ks* were deliberately chosen to make the synchronous correlations equal.

In Table 12.3 there is the Pearson-Filon test of the difference between synchronous correlations both with and without correction. It shows that for three of the pairs of the synchronous correlations there is a significant difference. All three of these pairs involve the Letter variable, which is not surprising since that variable has the communality ratio that is most different from one. After correcting for shifts in communality, these three significant differences between the synchronous correlations disappear. It then appears that the correction is helping and quasi-stationarity is indicated.

Table 12.4. Pearson-Filon Test of Equality of Crosslags

	No Correction	Correction ^a
Body Parts-Forms	-.275 ^a	-.628
Body Parts-Letters	1.107	.107
Body Parts-Numbers	.905	1.350
Forms-Letters	1.801	.770
Forms-Numbers	.625	1.611
Letters-Numbers	-2.015	1.095

^aThe value is positive if the first variable in the row "causes" the second, and negative if the second variable "causes" the first.

^bCorrection for shifts in communality.

In a similar fashion the difference between crosslags can be tested. Again there is the problem that it must be assumed that the *ks* are population values. In Table 12.4 there is the Pearson-Filon test of the difference between crosslags. There is one significant difference between crosslags before correction: with Letters and Numbers. Before the reader begins to concoct all sorts of explanations how the learning of numbers causes the learning of letters, note that after correction this difference disappears.

CONCLUSION

CLPC is a valuable technique for ruling out the plausible rival hypothesis of spuriousness. It should not be viewed as only an intuitive approach but as a formal method with assumptions. This chapter has inordinately emphasized alternative explanations of crosslag differences in order to present the reader with a list of problems much in the same way that Campbell and Stanley (1963) do for quasi-experimental designs.

CLPC is, however, largely an exploratory strategy of data analysis. My own suspicion is that its main use will be in uncovering simple causal relationships between uncontrolled variables. What would then follow is either the refinement of both the variables and the process in controlled settings or the estimation of causal parameters of the system by structural equation models. The career of a hypothesized causal relationship might be as follows: first, the consistent replication of a cross-sectional relationship; second, the finding of time-lagged relationships between cause and effect; third, the finding of cross-lagged differences; and fourth, an experiment in which the causal variable is manipulated. Obviously, these steps may often overlap, some may be omitted, and the order may be different. I hope to emphasize that CLPC plays only an intermediary role in social science, between the correlation and a well-elaborated structural model.

13

LOOSE ENDS

In this final chapter both technical and conceptual issues concerning causal modeling are covered. Hopefully some closure is brought to what is certainly a limitless topic. The chapter is divided into two sections. The first section considers technical issues of linearity, ordinal and nominal variables, and standardization. The second section considers conceptual problems in model testing. Let me now try to tie together the loose ends that remain.

TECHNICAL ISSUES

LINEARITY

In this context linearity refers to a straight-line functional relationship between the cause and effect and not the level of measurement of either variable or the additivity of effects. These last two topics are discussed later in this chapter. The issue of linearity then concerns itself with whether there is straight-line relationship between X and Y . The simplest and most intuitive test of linearity is to examine the scatterplot between X and Y . One should always remember that behind every correlation lies a scatterplot; a correlation or regression coefficient is only a number that tries to capture the complex meaning of the scatterplot. While the old eyeball method can often be very useful in determining nonlinearity, in marginal cases your left eyeball may disagree with your right eyeball. You may then need a more formal way to assess the degree of nonlinearity.

One common method of testing for nonlinearity is polynomial regression. The effect variable, Y , is regressed on X, X^2, X^3, \dots, X^n . If

any off these variables besides X significantly predicts Y , then nonlinearity is indicated. Be aware, of problems of large numbers (e.g., 25^5) and multicollinearity. It may be advisable to subtract off the mean of X before computing powers.

A well-known measure of relationship, both linear and nonlinear, is η^2 or eta squared, the correlation ratio. A second strategy for testing nonlinearity based on η^2 is to compute $\eta^2 - r^2$ which is a measure of the percentage of Y variance explained by nonlinearity. Operationally this becomes a bit computationally messy since η^2 can only be estimated when X is divided into discrete categories. If X is not naturally in such categories, it must be divided into categories, for example, for intelligence 80-89, 90-99, 100-109, and so on. The estimate of η^2 is obtained from a one-way analysis of variance with the discrete levels of X as the independent variable and Y as the dependent variable. The value for η^2 is then $SS_{BG}/(SS_{BG} + SS_{WG})$ where SS_{BG} is the sum of squares between groups and SS_{WG} the sum of squares within groups. Now ρ^2 should not be computed (as is sometimes mistakenly done) by correlating X with Y . Rather the means of the X s within categories are correlated with raw Y variable. It then follows that under the null hypothesis of $\eta^2 - r^2$

$$F(k-2, N-k) = \frac{(\eta^2 - r^2)SS_{TOT}/(k-2)}{MS_{WG}}$$

where k is the number of categories that were created. A more efficient method is analysis of covariance. The variable X is the covariate and discrete levels of X are the "treatments." If there is a main effect for treatments then nonlinearity is indicated. This method is more powerful than the analysis of variance method since the MS_{WG} is reduced due to the covariate.

The assumption of linearity should not be tested with just X and Y . It could very well be that the relationship between X and Y is nonlinear but once other variables are controlled the relationship

becomes linear. Thus, it is more appropriate to examine the relationship between X and residuals of Y , if they can be computed.

Given that the relationship between X and Y is nonlinear, what is to be done? The usual strategy is a transformation. Either the exogenous or endogenous variable is transformed into a new metric. Following Kruskal (1968), it is useful to distinguish between one- and two-bend transformations. As suggested by the name, one-bend transformations are functions whose curves have a single bend, while two-bend transformations are curves with two bends. Power transformations of the form $X_i = aX^p$, where X_i is the transform of X , are called one-bend transformations. Transformations like arcsin, logit, probit, and Fisher's Z are two-bend transformations. One-bend transformations are generally more appropriate for variables that have a lower limit of zero and no upper limit. Two-bend transformations are useful for variables that have a lower and upper limit, for instance zero and one.

The three most common one-bend transformations are logarithm, square root, and reciprocal. The logarithm is most useful for amounts, for example, dollars. More generally, the log transformation is useful when it is suspected that two variables combine multiplicatively to cause the dependent variable. For instance, the Cobb-Douglas law states that capital and labor multiply to cause gross national product. The square root transformation is most useful for variables that are assumed to be distributed as Poisson. Good candidates for a square root transformation, therefore, are counts, for example, the number of behaviors per unit time. The third one-bend transformation is the reciprocal or $1/X$. This transformation is especially useful for a variable like time to response or latency, where the transformed variable is now in the metric of speed. With any transformation it is important to be able to interpret the metric of the transformed variable.

The two-bend transformations are useful for variables that are proportions or could easily be turned into a proportion. For instance the number of items passed in an intelligence test could be

divided by the total number of items and would then become a proportion. The most radical transformation is the logit, which is the natural logarithm of the odds. The odds of a proportion p is simply defined as $p/(1 - p)$. Slightly less radical is probit. Probit, which is used by biologists, simply takes the probability and looks up the corresponding standard normal deviate (Z value) that would yield that probability. Since biologists do not seem to like negative numbers, it is customary to add five to this value. The least radical two-bend transformation is the arcsin, a transformation psychologists seem to prefer in analyzing proportions.

There is a final two-bend transformation that is commonly employed. It is Fisher's Z transformation, which is a cousin of the logit transformation. This transformation is usually applied to correlations.

One obvious point should be made here. One-bend transformations should not be made on negative numbers, the log and reciprocal on zero, and logit and probit on zero or one.

Tukey (1977) outlines a number of useful methods by which the appropriate transformation can be chosen empirically. It is beyond the scope of this book to restate these methods. A useful point, though, is that when exploring which transformation to apply, one should work with medians since the median of transformed set of data is the transform of the median while this does not necessarily hold for the mean.

It has been assumed throughout that the causes of an endogenous variable add up. Of course, in some cases nonadditive models may be just as plausible if not more plausible. The whole may not be the sum of its parts. Within the experimental tradition there are methods to distinguish whether the independent variables add or multiple. Also nonlinear least squares (Draper & Smith, 1966) is a feasible computing method, but in the nonexperimental case the functional form in which the exogenous variables combine is best determined a priori and not in any simple a posteriori manner.

