THE STRUCTURE OF INDIVIDUAL DECISIONS IN AMERICAN ELECTIONS:

THE INFLUENCE OF RELEVANT ALTERNATIVES

Thesis by

Kenneth Frank McCue

In Parital Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1984

(Submitted August 22, 1983)

Acknowledgements

It is traditional to thank individuals after the completion of a large piece of work, paying homage to both their wisdom and efforts on one's behalf. History will judge their wisdom, but I can summarize some of their efforts here. Mo Fiorina spent a good amount of time reading my earlier work and providing encouragement at a time when I needed it. John Ferejohn consented to argue on the subject of coattails with me and has even admitted on occasion that my way of thinking might have merit, though he still resists my arguments on far too many matters. Bruce Cain and Rod Kiewiet commented upon my drafts and provided considerable assistance with my style of presentation. Doug Rivers and Quang Vuong actually read in detail several of my more technical chapters and made voluminous (and valuable) comments, most of which I adopted into the text. Quang was also particularly helpful with his meticulous attention to my earlier drafts. Finally, Rebecca Rothenberg, while perhaps not offering academic advice, provided counsel and encouragement. Needless to say, any errors in this dissertation are my responsibility alone.

ABSTRACT

Studies in political behavior, when attempting to explain certain types of behavior, often concentrate upon those behaviors in the belief that other possible behaviors are irrelevant. This dissertation examines several forms of political behavior in American national elections with the intent of including or examing relevant alternatives which have an effect upon the primary behavior of interest. Three areas selected for examination are the question of political participation in American national elections (turnout), the influence of a race for one office upon the race for another office in American national elections (coattails), and the relationship between three forms of expression of political desire, one of which is the voting decision itself. Furthermore, an alternative model of describing decision making in human beings is discussed.

The results of these three examinations of political behavior may briefly be described as follows. With respect to turnout, by using a model which combines the voting decision with the decision to turnout, we show that the decline in turnout from 1960 to 1980 is strongly related to the way in which individuals translate their thoughts about politics into voting, rather than the changes which may have occured in their thoughts themselves. With respect to coattails, we posit that individuals tend to associate their vote for president and congressman in order to to overcome the separation of powers implicit in the American federal system, and provide strong evidence to support that hypothesis. With respect to the forms of expression of political desire, we show that the

variables known as feeling thermometers, the sums of open-ended evaluations of the candidates, and the voting decision itself are, in a well-defined and empirically verifiable manner, the same variable.

Contents

Acknowledgement		iii
Abstract		iv
List of Tables		viii
List of Figures		ix
Chapter I	Introduction	1
Chapter II	The Decline in Electoral Participation in Recent American Elections: A Reexamination	12
	The Analysis of Change over Time	13
	The Conditional Motivational Probit Model	21
	Selection of Data and Outline of Analysis	25
	A Comparision of Selected Elections from 1960 to 1980	34
	Conclusions	46
Chapter III	Presidential Coattails: A Structural Motivation	52
	A Simultaneous Decision System for Examining Coattail Voting	59
	A Test of Two Theories: the 1980 ANES	65
	The Coattails Effect and American Elections	86
Chapter IV	The Relationship of Thermometers and Open-ended Evaluations to the Vote	96
	Comparisions of Presther and Zcand with the Vote--the Zero-order Story	99
	The Definition of an Endogenous Proxy	107
	A Utility-maximizing Model of the Voting Decision	114
	Comparisions of Presther and Zcand with the Vote--a More Sophisticated Look	116
	Comparisions of Utility for Reagan and Carter	122
	An Analysis of the Utility Function of the Non-voter	127
	Comparisions of Thermometers with the Vote for the House and Senate (1978)	
	Conclusions	142

	Appendix I: The Derivation of Covariance Matrices of Estimators Derived by the Method of Maximum Likelihood	150
	Appendix II: The Wald Test for Similarity of Coefficients	166
Chapter V	A Model of Participatory Behavior: the Conditional Motivational Probit Model	169
	The Conditional Motivational Probit Model	170
	The Relationship Between Random Utility Maximization and the Conditional Motivational Probit	179
	Extensions to Simultaneous Equation Systems and "Economies to Scale"	184
	An Example of the Conditional Probit: Presidential Voting in 1980.	189
Chapter VI	A Simple Method for Constructing Consistent Estimates for the Covariance of Two Maximum Likelihood Estimators Obtained from the Same Set of Observations	196
	Derivation of Theoretical Results	196
	An Application to Amemiya's Principle	200
	An Application to Ordered Response Models	203
	An Application to Hausman Tests	210

List of Tables

Table 2.1	Point Estimates of the Relationship between Strength of Party Identification and Turnout, 1960-1980	20
Table 2.2	Conditional Motivational Probit Model for 1960, 1972,1976 and 1980	36
Table 2.3	Conditional Motivational Probit Model for 1960 and 1980, whites only	37
Table 2.4	Partisan Identification for Those Scoring Zero on CAND	43
Table 3.1	Momentum Model for Presidential Coattails	71
Table 3.2	Structural Model for Presidential Coattails	76
Table 3.3	Structural Model with Presidential Vote Broken Down by Party Coherence	80
Table 3.4	Structural Model with Presidential Charm	84
Table 4.1	Regression and Probit Equations for Vote, Presther, Zcand	120
Table 4.2	Regressions for Thermometers--1980	124
Table 4.3	Unrestricted and Restricted Sum of Square Residuals	132
Table 4.4	Regressions for Reagan Thermometers, Voter and Non-voters	134
Table 4.5	Regression and Probit Equations for Congressional Voting	137
Table 4.6	Regression and Probit Equations for Senatorial Voting	139

List of Figures

Figure 2.1	Turnout of 1960 White Weighted Sample Under Structural Coefficients from 1960, 1972, 1976, 1980	40
Figure 2.2	Distribution of CAND for White Samples, 1960 and 1980	44
Figure 3.1	Proportion of Respondents Voting for Carter (1980) by Presther	103
Figure 3.2	Proportion of Respondents Voting for Carter (1980) by Presther	104
Figure 3.3	Presther by Zcand	105
Figure 3.4	Standardized Zcand Minus Standardized Presther by Zcand	106
Figure 5.1	Hypothesized Cost Functions Associated with Underlying Motivation	173
Figure 5.2	Hypothesized Utility Functions for Random Utility Maximization	182

Chapter I: Introduction

The behavioral "revolution" in the social sciences has left behind it a legacy of conflict, and no more apparent is this conflict than in political science. It has been a long time since psychology students were expected to read Aquinas, but an understanding of Hobbes, Locke, and Rousseau and their theories of the state are expected to be juxtaposed with game theory in a political scientist's lexicon of professional knowledge. In practice, of course, no such thing occurs--those who understand the states of nature political philosophers presuppose are ignorant of the competitive solution, and *vice versa*. Nevertheless, such ignorance does not keep those of the various theoretical perspective from proclaiming the superiority of their approach and castigating alternative approaches.[1] "Never the 'twain shall meet" seems to be the order of the day, and what is of interest to one school will not be of interest to another. What then is the relevance of this thesis to any but the most dedicated behaviorist, since it is a work which many would say lies firmly in the behavioral tradition with which political science has such difficulties?

To be honest, I'm not sure. It depends largely upon the desires and goals of the non-behavioral reader. It has the advantages of a behavioral work--well developed hypotheses which are testable and falsifiable, statistical methodology which has its roots in recently developed methods of dealing with data which are of the type political scientists most

1. For an account of some of these phenomenon, see Parenti (1983).

commonly face, and communicability to other scientists.[2] Does it have the disadvantages: ignoring content and substance, addressing problems which are narrow, insignificant, and non-controversial? I will attempt in this chapter to defend the view that it does not. In doing so, I will define the interests I believe a non-behavioral reader would need in order to find this work of use to him.

The title of this thesis is *The Structure of Individual Decisions in American Elections: The Influence of Relevant Alternatives.* The primary concern is with modelling the individual's behavior in a manner which reflects the actual process of his decision making, and in doing so, taking into account alternatives usually discarded by political scientists when undertaking analyses. An example of this is the voting decision. The usual method of analysis is to look at one race for an office by itself and attempt to discern why the individual chose between two candidates running for that office. Two immediate extensions suggest themselves: first, to look at the reason for voting for either candidate for that office at all (the decision to turnout), and to look at the influence that races for other offices have upon the decision to vote for a candidate for that particular office.

The first extension, that of analyzing the addition of the decision to turnout to the decision of voting for candidates, is done by means of a

2. By this I mean simply that an astro-physicist can understand the methods and techniques that I use to test a hypothesis, even as I can understand his. Furthermore, he is capable of making judgment upon the methodology involved, even though the theories may be foreign to him, just as I am capable of understanding his methodology. Thus a standard technique of science which allows an easy exchange of new concepts and ideas is possible.

model which unifies the decision to turnout and the decision to vote in a parsimonious fashion, and does so with good predictive power. The model is known as the conditional motivational probit model. It is developed by postulating that individuals who face a situation in which they have the option of performing one of two opposite and mutually exclusive actions or of doing nothing have an underlying dimension of motivation which dictates the choice of their decision. Presumably, as the motivation for performing one action increases, the probability of an indivdiaul performing that action increases. Thus the term motivational in conditional motivational model. The term conditional comes from the fact that the model is estimated conditionally upon different costs (both economic and psychic) that the individual faces and upon different attitudes and personality traits that the individual possesses which may affect motivation. Finally, the term probit comes from the fact that if the estimation is made without the conditional part of the model, the estimation reduces to that of an n-chotomous probit (McKelvey and Zavoina (1975)). The full development of this model (which may be called an individual response model or a behavioral response model, since it takes the attributes of an individual and attempts to predict his behavior from those attributes), with justification and comparision with more traditional forms of individual response models (such as random utility maximization), is found in Chapter 5.

The substantive use to which this model is put, however, is found in Chapter 2, where it is used to address the puzzle of the decline of turnout. The focus of this investigation is the results from the Abramson and Aldrich piece in the September, 1982 *American Political Science Review*. There, the finding is that the decline in political efficacy and strength of partisanship over the last twenty years accounts for approximately 75 percent of the decline in white turnout for those years. With the use of the conditional motivational probit model, I conduct my own empirical analysis and conclude that the decline in political efficacy and strength of partisanship have had almost no effect upon the decline in turnout; rather, that there has been what I (and econometricians) refer to as structural change in the American citizen's decision process on how to vote and whether to vote. The exact nature of what I refer to as structural change has a technical meaning, but there is a simple way of explaining the nature of this change, one which is of interest both to the behavioral and non-behavioral reader.

I presume that each individual has a certain manner of organizing and acting upon the information, attitudes, and values which he possesses. I also presume that it is possible to reconstruct the individual's organization method from observed actions, and the method I use to do this is the model I have developed. What I do not presume, however, is that the individual's method of organizing the attributes he possesses stays constant over time. What I have referred to as structural change, then, merely refers to changes in the organization and action structure of the individual. His attitudes, information and whatnot are inputs to that process, whereas the process itself is heavily dependent upon the context of the society in which the individual is operating. If behavioral research is often said to ignore the context, here is an example for our non-behavioral reader where the presence of context can be detected by behavioral means. The context of the period 1960 to 1980, contains, of course, as Kousser (1983) has noted, "enough special circumstances to account for nearly any political phenomenon." Behavioral models do not operate in a vacuum, either, and, if properly specified and interpreted, can pick up the context of a decision as well as the process. Identifying exactly the manner in which the context of a decision can affect the process is a very difficult question; some work has recently been done in econometrics and the question is discussed briefly in Chapter 2. I do not think the objection, however, to behavioral research as ignoring context and emphasizing process is necessarily valid. Much research may exhibit such failings, but it is not a *sine qua non* on behavioral research.

The other extension to voting behavior deals with the addition to the analysis of an individual's voting behavior additional races, so that the effects of one race on another may be discerned. In national races, this is popularly known as a coattails effect, and it is usually assumed to occur because the personal attraction of the presidential candidate sways the voter to support the candidacy of other members of his party. If it is put that way, this is a very dangerous behavioral tendency for the American public to possess, as personal attraction may lead the crowd to support Fuehrers and other such individuals. Even in its less extreme forms, personal attraction is, to someone raised in the behavioral school with its emphasis on goal-oriented behavior, an odd motivational force. Suppose instead we posit that an individual has a behavioral tendency to overcome the separation of powers inherent in the American federal system by associating his vote for the various national offices?

In some ways this is a provocative thesis, for on the face of it it seemingly requires that individuals have a scholar's view of the American constitutional system. I do not think, however, that the informational content needed is all that much greater than that usually presupposed by a rational-choice perspective. After all, for policy voting, it is necessary that one believe that the individual running for an office have some ability to implement that policy. If one policy votes for President, and one policy votes for Congressman, then one would have some inkling that both candidates have some influence over whatever policies one is basing one's vote on, and that thus they would have to cooperate to some extent to implement this policy. The only question then remains the extent of policy voting in the American electorate, particularly for congressional races. This is an empirical question.

Another way it is thought to be a provocative question is that the question of justification for an individual associating his votes in this manner arises. This can be reduced to a form similar to another question, that of whether an individual votes rationally at all in national races, by the following method. Suppose an individual wants to choose two objects, each object coming from a different class. Say there are two classes, food and attire, so that the class of food includes, say, Mexican and French, and the class of attire, formal and informal. One would think there would be an interaction between the two choices, so that if one chose French food, one would be more likely to choose formal attire. Now suppose the individual had to vote, simultaneously, on both choices of food and attire. If we assume that there is no strategizing,[3] we would expect him to consider the interaction before voting. As the size of the group gets larger, we continue to assume that he acts as if his vote matters, just as we assume when analyzing a presidential or congressional race by itself.

In Chapter 3 I give technical meaning to the concepts of personal charm and a behavioral tendency to overcome the separation of powers. The empirical evidence I examine supports the hypothesis I have posited and rejects the personal charm hypothesis. Furthermore, further hypotheses from the hypothesis I have posited are generated and tested. This is once more an indication of the strength of the behavioral method, yet the results must be of interest to even the non-behaviorally inclined, for they indicate a source of strength and consolidation of power in a system of government which was deliberately designed not to allow such consolidation.

The last substantive essay deals with the choices an individual makes in a different context, not directly relevant to the election results, perhaps, but important to policital scientists. I refer here to the various

3. Strategic voting and an application is defined in Cain (1978).

intruments an individual is asked to respond to on the American National Election Studies (ANES). In particular, I ask the question "When are two variables the same variable?", for three very important variables to many election studies undertaken in the last thirty years: feeling thermometers, the open-ended evaluations of the candidates, and the voting decision itself. Each of these variables has been used in many different contexts and with many different claims as to their meaning. By the methodology I use (called random utility maximization in economics), I determine that each one of these variables is actually the same variable.

There is a sensible way of viewing the open-ended evaluations of the candidates, the vote, and feeling thermometers. An individual is sampled in several different ways about his attitudes and intentions towards the candidates, and at different times. While the instruments are somewhat different, all allow for a great deal of leeway in answering. Thermometers allow for a continous mark, CAND for an open-ended responding, and the vote for a one-zero binary decision on one's evaluation of the candidate. It is therefore not surprising that the individual displays a great deal of consistency in answering all three types of instruments.

The main advantage of this finding is that empirical researchers are now able to have a clear idea of what those variables mean when they attempt to make operational their conceptual models. The non-behavioral reader will not benefit directly from reading this piece, perhaps, but in the long run he should benefit from superior research generating from the behavioral side of the profession. Both behaviorists and non-behaviorists need to place their house in order before going naked to the world.

Chapter 6 is a chapter intended to bring an old technique to the attention of both political scientists and all other who use applied statistics. It primarily exposits on the distribution of two estimators derived by the method of maximum likelihood, demonstrates how simple it is to obtain consistent estimators of relevant variances and covariances, and demonstrates the theoretical usefulness of this method by developing a test of consistency for the n-chotomous probit model. The usefulness of this method is demonstrated by the fact that many of the statistical analyses in this dissertation were calculated using its results, saving myself many hours of programming time. It can save others many hours also.

This thesis represents the best that I could do in analyzing a number of problems of interest to the researcher who is behaviorally trained. The motivation for examing turnout was the previously mentioned Abramson and Aldrich article, for which I thought the results were artifactual and would not stand up under a more rigorous analysis. The motivation for examing the coattails phenomenon was my belief that the American voters are, to an amazing extent given the nature of their political system, able to judge their own interests and not allow the personal charm of a presidential candidate to overcome their interests. The motivation for studying the relationships between thermometers, the open-ended evaluations and the vote came from a remark in Mo Fiorina's

Retrospective Voting in American Elections, in which he explained his intuition about the nature of feeling thermometers and the vote. The motivation for the conditional motivational probit came from a desire to provide a unified theory of behavior which was closer to what I, trained in psychology, understood as actually occurring when a human being makes a decision, than economic theories such as random utility maximization. In all, I hope that the nature of the analysis here, which attempts to use methodological rigorous methods to explore important questions of political behavior, will be of some use to fellow researchers in political science.

Abramson, Paul R. and John H. Aldrich. 1982. "The Decline of Electoral Participation in American." *American Political Science Review*, 76, 3, 502-521.

Cain, Bruce E. 1978. Strategic Voting in Britain. *American Journal of Political Science*, 22, 639.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven, Conn: Yale University Press.

Kousser, Morgan J. 1983. "Suffrage and Political Participation." Social Science Working Paper 471, California Institute of Technology (forthcoming in the *Encyclopedia of American Political History*)

McKelvey, Richard and W. Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology*, 4, 103-120.

Parenti, Michael. 1983. "The State of the Discipline: One Interpretation of Everyone's Favorite Controversy." *PS*, 16, 2: 189-195.

Chapter II: The Decline in Electoral Participation in Recent American Elections: A Reexamination

The analysis of turnout in post World War Two American National Elections has consistently centered upon the presumed continuous decline[1] in participation from 1960 to 1980 (Ferejohn and Fiorina (1979), Abramson and Aldrich (1982), Shafer (1981), Brody (1978), Boyd (1981), Cassel and Hill (1981)). For a profession which cannot agree upon a theoretical structure for turnout,[2] it may seem somewhat presumptuous to examine the decline of participation, but scholarly attention continues to be riveted on what Kousser (1983) calls "the smallest and perhaps least permanent change [in turnout]--the decrease in non-southern participation since 1960." Aside from the natural interest of political behaviorists into what is, after all, the most common political act (Verba and Nie, 1972), a substantive reason often cited for interest in the decline of turnout is that American democracy is in some type of danger from non-participation.[3]

1. While it is clear that the percent of individuals of voting age voting in 1960 is greater than those in 1980, evidence will be presented later indicating that the assumption of a continuous decline may be in error.
2. A point we will return to later. It is crucial to any explanation of decline in participation which will withstand critical scrutiny. It should be noted that the lack of an accepted theoretical structure has not prevented a voluminous descriptive literature on the correlates of turnout. It is not that theoretical models have not been proposed (see Aldrich (1976), for a test of several models), it is more that the evidence is sufficiently indecisive to differentiate among models.
3. Abramson and Aldrich (p 502)state that the "dangers of electoral nonparticipation for American democracy have been pointed out by both journalists and scholars; in his 'crisis of confidence' speech, President Carter lamented that 'Two-thirds of our people do not even vote.'" The citing from that particular Carter speech is perhaps unfortunate, as some pundits (George Will in particular), have observed that the "crisis of confidence," reflected more Jimmy Carter's state of mind than any crisis of the American people.

We find the motivation from the standpoint of understanding political behavior sufficient, which is fortunate, for we remain unconvinced that a decline in participation in late twentieth century America has had any impact upon the nature of public policy. In this chapter, then, we will present a model which uses the concept of *motivation* to analyze turnout, and within the context of that model we will examine the decline of turnout in American National Elections, 1960 to 1980. In the next section we examine some elementary notions of how to measure change in the context of behavioral response models, and identify some errors which have been made in the literature with regards to that type of model. In light of that discussion, we then examine data from the 1960 and 1980 elections in an effort to understand why turnout declined between those two elections. Specifically, we present evidence rejecting the thesis that the decline in turnout is in any way related to the decline in partisan identification. We end by suggesting that the decline of turnout, as Kousser speculates, probably cannot be isolated from the "special circumstances" of 1960 to 1980.

2.1 The Analysis of Change over Time

Suppose we wish to examine the change in a variable over time. This variable (call it y^*) is assumed for all time periods to follow the behavioral relationship (which we will call an individual response model)

$$y_{it}^* = x_{it}\beta_t + u_{it}, \tag{2.1.1}$$

where the it indexes the variables and coefficients for the i^{th} case in the

t^{th} year. The usual interpretations apply: x_{it} is a vector of variables which are assumed to be uncorrelated with the disturbance term u_{it}, and β_t are vectors (conformable to x) which are the parameters in the problem to be estimated. The question of interest then becomes: if y^* is changing over time, what is the relation of that change to changes in the x and β?

A glance at equation (2.1.1) indicates there are two main types of change, if we assume $u_{it} = u$ for all time periods. The first type we will refer to as structural change and the second type we will refer to as distributional change. Considering the second type first, if we suppose $\beta_t = \beta$ for all time periods, it is obvious that the distribution of at least some of the variables making up the x vector must change. Structural change indicates that the weights β_t are different in some time periods than in other time periods, so that if r and t are different time periods, $\beta_r \neq \beta_t$. It is clear that both types of change may be taking place simultaneously.

Some researchers apparently do not believe in structural change, Abramson and Aldrich being a good example, for they state (p 504) "We are guided by two fundamental facts. First, political attitudes that have not changed in the aggregate cannot account for change in the behavior of the electorate." The acceptance of this "fact" implies that structural change cannot take place, only distributional change. The methodology we have outlined above is standard in econometrics, incidentally, going under such names as "tests of stability of statistical regimes" or "tests of structural change" (see Maddala (p 460)). Structural change can take place even though an exogenous variable (such as a political attitude) has not changed and can fully account for a change in behavior in the population under study (such as the electorate).[4]

Structural change in individual response mechanisms has not obtained the same type of attention that changes in the distribution of exogenous variables in an individual response mechanism have. One problem is, of course, that to study this type of change one has to first postulate an individual response mechanism.[5] Assuming that a mechanism has been specified, it is much simpler to analyze the change in a distribution of exogenous variables than it is to examine structural change, since the number of cases collected in any one time period t is typically much larger than the number of time periods.

A simple example will suffice to make this concept clear. Suppose a researcher in the year 2084 is interested in studying the effects of a presidential resignation on the structural weight of feelings of political efficacy in an individual response mechanism for turnout in an election. Suppose that the individual researcher posits a mechanism of the form of

4. We shall return to the Abramson and Aldrich article frequently, for several reasons. First, it is one of the latest articles to treat the subject of decline of turnout, and hence incorporates to a large extent the findings of previous researchers. Second, and more importantly, it is also a classic example of an approach which is completely oblivious to the difference between what we have termed behavioral response models and correlational analysis.

5. The most common type of individual response mechanism familar to political scientists is that generated by utility maximization. Given the existence of a utility function, an individual will make choices as if he were maximizing this function. Often this function is parametrized by certain weights, or what we have called the structural parameters of the individual response mechanism. Some work has been done in econometrics on structural change in individual response mechanism models, (see Heckman (1981)), but it is much too complicated to present here.

(2.1.1), and that he has data for elections from 1952 to 2080.[6] Let $\nabla\beta_t = \beta_t - \beta_{t-1}$. Then the researcher tests the equation

$$\nabla\beta_t = R_t + v_t,$$

where R_t is a dummy with a value of one if a president resigned in the last four years and zero otherwise. While this example is simplistic (presumably a researcher would use other methods to analyze structural change), it is clear that the analysis of structural change is a) possible and b) virtually unused in political science. For comparisions over time, however, structural change must be taken into account. Furthermore, the individual response mechanism must be correctly specified to make correct inferences, as the following example makes clear.

Suppose (2.1.1) is the true relationship between y^*_{it} and x_{it}, β_t and u_{it}, but under the impression that we do not need to explain all variation in y^*_{it} but only some part of it, we estimate

$$y^*_{it} = z_{it}\tau_t + v_{it} \qquad (2.1.2)$$

instead of (2.1.1). Then the relationship between the expectations of τ_t and β_t when (2.1.1) and (2.1.2) are estimated by ordinary least squares is well-known (see Maddala, 1981, p 461, for example), and is simply

$$E[\hat{\tau}_t] = A_t^{-1}B_t\beta,$$

where A is the matrix with entries made up of $Cov(z_{ikt}, z_{ilt})$, and B is the matrix with entries made up of $Cov(z_{ikt}, x_{ilt})$, where the double subscripts

6. Given the present level of commitment to funding of the American National Election Studies, a non-zero probability must be assigned to this event.

ik and il indicate particular elements of the z and x vectors.[7] A simple example with two variables will indicate the nature of the results we may expect with estimating (2.1.2) instead of (2.1.1). For two time periods $t = 1,2$, let $x_t = (x_{1t}, x_{2t})$ and $z_t = (x_{1t})$. Then we have

$$E[\hat{\tau}_1] = \beta_{11} + \frac{Cov(x_{11}, x_{12})}{Var(x_{12})}\beta_{12}$$

and

$$E[\hat{\tau}_2] = \beta_{21} + \frac{Cov(x_{21}, x_{22})}{Var(x_{22})}\beta_{22}.$$

Thus a change between $\hat{\tau}_1$ and $\hat{\tau}_2$ may occur because of several reasons. First, the β_{i1} may have changed (and this is presumably what the researcher who wants to examine only change in turnout would wish to claim). Second, the β_{22} may have changed. Third, $\frac{Cov(x_{i1}, x_{i2})}{Var(x_{i2})}$ may have also changed; that is, the relationship between the predictor variables may have changed. The lesson we may infer from this is clear: *If we are examining change over time it is important that we specify the model completely or offer convincing evidence that the omitted variables are independent of the variables in the estimation.*

Unfortunately, researchers such as Abramson and Aldrich do not heed this lesson. Specifically, they state (footnote 20)

7. We are assuming that the vectors $(y^*_{it}, x_{it}, z_{it}, u_{it})$ are independently and identically distributed for all observations and all i, a common assumption when data is obtained by sampling instead of experimental control (see White, (1982)). If the x and z are assumed to be non-random, the matrices of population moments would replace the matrices of covariances.

> In examining these trends over time, we do not present change among each efficacy level with controls for level of education. Cassel (1981), in a discussion an earlier version of this article (Abramson and Aldrich (1981)), argues that by not including level of education in our analysis we have not specified our model adequately. We disagree. Including education would be necessary if our goal were to explain all variations in turnout during the postwar years, but our goal is to account for the *decline* [their emphasis] of turnout.

But as we have seen in the above simple example, one cannot drop variables and then justify that procedure as an analysis of change rather than all variation. Unless Abramson and Aldrich sincerely believe that turnout is simply a function of partisan identification and efficacy (and that footnote indicates the contrary), they will not be able to sort out the effects of distributional change of one particular variable from, among other things, distributional change in other variables, a change in structural weights, or a change in the population covariances between two set of variables.

As an indication of how these changes can be confounded, consider Abramson and Aldrich's Table 2 (p 509). They fit the equivalent of the equation (their equation (3))

$$Pr(Y_{it} = 1) = \Phi(\mu_t + \beta_t P_{it})$$

to all white respondents for each election year between 1960 and 1980. Here, Y_{it} is one if the i^{th} respondent in the t^{th} year voted, zero if he did not, μ_t is the constant, P_{it} is a variable coded two if the i^{th} respondent in the t^{th} year is a strong identifier with some party, one if he identifies weakly with a party or leans towards a party, and 0 if he is an independent. β_t is then the structural or behavioral weight of the effects of partisanship upon turnout. Their results for this equation are presented in Table 2.1.

Table 2.1 Probit Estimates of the Relationship between Strength of Party Identification and Turnout, 1960-1980 (taken from Abramson and Aldrich, Table 2)

Constant in Election Year	
μ_{60}	.670**
μ_{64}	.385*
μ_{68}	.366*
μ_{72}	.127**
μ_{76}	.119**
μ_{80}	.113**
Slope in year	
β_{60}	.240**
β_{64}	.368
β_{68}	.363
β_{72}	.442*
β_{76}	.481**
β_{80}	.478**
-2*LLRatio	388.8 (χ^2_{11})**
n	9386

*$p < .05$. **$p < .01$.
(Significance measured for μ's from μ_{60} for μ_{64} to μ_{80} and for β's from β_{60} for β_{64} to β_{80})

As Abramson and Aldrich put it (p 509), "the strength of partisan affiliation is strongly and positively related to turnout in each election, and this relationship has grown over time." Examining Table 2.1, though, can any political scientist really believe that partisan identification was twice as important for turnout in 1980 as in 1960 (a coefficient of .478 as opposed to .240)? If we accept these structural weights as summaries of the behavioral propensities of certain sub-groups in the population to engage in a behavior, this is basically what is being claimed here. Another interpretation, aside from the claim of an increasing strength of relationship between partisan identification and turnout, is that a third variable, correlated with both the constant term and partisan identification, has been changing over time, and that this has produced the change in the structural coefficients. With the model we will introduce in the next section, we will see that the strength of partisan identification and its relationship to turnout, in the model we use, has undergone almost no change.

2.2 The Conditional Motivational Probit Model

Consider the individual who is faced with the following situation. First, there are two sides or choices of actions available to the individual associated with this situation, with the two actions being opposed to one another. Furthermore, the individual's motivation to engage in one action or the other lies along an underlying dimension of motivation to help one side or another, with neutrality lying at the zero point of motivation. Also, there are thresholds that must be passed by each individual if his underlying motivation is to be translated into action.

An example of this would be a civil war, in which the diametrically opposite actions are fighting with the rebels or fighting with the government forces. An individual would lie along the underlying dimension, either towards the side of joining the rebels or joining the government forces. The thresholds that the individual would have to pass indicates the point that motivation is turned into action. It is presumed that this threshold effect occurs because at the point where action is chosen over no action the cost to the individual (and here cost is used in a broad sense, encompassing such things as psychic cost), between choosing the action and not choosing the action is the same.

For our general exposition, then, we assume there exists an underlying dimension of motivation, and that an individual's motivation, y^*, is distributed along this dimension. We assume that the form of the relationship between y^*, the individual's motivation, and a set of exogenous variables x and their weights β may be written as

$$y^* = x\beta + u,$$

where

$$\beta = \beta(z, \tau),$$

where τ is some set of weights indexing z. The set z can be thought of as variables which have an effect on an individual's motivation, by either depressing or increasing the effects of the exogenous variables x. Such attributes as feelings or personality states would be prime candidates for such variables. In the case which we will use for our empirical estimation, we will let

$$\beta(z,\tau) = (z\tau)\beta.$$

The rule for motivation being translated into behavior is as follows. We assume there are three possible actions; A, B and do nothing. If the underlying level of motivation y^* is greater than some point α_2, A is performed, if it is less than α_1, B is performed, and if it is at or between those two points, nothing is done. It is also presumed that the subset of z affecting α_1 and α_2 is w, indexed by γ, so that

$$\alpha_1 = \alpha_1[w\gamma]$$

and

$$\alpha_2 = \alpha_2[w\gamma]\,,$$

or, in the usual case of a linear effect,

$$\alpha_1 = (w\gamma)\alpha_1$$

and

$$\alpha_2 = (w\gamma)\alpha_2.$$

The interpretation of the α's may be made in terms of cost–such an interpretation is made in Chapter 5.

The likelihood function for the conditional motivational probit model is easily written down, under our assumptions. There are three categories where an individual might be classified.[8] For ease of notation,

8. There may be more than three categories, and the extension is immediate and obvious. For example, in our civil war example, there could be five categories, ordered as follows: fight with the rebels, provide material aid to the rebels, do nothing, provide material aid to the government, fight with the government. The concept of ordering might well be con-

let $\alpha_3 = +\infty$ and $\alpha_0 = -\infty$. Let $y = 1$ if action A is taken, $y = 2$ if no action is taken, and $y = 3$ if action B is taken. Let F_σ be the cumulative distribution function of the error term u (we assume F has a continuous distribution). Then the probability of any action y, for $y = 1,2,3$, is simply

$$Pr\ (\ y = c\ |x,z,\beta,\tau\ ) = Pr\ (\ \alpha_c > y > \alpha_{c-1}\ |x,z,\beta,\tau\ ),$$

$$= Pr\ [\ \alpha_c(w,\gamma) > x\beta(z,\tau) + u$$

$$> \alpha_{c-1}(w,\gamma)\ ],$$

$$= Pr\ [\ x\beta(z,\tau) + u > \alpha_{c-1}(w,\gamma)\ ]$$

$$- Pr\ [\ x\beta(z,\tau) + u > \alpha_c(w,\gamma)\ ],$$

$$= F_\sigma[\ \alpha_c(w,\gamma) - x\beta(z,\tau)\ ]$$

$$- F_\sigma[\ \alpha_{c-1}(w,\gamma) - x\beta(z,\tau)\ ]. \qquad (2.2.1)$$

We will assume the simpler linear interactive specification for the effects of the z_i's, as given above. In this case, then,

$$Pr\ (\ y = c\ |x,z,\beta,\tau\ ) = F_\sigma[\ (w\gamma)\alpha_c - (z\tau)x\beta\ ]$$

$$- F_\sigma(\ (w\gamma)\alpha_{c-1} - (z\tau)x\beta\ )$$

Let

$$Z_{jc} = 1,\ \ y = c$$

sidered to exist here, and be conditional upon the same variables as the three category model.

$$= 0, \quad y \neq c$$

The likelihood may thus be written as

$$L(y, x, z, \beta, \tau) = \prod_{j=1}^{n} \prod_{c=1}^{3}$$

$$\Big\{ F_\sigma[\,(w\gamma)\alpha_c - (z\tau)x\beta\,]$$

$$- F_\sigma[\,(w\gamma)\alpha_{c-1} - (z\tau)x\beta\,] \Big\}^{Z_{jc}}.$$

The log of the likelihood function is then

$$\ell(y, x, z, \beta, \tau) = \sum_{j=1}^{n} \sum_{c=1}^{3} Z_{jc}$$

$$\ln \Big\{ F_\sigma[\,(w\gamma)\alpha_c - (z\tau)x\beta\,]$$

$$- F_\sigma[\,(w\gamma)\alpha_{c-1} - (z\tau)x\beta\,] \Big\}. \qquad (2.2.2)$$

Due to the multiplicative nature of the $z\tau$ upon the x's and the $w\gamma$ upon the α's, we put a constant one, with weight one, in both the z's and the w's.

2.3 Selection of Data and Outline of Analysis

Most scholars who have attempted to analyze the decline in turnout in American national elections have used the American National Election Studies (ANES) from the University of Michigan Survey Research Center (SRC). The main reason for doing this, of course, is that the ANES provide the only source for examining the attitudinal correlates of turnout at the individual level, and implicit in most political behavior research is the presumption that a change in attitudes (particularly the increase in cynicism towards and disillusionment in government and the decline in partisanship) have resulted in the decline in turnout.[9] In order to address the same issues which other researchers have, we will also use the ANES.

Four types of variables need to be specified for the motivational probit model which we outlined in the proceeding section. The first variable we need to specify is the dependent variable. For our motivational model, we need two actions which lie on opposite ends of a motivational continuum, with no action lying between the thresholds which motivation needs to surpass for any action to take place. The obvious actions, then, are a Republican vote at the negative end of the continuum of motivation and a Democratic vote at the positive end of the continuum, with no action lying between those two actions. Theoretically, we should have an axis of motivation for each particular office for which the individual is casting a ballot, and an approach for this method is worked out in Chapter 5.[10] In this examination, though, we will use the

9. One exception is the Ferejohn and Fiorina study (1979), which concludes that the decline in turnout closely tracks the decline in concern ("Generally speaking, would you say that ***you personally*** care a good deal which party wins the presidential election this fall, or that you don't care very much which party wins?"). Still, it can be claimed (see Abramson and Aldrich, p 519), that the concern variable and partisanship are closely intertwined, and hence the true decline is in attitudinal variables. While our analysis does not look at concern ***per se***, it looks at another variable which one would expect to play much the same role, the variable we will call CAND.

10. The main difficulty with estimating multiple offices is that treatment of the cost (which is represented in our model by the α's) becomes quite complicated, as the voter who turns out to vote for one office suddenly finds it much cheaper to vote for another office. Our use of the presidential vote only amounts to an assumption that an individual's vote for president represents the most important reason for his decision to turn out, and that the model is not biased by the motivations for other races.

presidential vote as the object of motivation which produces turnout.

The next set of variables which we need to specify are the variables which produce motivation. We use an adaptation of the decision rule first postulated by Kelly and Mirer (1974), utilizing the sum of the open-ended questions as to the likes and dislikes (or reasons for voting for or against) the presidential candidates to create a variable called CAND. The CAND variable was first used in Campbell, Gurin, and Miller (1954). It is constructed by using responses to the question "Is there anything in particular about (candidate) for president) that might make you want to vote (for/against) him?". The number of responses (there may be up to five in all SRC's survey's during an on-year except 1972, when there were three) for liking the Democratic candidate are added to the number of responses for disliking the Republican candidate, and then the number for disliking the Democratic candidate and liking the Republican candidate are added together and this number subtract from the first number. The number may thus run from -10 to 10.

