DOCUMENT RESUME

ED 186 704 CE 025 291

AUTHOR Hotchkiss, Lawrence

TITLE Differential Equation Methodology Applied to Career Decisions and Status Attainment Processes: Conceptualization and Calculation.

INSTITUTION Ohio State Univ., Columbus. National Center for Research in Vocational Education.

SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.

PUB DATE 79

GRANT OB-NIE-G-78-0211

NOTE 139p.: Some figures will not reproduce well due to small type. For a related document see CE 025 292.

EDRS PRICE MF01/PC06 Plus Postage.

DESCRIPTORS *Career Development; Concept Formation; Decision Making; Educational Research; *Expectation; *High School Students; Longitudinal Studies; *Mathematical Formulas; Mathematical Models; Research Design; *Research Methodology; Socioeconomic Status; Statistical Analysis; Theories

IDENTIFIERS *Differential Equations

ABSTRACT

This document was prepared in connection with a three-year longitudinal study of career expectations of high school students conducted to explain scientifically the process by which youth form career expectations, e.g., educational, occupational, and income expectations. Divided into six chapters, this document contains a theoretical rationale for a differential equation model of the process by which career expectations of youth evolve and presents a detailed explication of the technical information needed to use the model. Chapter 1 presents the purpose and overview of this report. Chapter 2 contains a theoretical and conceptual discussion of the use of differential equations to represent career planning processes. Chapter 3 presents the basic concepts of selected mathematical and statistical topics, which are discussed in further detail in chapters 4 and 5. Chapter 4 develops the mathematics of differential equations, presents a justification for using ordinary least squares in the statistical analysis, and describes a computer program that can be used to estimate coefficients of the differential equation system. The fifth chapter draws on the technical materials presented in chapters 3 and 4 to describe interpretations of differential equation systems applied to career expectation variables. A summary is presented in Chapter 6. (Author/BM)

DIFFERENTIAL EQUATION METHODOLOGY APPLIED TO CAREER DECISIONS AND STATUS ATTAINMENT PROCESSES:

CONCEPTUALIZATION AND CALCULATION

Lawrence Hotchkiss

The National Center for Research in Vocational Education
The Ohio State University
1960 Kenny Road
Columbus, Ohio 43210

1979

The National Center Mission Statement

The National Center for Research in Vocational Education mission is to increase the ability of diverse agencies, institutions, and organizations to solve educational problems relating to individual career planning, preparation, and progression. The National Center fulfills its mission by:

- Generating knowledge through research
- Developing educational programs and products
- Evaluating individual program needs and outcomes
- Installing educational programs and products
- Operating information systems and services
- Conducting leadership development and training programs

Report Prepared Under Contract Number

OB-NIE-G-78-0211

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

National Institute
of Education

TABLE OF CONTENTS

LIST OF FIGURES

FOREWORD

This volume is a text describing how to apply differential equations in panel studies of career-decision making processes. Two contributions are particularly noteworthy. First, the report shows how a dynamic mathematical model can be formulated to reflect the important theoretical idea that career expectations gradually evolve in a continuous process over time. Secondly, the report advocates use of the dynamic theoretical model to generate forecasts of career expectations. Such forecasts provide much stronger tests of theory than methods currently in general use. Procedures for calculating and evaluating the forecasts are described in detail in the text. The report is one product of a three-year longitudinal study of developing career expectations in which the differential-equation methodology will be applied.

Numerous persons deserve appreciation for their roles in completing the manuscript, particularly, the author, Larry Hotchkiss, and his staff, Lisa Chiteji, Alireza Rabbani, and Nancy Robinson. Exceptionally capable technical reviews of the report were submitted by Patrick Doreian, Edward Fink, and Michael Black. In addition, insightful comments and suggestions related to the report have come from several persons, particularly Evans Curry, Archibald Haller, Steven Picou, Richard Campbell, and Robert Leik. Frank Pratzner (Division Associate Director), Harry Drier (Program Director) and Robert Wise (NIE Project Officer) lent continuing support to the work.

ABSTRACT

This volume contains a theoretical rationale for a differential equation model of the process by which career expectations of youth evolve and presents a detailed explication of the technical information needed to use the model. Three important advantages of the differential equation model are reviewed. First, in the differential equation model, the dependent variables are rates of change over time in career expectations of youth; hence, theory expressed by differential equations automatically accommodates change over time. In contrast, most current models express static theory in which cross-sectional differences among individuals are the subject of inquiry. Secondly, the differential equation model expresses all causal feedback suggested by theoretical analysis; whereas, alternative models frequently neglect causal feedback. Finally, the differential equation model provides a built-in facility for carrying out empirical tests of theory by evaluation of the accuracy of forecasts made by the model.

The technical explication includes: (1) a review of mathematical and statistical material needed to understand the differential equation model, (2) derivation of calculating formulas for application of the differential equation model, and (3) description of a computer program written in conjunction with this volume. The computer program is designed to carry out calculations needed for application of the differential equation model.

CHAPTER 1

INTRODUCTION

Purpose of Volume

This volume is written in connection with a three-year longitudinal study of career expectations of high school students; the study is funded by the National Institute of Education. The intent of the longitudinal study is to contribute to scientific understanding of the process by which youth form career expectations such as educational, occupational, and income expectations. To help achieve this intent, the theory for the study is expressed by systems of differential-equations. The differential-equation methodology achieves three desirable outcomes:

(a) The methodology expresses verbal theory that career choice is a continuously evolving process rather than an event

(b) The methodology facilitates examination of two-directional cause-and-effect relationships

(c) The differential equation system expresses theory in such a manner that forecasting (prediction) to any point along a continuous time scale is an immediate consequence of the theory, thus, predictive tests of theory are encouraged.

Clearly, the differential-equation methodology holds a much broader appeal than the application to study of career expectations. It presents a vehicle for expressing the dynamics of natural systems in a manner that is unparalleled by standard practice in the application of statistical procedures and contains the essential elements for incorporating predictions (forecasting) into the expression and testing of theory. Examples of topics for which these features are useful include migration and economic growth, organizational behavior, voting behavior and political attitudes, and criminology. In spite of the potentially wide application of the methodological procedures, they have not been applied very often, primarily because of poor dissemination to substantive researchers, including those working on development of career expectations. This volume, therefore, is designed to present a clear explication of the differential-equation methodology written in a style permitting use by researchers with moderate amount of technical training. The intent is to avoid a purely technical exercise; rather, the focus centers on the juxtaposition of techniques with substantive issues.

A number of years ago Robert Merton wrote:

> This limited account has, at the very least, pointed to the need for a closer connection between theory and empirical research. The prevailing division of the two is manifest in marked discontinuities of empirical research, on the one hand, and systematic theorizing unsustained by empirical test, on the other (Merton, 1957: 99).

One goal of this volume is to help reduce the division between method and theory.

Outline of the Volume

There are six chapters in the volume. Chapter 2 contains a theoretical and conceptual discussion of the use of differential equations to represent career planning processes. Chapter 3 presents a review of selected mathematical and statistical topics needed for Chapter 4 and Chapter 5. The intent of Chapter 3 is to communicate the basic concepts, omitting rigorous proofs. Chapter develops the mathematics of differential equations, presents justification for using ordinary least squares in the statistic. analysis, and describes a computer program that can be used to estimate coefficients of the differential equation system. Chapter 5 draws on the technical material presented in Chapters 3 and 4 to describe interpretations of differential-equation systems applied to career expectation variables. Topics include comparison of interpretations of effects based on cross-lagged regressions to interpretations based on differential equations, analysis of the time path of career expectations, including oscillations and equilibrium conditions, development of standardization methods for the coefficients of differential equations, and presentation of a generalized correlation for assessing accuracy of forecasts. The final chapter summarizes the volume.

CHAPTER 2

CONCEPTUALIZATION AND INTERPRETATION

The purpose of this chapter is to introduce the idea that differential equations can be used to express some important concepts and theory related to development of educational and occupational expectations. The first section of the chapter explicates a simplified cross-sectional path model of career expectations. The path model serves as a reference helping to motivate the succeeding discussion of differential equations; the features of the differential equation systems are contrasted to those of the cross-sectional path model. A second section presents a brief theoretical rationale for using differential equations in research on development of career expectations. The third section develops a specific differential equation model to describe the variables included in the path model in section one. In the fourth section, interpretation of the differential equation system is discussed. The final section summarizes the chapter. The chapter is not intended to contain a thorough technical justification for every point. Technical descriptions are contained in Chapters 4 and 5.

A Cross-Sectional Path Model

Although the study of which this volume is a part addresses knotty theoretical and conceptual issues that cross the boundaries of academic disciplines, the guiding paradigm of the research is drawn from status attainment work in the field of sociology. A path model of the processes by which parental statuses are transmitted to offspring forms the fundamental point of departure. ~~This section~~ section describes a simplified version of such a model which can be used as a reference when examining a differential-equation model of the same variables. To simplify the exposition, the example is less complicated than many path models in the published empirical literature related to status attainment processes (e.g., Hauser, 1972). Nevertheless, the example does capture the major conceptual and theoretical ideas in the status attainment literature.

For illustrative purposes, in developing the model it will be helpful to identify a small set of variables commonly used in status attainment research and to classify them into three broad categories, as follows:

Exogenous variables

Parental socioeconomic status	(x_1)
Mental ability	(x_2)

Intervening variables

Academic performance	(y_1)
Significant other educational expectation of ego	(y_2)
Significant other occupational expectation of ego	(y_3)
Ego's own educational expectation	(y_4)
Ego's own occupational expectation	(y_5)

Attainment variables

Educational attainment	(z_1)
Occupational attainment	(z_2)

For purposes of illustrating the model, precise definitions and operational procedures for these variables are not needed, but some description of the variables and terminology is warranted, especially for scholars who are not familiar with status attainment research.

First, note that the variables are classified into three categories: exogenous variables, intervening variables, and attainment variables. Exogenous variables refer to background characteristics that affect career planning but are not affected by any other variables in the model. In this example parental socioeconomic status and mental ability are classified as exogenous variables. Parental socioeconomic status nearly always includes father's educational level and father's occupational status. Family income, mother's education and mother's occupational status also are included frequently. In most of the recent research, these status variables are treated empirically as distinct variables (e.g., Hauser, 1972; Sewell and Hauser, 1975; Featherman and Hauser, 1977), but for the illustration presented here, the simplicity of a single, aggregate parental-status variable (SES) is preferable. Mental ability is operationalized by use of a standardized ability test. In recent work it is treated as dependent on SES rather than as an exogenous variable, but this treatment is difficult to justify and does not alter most substantive conclusions very much. For current purposes, it is useful to present the example with two exogenous variables.

The intervening variables refer to educational and occupational expectations. Academic performance (school grades) is also included. To avoid confusion with the Freudian concept of "ego", it is important to note that in the status attainment literature the term ego refers to the individual on whom attention is centered; the term is used in contrast to the term "significant other" which refers to an individual other than ego who may exert some influence on ego's attitudes and/or behavior. For example, suppose Jack is a high school sophomore, the level of education that he expects to achieve illustrates the variable identified as "ego's" own educational expectation. The level of education that Jack's mother expects him to achieve is an example of the variable called "significant other educational expectation of ego."

The variables called attainment variables refer to the socioeconomic achievement of ego. Achievement should be carefully differentiated from aspiration or expectation.

It should be reiterated that the illustration is not a comprehensive model. In published research, parental SES is generally disaggregated into its components, a more detailed list of significant-other variables generally is included, and, increasingly, income attainments and, sometimes income expectations also are studied. The important theoretical, conceptual and methodological features of status-attainment research can be illustrated with this list of variables, however.

Figure 1 displays a path diagram of the basic theory. A straight, single-headed arrow denotes a cause-and-effect relation; whereas, the curved, double-headed arrows indicate unanalyzed correlations (i.e., no causal relationship specified). It should be noted that the two-directional arrows connecting pairs of intervening variables depart from conventional path diagrams. Normally, these arrows are omitted. They are included here because their presence more accurately reflects the state of the theory than their omission. Omitting the curved arrow connecting, say significant other educational expectation to significant other occupational expectation implies that neither one affects the other. It is more accurate to state that the theory lacks the power to specify a one-directional relationship, and since the statistical methodology demands the assumption of one-directional effects, the cause and effect relationship generating the correlation between the two variables is left unanalyzed.

Note especially three features of the model. First, there are no two-directional effects hypothesized. Secondly, the model for the intervening variables ignores change over time. Finally, the three subsets of variables comprise a "block-recursive" system. That is, exogenous variables may affect the other two "blocks" of variables but are unaffected by them, and intervening variables may affect attainment variables but are not affected by attainment variables. (The intervening variables normally are defined over the adolescent years.)

As represented in figure 1, there is no direct effect of the exogenous variables on attainment variables, because there are no arrows leading directly from parental SES or mental ability to attainment. This is an important theoretical feature of the model. By hypothesis, intervening variables comprise the mechanisms by which status is transmitted from parent to child. In addition, by hypothesis, educational achievement is the route by which parental SES, mental ability, and career expectations get translated into occupational status. In simplified form, the model may be diagrammed as follows.

SES / Mental Ability	→	Grades / Expectations	→	Education	→	Occupational Status

Of course, data seldom support this parsimonious viewpoint in every particular, but the theoretical model is approximated frequently in empirical study (e.g., Sewell and Hauser, 1975; Rehberg and Rosenthal, 1978; Duncan, Featherman, and Duncan, 1972; Alexander and Eckland, 1975; Cosby, et al., 1979).

The path diagram is a heuristic picture of an equation system. The major hypotheses are expressed by a set of linear equations, in the following manner:

(1a) $y_1 = a^*_{10} + a^*_{11} x_1 + a^*_{12} x_2 + u_1$

(1b) $y_2 = a^*_{20} + a^*_{21} x_1 + a^*_{22} x_2 + b^*_{21} y_1 + u_2$

(1c) $y_3 = a^*_{30} + a^*_{31} x_1 + a^*_{32} x_2 + b^*_{31} y_1 \quad u_3$

(1d) $y_4 = a^*_{40} + a^*_{41} x_1 + a^*_{42} x_2 + b^*_{41} y_1 + b^*_{42} y_2 + b^*_{43} y_3 + u_4$

(1e) $y_5 = a^*_{50} + a^*_{51} x_1 + a^*_{52} x_2 + b^*_{51} y_1 + b^*_{52} y_2 + b^*_{53} y_3 + u_5$

(1f) $z_1 = q^*_{10} + q^*_{11} x_1 + q^*_{12} x_2 + p^*_{11} y_1 + p^*_{12} y_2 + p^*_{13} y_3 + v_1$

(1g) $z_2 = q^*_{20} + q^*_{21} x_1 + q^*_{22} x_2 + p^*_{21} y_1 + p^*_{22} y_2 + p^*_{23} y_3 + s^*_{21} z_1 + v_2$

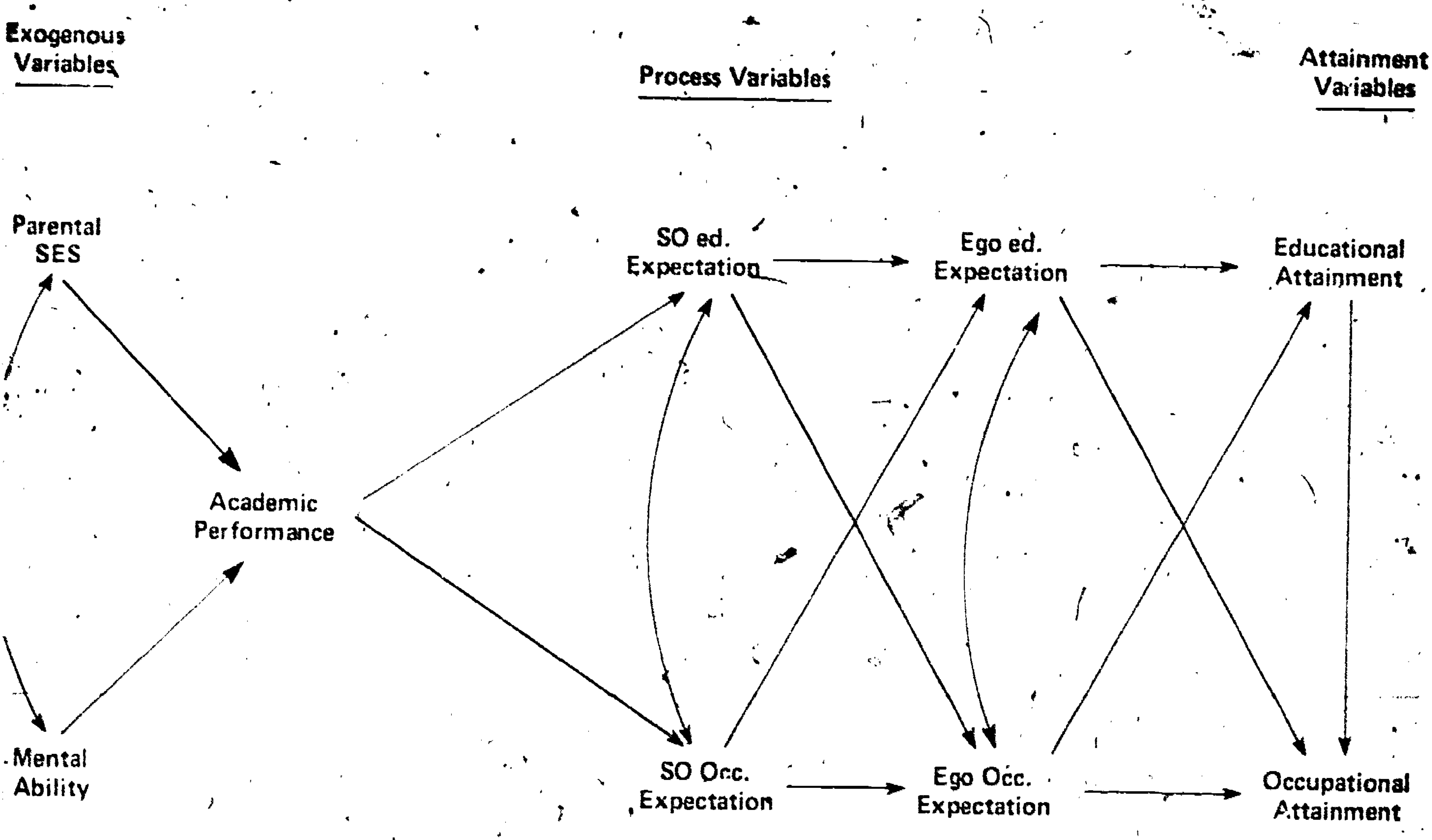

Figure 1. Illustrative Path Model of Status Attainment Processes.

Notes: SES = socioeconomic status

SO = significant other

ed. = education

Occ. = occupation

Where the a^*_{ij}, b^*_{ij}, q^*_{ij}, p^*_{ij}, and s^*_{ij} are constants (path coefficients) indexing the effect of variable j on variable i, the u_i and v_i are unmeasured "disturbance" variables. The variances of the disturbance variables index the degree to which dependent variables are controlled or influenced by the measured variables hypothesized to affect them; the smaller the variance of a disturbance variable, the larger the influence of the measured variables.

It is important to notice that equation system (1) does not correspond exactly to the path diagram in figure 1. First, for simplicity the disturbance variables are omitted from the path diagram; this omission marks a mild departure from convention. Secondly, there are more path coefficients (a*s, b*s, q*s, p*s and s*s in equation system (1)) than there are paths in figure 1. The path diagram oversimplifies to emphasize the theory; whereas, the equations include all paths that typically would be calculated in empirical application. An indication of how well the theory fits can be gained by observing how close to zero are the calculated paths corresponding to missing arrows in figure 1. If the theory were sufficiently convincing, the postulated zero paths could be assumed zero, and the information that some paths are zero could then be used to improve estimation of the other paths. In the literature, however, it is generally assumed that the theory is not strong enough to merit setting any paths to zero a priori.

There are two features of the model that are important in the present context. First, there are no two-directional cause-and-effect relationships hypothesized. This means that in no case is there a path coefficient included to index the effect of y on x if there is a path coefficient included to index the effect of x on y. Thus, for example, b^*_{21} is present in the model; its presence indicates a hypothesized effect of academic performance on the level of education that significant others expect ego to achieve. On the other hand, b^*_{12} is not present in the model; its absence indicates an implicit hypothesis that ego's academic performance is not influenced by the level of educational expectation held for him/her by significant others (such as parents).

The second important feature of the model is its static conceptualization of the process variables. Notice that changes over time in the process variables are not part of the model. The model is about cross-sectional differences among individuals not about changes over time for a particular individual. Although this point is not new, it does not appear to be widely understood. This distinction between change over time and cross-sectional differences among individuals forms an important part of the justification for applying differential equations to the study of career planning. In consequence, the next few paragraphs develop an algebraic treatment of the appropriate interpretation of path coefficients calculated from cross-sectional data.

The purpose of this discussion is to review the meaning of the phrase: statistical "control" by linear regression, when regression methods are applied to cross-sectional data. The commentary also applies to the concept of direct effect in path analysis. The fundamental idea is to examine the impact of a single independent variable on a dependent variable while holding all other independent variables "constant." For this exercise, assume that the dependent variable is ego's educational expectation (y_4) and that one wishes to examine the effect of significant others' educational expectation of ego (y_2) while "controlling" or "holding constant" the other independent variables affecting y_4.

To develop this idea, consider two individuals who have the same mental ability, same parental SES, same significant-other occupational expectation, and the same level of the "disturbance" variable, but who differ by <u>one unit</u> with respect to significant-other educational expectation. Suppose one wished to discover to what degree two such individuals differ in the level of educational expectation (y_4). Define the following elaboration of the notational scheme:

x_{11}, x_{12} = parental SES for person 1 and person 2, respectively

x_{21}, x_{22} = mental ability for person 1 and person 2, respectively

y_{11}, y_{12} = academic performance (grade average) for person 1 and person 2, respectively

y_{21}, y_{22} = significant-other educational expectation for person 1 and person 2, respectively

y_{31}, y_{32} = significant-other occupational expectation for person 1 and person 2, respectively

y_{41}, y_{42} = ego's own educational expectation for person 1 and person 2, respectively

u_{41}, u_{42} = disturbance term for person 1 and person 2, respectively

Keeping in mind the conditions just established, insert this notation into equation (1d) for person 1 and person 2:

$$(2a) \quad y_{41} = a^*_{40} + a^*_{41}x_{11} + a^*_{42}x_{21} + b^*_{41}y_{11} + b^*_{42}y_{21} + b^*_{43}y_{31} + u_{41} \text{ (person 1)}$$

$$(2b) \quad y_{42} = a^*_{40} + a^*_{41}x_{12} + a^*_{42}x_{22} + b^*_{41}y_{12} + b^*_{42}y_{22} + b^*_{43}y_{32} + u_{42} \text{ (person 2)}$$

Subtracting both sides of equation (2b) from equation (2a) yields:

$$(3) \quad y_{41} - y_{42} = (a^*_{40} - a^*_{40}) + a^*_{41}(x_{11} - x_{12}) + a^*_{42}(x_{21} - x_{22}) + b^*_{41}(y_{11} - y_{12}) + b^*_{42}(y_{21} - y_{22}) + b^*_{43}(y_{31} - y_{32}) + (u_{41} - u_{42})$$

Since we have constructed the situation so that all independent variables are "constant" except significant-other educational expectations of ego and have stipulated a unit difference between the two person's significant-other educational expectations, the following facts hold:

$$x_{11} - x_{12} = x_{21} - x_{22} = y_{11} - y_{12} = y_{31} - y_{32} = u_{41} - u_{42} = 0,$$

and $y_{21} - y_{22} = 1.0$. Equation (3) therefore reduces to:

$$y_{41} - y_{42} = b^*_{42}$$

In words, ceteris paribus, a unit difference between to persons on significant-other educational expectation generates a difference in ego's own educational expectation equal to the path coefficient (b^*_{42}). More generally, if $y_{21} - y_{22} \neq 1.0$ and is not zero, the ratio of the difference in the dependent variable to the difference in the independent variable, "everything else being constant" is the path coefficient, viz:

$$(y_{41} - y_{42})/(y_{21} - y_{22}) = b^*_{42}.$$

This is the meaning of statistical "control" by regression in cross-sectional data. For the present discussion, it is important to note that when path analysis is applied to cross-sectional data, the path coefficients <u>index differences between</u>

individuals, not changes over time. The relationship between individual differences and change over time is seldom explicit (see, however, Coleman, 1968 and Hout and Morgan, 1975). One of the virtues of differential-equation analysis is that it can explicate this relationship.

Theoretical Basis for a Differential Equation Model

Theoretical Background

The developmental theme of forming career orientations is pervasive in the vocational psychology literature. For example, the idea is repeatedly stated in the writings of Donald Super, who at one point expressed it in the following terms:

> PROPOSITION 1. Vocational development is an ongoing continuous, and generally irreversible process. Vocational preferences and competencies ... change with time and experience, making choice and adjustment a continuous process (Super, et al., 1957: 89, emphasis in the original).

Similarly, Ginzberg and associates state:

> Our basic assumption was that an individual never reaches the ultimate decision at a single moment in time, but through a series of decisions over a period of many years; the cumulative impact is the determining factor (Ginzberg, et al., 1951: 27).

This theme can also be found in the writings of many other theorists (e.g., Pietrofesa and Splete, 1976; Tiedeman, 1961; Rodgers, 1966; Blau, et al., 1956). After reviewing several "macro-theories," Picou, Curry and Hotchkiss indicated the following general characterization of the theoretical literature.

> The macro-theoretical approaches reviewed above have several common themes. First, all of the above theorists have implicitly or explicitly noted the developmental character of occupational choice and placement. The problems of career choice and attainments are clearly limited to a life-cycle framework. Labor market entry and career patterns tend to be viewed in conjunction with individual maturation and growth (Picou, Curry and Hotchkiss, 1976: 12).

It is undoubtedly obvious to most of the research community that career orientations are formed in a gradual process over time, yet operational procedures commonly found in empirical study do not reflect this obvious point. While the theoretical literature has been helpful in pointing out the dynamic nature of the process, a proposition stated in the general terms such as those used by Super and associates, quoted above, is of little use in empirical research. It is imperative that the general idea be translated into exact hypotheses. Again quoting Merton,

> Much of what is described in textbooks as sociological theory consists of general orientations toward substantive materials. Such orientations involve broad postulates which indicate types of variables which are somehow to be taken into account rather than specifying determinant relationships between particular variables. Indispensable though these orientations are, they provide only the broadest framework for empirical inquiry (Merton, 1957: 87-88).

Although most status-attainment models have been confined to one-directional causal systems (for notable exceptions see Hout and Mogran, 1976; Woelfel and Haller, 1971; and Nolle, 1973) there is ample reason to expect that several of the key process variables in career planning exercise reciprocal effects on each other. For example, as described in the preceeding section in most models of the process, academic performance and significant-other variables are assumed to affect youths' career orientations; effects of youths' career orientations on academic performance and significant other variables are assumed zero. Yet it is plausible that students adjust their academic efforts in response to their educational and occupational ambitions. It is also plausible that significant others, in part, adjust their expectations of youth to conform to the career orientations that the significant others know the youth hold (see Curry, et al., 1976 for a discussion of these issues). Additionally, career orientations probably exhibit reciprocal effects on each other; for example, educational expectation probably affects occupational expectation, and vice versa (see Kerckhoff, 1971). Consequently, the dynamic model must permit reciprocal effects among these variables.

In the next subsection, simultaneous differential equation models are proposed as a general methodology for expressing the dynamic, reciprocal aspects of forming career orientations. The discussion is presented via a simple example that parallels the example of a path model reviewed in the present section.

An Example of a Dynamic Model

Virtually all structural-equation models of status attainment processes have been linear, and the few available tests of linearity assumptions show only minor departures from linearity (e.g., Gasson, Haller and Sewell, 1972; and Wilson and Portes, 1975). Consequently, as a first effort to translate current models into dynamic two-directional models it is sensible to adhere to the linear postulate.[1]

The simultaneous, differential equation model assumes continuous time and continuous change in the endogenous (intervening and attainment) variables. There is little difficulty in conceptualizing the intervening variables as changing continuously; this is a major emphasis in the theoretical literature cited above. The attainment variables, on the other hand, cannot be so easily conceptualized as changing continuously. Occupational change obviously occurs in discrete jumps, and educational attainment exhibits similar discontinuous change especially at the end of degree programs. These difficulties cannot be resolved in this volume; hence, the illustration will be confined to the exogenous variables and the intervening variables, and specification of a model including attainment variables will be left for future work.

With the above considerations in mind, it is now plausible to formulate in dynamic terms the basic status attainment model. To do this some additional notation is needed. Let the variable symbols, x_i and y_i be identified by the preceding list of variables, let b_{ij} and c_{ij} be constants over time and over individuals, let $dy_i(t)/dt$ represent the instantaneous change in variable y_i for a small change in time, and let $u_i(t)$ be a disturbance term at time t. The dynamic model can now be written in the following form:

$$dy_1(t)/dt = a_{10} + a_{11}x_1 + a_{12}x_2 + b_{11}y_1(t) + b_{12}y_2(t) + b_{13}y_3(t) + b_{14}y_4(t) + b_{15}y_5(t) + u_1(t) \quad (4a)$$

$$dy_2(t)/dt = a_{20} + a_{21}x_1 + a_{22}x_2 + b_{21}y_1(t) + b_{22}y_2(t) + b_{23}y_3(t) + b_{24}y_4(t) + b_{25}y_5(t) + u_2(t) \quad (4b)$$

$$dy_3(t)/dt = a_{30} + a_{31}x_1 + a_{32}x_2 + b_{31}y_1(t) + b_{32}y_2(t) + b_{33}y_3(t) + b_{34}y_4(t) + b_{35}y_5(t) + u_3(t) \quad (4c)$$

$$dy_4(t)/dt = a_{40} + a_{41}x_1 + a_{42}x_2 + b_{41}y_1(t) + b_{42}y_2(t) + b_{43}y_3(t) + b_{44}y_4(t) + b_{45}y_5(t) + u_4(t) \quad (4d)$$

$$dy_5(t)/dt = a_{50} + a_{51}x_1 + a_{52}x_2 + b_{51}y_1(t) + b_{52}y_2(t) + b_{53}y_3(t) + b_{54}y_4(t) + b_{55}y_5(t) + u_5(t) \quad (4e)$$

It is important to describe how these equations reflect the highly generalized hypothesis that formation of career expectations is a continuous process, and the somewhat more specific arguments that reciprocal effects may be observed among the process

1. Clearly, this argument must be viewed as an analogy, for, as noted in the preceding section of the text, the cross-sectional path coefficients and the parameters of the differential-equation model are not equivalent. I wish to thank Professor Robert Leik for bringing this point to my attention.

variables. The fact that the dependent variables are instantaneous change rates $[dy_i(t)/dt]$ with respect to time reflects the continuous nature of the process. Observing equation (4a), one can see that the instantaneous change in academic performance at time t $[dy_1(t)/dt]$ is a linear combination of the exogenous variables, SES (x_1), and mental ability (x_2), and of all the current values of the intervening variables $[y_i(t)]$, including the current value of academic performance itself $[y_1(t)]$. The second equation (4b) indicates that the change in significant-other educational expectation for ego $[dy_2(t)/dt]$ is also a partial function of the current values of all the intervening variables. Hence, change in academic performance is affected by significant-other expectation of ego, and change in significant-other expectation for ego is affected by academic performance. This type of reciprocal pattern can be observed among all the intervening variables. On the other hand, the exogenous variables are assumed fixed and unaffected by any intervening variable.

Stated in the differential equation form, these hypotheses cannot be tested against data, since instantaneous change rates cannot be observed. Consequently, it is necessary to integrate the system in order to find the relationship between observable variables at two points in time separated by a finite time interval. The integration procedure is described in Chapter 4.

It should be emphasized that equation system (4) is only an example. Technically, it is termed first-order ordinary simultaneous linear differential equations with constant coefficients. The term first order means that no second or higher order derivatives appear in the equations.[2] The term "ordinary" means that change rates are taken with respect to only one variable, time. The term simultaneous means that all of the relations must hold at the same time -- thus the equations express the concept of a system of variables. The term linear means that the algebraic form of the equations is linear. Finally, the term constant coefficients means that all the b_{ij} are constant over time. Each of these features forms a part of the implicit assumptions of the model. These assumptions are fairly restrictive, but it seems reasonable to start with a simple model and add complicating features as empirical results indicate. The point is that by suggesting this model, exploration of the wealth of available technology for describing dynamic systems of career-expectation variables has only begun.

---

2. The concept of first, second, and higher order derivatives are described in Chapter 3.

Interpretation and Discussion

This section is divided into three subsections. The first subsection discusses the manner in which change over time is incorporated into the conceptualization of the differential-equation system. The second subsection reviews the use of differential equations for forecasting (prediction). The concept of equilibrium and oscillation are introduced briefly. The final subsection discusses the relationship between the change coefficients (a_{ij} and b_{ij}) in equation system (4) and the path coefficients (a^*_{ij} and b^*_{ij}) in equation systems (1).