ORDINAL AND NOMINAL VARIABLES

Just as every adolescent goes through an identity crisis, it seems that every researcher goes through a "measurement crisis." At some point he or she realizes that the interval assumptions of most social scientific measures are just assumptions. A little reflection then reveals that, at best, most measures are only ordinal. The researcher then vows to discard the correlation and the mean for the rank order coefficient and the median. But just like the crisis of an adolescent, it seems only to be important to have the crisis not to resolve it.

It is indeed true that there is a giant leap of faith in assuming that our measuring instruments are interval, but we are often forced to make just such an assumption. Unfortunately sometimes we are faced with a measure that is clearly ordinal. There is no meaningful rule for assigning numbers to objects. In such a case we might try to determine numbers by scaling methods, for example, multidimensional scaling methods.

If the exogenous variable is ordinal one can simply create a set of dummy variables as was done in Chapter 10. If both the endogenous and exogenous variables are ordinal, there is an interesting empirical scaling method that gives scale weights to the categories of both variables to maximize the correlation between them (Bonacich & Kirby, 1975).

In recent years there has been a great deal of statistical work on the analysis of nominal variables. The model that is analogous to the regression models in Chapter 4 is called logit analysis (Fienberg, 1977), and the model that is analogous to factor analysis is called latent structure analysis (Goodman, 1974). The two preceding sources should prove useful introductions to readers interested in either topic. To give the reader an example of the power of these methods consider the following example taken from Werts, Linn, and Jöreskog (1973). Given three measured dichotomies A , B , and C , let there be a dichotomy Z such that within levels of Z , variables A , B , and C are independent. This is a latent structure model similar to the single-factor

models discussed in Chapter 7. Werts, Linn, and Jöreskog show that not only can the loadings be estimated but the marginals of the underlying dichotomy can be as well. One potential use of this technique might be in clinical assessment. For instance, say three clinicians independently made judgments about whether each member of the patient population was schizophrenic or not. It could then be determined what is the actual percentage of the schizophrenics. Also, the hit rate of each clinician can be measured, and from the hit rates one can determine the probability a patient is a schizophrenic given the judgments of the clinicians. Of course, independence of judgments is of the essence.

INTERACTION

Unfortunately, the topic of interaction has received scant treatment in the causal modeling literature. To some degree researchers do not search for interactions because they are afraid to find them. The approach here is that an interaction implies that a causal law of the form " X causes Y " must be modified to " X causes Y depending on Z ." The variable Z is said "to modify the X causes Y relationship." This notion of interaction as a moderator of a causal law has led some to suggest to draw an arrow from Z to the path between X and Y . This view of interaction is somewhat different from the usual analysis of variance view of interaction in which the two components of the interaction are equal partners in the relationship. The view here is more like the view of Campbell and Stanley (1963) who define an interaction as a threat to external validity. The choice of whether Z modifies XY , the causal relationship between X and Y , or X modifies ZY relationship is arbitrary. For instance, if age and motivation interact in causing athletic ability, a developmentalist would say motivation modifies the developmental age trends and a motivational psychologist would say age modifies the effect of motivation. Both would be correct, but both would be aided by the idea of modification.

The usual way that interaction is modeled is by the inclusion of multiplicative terms in the model. For instance for our X, Y, Z example we would have

$$Y = aX + bXZ \quad [13.1]$$

To capture the idea of modification, we simply factor X out of 13.1 to obtain

$$Y = (a + bZ)X \quad [13.2]$$

Thus the structural coefficient for X is $a + bZ$. It is a variable since we do not know how much X affects Y without knowing Z . Take a simple case where Z can only take on the two values of 0 and 1 and $a + b = 0$. When Z is zero the coefficient is a , and when Z is 1 the coefficient is zero.

Researchers should be encouraged to introduce interaction terms. One should note that ordinarily the XZ and X will be highly correlated, depending largely on whether X has a nonzero mean. Thus, the researcher may have low power in testing interactions. Currently there is much interest in the addition of multiplicative terms to a linear model (Bohrstedt & Marwell, 1977; Cohen, 1978).

As was just stated, it may be useful to include multiplicative terms in an equation to test for specification error. The presence of a significant interaction term might indicate either a need to transform, a nonadditive combination of variables, or even the omission of important third variables.

STANDARDIZATION

This text has relied very heavily on the use of standardized coefficients. To many researchers who exclusively rely on correlation it may come as a surprise the statistician John Tukey has said:

Sweeping things under the rug is the enemy of good data analysis. Often, using the correlation coefficient is “sweeping under the rug” with a vengeance (1969, p. 89).

Others have called correlation and standardization the devil's tools (Birnbaum, 1973). It would then seem that standardization is highly controversial. Let us consider four criticisms of standardization.

1. Standardized measures are inappropriate for generalizing across populations.
2. The unstandardized metric is more interpretable.
3. Standardization will not allow certain specifications or at least make them clumsy.
4. Standardization introduces statistical complexities.

Consider the first point, that standardized coefficients are not as comparable across populations as unstandardized coefficients. It is more theoretically plausible that structural parameters may be invariant across different populations when they are in the unstandardized metric than in the standardized metric. Normally, invariance of the unstandardized metric does not imply invariance of the standardized metric. For instance, if the structural equations for two populations are identical, the standardized path coefficients will ordinarily differ if the variances of the endogenous variable differ. Thus, the well-known finding that beta weights and R^2 are usually larger for whites than blacks is often an artifact due ordinarily to the larger variance in whites than blacks. A good rule of thumb is that then comparing structural coefficients across populations, it is usually best to compare the unstandardized coefficients.

The preceding rule is easy to apply with causal models with measured variables. It is not as instructive for models with unobserved variables. Take, for instance, the simple case in which there is one factor that causes three measured variables with uncorrelated disturbances. We can take the two strategies. We could assume that

the variance of the unmeasured variable is invariant across populations and let the "factor loadings" be variable, or we could assume that the loadings were invariant. If we take the second strategy we must still fix one factor loading. Suffice it to say that we cannot in this case nor in most others allow both the factor loadings and variances to vary across populations.

The second criticism of standardization is that the unstandardized metric is a more interpretable one. If the raw metric has some natural meaning, then all is lost by standardization. It is nice to know how many dollars of earned income is caused by each year of education, the number of items learned for each season of Sesame Street that is watched, and number of pecks for each reinforcement. However, interpretability is often a subjective issue. Sadly, the metrics in most social science studies are nearly meaningless. It may be the case that knowing that if a unit is one standard deviation above the mean on X , then the unit is b units of a standard deviation from the mean on Y is more meaningful. Perhaps the biggest challenge to social scientists is to make measures more naturally interpretable.

The third criticism of standardization is that it sometimes prevents the researcher from making specifications that allow for identification. Consider an example. In Chapter 9 we discussed a model of Wiley and Wiley that stated that the error variance of a measure would remain stable over time while the true variance might change. If the researcher has standardized the variables and discarded the variances, it is impossible to make such a specification. If the variances are not lost it is still possible to standardize and then assume

$$\frac{1 - \rho_{X_1X_1}}{1 - \rho_{X_2X_2}} = \frac{V(X_2)}{V(X_1)}$$

It is always possible that one can standardize and estimate the parameters if the variances are saved. It is then critical always to compute and report variances, as well as means. Even if you are not interested in the means or variances, someone else may be.

It is also possible, though not common, that some specifications or overidentifying restrictions can occur for only the standardized metric. For instance, the fan spread hypothesis implies an overidentification that holds for treatment-effect correlations but not covariances.

In certain cases the unstandardized metric is much simpler. For instance, with standardization it is impossible to make the classical specification that a measured variable equals the true score plus error of measurement. This created complications for the analysis in Chapter 5 on measurement error. Yet it is was a simple matter to handle.

The final reason for unstandardized coefficients is that the standardized coefficients imply dividing by a sample estimate, the standard deviation. This should increase confidence intervals and alter tests of significance. Although these problems are beyond the scope of this text and the competence of its author, we should recognize that we have created a statistical problem by standardization. We leave it to the statisticians to develop solutions to these problems.

In general, unstandardized coefficients should be preferred. However, one should be flexible. Often times, I compute both. In Chapter 3 we gave a very simple formula to destandardize and another formula to standardize an unstandardized coefficient. This text has overly concentrated on the standardized metric. This was done for two reasons.