Having obtained CAND, we then code the seven-point partisan identification scale into seven zero-one dummy variables (called strong Democrat, weak Democrat, independent Democrat, independent, independent Republican, weak Republican, strong Republican), with strong Democrat being a one if the individual indicates (through the ANES questions) that his partisan affiliation is strongly Democratic (on the seven-point scale), and if *the value of CAND is zero* (the same is done for the other six dummy variables also). The reasons for doing this are several-fold. First, as Kelly and Mirer suggested, an excellent pattern of prediction is achieved if one looks at the sign of the CAND variable and, if that is zero, then use the individuals partisan affiliation to break ties. The second reason is explicated in some detail in chapter 4 of this dissertation, and that is the observation that CAND and the vote are, in some sense, the same variable.[11]

Of course, CAND is obtained prior to the election, whereas one's vote choice is obtained after the election. Thus, CAND is causally antecedent to the vote, and should not be correlated with the disturbance term for the vote. A sensible way of viewing CAND and the vote (and feeling thermometers, for that matter--see Chapter 4), is that an individual is sampled in several different ways about his attitudes and intentions towards the candidates, and at different times. While the instruments are somewhat different, all allow for a great deal of leeway in answering. Thermometers allow for a continous mark, CAND for an open-ended responding, and the vote for a one-zero binary decision on one's evaluation of the candidate. It is not surprising that the individual displays a great deal of consistency in answering all three types of instruments.

The reason for including CAND, then, is that as a casually antecedent variable it satisfies the statistical requirements of our model and as the

11. The technical result from chapter 4 is that if, as in a random utility maximization model, the vote is assumed to represent an underlying variable which represents the difference in utility in voting for two candidates, the continous variable represented by CAND and the underlying variable for the vote are produced by the same random utility maximization model. Thus, CAND and the underlying variable which is presumed to generate the vote are the same variable under this framework.

sum of open-ended responses it allows a determination of how the distribution of reasons for voting for or against the presidential candidates has changed. Presumably CAND is generated in part by partisanship, yet it must also be generated in part by the individual's assessment of reality (which is no doubt influenced by his partisan commitment). We know partisanship has declined, but the distribution of the CAND variable need not have changed, for instead of partisanship being the impetus behind the CAND variable, a concern over issues (as postulated by many writers, most specifically Nie, Verba and Petrocik (1976)) may have replaced it. On the other hand, the decline of partisanship may have resulted in a decline of the CAND variable, and we may also test for that. The other measure of partisanship we have affecting motivation are the dummy variables, and effects on the decline in partisanship may be also be obtained from them.

The next set of variables that we need to specify are the variables (which we have referred to as z) modifying the variables (the x's) which are assumed to create the underlying motivation. These are variables which are assumed to affect all motivational variables. The one clear choice, since it is one of the main two choices of Abramson and Aldrich and we wish to test their analysis under a model which we feel is more correctly specified, is the variable derived from the two questions which are sometimes referred to as external political efficiency.[12] Two one-zero dummies were created out of these two questions, with the variable low

12. The two questions are "I don't think public officials care much what people like me think" and "People like me don't have much say about what the government does."

efficacy being coded as a one if the individual responded with a low efficacy response (agreement is a low efficacy response) scored on both questions and the variable medium efficacy being coded as a one if the individual responded with a low efficacy response on one question and a high efficacy question on another variable. The reference category was thus high efficacy responses on both questions, or a high efficacy person.

The last set of variables we need to specify are the variables which modify the thresholds at which motivation is translated into action. The obvious choice for this is the variable which has the greatest zero-order correlation with voting, the variable derived from the questions which make up attitude toward the voting process.[13] The usual (see Campbell, Converse, Miller and Stokes (1960), p 105) four-point scale was created from these questions (zero being the highest duty and four being the lowest), and dummy variables were created. High duty was coded one if an individual had a score of one on the duty scale, medium duty was coded one if an individual had a score of two on the duty score, and low duty was coded one if an individual had a score of three or four on the duty scale (due to the small number of individuals having scores of three or four). The reference category was thus those individuals having very high duty. The rational behind using duty is that it would seem that its operation embodies psychic cost--an individual who has a strong sense of

13. The questions used are 1) "It isn't so important to vote when you know your party doesn't have a chance," 2) "So many other people vote in the national elections that it doesn't matter much to me whether I vote or not," 3) "If a person doesn't care how an election comes out he shouldn't vote in it," 4) "A good many local elections aren't important enough to bother with."

"citizen duty" pays a high cost for not turning out in an election, whereas one with a weak sense pays little cost.

These are the variables, then, which are used to analyze turnout for each year, being estimated according to equation (2.2.1). There remains the matter of estimating the amount of change which has occurred from year to year, and separating out what we have called structural and distributional change. There are several ways to do this. First, to separate out structural from distributional change, we may do the following. We may take the distributions of exogenous variables from one year, say 1960, and place them in the conditional motivational probit from another year, say 1980. We then obtain predictions for turnout for each individual voter in 1960 based on the conditional motivational probit weights for 1980.[14] The difference in these two predictions is the amount of structural change which has taken place (which can either be positive or negative). The remaining amount of change is distributional, and it may be calculated as follows. First, we may take the entire distribution from one year (say 1980), and then change the distribution of one variable from that distribution to that of another year (such as 1960). We then calculate the predicted actions of the individuals with that distribution, and compare it with the old one. In that manner, we can obtain estimates for each individual variable for distributional

14. These predictions are made using the maximum likelihood estimators obtained from maximizing equation (2.2.2) and then using them in equation (2.2.1), which gives the probability of any particular individual with a certain set of (x, z) voting for a Democratic presidential candidate, a Republican presidential candidate, or not voting. Two methods are then available for making predictions. The first is classifying the individual according to the action with the maximum probability. The second is averaging the probability for each action over all groups. We will use both methods.

change, under the new structural regime.

A word might be made here about the method of prediction used by Abramson and Aldrich. They basically estimate three different models for examining the effect of partisanship on turnout.[15]

$$Pr(Y_{it} = 1) = \Phi(\mu + \beta P_{it}) \quad (2.3.1)$$

$$Pr(Y_{it} = 1) = \Phi(\mu_t + \beta P_{it}) \quad (2.3.2)$$

$$Pr(Y_{it} = 1) = \Phi(\mu_t + \beta_t P_{it}) \quad (2.3.3)$$

Here Y_{it} is once again one if the i^{th} individual in the t^{th} year voted (while P_{it} is his partisan strength), while μ_t and β_t are year specific constants, with β_t constrained to be β in (2.3.2) for all years and both β_t and μ_t constrained to be β and μ for all years in (2.3.1). The simplicity of even the most general model for relating partisanship and turnout has its drawbacks: to quote Abramson and Aldrich (footnote 15).

> Since the estimates reported here predict that citizens in all categories of the independent variable in each election year have a higher probability of voting than abstaining, all are predicted to vote. Thus, such measures as the percentage of case predicted correctly are irrelevant.

We will not comment on the irrelevancy of the percentage of cases predicted correctly as a means of measuring change between years, but rather will look at the procedure they use. Their choice for predicting the proportion of cases is to use the proportion predicted in each category of

15. And similar models for efficacy and a set of similar models for both together, but since our point can be made with this one set of models we restrict ourselves to it.

the independent variable for equation (2.3.1) times the number of observations in that category. Equation (2.3.1), when using their estimates of the β, gives proportions of 60 percent of the independents vote, 75 percent of the weak or independent partisans vote, and 86 percent of the strong partisans vote. From putting in the distributions from each year, they graph predicted turnout against actual reported turnout from the ANES from 1960 to 1980, and (in their Figure 1), obtain a drop in predicted turnout of only 2.3 percent. As actual reported turnout drops 10.2 percent between 1960 and 1980, this implies to them that "approximately 30 percent of the drop in voting can be attributed to the decline in P_{it}."

Taking their method one step further, however, suppose we use model (2.3.3) to obtain our predictions from, a model which Abramson and Aldrich state (footnote 17) is a significant improvement over model (2.3.1) (a x^2 of 68.0 ($\alpha > 0.999$), by Aldrich's test). If we use (2.3.3), and use Abramson and Aldrich's proportion method, we find that predicted turnout exactly matches actual turnout, so that by their methodology, their best model explains exactly 100 percent of the decline in turnout. Naturally, there is a problem, since if the counterpart of (2.3.3) for efficacy is used, it too predicts the decline in turnout 100 percent. The only justification Abramson and Aldrich give for using equation (2.3.1) "Equation (1) [(2.3.1)] is the most direct test of the hypothesis." Perhaps it is. It does seem strange, however, to use a method of prediction, which, when used with a statistically superior model, gives results which are nonsense.

2.4 A Comparision of Selected Elections from 1960 through 1980

Given the basic set of variables discussed in the last section, we estimated the equation (2.2.1) using the ANES data for each of the years 1960, 1972, 1976 and 1980. The years 1964 and 1968 were omitted, as the questions used to derive the duty scale were not asked in those two years. The method of estimation was by maximum likelihood; a modified Newton-Ralphson algorithm was used.[16] All analyses on the conditional motivational probit models were done with unweighted data. Cases were excluded on the basis of missing data on the duty, efficacy, open-ended questions regarding likes and dislikes of the presidential candidates, partisan affiliation, and missing information on votes for presidential candidate (which could include refusals to give answers to who the respondent voted for). "Don't knows" on duty and efficacy were coded as providing a low duty or low efficacious response. Voters for third party candidates were also excluded, including voters for Anderson in 1980.

16. It is possible to think of the procedure as running a probit of a dependent variable on a vector of exogenous variables x, and then modifying this probit by a set of exogenous variables z. For 1980, a probit program was run on the x variables, and upon convergence of that procedure (which is well-behaved—see Pratt (1981), for a proof of the concavity of the log likelihood of an n-chotomous probit), the Newton-Ralphson algorithm was expanded to both the x and z variables. This then converged in several more iterations. For 1960, 1972 and 1976, the Newton-Ralphson procedure on both the x and z variables was run starting from the 1980 coefficients. Convergence was achieved in several iterations for both 1972 and 1976, but for 1972 the algorithm "blew up." The 1972 program was rerun with slightly different coefficients and this time it converged in several iterations. The problem seems to be on an unconstrained Newton-Ralphson the coefficients can become too large for the program to handle with numerical integrity. In general, several of the coefficients will become much larger in magnitude than their final value, but usually the numerical accuracy of the computer is not exceeded and they return to a reasonable size after one errant iteration. If the numerical accuracy is exceeded, then the program can "blow up".

The results of the analyses are given in Table 2.2. The cases for analyses include both whites and blacks, as we have no *a priori* reason to expect blacks and whites to have different individual response mechanisms. As a check on this assumption, however, analyses for whites only were run for 1960 and 1980 and those results are presented in Table 2.3. As can be seen from perusing both tables, there is essentially no structural difference between the two groups.

Table 2.2 Conditional Motivational Probit Model for 1960, 1972, 1976 and 1980

Variable Name	1960	1972	1976	1980
α_1	-0.2863***	-0.3256***	-0.3115***	-0.3650***
	(0.0397)	(0.0317)	(0.0284)	(0.0361)
α_2	0.1748***	0.3434***	0.3152***	0.4594***
	(0.0346)	(0.0329)	(0.0302)	(0.0471)
Strong Democrat	0.5119*	0.5518***	0.8659***	0.8417**
	(0.2419)	(0.1196)	(0.2724)	(0.2953)
Weak Democrat	0.6824***	0.0062	0.4827***	0.3569
	(0.1974)	(0.0866)	(0.1564)	(0.2012)
Independent Democrat	0.4355	0.2979*	0.8121**	0.3589
	(0.3461)	(0.1213)	(0.2877)	(0.3074)
Independent	0.3032	-0.2394**	-0.0253	0.0989
	(0.3492)	(0.1012)	(0.1793)	(0.2635)
Independent Republican	0.5199	-0.9896***	-1.2935***	-1.3254***
	(0.4922)	(0.1370)	(0.3401)	(0.3209)
Weak Republican	-0.3129	-1.0263***	-0.6605***	-0.9754***
	(0.3827)	(0.1191)	(0.2010)	(0.2969)
Strong Republican	-1.8175***	-1.5765***	-1.2246**	-0.9892**
	(0.3832)	(0.1618)	(0.4148)	(0.4791)
CAND	0.3398***	0.3275***	0.3350***	0.3353***
	(0.0258)	(0.0204)	(0.0243)	(0.0257)
Medium Efficacy	-0.3509***	-0.1706**	-0.2641***	-0.1365
	(0.0924)	(0.0727)	(0.0767)	(0.0905)
Low Efficacy	-0.2474	-0.2618***	-0.4006***	-0.3656***
	(0.1344)	(0.0701)	(0.0617)	(0.0743)
High Duty	0.4006	0.5068***	0.7305***	0.7685***
	(0.2318)	(0.1245)	(0.1635)	(0.1726)
Medium Duty	1.2696***	1.4727***	2.0070***	0.9418***
	(0.5047)	(0.2609)	(0.3364)	(0.2731)
Low Duty	5.4814***	2.5082***	3.1006***	1.8801***
	(1.1287)	(0.3567)	(0.4806)	(0.4267)
Wald (overall)	615.17	1861.71	1463.94	1007.84
Wald (β)	339.31	776.45	575.59	494.81
Wald (τ)	27.28	30.14	105.67	46.60
Wald (γ)	69.65	172.83	182.50	95.37
$\bar{R}^2$	.6135	.4356	.5318	.5495
n	1052	2211	1813	1222

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$

Table 2.3 Conditional Motivational Probit Model for 1960 and 1980, Whites only

Variable Name	1960	1980
α_1	-0.2459***	-0.3251***
	(0.0393)	(0.0363)
α_2	0.1379***	0.4843***
	(0.0330)	(0.0518)
Strong Democrat	0.3772	0.4088
	(0.2370)	(0.3249)
Weak Democrat	0.6527***	0.2807
	(0.2074)	(0.2133)
Independent Democrat	0.4037	0.1056
	(0.3359)	(0.3464)
Independent	0.5006	0.1735
	(0.4029)	(0.2863)
Independent Republican	0.5006	-1.5333***
	(0.4858)	(0.3748)
Weak Republican	-0.7257	-0.9981***
	(0.4693)	(0.3198)
Strong Republican	-1.8147***	-1.1250*
	(0.4357)	(0.5554)
CAND	0.3459***	0.3375***
	(0.0276)	(0.0275)
Medium Efficacy	-0.3609***	-0.1560
	(0.0948)	(0.0963)
Low Efficacy	-0.2369	-0.3810***
	(0.1459)	(0.0769)
High Duty	0.5878*	0.8693***
	(0.2943)	(0.1941)
Medium Duty	1.7437**	0.9541***
	(0.6466)	(0.3077)
Low Duty	5.6572***	1.9622***
	(1.3432)	(0.4951)
Wald (overall)	529.51	900.98
Wald (β)	298.16	424.99
Wald (τ)	25.36	46.73
Wald (γ)	58.78	88.46
$\hat{R}^2$	.6305	.5525
n	972	1080

*$p < .05$. **$p < .01$. ***$p < .001$

The coefficients in Tables 2.2 and 2.3 are of the sign and magnitude that we might expect.[17] The shift in the duty coefficients has been the most pronounced, with the coefficient for high duty gradually increasing over time, plus the thresholds α_1 and α_2 increasing (since the very high duty individuals are the reference category, we would expect changes in the behavior of those individuals to be picked up in a shift away from zero by the thresholds). Low duty has actually decreased over time, but given the small number of individuals in that category, this has little effect on the predicted number of individuals voting. The effects of the efficacy variables is also generally what one would expect, with medium efficacy reducing turnout less than low efficacy (except in 1960, when the order of magnitude is reversed, but the difference of the two coefficients is not statistically significant, having an asymptotic t-ratio of less than one). Partisanship bounces around quite a bit, but given the small proportion of the sample, has little effect on final predictions. Certain anomalies which one might expect from historical knowledge of various campaigns crop up--one example being the nearly zero coefficient on those professing a weak Democratic allegiance in 1972. Finally, CAND takes on exactly the same weight in all four election years.[18]

17. The Wald statistic is distributed as χ^2 with 15 degrees of freedom for the overall Wald, 8 degrees of freedom for the β Wald, 2 degrees of freedom for the τ Wald, and 3 degrees of freedom for the γ Wald. As indicated, each Wald is the χ^2 value for a test of the null hypothesis that for the particular set of coefficients, all coefficients are zero. They are all significant at the .01 level or greater. The $\bar{R}^2$ is a generalization of the McKelvey-Zavoina (1975) statistic of the same name, and if the conditional motivational probit model is estimated without the condtional variables (the ones associated with the τ and the β), reduces to that statistic.

18. As mentioned before, in 1972 only three responses on the open-ended likes and dislikes were coded. Thus, the CAND variable in 1972 only varies from -6 to 6. As we will see later, however, nearly all of the CAND responses fall between -6 and 6 in any case, so the similarity of 1972 to the other three years is not unexpected.

What then may we say about the amount of structural change between 1960 and 1980? The easiest manner in which to address this question is to take the 1960 sample and examine the percent of individuals who would have turned out if they had had the 1960 exogenous variables and the later years individual response functions. This is done in Figure 2.1 for the weighted 1960 sample, restricted to whites. The two methods of prediction mentioned before were utilized, the first method being classification on the basis of the action with the maximum probability and the second one being the average of the probabilities for each action. Both types of prediction show a drop of about fifteen percentage points for the 1960 white weighted sample. As the actual reported turnout from our sample is 81.47 percent,[19] it can be seen that this method overestimates turnout quite a bit for the first type of prediction and overestimates it by a small amount for the second type. Thus, while it is not possible to get an exact estimate of the structural change, it is evident that it is quite extensive and probably accounts for almost all of the drop in turnout. Confirmation for that conclusion may be obtained by examining the effects of the distributional change of the exogenous variables between 1960 and 1980, as the total change should be the sum of distributional and structural change.

19. This differs slightly from the Abramson and Aldrich figure (which was 83.3 percent), since we included the don't knows on the duty and efficacy questions and also because we excluded those who reported turn out but not their vote choice.

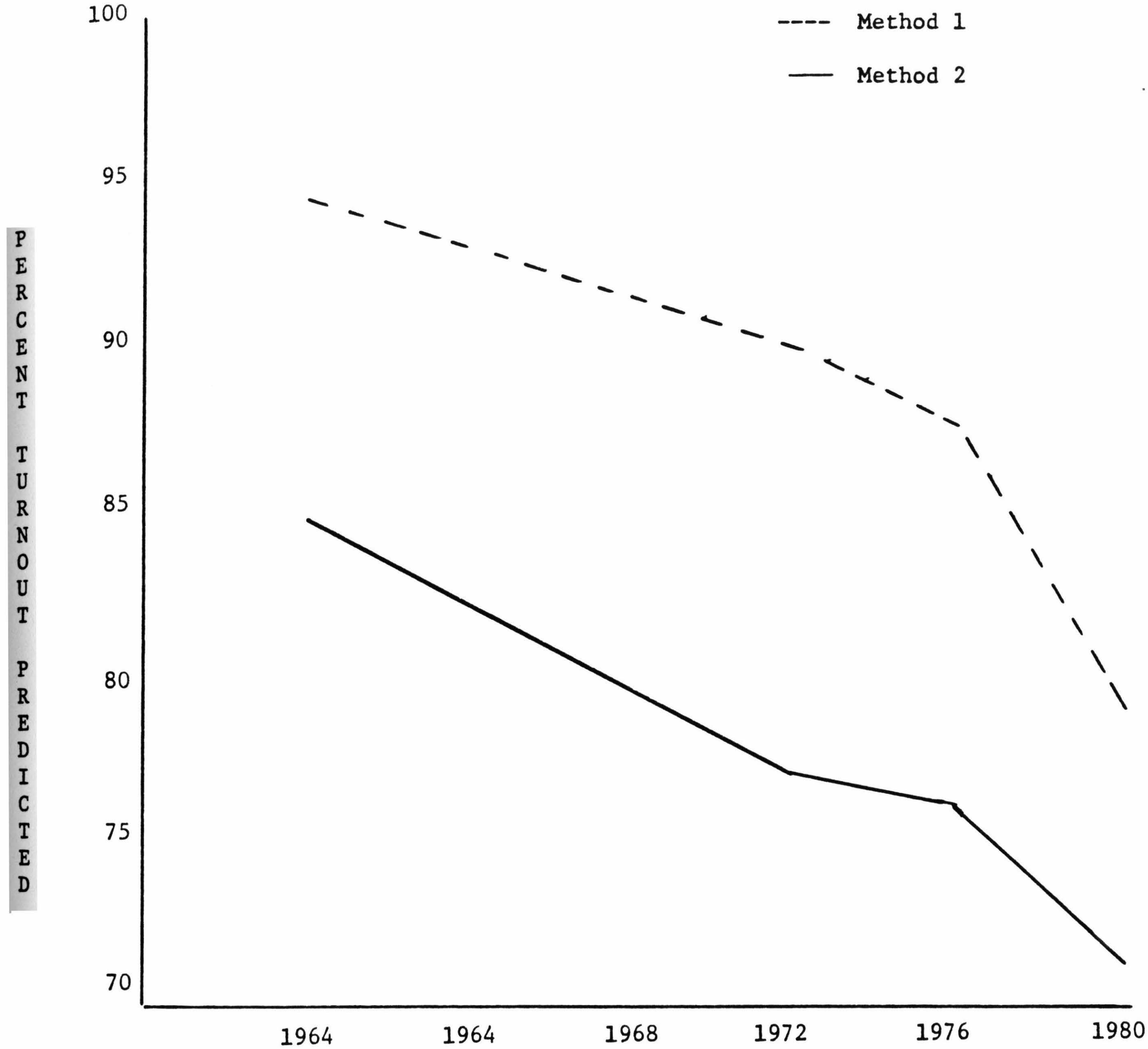


Figure 2.1 Turnout of 1960 White Weighted Sample Under Structural Coefficients from 1960, 1972, 1976, 1980

In our model we may look at change in four sets of variables: the variables dealing with attitudes of political efficacy, the variables dealing with attitudes towards the voting process (citizen duty), the variables dealing directly with the effects of partisan identification on the voting and turnout decision, and the variable which measures the respondent's attitudes directly on the presidential candidate, which probably is to some extent a function of partisan identification, but also is a function of other things. We are interested here in exploring distributional change; that is, change in the distribution of these variables which we have assumed to be exogenous to the voting decision and how they affect the decision to vote. We will continue with our weighted white sample from 1960, and when using 1980 data, use only the white sample there, keeping once more in line with the Abramson-Aldrich results.

We may look at duty first. The distribution of duty in 1960 among whites is as follows. The percent at the very highest level of duty was 51.85 (zero on our scale of citizen duty), the percent at a high level of duty was 36.83 (one on our scale), the percent at a medium level of duty was 6.69 (two on our scale), and the percent at the lowest levels was 4.63 (three to four on our scale). For 1980, the distribution was 50.92 percent at the highest level, 35.93 percent at the next highest level, 8.15 percent at a medium level, and 5.00 percent at the lowest levels. It seems manifestly clear that the distribution of duty, as noted many times before, has not changed. Thus there is no distributional change from a change in duty.

The next variable we look at is partisan identification, both through its direct effects in our seven dummy variables and its indirect effects through CAND. In Table 2.4 we present the distribution of partisan identification for those individuals who have a zero score on CAND for both 1960 and 1980 samples. The distributional differences between these two samples is once again minimal, so we may look for little change in voting and turnout patterns as a result of the direct changes in partisanship. As for indirect changes in partisanship, as reflected in the distribution of the CAND variable, we have plotted a frequency distribution of the 1960 CAND variable and the 1980 CAND variable in Figure 2.2. As can be seen, the distributions are once again very close. So the distribution of CAND has remained the same, even though the partisan content may have gone down.[20]

20. Wattenberg (1981) has given a good account of the increasing irrelevance of political parties in America; see also Miller and Miller (1976). Determining the exact nature of the partisan content of the candidate evaluations is beyond the scope of this chapter!

Table 2.4 Partisan Identification for Those Scoring Zero on CAND, 1960 and 1980

Variable Name	Percent in 1960	Percent in 1980
Strong Democrat	2.3	1.9
Weak Democrat	3.8	5.0
Independent Democrat	1.2	1.9
Independent	1.1	2.5
Independent Republican	0.2	1.6
Weak Republican	1.3	2.1
Strong Republican	0.7	0.5

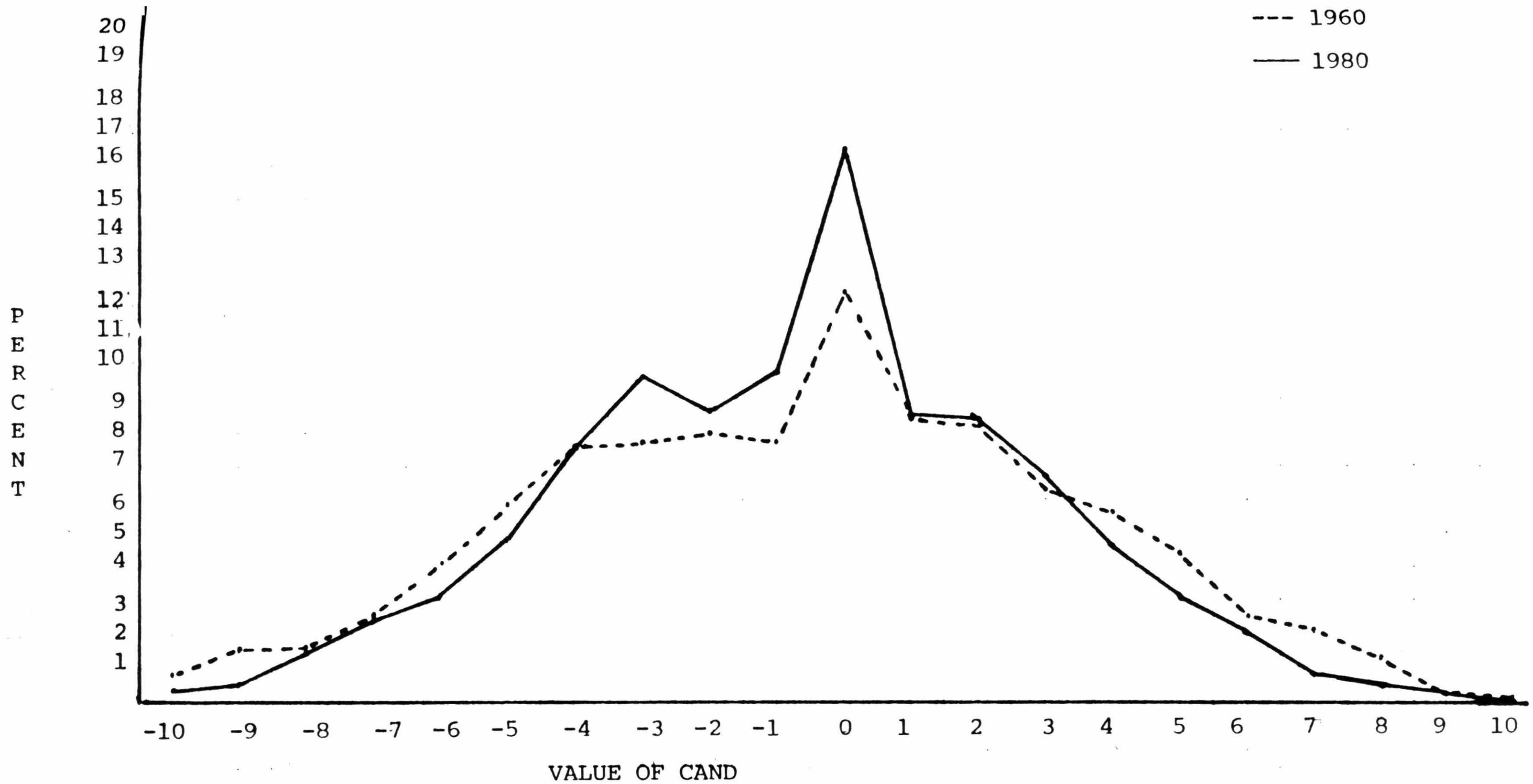


Figure 2.1 Distribution of CAND for White Samples, 1960 and 1980

This leaves the variable efficacy as the only variable for which there may be distributional change. In 1960, according to our weighted sample of whites, 63.2 percent were at the highest level of external efficacy, 22.9 percent were at a medium level of efficacy, and 13.7 percent were at the lowest level of efficacy. In 1980, only 34.7 percent were at the highest level, 33.9 percent were at a medium level, and 31.4 percent were at the lowest level. This is quite a difference in the distribution of this exogenous variable. To measure the effect of this distributional change, then, we need to estimate the predicted vote in 1980 with the distribution of efficacy in 1960, and subtract the predicted vote with the actual distribution of efficacy in 1980. This procedure, then, will give us a rough estimate of the effect of the change of efficacy on the change in vote.

We would be less than honest if we did not mention some problems with this procedure. The main problem is what cases is one to change the efficacy variable on? Presumably efficacy is related to other variables in our sample and this new distribution should retain those relations. This is a difficult problem to solve and we will not attempt to do so here. Rather, we take the cases in the sequence as they appear, setting the efficacy variables until the frequencies in 1980 emulate the frequencies in 1980. We also will estimate turnout with all values of efficacy set to their highest level. This will give us a maximum for the increase in turnout due to an increase in efficacy, and we can have perfect confidence in that figure as representing the maximum possible change in turnout due to distributional change of the efficacy.

When we do this, we obtain the following results. Call the method of classifying observations by predicting the behavior for which their is the greatest probability method one, and call the method of making predictions on the basis of probability averages method two. Then our baseline (predictions for 1980 using 1980 data) is 75.4 percent for method one and 68.1 percent for method two. When we set the efficacy frequencies equal to the 1960 figures, we get predictions of 77.6 percent by method one and 69.5 percent by method two, or an increase of around 1.3 percent by each method (1.2 percent and 1.4 percent by method one and two, respectively). When we set all efficacy to the maximum amount possible, we obtain a prediction of 78.5 percent for method one and 71.0 percent for method two, or around 3 percent. Given the nature of these figures, we are confident in concluding that the actual change in turnout given the change in the distribution of external political efficacy is considerably under 2 percent.

2.5 Conclusions

The results of the previous sections seem clear, but deserve restating: the decline in turnout in the American public from 1960 to 1980 has little or nothing to do with the distributional changes of different attributes but instead reflects structural changes in the manner in which citizens act with respect to those attributes. No attempt is made to explain the causality behind this strucutural change, and it may very well be related to the change in distributions of such variables as political efficacy, strength of partisanship, and more general attitudes such as cynicism and generalized disaffection. The period under study has also been tumultuous. To quote Kousser,

> The period since 1960 also contains enough special circumstances to account for nearly any political phenomenon. The assassinations of the two Kennedys at times when both were popular and very likely to be nominated by the Democratic party; the profound dislocations caused by the most sustained attack on racisms in America's history; the travail of the country's least popular and least successful war since 1815 and its longest war ever; the deliberate subterfuge of presidents and their advisers about Indochina and Watergate and their later open defiance of large segments of informed and intense public sentiment; and the two huge spurts in energy prices and the extreme economic consequences of the OPEC cartel's actions--no era of American history can match this one for a series of wrenching shocks to the national political consciousness.

Is there any doubt that the manner in which individual's make decisions over political acts such as voting would change in such an environment?

We mentioned earlier that we would provide some evidence that the decline in turnout has not been continuous from 1960 to 1980, and we will use this to indicate some areas of future research. It is true that Table 801 ("Participation in Elections for President and U.S. Representatives: 1932 to 1980") of the 1982 *United States Statistical Abstract* indicates a decline in turnout steadily from 1960 to 1980. Table 806, however, ("Participation in National Elections, 1964 to 1980"), asks individuals to self-report whether they voted or not, and according to this, there was no decrease or increase between 1976 and 1980. Furthermore, we now know there was an increase in turnout in 1982. The pattern of the decreases and increase from Table 801 is more interesting--there was a full 10.5 percent increase from the 1948 to the 1952 election, for example, and perhaps more importantly for the ANES period, the largest decreases were from 1968 to 1972 and 1972 to 1976 (the ANES data show an *increase* from 1972 to 1976 among whites), both on self-report (Table 806) and estimated proportion from the population (Table 801). These two periods were the periods when some of the most brutal of the events described by Kousser occurred–that there is a correlation seems obvious. It is to the effects of external political events on the methods by which voters make decisions that those interested in political behavior should turn, which, as indicated in the text, will not be a simple matter. It is one well worth the undertaking, however.

Abramson, Paul R. and John H. Aldrich. 1982. "The Decline of Electoral Participation in American." *American Political Science Review*, 76, 3, 502-521.

Aldrich, John H. 1976. "Some Problems in Testing Two Rational Models of Participation." *American Journal of Political Science*, 20, 713-733.

Boyd, Richard W. 1981. "Decline of U.S. Voter Turnout: Structural Explanations." *American Politics Quarterly*, 9, 133-159.

Brody, Richard A. 1978. "The Puzzle of Political Participation in America." In *The New American Political System*, ed. Anthony King, pp 287-324. Washington, D.C.: American Enterprise Institute.

Campbell, Angus, Gerald Gurin and Warren E. Miller. 1954. *The Voter Decides*. Evanston, Ill: Row, Peterson and Co.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. Chicago, Ill: University of Chicago Press.

Cassel, Carol A. and Hill, David B. 1981. "Explanations of Turnout Decline: a Multivariate Test." *American Politics Quarterly*, 9, 181-195.

Ferejohn, John A. and Morris P. Fiorina. 1979. "The Decline of Turnout in

Presidential Elections." Paper presented at the National Science Foundation Conference on Voter Turnout, San Diego.

Heckman, James J. 1981. "Statistical Models for Discrete Panel Data." In C. Manski and D. McFadden (eds.),*Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass. MIT Press.

Kelly, S. Jr. and T. W. Mirer. 1974. "The Simple Act of Voting." *American Political Science Review*, 61, 572-591.

Kousser, Morgan J. 1983. "Suffrage and Political Participation." Social Science Working Paper 471, California Institute of Technology (forthcoming in the *Encyclopedia of American Political History*)

Maddala, G.S. 1981. *Econometrics*. New York, New York: McGraw-Hill.

McKelvey, Richard and W. Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology*, 4, 103-120.

Miller, Arthur H. and Warren E. Miller. 1976. "Ideology in the 1972 Election: Myth or Reality--A Rejoinder". *American Political Science Review*, 70, 834.

Nie, Norman H., Sidney Verba and John R. Petrocik. 1976. *The Changing American Voter*. Cambridge, Mass: Harvard University Press.

Pratt, J. W. 1981. "Concavity of the Log-Likelihood." *Journal of the American Statistical Association*, 76, 137-139.

Shafer, Stephen D. 1981. "A Multivariate Explanation of Decreasing Turnout in Presidential Elections, 1960-1976." *American Journal of Political Science*, 25, 68-95.

Verba, Sidney and Norman Nie. 1972. *Participation in American.* New York: Harper and Row.

Wattenberg, Martin P. 1981. "The Decline of Political Partisanship in the United States: Negativity or Neutrality?" *American Political Science Review*, 75, 941-950.

White, Halbert. 1982. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica*, 50, 1, 1-25.

Chapter III: Presidential Coattails: A Structural Motivation

The supposed recent disappearance of coattails in American Federal elections, (Cook, 1976),[1] and its reappearance in the 1980 elections (Buchanan, 1980),[2] is a political mystery of the first order. For scholars writing before the 1980 election, the decline of what is considered "coattail" voting has either elicited no explanation from political scientists (Kritzer and Eubank (1979)), or been blamed as a consequence of the increase insulation of incumbents from electoral effects (Burnham, 1975). While it has been shown that the connection of the presidential and congressional vote has been decreasing over time, with the lowest level of connection in the post-war period (Calvert and Ferejohn (1980)), the relationship of this fact to the behavior of the individual voter is not at all clear. In this paper, then, we will show that previously offered explanations for the motivations for coattail voting are not supported by the evidence from the 1980 election and that the dimunition of "coattail" voting has, in all probability, not occurred. Rather, we will show that the motivation of the individual voter includes a desire by the voter to increase the chances of obtaining the policy goals he desires, and that this motivation creates what is usually referred to as "coattail" voting.

There exist in the literature at least three suggested motivations behind the individual voter's propensity to engage in coattail voting,

1. "There were no coattails to speak of in the 1976 presidential elections." (p 3135).
2. "The apparent national shift to the right combined with a variety of scandals, complacency by some incumbents and unusually strong coattails from Ronald Reagan gave the Democrats a net loss of 33 House seats." (p 3317).

which may be roughly characterized as the personal charm, the interest, and the candidate evaluations motivations.[3] The personal appeal, or, as it is sometimes referred to, the personal charm motivation, has been summarized by Miller (1955), as "the magnetism of the presidential candidate which can be translated into the sort of allegiance which commands the voter to do his bidding and gives support to his cohorts who follow him on the ballot".

It must be said, though, that there are problems with the notion of "personal charm". By anyone's casual observation, the most charming and charismatic presidential candidate of the post World War Two years, John F. Kennedy, provided no coattails for his party (and, in fact, he ran behind the Democratic congressional vote).[4] Miller also uses "appeal", which is certainly more fitting, as we can easily convince ourselves that a candidate's appeal is a function of the criteria that political scientists usually associate with a decision to vote for a candidate--issue positions, candidate characteristics, long-term partisan dispositions. Appeal is also very close to the definition of coattails motivation referred to as candidate evaluations, so it may be that one cannot think of charm, interest and candidate evaluations separately.[5]

3. Another theory, that of Kritzer and Eubank (1979), can be thought of as a modification of the interest, or "surge and decline" theory. They suggest that "very few party identifiers will vote for the other party's presidential candidate", whereas independents will, because of the low salience of presidential elections, vote for Congressman the way they voted for President. As they conclude that the data do not support this theory, we will not consider it specifically in what follows.
4. This may be a subjective opinion, but Kritzer and Eubank, based on their own subjective rankings of Nixon, Johnson and Carter, also conclude "the magnetism explanation fails to account for recent electoral trends."
5. We will endeavor, however, to test the presidential candidate charm theory in a way which distinguishes it from the interest and candidate evaluation theory.

