Individual Choice in Political Economy

Thesis by

Langche Zeng

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1993

(Defended August 10, 1992)

Acknowledgements

During my four years in Caltech, and particularly in the process of writing this dissertation, I have received enormous help and influence from many individuals. My gratitude to them can hardly be expressed in words. My deepest thanks go to my advisor Peter Ordeshook, who has shown great care for my personal and professional growth, has given me the support and protection needed for me to pursue my academic work, and has provided the intellectual inspiration and careful guidance vital to the carrying out of my research. He also suggested the title for this dissertation. In addition, he and Betty have extended warm hospitality to me, which has made my life in a foreign country much more enjoyable.

I am grateful to my committee members, Thomas Palfrey, Rod Kiewiet and Peter Bossaerts for their availability, support, and advice; and for inspiring me, challenging me and broadening my mind with their diverse research styles. I am indebted to John Ledyard for his understanding and help through my early years in Caltech. A special note of thanks goes to Morgan Kousser, who has always cared about my progress and has been a constant source of encouragement. He has also greatly helped me with my English writing.

Chapter 1 is based on my second year paper completed under the advice of Jeffrey Dubin. The research for Chapter 2 was carried out under the advice of Tom Palfrey and Peter Ordeshook, whose enthusiasm and endless flow of ideas and suggestions have improved the quality of this chapter, and have helped me to maintain a high level of motivation. I benefited from Peter Bossaerts's insightful comments and suggestions on the econometrics part of the chapter, and other helpful comments from Elisabeth Gerber, Rod Kiewiet, Morgan Kousser, and Keith Poole. Tom Palfrey and Keith Poole provided the data. Chapter 3 was completed under the close supervision of Rod Kiewiet. I wish to thank him for his valuable advice, for his encouraging optimism, and especially for his help in collecting and validating the data. I also thank David Grether, Tom Palfrey and Peter Ordeshook for their helpful comments on this chapter.

Words do not suffice in describing my gratitude to you, Gert. Your love has brightened up my days, and your patience, understanding, support and confidence in me have helped me through the many difficult times. And thank you for the many stimulating discussions about our work; for being my listener, my proof-reader, my computer consultant; and for your unlimited help whenever needed, especially for the hours spent in the Mac-lab before the deadlines.

I wish to thank Laurel Auchampough for her administrative support. Financial support from the California Institute of Technology through teaching and research assistantships, and from the John Randolph and Dora Haynes foundation through dissertation fellowships is gratefully acknowledged.

Finally, I thank you, God, for the eternal source of love, guidance, protection; for letting my life to happen as it does, and keeping it in perspective.

Individual Choice in Political Economy

Langche Zeng

Peter C. Ordeshook, Advisor

Abstract

This dissertation consists of three relatively independent chapters that study individual choice behavior from various angles. Chapter 1 is aimed at improving the existing discrete choice models. Of the commonly used models, the probit class is computationally infeasible for problems with more than a few alternatives, and the GEV class, including the widely used logit and nested logit models, suffers from the restriction of homoscedastic disturbances. We relax the homoscedasticity restriction on the GEV class to achieve both functional flexibility and computational feasibility. The heteroscedastic logit/nested logit models are of particular practical interest.

Chapter 2 studies voting behavior in mass elections using data from the 1968 and 1980 presidential elections. We discuss theoretical and methodological issues in the specification, comparative study, and empirical testing of the rational voter models, and explore the methodological treatment of voter heterogeneity. While the standard models do not predict voting turnout well, we obtain clear evidence of strategic voting in the candidate choice decision in three candidate elections.

The data suggest voter information as one source of voter heterogeneity which introduces heteroscedasticity. The heteroscedastic logit model developed in Chapter 1 is therefore applied and is shown to outperform the standard logit model and to reveal strong effects of voter information on the turnout decision.

Chapter 3 studies choice behavior in congressional career decisions. Previous research largely focuses on the binary choices of retiring vs. seeking reelection or seeking higher office vs. seeking reelection. Using data from the 80th through the 99th congresses, we rigorously explore the congressmen's choice from all available career options, and discuss the effects of variables on both pairwise comparisons of the alternatives and on the unconditional probabilities of choosing the congressional career options. Our findings suggest that formal positions held and previous vote margins do not figure into House members' career decisions, and being a Republican *per se* does not encourage progressive ambition. We also see that a number of factors previously identified as predisposing House members to seek higher office also affect retirement decision.

Table of Contents

List of Tables and Figures

Chapter 2

Chapter 3

Introduction

Making choices is a central feature of human existence. An individual constantly faces choice situations in all aspects of life. Decisions on labor force participation, occupation, residential and work location, travel modes, durable goods purchases, voting participation, and vote choices are but a few examples.

Understanding individual choice behavior, however, is not only important for its own sake, but also vital to the understanding of many interesting aggregate phenomena, because aggregate behavior is the result of individual decisions. Indeed, the aggregate level of voting turnout, the results of a presidential election, the market demand for a commodity, the composition of the labor force, and even the size of the population, etc., are all results of individual choices. Therefore, the understanding of individual choice behavior occupies a core position in the understanding of aggregate political, economic, and social phenomena.

This dissertation presents a collection of three essays that study individual choice behavior from different angles and in different settings. Each essay constitutes a chapter of the dissertation. The chapters are relatively independent in style and content, but are connected by the unifying theme of modeling individual choice behavior under certain rationality assumptions. Chapter 1 is methodological in nature, it concerns the improvement of the existing statistical models of individual choice behavior. Chapter 2 and 3 study choice behavior of two important groups of individuals: the voters and the legislators, respectively, who are largely responsible for the operation of the democratic system and the outcomes of public policy.

Chapter 1, more precisely, considers discrete choice models.[1] For a decision maker facing a finite, exhaustive set of mutually exclusive alternatives, a discrete choice model specifies the probabilities of choosing each alternative in terms of observable independent variables and a set of unknown parameters. The values of the parameters are estimated from a sample of observed choices made by the decision makers. The effects of the independent variables on the choice probabilities, and the changes in choice probabilities following a change in exogenous variables, can then be inferred based on the estimated model. Many choice situations can be described by discrete choice models. In fact, all the choice situations mentioned in the beginning of the introduction are examples where discrete choice models are potentially useful, so are the choice situations studied in Chapter 2 and Chapter 3. A good choice model should be functionally flexible (i.e., should allow general patterns of heterogeneity) and computational feasible, and should

[1]Chapter 1 is a modified and expanded version of my second-year paper presented to the Social Science Seminar, Division of Humanities and Social Sciences, Caltech, June 11, 1990.

have good explanatory power in empirical applications. Of the existing (parametric) models, the most commonly used ones are the probit, the logit and the nested logit models. The latter two belong to the family of the generalized extreme value (GEV) model. The probit model allows general structure of the covariance matrix of the disturbances, hence enjoys functional flexibility. However, it is computationally intractable for problems with more than a few alternatives. The GEV class is amenable to computation even for large choice sets, but it suffers from the restriction of homoscedastic disturbances, which is not plausible in many choice situations. When heteroscedasticity is present, application of the GEV class is problematic, resulting in inconsistent estimators, incorrect test statistics, and misleading forecasts. The goal of Chapter 1 is to relax this homogeneity restriction of the GEV model to obtain a class of choice models that are both functionally flexible and computationally feasible. The application of one of our new models, the heteroscedastic logit model, can be found in chapter 2, where it proves to be superior to the standard model in explaining voting turnout decisions.

Chapter 2 studies voters' choice behavior in mass elections. How voters make up their minds is one of the most thoroughly studied subjects in political science because it has important consequences for the operation of democratic systems. Despite all the research that has been done in this area, however, many important issues remain controversial, or ambiguous, or largely unexamined. Among them are some theoretical and methodological issues concerning the specification and comparative study of various rational voter models for the turnout decision, the formal testing of strategic voting behavior in the candidate choice decision in the presence of a minor party candidate, and the methodological treatment of voter heterogeneity. "Rational voter models" are an important class of voting models which assume voters to be "rational" in making their voting decisions. Such models provide explicit, precise theoretical bases for voting decisions and their analysis, and lead to testable hypotheses about voting behavior. Two major models proposed in this tradition are the Downsian model based on expected utility maximization, and the "minmax regret" model based on a different decision rule in which voters minimize their maximum "regrets." For the voting turnout decision, the first model implies, operationally, that the *interaction* of the closeness of the race and the utility difference terms matter in the voting decision; while in the second model, only the utility difference terms matter. For the candidate choice decision, the Downsian model permits strategic voting, while the minmax regret model predicts sincere voting. Previous research is neither complete nor satisfactory in the empirical testing of these models. In testing the Downsian model in turnout decisions, previous research often fails to maintain the theoretical specification of the model; in comparing the two different turnout models, it often ignores data problems and settles for unreliable conclusions. Strategic voting behavior predicted by the Downsian model in multiple candidate election is seldomly tested against empirical data at all. And voter heterogeneity is another important topic that has not received satisfactory treatment in previous research. It is the goal of Chapter 2 to explore these issues in detail and to improve our understanding of voting behavior. Using data from the 1968 and 1980 presidential elections, the chapter rigorously examines the performance of rational voter models in explaining both voting turnout and vote choice decisions, and discusses a methodological treatment of voter heterogeneity inherited from diverged information levels. Our results show that standard rational voter models do not predict turnout decisions well, rather, it is social-psychological factors that are largely responsible for the variations in turnout decision making. For the candidate choice decision, however, the Downsian model does have significant explanatory power, reflecting the distinct nature of turnout and vote choice decisions, and providing clear evidence of strategic voting in the presence of a third candidate. Concerning voter heterogeneity, we show that voter information is an important source of heterogeneity, which, operationally, renders the standard homoscedastic voting models inappropriate. A heteroscedastic model developed in Chapter 1 is therefore applied, and is shown to significantly outperform the standard model, revealing stronger effects of voter information on the turnout decisions.

Chapter 3 studies choice behavior in congressional career decisions. Except the few congressmen who died in office or were expelled or were appointed to other offices, at the end of each term most members of the House face a career choice situation: they can choose to run for reelection, or to retire, or to seek other office if such an opportunity exists. A congressman's career choice decision reveals, of course, his motivation and direction of ambition, which lie at the heart of politics. Moreover, in the face of decreasing electoral competition, it is how such choices are made that increasingly determines the overall level of House turnover and the composition of the House membership, which in turn bears on the nature of public policy. Previous research on congressional career decisions mostly focuses on the binary choices of either seeking reelection versus seeking higher office or seeking reelection versus retiring, excluding members who choose the third alternative from the sample. Therefore, these studies do not analyze the unconditional probabilities of members choosing to retire, to seek reelection, or to run for higher office, but rather the conditional probabilities of choosing from a subset of alternatives given that the third alternative is not chosen. Conditional probabilities only draw an incomplete picture of the choice situation. Moreover, previous researchers often interpret variables affecting the conditional probabilities as affecting the unconditional probabilities. Doing so causes misunderstandings on the roles of the variables, and leads to failure of identifying relevant variables that enter the choice calculation. In Chapter 3, we seek to overcome the weaknesses in previous research and to study congressional career decisions in a rigorous fashion. We formulate and estimate an integrated model of congressional career decisions using data from the 80th through the 99th Congresses. Under certain rationality assumptions, we explore rigorously the choices of the congressmen among all available career options. Employing a general choice model, we estimate and discuss not only the effects of relevant variables on pairwise comparisons of the alternatives (the conditional probabilities), but also the direction and extent of the effects of independent variables on the unconditional probabilities of choosing the congressional career options. While some of our findings confirm previous research or conform to expectations, others offer fresh insight into the nature of the decision making process. Notably, we find that formal positions held and previous vote margins do not figure into House members' career decisions, that being a Republican *per se* does not encourage progressive ambition, and that a congressman's age does not offer much information on his probability of choosing to run for reelection. We also see that a number of factors previously identified as predisposing House members to seek higher office also affect retirement decision.

As mentioned earlier, the chapters are relatively independent in style and content. Nevertheless, they all contribute to the understanding of individual choice behavior, either by improving the methodology, or by adding to the substantive knowledge of choice behavior in political economy. The methodologies adopted in this dissertation are also readily applicable to other fields of social science.

Chapter 1

A Heteroscedastic GEV Model

1.1 Introduction

Discrete choice models have found wide application in a variety of fields. They are receiving increasing attention in applied work because many empirically important decisions involve choices among discrete alternatives. Examples are decisions on labor force participation, occupation, residential and work location, travel modes, durable goods purchases, voting participation, and vote choices (see, for example, Barnard and Hensher 1989; Ben-Akiva and Lerman 1985; Dubin 1985; Dubin and McFadden 1984; Enberg et al. 1990; Lee and Cohen 1985; McFadden 1978, 1979; Palfrey and Poole 1987; and Train 1980).

For a decision maker facing a finite and exhaustive set of mutually exclusive alternatives, a choice model specifies the probabilities of choosing each alternative in terms of observable independent variables and a set of unknown parameters.

The values of the parameters are estimated from a sample of observed choices made by the decision makers. The effects of the independent variables on the choice probabilities, and the changes in choice probabilities following a change in exogenous variables, can then be inferred from the estimated model. Of the existing (parametric) models, the most commonly used ones are the probit, the logit and the nested logit models. The latter two belong to the family of the generalized extreme value (GEV) model. The probit model allows general structure of the covariance matrix of the disturbances, hence enjoys functional flexibility. However, it is computationally intractable for problems with more than a few alternatives. The GEV class is amenable to computation even for large choice set, but it suffers from the restriction of homoscedastic disturbances, which is not plausible in many choice situations. When heteroscedasticity is present, application of the GEV class is problematic, resulting in inconsistent estimators, incorrect test statistics, and misleading forecasts. In this chapter, we offer a generalization of the GEV model to avoid this restriction. We show that the resulting heteroscedastic models achieve both functional flexibility and computational feasibility. We then discuss the estimation techniques and statistical testing for violation of the assumptions of the standard models. Special attention is paid to the logit model and its properties are discussed in more detail, because the standard logit model is by far the most widely used model in applications and the heteroscedastic logit model has already proved its practical value in different fields (Chapter 2, Section 4; Dubin and Zeng

1991). The heteroscedastic nested logit model is also discussed.

The chapter unfolds as follows: Section 2 provides the motivation and background for the work in this chapter by briefly reviewing the existing models and discussing their properties. Section 3 presents the improved GEV model and addresses issues in the estimation and hypothesis testing of the new model. Section 4 focuses on the heteroscedastic logit model and discusses its properties in some detail. The heteroscedastic nested logit model is also briefly discussed in this Section. Finally, Section 5 offers concluding remarks.

1.2 Discrete Choice Models

Throughout this chapter, we consider choice models based on the assumption of stochastic utility maximization. Assume a sample of T decision makers, each choosing among I_t (for notational simplicity, we drop the subscript t hereafter) discrete alternatives. Each alternative i provides utility u_{it} to individual t. u_{it} consists of a deterministic component v_{it} (usually specified as $v_{it} = \sum_{k=1}^{K} x_{it}^{k}\beta^{k}$ with x the observed characteristics and β the unknown parameters to be estimated) and an unobserved disturbance ϵ_{it} : $u_{it} = v_{it} + \epsilon_{it}$. The stochastic utility maximization assumption states that an individual t chooses alternative i if and only if it maximizes his stochastic utility, *i.e.*, if and only if $u_{it} > u_{jt} \quad \forall j \neq i$. Therefore the probability that person t chooses alternative i from the set of all I alternatives is:

$$\begin{aligned} p_{it} &= Pr\{u_{it} > u_{jt} \quad \forall j \neq i\} \\ &= Pr\{v_{it} + \epsilon_{it} > v_{jt} + \epsilon_{jt}\} \\ &= Pr\{\epsilon_{jt} < v_{it} + \epsilon_{it} - v_{jt}\} \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{v_{it}+\epsilon_{it}-v_{1t}} \cdots \int_{-\infty}^{v_{it}+\epsilon_{it}-v_{It}} f(\epsilon_{1t}, \ldots, \epsilon_{It}) d\epsilon_{It} \ldots d\epsilon_{1t} d\epsilon_{it}, \end{aligned} \quad (1.1)$$

where $f(\epsilon_{1t}, \ldots, \epsilon_{It})$ is the joint density of the ϵ_{it}. Let $F(\epsilon_{1t}, \ldots, \epsilon_{It})$ be the cumulative distribution function of the disturbances. Equation (1.1) can be equivalently expressed as:

$$p_{it} = \int_{-\infty}^{+\infty} F_i(v_{it} + \epsilon_{it} - v_{1t}, \ldots, \epsilon_{it}, \ldots, v_{it} + \epsilon_{it} - v_{It}) d\epsilon_{it}, \quad (1.2)$$

where F_i is the partial derivative of F with respect to its i^{th} argument.

Equation (1.2) shows that a particular choice model is obtained by specifying the joint distribution of the disturbances.[1] Criteria for choosing the distribution function of the disturbances include:

1. Functional flexibility—it should allow general patterns of heterogeneity and interdependence of the disturbances;

[1] A choice model can also be obtained by directly specifying a set of choice probabilities. When the choice probabilities satisfy a set of compatibility conditions(Bösch-Supan,1990), they define a choice model compatible with stochastic utility maximization with an implied joint distribution of the disturbances. In this chapter, however, we proceed by specifying the distribution function of ϵ_{it} which generates the choice model.

2. Computational feasibility—the resulting model should be amenable to computation; and

3. Consistency with empirical evidence—it should have good explanatory power in empirical applications.

The *Probit* model is obtained by assuming that the disturbances follow a multivariate normal distribution. In the *random coefficients probit* model, proposed by Housman and Wise (1978), the covariance matrix of the error terms is parameterized such that the model allows correlation among the error terms and variation in tastes across individuals. The probit class thus allows general patterns of the covariance structure of the disturbances and hence enjoys great functional flexibility. However, probit choice probabilities involve multivariate integrals of dimension $I-1$ when there are I alternatives, a feature that makes the model computationally infeasible for more than four or five alternatives. The model is therefore useful only for choice situations involving a small number of alternatives, which is a serious limitation of the model.

The *logit* model results from the assumption that ϵ_{it} have an i.i.d. extreme value distribution:[2] $F(\epsilon_{it}) = e^{-e^{-\epsilon_{it}}}$ (extreme value distribution with $\xi = 0$ and $\theta = 1$).

[2] ϵ is (type 1) extreme value distributed if its distribution function is : $F(\epsilon) = exp\{-e^{-(\epsilon-\xi)/\theta}\}$, where ξ is a location parameter and θ is a positive scaling parameter. The variance of ϵ in this case is $1/6\pi^2\theta^2$.

Then p_{it} has the form:

$$p_{it} = e^{v_{it}} / \sum_{j=1}^{I} e^{v_{jt}}. \tag{1.3}$$

The logit model (1.3) allows a ready interpretation of the choice probabilities in terms of the relative representative utilities of the alternatives. The probabilities have closed form expressions and, hence, are easy to compute. Also, large number choice problems can be decomposed easily into smaller ones and estimated sequentially (with some loss of efficiency). Because of these advantages, the logit model has received wide use in many fields. However, this model requires that the disturbances are independent and identically distributed, and it suffers from the so-called IIA (independence from irrelevant alternatives) property. The IIA property holds if the ratio of the choice probabilities of any two alternatives is not affected by the presence of other alternatives. The IIA property is not a plausible assumption in many cases (*e.g.*, when the alternative sets contains choices that are close substitutes).

The *Nested logit* model is an extension of the logit model that is not subject to the IIA property. In the nested logit model, alternatives are grouped into subsets, with "similar" alternatives falling into the same subset. The disturbance terms within the subsets are allowed to be correlated. To establish that the nested logit model is consistent with stochastic utility maximization motivated the development of the *GEV* (generalized extreme value) model (McFadden 1978). The GEV model is obtained by assuming the following joint distribution function for the disturbances:

$$F(\epsilon_{1t}, \ldots, \epsilon_{It}) = exp\{-G(e^{-\epsilon_{1t}}, \ldots, e^{-\epsilon_{It}})\}, \tag{1.4}$$

where $G(Y_1, \ldots, Y_I)$ is a nonnegative, homogeneous of degree one [3] function of $Y_i \geq 0$ with the properties that $\lim_{Y_i \to \infty} G = +\infty$ and that the l^{th} order partial derivative of G with respect to any combination of distinct Y_i's is nonnegative if l odd and nonpositive if l even. McFadden(1978) shows that under these conditions, (1.4) defines a multivariate extreme value distribution with corresponding choice probabilities

$$p_{it} = e^{v_{it}} G_i(e^{v_{1t}}, \ldots, e^{v_{It}})/G(e^{v_{1t}}, \ldots, e^{v_{It}}), \tag{1.5}$$

where G_i is the partial derivative of G with respect to its i^{th} argument.

The GEV model is actually a class of models that include the logit and the nested logit models as special cases. It can be easily verified that if $G(Y_1, \ldots, Y_I) = \sum_{i=1}^{I} Y_i$, then (1.5) reduces to the standard logit model. The nested logit model results from $G(.)$ taking the form:

$$G(Y_1, \ldots, Y_I) = \sum_{k=1}^{J} (\sum_{i \in I_k} Y_i^{1/\sigma_k})^{\sigma_k}, \tag{1.6}$$

where $I_k \subset \{1, \ldots, I\}$, $\cup_{k=1}^{J} I_k = \{1, \ldots, I\}$, and $0 < \sigma_k \leq 1$.[4] Here the I alternatives

[3] This was relaxed to degree $\mu > 0$ later. See Ben-Akiva and Lerman(1985).