First, most researchers are more familiar with the standardized model. It seems to me simpler to motivate causal modeling by using a language and notation familiar to the audience. Second, in my opinion, and others differ here, the standardized metric is just much simpler to work with. One need not worry about normalization rules and the like. It seems to me that standardization is especially simpler when one takes a simple high school algebraic approach as was done in the text. The unstandardized metric is somewhat simpler when using a matrix algebra formulation.

Readers should be slowly weaned away from exclusive reliance on standardization. Other texts on this topic, both advanced

and beginning, almost exclusively employ analysis using the raw metric. One may need to first translate some of these analyses to the standardized metric, but eventually one should be able to handle the problem entirely in terms of the unstandardized metric.

Remember to be flexible. The important thing is to learn from data. One must learn only with what one knows, but one must be willing to learn new methods in order to learn even more.

CONCEPTUAL ISSUES

ESTIMATION VERSUS TESTING

Traditionally the major emphasis within the causal modeling literature has been on the estimation of structural parameters. An alternative and sometimes competing emphasis has been the testing and comparison of causal models. To some extent this distinction parallels the classic distinction between descriptive and inferential statistics.

The best way to illustrate the differences between the two emphases is to show how they differently handle overidentification. Let us say we have a model that is overidentified and there are two different researchers who approach it. The first researcher is interested in estimation. He or she wants the most efficient estimates of parameters. The researcher is overjoyed that the model is overidentified for the following reason. Since a given parameter estimate is overidentified, there are two or more estimates of the same parameter. The researcher can then pool these multiple estimates to obtain a more efficient estimate. There may be some technical problems on how to pool the multiple estimates (Goldberger, 1973), but it always involves a weighted sum of the estimates.

The second researcher is also overjoyed the model is overidentified, but for a very different reason. Since there is overidentification, it should be possible to find an overidentifying restriction. With a little imagination a statistical test can be found and it

can be determined whether the overidentifying restriction is satisfied by the data. If the overidentifying restriction is not satisfied by the data, then the model contains a specification error.

The aims of the two researchers are not incompatible since it is always possible to test if the overidentifying restriction is met in the sample, and if it is, to pool estimates to obtain more efficient estimates of parameters. Unfortunately, though, some researchers fail to test overidentifying restrictions and others do not obtain efficient estimates of parameters.

The historical reason for the greater emphasis on estimation is due to the fact that the growth of structural modeling occurred first among econometricians. The typical data set they use is a time series with usually no more than 25 observations. It would seem that this would be an incredibly hostile environment for causal modeling to grow, but grow it did. Faced with the problem of serial dependency, the issue of estimation became paramount. Moreover, given the small sample sizes and very high multicollinearity, the efficiency of estimation is even more critical. Readers of econometrics texts often fail to realize the co-linearity of economic data since usually only covariance matrices are given. For instance, for the inventory-demand model discussed by Goldberger (1964, p. 286) the median correlation is .92, and lowest correlation is .90. It is hardly surprising that econometricians worry about the efficiency of their estimates.

A second reason for the emphasis on estimation over testing is that econometricians were the pioneers of the theory of identification. The central question is whether a model is identified in the first place not whether it is overidentified. Thus, the focus was on identification and not overidentification.

A third reason for the emphasis on estimation within econometrics is the very natural metric that they possess: dollars. Economists, unlike psychologists, are unabashedly applied and as a matter of course want to know the effect in dollars and cents of the exogenous variables. Estimation of structural parameters is then very useful for economic policy.

Estimation is an important goal in causal modeling, but it should not blind one from considering the testing of models as also important. One should not neglect the testing of overidentifying restrictions. Also one should consider models in which none or few of the parameters are estimable, but the model is still overidentified. Examples of this are in Figure 3.6 and throughout Chapters 11 and 12. Normally, parameter estimates are complicated combinations of covariances but it may be that a simple comparison of two covariances or correlations reveals something very interesting about a model.

Still overidentifying restrictions are not the be all and end all of causal modeling. A major problem with an overidentifying restriction is that, if it fails, it may not be able to pinpoint where the specification error exists. Ordinarily if the restriction is not met, a whole multitude of specifications could be incorrect, either only one or all. Thus not meeting the restriction may not be sufficiently diagnostic about where the specification error is.

Also, even if the restriction is met the model is still not proved to be true. In the first place, the test of the overidentifying restriction may be of low power. For instance, if correlations are low, then the test of a vanishing tetrad is of low power. Second and even more fundamental, if the restriction holds, there are an infinite set of similarly over-identified models that would satisfy exactly the same set of restrictions.

We should never lose sight that causal models can never be proven to be true. Certainty must be left to the mystic. Causal models can be disconfirmed by the data but never is a single model confirmed, but rather a host of such models. No doubt that after finishing this text the reader has a sense of disillusionment. Before starting the text, it was perhaps hoped that one would learn magical methods to use correlations to test theory. Perhaps some hoped for a black box into which one would put the correlations and out would come the theory that explained the correlations. Such is not the case. It should now be clear that causal modeling requires the researcher to already know a

great deal about the processes that generated the data. Knowledge does not grow in a vacuum.

There is still an advantage in having a model that is overidentified. A model should say not only what should happen with the data but also what should not. If a zero path or an overidentification follows from a model, then the path must be zero or overidentification must hold. If they do not hold then the model is misspecified. One can, of course, rescue the model by adding an additional parameter, but, hopefully, honest researchers will lessen their confidence in the model.

Campbell and Fiske (1959) make a valuable contribution in their discussion of the validity of constructs by introducing convergent and discriminant validity. Metaphorically one can test the convergent and discriminant validity of a structural model. Convergent validity refers to the fact that a model replicates across different samples, operationalizations of constructs, and estimation methods. Duncan (1975, pp. 152-158) provides a very useful discussion of the replication of estimates of a structural parameter across different samples. Just as important is discriminant validation. The model should imply very different things than alternative competing models. This will usually take the form of zero paths and overidentifying restrictions. Ideally data should be able to discriminate between alternative structural models. One can never rule out all possible models but we should attempt to rule out the strongest competing alternative models.

For correlational studies the classic Campbell and Stanley (1963) distinction between internal and external validity breaks down. To some extent internal validity can be defined as all the threats to causal inference that are controlled for by a true experiment. Thus by definition a correlational study lacks internal validity. It would then seem that the more appropriate goals of a correlational study are convergent and discriminant validity.

MODEL SPECIFICATION

Two errors of model specification are common to beginning causal modelers. The first is the tendency to draw path models in which everything causes everything else. All variables are assumed to be measured with error and all the variables are involved in feedback relationships. When one reads theory papers on various topics, one usually finds a pseudo path diagram in which every variable causes the other. As the reader must now realize, although such a model may be theoretically interesting, it is most certainly underidentified; that is, it is empirically useless. The researcher and the theorists must learn to pare a theory down to a smaller number of paths. Such a simplification will involve distortion, but if the omitted paths are small then such distortion should be minor. Each of us should realize that all models are most certainly misspecified, but hopefully the specification errors are minor.

One common type of "overspecification" is measurement error in measured exogenous variables. Readers of Chapter 5 should not despair of identifying causal models when there are "errors in variables." If reliability is high the effect of measurement error will be small.

One still should be very careful to avoid an "ad hoc specification." An ad hoc specification is one that is chosen solely with the purpose of identification in mind. It strikes me that often the zero path assumption made in instrumental variable estimation is often made for the sake of expediency rather than being motivated by theory. One must avoid the trap of ad hoc specification.

Related to ad hoc specification is the need to respecify a model after estimation and testing. This is a normal practice and a researcher should not be embarrassed. What one should avoid in writing up and presenting results is discussion of only the final model as if it were the only model. The reader deserves to know and to be instructed about the initial model and the steps that were taken to obtain the final model.

An excellent example of the development of a final model is contained in Duncan, Haller, and Portes (1971).

The second problem that beginners have with model specification is underspecification. The model is nicely overidentified but it is clearly unrealistic. Be realistic about all possible linkages in a model. Read the literature and know which paths have been found in the past. It is indeed useful to always start with (or work to) the simplest and most parsimonious model, but realize that the world is not as simple as we might like it to be.