Conceptualization of Change

Consider the dependent variables in equation system (4); in each specific equation the dependent variable is a rate of change, $dy_i(t)/dt$. The numerator, $dy_i(t)$, stands for a change in the variable y_i over a very brief increment in time.

$$dy_i(t) = \lim_{t_1 \to t_0} y_i(t) - y_i(t_0)$$

Where $y_i(t_1)$ is the value of y_i at time t_1, $y_i(t_0)$ is the value of y_i at time t_0, and lim $t_1 \to t_0$ means in the limiting case as t_1 and t_0 are very close to being the same. Similarly, the denominator, dt, stands for a very small increment in time.

$$dt = \lim_{t_1 \to t_0} t_1 - t_0$$

Putting these two definitions together, one sees that the dependent variable is the ratio of change in y to change in time, as the change in time approaches, <u>but never quite reaches, zero</u>.

The hypothesis stated in the example is that this rate of change is a linear function of the exogenous variables (SES, MA) and the current value of <u>all</u> the intervening variables including the intervening variable whose change rate is the dependent variable. As indicated above, this particular functional form is hypothesized by analogy with current path models of the process. There are several obvious generalizations of the example in equation system (4). The simplest generalization is to permit the coefficients (a_{ij}, b_{ij}) to be functions of time, while maintaining the linear form. Thus, for example, if a_{51} were a positive number but tended to move to zero over time, then one would say that the influence of parental status on youth's occupational status expectation declines over the high school years. More generally, the change rates can be hypothesized to be arbitrary nonlinear functions of the current levels of the variables and of time. These generalizations will not be considered further in this volume.

Forecasting

Given the hypotheses in equation system (4) and observations at two points in time for a sample of individuals, it is possible to estimate all parameters of forecasting equations. Forecasts then can be expressed as functions of time alone. Procedures for determining numerical results are described in chapter 4. Once these forecasting functions are determined, a predicted value for each intervening variable for each individual can be generated for any point along a continuous time scale, prior to collecting data for the third and successive panels. These predictions, then, can be viewed as predictions derived from theory of the manner in which a system of career planning variables operate dynamically over time. The advantages of the built-in prediction formula are substantial. First, it provides the technique and justification for predictive tests of the theoretical model of status attainment processes that heretofore have not been carried out. Past work with longitudinal data has been confined to what might be termed "post-diction," that is, the values of the predicted variables have been used to help estimate the regression equations. Since the regression coefficients are chosen, ex post facto, to maximize the accuracy of the estimates of the dependent variables (in ordinary least squares) it is not surprising to find that moderately accurate estimates can be made. Projecting predictions prior to observation of the predicted variables could result in mean-square errors exceeding the cross-sectional variance of the dependent variables, but the cross-sectional variance is the maximum error variance for postdiction using OLS. Hence, the predictive test has a much stronger chance of failure and thus constitutes a much stronger test of the model (see chapter 5.)

In addition, the continuous-time capability of the predictive model permits the measurement intervals to vary. A period of one year might elapse between the first and second panels and the projections could be made for one and a half years, two years, or any time period beyond the second panel. This feature facilitates comparisons between studies using different measurement intervals, and also provides researchers freedom to determine measurement intervals to suit their needs.

It has been observed that a path analysis based on longitudinal data could also be used to generate predictions.[3] Certainly path analysis (or regression analysis) could be used in this way, but its application is seriously limited when compared to the differential equation model. Based on the path analysis alone, any predictions would have to be made over a time interval of exactly the same length as the length of time between the measurements on which the estimates of the path coefficients were calculated. For example, if measurements were taken on high school freshmen and again on the same individuals when they became juniors, there is no mechanism built into the theory of path analysis by itself to permit predicting to the senior year. Predictions would necessarily be to the year following the senior year, since estimates of path coefficients were based on a two-year interval. Similarly, if the first two measurements were taken during the freshman and sophomore years, predictions would be to the junior year. Predictions to the senior year based on path-analysis would be either (a) ad hoc, or (b) depend on extending the theory of path analysis. It seems unlikely that many researchers would extend path-analysis theory in the required manner. Further, the extension would generate a special case of the forecasting formula derived from differential equations without the corresponding conceptual benefits and flexibility of the differential equations.[4]

Interpretation of Effects

The change coefficients, a_{ij} and b_{ij} in equation system (4), can be used to indicate the effects of the intervening variables on each other (b_{ij}) and effects of the exogenous variables on the intervening variables (a_{ij}). For example, a_{51} indicates the effect of parental SES on the rate of change in occupational-status expectation of youth. Perhaps of more interest, the b coefficients can be used to assess the relative magnitudes of two-directional effects. To illustrate, b_{42} indicates the instantaneous effect of significant-other educational expectation

---

3. This point was made in a private communication from Professor Robert Leik.

4. The extension undoubtedly would follow a line of reasoning in which the path coefficients calculated over, say, a one year interval were applied successively to predicted values, starting one year after the second measurement used in calculating the path coefficients. A general prediction equation could then be developed by induction in a manner paralleling the theory of Markov chains over discrete time intervals. Such a development depends, however, on a model in which the same endogenous variables are measured at both time points, a practice not frequently followed in empirical studies.

for ego on the rate of change in ego's own educational expectation; conversely b_{24} shows the impact of ego's educational expectation on significant others' educational expectation of ego.

As conceptualized here, comparison of effects of one variable to effects of a second variable on a given dependent rate of change is awkward, because the independent variables may be measured in scales that are quite different from variable to variable. In path analysis this problem often is resolved by transforming all variables to a scale with zero mean and unit variance, thereby rendering rough comparisons between coefficients feasible. The simple strategy of transforming all variables to zero mean and unit varialce when the same variables are measured at more than one point in time is not advisable however, because the strategy artifically removes change over time in the mean and variance from the data. Methods of standardization that avoid this problem will be developed in Chapter 4.

Comparison to Cross-Sectional Path Coefficients. In the preceding discussion it was emphasized that cross-sectional path coefficients measure differences between individuals at a given time point rather than changes over time. Under certain circumstances, however, the cross-sectional paths (a^*_{ij} and b^*_{ij}) yield useful information about the change coefficients (a_{ij} and b_{ij}). If the system of career development variables (equations (4)) has reached an equilibrium state, i.e., stopped changing, then the cross-sectional path coefficients and the change coefficients for a given equation are equal, up to a constant of proportionality. This means that each path coefficient in a given equation can be multiplied by the same constant to yield the corresponding change coefficient (Coleman, 1968). Unfortunately, the constant required to convert from cross-sectional paths to change coefficients cannot be calculated from cross-sectional data. This interpretation, of course, depends on specifying the path model so that it permits nonzero paths everywhere that nonzero change coefficients are permitted (e.g., if the differential equation model permits nonzero b_{ij} and b_{ji}, the path model must permit nonzero b^*_{ij} and b^*_{ji}). It should be emphasized, however, that even in equilibrium, comparison of path coefficients between equations is not permissible, since different conversion constants are required for each equation. For example, comparing b^*_{41} to b^*_{43} is permissible, since these two coefficients appear in the same equation, but comparing b^*_{41} to b^*_{14} is not permissible. Comparisons between equations would be appropriate only under the assumption that the instantaneous effect of the level of each variable on its own change rate is the same for all variables ($b_{11} = b_{22} = \ldots = b_{JJ}$, where there are J equations). This is a restrictive assumption, so such comparisons should generally be avoided.

<u>Comparison to Cross-Lagged Path Coefficients</u>. Suppose two observations for each individual and each endogenous variable were available, and a path analysis were executed in which the time-one observations on the endogenous variables were among the independent variables and time-two observations were the dependent variables. Define all the same endogenous variables as in the previous examples, and let $y_i(0)$ be the observation for y_i at the beginning point, and $y_i(1)$ the corresponding observation at time 1. A cross-lagged path analysis might be written as follows:

$$
\begin{aligned}
y_1(1) &= a^*_{10} + a^*_{11}x_1 + a^*_{12}x_2 + b^*_{11}y_1(0) + \dots + b^*_{15}y_5(0) + u^*_1 \\
&\;\;\vdots \\
y_5(1) &= a^*_{50} + a^*_{51}x_1 + a^*_{52}x_2 + b^*_{51}y_1(0) + \dots + b^*_{55}y_5(0) + u^*_5
\end{aligned}
$$

(see, e.g., Heise, 1970). Generally, the coefficients a^*_{ij} and b^*_{ij} are interpreted as indicators of the magnitude of effects of one variable on another. But such interpretations of the cross-lagged path coefficients calculated over finite time intervals must be made with caution. The magnitudes of the b* coefficients depend on the length of time interval between measurements; further, the relative magnitudes of different b* coefficients also depend on the length of the measurement interval. This means that the relative effects of two variables on a given process variable (b^*_{ij} vs $b^*_{ij'}$, $j \neq j'$) depend on the length of the measurement interval, and the relative magnitude of reciprocal effects depend on the length of the measurement interval. Even the sign of the b* may depend on the length of the measurement interval.[5]

This observation will be justified in Chapter 4, but it should be noted here that the statement depends on the assumption that the differential equation system really does describe the process.

Summary

This chapter has developed the substantive rationale for applying differential equations to the study of development of career expectations. The conceptual framework is drawn from status-attainment research in sociology, but important theoretical insights from other research traditions were referenced also, and included in the model where possible.

5. Doreian and Hummon (1974; 1976) make similar observations about a single-equation model of status attainment processes.

The first section of the chapter reviews the status attainment theory and presents an illustration of a path model of this theory. The path model shows status background and mental ability affecting career-expectation variables. The career-expectation variables affect career attainments. The path model of this theory is a set of linear equations describing cross-sectional differences among individuals rather than changes over time.

The second section presents the theoretical basis for translating the cross-sectional path model into a dynamic model represented by a system of linear differential equations. Two points are made. First, theory suggests that career expectations develop gradually over continuous time. Secondly, most of the career-expectation process variables probably exhibit two-way effects on each other. Neither of these important features of career expectations are expressed by cross-sectional path models.

The third section presents an example of a differential equation model to express the dynamics of developing career expectations and two-way effects among career-expectation process variables. The differential equation model represents the rate of change over time in career-expectation variables (e.g., status level of occupational expectation) as linear functions of background variables (family status, mental ability) and of current levels of the career-expectation variables. Career attainments are omitted from the model because they change abruptly at isolated time points rather than continuously over time.

The fourth section discusses interpretation and applications of the differential-equation model. It is noted that the model refers to change at a point in time rather than over a finite interval of time. Use of the model to test theory by forecasting is emphasized. It is argued that accuracy of forecasts comprise a stronger test of theory than the usual post facto regression analyses in which correlations are used to index the accuracy of the model. In addition, interpretation of the change coefficients associated with the differential equation model is compared to interpretation of cross-sectional and cross-lagged path coefficients. It is noted that cross-sectional paths generally do not coincide with the change coefficeints but do estimate change coefficients up to a constant of proportionality if change in the system has ceased. Perhaps of more interest, the cross-lagged path coefficients calculated from two panels of observations on the same process variables depend on the length of the interval of time separating the two panels. In contrast, the change coefficients do not depend on the length of the measurement interval, assuming the differential-equation model describes the process under study. It is therefore recommended that interpretation of cross-lagged path coefficients proceed with caution and with the realization that the coefficients do depend on the length of the time interval separating panels. The change coefficients associated with the differential equation model are an alternative set of coefficients that researchers may wish to depend on for interpretations of effects.

CHAPTER 3

REVIEW OF SELECTED MATHEMATICAL AND STATISTICAL CONCEPTS AND RELATIONS

Thorough understanding of parts of this volume probably requires some previous exposure to calculus and to linear algebra. To reduce the burden on the reader, however, this chapter reviews some basic concepts and relationships in mathematical and statistical theory that are used repeatedly in the remaining chapters. The review is cursory and highly selective. In most instances, basic concepts are summarized in an intuitive manner, important theorems are illustrated numerically, and the theorems are stated in brief form without proofs. There are four main topics. First, elementary concepts and formulas in calculus are summarized. Secondly, complex numbers are discussed briefly. Thirdly, concepts and formulas in linear algebra and matrix equations are covered. This review is somewhat more extensive than the others because of the strong dependence of much of the material in this volume on theorems from linear algebra. Finally, a very brief and selective discussion related to inferential statistics is presented.[6]

Elementary Concepts and Formulas in Calculus

There are two major branches to the study of calculus -- the differential calculus and the integral calculus. A selective review of the differential calculus is presented first, then the integral calculus is reviewed. These reviews are not rigorous presentations of the mathematical theory. Rather, they summarize very briefly some basic results needed in the remaining chapters.

---

6. The material in this chapter is common knowledge or readily derivable from common knowledge in the mathematics or statistical literature. Numerous reference sources were used to assemble the information contained in the chapter. These include Goodman (1969), Freund (1971), Hohn (1972), Yamane (1968), Lancaster (1968), Fisher and Ziebur (1958), and Platt (1971). Readers desiring more rigorous and complete presentations of the theory are referred to these sources, among many others that are available.

Differential Calculus

The differential calculus is the study of slopes of continuous functions at isolated points on the curves defined by the functions. The slope of a straight line defined by $y = a + bx$ is familiar; the slope is the "rise over run" or the change in y as a ratio to the change in x. Let x_0 and x_1 be two values of x, and let y_0 and y_1 be the corresponding values of y determined by the linear equation, i.e., $y_0 = a + bx_0$, and $y_1 = a + bx_1$. Denote the change in x by $\Delta x = x_1 - x_0$ and the change in y by $\Delta y = y_1 - y_0$. The slope is then defined by

$$\frac{\Delta y}{\Delta x} = \frac{y_1 - y_0}{x_1 - x_0} = \frac{(a + bx_1) - (a + bx_0)}{x_1 - x_0} = \frac{b(x_1 - x_0)}{x_1 - x_0}$$

$$\frac{\Delta y}{\Delta x} = b$$

This is a familiar result, illustrated in the previous chapter; the slope of a straight line is the multiplier constant b in the linear equation $y = a + bx$.

For a linear equation, this result holds irrespective of the magnitude of Δx, but for a curved line, the slope does depend on the magnitude of Δx. To see this, observe the parabola in figure 2, defined by $y = a + bx^2$.

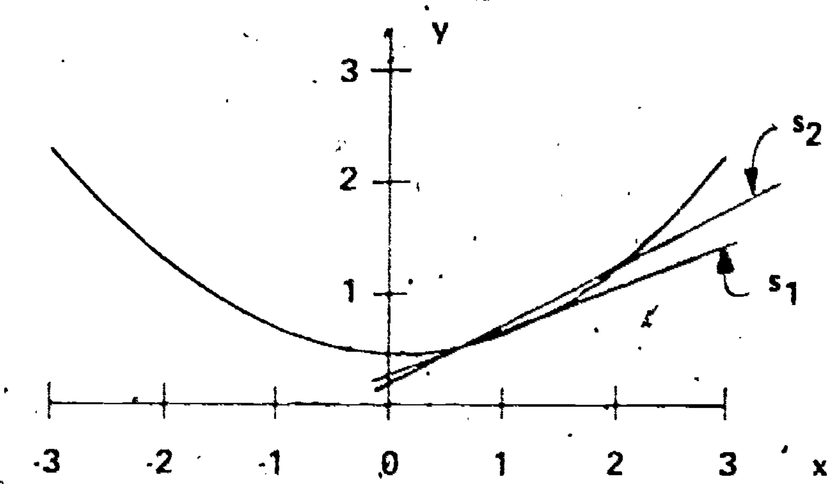

Figure 2. Slopes of a Curved Line

The graph is constructed with $a = .5$, and $b = .2$. Assume for example that one wishes to calculate the slope of the line defined by connecting the points on the graph corresponding to $x = 1.5$, and $x = .5$. For $x_1 = 1.5$, $y_1 = .5 + .2\ (1.5)^2 = .95$. For $x_0 = .5$, $y_0 = .5 + .2(.5)^2 = .55$. Hence, $\Delta y = y_1 - y_0 = .95 - .55 = .4$. Also, $\Delta x = x_1 - x_0 = 1.5 - .5 = 1$. The slope, therefore, is: $\Delta y/\Delta x = .4/1 = .4$. This slope is represented by

s_1 on the graph (figure 2). To show that this slope does not remain fixed for different amounts of change in x, carry out a second set of calculations with $x_1 = 2$, and $x_0 = .5$ One finds:

$$y_1 = .5 + .2(2)^2 = 1.3$$

$$y_0 = .5 + .2(.5)^2 = .55$$

$$\Delta y = y_1 - y_0 = 1.3 - .55 = .75$$

$$\Delta x = x_1 - x_0 = 2 - .5 = 1.5$$

Thus, slope $= \Delta y/\Delta x = .75/1.5 = .5 \neq .4$

The following point has been illustrated: From a given starting point, a change of one unit in the independent variable of a nonlinear function generates a different slope than does a change of one and a half units.

It may be useful to contrast this example to a parallel set of figures for a linear function. Consider the linear function $y = a + bx$ with $a = .5$ and $b = .2$. Note that the constants for this illustration are the same as for the previous example involving a nonlinear function. It was just shown that the slope of the linear function, $y = a + bx$, is b. In the present case, slope $= b = .2$. To illustrate this fact, take the same points for x which were used to illustrate the behavior of the nonlinear function. For $x_1 = 1.5$, and $x_0 = .5$;

$$\Delta y = y_1 - y_0 = (.5 + .2(1.5)) - (.5 + .2(.5)) = .2$$

$$\Delta x = 1.5 - .5 = 1$$

$$\text{slope} = \Delta y/\Delta x = .2/1 = .2$$

For the second pair of x values used in the nonlinear example, $x_1 = 2$, and $x_0 = .5$, one has

$$\Delta y = y_1 - y_0 = (.5 - .2(2)) - (.5 - .2(.5)) = .3$$

$$\Delta x = 2 - .5 = 1.5$$

$$\text{slope} = \Delta y/\Delta x = .3/1.5 = .2$$

As illustrated, the slope is constant for the linear function and equals the multiplier constant, $b = .2$.

It can also be seen that for the nonlinear function, the slope is a function of the starting value of x as well as of Δx. For example, let $\Delta x = (-.5) - (-1.5) = 1$. Note that $\Delta x = 1$ is the same as for the first example with the parabola. For the first example, however, the slope $= \Delta y/\Delta x = .4$, but a quick calculation for the present case shows that the slope is $-.4$. Similarly, for $\Delta x = (-.5) - (-2) = 1.5$, slope $= \Delta y/\Delta x = -.75/1.5 = -.5$. In the second example with the parabola, the change in x was also 1.5, but the slope was a positive one-half rather than negative one-half.

The differential calculus is the study of how slopes of continuous functions depend on x as Δx goes toward the limit of zero. Consider now the function $y = a + bx^2$ algebraically rather than graphically. The slope can be written

$$\frac{\Delta y}{\Delta x} = \frac{(a + bx_1^2) - (a + bx_0^2)}{\Delta x}$$

$$= \frac{b(x_1^2 - x_0^2)}{x_1 - x_0}$$

$$= \frac{b(x_1 + x_0)\cancel{(x_1 - x_0)}}{\cancel{x_1 - x_0}}$$

$$= b(x_1 + x_0)$$

In the limit as $\Delta x \to 0$, $x_1 \to x_0$ and this result can be written:

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = 2bx_0 = 2bx_1 = 2bx$$

This is the derivative of y with respect to x for a parabola of the form $y = a + bx$. In general, the derivative of y with respect to x is denoted by dy/dx. As can be seen by the example, derivatives do, in general, depend on the value of x.

The general definition of the derivative can be expressed as follows. Let y be a continuous function of x. Denote by y(a) the value of y when x = a. Now, by definition

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{y(x + \Delta x) - y(x)}{\Delta x}$$

In words, the derivative of a continuous function of x is the slope of the function at a specified point on the curve. This is not a rigorous definition, but it does offer an intuitive statement of the main idea.

The study of the differential calculus largely is concerned with finding algebraic expressions for the derivatives of algebraically expressed continuous functions.[7] Some of the basic results needed in the remaining chapters are given here without proof. It is assumed that y and z are continuous functions of x.

1. $\frac{d}{dx}(a + by) = b\frac{dy}{dx}$

2. $\frac{dy^n}{dx} = ny^{n-1}\frac{dy}{dx}$

3. $\frac{d}{dx}(y + z) = \frac{dy}{dx} + \frac{dz}{dx}$

4. $\frac{d \ln y}{dx} = \frac{1}{y}\frac{dy}{dx}$; $y > 0$; ln y stands for the natural log of y.

5. $\frac{dyz}{dx} = y\frac{dz}{dx} + z\frac{dy}{dx}$

6. $\frac{de^y}{dx} = e^y\frac{dy}{dx}$; e is the base of the natural logarithm

7. $\frac{d \sin y}{dx} = \cos y\frac{dy}{dx}$

8. $\frac{d \cos y}{dx} = \sin y\frac{dy}{dx}$

---

7. Certainly, not all functions can be expressed algebraically. For example, a string may be spread in a curved fashion across a pair of rectangular coordinates. So long as no line perpendicular to the horizontal axis crosses the string twice, the set of points defined by the path of the string forms a function. If there are no "kinks" in the string, it forms a continuous function and hence, is differentiable; yet it is unlikely that any algebraic expression describes the path of the string. It is, perhaps, a curious fact that the derivatives of many functions can be expressed algebraically even though the functions have no algebraic expression.

In all these cases, the fact that y and z are continuous functions of x is important to a proper interpretation of the formulas. First, illustrate the formulas in which z does not appear. Assume the simplest possible function between y and x by setting $y = x$. For this instance $dy/dx = 1$. Thus, for example, formula 1 simplifies to

$$\frac{d}{dx}(a + by) = \frac{d}{dx}(a + bx) = b, \text{ as before}$$

Formula 2 becomes

$$\frac{dy^n}{dx} = \frac{dx^n}{dx} = nx^{n-1}$$

Formulas 4, 6, 7, and 8, respectively, simplify to:

$$\frac{d \ln x}{dx} = \frac{1}{x}$$

$$\frac{de^x}{dx} = e^x$$

$$\frac{d \sin x}{dx} = \cos x$$

$$\frac{d \cos x}{dx} = -\sin x$$

To illustrate the addition formula 3 and product formula 5, let $y = e^x$, and $z = \ln x$. From formulas 6 and 4, respectively, one finds $de^x/dx = e^x$, and $d \ln x/dx = \frac{1}{x}$, $x > 0$. Hence, the addition formula specializes to

$$\frac{d(y + z)}{dx} = \frac{d(e^x + \ln x)}{dx} = e^x + \frac{1}{x},\ x > 0$$

Similarly, with these two functions (y, z), the product formula becomes

$$\frac{dyz}{dx} = \frac{d(e^x \ln x)}{dx} = e^x \frac{d \ln x}{dx} + \ln x \frac{de^x}{dx}$$

$$= e \frac{1}{x} + (\ln x)\ e^x$$

$$= e^x(\frac{1}{x} + \ln x)$$

To illustrate a more complicated case than $y = x$ for the formulas not involving z, set $y = \frac{1}{2}x^2$. In this case, formula 1, for example, indicates

$$\frac{d}{dx}(a + by) = b\,\frac{d\frac{1}{2}x^2}{dx} = bx$$

Attention now shifts to defining higher-order derivatives. Consider a simple function and its derivative:

$$y = (1/6)x^3$$

$$dy/dx = (3/6)x^2 = (1/2)x^2 \quad \text{(rule 2)}$$

Let $z = dy/dx = \frac{1}{2}x^2$; then $dz/dx = x$ (by rule 2 again). This can be conceptualized by the following notation.

$$\frac{dz}{dx} = \frac{d}{dx}\left\{\frac{dy}{dx}\right\}$$

$$= \frac{d}{dx}\left\{\frac{1}{2}x^2\right\} = \frac{d\,\frac{1}{2}x^2}{dx}$$

Now, since dz/dx is the function defined by differentiating $y = (1/6)x^3$ twice, it is termed the second derivative of y or the derivative of the second order. A common notation for this is $d^2(1/6\ x^3)/dx^2 = x$. More generally, let $f(x)$ be a smooth function of x, then its nth order derivative is denoted by $d^nf(x)/dx^n$. It is defined by differentiating the function $f(x)$ n times.

The Integral Calculus

There are two types of integrals -- the indefinite integral (or antiderivative), and the definite integral. It is important to maintain the conceptual distinction between the two integrals. Consider a continuous function of x, $f(x)$. The indefinite integral of $f(x)$ is defined to be a second continuous function, say, $F(x)$, whose derivative equals $f(x)$. The indefinite integral of $f(x)$ is denoted by the symbols $\int f(x)dx$. Let

$$\int f(x)dx = F(x)$$

then by definition

$$\frac{dF(x)}{dx} = f(x)$$

The definite integral is somewhat more difficult to define intuitively, but it can be interpreted as the area under a continuous curve from one specific point on the x axis to a second specific point on the x axis. For example, the shaded area in figure 3 depicts the definite integral from point a to point b.

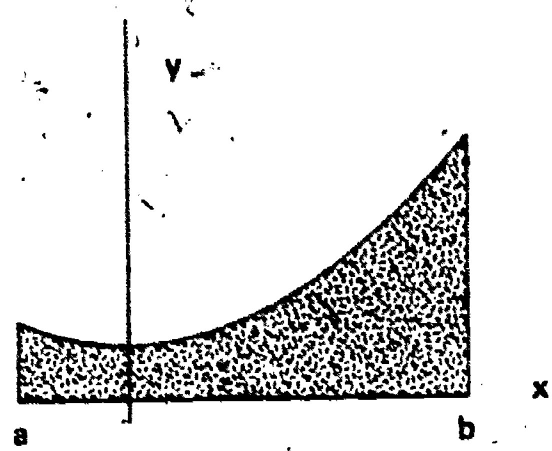

Figure 3. Area Under A Continuous Curve

An approximation of an area under a curved line can be found by adding the areas of narrow strips under the curve, as illustrated in figure 4.

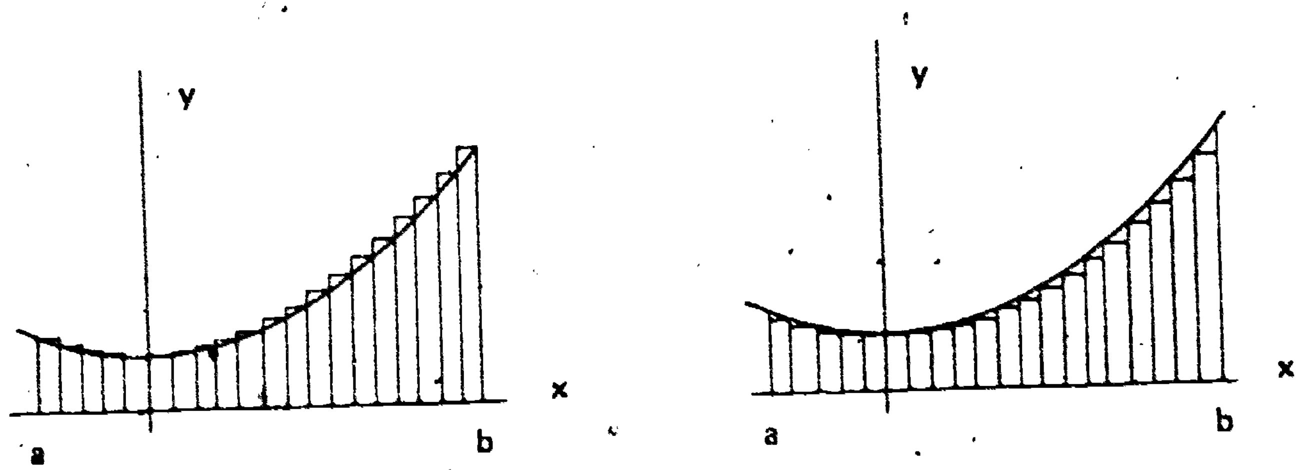

Figure 4. Approximations to the Area Under a Smooth Curve

Note that the area A of each rectangle is given by the product of its height and width. If its height is $y = f(x)$ and width is the change in x - - Δx, then $A = f(x)\ \Delta x$. Intuitively, it is clear that in the limit as Δx goes to zero, the sum of the areas of the rectangles converges on the area under the curve. The definite integral from a to b of $f(x)$ is denoted by

$$\int_a^b f(x)\,dx$$

It can be viewed heuristically as the sum of the areas of narrow strips whose heights are $f(x)$ and widths are Δx, in the limit as $\Delta x \to 0$.

Now, let $F(x)$ be the indefinite integral of $f(x)$, so that

$$F(x) = \int f(x)dx$$

One of the most fascinating and useful facts in mathematics is the connection between the indefinite integral $F(x)$ and the definite integral, to wit,

$$\int_a^b f(x)dx = F(b) - F(a)$$

This is known as the fundamental theorem of calculus. It says that the area under a section of a continuous curve, that is, the definite integral, can be calculated by finding the difference between the indefinite integral evaluated at the end points, a and b, of the section.

The integral calculus is the basis for translating hypotheses concerning instantaneous changes over time into forms that can be studied empirically. Since such translations are the focus of the differential-equation model of career expectations, it is useful to see how the translation can be achieved. Suppose one forms a hypothesis about changes in y over time:

$$\frac{dy}{dt} = f(y,t)$$

A derivative cannot be observed since it is a slope over an infinitesimally short period. Let $F(t)$ be a function of t such that $F(t)$ is the indefinite integral of $f(y,t)$, and note that

$\int_a^b dy = y_b - y_a$. Taking the definite integral on both sides of $dy = f(y,t)dt$, therefore, yields

$$y_b - y_a = F(t_b) - F(t_a)$$

$$y_b = F(t_b) + (y_a - F(t_a))$$

where t_a, t_b are two specific time points at which $y = y_a$ and $y = y_b$, respectively. The equation involving derivatives has been converted into one in which all elements are observable. One can view t_a at a fixed initial time point and let t_b vary; in these circumstance y_a is a fixed initial value of y and y_b

varies with t_b. Hence, from the initial hypothesis concerning continuous change, a function giving y in terms of t has been derived. The function can be used in conjunction with data.

The above result was derived with the definite integral. The same outcome also can be derived using the indefinite integral. Since it is sometimes more convenient to use the indefinite integral, the alternative derivation is presented. Again, consider the differential equation: $dy/dt = f(y,t)$.

$$dy/dt = f(y,t)$$

$$dy = f(y,t)dt$$

$$\int dy = \int f(y,t)dt$$

Suppose that $F(y,t)$ is a solution to the differential equation, so that, by definition, $dF(y,t)/dt = f(y,t)$. If $dF(y,t)/dt = f(y,t)$, then so does $d[F(y,t) + c]/dt = f(y,t)$, where c is any constant over t. This is an essential aspect of integration; c sometimes is called the constant of integration. Applying this concept, one finds:

$$\int dy = \int f(y,t)dt$$

$$y = F(y,t) + c$$

To check this result, differentiate both sides with respect to t:

$$dy/dt = d[F(y,t) + c]/dt$$

$$= dF(y,t)/dt + dc/dt$$

$$= f(y,t) + 0$$

since $dF(y,t)/dt = f(y,t)$ (by construction), and the derivative of a constant is always zero.

For empirical work, the difficulty with the result that $y = F(y,t) + c$ is that c is an unknown constant. To find c one needs an observation on y at a given time point, say t_a. With the observation point (t_a,y_a), one has

$$y_a = F(t_a) + c$$

$$c = y_a - F(t_a)$$

Putting this result into the indefinite integral yields the previous conclusion based on the definite integral:

$$y = F(t) + (y_a - F(t_a))$$

Sometimes an alternative notation for finding a definite integral over a specified interval is convenient. In the notation used so far

$$\int_{t_a}^{t_b} dy = \int_{t_a}^{t_b} f(y,t)dt = F(t_a) - F(t_b)$$

Suppose that one desired to denote the initial time point by 0 (time zero) and the second time point by t. The notation $\int_0^t f(y,t)dt$ is an awkward notation to represent the desired result because the symbol t is used in the same expression to indicate the end point of the interval and all the points in between. A much clearer notation is:

$$\int_0^t dy = \int_0^t f(y,\tau)d\tau = F(t) - F(0)$$

Now, t denotes the end point of the interval, and τ indicates the variable which assumes all values in the interval $(0,t)$. The variable here denoted by τ is sometimes called a dummy variable, but it should not be confused with the same term used to indicate representations of categorical variables in regression analysis.

A large part of the study of the integral calculus entails finding algebraic expressions for indefinite integrals. Several of these results are listed without proof below, for reference in later chapters. It is presumed y and z are continuous functions of x.