[4] Equation (1.6) defines a two-level nested logit model. The results of this chapter are easily generalized to more than two-level tree structures.

are grouped into J subsets. The parameter σ_k can be interpreted as an index of the similarities of the disturbances within subset I_k . $0 < \sigma_k \leq 1$ is required for G to satisfy the conditions of the GEV model (McFadden 1978). When $\sigma_k \equiv 1$, the model reduces to the multinomial logit model.

The GEV class preserves the computational convenience of the logit model in that it still has closed form expressions. In addition, it allows a general pattern of dependence among alternatives. Note that the marginal distribution for ϵ_{it} is given by

$$\begin{aligned}\lim_{\epsilon_{jt}\to\infty\forall j\neq i} F_{it} &= exp\{-G(0,0,\ldots,e^{-\epsilon_{it}},0,\ldots,0)\} \\ &= exp\{-a_i e^{-\epsilon_{it}}\} \\ &= exp\{-e^{(\epsilon_{it}-\ln a_i)}\}, \qquad (1.7)\end{aligned}$$

where $a_i = G(0,0,\ldots,0,\overset{ith}{1},0,\ldots,0)$. In general, therefore, the disturbances are not identically distributed. Nor are they independent.

Despite its advantages, the GEV class (including the logit, the nested logit models, etc) imposes a serious restriction on the disturbances. That is, all disturbances have identical variances (homoscedasticity). This is because (1.7) is an extreme value distribution with $\xi = \ln a_i$ and $\theta = 1$, and hence the variance of ϵ_{it} is:

$$Var(\epsilon_{it}) = 1/6\pi^2, \ \forall i, \ \forall t.$$

Furthermore, this fact holds regardless of the functional form of G.

Other efforts to allow more general covariance structure of the logit disturbances include Steckel and Vanhonacker (1988)'s "heterogeneous conditional logit model", Housman and Ruud (1987)'s "heteroscedastic rank ordered logit model" and Small (1987)'s "ordered GEV" model. Steckel and Vanhonacker's model assumes the error terms follow the following distribution:

$$F(\epsilon_j) = exp[-\tau_i \beta exp(-\frac{\epsilon_i - \alpha}{\beta})],$$

where i indicates individuals, j indicates alternatives. $\tau_i > 0$ is assumed to be distributed across the population according to a gamma distribution, and $\beta > 0$ is a scale parameter. The choice probabilities obtained from this distribution have quite complicated expressions and, although the model avoids IIA, it does not get around the homoscedasticity of the error terms.

The rank ordered logit model proposed by Housman and Ruud is a generalization of the standard rank ordered logit model that takes into consideration the fact that people might rank the most preferred choice more carefully than least preferred ones. Ranking the most preferred alternative is thus based on a logit model that has a larger *common variance* of the error terms than the logit models used for ranking less preferred alternatives. We see that the basic logit models used here are still homoscedastic.

Small's ordered GEV model is an extension of the nested logit model that allows stochastic correlation among alternatives in close proximity. Because it belongs to the GEV family, it is also homoscedastic.

From the above review we see that all GEV models and their variations are subject to the restriction of homoscedastic error terms. The use of these models may cause serious problems in parameter estimation, statistical inference and forecasting when the true model is actually heteroscedastic (which is more often the case). In general, estimates of both the unknown parameters and the choice probabilities will be inconsistent. Test statistics will be incorrect, and forecasting can be misleading. Horowitz (1981) demonstrates the consequences of this specification error using hypothetical data. It is shown that the damage is often serious enough to destroy the practical value of the model. The lack of a theoretical specification for a heteroscedastic logit model also causes difficulties in developing specification tests for the standard logit model. For example, Davidson and Mackinnon (1984) perform a LM test for heteroscedasticity in binary logit models. Because they do not have a theoretical framework of a heteroscedastic logit model as the alternative model, they base the test on a specification which they think "cannot properly be called a specification of heteroscedasticity" (Davidson and Mackinnon 1984, p.247).[5] Hence Greene (1990, p.685) notes that the LM test for heteroscedasticity as described in Davidson and Mackinnon "is not well suited to the logit model." It is therefore clear that, to provide models that are better suited than the stan-

[5] Interesting enough, it turns out that their specification *is* a proper heteroscedastic logit model in our framework developed in the next section, and therefore the test they develop *is* a proper test.

dard GEV class in many empirical situations, and to facilitate the specification test for the standard models, the development of heteroscedastic GEV models, and a heteroscedastic logit model in particular, is necessary.

1.3 A Heteroscedastic GEV Model

We now present the heteroscedastic models. Intuitively, if $var(\epsilon_{it}) = C$ is a constant, then $var(\epsilon_{it}\theta_{it}) = \theta_{it}^2 C$ will be varying with θ_{it}. Therefore we can modify the generalized extreme value distribution in a simple and straightforward way to introduce heteroscedasticity:

A Heteroscedastic GEV Model

Theorem 1 *Let $G(Y_1, \ldots, Y_I)$ be the generating function in the GEV model, i.e., G is a nonnegative function of $(Y_1, \ldots, Y_I) \geq 0$ with the following properties:*

1. *G is homogeneous of degree $\mu > 0$;*

2. *$lim_{Y_i \to \infty} G(Y_1, \ldots, Y_I) = +\infty$;*

3. *For any distinct $(i_1, \ldots, i_k)$, $\partial^k G / \partial Y_{i_1} \ldots \partial Y_{i_k}$ is nonnegative if k is odd and nonpositive if k is even.*

Then,

$$\widetilde{F}(\epsilon_{1t}, \ldots, \epsilon_{It}) = exp\{-G(e^{-\epsilon_{1t}\theta_{1t}}, \ldots, e^{-\epsilon_{It}\theta_{It}})\} \tag{1.8}$$

is a multivariate extreme value distribution function if $\theta_{it} > 0$, $\forall i$, $\forall t$; *and*

$$p_{it} = \int_{-\infty}^{\infty} \theta_{it} e^{-\epsilon\theta_{it}} G_i(e^{-(v_{it}+\epsilon-v_{1t})\theta_{1t}}, \ldots, e^{-\epsilon\theta_{it}}, \ldots, e^{-(v_{it}+\epsilon-v_{It})\theta_{It}}) e^{-G(.)} d\epsilon \qquad (1.9)$$

defines a probability choice model consistent with stochastic utility maximization. When $\theta_{it} = \theta_t$,

$$p_{it} = \frac{1}{\mu} e^{v_{it}\theta_t} G_i(< e^{v_{it}\theta_t} >)/G(< e^{v_{it}\theta_t} >) \qquad (1.10)$$

where $< e^{v_{it}\theta_t} >$ *denotes a vector with its* i^{th} *element being* $< e^{v_{it}\theta_t} >$.

Proof: We first prove that $\widetilde{F}$ in (1.8) is a multivariate extreme value distribution. We need to show that, first, if $< \epsilon_{it} > \rightarrow -\infty$, then $\widetilde{F} \rightarrow 0$; second, if $< \epsilon_{it} > \rightarrow +\infty$, then $\widetilde{F} = 1$ and third, $\partial^k \widetilde{F}/\partial\epsilon_{i_1 t} \ldots \partial\epsilon_{i_k t} \geq 0$. For notational simplicity, the subscript t's are dropped in this part. But the proof is valid for all t.

Note that $\widetilde{F}(\epsilon_1, \ldots, \epsilon_I) = F(\epsilon_1\theta_1, \ldots, \epsilon_I\theta_I)$ where F is the distribution function in equation (1.4). So we have:

$$\begin{aligned}
\lim_{\epsilon_i \rightarrow -\infty \ \forall i} \widetilde{F} &= \lim_{\epsilon_i \rightarrow -\infty} F(\epsilon_1\theta_1, \ldots, \epsilon_i\theta_i, \ldots, \epsilon_I\theta_I) \\
&= F(-\infty, \ldots, -\infty) = 0 \\
\lim_{\epsilon_i \rightarrow \infty \ \forall i} \widetilde{F} &= F(\infty, \ldots, \infty) = 1 \\
\partial^k \widetilde{F}/\partial\epsilon_{i_1} \ldots \partial\epsilon_{i_k} &= \theta_{i_1} \ldots \theta_{i_k} F_{i_1 i_2 \ldots i_k} \geq 0. \qquad (1.11)
\end{aligned}$$

Equation (1.11) is valid for all i_k. Hence $\widetilde{F}$ is a proper cumulative distribution function.

The marginal distribution function for ϵ_{it} is the $\widetilde{F}$ valued at $\epsilon_{jt} = \infty \; \forall j \neq i$, i.e.,

$$\begin{aligned} \widetilde{F}_{it} &= \widetilde{F}(\infty, \ldots, \epsilon_{it}, \ldots, \infty) \\ &= exp\{-G(0, \ldots, e^{-\epsilon_{it}\theta_{it}}, \ldots, 0)\} \\ &= exp\{-a_i e^{-\epsilon_{it}\mu\theta_{it}}\} \\ &= exp\{-e^{-\frac{\epsilon_{it}-(\ln a_i)/\mu\theta_{it}}{1/(\mu\theta_{it})}}\}, \end{aligned} \tag{1.12}$$

where the third equality uses the fact that G is homogeneous of degree μ. Since (1.12) is an extreme value distribution with location parameter $\xi = (\ln a_i)/(\mu\theta_{it})$ and scale parameter $\theta = 1/(\mu\theta_{it})$, $\widetilde{F}$ is a multivariate extreme value distribution.

Note that equation (1.12) implies that the variance of ϵ_{it} is $\frac{1}{6}(\frac{\pi}{\theta_{it}\mu})^2$. Therefore, the marginal distributions for ϵ_{it} are no longer homoscedastic.

Next we derive the choice probabilities (1.9) and (1.10). Assume the disturbances follow the distribution in (1.8). Under the stochastic utility maximization assumption, the choice probabilities are given by (1.2). In our case,

$$\widetilde{F}_i = \theta_{it} e^{-\epsilon_{it}\theta_{it}} G_i exp\{-G(< e^{-\epsilon_{it}\theta_{it}} >)\}.$$

Substituting this into equation (1.2), we have

$$\begin{aligned} p_{it} &= \int_{-\infty}^{\infty} \tilde{F}_i(v_{it}+\epsilon - v_{1t}, \ldots, \epsilon, \ldots, v_{it}+\epsilon - v_{It}) d\epsilon \\ &= \int_{-\infty}^{\infty} \theta_{it} e^{-\epsilon\theta_{it}} G_i(e^{-(v_{it}+\epsilon-v_{1t})\theta_{1t}}, \ldots, e^{-\epsilon\theta_{it}}, \ldots, e^{-(v_{it}+\epsilon-v_{It})\theta_{It}}) e^{-G(.)} d\epsilon \end{aligned}$$

which is equation (1.9).

When $\theta_{it} = \theta_t$, i.e., when there is heteroscedasticity across individuals only, we have:

$$\begin{aligned} p_{it} &= \int_{-\infty}^{\infty} \theta_t e^{-\epsilon\theta_t} e^{-(v_{it}+\epsilon)\theta_t(\mu-1)} G_i(< e^{v_{it}\theta_t} >) exp\{-e^{-(v_{it}+\epsilon)\theta_t\mu} G(< e^{v_{it}\theta_t} >)\} d\epsilon \\ &= e^{-(\mu-1)\theta_t v_{it}} \theta_t G_i(.) \int_{-\infty}^{\infty} e^{-\epsilon\theta_t} e^{-(\mu-1)\epsilon\theta_t} exp\{-G(.) e^{-\mu v_{it}\theta_t} e^{-\mu\epsilon\theta_t}\} d\epsilon. \end{aligned}$$

Define $Q = e^{-(\mu-1)\theta_t v_{it}} \theta_t G_i(.)$, then

$$\begin{aligned} p_{it} &= Q \int_{-\infty}^{\infty} e^{-\mu\epsilon\theta_t} exp\{-G(.) e^{-\mu v_{it}\theta_t} e^{-\mu\epsilon\theta_t}\} d\epsilon \\ &= Q \int_{-\infty}^{\infty} \frac{1}{\mu\theta_t e^{\mu v_{it}\theta_t} G} d\left(exp\{-e^{-\mu v_{it}\theta_t} G e^{-\mu\epsilon\theta_t}\}\right) \\ &= Q \frac{1}{\mu\theta_t e^{-\mu v_{it}\theta_t} G} \\ &= \frac{e^{-(\mu-1)\theta_t v_{it}} G_i(.)}{\theta_t \frac{\mu}{\theta_t} e^{-\mu\theta_t v_{it}} G} \\ &= \frac{1}{\mu} e^{v_{it}\theta_t} G_i(< e^{v_{it}\theta_t} >)/G(< e^{v_{it}\theta_t} >). \end{aligned}$$

The first equation uses the homogeneity of degree μ of G and (hence) homogeneity of degree $\mu - 1$ of G_i. **Q.E.D.**

When $\mu = 1$ and $\theta_{it} = 1$, we get back the GEV model (1.5). Therefore the GEV model is a special case of our heteroscedastic model. So is the logit model, the nested logit model, etc. Take any G function that satisfies the conditions of the theorem, if we plug it into model (1.5), we obtain a homoscedastic model; if we plug it into (1.10), we get a model with heteroscedasticity across population only; and if we plug it into (1.9), we obtain a model with heteroscedasticity across the population as well as across alternatives.

The heteroscedastic models developed in theorem 1 allows a very general covariance structure for the disturbances. It allows correlation and heteroscedasticity across alternatives as well as across individuals. Thus it is more flexible than the GEV model. In terms of computational convenience, the model has a similar closed form as the standard GEV model for $\theta_{it} = \theta_t$. In the more general case in which $\theta_{it} \neq \theta_t$, there is not a single closed form for arbitrary G. However, the probabilities p_{it} only involve *one* dimensional integrals because $\widetilde{F}_i$ itself has a closed form. Moreover, the integrand has very good properties: it presents exponential decay and possesses a rather high degree of smoothness. Hence, the methods of numerical integration can be easily and efficiently applied, and we can expect high quality approximation and fast convergence (Davis and Robinowitz 1975).

The heteroscedastic GEV models are nearly as flexible as the probit model, but have great advantages over the latter in terms of computational cost. When the number of alternatives is large (e.g.,Daganzo and Kusnic 1990, where the number of alternatives is over 100), the probit model is not feasible and the heteroscedastic GEV model may be the only practical alternative that allows general covariance structure for the disturbances.

Estimation

The first issue in estimating the heteroscedastic GEV model is the parameterization of the unknown parameters, $\{\theta_{it}\}$. There are a large number of parameters in the new model. We have one θ_{it} for each alternative i and each individual t. We need to reduce the number of unknown parameters to a reasonable magnitude in order to allow meaningful estimation of the parameters. In econometrics work parameterization is a common practice to reduce the number of parameters. Examples can be found in Housman and Wise (1978), Davidson and Mackinnon (1984) and Engle (1984). In this study, we suggest the following parameterizations:

$$\theta_{it} = (1 + \alpha_1' Z_i + \alpha_2' Z_t)^2 \tag{1.13}$$

or

$$\theta_{it} = e^{\alpha_1' Z_i + \alpha_2' Z_t}, \tag{1.14}$$

where Z_i is a vector of m observable attributes that vary with alternatives; Z_t is a vector of n attributes that vary with t. α_1 is a vector of m parameters; and α_2 is a vector of n parameters. Z_i and Z_t can be subsets of $\{X_{it}\}$, or can consist of composite variables based on $\{X_{it}\}$, depending on the nature of the specific choice situation. In any case we expect that $m << IT$ and $n << IT$. Thus the total number of parameters is greatly reduced. Note that equation (1.13) and (1.14) satisfy the condition that all $\{\theta_{it}\}$ are positive.

The estimation of the heteroscedastic models (1.9) or (1.10) can be carried out using the maximum likelihood method. Suppose we have a random sample with observations (c_{it}, x_{it}), where $c_{it} = 1$ if individual t chooses alternative i, and $c_{it} = 0$ otherwise. The log-likelihood of observation t is:

$$l(c_t, x_t, \theta) = \sum_{i=1}^{I} c_{it} \ln p_{it},$$

where θ is the vector of all unknown parameters of p_{it}:

$$\theta = (\beta', \alpha_1', \alpha_2', \phi')', \tag{1.15}$$

where ϕ is the vector of possible parameters in G. In the case of the nested logit model (1.6), for example, $\phi = (\sigma_1, \ldots, \sigma_J)'$.

The sample log-likelihood is:

$$L(\theta) = \sum_{t=1}^{T} l(c_t, x_t, \theta).$$

The maximum likelihood estimators are obtained by solving $\partial L(\theta)/\partial\theta_i = 0$ using standard numerical maximization methods. McFadden(1984) provides a set of regularity conditions under which MLE of probability choice yields consistency and asymptotic normality. Of these conditions the most important one is that p_{it} be continuous and differentiable in the unknown parameters θ. Our generalization of the GEV class preserves this condition.

Hypothesis Testing

All econometric models are subject to potential specification errors which can damage the desired properties of the estimators. Models that are estimated by the maximum likelihood method are particularly sensitive to specification errors. Hence it is very important to develop specification tests to check the validity of the model assumptions. We now discuss a test for the presence of heteroscedasticity. The test can help us to decide whether the heteroscedastic models should be used instead of the standard, homoscedastic models. As the homoscedastic models are parametric special cases of the heteroscedastic models, LM tests can be easily constructed.

From (1.13) and (1.14) we see that when $\alpha_1 = 0$, $\theta_{it} = \theta_t$. That is, there is no heteroscedasticity across alternatives. When $\alpha_2 = 0$, then $\theta_{it} = \theta_i$, that is, there is no heteroscedasticity across individuals. Finally, when $\alpha_1 = \alpha_2 = 0$, then the model is homoscedastic. Therefore, we can test the presence of heteroscedasticity by testing the values of the α's. Write θ in (1.15) as $(\theta_1', \theta_2')'$, where $\theta_1 = \alpha_1$ for the purpose of testing $\alpha_1 = 0$; $\theta_1 = \alpha_2$ for testing $\alpha_2 = 0$; and $\theta_1 = (\alpha_1', \alpha_2')'$ for testing $\alpha_1 = \alpha_2 = 0$.

Then we want to test the null hypothesis

$$H_0: \quad \theta_1 = 0$$

Let $\widetilde{\theta} = (0', \widetilde{\theta}_2')'$, where $\widetilde{\theta}_2$ is the MLE of θ_2 under the null hypothesis. Denote $L_{\theta_1} = \partial L/\partial\theta_1$, $L_{\theta_2} = \partial L/\partial\theta_2$, $V_{\theta_1\theta_1} = EL_{\theta_1}L'_{\theta_1}$, $V_{\theta_2\theta_2} = EL_{\theta_2}L'_{\theta_2}$, $V_{\theta_1\theta_2} = EL_{\theta_1}L'_{\theta_2}$, *and* $V_{\theta_2\theta_1} = EL_{\theta_2}L'_{\theta_1}$, then the general formula of the LM test is[6]:

$$\xi_{LM} = \widetilde{L}'_{\theta_1}[\widetilde{V}_{\theta_1\theta_1} - \widetilde{V}_{\theta_1\theta_2}\widetilde{V}^{-1}_{\theta_2\theta_2}\widetilde{V}_{\theta_2\theta_1}]^{-1}\widetilde{L}_{\theta_1} \tag{1.16}$$

which follows a limiting χ^2 distribution with degree of freedom equal to the number of constraints. In (1.16) the tilde means "evaluated at $\widetilde{\theta}$."

1.4 A Heteroscedastic Logit Model

This section focuses on the heteroscedastic logit model, a special case of the heteroscedastic GEV class which is potentially useful in many applications. We discuss the specification, estimation, hypothesis testing, and properties of the choice probabilities of the model. We also briefly discuss the heteroscedastic nested logit model, which is closely related to the heteroscedastic logit model.

A Heteroscedastic Logit Model

---

[6] See Engle (1984).

Consider the case $G(Y_1, \ldots, Y_I) = \sum_{i=1}^{I} Y_i$ (the generating function for the multinomial logit model). We have $G_i = 1 \; \forall i$, and model (1.9) becomes:

$$p_{it} = \int_{-\infty}^{\infty} \theta_{it} e^{-\epsilon\theta_{it}} exp\{-\sum_{j=1}^{I} e^{-(v_{it}+\epsilon-v_{jt})\theta_{jt}}\} d\epsilon. \tag{1.17}$$

This is a heteroscedastic logit model with heteroscedasticity across population as well as across alternatives. Note that this model need not exhibit the IIA property.

If $\theta_{it} = \theta_t$, then from equation (1.10) we have:

$$p_{it} = e^{v_{it}\theta_t} / \sum_{j=1}^{I} e^{v_{jt}\theta_t}. \tag{1.18}$$

This is a heteroscedastic logit model with heteroscedasticity across population only. This model is very attractive because of the simplicity of the probability expressions. The probabilities are similar to the standard logit probabilities, except that the deterministic part of the utility functions, the v_{it}'s, are weighted by θ_t's that are inversely related to the standard error of the error terms. Intuitively this means that more weight is given to the utility functions of those individuals with smaller variances of the error terms. In the following discussions we focus on this model. It can be easily verified, however, that model (1.18) is subject to the IIA property.

If $\theta_{it} = 1, \; \forall i, \; \forall t$, then (1.18) reduce to the standard logit model, which is subject to the IIA restriction as well as homoscedasticity of the disturbances.