The next suggestion for the individual's motivation for coattail voting is what we will call "interest" coattails, or those coattails produced by the fact that in a presidential campaign, the focus of the media is predominantly put upon the presidential candidates of both parties. Miller (1955, p 359) states it as follows,

> "In addition to measuring motivations relevant to the presidential and congressional elections, a third problem thus arises from the necessity to determine whether any of the motivations found to be focused on the congressional elections were in fact stimulated, activated or, if you will, produced by the presidential campaign. People exhibiting congressionally oriented motivations of this type, motivations which owe their existence to the presidential candidate, are just as clearly coattails voters as are their cohorts who give no evidence of motivations uniquely related to voting for congressional candidates."

This theory has, on the surface, a plausibility which is missing from the personal charm theory. It is obvious that in an on-year election, the great bulk of the media attention and voter interest is directed towards the presidential race, to the detriment of coverage and interest in other races. Thus it would seem possible that motivations excited by the presidential race would affect the congressional race.[6]

6. One point that seems to have been missed by political scientists is that if all of the attention is placed upon the presidential campaign, then the presidential candidates have greater control of the agenda of national politics, so to speak. Thus the presidential candidates of both parties have an opportunity to place a large range of issues on the national agenda which could have an impact upon congressional contests. Indeed, when one thinks back to the campaigns since the end of the war, the surface plausibility of this type of motivation seems strong, since one would expect the strongest coattails effect when the individual was either strongly identified with a party or ran a campaign which was highly partisan. Both the 1948 and 1960 elections were very close, but excluding them, the two elections which on casual observation had large coattail effects involved Johnson, who was a longtime Democratic Senator, and Reagan, who ran a highly partisan campaign. Eisenhower, on the other hand, was sought after to be the nominee of both parties, Nixon allowed the Democrats to fight among themselves in 1968 without raising the one issue which would have had highly partisan content, ran completely independently of his party in 1972, and Carter ran as an outsider and, indeed, captured the nomination completely independently of the Democratic Party's elite. The problem with such *ad hoc* theorizing, of course, is that one needs to look

The third type of motivation, apparently first suggested by Moreland (1973) (though very similar to the appeal theory posited earlier), is that of positive candidate evaluations creating a coattails effect. As Moreland puts it, "the coattails phenomenon appears when the positive identification with a presidential candidate is carried over in the evaluation of other candidates of the same party, with the assumed result that these candidates gain votes they would not otherwise receive" (p 171). Calvert and Ferejohn (1983) phrase it in the following manner,

> "Of course, this dependence [of the congressional vote on the presidential vote] might be based on factors directly attributable to the presidential campaign or on factors more or less extraneous to it. For example, partisan affiliations and reactions to the general economic situation could induce a certain correlation between the electoral results at the presidential and congressional levels, having little to do with the candidates for presidential office. On the other hand, evaluations of the attributes of the presidential candidates (including their issue positions) might have a direct effect on voting for congressional candidates independently of the effects of party, the general economic situation or other factors. It is this direct effect which is sometimes called the 'coattail' vote."

This motivation, then, focuses on the voter's evaluation of the presidential candidate and assumes that it directly affects the congressional vote in the case of coattails.

We can classify all three of these hypotheses of motivation as magnitude motivations, i.e., the more the person feels about a presidential candidate or candidacy (charmed, interested or having a

---

for data to back up one's assertions, and arguments could be made with all of the classifications made above. Still, we will present evidence later on in this paper that the effect we have proposed in this footnote may indeed be present.

positive evaluation), the greater is his motivation to support the congressional candidate of the party for which he casts his presidential vote. It is not clear why any of these motivations should have an effect on the voting behavior of the individual. Certainly, from the standpoint of a government which is able to enact effective policies in a timely manner, it is desirable to have a legislative and executive branch under the control of the same party. Why should the individual be concerned about this, however? One immediate answer is that when the different branches are under the control of different parties, those effective and timely policies become much less likely. Thus, if the individual values the implementation of effective policies, he will have a desire to see that both the legislative and executive branches are under the control of the same party. One way for him to implement his desire is to associate his vote for the different offices. This explanation, of course, has nothing to do with the magnitude for a presidential candidate; rather, it is a motivation related to his choice of presidential candidate (or congressional candidate, for it is clear this type of motivation is symmetric).

In addition to the three hypotheses of motivation suggested in the literature, then, we suggest a fourth hypothesis, that of the desire of the individual to achieve desired policies by associating his vote for one office with another in order to overcome the lack of cooperation implicit in the Federal system. We thus posit the existence of a tendency for the voter to associate votes for different Federal offices in order to increase cooperation by office-holders of the same party. It is this tendency that we will refer to as the structural motivation of coattails. The suggestion of the voter's cognizance of the separated nature of the Federal system as being the motivation behind coattails is new, and we need to offer some motivation as to why it might be a reasonable criteria for a elector to consider in his voting decision.

Any discussion of coattails must acknowledge the separation of powers with which the Federal government was created and still maintains, to an amazing degree, today. The reason for this deliberate effort was perhaps best stated by Alexander Hamilton (*The Federalist*, Number 60)

> The dissimilarity in the ingrediants which will compose the national government, and still more in the manner in which they will be brought into action in its various branches, must form a powerful obstacle to a concert of views in any partial scheme of elections...the House of Representatives being to be elected immediately by the people, the Senate by the State legislatures, the President by elector chosen for that purpose by the people, there would be little probability of a common interest to cement these different branches in a predilection for any particular class of electors.

As is evident from the passage, one of the main concerns of the authors of *The Federalist Papers* was separating the influence of the three classes of electors by providing each one with a different office of the Federal system to vote for. Only in this way, was it thought, would one class of electors (with one set of interests) be prevented from dominating the government and produce the tyranny of the majority which was so feared.

Actual history, of course, has taken a different turn from the system envisioned by the constitutional convention; not only do the common voters participate directly in voting for President, Senator and Congressman, but there exists that other anathema of the authors of *The Federalist Papers*, parties, to guide the electors in their voting decision. The voter now has the choice, when entering the voting booth, to vote for an entire slate of individuals dedicated, to some degree or another, to cooperating in passing legislative programs. It is self-evident that if he votes for a President of one party and a Congressman of another party, he is lowering the amount of cooperation on a legislative program that those two, if elected, will engage in than if he votes for a President of one party and a Congressman of the opposite party.[7] Thus, even if a voter decides that a President of one party and a Congressman of another party are both more preferred by whatever criteria he employs in making his decisions, he may still vote for a President of one party and a Congressman of the same party simply because of the cooperation that the candidates for different offices may engage in.[8]

7. This implicitly assumes that party is a valid label on which to judge cooperation with individuals of the other branches of the Federal government. While the decline of party discipline has been much lamented in recent years, this assumption is probably still accurate, though not to the degree that it once was (see Cummings, (1966), page 2 for a similar argument). Another possible objection is that the individual needs to take into account the composition of the House, say, when making his voting decision. It is obvious, though, that even if the House was heavily Democratic and the voter desired to vote for a Republican President, this Republican President would have greater opportunity with an additional Republican Congressman than without one.

8. This is not the place to enter into the debate over why people vote or whether it matters how they do vote. It should be noted, though, that there are a whole host of studies which indicate that people vote not only in accordance with what they perceive their own interests, but perhaps more importantly, vote in accordance with what political scientists see as the voters' own interests. Thus, even if their reasons for going to the polls is a result of some sociological or psychological process, there is ample reason to believe that they behave "rationally" once they are there. Another problem which we do not discuss is that of "strategic voting" (see Cain (1979)) for a discussion of this phenomenon). Strategic voting is voting against one's preferred candidate in favor of a less preferred candidate for reasons (usually) having to do with the context that the election is held in (more than two candidates, for example). As will presently become clear, though, strategic voting, if it does exist, will not be a problem for our examination.

We have set up the broad outlines of our research strategy, then. There exist four theories which purport to explain the motivation behind coattail voting, three of them very familiar, one of them new to this article. To examine the incidence of these four motivations we turn to the one complete data set dealing with both presidential and congressional elections, the 1980 American National Election Study. It is the task of developing the research strategy which will allow the examination of these four competing theories of coattails that we turn to in the next section.

3.1 A Simultaneous Decision System for Examining Coattail Voting

In this section we consider a statistical model for coattails voting which will allow us to test the four theories of the motivation behind coattails examined in the previous section. The advantages of the model we consider include the ability to distinguish between the various hypotheses of motivation for coattails voting while allowing the specification of voting models which are consistent with what is known about voting in congressional and presidential races. The only disadvantage is that it does not allow simultaneous estimation of turnout, but this is immaterial to the testing of hypotheses of the motivations behind coattail voting.[9]

We consider a simultaneous equation system involving two

9. As opposed to the overall effect of coattail voting, where as Miller notes (p 366n.), voters may be mobilized to go to the polls by a particularly attractive candidate. Those voters (the number of which one would guess would be low in the post World War Two period), are assumed to behave in the same manner as other voters once they reach the polls.

underlying variables observed only discretely.[10] Let y_1^* and y_2^* be unobserved continuous variables and let y_1 and y_2 be observed if

$$y_1 = 1 \quad y_1^* > \alpha_1$$

$$= 0 \quad y_1^* \leq \alpha_1,$$

and

$$y_2 = 1 \quad y_2^* > \alpha_2$$

$$= 0 \quad y_2^* \leq \alpha_2.$$

This is of course the probit specification (usually α_1 and α_2 are set equal to zero). Consider perhaps the most general simultaneous linear model with these probit variables,

$$y_1^* = \gamma_{11}y_1 + \gamma_{21}y_2 + \delta_1 y_2^* + X_1\beta_1 + \varepsilon_1$$

and

$$y_2^* = \gamma_{12}y_1 + \gamma_{22}y_2 + \delta_2 y_1^* + X_2\beta_2 + \varepsilon_2.$$

As Schmidt (1981) shows, the model is internally consistent only if the following conditions hold:

$$\gamma_{11} + \delta_1\gamma_{12} = 0,$$

$$\gamma_{22} + \delta_2\gamma_{21} = 0,$$

and

10. Calvert and Ferejohn (1983) also considered a a simulataneous decision system, but they assumed that the coattails effect was a result of presidential evaluations. Our system is somewhat different in order to allow the testing of the four motivation hypotheses.

$$\gamma_{21}\gamma_{12} = 0.$$

Internally consistent in this context means that a given set of exogenous variables and disturbances yield a unique solution for the endogenous variables. Demonstration of these conditions in this set of equations depends upon a theorem of Schmidt, but the reason for such internal consistency conditions can be easily demonstrated with a simpler set of equations. Let

$$y_1^* = \gamma_{21}y_2^* + X_1\beta_1 + \varepsilon_1$$

and

$$y_2^* = \gamma_{12}y_1 + X_2\beta_2 + \varepsilon_2.$$

Putting the expression for y_2^* from equation (5.1.10) into the expression for y_2^* in equation (5.1.11), we have that

$$y_1^* = \gamma_{21}\gamma_{12}y_1 + X_1\beta_1 + \gamma_{21}X_2\beta_2 + \gamma_{21}\varepsilon_2 + \varepsilon_1.$$

Then, noting that $y_1 = 0$ if $y_1 \leq \alpha_1$, we have

$$y_1 = 0 \text{ if } \gamma_{21}\varepsilon_2 + \varepsilon_1 \leq \alpha_1 - X_1\beta_1 + \gamma_{21}X_2\beta_2$$

and

$$y_1 = 1 \text{ if } \gamma_{21}\varepsilon_2 + \varepsilon_1 > \alpha_1 - X_1\beta_1 + \gamma_{21}X_2\beta_2 + \gamma_{21}\gamma_{12}.$$

Clearly $\gamma_{21}\gamma_{12}$ must equal zero for this to have a unique solution, and this condition is then the internal consistency condition for (5.1.10) and (5.1.11).

Consider what (5.1.5) and (5.1.6) mean in a voting choice context.

Interpreting the underlying variable as the utility of a voting choice or as the desire to vote for one candidate, it is clear that there is no political interpretation for the underlying utility of a vote for a candidate to be a function of that actual vote itself.[11] Thus γ_{11} and γ_{22} should be equated to zero. Once having done that, we are left with the three conditions to be satisfied: $\delta_1\gamma_{12} = 0$, $\delta_2\gamma_{21} = 0$, and $\gamma_{12}\gamma_{21} = 0$. Before we attempt to decide how to satisfy these conditions, let us consider the interpretation of the endogenous variables on the right-hand side of the model.

Consider the underlying variable y_1^* first. Its usual interpretation is that of the desire to perform some dichotomous action (in this case, vote for one candidate). When y_1^* is above a certain point (called a cutpoint), the action is observed (voting for a Democratic congressional candidate, for example). When y_1^* is below that point, the action is not observed (not voting for a Democratic congressional candidate, which, if one restricts oneself to voters who voted and had only a choice of a Democratic or Republican congressional candidate, is equivalent to voting for a Republican congressional candidate). The interpretation of the value of y_1^* as utility (or desire to perform an action) comes from the interpretation of the probit model as a random utility maximizer.[12] While

---

11. There are theories (see Hinich, 1981 in particular) which posit that the utility of a vote choice comes partially from voting for the winning candidate, but that is different from utility for a candidate being derived from a vote for that candidate.

12. An individual is a random utility maximizer if he maximizes his utility subject to some random error term. In the probit case, let the utility of alternative one be $u(1|x,\beta) = x\beta_1 + \varepsilon_1$ and that of alternative two be $u(2|x,\beta) = x\beta_2 + \varepsilon_2$. Then $u(1|x,\beta) - u(2|x,\beta) = x(\beta_1 - \beta_2) + \varepsilon_1 - \varepsilon_2 = x\beta + u$. This is, of course, the usual probit model, when we observe only indicators of utility (the choice or lack of choice) of the individual. Then the difference in utility between the two actions may be estimated by $x\beta$, which is a consistent estimator of y_1^*. Thus the interpretation of the underlying variable as utility.

y_1^* is not observed, it may be consistently estimated from the probit equation.

We will, in this paper, assume that the values of the underlying variables correspond to the magnitude of motivation for a candidate. If the equation is well specified, one would expect the values of the underlying variable to reflect what Moos (p 5) calls "the strong momentum generated by the presidential campaign". Whether one thinks of it as being generated by personal charm, interest, candidate evaluations or whatever, it is implicit in the formulation of either hypothesis that it is a magnitude effect, that is, the larger the magnitude of the motivation of the individual voter in the presidential race to vote for one candidate or the other, the larger the coattails effect. If we find that a large magnitude in a presidential race affects the choice of a congressional candidate, we will then need to inquire as to whether it is the appeal of the presidential candidate or the additional interest of the presidential campaign which is the motivation behind coattails voting. On the other hand, a small or non-existent effect will cast doubt on the whole hypothesis of any type of "momentum" generating coattails effects.

Now consider the observed value of the underlying variable, y_1, presumed to have been generated by the underlying variable y_1^* in the manner described above. The interpretation of this variable is simply that of choice. If the effect of y_1 is significantly related to the choice of y_2, this would be an indication of the validity of the assumption of the structural motivation for coattails, for then it would be the final choice itself, which directs the coattail vote. In other words, independently of the motivation to vote for a presidential candidate (say), the fact of a choice for the presidential candidate resulting in a direct influence on the choice of a congressional candidate would be a a substantiation of the structural theory.

Naturally, both or neither theory may be working in the electorate. Suppose neither was present. This would say, basically, that there were no coattails. If a strong effect was found of the effect of personal appeal on the votes for other offices but not on the actual choice, then this would indicate the appeal motivation was predominant and that the structural motivation was low or non-existent. A reversal of these results would indicate a reversal of the conclusion, whereas a strong effect from both would indicate that both theories would hold. They are not mutually exclusive, conceptually.

To make operational the test of the theories we have proposed, however, we need to consider once again the constraints $\delta_1\gamma_{12} = 0$, $\delta_2\gamma_{21} = 0$, and $\gamma_{12}\gamma_{21} = 0$. First, it is clear that either $\gamma_{21} = 0$ or $\gamma_{12} = 0$ (or both). Thus, intuitively enough, one cannot have a chain where one discrete choice predicts the other discrete choice which predicts the first discrete choice, etc. Set $\gamma_{21} = 0$ and let γ_{12} be non-zero. Then the condition $\delta_1\gamma_{12} = 0$ implies that $\delta_1 = 0$. We have thus imposed a recursitivity on the system, which looks like the following:

$$y_1^* = X_1\beta_1 + \varepsilon_1$$

and

$$y_2^* = \gamma_{12}y_1 + \delta_2 y_1^* + X_2\beta_2 + \varepsilon_2,$$

where both δ_2 and γ_{12} may be non-zero. We will call this the structural model, since it is the only model which allows the presence of a discrete choice on the right-hand side of an equation and still satisfies the internal consistency conditions of Schmidt.

Now, suppose we set γ_{12} and γ_{21} to zero. This then allows both δ_1 and δ_2 to be non-zero, making the formulas given in (5.1.5) and (5.1.6) to be

$$y_1^* = \delta_1 y_2^* + X_1\beta_1 + \varepsilon_1$$

and

$$y_2^* = \delta_2 y_1^* + X_2\beta_2 + \varepsilon_2.$$

We will call this the momentum or motivational model, following the discussions in the proceeding paragraphs. Testing of these two models will thus allow us to determine which theory or theories, if any, best fit the behavior of voters in American Federal elections.

3.2 The Test of the Theories: the 1980 ANES

To test the theories proposed in the last section we turn to the 1980 ANES. In the last section we proposed two basic models of possible congressional behavior, the momentum model and the structural model. The models differ in that the endogenous variables are assumed to enter in a different fashion in each, but both assume the entry of a number of exogenous variables in each. These exogenous variables are assumed to be related to the voting decision but not to be correlated with the error term, playing the role of causal precursors which allow the calculation of the underlying desire to vote for a candidate for any individual voter.

In order to estimate a model such as (3.1.15) and (3.1.16), or (3.1.17) and (3.1.18), the system needs to be identified. Identification of a system of simultaneous equations can be made by either covariance restrictions or linear (usually exclusion) restrictions, or a combination of both. Both type of restrictions work on the assumption that the researcher has prior knowledge on the nature of the problem which allows him to make the restrictions. In general, without a very well formulated theory, covariance restrictions are contraindicated, so we will not consider those here. Rather, we will use restrictions on the inclusion of exogenous variables. This requires that we know a) the set of variables which are assumed to be in both equations, b) the set of variables which are assumed to be in only the presidential race equation, and c) the set of variables which are assumed to be in only the congressional equation.

Dealing with the presidential campaign first, then, in the context of one election, it is clear that in order to test "whether any of the motivations found to be focused on the congressional elections were in fact stimulated, activated, or, if you will, produced by the presidential campaign", we will need to decide upon a method of classifying those issues which are in some way stimulated by the presidential campaign and decide upon some way of letting them affect congressional races. We will gain some help if we consider the fact that in every election except one since 1928, the President or Vice-president of the administration in power at the time of that election was running for President (and in 1952, "Communism, Corruption and Korea" certainly related to the Truman administration). This ensures that the campaign will focus on the activities of the previous administration, and it is likely that most of it will focus on three factors of great importance in American politics, unemployment, inflation and war. Thus for issues facing the public which can be considered to induce motivations on congressional races from presidential races, these three and the public's perception of the incumbent administration's policy on them can be considered to be issues which have an indirect effect on congressional races through the presidential campaign, rather than a direct effect. For our estimation, then, we will use as variables unique to the presidential race the voter's judgment of the previous administration's policy on those three items, as evidenced by the voter's responses to the questions on whether the voter approves or disapproves of Jimmy Carter's performance on inflation, unemployment and the Iranian crisis (Iran was used because it dealt with a situation which was the closest to a war that the US was involved in during the Carter administration).[13]

13. Questions were used as follows to create dummy variables used in the analyses. A typical question was the one on inflation, question 197, "In general, do you approve or disapprove of the way Jimmy Carter is handling inflation?" There would then be four cateories of answer—approve, disapprove, don't know and not applicable. Two dummy variables would then be created, the first coded one if the individual approved, zero otherwise, the other one if the individual disapproved and zero otherwise. The reference category would then be don't know. Analyses were run with both missing data (the not applicable) included in the reference category (and hence those cases were left in the analysis), and with cases with missing data excluded. The results were comparable under both procedures, so that the analyses with the missing data included in the reference category are reported in this paper. It would be expected in any case that this type of procedure would reduce the size of the coefficient for the dummy variable with a reference category which included missing data.

The 1980 study is also unique in that it contains a number of questions about congressional races which had previously only been asked in the 1978 study, questions about such things as casework and issues in House races. The view of House races as largely party-oriented affairs (Stokes and Miller (1962)) has fallen by the wayside, thanks in particular to the 1978 ANES (Mann and Wolfinger (1980), Fiorina (1981), Yiannakis (1981), Abramowitz (1980)). Instead, it is generally realized that casework in particular makes a difference, and may actually explain most of the phenomenon of the overwhelming advantage of incumbent congressmen in reelection contests (see Erikson (1971) or Hinckley (1981) for a description of the incumbency advantage, Fiorina (1977) for a particularly strong argument that casework was responsible). For our purposes, this previous research indicates that variables such as casework and issues are important for a well-specified model of congressional elections. Therefore, variables such as contacts with the candidates and preference between candidates based on an issue in the congressional race were used. Dummy variables were also used for satisfaction with a contact with the incumbent candidate (a contact to express an opinion, inquire after information, or request help) and a dummy for incumbency. These then were the variables which were decided were unique to the congressional race.

That leaves the variables which are assumed to influence both races. Immediately one thinks of long-term partisan identification, represented by the respondent's self-identified partisan identification. These were coded as one-zero dummies for a strong Democrat, weak Democrat, and independent Democrat, and as one-zero dummies for strong Republican, weak Republican and independent Republican, the reference category being independent. Also, going back to the themes of inflation, unemployment and war, dummies were created for the questions relating to the respondent's beliefs as to which party was better on each of these themes, the reference category in each case being those who supposed no difference or claimed not to know. Finally, sex and race were included in each equation, the former to pick up any of the much-remarked sex difference (see *Public Opinion*, (1981)), the latter to pick up any results of race.[14]

Having chosen our variables, then, we turn to testing the momentum model, the results of which are shown in Table 3.1. Some explanation needs to be made for those unfamiliar with the simultaneous equation methods, particularly those involving probit estimation. Such a method requires estimation in more than one step, unless one is doing full information maximum likelihood. In this procedure, we use Amemiya's principle (1978), which requires a first stage estimation of the reduced form (the endogenous variable regressed on all of the exogenous variables), and then uses these reduced form parameters to solve for the parameters of the structural form (or, in this model, simply (3.1.17) and

14. It is likely that the significant tendency of blacks to favor Democratic candidates (reported later), even controlling for all the other factors that we have mentioned, proxies the fact that blacks as a group receive much greater benefits under Democratic administrations than Republican ones. Thus, we are really picking up an economic effect by specifying blacks.

(3.1.18)) by the method of linear regression. This method gives consistent estimates of the reduced form parameters and these estimates will be more efficient than the usual two-stage method (see Heckman (1978) for a description of those methods) if generalized least squares is used with Amemiya's principle. Therefore, there are none of the usual summary statistics available for the estimation of the structural equations (as there would be in a two-stage procedure), so we report the first stage statistics. For tests of significance of the coefficients, we use the usual procedure of dividing the coefficient by its asymptotic standard error to obtain what is called an asymptotic t-ratio (though it is asymptotically normally distributed).

Table 3.1 Momemtum Model for Presidential Coattails

Independent variable	Dependent variable: Presidential Vote	Dependent variable: Congressional Vote
Strong Democrat	0.9615***	0.4393
Weak Democrat	0.8981***	0.1811
Independent Democrat	0.7614***	0.0197
Independent Republican	0.2288	-0.4626
Weak Republican	-0.4678	-0.6078**
Strong Republican	-0.5621	-0.7533***
Inflation		
Democrats better	0.3212	-0.2243
Republicans better	-0.7156***	-0.1867
Approve of Jimmy Carter's performance	0.0766	
Disapprove of Jimmy Carter's performance	-0.1183	
Unemployment		
Democrats better	0.7222***	0.5635**
Republicans better	-0.5221**	-0.2124
Approve of Jimmy Carter's performance	0.2433	
Disapprove of Jimmy Carter's performance	-0.4170**	
War		
Democrats better	0.2292	0.4803***
Republicans better	0.1621	0.0320
Approve of Jimmy Carter's performance on Iran	0.2756	
Disapprove of Jimmy Carter's performance on Iran	-0.5426*	
Black	1.0429***	0.7386**
Male	-0.2894*	0.0230
Opinions of and contacts with House candidates—Democratic		
General contact with House candidate		0.8192***
Wrote opinion and was satsified with response		0.3657

Asked information and was satisfied with response		0.0330
Asked for help and was satisfied with response		1.4126***
Preferred candidate because of issue in race		1.0701**
Incumbent candidate running for office		0.4472*
Opinions of and contacts with House candidates—Republican		
General contact with House candidate		-0.5775***
Wrote opinion and was satsified with response		-0.8245
Asked information and was satisfied with response		-0.8936
Asked for help and was satisfied with response		-0.6509*
Preferred candidate because of issue in race		-2.9033***
Incumbent candidate running for office		-0.0837
Underlying motivation to vote for a Democratic President		0.0425
Underlying motivation to vote for a Democratic Congressman	0.0548	
First stage $\hat{R}^2$	.784	.773
First stage -2*LLRatio	565.05	461.61
n	677	677
Democratic votes correctly predicted	227 (86.6%)	291 (83.9%)
Democratic votes incorrectly predicted	35 (13.4%)	56 (16.1%)
Republican votes correctly predicted	371 (89.4%)	273 (83.0%)
Republican votes incorrectly predicted	44 (10.6%)	56 (17.0%)

*$p < .10$. **$p < .05$. ***$p < .01$.

The first conclusion we may draw from the data in Table 3.1 is that the momentum model has no support. The coefficient for the underlying desire to vote for a Democratic President has no effect on the decision to vote for a Democratic Congressman, similarly, the coefficient for the underlying desire to vote for a Democratic Congressman has no effect upon the decision to vote for a Democratic President. Both of these coefficients are statistically insignificant (indeed, their asymptotic t-ratio's are both one or less), but even if they were statistically significant, the size of the coefficients would indicate that they would not play much effect in changing the probability of a voter voting for one candidate or another.[15]

Another conclusion to be drawn from table 3.1 is that the model fits the data very well. The $\hat{R}^2$ is close to .8 in both models, and the rate of predicition is around 87% for the presidential race and 83% for congressional races (with misclassifications falling about equally between Democratic voters being classified as Republican voters and Repbublican voters being classified as Democratic voters, for both types of race). The importance of the model fitting well for the purposes of testing the momentum model is that if the model does fit well, in terms of prediction,

15. Obviously, not only the size of the coefficients but also the range of the exogenous variable must be considered. Calculation of the effect of an independent variable on the change in probability in a probit model is perhaps most easily accessible in Wolfinger and Rosenstone (1980, p 121-123). Basically, to obtain the overall increase in probability from the effect of one variable, they suggest using the weights calculated in the probit to estimate the probability of an individual voting with the lowest possible category of the exogenous variable, then his actual probability, and subtracting the former from the latter. This gives a change in probability for every member of the population for that exogenous variable. Aggregation can then be made for any particular sub-group or the whole population, if one wishes.

we can have some confidence that we are really picking up the underlying desire to vote for a Democratic, President or Democratic Congressman[16]. That being the case, we can then be assured that we have given the momentum model a reasonable test.

Other conclusions that might be drawn from the data and the estimations will not be pursued here, for the reason that we are attempting to test two competing theories of coattails rather than examine the voter's decision calculus. We might note that in our sample, though, which excludes uncontested seats, Republican candidates for Congress actually gathered more votes than Democratic candidates for Congress (there were no seats with a Republican unopposed which were in the sample). Also, there is an asymmetric role of long-term partisan attachments in the voting for Congress and voting for President, with Democratic attachments playing a much greater role in voting for President and Republican attachments a much greater role in voting for Congress. Finally, both issues, contacts and (successful) casework make a difference in House races, substantiating the research quoted above.

We turn next to testing the structural motivation model, given in equations (3.1.15) and (3.1.16). For this model, we are allowed to have only one endogenous variable on the right-hand side, which we will assume to be the presidential. This is in accord with the usual definition of the "choice of candidate for a major office [tending] to decide his final

16. A voter is predicted to vote Democratic if $\sum_j x_j \hat{\beta}_j$ is greater than the cutpoint, and to vote Republican if it is less than or equal to the cutpoint. These scores are, however, the same ones which make up the underlying desire to vote instruments.

choices for minor offices" (Meyer, p 53). Given the results of the previous model, we also set the coefficient on the underlying desire to vote for a Democratic President to zero. This leaves us with a model to estimate of the following form;

$$y_1^* = X_1\beta_1 + \varepsilon_1$$

and

$$y_2^* = \gamma_{12}y_1 + X_2\beta_2 + \varepsilon_2,$$

where y_1^* is the underlying desire to vote for a Democratic President, y_1 is coded one if the respondent voted for Carter and zero if the respondent voted for Reagan, y_2^* is the underlying desire to vote for a Democratic Congressman, and X_1 and X_2 are the exogenous variables used above.

The natural impulse in a case like this is to put in the vote for President in (3.2.2) and estimate the model using a probit model. Unfortunately, y_1 may be correlated with the error term ε_2, so a direct substitution may result in biased estimates. The suggestion in the literature (Heckman) is to use an instrument, $\Phi(X_1\hat{\beta}_1)$, for y_1, where Φ is the cumulative distribution function of the normal distribution. The problems with using an instrument is that one can lose a large amount of efficiency. The problem is resolved for us if we estimate (3.2.2) twice, once using y_1 and once using the instrument $\Phi(X_1\hat{\beta}_1)$. These results are reported in Table 3.2.

Table 3.2 Structural Model for Presidential Coattails

Independent variable	Dependent variable – Congressional Vote: Instrument	Non-instrument
Strong Democrat	0.2798	0.3108
Weak Democrat	0.0689	0.1060
Independent Democrat	-0.0267	0.0274
Independent Republican	-0.4747	-0.4495
Weak Republican	-0.6220**	-0.6040**
Strong Republican	-0.7981***	-0.7921***
Inflation		
Democrats better	-0.2617	-0.2624
Republicans better	-0.0325	-0.0355
Unemployment		
Democrats better	0.5302**	0.5635**
Republicans better	-0.2109	-0.2124
War		
Democrats better	0.3867**	0.3668**
Republicans better	0.0664	0.0606
Black	0.5966*	0.5878*
Male	0.0564	0.0650
Opinions of and contacts with House candidates–Democratic		
General contact with House candidate	0.8382***	0.8669***
Wrote opinion and was satsified with response	0.3017	0.3058
Asked information and was satisfied with response	-0.2713	-0.2392
Asked for help and was satisfied with response	1.4069**	1.4434**
Preferred candidate because of issue in race	0.9627**	0.9258**
Incumbent candidate running for office	0.4222*	0.4128*
Opinions of and contacts with House candidates–Republican		
General contact with House candidate	-0.5686***	-0.5917***

Wrote opinion and was satsified with response	-1.0771*	-1.0411*
Asked information and was satisfied with response	-0.4173	-0.5210
Asked for help and was satisfied with response	-0.6129	-0.6527
Preferred candidate because of issue in race	-2.7355***	-2.7833***
Incumbent candidate running for office	-0.1755	-0.1931
Instrument for vote for Democratic President	0.6182	
Actual vote for Democratic President		0.6206***
$\hat{R}^2$	.763	.769
-2*LLRatio	456.88	465.08
n	677	677
Democratic votes correctly predicted	270 (81.8%)	274 (83.0%)
Democratic votes incorrectly predicted	60 (18.2%)	56 (17.0%)
Republican votes correctly predicted	295 (84.0%)	295 (83.0%)
Republican votes incorrectly predicted	52 (16.0%)	52 (17.0%)

*$p < .10$. **$p < .05$. ***$p < .01$.

The results of Table 3.2 makes it clear that y_1 is not correlated with the error term for (3.2.2), as the coefficients for both y_1 and $\Phi(X_1\hat{\beta}_1)$ are nearly identical, as are all other coefficients. This implies that inferences we draw from using the presidential vote in the (3.2.2) are as valid as those drawn from using the instrument (the instrument itself is nearly significant at the .10 level, having an asymptotic t-ratio of 1.5. With a one-tailed rather than a two-tailed test, it would be significant at the .10 level). This, then, is strong support for the hypothesis of a structural motivation for coattail voting.

We can also test the hypothesis that individuals tend to associate votes because they see the parties as representing, to some extent at least, entities which have particular policy goals which they will work together in order to carry out. In 1980, five questions were asked with regard to individuals attitudes towards the political parties. While all bear on some degree or another on this concept, the one which bears the closest resemblence is question 359, "The parties do more to confuse the issues than to provide a clear choice on the issues." It follows that if an individual thinks that the parties do provide a clear choice, he will believe that their is a certain congruence of goals between the members of the party in the legislative and executive branch (this does not necessarily mean he thinks they will work together, but it is a necessary precondition). On the other hand, if he feels that the parties do not provide a clear choice on the issus, there is less reason for him to associate his vote.[17]

17. A question (number 360) such as "It would be better, if, in all elections, we put no party labels on the ballot" could very well have a completely different response pattern from ques-

We carry out such an estimation as follows. The question on the coherence of the party system has seven categories, and (given the way the question is phrased), those disagreeing with the question support the idea of party coherence the most and those agreeing with it least support that concept. Therefore, we broke the responses down into three categories; one category for those who thought the parties provided a clear choice on the issues, one for those who were exactly in the middle, and one for those who did not think the parties provided a clear choice.[18] These dummy variables were then multiplied by the actual presidential vote (one if a vote for Carter, zero if a vote for Reagan), and entered into the probit estimation of (3.2.2). The same equation was then estimated with presidential vote coded as one if the respondent voted for Reagan and zero if he voted for Carter. This was done so both voters for Carter and for Reagan could be included in the analysis.

The results are given in Table 3.3.

---

tion 359, since simply because one thinks that the parties do offer a clear choice on the issues does not mean that one thinks the labels should be on the ballot (one may feel that citizens should inform oneself on the positions of each candidate before the election, for example, and not allow poor candidates to "hide" behind a party label). Similar objections apply to the other three questions of attitudes towards the parties.

18. A special category was created for those who were in the middle of the seven-point scale, i.e., those who responded four, primarily because of suspicions that these individuals might include not only those who were halfway on the matter but others who had no real opinion but were merely answering the question.

Table 3.3 Structural Model with Presidential Vote Broken Down by Party Coherence

Independent variable	Dependent variable—Congressional Vote: Democratic One	Dependent variable—Congressional Vote: Republican One
Strong Democrat	0.3529	0.3664
Weak Democrat	0.1096	0.1616
Independent Democrat	0.0231	0.0521
Independent Republican	-0.4499	-0.4429
Weak Republican	-0.6147**	-0.6212**
Strong Republican	-0.7808***	-0.7832***
Inflation		
Democrats better	-0.2205	-0.2274
Republicans better	-0.0492	-0.0524
Unemployment		
Democrats better	0.5469**	0.5714***
Republicans better	-0.2090	-0.2168
War		
Democrats better	0.3500*	0.4106**
Republicans better	0.0683	0.0618
Black	0.6070	0.6498*
Male	0.0746	0.0442
Opinions of and contacts with House candidates—Democratic		
General contact with House candidate	0.8894***	0.8573***
Wrote opinion and was satsified with response	0.2566	0.2153
Asked information and was satisfied with response	-0.2642	-0.1784
Asked for help and was satisfied with response	1.4794**	1.4804***
Preferred candidate because of issue in race	0.9297**	0.9979**
Incumbent candidate running for office	0.3771	0.4504*
Opinions of and contacts with House candidates—Republican		
General contact with House candidate	-0.6030***	-0.5837***

Wrote opinion and was satsified with response	-1.1750*	-0.9894*
Asked information and was satisfied with response	-0.6366	-0.5089
Asked for help and was satisfied with response	-0.7346	-0.6393
Preferred candidate because of issue in race	-2.8547***	-2.7953***
Incumbent candidate running for office	-0.2263	-0.1743
Vote for Democratic President—parties offer choice	0.8580***	
Vote for Democratic President—parties maybe offer choice	0.6566**	
Vote for Democratic President—parties don't offer choice	0.4528**	
Vote for Republican President—parties offer choice		-0.5920***
Vote for Republican President—parties maybe offer choice		-0.5396***
Vote for Republican President—parties don't offer choice		-0.3475**
$\hat{R}^2$	.777	.770
-2*LLRatio	466.80	462.23
n	677	677
Democratic votes correctly predicted	276 (83.6%)	273 (82.7%)
Democratic votes incorrectly predicted	54 (16.4%)	57 (17.3%)
Republican votes correctly predicted	297 (85.6%)	300 (86.5%)
Republican votes incorrectly predicted	50 (14.4%)	47 (13.5%)

*$p < .10$. **$p < .05$. ***$p < .01$.

The results are supportive of the hypothesis, namely, people who see party's as offering a choice are more likely to associate their vote for President and Congressman. The signs on the coefficients for the "parties offer choice" dummy times the presidential vote is the largest in magnitude, decreasing for the "maybe offer choice" category and having its smallest value for the "parties don't offer choice" category. A test of the differences of the coefficients between the "parties offer choice" and "parties don't offer choice" gives an asymptotic t-ratio of about 1.2 for both voters for Carter and for Reagan. Since we are imposing this test on top of partisan affiliations already being controlled for in the probit equation, it is likely that the effect we are attempting to identify is very small. One does have more confidence in it since the sign pattern that is expected shows up in both equations and the t's in both are close to significance.