IN SEARCH OF HIGH R^2

The traditional criterion for evaluating a regression equation is R^2 . It is generally thought that the larger the R^2 , the better the model is. It would be nice to be able to predict 100% of the variance, but this may not be feasible or even possible. As suggested in Chapter 1, one may take the ontological position that empirical phenomena are not perfectly predictable in both theory and practice. As a practical rule there often appears to be an upper limit to R^2 .

In the first place the researcher should avoid questionable statistical methods that capitalize on chance. The R^2 should be adjusted downward to take into account any such capitalization. Moreover, there are three guaranteed methods of increasing R^2 that involve no real theoretical insight: using an alternative measure, a lagged value, and aggregation of units. The alternative measure method is simply to use another measure of the endogenous variable as an exogenous variable. In such a case the R^2 will be more a measure of reliability than predictability. A second and related method is to use a lagged value as an exogenous variable; if one is attempting to predict intelligence at age 10, use intelligence at age 9 as a predictor. The third method of achieving a high R^2 is to aggregate the data. Combine subjects into groups and make groups the unit of analysis. As a rule aggregated data yield a higher R^2 than individual level data. But beware of the ecological fallacy. The inferences for the aggregated data refer to the

aggregated units and not the individual. The goal of causal modeling is not to maximize R^2 but to test theory. One would suppose that better theories have higher R^2 but this does not necessarily follow.

One can legitimately increase R^2 by having models that include unmeasured variables. For instance, let X_1 and X_2 be equally reliable indicators of a latent construct F with reliability a^2 and assume that F causes X_3 . If $r_{13} = r_{23} = b$ then the multiple correlation squared of X_1 and X_2 with X_3 is

$$\frac{2a^2b^2}{1+a^2} \quad [13.3]$$

whereas the correlation squared of F with X_3 is b^2/a^2 . Note for any value of b the correlation of F with X_3 is always greater than Equation 13.3 ($a^2 < 1$) and that for a fixed b , as parameter a decreases the correlation of F with X_3 increases. Thus, for a fixed set of correlations between indicators and an endogenous variable, the correlation of the factor with that variable increases as the reliabilities of the measures decrease!

Although a high R^2 is not the absolute goal, some correlation is necessary for causal analysis. Please take my advice and do not vainly attempt to salvage a matrix of essentially zero correlations. There are tests of significance that triangular and rectangular matrices have only zero correlations. Such tests should be used more often to discard useless data sets.

NEGLECTED TOPICS AND PROCEDURES

This text is meant to be only an introduction to causal modeling. Although only introductory, it has taken a rather broad sweep. Some of these topics have been covered rather superficially and tangentially. The hope is that the broad exposure will whet the appetite for more. But remember, a little sip is dangerous, so drink deep.

There has been no extensive discussion of time series. This would be a text in and of itself. Also nowhere is the estimation method of generalized least squares. This estimation technique is useful for overidentified models, "seemingly unrelated equations," and time series. Not enough attention has been paid to various statistical assumptions (e.g., distributional) and their implications to statistical estimates. Finally and perhaps most seriously, the text has not employed matrix algebra. Thus, many of the methods discussed have not been adequately explained for the general case. However, this book is an introduction. The concern here is for initial comprehension of a simple case. Then once that is learned, the reader will be motivated to learn the general case. The book, to use a metaphor of Wittgenstein, is a ladder with which one can climb to new heights, but once one has climbed as high as the ladder will take one, the ladder must be cast aside.

REFERENCES

Althausen, R. P. (1971). Multicollinearity and nonadditive regression models. In H. M. Blalock (Ed.), *Casual models in the social sciences* (pp. 453-472). Chicago: Aldine-Atherton.

Alwin, D. F. (1974). Approaches to the interpretation of relationships in the multitrait-multimethod matrix. In H. L. Costner (Ed.), *Sociological methodology 1973-1974* (pp. 79-105). San Francisco: Jossey-Bass.

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review*, *40*, 37-47.

Alwin, D. F., & Tessler, R. C. (1974). Casual models, unobserved variables, and experimental data. *American Journal of Sociology*, *80*, 58-86.

Anscombe, F. J. (1968). Statistical analysis: Outliers. In D. L. Sills (Ed.), *International encyclopedia of the social sciences* (pp. 178-181). New York: Macmillan & The Free Press.

Applebaum, M. I., & Cramer, E. M. (1974). Some problems in the nonorthogonal analysis of variance. *Psychological Bulletin*, *81*, 335-343.

Bachman, J. G., & Jennings, M. K. (1975). The impact of Vietnam on trust in government. *Journal of Social Issues*, *31*, 141-155.

Ball, S., & Bogatz, G. A. (1970). *The first year of Sesame Street: An evaluation*. Princeton, NJ: Educational Testing Service.

Berman, J. S., & Kenny, D. A. (1976). Correlational bias in observer ratings. *Journal of Personality and Social Psychology*, *34*, 263-273.

Birnbaum, M. H. (1973). The devil rides again: Correlation as an index of fit. *Psychological Bulletin*, *79*, 239-242.

Blalock, H. M. (1964). *Casual inferences in nonexperimental research*. Chapel Hill: University of North Carolina.

Blalock, H. M. (1969). *Theory construction: From verbal to mathematical formulations*. Englewood Cliffs, NJ: Prentice Hall.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.

Bohrstedt, G. W. (1969). Observations on the measurement of change. In E. F. Borgatta (Ed.), *Sociological methodology 1969* (pp. 113-133). San Francisco: Jossey-Bass.

Bohrstedt, G. W., & Marwell, G. (1978). The reliability of products of two random variables. In K. F. Schuessler (Ed.), *Sociological methodology 1978* (pp. 254-273). San Francisco: Jossey-Bass.

Bonacich, P., & Kirby, D. (1975). Using assumptions of linearity to establish a metric. In D. R. Heise (Ed.), *Sociological methodology 1976* (pp. 230-249). San Francisco: Jossey-Bass.

Brewer, M. B., Campbell, D. T., & Crano, W. D. (1970). Testing a single factor model as an alternative to the misuse of partial correlations in hypothesis-testing research. *Sociometry*, *33*, 1-11.

Byrk, A. S., & Weisberg, H. I. (1977). Use of the nonequivalent control group design when subjects are growing. *Psychological Bulletin*, *84*, 950-962.

Calsyn, R. J. (1973). *The casual relation between self-esteem, locus of control, and achievement: A cross-lagged panel analysis*. Unpublished doctoral dissertation, Northwestern University.

Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology*, *69*, 136-145.

Campbell, D. T. (1963). From description to experimentation: Interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 212-242). Madison: University of Wisconsin.

Campbell, D. T. (1969). *The effect of college on students: Proposing a quasi-experimental approach*. Research report, Northwestern University.

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, *24*, 409-429.

- Campbell, D. T. (1971). Temporal changes in treatment-effect correlations: A quasi-experimental model for institutional records and longitudinal studies. In G. V. Glass (Ed.), *The promise and perils of educational information systems* (pp. 93-110). Princeton, NJ: Educational Testing Service.
- Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schlipp (Ed.), *The philosophy of Karl R. Popper* (pp. 413-463). LaSalle, IL: Open Court Publishing.
- Campbell, D. T., & Erlebacher, A. (1970). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), *Compensatory education: A national debate. The disadvantaged child* (Vol. III, pp. 185-225). New York: Brunner-Mazel.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago: Rand McNally.
- Clyde, D. J., Cramer, E. M., & Sherin, R. J. (1966). *Multivariate statistical programs*. Coral Gables, FL: University of Miami.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic.
- Cohen, J. (1978). Partialled products are interactions; partialled powers are curve components. *Psychological Bulletin*, *85*, 858-866.
- Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Cook, T. D., Appleton, H., Conner, R. F., Shaffer, A., Tamkin, G., & Weber, S. J. (1975). *"Sesame Street" revisited: A case study in evaluation research*. New York: Russell Sage.
- Cook, T. D., & Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational research* (pp. 223-326). Chicago: Rand McNally.
- Costner, H. L. (1969). Theory, deduction, and rules of correspondence. *American Journal of Sociology*, *75*, 245-263.
- Costner, H. L., & Schoenberg, R. (1973). Diagnosing indicator ills in multiple indicator models. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 167-199). New York: Seminar.
- Crano, W. D., Kenny, D. A., & Campbell, D. T. (1972). Does intelligence cause achievement? A cross-lagged panel analysis. *Journal of Educational Psychology*, *63*, 258-275.
- Cronbach, L. J., & Furby, L. (1970). How should we measure "change"- or should we? *Psychological Bulletin*, *70*, 68-80.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.
- Cronbach, L. J., Rogosa, D. R., Floden, R. E., & Price, G. G. (1978). *Analysis of covariance in nonrandomized experiments: Parameters affecting bias*. Evaluation Consortium, Stanford University.
- Cudeck, R. (1989). The analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, *105*, 317-327.
- Director, S. M. (1974). *Underadjustment bias in the quasi-experimental evaluation of manpower training*. Unpublished doctoral dissertation, Northwestern University.
- Draper, N. R., & Smith, H. (1966). *Applied regression analysis*. New York: Wiley.
- Duncan, O. D. (1966). Path Analysis: Sociological examples. *American Journal of Sociology*, *72*, 1-16.
- Duncan, O. D. (1969). Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin*, *72*, 177-182.