1. $\int b dx = bx + c$
2. $\int x^n dx = \frac{1}{n+1}x^{n+1} + c \text{ , } n \neq -1$
3. $\int (y + z)dx = \int ydx + \int zdx + c$
4. $\int \frac{dx}{x} = \ln x + c$
5. $\int e^y dx = e^y \frac{dx}{dy} + c$
6. $\int \cos x \, dx = \sin x + c$
7. $\int \sin x \, dx = -\cos x + c$

where c is a constant.

It should be emphasized that many integrals exist that cannot be expressed by an algebraic statement. In fact, even for functions given in algebraic terms, existence of an integral expressable by an algebraic statement is the exception rather than the rule. When algebraic expression of integrals cannot be found, numerical solutions are a viable alternative, by use of electronic computers.

Complex Numbers

Some of the results of the section on linear algebra depend on knowledge of complex numbers; hence, a very brief summary of the important ideas is given here. The concept of a complex number is more meaningful if set in the context of other number sets. The positive integers are discussed first, and the set of positive integers is gradually expanded through the real numbers, and, finally, the complex numbers.

Consider the number set defined by the positive integers: 1, 2, ... One can add and multiply two positive integers together and end up with a positive integer for the result. The positive integers therefore are said to be closed in addition and multiplication. Subtracting one positive integer from another positive integer does not necessarily yield a positive integer as the result, however. For example, $3 - 5 = -2$, or $4 - 4 = 0$. This difficulty can be overcome by defining an expanded number set, to include the positive integers, zero, and negative integers. This new number set is closed for addition, multiplication and subtraction, but not for division. For example, $6 \div 2 = 3$, but $5 \div 2 = 2.5$, which is not an integer. Expanding the number set to include rational fractions as well as all integers and zero generates a number set which is closed to the four arithmetic operations -- addition, multiplication, subtraction, and division.

Even the rational fractions do not correspond to every point on a straight line. This may be counter-intuitive; to see why it is true, consider the point p in figure 5 below.

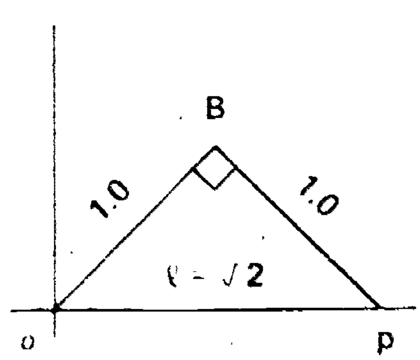

Figure 5. Real numbers and points on a line

4

The positively sloping line from the origin to point B and the negatively sloping line from B to point p are at right angles and each is one unit in length. The problem is to find the distance from the origin to point p. This distance is the length of the hypotenuse of a right triangle. By the Pythagorean theorem, therefore, the square of the length from the origin to p is $2 = 1^2 + 1^2$. The length, ℓ, therefore, is $\ell = \sqrt{2}$. But the square root of two is not a rational fraction, as shown by the following argument.

Since $0^2 = 0$, $1^2 = 1$, and the square of any integer larger than 1 exceeds 2, try letting $\sqrt{2} = a/b$, where a/b is a rational fraction reduced to its lowest common denominator. If a/b is a rational fraction reduced to its lowest common term (all rational numbers can be so represented), then either a or b must be an odd integer, because if they were both even, they would be divisable by 2, and therefore, not be reduced to the lowest common term. Since we are assuming $\sqrt{2} = a/b$; $a^2 = 2b^2$; hence, $a = 2(b/\sqrt{2})$, thus showing that a is an even integer. Since a is an even integer, one can represent it by the product, 2c, where c is any integer. One finds

$$4c^2 = 2b^2$$

$$2c^2 = b^2$$

Therefore, b^2 is also an even integer, implying b is an even integer. Thus, the presumption that a/b is a rational fraction reduced to its lowest common denominator is contradicted. It is concluded, therefore, that $\sqrt{2}$ is not a rational number. Yet it is a point on the line, as illustrated in figure 5.

The set of numbers must be expanded to include some numbers other than the integers and rational fractions if all points on a line are to correspond to a number. The real numbers comprise a set of numbers with a one-to-one correspondence to points on a line of infinite distance in either direction from the origin. Yet all operations on the real numbers do not yield real numbers as answers. For example, the equation $\iota^2 + 1 = 0$ gives $\iota^2 = -1$, but no real number satisfies this relation, since the product of any real number with itself is positive. The last expansion of the number set is the complex numbers. Complex numbers are written in the following form:

$$x = a + b\iota,$$

where $\iota^2 \equiv -1$ is the "imaginary" unit. (The symbol $\equiv$ means is defined by.) The real number a is called the "real" part of x and the real number b is termed the "imaginary" part of x. The complex conjugate of x is denoted by $\bar{x}$ and is defined by:

$$\bar{x} = a - b\iota$$

All algebraic operations are defined on the set of complex numbers in the same way as for real numbers, except where $\imath^2$ is encountered it is replaced with -1. Thus, addition, subtraction, multiplication and division of $x_1 = a_1 + b_1\imath$ and $x_2 = a_2 + b_2\imath$ are defined as follows.

$$x_1 + x_2 = a_1 + b_1\imath + a_2 + b_2\imath = (a_1 + a_2) + (b_1 + b_2)\imath$$

$$x_1 - x_2 = a_1 + b_1\imath - a_2 - b_2\imath = (a_1 - a_2) + (b_1 - b_2)\imath$$

$$x_1x_2 = (a_1 + b_1\imath)(a_2 + b_2\imath) = a_1a_2 + a_1b_2\imath + a_2b_1\imath - b_1b_2$$

$$= (a_1a_2 - b_1b_2) + (a_1b_2 + a_2b_1)\imath$$

$$x_1/x_2 = (a_1 + b_1\imath)/(a_2 + b_2\imath) = \frac{(a_1 + b_1\imath)\ (a_2 - b_2\imath)}{(a_2 + b_2\imath)\ (a_2 - b_2\imath)}$$

$$= \frac{(a_1 + b_1\imath)\ (a_2 - b_2\imath)}{a_2^2 - b_2^2\imath^2}$$

$$= \frac{(a_1a_2 + b_1b_2) - (a_1b_2 - a_2b_1)\imath}{a_2^2 + b_2^2}\text{, since } \imath^2 = -1$$

$$x_1/x_2 = \frac{a_1a_2 + b_1b_2}{a_2^2 + b_2^2} - \frac{a_1b_2 - a\ b_1}{a_2^2 + b_2^2}\ .\imath$$

Hence one sees that the complex numbers are closed in addition, subtraction, multiplication and division, since the results of these operations with complex numbers are, themselves, complex numbers.

A complex number is a pair of real numbers a, b, and can be graphed as a point on a plane, as in figure 6.

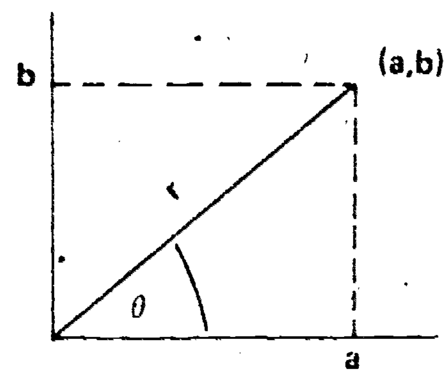

Figure 6. Graphical representation of a complex number

The absolute value of $x = a + b\imath$ is defined by $|x| = r = \sqrt{a^2 + b^2}$. The angle θ is termed the argument of x. Note that x can be written in trigonometric form:

$$x = a + b\imath = r(\cos\theta + (\sin\theta)\imath)$$

since $\sin\theta = b/r$, and $\cos\theta = a/r$.

For multiplication, division, finding powers and roots, and exponentiation, the trigonometric form is very useful. The following results follow directly from the definitions of complex numbers and the operations; they are listed without proof. Let $x_1 = r_1(\cos\theta_1 + (\sin\theta_1)\imath)$, $x_2 = r_2(\cos\theta_2 + (\sin\theta_2)\imath)$, and $x = r(\cos\theta + (\sin\theta)\imath)$, then:

$$x_1x_2 = r_1r_2\{\cos(\theta_1 + \theta_2) + (\sin(\theta_1 + \theta_2)\imath\}$$

$$x_1/x_2 = (r_1/r_2)\{\cos(\theta_1 - \theta_2) + (\sin(\theta_1 - \theta_2))\imath)$$

$$x^n = r^n\{\cos(n\theta) + (\sin(n\theta)\imath\}$$

$$x^{1/n} = r^{1/n}\{\cos[(\theta + 2k\pi)/n] + (\sin[(\theta + 2k\pi)/n]\imath\}$$

where θ is given in radians and $k = 0, 1, \ldots, n-1$. There are n roots to $x^{1/n}$. Exponentiation and natural logs can be written:

$$e^x = e^a[\cos b + (\sin b)\imath], x = a + b\imath$$

$$\ln x = \ln r + (\theta + 2\pi k)\imath$$

where k is any integer.

The exponential and logarithm of complex numbers enter into systems of linear differential equations in an important way. The principal branch of ln x is defined by setting k = 0:

$$\ln x = \ln r + \theta i \text{ ; if } k = 0$$

Hereafter in this volume it is assumed that the principal branch of the logarithm is to be used.

Elements of Linear Algebra

There are three aspects of linear algebra and related matrix equations which are reviewed here. First, basic concepts related to systems of linear relations are summarized. Coverage includes the concepts of a solution to a system of linear equations, rank of a system, and determinants. Secondly, characteristic equations are reviewed, including eigenvalues and eigenvectors of asymmetric real matrices. Thirdly, matrix functions involving eigenvalues and eigenvectors are introduced.

Systems of Linear Equations

Consider t[illegible]ollowing linear equation in two unknowns, x and y.

$$x + 2y = 3$$

One can solve for y as a function of x or solve for x as a function of Y;

$$y = 1/2(3 - x)$$

$$x = 3 - 2y$$

The equation does not contain enough information to yield a unique value for x and y; any point along the line defined by $y = 1/2(3 - x)$ satisfies the equation. Now, add a second equation so that a pair of equations must be satisfied:

$$x + 2y = 3$$

$$2x + 2y = 4$$

If the first equation is subtracted from the second, the result is $x = 1$. Setting $x = 1$ in the first equation and solving for y gives $y = 1$. There is no other pair of values x, y that render both equations true at the same time; hence, $x = 1$ and $y = 1$ constitutes a unique solution to the pair of simultaneous linear equations.

Suppose that, for the second equation, instead of $2x + 2y = 4$, one had $3x + 6y = 9$. The new pair of equations is

$$x + 2y = 3$$

$$3x + 6y = 9$$

There is no way to solve this pair of equations to obtain unique values of x and y. If one attempts to remove y from the second equation by multiplying the first by 3 and subtracting the result from the second, both x and y are eliminated, and so is the constant on the right:

$$\begin{array}{r} 3x + 6y = 9 \\ \underline{-3(x + 2y = 3)} \\ 0 + 0 \ = 0 \end{array}$$

The reason for this result is that the second equation is a simple transformation of the first; it was obtained by multiplying both sides of the first equation by 3. Hence, the two equations are <u>linearly dependent</u>.

Now, return to the first pair of equations and add a third:

$$x + 2y = 3$$

$$2x + 2y = 4$$

$$x - y = 2$$

It has already been found that the first two equations imply that $x = 1$ and $y = 1$. But, if the third equation is subtracted from the first, one concludes that $y = 1/3$. Substituting $y = 1/3$ into the third equation gives $x = 7/3$. So, $x = 1$, $y = 1$ satisfies the first and second equation, and $x = 7/3$, $y = 1/3$ satisfies the first and third. The three equations are <u>inconsistent</u>. The third equation would be consistent with the other two if it stated $x - y = 0$, or some multiple thereof.

It is useful to represent these operations as matrix equations. A matrix is a rectangular table filled with numbers (or symbols representing numbers). If $\underline{A}$ symbolizes the matrix, then a_{ij} represents the entry in the ith row and jth column of $\underline{A}$. The entries for which the row and column indexes are equal are called diagonal entries and the diagonal entries comprise the diagonal of the matrix.

For example, if

$$\underline{A} = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}$$

Then $a_{11} = 1$, $a_{21} = a_{12} = a_{22} = 2$. If all entries except the diagonal entries of a matrix are zero, the matrix is called a diagonal matrix. The order of the matrix is the number of rows by the number of columns. $\underline{A}$ is a 2 x 2 matrix. If the number of columns is one, the matrix is a column vector; if the number of rows is one, the matrix is a row vector. The transpose of a matrix $\underline{A}$ is defined by the matrix found when rows of $\underline{A}$ form the columns of the new matrix. The transpose of $\underline{A}$ is denoted by $\underline{A}'$ or $\underline{A}^T$. If $\underline{A} = \underline{A}'$, $\underline{A}$ is said to be symmetric.

Matrix addition and subtraction are defined as the sum/difference of the individual elements. For example, if $\underline{C} = \underline{A} + \underline{B}$, then $c_{ij} = a_{ij} + b_{ij}$. Matrix addition and subtraction are not defined unless the order of the two matrices to be added or subtracted is the same.

Matrix multiplication and division are not defined as element-wise multiplication and division. If $\underline{C} = \underline{AB}$, $c_{ij} \neq a_{ij}b_{ij}$. Matrix multiplication is defined to represent systems of simultaneous linear equations. If $\underline{C} = \underline{AB}$, then, by definition,

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{iK}b_{Kj}$$

$$= \sum_{k=1}^{K} a_{ik}b_{kj}$$

where K is the column order of $\underline{A}$ and row order of $\underline{B}$.

Note that, in general $\underline{AB} \neq \underline{BA}$. In fact, it is possible that $\underline{AB}$ is defined when $\underline{BA}$ is not, since the column order of $\underline{A}$ may match the row order of $\underline{B}$ when the column order of $\underline{B}$ does not match the row order of $\underline{A}$. For example, $\underline{A}$ may be a 2 x 3 matrix, and $\underline{B}$ a 3 x 3 matrix. Consequently, it is important to designate premultiplication or postmultiplication. Premultiplication of $\underline{B}$ by $\underline{A}$ indicates $\underline{AB}$, and postmultiplication of $\underline{B}$ by $\underline{A}$ means $\underline{BA}$. To be conformable for multiplication, the column order of the premultiplier must match the row order of the postmultiplier.

To illustrate matrix representation of linear-equation systems, reconsider the first pair of equations among the preceding illustrations:

$$x_1 + 2x_2 = 3$$

$$2x_1 + 2x_2 = 4$$

where x_1 replaces the symbol x, and x_2 replaces y.

Define the matrix $\underline{A}$ and column vectors $\underline{x}$ and $\underline{b}$, as follows.

$$\underline{A} = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}, \quad \underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \underline{b} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

Now, the pair of linear equations can be written compactly

$$\underline{Ax} = \underline{b}$$

$$\begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

Suppose one premultiplied both sides of this matrix equation by the following matrix

$$\underline{P} = \begin{pmatrix} -1 & 1 \\ 1 & -\frac{1}{2} \end{pmatrix}$$

as follows:

$$\begin{pmatrix} -1 & 1 \\ 1 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

Carrying out the indicated operations yields

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Thus, premultiplication of the matrix equation $\underline{Ax} = \underline{b}$ by the matrix $\underline{P}$ yields $x_1 = 1$, $x_2 = 1$, as before. $\underline{P}$ is the inverse of $\underline{A}$ and is written $\underline{A}^{-1}$; it represents the matrix generalization of division. Suppose, a, b, and c are scalars, i.e., single numbers rather than matrices, and ab = c. Then

$$a^{-1}ab = a^{-1}c$$

$$1 \cdot b = a^{-1}c$$

$$b = c/a$$

The matrix inverse is defined in the same manner. If $\underline{AB} = \underline{C}$ ($\underline{A}$, $\underline{B}$, and $\underline{C}$ comformable matrices for the indicated operations), then $\underline{A}^{-1}$ is defined such that $\underline{B} = \underline{A}^{-1}\underline{C}$. Let a special matrix be defined for which each diagonal entry is 1.0 and the remaining entries are zero; this is called the <u>identity matrix</u> and generally is represented by the letter $\underline{I}$. For $\underline{B} = \underline{A}^{-1}\underline{C}$ to be a solution to $\underline{AB} = \underline{C}$, one must have $\underline{A}^{-1}\underline{A} = \underline{I}$. Hence, this relation defines the matrix inverse. The matrix inverse is defined only for square matrices but may not always exist even for square matrices. If $\underline{A}^{-1}$ exists, then $\underline{AA}^{-1} = \underline{A}^{-1}\underline{A} = \underline{I}$, and $\underline{A}^{-1}$ is unique.

To gain some insight into the relationship between the existence of the matrix inverse and solvability of systems of linear equations, reconsider the second pair of linear equations in the above illustrations:

$$x_1 + 2x_2 = 3$$

$$3x_1 + 6x_2 = 9$$

Recall that no unique solution for x_1 and x_2 could be found for this pair of equations (numerous solutions exist). Now rewrite the pair in matrix notation.

$$\underset{\underline{Q}}{\begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}} \underset{\underline{X}}{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} = \underset{\underline{P}}{\begin{pmatrix} 3 \\ 9 \end{pmatrix}}$$

It can be shown that for this system $\underline{Q}^{-1}$ does not exist; finding $\underline{Q}^{-1}$ is like finding the reciprocal of zero.

To see this, define the <u>determinant</u> of a 2 x 2 matrix, $\underline{A}$, by $|\underline{A}| = a_{11}a_{22} - a_{12}a_{21}$, where $|\underline{A}|$ denotes the matrix determinant. It can be shown that the inverse of $\underline{A}$ for $\underline{A}$ of order 2 x 2 can be found as follows.

$$\underline{A}^{-1} = \frac{1}{|\underline{A}|}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

Clearly, the inverse of a 2 x 2 matrix exists if and only if $|\underline{A}| \neq 0$. Calculate next the determinant of $\underline{Q} = |Q| = q_{11}q_{12} - q_{12}q_{21} = 1 \times 6 - 2 \times 3 = 0$. Thus, $\underline{Q}^{-1}$ does not exist, reflecting the fact that the corresponding pair of equations does not have a unique solution.

An accurate definition of the determinant of a general matrix requires more space than can be justified here. The main results needed for later chapters are: (a) the matrix inverse of $\underline{A}$ exists and is unique if and only if $|\underline{A}| \neq 0$, and (b) a system of simultaneous linear equations has a unique solution if and only if the determinant of the associated coefficient matrix, $\underline{A}$, is not zero --- $|\underline{A}| \neq 0$. These are two ways to say the same thing.

If in the system of equations $\underline{Ax} = \underline{b}$, the row order of $\underline{A}$ is less than the column order, the system is underidentified -- any number of vectors $\underline{x}$ can be found such that $\underline{Ax} = \underline{b}$. If $\underline{A}$ is square and $|\underline{A}| = 0$, then $\underline{Ax} = \underline{b}$ also is underidentified. If the row order of $\underline{A}$ exceeds the column order, the equation system $\underline{Ax} = \underline{b}$ is overidentified, meaning in general, no $\underline{x}$ can be found to satisfy the relation $\underline{Ax} = \underline{b}$. Each of these cases has been illustrated above.

The rank of a matrix is defined as the maximum number of rows or columns that are not linearly dependent. For a square matrix, if its rank equals its order it is termed full rank. If it is not full rank it is called singular. For example

$$\underline{A} = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}$$

is rank 2; its rows and columns are linearly independent; hence, it is full rank. In contrast, the rows (columns) of $\underline{Q}$ are linearly dependent:

$$\underline{Q} = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$$

because row 2 = 3×(row 1):

$$3(1\ \ 2) = (3\ \ 6)$$

Also, column 2 = 2×(column 1):

$$2\times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$$

This can be written

$$2 \begin{pmatrix} 1 \\ 3 \end{pmatrix} + (-1) \begin{pmatrix} 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This illustrates the general definition of linear dependence. The columns of $\underline{A}$ are defined to be linearly dependent if

$$\underline{Ax} = \underline{0}$$

with not all $\underline{x}$ zero. In the example,

$$\underline{x} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

note that for a scalar k

$$\underline{z} = k\underline{x},\ k \neq 0$$

is also a nonzero solution to $\underline{Ax} = 0$. This illustrates the general fact that if a nonzero (often termed nontrivial) solution to $\underline{Ax} = 0$ exists, it is determined only to a constant of proportionality, k. Note that $\underline{x}$ and $\underline{z} = k\underline{x}$ are linearly dependent solutions, since $k\underline{x} - \underline{z} = \underline{0}$. An important theorem in linear algebra states that a nontrivial solution to $\underline{Ax} = \underline{0}$ exists if and only if $|\underline{A}| = 0$ (for square $\underline{A}$). For future reference it is noted that $\underline{Ax} = \underline{0}$ is a set of homogeneous linear equations, and $\underline{Ax} = \underline{b}$, $\underline{b} \neq \underline{0}$ is a set of nonhomogeneous linear equations.

Characteristic Equations

Few facts about matrices find broader application in the social sciences than the theory of characteristic equations. Factor analyses, canonical correlation and certain estimation methods in econometrics depended on characteristic-equation theory. As shown later in the present volume, solutions to linear differential equation systems with constant coefficients also depend on characteristic equations.

One of the fascinating facts about matrices is that some constant λ can be subtracted from each diagonal entry, and the determinant of the resulting matrix is zero, even if the original matrix $\underline{A}$ is full rank viz, $|\underline{A}| \neq 0$. Consider as an example, the matrix $\underline{A}$ presented earlier. It is desired to find a constant λ that can be subtracted from each diagonal element of $\underline{A}$ so that the resulting matrix is singular. Now $\lambda\underline{I}$ is a matrix whose diagonal entries are λ and off diagonals are zero; hence, $\underline{A} - \lambda\underline{I}$ is a matrix whose diagonal elements are $a_{jj} - \lambda$ and the remaining elements are a_{ij}. For $\underline{A} - \lambda\underline{I}$ to be singular, $|\underline{A} - \lambda\underline{I}| = 0$.

Written out, this becomes

$$|\underline{A} - \lambda\underline{I}| = \begin{vmatrix} 1 - \lambda, & 2 \\ 2, & 2 - \lambda \end{vmatrix} = 0$$

$$0 = |\underline{A} - \lambda\underline{I}| = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21}$$

$$= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21}$$

$$0 = \lambda^2 - 3\lambda - 2 \text{ (for the numerical example)}$$

which is a quadratic equation and has two solutions: $\lambda_1 = 3.5616$, $\lambda_2 = -.5616$.[8] Substituting these for λ in the determinental equation $|\underline{A} - \lambda\underline{I}| = 0$ shows the result to be zero. The λs are called <u>eigenvalues</u>, <u>characteristic values</u>, or <u>characteristic roots</u> of $\underline{A}$. The equation $|\underline{A} - \lambda\underline{I}| = 0$ is termed the <u>characteristic equation</u> of $\underline{A}$.

It was seen that the characteristic equation of a 2 x 2 matrix is a second degree polynomial. This instance generalizes. The characteristic equation of an n x n matrix is a polynomial of degree n. There are n roots λ of the equation, but some of the roots may equal other roots. There are at most n distinct roots.

Consider the previously defined matrix $\underline{Q}$ as a second example of a characteristic equation. It is desired to find λ such that $|\underline{Q} - \lambda\underline{I}| = 0$. One has

$$\begin{vmatrix} 1 - \lambda, & 2 \\ 3, & 6 - \lambda \end{vmatrix} = 0$$

$$(1 - \lambda)(6 - \lambda) - 6 = 0$$

$$\lambda^2 - 7\lambda + (6 - 6) = 0$$

$$\lambda^2 - 7\lambda = 0$$

$$\lambda - 7 = 0, \lambda \neq 0$$

$$\lambda_1 = 7$$

$$\lambda_2 = 0$$

Checking, one finds $(1 - \lambda_1)(6 - \lambda_1) - 6 = (1 - 7)(6 - 7) - 6 = -6(-1) - 6 = 0$, and $(1 - \lambda_2)(6 - \lambda_2) - 6 = (1 - 0)(6 - 0) - 6 = 0$. This illustration is a special case of an important theorem: the rank of a matrix equals the number of nonzero eigenvalues. Also, the determinent of a matrix can be found by the running product of its eigenvalues:

$$|\underline{A}| = 3.5616(-.5616) = -2 \text{ (discrepancy due to rounding)}$$

$$|\underline{Q}| = 7 \times 0 = 0$$

8. Throughout the text matrix elements involving some operation such as $1 - \lambda$ are separated from other elements by commas, but single numbers or characters are not separated by commas.

Clearly, if one or more λs is zero, the matrix is singular.

Consider a further example. Let

$$\underline{B} = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}.$$

Now, the characteristic polynomial is:

$$|\underline{B} - \lambda \underline{I}| = 0$$

$$(2 - \lambda)(2 - \lambda) + 1 = 0$$

$$(2 - \lambda)^2 + 1 = 0$$

There is no real number that is a solution to this equation, since $(2 - \lambda)^2 = -1$, and the square of a real number is never negative. Evidently, $(2 - \lambda) = \pm \iota$, where, as before $\iota^2 \equiv -1$. Hence, $\lambda_1 = 2 + \iota$, and $\lambda_2 = 2 - \iota$.

This last example illustrates an important fact: the characteristic roots (eigenvalues) of a real matrix may be complex. However, if the real matrix is symmetric, as is $\underline{A}$ in the example, the eigenvalues are all real -- if the matrix is real, it has complex roots only if it is asymmetric. It is not always true that every or even any of the eigenvalues of an asymmetric real matrix are complex, however. Another important theorem has also been illustrated: If $\lambda = \lambda_R + \lambda_I \iota$ is a root of $\underline{A}$, then so is its complex conjugate $\bar{\lambda} = \lambda_R - \lambda_I \iota$.

Since $|\underline{A} - \lambda \underline{I}| = 0$, the homogeneous linear system of equations $(\underline{A} - \lambda \underline{I})\underline{x} = \underline{0}$ ($\underline{x}$ a conformable vector) has a nontrivial solution. The vector $\underline{x}$ is called the <u>eigenvector</u> or <u>characteristic vector</u> of $\underline{A}$ associated with λ. As before, if $(\underline{A} - \lambda \underline{I})\underline{x} = \underline{0}$, so does $(\underline{A} - \lambda \underline{I})\underline{z}$, with $\underline{z} = k\underline{x}$ (k a scalar). If the eigenvalues of $\underline{A}$ are distinct, meaning no two λs have the same value, then there is a characteristic vector associated with each λ and the set of all characteristic vectors is linearly independent. Let $\underline{\Lambda}$ be a diagonal matrix with diagonal elements set equal to the eigenvalues of $\underline{A}$ -- λs (and offdiagonal elements zero), and let $\underline{X}$ be a matrix whose columns form the eigenvectors of $\underline{A}$ associated with $\underline{\Lambda}$. That is $\underline{X} = [\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n]$, where $\underline{x}_1$ is the eigenvector of $\underline{A}$ associated with λ_1, etc. and

$$\underline{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

where $\underline{A}$ is square with order n. Now, if $(\underline{A} - \lambda_j\underline{I})\underline{x}_j = 0$, all $j = 1, \ldots, n$, so that $\underline{Ax}_j = \lambda_j\underline{x}_j$, then

$$\underline{AX} = \underline{X\Lambda}$$

Since the eigenvectors are independent, $\underline{X}^{-1}$ exists, and one finds

$$\underline{X}^{-1}\underline{AX} = \underline{\Lambda}$$

$\underline{A}$ can be transformed to a diagonal matrix by the given formula. In this case, $\underline{A}$ is said to be diagonalizable. One also has

$$\underline{A} = \underline{X\Lambda X}^{-1}$$

This relation forms the basis for some important results in linear differential equation systems. It should be noted that, if some of the roots (λs) of $\underline{A}$ are repeated (e.g., $\lambda_2 = \lambda_3$), $\underline{A}$ may not be diagonalizable, although it may sometimes be diagonalizable even if the λs are not distinct. Most matrices in empirical work will have n distinct roots and, hence, be diagonalizable.

If $\underline{A}$ is a real symmetric matrix, and $\underline{AX} = \underline{X\Lambda}$, then the transpose relationship also holds: $\underline{X}'\underline{A}' = \underline{\Lambda}'\underline{X}'$; $\underline{X}'\underline{A} = \underline{\Lambda X}'$, since $\underline{\Lambda}' = \underline{\Lambda}$. If $\underline{A}$ is not symmetric, then $\underline{AX} = \underline{X\Lambda} \rightarrow \underline{X}^{-1}\underline{A} = \underline{\Lambda X}^{-1}$. Thus $\underline{X}^{-1}$ plays the same role for asymmetric $\underline{A}$ that $\underline{X}'$ does for symmetric $\underline{A}$. One may calculate the solutions to the equation $\underline{y}'(\underline{A}-\mu\underline{I}) = 0$. Where $\underline{y}'$ is a 1 x n row vector and $|\underline{A} - \mu\underline{I}| = 0$. It turns out that the μs are the same as the λs. Forming a matrix whose columns consist of ys associated with λs, one has $\underline{Y} = [\underline{y}_1, \ldots, \underline{y}_n]$. It follows that $\underline{Y}'\underline{A} = \underline{\Lambda Y}'$. $\underline{Y}$ is called the matrix of left eigenvectors of $\underline{A}$ and $\underline{X}$ is the matrix of right eigenvectors. If $\underline{A}$ is symmetric, $\underline{Y} = \underline{X}$; otherwise, $\underline{Y}' = \underline{X}^{-1}$. Hence, the distinction between right and left eigenvectors is necessary only if $\underline{A}$ is asymmetric.

Recall that if $\underline{A}$ has a complex root, $\lambda = \lambda_R + \lambda_I i$, then the complex conjugate of λ is also an eigenvalue of $\underline{A}$ -- $\bar{\lambda} = \lambda_R - \lambda_I i$. Let $\underline{x}$ be the right eigenvector associated with complex λ, then $\underline{x} = \underline{x}_R + \underline{x}_I i$ is also complex, and its complex conjugate $\underline{\bar{x}} = \underline{x}_R - \underline{x}_I i$ is the eigenvector associated with $\bar{\lambda}$.

To illustrate these facts $\underline{\Lambda}$, $\underline{X}$ and $\underline{X}^{-1}$ are calculated for

$$\underline{A} = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}$$

$$\underline{Q} = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$$

$$\underline{B} = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$$

all of which have been used in previous examples. For the matrix $\underline{A}$,

$$\underline{\Lambda} = \begin{pmatrix} 3.5616 & 0 \\ 0 & -.5616 \end{pmatrix}$$

$$\underline{X} = \begin{pmatrix} 1.0 & -1.2808 \\ 1.2808 & 1.0 \end{pmatrix}$$

$$\underline{X}^{-1} = \begin{pmatrix} .3787 & .4851 \\ -.4851 & .3787 \end{pmatrix}$$

Checking the calculations, one finds

$$\underline{A} = \overset{\underline{X}}{\begin{pmatrix} 1.0 & -1.2808 \\ 1.2808 & 1.0 \end{pmatrix}} \overset{\underline{\Lambda}}{\begin{pmatrix} 3.5616 & 0 \\ 0 & -.5616 \end{pmatrix}} \overset{\underline{X}^{-1}}{\begin{pmatrix} .3787 & .4851 \\ -.4851 & .3787 \end{pmatrix}} = \overset{\underline{A}}{\begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}}$$

For the matrix $\underline{Q}$, one finds

$$\underline{\Lambda} = \begin{pmatrix} 7 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\underline{X} = \begin{pmatrix} 1 & -2 \\ 3 & 1 \end{pmatrix}$$

$$\underline{X}^{-1} = \begin{pmatrix} 1/7 & 2/7 \\ -3/7 & 1/7 \end{pmatrix}$$

with

$$\underline{Q} = \overset{\underline{X}}{\begin{pmatrix} 1 & -2 \\ 3 & 1 \end{pmatrix}} \overset{\underline{\Lambda}}{\begin{pmatrix} 7 & 0 \\ 0 & 0 \end{pmatrix}} \overset{\underline{X}^{-1}}{\begin{pmatrix} 1/7 & 2/7 \\ -3/7 & 1/7 \end{pmatrix}} = \overset{\underline{Q}}{\begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}}$$

For $\underline{B}$, one finds

$$\underline{\Lambda} = \begin{pmatrix} 2 + i, & 0 \\ 0 & ; 2 - i \end{pmatrix}$$

$$\underline{X} = \begin{pmatrix} 1 + i, & 1 - i \\ -(1 - i), & -(1 + i) \end{pmatrix}$$

$$\underline{X}^{-1} = \begin{pmatrix} \tfrac{1}{4}(1 - i), & -\tfrac{1}{4}(1 + i) \\ \tfrac{1}{4}(1 + i), & -\tfrac{1}{4}(1 - i) \end{pmatrix}$$

Checking the calculations, again one finds:

$$\underline{B} = \overset{\underline{X}}{\begin{pmatrix} 1 + i, & 1 - i \\ -(1 - i), & -(1 + i) \end{pmatrix}} \overset{\underline{\Lambda}}{\begin{pmatrix} 2 + i & 0 \\ 0 & 2 - i \end{pmatrix}} \overset{\underline{X}^{-1}}{\begin{pmatrix} \tfrac{1}{4}(1 - i), & -\tfrac{1}{4}(1 + i) \\ \tfrac{1}{4}(1 + i), & -\tfrac{1}{4}(1 - i) \end{pmatrix}}.$$

$$\underline{B} = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$$

Note that column 2 of $\underline{X}$ is the complex conjugate of column 1, just as $\lambda_2 = \bar{\lambda}_1$.