Finally, we may attempt a direct test of two of the three motivations suggested for coattails voting. The two are whether a presidential candidate's charm is an important part of the decision process of a voter, and whether one's evaluation of the presidential candidate has a coattail effect. For the charm motivation, we turn to the questions in the 1980 ANES on candidate qualities. Questions are asked on such qualities as "moral", "dishonest", "weak", "knowledgable", "power-hungry", "inspiring", "he would solve our economic problems", "he would provide strong leadership", and "he would develop good relations with other countries". Of these qualities, the one that best seems to fit the traditional theory of coattails as personal appeal or charm was that of "inspiring". Therefore, the variables inquiring as to whether the particular candidate was inspiring were made into dummy variables and added to the structural motivation model. For the question of whether the candidate evaluations are a significant part of the decision process, we construct Calvert and Ferejohn's (1983) CAND variable and estimate the structural model with it included.[19]

The results of these two direct tests of suggested motivations are given in Table 3.4. For the charm test, none of the inspiration variables were near to significant (some had the wrong signs), and the coefficients on the other exogenous variables in the model did not change significantly. For the candidate evaluation test, the coefficient was insignificant for CAND and quite small, and the coefficients on the other exogenous variables remained comparable to previous specifications. In all cases, presidential vote remained significant.[20] We therefore see these results as confirming our previous rejection of the magnitude hypotheses of coattails.

---

19. The CAND variable was first used in Campbell, Gurin, and Miller (1954). It is constructed by using responses to the question "Is there anything in particular about (candidate) for president) that might make you want to vote (for/against) him?". The number of responses (there may be up to five in all SRC's survey's during a on-year except 1972, when there were three) for liking the Democratic candidate are added to the number of responses for disliking the Republican candidate, and then the number for disliking the Democratic candidate and liking the Republican candidate are added together and this number subtract from the first number. The number may thus run from -10 to 10. Calvert and Ferejohn truncate this at -6 or +6, but we allow it to run over the full range.
20. Though the standard error on presidential vote did increase a good deal when CAND was entered, indicating that they do share some common factor. Given the results of Chapter 6, which indicates that CAND may be considered an endogenous proxy for vote, this would not be surprising.

Table 3.4 Structural Model with Presidential Charm

Independent variable	Dependent variable—Congressional Vote	
	Charm included	Cand included
Strong Democrat	0.3281	0.2984
Weak Democrat	0.1516	0.1014
Independent Democrat	-0.0094	0.0369
Independent Republican	-0.4650	-0.4419
Weak Republican	-0.5655*	-0.6080**
Strong Republican	-0.7405**	-0.7655***
Inflation		
Democrats better	-0.2548	-0.2798
Republicans better	-0.0589	0.0058
Unemployment		
Democrats better	0.4336**	0.5465**
Republicans better	-0.2448	-0.1913
War		
Democrats better	0.4429**	0.3229*
Republicans better	0.0753	0.0882
Black	0.5633	0.5907*
Male	0.0705	0.0738
Opinions of and contacts with House candidates—Democratic		
General contact with House candidate	0.9116***	0.8837***
Wrote opinion and was satisfied with response	0.3937	0.3038
Asked information and was satisfied with response	-0.2751	-0.2874
Asked for help and was satisfied with response	1.4888**	1.4614**
Preferred candidate because of issue in race	1.0326**	0.9233**
Incumbent candidate running for office	0.3870	0.3956
Opinions of and contacts with House candidates—Republican		
General contact with House candidate	-0.6483***	-0.6111***

Wrote opinion and was satsified with response	-0.8383	-1.0613*
Asked information and was satisfied with response	-0.5631	-0.5402
Asked for help and was satisfied with response	-0.7315	-0.6414
Preferred candidate because of issue in race	-2.7139***	-2.7820***
Incumbent candidate running for office	-0.2410	-0.1964
Vote for Democratic President	0.5455***	0.5331***
Inspiring describes Reagan extremely or quite well	-0.2734	
Inspiring describes Reagan not too well or not well at all	0.1402	
Inspiring describes Carter extremely or quite well	-0.5431	
Inspiring describes Carter not too well or not well at all	-0.1926	
Cand		0.0301
$\hat{R}^2$	.776	.771
-2*LLRatio	478.32	466.24
n	677	677
Democratic votes correctly predicted	273 (82.7%)	276 (83.6%)
Democratic votes incorrectly predicted	57 (17.3%)	54 (16.4%)
Republican votes correctly predicted	300 (85.4%)	300 (85.4%)
Republican votes incorrectly predicted	47 (13.6%)	47 (13.6%)

*$p < .10$. **$p < .05$. ***$p < .01$.

3.3 The Coattails Effect and American Elections

We have demonstrated with some degree of plausibility the existence of another motivation for coattail voting by individuals in the American electorate; that of a motivation to overcome the separation of powers in the American Federal system. Hamilton, among others, saw the possiblity of one electorate controlling all representative functions, both legislative and executive, and the Constitutional Convention took steps to prevent the occurence. If one group of electors did obtain control of the election of both the legislative and executive branches, in contravention of the plans of the Constitutional Convention, is it likely that they would dilute their voting power by voting for representatives with differing policy goals for different branches without good reason?

Hamilton feared that they would not, certainly, yet the amount of split ticket voting is today at an all-time high. Before we remark upon this, we might consider Miller's conditions for an analysis of coattail voting–"(a) those straight ticket [that is, those who voted for a President and Congressman of the same party] voters whose presidential voting motivations (b) included an attraction for the presidential candidate, and whose congressional motivations either (c) did not include a preponderance of factors uniquely related to the congressional race *or* (d) did include only factors impelled by the existence of the presidential candidate and his campaign." Looking at these conditions, we would claim that we have fulfilled them, something Miller was unable to do at the time (1955) he wrote his article, since the technology did not exist then. What we have shown is that attraction does not matter--rather, choice is what produces the association of presidential and congressional vote. As this association tends to increase as one perceives the parties offering a clear choice on the issues, we then conclude that the individuals do take into account the Federal nature of the American system when making their voting choice.

They take account of much more, though, as our congressional specification makes clear. Issues are very important, but only 16% of our sample prefers one candidate or another because of the issues (it might be pointed out that that almost all of the voters who had a candidate preference because of an issue in the House race voted for the President of the same party as that preference in the House race). This means that the other 84% of the voters in our sample must get along with criteria such as contacts, partisan cues, and perceptions of the abilities of the parties to bring about desirable states of the world. Contacts become a very important feature, for if one is helped by a Democratic incumbent, one is almost certain to vote for the Democratic incumbent. The incumbency effect itself, while still significant for Democratic incumbents, does not have a large sign for either Democratic or Republican incumbents. We would expect that in a completely specified model the coefficients on incumbency would go to zero.

From the viewpoint of political science as a science, it might be interesting to speculate on why it is believed, both among political professionals and political scientists, that magnitude is the basis of coattails voting. It would be difficult to perceive of a policy-oriented reason for allowing one's perhaps overwhelming attraction to one candidate to influence one's vote for another. A reason might be proposed on the basis of some social psychological dissonance theory, but it should be remembered that cognitive dissonance depends above all on choice, not underlying motivation. The association of the vote that we have found in American Federal elections should, in any case, have at least the possibility of existing in all types of elections where there is more than one office to be filled at any one time. We are aware of no theoretical work, nor empirical work aside from the work on coattails, which even addresses this problem. It would seem to be a useful field in political science to examine.

We spoke earlier (footnote 6) of the ability of the presidential candidates to put issues on the national agenda, given that the large majority of attention of the media is placed upon them. If the presidental candidate and his supporters emphasize the partisan nature of the contest, it would be expected that the number who see significant differences between the parties would increase. This would increase, then, the coattails effect, through the process examined in Table 3.3. If the presidential candidates chose to ignore the partisan aspects of the campaign and instead ran on a campaign devoted to personalities, one would expect the number seeing significant differences in the parties to decrease. The change in campaign styles, along with the insularity of congressmen from national trends noted by Burnham, has probably contributed in the change in coattails. It should be noted that while in 1964, 37 Republican incumbents were defeated, in 1980, 27 Democratic incumbents were defeated. Not as many, perhaps, but still a considerable number. The evidence given in this paper indicates that it is likely that the nature of the campaign increased the saliency of party differences and helped bring about a slight increase in "structural" voting through increased perceptions of differences in the parties.

Another point relating to the saliency of the campaign and the issues within it is the degree to which issues become incorporated in partisan attachments. In Table 3.1, disapproval of Jimmy Carter's performance on both unemployment and Iran and perceptions of the Republicans being better on inflation and unemployment all had highly significant coefficients which would increase the probability of voting for Reagan. On the other hand, the only coefficient which was positive and significant on the Jimmy Carter approval and Democrats better was the Democrats being better on unemployment. Rather, the coefficients for long-term partisan attachments (self-placement as a strong Republican, weak Republican, etc.) were much greater for Democrats in the presidential campaign than in the congressional campaign, where the coefficients on long-term partisan attachments were shifted to the Republican favor. Are different things happening? It seems unlikely; rather, individuals just had so many avenues to express their dissatisfactions on the Carter presidency that they associated their discontent with particular issues when voting for President. In the congressional case, that dissatisfaction did not express itself in particular issue positions but rather in partisan attachments, with the coefficients on the long-term partisan attachments favoring the Republicans much more than in the presidential races. If we recall Miller's definition of coattails as including those who had congressional motivations which were produced by the presidential campaign, we cannot rule out that individuals have been affected in their general orientations by that campaign. This is obviously a subject requiring more study, though it is difficult to see how one would differentiate these effects, given the broad nature of Miller's hypothesis. What we have done here is rule out several other hypotheses of the nature of coattail voting, but this one still stands as one in need of further study.

While obviously this analysis, as it stands now, cannot be extended back across time, we can perhaps engage in some informed speculation. First, it is clear that structural motivation effect, as we have called it, is not insignificant. Say a voter was an independent Republican who took on no other values on the other exogenous variables. Then the probabilty that this voter casts a ballot for a Republican Congressman, given that he has voted for Reagan, is simply $Pr[-.43 + u < .11]$ (where .11 is the cutpoint), or around .7. On the other hand, if this elector cast a ballot for Carter, the probability would be only around .48. Given a large electorate, this type of behavior will add up. On the other hand, if this individual has been helped by a Democratic incumbent and he voted for Reagan, the probability of his voting for a Republican Congressman is only around 9%. Even contact with a Democratic Congressman or candidate without a corresponding contact with the Republican candidate means that our independent Republican, having voted for Reagan, will have only a 16% probability of voting for a Republican Congressman.

So it is not that the basis of coattails has vanished, though it has probably been decreasing over time. We can easily envision a vicious cycle where voters see party not as important in presenting alternatives simply because it isn't, given that Congressmen are concentrating on enhancing their chances for reelection through casework. Thus their propensity to associate their congressional vote with their presidential vote will decrease over time, creating a Congress even more independent of government policy, leading voters to view the parties as even less meaningful in presenting clear choices. As of this moment, though, there is still a strong and meaningful tendency for voters to associate their Congressional and Presidential votes. The problem is that it is so easily overwhelmed by other factors unrelated to policies under which the country is run.

Abramowitz, Alan I.. "A Comparision of Voting for U.S. Senator and Representative in 1978." *American Political Science Review* 74 #3 (1980): 633-640.

Amemiya, Takeshi. "The Estimation of a Simultaneous Equation Generalized Probit Model." *Econometrica* 46, # 5 (September, 1978): 1193-1206.

Cain, Bruce. "Strategic Voting in Britain." *American Journal of Political Science* 22 #3 (1978): 639-655.

Buchanan, Christopher. "Republicans Make Substantial House Gains." *Congressional Quarterly Weekly Report* 38, # 45 (November 8, 1980): 3317-3321.

Burnham, Walter Dean. "Insulation and Responsiveness in Congressional Elections." *Political Science Quarterly* 9, # 3 (Fall, 1975): 411-435.

Calvert, Randall L. and Ferejohn, John A. "Presidential Coattails in Historical Perspective." Social Science Working Paper # 343, California Institute of Technology, 1980 (forthcoming in the *American Political Science Review*).

Calvert, Randall L. and Ferejohn, John A. "Coattail Voting in Recent Presidential Elections." (forthcoming in the *American Journal of Political Science*).

Cook, Rhodes. "Presidential Coattails of Little Benefit." *Congressional Quarterly Weekly Report*, 34, # 45 (November 6, 1976): 3135.

Cummings, Milton. *Congressmen and the Electorate: Elections for the U.S. House and the President, 1920-1964.* New York: Free Press, 1966.

Erikson, Robert. "The Advantage of Incumbency in Congressional Elections." *Polity* 3 (Spring 1971): 395-405.

Fiorina, Morris P. *Congress: Keystone of the Washington Establishment.* New Haven: Yale University Press (1977).

Fiorina, Morris P. "Some Problems in Studying the Effects of Resource Allocation in Congressional Elections." *American Journal of Political Science* 25 #3 (1981): 543-567.

Heckman, J. J. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46 (1978): 931-959.

Hinckley, Barbara. *Congressional Elections.* Washington, DC: Congressional Quarterly Press (1981).

Hinich, Mel. "Voting as an Act of Contribution." *Public Choice* 36 # 1 (1981): 135-140.

Kritzer, Herbert M. and Robert B. Eubank. "Presidential Coattails Revisited: Partisanship and Incumbency Effects." *American*

Journal of Political Science 23 (1979): 615-626.

Mann, Thomas E. and Raymond E. Wolfinger. "Candidates and Parties in Congressional Elections." *American Political Science Review* 74 #3 (1980): 617-632.

Meyer, John W. "A Reformulation of the 'Coattails' Problem." In William N. McPhee and William A. Glaser, eds. *Public Opinion and Congressional Elections*. New York: Free Press, 1962.

Miller, Warren. "Presidential Coattails." *Public Opinion Quarterly* 19 (1955): 353-368.

Moos, Malcolm. *Politics, Presidents, and Coattails*. Baltimore: John Hopkins Press, 1952.

Moreland, William B. "Angels, Pinpoints and Voters: The Pattern for a Coattails." *American Journal of Political Science* 17 (1973): 170-176.

Schmidt, Peter. "Constraints on the Parameters in Simultaneous Tobit and Probit Models." In Charles F. Manski and Daniel McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Massachusetts: MIT Press, 1981.

Stokes, Donald and Miller, Warren. "Party Government and the Saliency of Congress." *Public Opinion Quarterly* 26 (1962): 531-546.

Yiannakis, Diana Evans. "The Grateful Electorate: Casework and Congressional Elections." *American Journal of Political Science* 25 #3 (1981): 568-580.

Chapter IV: The Relationship of Thermometers and Open-ended Evaluations to the Vote

The selection of variables with which to test a model in political science is always among the most difficult problems a researcher faces. One must decide which are to be the endogenous (or dependent) variables, what are to be the exogenous (or independent) variables, and the statistical regime by which these variables are related. The only aspect of the situation which makes the venture possible at all is that presumably there is a model, generated by theory, with which one wishes to test certain hypotheses.[1] There are often aspects of this model which may be exploited in order to obtain some insight into the central problem raised above, that of variable selection. It is the pupose of this chapter to introduce a general technique for examining the relationship of variables to one another within the context of some commonly used models in political science. Specifically, it is a technique for testing whether one variable which is assumed to be endogenous may be used as a proxy for another variable which is assumed also to be endogenous. The word proxy is used here in the sense of substitute.[2] With this technique, an

1. One may argue convincingly that there can be no demonstration of causality without experimental control but it would seem to be indicative of causality if one were to hypothesize a model, apply the statistical analyses which are indicated by that model, and observe that the results of those analyses are consonant with that model.

2. Another possible term is an instrument, but ordinarily in econometrics, an instrumental variable is an "independent" variable which is highly correlated with another "independent" variable but uncorrelated with the residual term. Two-stage least square theory may be developed using this concept, for example. A proxy variable is also usually taken to mean an "independent" variable also (one that can be correlated with the error—see Maddala, p 158), but it does not seem an abuse of nomenclature to refer to such variables as proxy variables. It is certainly less objectable than instrument, which has a specific econometric meaning.

examination will be conducted as to whether certain variables in the American National Election Survey (ANES) may be considered proxies for others, specifically, as to whether suitable functional forms of feeling thermometers and the open-ended questions relating the candidates may be considered as proxies for the voting decision. The results of such an investigation should prove to be worthwhile for the simple reason that researchers have used the aforementioned variables both as proxies and as independent variables in linear regressions, and as should be obvious and will be demonstrated later on, they cannot be both.

"No one knows what thermometer scores measure" states Fiorina (1981,p154).[3] As may be, but this has not kept researchers from using such variables. The 1968 ANES was the first such survey to introduce the social psychological tool known as a "feeling thermometer". Since then, feeling thermometers have been used as a measure of affect (Miller and Miller (1976)), as a proxy for candidate evaluations (Repass, (1976)), as a proxy for candidate preference (Brody (1976)), Black (1976), Kiewiet (1980), Adkison (1982)), and as a proxy for the vote in a simultaneous equation system (Page and Jones (1977)). Similarly, the open-ended questions have also seen yeoman's duty, being used in various manners; as a measure of candidate evalutaions by Repass (1976), as independent

3. The actual context: "Some colleagues have suggested, in particular, that addition of the CPS thermometer scores to the vote equations that follow would provide an extra control for uncaptured references relating to candidate personality. No doubt it would. It would also provide an extra control for retrospective judgments, party influence, issue positions, and anything else that contributes to evaluations of candidates. No one know what thermometer scores measure. While they may be of use as continuous surrogates for vote choice or as measures of candidate preference, they raise grave difficulties when used as explanatory variables. Their contribution to explanations of voting behavior is purely statistical, not substantive."

variables (as the basis of the SRC's famous six-factor model of voting behavior, see Campbell, Converse, Miller and Stokes (1960)), as functions of underlying unobserved variables which determine the vote (Miller and Miller (1976)), and as a measure of presidential candidate attractiveness in house races (Calvert and Ferejohn (1982)). Clearly, the use of such variables is extremely popular--the question remains, which uses are valid and which are not?

To answer that question, one must examine the following. First, a discussion of what a proxy is and how one endogenous variable may be tested as a proxy for another endogenous variable is pursued. It is intuitively clear that in some sense a proxy must contain no different information than the original variable, and this intuition is given statistical meaning. Given this definition, then, different methods of hypothesis testing are examined. One type of test, based on the results of Cox (1961) and White (1982), are derived in this paper for the first time and it is argued that this type of test is more versatile than other types of tests.

Second, the voting decision and the assumptions implicit when it is modelled by political researchers must be examined. Frequently, however, individual researchers do not bother to exposit the underlying theoretical model which their work assumes, or their exposition is such that other scholars are less than sanguine about correctly interpreting it.[4] This problem is handled by assuming a model which is consistent with

4. Popkin, et. al. (1976, p 779, ft 1) quote Burnham to the effect that "'it is fair to say that there is no single place in any of this immense (SRC) corpus of work where a comprehensive

practically all of the research on voting behavior which uses the general linear model or some variant of it (such as probit analysis). Emprical estimation is then carried out using the methods of estimation chosen and the statistical tests applied.

Third, given the results of the tests carried out in this chapter, recommendations are made to the researcher using the ANES data sets. These recommendations can be summarized in one sentence: the use of certain functional forms of variables such as thermometers and the open-ended questions as exogenous variables is contraindicated. This is because they behave so much like the vote that any substantive interpretation of their effect is impossible. An examination of previous research which has used the questionable functional forms of such variables is examined, with an eye to interpretation in light of these results.

4.1 Comparisions of Presther and Zcand with the Vote (1980)--the Zero-order Story

In this chapter an examination is made as to the relationship between feeling thermometers, zcand and the voting decision in the 1980

set of theoretical propositions concerning voting behavior is laid down, with each proposition linked with the next in a closed analytical framework. The problem therefore arises that any effort by an outside scholar to provide an approximation to such a summary is likely to be challenged as misleading or incomplete' (Burnham ms. pp. 4-5)." Popkin et. al.'s strategy is obvious--they are trying to avoid this challenge by observing that other scholars have had the same problem that they have in reconstructing the SRC work (from the same footnote, "In citing this and other works from the SRC corpus, we are attempting to outline the ***central verbal model*** underlying the work.") Incidentally, it didn't work. See Miller and Miller (1976, p833), "This claim is something of a misrepresentation of the facts and it certainly ignores the original version of "Majority Party" as well as previously published CPS research on the 1972 election."

election. An analysis file of the 1980 ANES survey created by Mo Fiorina was used. Voters for Anderson were excluded from the analysis, as were non-voters. Presther is defined throughout as the difference of the two thermometer scores for president, Carter minus Reagan, so the maximum score possible was one hundred and the minimum score was minus one hundred. Ninety-seven percent of the sample fell between the maximum and the minimum.

The third variable, zcand, is seemingly less familiar and deserves a more detailed examination. Zcand is a variable constructed from the likes and dislikes questions asked about the presidential candidates. The sum of likes for the democratic candidate and dislikes for the republican candidate, minus the sum of likes for the republican candidate and dislikes for the democratic candidate, is zcand. What is being referred to here as zcand has also served in some forms as the basis for the candidate evaluation variable in the famous SRC six-factor model of voting behavior. A description is given in Miller and Miller (1976,Rejoinder, ft 8)

> "The candidate and party affect measures used in the regressions were based on a series of questions asking respondents what they liked and disliked about each party and candidate. The positive and negative comments about each party and again for each candidate were summed to provide two indices, one for the parties and the other for the candidates. The final indices are 'net' affect measures, that is, they indicate the preponderance of positive or negative attitudes toward the Republican candidate or party over those toward the Democratic candidate or party."

This is also the candidate evaluation variable used in ***The American***

Voter.[5]

As a means of preliminary examination, the proportion voting for Carter was graphed against the Carter minus Reagan thermometer and zcand. For presther, responses were coded into categories, with the first category consisting of all individuals scoring a minus 100, the second interval all those between and including minus 99 and minus 90, and so forth, with the last category all of those scoring 100. This graph is shown in Figure 4.1. As can be seen, the graph looks very much like the cdf of a symmetric distribution which has narrow tails (such as a normal cdf).

Similarly, the proportion voting for Carter was graphed against zcand, as is shown in Figure 4.2. Here, as zcand takes on only twenty-one possible values, each category was simply the value of zcand. The graph of this looks even more like a normal cdf, with fewer deviations from the curve. Finally, in Figure 4.3, presther is plotted against zcand in a scattergram. The scatter of these points is nearly perfectly symmetrical around the regression line, the only deviations coming where there are large values of the thermometer but the zcand score score is not quite as large. In Figure 4.4, the standardized presther (standardized here means an observation has the mean of the observations subtracted from the observation and this figure is divided by the standard deviation of the observations) is subtracted from the standardized zcand and plotted against zcand. The symmetry of these "residuals" is once again evident,

5. By inference, anyway. See Chapter 3 and 4 on "Perceptions of Parties and Candidates" and "Partisan Choice".

but the fact that zcand and presther are both bounded means that for large positive values of zcand, negative values of the "residuals" are impossible, while for large negative values of zcand, positive values of the "residuals" are not possible. If one merely looks at the "residuals" for values of zcand between and including -3 and 3, it can be seen the values of the "residuals" are nearly uncorrelated with zcand (pearson r equal to .30 for all zcand (n=832), -.10 for zcand between and including -3 and 3 (n=505)).

From the figures presented in this section, then, it is simple to see that there is a great deal of similarity between the vote, presther, and zcand. The question then becomes what is the nature of that similarity and what conclusions may be drawn from that similarity? To answer those questions, it is necessary to define what is meant by a proxy when both variable are endogenous. That task is performed in the next section.

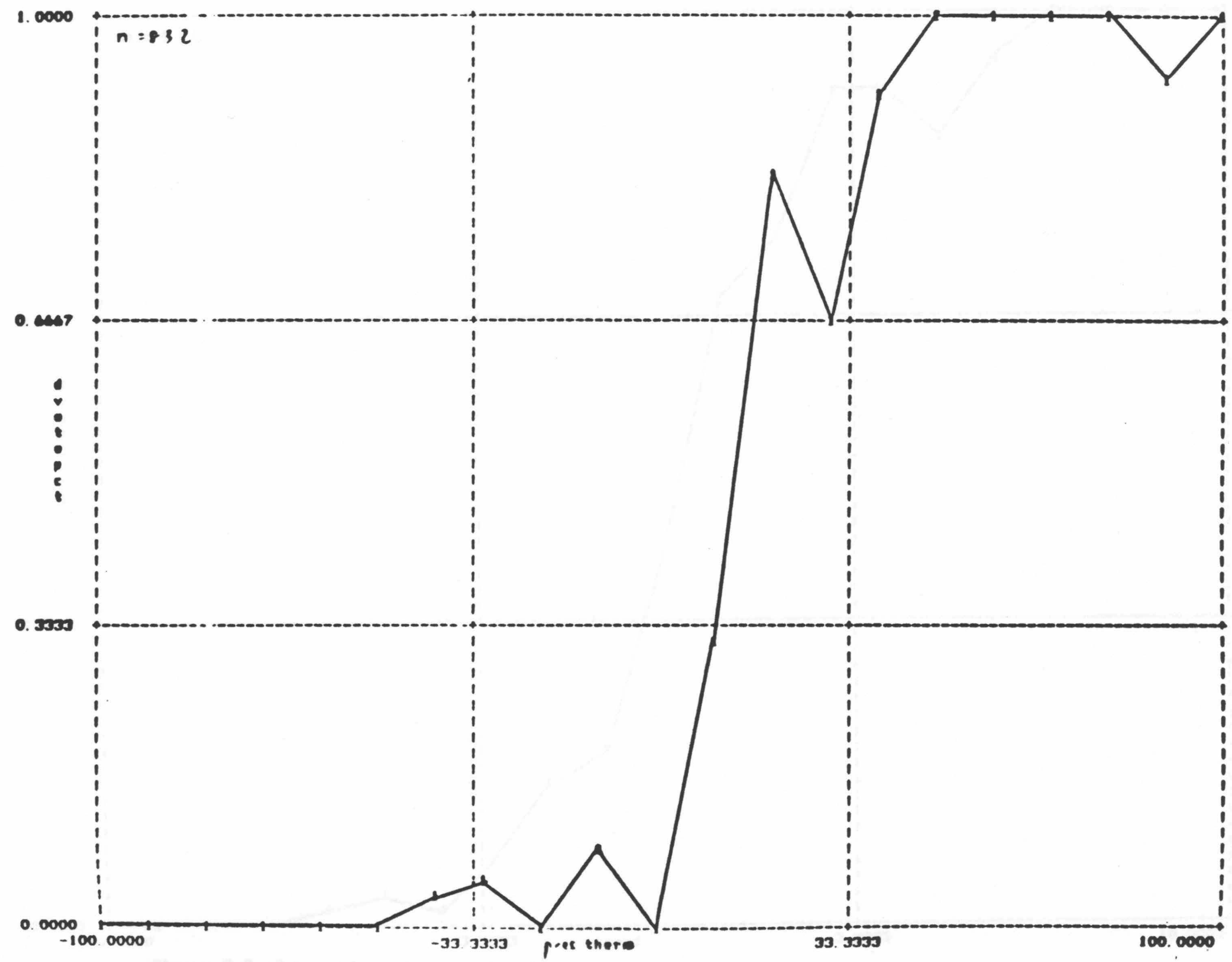


Figure 4.1 Proportion of respondents voting for Carter (1980) by presther

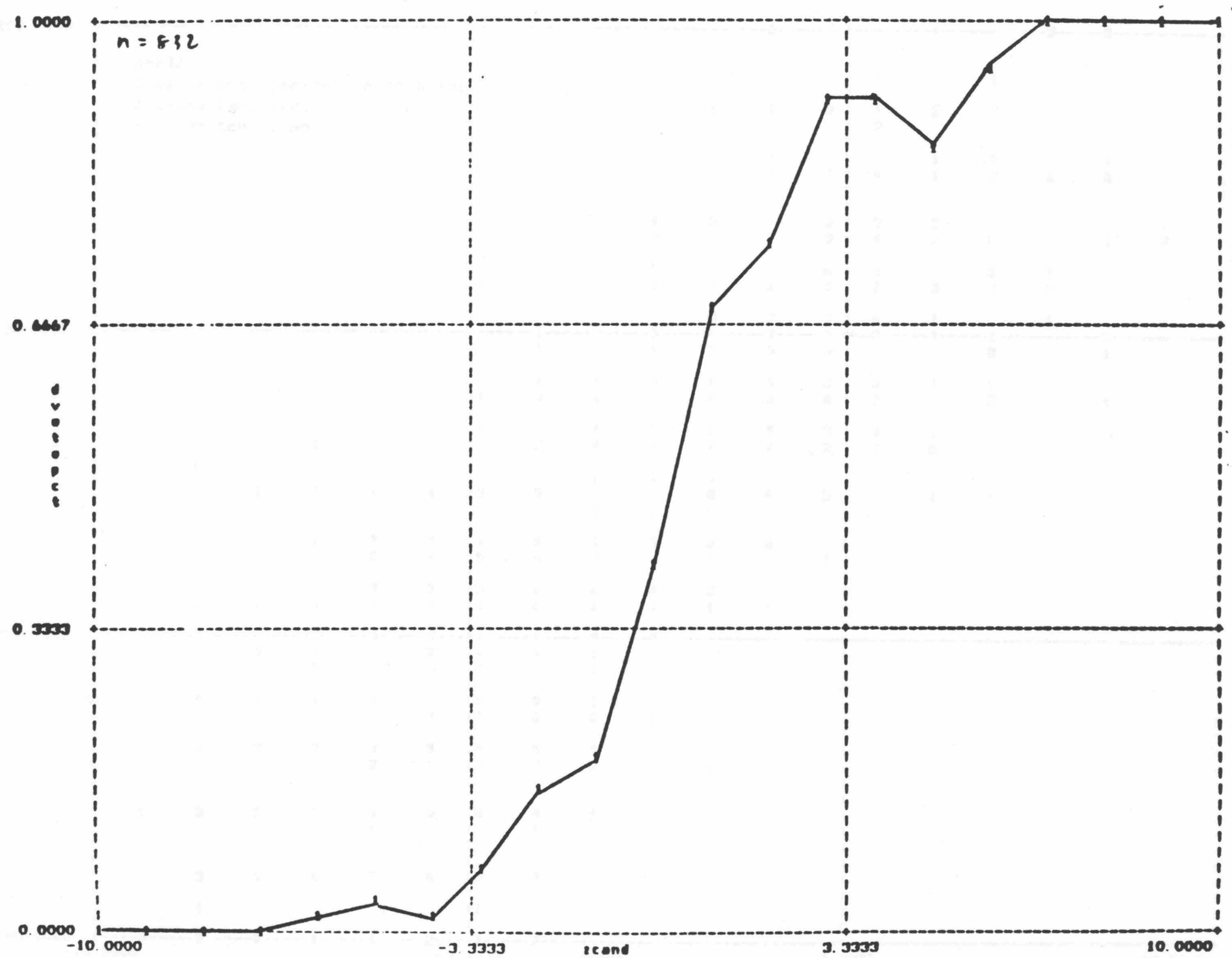


Figure 4.2 Proportion of respondents voting for Carter (1980) by zcand

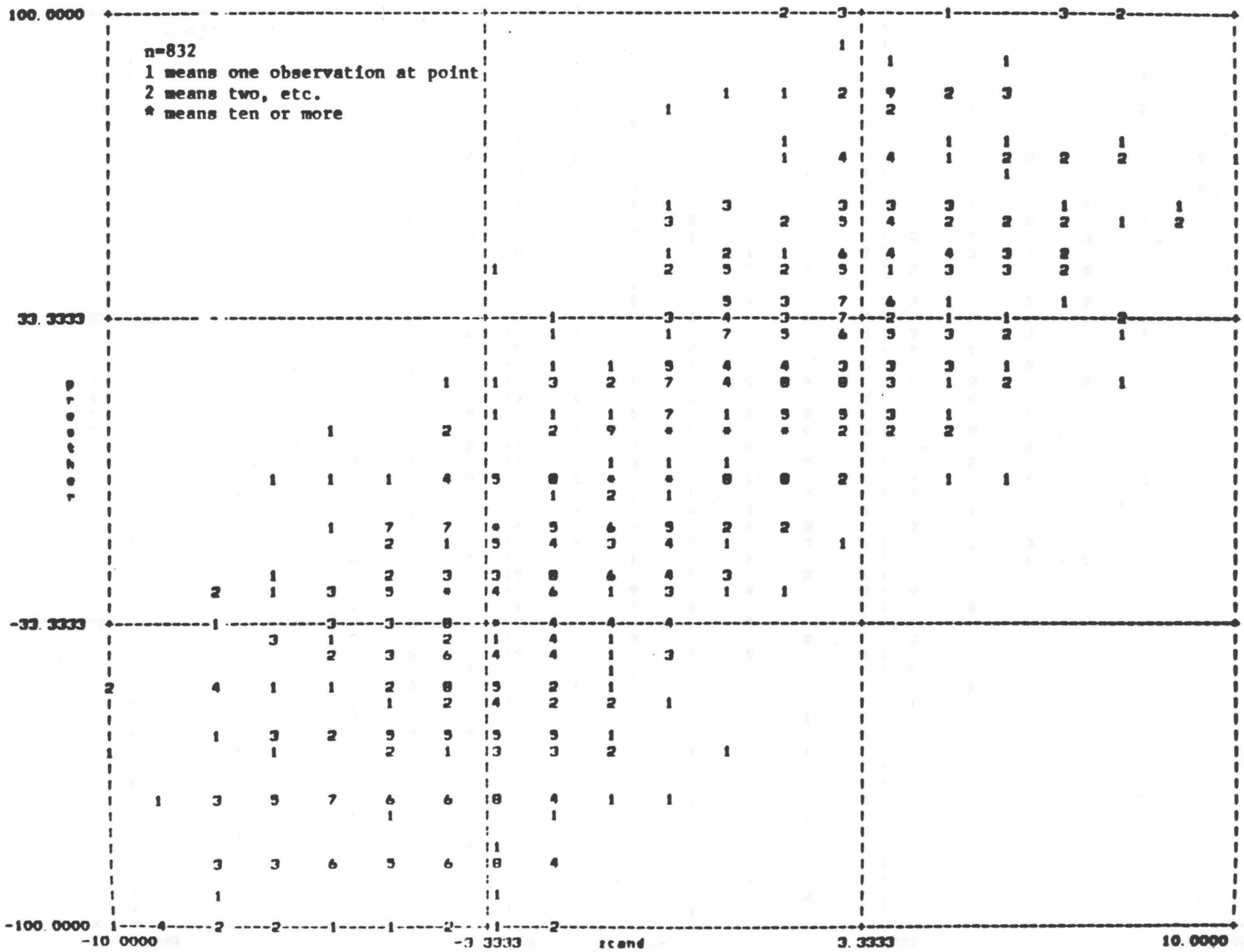


Figure 4.3 Presther by zcand

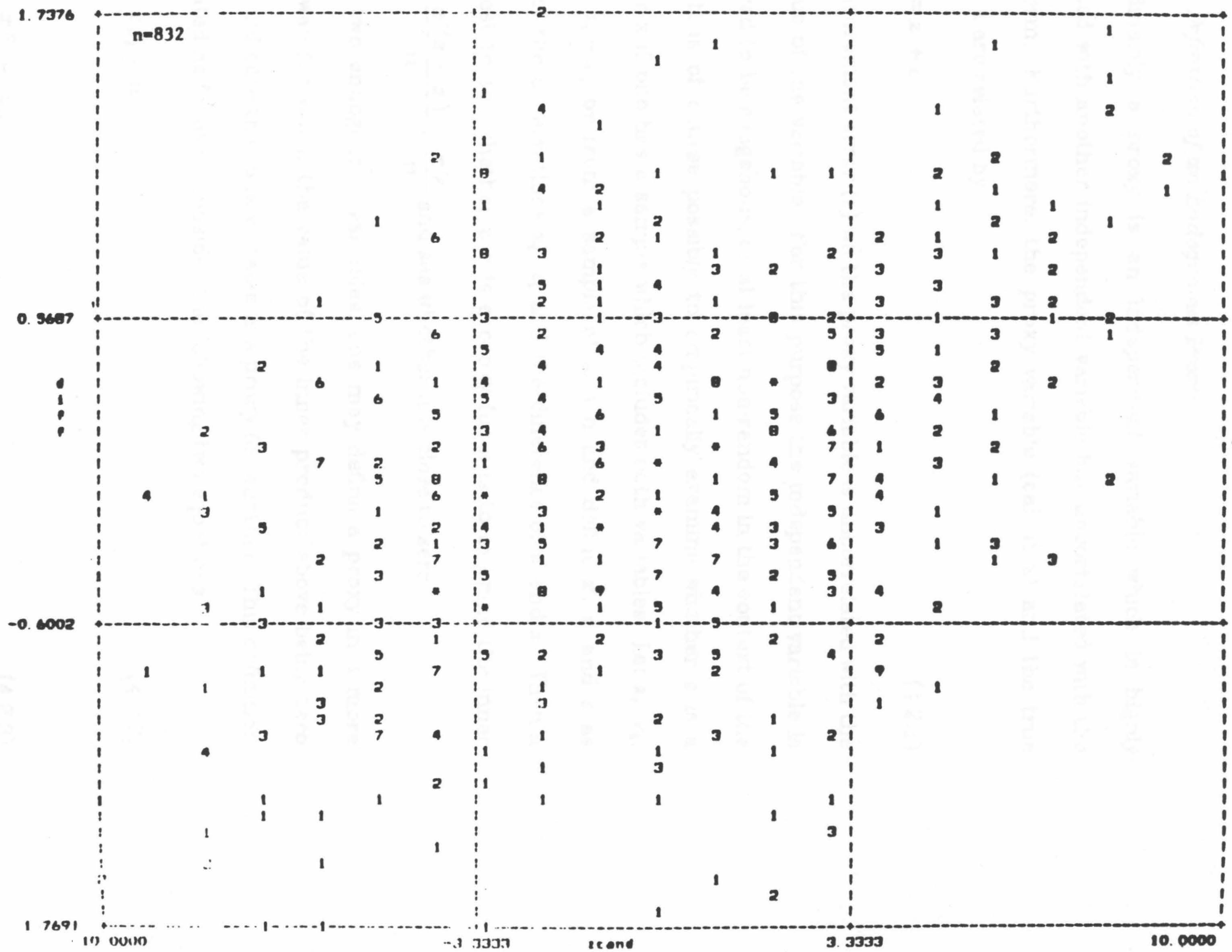


Figure 9.4 Standardized zcand minus standardized presther by zcand

4.2 The Definition of an Endogenous Proxy

Ordinarily a proxy is an independent variable which is highly correlated with another independent variable but uncorrelated with the error term. Furthermore, the proxy variable (call it z) and the true variable x are related by

$$z = x + \varepsilon \tag{4.2.1}$$

The measurement error (ε) on the proxy variable is uncorrelated with the true value of the variable. For this purpose the independent variable is considered to be exogenous, or at least non-random in the context of the model. It is of course possible to empirically examine whether z is a proxy for x if one has a sample which includes both variables. Let x_i, z_i, and $\varepsilon_i = z_i - x_i$ be from a sample of size n and define x, z and ε as vectors of the observations x_i, z_i and the difference of z and x. Then a simple test to see whether a z is a proxy for x is to examine the inner product $\frac{x'(z - x)}{n} = \frac{x'\varepsilon}{n}$ and see whether it is close to zero.