- Duncan, O. D. (1970). Partial, partitions, and paths. In E. F. Borgatta & G. W. Bohrnstedt (Eds.), *Sociological methodology 1970* (pp. 38-47). San Francisco: Jossey-Bass.
- Duncan, O. D. (1972). Unmeasured variables in linear models for panel analysis: Some didactic examples. In H. L. Costner (Ed.), *Sociological methodology 1972* (pp. 36-82). San Francisco: Jossey-Bass.
- Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic.
- Duncan, O. D. (1976). Some linear models for two-wave, two-variable panel analysis with one-way causation and measurement error. In H. M. Blalock et al. (Eds.), *Mathematics and sociology* (pp. 285-306). New York: Academic.
- Duncan, O. D., Haller, A. O., & Portes, A. (1971). Peer influences on aspirations: A reinterpretation. In H. M. Blalock (Ed.), *Causal models in the social sciences* (pp. 219-244). Chicago: Aldine-Atherton.
- Fienberg, S. E. (1977). *The analysis of cross-classified data*. Cambridge, MA: MIT Press.
- Fine, G. A. (1976). Psychics, clairvoyance, and the real world: A social psychological analysis. *Zetetic*, 1, 25-33.
- Finney, J. M. (1972). Indirect effects in path analysis. *Sociological Methods & Research*, 1, 175-186.
- Goldberger, A. S. (1964). *Econometric theory*. New York: Wiley.
- Goldberger, A. S. (1970). On Boudon's method of linear causal analysis. *American Sociological Review*, 35, 97-101.
- Goldberger, A. S. (1971). Econometrics and psychometrics. A survey of communalities. *Psychometrika*, 36, 83-107.
- Goldberger, A. S. (1973). Efficient estimation in overidentified models: An interpretive analysis. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 131-152). New York: Seminar.
- Goodman, L. A. (1973). Causal analysis of data from panel studies and other kinds of surveys. *American Journal of Sociology*, 78, 1135-1191.
- Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I. A modified latent structure approach. *American Journal of Sociology*, 79, 1179-1259.
- Gordon, R. A. (1968). Issues in multiple regression. *American Journal of Sociology*, 73, 592-616.
- Gorsuch, R. L. (1974). *Factor analysis*. Philadelphia: W. B. Saunders.
- Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.
- Hamilton, V. L. (1975). *The crime of obedience: Jury simulation of a military trial*. Unpublished doctoral dissertation, Harvard University.
- Hargens, L. L., Reskin, B. F., & Allison, P. D. (1976). Problems in estimating measurement error from panel data: An example involving the measurement of scientific productivity. *Sociological Methods & Research*, 4, 439-458.
- Harmon, H. H. (1967). *Modern factor analysis*, 2nd ed. Chicago: University of Chicago.
- Harris, R. J. (1975). *A primer of multivariate statistics*. New York: Academic.
- Hauser, R. M. (1973). Disaggregating a social-psychological model of educational attainment. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 255-284). New York: Seminar.
- Hauser, R. M., & Goldberger, A. S. (1971). The treatment of unobservable variables in path analysis. In H. L. Costner (Ed.), *Sociological methodology 1971* (pp. 81-117). San Francisco: Jossey-Bass.
- Hays, W. L. (1963). *Statistics*. New York: Holt, Rinehart, & Winston.

- Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, 34, 93-101.
- Heise, D. R. (1970). Causal inference from panel data. In E. F. Borgatta & G. W. Bohrnstedt (Eds.), *Sociological methodology 1970* (pp. 3-27). San Francisco: Jossey-Bass.
- Heise, D. R. (1975). *Causal analysis*. New York: Wiley-Interscience.
- Huck, S. W., & McLean, R. A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin*, 82, 511-518.
- Humphreys, L. G. (1960). Investigations of a simplex. *Psychometrika*, 25, 313-323.
- Insko, C. A., Thompson, V. D., Stroebe, W., Shaud, K. F., Pinner, B. E., & Layton, B. D. (1973). Implied evaluation and the similarity-attraction effect. *Journal of Personality and Social Psychology*, 25, 297-308.
- Jaccard, J., Weber, J., & Lundmark, J. (1975). A multitrait-multimethod analysis of four attitude assessment procedures. *Journal of Experimental Social Psychology*, 11, 149-154.
- Jacobson, J. (1977). *The determinants of early peer interaction*. Unpublished doctoral dissertation, Harvard University.
- Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality*. New York: Basic Books.
- Jessor, R., & Jessor, S. (1977). *Problem behavior and psychosocial development*. New York: Academic.
- Jöreskog, K. G. (1970). Estimation and testing simplex models. *British Journal of Mathematical and Statistical Psychology*, 23, 121-145.
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.
- Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York: Seminar.
- Jöreskog, K. G., & Sörbom, D. (1976). *LISREL III- Estimation of linear structural equation systems by maximum likelihood methods*. Chicago: National Educational Resources.
- Jöreskog, K. G., & Sörbom, D. (1977). Statistical models and methods for the analysis of longitudinal data. In D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socioeconomic models* (pp. 285-325). Amsterdam: North-Holland.
- Kelley, T. L. (1935). Essential traits of mental life. *Harvard Studies in Education*, 26, 146.
- Kenny, D. A. (1973). Cross-lagged and synchronous common factors in panel data. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 153-166). New York: Seminar.
- Kenny, D. A. (1974). A test of vanishing tetrad: The second canonical correlation equals zero. *Social Science Research*, 3, 83-87.
- Kenny, D. A. (1975a). A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. *Psychological Bulletin*, 82, 345-362.
- Kenny, D. A. (1975b). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, 82, 887-903.
- Kenny, D. A., & Cohen, S. H. (1980). An reexamination of selection and growth processes in the nonequivalent control group design. In K. F. Scheussler (Ed.), *Sociological methodology 1980* (pp. 290-313). San Francisco, CA: Jossey-Bass.
- Kerchoff, A. C. (1974). *Ambition and attainment*. Rose Monograph Series.
- Kidder, L. H., Kidder, R. L., & Synderman, P. (1974). *Do police cause crime? A cross-lagged panel analysis*. Duplicated research report, Temple University.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