Matrix Functions

One of the most remarkable aspects of diagonalizable matrices is that interesting matrix functions may be defined by use of the eigenvalues and eigenvectors. Consider an elementary example. In the preceding section the inverse of the matrix:

$$\underline{A} = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}$$

was found to be:

$$\underline{A}^{-1} = \begin{pmatrix} -1 & 1 \\ 1 & -\tfrac{1}{2} \end{pmatrix}$$

and the eigenvalues and eigenvectors calculated to be:

$$\underline{\Lambda} = \begin{pmatrix} 3.5616 & 0 \\ 0 & -.5616 \end{pmatrix}, \quad \underline{X} = \begin{pmatrix} 1.0 & -1.2808 \\ 1.2808 & 1.0 \end{pmatrix}$$

$$\underline{X}^{-1} = \begin{pmatrix} .3787 & .4851 \\ -.4851 & .3787 \end{pmatrix}$$

Now, the inverse of $\underline{A}$ can be calculated by the following formula:

$$\underline{A}^{-1} = \underline{X\Lambda^{-1}}\underline{X}^{-1}$$

$$\underline{A}^{-1} = \begin{pmatrix} 1.0 & -1.2808 \\ 1.2808 & 1.0 \end{pmatrix} \begin{pmatrix} 1/3.5616 & 0 \\ 0 & -1/.5616 \end{pmatrix} \begin{pmatrix} .3787 & .4851 \\ -.4851 & .3787 \end{pmatrix} =$$

$$= \begin{pmatrix} -1 & 1 \\ 1 & -\frac{1}{2} \end{pmatrix}$$

The proof that this is a general result is simple and instructive. Let $\underline{A}$ be a nonsingular, diagonalizable matrix, so that $\underline{A} = \underline{X\Lambda X}^{-1}$, where Λ and X are the characteristic roots and vectors of $\underline{A}$, as defined previously. If $\underline{A}$ is full rank (nonsingular) then, it has been noted, all the eigenvalues of $\underline{A}$ are nonzero; hence $\underline{\Lambda}^{-1}$ exists. Let $\underline{B} = \underline{X\Lambda}^{-1}\underline{X}^{-1}$, then

$$\underline{AB} = (\underline{X\Lambda X}^{-1})(\underline{X\Lambda}^{-1}\underline{X}^{-1})$$

$$= \underline{X\Lambda\Lambda}^{-1}\underline{X}^{-1}$$

$$\underline{AB} = \underline{I}$$

Hence, $\underline{B}$ must be the inverse of $\underline{A}$

The square root of a matrix may be defined in an analogous way. Let $\underline{A}^{\frac{1}{2}} = \underline{X\Lambda}^{\frac{1}{2}}\underline{X}^{-1}$ with $\Lambda \geq 0$ (meaning every element of $\underline{\Lambda}$ is nonnegative). Then $\underline{A}^{\frac{1}{2}}\underline{A}^{\frac{1}{2}} = (\underline{X\Lambda}^{\frac{1}{2}}\underline{X}^{-1})(\underline{X\Lambda}^{\frac{1}{2}}\underline{X}^{-1}) = \underline{X\Lambda}^{\frac{1}{2}}\underline{\Lambda}^{\frac{1}{2}}\underline{X}^{-1} = \underline{X\Lambda X}^{-1} = \underline{A}$. This idea can be generalized to any function that can be represented by a series expansion. Examples include the trigonometric functions, logarithm and exponential functions. The matrix exponential and log functions are two important special cases, as they play an important role in the next chapter. They can be defined by

$$e^{\underline{A}} = \underline{X}e^{\underline{\Lambda}}\underline{X}^{-1}$$

$$\ln \underline{A} = \underline{X}\ (\ln \underline{\Lambda})\underline{X}^{-1}$$

Using series expansion of e^{λ_j} and $\ln \lambda_j$, it can be found that

5.

$$e^{\underline{\Lambda}} = \begin{pmatrix} e^{\lambda_1} & 0 \ldots 0 \\ 0 & e^{\lambda_2} & \vdots \\ \vdots & \vdots \\ 0 & 0 \ldots e^{\lambda n} \end{pmatrix}, \quad \ln \underline{\Lambda} = \begin{pmatrix} \ln\lambda_1 & 0 & \ldots & 0 \\ 0 & \ln\lambda_2 & & \vdots \\ \vdots & \vdots \\ 0 & 0 \ldots & \ln\lambda_n \end{pmatrix}$$

Hence,

$$e^{\underline{A}} = \begin{pmatrix} x_{11} & \ldots & x_{1n} \\ \vdots & & \vdots \\ x_{n1} & \ldots & x_{nn} \end{pmatrix} \begin{pmatrix} e^{\lambda_1} & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & e^{\lambda n} \end{pmatrix} \begin{pmatrix} x_{11} & \ldots & x_{1n} \\ \vdots & & \vdots \\ x_{n1} & \ldots & x_{nn} \end{pmatrix}^{-1}$$

$$\ln \underline{A} = \begin{pmatrix} x_{11} & \ldots & x_{1n} \\ \vdots & & \vdots \\ x_{n1} & \ldots & x_{nn} \end{pmatrix} \begin{pmatrix} \ln\lambda_1 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & \ln\lambda_n \end{pmatrix} \begin{pmatrix} x_{11} & \ldots & x_{1n} \\ \vdots & & \vdots \\ x_{n1} & \ldots & x_{nn} \end{pmatrix}^{-1}$$

Obviously, $e^{\ln \underline{A}} = \ln(e^{\underline{A}}) = \underline{A}$.

Consider a matrix $\underline{A}(t)$ whose elements are a function of t. If $\underline{A}(t)$ is diagonalizable such that the characteristic vectors, $\underline{X}$, are the same for all t, then the derivative and integral of $\underline{A}(t)$ with respect to (t) can be written in terms of eigenvalues and eigenvectors. An example of particular importance in the next chapter is the integral $\int e^{\underline{A}t}dt$. For a sense of closure and general interest, both the derivative and integral of $e^{\underline{A}t}$ are given:

$$\frac{de^{\underline{A}t}}{dt} = \underline{X}\left(\frac{de^{\underline{\Lambda}t}}{dt}\right)\underline{X}^{-1} = \underline{X}(e^{\underline{\Lambda}t}\underline{\Lambda})\underline{X}^{-1}$$

$$= (\underline{X}e^{\underline{\Lambda}t}\underline{X}^{-1})(\underline{X\Lambda X}^{-1})$$

$$= e^{\underline{A}t}\underline{A} = \underline{A}e^{\underline{A}t}, \text{ since } e^{\underline{\Lambda}t}\underline{\Lambda} = \underline{\Lambda}e^{\underline{\Lambda}t}$$

Also,

$$\int e^{\underline{A}t}dt = \underline{X}(\int e^{\underline{\Lambda}t}dt)\underline{X}^{-1} = \underline{X}e^{\underline{\Lambda}t}\underline{\Lambda}^{-1}\underline{X}^{-1}, \quad |\underline{A}| \neq 0$$

$$= (\underline{X}e^{\underline{\Lambda}t}\underline{X}^{-1})(\underline{X\Lambda}^{-1}\underline{X}^{-1})$$

$$= e^{\underline{A}t}\underline{A}^{-1} = \underline{A}^{-1}e^{\underline{A}t}, \quad |\underline{A}| \neq 0$$

Statistical Concepts

It is assumed that the reader is familiar with most of the important concepts in statistics needed for this volume, including, in particular, the population/sample mean, variance, covariance, correlation, least-squares regression, and significance tests for statistics such as these. There are some fundamental ideas and concepts in inferential statistics, however, that appear to require some explication. In particular, these include the concepts of expected value of a sample statistic, its variance, bias, and consistency. The following paragraphs attempt to describe these concepts so that their meaning will be clear when they are used in the next chapter.

The distinctions between sample distribution, population distribution and sampling distribution are critical to thorough understanding of elementary concepts in inferential statistics. Suppose one has a population or universe defined by the set of all cases that a research question addresses. Assume that for each case in the population it is theoretically possible to observe a value on a variable called x. Each value of x has a probability associated with it, p_x giving the relative frequency of that value of x in the population.[9] The set of all pairs (x,p_x) is the <u>population</u> distribution of x. The <u>sample</u> distribution of x refers to the relative frequencies associated with each value (or range of values) of x observed in the particular sample. The term <u>sampling</u> distribution generally is not applied to x at all. Rather, it is applied to some function of the sample x s, such as the sample mean, variance, or median. Take the sample mean as an example, and consider calculating the sample mean from a very large number of samples, each of size n. The distribution of these means over all possible samples is termed the sampling distribution of the mean. One of the most remarkable theorems in statistics is that, no matter what the population distribution of x, as the sample size gets large, the sampling distribution of the mean tends to the normal distribution. This fact forms part of the central limit theorem.

The expected value of a sample statistic is defined as the mean of that statistic calculated over all possible samples. Clearly it is desirable that the expected value of a sample statistic equal the value of the population parameter that it is intended to estimate. The <u>standard error</u> of a sample statistic is defined as the standard deviation of that statistic calculated

9. This is not a precise definition, but is sufficient for present purposes. The concept of probability element needed for continuous density functions would complicate unduly the presentation.

CHAPTER 4

MATHEMATICAL AND STATISTICAL THEORY IN THE APPLICATION OF LINEAR DIFFERENTIAL EQUATIONS

This chapter is divided into three main sections. The first section contains a brief discussion of general differential equation systems. The purpose of this discussion is to show how the particular application of linear systems with constant coefficients fits with the general theory, and thereby, to communicate an idea of the untapped potential of differential-equation systems for representing social processes. Section two reviews the mathematical and statistical theory necessary to convert a model written as differential equations into a form suitable for empirical study. The third section describes a computer program package that can be used to carry out calculations needed to estimate parameters of the differential-equation model.

General Differential-Equation Systems

In order to gain some understanding of general differential equation systems it is useful to consider specific cases first and work up gradually to the general case. A very simple case arises for the continuous population-growth model. Its differential equation is of the following form.

(5a) $dy/dt = by$

Where y is population size, b is the constant rate of population growth, and, as before, t is time. It is useful to summarize the meaning of this simple differential equation, because in most important respects the meaning generalizes to more complicated differential equations. One might view the lefthand side (dy/dt) as the "dependent variable." The dependent variable, then, is the rate of change in population size (y) at a given instant in time. The righthand side of the equation may be interpreted as a hypothesis about the manner in which the rate of change in population size occurs. In this simple example, it is hypothesized that rate of change in population size is a constant number (b) multiplied by the current population size (y). More generally, the righthand side will contain the current value of y, time, and any number of additional independent variables. The functional form of the right side may vary depending on the substantive application.

Integrating equation (5a) generates the familiar exponential population growth curve, as follows:

(5b) $y_t = y_o e^{bt}$

where t is a specific point in time, and y_t, y_o are population size at time t and time 0, respectively. Equation (5b) also describes the accumulation of principal on an investment with a constant rate of interest (return) of b and continuous compounding of the interest.

Equation (5a) is a single linear differential equation. It becomes nonlinear if the function of y on the right side is nonlinear. For example,

$$(6) \quad dy/dt = b \sin y$$

is a single, nonlinear differential equation.

If there is more than one equation, then one has a system of simultaneous differential equations. For example, the following pair of linear differential equations form a special case of the model of career expectations given in equations (4).

$$(7a) \quad dy_1/dt = b_{11}y_1 + b_{12}y_2$$

$$(7b) \quad dy_2/dt = b_{21}y_1 + b_{22}y_2$$

where y_1, y_2 are two variables, the b_{ij} are constants, and t is time. Since there is no constant intercept in equations (7), they form a pair of homogeneous linear differential equations. Adding an intercept makes the system nonhomogeneous and parallel in every technical respect to the dynamic model of career expectations in equations (4). Equations (8) show an example of linear nonhomogeneous system composed of two equations.

$$(8a) \quad dy_1/dt = a_1 + b_{11}y_1 + b_{12}y_2$$

$$(8b) \quad dy_2/dt = a_2 + b_{21}y_1 + b_{22}y_2$$

where a_1 and a_2 are constant intercepts, and the other notations are defined as in (7).

Equations (7) and (8) are linear systems with constant coefficients. More generally, the coefficients of linear systems may be taken as functions of time. For example, suppose that y_1 represents ego's level of educational expectation and y_2 is significant other's level of educational expectation for ego. The idea that the influence of the significant other on ego's educational expectation level declines with time could be expressed by making b_{12} a declining function of time, say $b_{12} = e^{-qt}$, q a positive constant.

So far, all the examples have been first-order differential equations, because only the first derivative appears in them. Derivatives of any order may appear in a differential equation. The order of the highest derivative in a given differential equation defines the order of the equation. The order of a system of differential equations is defined by the sum of the orders of the equations forming the system. Meaningful examples of second and higher order differential equations involving social or psychological variables are scarce, but there are numerous physical examples. Probably the most common physical example relates to distance traveled as a function of time. Speed is the first derivative and acceleration is the second derivative of distance with respect to time. For example, if y is distance and t is time, one might posit that acceleration of a vehicle is a positive but declining function of time, say

(9a) $d^2y/dt^2 = ce^{-bt}$

where c and b are positive constants. With this equation one may find the distance traveled from a standing start as a function of time, as follows

(9b) $y = (c/b)t + (c/b^2)(e^{-bt} - 1)$

where y_t is distance traveled from a standing start after t minutes have elapsed. Thus, the single second order differential equation (9a) can be operated on to produce a prediction about distance traveled from a standing start at every point in time. These predictions could be compared to data to see if the acceleration hypothesis is correct.

Although social science definitions are seldom given as second derivatives, there are some concepts that might be defined fruitfully in this manner. For example, one might define learning speed as the change in information divided by the change in time. At an instant in time, then, learning could be defined as the derivative of information with respect to time. A change in learning speed, therefore, would be the second derivative of information with respect to time. Thus, the concept of accelerated learning could be given a precise definition quite analogous to the physical concept of acceleration. Similarly, "vertical" occupational mobility could be defined as the derivative of occupational status with respect to time, so that change in mobility rate would be a second derivative. A hypothesis about changing rate of mobility, analogous to the hypothesis about acceleration in equation (9a), could be used to generate predictions about occupational status level at any point in time.

In all the examples, derivatives have been with respect to time alone. Whenever differential equations involve derivatives with respect to only one variable they are termed ordinary differential equations. If derivatives with respect to more than one variable are included one has partial differential equations.

A very general first-order, ordinary differential equation system can be written as follows:

$$
(10)\quad
\begin{aligned}
dy_1/dt &= f_1(y_1, y_2, \ldots, y_k, t)\\
dy_2/dt &= f_2(y_1, y_2, \ldots, y_k, t)\\
&\vdots\\
dy_k/dt &= f_k(y_1, y_2, \ldots, y_k, t)
\end{aligned}
$$

where $y_1, y_2, \ldots, y_k$ comprise a set of k variables, t is time, and f_i are any continuous functions of the ys and of t. More general equation systems involving higher order derivatives and partial derivatives can also be written but are not needed for the present discussion.

The next section of this chapter develops the technical aspects of linear systems of ordinary differential equations with constant coefficients. As illustrated in the present section there are numerous ways in which the restrictive assumptions associated with such systems can be relaxed.

Technical Aspects of Ordinary Linear Differential Equation Systems with Constant Coefficients

It is impossible to observe all the terms in a differential equation, since a differential equation always contains at least one derivative, the derivative being a slope at a single point along a smooth curve. Take equation (5a) as an example. The dependent variable is the instantaneous rate of change in population size --

$$dy/dt = \lim_{t_1 \to t_0} \frac{y_1 - y_0}{t_1 - t_0}$$

where t_1 and t_0 are the two points in time that are very close together and y_1 and y_0 are the population sizes at those time points. Of course, it might be possible in theory to approximate the derivative at several time points by drawing a sequence of observation pairs with the time points of each pair spaced a finite but short distance apart. In practice, however, this strategy would be difficult or impossible to carry out for most social or psychological topics. A much more practical strategy is to operate on qeuation (5a) to produce (5b), which can be compared readily to observations.

Note that if one differentiates (5b) with respect to t, one finds

$$dy_t/dt = b(y_0 e^{bt}) = by_t, \text{ since } y_t = y_0 e^{bt}$$

This exercise shows that if one differentiates (5a) one obtains equation (5b). Equation (5b) is, therefore, the antiderivative or indefinite integral of equation (5a). Equation (5b) is termed, therefore, a solution of equation (5a). Referring to the example regarding acceleration given by equations (9), the relationship between (9a) and (9b) is parallel to the relationship between (5a) and (5b); equation (9b) is a solution of the second order differential equation, equation (9a), meaning that, if (9b) is differentiated twice, one gets (9a) back:

$$\frac{d^2y}{dt^2} = \frac{d}{dt}\left[\frac{dy}{dt}\right]$$

$$= \frac{d}{dt}\ \frac{d}{dt}\left[\frac{c}{b}t + \frac{c}{b^2}(e^{-bt} - 1)\right]$$

$$= \frac{d}{dt}\left[\frac{c}{b} - \frac{c}{b}\ e^{-bt}\right]$$

$$\frac{d^2y}{dt^2} = ce^{-bt}$$

As does equation (5b), equation (9b) gives predictions that are observable. In general, finding a solution to a differential equation may be interpreted as the process of converting hypotheses about change at every instant in time into a form suitable for use with empirical data.

The primary objective of this section is to describe practical methods for converting the dynamic model of career expectations given by equations (4) into a form suitable for use in empirical work. The section is subdivided into two subsections. The first subsection develops the mathematical theory necessary for finding integrals of linear systems, and the second subsection considers statistical issues related to estimating parameters of the integral equation.

Mathematical Theory

This subsection contains a review of mathematical solutions to systems of ordinary linear differential equations with constant coefficients; it does not purport to develop a rigorous statement of the mathematical theory. The review unfolds gradually, developing first the basic theory for homogeneous systems, then adding constant intercepts, and, finally, including a set of disturbance variables.

Consider the following equation expressing the general model of a linear differential equation system.

(11) $d\underline{y}/dt = \underline{a}(t) + \underline{B}(t)\underline{y}$

where

$\underline{y}$ = a K x 1 vector of variables

$\underline{a}(t)$ = a K x 1 vector of intercepts that are continuous functions of time

$\underline{B}(t)$ = a K x K matrix with each entry a continuous function of time

$d\underline{y}/dt$ = a K x 1 vector of derivatives of $\underline{y}$ with respect to time

If $\underline{a}(t) \equiv 0$ for all t, then the system is termed a <u>homogeneous system</u>. If every entry of $\underline{B}(t)$ is a constant over time the system is a linear system with <u>constant coefficients</u>.

Attention now focuses on homogeneous linear systems with constant coefficients, because essential aspects of the results for nonhomogeneous systems can be derived from the theory for homogeneous systems. The homogeneous system is written

(12) $\dot{\underline{y}} = \underline{B}\underline{y}$

where $\dot{\underline{y}} = d\underline{y}/dt$, and $\underline{B}$ is the K x K matrix of constant coefficients. The goal is to operate on (12) to produce a function giving each element of the vector $\underline{Y}$ as a function of time, subject to the restriction that equation (12) holds.

Consider first a one-element vector and matrix, so that (12) becomes a scalar equation rather than a matrix equation.

(13) $dx/dt = bx$

where x,b are scalars. The simplest way to solve (13) is to divide both sides by x and multiply by dt, producing

$$dx/x = bdt$$

integrating both sides yields

$$\int \frac{dx}{x} = \int bdt$$

(14) $\ln x = bt + q$

$$x = e^{bt+q} = e^{bt}(e^{q})$$

e is the mathematical constant which is the base of the natural logarithm, q is an arbitrary constant, and ln is the standard notation for the natural logarithm used throughout this volume. Notice that if the constant q were known, equation (14) achieves the goal for scalars -- x is expressed as a function of time. Note also that the derivative of (14) satisfies the differential equation (13), for

$$dx/dt = be^{bt}(e^{q}) = bx, \text{ since from (14), } x = e^{bt}(e^{q})$$

This simple method for finding a solution to (13) does not generalize conveniently to linear systems, since one cannot divide by a vector, y. It is useful, therefore, to develop an alternative derivation of the solution (14). To do so, subtract bx from both sides of the scalar differential equation (13); this yields:

$$dx/dt - bx = 0$$

Multiply both sides by e^{-bt} to get:

$$e^{-bt}(dx/dt - bx) = 0$$

Notice that $de^{-bt}x = e^{-bt}dx/dt - bxe^{-bt}$, by the multiplication rule for derivatives; hence, the above result can be rewritten

$$de^{-bt}x = 0$$

Integrating both sides yields

$$e^{-bt}x = c = e^{q}$$

$$x = e^{bt}c$$

where c is an arbitrary constant.

This result matches (14) with $c = e^q$ and, as it turns out, this second method of arriving at (14) generalizes readily to systems of linear, homogeneous differential equations.

To find the constant c, let x_t designate the value of x at a point in time, t. Now, note that, from equation (14), at the zero point on the time scale x may be found by

$$x_o = e^{b.0}(e^q) = e^{b.0}c$$

$$x_o = c$$

Hence, c can be set to the observed value of x at an initial panel in a longitudinal study. The value of x can then be made a function of time by the following formula.

$$(15) \quad x_t = e^{bt}c = e^{bt}e^q$$

$$x_t = e^{bt}x_0$$

where t is any point in time. This result is, of course, the equation for the exponential population-growth model noted earlier in this chapter.

Equation (15) could be used in conjunction with a two-panel longitudinal study with one (or more) observations on x at $t_o = 0$ and at $t_1 = t$. Define $b^* = e^{bt}$ and rewrite (15), as follows.

$$(15a) \quad x_t = b^*x_o$$

Obviously from (15a), $b^* = x_t/x_o$. One can calculate b from b^* as follows: $b = (\ln b^*)/t$. Substituting for b^*, one can calculate b directly.

$$(16) \quad b = [\ln(x_t/x_o)]/t$$

Hence, if there is no disturbance (error) term, an observation at two time points (t_o, t_1) for a single case suffices to generate exact predictions at every point in time -- by substituting the value of b calculated from (16) into the prediction equation (15).

The groundwork has now been developed for finding a solution to the linear homogeneous system, equation (12). To arrive at a solution, retrace the steps used in the second derivation of equation (14). Start with (12) and subtract By from both sides.

$$\dot{\underline{y}} = \underline{B}\underline{y}$$

$$\dot{\underline{y}} - \underline{B}\underline{y} = \underline{0}$$

Premultiply both sides by e^{-Bt}, where e^{-Bt} is a matrix exponential,

$$e^{-\underline{B}t}(\dot{\underline{y}} - \underline{B}\underline{y}) = \underline{0}$$

It can be shown that the derivative of the matrix product, $e^{-\underline{B}t}\underline{y}$, is equal to $e^{-\underline{B}t}\dot{\underline{y}} - e^{-\underline{B}t}\underline{B}\underline{y}$. Consequently, the above expression can be written as follows

$$de^{-\underline{B}t}\underline{y} = 0$$

just as for the scalar case. Integrating both sides yields

$$(17) \quad \begin{aligned} e^{-\underline{B}t}\underline{y} &= \underline{c} \\ \underline{y} &= e^{\underline{B}t}\underline{c} \end{aligned}$$

where $\underline{c}$ is a K x 1 vector of arbitrary constants. To find the vector of constants, as for the scalar case, let $t_o = 0$, $t_1 = t$, $\underline{y}_t$ = the value of the vector y at time t.

Then from (17)

$$\underline{y}_o = e^{\underline{B}\cdot 0}\underline{c} = \underline{I}\cdot c, \ \underline{I} = K \times K, \text{ identity matrix}$$

$$\underline{c} = \underline{y}_o$$

Substituting $\underline{c} = \underline{y}_o$ into (17) yields the desired prediction equation:

$$(18) \quad \underline{y}_t = e^{\underline{B}t}\underline{y}_o$$

Again, as in the scalar case, the value of $\underline{B}$ can be found from longitudinal data. There are at least two possibilities. First, one might have at least K cases at two time points (K = the number of variables). Secondly, one might have at least K + 1 observations for a single case spaced at equal time intervals. For the first instance, let $\underline{Y}_t$ be a K × K matrix of observations on the K variables for K cases, at time t, and let $\underline{Y}_o$ be the analogous matrix at time zero. Assuming $\underline{Y}_o$ is invertable, one finds

$$e^{\underline{B}t} = \underline{Y}_t\underline{Y}_o^{-1}$$

(19) $$\underline{B} = [\ln(\underline{Y}_t\underline{Y}_o^{-1})]/t, \quad |\underline{Y}_t\underline{Y}_o^{-1}| \neq 0$$

($\ln(\underline{Y}_t\underline{Y}_o^{-1})$ is a matrix logarithm, not an elementwise log). If there are more than K cases, any subset of K cases can be taken so long as those selected are linearly independent, so that $\underline{Y}_o^{-1}$ exists.

For the second isntance in which K + 1 or more observations on a single case are available, $\underline{Y}_o$ may be taken as K column vectors, each column being an observation on K variables at a single time point, starting with $t_o = 0$; $\underline{Y}_t$ can be defined analogously, but with the first column of $\underline{Y}_t$ being a set of observations at the second time point, that is

$$\underline{Y}_o = [\underline{y}_o, \underline{y}_1, \ldots, \underline{y}_K]$$

$$\underline{Y}_t = [\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_{K+1}]$$

The length of time between adjacent observations must all be the same.

Recall the definitions of the matrix logarithm and matrix exponential given in the mathematical review, chapter 3, viz.

(20) $\ln \underline{M} = \underline{V}(\ln \underline{\Lambda})\underline{V}^{-1}, \quad |\underline{M}| \neq 0$

(21) $e^{\underline{M}} = \underline{V}e^{\underline{\Lambda}}\underline{V}^{-1}$

where

$\underline{M}$ = a diagonalizable K x K matrix, $|\underline{M}|$ = determinant of $\underline{M}$

$\underline{\Lambda}$ = a K x K diagonal matrix with diagonal entries containing the characteristic values of $\underline{M}$

$\underline{V}$ = a K x K matrix whose columns are the characteristic vectors of $\underline{M}$ associated with $\underline{\Lambda}$

The functions $\ln \underline{\Lambda}$ and $e^{\underline{\Lambda}}$ are defined on each diagonal element of $\underline{\Lambda}$, i.e.,

$$\ln \underline{\Lambda} = \begin{pmatrix} \ln \lambda_1 & 0 & \dots & 0 \\ 0 & \ln \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \ln \lambda_K \end{pmatrix}$$

$$e^{\underline{\Lambda}} = \begin{pmatrix} e^{\lambda_1} & 0 & \dots & 0 \\ 0 & e^{\lambda_2} & & \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & e^{\lambda_K} \end{pmatrix}$$

(Recall from chapter 3 that these definitions do fit the general definition of matrix logarithm and of matrix exponential. That is, $\ln(e^{\underline{\Lambda}}) = e^{\ln \underline{\Lambda}} = \underline{\Lambda}$).

If every element of a matrix is multiplied by a scalar, as in equation (18) where $\underline{B}$ is multiplied by the scalar t, then the characteristic values of the resulting matrix, $\underline{B}t$, are just the characteristic values of the matrix $\underline{B}$ multiplied by the scalar t. Thus, if $\underline{\Lambda}$ is the diagonal matrix of eigenvalues, of $\underline{B}$, then $\underline{\Lambda}t$ is the diagonal matrix of eigenvalues of the matrix $\underline{B}t$. Hence, equation (21) can be written as follows, for the matrix $\underline{B}t$.

$$(22) \quad e^{\underline{B}t} = \underline{V}e^{\underline{\Lambda}t}\underline{V}^{-1}$$

where $\underline{V}$, $\underline{\Lambda}$ are now taken as eigenvector and eigenvalue matrices, respectively, of $\underline{B}$.

The matrix derivative $\frac{d}{dt}e^{\underline{B}t} = e^{\underline{B}t}\underline{B}$ was used to derive the solution to the homogeneous linear system, equation (12), the solution being given by equation (17). Equation (22) can be used conveniently to derive the formula for the derivative of the matrix exponential, $e^{\underline{B}t}$.

$$\frac{d}{dt}e^{\underline{B}t} = \frac{d}{dt}[\underline{V}e^{\underline{\Lambda}t}\underline{V}^{-1}]$$

$$= \underline{V}(\frac{d}{dt}e^{\underline{\Lambda}t}\underline{V}^{-1})$$

Note that if no diagonal entry in $\underline{\Lambda}$ is zero, the derivative of the diagonal matrix, $e^{\underline{\Lambda}t}$, is

$$\frac{d}{dt}e^{\lambda t} = \begin{pmatrix} de^{\lambda_1 t}/dt & d0/dt & \dots & d0/dt \\ d0/dt & de^{\lambda_2 t}/dt & \dots & d0/dt \\ \vdots & \vdots & & \\ d0/dt & d0/dt & \dots & de^{\lambda_K t}/dt \end{pmatrix}$$

$$\begin{pmatrix} e^{\lambda_1 t}\lambda_1 & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t}\lambda_2 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & e^{\lambda_K t}\lambda_K \end{pmatrix}$$

Consequently,

$$\begin{aligned} \frac{d}{dt}e^{\underline{B}t} &= \underline{V}e^{\underline{\Lambda}t}\underline{\Lambda}\underline{V}^{-1} & &= \underline{V}\underline{\Lambda}e^{\underline{\Lambda}t}\underline{V}^{-1} \\ &= \underline{V}e^{\underline{\Lambda}t}\underline{V}^{-1}\underline{V\Lambda V}^{-1} & &= \underline{V\Lambda V}^{-1}\underline{V}e^{\underline{\Lambda}t}\underline{V}^{-1} \\ &= (\underline{V}e^{\underline{\Lambda}t}\underline{V}^{-1})(\underline{V\Lambda V}^{-1}) & &= (\underline{V\Lambda V}^{-1}(\underline{V}e^{\underline{\Lambda}t}\underline{V}^{-1}) \\ &= e^{\underline{B}t}\underline{B} & &= \underline{B}e^{\underline{B}t} \end{aligned}$$

Since $\underline{B} = \underline{V\Lambda V}^{-1}$,

Having found the solution to a homogeneous system of linear differential equations with constant coefficients, one is prepared to deal with nonhomogeneous systems with constant coefficients. The mathematical theory is presented first, then it is applied to the model of developing career aspirations given in equation (4).

The nonhomogeneous system with constant coefficients is written as follows.

(23) $\dot{\underline{y}} = \underline{a}(t) + \underline{B}\underline{y}$

where $\underline{a}(t)$ may be a variable function of time, but the matrix $\underline{B}$ is fixed over time. To find the solution to (23), trace through the same steps used to solve the homogeneous system.

$$\underline{\dot{y}} - \underline{B}\underline{y} = \underline{a}(t)$$

$$e^{-\underline{B}t}(\underline{\dot{y}} - \underline{B}\underline{y}) = e^{-\underline{B}t}\underline{a}(t)$$

$$d(e^{-\underline{B}t}\underline{y}) = e^{-\underline{B}t}\underline{a}(t)dt$$

$$\int d(e^{-\underline{B}t}\underline{y}) = \int e^{-\underline{B}t}\underline{a}(t)dt + \underline{c}^*$$

$$e^{-\underline{B}t}\underline{y}_t = \int_o^t e^{-\underline{B}\tau}\underline{a}(\tau)d\tau + \underline{c}$$

$$(24) \qquad \underline{y}_t = e^{\underline{B}t}\underline{c} + e^{\underline{B}t}\int_o^t e^{-\underline{B}\tau}\underline{a}(\tau)d\tau$$

where τ is the dummy variable of integration as explained on page 31. Setting $t_o = 0$ and solving for the constant vector, $\underline{c}$, leads to the conclusion that $\underline{c} = \underline{y}_o$. Inserting $\underline{c} = \underline{y}_o$ into (24) yields

$$(25) \quad \underline{y}_t = e^{\underline{B}t}\underline{y}_o + e^{\underline{B}t}\int_o^t e^{-\underline{B}\tau}\underline{a}(\tau)d\tau$$

Notice that the term $e^{\underline{B}t}\int_o^t e^{-\underline{B}\tau}\underline{a}(\tau)d\tau$ is a K x 1 vector.