For two endogenous variables, one may define a proxy in a more general way for which the value of the inner product above being zero does not indicate that one variable is a proxy for another. This definition is motivated as follows. Consider the following two equations

$$t - x\beta_t = u_t \tag{4.2.2}$$

and

$$v - x\beta_v = u_v, \tag{4.2.3}$$

where x is a 1 by k vector, and β_t and β_v are k by 1. Call t a proxy for v if $\beta_t = \beta_v$. This is the definition used in this paper to test whether one variable may be considered an endogenous proxy for another. The selection of this criteria can be defended by noting that this is what reseachers actually assume when they replace one endogenous variable by another, i.e., they are assuming that the inferences they draw from the effect of the exogenous variables upon the endogenous proxy is the same as the effect of the exogenous variables upon the endogenous variable itself. It will be shown that the method demonstrated above of the testing whether one variable may be used as a proxy for another is not sufficient for testing whether one variable is an endogenous proxy for another.

Define ε by

$$\varepsilon = v - t = x(\beta_v - \beta_t) + u_v - u_t.$$

Now,

$$\frac{v'\varepsilon}{n} = \frac{\left[X\beta_v + u_v\right]'\left[X(\beta_v - \beta_t) + u_v - u_t\right]}{n},$$

$$= \beta'\frac{X'X}{n}(\beta_v - \beta_t) + \beta'\frac{X'u_v}{n} - \beta'\frac{X'u_t}{n}$$

$$+ \frac{u'_vX}{n}(\beta_v - \beta_t) + \frac{u'_vu_v}{n} - \frac{u'_vu_t}{n}$$

Under the assumption that $plim\ \frac{X'u_v}{n} = 0$, $plim\ \frac{X'u_t}{n} = 0$, $plim\ \frac{u'_vu_t}{n} = plim\ \frac{u'_vu_v}{n}$, and $plim\ \frac{X'X}{n} = Q_{xx}$, one has

$$plim \frac{v'\varepsilon}{n} = \beta'_v Q_{xx}(\beta_v - \beta_t).$$

If this expression is equated to zero, then, assuming β_v fixed, one has a solution set of the form

$$\left\{\beta_t \,|\, \beta'_v Q_{xx}(\beta_v - \beta_t) = 0\right\}.$$

This set includes the point $\beta_t = \beta_v$ and has dimensionality k-1.

In general, then, $plim \frac{v'\varepsilon}{n} = 0$ will be necessary but not sufficient to test whether a variable may be used as a endogenous proxy for another. The independent variables $x_1, \ldots, x_k$, however, will be correlated with the residuals ε_i, since

$$plim \frac{x'_i\varepsilon}{n} = plim \frac{x'_i X(\beta_v - \beta_t)}{n} + plim \frac{x'_i(u_v - u_t)}{n},$$

or

$$plim \frac{x'_i\varepsilon}{n} = plim \frac{x'_i X(\beta_v - \beta_t)}{n}$$

as the second plim on the right hand side may be expected to go to zero. Thus, aother method of discerning whether one variable may be used as a proxy for another is to examine the correlations between the indepedent variables and ε.

There are problems with that method. An independent variable, while it may have an identical coefficient in both equations, will still, unless it is uncorrelated with every other variable in the equation, will still be correlated with the residuals if the null hypothesis is false. Or, the pattern of correlations with the residuals may be such that an independent variable which has different coefficients in the two models might still have a zero correlation with the independent variabless. Also, a variable z which does not belong in either of the two equations will, if it is correlated with any variable (call it x_j) in the equation for which $\beta_{v_j} \neq \beta_{t_j}$, be correlated with the residuals. Thus examining the residuals, while allowing detection of falsity of the null hypothesis, has two main drawbacks: they do not allow determination as to which coefficients are identical between the two equations, and they do not allow determination as to which variables should be in the equation.

The obvious way around this problem is to run the two regressions and compare the coefficients. Then the step of determining whether the null hypothesis should be rejected or accepted can be tested with this step as well, thus, one may as well run this test originally without examining the residuals. Consider the case of a correctly specified model first. Then

$$\widehat{\beta}_v = \beta_v + (X'X)^{-1}X'u_v$$

and

$$\widehat{\beta}_t = \beta_t + (X'X)^{-1}X'u_t,$$

so, under the assumption that $u_t = u_v + \varepsilon$ and $Cov(u_v, \varepsilon) = 0$, one has

$$E(\widehat{\beta}_v - E[\widehat{\beta}_v], \widehat{\beta}_t - E[\widehat{\beta}_t]) = (X'X)^{-1}X'E[u_t u_v]X(X'X)^{-1}$$

$$= \sigma_v^2 (X'X)^{-1},$$

or the same as the covariance of $\hat{\beta}_v$, which anyone familiar with the results of Cox (1961) or Hausman (1978) would not find surprising. Thus the covariance matrix of the two estimators together is

$$Var\begin{bmatrix}\hat{\beta}_v\\ \hat{\beta}_t\end{bmatrix} = \begin{bmatrix}\sigma_v^2(X'X)^{-1} & \sigma_v^2(X'X)^{-1}\\ \sigma_v^2(X'X)^{-1} & \sigma_t^2(X'X)^{-1}\end{bmatrix}$$

Thus the simple application of a Wald-type statistic (see Appendix II for definitions of and derivations of Wald statistics used in this chapter) will provide a statistical test about the probability of null hypothesis of $\beta_v = \beta_t$ being true.

What about the case of misspecification? Supposing one estimates the models

$$v = z\tau_v + \nu_v$$

and

$$t = z\tau_t + \nu_t$$

insteaad of the true models given by (4.2.2) and (4.2.3). Then one has

$$\hat{\tau}_v = (Z'Z)^{-1}Z'\beta_v + (Z'Z)^{-1}Z'u_v$$

and

$$\hat{\tau}_t = (Z'Z)^{-1}Z'\beta_t + (Z'Z)^{-1}Z'u_t,$$

so one may construct a covariance matrix the same as the one above. The problem with this covariance matrix is that this method of estimating it requires an unbiased estimator of σ_v and σ_t, for which, as is well-known, the usual estimators

$$\sigma_v = \frac{1}{n}\sum_{i=1}^{n} (v_i - z_i\hat{\tau}_v)^2$$

and

$$\sigma_t = \frac{1}{n}\sum_{i=1}^{n} (t_i - z_i\hat{\tau}_t)^2$$

will produce (upward biased) estimators. This will result in erroneously larger confidence intervals than the true ones, which means in practical terms that one will not reject the null hypothesis when it is false. For most research in political science, one would rather not reject the null hypothesis than reject it erroneously. For the present line of inquiry, one would prefer a consistent estimator of the covariance of the two estimators (and also the respective covariances matrices of the estimators themselves). The general distribution of two maximum liklihood estimators is derived in Appendix I. Using these results, one may obtain a consistent estimator of the covariance of $\hat{\tau}_v$ and $\hat{\tau}_t$ by using the expression (the details of this derivation are given in the appendix)

$$Cov(\hat{\tau}_v - E[\hat{\tau}_v], \hat{\tau}_t - E[\hat{\tau}_v]) = (Z'Z)^{-1}Z'\hat{\Lambda}_{vt}Z(Z'Z)^{-1},$$

where $\hat{\Lambda}_{vt}$ is a diagonal matrix with elements along the diagonal

$$\hat{\lambda}_i = (v_i - z_i\hat{\tau}_v)(t_i - z_i\hat{\tau}_t).$$

Similarly, the covariance matrices of the two estimators $\hat{\tau}_v$ and $\hat{\tau}_t$ are

$$Cov(\hat{\tau}_v, \hat{\tau}_v) = (Z'Z)^{-1}Z'\hat{\Lambda}_{vv}Z(Z'Z)^{-1},$$

where $\hat{\Lambda}_{vv}$ is a diagonal matrix with elements along the diagonal

$$\hat{\lambda}_i = (v_i - z_i\hat{\tau}_v)^2$$

and

$$Cov(\hat{\tau}_t, \hat{\tau}_t) = (Z'Z)^{-1} Z' \hat{\Lambda}_{tt} Z (Z'Z)^{-1},$$

where $\hat{\Lambda}_{tt}$ is a diagonal matrix with elements along the diagonal

$$\hat{\lambda}_i = (t_i - z_i \hat{\tau}_t)^2.$$

Note that these matrices do not depend upon the unobserved parameters σ_v^2 or σ_t^2.

Thus for the case of continuous variables, it can be seen that it is fairly simple to derive the joint covariance matrix of the estimators for the two statistical regimes. A Wald-type test may then be applied to derive a statistic to test the hypothesis $\beta_v = \beta_t$. As is shown in Appendix I, the same method may be used for non-continuous variables, such as vote, and the same general type of Wald statistic may be used to test $\beta_v = \beta_t$.

The advantages of this method are plain. First, it allows the development of one statistic to test whether one variable may be used as an endogenous proxy for another. Second, if the null hypothesis of equal coefficients in the two linear models is rejected, it allows the determination of which coefficients lead to the rejection of this model. The alternative method is to examine the residuals and construct tests based upon the residuals. As pointed out before, however, that method does not allow a determination of which coefficients are different between the two regimes, whereas confidence intervals for the differences of the two coefficients may be easily derived from the covariance structure, thus allowing a determination over which coefficients have changed if the null hypothesis is rejected. For these reasons, this is the approach used in this chapter to examine whether certain functional forms of thermometers and the open-ended questions may be used as endogenous proxies for the vote.

4.3 A Utility-maximizing Model of the Voting Decision

As noted before, discerning what individual researchers have intended when they model the voting decision is a difficult task. For the purposes of this exploration, then, a model is developed which previous modelling (which has used the general linear model or some variant of it) may be interpreted. The introduction of such an underlying model should allow a common framework of reference within which the results of this chapter may be interpreted.

What follows, then, is a fairly standard method of creating a model of utility maximizing behavior when a choice between two objects must be made (see Fischer and Nagin (1981) for a description of this type of model, some extensions, and more extensive references). First, if there are two candidates in the race, and one is Republican and one is Democratic, the utility (as a linear function of exogenous variables) for a Democratic vote may be written as

$$v_d = x\beta_d + u_d, \tag{4.3.1}$$

whereas the desire for a Republican vote may be written as

$$v_r = x\beta_r + u_r. \tag{4.3.2}$$

Taken together, then, one obtains

$$v_d - v_r = x(\beta_d - \beta_r) + u_d - u_r. \quad (4.3.3)$$

Under the assumption that u_d and u_r are distributed normally and serially independently (though not necessarily independently across equations), the term $u_d - u_r$ is distributed normally with variance $Var(u_d) + Var(u_r) - Cov(u_d, u_r)$.

The variable $v_d - v_r$ is the utility difference between a vote for a Democratic candidate and a vote for a Republican candidate. If a Democratic vote is coded as a one and a Republican vote as a two, say, then a probit procedure may be utilized to estimate (4.3.3). Note that knowledge of the voting decision does not allow estimation of (4.3.1) or (4.3.2). To do that, one needs knowledge of the utility one holds for the Democratic candidate and the Republican candidate. This is what thermometers are purported to do. Several researchers have used thermometers just as that--indicators of the utility an individual has for a particular candidate. This formulation has been made explicit by Kiewiet (1980), and Black (1978).[5]

There is an conceptual problem here which it is important to delineate. When the vote is observed, it is assumed that the individual has made his utility-maximizing choice. This is what an economist would

5. Both of these authors use the candidates thermometers to rank candidates in a three-way race. Obviously, if one must choose one candidate, one may not obtain rankings over the candidates by observing the choice of one candidate. The central thesis of both the Kiewiet and Black works, incidentally, is that individuals sometimes make "strategic" choices in their voting behavior, a phenomenon which is discussed empirically in Cain (1978) and theoretically in Farquharson (1969) (under the name sincere voting). In the context of this discussion, a choice would be strategic if it is not the choice which would provide the greatest utility. As the contests in this paper are restricted to two choices, the problem of strategic voting does not arise.

call revealed preference, that is, an individual, when he chooses between two objects, chooses the one which provides him with the highest utility. If one accepts the principle of revealed preference, then the best indication of utility is the choice itself. Other measures (such as thermometers), being rankings of utility rather than the choice behavior itself, must necessarily be more prone to error than the choice itself.[6]

Following the general notion of revealed preference, it will be assumed in this chapter that the "best" measure of the utility difference between two candidates is, indeed, the actual voting decision. Under the null hypothesis of thermometers indeed being a proxy for the utility difference between two candidates, it will be assumed that the difference in thermometer scores is the poorer measure of that utility difference. All tests will be conducted under these assumptions, though it will be shown that the statistical tests themselves are somewhat robust to the above assumptions.

4.4 Comparisions of Presther and Zcand with the Vote (1980)--a More Sophisticated Look

In this section an examination is made as to the relationship between feeling thermometers, zcand and the voting decision in the 1980 election, using the theoretical framework which has been developed in the

6. It is important to distinguish between two types of errors, the error from using an proxy and the error term of the linear model. The error term of the linear model is presumed to occur simply because one is unable to track down all influences on the decision. If one has knowledge of more influences, one has less error. The error from the proxy occurs simply because it is an proxy--it is impossible to measure a phenomenon better than that actual phenomenon itself and in all likelihood one has measured it worse.

preceding part of this chapter. The method of analysis is as follows: each of these three variables is used as the dependent variable in the fundamental model,

$$v = x\beta + u$$

Using the covariance matrix derived in the appendix, the coefficients of the independent variables from the regressions of zcand and the difference in feeling thermometers are then compared with the coefficients of the independent variables from the probit with vote dependent.

The sample used needs to be examined. As mentioned before, an analysis file of the 1980 ANES survey created by Mo Fiorina was used. Sixteen independent variables were chosen, with a thought towards capturing differing dimensions of the voters choice. First, party identification was coded into six zero-one dummy variables, using the familiar ANES seven-point scale. The six were strong democratic, weak democratic, independent democratic, independent republican, weak republican, and strong republican, with independents being the reference category. Attitudes towards inflation and unemployment were measured by the question dealing with which party was better able to handle the unemployment problem and a similar question dealing with inflation ("Do you think that inflation would be handled better ..."). There were four possible responses to this question, the republicans, the democrats, don't know, or no difference,. Of these, the first three were made into zero-one dummy variables leaving no difference as the reference category for both inflation and unemployment. For a foreign policy issue, the question on Iran was used, with the two categories made into zero-one for the Iranian crisis being handled well or poorly and the reference category being don't know. Finally, a dummy for an individual being a member of a union was used, as well as a dummy for whether an individual was male.

The three dependent variables were vote, zcand, and presther. Vote was taken to be zero if a vote for Reagan and one if a vote for Carter. Voters for Anderson were excluded from the analysis, as were non-voters. Individuals with missing data on any of the above independent variables were also excluded.[7]

The estimation procedure for this problem was done as follows. Both thermometers and zcand were treated as continuous ordinal variables and estimated by OLS. There are difficulties with this procedure. First, zcand takes on only 20 different values, so in reality a probit procedure should be used. On the other hand, that large a number of categories will tax any probit program.[8] In any case, it was considered that using an OLS procedure upon zcand when multichotomous probit was called for was, if

7. Actually, if data for them were missing on a set of sixty-three variables, they were excluded. A number of specifications were tried using the above and other exogenous variables, until finally the above were utilized. No Wald tests were run except on the variables chosen above—all selection choices were made from the probits with vote dependent. The logic of this procedure is discussed in more detail later.

8. In actual fact, an attempt was made to use a probit procedure on the problem, with no success. Two multichotomous probit programs were available, the (by now) standard McKelvey one, and one written by the author. The main conceptual difference between the two programs is that the McKelvey program uses a constrained maximization proceudre, whereas the author's program uses a the Newton-Ralphson algorithim. The reason a Newton-Ralphson algorithim should converge is that Pratt (1981) offers a proof that the likelihood function is concave no matter what the number of categories. When both programs were run, the McKelvey program "blew up" on the first iteration, whereas the author's program did not converge but rather took off in a direction which made it seem likely that it would not converge (though none of the. constraints on the cutpoints were violated for the 90 iterations it was allowed to run).

anything, a conservative procedure, as one is in reality making an assumption of equal category size. This is also the assumption made in the literature when using zcand on the right-hand side, so this would be another reason for treating it as a continuous variable. For presther, the same problem presents itself. First, anyone who has perused a frequencies of a thermometer knows that most of the responses pile up at five point intervals (0,5,10,15, etc.). Second, they are limited by zero to one hundred, so that presther is limited by minus one hundred and plus one hundred. Thus a two-limit probit is called for.[9]

The results can be seen in Table 4.1. As is easily observed, the Wald test for the similarity of the coefficients is well below significance in both cases, indicating that the null hypothesis of both the difference of thermometers and zcand as proxies for the vote cannot be rejected. Two important points should be discussed immediately: one, how does one interpret the results of such a test, and two, what occurs if a proxy variable is used on the right-hand side of the equation?

9. Two-limit probit (see Rosset and Nelson (1975) for a description of this type of procedure) was run on the difference of thermometers, with negligible changes in the size of the coefficients. Since there are very few observations in either of the limit classes (roughly three percent), there is little lost in using an OLS procedure.

Table 4.1 Regression and Probit Equations for Vote, Presther, Zcand

Independent variable	Presidential vote	Presther	Zcand
	Dependent variable		
Strong Democrat	1.387***	23.63***	1.695***
Weak Democrat	.809***	17.54***	1.063**
Independent Democrat	.712**	10.67**	.89
Independent Republican	.000	-1.15	.372
Weak Republican	-.630	-3.80	.350
Strong Republican	-.809	-13.19**	-.747
Inflation			
Democrats better	.727**	9.11**	.647
Republicans better	-1.024***	-24.18***	-2.47***
Don't know which is better	-.016	-0.45	-.561
Unemployment			
Democrats better	0.450*	10.65***	1.16***
Republicans better	-.521*	-15.67***	-.604*
Don't know which is better	-.685	-10.02	-.399
Iran			
Carter performance good	.378	19.50***	.972
Carter performance bad	-.691*	-6.21	-1.17*
Union member	.414**	-0.37	.124
Male	-.217	-3.21	-.285
$\hat{R}^2$	.747		
-2*LLRatio	421.7		
F		59.01	34.33
$\bar{R}^2$		.636	.503
$\hat{\sigma}$		27.5	2.71
Wald (df)		11.13 (15)	10.75 (15)
n	588	588	588

*$p < .10$. **$p < .05$. ***$p < .01$.

The interpretation of a test of this sort is always difficult, for the simple reason that there is no well-defined alternative hypothesis. This type of test is sometimes labelled non-constructive, that is, the rejection of the test does not imply that the null hypothesis is false.[10] Rather, this is a test of the consistency of the assumption that a certain variable may be used as an endogenous proxy for another variable. This assumption implies certain things, one of which (and one which is testable) is that the coefficients of both regressions when estimated by the correct estimating procedure, should be similar. The statistical test that has been applied states that the hypothesis they are similar cannot be rejected on the data alone. In a practical sense, the non-rejection of the null hypothesis means that a researcher can use either of the three variables and derive much the same conclusion. Indeed, looking at Table 7.1, this is seen to be true.[11]

The question of what occurs when the proxy variable is used on the right-hand side is a question that is easily handled in the same manner that the problem of omitted variables usually is. Let the true model be

$$v = x\beta_v + u_v$$

and instead,

10. The Hausman test (1978) is another example of this type of test. It is important to remember that the null hypothesis one would like to test is that the difference of thermometers can be used as a proxy for the vote. A rejection of the null hypothesis of the test simply states that it is unlikely that the ratios of the coefficients are all equal to the same parameter, which could come about because the model was misspecified in some way (i.e., omitted variables, non-linearities, etc.)

11. For the suspicious at heart (as I am), if one replaces the coefficient on Sdem with the negative of the coefficient, one obtains a chi-square statistic which indicates the null hypothesis should be rejected at the .95 confidence level. So one reversal in sign on a highly significant variable would be enough to cause rejection of the null hypothesis.

$$v = z\tau + u_\tau$$

is estimated. Then, if OLS is used,

$$\tau = (Z'Z)^{-1}Z'v$$

$$= (Z'Z)^{-1}Z'X\beta + (Z'Z)^{-1}Z'u_v.$$

Since the proxy variable is included in the z variables, that variable will be correlated with the residual term u_v, so that the expectation of the term $(Z'Z)^{-1}Z'u_v$ will be non-zero. Thus, even if the other variables in the x and z are identical, so that the only additional variable in z is the proxy, all of the coefficients will be biased.

4.5 Comparisions of Utility for Reagan and Carter

It was noted earlier in this chapter that the individual utility for each candidate, when expressed as a linear function of exogenous variables, was not estimable when one has only the voting decision with which to estimate it. Rather, in that case, one was estimating the difference of the utility functions, as in equation (4.3.3). Perforce, then, (4.3.1) and (4.3.2) were non-estimable.

With thermometers and zcands established as valid proxies for the voting decision, however, (4.3.1) and (4.3.2) may indeed be estimated with the additional assumption that the difference of the utilities displayed in (4.3.3) was derived from (4.3.1) and (4.3.2). As the validity of using thermometers for estimating (4.3.3) was established above, it seems natural to make the additional assumption that the desire for a vote for

Reagan may be expressed as the score on the feeling thermometer for Reagan and that the desire for a vote for Carter may be expressed as the score on the feeling thermometer for Carter. This is, in fact, the procedure that is followed.

The basic story for 1980 is told in table 4.2. In this table, the coefficients for the exogenous variables from the three regressions corresponding to equations (4.3.3), (4.3.1) and (4.3.2) are given, respectively, as well as the standard error and the ratio of the coefficient over the standard error. The results are for the individual utilities are somewhat surprising.

Table 4.2 Regressions for thermometers--1980

Independent variables	Prestherm	Carter	Reagan
	Dependent variable		
Strong Democrat	23.63***	15.82***	-7.80**
Weak Democrat	17.54***	12.14***	-5.39
Independent Democrat	10.67**	8.63**	-2.03
Independent Republican	-1.15	2.26	3.42
Weak Republican	-3.80	3.89	7.69**
Strong Republican	-13.19**	-0.82	12.37***
Inflation			
Democrats better	9.11**	5.82*	-3.28
Republicans better	-24.18***	-16.96***	7.21***
Don't know which is better	-0.45	-0.62	-0.16
Unemployment			
Democrats better	10.65***	0.528	-10.12***
Republicans better	-15.67***	-8.854***	6.82***
Don't know which is better	-10.02	-2.88	7.14
Iran			
Carter performance good	19.50***	9.993**	-9.50**
Carter performance bad	-6.21	-7.24	-1.02
Union member	-0.37	-0.940	-0.56
Male	-3.21	-1.200	2.00
F	59.01	34.67	23.46
R^2	.636	.491	.409
σ	27.54	20.49	19.44
n	588	588	588

*$p < .10$. **$p < .05$. ***$p < .01$.

One way to analyze the problem is to note that the coefficients for the individual thermometers are either fairly symmetrical or asymmetrical (note that the coefficient in the Carter column minus the coefficient in the Reagan column is equal to the coefficient in the difference column). The asymmetric variables are all of the party identification ones, the dummy for those who thought inflation was handled better by the republicans, the dummies for those who thought unemployment was better handled by the democrats or those who didn't know which party handled it better, and the dummy for those who thought Carter had handled Iran badly. The symmetric ones, on the other hand, are the dummies for those who thought inflation was handled better by the democrats, unemployment handled better by the republicans, and the dummy for male. Note that weak republican (and to some extent, independent republican) is the only variable which "cancels" out, that is, it is fairly important for both the individual utilities for Carter and Reagan but insignificant for the prediction of the difference. A party identification of of anything but strong republican increased the utility for Carter (and strong republican is insignificant), indicating that independents were particularly hard on Carter (which is not discernable from the regression of differences).

Carter was burned badly on the inflation issue, as seen by the asymmetry on the by those who thought that the republicans handled inflation better, and Reagan was also hurt badly by those who thought that unemployment was handled better by the democrats, retaining a nice symmetry between the parties which seems to bear out the contention of party asymmetry theorized by Fiorina (1981).[12] The Iranian issue broke perfectly symmetrically between those who thought a good job had been done handling the issue, with a marked asymmetry for those who felt that Carter had done a poor job. The male-female bias much remarked upon in the media presents itself as having a small and expected effect-- even after controlling for the Iranian issue.

A methodological question which may be worthwhile to explore is that of the effect of misspecification on the relationship between the estimates of the regression weights from the regressions of the individual thermometers and the estimates of the regression weights from the regression of the difference in thermometers. It is clear from the discussion dealing with equations (4.3.1), (4.3.2) and (4.3.3) that if the model is correctly specified, then $\beta_d - \beta_r = \beta$. Suppose the model is misspecified. What can be said about the relationship between the estimators for the misspecified equations dealing with the individual candidates, and the estimator for the misspecified equation dealing with the difference of the thermometers?

Not surprisingly, they are the same. Consider the following derivation. Let the true equations be as in (4.3.1), (4.3.2) and (4.3.3).

12. As Fiorina puts it (p 107), "Inflation is up under a Republican administration. Would it have been even higher under a Democratic one? National unemployment under a Democratic administration is running at 6 percent. Would it have been 8 pecent under a Republican administration?" The fact that for those who think inflation is better handled by the Repbulicans punish Carter badly while those who think that inflation is better handled by the Democrats do not reward him equally (and there is a similar pattern for unemployment for Reagan, thogh not as much) would seem to indirectly support this type of reasoning.

Then for the misspecified equations, one is estimating

$$v_d = z\tau_d + v_d$$

$$v_r = z\tau_r + v_r$$

and

$$v = z\tau + v_d$$

The least squares estimates (or maximum likelihood estimates, under the assumption of normality) are as follows

$$\hat{\tau}_d = (Z'Z)^{-1}Z'(X\beta_d + v_d)$$

$$\hat{\tau}_r = (Z'Z)^{-1}Z'(X\beta_r + v_r)$$

and

$$\hat{\tau} = (Z'Z)^{-1}Z'X(v_d - v_r)$$

$$= (Z'Z)^{-1}Z'X(x\beta_d - x\beta_r + u_d - u_r).$$

Thus the differences of the individual coefficients sum to the coefficient derived from estimating the difference of v_d and v_r, whether or not the model is correctly estimated or not, if, as is assumed in this paper, the function is taken to be linear in these arguments.

4.6 An Analysis of the Utility Function of the Non-voter

As indicated before in the discussion leading to (4.3.1) through (4.3.3), if one can find a valid proxy for the voting decision, and if this proxy is also available for non-voters, then it is possible to examine the decision "calculus" for the non-voter as well as for the voter. This then has substantive interest as well as methodological interest, for it may shed some light upon some issues in voting behavior.

The classical formulation of rational voting behavior is that of Downs (1957), with his model

$$u = pb - c,$$

where the utility from an individual's voting is a function of p (the probability of his vote affecting the race) times b (the benefit of the individual from a victory in that race), minus c, the cost of voting in that race (an extension of this model to a multi-candidate race is given in McKelvey and Ordeshook (1972)). This model has been through some reformulations,[13] but one (heretofore) untested assumption of this model as empirically estimated is that individuals have basically the same utility function, whether they are voters or not. Naturally, in the pristine formulation, for every voter i, one would have a utility function

$$u_i = p_i b_i - c_i,$$

but empirical researchers of course lack the information to model the decision in that manner. Rather, empirical reasearchers who are modelling rational choice (see Repass (1971), Popkin et. al. (1976), Fiorina (1978,1981)) assume a population utility function b which is a function of a number of exogenous variables $x_1, \ldots, x_k$, usually linear.

For an argument that there is in fact a rational choice basis for

13. The one that seems to have stuck is the one by Riker and Ordeshook, who added a "D" term to the model for duty. Citizen duty is well-known to have the largest zero-order effect upon turnout (see Campbell, Converse, Miller, and Stokes (1960))

voting (and in particular, turnout), this utility function should be the same for both voters and non-voters. Thus an essential piece of groundwork for the rational choice model to be established is that the model apply equally to both voters and non-voters. If it can be demonstrated that thermometer differences in presidential races can indeed be taken in this manner, this provides another item of evidence for the rational choice modelling of voting in American presidential elections.

The task, then, is to see whether the decision calculus of the voter may be taken to be the same as that of the non-voter. For this analysis, the following course of action was taking. In the initial analysis strategy, a large set of variables which could be assumed to be more or less exogenous[14] was utilized. Naturally, with that large of number of cases, one would expect there to be a great many missing values, and as cases with missing values were excluded from the analysis, the result was that there were 558 cases in the analysis. This was the set that was used to run the equivalency tests between the differences of thermometers and the voting decision. As mentioned before, all preliminary model selection was done regressing the vote upon different subsets of these variables, and then a test was made of the use of the difference of thermometers as a proxy for the vote using that subset. This was done in an effort to avoid bias.[15]

14. Naturally, many of these variables have been used as endogenous variables at one time or another, particularly the party identification variables (see Jackson (1974), Page and Jones (1977), and Fiorina (1981)).
15. Such as using one set of variables, then observing that the null hypothesis was rejected, then using another set of variables, and so forth, until finally success was achieved.

On the other hand, for an examination as to whether non-voters and voters have the same decision rules (i.e., the coefficients in the linear regression are identical), there seems to be no good reason not to use all individuals who do not have values missing on the sixteen independent variables and the two thermometers. When this is done, the number of cases available for voters is 817, and the number for non-voters is 651. This would seem to be a satisfying number to test the hypothesis that voters and non-voters utilize the same decision calculus.[16]

The actual statistic that is used is the Chow test. Let y_1 be the vector of differences in thermometer scores for the voters and y_2 be the vetor of differences in thermometer scores for the non-voters. Let X_1 be the matrix of of exogenous variables for voters and X_2 be the matrix of exogenous variables for non-voters. Let β_1 be the vector of weights for the voters and β_2 be the vector of weights for non-voters. Then write the two systems of equations as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

where u_1 and u_2 are the error terms associated with the regime for voters and non-voters, respectively. Then the Chow test is simply to estimate the above system in two different manners: one, setting

16. Wald statistics for the similarity of both thermometers and zcand for this augmented data set were calculated. Both were non-significant at well under the .05 level, with 17.38 for thermometers and 13.36 for zcand being the values of the χ^2 variables. While thus both statistics are very far from being significant (a value of 25.0 is required for significance at the .05 level for a χ^2 variable with fifteen degrees of freedom), there is a notable "creep" upward from the previous Wald statistics. The question of what this indicates for the validity of the test is discussed in the last part of this chapter, but a strong claim can be made that this type of phenomenon does not invalidate the test.

$\beta_1 = \beta_2 = \beta_*$ (the restricted estimator), and two, allowing β_1 and β_2 to vary freely (the unrestricted estimator). Defining the various residual sum of squares as

$$RRSS = (y_1 - x_1\hat{\beta}_*)'(y_1 - x_1\hat{\beta}_*) + (y_2 - x_2\hat{\beta}_*)'(y_2 - x_2\hat{\beta}_*)$$

$$URSS = (y_1 - x_1\hat{\beta}_1)'(y_1 - x_1\hat{\beta}_1) + (y_2 - x_2\hat{\beta}_2)'(y_2 - x_2\hat{\beta}_2)$$

Then if n_1 is the number of elements (cases) in y_1, n_2 is the number of elements (cases) in y_2 and k is the number of elements in β_1 (which is equal to the number of elements in β_2), then

$$\frac{\dfrac{RRSS - URSS}{k}}{\dfrac{URSS}{n_1 + n_2 - 2k}}$$

is distributed as a F_{k-1, n_1+n_2-2k} random variate (this fact is shown in Maddala, p460) under the null hypothesis that $\beta_1 = \beta_2$. This then is the statistic that is used for testing whether or not the decision calculus for voters is the same as that for non-voters.

The actual results of these Chow tests can be seen in Table 4.3. There, the different residual sum of squares are displayed and the F statistics are displayed. Both the F statistic for the thermometer difference and the F statistic for the Carter thermometer are not significant at the .05 level. The same, however, cannot be said of the Reagan thermometer. Rather, it is significant at the .01 level. How can these results be interpreted?

Table 4.3 Unrestricted and restricted sum of squared residuals

	Dependent variable		
	Presther	Carter	Reagan
URSS (n=1468)	1190218	623891	660699
Voters (n=817)	623755	323351	309414
Non-voters (n=651)	545965	294512	333998
RRSS	1169720	617863	643412
RRSS - URSS	20498	6028	17287
F	1.452	.851	2.207

The first and most important result is that the null hypothesis of the decision calculus based upon the differences of the thermometers cannot be rejected. In practical terms, the researcher may take them to be the same. This can be seen as a partial confirmation of the rational choice modelling of presidential elections. The proxy for the utility difference that it is assumed the vote is cast on the basis of extends in the manner that would be predicted to non-voters. Similarly for the Carter thermometer, it can be seen that the differences between voters and non-voters is minimal, and they may also be taken to be the same.

On the other hand, there is the Reagan thermometer. In it, the two regimes are different. The interpretation of this is perhaps best handled by a perusal of table 4.4. Inspecting this table, it is hard to escape the conclusion that the prediction for Ronald Reagan for those that do not vote is simply very poor. The only two variables which vary dramatically between voters and non-voters are the constant term and the uedembet variable. The R^2 is very poor, though, with a value of .174. The R^2 for the non-voters on the Carter equation is .337, or twice as great, whereas the R^2 for the voters on the Carter thermometer equation is .510, not all that much greater than the .414 for the Reagan thermometer equation restricted to voters.

Table 4.4 Regressions for Reagan Thermometers, Voter and Non-voters

Independent variable	Dependent variable: Both	Voters	Non-voters
Constant	55.10	59.50	54.00
Strong Democrat	-6.03***	-9.70***	-5.38***
Weak Democrat	-2.86	-6.76**	-1.68
Independent Democrat	-3.85*	-4.49	-4.77
Independent Republican	7.70***	2.03	12.07***
Weak Republican	8.21***	5.66*	7.52**
Strong Republican	16.17***	12.10***	14.40***
Inflation			
Democrats better	-6.15***	-4.44*	-7.22**
Republicans better	8.15***	7.608***	7.21**
Don't know which is better	2.70	6.03	3.39
Unemployment			
Democrats better	-6.47***	-10.71***	-2.00
Republicans better	6.03***	5.87***	4.58
Don't know which is better	-0.19	2.17	-1.56
Iran			
Carter performance good	-5.48**	-6.27*	-4.79
Carter performance bad	-0.54	1.36	-2.47
Union member	-1.43	-0.44	-2.76
Male	0.75	0.87	0.15
F	39.69	35.34	8.36
R^2	.304	.414	.174
σ	21.33	19.66	22.95
n	1468	817	651

$^{*}p < .10$. $^{**}p < .05$. $^{***}p < .01$.

4.7 Comparisions of Thermometers with the Vote for the House and Senate (1978)

A logical extension to the analysis presented so far is to attempt to discern whether the difference of thermometers, which have been seen to be a valid proxy for the presidential vote, can be considered a valid proxy for house and senatorial contests. The data set which was chosen for this was the 1978 ANES data set, in part because it offers a wider variety of information about senatorial and congressional races than most other ANES data sets.[17]

The data is summarized in tables 4.5 (the congressional) and 4.6 (the senatorial). The format of the data is similar to that in table 4.1. There are some obvious differences. First, there are two columns of coefficients for the probit with vote as the dependent variable. In the first column are the coefficients for the four category vote variable, in the second column the coefficients for the two category vote column. The four category vote variable was constructed by using the strong and not strong preference questions in the 1978 survey.[18] If an individual voted for a republican House candidate and his preference was strong, he was coded as being in the lowest category, if his preference was not strong, he was coded as being in the next highest category (category 2). On the other hand, if he voted for a democratic House candidate and his

17. Once again this data set was designed by Mo Fiorina. (Though I personally did much of the computer work for both the 1980 data set and the 1978 data set while I was employed as a graduate research assistant at the California Institute of Technology).
18. The question asked is "Would you say that your preference for this House candidate <Respondent voted for> was strong or not strong?"

preference was strong, he was placed in the highest category, whereas if his preference was not strong, he was placed in a category one lower (category 3).