- Kruskal, J. B. (1968). Statistical analysis, the transformation of data. In D. L. Sills (Ed.), *International encyclopedia of the social sciences* (pp. 182-193). New York: Macmillan & The Free Press.
- Lawley, D. N., & Maxwell, A. E. (1963). *Factor analysis as a statistical method*. London: Butterworths.
- Lazarsfeld, P. F. (1972). Mutual relations over time of two attributes: A review and integration of various approaches. In M. Hammer, K. Salzinger, & S. Sutton (Eds.), *Psychopathology*. New York: Wiley.
- Levin, J. R., & Marascuilo, L. A. (1972). Type IV error and interactions. *Psychological Bulletin*, 78, 368-374.
- Linn, R. L., & Werts, C. E. (1977). Analysis implications of the choice of a structural model in the nonequivalent control group design. *Psychological Bulletin*, 84, 229-234.
- Lord, J. M. (1960). Large-scale covariance analysis when the controlled variable is fallible. *Journal of the American Statistical Association*, 55, 307-321.
- Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, 68, 304-305.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Magidson, J. (1977). Toward a causal model approach for adjusting for preexisting differences in nonequivalent control group situation: A general alternative to ANCOVA. *Evaluation Quarterly*, 1, 399-420.
- McNemar, Q. (1969). *Psychological Statistics*, 4th ed. New York: Wiley.
- Milburn, M. A. (1978). *Process analysis of mass media campaigns*. Unpublished doctoral dissertation, Harvard University.
- Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. (1975). *SPSS: Statistical package for the social sciences*, 2nd ed. New York: McGraw-Hill.
- Overall, J. E., & Spiegel, D. K. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin*, 72, 311-322.
- Overall, J. E., Spiegel, D. K., & Cohen, J. (1975). Equivalence of orthogonal and nonorthogonal analysis of variance. *Psychological Bulletin*, 82, 182-186.
- Overall, J. E., & Woodward, J. A. (1977). Nonrandom assignment and the analysis of covariance. *Psychological Bulletin*, 84, 588-594.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University Press.
- Pelz, D. C., & Andrews, F. M. (1964). Detecting causal priorities in panel study data. *American Sociological Review*, 29, 836-848.
- Peters, C. C., & VanVoorhis, W. R. (1940). *Statistical procedures and the mathematical bases*. New York: McGraw-Hill.
- Porter, A. C., & Chibucos, T. R. (1974). Selecting analysis strategies. In G. D. Borich (Ed.), *Evaluating educational programs and products* (pp. 415-464). Englewood Cliffs, NJ: Educational Technology Publications.
- Rickard, S. (1972). The assumptions of causal analyses for incomplete causal sets of two multilevel variables. *Multivariate Behavioral Research*, 7, 317-359.
- Rosengren, D. E., Windahl, S., Hakansson, P. A., & Johnsson-Smargdi, V. (1976). Adolescents' TV relations: three scales. *Communication research*, 3, 347-366.
- Rosenthal, R., & Rosnow, R. L. (1975). *The volunteer subject*. New York: Wiley-Interscience.
- Rozelle, R. M., & Campbell, D. T. (1969). More plausible rival hypothesis in the crossed-lagged panel correlation technique. *Psychological Bulletin*, 71, 74-80.
- Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2, 1-26.

- Sartre, J. P. (1968). *Search for a method*. New York: Vintage Books.
- Schwartz, E., & Tessler, R. C. (1972). A test of a model for reducing measured attitude-behavior discrepancies. *Journal of Personality and Social Psychology*, 24, 225-236.
- Sibley, L. (1976). *An empirical study of the effects of context and linguistic training on judgments of grammaticality*. Unpublished paper, Harvard University.
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcomes studies. *American Psychologist*, 32, 752-760.
- Snedecor, G. W., & Cochran, W. G. (1967). *Statistical methods*, 6th ed. Ames, IA: Iowa State University.
- Spearman, C., & Holzinger, K. (1924). The sampling error in the theory of two factors. *British Journal of Psychology*, 15, 17-19.
- Stanley, J. C. (1961). Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. *Psychometrika*, 26, 205-219.
- Staw, B. M. (1974). Attitudinal and behavioral consequences of changing a major organizational reward: A natural field experiment. *Journal of Personality and Social Psychology*, 6, 742-751.
- Steiner, I. D. (1970). Perceived freedom. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 5, pp. 187-249). New York: Academic.
- Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North-Holland.
- Tatsuoka, M. M. (1971). *Multivariate analysis*. New York: Wiley.
- Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51, 309-317.
- Tukey, J. W. (1969). Data analysis: Sanctification or detective work. *American Psychologist*, 24, 83-91.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Tyron, R. C., & Bailey, D. C. (1970). *Cluster analysis*. New York: McGraw-Hill.
- Weiss, C. H. (1972). *Evaluation research*. Englewood Cliffs, NJ: Prentice-Hall.
- Werts, C. E., & Linn, R. L. (1970). A general linear model for studying growth. *Psychological Bulletin*, 73, 17-22.
- Werts, C. E., Linn, R. L., & Jöreskog, K. G. (1973). A congeneric model for platonic true scores. *Educational and Psychological Measurement*, 33, 531-544.
- Werts, C. E., Jöreskog, K. G., & Linn, R. L. (1971). Comment on the estimation of measurement error in panel data. *American Sociological Review*, 36, 110-112.
- Werts, C. E., Rock, D. A., Linn, R. L., & Jöreskog, K. G. (1976). Comparison of correlations, variances, covariances, and regression weights with or without measurement error. *Psychological Bulletin*, 83, 1007-1013.
- Wiley, D. E., & Wiley, J. A. (1970). The estimation of measurement error in panel data. *American Sociological Review*, 35, 112-117.
- Williams, E. J. (1959). The comparison of regression variables. *Journal of Royal Statistical Society Series B*, 21, 396-399.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557-585.
- Yee, A. H., & Gage, N. L. (1968). Techniques for estimating the source and direction of causal inference in panel data. *Psychological Bulletin*, 70, 115-126.

I N D E X

- Acland, H., 336
 Allison, P. D., 215, 223, 335
 Althausen, R. P., 235, 330
 Alwin, D. F., 54, 90, 188, 250, 330
 Analysis of covariance, 258, 262, 263, 264, 265, 266, 268, 269, 271, 273, 277, 278, 279, 280, 313, 339
 Analysis of variance, 67, 162, 184, 232, 234, 239, 252, 253, 254, 262, 276, 313, 330, 339
 multivariate, 248, 249, 250, 253
 repeated measures, 247, 248, 250, 262
 unweighted means, 242, 243, 244
 Andrews, F. M., 282, 339
 Anscombe, F. J., 17, 330
 Appelbaum, M. I., 330
 Appleton, H., 303, 332
 Attenuation, correction for, 97, 98, 101
 Autoregressive process, 215, 216, 220, 222, 292, 299
- Bachman, J. G., 284, 330
 Bailey, D. C., 173, 340
 Ball, S., 303, 330
 Bane, M. J., 336
 Bent, D. H., 338
 Berman, J. S., iv, 135, 330
 Birnbaum, M. H., 319, 330
 Blalock, H. M., 28, 71, 330, 334
 Bock, R. D., 62, 231, 234, 239, 249, 331
 Bogatz, G. A., 303, 330
 Bohrnstedt, G. W., 282, 318, 331, 333, 336
 Bonacich, P., 316, 331
 Brewer, M. B., 158, 331
 Brush, L., 107, 108
- Byrk, A. S., 331
 Byrne, D., 240, 241
- Calsyn, R. J., 296, 331
 Campbell, D. T., 2, 11, 84, 86, 135, 158, 185, 187, 189, 193, 196, 202, 205, 230, 256, 259, 265, 267, 275, 279, 280, 282, 284, 289, 292, 293, 294, 296, 298, 308, 317, 325, 331, 332, 333, 339, 340
 Causation, 3-5, 10
 Cause, composite, 223-225
 Chain, casual, 79, 215-216, 254
 Change score analysis, 260, 262, 263, 264, 265, 275, 276, 278
 Chibucos, T. R., 262, 263, 271, 339
 Clyde, D. J., 332
 Cochran, W. G., 89, 340
 Cohen, D., 336
 Cohen, J., 85, 239, 246, 318
 Cohen, P., 239
 Cohen, S. H., iv, 276
 Communalities, 139, 152, 154, 173, 301, 302, 305, 306, 307
 ratio, 301, 302, 305, 306, 307
 Conner, R. F., 303, 332
 Consistency of epistemic correlations, 175, 184, 186, 214
 Constraint, equality, 154, 183, 209, 212, 213, 214
 proportionality, 50, 170
 Contrast, 195, 234
 Cook, T. D., 196, 230, 289, 303, 332, 333
 Correlation, 23
 canonical, 47, 146, 147, 148, 177, 180-182, 187, 218, 224, 225, 249, 253, 273, 337
 cross-, 175, 180
 cross-lagged, 282, 287, 288, 295, 302
 epistemic, 172, 173, 174, 175, 184, 186, 214
 factor, 139, 153, 154, 170, 173, 174, 190, 191, 193, 199