Denote it by $\underline{a}^* = e^{\underline{B}t}\int_o^t e^{-\underline{B}\tau}\underline{a}(\tau)d\tau$, and let $\underline{B}^* = e^{\underline{B}t}$.

Equation (25) can now be written

$$(25b) \quad \underline{y}_t = \underline{a}^* + \underline{B}^*\underline{y}_o$$

For two fixed points in time, $t_o = 0$, and $t_1 = t$, equation (25b) is linear across observations. Thus, if k + 1 or more observations at $t_o = 0$ and $t_1 = t$ are available one may form the matrices

$$\underline{Y}_t = \begin{pmatrix} 1 & \cdots & 1 \\ y_{11t} & \cdots & y_{1Kt} \\ y_{21t} & \cdots & y_{2Kt} \\ \vdots & & \vdots \\ y_{K1t} & \cdots & y_{KKt} \end{pmatrix}, \text{ and}$$

$$\underline{P} = [\underline{a}^*, \underline{B}^*]$$

Then for two time points, $t_o = 0$, $t_1 = t$,

$$(26) \quad \underline{Y}_t = \underline{P}\underline{Y}_o$$

$$\underline{P} = \underline{Y}_t\underline{Y}_o^{-1} \quad |\underline{Y}_o| \neq 0$$

where $|\underline{Y}_o|$ stands for the determinant of $\underline{Y}_o$.

As with the homogeneous case, if more than the minimum number of cases are available at t_o and at t_1, then K + 1 linearly independent cases can be selected arbitrarily. Also, if K + 2 time points on a single case are available, $\underline{Y}_o$ can be defined by the first K + 1 observations and $\underline{Y}_t$ by the second through the K + 2 cases; then equation (26) still holds.

Once $\underline{P} = [\underline{a}^*, \underline{B}^*]$ is found, $\underline{a}^*$ and $\underline{B}^*$ can be used separately to find the parameters of the nonhomogeneous differential equation (23). Since $\underline{B}^* = e^{\underline{B}t}$, one has

$$(27a) \quad \underline{B} = (\ln \underline{B}^*)/t, \quad |\underline{B}^*| \neq 0$$

where $\ln \underline{B}^*$ is the matrix natural logarithm.

Note that this is the same solution as for the homogeneous case.

Appropriate use of $\underline{a}^*$ depends on the nature of the function $\underline{a}(t)$. Let

$$\underline{q}(t, \alpha) = \int_o^t e^{-\underline{B}\tau}\underline{a}(\tau)d\tau$$

where α is a set of parameters of the function a(t). One has

$$\underline{a}^* = e^{\underline{B}t}\underline{q}(t, \alpha)$$

(27b) $$e^{-\underline{B}t}\underline{a}^* = \underline{q}(t, \alpha)$$

Under fairly general circumstances, equation (27b) can be solved for the unknown parameters of the differential equation. To illustrate, two cases may be useful. First, let the function $\underline{a}(t)$ be constant over t, say $\underline{a}(t) = \underline{\alpha}$, $\underline{\alpha}$ a set of constant coefficients. Then equation (27b) becomes

$$e^{-\underline{B}t}\underline{a}^* = \int_0^t e^{-\underline{B}\tau}\underline{\alpha}\, d\tau = [-e^{-\underline{B}\tau}\underline{B}^{-1}\underline{\alpha}]_o^t = (\underline{I} - e^{-\underline{B}t})\underline{B}^{-1}\underline{\alpha}$$

$$\underline{\alpha} = \underline{B}(e^{\underline{B}t} - \underline{I})^{-1}\underline{a}^*, \quad |e^{\underline{B}t} - \underline{I}| \neq 0 \leftrightarrow |\underline{B}| \neq 0$$

For a second example, consider a more involved instance. Suppose that each element of $\underline{a}(t)$ is the cosine of $t + \alpha_i$, and denote the entire vector by $\underline{a}(t) = c(\underline{\alpha}+t)$. In this case, the vector $\underline{a}(t)$ oscillates over time, but the different elements of $\underline{a}(t)$ do not oscillate in phase unless all α_i are the same. The function $\underline{q}(t,\underline{\alpha}) = \int e^{-\underline{B}t}c(\underline{\alpha}+t)dt = (I + \underline{B}^2)^{-1}\{e^{-\underline{B}t}[s(\underline{\alpha}+t) - \underline{B}c(\alpha+t)] - [s(\alpha) - Bc(\alpha)]\}$, where the elements of $s(\alpha+t)$ are $\sin(\alpha_i+t)$. This result can be substituted on the right side of (27b), as follows:

$$e^{-\underline{B}t}\underline{a}^* = (\underline{I} + \underline{B}^2)^{-1}\{e^{-\underline{B}t}[s(\underline{\alpha}+t) - \underline{B}c(\underline{\alpha}+t)]$$
$$- s(\underline{\alpha}) + \underline{B}c(\underline{\alpha})\}$$

$$(\underline{I} + \underline{B}^2)\underline{a}^* = s(\underline{\alpha}+t) - \underline{B}c(\underline{\alpha}+t) - e^{\underline{B}t}[s(\underline{\alpha}) - \underline{B}c(\underline{\alpha})]$$

since $(\underline{I} + \underline{B}^2)^{-1}$ commutes with $e^{\underline{B}t}$. Because the sine and cosine functions are periodic both with period 2π, if t is judiciously selected so that $t = 2n\pi$, n a positive integer, then this formula reduces to

$$(\underline{I} + \underline{B}^2)\underline{a}^* = (\underline{I} - e^{\underline{B}t})[s(\underline{\alpha}) - \underline{B}c(\underline{\alpha})]$$

In either form, it is unlikely that a simple algebraic solution can be found for $\underline{\alpha}$. Nevertheless, for any application where $\underline{B}$ and $\underline{a}^*$ can be calculated and t is known, it is likely that a numerical solution can be found. Use of numerical analysis to find solutions to equations that are intractable or insoluable algebraically is an option that is not appreciated by many social researchers. One of the purposes of this example is to emphasize that algebraic solutions are not necessary to the conduct of empirical research.

The mathematical theory required for the model of career expectations given by equation system (4) has been summarized in the preceding discussion. It now remains to translate the general results into a specific form suitable for use with the substantive theory of career expectations. To achieve this translation it is desirable to rewrite equation system (4) in matrix notation, as follows.

(28) $\dot{\underline{y}} = \underline{Ax} + \underline{By} + \underline{u}(t)$

where

$$\dot{\underline{y}} = \begin{pmatrix} dy_1/dt \\ dy_2/dt \\ \vdots \\ dy_K/dt \end{pmatrix}$$ = a $K \times 1$ vector of derivatives of $\underline{y}$ with respect to time

$$\underline{A} = \begin{pmatrix} a_{10} & a_{11} \cdots a_{1L} \\ a_{20} & a_{21} \cdots a_{2L} \\ \vdots & \vdots \quad\quad \vdots \\ a_{K0} & a_{K1} \cdots a_{KL} \end{pmatrix}$$ = K x (L + 1) matrix of intercepts (a_{10}) and effect coefficients for exogeneous variables (a_{ij}, $j>0$). These coefficients are constant over time

$$\underline{x} = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_L \end{pmatrix}$$ = an (L + 1) x 1 vector, with the first element being the constant 1.0 and other elements being the L exogeneous variables

$$\underline{B} = \begin{pmatrix} b_{11} & b_{12} \cdots b_{1K} \\ b_{21} & b_{22} \cdots b_{2K} \\ \vdots & \vdots \quad\quad \vdots \\ b_{K1} & b_{K2} \cdots b_{KK} \end{pmatrix}$$ = a K x K matrix of effect coefficients indicating the impact of the endogenous variables on each other. These coefficients are constant over time.

$$\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{pmatrix}$$ = a K x 1 vector containing endogenous variables

$$\underline{u}(t) = \begin{pmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_K(t) \end{pmatrix}$$ = a K x 1 vector of disturbance functions of time

Note that $\underline{Ax} + \underline{u}(t)$ is a K x 1 vector which is a function of time; hence, one may define $a(t) = Ax + u(t)$ and apply the mathematical theory of the preceding discussion. The result of this application is an equation in which time t values of y are shown as a linear combination of time-zero values of y and of the predetermined variables, plus "disturbance" or error variables (u^*), as follows:

(29) $\underline{y}_t = \underline{A}^*\underline{x} + \underline{B}^*\underline{y}_o + \underline{u}^*$

where, by use of equation (27a)

(30a) $\underline{B}^* = e^{\underline{B}t}$, and

(30b) $\underline{B} = (\ln \underline{B}^*)/t \ , \ |\underline{B}^*| \neq 0$

To find $\underline{A}^*$ and $\underline{u}^*$, one may apply equation (27b). For this application, the definite integral on the right of (27b) can be separated into two additive parts:

$$\int_0^t e^{-\underline{B}\tau}(\underline{Ax} + \underline{u}(\tau))d\tau = \int_0^t e^{-\underline{B}} \underline{Ax}d\tau + \int_0^t e^{-\underline{B}\tau}\underline{u}(\tau)\, d\tau$$

The first term on the right of this equation has an explicit algebraic solution which depends on the rank of the matrix $\underline{B}$. If $\underline{B}$ is full rank ($|\underline{B}| \neq 0$), the following solution holds:

$$\int_0^t e^{-\underline{B}\tau}\underline{Ax}d\tau = (\underline{I} - e^{-\underline{B}t})\underline{B}^{-1}\underline{Ax}, \quad |\underline{B}| \neq 0$$

Inserting this result in (27b) and solving for $\underline{A}^*$ yields

$$\text{(31a)} \quad \underline{a}^* = (e^{\underline{B}t} - \underline{I})\underline{B}^{-1}\underline{Ax} + e^{\underline{B}t}\int_0^t e^{-\underline{B}\tau}\underline{u}(\tau)d\tau, \quad |\underline{B}| \neq 0$$

If the coefficient matrix B is not full rank ($|\underline{B}| = 0$), some additional notation is needed to define the solution. Let $\underline{\Lambda}$ be the $K \times K$ diagonal matrix with diagonal elements equal to the eigenvalues of $\underline{B}$. If $|\underline{B}| = 0$, at least one diagonal element of $\underline{\Lambda}$ is zero. Let $\underline{V}$ be the matrix whose columns are the right eigenvectors of $\underline{B}$ associated with $\underline{\Lambda}$, so that, if $\underline{B}$ is diagonalizable, $\underline{B} = \underline{V\Lambda V}^{-1}$. Assume there are K_1 nonzero eigenvalues (λs) in $\underline{\Lambda}$ and K_2 zero λs, $K_1 + K_2 = K$, and let $\underline{\Lambda}_1$ be the $K_1 \times K_1$ diagonal matrix with nonzero λ_i, $i = 1, \ldots, K_1$, in the diagonal. Define $\underline{\Lambda}_2$ to be the $K_2 \times K_2$ null matrix containing the zero λs. Also, let $\underline{V}_1$ be the K_1 right eigenvectors associated with $\underline{\Lambda}_1$, and $\underline{V}_2$ the $\underline{K}_2$ right eigenvectors associated with $\underline{\Lambda}_2$. Similarly, let $\underline{V}^{-1}$ be partitioned into $\underline{V}^{(1)}$, a $K_1 \times K$ matrix of transposed left eigenvectors of B associated with $\underline{\Lambda}_1$, and $\underline{V}^{(2)}$, a $K_2 \times K$ matrix of transposed left eigenvectors of $\underline{B}$ associated with the zero λs; $\underline{\Lambda}_2$, $\underline{\Lambda}$, $\underline{V}$ and $\underline{V}^{-1}$ are, thus, partitioned as follows.

$$\underline{\Lambda} = \begin{pmatrix} \underset{(K_1 \times K_1)}{\underline{\Lambda}_1} & \underline{0} \\ \underline{0} & \underset{K_2 \times K_2}{\underline{\Lambda}_2} \end{pmatrix} = \begin{pmatrix} \underline{\Lambda}_1 & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix}$$

$$\underline{V} = \left(\underset{(K\times K_1)}{\underline{V}_1} \quad , \quad \underset{(K\times K_2)}{\underline{V}_2} \right)$$

$$\underline{V}^{-1} = \begin{pmatrix} \underset{(K_1\times K)}{\underline{V}^{(1)}} \\ \\ \underset{(K_2\times K)}{\underline{V}^{(2)}} \end{pmatrix} ,$$

Thus,

$$\underline{B} = \underline{V\Lambda V}^{-1} = (\underline{V}_1, \underline{V}_2) \begin{pmatrix} \underline{\Lambda}_1 & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix} \begin{pmatrix} \underline{V}^{(1)} \\ \underline{V}^{(2)} \end{pmatrix} = \underline{V}_1\underline{\Lambda}_1\underline{V}^{(1)}$$

and

$$e^{\underline{B}t} = \underline{V}e^{\underline{\Lambda}t}\underline{V}^{-1} = (\underline{V}_1, \underline{V}_2) \begin{pmatrix} e^{\underline{\Lambda}_1 t} & \underline{0} \\ \underline{0} & \underline{I} \end{pmatrix} \begin{pmatrix} \underline{V}^{(1)} \\ \underline{V}^{(2)} \end{pmatrix}$$

$$= \underline{V}_1 \, e^{\underline{\Lambda}_1 t} \, \underline{V}^{(1)} + \underline{V}_2\underline{V}^{(2)}$$

Using these partitions of $\underline{\Lambda}$, $\underline{V}$ and $\underline{V}^{-1}$, the desired definite integral for the case when $\underline{B}$ is less than full rank and diagonalizable can be written

$$\int_0^t e^{-\underline{B}\tau}\underline{Ax}d\tau = [(\underline{I} - e^{-\underline{B}t})\underline{V}_1\underline{\Lambda}^{-1}\underline{V}^{(1)} + t\,\underline{V}_2\underline{V}^{(2)}]\underline{Ax}$$

Substituting this result into the right side of (27b) and solving for $\underline{a}^*$ yields

$$\underline{a}^* = e^{\underline{B}t}[(\underline{I} - e^{-\underline{B}t})\underline{V}_1\underline{\Lambda}_1{}^{-1}\underline{V}^{(1)} + t\, e^{\underline{B}t}\underline{V}_2\underline{V}^{(2)}]\underline{Ax}$$
$$+ e^{\underline{B}t}\int_0^t e^{-\underline{B}\tau}\underline{u}(\tau)d\tau$$

which, after simplification, can be written

$$(31b)\quad \underline{a}^* = [(e^{\underline{B}t} - \underline{I})\underline{V}_1\underline{\Lambda}_1^{-1}\underline{V}^{(1)} + t\cdot\underline{V}_2\underline{V}^{(2)}]\underline{Ax} + e^{\underline{B}t}\int_0^t e^{-\underline{B}\tau}\underline{u}(\tau)d\tau,\ |\underline{B}| = 0$$

since it can be shown that $e^{\underline{B}t}\underline{V}_2\underline{V}^{(2)} = \underline{V}_2\underline{V}^{(2)}$.

The unsolved integral in (31a) and (31b) has no solution unless the function u(t) is defined. Fortunately, the function can be left unspecified and treated as part of the disturbance term in the statistical analysis:

$$\underline{u}^* = e^{\underline{B}t}\int_0^t e^{-\underline{B}\tau}\underline{u}(\tau)d\tau$$

where u* is the notation for the disturbance variable (see equation [29]).

In equation (31a), the term $(e^{\underline{B}t} - \underline{I})\underline{B}^{-1}\underline{A}$ is a K x (L + 1) matrix. Its elements can be considered coefficients of the exogenous variables. In equation (31b), the matrix $[(e^{\underline{B}t} - \underline{I})\underline{V}_1\underline{\Lambda V}^{(1)} + t\cdot\underline{V}_2\underline{V}^{(2)}]\underline{A}$ is also a K x (L + 1) matrix of coefficients of the exogenous variables. In both cases, the coefficient may be denoted by $\underline{A}^*$ and used in the integral equation (29). Thus, one has

$$(31c)\quad \underline{A}^* = (e^{\underline{B}t} - \underline{I})\underline{B}^{-1}\underline{A},\ \text{if } |\underline{B}| \neq 0.$$

$$(31d)\quad \underline{A} = \underline{B}(e^{\underline{B}t} - \underline{I})^{-1}\underline{A}^*$$

$$(31e)\quad \underline{A}^* = [(e^{\underline{B}t} - \underline{I})\underline{V}_1\underline{\Lambda}^{-1}\underline{V}^{(1)} + t\cdot\underline{V}_2\underline{V}^{(2)}]\underline{A},\ \text{if } |\underline{B}| = 0$$

$$(31f)\quad \underline{A} = [(e^{\underline{B}t} - \underline{I})\underline{V}_1\underline{\Lambda}_1^{-1}\underline{V}^{(1)} + t\cdot\underline{V}_2\underline{V}^{(2)}]^{-1}\underline{A}^*$$

It may be helpful to summarize these procedures. The steps in the analysis are listed and described below.

Step 1. Form a hypothesis about continuous change over time by writing a nonhomogeneous system of linear differential equations with constant coefficients, written as follows:

$$(32)\quad \dot{\underline{y}} = \underline{Ax} + \underline{By} + \underline{u}(t)$$

where

$\dot{\underline{y}}$ = a K × 1 vector of derivatives of y_i with respect to time

$\underline{x}$ = an (L + 1) × 1 vector of exogenous variables with the first element set to the constant value 1.0

$\underline{y}$ = a K × 1 vector of endogenous variables

$\underline{u}(t)$ = a K × 1 vector function of time, considered to be a residual disturbance vector

$\underline{A}$ = a K × (L + 1) matrix of constant coefficients associated with the exogenous variables, the first column representing intercepts

$\underline{B}$ = a K × K matrix of constant coefficients indexing effects of endogenous variables on changes in endogenous variables

Step 2. Integrate the differential equation (32) to form a prediction equation:

$$(33) \quad \underline{y}_t = \underline{A}^*\underline{x} + \underline{B}^*\underline{y}_o + \underline{u}^*$$

where the subscript attached to $\underline{y}$ designates a specific point in time, and $\underline{A}^*$ and $\underline{B}^*$ are K × (L + 1) and K × K coefficient matrices, respectively.

Step 3. Estimate the coefficient martices A* and B* by some type of regression methodology applied to longitudinal data. The longitudinal data must contain at least one observation per case for each exogenous variable x and two sets of observations per case for each endogenous variable -- one set of observations at t_o and one at t_1. Thus a two-panel design is sufficient for empirical estimates of the coefficients. Equation (33) follows the form of simultaneous structural equations as given in the econometric literature. If one assumes $\underline{x}$ and $\underline{y}_o$ are uncorrelated with $\underline{u}^*$, then "ordinary least squares" regression can be applied. Otherwise, some different regression method should be used. Statistical methods will be discussed in the next subsection of this chapter.

Step 4. Calculate estimates of the coefficients of the differential equation system (32) by using the following formulae:

$$(34)\quad \underline{B} = (\ln \underline{B}^*)/t$$

$$(35a)\quad \underline{A} = \underline{B}(\underline{B}^* - \underline{I})^{-1}\underline{A}^*, \text{ if } |\underline{B}| \neq 0$$

$$(35b)\quad \underline{A} = [(e^{\underline{B}t} - \underline{I})\underline{V}_1\underline{\Lambda}_1^{-1}\underline{V}^{(1)} + t\cdot\underline{V}_2\underline{V}^{(2)}]^{-1}\underline{A}^*, \text{ if } |\underline{B}| = 0$$

where ln $\underline{B}^*$ is a matrix logarithm defined by equation (20).

Once the matrices A and B have been estimated, they can be applied to generate predictions to any point along a continuous time scale. These predictions can be compared to observations as a test of the theoretical model.

Statistical Estimation

This subsection briefly summarizes statistical theory related to estimating coefficients of the integral equation (33) by regression methods. Unlike the mathematics of differential equations, the relevant statistical theory is familiar to many social scientists via the econometric and, increasingly, the sociological and political science literatures (see, e.g., Johnston, 1963; Goldberger, 1964; Duncan, 1965; Heise, 1975; Goldberger and Duncan, 1973; Ostrom, 1978; Asher, 1976; Boardman and Murnane, 1979). In consequence, the treatment of statistical theory in this volume need not be as detailed as the treatment of mathematical theory. In particular, this subsection develops the assumptions underlying application of ordinary least squares (OLS) to two panels of data in order to secure estimates of the coefficient matrices A* and B* in equation (33). Readers interested in differential equation systems for general applications should be aware that numerous alternative estimation techniques are available; these will be summarized briefly toward the end of this subsection. The present focus on OLS applied to two panels of data reflects likely needs in status attainment research for the immediate future. First, time series of data with numerous time points seldom are available in career-decision making research. Secondly, application of differential equation models in career-decision making research is new, a variety of statistical methods may be tried out as experience dictates.

Equation (33) may be interpreted as a special case of a simultaneous equation system, and one may consult the econometric literature for investigations of appropriate statistical analysis. Before addressing the specific task of statistical estimation of equation (33), it is useful to review the broad outlines of econometric theory of simultaneous linear systems. In this theory, sets of linear equations are the object of study; each equation expresses a hypothesis about the substance of one's topic. Consider the following pair of structural equations as an example.

over samples. The variance of the statistic, as with all variances, is just the square of the standard deviation (standard error). Certainly, the smaller the variance of a statistic the better, ceteris paribus. If one sample statistic has a smaller variance than a second statistic estimating the same population parameter, the former is termed more efficient than the latter. As an example, part of the central limit theorem states that the expected value of the sample mean is the population mean, and that the variance of the sample mean of x is the population variance of x, divided by the sample size. Similarly, the mean over samples of a sample regression coefficient (in the fixed-effects model) is the corresponding population regression coefficient, and the variance of the sample regression coefficient is a function of the standard error of estimate and a diagonal element of the inverse of the matrix of cross products among the independent variables.

Two concepts of major importance in structural-equation analysis are bias and consistency. Let s by any sample statistic, such as the mean or sample regression coefficient, and let the corresponding population parameter be denoted by S. The expected value of s is denoted by E(s). The bias of s is the difference between its expected value and the value of the corresponding population parameter:

$$\text{Bias}(s) = E(s) - S$$

If $E(s) = S$, s is termed unbiased. If $E(s) \neq S$ but Es approaches S as the sample size goes to infinity, s is described as asymptotically unbiased. If the variance of a statistic goes to zero as the sample size goes to infinity, then the statistic is termed consistent. A consistent statistic is asymptotically unbiased, but asymptotic unbiasedness obviously does not assure consistency. Unbiasedness is a desirable characteristic of a statistic, but frequently, it is necessary to accept a sample statistic that is consistent; this is the case, for example, with many econometric methods such as two-stage least squares, limited-information, maximum-likelihood estimation, and so forth.

(36a) $y_1 = p_{10} + p_{11}z + q_{12}y_2 + v_1$

(36b) $y_2 = p_{20} + p_{21}z + q_{21}y_1 + v_2$

where y_1, y_2, z are observed variables, p_{ij} and q_{ij} are constants, and v_1, v_2 are unobserved disturbance variables. In (36a) y_1 is hypothesized to be affected by z and y_2, and in (36b) y_2 is considered an effect of z and of y_1. Thus, y_1 and y_2 exhibit feedback effects on each other, but z is not affected by any variable in the system. Variables such as z are termed exogenous variables; variables such as y_1 and y_2 are endogenous. More generally, endogenous variables are variables whose values are determined, in part, by other variables included in the system; whereas, exogenous variables are those whose values are determined outside the system, i.e., the theory does not account for their values.

For statistical applications, the most important distinction between exogenous and endogenous variables is that the exogenous variables are assumed to be uncorrelated with the disturbance variables; whereas, the endogenous variables are permitted to be correlated with the disturbance variables. It is argued that if y_1 affects y_2, then the disturbance for y_1 is likely to be correlated with y_2, and vice versa. This is not a deductive argument, however; rather it is based on common sense. This point does not seem to be understood clearly in the literature. Often, readers might acquire the impression from published accounts that there is a deductive argument demonstrating conclusively that if reciprocal causation occurs between y_1 and y_2, then a correlation necessarily must arise between y_1 and the disturbance associated with y_2, and analogously, that y_2 necessarily must be correlated with the disturbance on y_1. Neither of these correlations follow deductively from the assumption of reciprocal causation. What does follow deductively is that, once nonzero correlations are assumed between independent (measured) variables and disturbance variables, OLS yields inconsistent estimators of the effect parameters in the structural equations.

Equations (36a) and (36b) each appear as a linear equation with a stochastic error term (disturbance variable); hence, one might contemplate using ordinary least squares regression to estimate the parameters -- p_{ij}, q_{ij}. The formula for OLS estimation of equation (36a) is

(37) $[\hat{p}_{10}, \hat{p}_{11}, \hat{q}_{12}] =$

$$[\sum_j y_{1j}, \sum_j z_j y_{1j}, \sum_j y_2 y_{1j}] \begin{bmatrix} N & , \sum_j z_j & , \sum_j y_{2j} \\ \sum_j z_j & , \sum_j z_j^2 & , \sum_j z_j y_{2j} \\ \sum_j y_{2j}, & \sum_j y_{2j} z_j & , \sum_j y_{2j}^2 \end{bmatrix}^{-1}$$

where $\hat{p}_{10}$, $\hat{p}_{11}$, and $\hat{q}_{12}$ are sample estimates of the corresponding population values p_{10}, p_{11}, and q_{12}, respectively, the j subscript stands for the jth case, and N = the number of cases.

To simplify the presentation, it is convenient to shift to a matrix notation.

Accordingly, let

$\underline{y}$ = an $1 \times N$ vector of observations on y_1

$\underline{W}$ = a $3 \times N$ matrix with the first row containing N constant values of 1.0, the second row containing N observations on $\underline{z}$, and the third row containing N observations on y_2. Each column of $\underline{W}$ represents a single case

$\underline{v}$ = an $1 \times N$ vector of values for v_1

$\underline{r}$ = a 3×1 vector of coefficients;

$\underline{r} = (p_{10}, p_{11}, q_{12})'$

where the prime stands for transpose.

Equation (37) can now be written

(37a) $\hat{\underline{r}}' = \underline{y}\underline{W}'(\underline{WW}')^{-1}$

where $\hat{\underline{r}}$ is the OLS estimate of $\underline{r}$. Note that (37a) is the transpose of most notations for simultaneous linear systems; the transposed notation is used to maintain consistency with the notation for the differential equation system. The differential equations are presented in notation that is prevalent in the mathematical literature.

To evaluate this estimate, write (36a) in the matrix notation for N observations --

(38) $\underline{y} = \underline{r}'\underline{W} + \underline{v}$

and postmultiply by the transpose of the independent variable matrix ($\underline{W}'$) --

(39a) $\underline{yW}' = \underline{r}'\underline{WW}' + \underline{vW}'$

(39b) $\underline{r}' = \underline{yW}'(\underline{WW}')^{-1} - \underline{vW}'(\underline{WW}')^{-1}$

assuming $\underline{WW}'$ is nonsingular. For a given sample, subtracting (39b) from (37a) gives the sampling error:

(40) $\hat{\underline{r}}' - \underline{r}' = \underline{vW}'(\underline{WW}')^{-1}$

Thus, $\tilde{r}$ is unbiased only if $E\underline{vW}'(\underline{WW}')^{-1}$ is zero. In general this expected value is not zero. Even worse, it does not tend to zero as the sample size increases. Its probability limit is:

$$\text{plim}[\underline{vW}'(\underline{WW}')^{-1}] = \text{plim}[\underline{vW}'][\text{plim}(\underline{WW}')^{-1}] =$$
$$E(v_1\underline{w}')[E(\underline{ww}')]^{-1} \neq 0$$

where plim stands for probability limit, and $\underline{w}$ is a column from $\underline{W}$. Written out this becomes

$$E(v_1\underline{w}')[E(\underline{ww}')]^{-1} = [Ev_1,\ Ev_1z,\ Ev_1y_2]\begin{bmatrix} 1 & Ez & Ey_2 \\ Ez & Ez^2 & Ezy_2 \\ Ey_2 & Ey_2z & Ey_2^2 \end{bmatrix}$$

Since, by assumption Ev_1, Ev_1z are zero, the difficulty arises from the nonzero expected value, Ev_1y_2. As noted, this nonzero assumption is made because y_2 is hypothesized to be affected by y_1, and it therefore seems unreasonable to include in the assumptions the postulate that the disturbance for y_1 is uncorrelated with y_2. (Note that Ev_1y_2 is the numerator to the correlation between v_1 and y_2, since $Ev_1 = 0$.)

This argument was developed for equation (36a). An analogous argument applies to (36b). Clearly, the same argument is quite general, applying to every equation in systems of any dimentions. Thus, it is concluded that OLS estimates are inconsistent (i.e., do not converge on the population parameters they purport to estimate as N goes to infinity), when one or more of the regressors is correlated with the disturbance variable in the associated structural equation.

Consistent estimation of the structural parameters, p_{ij} and q_{ij} is approached by way of a <u>reduced form</u> set of coefficients. To explore methods based on the reduced form, it is instructive to rewrite equations (36) in the following way.

$$0 = p_{10} + p_{11}z - y_1 + q_{12}y_2 + v_1$$

$$0 = p_{20} + p_{21}z + q_{21}y_1 - y_2 + v_2$$

This pair of equations can be compactly written in matrix notation for L exogenous variables, K endogenous variables, and N observations:

(41) $\underline{O} = \underline{PZ} + \underline{QY} + \underline{V}$

where

$\underset{(L+1)\times N}{Z}$ = a matrix whose first row contains all ones and the remaining rows contain N observations on each of the L exogenous variables

$\underset{K\times N}{Y}$ = a matrix whose rows contain N observations on each of K endogenous variables

$\underset{K\times N}{V}$ = a matrix of K disturbance variables for N cases

$\underset{K\times (L+1)}{P}$ = a matrix of constant coefficients including the intercepts for each equation in the system and the parameters indexing the effects of each of the L exogenous variables on each of the K endogenous variables

$\underset{K\times K}{Q}$ = a nonsingular matrix of coefficients indexing effects of the endogenous variables on each other. The diagonal elements of $\underline{Q}$ can be defined to equal -1.0 to maintain consistency between (41) and (36).

$\underline{O}$ = a conformable null matrix.

Again, note that, to maintain consistency with the previous subsection, equation (41) is the transpose of notation most prevalent in the econometric literature.

The reduced form of the system is obtained by premultiplying (41) by $\underline{Q}^{-1}$ (assumed to exist), and solving for Y.

$$\underline{O} = (\underline{Q}^{-1}\underline{P})\underline{Z} + \underline{Y} + \underline{Q}^{-1}\underline{V}$$

$$\underline{Y} = -(\underline{Q}^{-1}\underline{P})\underline{Z} - \underline{Q}^{-1}\underline{V}$$

$$(42) \quad \underline{Y} = \underline{\Pi Z} + \underline{V^*}$$

where

$$(43) \quad \underline{\Pi} = -\underline{Q}^{-1}\underline{P} = \text{the reduced-form coefficients matrix}$$

$$(44) \quad \underline{V^*} = -\underline{Q}^{-1}\underline{V}$$

Since $\underline{Z}$ is uncorrelated asymptotically with $\underline{V}$, $(1/N)\text{plim } \underline{V^*Z'} = -1/N \text{ plim } Q^{-1}\underline{VZ'} = (-1/N)Q^{-1} \text{ plim } \underline{VZ'} = O$. Hence OLS can be used to estimate consistently $\underline{\Pi}$ in equation (42). The estimate is $\hat{\underline{\Pi}} = \underline{YZ'}\ (\underline{ZZ'})^{-1}$.

Estimation of the structural parameters $\underline{P}$ and $\underline{Q}$ may now proceed by making use of (43). Premultiplying by $-\underline{Q}$ yields

$$(45) \quad -\underline{Q\Pi} = \underline{P}$$

Given a sample estimate of the reduced form matrix, $\hat{\underline{\Pi}}$, one may write the sample version of (45), and investigate how to use the relation for finding $\hat{\underline{P}}$ and $\hat{\underline{Q}}$;

$$(45a) \quad -\hat{\underline{Q}}\hat{\underline{\Pi}} = \hat{\underline{P}}$$

In general, (45a) is not identified. There are K(L + 1) known values in $\hat{\underline{\Pi}}$, but K(K - 1) unknowns (excluding diagonal elements, set to -1) in $\hat{\underline{Q}}$ plus K(L + 1) unknowns in $\hat{\underline{P}}$. Obviously, one must draw on substantive theory to constrain the values of $\underline{P}$ and/or Q. Generally, certain elements of P and Q are assumed zero, and the altered relation implied by (45a) and the zero coefficients is examined to see if the remaining unknowns can be calculated. Numerous methods are available for such calculations (see Goldberger, 1964 and Johnston, 1963 for thorough reviews).