The other change is in the variable list. While some variables are the same (such as strong democrat, weak democrat, etc.), others have been changed to in order to more accurately describe house and senate races, as opposed to presidential races. The rating of Jimmy Carter was included, as were contacts with the candidates and the incumbency status of the candidates. Also, the inflation and unemployment variables now include the "don't knows" as part of the surpressed reference category.

Table 4.5 Regression and Probit Equations for Congressional Voting

Independent variables	Dependent variable: Cong vote 4 cat	Cong vote 2 cat	Congther
Strong Democrat	1.121***	1.136***	18.57***
Weak Democrat	0.446*	0.73**	16.80***
Independent Democrat	0.431	0.319	14.26**
Independent Republican	-0.258	-0.405	-4.39
Weak Republican	-0.226	-0.459	-2.90
Strong Republican	-0.602**	-1.128***	-12.70*
Inflation			
Democrats better	-0.118	-0.119	1.71
Republicans better	0.001	-0.145	-0.35
Unemployment			
Democrats better	0.043	0.142	3.25
Republicans better	-0.065	0.109	-6.76
Carter performance rating			
Good	-0.038	0.062	-4.84
Bad	-0.561***	-0.527**	-15.25***
Incumbency status of candidate			
Democrat	0.519***	0.588**	16.84***
Republican	-0.712***	-0.754***	-7.97*
Union	0.348**	0.526***	3.947
Candidate contacts			
Democratic	0.267	0.240	6.11
Republican	-0.305*	-0.520**	-9.37**
$\hat{R}^2$	.486	.603	
-2*LLRatio	203.9	195.4	
F			13.96
R^2			.406
Wald for two categories			15.2
Wald for four categories			24.7
df			16

n	364	364	364

*$p < .10$. **$p < .05$. ***$p < .01$.

Table 4.6 Regression and Probit Equations for Senatorial Voting

Independent variables	Dependent variable Sen vote 4 cat	Sen vote 2 cat	Senther
Strong Democrat	0.882***	1.085***	18.13**
Weak Democrat	0.525**	0.599**	5.42
Independent Democrat	0.561**	0.603**	1.84
Independent Republican	-0.401	-0.872**	-11.94
Weak Republican	-0.276	-0.346	-8.33
Strong Republican	-0.775***	-1.377***	-24.23***
Inflation			
Democrats better	-0.037	0.066	2.99
Republicans better	-0.200	-0.049	-9.04*
Unemployment			
Democrats better	0.313*	0.374*	6.41
Republicans better	-0.273	-0.447	-4.45
Carter performance rating			
Good	0.038	0.112	6.69
Bad	0.135	0.250	-2.06
Incumbency status of candidate			
Democrat	0.181	0.308	3.52
Republican	-0.138	-0.511	14.98
Union	0.157	0.199	1.726
Candidate contacts			
Democratic	0.173	-0.109	19.67*
Republican	0.091	0.3338	1.59
$\hat{R}$	.338	.491	
Rq	.274	.253	
-2*LLRatio	129.9	155.3	
F			7.24
$\bar{R}^2$			.26
Wald for two categories			18.27

Wald for four categories			11.40
df			16
n	368	368	368

$^{*}p < .10$. $^{**}p < .05$. $^{***}p < .01$.

The results of the probits and regression are not surprising. Partisanship plays a strong role in the voting decision in both voting decisions. Opinions as to which party is the better handler of inflation and unemployment are insignificant in both cases, whereas union status directs an individual towards voting democratic. On the congressional side, contacts with the candidates, incumbency status and ratings of Carter's performance all had effects in the expected direction (with the exception that a good rating of his performance was insignificant), but on the senatorial side these variables were all insignificant (and sometimes of the wrong sign). The R^2 for the congressional thermometer was quite a bit higher (.41) than that for the senatorial thermometer (.26), indicating that the fit for congressional races were quite a bit better than that for the senatorial races. Doubtlessly, given the greater salience of senatorial races, important indicators such as issue positions have been omitted.[19]

A perusal of Table 4.5 and Table 4.6 raises another interesting question: are the four-category and two-category probits for both the congressional and senatorial races equivalent? This is not a frivolous question, as a glance at the tables indicates that there is a difference in the coefficients, the only question being, is there enough of a difference for the researcher to conclude that the probits, when run with the four-category or two-category dependent variable, are identical? One way to

19. It should be noted parenthetically that there are a large number of missing cases for these measures, particularly the thermometers. The relevant missing figures were for house and senate incumbents, 22 percent and 15 percent, respectively, for house and senate democratic candidates who were not incumbents, 47 percent and 21 percent, and for house and senate republicans who were not incumbents, 60 percent and 33 percent.

answer that question from a statistical standpoint is to test the hypothesis $H_0 = \{\beta_4 - \beta_2 = 0\}$.

The application of this test obviously requires knowledge of the covariance matrix of the statistic $\beta_4 - \beta_2$. The derivation of this matrix, when both parameters are estimated by the method of maximum likelihood, is given in Chapter II. As it turns out, it is simply the matrix covariance matrix of β_2 minus the covariance matrix of β_4. Denote the difference of these covariance matrices as R. Then a Wald test of the difference of the two estimators is simply $(\beta_4 - \beta_2)'R(\beta_4 - \beta_2)$, which, under the null hypothesis H_0, will be distributed asymptotically χ^2 with k-1 degrees of freedom, where k is the number of elements in the vector $\beta_4 - \beta_2$.

Such test was run for both the congressional and senate races. For the congressional races, the Wald statistic had a value of 23.69 with sixteen degrees of freedom. As the .05 significance level for a χ^2 variable with sixteen degrees of freedom requires a value of 27.5, one may not reject the null hypothesis at the .05 level. Similarly, for the senatorial races, the Wald statistic had a value of 24.553. This statistic, also, does not achieve a significant value and so may not be rejected at the .05 level.

4.8 Conclusions

This inquiry, which focuses upon the relationship of various variables to one another, is yet another indication of the importance of models in political science. Candidate evaluations, when summed in a crude manner, have been seen to give no different substantive results when used as the endogenous variable in an equation than the vote itself. Thermometers too have this property, thus bearing out Fiorina's advice (see footnote 3) on the desirability of not utilizing such variables as exogenous variables. Doubtlessly other variables, or constructions bases upon other variables, have the property that they too may be used as a proxy for the vote. It is certainly not inconsistent with many psychological theories (such as cognitive consistency or dissonance reduction theories) that individuals will have one central cognition or attitude and reexpress it in different manners under a variety of prompts. Should it be so surprising that individuals would do the same with the voting decision?

Whether or not this is what is happening, it is clear that, given the specifications in this paper, the difference of thermometers or the sum of the likes and dislikes, when treated as the endogenous variable, produce beta weights which are similar to those produced by using the voting decision itself as the voting decision. It has been argued that the results are fairly robust to the choice of voting decision, but there is also another argument, one based on the fact that there was no well-defined alternative hypothesis in the tests that were presented in this paper. If, it is claimed, thermometers (or zcand) measure some aspect of the voting decision which is unique from the decision itself, then surely there must be some exogenous predictor of that difference which would not matter in the voting decision but would matter in, say, the thermometer score. In that case, that variable or set of variables could easily be placed in the specification and the similarity of the weight or weights associated with it tested, in the same manner that they were tested here. This would provide, incidentally, the well-defined alternative hypothesis which was lacking before.

It would seem that the results presented in this paper can be seen as giving a vote of confidence to previous research which has used zcand or thermometers as endogenous variables, or as predictors of the vote. In particular, the fact that the actual metric of the thermometers proxies the voting decision would seem to validate the use of, say Black, who uses the differences of the thermometers to indicate utility differential. The use of thermometers as proxies for the vote in a simultaneous equation system by Page and Jones would likewise seem reasonable. The factor analysis of the thirty-four different categories of open-ended questions by Miller and Miller acquires new validity, for if the sum of the open-ended questions can be used as a proxy for the vote, then (leaving aside the desirability of factor analysis as a method of statistical analysis) all the factors should be included in the thirty-four open-ended questions. Miller and Miller, however, claimed that it measured the major personality dimensions that voters had used in evaluating the candidates (page 634). It would seem that it would measure more. Similarly, they interpret thermometer scores as affect (see their footnote 33 in "Majority Party"). Their table 3 (p 771), where they tabulate thermometer scores against policy orientation, rather than having an interpretation as affect, have the simple and straightforward interpretation as a utility score. It is not surprising, in that case, that Nixon would be seen "coldly" by those with a liberal persuasion-- he provides them less utility. Of course, they may actually "feel" less "warmly" towards Nixon as well, but it is not clear what exactly that means. The results presented here seem to indicate that there is no difference between feelings and utility, at least as measured by present instruments.

If these results give votes of confidence for the use of thermometers and zcand as proxies, they must also give a vote of no confidence to the use of such variables as exogenous variables. It is shown in the paper the statistically results of using an endogenous variable as an exogenous variable, but what are the substantive results? The primary one, following the comments by Fiorina, would simply be that there can be no valid substantive interpretation to such variables entered on the right-hand side of the regression equation. Consider zcand, the sum of the open-ended questions. Taken as a sum, these variables proxy the vote. Should they not be considered separately as exogenous variables rather than taken as a sum? The individual comments reflect a variety of concerns, and taken individually, they may offer much in understanding about the voting decision. But taken together, they seem to indicate that what individuals say about a candidate is consistent with their voting decision. It would probably be more surprising if they were not.

Adkison, Danny M. 1982. The Electoral Significance of the Vice Presidency. *Presidential Studies Quarterly*, Vol 12, # 3 (Summer 1982): 330-336.

Black, Jerome H. 1978. The Multicandidate Calculus of Voting: Application to Canadian Federal Elections. *American Journal of Political Science*, Vol 22, #3 (August 1978): 609-638.

Cain, Bruce E. 1978. Strategic Voting in Britain. *American Journal of Political Science*, Vol 22, #3 (August 1978): 639- .

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. Chicago: University of Chicago Press.

Cox, D. R. 1961. Tests of Separate Families of Hypothesis. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol 1 Berkeley. University of California Press, 105-123.

Cox, D. R. 1962. Further Results on Tests of Separate Families of Hypothesis. *Royal Statistical Society Journal*. Series B, Vol 24, # 2 : 406-424.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper & Row.

Farquharson, Robin. 1969. *Theory of Voting*. New Haven. Yale University

Press.

Ferejohn, John A. and Morris P. Fiorina. 1974. The Paradox of Not Voting: A Decision Theoretic Analysis. *American Political Science Review*, Vol 68, #2 (June 1974): 525-536.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven: Yale University Press.

Fischer, Gregory W. and Daniel Nagin. 1981. Random versus Fixed Coefficient Quantal Choice Models. In *Structural Analysis of Discrete Data with Econometric Applications*. Manski, Charles F. and Daniel McFadden. Cambridge, Massachusetts: MIT Press.

Hausman, J. A. 1978. Specification Tests in Econometrics. *Econometrica*, Vol 46, # 6, (November, 1978): 1251-1271.

Huber, Peter J. 1967. The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. In *Proceedings of the Fifth Berkeley Symposium*, Vol 1. Berkeley. University of California Press: 221-233.

Hogg, Robert V. and Allen T. Craig. 1970. *Introduction to Mathematical Statistics*, 3rd edition. New York, New York. Macmillan Publishing Co, Inc.

Kiewiet, D. Roderick. 1980. Approval Voting: The Case of the 1968 Election. *Polity*

Maddala, G.S. 1977. *Econometrics*. New York, New York. McGraw-Hill Book Company.

McKelvey, Richard D. and Peter C. Ordeshook. 1972. A general theory of the calculus of voting. In J. F. Herndon and J. L. Bernd., eds, *Mathematical Applications in Political Science*, VI. Charlottesville: University Press of Virginia, pp. 32-78.

Miller, Arthur H., Warren E. Miller, Alden S. Raine and Thad A. Brown. 1976. A Majority Party in Disarray: Policy Polarization in the 1972 Election. *American Political Science Review*, Vol 70, #3, (Sept,1976): 753-778.

Miller, Arthur H. and Warren E. Miller. 1976. Ideology in the 1972 Election: Myth of Reality—A Rejoinder. *American Political Science Review*, Vol 70, #3, (Sept,1976): 832-849.

Morrison, Donald F. 1976. *Multivariate Statistical Methods*. New York, New York. McGraw-Hill Book Company.

Page, Benjamin I and Calvin C. Jones. 1977. Reciprocal Effects of Policy Preferences, Party Loyalities and the Vote. *American Political Science Review*, Vol 73, : 1071-1089.

Popkin, Samuel, John W. Gorman, Charles Phillips, and Jeffrey A. Smith. 1976. Comment: What Have You Done for Me Lately? Toward an Investment Theory of Voting. *American Political Science Review*, Vol 70, #3, (Sept,1976): 779-805.

Repass, David E. 1971. Issue Salience and Party Choice. *American Political Science Review*, Vol 65: 389-400.

Riker, William H. and Peter C. Ordeshook. 1968. A Theory of the Calculus of Voting. *American Poltical Science Review*, Vol 62, # 1, (March 1968): 25-42.

Rosset, Richard N. and Forrest D. Nelson. 1975. Estimation of the Two-limit Probit Regression Model. *Econometrica*, Vol 43, # 1 (January, 1975): 141-146.

White, Halbert. 1981. Consequences and Detection of Misspecified Nonlinear Regression Models. *Journal of the American Statistical Association*, Vol 76, # 374, (June, 1981): 419-433.

White, Halbert. 1982. Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, Vol 50, # 1 (January, 1982): 1-25.

Appendix I: The Derivation of Covariance Matrices of Estimators Derived by the Method of Maximum Likelihood

For the derivation of various covariance matrices displayed in the exposition it is necessary to take a brief journey through maximum likelihood estimation theory. The joint distribution of two estimators derived by the method of maximum likelihood was apparently first considered in a general framework by Cox (1961), who derived in a heuristic manner the asymptotic covariance matrix of two maximum likelihood estimators based upon the same data when one of the estimators was derived under the true probability law and the other estimator was derived under some untrue probability law. Huber (1965) derived in a rigorous manner the asymptotic distribution of a maximum likelihood estimator when it is estimated under some untrue probability law, and White (1982) derived rigorously the asymptotic covariance matrix derived by Cox. In this section the joint asymptotic normality of any two maximum likelihood estimators is shown in a a heuristic manner, a result implicit in all of the above results but never (to the knowledge of this author) explicitly proven. This result then justifies the assumption of asymptotic multivariate normality of any two estimators derived by the method of maximum likelihood.

Let there be two probability laws $f(v \mid x, \beta_v)$ and $g(t \mid z, \beta_t)$, where z is used to emphasize that the two probability laws may be conditional on different sets of variables. Assume

$$plim \ \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \ln f(v_i \mid x_i, \beta_v)}{\partial \beta_{v_k} \partial \beta_{v_l}} = \lim \frac{1}{n}\sum_{i=1}^{n} E\left[\frac{\partial^2 \ln f(v_i \mid x_i, \beta_v)}{\partial \beta_{v_k} \partial \beta_{v_l}}\right] = Q_{\beta_{v_{k,l}}}$$

and

$$plim \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \ln g(t_i | z_i, \beta_t)}{\partial \beta_{t_k} \partial \beta_{t_l}} = \lim \frac{1}{n}\sum_{i=1}^{n} E\left[\frac{\partial^2 \ln g(t_i | z_i, \beta_t)}{\partial \beta_{t_k} \partial \beta_{t_l}}\right] = Q_{\beta_{t_{k,l}}}.$$

for all k, l. The first part of each equality is simple an assertion that the law of large numbers holds, the second part that there is convergence to a well-defined limit. Let Q_{β_v} be the matrix made up of the $Q_{\beta_{v_{k,l}}}$, while Q_{β_t} is the matrix made up of the $Q_{\beta_{v_{k,l}}}$. If there are k_v elements in β_v, Q_{β_v} will be k_v by k_v, and there being k_t elements in β_t, implies that Q_{β_t} will be k_t by k_t. It is assumed that these two matrices are invertible.

Let $\left[\frac{\partial \ln f(v_i | x_i, \beta_v)}{\partial \beta_v}\right]$ be the k_t by 1 matrix made up of $\frac{\partial \ln f(v_i | x_i, \beta_v)}{\partial \beta_{v_l}}$, and similarly, let $\left[\frac{\partial \ln g(t_i | x_i, \beta_t)}{\partial \beta_t}\right]$ be the k_v by 1 matrix made up of $\frac{\partial \ln g(t_i | z_i, \beta_v)}{\partial \beta_{t_l}}$. Then, following Cox (equation 31), one has that asymptotically

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left[\frac{\partial \ln f(v_i | x_i, \beta_v)}{\partial \beta_v}\right] + \sqrt{n}\, Q_{\beta_v}[\hat{\beta}_v - \beta_v] = 0$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left[\frac{\partial \ln g(t_i | z_i, \beta_v)}{\partial \beta_v}\right] + \sqrt{n}\, Q_{\beta_t}[\hat{\beta}_t - \beta_t] = 0.$$

Then, asymptotically,

$$\sqrt{n}\, Q_{\beta_v}[\hat{\beta}_v - \beta_v] = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left[\frac{\partial \ln f(v_i | x_i, \beta_v)}{\partial \beta_v}\right]$$

and

$$\sqrt{n}\,Q_{\beta_t}[\hat{\beta}_t - \beta_t] = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\partial \ln f\,(t_i\,|\,z_i,\beta_v)}{\partial \beta_v}\right].$$

Let c_v and c_t be any two constants such that c_v is k_v by 1 and c_t is k_t by 1. It is necessary and sufficient to show that

$$\sqrt{n}\,c'_v[\hat{\beta}_v - \beta_v] + \sqrt{n}\,c'_t[\hat{\beta}_t - \beta_t]$$

is asymptotically normal for any c_v and c_t, for the asymptotically normal estimators to have a joint normal distribution. Let a_v be such that $c'_v = a'_v Q_{\beta_v}$ and a_t such that $c'_t = a'_t Q_{\beta_t}$. The condition that Q_{β_t} and Q_{β_v} be invertible ensures that such a_v and a_t may be found. Then,

$$\sqrt{n}\,c'_v[\hat{\beta}_v - \beta_v] + \sqrt{n}\,c'_t[\hat{\beta}_t - \beta_t]$$

$$= \sqrt{n}\,a'_v Q_{\beta_v}[\hat{\beta}_v - \beta_v] + \sqrt{n}\,a'_t Q_{\beta_t}[\hat{\beta}_t - \beta_t]$$

$$= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{a'_v\left[\frac{\partial \ln f\,(v_i\,|\,x_i,\beta_v)}{\partial \beta_v}\right] + a'_t\left[\frac{\partial \ln f\,(t_i\,|\,x_i,\beta_t)}{\partial \beta_t}\right]\right\}.$$

This is, however, the sum of n independent random variables divided by the square root of n. Thus by a variant of the central limit theorem, this approaches a normal variate in probability. As this is true for any c_v and c_t, the asymptotic distribution of the two maximum likelihood estimators is jointly normal.

To derive the covariance matrix of two estimators derived by the method of maximum likelihood, then, the above results allow the following asymptotic equality

$$\lim_{n\to\infty} E\left[\sqrt{n}\,Q_{\beta_v}[\hat{\beta}_v - \beta_v][\hat{\beta}_t - \beta_t]'Q_{\beta_t}\sqrt{n}\right]$$

$$= \lim_{n\to\infty} E\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial \ln f\,(v_i\,|\,x_i,\beta_v)}{\partial\beta_v}\right]\sum_{i=1}^{n}\left[\frac{\partial \ln g\,(t_i\,|\,z_i,\beta_t)}{\partial\beta_t}\right]'\right\} \qquad (4.A.1)$$

If v_i and t_i are independent of one another for all i, this may be rewritten as

$$= \lim_{n\to\infty} E\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\partial \ln f\,(v_i\,|\,x_i,\beta_v)}{\partial\beta_v}\right]\right\}E\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\partial \ln g\,(t_i\,|\,z_i,\beta_t)}{\partial\beta_t}\right]'\right\}.$$

As $n\to\infty$ both these terms approach normal variates with means zero and covariance matrices Q_{β_v} and Q_{β_t}, respectively. Thus the asymptotic expectation of the covariance is zero for maximum likelihood estimators derived from independent random variables, as one would have expected.

Usually, though, t_i and v_i will be dependent upon one another for the same observation, so that one may not simply take expectations in the manner above. Rather, since it is assumed that the observations are independent of one another, write

$$\lim_{n\to\infty} E\left[\sqrt{n}\,Q_{\beta_v}[\hat{\beta}_v - \beta_v][\hat{\beta}_t - \beta_t]'Q_{\beta_t}\sqrt{n}\right]$$

$$= \lim_{n\to\infty} E\left\{\frac{1}{n}\sum_{i=j}\left[\frac{\partial \ln f\,(v_i\,|\,x_i,\beta_v}{\partial\beta_v}\right]\left[\frac{\partial \ln g\,(t_i\,|\,z_i,\beta_t}{\partial\beta_t}\right]'\right\}$$

$$+ \lim_{n\to\infty} E\left\{\frac{1}{n}\sum_{i\neq j}\left[\frac{\partial \ln f\,(v_i\,|\,x_i,\beta_v}{\partial\beta_v}\right]\left[\frac{\partial \ln g\,(t_i\,|\,z_i,\beta_t}{\partial\beta_t}\right]'\right\}$$

Note that the limit of the second sum is zero if the observations are independent of one another and the the expectation of the first partial with respect to the parameters is zero. This will be the case if the model is correctly specified, or if the endogenous and exogenous variables are jointly independently identically distributed (instead of the exogenous variables being fixed, as in, say, an experimental situation).

So let

$$Q_{vt} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=j} E\left[\frac{\partial \ln f(v_i \mid x_i, \beta_v}{\partial \beta_v}\right]\left[\frac{\partial \ln g(t_j \mid z_j, \beta_t}{\partial \beta_t}\right]',$$

where Q_{vt} represents the (assumed to be defined) limit of the above sum. Usually the form of the expectation may be inspected to to determine what assumptions are necessary on the x's and z's to ensure convergence. Hence

$$plim\ \sqrt{n}\,[\hat{\beta}_v - \beta_v]\sqrt{n}\,[\hat{\beta}_t - \beta_t]' = Q_{\beta_v}^{-1} Q_{vt} Q_{\beta_t}^{-1}$$

For the general linear model estimated by maximum likelihood with normal error terms, then

$$\frac{\partial^2 \ln f(v_i \mid x_i, \beta_v)}{\partial \beta_{v_l} \partial \beta_{v_k}} = -\frac{1}{\sigma_v^2} x_{l,i} x_{k,i}$$

and

$$\frac{\partial \ln f(v_i \mid x_i, \beta_v)}{\partial \beta_{v_l}} \frac{\partial \ln f(v_i \mid x_i, \beta_v)}{\partial \beta_{v_k}} = -\frac{1}{\sigma_v^4}(v_i - x_i \beta_v) x_{l,i} x_{k,i}$$

This expression of the second partials is non-random whereas the second expression is random and has expectation, if $v_i - x_i\beta_v$ is distributed normally with mean zero and variance σ_v^2,

$$E\left[\frac{\partial \ln f(v_i \mid x_i, \beta_v)}{\partial \beta_{v_l}} \frac{\partial \ln f(v_i \mid x_i, \beta_v)}{\partial \beta_{v_k}}\right] = \frac{1}{\sigma_v^2} x_{l,i} x_{k,i}.$$

or the minus the value of the second partials.

Then

$$\sum_{i=1}^{n} \frac{\partial \ln f(v_i | x_i, \beta_v)}{\partial \beta_{v_l}} \frac{\partial \ln f(v_i | x_i, \beta_v)}{\partial \beta_{v_k}} = \frac{1}{\sigma_v^4} \sum_{i=1}^{n} (v_i - x_i \beta_v)^2 x_{k,i} x_{l,i}$$

Or, using the matrix notation for the partials again, then one has

$$\sum_{i=1}^{n} \left[\frac{\partial \ln f(v_i | x_i, \beta_v)}{\partial \beta_{v_l}} \right] \left[\frac{\partial \ln f(v_i | x_i, \beta_v)}{\partial \beta_{v_k}} \right]' = \frac{1}{\sigma_v^4} \sum_{i=1}^{n} X' \Lambda_{vv} X,$$

where Λ_{vv} is diagonal with elements $\lambda_i = (v_i - x_i \beta_v)^2$. Then

$$\frac{1}{n\sigma_v^4} X' \Lambda_{vv} X \to \frac{1}{n\sigma_v^4} X' E[\Lambda_{vv}] X \to Q_{vv},$$

where Q_{vv} is the limit of the expectation, which is assumed to exist.

So, with no misspecification,

$$Q_{\beta_v}^{-1} Q_{vv} Q_{\beta_l}^{-1} = \sigma_v^2 (X'X)^{-1} \frac{X'X}{\sigma_v^2} \sigma_v^2 (X'X)^{-1}$$

$$= \sigma_v^2 (X'X)^{-1},$$

which is the usual estimator. To derive a consistent estimator from this one would replace σ_v^2 with a consistent estimator. Now, in the case of misspecification, one has

$$Q_{\beta_v}^{-1} Q_{vv} Q_{\beta_l}^{-1} = \sigma_v^2 (X'X)^{-1} \frac{X' E[\Lambda_{vv}] X}{\sigma_v^2} \sigma_v^2 (X'X)^{-1}$$

$$= (X'X)^{-1} X' E[\Lambda_{vv}] X (X'X)^{-1}.$$

Once again, one would replace $E[\Lambda_{vv}]$ with a consistent estimator such as $\hat{\Lambda}_{vv}$, a diagonal matrix with elements $\hat{\lambda}_i = (v_i - x_i\hat{\beta}_v)$, to derive a consistent estimator for the above covariance matrix. An intuitive thought, at this point, is to test whether the model is misspecified by trying to discern how "close" $X'\hat{\Lambda}_{vv}X$ is to $X'X$. Tests of that type are considered in White (1982).

Usually the variance term σ_v^2 must be estimated, and hence it may be thought that the estimation of this parameter would have an effect on the inverse of Q_{β_v}. As it turns out, this is not the case. Noting that

$$\frac{\partial \ln f(v_i \mid x_i\beta_v)}{\partial \beta_{v_l} \partial \sigma} = -\frac{2}{\sigma_v^3}(v_i - x_i\beta_v)x_{l,i},$$

and assuming that the expectation of v_i is $w_i\tau$, then

$$\frac{1}{n}\sum_{i=1}^{n} E\left[\frac{\partial \ln f(v_i \mid x_i\beta_v)}{\partial \beta_v \partial \sigma}\right] = -\frac{2}{\sigma_v^3}\left[\frac{X'W}{n}\tau - \frac{X'X}{n}\beta_v\right]$$

will equal zero, since $\beta_v = \left[\frac{X'X}{n}\right]^{-1}\frac{X'W}{n}\tau$. Hence Q_{β_v} is of the form

$$Q_{\beta_v} = \begin{bmatrix} R & 0 \\ 0 & S \end{bmatrix},$$

so that

$$Q_{\beta_v}^{-1} = \begin{bmatrix} R^{-1} & 0 \\ 0 & S^{-1} \end{bmatrix}.$$

So one may ignore the covariance of $\hat{\beta}_v$ in the event of this type of misspecification.

Now, it is easy to show that

$$E\left[\frac{\partial \ln f\,(v_i\,|\,x_i\beta_v)}{\partial \beta_{v_i}}\frac{\partial \ln g\,(t_i\,|\,x_i\beta_t)}{\partial \beta_{t_i}}\right] = \frac{1}{\sigma_t^2}x_{k,i}x_{l,i}$$

under the assumption that $v_i - x_i\beta_v = u_v$, $t_i - x_i\beta_t = u_t$, $u_t = u_v + \varepsilon$, $E[u_v\varepsilon] = 0$, and the u_t and u_v are identically distributed (note that here it is assumed x=z). So under no misspecification, one obtains the same covariance matrix that was derived in the body of the chapter, though for this derivation, incidentally, normality of the disturbance terms is not needed to show asymptotic normality. Thus

$$Q_{\beta_v}^{-1}Q_{vt}\,Q_{\beta_t}^{-1} = \sigma_v^2(X'X)^{-1}.$$

Now, under misspecification, the matrix

$$Q_{\beta_v}^{-1}Q_{vt}\,Q_{\beta_t}^{-1} = \sigma_v^2(X'X)^{-1}\frac{X'E[\Lambda_{vt}]Z}{\sigma_v^2\sigma_t^2}\sigma_t^2(Z'Z)^{-1}$$

$$= (X'X)^{-1}X'E[\Lambda_{vt}]Z(Z'Z)^{-1},$$

where a consistent estimator is derived by replacing the expectation with $\widehat{\Lambda}_{vt}$, where the diagonal elements of this matrix are of the form $(v_i - x_i\widehat{\beta}_v)(t_i - z_i\widehat{\beta}_t)$.

For completeness, note

$$Q_{\beta_t}^{-1}Q_{tt}\,Q_{\beta_t}^{-1} = (Z'Z)^{-1}Z'E[\Lambda_{tt}]Z(Z'Z)^{-1},$$

where Λ_{tt} is diagonal with elements $(t_i - z_i\beta_t)^2$.

Now the covariance matrix of the estimators from a probit equation and the estimators from a continuous variable which is assumed to be a proxy for that variable is derived. Assume both u_v and u_t are normally distributed with means zero and variances σ_v^2 and σ_t^2, respectively. The assumption that t is a proxy for v translates then, in terms of this model, into the assumption that $\sigma_t^2 > \sigma_v^2$. The assumption is also made that the additional error in the thermometer variable is normally distributed with mean zero and variance σ_ε^2 and is uncorrelated with u_v, so that

$$u_t = u_v + u_\varepsilon. \tag{4.8}$$

The covariance of u_v and u_t is calculated by multiplying (4.8) through by u_v, taking expectations, and noting that u_ε is assumed to be independent of u_v. Then

$$Var\begin{bmatrix} u_v \\ u_t \end{bmatrix} = \begin{bmatrix} \sigma_v^2 & \sigma_v^2 \\ \sigma_v^2 & \sigma_t^2 \end{bmatrix}$$

The marginal distribution of u_v and u_t will be normal, as they are distributed bivariately normally. In particular, let f denote the marginal distribution of u_v and g denote the marginal distribution of u_t. Then

$$f\left(\frac{u_v}{\sigma_v}\right) = f\left(\frac{v - x\beta_v}{\sigma_v}\right)$$

and

$$g\left(\frac{u_v}{\sigma_t}\right) = g\left(\frac{t - x\beta_t}{\sigma_t}\right)$$

are distributed as normal variates with means zero and variances one.

Let the unobserved variable v be observed as $\tilde{v} = 2$ if $v > \alpha$, and as $\tilde{v} = 1$ if $v \leq \alpha$. The probability density function of $\tilde{v}_j$ may be derived from that of v_j, by the following manipulations:

$$Pr\ (\tilde{v}_j = 2\ |x_j\beta_v\ ) = Pr\ (\infty > v_j > \alpha\ |x_j\beta_v)$$

$$= Pr\ (\ \infty > x_j\beta_v + u_j > \alpha)$$

$$= Pr\ (\ \infty > x_j \frac{\beta_v}{\sigma_v} + \frac{u_j}{\sigma_j} > \frac{\alpha}{\sigma_v})$$

$$=1 - \Phi(\ \frac{\alpha - x_j\beta_v}{\sigma_v}),$$

and similarly

$$Pr\ (\ \tilde{v}_j = 1\ |x_j\beta_v\ ) = \Phi(\ \frac{\alpha - x_j\beta_v}{\sigma_v}).$$

With these results and the results of chapter II, the covariance of $\hat{\beta}_v$ and $\hat{\beta}_t$, where the $\hat{\beta}$'s are estimated by maximum likelihood, may be calculated. First, note that

$$\frac{\partial \ln f\ (\tilde{v})}{\partial \beta_v} = \frac{\varphi(\dfrac{\alpha - x_j\beta_v}{\sigma_v})}{1 - \Phi(\dfrac{\alpha - x_j\beta_v}{\sigma_v})} x_j, \quad \tilde{v} = 2$$

$$= -\frac{\varphi(\dfrac{\alpha - x_j\beta_v}{\sigma_v})}{\Phi(\dfrac{\alpha - x_j\beta_v}{\sigma_v})} x_j, \quad \tilde{v} = 1$$

$$\frac{\partial \ln g\,(t)}{\partial \beta_t} = (\frac{t - x_j\beta_t}{\sigma_t^2})x_j$$

$$\frac{\partial \ln g\,(t)}{\partial \sigma_t} = -\frac{1}{\sigma_t} + \frac{(t - x_j\beta_t)^2}{\sigma_t^3}$$

Letting h(t,v) represent the joint distribution of t and v (which will be bivariately normal), one has

$$E\left[\frac{\partial \ln f}{\partial \beta_{v_k}} \frac{\partial \ln g}{\partial \beta_{t_l}}\right]$$

$$= x_k x_l \left[\int_{\alpha}^{+\infty} \int_{-\infty}^{+\infty} \frac{\varphi\left(\frac{\alpha - x\beta_v}{\sigma_v}\right)}{1 - \Phi\left(\frac{\alpha - x\beta_v}{\sigma_v}\right)} \left(\frac{t - x\beta_t}{\sigma_t^2}\right) h(t,v)\, dt\, dv \right.$$

$$\left. - \int_{-\infty}^{\alpha} \int_{-\infty}^{+\infty} \frac{\varphi\left(\frac{\alpha - x\beta_v}{\sigma_v}\right)}{\Phi\left(\frac{\alpha - x\beta_v}{\sigma_v}\right)} \left(\frac{t - x\beta_t}{\sigma_t^2}\right) h(t,v)\, dt\, dv \right].$$

Let

$$\nu_t = \frac{t - x_j\beta_t}{\sigma_t}$$

and

$$\nu_v = \frac{v - x_j\beta_v}{\sigma_v}$$

Then the joint distribution of ν_t and ν_v will be bivariately normal with means zero and covariance matrix

$$Var\begin{bmatrix}\nu_v \\ \nu_t\end{bmatrix} = \begin{bmatrix} 1 & \frac{\sigma_v}{\sigma_t} \\ \frac{\sigma_v}{\sigma_t} & 1 \end{bmatrix}.$$

Making the change of variables given above, one has

$$E\left[\frac{\partial \ln f}{\partial \beta_{v_k}} \frac{\partial \ln g}{\partial \beta_{t_l}}\right]$$

$$= x_k x_l \left[\frac{\varphi(\frac{\alpha - x\beta_v}{\sigma_v})}{1 - \Phi(\frac{\alpha - x\beta_v}{\sigma_v})} \int_{\frac{\alpha - x\beta_v}{\sigma_v}}^{+\infty} \int_{-\infty}^{+\infty} \frac{\nu_t}{\sigma_t} \varphi(\nu_v, \nu_t) d\nu_v d\nu_t \right.$$

$$\left. - \frac{\varphi(\frac{\alpha - x\beta_v}{\sigma_v})}{\Phi(\frac{\alpha - x\beta_v}{\sigma_v})} \int_{-\infty}^{\frac{\alpha - x\beta_v}{\sigma_v}} \int_{-\infty}^{+\infty} \frac{\nu_t}{\sigma_t} \varphi(\nu_v, \nu_t) d\nu_v d\nu_t . \right]$$

Rewriting $\varphi(\nu_t, \nu_v)$ as $\varphi(\nu_v)\varphi(\nu_t | \nu_v)$, where $\varphi(\nu_t | \nu_v)$ is the conditional density of ν_t given ν_v, one has, as demonstrated in any introductory text on probability (see Hogg & Craig, p 112 , for example), that

$$\int_{-\infty}^{+\infty} \frac{\nu_t}{\sigma_t} \varphi(\nu_t | \nu_v) d\nu_t = \rho \frac{\sigma_t}{\sigma_v} \frac{\nu_v}{\sigma_t}$$

$$= \frac{\sigma_v^2}{\sigma_t \sigma_v} \frac{\sigma_t}{\sigma_v} \frac{\nu_v}{\sigma_t}$$

$$= \frac{\nu_v}{\sigma_t}$$

This implies that

$$E\left[\frac{\partial \ln f}{\partial \beta_{v_k}} \frac{\partial \ln g}{\partial \beta_{t_l}} \right]$$

$$= x_k x_l \left[\frac{\varphi(\frac{\alpha - x\beta_v}{\sigma_v})}{1 - \Phi(\frac{\alpha - x\beta_v}{\sigma_v})} \int_{\frac{\alpha - x\beta_v}{\sigma_v}}^{+\infty} \frac{\nu_v}{\sigma_t} \varphi(\nu_v) d\nu_v \right.$$

$$\left. - \frac{\varphi(\frac{\alpha - x\beta_v}{\sigma_v})}{\Phi(\frac{\alpha - x\beta_v}{\sigma_v})} \int_{-\infty}^{\frac{\alpha - x\beta_v}{\sigma_v}} \frac{\nu_v}{\sigma_t} \varphi(\nu_v) d\nu_v . \right]$$

and, as the anti-derivative of $\varphi(\nu_v)$ is $-\nu_v\varphi(\nu_v)$, the above may be integrated directly to give

$$= \frac{x_k x_l}{\sigma_t}\left[\frac{\varphi(\frac{\alpha - x\beta_v}{\sigma_v})}{1 - \Phi(\frac{\alpha - x\beta_v}{\sigma_v})} (-\varphi(\tau))\Big|_{\frac{\alpha - x\beta_v}{\sigma_v}}^{+\infty} \right.$$

$$\left. - \frac{\varphi(\frac{\alpha - x\beta_v}{\sigma_v})}{\Phi(\frac{\alpha - x\beta_v}{\sigma_v})} (-\varphi(\tau))\Big|_{-\infty}^{\frac{\alpha - x\beta_v}{\sigma_v}} \right]$$

$$= \frac{x_k x_l}{\sigma_t}\left[\frac{(\varphi(\frac{\alpha - x\beta_v}{\sigma_v}))^2}{1 - \Phi(\frac{\alpha - x\beta_v}{\sigma_v})} + \frac{(\varphi(\frac{\alpha - x\beta_v}{\sigma_v}))^2}{\Phi(\frac{\alpha - x\beta_v}{\sigma_v})} \right]$$

$$= \frac{x_k x_l}{\sigma_t} \frac{(\varphi(\frac{\alpha - x\beta_v}{\sigma_v}))^2}{\Phi(\frac{\alpha - x\beta_v}{\sigma_v})(1 - \Phi(\frac{\alpha - x\beta_v}{\sigma_v}))}$$

Thus,

$$\sum_{j=1}^{n} E\left[\frac{\partial \ln f_j}{\partial \beta_{v_k}} \frac{\partial \ln g_j}{\partial \beta_{t_l}} \right]$$

$$= \sum_{j=1}^{n} \left[\frac{(\varphi(\frac{\alpha - x_j\beta_v}{\sigma_v}))^2}{\Phi(\frac{\alpha - x_j\beta_v}{\sigma_v})(1 - \Phi(\frac{\alpha - x_j\beta_v}{\sigma_v}))} \right] \frac{x_{j,k} x_{j,l}}{\sigma_t}$$

$$= \sum_{j=1}^{n} \lambda_j \frac{x_{j,k} x_{j,l}}{\sigma_t},$$

where λ_j represents the quantity in brackets.