If $I = 2$ and θ_t is parameterized as $1/exp(z_t\beta_2)$, then model (1.18) is the specification used by Davidson and Mackinnon (1984) in developing the LM test for heteroscedasticity in the probit and the standard logit model. Therefore, as mentioned in Section 2, it is a mistake to say that the specification "cannot be properly called a specification of heteroscedasticity" and the LM test "is not well suited to the logit model."

Estimation

To estimate model (1.18), parameterize θ_t as (for example): $\theta_t = e^{\alpha' z_t}$, where z_t is a vector of n attributes that vary with t. Then,

$$\begin{aligned}
\ln p_{it} &= \beta' x_{it} e^{\alpha' z_t} - \ln \sum_{j=1}^{I} e^{\beta' x_{it} e^{\alpha' z_t}} \\
\partial \ln p_{it} / \partial \beta &= x_{it}\theta_t - \sum_{j=1}^{I} x_{jt}\theta_t p_{jt} \\
&= \theta(x_{it} - x_t), \ where \ \ x_t = \sum_{j=1}^{I} x_{jt} p_{jt} \\
\partial \ln p_{it} / \partial \alpha &= \beta' x_{it}\theta_t z_t - \sum_{j=1}^{I} \beta' x_{jt}\theta_t z_t p_{jt} \\
&= \theta_t z_t \beta'(x_{it} - x_t).
\end{aligned}$$

The log likelihood is $L = \sum_t \sum_i c_{it} \ln p_{it}$. Therefore the MLE estimators are obtained by solving:

$$\partial L / \partial \beta = \sum_t \sum_i c_{it} e^{\alpha' z_t}(x_{it} - x_t) = 0 \quad and \tag{1.19}$$

$$\partial L / \partial \alpha = \sum_t \sum_i c_{it} e^{\alpha' z_t} z_t \beta'(x_{it} - x_t) = 0. \tag{1.20}$$

Hypothesis Testing

To test the presence of heteroscedasticity, i.e., to test $\alpha' = 0$, we derive the test statistic from the general formula (1.16). From (1.19) and (1.20), we have:

$$\begin{aligned}
V_{\beta\beta} = EL_\beta L'_\beta &= \sum_t \sum_i p_{it}\theta_t^2(x_{it} - x_t)(x_{it} - x_t)' \\
V_{\alpha\alpha} = EL_\alpha L'_\alpha &= \sum_t \sum_i p_{it}\theta_t^2[\beta'(x_{it} - x_t)]^2 z_t z_t' \\
V_{\alpha\beta} = EL_\alpha L'_\beta &= \sum_t \sum_i p_{it}\theta_t^2 \beta'(x_{it} - x_t) z_t (x_{it} - x_t)' \\
V_{\beta\alpha} = EL_\beta L'_\alpha &= \sum_t \sum_i p_{it}\theta_t^2 \beta'(x_{it} - x_t)(x_{it} - x_t) z_t'.
\end{aligned}$$

Evaluating these expressions and the partial derivatives of L at $\alpha = 0$, $\beta = \widetilde{\beta}$ and $p_{it} = \widetilde{p}_{it}$ estimated from the homoscedastic model (i.e., the standard logit model), we get:

$$\begin{aligned}
\widetilde{L}_\beta &= \sum_t \sum_i c_{it}(x_{it} - x_t) \\
\widetilde{L}_\alpha &= \sum_t \sum_i c_{it} z_t \widetilde{\beta}'(x_{it} - x_t) \\
\widetilde{V}_{\beta\beta} &= \sum_t \sum_i \widetilde{p}_{it}(x_{it} - x_t)(x_{it} - x_t)' \\
\widetilde{V}_{\alpha\alpha} &= \sum_t \sum_i \widetilde{p}_{it}[\widetilde{\beta}'(x_{it} - x_t)]^2 z_t z_t' \\
\widetilde{V}_{\alpha\beta} &= \sum_t \sum_i \widetilde{p}_{it}\widetilde{\beta}'(x_{it} - x_t) z_t (x_{it} - x_t)' \quad and \\
\widetilde{V}_{\beta\alpha} &= \sum_t \sum_i \widetilde{p}_{it}\widetilde{\beta}'(x_{it} - x_t)(x_{it} - x_t) z_t'.
\end{aligned}$$

Substituting these values into equation (1.16) (with $\theta_1 = \alpha$ and $\theta_2 = \beta$), we obtain the LM test statistic for presence of heteroscedasticity of the multinomial logit model, with model (1.18) as the alternative heteroscedastic model. Furthermore, the test statistic ξ_{LM} has the same expression if we parameterize θ_t as $(1 + \alpha' z_t)^2$. We conjecture that this will be the case for any parameterization $h(\alpha' z_t)$ of θ_t such that h is second order continuous and $h(0) = 1$.

Note that if we parameterize θ_t as a linear (in parameters) function of observable variables (provided that $\theta_t > 0$ is guaranteed), then the parameters in θ_t cannot be identified from β. The estimation of the heteroscedastic logit model can then be easily carried out using a standard logit package after properly transforming the independent variables.

Other interesting tests for the heteroscedastic logit model include the test of the IIA property of model (1.18). There is a rich literature on testing the IIA assumption for the *standard* multinomial logit model (e.g., Housman and McFadden 1984; McFadden 1987; McFadden, Train and Tye 1976; Small and Hsiao 1985; and Wills 1987). We expect that with some appropriate adjustment, results from this research can be readily applied to the heteroscedastic model.

Derivatives and elasticities of choice probabilities

One purpose of estimating a choice model is to know how the choice probabilities change in response to a change in some observed variables, or how an observed factor affects the probabilities of choosing each alternative. For this purpose the derivatives and/or elasticities of the choice probabilities need to be derived. Unlike linear models in which the derivatives of the dependent variable with respect to the independent variables are directly observable from the estimated coefficients, most probability choice models are non-linear models and the expressions for the derivatives/elasticities can be quite complicated. We now derive the derivatives and elasticities of the choice probabilities for our heteroscedastic logit model.

Assume the deterministic utility functions v_i are linear in parameters,[7] that is,

$$v_i = \sum_{m=1}^{M} \beta_m x_m^i + \sum_{n=1}^{N} \gamma_n^i y_n, \tag{1.21}$$

where x denotes alternative-specific variables and y denotes individual-specific variables. In studying how consumers choose from different types of automobiles, for example, x might include such variables as the price of each type and the gas-efficiency of each type, with x^i indicating the corresponding price or gas-efficiency of type i automobile; and y might include the buyer's income, which affect different v_i's by having different coefficients. The probability expressions of the heteroscedastic logit model are as in (1.18). Then,

$$\frac{\partial p_i}{\partial x_k^s} = \sum_j \frac{\partial p_i}{\partial \theta v_j} \frac{\partial \theta v_j}{\partial x_k^s}. \tag{1.22}$$

It can be verified that

$$\frac{\partial p_i}{\partial \theta v_j} = p_i(D_1 - p_j), \tag{1.23}$$

[7] Hereafter we drop the subscript t in v_{it} for notational simplicity.

where $D_1 = 1$ if $j = i$, and $D_1 = 0$ otherwise. From (1.21), we have

$$\frac{\partial \theta v_j}{\partial x_k^s} = \frac{\partial \theta}{\partial x_k^s} v_j + \frac{\partial v_j}{\partial x_k^s}\theta = \frac{\partial \theta}{\partial x_k^s} v_j + D_2 \beta_k \theta, \tag{1.24}$$

where $D_2 = 1$ if $j = s$, and $D_2 = 0$ otherwise. Combining (1.23) and (1.24), we have

$$\begin{aligned}
\frac{\partial p_i}{\partial x_k^s} &= \sum_j p_i (D_1 - p_j)[\frac{\partial \theta}{\partial x_k^s} v_j + D_2 \beta_k \theta] \\
&= p_i [\frac{\partial \theta}{\partial x_k^s} \sum_j (D_1 - p_j) v_j + \theta \beta_k \sum_j D_2 (D_1 - p_j)] \\
&= p_i [\frac{\partial \theta}{\partial x_k^s} (v_i - \sum_j p_j v_j) + \theta \beta_k (D - p_s)],
\end{aligned} \tag{1.25}$$

where $D = 1$ if $s = i$, and $D = 0$ otherwise.

Next we derive $\frac{\partial p_i}{\partial y_k}$ in a similar fashion. We have

$$\frac{\partial p_i}{\partial y_k} = \sum_j \frac{\partial p_i}{\partial \theta v_j} \frac{\partial \theta v_j}{\partial y_k}. \tag{1.26}$$

We already know that

$$\frac{\partial p_i}{\partial \theta v_j} = p_i (D_1 - p_j) \tag{1.27}$$

where $D_1 = 1$ if $j = i$, and $D_1 = 0$ otherwise. From (1.21), we also know that

$$\frac{\partial \theta v_j}{\partial y_k} = \frac{\partial \theta}{\partial y_k} v_j + \frac{\partial v_j}{\partial y_k}\theta = \frac{\partial \theta}{\partial y_k} v_j + \gamma_k^j \theta. \tag{1.28}$$

Combining (1.27) and (1.28), we have

$$\begin{aligned}
\frac{\partial p_i}{\partial y_k} &= \sum_j p_i(D_1 - p_j)[\frac{\partial \theta}{\partial y_k} v_j + \gamma_k^j \theta] \\
&= p_i[\frac{\partial \theta}{\partial y_k} \sum_j (D_1 - p_j) v_j + \theta \sum_j (D_1 - p_j)\gamma_k^j] \\
&= p_i[\frac{\partial \theta}{\partial y_k}(v_i - \sum_j p_j v_j) + \theta(\gamma_k^j - \sum_j p_j \gamma_k^j)]. \qquad (1.29)
\end{aligned}$$

The elasticities of choice probabilities can be easily obtained from the derivatives (1.25) and (1.29):

$$\begin{aligned}
E_{p_i x_k^s} &= \frac{x_k^s}{p_i}\frac{\partial p_i}{\partial x_k^s} = x_k^s[\frac{\partial \theta}{\partial x_k^s}(v_i - \sum_j p_j v_j) + \theta \beta_k (D - p_s)] \qquad (1.30) \\
E_{p_i y_k} &= \frac{y_k}{p_i}\frac{\partial p_i}{\partial y_k} = y_k[\frac{\partial \theta}{\partial y_k}(v_i - \sum_j p_j v_j) + \theta(\gamma_k^j - \sum_j p_j \gamma_k^j)]. \qquad (1.31)
\end{aligned}$$

When $\theta = 1$, and therefore $\frac{\partial \theta}{\partial x_k^s} = \frac{\partial \theta}{\partial y_k} = 0$, (1.25) and (1.29) become the derivatives of the standard logit choice probabilities:

$$\begin{aligned}
\frac{\partial p_i}{\partial x_k^s} &= p_i \beta_k (D - p_s) \qquad (1.32) \\
\frac{\partial p_i}{\partial y_k} &= p_i(\gamma_k^j - \sum_j p_j \gamma_k^j) \qquad (1.33)
\end{aligned}$$

and (1.30) and (1.31) become the elasticities of the standard logit choice probabilities:

$$\begin{aligned}
E_{p_i x_k^s} &= x_k^s \beta_k (D - p_s) \qquad (1.34) \\
E_{p_i y_k} &= y_k(\gamma_k^j - \sum_j p_j \gamma_k^j). \qquad (1.35)
\end{aligned}$$

A Heteroscedastic Nested Logit Model

We have seen that the heteroscedastic logit model is a special case of the heteroscedastic GEV model when $G(Y_1, \ldots, Y_I) = \sum_{i=1}^{I} Y_i$. Another special case of the heteroscedastic GEV model is the heteroscedastic nested logit model, resulted from taking the G function to be[8]

$$G(Y_1, Y_2, Y_3) = Y_1 + (Y_2^{1/\sigma} + Y_3^{1/\sigma})^{\sigma}. \tag{1.36}$$

Substituting this G function and its partial derivatives into model (1.10), we can derive the choice probabilities of a nested logit model with heteroscedasticity across individuals:

$$p_1 = \frac{e^{\theta v_1}}{e^{\theta v_1} + (e^{\theta v_2/\sigma} + e^{\theta v_3/\sigma})^{\sigma}} \tag{1.37}$$

$$p_{(2,3)} = \frac{(e^{\theta v_2/\sigma} + e^{\theta v_3/\sigma})^{\sigma}}{e^{\theta v_1} + (e^{\theta v_2/\sigma} + e^{\theta v_3/\sigma})^{\sigma}} \tag{1.38}$$

$$p_{j|(2,3)} = \frac{e^{\theta v_j/\sigma}}{e^{\theta v_2/\sigma} + e^{\theta v_3/\sigma}}. \tag{1.39}$$

Define the *inclusive value* I as

$$I = \log(e^{\theta v_2/\sigma} + e^{\theta v_3/\sigma}). \tag{1.40}$$

[8]For notational simplicity, we consider the G function for a three alternative, two-level-tree nested logit model, with the second and the third alternative being the "similar" ones (see, for example, Maddala 1983 for a discussion of the nested logit model). The results here can be readily generalized to cases with more alternatives and higher level tree structures.

Then we can rewrite (1.37) and (1.38) as

$$p_1 = \frac{e^{\theta v_1}}{e^{\theta v_1} + e^{\sigma I}} \tag{1.41}$$

$$p_{(2,3)} = \frac{e^{\sigma I}}{e^{\theta v_1} + e^{\sigma I}}. \tag{1.42}$$

Careful examination of (1.39),(1.40),(1.41), and (1.42) shows that, similar to the standard nested logit model, the heteroscedastic nested logit model can be estimated sequentially (with some loss of efficiency). We can first estimate the coefficients in v_2, v_3 and θ from the conditional choice model (1.39), which is a heteroscedastic logit model.[9] We then calculate the inclusive value I by (1.40). Finally, parameters in v_1 and σ can be estimated from (1.41) and (1.42), which are probabilities of a standard logit model. The reason why (1.41) and (1.42) are standard logit probabilities is as follows: if the parameters in θ can be identified from those in v (e.g., when θ is parameterized as $e^{\alpha z}$), then they can be estimated in the first stage, and in the second stage θ is like an observed independent variable that can be easily incorporated into v_1; if the parameters in θ cannot be identified from those in v (e.g., when θ is parameterized as αz), then model (1.41) and (1.42) can be estimated by a standard logit package after proper transformation of the independent variables.

The nested logit model avoids the IIA restriction of the standard logit model, the heteroscedastic nested logit model further relaxes the assumption of homoscedastic

[9] Note that σ cannot be identified from the coefficients in v_2 and v_3.

errors in the nested logit model, and therefore achieves great functional flexibility. Its estimation can be carried out either with a full information maximum likelihood procedure, or, at some cost of efficiency, by the sequential procedure outlined above. Because of the functional similarity between a standard nested logit model and the heteroscedastic one, we expect that many results relevant to the standard nested logit model can be easily extended to the heteroscedastic model. Examples include results on correcting the errors in the covariance matrix estimated at the second stage of the sequential estimating procedure (Amemiya 1978) and results on specification tests of the nested logit model (e.g., Horowitz 1987).

1.5 Conclusions

This chapter relaxes the homoscedasticity restriction of the GEV models. The resulting heteroscedastic models preserve the computational ease of the GEV family, and allow general patterns of covariance structure of the disturbances. The heteroscedastic GEV model is nearly as flexible as the probit model, but has great advantage over the latter because it can handle problems with a large number of alternatives, while the probit model is computationally infeasible when there are more than a few alternatives. Of particular practical interest are the heteroscedastic logit and nested logit models. Both models have simple expressions for the choice probabilities that are similar to their standard counterparts. Estimation of these new models are implementable with standard maximum likelihood technique and straightforward tests have been developed to check for violations of the standard maintained hypothesis of homoscedasticity across alternatives and individuals in the GEV family. An empirical application of the heteroscedastic logit model can be found in Chapter 2 that demonstrates the practical value of the model. Future research can extend many interesting results on specification tests for the standard logit/nested logit models to the heteroscedastic ones, and can further explore the empirical validity of the various heteroscedastic models.

References

Amemiya, T. (1978). "On a Two-Step Estimation of a Multivariate Logit Model." *Journal of Econometrics*, 8(1): 13-21.

Barnard, P.O. and Hensher, D.A. (1989). "Joint Estimation Of a Polychotomous Discrete-Continuous Choice System: An Analysis of the Spatial Distribution of Retail Expenditures." Working Paper.

Ben-Akiva,M. and Lerman,S.(1985). *Discrete Choice Analysis: Theory and Application to Travel Demand.* The MIT Press.

Bösch-Supan,A.(1990). "On the Compatibility of NMNL Models With Utility Maximization." *Journal of Econometrics.* 43, 373-388. North-Holland.

Breusch, T.S. and Pagan, A.R. (1980). "The Lagrange Multiplier Test And Its Applications To Model Specification in Econometrics." *Review Of Economic Studies.* XLVII, 239-253.

Daganzo,C. and Kusnic, M. (1990) "Large-Scale Nested Logit Models: Theory and Experience." Draft.

Davidson, R. and MacKinnon, J. G. (1984). "Convenient Specification Tests For Logit And Probit Models." *Journal Of Econometrics.* 25. 241-262.

Davis, J. and Robinowitz, P. (1975). *Methods of Numerical Integration.* Academic Press.

Dubin, J.A. (1985). *Consumer Durable Choice and the Demand for Electricity.* North-Holland.

Dubin, J.A. and McFadden, D.(1984). " An Econometric Analysis of Residential Electric Appliance Holdings And Consumption." *Econometrica*. Vol. 52, No.2.

Dubin, J.A. and Zeng, L. (1991). "A Heterogeneous Logit Model." Working Paper No. 759, Caltech.

Enberg, J. et al. (1990)."A Random-Effect Logit Model Of Work-Welfare Transitions." *Journal Of Econometrics*. 43, 63-75.

Engle, R.(1984). "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics." Chap.13 of *Handbook of Econometrics*.

Greene, W.H. (1990). *Econometric Analysis*. Macmillan Publishing Company, 1990.

Horowitz, J.L. (1981). "Identification and Diagnosis Of Specification Errors in the Multinomial Logit Model." *Transportation Research B*, Vol. 15B, No.5, 345-360.

Horowitz, J. L. (1983). "Statistical Comparison Of Non-Nested Probabilistic Discrete Choice Models." *Transportation Science*. Vol.17, No.3. 319-350.

Horowitz, J. L. (1987). "Specification Tests For Nested Logit Models." *Environment and Planning A*. Vol. 19. 395-402

Housman, J.A. and Wise, D.A. (1978). "A Conditional Probit Model For Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica*. Vol.46, No.2, 403-426.

Housman, J. and McFadden, D. (1984). "Specification Tests For The Multinomial Logit Model." *Econometrica*. Vol. 52. No.5. 1219-1240

Housman, J.A. and Ruud, P.A. (1987). " Specifying And Testing Econometric

Models For Rank-Ordered Data." *Journal Of Econometrics*. 34, 83-104.

Johnson, N and Kotz, S.(1970). "Extreme Value Distributions." Chap.21 of *Continuous Univariate Distributions*.

Lee, H.L. and Cohen, M.A. (1985). "A Multinomial Logit Model For The Spatial Distribution Of Hospital Utilization." *Journal Of Business and Economic Statistics.*. Vol. 3, No.2.

McFadden, D.(1973). "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka(ed.) *Frontiers in Econometrics*. New York. Academic Press.

McFadden, D., Train, K. and Tye, W.B. (1976). " An Application of Diagnostic Tests For The Independence From Irrelevant Alternatives Property Of The Multinomial Logit Model." *Transportation Research Record*. 637. 39-46.

McFadden, D. (1978). "Modeling the Choice of Residential Location." In A.Kaelquist et al.(eds.) *Spatial Interaction Theory and Residential Location*, pp.75-76. Amsterdam: North-Holland.

McFadden, D. (1979). "Quantitative Methods for Analyzing Travel Behavior of Individuals." In D. Hensher and P. Stopher, eds., *Behavioral Travel Modeling*. Croom-Helm.

McFadden, D. (1981). "Econometric Models of Probabilistic Choice." in C. Manski and D. McFadden, eds.,*Structural Analysis of Discrete Data*, Cambridge, Mass. MIT Press.

McFadden, D. (1984). "Econometric Analysis of Qualitative Response Models."

Chap.24 of *Handbook of Econometrics*. North-Holland.

McFadden, D. (1987). "Regression-Based Specification Tests For The Multinomial Logit Model." *Journal Of Econometrics*. Vol.34. 63-82.

McFadden, D. (1989). "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." *Econometrica*. Vol.57, No.5. 995-1026.

Maddala, G.S.(1983). *Limited Dependent and Qualitative Variables in Econometrics* . Econometric Society Monographs. Cambridge Univ. Press.

Newey, W. K. (1985). "Maximum Likelihood Specification Testing and Conditional Moment Tests." *Econometrica*. Vol. 53, No.5. 1047-1070.

Pagan, A. and Vella, F. (1989). "Diagnostic Tests For Models Based On Individual Data : A Survey." *Journal Of Applied Econometrics*. Vol. 4. 529-559.

Palfrey, T.R. and Poole,K.T. 1987. "The Relationship Between Information, Ideology and Voting Behavior." *American Journal of Political Science,* Vol.31, No.3:511-530.

Quandt, R.(1984). "Computational Problems and Methods." Chap.12 of *Handbook of econometrics*. North-Holland.

Ruud, P. A. (1986). "Consistent Estimation of Limited Dependent Variable Models Despite Misspecification Of Distribution." *Journal Of Econometrics*. Vol.32. 157-187.

Small, K. and Hsiao, C. (1985). "Multinomial Logit Specification Tests." *International Economic Review*. Vol. 26, No. 3. 619-627.