The immediate goal is to find suitable estimates of the matrices A^* and B^* in the integral equation (33). This estimation problem may be interpreted as one of estimating the reduced form of a general simultaneous structural equation system. To see this, it is necessary to introduce some additional terminology. The variables of a simultaneous system frequently are classified into one group termed predetermined variables and a second group called jointly dependent variables. Predetermined variables include exogenous variables and lagged values of the endogenous variables, and the jointly dependent variables include only current values of the endogenous variables. For instance, in the integral equation system (33), $\underline{x}$ represents exogenous variables and the lagged endogenous variables are denoted by $\underline{y}_o$; the predetermined variables include $\underline{x}$ and $\underline{y}_o$. The jointly dependent variables are denoted by $\underline{y}_t$. In statistical applications, the matrix Z contains predetermined variables, and $\underline{Y}$ contains jointly dependent variables.

To pursue the argument, shift equation (33) into the notation of structural equations developed above. Let $\underline{y}_{tj}$ represent a $K \times 1$ vector of values on K endogenous variables at time t for case j, and let $\underline{x}_j$ be a $(1 + L) \times 1$ vector with its first element equal to 1.0 and remaining elements containing values of L exogenous variables for a single case. Also, let $\underline{v}_j$ represent a column vector of K disturbance variables for a single case. Now, define

$$\underset{(1+L+K)\times N}{Z} = \begin{pmatrix} \underline{x}_1 & \cdots & \underline{x}_N \\ \underline{y}_{01} & \cdots & \underline{y}_{0N} \end{pmatrix}$$

$$\underset{K\times N}{\underline{Y}} = \begin{pmatrix} \underline{y}_{t1} & \cdots & \underline{y}_{tN} \end{pmatrix}$$

$$\underset{K\times N}{V} = \begin{pmatrix} \underline{v}_1 & \cdots & \underline{v}_N \end{pmatrix}$$

$$\underset{K\times(1+L+K)}{P} = \begin{pmatrix} \underline{A}^* & \underline{B}^* \end{pmatrix}$$

$$\underset{K\times K}{Q} = \underset{K\times K}{-I} = \text{The negative of a } K \times K \text{ identity matrix}$$

Now the integral equation (33) can be written

$$(46)\quad \underline{Y} = \underline{PZ} + \underline{V}$$

$$\underline{O} = \underline{PZ} - \underline{Y} + \underline{V} = \underline{PZ} + (-\underline{I})\underline{Y} + \underline{V}$$

$$(46a)\quad \underline{O} = \underline{PZ} + \underline{QY} + \underline{V}$$

Equation (46a) is precisely the same form as the general structural system given in (41), with $\underline{Q} \equiv -\underline{I}$. Estimation via reduced form may proceed directly by reference to equation (45). Clearly, when $\underline{Q} = -\underline{I}$, (45) yields:

$$(47)\quad \underline{P} = \underline{\Pi}$$

Since $\underline{P}$ contains the structural coefficients for the predetermined variables and Π is the matrix of reduced-form coefficients, (47) establishes the desired result: Estimation of the parameters of the integral equation (33) may be viewed as estimation of the reduced form of a general simultaneous structural-equation model. (See Goldberger, 1964: 373ff for a similar interpretation.) Thus one may consistently estimate A* and B* by the following formulas.

$$[\underline{\hat{A}}^{*}, \underline{\hat{B}}^{*}] = \underline{\hat{P}}$$

$$(48)\quad \underline{\hat{P}} = \underline{YZ}'(\underline{ZZ}')^{-1}$$

The asymptotic variance-covariance matrix for each row of P (equation of [33]) of these estimates is given by

$$(49)\quad \underline{C}_{\hat{p}_i \hat{p}_i} = 1/N\ s_{v_i}^{2}\ [E(\underline{zz}')]^{-1}$$

where

$\underline{C}_{\hat{p}_i \hat{p}_i}$ = the covariance matrix for the sample estimates in the i'th row of $\underline{\hat{P}}$ -- $\hat{\underline{p}}_i$

$s_{v_i}^{2}$ = the variance of the disturbance for equation i

$E(\underline{zz}')$ = the covariance matrix among the predetermined variables.

Generally, the quantities in (49) being population values, are unknown. However, consistent estimates can be formed by calculation of analogous sample values:

$$(49a)\quad \hat{\underline{C}}_{\hat{p}_i \hat{p}_i} = (1/N)\hat{s}_{v_i}^{2}\ [(1/N)(\underline{ZZ}')]^{-1}$$

$$= \hat{s}_{v_i}^{2}\ (\underline{ZZ}')^{-1}$$

89

where the circumflex indicates a sample value for the corresponding population statistic. Note that $s_{v_i}^2 = (1/N)\sum_{j=1}^{N} v_{ij}^2 = \hat{s}_{y_i}^2 (1 - \hat{R}_i^2)$, where $\hat{s}_{y_i}^2$ is the sample variance of the ith jointly dependent variable, and $\hat{R}_i^2$ is the square of the multiple correlation for equation i. A slightly improved estimate is given by replacing the reciprocal 1/N with 1/(N-K-1), where K is the number of regressors.

As noted in the beginning of this subsection, there are numerous methods for estimating A* and B* other than application of OLS to two panels of data. Consequently, it is important to identify the key assumptions underlying this application, and to summarize briefly alternative methods. The first two assumptions are substantive rather than statistical. They are:

Assumption 1: All individuals are governed by the same differential-equation structure. This means that the matrices A and B (and, therefore, A* and B*) are the same for all persons.

Assumption 2: The differential-equation structure is stationary over time. This means that all values in the coefficient matrices A and B are fixed over time.

It is possible, though improbable, that statistical assumptions, to be reviewed presently, could hold even though assumption 1 and/or assumption 2 are not valid. In such a case, one would produce good statistical estimates of parameters that provide seriously incomplete description of the substantive topic.

The next two assumptions are necessary (and sufficient) for the statistical estimates to be consistent.

Assumption 3: The means of all disturbance variables v are zero.

Assumption 4: The independent variables z are not correlated with the disturbance variables v. This means that all exogenous variables x and lagged endogenous variables y_o are not correlated with any unmeasured variables v.

If assumption 3 and assumption 4 hold, OLS estimates are consistent, meaning that: (a) they are unbiased in the limit as N approaches infinity (i.e., asymptotically unbiased), and (b) their sampling variance approaches zero as N approaches infinity. (The latter feature implies the former.)

The next assumption is required for derivation of the variance covariance matrix among the regression coefficients of a given equation (equation [49]).

> Assumption 5: The variance over samples of each disturbance term is the same for all observations and it does not depend on z.

If assumption 5 is not met it does not imply necessarily that OLS is inconsistent. Rather, it means that statistical tests of significance should not be applied using formula (49a) to estimate standard errors. In this case, however, generalized linear regression (GLR) may be preferable to OLS, because, while both OLS and GLR are consistent, the sampling variability of GLR is less than for OLS (i.e., GLR is more efficient, see Goldberger, 1964).

Any one of these assumptions poses a serious threat to the valid use of OLS to estimate A* and B*, because status attainment theory is not sufficiently developed to generate confidence that any of the assumptions are met. It should be emphasized, however, that no statistical method will compensate for incomplete theory. Thus, one should not expect to find the situation much improved by resort to alternative statistical procedures.

In the statistical literature on simultaneous structural equations, assumption 4 has been of major interest. As reviewed above, when assumption 4 fails (i.e., regressors are correlated with the disturbance), OLS is biased and inconsistent. If OLS estimates are inconsistent, then numerous alternative estimation methods might be substituted; examples of alternative methods include instrumental variable estimation (IVE), indirect least squares (ILS), two-stage least squares (2-SLS), full-information maximum likelihood (FIML) estimation, and three-stage least squares (3-SLS). All of these methods depend on measuring a set of exogenous variables (or instrumental variables) that have no direct effect on the jointly dependent variables but are uncorrelated with the disturbance variables.

The bias and inconsistency of OLS when it cannot be assumed that regressors are uncorrelated with the disturbance is well known, but a similar result for alternatives to OLS has not been widely publicized. To emphasize the point that statistical methods cannot substitute for mature theory, the case of instrumental variable estimation (IVE) is examined below. Not surprisingly, it is found that IVE estimates are biased and inconsistent when it is assumed erroneously that direct effects of the instruments are zero. It should be emphasized that IVE is equivalent to indirect least squares, and for "just identified" systems, also equivalent to other methods such as two-stage least squares.

Instrumental variables for a given structural equation are variables that are uncorrelated with the disturbances for that equation and are correlated with the independent variables in the equation. Suppose that $\underline{W}$ is an $(L + 1) \times N$ matrix with the first row containing all ones and the remaining rows containing N observations on L instrumental variables. To avoid defining additional notation, assume that $\underline{W}$ is uncorrelated with all disturbance variables; in this case $\overline{EVW'} = 0$. Referring to equation (41) and using the fact that $\underline{Q} = -\underline{I}$, one may write

$$\underline{O} = \underline{PZ} - \underline{Y} + \underline{V}$$

$$\underline{Y} = \underline{PZ} + \underline{V}$$

The instrumental variable estimation of $\underline{P}$ is $\hat{\underline{P}} = \underline{YW}'(\underline{ZW}')^{-1}$; by the assumption $(\underline{ZW}')^{-1}$ exists. The sampling error for $\hat{\underline{P}}$ is

$$\begin{aligned}\hat{\underline{P}} - \underline{P} &= \underline{YW}'(\underline{ZW}')^{-1} - \underline{P} = (\underline{PZ} + \underline{V})\underline{W}'(\underline{ZW}')^{-1} - \underline{P} \\ &= \underline{P} + \underline{VW}'(\underline{ZW}')^{-1} - \underline{P} \\ &= \underline{VW}'(\underline{ZW}')^{-1}\end{aligned}$$

Taking probability limits, one finds

$$\begin{aligned}\operatorname{plim}(\hat{\underline{P}}-\underline{P}) &= \operatorname{plim} \underline{VW}'(\underline{ZW}')^{-1} = \operatorname{plim} \underline{VW}' \operatorname{plim}(\underline{ZW}')^{-1} \\ &= E\underline{vw}'(E\underline{zw}')^{-1} \\ &= \underline{O}\end{aligned}$$

where $\underline{z}$, $\underline{w}$ are columns from $\underline{Z}$ and $\underline{W}$, respectively.

Hence, under the stated assumptions, instrumental variable estimation (and indirect least squares) is consistent, and, as shown earlier, OLS is not. This is a basic result in the econometric literature.

An implicit assumption of the IVE is that the instrumental variables exercise no direct effect on the dependent variables. To see this, suppose that some or all the instruments do exercise some effect on the dependent variables and let $\underline{S}$ be a conformable matrix containing the effect parameters. The observations now are generated by the following equation.

$$\underline{Y} = \underline{PZ} + \underline{SW} + \underline{V}$$

Suppose that one assumes mistakenly that $\underline{S} = \underline{O}$ and forms IVE estimates, $\hat{\underline{P}} = \underline{YW}'(\underline{ZW}')^{-1}$. The sampling error for $\hat{\underline{P}}$ now becomes

$$\begin{aligned}\hat{\underline{P}} - \underline{P} &= \underline{YW}'(\underline{ZW}')^{-1} - \underline{P} \\ &= (\underline{PZ} + \underline{SW} + \underline{V})\underline{W}'(\underline{ZW}')^{-1} - \underline{P} \\ &= \underline{P} + \underline{S}(\underline{WW}')(\underline{ZW}')^{-1} + (\underline{VW}')(\underline{ZW}')^{-1} - \underline{P} \\ &= \underline{S}(\underline{WW}')(\underline{ZW}')^{-1} + (\underline{VW}')(\underline{ZW}')^{-1}\end{aligned}$$

taking probability limits gives

$$\text{plim } (\hat{\underline{P}} - \underline{P}) = \underline{S}E(\underline{ww}')(E\underline{zw}')^{-1} \neq \underline{O}$$

Thus, it is clear that instrumental variable estimation generates inconsistent estimation of $\underline{P}$ when it is assumed incorrectly that the instruments have no direct effect on the dependent variable (i.e., incorrectly assume $\underline{S} = O$).

This is hardly a surprising result, but it does emphasize the point that methodological techniques cannot substitute for adequate theory. Whatever method is used, assumptions that cannot be tested against data must be made. Confidence in the result of the calculations must necessarily rest heavily on confidence in the theory.

The above discussion introduces the general question of identification. Before any statistical method can be applied, theoretical assumptions must be imposed. A thorough presentation of the identification problem is beyond the scope of this monograph, but a brief treatment is appropriate.

Assume that y represents one case on one of the dependent variables (a scalar). Also, let $\underline{p}$ represent one row of P corresponding to y, and let v be the corresponding scalar disturbance. Let $\underline{z}$ be one column from the matrix $\underline{Z}$. With this notation, the model can be written:

$$y = \underline{pz} + v$$

postmultiplying by $\underline{z}'$ and taking expectations on both sides, one may derive estimating equations.

$$Ey\underline{z}' = \underline{p}E\underline{zz}' + Ev\underline{z}'$$

or

$$E\underline{z}y = (E\underline{zz}')\underline{p}' + E\underline{z}v$$

The second form expresses the system of estimating equations in conventional column-vector form. In the system, p' and Ezv are unknown, and Ezy and Ezz' are assumed known. There are L + 1 rows, hence L + 1 equations. There are L + 1 unknown elements in p' and L + 1 unknown elements in Ezv (including the mean of v). Letting I represent an identity matrix of order L + 1, the system can be written as a single system:

$$Ezy = [Ezz' \ \vdots \ I] \begin{bmatrix} p' \\ Ezv \end{bmatrix}$$

Thus, we have a system of L + 1 linear equations in 2(L + 1) unknowns. The maximum rank of the supermatrix of known coefficients [Ezz': I] is, therefore, L + 1. The identification problem resolves into the process of drawing on theory to place at least L + 1 independent restrictions on the system. Assuming the rank of the system is L + 1, when at least L + 1 additional linear restrictions are imposed, the system becomes identified. As it stands, it is underidentified. As the previous discussion implies, the most common type of restriction is to assume certain elements in p' and/or Ezv are zero. A total of at least L + 1 such assumptions are necessary (but not sufficient). When it is assumed that all elements in Ezv are zero, OLS are consistent estimators, but when some combination of elements from p' and Ezv are set to zero, some different method must be used. If exactly L + 1 values are set to zero, then the system is just identified, and, if some nonzero elements are in Ezy, methods such as indirect least squares or two-stage least squares are identical to instrumental variable estimation. If more than L + 1 coefficients are set to zero, then the system is over identified. If some of the nonzero elements are in Ezv, then methods such as two-stage least squares, three-stage least squares, or full-information maximum likelihood must be used. The reader is referred to standard texts in econometrics (e.g., Johnston, 1963; Goldberger, 1964) for exposition of these methods and a thorough treatment of the identification question (on the latter issue, see also Fisher, 1976).

Although economic theory sometimes may be powerful enough to specify more than the minimum L + 1 zero coefficients, status attainment theory does not justify even the minimum L + 1 assumptions with sufficient authority to generate strong confidence in any estimation technique (see the discussion of a paper by Haller and Woelfel [1971a] -- Land, 1971; Henry and Hummon, 1971; and Woelfel and Haller, 1971b). Hence, it seems unlikely that there will be in the immediate future much justification for applying methods in status attainment work such as two-stage least squares that are designed for overidentified systems.

To illustrate these comments, consider again the system represented by equations (36). Let z be parental status, y_1

represent significant other's occupational expectation of ego, and y_2 indicate ego's own occupational expectation. Because there is causal feedback in the system, the standard presumption is that y_1 is correlated with v_2, and y_2 is correlated with v_1; hence, OLS are inconsistent and should not be used. Although it does seem plausible that $Ev_1y_2 \neq 0$, and $Ev_2y_1 \neq 0$, these statements do not follow deductively from equations (36). On the other hand, theory suggests, but certainly does not confirm, that the effects of parental background on ego's occupational expectation operate indirectly via significant-other occupational expectation of ego. If one is willing to make this assumption and set $p_{21} = 0$, then consistent estimates of p_{20}, q_{21}, and Ev_1v_2 can be secured through indirect least squares.

This situation is a specific example of the general conclusion reviewed above. Without some restrictive assumptions, a unique solution for all unknowns in the equation (36b), $\hat{q}_{20}$, $\hat{q}_{21}$, $\hat{p}_{21}$, $\hat{E}v_2y_1$, cannot be found. It is plausible that $q_{21} = 0$ and, $Ev_2y_1 \neq 0$. If these assumptions are accepted, then indirect least squares are appropriate. It is almost as plausible however, that $q_{21} \neq 0$, and $Ev_2y_1 = 0$, thus indicating use of OLS. The most likely case is that neither coefficient is zero.

The example is typical of the difficulty of identifying structural equation models in status attainment research (see Nolle, 1973, for example). All published work includes a larger set of variables than the example, but each independent variable added to an equation adds two unknowns, but only one new equation to the estimating-equation set. Hence, adding variables is not sufficient to identify the parameters of any of the structural equations. What is required is better theory, but good theory takes time to develop. In the meantime, exploratory analysis based on available theory and data is necessary. As indicated here, application of OLS in such exploratory analysis frequently may be as easy to justify as alternative methods such as indirect least squares, presence of feedback loops in the system not withstanding. It does appear, however, that use of more than one estimation method frequently could contribute some insight into the substantive questions under study and serve to emphasize the exploratory nature of the analysis.

One of the main themes of this volume is that tests of theory by checking the accuracy of forecasts derived from theory constitutes a promising tool for improving the quality of theory. First, forecast accuracy is a much more stringent test of theory than most techniques in general use currently. If theory fails to generate accurate forecasts, then scholars are led to reformulate the theory. If theory consistently does generate accurate forecasts, then confidence in it will be strengthened considerably over what can be justified by reliance on multiple correlations. Secondly, empirical tests using forecasts should stimulate formulation of dynamic theory to account for the behavior of all variables at every instant on the continuous time scale. Such dynamic formulation represents a marked improvement over cross-sectional models that are so prevalent in the literature.

Much confusion is associated with selection of an appropriate estimation method. For example, one paper on statistical estimation with panel data states:

> Whenever a lagged dependent variable is included in the model, the errors will be correlated with at least one of the regressors, and ordinary least squares regression will be biased and inconsistent (Hannan and Young, 1977).

As noted above, this statement is in error. It does not follow deductively that lagged endogenous variables are correlated with any disturbance variable, including the disturbance associated with the current value of the same variable. Such a correlation might be a likely hypothesis, but it does not automatically follow from the model.

Two reasons emerge from the above discussion suggesting that OLS estimation can be recommended as an initial methodology for estimation with two panels of data. First, the required assumptions have been made explicit and are at least as plausible as alternative assumptions. With OLS, one need not assume any of the coefficients in A^* or B^* are zero but must assume all covariances between the disturbance variables $\underline{v}$ and the exogenous variables and lagged endogenous variables are zero. Non OLS methods require that some elements of A^* and/or B^* be assumed zero, but permit estimation of some covariances involving disturbance variables. Secondly, OLS selects parameter estimates to minimize the variances in the disturbance variables. If the theoretical model is approximately accurate, then the minimization criterion should generate good forecasts. If it does not, then alternative estimation methods should be tried.

Probably the main threat to consistent estimation using OLS is the autocorrelation of disturbances, meaning that the unmeasured variables are correlated over time. Autocorrelation certainly would arise, for example, if one or more important variables is omitted from the analysis and the omitted variables are correlated over time. Given a first order autocorrelation process for the disturbance variables, the correlation matrix between lagged endogenous variables $\underline{y}_{t-1}$ and current disturbance $\underline{v}_t$, can be written,

$$E\underline{v}_t\underline{y}'_{t-1} = \underline{\Gamma\Delta} + \underline{\Gamma}^2\underline{\Delta B^{*\prime}} + \underline{\Gamma}^3\underline{\Delta}(\underline{B}^{*\prime})^2 + \ldots$$

where Γ is the matrix of autocorrelations, $\underline{\Delta}$ is the matrix of cross-sectional variances and covariances among the disturbances; $\underline{\Delta} = E\underline{v}_t\underline{v}_t'$, all t. It is assumed that $\underline{\Delta}^2 < 1$, and that the process has operated over n periods with n sufficiently large so that $\Gamma^n \rightarrow 0$. This covariance is unlikely to be zero, but note that it approaches zero as either Γ (the matrix of autocorrelations) approaches zero or as $\underline{\Delta}$ (the cross-sectional variance-covariance matrix among residuals) approaches zero. Since the diagonal elements of Δ are error variances, one sees that a model accounting for a large proportion of variance helps to keep the asymptotic bias of OLS low. A short measurement interval probably contributes to small $\underline{\Delta}$ but tends to generate large $\underline{\Gamma}$; hence, it is difficult to draw conclusions about the appropriate length of the measurement interval.

This discussion has been confined to estimation methods based on two panels of data, because this seems like the most likely data base in status attainment work in the immediate future. The reader should be reminded, however, that estimation is possible with a time series on a given case extending at least to one more observation than number of variables in the model. For presentation of estimation methods with time series data, see Ostrom (1978) or Box and Jenkins (1970). Econometric texts show how to handle time series data. Doreian and Hummon (1976) give examples of estimation with time series. In addition, pooling of time series of cross sections sometimes produces improved estimation (see Hannon and Young [1977] for a review and survey of recent literature).

In sum, it appears that use of OLS to estimate the coefficients of the integral equation (33) is justified as an initial strategy, but that alternative specification might fruitfully be investigated in future research. In particular, it seems advisable to explore how the dynamic quality of the differential equation model can be exploited to address the statistical issues. (See Doreian and Hummon (1976) for several illustrations of estimation techniques tailored to specific substantive questions.) Also, specification of a model accounting for measurement error is desirable (see, e.g., Coleman, 1968; Joreskog, 1973; Wheaton, et al., 1977), but it is beyond the scope of the present volume.

Computer Calculations

The main purpose of this section is to describe a computer program package which is available for converting raw-data input into estimates of the coefficients of the differential equation (28) (the matrices A and B) by using OLS to estimate the parameters of the corresponding integral equation (29) (the matrices A* and B*, contained in the matrix P). The presentation proceeds in four parts. First, a brief discussion of the calculating task is presented. Secondly, the input required to use the computer program is described. Thirdly, the output of the program is described. Finally, provision in the program package for carrying out data transformations is described.

The Calculating Task

The logic of the program is quite simple. The calculations proceed in three steps:

1. Calculate means, standard deviations, and correlations

2. Calculate OLS estimates of $\underline{P} = [\underline{A}^*, \underline{B}^*]$, using the means, standard deviations, and correlations as input

3. Calculate estimates of the coefficient matrices $\underline{A}$ and $\underline{B}$ from $\underline{A}^*$ and $\underline{B}^*$, using equations (30) and (31)

Calculation of means and standard deviations by the program is executed by omitting missing observations on each variable and dividing the accumulated sums by the number of observations present for each variable. Correlations are calculated using all data present for each pair of variables.

The only unusual calculations are associated with finding the matrix logarithm demanded by equation (30b). This calculation can be accomplished by finding the characteristic roots and vectors of the matrix $\underline{B}^*$ and applying formula (20). Since $\underline{B}^*$ is not a symmetric matrix, its characterhstic roots and vectors may sometimes be complex numbers; hence, complex-number arithmetic is required. To find the eigenvalues and eigenvectors, the program package uses subroutines from an eigenvalue-eigenvector package called EISPAC distributed by The National Software Center at the Argonne National Laboratories in Chicago. All programming is written in FORTRAN IV, Level G.

The calculations are executed on the assumption that the matrix $\underline{B}$ is full rank. This is not a serious limitation, however, since it is rare to find a singular matrix in empirical data.

Input to the Program

To operate the program, three types of input are required: a) data containing N cases for all variables, b) control cards describing the data, and c) system-information cards.

Data Input

The program assumes that observations for all variables are grouped together for each case (e.g., individual respondent), and that the position of each variable on the cards, tape or disc is the same for all cases. To illustrate the proper arrangement of data, figure 7 shows the arrangements on cards for three variables and five cases, each case being a person.

Col. No.:	1234567890 1234567 . . . 80		
card 1, person 1	75	6.9	.46
card 2, person 2	2	.3	.82
card 3, person 3	32	3.8	.19
card 4, person 4	65	7.4	.56
card 5, person 5	42	.9	.94

Figure 7. Illustration of organization of Input Data

(Of course, the decimal points shown in the illustration need not be punched on the cards, as a FORMAT statement can be used to position the decimal for each number.) If there are too many variables to fit on one card, continuation cards can be used for each case. There may be any number up to and including 1000 variables in the data set. The program selects the variables needed for a given analysis in whatever order designated by the user, according to information supplied by the control cards. The data may be stored according to a FORTRAN FORMAT, or they may be unformatted, using FORTRAN unformatted input-output. The mode of data storage and FORMAT statement, if needed, form part of the information supplied to the program by the control cards.

Control Cards

There are three control cards which supply information about the data to the program. The three cards are a) RUN CARD,

b) DATA CARD, and c) FORMAT CARD. The content of these three types is described below.

One RUN CARD is required for each analysis, and it must appear first. The RUN CARD supplies seven parameters to the program; the information, location on the RUN CARD and default value for each parameter are described below.

	Location (Column Numbers)	Purpose	Default Value
parameter 1 (NV):	1 - 5	Indicates the number of variables in the analysis	None
parameter 2 (NIV):	6 - 10	Indicates the number of exogenous variables in the differential equation system	None
parameter 3 (NR):	11 - 15	Indicates the number of variables in the data set from which data is to be read, or the number accounted for in the FORMAT CARD, if the latter is less than the former	Set equal to parameter 1
parameter 4 (ITYP):	16 - 20	Indicates whether the data are formated or unformated; 0 = unformated; nonzero indicates the number of cards, up to six, on which the FORMAT is punched	0
parameter 5 (NU):	21 - 25	Indicates the unit number associated with the READ statement used for data input. Selection of the unit is described in the text under system-information cards	11

	Location (Column Numbers)	Purpose	Default Value
parameter 6 (T):	26 - 30	Indicates the number of months elapsed between panel 1 and panel 2	12
parameter 7 (MIS):	31 - 40	Indicates the numerical value used to indicate missing data (if the missing data code is not the same for all variables, transformations must be executed to change all missing data codes to the same value. The last subsection, entitled "Data Transformations" illustrates how this can be done). Zero is an invalid missing data code.	10^9

At least one DATA CARD is required for each analysis, and should follow the RUN CARD. The DATA CARD indicates the variable sequence number of each variable to be used in the analysis. Each sequence number is right justified in a five column field -- the first sequence number appearing in columns 1 - 5, the second appearing in columns 6 - 10, etc. (Any number of continuation cards may be used.) If the data are formatted ($1 \leq ITYP \leq 6$; ITYP = parameter 4 on the RUN CARD), then the sequence number gives the order in which the variable appears in the FORMAT statement supplied by the FORMAT CARD. If the data are unformatted (ITYP = 0), then the variable sequence number indicates the order in which the variable appears in the input data set. If for a formatted data set, a FORMAT statement is supplied describing all variables in the input data set, then the sequence number gives the order of the variable in the data set for formatted data, just as it does for unformatted data. If the data are formatted, it is recommended that a standard format describing all variables in the data set be supplied; this procedure avoids the need to construct a new FORMAT CARD for each analysis, thereby reducing the numerous chances for error inherent in constructing FORMAT statements.

In each analysis, the variables are ordered in the same order in which they appear on the DATA CARD. The program assumes that exogenous variables are listed first, followed by the endogenous variables for panel one, then by the endogenous variables for panel two. The endogenous variables for the two panels should be listed in the DATA CARD in the same order as the order of the endogenous variables for panel one. If all variables in the FORMAT statement are to be used in the order in which they occur in the FORMAT, then one may insert a single blank DATA CARD.

One FORMAT CARD is required if the input data are formatted, but should be omitted if the input data are unformatted. If included, the FORMAT CARD follows the DATA CARD. The FORMAT statement is limited to six cards. All the rules for FORTRAN FORMAT statements apply, and, since these rules are widely available in other sources (e.g., the SPSS Manual [Nie, et al., 1975], and numerous texts on FORTRAN programming) and familiar to most readers, they are not reviewed here. It should be noted, however, that the word FORMAT does not appear on the FORMAT CARD.

To illustrate the use of control cards, assume the input data set is formatted and contains 100 variables. For the current run, suppose there are two exogenous variables and five endogenous variables, each of the latter measured twice -- once at t_o and once at t_1, giving twelve variables in the current run. Assume that all 100 variables appear on cards in adjacent five-column fields (thus requiring seven cards per case), and that the twelve variables for the current run are arranged as follows on the cards:

Symbol		Name	Card No.	Columns	Sequence Numbers
x_1	=	first exogenous variable	1	6 -10	2
x_2	=	second exogenous variable	1	21-25	5
y_{o1}	=	first endogenous variable, t_o	1	51-55	11
y_{o2}	=	second endogenous variable, t_o	1	41-45	9
y_{o3}	=	third endogenous variable, t_o	2	21-25	21
y_{o4}	=	fourth endogenous variable, t_o	2	61-65	29
y_{o5}	=	fifth endogenous variable, t_o	3	11-15	35
y_{11}	=	first endogenous variable, t_1	3	26-30	38

Symbol		Name	Card No.	Columns	Sequence Numbers
y_{12}	=	second endogenous variable, t_1	3	46-50	42
y_{13}	=	third endogenous variable, t_1	3	61-65	45
y_{14}	=	fourth endogenous variable, t_1	5	31-35	71
y_{15}	=	fifth endogenous variable, t_1	3	31-35	39

In this case the sequence number follows directly from the card and column numbers. Assuming every five-column field contains one and only one of the 100 variables, the sequence number was calculated by: sequence # = 16 × (card # -1.0) + (last col. # in field)/5. These sequence numbers give the order in which the variables appear in the input data. Further, suppose that any missing data is coded 10^9 (1.E9 on the cards), that twelve months separate t_o from t_1, and that it is desired that the variables appear in the analysis in the order as listed in the above tabulation. The control cards shown in figure 8 can be used to operate the program.

```
12   2  100   1                                                (RUN CARD)
 2   5   11   9   21   29   35   38   42   45   71   39      (DATA CARD)
(16F5.0)                                                    (FORMAT CARD)
```

Figure 8. Control cards with FORMAT for all variables

The RUN CARD indicates 12 variables in the analysis, 2 exogenous variables, 100 variables accounted for by the FORMAT CARD, and formatted data with one FORMAT CARD. Default values for the other parameters on the RUN CARD are assumed by the program. The DATA CARD lists the sequence number of each variable. The sequence numbers appear in the order in which they will be used for the analysis. The FORMAT card indicates the FORMAT of each of the five cards per case. FORTRAN assumes the same FORMAT is to be used for each card. The FORMAT indicates five column fields with the decimal placed to the right of all digits. Decimals punched on the card override the FORMAT.

Alternatively, the cards in figure 9 accomplish the same purpose. In this case, the FORMAT CARDS account for only the twelve variables to be used in the analysis; hence the default value can be used for parameter 3 on the RUN CARD, setting

```
12    2         2                                                    (RUN CARD)
 1    2    4    3    5    6    7    8    10   11   12    9          (DATA CARD)
(5X,F5.0,10X,F5.0,15X,F5.0,5X,F5.0/20X,F5.0,35X,                      (FORMAT
 F5.0/10X,F5.0,10X,2F5.0,10X,F5.0,10X,F5.0//30X,F5.0)                  CARDS)
```

Figure 9. Control cards with FORMAT for variables in current analysis

parameter 3 equal to parameter 1. This illustrates the point that the number of variables accounted for by the FORMAT should be the value of parameter 3. Note also, that the variable sequence numbers also reflect the sequence number of each variable which is established by the FORMAT CARD.

If the data were unformatted, the first alternative is appropriate, except that parameter 4 on the RUN CARD should be zero (or blank) and the FORMAT CARD should be omitted.

System Information Cards

The system cards required vary among installations. The program described here was written and tested on the AMDAHL system at the Instructional and Research Computing Center, Ohio State University. The AMDAHL works like an IBM machine in most important respects, including the main features of JOB CONTROL LANGUAGE (JCL) used by IBM. The description of system cards given here is confined to the required JCL; it is assumed that users know or have access to general rules for JCL. Users of systems not accepting JCL must consult with personnel of their facilities to learn about the required system cards.

A schematic view of the type and placement of JOB CONTROL cards is shown below.

JOB statement

EXEC statement
DATA DEFINITION statement for FORTRAN program

[FORTRAN program deck, followed by a card with /* in cols. 1 and 2]

DATA DEFINITION statement for input data

[input data, if on cards, followed by a card with /* in cols. 1 and 2]

DATA DEFINITION statement for program CONTROL CARDS

[Program CONTROL CARDS, followed by a card with /* in cols. 1 and 2]

If the data are on cards the last two DATA DEFINITION (DD) cards can be in reverse order. The information not in brackets is JCL; the bracketed information is not JCL. The JCL cards are described below in the order listed.