Calculation of the expected value of the mixed partials when the derivative of ln g is with respect to σ_t is done in the same manner. In this case, one has

$$E\left[\frac{\partial \ln f}{\partial \beta_{v_k}}\frac{\partial \ln g}{\partial \sigma_t}\right]$$

$$= x_k\left[\frac{\varphi(\frac{\alpha - x\beta_v}{\sigma_v})}{1-\Phi(\frac{\alpha - x\beta_v}{\sigma_v})}\int_{\frac{\alpha - x\beta_v}{\sigma_v}}^{+\infty}\int_{-\infty}^{+\infty}(-\frac{1}{\sigma_t}+\frac{\nu_t^2}{\sigma_t^3})\varphi(\nu_v,\nu_t)d\nu_v d\nu_t\right.$$

$$\left.-\frac{\varphi(\frac{\alpha - x\beta_v}{\sigma_v})}{\Phi(\frac{\alpha - x\beta_v}{\sigma_v})}\int_{-\infty}^{\frac{\alpha - x\beta_v}{\sigma_v}}\int_{-\infty}^{+\infty}(-\frac{1}{\sigma_t}+\frac{\nu_t^2}{\sigma_t^3})\varphi(\nu_v,\nu_t)d\nu_v d\nu_t\right].$$

As before, rewrite $\varphi(\nu_t,\nu_v)$ as $\varphi(\nu_v)\varphi(\nu_t\,|\,\nu_v)$. Then, as before, it is well known (see Hogg & Craig, p 112 , again), that

$$\int_{-\infty}^{+\infty}\frac{\nu_t^2}{\sigma_t}\varphi(\nu_t\,|\,u_v)d\nu_t = \sigma_t(1-\rho^2)$$

The fact that the result of this integration does not depend upon ν_v implies directly (since $E\left[\frac{\partial \ln f}{\partial \beta_{v_k}}\right] = 0$) that the expectation of the above is zero, since

$$E\left[\frac{\partial \ln f}{\partial \beta_{v_k}}\frac{\partial \ln g}{\partial \sigma_t}\right] = \sigma_t(1-\rho^2)\left[E\left[\frac{\partial \ln f}{\partial \beta_{v_k}}\right]\right] = 0.$$

Or, the above may be worked out directly to provide the same result.

Thus one may ignore the expectation of the mixed partials involving σ_t when calculating the covariance matrix.

Returning to the mixed partials with respect to β_{v_k} and β_{t_l}, as shown in Chapter 1, these observations may be "stacked" and then one has

$$\sum_{j=1}^{n} E\left[\frac{\partial \ln f_j}{\partial \beta_{v_k}} \frac{\partial \ln g_j}{\partial \beta_{t_l}}\right] = \frac{X'\Lambda X}{\sigma_t},$$

where Λ is a diagonal matrix with the individual λ_j's as the elements. Now, the full form of the covariance matrix of $\hat{\beta}_v - \beta_v, \hat{\beta}_t - \beta_t$, as given in Cox (1961) or as displayed in Chapter 2, is

$$Cov\left[\hat{\beta}_v - \beta_v, \hat{\beta}_t - \beta_t\right]$$

$$= \left[\sum_{j=1}^{n} E\left[\frac{\partial^2 \ln f_j}{\partial \beta_{v_k} \partial \beta_{t_l}}\right]\right]^{-1} \sum_{j=1}^{n} E\left[\frac{\partial \ln f_j}{\partial \beta_{v_k}} \frac{\partial \ln g_j}{\partial \beta_{t_l}}\right]\left[\sum_{j=1}^{n} E\left[\frac{\partial^2 \ln g_j}{\partial \beta_{v_k} \partial \beta_{t_l}}\right]\right]^{-1}.$$

Using the results of chapter one, the results above, and the well-known properties of the general linear model, this is equivalent to

$$= \left[X'\Lambda X\right]^{-1}\left[\frac{X'\Lambda X}{\sigma_t}\right]\left[\sigma_t^2 (X'X)^{-1}\right]$$

$$= \sigma_t (X'X)^{-1}.$$

This, then, is the covariance between the two coefficients. The entire covariance structure may be written as

$$Cov\begin{bmatrix}\hat{\beta}_v - \beta_v \\ \hat{\beta}_t - \beta_t\end{bmatrix} = \begin{bmatrix}(X'\Lambda X)^{-1} & \sigma_t (X'X)^{-1} \\ \sigma_t (X'X)^{-1} & \sigma_t^2 (X'X)^{-1}\end{bmatrix}$$

This is the covariance matrix that will be used in constructing the Wald test, following the results of Chapter 3.

Appendix II: The Wald Test for Similarity of Coefficients

The Wald test used here is more fully described in chapter 3. Let $\frac{\beta_{v_j}}{\sigma_v} = \gamma_{vj}$ and $\beta_{t_j} = \gamma_{tj}$, where β_{v_j} and β_{t_j} are the j^{th} coefficients from the models above. Then if $\beta_v = \beta_t$,

$$\frac{\gamma_{11}}{\gamma_{21}} = \frac{\gamma_{12}}{\gamma_{22}} = \cdots = \frac{\gamma_{1k}}{\gamma_{2k}}. \tag{7.9}$$

This suggests rearranging this into k-1 equations of differences. Let $\gamma = (\gamma_{11}, \ldots, \gamma_{2k})$, and define $g(\gamma) = (g_1(\gamma), \ldots, g_{k-1}(\gamma))$ by

$$\begin{aligned} g_1(\gamma) &= \gamma_{11}\gamma_{22} - \gamma_{12}\gamma_{21}, \\ g_2(\gamma) &= \gamma_{12}\gamma_{23} - \gamma_{22}\gamma_{13}, \\ &\vdots \\ g_{k-1}(\gamma) &= \gamma_{1(k-1)}\gamma_{2k} - \gamma_{1k}\gamma_{2(k-1)}. \end{aligned}$$

Under the null hypothesis (7.9) each g_j is zero. Denote the covariance matrix of γ by R. This covariance matrix was derived above and so

$$R = \begin{bmatrix} (X'\Lambda X)^{-1} & \sigma_t(X'X)^{-1} \\ \sigma_t(X'X)^{-1} & \sigma_t^2(X'X)^{-1} \end{bmatrix}$$

The probability structure was particularly simple in this case, thus allowing a theoretical derivation. It should be emphasized, however, that there is always a consistent estimator of the covariance matrix between two maximum likelihood estimators available, following the results of Cox.

The following theorem is used to construct a test.

Theorem

Suppose that a_n is a sequence of functions tending to ∞, b is a fixed number, and $a_n(Z_n - b)$ approaches in law some random variable X. Let g be a function of a real variable which is differentiable and whose derivative g' is continuous at b. Then

$$a_n[g(Z_n) - g(b)]$$

approaches g'(b)X in law.

(For a proof of this, see Bickel and Doksum, p 461).

In this case, $a_n = n^{\frac{1}{2}}$, the $Z_n = \hat{\gamma}_n$, and the b is simply γ. Calculation of the matrix

$$\left[\frac{\partial g(\gamma_{mi})}{\partial \gamma_{mj}}\right]$$

is necessary before the covariance of $g(\gamma)$ may be calculated.

The derivative for any component of the vector $g_i(\gamma)$ is simply

$$\frac{\partial g_i(\gamma)}{\partial \gamma_{1j}} = \gamma_{2,i+1}, \quad i=j$$

$$= -\gamma_{2,i}, \quad i=j+1$$

$$= 0, \quad otherwise,$$

and

$$\frac{\partial g_i(\gamma)}{\partial \gamma_{2j}} = \gamma_{1,i+1}, \quad i=j$$

$$= -\gamma_{1,i}, \quad i=j+1$$

$= 0$, *otherwise*

Let the covariance matrix of $g(\gamma)$ calculated in the above manner be denoted as W. Then the statistic

$$T_Q = g(\gamma)' W g(\gamma)$$

is distributed as χ^2 with k-1 degrees of freedom. This is the test statistic that is used in determining whether the coefficients of the linear regression with presther, congther, senther or zcand dependent are the same as the coefficients of the probit with the dependent variable being the vote (for president, congressman, senator or president, respectively).

Chapter 5: A Model of Participatory Behavior: the Conditional Motivational Probit Model[1]

The modelling of individual choices among discrete alternatives has recently attracted much attention in econometrics. In general, the development of these models assumes that the individual choosing among these alternatives random utility maximizes. He chooses, as in traditional economics, the alternative which obtains for him the greatest utility, but that utility consists of two parts: one, a function of various characteristcs relating to the individual and the particular alternative, and two, a random error term. The random error term, which represents influences and factors left out of the model, allows individuals with identical characteristics to choose different alternatives, which would invalidate a traditional economic utility-maximizing consumer model. The function of characteristics (which includes such things as attitudes, cost, socio-demographic characteristics, etc.) allows a characterization of choice demand in terms of identifiable criteria, and is assumed to represent the non-stochastic portion of the utility function. The utility function, in turn, represents an ordering of preferences the individual might hold over the different alternatives.

Explained in this manner, it is clear that the interpretation given to the function of individual and choice characteristics is somewhat arbitrary. While we are, after all, observing a behavioral propensity

1. The relationship of the conditional motivational probit model to the conditional logit and other models derived from the hypothesis of random utility maximization are described below.

associated with the various characteristics, there is no general theory behind why we might find such a propensity. Naturally, in specific applications, there may very well be a theory of the choice. For example, if the decision is to choose between several types of water heating systems, there is naturally a consideration by the individual of capital outlays and operating cost, and there is undoubtedly a tradeoff between the two, indexed by a discount rate, and that being a function both of income and of the individual's attitudes. Thus, prediction of the discount rate may be important for both studies of proposed policy changes and an understanding of the behavioral process of the individual decision making. There may often, however, be no more theory than that of modelling a particular choice process with some "likely looking" variables which are assumed to be correlated with the choice decision under study.

In particular, there is often no theory behind the relation of underlying psychological states to the individual's choice of a particular alternative. In the choice of water heating system given above, an understanding of an individual's underlying psychological state is of no doubt little interest, given the lack of emotional investment most individuals have regarding market decisions. In other types of behavior, however, an understanding of how particular attitudes and characteristics affect motivations to perform particular behaviors may be of the greatest interest for the researcher. It is to a method of modelling those that we turn to in the next section.

5.1 The Conditional Motivational Probit Model

Consider the individual who is faced with the following situation. First, there are two sides or choices of actions available to the individual associated with this situation, with the two actions being opposed to one another. Furthermore, the individual's motivation to engage in one action or the other lies along an underlying dimension of motivation to help one side or another, with neutrality lying at the zero point of motivation. Also, there are thresholds that must be passed by each individual if his underlying motivation is to be translated into action.

An example of this would be a civil war, in which the diametrically opposite actions are fighting with the rebels or fighting with the government forces. An individual would lie along the underlying dimension, either towards the side of joining the rebels or joining the government forces. The thresholds that the individual would have to pass indicates the point that motivation is turned into action. It is presumed that this threshold effect occurs because at the point where action is chosen over no action the cost to the individual (and here cost is used in a broad sense, encompassing such things as psychic cost), between choosing the action and not choosing the action is the same.

For our general exposition, then, we assume there exists an underlying dimension of motivation, and that an individual's motivation, y^*, is distributed along this dimension. We adopt the following axioms.

Axiom 1: Existence of Cost Functions

Let there be two possible actions, A and B, and let there be the possiblity of no action, call it N. Then there exists a cost function for each of the three possible actions, which are indexed by the motivation, y^*, and by exogenous variables w. Call these functions $C_A(y^*,w)$, $C_B(y^*,w)$, and $C_N(y^*,w)$. Then it is assumed that these functions are differentiable and have the following signs on their derivatives.

$$\frac{\partial C_A(y^*,w)}{\partial y^*} > 0, \tag{5.1.1}$$

$$\frac{\partial C_B(y^*,w)}{\partial y^*} < 0, \tag{5.1.2}$$

$$\frac{\partial C_N(y^*,w)}{\partial y^*} > 0, \quad y^* > 0, \tag{5.1.3}$$

and

$$\frac{\partial C_N(y^*,w)}{\partial y^*} < 0, \quad y^* < 0. \tag{5.1.4}$$

In addition, $C_N(0,w) = 0$, $C_A(0,w) = C_B(0,w) > 0$, and

$$\lim_{y^* \to -\infty} C_A(y^*,w) = 0, \tag{5.1.5}$$

$$\lim_{y^* \to +\infty} C_B(y^*,w) = 0. \tag{5.1.6}$$

The functions are displayed in Figure 5.1.

[Put Figure 5.1 about here]

The intuition behind these conditions on the cost functions is as follows. The cost of doing nothing, $C_N(y^*,w)$, is zero when there is no motivation and increases as motivation increases in magnitude. The cost associated with action A is increasing as motivation moves toward the positive end of the underlying motivational dimension, while the cost

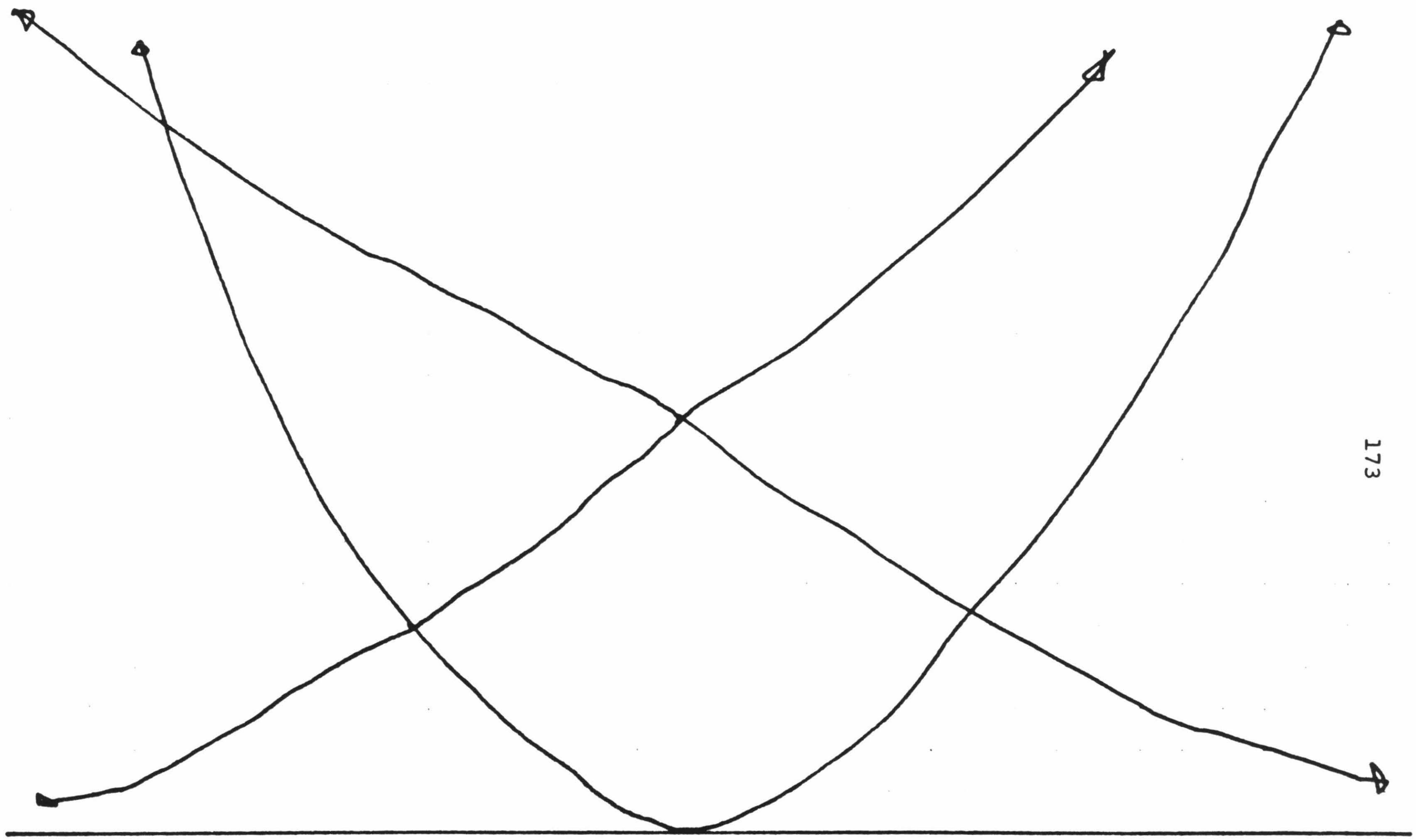


Figure 5.1 Hypothesized Cost Functions Associated with Underlying Motivation

associated with action B is increasing as motivation moves towards the negative end of the underlying motivational dimension. The conditions on the values of the various cost functions at zero states that at no level of motivation, the least costly action is doing nothing and that there is some cost to doing some action. The conditions on the limits of C_A as y^* approaches $-\infty$ and C_B as y^* approaches $+\infty$ are the parallel conditions on the zero cost of doing nothing when motivation is zero, simply saying that when motivation approaches infinity the cost of an action becomes trivial.

The next axiom gives conditions under when we may expect behavior based upon these cost functions.

Axiom 2: The Individual is a Cost Minimizer

The individual, given his level of motivation y^* and the levels of the variables w, chooses the action A,B or N which minimizes his cost.

His cost-minimizing decision will be based on the existence of the cost functions $C_N(y^*,w)$, $C_A(y^*,w)$ and $C_B(y^*,w)$ assumed by Axiom 1. The following lemma describes the nature of the cost-minimizing decison with respect to the underlying level of motivation and the conditioning variables w.

Lemma 1

For any level of the conditioning variables w, there exists points $\alpha_1(w)$ and $\alpha_2(w)$ such that the following holds: if $y^* < \alpha_1(w)$, action A is the cost-minimizing decision, if $y^* > \alpha_2(w)$, action B is the cost-minimizing decision, and if $\alpha_1(w) < y^* < \alpha_2(w)$, no action is the cost-minimizing decision. Furthermore, $\alpha_1(w) < 0$ and $\alpha_2(w) > 0$.

Proof

From Axiom 1, the conditions

$$\frac{\partial C_A(y^*,w)}{\partial y^*} > 0,$$

$$\frac{\partial C_B(y^*,w)}{\partial y^*} < 0,$$

imply that

$$\frac{\partial (C_A(y^*,w) - C_B(y^*,w)}{\partial y^*} > 0.$$

Now,

$$C_A(0,w) - C_B(0,w) = 0,$$

which implies

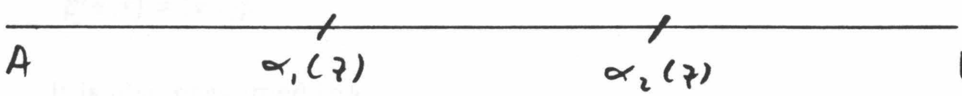
$$C_A(y^*,w) \leq C_B(y^*,w), \quad y^* < 0$$

$$C_A(y^*,w) = C_B(y^*,w), \quad y^* = 0$$

$$C_A(y^*,w) \geq C_B(y^*,w), \quad y^* > 0,$$

so the cost of action A is always less than the cost of action B on $(-\infty,0)$. We may perform a similar calculation for the difference of the $C_A(y^*,w)$ and $C_N(y^*,w)$, and determine that they intersect at one point less than zero, below which A is the cost-minimizing action and above which (at least between this point and zero), N is. This point gives us our $\alpha_1(w)$. When we perform a similar proof for $C_B(y^*,w)$ and $C_N(y^*,w)$, we obtain $\alpha_1(w)$ and the theorem.

We might note that if C_A and C_B are symmetric, this implies that $\alpha_1(w) = \alpha_2(w)$. In general, we will assume this symmetry, as there is no reason to believe otherwise. Note that we have assumed, with our conditions on the cost functions in Axiom 1, that behavior consistent with motivation is cheaper than behavior inconsistent with motivation. This makes sense from the standpoint of theories of cognitive consistency and also from a simple common sense point of view that an individual behaves in a manner which he is motivated. We may display our underlying dimension of motivation as follows.



We now introduce an axiom on the creation of motivation.

Axiom 3: Construction of Motivation

The individual has a motivational function (a function which produces his motivation on the underying motivational dimension) of the form

$$y^* = V(x,z) + u$$

where V is a non-random function of various exogenous variables x and z

while u is a stochastic error term, assumed to exist because of the omission of small influences uncorrelated with x and z.

This axiom, along with the previous two axioms, will allow us to estimate parameters in the conditional motivational probit model.

We assume that the form of the function $V(x,z)$ in the above motivational function may be written as

$$y^* = x\beta + u,$$

where

$$\beta = \beta(z,\tau),$$

where τ is some set of weights indexing z, which are assumed to affect the x's. In the case which we will use for our empirical estimation, we will let

$$\beta(z,\tau) = (z\tau)\beta.$$

It is also presumed that

$$\alpha_1 = \alpha_1[w\gamma]$$

and

$$\alpha_2 = \alpha_2[w\gamma],$$

or, in the usual case of a linear effect,

$$\alpha_1 = (w\gamma)\alpha_1$$

and

$$\alpha_2 = (w\gamma)\alpha_2.$$

The likelihood function for the conditional motivational probit model is easily written down, under our assumptions. There are three categories where an individual might be classified.[2] For ease of notation, let $\alpha_3 = +\infty$ and $\alpha_0 = -\infty$. Let $y = 1$ if action A is taken, $y = 2$ if no action is taken, and $y = 3$ if action B is taken. Let F_σ be the cumulative distribution function of the error term u, where F is a continuous distribution. Then the probability of any action y, for $y = 1,2,3$, is simply

$$Pr\ (\ y = c\ |x,z,w,\beta,\tau,\gamma\ ) = Pr\ (\ \alpha_c > y \geq \alpha_{c-1}\ |x,z,w,\beta,\tau,\gamma\ ),$$

$$= Pr\ [\ \alpha_c(w,\gamma) > x\beta(z,\tau) + u$$

$$\geq \alpha_{c-1}(w,\gamma)\ ],$$

$$= Pr\ [\ x\beta(z,\tau) + u > \alpha_{c-1}(w,\gamma)\ ]$$

$$-\ Pr\ [\ x\beta(z,\tau) + u \geq \alpha_c(w,\gamma)\ ],$$

$$= F_\sigma[\ \alpha_c(w,\gamma) - x\beta(z,\tau)\ ]$$

$$-\ F_\sigma[\ \alpha_{c-1}(w,\gamma) - x\beta(w,\gamma)\ ].$$

We will assume the simpler linear interactive specification for the effects of the z_i's, as given above. In this case, then,

$$Pr\ (\ y = c\ |x,z,w,\beta,\tau,\gamma\ ) = F_\sigma[\ (w\gamma)\alpha_c - (z\tau)x\beta\ ]$$

2. There may be more than three categories, and the extension is immediate and obvious. For example, in our civil war example, there could be five categories, ordered as follows: fight with the rebels, provide material aid to the rebels, do nothing, provide material aid to the government, fight with the government. The concept of ordering might well be considered to exist here, and be conditional upon the same variables as the three category model. A slight reworking of the axioms relating to the cost functions would be necessary.

$$- F_{\sigma}((w\gamma)\alpha_{c-1} - (z\tau)x\beta)$$

Let

$$Z_{jc} = 1,\ y = c$$

$$= 0,\ y \neq c$$

The likelihood may thus be written as

$$L(y,x,z,w,\beta,\tau,\gamma) = \prod_{j=1}^{n} \prod_{c=1}^{3}$$

$$\Big\{ F_{\sigma}[(w\gamma)\alpha_c - (z\tau)x\beta]$$

$$- F_{\sigma}[(w\gamma)\alpha_{c-1} - (z\tau)x\beta] \Big\}^{Z_{jc}}.$$

The log of the likelihood function is then

$$\ell(y,x,z,w,\beta,\tau,\gamma) = \sum_{j=1}^{n} \sum_{c=1}^{3} Z_{jc}$$

$$\ln \Big\{ F_{\sigma}[(w\gamma)\alpha_c - (z\tau)x\beta]$$

$$- F_{\sigma}[(w\gamma)\alpha_{c-1} - (z\tau)x\beta] \Big\}.$$

Due to the multiplicative nature of the $z_1\tau_1$ upon the x's and the $w\gamma$ upon the α's, we put a constant one, with weight one, in both the z's and the w's.

6.2 The Relationship between Random Utility Maximization and the

Conditional Motivational Probit

In both economics and psychology, choice behavior is often modelled as being performed under the Random Utility Maximization (RUM) hypothesis (McFadden, 1973, Tversky, 1971). The basic concept behind RUM is that the individual, if given the choice between k alternatives, will choose the i^{th} alternative if and only if

$$U_i > \max_{j \neq i} U_j.$$

Here, the U_j are utility functions which are assumed to have a random component (by far the most popular definition is $U_i = V_i(x,\beta) + \varepsilon_i$, where $V_i(x,\beta)$ is a non-stochastic function of exogenous variables x and their weights β, while ε_i is an error term which captures unspecified exogenous variables, errors in perception and optimization by the consumer, and so forth). Thus the RUM hypothesis makes operational the economic concept of the utility-maximizing consumer. A natural question which arises, then, what is the relationship between the conditional motivational probit model introduced in this chapter and the RUM hypothesis?

The first thing to note is that the conditional motivational probit as it is modelled does not satisfy the RUM hypothesis. One does not choose an action i out of k possible actions if

$$U_i > \max_{j \neq i} U_j;$$

rather, one chooses A if one's motivation is less than α_1, one does nothing if one's motivation is between α_1 and α_2, and one chooses B if one's motivation is greater than α_2. The difference is with the conditional motivational probit one is modelling motivations, rather than utility. Utility functions would then be presumed to be a function of motivations. The question thus becomes, are there a set of non-random utility functions (call them V_A for action A, V_B for action B, and V_N for no action) which are consistent with the hypothesis of an underlying level of motivation but which allow the choice process to be validly modelled by RUM?

The answer is yes, such functions do exist, and a sample are displayed in Figure 5.2.

[Put Figure 5.2 about here]

Let

$$\chi_A = \{x\beta \mid x\beta < \alpha_1\},$$

$$\chi_N = \{x\beta \mid \alpha_1 < x\beta < \alpha_2\},$$

and

$$\chi_B = \{x\beta \mid \alpha_2 < x\beta\}.$$

Then on χ_A, V_A must be greater than V_B and V_N, on χ_N, V_N must be greater than V_A and V_B, and on χ_B, V_B must be greater than V_A and V_N. The addition of a random component is then all that is necessary. Note that the property of Independence of Irrelevant Alternatives obviously cannot hold here; rather, a multinomial probit or a proximate covariance model is called for, rather than a multinomial logit. Also, in some sense, the conditional motivational probit model is more primitive, as knowledge (or

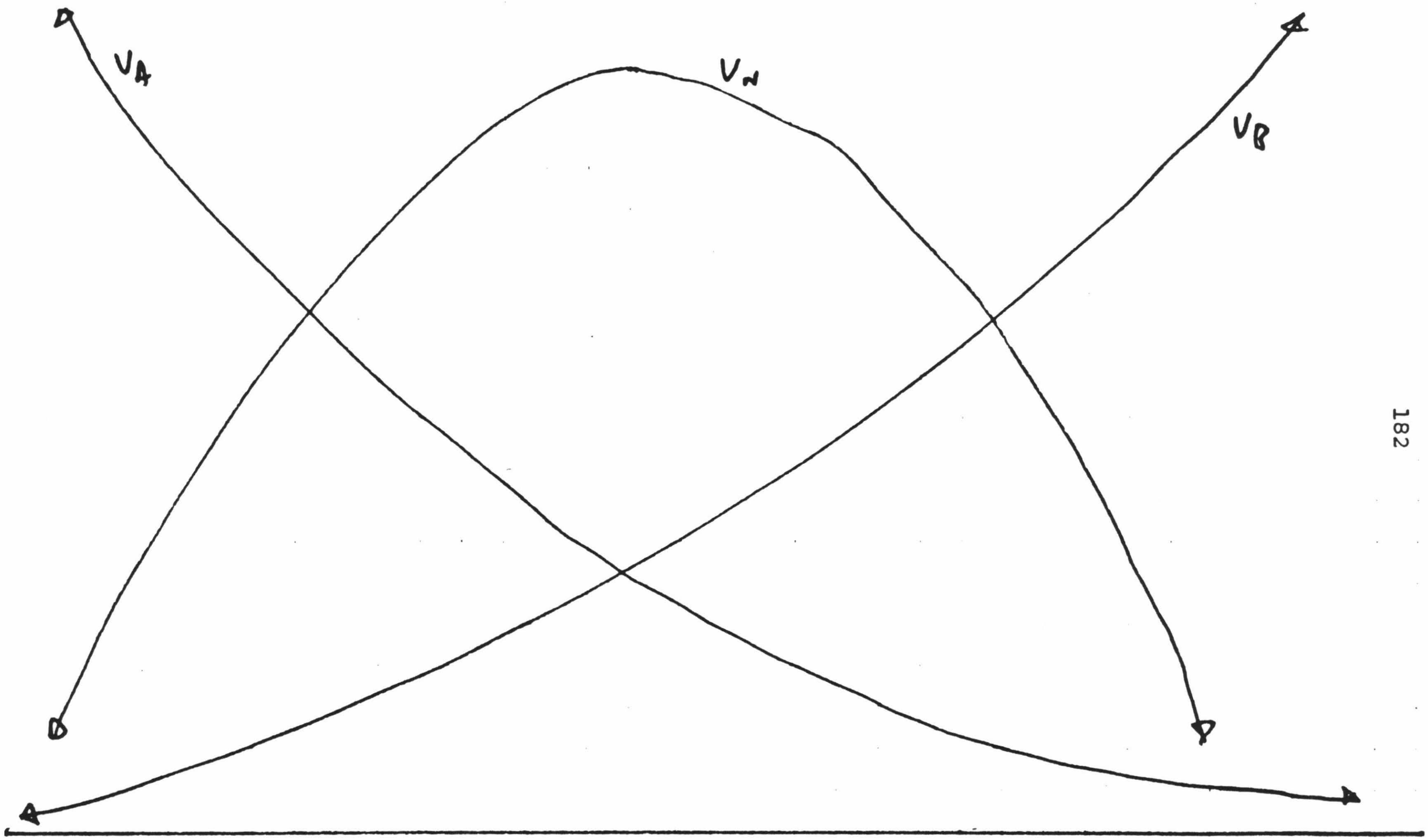


Figure 5.2 Hypothesized Utility Functions for Random Utility Maximization

assumption) of the form of the utility functions is not necessary. Rather, the assumption is that there is an underlying motivation dimension which gives rise to behavior, and that this underlying dimension can be predicted accurately.

One factor about Figure 8.1 is that the utility function for the choice of no action, V_N, cannot be linear unless it is constant across all levels of motivation.[3] Otherwise, if V_A and V_B are linear, V_N must be non-linear. The usual method in multinomial logit (or probit, which we will assume is being used here), for example, is to assume that the utility functions are linear, so a serious misspecification would result. This can be seen by setting

$$V_A = W\beta_A,$$

$$V_B = W\beta_B,$$

and

$$V_N = W\beta_N.$$

Now, suppose both V_A and V_B are linear as described above. Then if V_N was treated as the reference category, the assumption would be that V_N is described as above and therefore,

$$V_A - V_N = W(\beta_A - \beta_N).$$

If, however, as seems clear from the above exposition, $V_N = H(W, \beta_N)$, where H is some non-linear function, then the usual device of subtracting

3. This seems unlikely. That implies that an individual gets as much utility out of no action when he is highly motivated to act as when he is not motivated at all to act. Theories of cognitive dissonance would indicate this is a highly undesirable state.

off the reference category coefficients will not be adequate. Rather, as shown in McFadden (1981, footnote 26) a linear approximation to V_N can be used, which, as McFadden notes, "this justification from approximation theory for a linear-in-parameters form does not imply that this approach is efficient, or even pratical, for all applications." Whether the approximation is practical or not, it is clear that the coefficients on the exogenous variables will be different in the case of the multinomial probit than in the conditional motivational probit.

5.3 Extension to Simultaneous Equation Systems and "Economies to Scale"

Much work has been done in recent years on simultaneous equation systems involving categorical dependent variables (Maddala (1983) is the best reference). The conditional motivational probit model presented in this chapter may also be extended to systems of equations where more than one participatory behavior is being considered at a single moment in time. Furthermore, the situation where one behavior, once undertaken, makes further behaviors less costly, will also be considered and a model which describes that situation will be presented.

The usual simultaneous equation system of g equations is written in the form

$$By^* + \Gamma x = u$$

where y^* is g by 1, B is g by g, X is k by 1, Γ is g by k, and u is g by 1. Under the assumption that B is invertible, this is equivalent to

$$y^* = -B^{-1}\Gamma x + B^{-1}u,$$

$$= \Pi x + v.$$

Now, for any individual equation in this system of g equations, we have

$$y_i = \pi_i x + v_i,$$

where, π_i the i^{th} row of Π. For a simultaneous equation system for the conditional motivational probit, then, the above equation may be estimated conditional on z and τ, as outlined in the preceeding section. When done for all g equations, this will give consistent estimates of Π. These estimates may then be used to obtain, by Amemiya's method, consistent estimates of the structural parameters. Amemiya's method and a simple way of obtaining more efficient estimates by use of it is discussed in chapter 6. Computer programs for the actual calculations are discussed in chapter 9.

It is important to note that the estimation of the cutpoints for the i^{th} equation, α_{i1} and α_{i2}, come out of the reduced form estimation and are not estimated by Amemiya's method. What does come out of Amemiya's method are estimates of the structural parameters. Given our assumptions about the nature of the effect of the conditional variables z upon the behavioral weights β (a simple interactive form), the original specification may be recovered with point estimates and standard errors for all parameters (the point estimates and standard errors for the z variables come from the first stage). More complex specifications may not be amenable to a procedure such as Amemiya's and may require maximum likelihood methods. The modelling of the case when multiple behaviors have "returns to scale" requires just that.

A good example of a "returns to scale" problem is that of voting in American federal elections with the choice of choosing multiple candidates in multiple races. The main cost here is of course going to the polls, and once that price is paid, then voting for any candidates is simply a matter of pulling another lever rather than making an extra trip to the polls. If we retain our hypothesis of a single dimension for each behavior, however, with the α's retaining the interpretation of cutpoints dependent upon cost, it can be seen that going to the polls in order to vote for one candidate will make the cost of voting for other candidates much cheaper. Furthermore, since it is not clear which behavior motivates the individual to go to the polls and allows the other voting choices to be accomplished much more easily, the process must be modelled simultaneously.

Let α_{i1}, α_{i2} be the cutpoints of the i^{th} equation. Introduce the function $\chi_1, \ldots, \chi_g$, where

$$\chi_i(y^*) = 1 \text{ if } y^{j^*} < \alpha_{j1} \text{ or } y^{j^*} > \alpha_{j2}, \text{ for any } j \neq i.$$

$$= 0, \text{ otherwise}$$

Thus, for motivational dimension i, the function χ_i takes on a value of one if any of the other underlying motivations on the other motivational dimensions indicates that the behavior is to be performed. Once the behavior is to be performed on any other motivational dimension, the cost of performing a behavior for the i^{th} dimension is cheaper to some extent. We make this operational by including the χ_i in the z_{i1} variables, as

$$z_{i1}\tau_{i1} = \tilde{z}_{i1}\tilde{\tau}_{i1} + \theta_{i1}\chi_i(y^*)$$

This "dummying in" of the effects of a decision to turn out should take into account the "economies of scale" of the turnout problem. At present, the only way of estimating the likelihood is through the method of maximum likelihood. This limits the number of equations which may be estimated if the error terms are assumed normal, as maximum-likelihood techniques become increasingly unreliable as the number of equations increase.