Small, K.A. (1987). "A Discrete Choice Model For Ordered Alternatives." *Econometrica*, Vol. 55, No.2. 409-424.

Smith, T.E.(1984). "A Choice Probability Characterization of Generalized Extreme Value Models." *Applied Mathematics And Computation*. 14:35-62.

Steckel, J.H. and Vanhonacker, W.R. (1988). "A Heterogeneous Conditional Logit Model Of Choice." *Journal Of Business and Economic Statistics*. Vol. 6, No.3. 391-398.

Train, K. (1980). "A Structured Logit Model Of Auto Ownership and Mode Choice." *Review of Economic Studies*. XLVII, 357-370.

Tse, Y.K. (1987). "A Diagnostic Test For The Multinomial Logit Model." *Journal Of Business and Economic Statistics*. Vol. 5, No.2, 283-286.

Vuong, Q.H. (1983) "Misspecification and Conditional Maximum Likelihood Estimation." Social Science Working Paper 503. Caltech.

White, H. (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator And A Direct Test For Heteroskedasticity." *Econometrica*. Vol. 48, 817-838.

White, H. (1982). "Maximum Likelihood Estimation Of Misspecified Models." *Econometrica*. Vol. 50, No.1. 1-25.

Wills, H. (1987). "A Note On Specification Tests for The Multinomial Logit Model." *Journal Of Econometrics*. Vol. 34, 263-274.

Zeng, L. (1990). "A Heteroscedastic Logit Model." Paper Presented to the Social Science Seminar, Caltech. June 11, 1990.

Chapter 2

Rational Voter Models

2.1 Introduction

Voting behavior in mass elections is one of the most thoroughly studied subjects in political science because it has important consequences for the operation of democratic systems. An important class of voting models which assume voters to be "rational" in making their voting decisions are the so-called rational voter models. An advantage of these models is that they provide explicit, precise theoretical bases for voting decisions and their analysis, and lead to testable hypotheses about voting behavior. They have therefore received much attention in the literature and have become quite popular. Two major models proposed in this tradition are (1) the expected utility maximization model or the Downsian model (Downs 1957; Riker and Ordeshook 1968; McKelvey and Ordeshook 1972), which states that voting is (partly) an investment decision and that voters choose the actions which give them greater expected benefits involving the PB terms;[1] and (2) the "minmax regret" model (Ferejohn and Fiorina 1974) which posits that voters choose their actions so as to minimize their maximum regrets.[2] For the voting turnout decision, the first model implies, operationally, that the *interaction* of the closeness of the race and the utility difference terms matter in the voting decision (in addition to the fixed benefit and cost terms), while in the second model, only the utility difference terms matter. For the candidate choice decision, the Downsian model permits strategic voting, while according to the minmax regret model voters should always vote for the top preference. These models have invoked an intense debate among political scientists holding different views on the interpretation and modelling of rational behavior, and many attempts to test them empirically can be found in the literature (see, for example, Rosenthal and Sen 1973; Ferejohn and Fiorina 1975; Ashenfelter and Kelly 1975; Black 1978; Foster 1984; Hansen, Palfrey and Rosenthal 1987).

Despite all the research that has been done in this area, however, many important issues remain controversial, or ambiguous, or largely unexamined. Among them are some theoretical and methodological issues concerning the specification and comparative study of various rational voter models for the turnout decision, the formal testing of strategic voting behavior implied by the Downsian model for

[1] P is the probability that by casting the vote the voter can make a difference in the election outcome, and B is the utility gain (loss) the voter receives if he makes a difference.

[2] Here "regret" is defined as the voter's utility loss resulting from taking an action that may not be optimal given the "true" state of the world, which is unknown to the voter. The loss may be different under different "true state" of the world, and the voter is supposed to choose an action such that the maximum regret associated with this action is the minimum among all actions.

the candidate choice decision in the presence of a minor party candidate, and the methodological treatment of voter heterogeneity. These issues are explored in the current study.

Concerning the specification and comparative study of the turnout models, there are two problems in the previous research. First, although the Downsian model clearly specifies the expected utility maximization rule, it is often taken to mean a model in which "closeness matters" and just that. Hence, it is not unusual that, in testing the Downsian model, scholars focus only on the effect of closeness of the race, that is, the P terms, rather than on the interaction of the closeness of the race and the utility differences, that is, the PB terms (e.g., Ashenfelter and Kelly 1975; Black 1978). However, as we discuss in the next section, a model in which P but not PB matters is no longer an *expected* utility maximization model. Rather, it is a distinct model based on utility maximization which involves no risk.[3] To claim empirical support for the Downsian model based on the significance of the P terms only is therefore not justified. That the two models are different can be further shown by demonstrating empirically that the significance of the P term does not imply that of the PB term, and vice versa. Indeed, analysis of the 1968 and 1980 data will show that, for the turnout decision, it is the P terms, but not the PB terms, that significantly enter the calculus, while for the candidate choice decision, the opposite is true.

[3]We hereafter refer this model as the "closeness model" or the "P model" for convenience.

Second, previous researchers ignore data problems such as multicollinearity.[4] Although different rational voter models reflect different hypotheses about decisionmaking, the variables specified by these models may be closely correlated by construction. For instance, the PB term of the Downsian model is constructed as the product of P of the closeness model and B of the minmax regret model. Hence multicollinearity is likely to be present if the models are compared by including all variables in the same equation. When multicollinearity exists, desirable properties of estimators obtained by standard procedures are damaged and inferences based on such estimators are unreliable and misleading. It is nevertheless a common practice to test different theories by including all the possible variables in one general equation, and testing for the significance of coefficients (e.g., Ashenfelter and Kelly 1975; Ferejohn and Fiorina 1975).[5] Because of this, and because the problem of multicollinearity in the logit model is far less studied than in the linear regression model, we feel it is useful to discuss the problem and related techniques in some detail in this essay. We will show through formal tests that severe problems of multicollinearity do occur in, for example, Ferejohn and Fiorina's (1975) analysis, and therefore conclusions based on standard estimators are questionable. We then briefly discuss several alternative estimators such as Stein, ridge, and principle component estimators for the logit model. As these estimators are biased and their

[4]Foster (1984) is an exception, discussing the problem in the context of the linear model.

[5]One reason for doing so, according to Ferejohn and Fiorina (1975), is that to compare models involving different variables by estimating them separately would require non-nested hypothesis testing techniques, which are not as well-developed and as familiar to many as nested hypothesis testing. We will turn to this problem later.

statistical properties are ambiguous, we do not base our inferences about model comparison on them. Rather, we estimate different models separately and apply non-nested hypothesis testing to compare their performance. The results show that neither the Downsian nor the minmax regret model predict voting turnout well, while the closeness model is a slight improvement over the two standard ones.

In studying the performance of rational voter models for the candidate choice decision, we show that unlike turnout, expected utility terms as specified in the Downsian model do significantly affect vote choice. We believe this difference between the turnout and the vote calculation reflects the distinct nature of the two decisions. The results also provide clear evidence of strategic voting in the presence of a third candidate. Strategic voting in three-candidate races that avoids the "wasted vote" and causes a third party "squeeze" has been discussed theoretically (Duverger 1967; Downs 1957; Farquharson 1969; Myerson and Weber 1988; Palfrey 1989) and tested both empirically (Black 1978; Cain 1978; Laver 1987) and in laboratory experiments (Plott 1991). In the U.S., two presidential elections in recent history involve a prominent third candidate—the 1968 and the 1980 presidential elections. Although some evidence of strategic voting behavior in these elections exists (Kiewiet 1979; Abramson *et al.* 1982), no formal estimate of the relationship between candidate viability and strategically rational voting is offered. Our results show that strategic voting does occur as theory predicts. Furthermore, we show that strategic considerations affect voters favoring the minor party candidate much more significantly than major party voters. For the latter, the vote decision is determined largely by such variables as partisanship and candidate evaluations.

We lastly address the issue of voter heterogeneity. Heterogeneity means, roughly, differences. Voters differ from one another in various aspects. Some of these differences are observable and measurable, as captured by the values of various explanatory variables. However, there may exist other types of differences beyond this. One example is that the error terms of the utility functions, which include the summary effects of all excluded relevant variables, may have different variances (heteroscedasticity) because their effects on voting behavior may be different in nature and magnitude across voters. Another example is that the measured attributes included in the model may affect different voters in different ways, leading to varying coefficients across voters.[6] When heterogeneity exists, standard models such as simple logit are not strictly appropriate. In this study we concentrate on heteroscedasticity. Specifically, we are interested in testing the notion of "effective rationality" (Goldberger 1969; Black 1978) which embodies the idea that even though all individuals may try to act rationally, they may not be able to avoid errors in their calculations and they may differ in the accuracy of their calculations. One source of this variation is voter information (Black 1978; Palfrey

[6]This type of heterogeneity is discussed in more detail in Rivers (1988), where a varying coefficients logit model using rank-ordered data is proposed as a solution. A crucial limitation of this model is that the number of unknown parameters increases with sample size N, and hence the maximum likelihood estimation is not consistent. The random coefficient models, assuming coefficients are drawn from a certain distribution, can avoid this "infinite parameter" problem. See, for example, Manski and McFadden 1981.

and Poole 1987). Operationally this translates to heteroscedasticity and renders the widely used standard logit model inappropriate. Therefore the application of a heteroscedastic logit model developed in Chapter 1 is explored as an alternative. We demonstrate that the heteroscedastic logit model can be easily applied, and its estimation can often be carried out using the standard logit procedure. The new model is shown to significantly outperform the standard logit. Findings from the data support the notion of effective rationality, and reveal strong effects of information on the turnout decision.

In what follows, Section 2 discusses the specification of rational turnout models, and presents empirical findings on the comparative merits of these models; Section 3 focuses on strategic voting behavior; Section 4 concerns the methodological treatment of voter heterogeneity; and Section 5 offers concluding remarks. Technical details on some methodological topics are contained in the appendix. Note that the 1980 election data are validated, but the 1968 data are not. Hence more weight should be placed on results from the 1980 data.

2.2 Rational Voter Models: Comparative Studies

Two major rational choice models have been proposed to explain voting turnout behavior: the expected utility maximization model and the minmax regret model. The original version of the expected utility model (Downs 1957) views voting as a pure investment decision: a voter compares the costs of voting in the present with the *expected* utility gain in the future and votes if the expected benefit exceeds costs. The net return of voting can be expressed as:

$$R = PB - C, \tag{2.1}$$

where P is the probability that by voting the voter can make a (favorable) difference on the election outcome. This probability is positively related to the closeness of the race between the candidates. B is the benefit the voter receives from this difference, and C is the cost of voting. Assuming that $C > 0$, this model fails to provide a satisfactory explanation of turnout in mass elections because the probability that a single vote can make a difference on the election outcome is practically zero and hence the PB term is infinitesimal. So it is irrational for voters with any significantly positive voting cost to turn out (Ledyard 1984; Palfrey and Rosenthal 1985).

Bothered by this feature of the Downsian model and observing that Downs does not include "non-political" benefits of voting appeared important in empirical investigations, Riker and Ordeshook (1968) introduced a D term into the calculus, which registers sources of positive satisfaction with voting such as the compliance with the ethic of voting ("citizen duty") . Because the benefits in D can be substantial, it may therefore be rational for people to vote. Equation (2.1) thus becomes:

$$R = PB + D - C \tag{2.2}$$

which is generalized for multicandidate elections by McKelvey and Ordeshook (1972) to

$$E^k - E^0 = B_{k1}P_{k1} + B_{k2}P_{k2} + \ldots + D - C \tag{2.3}$$

assuming only two-way ties are likely. In (2.3), $E^k - E^0$ is the expected utility difference between voting for candidate k and abstaining, and the BP terms on the right hand side represent pairwise comparisons between candidate k and the other candidates.

Assuming that the probability terms are unknown or unknowable, Ferejohn and Fiorina (1974) consider an alternative decision rule, namely the minmax regret decision rule, which specifies that the voter should choose the act that minimizes his maximum regret. They show that in this model only the B terms enter the calculation of the turnout decision.

For the expected utility maximization model, the empirical significance of the D term seems well established (Riker and Ordeshook 1968; Katosh and Traugott 1982; and Ashenfelter and Kelly 1975). However, the importance of the PB term is questionable. We are not aware of any strong empirical support for the significance of the PB term in turnout decisions in mass elections, although there are a number of studies that find that "closeness matters," *i.e.*, closeness alone as an independent variable matters (*c.f.*, Cohen and Uhlaner 1991; Foster 1984; Palfrey and Rosenthal 1987; Rosenthal and Sen 1973). But does the fact that "closeness matters" establish empirical support for the Downsian model? In other words, is "closeness matters" equivalent to "expected utility terms matter"? If not, what do these empirical findings suggest?

Closeness Matters. So? To answer the above questions, let us recall that "expected utility" refers to the utility derived from possible future benefits (and/or costs), or more accurately, utility over *lotteries* (*c.f.*, Varian 1984). Roughly speaking, to maximize expected utility means to choose the action that generates the greatest future net benefits *discounted* by the probabilities that these benefits will be realized. Note that the main body of the object of maximization is benefits and costs, and that probabilities of the states of the world only serve as *discount factors*, which generate no utility on their own. Voters get no direct utility from breaking ties, but only from the benefits they receive when their candidate wins and poses programs they favor. Hence, to show that voters are maximizing *expected utility*, we need to show that the *discounted* future *benefits* (*i.e.*, PB) matter, not that the discount factor (P) itself matters. P does not constitute an expected utility term, and if it is significant in its own sake, it must be that it is playing a different role than a probability discount factor, and bears its own substantive meanings. Therefore, if PB does not enter the calculus significantly, then the voter is no longer maximizing *expected* utility, even if "closeness matters."

Note that if P enters the model as an independent variable that bears its own substantive meaning (we discuss the meaning later), but PB does not, then the decision no longer involves any *risk;* and all benefits and costs come with certainty, without any discounts. In this situation, a voter's decision involves no *investment* elements, in the sense that we cannot infer that he or she is looking to *future outcomes*. And therefore, he is an *utility* maximizer, not an *expected utility* maximizer. That "P matters" and "PB matters" are different things can also be justified empirically by showing that the significance of P does not imply the significance of PB, and vice versa. Such findings are indeed reported in Rosenthal and Sen (1973) in studying the "heuristic model" of the Downsian specification. Our own empirical results shall provide another piece of evidence.

So previous research which concludes that "closeness matters" but PB does not actually provides empirical support not for the Downsian model, but for another distinct rational voter model, a pure consumption model based on utility maximization. Let us formulate this model as:[7]

$$R = f(P) + D - C \tag{2.4}$$

or, for multi-candidate competitions,

$$R = f(P_{ij}, \forall i \neq j) + D - C \tag{2.5}$$

where f is some (possibly increasing) function which is specified as linear in our

[7] We can also write the model as $R = D_1 - C$, because P generates *consumption value*, just as D does.

empirical study for simplicity.[8]

As for the substantive meaning of P in (2.4) or (2.5), Rosenthal and Sen (1973) hypothesize that the level of the competition is positively correlated with a voter's interest in a campaign, which in turn stimulates turnout. We conjecture that closeness of the race is also positively related to a voter's feeling of efficacy when voting. Note that compared with its role in expected utility considerations, the closeness of the race need not be very high to stimulate interest or to introduce efficacy. In Ferejohn and Fiorina's language, "for expected utility maximization the probability...is very important," whereas as a stimulating source, "the mere logical possibility of such an event (that an individual has any impact) is enough." (Ferejohn and Fiorina 1975).

Now we have three different voting turnout models, and one way to compare them empirically is to include the variables specified by each model into one general equation, and see which of them are significant, and which are not. This is the method employed by several researchers (*e.g.*, Ashenfelter and Kelly 1975; Ferejohn and Fiorina 1975). In doing so, however, we encounter a serious problem with multicollinearity. Hence a discussion of the problem of multicollinearity is in order.

---

[8]Note that theoretically the Downsian model and the P model do not exclude each other, for closeness may both generate utility itself *and* serve as a proxy of probability factor. Hence we may include the P terms into the Downsian framework and write:

$$E^k - E^0 = B_{k1}P_{k1} + B_{k2}P_{k2} + \ldots + f(P_{ij}, \forall i \neq j) + D - C \tag{2.6}$$

where $f(P_{ij}, \forall i \neq j)$ can be regarded much as another D term.

Data Problem: Multicollinearity Multicollinearity arises when two or more independent variables or combinations of variables are highly but not perfectly correlated. For the linear regression model, the effects, diagnosis, and remedy of this problem are thoroughly studied (see, for example, Judge *et al.* 1985). Although little work has been done for non-linear models in general, recent advances in the statistics literature do offer information about the binary logit model (*e.g.*, Marx and Smith 1990; Schaefer 1986; Schaefer, Roi and Wolfe 1984), which is the most widely used model in voting studies. The effect of multicollinearity in the logit model is found to be similar to that in the linear regression. Namely, it causes extreme sensitivity of parameters to small perturbations in the explanatory variables, extremely large variances of coefficients and predictions, and bad properties of statistical testing on parameters. Because of this, diagnosis of the problem where it is likely to be present is necessary, and action should be taken accordingly.

Diagnosis Some useful diagnostic tests are offered by Marx and Smith (1990), which should be consulted for technical details. Here we briefly discuss the ideas behind the tests. Note that the covariance matrix of the coefficients estimated by the maximum likelihood procedure is approximated by the inverse of the (estimated) information matrix, hence the properties of the information matrix will directly affect that of the estimators. If it is ill-conditioned, *i.e.*, if it is near singular, or its square root has nearly linearly dependent columns,[9] then the undesirable proper-

[9]The square root of a matrix A is a matrix D such that $D'D = A$.

ties discussed above would be present. Denote the estimated information matrix by $\widehat{F}$, and its square root by S. Denote S^* the scaled S such that its columns have unit length, and denote $F^* = S^{*\prime}S^*$. Then, diagnosis of ill-conditioning of $\widehat{F}$ can be based on the following:

1. The correlations between columns of S: high correlations indicate ill-conditioning of $\widehat{F}$. The correlation matrix is given by F^*.

2. The variance inflation measures: diagonal elements of the inverse of F^* give a measure of inflation of the asymptotic variances of the estimated parameters. These measures are 1 when the columns of S are perfectly orthogonal (ideal condition).

3. The condition numbers of F^*: defined as the square root of the ratio of the largest eigenvalue of F^* over the others, indicate ill-conditioning of F if any of them is "large."[10]

4. Variance proportion: each eigenvalue of F^* contributes to the variance of the estimated coefficients. If the proportion of the variance of a coefficient attributed to a small eigenvalue (associated with a large condition number) is big (say, exceeds 50 percent), that coefficient is likely affected by multi-collinearity, and its value is unreliable.

[10]There is no strict criteria as how large is "large." Johnston (1984) sets the critical value to 20, Marx and Smith to 30. Our own experience suggests that caution should be applied when any conditional number exceeds, say, 10.

To show that the problem of multicollinearity may exist when the rational voting turnout models are compared by including all variables in the same equation, we apply the above tests to the following model, in which the PB, B, and P terms and some D and C terms are all included. Specifically, the model is:

$$\begin{aligned} logit(p[vote]) &= \{Const., PB_{12}, PB_{13}, PB_{23}, P_{12}, P_{13}, P_{23}, B_{12}, B_{13}, \\ &\quad U_1, Info., Educ., Party\} * \beta. \end{aligned} \tag{2.7}$$

In the above model,[11] $logit$ specifies the logit operation (a logit model is used given the discrete nature of the dependent variable), $p[vote]$ is the probability of voting turnout, $Const.$ is a constant term, P_{ij} and B_{ij} denote the closeness of the race and utility difference between the respondents' ith and the jth ranked preference[12] respectively, $PB_{ij} = P_{ij} * B_{ij}$, U_1 is the utility for the first preference,[13] *Info.* is an index of voter information (Palfrey and Poole 1987), *Educ.* education level, and *Party* party strength. The last four variables are included as proxies of the D or C term.[14] Lacking a cardinal subjective probability index in our data set, the actual

[11]Note that the multi-candidate models are used here because we have three candidates for both 1968 and 1980 elections. As the three B terms are perfectly collinear ($B_{13} = B_{12} + B_{23}$), only two of them are included. It does not matter which two are included. The third provides no information.

[12]Preference orderings are obtained from the thermometer scores, with ties broken by party scores.

[13]U_1 is included to test the alienation hypothesis (Riker and Ordeshook 1968). The prediction is that the higher U_1, the more likely individuals will vote, as they feel more positive about the alternatives they face.

[14]Note that information may reflect voting costs. Other C category variables, like registration laws, are not included because of lack of data.

vote return in the respondent's state is used to construct the closeness terms.[15] Following Myerson and Weber (1988), we operationalize the probability of a tie between the j^{th} and the k^{th} preference as: $P_{jk} = v_j * v_k$, where v_j denotes the vote percentage received by the voter's j^{th} preference which is positively related to the marginal probability that his j^{th} preference is in first place.