- multiple, 61, 62, 64, 66, 67, 68, 70, 73, 75, 84-88, 93, 132, 139, 152, 232, 328
- partial, 47, 55, 70-75, 93, 119, 143, 215, 225, 249, 252, 264, 282, 284, 298, 303
- point biserial, 236, 238, 259
- test-retest, 100, 222, 293, 235
- treatment-effect, 259, 260, 263, 264, 273, 276, 277, 278, 321
- Correlational inference, 1-2
- Costner, H. L., 141, 160, 173, 330, 333, 334, 335
- Covariance, 1, 14, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 121, 127
 - algebra, 20-26
 - matrix, 26, 49, 170, 182, 183, 195, 196, 198, 199, 201, 219, 220
- Cramer, E. M., 162, 246, 330, 332
- Crano, W. D., 158, 185, 202, 296, 298, 331, 333
- Cronbach, L. J., 100, 165, 166, 196, 262, 265, 280, 333
- Cudeck, R., 89
- Cycle, 18

- Design, experimental, iii, 28, 29, 84, 229, 246, 280, 284, 308, 332
 - factorial, 29, 84, 231, 232, 235, 236, 238, 246
 - nonequivalent control group, 256-281
- Director, S. M., 278, 333
- Disturbance, 34-37, 43, 44, 49, 50, 53, 65, 66, 67, 68, 72, 73, 74, 81, 91, 93, 94, 101, 104, 112, 115, 117, 118, 119, 130, 132, 140, 142, 145, 160, 184, 195, 196, 198, 201, 204, 212, 221, 223, 229, 230, 253
 - absence of, 195, 223, 253
 - correlated, 41, 71, 79, 115, 159-161, 178, 179, 188, 214, 319
- Draper, N. R., 315, 333
- Duncan, O. D., 33, 50, 53, 84, 116, 133, 134, 136, 146, 148, 209, 211, 212, 213, 282, 325, 327, 333, 334, 335, 336, 337
- Effect, indirect, 89, 90, 91, 92, 245, 246, 334
- Einstein, A., 11, 52
- Equation, structural, 14, 26, 28, 29, 30, 31, 32, 33, 39, 42, 43, 48, 54, 55, 56, 69, 77, 82, 86, 94, 102, 112, 114, 174, 190, 194, 197, 200, 201, 204, 211, 212, 215, 221, 222, 224, 226, 227, 251-253, 258, 286, 287, 297, 301, 308, 319, 333, 334, 335, 336, 337
- Equilibrium, 66, 122, 123, 136
- Erlebacher, A., 205, 265, 279, 332
- Error, measurement, 18, 19, 94, 95, 97, 98, 101, 102, 104, 113, 119, 140, 145, 180, 190, 195, 196, 198, 216, 222, 224, 258, 262, 263, 321, 326, 334, 335, 341
 - sampling, 45, 68, 116, 143, 149, 181, 219, 230, 286, 305, 340
 - type IV, 7, 338
- Estimation, 53-54

- Factor, common, 138, 139, 143, 265, 301
 - loading, 139, 141, 151, 160, 165, 170, 172, 191, 320,
 - second-order, 196
 - specific, 138, 196, 301
 - unique, 139, 195, 196, 301
- Factor analysis, i, iii, 6, 7, 35, 136, 138, 141, 143, 148, 153, 154, 155, 157, 164, 170, 174, 177, 180, 196, 200, 217, 247, 250, 282, 316, 335, 338
 - cluster analysis, 173, 340
 - confirmatory, 154, 180, 200, 250
 - diagonal, 216
 - maximum likelihood, 153, 154, 166, 174, 177, 180
 - multiple group solution, 173
 - principal components, 114, 149, 151, 152, 173, 216
 - principal factors, 149, 152
 - square root method, 170
- Fienberg, S. E., 162, 316, 334
- Fine, G. A., 82, 334
- Finney, J. M., 54, 334
- First law, 33, 36-42
- Fiske, D. W., 86, 134, 186, 188, 193, 325, 332
- Floden, R. E., 280, 333
- Freedom, 10-12

- Furby, L., 262, 265, 280, 333
- Gage, N. L., 283, 292, 332, 341
- Gintis, H., 336
- Glass, G. V., 8, 332, 340
- Gleser, G. C., 333
- Goldberger, A. S., 45, 78, 147, 225, 282, 322, 323, 333, 334, 335, 336, 337
- Goodman, L. A., 283, 316, 334, 335
- Gordon, R. A., 84, 335
- Gorsuch, R. L., 173, 335
- Growth, 223, 275, 281, 297, 337, 341
- Gruvaeus, G. T., 174
- Guilford, J. P., 189, 335
- Hakansson, P. A., 192, 339
- Haller, A. O., 136, 209, 212, 213, 327, 334
- Halo, 136
- Hamilton, V. L., 156, 166, 335
- Hargens, L. L., 215, 223, 335
- Harmon, H. H., 53, 143, 150, 170, 173, 335
- Harris, R. J., 84, 331, 335
- Hauser, R. M., 54, 90, 147, 224, 225, 330, 335
- Hays, W. L., 53, 335
- Head Start, 205, 206, 207, 208
- Heise, D. R., 123, 136, 217, 219, 282, 335, 336
- Heyns, B., 336
- Heywood case, 143
- Holzinger, K., 146, 340
- Homogeneity, between constructs, 175, 176, 177, 181
within constructs, 174, 176, 177, 179, 181, 186
- Homogeneous stability, 222
- Homoscedasticity, 62, 65
- Hotelling-Williams test, 277
- Huck, S. W., 262, 336
- Hull, C. H., 338
- Humphreys, L. G., 215, 336
- Identification, 43, 52
minimum condition of, 43, 44, 45, 49, 170, 171, 173, 190, 252
single factor model, 53, 140, 142, 144, 149, 154, 156, 159, 160, 161, 163, 164, 178, 190, 331
- Indicators, 86, 87, 98, 105, 140, 141, 143, 144, 158, 159, 164, 168, 171, 172, 173, 174, 175, 176, 177-181, 184-187, 194, 196, 199, 201-203, 205, 206, 207, 209, 211, 212, 213, 224, 249, 328
- Insko, C. A., 251, 336
- Interaction, 18, 20, 231, 232, 235, 236, 238, 239, 244, 246, 247, 250, 252, 253, 262, 269, 275, 278, 317, 318, 332, 336, 338
- Jaccard, J., 189, 336
- Jacobson, J., 20, 336
- Jencks, C., 54, 336
- Jenkins, J. G., 338
- Jennings, M. K., 284, 330
- Jessor, R., 284, 298, 336
- Jessor, S., 284, 298, 336
- Johnsson-Smargdi, V., 192, 339
- Jöreskog, K. G., 89, 108, 153, 174, 186, 188, 195, 200, 217, 283, 316, 317, 336, 337, 341
- Just-identification, 43, 44, 46
- Kant, I., 5
- Kelley, T. L., 169, 337
- Kenny, D. A., 51, 54, 135, 146, 179, 185, 193, 202, 263, 275, 276, 282, 296, 298, 301, 303, 305, 330, 331, 333, 337
- Kerchoff, A. C., 60, 67, 68, 88, 91, 118, 337
- Kidder, L. H., 293, 337
- Kidder, R. L., 293, 337
- Kirby, D., 316, 331