It is sometimes desired in simultaneous equation systems to put in the observed dichotomous indicator of the underlying continuous variable as a right-hand side variable. This approach was first suggested by Heckman (1978), and results in certain constraints for the model to be "internally valid". Basically, unless these constraints are met, some values of the underlying continuous variable will not have an observable response or multiple values of the observable response will be generated by the model. This problem is exactly the same in the use of simultaneous equation methods in this model, and similar restrictions must be observed. The most complete treatment of this type of restriction is given in Schmidt (1981), and the reader is referred to that.

It may be perhaps wise at this point to quote Schmidt's warning on the problem of simultaneous equation systems in economics.

> Another point worth making is that the necessity of the constraints discussed here may raise question of whether we ought to even consider simultaneous tobit and probit models at all. It is by no means clear that the best way to model truncated and/or qualitative variables is to embed them (essentially by analogy) into the usual simultaneous equations model. At this point it is not clear what the alternatives would be. However, the difficulties one runs into, especially in the probit case, do indicate that the analogy to the usual simultaneous equations model is not completely straightforward.

Implicit in Schmidt's comments is the realization that simultaneous equation methods were developed in economics in order to model macro-economic models of the economy, where existing theory predicted, basically, a structural system of equations. The variables were, of course, all continous variables and the restriction conditions that Schmidt derives in his article do not apply to them, but his main point is clear: should we model this process in this way at all?

We think this method is reasonable for the following reasons. First, the processes developed here are not "essentially by analogy"; theories of cognitive consistency do suggest to the researcher that there are indeed simultaneities in the decision process. Furthermore, it is often clear that an individual performing an action in one sphere will indeed have an effect upon one's action in another sphere. Nowhere is this more evident than in the example when there are clear returns to scale to performing both actions simultaneously, as in voting for candidates for different offices once one has gone to the polls. Modelling the "shared" cost is a necessary part of understanding voting turnout, and that simply cannot be done in a non-simultaneous framework. Furthermore, it is not unreasonable to expect motivations for one behavior to effect motivations for another behavior. The presence of restrictions in a simultaneous system is an unfortunate fact of life. Whether or not useful modelling can be made in the presence of those restrictions is a decisions for the applied researcher to make.

5.4 An Example of the Conditional Motivational Probit: Presidential Voting in 1980

For this example of the use of the conditional motivational probit model we turn to the 1980 American federal election. For purposes of this example we will use data from the 1980 American National Election Study. The behavior which we have referred to as A in the previous sections will be voting for Ronald Reagan, the behavior which we have referred to as B will be voting for Jimmy Carter, and no behavior will be not voting. Anderson voters are excluded from this analysis.

The use of the conditional motivational probit demands the specification of two types of variables: the x's, which are the variables which are presumed to effect the motivation of the individual, and the z's, which are the variables for which the relation $y = x\beta + u$ is to be estimated conditionally upon. For the purposes of this example the x's used will be a respondent's self-placement on a scale of partisan identification (seven categories, ranging from strong Democrat to strong Republican, with independent being the reference category), and a variable called cand, which is the sum of open-ended likes and dislikes towwards the Democratic and Republican presidential candidates[4].

4. The CAND variable was first used in Campbell, Gurin, and Miller (1954). It is construoted by using responses to the question "Is there anything in particular about (candidate) for

Candidates self-placement on the partisan scale is non-zero only if the value of the cand variable is zero, as it is assumed that a large proportion of the cand variance is explained (and is casually anetecedent to) by the same partisanship as is the self-placement on the partisan scale.

For the *z*'s which are assumed to condition the motivational relationship, we use the four "Attitudes towards the voting process" questions which are available in the ANES survey.[5] These questions ask whether an individual agrees or disagrees with a statement designed to elicit certain attitudes towards the act of voting, presumably normative attitudes towards the responsibility of a citizen to vote in a democracy. From these questions we construct a four point scale, which take a value of one if a person disagrees with all four statements, two if a person disagrees with all but one statement, three if a person disagrees with two statements, and four if a person disagrees with one or zero statements.[6] This scale (with which a value of one indicates the highest category of what is referred to in the political science literature as "citizen duty") is then used to create three one-zero dummy variables, one for a value of one on the scale, one for a vaue of two on the scale, one for a value of

---

president) that might make you want to vote (for/against) him?". The number of responses (there may be up to five in all SRC's survey's during a on-year except 1972, when there were three) for liking the Democratic candidate are added to the number of responses for disliking the Republican candidate, and then the number for disliking the Democratic candidate and liking the Republican candidate are added together and this number subtract from the first number. The number may thus run from -10 to 10.

5. The questions used are 1) "It isn't so important to vote when you know your party doesn't have a chance," 2) "So many other people vote in the national elections that it doesn't matter much to me whether I vote or not," 3) "If a person doesn't care how an election comes out he shouldn't vote in it," 4) "A good many local elections aren't important enough to bother with."

6. The reason for including one or zero statements in the fourth category is the few number of respondents in either one of those categories. Incidentally, this scale is actually a Guttman scale for the vote.

three on the scale, with the reference category being those individuals who scored four on the scale. We will refer to these dummies as High duty, Moderately high duty, and Moderate duty, with the reference category being Low duty.

The choice of z_1 and z_2 was to make both of them the three dummy variables listed above. This was done on the lack of previous knowledge on how to incorporate them into the specification. While there has been a massive amount of research on the determinants of turnout, none of it suggested any means of discerning how attitudes towards the voting process should enter into the conditional motivational probit model being tested here. Thus the inclusion of both sets of variable into both z's. Additionally, τ_1 was allowed to vary for this problem for both cutpoints (α_1 and α_2), so in reality there was a τ_{11} and τ_{12}. This was done primarily due to a suspicion that the model itself was misspecified, so that variables which relate to may have been ommitted from the model. This would manifest itself in a asymmetry in the cutpoints about zero. The results in table 5.1 are thus reported with a separate set of cutpoints for each value of the scale, and the coefficients for the three dummies for the z_2's, and the values of the coefficients for the x's.

The results are given in table 5.1

Table 5.1: Turnout Example for Conditional Motivational Probit Model (1980)

Coefficient	Maximum Likelihood Estimate	Standard Error	Asymptotic T-Ratio
Alpha 1 (High Duty)	-0.4582	0.0540	-8.4918
Alpha 2 (High Duty)	0.7116	0.0582	12.2306
Alpha 1 (Moderately High Duty)	-0.7386	0.0675	-10.9430
Alpha 2 (Moderately High Duty)	0.9824	0.0723	13.5814
Alpha 1 (Moderate Duty)	-1.0726	0.1460	-7.3491
Alpha 2 (Moderate Duty)	0.8280	0.1257	6.5892
Alpha 1 (Low Duty)	-1.4265	0.2086	-6.8376
Alpha 2 (Low Duty)	1.3084	0.1979	6.6119
Strong Democrat	0.5806	0.2392	2.4267
Weak Democrat	0.2443	0.1669	1.4637
Independent Democrat	-0.0719	0.2224	-0.3233
Independent Republican	-0.7710	0.2127	-3.6246
Weak Republican	-0.7135	0.2408	-2.9633
Strong Republican	-0.8358	0.4444	-1.8807
Cand	0.2220	0.0108	20.6156
High duty	1.1113	0.0102	10.8823
Moderately High duty	1.1122	0.0158	7.1012
Moderate Duty	0.9703	0.0572	0.5192

The results of Table 5.1 are in some senses gratifying. The signs on all of the x's are of the "correct" sign and are of the expected magnitude. The α's are somewhat asymmetric, but they have the correct sign also. The effects of the underlying conditional variables are clearly greatest upon the cutpoints in this example, effecting the threshold for which motivation needs to attain before being trnaslated into action, but not affecting the motivation itself in a substantial manner[7]. The asymptotic t-ratio's are given for the difference of the coefficient from one, rather than zero.

It might also be of interest to see how well we are predicting the various categories of behavior, given the underlying level of duty. These values are presented in Table 5.2. Given the somewhat simplistic model presented here, the fit is surprisingly good.

7. The coefficient of the omitted Low Duty dummy variable for the x's is assumed to be one, since this is a multiplicative rather than an additive specification. Thus the coefficients for the other dummies need to be compared with one rather than zero.

Table 5.2: Prediction Versus Actual Behavior

High Duty

Predicted	Voted for Reagan	No Vote	Voted for Carter
Vote for Reagan	235	95	19
No Vote	32	75	75
Vote For Carter	9	63	122

Moderately High Duty

Predicted	Voted for Reagan	No Vote	Voted for Carter
Vote for Reagan	103	49	2
No Vote	58	170	67
Vote For Carter	3	23	32

Moderate Duty

Predicted	Voted for Reagan	No Vote	Voted for Carter
Vote for Reagan	6	5	0
No Vote	18	63	23
Vote For Carter	0	5	10

Low Duty

Predicted	Voted for Reagan	No Vote	Voted for Carter
Vote for Reagan	1	1	0
No Vote	8	67	11
Vote For Carter	0	0	0

Campbell, Angus, Gerald Gurin and Warren E. Miller. 1954. *The Voter Decides*. Evanston, Ill: Row, Peterson and Co.

Heckman, J.J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica*. 46; 931-959.

Maddala, G.S. 1983. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge, Mass. Cambridge University Press.

McFadden, Daniel. 1973. "Conditional Logit Analysis of Qualitative Choice." In P. Zarembka, (ed.), *Conditional Logit Analysis of Qualitative Choice Behavior*. New York: Academic Press.

McFadden, Daniel. 1981. "Econometric Model of Probabilistic Choice." In Charles F. Manski and Daniel McFadden,(eds.), *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Massachusetts: MIT Press.

Schmidt, Peter. 1981. "Constraints on the Parameters in Simultaneous Tobit and Probit Models." In Charles F. Manski and Daniel McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Massachusetts: MIT Press.

Tversky, A. 1972. "Elimination-by-Aspects: A Theory of Choice." *Psychological Review*, 79, 281-299.

Chapter 6: A Simple Method for Constructing Consistent Estimates for the Covariance of Two Maximum Likelihood Estimators Obtained from the Same Set of Observations

Situations often occur where it is desirable to obtain a consistent estimator of the covariance of two maximum likelihood estimators. Tests of significance involving two maximum likelihood estimates, for example, require the derivation of the covariance between the two estimators, as well as the more familiar derivation of the variance of the estimator. Situations where such derivations are necessary arise frequently in both theory and application, such as Hausman tests (1978), derivation of structural form coefficients from reduced form coefficients in simultaneous equation models (Amemiya (1978), Lee (1981)), and, as will be demonstrated in this chapter, in tests of consistency in ordered response models (Aitchison and Silvey (1957), Maddala (1983)). It will be shown in this chapter that the method of calculating consistent estimates between two different maximum likelihood estimators derived from the same data is valid, under certain assumptions, even in the presence of misspecification and that it is computationally simple.

6.1 Derivation of theoretical results

We assume that either the conditions of Huber (1965) or White (1982) obtain. In particular, this includes the assumption that the observations are independently and identically distributed (iid). We use a theorem of Varadarajan, quoted in Rao (1973,p 128).

Theorem

Let

$$F_n = F_n(X_n^{(1)}, \ldots, X_n^{(k)})$$

and

$$F_{\lambda n} = F_{\lambda n}(\lambda_1 X_n^{(1)} + \cdots + \lambda_k X_n^{(k)})$$

Then $F_n \to F$ iff $F_{\lambda n} \to F_\lambda$ for all $\lambda \varepsilon R^k$.

Let us consider the case of two estimators derived from the same data, call them $\hat{\beta}_1$ and $\hat{\beta}_2$. Let these estimators be defined in such a way that

$$\lambda_1 \hat{\beta}_1 + \lambda_2 \hat{\beta}_2$$

$$= \frac{1}{\sqrt{n}} \lambda_1 q_1 \sum_{i=1}^{n} S_{1i} + \frac{1}{\sqrt{n}} \lambda_2 q_2 \sum_{i=1}^{n} S_{2i}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\lambda_1 q_1 S_{1i} + \lambda_2 q_2 S_{2i})$$

It is clear from this formulation that if S_{1i} and S_{2i} are identically independently distributed, the sum above satisfies the conditions of the Lindeberg-Levy central limit theorem, and has a limiting distribution which is normal. If the expectations of S_{1i} and S_{2i} are both zero, then the variance of any term in the above sum may be written as

$$Var(\lambda_1 q_1 S_{1i} + \lambda_2 q_2 S_{2i})$$

$$= \lambda_1^2 q_1^2 Var(S_{1i}) + \lambda_2^2 q_2^2 Var(S_{2i}) + 2\lambda_1\lambda_2 q_1 q_2 E(S_{1i}S_{2i}).$$

$$= \begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix}^t \begin{bmatrix} q_1^2 Var(S_{1i}) & q_1 q_2 E(S_{1i}S_{2i}) \\ q_1 q_2 E(S_{1i}S_{2i}) & q_2^2 Var(S_{1i}) \end{bmatrix} \begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix}$$

By the theorem quoted above, since the above holds for all λ, the center matrix is our covariance matrix for $\hat{\beta}_1$ and $\hat{\beta}_2$.

Any type of estimators which are the sum of a function of realizations from some iid probability law follow this theorem. Usually instead of q_1 and q_2, we will have functions of the observations which approach in probability q_1 and q_2, but these estimators will have the same limiting distribution, by Rao (2.c.x.b). It becomes simple to calculate the covariance between two estimators, then, when they are of this form, and even simpler to obtain consistent estimators. Maximum likelihood estimators are the prime candidate for this treatment, but method of moments estimators also fall into this category.

This way of expressing the variance is important from the standpoint of maximum likelihood estimation, also. In maximum likelihood estimation in a correctly specified model, $\frac{1}{q_1} = Var(S_{1i})$ and $\frac{1}{q_2} = Var(S_{2i})$. In an incorrectly specified model, as apparently first noted by Cox (1961) and proven rigorously by Huber (1965) and White (1982), the diagonals of the center matrix are the correct variances for the individual maximum likelihood estimates. Note that the covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$ is the same expression whether or not the probability laws with respect to which they were derived are correctly specified or not.

This approach may obviously be generalized to the case where vectors of maximum likelihood estimators are being estimated, but in this case the algebra in deriving the covariance matrix is extremely tedious. The general result is the same, however. We may state it as follows.

Fact

The covariance between any two estimators $\hat{\beta}_i$ and $\hat{\beta}_j$ derived from the same (iid) data by the method of maximum likelihood may be consistently estimated by

$$nQ_i S'_i S_j Q_j,$$

where Q_i and Q_j are the matrices of minus the inverse of the second partials and S_i and S_j are matrices of the first partials.

If there are k_1 x 1 parameters in $\hat{\beta}_i$, k_2 x 2 parameters in $\hat{\beta}_j$, and n observations, then Q_i is k_1 x k_1, Q_j is k_2 x k_2, S_i is n x k_1, and S_j is n x k_2. This formula holds, incidentally, even if $\hat{\beta}_i$ and $\hat{\beta}_j$ are the same estimator. If they are the same estimator, and the model is correctly estimated, then the S'_iS_i and Q_i terms cancel, leaving the familiar nQ_i as the estimator of the covariance of estimators. Tests which make use of the fact that these two matrices do not cancel in the presence of misspecification are discussed in White (1982).

The above result, while existing in one form or another the maximum likelihood literature for years, has never, to this author's knowledge, been stated in the explicit manner above. The result, though, is not surprising; rather, it is the utility of its applications to both previously proposed problems involving maximum likelihood estimation and to new theoretical problems which is interesting. It is to that task that we now turn.

6.2 An Application to Amemiya's Principle

Amemiya (1978) suggests for estimating the a system of simultaneous equations the following principle. Let

$$y_1 = \alpha_1 y_2 + x_1\beta_1 + u_1$$

and

$$y_2 = \alpha_2 y_1 + x_2\beta_2 + u_2.$$

Let a reduced form regression (or probit, or tobit, or whatever) of y_1 and y_2 on all the exogenous variables be performed,

$$y_1 = x\pi_1 + v_1,$$

$$y_2 = x\pi_2 + v_2,$$

and then solve

$$\pi_1 = \pi_2\alpha_1 + J_1\beta_1$$

and

$$\pi_2 = \pi_1\alpha_2 + J_2\beta_2,$$

where J_1 and J_2 solve $x_1 = xJ_1$ and $x_2 = xJ_2$. π_1 and π_2 are not observed, but $\hat{\pi}_1$ and $\hat{\pi}_2$ (obtained from the reduced-form estimations) are, so the above may be rewritten as

$$\hat{\pi}_1 = \hat{\pi}_2\alpha_1 + J_1\beta_1 + w_1$$

and

$$\hat{\pi}_2 = \hat{\pi}_1\alpha_2 + J_2\beta_2 + w_2,$$

where

$$w_1 = \hat{\pi}_1 - \pi_1 - \alpha_1(\hat{\pi}_2 - \pi_2)$$

and

$$w_2 = \hat{\pi}_2 - \pi_2 - \alpha_2(\hat{\pi}_1 - \pi_1).$$

Amemiya suggests generalized least squares applied on each equation separately and shows that they are more efficient than Heckman's two-stage estimator, while Lee (1981) shows that application of Amemiya's principle is more efficient than two-stage procedures in a wide range of problems involving limited and qualitative dependent variables.

To perform this generalized least squares, it is necessary to calculate the covariance matrix of w_1 and w_2. Amemiya (and Lee) do so for the cases they consider, but the technique proposed in the last section can produce consistent estimators of the covariance matrix which allow both the calculation of the covariance between w_1 and w_2, and produce consistent estimators of the covariances even in the presence of misspecification (presuming the conditions discussed earlier in this chapter obtain). Letting Q_n^{-1} be minus the inverse of second partials of the likelihood function for the reduced-form regression for y_1, P_n^{-1} be the same for y_2, and letting S_n be the matrix of first partials for y_1 and T_n be the same for y_2, we have that

$$Var\begin{bmatrix}\hat{\pi}_1 - \pi_1 \\ \hat{\pi}_2 - \pi_2\end{bmatrix} = \begin{bmatrix}Q_n^{-1}S'_n S_n Q_n^{-1} & Q_n^{-1}S'_n T_n P_n^{-1} \\ Q_n^{-1}S'_n T_n P_n^{-1} & P_n^{-1}T'_n T_n P_n^{-1}\end{bmatrix}$$

$$= \Sigma,$$

and letting

$$A = \begin{bmatrix}1 & -\alpha_1 \\ -\alpha_2 & 1\end{bmatrix},$$

we have

$$Var\begin{bmatrix}w_1 \\ w_2\end{bmatrix} = A\Sigma A'.$$

Obviously, consistent estimators of α_1 and α_2 must be obtained for A. There are various ways of doing this, but the simplest seems to be to run the estimations in (6.2.7) and (6.2.8) as ordinary least squares regressions. This will give consistent estimates of α_1 and α_2, which may then be used in A.

The obtaining of the covariance matrix of w_1 and w_2 means that we may apply a generalized least squares procedure to both (6.2.7) and (6.2.8) simultaneously, resulting in a more efficient estimation procedure than generalized least squares applied to each separately. As noted before, these estimates are also valid estimators of the covariances even in the presence of misspecification. The ease of numerical computation with modern computers (essentially all that is happening is that derivatives are being multiplied together) makes this technique simple to program.

6.3 An Application to Ordered Response Models

Consider a model where a variable y is observed as a three when some underlying variable $y^* \geq \alpha_2$, observed as a 2 when $\alpha_2 > y^* \geq \alpha_1$, and observed as a 1 when $\alpha_1 > y^*$. Let the various probabilities associated with these events be given by

$$Pr(y = 3) = 1 - F[\alpha_2 - x\beta],$$

$$Pr(y = 2) = F[\alpha_2 - x\beta] - F[\alpha_1 - x\beta],$$

and

$$Pr(y = 1) = F[\alpha_1 - x\beta].$$

This, of course, is simply a generalization of the probit model, for which a number of variations (Amemiya(1975), McKelvey and Zavoina(1975), Atchison and Silvey (1958)) have appeared.

Now, consider a "collapsed" specification to two categories by creating a variable w, which has the property that $w = 2$ if $y = 2$ or $y = 3$, while $w = 1$ if $y = 1$. Then

$$Pr(w = 2) = 1 - F[\alpha_1 - x\beta],$$

and

$$Pr(w = 1) = F[\alpha_1 - x\beta].$$

Now, w and y have a joint density as follows:

		y		
		1	2	3
w	1	$F[\alpha_1 - x\beta]$	0	0
	2	0	$F[\alpha_2 - x\beta] - F[\alpha_1 - x\beta]$	$1 - F[\alpha_2 - x\beta]$

where the entries in this matrix are the probabilities of an event occurring.

It should be noted that by the standard results of maximum likelihood theory, if F is specified correctly (it is usually the standard normal distribution function), the maximum likelihood estimation using either w or y will produce consistent estimators. The difference between the coefficients of the β's using w and y may be seen as a test of misspecification, then, as it should go to zero under a correctly specified model and diverge from zero in an incorrectly specified model. In particular, if the distribution of the error term is correct but the assumptions on the order of the responses is incorrect for y but correct for w, then the difference of the two coefficients should diverge from zero.[1]

1. A good example of where this type of test might be useful occurs in political science over the concept of partisan or party identification, or attachment or identification with a party. The standard survey (American National Election Survey) used by most political scientists in studying partisan identification creates a seven-point scale with extreme attachment to the Democratic party at one end and extreme attachment to the Republican party at the other end. This scale is then supposed to be uni-dimensional, an interpretation which has been recently challenged (Weisberg, (1980)). It has also been suggested (Wolfinger, et. al. (1977)) that categories along this scale have been "switched", so that more extreme attachment to one party or another is incorrectly classified as less extreme attachement. Specifically, the Wlofinger hypothesis is that categories 2 and 3 and 5 and 6 on this partisan identification scale have been switched. This conjecture may also be examined by the type of test proposed here, simply by estimating the full category model and then the model with three categories, collapsing categories 1, 2 and 3 into 1, 4 into 2, and 5, 6 and 7 into 3, then testing the difference of the coefficients obtained from each estimation.

Using the results of the last section, we may evaluate the covariance of the estimators from the three-category estimation (call them $(\hat\beta^{(3)}, \alpha_1^{(3)}, \alpha_2^{(3)}) = \hat\theta^{(3)}$), and that from the two-category estimation $((\hat\beta^{(2)}, \alpha_1^{(2)}) = \hat\theta^{(2)})$. It is not known *a priori* whether the estimation with the w or the y will produce more efficient estimators, since an additional parameter α_2 must be estimated when using y. We will show that the estimation of the three-category model is asymptotically more efficient than that of the two category, even with the estimation of the additional parameter.

Now, the various derivatives with respect to the parameters are

$$\frac{\partial \ln prob\,(w=2\,|\,x,\theta^{(2)})}{\partial\beta^{(2)}} = \frac{f\,(\alpha_1^{(2)} - x\beta^{(2)})}{1 - F(\alpha_1^{(2)} - x\beta^{(2)})}x$$

$$\frac{\partial \ln prob\,(w=1\,|\,x,\theta^{(2)})}{\partial\beta^{(2)}} = -\frac{f\,(\alpha_1^{(2)} - x\beta^{(2)})}{F(\alpha_1^{(2)} - x\beta^{(2)})}x$$

and

$$\frac{\partial \ln prob\,(y=3\,|\,x,\theta^{(3)})}{\partial\beta^{(3)}} = \frac{f\,(\alpha_2^{(3)} - x\beta^{(3)})}{1 - F(\alpha_2^{(3)} - x\beta^{(3)})}x$$

$$\frac{\partial \ln prob\,(y=2\,|\,x,\theta^{(3)})}{\partial\beta^{(3)}} = -\frac{f\,(\alpha_2^{(3)} - x\beta^{(3)}) - f\,(\alpha_1 - x\beta^{(3)})}{F(\alpha_2^{(3)} - x\beta^{(3)}) - F(\alpha_1^{(3)} - x\beta^{(3)})}x$$

and

$$\frac{\partial \ln prob\,(y=1\,|\,x,\theta^{(3)})}{\partial\beta^{(3)}} = -\frac{f\,(\alpha_1^{(3)} - x\beta^{(3)})}{F(\alpha_1^{(3)} - x\beta^{(3)})}x$$

So, under the assumption that the model is specified correctly, we have

$$E\left[\frac{\partial \ln prob\,(y\,|\,x,\theta^{(3)})}{\partial \beta^{(3)}}\frac{\partial \ln prob\,(w\,|\,x,\theta^{(2)})}{\partial \beta^{(2)}}\right]$$

$$=\frac{f\,(\alpha_1^{(2)}-x\beta^{(2)})}{1-F(\alpha_1^{(2)}-x\beta^{(2)})}\frac{f\,(\alpha_2^{(3)}-x\beta^{(3)})}{1-F(\alpha_2^{(3)}-x\beta^{(3)})}$$

$$[1-F(\alpha_2^{(3)}-x\beta^{(3)})]x^2$$

$$+\frac{f\,(\alpha_1^{(2)}-x\beta^{(2)})}{1-F(\alpha_1^{(2)}-x\beta^{(2)})}\frac{f\,(\alpha_2^{(3)}-x\beta^{(3)})-f\,(\alpha_1^{(3)}-x\beta^{(3)})}{F(\alpha_2^{(3)}-x\beta^{(3)})-F(\alpha_1^{(3)}-x\beta^{(3)})}$$

$$[F(\alpha_2^{(3)}-x\beta^{(3)})-F(\alpha_2^{(3)}-x\beta^{(3)})]x^2$$

$$+\frac{f\,(\alpha_1^{(2)}-x\beta^{(2)})}{F(\alpha_1^{(2)}-x\beta^{(2)})}\frac{f\,(\alpha_1^{(3)}-x\beta^{(3)})}{F(\alpha_1^{(3)}-x\beta^{(3)})}[F(\alpha_2^{(3)}-x\beta^{(3)})]x^2$$

$$=\frac{[f\,(\alpha_1^{(2)}-x\beta^{(2)})]^2}{F(\alpha_1^{(2)}-x\beta^{(2)})[1-F(\alpha_1^{(2)}-x\beta^{(2)})]}x^2$$

$$=\lambda_i x^2$$

Also, by a similar process,

$$E\left[\frac{\partial \ln prob\,(y\,|\,x,\theta^{(3)})}{\partial \alpha_1^{(3)}}\frac{\partial \ln prob\,(w\,|\,x,\theta^{(2)})}{\partial \alpha_1^{(2)}}\right]$$

$$=\frac{[f\,(\alpha_1^{(2)}-x\beta^{(2)})]^2}{F(\alpha_1^{(2)}-x\beta^{(2)})[1-F(\alpha_1^{(2)}-x\beta^{(2)})]}$$

$$=\lambda_i$$

and

$$E\left[\frac{\partial \ln prob\,(y\,|\,x\,,\theta^{(3)})}{\partial \alpha_1^{(3)}}\,\frac{\partial \ln prob\,(w\,|\,x\,,\theta^{(2)})}{\partial \beta^{(3)}}\right]$$

$$= E\left[\frac{\partial \ln prob\,(y\,|\,x\,,\theta^{(3)})}{\partial \beta^{(3)}}\,\frac{\partial \ln prob\,(w\,|\,x\,,\theta^{(2)})}{\partial \alpha_1^{(2)}}\right]$$

$$= \frac{[f\,(\alpha_1^{(2)} - x\beta^{(2)})]^2}{F(\alpha_1^{(2)} - x\beta^{(2)})[1 - F(\alpha_1^{(2)} - x\beta^{(2)})]}x$$

$$=\lambda_i x$$

We need to inspect the derivatives with respect to α_2 to complete our calculation of the covariances of the maximum likelihood estimators from the two models. We have

$$\frac{\partial \ln prob\,(y=3\,|\,x\,,\theta^{(3)})}{\partial \alpha_2^{(3)}} = -\frac{f\,(\alpha_2^{(3)} - x\beta^{(3)})}{1 - F(\alpha_2^{(3)} - x\beta^{(3)})}$$

$$\frac{\partial \ln prob\,(y=2\,|\,x\,,\theta^{(3)})}{\partial \alpha_2^{(3)}} = \frac{f\,(\alpha_2^{(3)} - x\beta^{(3)})}{F(\alpha_2^{(3)} - x\beta^{(3)}) - F(\alpha_1^{(3)} - x\beta^{(3)})}x$$

and

$$\frac{\partial \ln prob\,(y=1\,|\,x\,,\theta^{(3)})}{\partial \alpha_2^{(3)}} = 0$$

So,

$$E\left[\frac{\partial \ln prob\,(y\,|\,x\,,\theta^{(3)})}{\partial \alpha_2^{(3)}}\,\frac{\partial \ln prob\,(w\,|\,x\,,\theta^{(2)})}{\partial \beta^{(2)}}\right]$$

$$= \frac{f(\alpha_1^{(2)} - x\beta^{(2)})}{1 - F(\alpha_1^{(2)} - x\beta^{(2)})} \frac{f(\alpha_2^{(3)} - x\beta^{(3)})}{1 - F(\alpha_2^{(3)} - x\beta^{(3)})}$$

$$[1 - F(\alpha_2^{(3)} - x\beta^{(3)})]x$$

$$- \frac{f(\alpha_1^{(2)} - x\beta^{(2)})}{1 - F(\alpha_1^{(2)} - x\beta^{(2)})} \frac{f(\alpha_2^{(3)} - x\beta^{(3)})}{F(\alpha_2^{(3)} - x\beta^{(3)}) - F(\alpha_1^{(3)} - x\beta^{(3)})}$$

$$[F(\alpha_2^{(3)} - x\beta^{(3)}) - F(\alpha_1^{(3)} - x\beta^{(3)})]x$$

$$= 0$$

Similarly, if we calculate the derivatives of $\ln prob\,(w\,|\,x,\beta^{(2)})$ with respect to $\alpha_1^{(2)}$, one may follow a similar process to obtain

$$E\left[\frac{\partial \ln prob\,(y\,|\,x,\theta^{(3)})}{\partial \alpha_2^{(3)}} \frac{\partial \ln prob\,(w\,|\,x,\theta^{(2)})}{\partial \alpha_1^{(2)}}\right] = 0$$

Hence one over the number of observations times the matrix of the expectations with respect to the parameters for the cross-partials may be written as

$$\frac{1}{n}E\left[\frac{\partial \ln prob\,(y\,|\,x,\theta^{(2)})}{\partial \theta^{(3)}} \frac{\partial \ln prob\,(w\,|\,x,\theta^{(2)})}{\partial \theta^{(2)}}\right] = \begin{bmatrix} \frac{X'\Lambda X}{n} & \frac{1'\Lambda X}{n} & 0 \\ \frac{1'\Lambda X}{n} & \frac{1'\Lambda 1}{n} & 0 \end{bmatrix}$$

where Λ is diagonal with entries λ_i, 1 is a n by 1 vector of ones, and X is a n by 1 matrix of the exogenous variables.

It is not hard to calculate the matrix of the expectations of the second partials for $(\alpha_1^{(2)},\beta^{(2)})$.[2] That matrix is simply

2. It is stated in Lee (1981, p 350), for example (there is a typographical error in that statement, incidentally).

$$P = \begin{bmatrix} \dfrac{X'\Lambda X}{n} & \dfrac{1'\Lambda X}{n} \\ \dfrac{1'\Lambda X}{n} & \dfrac{1'\Lambda 1}{n} \end{bmatrix},$$

or the same as the matrix of the expectations of the product of the first partials with respect to $(\beta^{(2)},\alpha_1^{(2)})$ and $(\beta^{(3)},\alpha_1^{(3)})$. Call the above matrix P. Call Q the matrix of the expectations of the second partials for $(\alpha_1^{(3)},\beta^{(3)})$. Then by the results in the first section of this chapter,

$$Cov[\sqrt{n}\,(\hat{\theta}^{(3)} - \theta^{(3)}), \sqrt{n}\,(\hat{\theta}^{(2)} - \theta^{(2)})]$$

$$= nP^{-1}[P \;\; 0]Q$$

$$= n[I \;\; 0]Q,$$

where I is a 2 by 2 identity matrix.

Let

$$\hat{\gamma}^{(3)} - \gamma^{(3)} = \begin{bmatrix} \hat{\beta}^{(3)} - \beta^{(3)} \\ \hat{\alpha}_1^{(3)} - \alpha_1^{(3)} \end{bmatrix}$$

This implies in particular that

$$Var[\sqrt{n}\,(\hat{\gamma}^{(3)} - \gamma^{(3)}) - \sqrt{n}\,(\hat{\theta}^{(2)} - \theta^{(2)})]$$

$$= Var[\sqrt{n}\,(\hat{\gamma}^{(3)} - \hat{\gamma}^{(3)}] + Var[\sqrt{n}\,(\hat{\theta}^{(2)} - \hat{\theta}^{(2)}]$$

$$- Cov[\sqrt{n}\,(\hat{\gamma}^{(3)} - \hat{\gamma}^{(3)}), \sqrt{n}\,(\hat{\theta}^{(2)} - \hat{\theta}^{(2)})]$$

$$= nQ + nP - 2nQ = nP - nQ \geq 0,$$

where $\geq$ is used in the sense of positive definition, since $P - Q$ is the variance of a random variable. This shows that the three category estimation is more efficient than the two category estimation (the proof may easily be extended to show that k_1 category estimation is more efficient than k_2 estimation, if $k_1 > k_2$ and the estimation with the fewer categories represents a "collapsing" of the estimation with larger categories). It also gives us the covariance matrix for the test of misspecification.

6.4 Application to Hausman Tests

One popular type of test of misspecification is what is known as the Hausman test (Hausman, 1978). The test consists of supposing two different probability laws generating relationships among the variables from the same set of data, one of which is supposed under the null hypothesis to be both true and asymptotically efficient (achieve the asymptotic Cramer-Rao lower bound), while the other one is consistent under the alternative hypothesis. In practice this consists of estimating two sets of parameters from the same data, with $\hat{\beta}_0$ being efficient under the null hypothesis and $\hat{\beta}_1$ being consistent for β_0 under both hypothesis.[3] The null hypothesis of no misspecification may then be tested by examining the statistic $\hat{q} = \hat{\beta}_0 - \hat{\beta}_1 = 0$

The advantage of the Hausman procedure is the simple calculation of the covariance matrix of $\hat{q}$. To quote Hausman (p 1253)

3. As Nelson (1981, footnote 8) points out, the condition that $\hat{\beta}_1$ be consistent for β_0 under the alternative hypothesis is not necessary for the logical consistency of the test; what is necessary is that β_1 and β_0 approach different limits under the alternative hypothesis.

> In constructing tests based on $\hat{q}$, an immediate problem comes to mind. To develop tests not only is the probability limit of $\hat{q}$ required, but the variance of the asymptotic distribution of $\sqrt{T}\hat{q}$, $\widehat{V}(q)$, must also be determined. Since $\hat{\beta}_0$ and $\hat{\beta}_1$ use the same data, they will be correlated which could lead to a messy calculation for the variance of $\sqrt{T}\hat{q}$. Luckily, this problem is resolved easily and, in fact, $\widehat{V}(q) = V(\hat{\beta}_1) - V(\hat{\beta}_0 = V_1 - V_0$ under the null hypothesis of no misspecification. Thus, the construction of specification error tests is simplified, since the estimators may be considered separately because the variance of the difference $\sqrt{T}\hat{q} = \sqrt{T}(\hat{\beta}_1 - \hat{\beta}_0)$ is the difference of the respective variances.

As is pointed out in the first section, however, consistent estimators of the asymptotic covariance of two estimators derived from the same set of data are (usually) easily obtained. Furthermore, one would want to use consistent estimators of the variance for $\hat{q}$, for the matrix $V_1 - V_0$ may not be positive definite under the alternative hypothesis, which could lead to erroneous inferences when the Wald test is applied.

Atchison, J. and S. Silvey. 1957. The Generalization of Probit Analysis to the Case of Multiple Response. *Biometrika*, Vol 44. 131-140.

Amemiya, Takeshi. 1978. The Estimation of a Simultaneous Equation Generalized Probit Model. *Econometrica*, Vol 46, # 5, (September, 1978): 1193-1206.

Cox, D. R. 1961. Tests of Separate Families of Hypothesis. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol ? Berkeley. University of California Press, 105-123.

Hausman, J. A. 1978. Specification Tests in Econometrics. *Econometrica*, Vol 46, # 6, (November, 1978): 1251-1271.

Huber, Peter J. 1967. The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions. In *Proceedings of the Fifth Berkeley Symposium*, Vol 1. Berkeley. University of California Press: 221-233.

Lee, L.F. 1981. Simultaneous Equations Models with Discrete and Censored Variables. In C. Manski and D. McFadden (eds.),*Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass. MIT Press.

McKelvey, Richard and W. Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology*, 4, 103-120.

Maddala, G.S. 1983. *Limited-dependent and Qualitative Variables in*

Econometrics. Cambridge, Mass. Cambridge University Press.

Nelson, Forrest D. A Test for Misspecification in the Censored Normal Model. *Econometrica*, Vol 49, # 5, (September, 1981): 1317-1329.

Rao, C. Radhakrishna. 1973. *Linear Statistical Inference and Its Applications*. New York, New York. John Wiley & Sons.

White, Halbert. 1982. Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, Vol 50, # 1 (January, 1982): 1-25.