The first sign of multicollinearity in this model is a dramatic change in the estimated parameters if some terms are dropped from the full model. For example, in estimating the 1968 sample, the full model estimates show that the PB terms are significant. However, if the P and B terms are dropped, then PB is no longer so. Formal test results of the problem are reported in Table 1.1, under "model 1." We see that for both the 1968 and the 1980 data, severe multicollinearity exists, with the estimates of the coefficients for P terms and for some of the PB and B terms most likely to be affected. The estimates of the P terms are especially unreliable, given that more than 90% of their variances are attributed to the smallest eigenvalue (associated with the least important eigenvector), and that these variances are highly inflated. For both years, the maximum correlation coefficients are well above 0.5, and the conditional numbers are close to 20. Hence, it is very misleading if the data problem is ignored and comparison of the models is solely based on the

---

[15]The use of the state level data is based on two considerations. First, it is more legitimate under the electoral college system (see, *e.g.*, Kiewiet 1979). We tried the national data and found them insignificant, which we interpret as evidence of the impact of the electoral college system. Second, as reported by Field (1981), "the amount of 1980 polling conducted at the state level was more extensive than ever before in a presidential election," and closely matched the actual vote return patterns. This means, of course, that the actual return data is close to the state poll data and, therefore, that actual returns can be used to construct the subjective probability terms.

maximum likelihood estimates of model (2.7).

Next we are curious to see whether the problem would exist for a simple model such as the one specified in Ferejohn and Fiorina (1975), where only the constant, one PB and one B term are included (here we include PB_{12} and B_{12}). The test results are shown in Table 2.1 under "model 2." The data are striking, especially for the year 1980. The condition number of the information matrix is as high as 83; the variances of the estimated coefficients are extremely inflated and are almost solely attributed to the smallest eigenvalue. Hence, the estimation of the model does not provide *any* information, and it is wholly inappropriate to draw any conclusion from such estimates.

It is clear, then, that we cannot base the comparison of the rational voter models on the MLE of model (2.7). We have two alternatives. First, we can derive alternative estimators for model (2.7). Second, we can estimate the various models separately, and perform non-nested hypotheses testing.

Alternative estimators Some alternative estimators, namely ridge, principle components and Stein estimators are proposed when multicollinearity is present (*e.g.*, Schaefer 1986; Schaefer, Roi and Wolfe 1984).[16] These estimators are aimed at improving the precision of the MLE estimates by adjusting the MLE in certain ways. Denote by β_{mle}, β_{pc}, β_s, and β_r the MLE, principle component, Stein, and

[16]These estimators are similar to their counterparts in the linear regression model, only that the focus of the operations is the information matrix rather than the matrix $X'X$ (assuming X is the matrix of independent variables).

ridge estimators, respectively, and recall that $\widehat{F}$ is the estimated information matrix. Then:

$$\begin{aligned}\beta_{pc} &= \widehat{F}^{+}\widehat{F}\beta_{mle}\\ \beta_{s} &= c\beta_{mle}\\ \beta_{r} &= (\widehat{F}+kI)^{-1}\widehat{F}\beta_{mle},\end{aligned}$$

where $\widehat{F}^{+}$ is the generalized inverse of $\widehat{F}$, i.e., the inverse of $\widehat{F}$ with its "minor" component deleted (its small eigenvalues zeroed); c the Stein shrinkage parameter, chosen to minimize the mean square error; k is the ridge parameter,[17] and I the identity matrix. The idea behind the principle component estimator is that when multicollinearity exists, some of the components of $\widehat{F}$ (those associated with small eigenvalues) supply little information but contribute much to the variance of the estimates, hence deleting them should reduce the variance of parameter estimates without losing much information. The Stein estimator is obtained by shrinking both the MLE parameters and their variances by the same factor, which is chosen to minimize the mean squared error of the Stein estimator. The ridge estimator works by altering the data to counteract near singularity (*c.f.*, Johnston 1984), while in a Bayesian interpretation, it implicitly incorporates information from outside the sample (Chow 1983).

[17]The choice of k is subjective. Schaefer suggests the use of $k = K/(\beta'_{mle}\beta_{mle})$, where K is number of parameters in the model. We use this formula here.

Coefficient estimates from the three alternative estimators, together with the original maximum likelihood estimates, are presented in Table 2.2-1 (1968) and Table 2.2-2 (1980). Also reported are the standard errors of the estimates and the t-statistics.[18] For the Stein estimator, only the shrinkage parameters are reported, because the coefficients and their standard errors are simply the MLE weighted by this parameter, and the "t-statistic" is the same as that of the MLE.

Note that all these alternative estimators shrink the variance of MLE, and hence are more "precise." The differences between the MLE and the ridge and principle components estimators mainly show up in the PB, P or B terms that are most affected by multicollinearity. This proves that these alternative estimators at least hit the problem. In this regard, the Stein estimator is the least satisfactory. Also, the ridge and principle components estimators in general have smaller variances than the Stein estimator. Hence, we do not recommend the Stein estimator.[19] In terms of fitting (log likelihood), the PC estimators are the worst, intuitively because they work as if by deleting variables, and hence is using less information from the data.[20] In terms of mean squared error, however, all four estimators are about the same: for the 1968 data, all of them generate the same MSE of 0.29; while for the 1980 data, the MLE, Stein and ridge have a MSE of 0.21, and the PC estimator, 0.22.

Although the alternative estimators all have a smaller variance than the stan-

[18]The t statistics for the alternative estimators are only heuristic and serve as a rough reference, as the statistical distribution of these estimators are actually unknown.

[19]Marx and Smith (1990) give similar recommendations.

[20]The MLE, of course, gives the best log likelihood because it is obtained by maximizing the likelihood.

dard logit estimators, they are nevertheless biased and with complicated, ambiguous sampling properties. The interpretation of the estimates can be difficult and statistical inferences about the true parameters cannot be made (Johnston 1984). Whether they "work well" in a specific application can only be judged by whether they intuitively "look" reasonable, *e.g.*, shrink the MLE "enough" (see Marx and Smith 1990 for another example). They are indeed helpful when, for some theoretical reason, none of the variables can be dropped from the equation. In our case, however, our primary goal is to compare *different* models. Hence instead of relying on unreliable estimates from the full model, we can estimate the models separately and base the comparison on rigorous non-nested hypothesis testing techniques.

Comparative studies: empirical results. We therefore estimate separately four models: the Downsian model, obtained by dropping the P and B terms from (2.7); the closeness model, obtained by dropping the PB and B terms from (2.7); the minmax regret model, obtained by dropping the PB and P terms from (2.7); and, a virtually "social-psychological" model formed by dropping all the PB, P and B terms from (2.7). The results are presented in Table 2.3-1 (1968) and Table 2.3-2 (1980). The Downsian model, the closeness model and the minmax regret model are not nested (neither is a parametric special case of the other), therefore classical tests such as the likelihood ratio test are not appropriate.[21] Instead, non-nested hypothesis testing should be performed. In the appendix we discuss the testing of

[21]The likelihood ratio test is applied in comparing the fourth model with others.

non-nested hypotheses and produce a convenient critical value table for one such test. To illustrate the use of the table, take the comparison between the P model and the Downsian model for the 1968 data. The two models have the same number of coefficients, but the absolute value of the log likelihood of the P model is 4.06 lower than that of the Downsian model ($l_1 - l_2 = 4.06$) which exceeds the critical value of 3.32 for the P model to be the better model with probability 99%. Test results show that the P model outperforms all other models in both years with a probability of at least 90%, whereas the Downsian model is a "Condorcet loser" in both years. The minmax regret model is slightly better than the social-psychological model (model 4) only for the 1968 data.[22]

In both years, *Info.*, *Educ.* and *Party* are highly significant. This confirms findings in previous research that social-psychological factors are important predictors of voting behavior. U_1 (*alienation*) and B_{13} (*indifference*) matters in the 1968 data, but not in 1980. The PB terms are not significant in either year, whereas the P terms are — thereby indicating that the Downsian model and the P model are substantially, as well as substantively different. In turnout, it is the closeness of the race itself that affects the decision, and not expected utility considerations. Intuitively, this finding is consistent with the argument that in a mass election the probability that any single vote can make any difference to the outcome is very small and hence

[22]Note that the testing technique described in the appendix is designed for pairwise comparisons. When there are more than two models to be compared, conflicts in pairwise comparisons may occur (e.g., model 1 is preferred to model 2, model 3 is preferred to model 1, but model 3 is rejected when compared with model 2). However, it is not the case here.

so is PB; however, the closeness of the race, as a stimulating source that generates some psychological value of voting, need not be that high to have an impact on voting behavior.

Note that the values of the P terms are not reliable because tests show that they are seriously affected by multicollinearity in both years.[23] However, the precision of the estimation of each P term is not an important issue here because we are mainly interested in whether the P terms matter as a whole group, and statistical testings on the significance of the whole group is not affected by the precision of each parameter within the group. Therefore, the comparison among different models is valid and we conclude that in predicting turnout, the P model outperforms the other models, although precise estimates of the parameters are unknown.

To have some intuitive sense of how much the P model outperforms the rest, Figure 2.1-1 (1968) and Figure 2.1-2 (1980) demonstrate graphically the fit of the models. A model is better if for voters it gives a higher predicted probability of voting, and for non-voters it gives a lower predicted probability. The figures do show that the P model performs better than the others. The pattern is especially clear for the 1980 data. However, the differences among the four models are not dramatic, and it seems that we do not lose much by using the simple social-psychological model.

[23]We performed the diagnostic tests for all models and the results suggest that the problem is present in the P model, with the P terms being affected. Intuitively this is because the three P terms are correlated. For example, if P_{12} is high, then 1 and 2 are likely the major candidates and 3 the minor candidate; hence, P_{13} and P_{23} are likely to be low.

2.3 Candidate Viability and Strategic Voting

In the last Section we compare the performance of several models in accounting for voting turnout behavior. Now we move on to the second stage of decision making, the decision to cast a specific vote given that a person goes to the polls. As mentioned earlier, the minmax regret model would predict that voters never vote for their second choice (see Ferejohn and Fiorina 1974 for proof). The "closeness model" provides no prediction for the candidate choice decision: it only states that closeness of the race would affect the turnout decision much the same way as the D term, which generates the same consumption value of *voting itself* regardless of who you vote for. The Downsian model, however, does provide some theory for the candidate choice decision. Consider the vote choice between first and second preferences.[24] From equation (2.3) or (2.6), the utility difference between voting for a first versus a second preference is:[25]

$$E^1 - E^2 = 2P_{12}B_{12} + P_{13}B_{13} - P_{23}B_{23}, \tag{2.8}$$

so a rational voter will vote for candidate 1 if $E^1 - E^2 > 0$. Equation (2.8) indicates that the vote decision is affected by such strategic considerations like candidate viability, as reflected in the P terms. For example, if the first-ranked candidate is not viable— if P_{12} and P_{13} are small— then the voter is more likely to vote for his

[24]Because third preference voting is *a priori* irrational as it is a dominated strategy, we only consider vote choice between the first and the second preferences.

[25]See Black (1978, pp.612-613) for the algebra.

second choice, especially when the race between the second and third preferences is close and the utility difference between them is substantial (so that $P_{23}B_{23}$ is high). Similar formulae can also be derived by directly imposing the above rationale as the basis of the decision rule in the vote choice decision (see, for example, Palfrey 1989).

So, testing the Downsian model in the candidate choice decision when more than two candidates are present is in fact a test of the notion of "strategic voting," which is itself an important topic because of its bearing on the stability of the two-party system. The failure of the two-party system to respond fully to the imperatives of electoral competition from time to time has been evident, and periods of third party strength are indicators of this failure (*c.f.*, Rosenstone *et al.* 1984). However, the two-party system persists and a major party candidate has won nearly every presidential election. In addition to the constraints on third party candidates arising from historical and institutional factors and from voters' loyalty to the two major parties, an important source of third party candidates' failure in presidential elections is strategic voting behavior based on the "wasted vote" logic. Abramson *et al.* (1982) write, "for the candidate with little chance of victory, the wasted vote logic becomes a self-fulfilling prophecy. Because people perceive the candidate unlikely to win, they don't vote for him. Consequently, the candidate cannot win"(p.176). A quick look at the 1980 data reveals convincing evidence of this effect: among voters who ranked Reagan first in their thermometer scores, 98%

voted for their first preference, and of those who most preferred Carter, 85%; but of those who most preferred Anderson, only 41% voted for him!

Two attempts have been made to formally test for strategic behavior by Cain (1978) and Black (1978), using data from Britain and Canada, respectively. However, although expression (2.8) forms the theoretical basis of both of these studies, neither of them actually tested the influence of the PB terms in ways that are described by this model. Black uses the P terms alone (because, he claims, they are the terms of "prime interest"), whereas Cain drops some relevant terms without any explanation.[26] Thus, the theoretical interpretations of their empirical findings are blurred. In the current analysis, we use the exact variables that are predicted to enter the utility calculation by the theory. Namely, we include $P_{12}B_{12}$, $P_{13}B_{13}$ and $P_{23}B_{23}$ as the independent variables.

Table 2.4-1 reports the estimation results for the Downsian model. The signs of the PB terms are exactly as predicted by theory, and the parameters are all significant above the 0.05 level. The predictive power of the theory is indeed impressive. Recall, however, that in Section 2, we found the expected utility terms to be insignificant in predicting voting turnout. The same indices are used here, and the expected utility terms are highly significant. This reflects the distinctive nature of the two decision making situations. When making the turnout decision, a person compares the benefits of voting with the *costs* of voting. Given that the costs

---

[26]The terms Cain drops are the U_3 terms, or the utilities for the third candidate. See Cain (1978), p.647.

of voting may not be negligible — voting at least costs the voter *time* with which he could engage in some possibly more enjoyable activity — and that the voter will have to pay the costs with certainty, any factor to be seriously considered as "benefits" will have to be of certain possibility and magnitude. Hence the expected utility terms cannot significantly affect the voter's decision because *compared with the costs,* they are immaterial. Rather, it is the social and psychological factors, the "consumption" value of voting, that outweighs the costs.

However, when making the vote choice decision, the calculation of the voter is based solely on the comparison of the net benefits of *voting for alternative candidates,* with both the D terms (including the P terms) and the C terms in the turnout calculation being left out of the consideration because whoever he votes he pays the same costs and receives the same consumption value. So the question becomes: to what extent do the expected utility terms constitute the net benefits of voting for a specific candidate? Their role is thus much more significant here than in the turnout equation.

Our last question, to what extent do the expected utility terms constitute the net benefits of voting for a specific candidate, immediately prompt another consideration: are the expected utility terms, as operationalized, the *only* factors that enter the net benefits of voting for a candidate? The vote decision model (2.8) is a pure "investment" model, in which such important variables as partisanship play no role. Partisanship may enter the model by affecting the candidate utility indices. However, research on the American electoral behavior has clearly concluded that partisanship is a more stable factor than candidate evaluations and has an independent effect on the vote decision (*cf.* Niemi and Weisberg 1984). But then *how* do partisanship and other factors whose importance in the vote decision is well established, for instance, issue position and candidate characteristics, enter the calculus of vote decision in the expected utility maximization framework?

Reflections on the development of the rational turnout models suggest a straightforward solution: one more term should be included in equation (2.3) or (2.7), a term that represents the *consumption value of voting for the specific candidate,* in addition to the D term, which is concerned with the consumption value of *turnout.* Let us denote this term by D_k. Feelings towards the candidates, party ID, distance from the issue position of the candidate are examples of factors that may enter D_k. For example, the more strongly the voter identifies with candidate k's party, the more utility the voter will receive from voting for k because of the effects of party loyalty. We can therefore rewrite equation (2.7) as:

$$E^k - E^0 = B_{k1}P_{k1} + B_{k2}P_{k2} + \ldots + f(P_{ij}, \forall i \neq j) + D - C + D_k \qquad (2.9)$$

and simple algebra leads us to the counterpart of model (2.8):

$$E^1 - E^2 = 2P_{12}B_{12} + P_{13}B_{13} - P_{23}B_{23} + D_1 - D_2. \qquad (2.10)$$

Estimation results using equation (2.10) are reported in Table 2.4-2. Included in $D_1 - D_2$ are the thermometer scores for the first and the second preference (U_1 and U_2), and the difference of party ID's relative to the first and the second preference ($Party_1 - Party_2 = Party$).[27] The prediction is that U_1 and $Party$ should positively affect first preference voting, and U_2, negatively affect it. All coefficients have the predicted signs, and the Party variable and U_1 are significant. Tests of multicollinearity show that the precision of U_1 and U_2 in 1968 and U_2 in 1980 are affected, but all other parameters are reliable. In comparison with Table 2.4-1, we see that the model significantly increases the log likelihood for both years, and likelihood ratio tests[28] show that the modified model outperforms the Downsian model.

Figure 2.2-1 and Figure 2.2-2 show, using data from the 1968 and 1980 election respectively, the graphical comparison of three vote decision models: the Downsian model (2.8), the modified model (2.10) and a model containing only the D terms. A model is better if for first preference voters it gives a higher predicted probability of voting for the first preference, and for second preference voters, a lower prediction. From the figures we see that the modified model performs much better than the

[27]The variables $Party_1$ and $Party_2$ are constructed from the 7 point party ID variable. Suppose a voter's first preference is Reagan. Then $Party_1 = 4$ if the voter's party ID is 6 (strong Republican); $Party_1 = 3$ if his party ID is 5 (weak Republican); $Party_1 = 2$ if party ID is 4 or 3 (independent leaning Republican and Independent) and $Party_1 = 1$ for all other party ID scores. The construction is similar if his first preference is Carter. If his first preference is Anderson, then $Party_1 = 4$ if party ID is 3 (Independent); 3 if party ID is 2 or 4 (Partisan independent); 2 if party ID is 1 or 5 (weak partisan) and 1 if party ID is 0 or 6 (strong partisan). $Party_2$ is obtained in a similar fashion. This construction measures both the direction and intensity of party ID relative to each preference.

[28]Note that the Downsian model is nested in the modified model.

other two for second preference voters, while for first preference voters the models are not dramatically different. The comparison between the modified model and the model without the PB terms suggests the importance of the PB terms in vote decision. In contrast to turnout, we see that the difference is not trivial. This again reflects the different nature of the two decisions.

Model (2.10) actually tells us more. Note that if the first preference is a third party candidate, then $P_{12}B_{12}$ and $P_{13}B_{13}$ are both practically zero, as the probability of first place ties between a third party candidate and a major party candidate is basically nil. Therefore, voters who most prefer the third party candidate will vote for their first preference if:

$$E^1 - E^2 = -P_{23}B_{23} + D_1 - D_2 > O$$

which implies that if the race between their second and last choice is close and they like the second choice a lot more than the last, then they are more likely to vote for their second choice than for their first choice.

For voters preferring a major party candidate, P_{23} would be practically zero and they should vote for the first choice if:

$$E^1 - E^2 = P_{12}B_{12} + P_{13}B_{13} + D_1 - D_2 > 0.$$

Table 2.4-3 presents the comparison. For the 1980 data, all voters favoring Anderson are considered favoring a third party candidate, and for the 1968 data, those favoring Wallace in states where Wallace finished third are classified as such voters. The estimates reveal substantial differences between voters with preference for major and minor party candidates. For voters favoring a major party candidate ("others"), the only variable that significantly affects their choice decision is party identification, and the PB terms obviously do not play an important role. For voters favoring a third candidate, however, $P_{23}B_{23}$ clearly enters their calculations. This means that in states where the race between a voter's second preference and her third preference is close, and if the voter prefers her second choice much more than she does her last, she would rationally vote for her second choice. This is precisely the evidence of strategic voting that causes the third party "squeeze." Figure 2.3-1 (1968) and Figure 2.3-2 (1980) further show that for voters favoring a third party candidate, the strategic voting model performs better for second preference voters; whereas, for other voters, the model predicts first preference voters better.[29]

2.4 Information and Voter Heterogeneity

In this section we turn to the topic of voter heterogeneity. Black (1978) notes that one major reason for the variation among voters is different levels of voter information. Better informed citizens, he asserts, are more effectively rational and their voting behavior more predictable. Both Black (1978) and Palfrey and Poole

[29]This is partly due to the fact that there are more observations of second preference vote for voters favoring a third candidate than for the "other" voters.

(1987) find evidence for this notion in candidate choice behavior, conditional on voting. However, we know of no previous study that tests this notion at the voting turnout stage, nor any that discusses the methodological implications of this phenomenon and proposes a solution. We explore these issues in the current study, with a concentration on the methodological treatment. All results in this section are based on the 1980 data. We did not find clear evidence from the 1968 data. This is because the objective information index as developed in Palfrey and Poole (1987) is not as informative for the 1968 data as for the 1980 data. The 1980 election study provides data on the candidates' locations (reported by voters) on the seven-point scales of *nine* issues, while the 1968 study only has data on *two* issues. Also, the 1968 data are not validated.

Define "better informed voters" as those whose information scores exceed the average, and "less informed voters" as the rest of the the electorate. Table 2.5-1 reports the estimation results of the turnout models for the two groups. The comparison reveals substantial differences between the two groups. First, the probability indices are significant only for the better informed group. The reason is simple: it is the *subjective* estimates of the probabilities that effectively enter the utility calculation, and because the better informed voters' estimates are more accurate, *i.e.*, closer to the objective data, the objectively constructed indices are closer to the subjective estimates of the better informed and hence affect their decision more significantly. Another difference between the two groups is that the better informed are more predictable. The model for this group has a much better fit and results in a much higher percentage of correct predictions.[30] This can be best seen from Figure 2.4. The model distinguishes voters from non-voters much better for the better informed group than for the less informed group. Our results thus provide clear evidence of the effective rationality notion in turnout decisionmaking.

We also conducted the test on the candidate choice data and the evidence is not strong. We believe that this is because on average voters who participate are better informed than those who do not, and the variations among voters in the turnout group is small relative to that in the full sample.