- Kruskal, J. B., 314, 337
- Lawley, D. W., 153, 338
- Layton, B. D., 336
- Lazarsfeld, P. F., 283, 338
- Least squares, iii, 59, 61, 65, 78, 94, 112, 113, 114, 116, 118, 130, 132, 200, 315, 329
 generalized, 329
 two-stage, iii, 94, 112, 113, 114, 118, 132, 200
- Levin, J. R., 7, 338
- Linearity, 246, 269, 312, 313, 323, 331
- Linn, R. L., 89, 217, 262, 276, 316, 317, 338, 341
- LISREL, 200, 201, 212, 221, 253, 337
- Lord, F. M., 81, 262, 265, 271, 272, 338
- Lundmark, J., 189, 336
- McLean, R. A., 262, 336
- McNemar, Q., 47, 296, 338
- Magidson, J., 205, 338
- Malt, B., 148
- Marascuilio, L. A., 7, 338
- Marwell, G., 318, 331
- Matrix, covariance, 25, 26, 49, 170, 182, 183, 195, 196, 198, 199, 201, 219, 220
 design, 231
 identity, 198
 lower-triangular, 197, 198
 multitrait-multimethod, 50, 168, 193, 195, 330, 332
- Maxwell, A. E., 153, 338
- Measures, congeneric, 99, 100
 parallel, 100
 tau equivalent, 100
- Mediation, 54
- Michelson, S., 336
- Milburn, M. A., 181, 338
- Models, structural, 29, 30, 32, 53, 56, 66, 161, 194, 198, 201, 202, 203, 209, 252
- Multicollinearity, 51, 76, 84, 85-87, 232, 242, 246, 251, 313, 323, 330
- Nanda, H., 333
National Enquirer, 82
- Nie, N. H., 338
- Novick, M. R., 81, 338
- Outliers, 17, 18, 19, 20, 330
- Overall, J. E., 244, 246, 268, 338, 339
- Overidentification, 49, 78, 116, 117, 199, 321, 322, 323, 325
- Overidentifying restriction, 45, 48, 49, 56, 78, 79, 114, 116, 117, 119, 134, 146, 160, 161, 169, 218, 219, 275, 299, 322, 323, 324
- Path, analysis, ii, 24, 26, 33, 35, 36, 40, 95, 96, 108, 118, 330, 333, 334, 335
 coefficient, 43, 66, 90, 102, 239
 diagram, 33, 34, 35, 39, 42, 46, 52, 56, 67, 70, 73, 81, 96, 98, 105, 120, 124, 125, 126, 133, 135, 189, 193, 194, 229, 239, 249, 250, 252, 326
- Pearson-Filon test, 306-307
- Pelz, D. C., 282, 339
- Peters, C. C., 296, 339
- Pinner, B. E., 336
- Porter, A. C., 262, 263, 271, 339
- Portes, A., 136, 209, 212, 213, 327, 334
- Power, 76, 84, 85, 87, 88, 114, 116, 181, 184, 241, 246, 250, 276, 296, 314, 316, 318, 324, 332
- Price, G. G., 280, 333
- Rajaratnam, N., 333
- Random assignment, 1, 29, 228, 229, 285

- Reduced form, 54, 56, 90, 91, 92, 93, 125, 126, 136
 Regression coefficient, 23, 24, 59, 61, 72, 139, 234, 239, 262, 263, 264, 265, 271, 272
 stepwise, 68, 74
 Reliability, 95, 96, 98, 100-105, 113, 114, 139, 164, 165, 166, 196, 219, 222, 223, 251, 258, 260, 263, 264, 265, 266, 271-273, 277, 278, 279, 294, 296, 301, 326, 327, 328, 331, 335
 Reskin, B. F., 215, 223, 335
 Rickard, S., 282, 339
 Rock, D. A., 89, 341
 Rogosa, D. R., 280, 333
 Rosengren, D. E., 192, 193, 339
 Rosenthal, R., 274, 339
 Rosnow, R. L., 274, 339
 Rotation, 7, 139, 153, 170, 173
 Rozelle, R. M., 282, 289, 292, 293, 294, 339
 Rubin, D. B., 268, 284, 339
 Rubin, Z., 268, 284, 339
- Sartre, J. P., 13, 339
 Schoenberg, R., 160, 333
 Schwartz, S. H., 158, 159, 340
Sesame Street, 303, 332
 Shaffer, A., 303, 332
 Shaud, K. F., 336
 Sherin, R. J., 332
 Sibley, L., 235, 239, 340
 Simplex, 215, 216, 336
 quasi-, 216
 Simultaneity, 120
 Smith, H., 315, 333
 Smith, M., 336
 Smith, M. L., 8, 340, 336
 Snedecor, G. W., 89, 340
- Snyderman, P., 293
 Sörbom, D., 283, 337
 Spearman, C., 146, 157, 158, 166, 340
 Specification, 11, 13, 28, 29, 35, 51, 56, 70, 76, 78, 80, 81, 101, 102, 106, 113-115, 117, 138, 142, 143, 157, 159, 161, 162, 173, 177, 178, 179, 184, 195, 200, 204, 208, 214, 222, 226, 230, 249, 250, 259, 267, 318, 320, 321, 323, 324, 326, 327
 error, 29, 51, 56, 70, 78, 80, 81, 117, 138, 142, 143, 159, 161, 177, 179, 184, 200, 204, 208, 214, 222, 318, 323, 324
 Spiegel, D. K., 238, 239, 244, 246
 Spuriousness, 5, 90, 284, 285-290, 293-296, 298, 303, 308
 Standardization, 24, 33, 318-322
 Standardized change score analysis, 260, 263, 264, 265, 276
 Stanley, J. C., 2, 84, 135, 230, 256, 280, 284, 308, 317, 325, 332, 40
 Stationarity, 273, 275, 281, 287-299
 quasi-, 222, 301, 302, 307
 Staw, B. M., 228, 340
 Steinbrenner, K., 338
 Steiner, I. D., 11, 340
 Stroebe, W., 251, 336
 Structural modeling, limits of, 8-10
 need for, 5-8
 Structural parameter, 30, 51, 53, 98, 116, 225, 228, 282, 319, 322, 323, 325
 Suppes, P., 4, 284, 340
 Synchronicity, 288-290, 294
- Tamkin, G., 303, 332
 Tatsuoka, M. M., 147, 340
 Tessler, R. C., 158, 159, 250, 330, 340
 Thistlewaite, D. L., 267
 Thompson, V. D., 336
 Tracing rule, 33, 37, 38, 39-42, 48, 53, 68, 70, 73, 91, 92, 105, 123-127, 130, 141, 172

- modified, 125, 127, 133
- Transformation, 65, 241, 276, 296, 314, 315, 337
- Trimming, 19
- Tukey, J. W., 19, 77, 315, 318, 340
- Tyron, R. C., 173, 340

- Underidentification, 49-51, 220
 - empirical, 51, 52, 86, 106, 115, 132

- Validity, construct, 230, 290
 - convergent, 188, 325
 - discriminant, 77, 86, 188, 325
 - external, 84, 317, 325
 - internal, 325
- Vanishing tetrad, 116, 119, 145, 146, 147, 161, 175, 185, 208, 218, 219, 272, 273, 299, 324, 337
- Van Voorhis, W. R., 296
- Variable, dependent, 31, 62, 158, 228, 229, 234, 236, 238, 240, 241, 245, 247, 249, 250, 251, 254, 257, 260, 262, 263, 264, 276, 277, 285, 314
 - dummy, 82, 206, 229, 232, 236, 238, 239
 - endogenous, 31, 32, 34, 36, 37, 54, 55, 56, 65, 68, 72, 74, 82, 89, 93, 101, 102, 104, 112, 113, 115, 117, 118, 145, 197, 201, 213, 228, 300, 314, 315, 319, 327, 328
 - exogenous, 33, 36, 49, 50, 56, 87, 103, 124, 132, 198, 200, 228, 229, 300, 316, 327
 - independent, 14, 31, 228, 229, 230, 231, 233, 236, 238, 240, 241, 243, 249, 250, 251, 285, 313
 - instrumental, 50, 105, 106, 107, 108, 112, 114, 115, 117, 119, 199, 326
 - marker, 157
 - nominal, 233, 312, 316
 - ordinal, 246, 312, 316
 - suppressor, 81, 245
 - unmeasured, iii, 35, 105, 138, 143, 147, 166, 168, 193, 194, 195, 199, 205, 249, 258, 272, 287, 301, 320, 328, 334
- Variance, 21-24
 - common, 152
 - error, 33, 43, 95, 100, 139, 157, 174, 196, 219, 220, 221, 222, 320
 - factor, 153, 154, 170, 173, 183
 - method, 177, 188, 189, 193
 - partitioning,
 - specific, 301

- Weber, J., 189, 303, 332, 336
- Weber, S. J., 189, 303, 332, 336
- Weight, beta, 47, 60, 62, 64, 66, 67, 68, 75, 86, 93, 101, 139, 163, 319
 - unit, 164, 165, 166, 180, 205
- Weisberg, H. I., 276, 331
- Weiss, C. H., 8, 256, 341
- Werts, C. E., 89, 217, 262, 276, 316, 317, 338, 341
- Wiley, D. E., 33, 217, 219, 320, 338, 339, 340, 341
- Wiley, J. A., 33, 217, 219, 320, 338, 339, 340, 341
- Williams, E. J., 277, 341
- Windahl, S., 192, 339
- Wittgenstein, L., 329
- Woodward, J. A., 268, 339
- Wright, S., 33, 341

- Yee, A. H., 283, 292, 341
- Yerkes-Dodson law, 20