The job statement takes the following form

//jobname JOB operands

The double slash always appears in the first two columns of all job control cards; no spaces may appear except where shown. The jobname is a one-to-eight character alphameric name generally selected by the user. It must begin with an alphabetic character or $. Some installations may supply the job card, or the first card of two or more cards required to complete the job statement. (When A JCL statement spans more than one card, the last character on all but the last card of the statement is a comma.) The operation "JOB" must be punched as is. The operands vary according to the job and installation.

The execute (EXEC) statement takes the following form

//stepname EXEC procname,operands

The stepname is an alphameric name of up to eight characters supplied by the user. It may be omitted, but one or more blanks must separate the // from the EXEC if the stepname is omitted. The EXEC must be punched as is. The procname refers to the name of a catalogued procedure supplying JCL for the FORTRAN program. Procedures may vary among installations. Operands are optional and generally can be omitted. An example with stepname STEP1, procname FORTRUN, and no operands follows:

//STEP1 EXEC FORTRUN

The DATA DEFINITION statement for the FORTRAN deck takes the following form:

//ddname DD *

It tells the computer that the FORTRAN deck follows. An example using FORTRUN follows

```
//STEP1 EXEC FORTRUN
//CMP.SYSIN DD *
```

where the ddname is CMP.SYSIN.

The DD statement for the input data takes the following form if the data are on cards:

```
//ddname DD *
```

where ddname is the data-definition statement name and depends on the catalogued procedure. For FORTRUN, the following statement is correct:

```
//GO.FT11F001 DD *
```

The eleven following FT designates the unit number and should match parameter 4 on the RUN CARD. It need not be eleven, so long as it matches parameter 4 and is not the same as unit numbers designated for READ, WRITE, and punch by the catalogued procedure. Units 5, 6, and 7, respectively, are typically reserved for these operations.

If the data are stored on tape or disc, the * following the "DD" must be replaced with information describing the location of the data, such as the data set name (DSN) and DCB information. The variety of such information is too great to permit description here. In order to create the tape or disc file, the user must know the required information that can be used to access the data by this program.

The DD statement for the control cards has the same general form as that for input data on cards. This general form typically assumes the specific form of

```
//GO.SYSIN DD *      or
//SYSIN DD *
```

A complete run using the FORTRUN procedure and unit 11 for input data on cards is illustrated below.

```
//A1000 JOB REGION=200K,TIME=1
//STEP1 EXEC FORTRUN
//CMP.SYSIN DD *
      FORTRAN program
/*
//GO.FT11F001 DD *
      data cards
/*
//GO.SYSIN DD *
     CONTROL CARDS
/*
//
```

The double slash in the first two columns of the last card designates the end of the job. Note that t[illegible] jobname is A1000, and two operands are specified, indicating [illegible]K bytes of storage and a one-minute time limit for the job.

It should be noted that the program assumes that unit 5 is reserved for reading control cards and unit 6 for printing. If the catalogued procedure makes a different allocation of unit numbers, the following DD statement should follow immediately the program deck:

```
//step.FT06F001 DD SYSOUT=A,DCB=RECFM=FA
```

and the following card should precede the CONTROL CARDS:

```
//step.FT05F001 DD *
```

where step refers to a step name (e.g., GO) within the catalogued procedure.

Program Output

There are four types of output generated by the programs, each type beginning on a new page of the printout. These four types are: (a) a record of the information contained on the CONTROL CARDS and default values, (b) univariate and bivariate statistics, (c) OLS multiple regression statistics, and (d) estimates of parameters of the differential equation. These four types of output are described briefly below.

Record of control cards

A sample of this output is shown below. The output is labeled with terminology closely matching that used in this section. Most of the output is self explanatory due to the

```
RUN-CARD PARAMETERS --

   NV =      8
  NIV =      2
   NR =      8
 ITYP =      0
   NU =     12
    T =   12.0
  MIS =  .1000E+10

VARIABLE   SEQUENCE
  ORDER     NUMBER

    1          1
    2          2
    3          3
    4          4
    5          5
    6          6
    7          7
    8          8
```

labels. Parameters of the RUN CARD are clearly labeled. When default values are used by the program, these are printed. Note that information on the DATA CARD(s) is given under the column titled "SEQUENCE NUMBER." The corresponding order of the variable used in the current run is printed in the column just to the left, labeled "VARIABLE ORDER." The FORMAT CARD(s) is(are) printed exactly as it(they) appear in the input. In the example, there is no FORMAT card.

Univariate and bivariate statistics

Univariate statistics are calculated for each variable by including all data present for that variable, and bivariate statistics are calculated by including all cases for which data are present on both variables in the pair. Thus the sample size and univariate statistics may differ for a given variable depending on the other variable with which it is paired. Consequently, the format for writing the univariate and bivariate statistics allows for a different sample size, and univariate statistics for each variable pair. The following information is printed for each variable pair. A sample of the output is reproduced on the following page.

Column Heading	Content
I	Variable order number of first variable in the pair
J	Variable order number of second variable in the pair
N(I,J)	Number of observations present for variable pair (I,J)
XBR(I)	Mean of variable I when paired with variable J
XBR(J)	Mean of variable J when paired with variable I
SD(I)	Standard deviation of variable I when paired with variable J
SD(J)	Standard deviation of variable J when paired with variable I
COV(I,J)	Covariance between variable I and variable J using all cases in which information is present for both variables, the number of such cases being N(I,J)
R(I,J)	Correlation between variable I and variable J using all cases in which information is present for both variables, R(I,J)=COV(I,J)/(SD(I)*SD(J))

UNIVARIATE AND BIVARIATE STATISTICS, N = 100

I	J	N(I,J)	XBR(I)	XBR(J)	SD(I)	SD(J)	COV(I,J)	R(I,J)
1	1	100	-4.437042	-4.437042	391.8011	391.8011	153508.1	1.000000
1	2	100	-4.437042	-80.33465	391.8011	243.4670	-62198.64	-.6520412
2	2	100	-80.33465	-80.33465	243.4670	243.4670	59276.20	1.000000
1	3	100	-4.437042	55.25454	391.8011	8.118765	-142.9749	-.4494734D-01
2	3	100	-80.33465	55.25454	243.4670	8.118765	-21.75362	-.1100529D-01
3	3	100	55.25454	55.25454	8.118765	8.118765	65.91434	1.000000
1	4	100	-4.437042	-10.07497	391.8011	44.07031	-9090.164	-.5264537
2	4	100	-80.33465	-10.07497	243.4670	44.07031	1559.937	.1453855
3	4	100	55.25454	-10.07497	8.118765	44.07031	129.9071	.3627959
4	4	100	-10.07497	-10.07497	44.07031	44.07031	1942.192	1.000000
1	5	100	-4.437042	129.5608	391.8011	400.9171	33915.69	.2159138
2	5	100	-80.33465	129.5608	243.4670	400.9171	10343.66	.1059692
3	5	100	55.25454	129.5608	8.118765	400.9171	835.3142	.2566288
4	5	100	-10.07497	129.5608	44.07031	400.9171	-3823.602	-.2164074
5	5	100	129.5608	129.5608	400.9171	400.9171	160734.5	1.000000
1	6	100	-4.437042	55.83353	391.8011	5.383714	-691.7146	-.4227449
2	6	100	-80.33465	55.83353	243.4670	5.383714	359.2122	.2740496
3	6	100	55.25454	55.83353	8.118765	5.383714	21.72343	.4970003
4	6	100	-10.07497	55.83353	44.07031	5.383714	133.4654	.5633465
5	6	100	129.5608	55.83353	400.9171	5.383714	-308.5620	-.1429573
6	6	100	55.83353	55.83353	5.383714	5.383714	28.98436	1.000000
1	7	100	-4.437042	-9.380658	391.8011	17.85116	81.71739	.1163375D-01
2	7	100	-80.33465	-9.380658	243.4670	17.85116	-1720.329	-.3958266
3	7	100	55.25454	-9.380659	8.118765	17.85116	2.503508	.1727398D-01
4	7	100	-10.07497	-9.380658	44.07031	17.85116	424.0796	.5390571
5	7	100	129.5608	-9.380658	400.9171	17.85116	-3469.643	-.4848012
6	7	100	55.83353	-9.380658	5.383714	17.85116	14.36941	.1495170
7	7	100	-9.380658	-9.380659	17.85116	17.85116	318.6638	1.000000
1	8	100	-4.437042	162.3972	391.8011	156.4833	14095.62	.2297430
2	8	100	-80.33465	162.3972	243.4670	156.4833	-4027.212	-.1057052
3	8	100	55.25454	162.3972	8.118765	156.4833	261.0605	.2054865
4	8	100	-10.07497	162.3972	44.07031	156.4833	-638.1230	-.9253164D-01
5	8	100	129.5608	162.3972	400.9171	156.4833	49678.15	.7918495
6	8	100	55.83353	162.3972	5.383714	156.4833	-130.6222	-.1550482
7	8	100	-9.380658	162.3972	17.85116	156.4833	-607.0796	-.2173257
8	8	100	162.3972	162.3972	156.4833	156.4833	24497.04	1.000000

OLS multiple regression

The OLS regression yields the matrix $\hat{\underline{P}} = [\hat{\underline{A}}^*, \hat{\underline{B}}^*]$, in standardized and unstandardized form. It also calculates multiple correlations. The output is labeled in a fairly clear manner. The first line of output includes the total sample size (N) the degrees of freedom for numerator and denominator of F ratios associated with each regression, and error code (IER). If IER = 0, calculations proceeded normally. If IER = -1, matrix inversion did not occur, possibly because the correlation matrix is not positive definite. If IER > 0, excessive rounding error may have occurred.

The regression coefficients (standardized and unstandardized) are arranged so that dependent variables (DV) cross rows and independent variables cross columns, as labeled. The integer numbers to the left of each row and at the top of each column of these matrices give the variable sequence number (not the order number). A sample of the output is shown on the following page.

Differential equations output

There are three types of output related to the differential equation, each type beginning on a new page. First, the unstandardized matrix $\hat{\underline{P}}$ is subdivided into its component parts $\hat{A}^*$ and $\hat{B}^*$, and these components are printed without changing their values. A sample of the output is reproduced below.

```
MATRIX OF A* COEFFICIENTS --
              0              1               2
6       40.759       -.86617D-03     .47827D-02
7       11.220        .10678D-01    -.22872D-01
8       185.67       -.27297D-01    -.16887
MATRIX OF B* COEFFICIENTS --
              3              4               5
6       .29262        .36038D-01    -.27081D-02
7      -.32118        .28055        -.14024D-01
8      -1.3928        .44887         .34361
```

The matrices $\underline{A}^*$ and $\underline{B}^*$ are labeled in the output. Again, dependent variables cross rows and independent variables ($\underline{x}$ for $\underline{A}^*$ and y_o for $\underline{B}^*$) cross columns. The integer numbers beside each row and above each column identify the variable sequence number. The "0" above column one of $\underline{A}^*$ designates the intercepts.

OLS MULTIPLE REGRESSION

N = 100 NUMERATOR DF = 5 DENOMINATOR DF = 94 RETURN CODE FROM INVS, IER = 0

STANDARIZED REGRESSION COEFFICIENTS--

DV	INDEPENDENT VARIABLES-- 1	2	3	4	5
6	-0.06304	0.21629	0.44128	0.29500	-0.20167
7	0.23436	-0.31194	-0.14607	0.69262	-0.31497
8	-0.06835	-0.26273	-0.07226	0.12642	0.88035

UNSTANDARDIZED REGRESSION COEFFICIENTS--

DV	INTERCEPT 0	INDEPENDENT VARIABLES -- 1	2	3	4	5
6	40.759	-.86617D-03	.47827D-02	.29262	.36038D-01	-.27081D-02
7	11.220	.10678D-01	-.22872D-01	-.32118	.28055	-.14024D-01
8	185.67	-.27297D-01	-.16887	-1.3928	.44887	.34361

MULTIPLE R'S--

6	7	8
0.70729	0.80607	0.82621

The eigenvalues and eigenvectors of $\underline{B}^*$ comprise the <u>second</u> type of output related to the differential equation. A sample of the output is shown below.

```
REAL PART OF EIGENVALUES OF B* --

         1            2            3
   .26545       .26545       .38588

IMAGINARY PART OF EIGENVALUES OF B* --

         1            2            3
   .13346      -.13346       .0

EIGENVECTORS OF B* FOLLOW.  THEY ARE STORED SUCH THAT
IF ADJACENT VECTORS (K,K+1) ARE CONJUGATE PAIRS, THE KTH
COLUMN CONTAINS THE REAL PART, AND THE (K+1)TH COLUMN
CONTIANS THE IMAGINARY PART

RIGHT EIGENVECTORS --

             1              2              3
 1    .17434D-01     .46292D-02    -.32109D-01
 2   -.12854D-01     .67787D-01    -.17795D-01
 3    .23198         .89314D-01     .86900

LEFT EIGENVECTORS --

             1              2              3
 1    18.568         -2.1615       -10.358
 2    -2.1149        -6.9361       -.29661
 3    .64275         -.22190        .76196
```

The first row of output displays the real part of the eigenvalues of $\underline{B}^*$, and the second row displays the imaginary component of the corresponding eigenvalues.

Let $\underline{\Lambda}$ be a diagonal matrix with diagonal elements equal to the eigenvalues of $\underline{B}^*$, and let $\underline{V}$ be the matrix whose columns are eigenvectors associated with $\underline{\Lambda}$, then

$$\underline{B}^*\underline{V} = \underline{V\Lambda}$$

Assuming B^* is diagonalizable, $\underline{V}$ is the matrix of right eigenvectors of $\underline{B}^*$, and $(\underline{V}^{-1})'$ is the matrix whose columns are composed of the left eigenvectors of $\underline{B}^*$. The output labeled RIGHT EIGENVECTORS gives $\underline{V}$, and the output labeled LEFT EIGENVECTORS gives $(\underline{V}^{-1})'$.

The estimates of the parameter matrices of the differential equation (28) ($\hat{A}$ and $\hat{B}$) comprise the <u>third</u> type of output related to the differential equation. A sample of the output is reproduced below.

```
INTERCEPTS & COEFFICIENTS OF PREDETERMINED VARIABLES,     A --
T1     INTERCEPT        INDEPENDENT VARIABLES --
           0                1             2
 6      68.492         -.24731D-02    .94802D-02
 7      57.476          .16312D-01   -.39912D-01
 8      375.44         -.57782D-01   -.23550
COEFFICIENTS OF ENDOGENOUS VARIABLES, B --
T1          TO. VARIABLES --
              3              4             5
 6      -1.1717          .12428       -.55639D-02
 7      -1.1504        -1.1618        -.46798D-01
 8      -3.4310         1.6079        -1.0461
```

The format of the output follows precisely the format for printing A* and B*. The intercept and exogenous variables (x) appear across the columns of A. The t_0 (T0) endogenous variables appear across the columns of B, and t_1 (T_1) values vary across rows, as labeled.

Data Transformation

A subroutine named DATA is called after each line of data is read. This subroutine can be used to carry out data transformations. The general use of the subroutine is described below, and an illustration is presented in which varying missing data codes are recoded to a single value.

If no data transformations are desired, the subroutine is defined as follows:

```
      SUBROUTINE DATA(X,R,M,IPK)
      REAL X(1), R(1)
      INGEGER*2 IPK(1)
      DO 1 J = 1, M
         X(J) = R(IPK(J))
 1       CONTINUE
      RETURN
      END
```

This version of DATA is included in the standard version of the program package, and nothing need be added if the user desires no data transformations. The variables are defined as follows:

X = a one-dimensional array which receives the values of the variables to be used in the current analysis

R = a one-dimensional array which contains values of all variables in the data set, or all variables accounted for by the FORMAT statement

M = a scalar giving the number of variables to be used in the current analysis

IPK = a one-dimensional array (length two integer) whose elements contain the sequence numbers of the variables to be used in the current analysis. These are listed in the order in which they appear in the analysis on the RUN CARD

The values of R, M, and IPK are passed to DATA by the subprogram which accumulates sums for the correlation calculations; the variable X is returned to the calling program.

If data transformations are desired, one may replace the above version of DATA with a user-written version. The user-written routine must contain the same variables in the calling list that are shown above (i.e., X, R, M, IPK) and must define each element of X. An example of a possible user-written version with six variables in the current analysis is shown below. The example recodes a variety of missing data codes to the standard value of 10^9 (1.E9).

```
      SUBROUTINE DATA(X,R,M,IPK)
      REAL X(1), R(1), XMIS(6)/9.,99.,9.,9.,9.,999./
      INTEGER*2 IPK(1)
      DO 1 J=1,M
         X(J) = R(IPK(J))
1        IF(X(J).EQ.XMIS(J)) X(J)=1.E9
      RETURN
      END
```

Note that the array XMIS contains the missing data codes for each X value. These are defined by the declaration statement: REAL X(1), R(1), XMIS(6)/9.,99.,9.,9.,9.,999./.

CHAPTER 5

ADDITIONAL INTERPRETATIONS OF CHANGE COEFFICIENTS

The material in the preceding chapter supplies the background necessary for a more thorough examination of substantive interpretations of differential equation systems than could be presented in chapter 2. There are four topics to be examined. First, interpretations of the coefficients of the differential equation (32) are compared to interpretations of the corresponding integral equation (33). Secondly, the time path of the career expectation variables implied by the differential equations is examined. This discussion includes consideration of equilibrium and oscillating systems. Thirdly, a method of standardizing the coefficients of the differential equation system is presented and the advantages and disadvantages of interpreting the standardized coefficients are discussed. Finally, a brief presentation discusses a generalized correlation for assessing accuracy of forecasts.

Coefficients of the Differential Equation and Cross-Lagged Regression Coefficients

The OLS estimates of A* and B* in the integral equation (33) are cross-lagged regression coefficients. It has been suggested that cross-lagged coefficients be used to assess the relative effects of one variable on another (e.g., Heise, 1970). In the present context, interpretation of effects by reference to the parameters A and B of the differential equation system (32) is an obvious alternative. As noted in chapter 2, conclusions about causal relationships may depend on which coefficient matrices are used in the interpretations.

To facilitate the discussion, some of the basic relationships are reproduced here, and the notation is altered slightly.

(50) $\dot{\underline{y}} = \underline{A}\underline{x} + \underline{B}\underline{y} + \underline{u}$ (differential equation)

(51) $\underline{y}_t = \underline{A}^*(t)\underline{x} + \underline{B}^*(t)\underline{y}_o + \underline{u}^*(t)$ (integral equation)

where

$\underline{x}$ = (1 + L) x 1 vector of exogenous variables

$\underline{y}$ = K x 1 vector of endogenous variables

$\underline{y}_t$ = y at time t

$\underline{u}$ = K x 1 vector of disturbances

$\underline{u}^*(t)$ = K x 1 vector of disturbances

$\underline{A}$ = K x (1 + L) matrix of coefficients of the exogenous variables

$\underline{B}$ = K x K matrix of coefficients of the endogenous variables

$\underline{A}^*(t)$ = K x (1 + L) matrix of coefficients of the exogenous variables for a discrete interval of time equal to t

$\underline{B}^*(t)$ = K x K matrix of coefficients of the endogenous variables for a discrete interval of time equal to t

The matrices $\underline{A}^*(t)$, $\underline{B}^*(t)$, and the vector $\underline{u}^*(t)$ are denoted here explicitly as functions of the time interval t between $t_o = 0$ and $t_1 = t$; the fact that these elements depend on time is an important aspect of their interpretation. The following relationships hold between $\underline{A}$, $\underline{B}$ and $\underline{A}^*(t)$, $\underline{B}^*(t)$.

$$\text{(52a)} \quad \underline{B}^*(t) = e^{\underline{B}t}$$

$$\text{(52b)} \quad \underline{B} = [\ln \underline{B}^*(t)]/t$$

$$\text{(53a)} \quad \underline{A}^*(t) = (e^{\underline{B}t} - \underline{I})\underline{B}^{-1}\underline{A}, \quad \text{if } |\underline{B}| \neq 0$$

$$\text{(53b)} \quad \underline{A} = \underline{B}(e^{\underline{B}t} - \underline{I})^{-1}\underline{A}^*(t), \text{ if } |\underline{B}| \neq 0$$

$$\text{(54a)} \quad \underline{A}^*(t) = [(e^{\underline{B}t} - \underline{I})\underline{V}_1\underline{\Lambda}_1^{-1}\underline{V}^{(1)} + t\cdot\underline{V}_2\underline{V}^{(2)}]\underline{A}, \text{ if } |\underline{B}| = 0$$

$$\text{(54b)} \quad \underline{A} = [(e^{\underline{B}t} - \underline{I})\underline{V}_1\underline{\Lambda}_1^{-1}\underline{V}^{(1)} + t\cdot\underline{V}_2\underline{V}^{(2)}]^{-1}\underline{A}^*(t),$$

$$\text{if } |\underline{B}| = 0$$

In a sense, the coefficients in $\underline{A}$ and $\underline{B}$ are fundamental to the theory because they express the instantaneous impact of $\underline{x}$ and $\underline{y}$ on changes in $\underline{y}$ over time. Also, $\underline{A}$ and $\underline{B}$ are simple to interpret because they are constant over time. In contrast, as revealed by equations (52a) and (53a) or (54a), $\underline{A}^*$ and $\underline{B}^*$ are matrix functions involving time. Consequently, simplistic comparisons among entries of A^* and B^* to index relative effects of different variables should be avoided. To see why, examine the relation $\underline{B}^*(t) = e^{\underline{B}t}$; this implies that

$$\text{(54)} \quad \underline{B}^*(t) = [\underline{B}^*(1)]^t$$

where $\underline{B}^*(1)$ is the coefficient matrix of the integral equation over a single time period, say one year. Thus, for example, the matrix of cross-lagged regression coefficients over a two-year interval is just the square of the matrix for a one-year interval. From this fact alone it can be seen that the elements of $B^*(t)$ for different t do not bear a simple relationship to each other.

Consider a four-variable example with the cross-lagged coefficients for a one year interval shown in the following matrix

$$\underline{B}^*(1) = \begin{pmatrix} .6 & .4 & .5 & .13 \\ .2 & .8 & .11 & .4 \\ .05 & .1 & .9 & .5 \\ .3 & .01 & .4 & .7 \end{pmatrix}$$

If the same system were studied over a two-year interval, the cross-lagged regression would be

$$\underline{B}^*(2) = [\underline{B}^*(1)]^2 = \begin{pmatrix} .504 & .611 & .846 & .579 \\ .406 & .735 & .447 & .681 \\ .245 & .195 & 1.046 & .847 \\ .412 & .175 & .791 & .733 \end{pmatrix}$$

Clearly, interpretations drawn from $B^*(1)$ would differ from those based on $\underline{B}^*(2)$. First, observing $B^*(1)$, one would conclude that the effects of the lagged value of each endogenous variable on its current value dominate the system because in every case the diagonal elements of $\underline{B}^*(1)$ are substantially greater than the off-diagonal elements. This observation does not hold for $\underline{B}^*(2)$, however. In fact, observing $\underline{B}^*(2)$, one might be impressed by the fact that "effects" other than those of the lagged endogenous variables on themselves are so strong. Also, the relative magnitudes of the coefficients have shifted, the most dramatic example being the change in relative magnitude of the (1, 4) and (4, 1) elements: for the one-year interval, $b^*_{14}(1) = .13$, and $b^*_{41}(1) = .3$, the difference being $.13 - .3 = -.17$. For the two-year interval, the direction of the difference has changed, $b^*_{14}(2) = .579$, and $b^*_{41}(2) = .412$, so that the difference now is $.579 - .412 = +.167$. In the first case, one would conclude that the effect of variable 1 on variable 4 far exceeds the reverse effect; whereas, in the second case just the opposite conclusion is suggested. Numerous other changes in the relative magnitudes of the coefficients can be observed when comparing $\underline{B}^*(1)$ to $\underline{B}^*(2)$, but most are substantively inconsequential. One interpretation of the coefficients in $\underline{B}^*(t)$ is that they index accumulated total effects -- indirect plus direct effects -- over the time interval.

Observing equation (5?a) or (54a) it is easy to surmise that the conclusions regarding interpretation of $\underline{B}^*(t)$ extend to interpretation of $\underline{A}^*(t)$. For the sake of brevity, however, no example is presented for $\underline{A}^*(t)$.

The main conclusion to be drawn from this discussion is not necessarily that the coefficient matrices of the integral equation hold no interpretative value. Rather, interpretations should proceed with caution in full knowledge of the manner in which $\underline{A}^*(t)$ or $\underline{B}^*(t)$ depend on time. One useful interpretation of these matrices is as change coefficients of a difference equation over a finite time interval, t. To motivate this interpretation, subtract y_o from both sides of equation (51):

$$(55)\quad \underline{y}_t - \underline{y}_o = \underline{A}^*(t)\underline{x} + [\underline{B}^*(t) - \underline{I}]\underline{y}_0 + \underline{u}^*(t)$$

The left side of this result is a change vector in $\underline{y}$ ($\Delta \underline{y} = \underline{y}_t - \underline{y}_o$) over a finite time interval, t. It can be seen from the right side of the equation that the coefficients of $\underline{x}$ and off-diagonal coefficients of $\underline{y}_o$ are precisely those of the integral equation (51). The diagonal coefficients in the change equation can be calculated simply from the diagonal coefficients in the integral equation. If the time interval t is taken to the limit of zero and division by dt is effected, equation (55) reduces to the differential equation (as required by all solutions to differential equations). Thus, one view of the differential equation is that it is a special case of the difference equation. The question therefore may be raised: why consider the coefficients of the differential equation more fundamental than those of the integral equation? The answer rests with the initial hypothesis that effects are instantaneous. If one doubts the hypothesis of instantaneous effects, then the parameters of the differential equation might not be viewed as fundamental. Even if the hypothesis of instantaneous effects is not tenable, however, it does not necessarily follow that $\underline{A}^*(t)$ and $\underline{B}^*(t)$ for some unknown t give the fundamental parameters of the system. A viable alternative may be to alter the functional form of the differential equation system, thus retaining the conceptual advantages of continuous rather than discrete time.

Time Path of Career Expectations

Omitting the disturbance term from the integral equation (51) leaves an equation that can be used to forecast the values of all career expectation variables and all individuals at any point in time:

(56) $\hat{\underline{y}}_t = \underline{A}^*(t)\underline{x} + \underline{B}^*(t)\underline{y}_0$

where $\hat{\underline{y}}_t$ is the vector of forecasts. The time paths of the elements of $\hat{\underline{y}}_t$ expresses the development over time of career expectations predicted by the theory; the patterns associated with $\underline{u}^*(t)$ are random and not accounted for by the theory. The purpose of this section is to examine the behavior of the predicted vector over time.

The behavior of the time path of predictions can be studied readily if one substitutes the expressions for $\underline{A}^*(t)$ and $\underline{B}^*(t)$ in terms of $\underline{A}$ and $\underline{B}$ into (56):

(56a) $\hat{\underline{y}}_t = (e^{\underline{B}t} - \underline{I})\underline{B}^{-1}\underline{Ax} + e^{\underline{B}t}\underline{y}_0 \qquad |\underline{B}| \neq 0$

(56b) $\hat{\underline{y}}_t = -\underline{B}^{-1}\underline{Ax} + e^{\underline{B}t}\underline{B}^{-1}\underline{Ax} + e^{\underline{B}t}\underline{y}_0 \qquad |\underline{B}| \neq 0$

These results hold if $\underline{B}$ is full rank; if $\underline{B}$ is not full rank:

(56c) $\hat{\underline{y}}_t = [(e^{\underline{B}t} - \underline{I})\underline{V}_1\underline{\Lambda}_1^{-1}\underline{V}^{(1)} + t\cdot\underline{V}_2\underline{V}^{(2)}]\underline{Ax} + e^{\underline{B}t}\underline{y}_0,$

$|\underline{B}| = 0$

(56d) $\hat{\underline{y}}_t = -(\underline{V}_1\underline{\Lambda}_1^{-1}\underline{V}^{(1)}\underline{Ax} + t\cdot\underline{V}_2\underline{V}^{(2)}\underline{Ax})$

$+ e^{\underline{B}t}(\underline{V}_1\underline{\Lambda}_1^{-1}\underline{V}^{(1)})\underline{Ax} + e^{\underline{B}t}\underline{y}_0, \quad |\underline{B}| = 0$

where, as before, $\underline{\Lambda}_1$ is the submatrix of $\underline{\Lambda}$ containing nonzero roots, $\underline{V}_1$, $\underline{V}^{(1)}$ are the submatrices of $\underline{V}$, $\underline{V}^{-1}$, respectively, associated with $\underline{\Lambda}_1$, and $\underline{V}_2$, $\underline{V}^{(2)}$ are the submatrices of $\underline{V}$, $\underline{V}^{-1}$, respectively, associated with the zero roots of $\underline{B}$.

It is clear from equations (56a) through (56d) that the time path of $\hat{\underline{y}}_t$ depends in an important way on the behavior of $e^{\underline{B}t}$ over time. If $e^{\underline{B}t}$ grows without bound, so does $\hat{\underline{y}}_t$, irrespective of whether $\underline{B}$ is full rank. If $e^{\underline{B}t}$ stabilizes or goes to zero over time, the predicted vector of career expectations $\hat{\underline{y}}_t$

stabilizes, if $\underline{B}$ is full rank ($|\underline{B}| \neq 0$). In general, $\underline{Y}_t$ grows linearly without bound if $\underline{B}$ is singular ($|\underline{B}| = 0$). This is clear by observing the term in (56d) which is linear in t, (i.e., $t \cdot (\underline{V}_2 \underline{V}^{(2)}) \underline{AX}$). If all eigenvalues and eigenvectors of $\underline{B}$ are real numbers, then the following matrix representation is useful:

$$e^{\underline{B}t} = \underline{V} e^{\underline{\Lambda} t} \underline{V}^{-1}$$

where Λ is the diagonal matrix with diagonal entries equal to eigenvalues of $\underline{B}$, and $\underline{V}$ is the corresponding matrix of eigenvectors. This result does not depend on Λ being real, but it does not yield so much insight when Λ is complex. When all eigenvalues (Λ) are real numbers, it is clear that $e^{\underline{B}t}$ stabilizes only if $\underline{\Lambda} \leq \underline{0}$.

For general matrices (real or complex eigenvalues), an alternative development is useful. For this development, the elements of $e^{\underline{B}t}$ are denoted by b^*_{ij}. A tedious algebraic argument leads to the following result.

$$(57) \quad b^*_{ij} = \sum_k \gamma_{ijk} e^{\lambda_{Rk} t} \sin(\lambda_{Ik} t + \delta_{ijk})$$

where

λ_k = the kth eigenvalue of $\underline{B}$ and, in general, is a complex number; $\lambda_k = \lambda_{Rk} + \lambda_{Ik} \iota$, with $\iota^2 \equiv -1$

$$\gamma_{ijk} = \sqrt{(v_{Rik} v_R^{kj} - v_{Iik} v_I^{kj})^2 + (v_{Iik} v_R^{kj} - v_{Rik} v_I^{kj})^2}$$

$$\delta_{ijk} = \omega\pi - \tan^{-1}(v_{Rik} v_R^{kj} - v_{Iik} v_I^{kj}) / (v_{Iik} v_R^{kj} + v_{Rik} v_I^{kj})$$

v_{Rik}, v_{Iik} = The i,kth real and imaginary parts, respectively, of the i,kth cell in $\underline{V}$ (right eigenvectors of $\underline{B}$)

v_R^{kj}, v_I^{kj} = The k,jth real and imaginary parts, respectively, of the k,jth cell in $\underline{V}^{-1}$ (left eigenvectors of $\underline{B}$)

ω = 1 if $v_{Iik} v_R^{kj} + v_{Rik} v_I^{kj} \geq 0$

0 otherwise

If $v_{Rik} v_R^{kj} - v_{Iik} v_I^{kj} = v_{Iik} v_R^{kj} + v_{Rik} v_I^{kj} = 0$, then $\delta_{ijk} \equiv -\lambda_{Ik} t$.

The formula given for δ_{ijk} does not define δ_{ijk} if $v_{Iik}v_R^{kj} + v_{Rik}v_I^{kj} = 0$, but δ_{ijk} goes to $\pm\Pi/2$ as this expression approaches zero. Hence, δ_{ijk} will be defined as $\pm\Pi$ if $v_{Iik}v_R^{kj} + v_{Rik}v_I^{kj} = 0$. Note that if the kth root of $\underline{B}$, λ_k, is real, then λ_{Ik}, and v_I^{kj} are all zero, and one finds

$$\left.\begin{array}{l} \gamma_{ijk} = v_{Rik}v_R^{kj} \\ \delta_{ijk} = \pm\Pi/2 \end{array}\right\} \text{for } \lambda_k \text{ a real number}$$

so that, for λ_k real one has $\gamma_{ijk}e^{\lambda_{Rk}t}\sin(\lambda_{Ik}t + \delta_{ikj}) = v_{Rik}\bar{v}_R^{kj}e^{\lambda_{Rk}t}$. If all roots of $\underline{B}$ are real, (57) specializes to

$$\text{(57a)} \quad b^*_{ij} = \sum_k v_{ik}e^{\lambda_k t}v^{kj}, \text{ if all } \lambda \text{ are real}$$

Observing (57a), it is clear that if the maximum λ_k (Λ real) is positive, every element of $e^{\underline{B}t} = \underline{B}^*(t)$ grows exponentially without bound. If the maximum λ_k is negative, each element of $e^{\underline{B}t}$ goes to zero by an exponential decay. If maximum λ_k is zero, the elements of $e^{\underline{B}t}$ stabilize over time. Referring to (57), these conclusions also apply if one or more λ_k is complex; here the critical variable is the real part of λ_k. Also, due to the sinusoidal function $\sin(\lambda_{Ik}t + \delta_{ijk})$, the entries of $e^{\underline{B}t}$ oscillate over time.