The existence of voter heterogeneity, though, raises questions about the appropriateness of modelling turnout by the standard logit model. The logit model is now widely used because of the simplicity and intuitive interpretation of the probability expressions and ease of computation. However, this model is subject to serious restrictions, among them the homoscedasticity assumption. The homoscedasticity assumption is not plausible in many situations. We have shown that voters vary in the accuracy of their utility calculations, which directly implies that the error terms in the utility functions are heteroscedastic. In general, estimates of both the parameters and the choice probabilities using a homoscedastic

[30]One goodness-of-fit index is: $R^2 = 1 - \frac{L(\widehat{\beta})}{L(0)}$, where $L(0)$ and $L(\widehat{\beta})$ are the initial log-likelihood and that at convergence respectively (*c.f.*, Ben-Akiva and Lerman 1985). Correct classification results when the predicted choice (that having the highest predicted probability) by the model matches the real choice.

model will be inconsistent if the true model is heteroscedastic. Another reason why the problem of heteroscedasticity should receive attention is given by Downs and Rocke (1979), who demonstrate that the pattern of heteroscedasticity may hold important implications for substantive insights. Therefore, given the clear evidence of heteroscedasticity in the turnout model, a heteroscedastic model should be employed.

A class of heteroscedastic choice models is proposed in Chapter 1, with the heteroscedastic logit model being a special case. Dubin and Zeng (1991) explore the application of the heteroscedastic logit model using a economic data set, and find that the model performs better than a standard logit model when heteroscedasticity is present. Assuming the utility that alternative i provides to individual t is u_{it}, which consists of a deterministic component v_{it} (usually specified as $v_{it} = \sum_{k=1}^{K} x_{it}^{k}\beta^{k}$ with x the observed independent variables and β the unknown parameters to be estimated) and an unobserved error term $\epsilon_{it} : u_{it} = v_{it} + \epsilon_{it}$, and assuming the error terms follow an extreme value distribution with variances $Var(\epsilon_{it}) = 1/6 * \frac{\pi^2}{\theta_t^2}$,[31] the choice probabilities given by the heteroscedastic logit model are:

$$p_{it} = e^{v_{it}\theta_t} / \sum_{j=1}^{I} e^{v_{jt}\theta_t} \tag{2.11}$$

where $\theta_t > 0 \ \forall t$.

Expression (2.11) is similar to the standard logit probabilities, except that the

[31] In the standard logit model, $Var(\epsilon_{it}) = 1/6 * \pi^2 \ \forall i; \forall t$.

deterministic part of the utility functions, the v_{it}'s, are weighted by θ_t's that are inversely related to the standard error of the error terms. Intuitively this means that more weight is given to the utility functions of those individuals with smaller variances of the error terms. In our turnout problem, we assume that information is inversely related to the standard errors, and thus use it in place of θ_t's (after transformed to take positive values only). Doing so we give more weight to the utility functions of the better informed voters.[32]

As pointed out in Chapter 1, although in general the heteroscedastic logit model may be nonlinear in the parameters, if θ_t's are linear in parameters, as is the case here,[33] then the heteroscedastic model can be estimated using a standard logit package after transformations of the independent variables. In our case, we only need to interact *all* explanatory variables including the constant term with information and use the resulting values as the new explanatory variables to estimate the model with a logit package.[34]

Table 2.5-2 reports the estimation results of the heteroscedastic logit turnout model as well as the standard logit turnout model[35] for comparison. The results

---

[32] Note that weighting the utility functions this way not only introduces heteroscedasticity, but also allows correlation between the error terms, as voters with similar levels of information are weighted similarly.

[33] To use information in places of θ_t's is equivalent to assuming $\theta = \alpha * Info.$, where α is an unknown parameter which cannot be identified from β's in v_{it}.

[34] It is important that for the model to be an appropriate random utility maximization model, θ_t must be positive and must weigh the entire v_{it} component. See the derivation of the heteroscedastic GEV class in Chapter 1 for an explanation of this. We emphasize this fact here because our turnout problem is one of the situations in which, at first glance, we would consider interacting θ with only some of the explanatory variables, because it seems that information only affects the accuracy of the P terms and hence interacts with these terms only. We actually estimated such a model and it is no improvement over the original model at all; the log likelihood is even slightly lower.

[35] The P model is used as it is the best. See the discussion in Section 2.

are obtained only for turnout because, as described earlier, the variation in information is much smaller within the voting group, and hence the effect of information on candidate choice decision is not strong. Using non-nested hypothesis testing, as discussed in the appendix, we see that given the two models have the same number of parameters and that since the difference in the log likelihood is higher than the critical value of 1.92 at a significance level of 0.05, we can assert that the heteroscedastic model is the better model with probability greater than 95%. A graphical comparison of the models is presented in Figure 2.5. A major difference between the two models is that the heteroscedastic model performs better in extreme probability margins. For example, for people who actually voted, the heterogeneous logit model predicts more of them to vote with probability higher than 0.8 than the standard logit model does, and fewer of them to vote with probability lower than 0.4. For non-voters, it predicts fewer of them to vote with probability higher than 0.6, and more of them to have a probability of voting below 0.2. This suggests that the heteroscedastic model separate out the electorate more precisely than the standard logit.

Next we wish to compare the implications of the two models for the importance of the independent variables in the turnout decision equation. We are especially interested in seeing how the role of the information index changes, if it does change from one model to another. For this purpose we need to compute the (estimated) partial derivatives or elasticities of the choice probabilities with respect to the independent variables. We prefer using elasticities over partial derivatives because the former are normalized for the variables' units. An elasticity is the percentage change in one variable that is associated with a percentage change in another variable. So the elasticity of the turnout probability with respect to information is the percentage change in the turnout probability associated with a percentage change in the information level. Different individuals have different choice probabilities and hence different elasticities. We report the *average* elasticities and the *aggregate* elasticities of the turnout probabilities in Table 2.5-3.[36]

Examination of Table 2.5-3 reveals surprisingly large differences in the effects of information on turnout probabilities suggested by the two different models, while the effects of other explanatory variables do not vary much from one model to another. Both the average and aggregate elasticities of turnout probabilities with respect to information in the logit model round up to 0.16; whereas in the heteroscedastic model they are 0.54 and 0.57 respectively. Since the heteroscedastic model is the superior model (with probabilities greater than 0.95), this result strongly suggests that information plays a more important role in turnout than previous analysis indicates. Downs and Rocke (1979) are indeed correct to assert that heteroscedasticity not only affects accuracy of estimation, but its analysis "can also provide political scientists with significant substantial information that would

[36]Aggregate elasticity measures the effect of a change in an independent variable on the *expected share* of a group choosing an alternative. Ben-Akiva and Lerman (1985, pp.112-113) showed that the aggregate elasticity associated with a uniform percent change in some independent variable across all members of a group is a weighted average of the individual level elasticities using the choice probabilities as weights.

ordinarily go undetected as well as evidence of model misspecification"(p.816).

2.5 Summary and Conclusion

This chapter addresses some theoretical and methodological issues regarding rational voter models of voting behavior in presidential elections, and discusses a methodological treatment of voter heterogeneity. First, we distinguish models in which closeness of the race alone enters the voting calculus from the standard expected utility maximization models, arguing that the former are actually utility maximization models (pure "consumption" models) in which closeness is *not* a probability discount factor but an independent variable bearing its own substantive meaning. This rationale not only formulates a distinct rational turnout model, but also implies that empirical support for the Downsian model based on the significance of the P term is not valid. In comparing the various rational turnout models, we demonstrate through formal tests that, if variables specified by different models are included in one general model, then severe multicollinearity exists that can damage the desirable properties of the maximum likelihood estimator, and inferences based on such estimators can be unreliable and misleading. Some alternative estimators are computed and their properties briefly discussed. We then estimate the turnout models separately and base our comparison of the models on rigorous non-nested hypothesis testing. For the turnout decision, empirical analysis of the 1968 and 1980 national election data suggests weak support for the utility maximization model against the expected utility maximization model and the minmax regret model.

We note that the decision to participate and the decision to cast a vote for a particular candidate are of distinct types, and the importance of various variables, especially the expected utility terms, would be different between these decisions. This view is supported by the data, and we show that in the decision to cast a specific vote, unlike the decision to turnout, expected utility terms significantly affect the calculus. Furthermore, voters who favor a minor party candidate are more affected by such strategic considerations than the other voters, whose decisions are basically determined by such social-psychological variables as candidate evaluation and partisanship. These findings provide strong evidence for the notion of strategic voting in three candidate elections.

Lastly, we address the question of voter heterogeneity. One source of heterogeneity in voting behavior is different levels of voter information, which introduces heteroscedasticity and makes the widely used standard logit models inappropriate. The application of a heteroscedastic logit model is therefore discussed and is shown to significantly outperform the standard logit and to reveal stronger effects of voter information on voting decisions. A useful non-nested hypothesis testing technique is also discussed and its application simplified.

Before ending, we offer some general remarks on the current study and its position in the literature. How voters make up their minds is one of the most thoroughly studied topics in political science as well as one of the most controversial ones. "There are countless ways of understanding voting," write Niemi and Weisberg (1984). We have a full range of all these seemingly different and even opposing models: the "social-historical" models, the "sociological" models, the "social-psychological" models, and the "rational voter" models. Within the rational voter category, our conclusion in Section 2 seemingly adds another "utility maximization" turnout model to the existing expected utility maximization model and the minmax regret model. However, viewed from a higher level, the controversies and the classification of research methods are to a large extent artificial. Indeed, different methods emphasize different variables that affect voting behavior. But all relevant variables are an integrated part in determining how a voter acts. A voter is at the same time a "social-historical" being, a "psychological" being, and a "rational" being ("rational" in the narrow sense as understood in the rational voter model context). It is only that in different situations and for different individuals, a different being may dominate. All these "beings" are but different *aspects* of the voter's characteristics. And thus viewed, all feelings, ideas, decisions, and actions, *etc.* are *rational* in the sense that they bring to him certain *utilities*, if utility is understood in a broad sense that includes all types of fulfillment and satisfactions (including satisfaction derived from, possibly, pain, for example). Using this utility concept we readily obtain a "unifying" model: a model in which voters maximize this "gross" utility. All the seemingly different models are just different faces of this same one because the factors considered in each can all potentially be part of this gross utility. And the only question becomes in various situations exactly which of these factors may play a more significant role.

In this view, our study does not propose any new models beyond this gross utility maximization model (no studies really do), and it "merely" obtains empirical evidence of factors contributing to gross utility that are not tested or tested satisfactorily before. Our findings suggest that, in turnout decision making, closeness of the race is a factor that independently contributes to the gross utility of voting as a measure of efficacy. Other potential contributors, such as the expected return from the election outcome or the utility difference indices associated with regrets, do not appear significant. In contrast, in candidate choice decision making, the expected utility factors do play a significant role, and they affect different voters in different ways.

Although we believe there *is* an unified way of examining voting behavior, we do not think we are close to a perfect understanding of it. Niemi and Weisberg (1984) write that "...we do not yet fully understand voting and elections." To this we might add that, "we can *never fully* understand voting," given the diversity and heterogeneity in any large population. Our results on voter heterogeneity provide clear evidence of this variation and reveal one source— information— for it.

Appendix: Non-Nested Hypothesis Testing

Many of the familiar hypothesis testing situations involve *nested* models, *i.e.*, one model (the *restricted model*) can be obtained from the other model (the *unrestricted model*) under the null hypothesis about the parameters. The t test on the significance of a parameter of a model, for example, is a testing of nested hypothesis, with the unrestricted model being the model estimated, the restricted model being a special case of this model when one of the parameters is zero. Statistical testing techniques are well developed for this type of hypothesis testing. The classical Wald, Lagrange Multiplier and likelihood Ratio tests are all for nested hypotheses. However, there are instances where we wish to compare two models, and one is not a special case of the other. The comparison of the various rational voter models is such a case. Neither can be obtained as a (parametric) special case of the other. Comparison of such models need to be carried out by *non-nested* hypothesis testing techniques. Non-nested hypothesis testing is not as mature as its nested hypothesis counterpart, however, we show below that one of such tests described by Ben-Akiva and Lerman (1985, pp.171-172) can be simple to use and we produce a mini-table of the critical values of the test.

Suppose we need to compare model 1 and model 2 which are not nested. It was shown that, under the null hypothesis that model 1 is true, the following holds asymptotically:

$$Pr(\rho_2^2 - \rho_1^2 > z) \leq \{-[-2zL(0) + (K_2 - K1)]^{1/2}\}, \quad z > 0, \tag{2.12}$$

where is the standard normal cumulative distribution function, ρ_h^2 is the adjusted likelihood ratio index for model $h = 1, 2$, $\rho_h^2 = 1 - \frac{L_h(\widehat{\beta}) - K_h}{L(0)}$, $L(0)$ is the initial log likelihood of the models, and $L_h(\widehat{\beta})$ are the log likelihood at convergence of model $h = 1, 2$, K_h is the number of parameters of model $h = 1, 2$. Expression (2.12) means that the probability that model 1, which has the lower adjusted likelihood ratio index, is the true model is bounded by the right hand side of (2.12). Note that $\rho_2^2 - \rho_1^2 = \frac{[l(2) - l(1)] + [K_2 - K_1]}{l(0)}$, where $l(h) =\mid L_h(\widehat{\beta}) \mid$ and $l(0) =\mid L(0) \mid$. Hence the right hand side of (2.12) becomes: $\{-[2(l(1) - l(2) + (K_1 - K_2)]^{1/2}\}$. Now suppose we want to find the critical values of the (two-tailed) test at significance level 0.05. Because $(-1.96) = .025$, we set $[2(l(1) - l(2)) + (K_1 - K_2)]^{1/2} = 1.96$, then:

$$l(1) - l(2) = 1.92 - 1/2(K_1 - K_2).$$

This means, for example, that if the two models have the same number of parameters, but model 2's log likelihood is more than 1.92 lower than model 1's in absolute values, then the probability that model 1 is the true model is less than 0.05. Table A lists the critical values for some commonly used significance levels and various $K_1 - K_2$ values.

References

Abramson, P.R. *et al.* 1982. *Changes and Continuity in the 1980 Elections.* Congressional Quarterly Press.

Ashenfelter, O. and Kelly, S. Jr. 1975. "Determinants of Participation in Presidential Elections." *The Journal of Law and Economics.* 18:695-733.

Ben-Akiva,M. and Lerman,S. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* The MIT Press.

Bennett, S.E. and Resnick, D. 1990. "The Implications of Nonvoting for Democracy in the United States." *American Journal of Political Science,* Vol.34, No.3:771-802.

Black, J.H. 1978. "The Multicandidate Calculus of Voting: Application to Canadian Federal Elections." *American Journal of Political Science,* Vol.22, No.3, 609-638.

Cain, B.E. 1978. "Strategic Voting in Britain." *American Journal of Political Science,* Vol. 22, No.3, 639-655.

Chow, G.C. 1983. *Econometrics.* McGraw-Hill Boo Company.

Cohen, L.R. and Uhlaner, C.J. 1991. "Participation, Closeness, and Economic Voting in the 1960 Presidential Election." Prepared for the 1991 Public Choice Society Annual Meeting, New Orleans, March 1991.

Downs, A. 1957. *An Economic Theory of Democracy.* New York, Harper and Row.

Downs, G.W. and Rocke, D.M. 1979. "Interpreting Heteroscedasticity." *American Journal of Political Science.* Vol.23, No.4:816-828.

Dubin, J.A. and Zeng, L. 1991. "A Heterogeneous Logit Model." *Social Science*

Working Paper, California Institute of Technology. Feb. 1991.

Duverger, M. 1967. *Political Parties.* London: Lowe and Brydone.

Farquharson, R. 1969. *The Theory of Voting*. New Haven: Yale University Press.

Ferejohn, J.A. and Fiorina, M.P. 1974. "The Paradox of Not Voting: a Decision Theoretical Analysis." *American Political Science Review*. 68:525-536

Ferejohn, J.A. and Fiorina, M.P. 1975. "Closeness Counts Only in Horseshoes and Dancing." *American Political Science Review*, 69:920-925.

Field, M. 1981. "Presidential Election Polling: Are the States Righter?" *Public Opinion*, Vol.4, No. 5. Oct/Nov. 1981.

Foster, C.B. 1984. "The Performance of Rational Voter Models in Recent Presidential Elections." *American Political Science Review*. 78:678-690.

Goldberger, A.S. 1969. "Social Determinism and Rationality as Base of Party Identification." *American Political Science Review*, 63:5-25.

Hansen, S., Palfrey, T.R. and Rosenthal, H. 1987. "The Downsian Model of Electoral Participation: Formal Theory and Empirical Analysis of the Constituency Size Effect." *Public Choice*. 52: 15-33.

Johnston, J. 1984. *Econometric Methods*. McGraw-Hill Book Company.

Judge, G. *et al.* 1985. *The Theory and Practice of Econometrics*. New York: Wiley.

Katosh, J.P. and Traugott, M.W. 1982. "Costs and Values in the Calculus of Voting." *American Journal of Political Science*, Vol.26, No.2:361-376.

Kiewiet, D.R. 1979. "Approval Voting: the Case of the 1968 Election." *Polity*, 12:170-181.

Laver, M. 1987. "The Logic of Plurality Voting in Multi-Party Systems." In M.J. Holler(ed.): *The Logic of Multi-party Systems.* Copyright 1987 by Martinus Nijhoff Publishers, Dordrecht, The Netherlands.

Ledyard, J.O. 1984. "The Pure Theory of Two-Candidate Elections." *Public Choice,* 44:7-41

Lee, L. 1982. "Specification Error in Multinomial Logit Models: Analysis of the Omitted Variable Bias." *Journal of Econometrics,* Vol.20, pp. 197-209.

Manski, C. and D. McFadden, eds., 1981. *Structural Analysis of Discrete Data,* Cambridge, Mass. MIT Press.

Marx, B.D. and Smith, E.P. 1990. "Weighted Multicollinearity in Logistic Regression: Diagnostics and Biased Estimation Techniques with an Example from Lake Acidification." *Can. J. Fish. Aquat. Sci.,* Vol. 47, No.6, pp. 1128-1135.

McKelvey, R.D. and Ordeshook, P.C. 1972. "A General Theory of the Calculus of Voting." In J. F.Herdon and J. L. Bernd, eds., *Mathematical Applications in Political Science,* VI. Charlottesville: University Press of Virginia, pp. 32-78.

Myerson, R.B. and Weber, R.J. 1988. "A Theory of Voting Equilibria." *Discussion Paper NO. 782.* The Center for Mathematical Studies in Economics and Management Science Northwestern University, Evanston, Illinois 60208.

Niemi, R.G. and Weisberg H.F. 1984. *Controversies in Voting Behavior.* Second Edition. Congressional Quarterly Press.

Palfrey, T.R. 1989. "A Mathematical Proof of Duverger's Law." In P.C. Ordeshook(ed.): *Models of Strategic Choice in Politics.* Ann Arbor, Michigan: University of Michigan.

Palfrey, T.R. and Poole,K.T. 1987. "The Relationship Between Information, Ideology and Voting Behavior." *American Journal of Political Science,* Vol.31, No.3:511-530.

Palfrey, T.R. and Rosenthal, H. 1985. "Voter Participation and Strategic Uncertainty." *American Political Science Review.* 79:62-78.

Plott, C.R. 1991. "A Comparative Analysis of Direct Democracy, Two-Candidate Elections, and Three Candidate Elections in an Experimental Environment." In T.R. Palfrey(ed.):*Contemporary Laboratory Research in Political Economy.* Ann Arbor: University of Michigan Press.

Poole, K.T. and Rosenthal H. 1984. "U.S. Presidential Elections 1968-1980: a Spatial Analysis." *American Journal of Political Science,* Vol.28, No.2:282-312.

Riker, W. H. and Ordeshook, P.C. 1968. "A Theory of the Calculus of Voting." *American Political Science Review.* 62:25-42.

Rivers, D. 1988. "Heterogeneity in Models of Electoral Choice." *American Journal of Political Science.* Vol.32, No.3:737-757.

Rosenstone, S. J. *et al.* 1984. *Third Parties in America.* Princeton University Press.

Rosenthal, H. and Sen, S. 1973. "Electoral Participation in the French Fifth Republic." *American Political Science Review.* 67: 29-54.

Schaefer, R.L. 1986. "Alternative Estimators in Logistic Regression When the Data

Are Collinear." *J. Statist. Comput. Simul.*, 25:75-91.

Schaefer, R.L.,Roi, L.D. and Wolfe, R.A. 1984. "A Ridge Logistic Estimator." *Communications in Statistics:* Theory and Methods. Vol. 13, No. 1. 99-113.

Varian, H. 1984. *Microeconomic Analysis.* New York: W.W. Norton and Company.

Table 2.1
Multicollinearity Diagnostics

Variables	Variance Inflation				Variance Proportion†			
	Model 1		Model 2		Model 1		Model 2	
	1968	1980	1968	1980	1968	1980	1968	1980
Const	17.7	4.0	40.0	1508	.92 *	.76 *	.99 *	.99 *
PB12	8.5	4.5	46.8	1370	.63 *	.14	.99 *	.99 *
PB13	9.0	11.2	—	—	.64 *	.37	—	—
PB23	5.0	4.5	—	—	.05	.09	—	—
P12	21.8	32.9	—	—	.92 *	.99 *	—	—
P13	23.7	43.0	—	—	.94 *	.99 *	—	—
P23	9.2	28.1	—	—	.95 *	.99 *	—	—
B12	11.1	3.7	2.29	1530	.60 *	.00	.36	99 *
B13	7.1	10.2	—	—	.10	.24	—	—
Alien.	3.4	2.2	—	—	.08	.14	—	—
Info.	1.4	1.5	—	—	.10	.21	—	—
Educ.	1.3	1.8	—	—	.02	.24	—	—
Party	1.4	1.6	—	—	.07	.20	—	—
Max. Corr. Coeff. (Abs. value)	.72	.68	.98	.55				
Max. Condition Number	17	19	15	83				

† These are the variance proportions attributed to the smallest eigenvalues.

* The corresponding coefficients are likely affected by multicollinearity.

Table 2.2-1

The MLE, Ridge, Principle Component and Stein Estimators: the 1968 Data

Var	Coefficients			Standard Error			*t*-Statistic †		
	MLE	Ridge	PC	MLE	Ridge	PC	MLE (& Stein)	Ridge	PC
Const.	0.56	0.05	-1.09	1.34	0.45	0.64	0.42	0.12	-1.71
PB12	12.42	4.90	7.00	6.55	3.60	4.24	1.90 **	1.32	1.65 *
PB13	12.51	7.40	7.26	6.04	3.99	4.38	2.07 **	1.86 *	1.66 *
PB23	-5.32	0.15	-2.11	6.80	4.99	6.58	-0.74	0.02	-0.32
P12	-13.23	-9.27	-4..11	4.70	2.24	2.24	-2.82 ***	-4.12 ***	-1.83 *
P13	-14.86	-1.14	-5.29	5.14	2.51	2.85	-2.90 ***	-4.55 ***	-1.86 *
P23	-9.86	-9.18	-4.68	3.90	2.42	3.31	-2.52 **	-3.80 ***	-1.42
B12	-2.83	-1.00	-1.46	1.27	0.46	0.81	-2.24 **	-2.17 **	-1.81 *
B13	0.17	0.14	0.33	0.75	0.45	0.74	0.23	0.31	0.46
Alien.	1.86	1.30	1.93	0.61	0.42	0.61	3.03 ***	3.09 ***	3.16 ***
Info.	0.50	0.49	0.32	0.07	0.07	0.07	6.87 ***	6.87 ***	4.92 ***
Educ.	0.23	0.23	0.32	0.07	0.07	0.07	3.15 ***	3.21 ***	4.48 ***
Party	0.30	0.30	-0.10	0.08	0.08	0.04	3.83 ***	3.97 ***	-2.71 ***
LL at conv.:	-592.4	-593.8	-612.6	Initial LL: -874.1			Stein shrinkage parameter: $c = 0.74$.		
	-594.5 (Stein)			$N = 1261$					

† heuristic for the ridge, PC and Stein estimators.
* significant at .10 level.
** significant at .05 level.
*** significant at .01 level.

Table 2.2-2

The MLE, Ridge, Principle Component and Stein Estimators: the 1980 Data

Var	Coefficients			Standard Error			*t*-Statistic †		
	MLE	Ridge	PC	MLE	Ridge	PC	MLE (& Stein)	Ridge	PC
Const.	-5.42	-2.45	0.06	1.72	0.79	0.19	-3.15 ***	-3.08 ***	0.32
PB12	2.83	4.05	6.92	6.50	5.83	6.37	0.44	0.69	1.09
PB13	-1.97	-0.70	1.98	4.69	3.76	4.52	-0.42	-0.20	0.44
PB23	0.70	1.95	4.26	6.08	5.13	5.97	0.12	0.38	0.71
P12	11.12	2.40	-5.50	5.97	3.45	2.94	1.86 *	0.70	-1.87 *
P13	13.67	4.76	-3.62	6.06	3.40	2.73	2.26 **	1.40	-1.33
P23	12.70	4.12	-3.52	5.75	3.37	2.72	2.21 **	0.22	-1.30
B12	-0.59	-0.44	-0.42	0.77	0.61	0.77	-0.76	-0.72	-0.54
B13	-0.16	-0.29	-0.82	1.02	0.71	1.00	-0.16	-0.41	-0.82
Alien.	1.74	0.94	0.56	0.83	0.62	0.75	2.08 **	1.51	0.74
Info.	0.96	0.86	0.83	0.26	0.25	0.25	3.73 ***	3.47 ***	3.24 ***
Educ.	0.27	0.25	0.23	0.10	0.10	0.10	2.66 ***	2.48 **	2.29 **
Party	0.30	0.29	0.27	0.09	0.09	0.09	3.21 ***	3.09 ***	2.90 ***
LL at conv.:	-413.8 -414.7 (Stein)	-415.4	-419.0	Initial LL: -476.8 $N = 675$			Stein shrinkage parameter: $c = 0.86$		

† heuristic for the ridge, PC and Stein estimators.
* significant at .10 level.
** significant at .05 level.
*** significant at .01 level.

Table 2.3-1

Comparison of Rational Turnout Models: the 1968 Data

Var.	MLE				*t*-Statistic			
	Model 1	Model2	Model3	Model4	Model 1	Model2	Model3	Model4
Const.	-3.18	-1.13	-3.16	-3.30	-6.23***	-1.04	-6.20***	-6.70***
PB12	2.19	—	—	—	0.89	—	—	—
PB13	0.65	—	—	—	0.31	—	—	—
PB23	0.96	—	—	—	0.29	—	—	—
P12	—	-7.03	—	—	—	-1.86*	—	—
P13	—	-7.80	—	—	—	-2.16**	—	—
P23	—	-8.23	—	—	—	-2.55**	—	—
B12	—	—	-0.50	—	—	—	-1.12	—
B13	—	—	0.78	—	—	—	2.26**	—
Alien.	1.79	2.20	1.56	2.05	3.09***	4.41***	2.61***	4.21***
Info.	0.50	0.50	0.50	0.50	6.96***	6.97***	6.93***	6.98***
Educ.	0.28	0.25	0.26	0.28	3.75***	3.43***	3.53***	3.86***
Party	0.31	0.31	0.31	0.31	4.01***	4.04***	4.05***	4.13***
LL at conv.:	-601.04	-596.98	-598.82	-601.45	Initial LL: -874.06		$N = 1261$	

Model 1: the Downsian model.
Model 2: the P-model.
Model 3: the minmax regret model.
Model 4: D ans C terms only.

* significant at .10 level.
** significant at .05 level.
*** significant at .01 level.

Table 2.3-2

Comparison of Rational Turnout Models: the 1980 Data

Var.	MLE				*t*-Statistic			
	Model 1	Model2	Model3	Model4	Model 1	Model2	Model3	Model4
Const.	-1.75	-5.02	-1.92	-1.51	-2.98***	-3.04***	-3.19***	-2.80***
PB12	-3.31	—	—	—	-0.92	—	—	—
PB13	-1.51	—	—	—	-0.87	—	—	—
PB23	0.63	—	—	—	0.20	—	—	—
P12	—	12.03	—	—	—	2.10**	—	—
P13	—	12.90	—	—	—	2.27**	—	—
P23	—	13.12	—	—	—	2.35**	—	—
B12	—	—	-0.52	—	—	—	-0.93	—
B13	—	—	-0.31	—	—	—	-0.67	—
Alien.	1.23	0.93	1.60	0.78	1.58	1.47	1.94*	1.25
Info.	0.92	0.88	0.96	0.89	3.61***	3.49***	3.75***	3.55***
Educ.	0.26	0.27	0.27	0.27	2.64***	2.66***	2.73***	2.71***
Party	0.30	0.31	0.29	0.28	3.20***	3.26***	3.10***	3.08***
LL at conv.:	-417.87	-415.45	-417.43	-418.63	Initial LL: -467.87			*N* = 675

Model 1: the Downsian model.
Model 2: the P-model.
Model 3: the minmax regret model.
Model 4: D ans C terms only.

* significant at .10 level.
** significant at .05 level.
*** significant at .01 level.

Table 2.4-1

Strategic Voting: Estimation of the Downsian Model

Variables	MLE		*t*-Statistic	
	1968	1980	1968	1980
Const.	1.28	0.08	4.34***	0.23
PB12	47.59	47.66	5.48***	2.89***
PB13	11.28	56.98	2.35**	4.32***
PB23	-12.57	-27.35	-2.87***	-5.11***
Initial LL:	-583.63	-237.75		
LL at Conv.:	-170.88	-97.55		
N:	842	343		

* significant at .10 level.
** significant at .05 level.
*** significant at .01 level.

Table 2.4-2

Strategic Voting: Estimation of the Modified Model

Variables	MLE		*t*-Statistic	
	1968	1980	1968	1980
Const.	-0.22	-1.65	-0.27	-1.36
PB12	27.49	24.55	1.93*	1.37
PB13	5.13	54.03	0.90	3.75***
PB23	-12.87	-29.95	-2.68***	-4.78***
U1	5.78	4.47	2.30**	1.98**
U2	-4.19	-2.10	-1.77*	-1.02
Party	0.37	0.33	4.19***	2.86***
Initial LL:	-586.63	-237.75		
LL at Conv.:	-159.38	-90.69		
N:	842	343		

* significant at .10 level.
** significant at .05 level.
*** significant at .01 level.

Table 2.4-3

Strategic Voting: Comparison of Voters Favoring a Third Candidate and Other Voters

Var.	MLE				t-Statistic			
	1968		1980		1968		1980	
	W.-voters †	Others	A.-voters	Others	W.-voters	Others	A.-voters	Others
Const.	-1.17	0.83	-2.63	0.46	-0.46	0.78	-1.82*	0.19
PB12	18.10	4.00	-74.89	-85.86	0.13	0.10	-0.64	-0.97
PB13	22.60	-6.22	49.36	12.43	0.63	-0.44	1.30	0.43
PB23	-21.33	1.76	-23.78	23.49	-1.90*	0.10	-2.33**	0.28
U1	7.97	0.11	8.94	27.52	1.12	1.42	2.14**	1.32
U2	-5.07	-0.12	-5.90	-28.12	-0.71	-1.45	-1.46	-1.39
Party	0.43	0.46	0.21	0.57	1.55	3.85***	1.49	2.25**
Initial LL:	-44.4	-470.	-74.9	-163.				
LL at conv.:	-22.2	-99.9	-57.6	-28.9				
N:	64	678	108	235				
% correct class.:	85.9	95.1	73.2	95.8				

† voters who prefer Wallace most in states where Wallace finished third.
* significant at .10 level.
** significant at .05 level.
*** significant at .01 level.

Table 2.5-1

Turnout Models for the Better and the Less Informed Voters

Variables	MLE		*t*-Statistic	
	Better Informed	Less Informed	Better Informed	Less Informed
Const.	-6.89	-3.32	-3.00***	-1.30
P_{12}	16.68	6.89	2.19**	0.77
P_{13}	16.40	8.98	2.20**	1.01
P_{23}	16.91	8.70	2.29**	1.00
Party	0.32	0.30	2.46**	2.17**
Alien.	1.17	0.52	1.27	0.60
Info.	1.77	-0.14	1.63	-0.27
Educ.	0.25	0.23	1.79*	1.67*
Initial LL:	-273.8	-194.1		
LL at conv.:	-224.3	-184.5		
% Correct Class.:	72.9	57.5		
N:	395	280		

* significant at .10 level.
** significant at .05 level.
*** significant at .01 level.

Table 2.5-2

Comparison of the Standard and the Heteroscedastic turnout Models

Variables†	MLE		*t*-Statistic	
	Logit	Heteros. Logit	Logit	Heteros. Logit
Const.	-5.02	-5.15	-3.04***	-3.54***
P12	12.03	11.12	2.10**	2.30**
P13	12.90	11.31	2.27**	2.38**
P23	13.12	11.72	2.35**	2.49**
Party	0.31	0.24	3.26***	3.05***
Alien.	0.92	0.95	1.46	1.67*
Info.	0.88	0.87	3.49***	2.88***
Educ.	0.26	0.21	2.66***	2.52**
LL at conv.:	-415.5	-412.8	LL init.: -467.9	*N* = 675

† For the heteroscedastic logit model all independent variables have been multiplied by the information index.
* significant at .10 level.
** significant at .05 level.
*** significant at .01 level.

Table 2.5-3

Elasticities of Choice Probabilities of the Logit and the Heteroscedastic Logit Turnout Models w.r.t. Independent Variables

Variables	Average Elasticity		Aggregate Elasticity	
	Logit	Heteros. Logit	Logit	Heteros. Logit
P_{12}	0.34	0.32	0.30	0.30
P_{13}	0.54	0.53	0.52	0.52
P_{23}	0.32	0.32	0.30	0.31
Party	0.18	0.16	0.18	0.16
Alien.	0.25	0.27	0.24	0.27
Info.	0.16	0.54	0.16	0.57
Educ.	0.16	0.14	0.16	0.14

Table A

Non-Nested Hypothesis Testing: Critical Values of $(l_1 - l_2)$

α \ k_1-k_2	-5	-4	-3	-2	-1	0	1	2	3	4	5
0.01	5.82	5.32	4.82	4.32	3.82	3.32	2.82	2.32	1.82	1.32	0.82
0.05	4.42	3.92	3.42	2.92	2.42	1.92	1.42	0.92	0.42	-0.08	-0.58
0.10	3.86	3.36	2.86	2.36	1.86	1.36	0.86	0.36	-0.14	-0.64	-1.14

α: significance level.
$k_{1,2}$: Number of Parameters in model 1, 2.
Formula: $l_1 - l_2 = c - (k_1 - k_2)/2$
where $c = 3.32$ for $\alpha = 0.01$; $c = 1.92$ for $\alpha = 0.05$ and $c = 1.36$ for $\alpha = 0.10$.