Combining these observations about the behavior of the elements of $e^{\underline{B}t}$ with equations (56) permit some conclusions regarding the time path of the career expectation variables:

1. If the real part of the largest eigenvalue of $\underline{B}$ (λ_R max) is negative, then from (56b) one sees that the equilibrium of $\underline{y}_t$ is

$$\hat{\underline{y}}_t(\text{equil.}) = -(\underline{B}^{-1}\underline{A})\underline{x}$$

For $\lambda_{R\ max} < 0$, therefore, individuals gradually revert to a set of career expectations determined by their status origins and mental ability. If $\lambda_{R\ max}$ is near zero, however, this may occur so slowly that, by adulthood, career expectations may still be far from equilibrium.

2. If the real part of the largest eigenvalue of $\underline{B}$ is positive, career expectations increase without bound. Again, if $\lambda_{R\ max}$ is close to zero, the growth may be slow in the short run.

3. If the largest eigenvalue of $\underline{B}$ is precisely zero, $\underline{B}$ is singular and one must reference equation (56c) or (56d). From either of these equations it is clear that for $\lambda_{R\ max} = 0$, career expectations, in general, grow linearly without bound.

4. If any of the roots (λ_k) of $\underline{B}$ are complex, an oscillation is introduced into the time path of career expectations. For example, the level of job status expectations waxes and wanes over time. The amplitude of the oscillations depends on the size of the real part of the eigenvalue; the larger the real part of the complex eigenvalues, the larger the amplitude.

5. Irrespective of whether $\underline{B}$ is full rank, if $\underline{Ax} = \underline{0}$, then the time path of $\hat{\underline{y}}_t$ is $e^{\underline{B}t}\underline{y}_o$. This means that the time path of career expectations does not depend on the status origins and mental ability of the individual. This is not a general conclusion, however. For some matrices $\underline{A}$ there may be no vector $\underline{x}$ except $\underline{x} = \underline{0}$ that makes $\underline{Ax} = 0$. In fact, $\underline{Ax} = \underline{0} \rightarrow \underline{x} = \underline{0}$ generally if there are more endogenous (career expectation) variables than exogenous variables (e.g., parental status variables). Since the first element of $\underline{x}$ is fixed at 1.0, $\underline{x}$ cannot be zero. If a nontrivial (nonzero) solution to $\underline{Ax} = \underline{0}$ exists, it means that there is at least one peculiar pattern of the status background and ability variables that frees the time path of career expectations from status background and ability. A particularly interesting case arises when the maximum λ_R is zero but its imaginary part is nonzero. In this instance, if $\underline{Ax} = \underline{0}$, then career expectations tend to a periodic function, i.e., the career expectations wax and wane

indefinitely with a constant amplitude between the high and low extremes. It should be emphasized, however, that this can happen only for individuals whose ability and status background follow the peculiar pattern satisfied by $\underline{Ax} = \underline{0}$, and the periodic pattern is impossible unless $\underline{A}$ is such that a nontrivial solution of $\underline{Ax} = \underline{0}$ exists.

Standardization Procedures

Since few of the variables used in career-decision making research are measured on a "natural" scale, i.e., one that is universally familiar, path analysis applications to career decisions frequently are presented in standardized form, to facilitate comparisons among coefficients associated with variables measured on different scales. Even when variables measured on "natural" scales are the subject of study, standardized coefficients can fascilitate interpretations, either as a supplement for the unstandardized coefficients or as the primary focus of attention (Wright, 1960). Consequently, it may be useful to present a standardization procedure for the differential equation system.

The standardization discussed in this section is confined to linear functions of single x and y variables. Note that the usual standardization to zero mean and unit variance is such a linear transformation, but we wish to study more general standardizations here.

If one makes a linear transformation that is constant over time on each x and each y variable to effect a standardization, the structure of the differential equation remains intact. To see this, define the following linear transformations on $\underline{x}$ and $\underline{y}$.

$$\underline{y}_1 = \underline{D}_y(\underline{y} + \underline{c}_y)$$

$$\underline{x}_1 = \underline{D}_x(\underline{x} + \underline{c}_x)$$

where $\underline{D}_y$ is an arbitrary diagonal matrix of order K, $\underline{D}_x$ is an arbitrary diagonal matrix of order L + 1, $\underline{c}_y$ and $\underline{c}_x$, are respectively, K x 1 and (L + 1) x 1 vectors; $\underline{D}_x$, $\underline{D}_y$, $\underline{c}_x$, and $\underline{c}_y$ are constant over time. In terms of the original variables, the differential equation is written

$$(62) \quad d\underline{y}/dt = \underline{Ax} + \underline{By} + \underline{u}$$

Now, since $\underline{D}_y$ and $\underline{c}_y$ are constant over time,

$$dy_1/dt = dD_y(y + c_y)/dt$$
$$= D_y dy/dt$$

Thus,

$$dy_1/dt = D_y Ax + D_y By + D_y u$$
$$= D_y AD_x^{-1} D_x A(x + c_x) - D_y Ac_x + D_y BD_y^{-1} D_y (y + c_y)$$
$$- D_y Bc_y + D_y u$$
$$dy_1/dt = -(D_y Bc_y + D_y Ac_x) + (D_y AD_x^{-1}) x_1 + (D_y BD_y^{-1}) y_1 + u_1$$

where $u_1 = D_y u$. Now, assume that the (1,1) element of D_x is unity and the first element of c_x is zero, so that the first element of x_1 is also unity. Now, let the first column of $D_y AD_x^{-1}$ be added to $-(D_y Bc_y + D_y Ac_x)$ and denote the resulting sum by α_0. Denote the remaining L columns of $D_y AD_x^{-1}$ by α_1, and form the supermatrix $A_1 = [\alpha_0, \alpha_1]$. The differential equation system for the standardized variables can now be written

$$(63) \quad dy_1/dt = A_1 x_1 + B_1 y_1 + u_1$$

with $B_1 \equiv D_y BD_y^{-1}$. This obviously has the same form as (62); hence, it is concluded that linear standardizations of single variables in the system preserve the structure of the system. It should be emphasized, however, that this argument does not justify standardization of all variables in the regression analysis to zero mean and unit variance. The t_o and t_1 endogenous variables are both used in the regression analysis. If both are standardized, the assumption that D_y and c_y are fixed over time is violated, since the mean and standard deviations of the y variables generally will shift over time, the D_y and c_y applied to y values at t_o would differ from those for t_1.

A judicious choice of the standardization constants D_y, c_y, D_x, and c_x is required in order to simplify interpretation of the standardized coefficients. For D_x and c_x, there is little difficulty. Since the mean and variance of the x variables are fixed in the analysis, D_x and c_x can be defined by the means and standard deviations of $\underline{x}$. Let the (1,1) element of $\underline{D}_x$ equal 1.0 and the remaining diagonal elements be defined by the reciprocal of the standard deviations of $\underline{x}$. Also, let the first element of $\underline{c}_x$ be zero and the remaining entries be defined by the negative of the means of the corresponding xs. Then $\underline{x}_1$ is the vector with element one equal to 1.0 and the remaining elements equal to standardized $\underline{x}$ variables with zero means and unit variances.

For the endogenous variables, choice of the standardization constants is somewhat more ambiguous than for the exogenous variables. For the exogenous variables, the strong precedent of standardizing to zero mean and unit variance was followed, but as noted above, this option is not available for the endogenous variables because it violates the presumption that $\underline{D}_y$ and $\underline{c}_y$ are constant over time. Nevertheless, a procedure analogous to standardization to zero means and unit variances might be used. One might pick the means and standard deviations of the y variables at a specified point in time, say T, and use means and standard deviations calculated from the y variables at the selected time point. The transformation constants would be $\underline{D}_y = \underline{S}_T^{-1}$, and $\underline{c}_y = -E(\underline{y}_T)$, where $\underline{S}_T$ is a diagonal matrix with diagonal elements equal to standard deviations of y at time T, and $E(\underline{y}_T)$ is the mean (expected value) over observations of $\underline{y}$ at time T.

The selection of T is arbitrary. If the system has a stable equilibrium, T may be set to infinity and the equilibrium means and standard deviations used to define the standardization constants. Interpretations could then be made in terms of movement toward the mean at equilibrium. If the system has no stable equilibrium, then one might set $T = t_o$ and interpret changes away from the mean at the starting point. Other options are, of course, defensible. One may set T to the senior year in high school and interpret changes toward the mean during the senior year.

As with path analysis, one should treat the standardization chiefly as a heuristic device -- certainly, the relative magnitudes of the standardized change coefficients ($\underline{A}_1$ $\underline{B}_1$) can be shifted at will by choice of the standardization constants. There is no clear reason other than precedent for using standard deviations rather than some other measure of dispersion such as the average deviation, empirical range, or permissible (theoretical) range of each variable. Similarly, choice of the mean rather than some other measure of central tendency is based primarily on precedent (see Hotchkiss [1976] for an extended discussion of this issue). The choice of $\underline{c}_x$, $\underline{c}_y$ affects only the intercept, but the selection of $\underline{D}_x$, $\underline{D}_y$ affects the relative magnitudes of the change coefficients associated with the variables.

Forecasting and Measures of Assciation

The integral equation (51) can be used to estimate the parameters of the differential equation, as described in the preceding chapter, or, once the coefficients have been estimated, the integral equation can be used to forecast or predict a value for each endogenous variable and each case, given initial observations on the vector of endogenous variables (y_o). Equation (56) is written explicitly as a forecasting equation. When OLS is applied to equation (51) a multiple correlation or its square (R-square) for each equation is a standard measure of association and is a normal part of the output of most multiple-regression computer programs. When the integral equation is used for forecasting, however, there is no standard measure of association. The purpose of this section is to discuss a generalized correlation that can be compared readily to correlations calculated from least-squares regression.

As preparation for defining a goodness-of-fit measure to assess the accuracy of forecasts, it is useful to review the interpretation of R-square as a proportional-reduction-of-error (PRE) measure. Let

$$y = p_o + \sum_{j=1}^{K} p_j z_j + v$$

where y is the dependent variable, the z_j are K independent variables, v is the error, and the ps are constants. The following formula for R-square offers considerable heuristic appeal:

$$(64) \quad R^2 = 1 - s_v^2/s_y^2$$

where R^2 denotes the square of the multiple correlation (R-square), and s_v^2, s_y^2 indicate the variance of v and of y, respectively. The OLS estimates of the ps insure that the mean of $v = y - \overset{*}{y}$ is zero, where $\overset{*}{y}$ is the value of y predicted from the regression equation. Hence, s_v^2 is a variance of the errors of prediction from the linear regression:

$s_v^2 = E(y - \overset{*}{y})^2$, where E denotes expected value.

The denominator of the ratio in equation (64) is the variance of the dependent variable:

$$s_y^2 = E(y - Ey)^2$$

This variance can be interpreted as a measure of error in the absence of information about the independent variables, where the mean of y (Ey) is used as a constant predicting the entire set of y values. Recall that, in fact, the mean of y is the best constant predictor of all ys, in the sense that the mean-square-error is minimized when the constant is the mean of the disbribution.[10]

With this background, it is clear that the ratio s_v^2/s_y^2 is a ratio of mean-square errors -- the numerator summarizing prediction errors from the regression equation, and the denominator summarizing prediction errors when all values of y are predicted to be the mean of y. Hence s_v^2/s_y^2 can be viewed as a PRE measure and, ipso factor, so can $R^2 = 1 - s_v^2/s_y^2$. When OLS estimates of the ps are used, the minimum R^2 is zero and its maximum is one.

A straightforward generalization of (64) provides a useful statistic for summarizing the accuracy of forecasts.

---

10. Let p_o be a constant over the y values and form the mean-square-error -- MSE = $E(y - p_o)^2$. Differentiating with respect to p_o, setting to zero and solving yields, $p_o = Ey$. The second derivative of MSE with respect to p_o is the positive constant, 2.0. Hence, a necessary and sufficient condition for a minimum is present when $p_o = E_y$. Alternatively, one might postulate $p_o = Ey + q$ = any constant, and develop the following argument:

$$E(y - Ey + q)^2 = E(y - Ey)^2 + q^2$$

Clearly, this expression is minimum if and only if $q = 0$, so that $p_o = Ey$.

Define

$$(65) \quad R^2_{Fc} = 1 - MSE/s^2_t$$

where R^2_{Fc} denotes an R-square for forecasts, MSE stands for mean-square error, and s^2_t is the variance of the predicted endogenous variable at time t:

$$MSE = E(y_t - \hat{y}_t)^2$$

$$s^2_t = E(y_t - Ey_t)^2.$$

where $\hat{y}_t$ is the forecasted value of the scalar y_t. It is important to distinguish $\hat{y}$, the forecasted y, from $\overset{*}{y}$, the post facto estimated value of y from a regression analysis. Note that, in general $E(y_t - \hat{y}_t) \neq 0$; therefore, MSE will not be a variance as it is in multiple regression.

The maximum value of R^2_{Fc} is 1.0 and occurs if and only if forecasts of y are correct for every observation. On the other hand, the minimum R^2_{Fc} is not zero; R^2_{Fc} may be negative (indicating R^2_{Fc} should be interpreted as an imaginary number). When used in conjunction with non OLS regression estimates, R^2 may also be negative, and it has been objected to on these grounds (Fox, 1979: 145, Baseman, 1962). The fact that R^2_{Fc} may be negative, however, is not a strong objection, since a negative R^2_{Fc} has a straightforward interpretation; the negative value indicates that more accurate estimates of all ys result when each y is estimated to be the mean of y than when the ys are forecast from the model.

Numerous alternatives to (65) are available (see Fox, 1979; Ostrom, 1978). Of those proposed, one seems particularly appealing. Ostrom (1968: 67) proposes that s^2_t in (65) be replaced by the mean-square of the differences between t_0 and t_1 values of y. The implicit hypothesis of this mean-square-error is that the y values don't change; a stable equilibrium has been reached. The resulting measure retains its PRE characteristic, since it is based on a ratio of a MSE due to the model to a MSE derived from a "naive" model. A more general definition of

R^2_{Fc} is therefore suggested (see Ostrom, 1978: 68);

$$(66) \quad R^2_{Fc} = 1 - \frac{MSE(M)}{MSE(N)}$$

Where MSE(M) is the mean-square error for the theoretical model and MSE(N) is the mean-square error for a "naive" model (naive is Ostrom's term). One important advantage of defining $MSE(N) = s^2_t$, as in (65), is that R^2_{Fc} is more readily compared to R-square than with other definitions of MSE(N); nevertheless, a variety of MSE(N) might prove useful, depending on the circumstances.

The mean-square errors in (66) might be replaced by average deviations. One advantage of so doing is that the error summaries are based on absolute values of errors, thus preserving the metric of the dependent variable rather than transforming that metric by squaring all errors.

Reliance on the bivariate correlation between y and $\hat{y}$ (Fox, 1979) to assess forecast accuracy should be used with caution. With OLS regression, this bivariate correlation is the same as the multiple correlation, but this equivalence does not generalize to R_{Fc}. The difficulty with the correlation between y and $\hat{y}$ is that it presumes that two regression constants, in addition to the parameters of the model, are utilized to make the predictions. Thus, systematic error in the forecasts could easily be masked by the correlation between y and $\hat{y}$. It is theoretically possible that the R^2_{Fc} is negative even when the bivariate correlation between y and $\hat{y}$ is high. If this fact is recognized, however, the correlation might be used in conjunction with R_{Fc} to assess the degree to which forecasting errors are systematic.

The distinction bewteen R-square (from OLS) and R_{Fc}-square calculated to assess the accuracy of forecasts is of fundamental importance in the assessment of theory. R-square assesses accuracy derived from a model for which the parameters are determined post facto according to the explicit criterion of maximixing R-square. R-square cannot be calculated until after the dependent variable is observed and incorporated into calculation of the regression coefficients. In contrast, R_{Fc}-square calculated from forecasts assesses prediction in the strict meaning of the term, because the forecasts are made prior to observing the endogenous variables at the time point for which predictions are targeted. The values of the dependent variable at the target date of the prediction do not enter into calculation of the prediction equation. One may conclude, therefore, that forecasts comprise a much stronger test of theory than regression studies. This interpretation is reflected by the range of R-quare due to OLS compared to the range of R_{Fc}-square from forecasts.

It should also be noted that most of these comments also apply to cross-validation studies. In fact, the most convincing evidence in support of theory may be derived from cross-validation studies in which parameters of a differential equation model (or analogous dynamic model) are estimated from one data set and forecasts are assessed on a different data set -- one collected <u>independently</u> from the data used to estimate the parameters. Certainly, little is gained by splitting at random a single sample, since discrepancies between the two halves (or several parts) must be due to sampling error alone and, therefore do not test the robustness of the model under variations due to the many detailed ways in which data-collection procedures may vary among data sets. In most cases, knowledge of sampling error is clearly stated in theoretical statistics; hence, little can be gained by a few observations of sampling error in a particular sample.

CHAPTER 6

SUMMARY AND COMMENTARY

This volume is about the application of differential equations to the development of career expectations. The starting point for the substantive work is path modeling of status attainment as contained in the sociology literature. The differential-equation model of career expectations may be viewed as a revision of path models of the same process. While status-attainment research in sociology provides the theoretical framework, important theoretical insights from other research traditions are referenced, in part to justify the differential equation model. More eclectic merging of ideas from a variety of sources is needed than is presented in this work, however.

The material in the preceding chapters divides conveniently into two main subtopics: (a) interpretations of the application of differential equations to the theory of career expectations, and (b) technical explication of a particular differential-equation model of career expectations.

Interpretation and Theory

Chapter 2 presents a justification for conceptualizing the process of developing career expectations as a simultaneous linear differential equation system. An example is presented which contains a set of two "exogenous" variables, parental status, and measured mental ability, and a set of five "endogenous" variables: school grades, significant-other educational expectation of ego, significant other occupational expectation of ego, ego's educational expectation for self, and ego's occupational expectation for self. The rates of change over time in the endogenous variables are hypothesized to be linear functions of the exogenous variables and of the current level of all the endogenous variables. The linear form of the model is viewed as an initial hypothesis, possibly to be modified later on the basis of theory and/or empirical evidence.

Three advantages accompany application of the differential-equation systems to describe developing career expectations. First, the continuous dynamic character of the development of career expectations is expressed by the differential equations; this feature of the differential equations is not shared by alternative models of career expectations. Secondly, all possible two-directional effects among the endogenous variables are included in the differential-equation model. While two-directional effects are not a unique feature of differential equations, current path models of career expectations geneally omit two-directional effects. Thirdly, since forecasts to any point along a continuous time scale are a natural aspect of the differential-equation model, strong tests of the model are encouraged. Observation of the exogenous variables on one occasion and each endogenous variable on two different occasions for each case provides enough data to estimate the parameters of the model. The estimated parameters can then be used to forecast a value for each endogenous variable and each person at any point in time. The accuracy of these forecasts offers a difficult test for the theory to pass, a test not available by observation of multiple correlations in a regression study.

The importance of using forecasts as tests of theory extends beyond the study of career expectations. Most social-science research is nonexperimental, that is, social scientists study natural systems of variables with little or no ability to manipulate important variables. Further, measurements of key concepts in social-science research often strain credibility. Thus, the barriers to scientific study of social phenomena are high. Consequently, credibility of findings from social research demands the strongest empirical tests that can be mustered.

While the linear differential-equation model of career expectations offers advantages over current models, there are important limitations of the differential-equation model which should be explicit. First, the differential equation model presumes continuous functions of time; hence, it is not readily generalizable to status attainments (e.g., occupation, education, income), because attainment variables exhibit abrupt shifts at isolated time points rather than continuous change. It is possible that career expectations manifest abrupt changes as well, although an apparently sudden change in expectation may be viewed as a continuous curve with short-radius turns. It is possible that abrupt changes can be modeled by a relatively new mathematical method called "catastrophe theory" (Zeeman, 1977).

A second limitation of the differential-equation model presented here is that it assumes a linear form with the linear coefficients constant over time. This assumption can be relaxed should experience warrant, but the technical features of estimating a general nonlinear model or a linear model with nonconstant coefficients have not been presented in the literature. A third limitation of the model is that it probably is far too simplistic to capture even a reasonable approximation of the complexities of forming career expectations. For example, the role of uncertainty in forming career expectations is ignored altogether in the model. Adding realistic complexities to the model while preserving some semblance of parsimony should be a primary aspect of the research agenda in the coming decade.

Chapter 5 addresses four aspects of interpreting linear differential equation systems in conjunction with the process of forming career expectations. First, it is noted that the "change coefficients" associated with the differential equations provide useful basis for interpreting effects of variables on each other. The change coefficients express the basic theory. Further, they exhibit the virtue of being constant with respect to the length of the time interval between measurements. In contrast, the cross-lagged path-coefficients (or regression coefficients) depend in a complex way on the length of the time interval between measurements. Secondly, the structure of the matrix ($\underline{B}$) of change coefficients associated with the endogenous variables determines the time-path of developing career expectations. If the largest real part of the characteristic roots (λs) of B is positive, career expectations increase exponentially. If the largest real part of the λs is zero or negative, and $\underline{B}$ is full rank, career expectations seek an equilibrium that depends on the background (exogenous) variables. If one or more of the λs is a complex number, career expectations oscillate over time; the amplitude of the oscillations expand, remain constant, or dampen over time, depending on whether the largest real part of the λs is positive, zero, or negative, respectively. If all λs are real, no oscillation occurs.

Chapter 5 also develops a standardization methodology for differential equation models that parallels standardization in path analysis. It should be noted that standardized path coefficients should not be used in the analysis because they artifically remove changes over time in means and variances from the data. The standardization methods developed in Chapter 5 retain changes in means and variances and fascilitate comparisons among coefficients associated with variables measured on different scales. One should be cautioned, however, not to overinterpret these coefficients. There chief value is heuristic.

The final topic in chapter 5 is assessment of forecast (prediction) accuracy. Several proportional-reduction-in error measures of strength of relationship are discussed. One particular measure involving the ratio of the mean-square error due to forecasts to the cross-sectional variance represents a generalization of the square of the multiple correlation (R-square); hence, it is readily comparable to R-squares. It is noted in the chapter that the generalized coefficient has no lower bound; it can assume negative values. This fact is not grounds for rejecting the measure, however, since negative values indicate useful information, viz, that the cross-sectional mean of the dependent variable is a more accurate estimate of all the values of the dependent variable than are the forecasts. It is concluded that the negative range of the generalized correlation reflects the fact that empirical tests based on forecasts offer more stringent tests of theory than do regression studies.

Technical Information

In addition to the interpretative contents of this volume, it also contains the technical information needed to apply linear differential equation systems with constant coefficients to the study of developing career expectations. Chapter 3 summarizes important concepts and theorems from mathematics that are used in the rest of the monograph. Coverage includes basic ideas in calculus, concepts and operations related to complex numbers, and aspects of linear algebra. The treatment is extremely brief and intuitive, the purpose being to summarize key concepts and theorems in a manner that is comprehensible to readers with little mathematical training. Chapter 4 presents development of the theory of linear simultaneous equation systems with constant coefficients, proposes a justification for viewing the statistical estimation of the model as a "reduced-form" system for which "ordinary least squares" (OLS) are appropriate, and describes use of a computer program which is available to carry out the calculations. In the program the exogenous variables and the time-zero endogenous variables are the predetermined variables and comprise the set of regressors. The time-one endogenous variables comprise the set of dependent variables. One OLS regression is calculated for each dependent variable using the same set of regressors for each regression equation. The resulting regression coefficients are inputs to the calculation of the parameters of the differential-equation model.

Commentary

A large part of the contents of this volume apply to topics other than development of career expectations. Any system of variables for which a linear differential equation system supplies a good initial hypothesis can be studied in the manner suggested in this volume. Examples include development of political preferences, racial prejudice, and prediction of migration. If two panels of data are available on a system of variables, the estimation techinique suggested in this volume may be applied.

The potential contribution of theory testing based on forecasts has been emphasized. It seems likely, however, that attempts to verify theory using even short-term forecasts will fail to yield convincing support of theory. Hopefully, such failures will stimulate imaginative revision of theory in which more nuances of the processes under study are incorporated into formal theoretical statements expressed in dynamic form.

BIBLIOGRAPHY

Alexander, Karl L., and Eckland, Bruce K. "Basic Attainment Processes: A Replication and Extension." Sociology of Education 48(1975):457-495.

Asher, Herbert B. Causal Modeling. Beverly Hills, CA: Sage Publications, Inc., 1976.

Basmann, R.L. "Letter to the Editor." Econometrica 30(1962): 824-826.

Blau, Peter M.; Gustad, N.W.; Jessor, R.; Parnes, Herbert S.; and Wilcox, R.C. "Occupational Choice: A Conceptual Framework." Industrial and Labor Relations Review 9 (1956):531-543.

Boardman, Anthony E., and Murnane, Richard J. "Using Panel Data to Improve Estimates of the Determinants of Educational Achievement." Sociology of Education 52(1979):113-128.

Box, G.E., and Jenkins, G.M. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1970.

Christ, C.F. Econometric Models and Methods. New York: John Wiley, 1966.

Coleman, James S. "The Mathematical Study of Change." In Methodology in Social Research, edited by Hubert M. Blalock, Jr., and Ann B. Blalock. New York: McGraw-Hill Book Company, 1968.

Curry, Evans W.; Picou, J. Steven; Hotchkiss, H. Lawrence; Scritchfield, Shirley A.; and Stahura, John M. Significant Other Influence and Career Decisions: Volume I. Black and White Male Urban Youth. Columbus, Ohio: The Ohio State University, The Center for Vocational Education, 1976.

Doreian, Patrick, and Hummon, Norman P. "Estimates for Differential Equation Models of Social Phenomena." In Sociological Methodology, edited by David R. Heise. San Francisco: Jossey-Bass, Inc., 1977.

Doreian, Patrick, and Hummon, Norman P. Modeling Social Processes 8. New York: Elsevier Scientific Publishing Company, Inc., 1976.

Doreian, Patrick, and Hummon, Norman P. "Models of Stratification Processes." Quality and Quantity 8(1974):327-345.

Duncan, Otis D.; Featherman, David L.; and Duncan, Beverly. Socioeconomic Background and Achievement. New York: Seminar Press, 1972.

Falk, William W.; Howard, William G.; Howell, Frank M.; McDermott, Virginia P.; Picou, J. Steven; and Thomas, John K. Education and Work in Rural America: The Social Context of Early Career Decision and Achievement, edited by Arthur G. Cosby and Ivan Charner. Houston, Texas: Stafford, Lowdon, 1978.

Fisher, Franklin M. The Identification Problem in Econometrics. Huntington, NY: Robert E. Kriger Publishing Company, 1976.

Fox, John. "Simultaneous Equation Models and Two-Stage Least Squares." In Sociological Methodology, edited by Karl F. Schuessler. San Francisco: Jossey-Bass, Inc., 1979.

Gasson, Ruth M.; Haller, Archibald O.; and Sewell, William H. Attitudes and Facilitation in the Attainment of Status. The Arnold M. and Caroline Rose Monograph Series, American Sociological Association, 1972.

Ginzberg, Eli; Ginzburg, Sol W.; Axelrad, Sidney; and Herma, John L. Occupational Choice: An Approach to a General Theory. New York: Columbia University Press, 1951.

Haller, Archibald O., and Woelfel, Joseph. "Significant Others, the Self-Reflexive Act and the Attitude Formation Process." American Sociological Review 36(1971):74-86.

Haller, Archibald O., and Woelfel, Joseph. "Reply to Land, Henry and Hummon." American Sociological Review 36(1971): 1102-1103.

Goldberger, Arthur S. Econometric Theory. New York: John Wiley, 1964.

Hannan, Michael T., and Young, Alice A. "Estimation in Panel Models: Results on Pooling Cross-Sections of Time Series." In Sociological Methodology, 1977, edited by David R. Heise. San Francisco: Jossey-Bass, 1977.

Hauser, Robert M. "Disaggregating a Social-Psychological Model of Educational Attainment." Social Science Research 1 (1972):159-188.

Heise, David R. "Causal Inference from Panel Data." In Sociological Methodology 1970, edited by Edgar F. Borgatta and George W. Bohrnstedt. San Francisco: Jossey-Bass, Inc., Publishers, 1970.

Henry, Neil W., and Hummon, Norman P. "An Example of Estimation Procedures in a Nonrecursive System." American Sociological Review 36(1971):1099-1102.

Hotchkiss, Lawrence. "A Technique for Comparing Path Models Between Subgroups: Standardized Path-Regression Coefficients." Sociological Methods and Research 5(1976):53-76.

Hout, Michael and Morgan, William R. "Race and Sex Variations in the Causes of the Expected Attainments of High School Seniors." American Journal of Sociology 81(1975):364-394.

Johnston, J. Econometric Methods. New York: McGraw-Hill, 1963.

Joreskog, Karl G. "A General Method for Estimating a Linear Structural Equation System." In Structural Equation Models in the Social Sciences, edited by Arthur S. Goldberger and Otis Dudley Duncan. New York: Seminar Press, Inc., 1973.

Kerckoff, Alan C. Educational, Familial, and Peer Group Influences on Occupational Achievement. Durham, NC: Duke University, 1971.

Land, Kenneth C. "'Significant Others, the Self-Reflexive Act and the Attitude Formation Process': A Reinterpretation." American Sociological Review 36(1971):1085-1098.

Malinvaud, E. Statistical Methods of Econometrics. Chicago: Rand McNally, 1966.

Merton, Robert K. Social Theory and Social Structure. Glencoe, Illinois: The Free Press and London: Collier-Macmillan Limited, 1957.

Nie, Norman H.; Hull, C. Hadlai; Jenkins, Jean G.; Steinbrenner, Karin; and Bent, Dale H. Statistical Package for the Social Sciences, 2nd ed. New York: McGraw-Hill, 1975.

Nolle, David B. "Alternative Path Analytic Models of Student-Teacher Influence: The Implications of Different Strokes for Different Folks." Sociology of Education 46(1973):1360-1383.

Ostrom, Charles W., Jr. Time Series Analysis: Regression Techniques 9, edited by Eric M. Uslaner. Beverly Hills, CA.: Sage Publications, Inc., 1978.

Picou, J. Steven; Curry, Evans W.; and Hotchkiss, H. Lawrence. Significant Other Influence, Career Choice and Achievement: Selected Theoretical and Conceptual Approaches. Columbus, OH: The Ohio State University, The Center for Vocational Education, 1976.

Pietrofesa, John J., and Splete, Howard. Career Development: Theory and Research. New York: Grune and Stratton, 1975.

Rehberg, Richard A., and Rosenthal, Evelyn R. Class and Merit in the American High School, edited by Don Adams. New York: Longman, Inc., 1976.

Rodgers, Roy H. "The Occupational Role of the Child: A Research Frontier in Development Conceptual Framework." Social Forces 45(1966):217-224.

Sewell, William H., and Hauser, Robert M. Education, Occupation and Earnings: Achievement in the Early Career. New York: Academic Press, 1975.

Super, Donald E.; Crites, John O.; Hummel, Raymond C.; Moser, Helen P.; Overstreet, Phoebe L.; and Warnath, Charles F. Vocational Development: A Framework for Research. New York: Teachers College Press of Columbia University, 1957.

Tiedeman, David V. "Decision and Vocational Development: A Paradigm and its Implications." Personnel and Guidance Journal 40(1961):15-21.

Wheaton, Blair; Muthin Bengt; Alwin, Duane F; and Summers, Gene. "Assessing Reliability and Stability in Panel Models." In Sociological Methodology, 1977, edited by David R. Heise. San Francisco: Jossey-Bass, 1977.

Wilson, Kenneth L., and Portes, Alejandro. "The Educational Attainment Process: Results from a National Sample." American Journal of Sociology 81(1975):343-363.

Wright, Sewell. "Path Coefficients and Path Regressions: Alternative or Complementary Concepts." Biometrica 16(1960): 189-202.

Zeeman, E.C. Catastrophe Theory: Selected Papers, 1972-1977. Reading, Mass: Addison-Wesley Publishing Company, 1977.