Figure 2.1-1

Performance of Rational Turnout Models: the 1968 Data

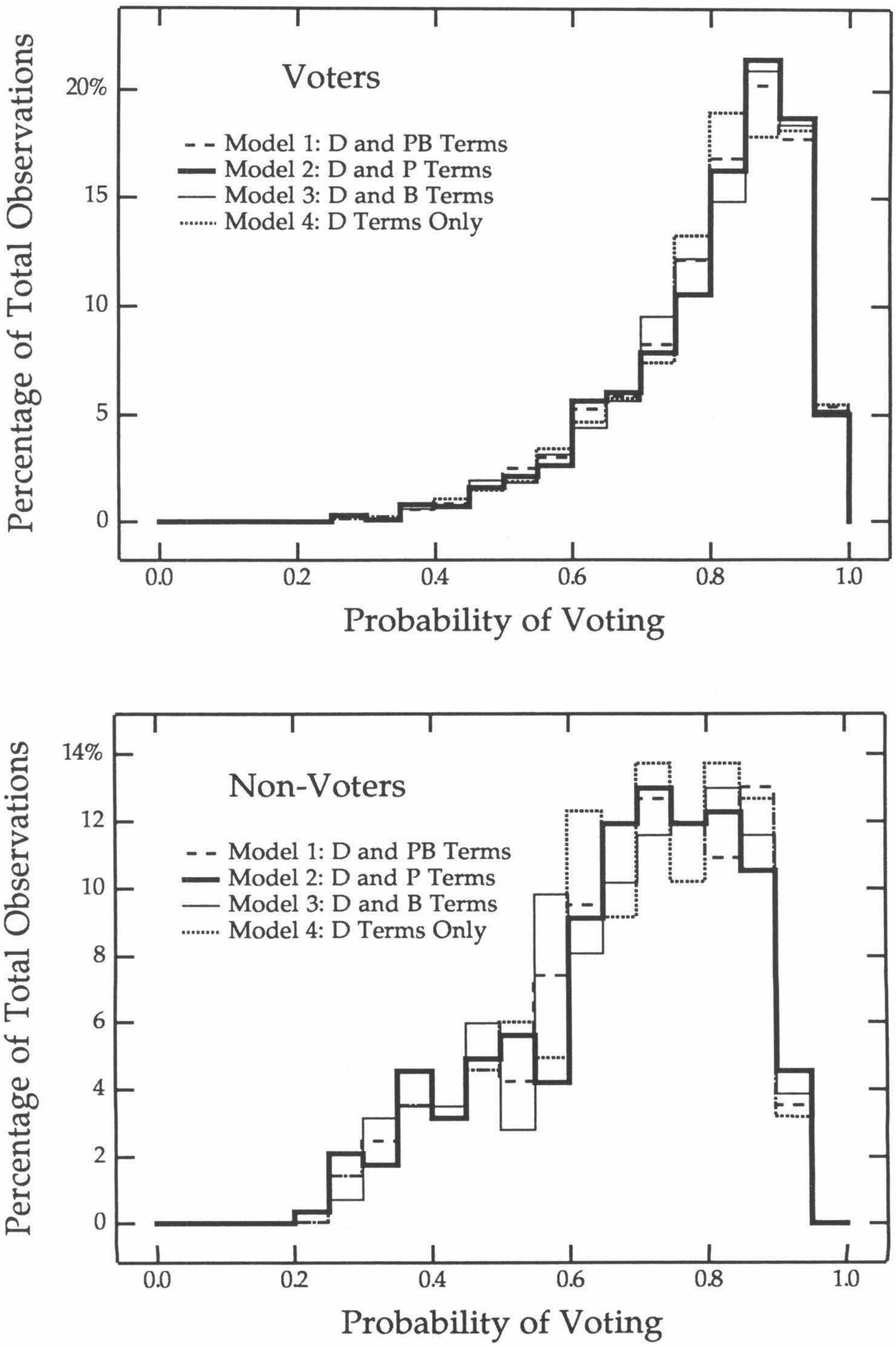


Figure 2.1-2

Performance of Rational Turnout Models: the 1980 Data

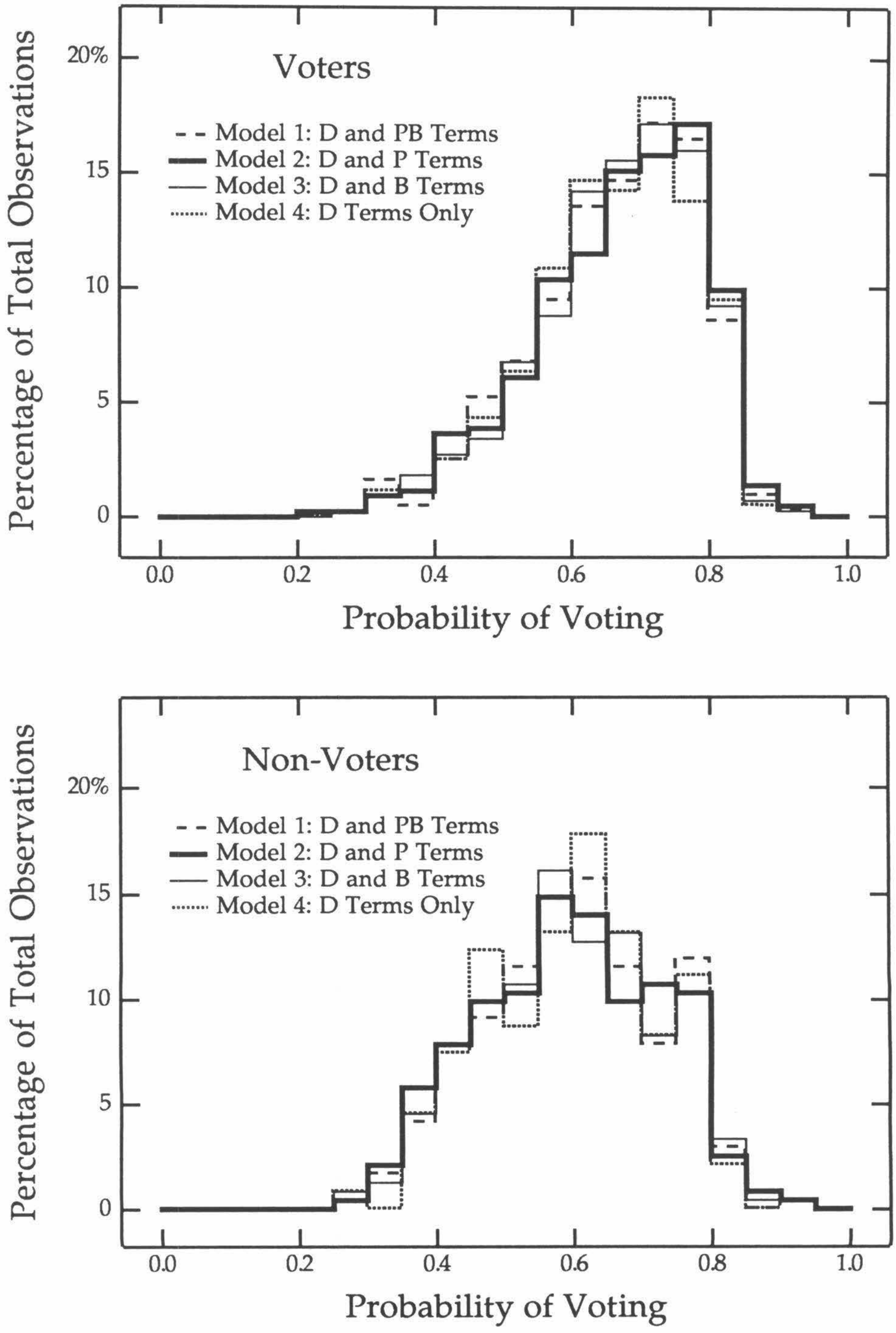


Figure 2.2-1

Performance of Strategic Voting Models: the 1968 Data

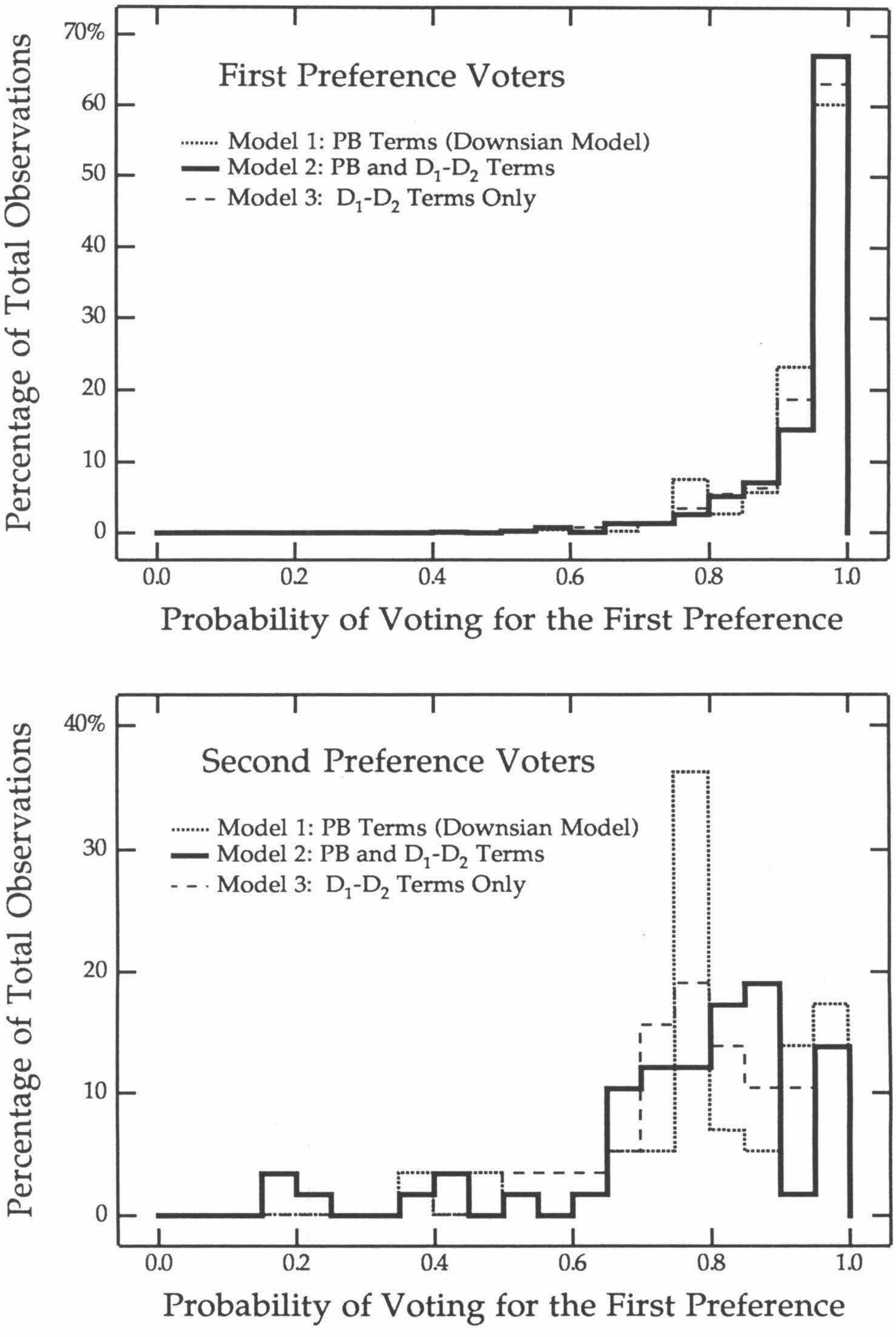


Figure 2.2-2

Performance of Strategic Voting Models: the 1980 Data

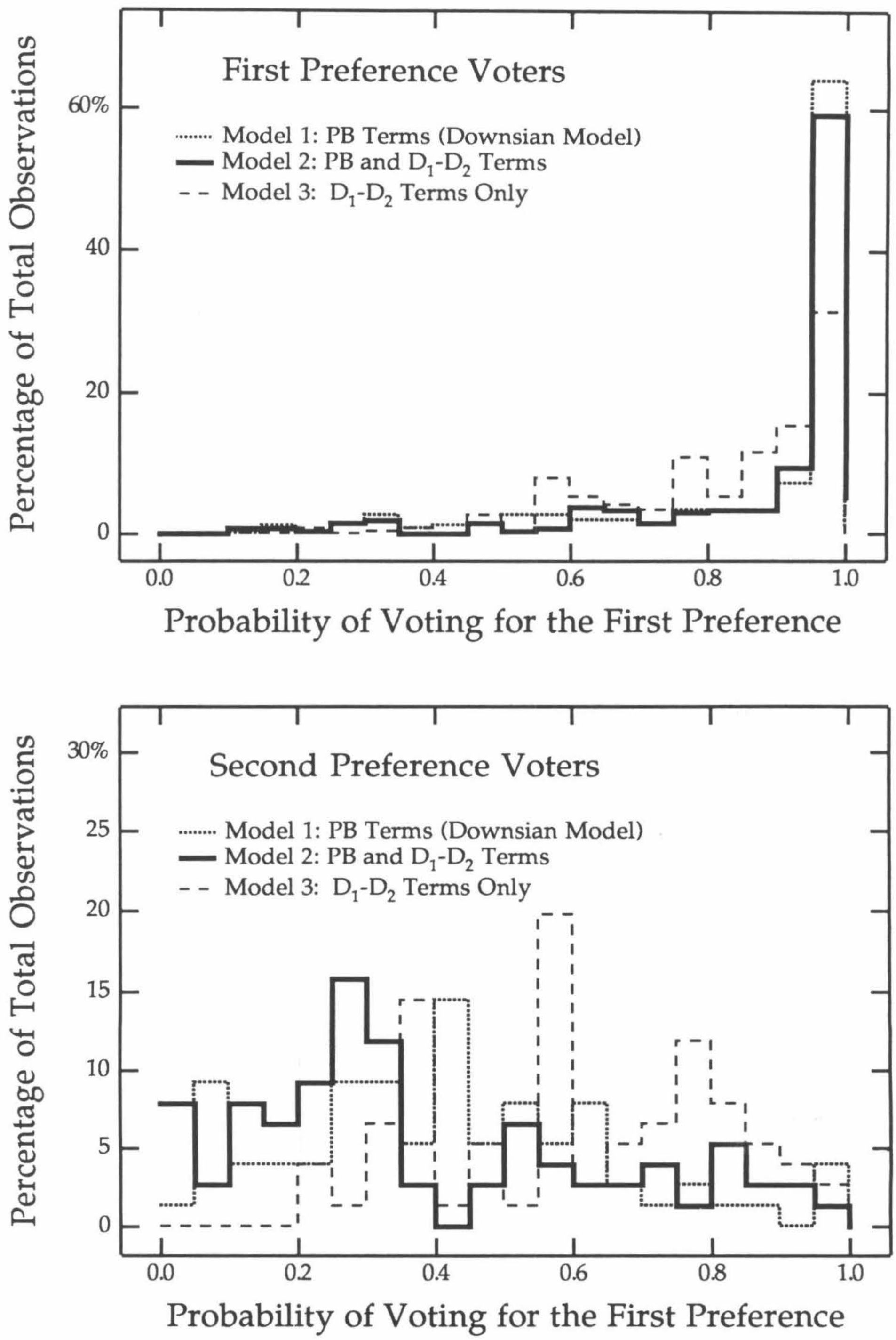
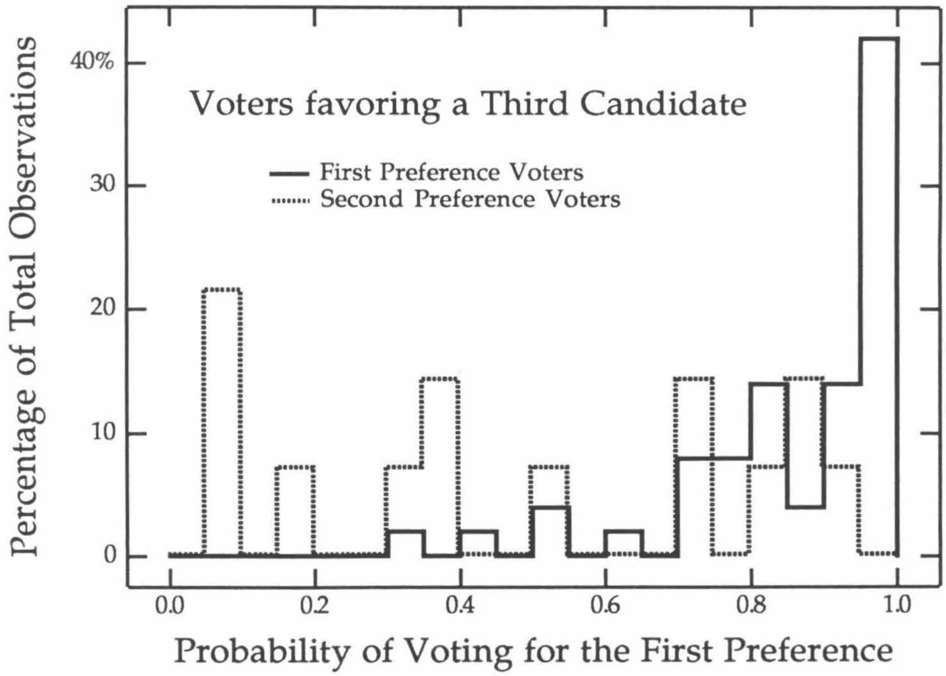


Figure 2.3-1

Predictability of Voters Favoring a Third Candidate in Comparison with Others: the 1968 Data



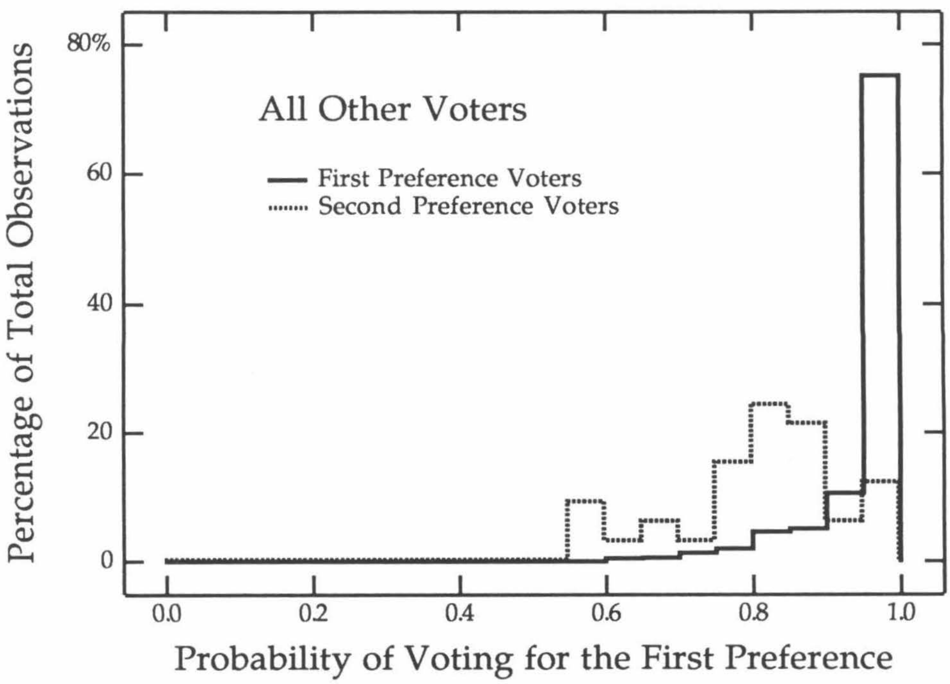
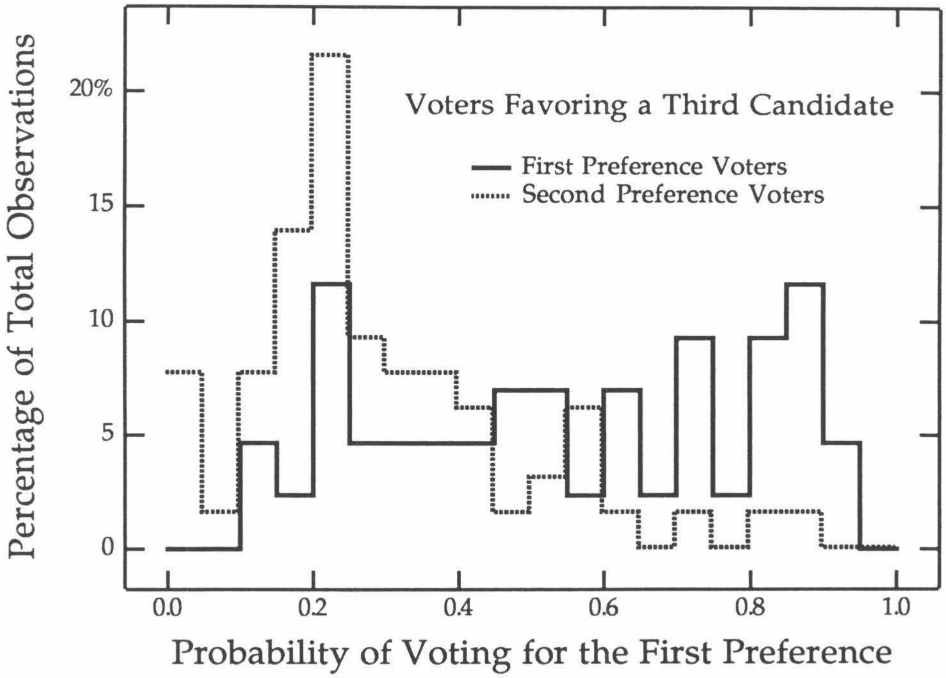


Figure 2.3-2

Predictability of Voters Favoring a Third Candidate in Comparison with Others: the 1980 Data



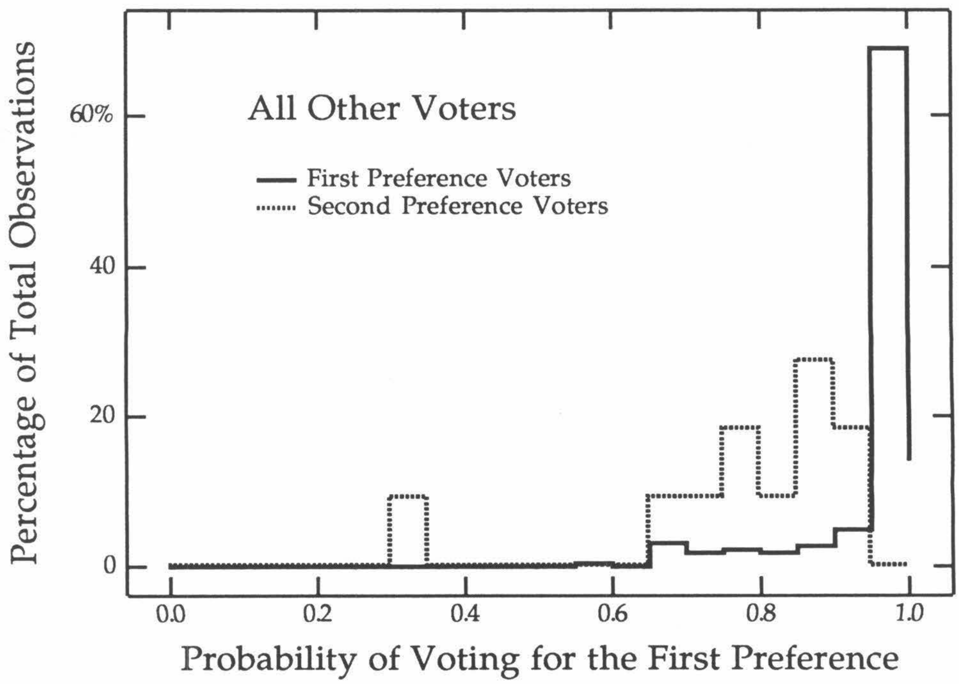


Figure 2.4

Predictability of Turnout for Better vs. Less Informed Electorates: the 1980 Data

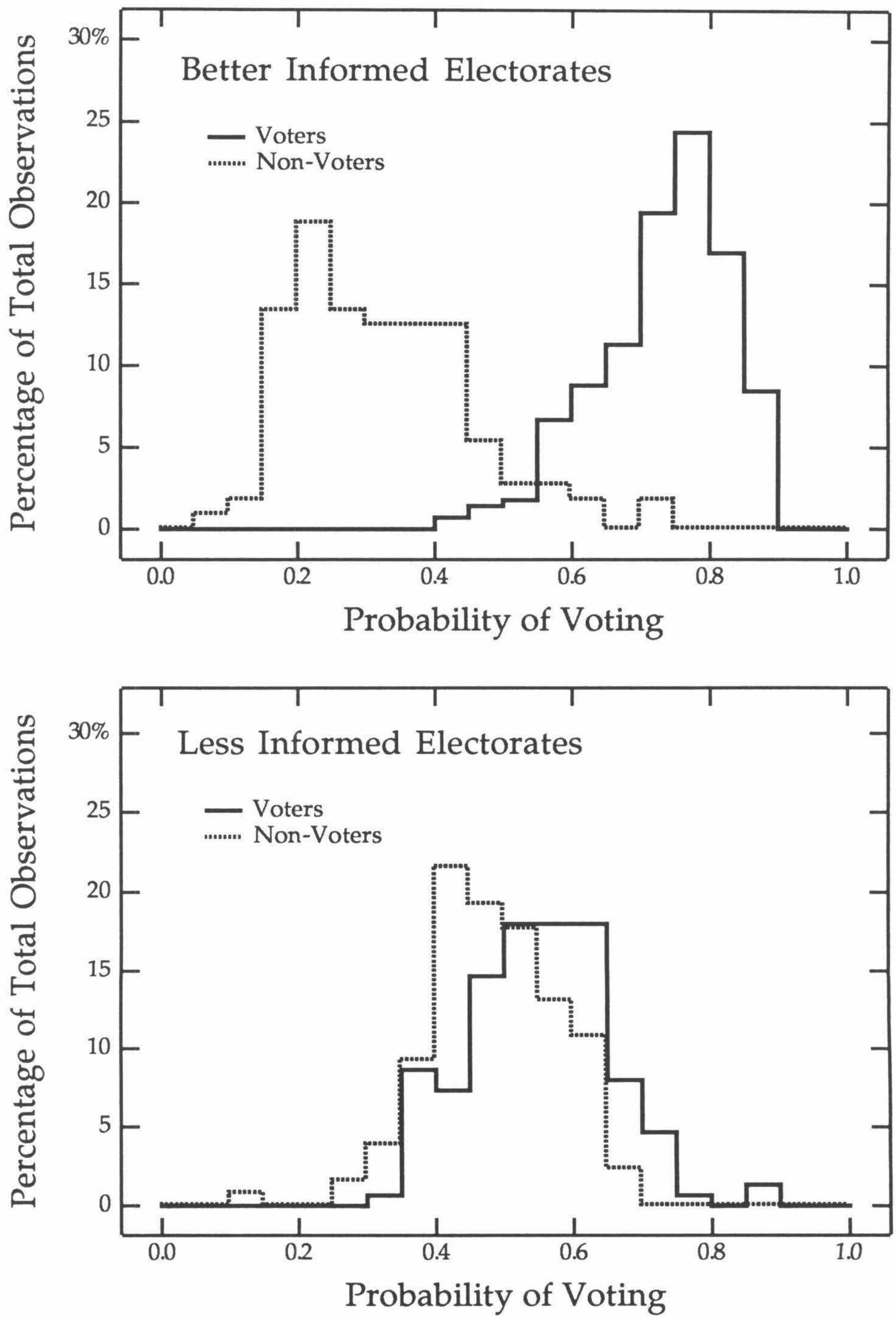
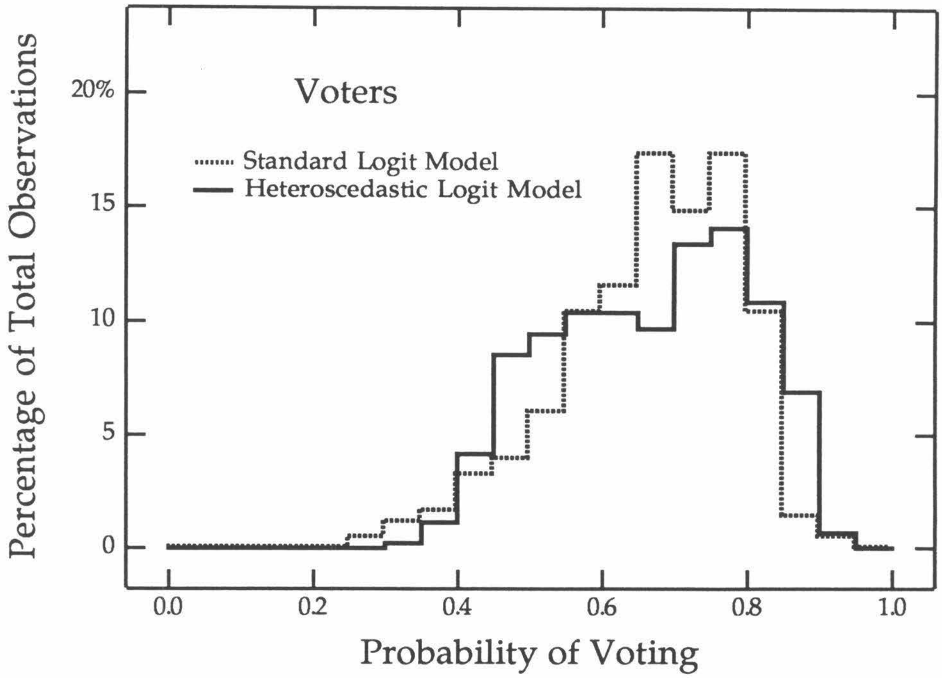
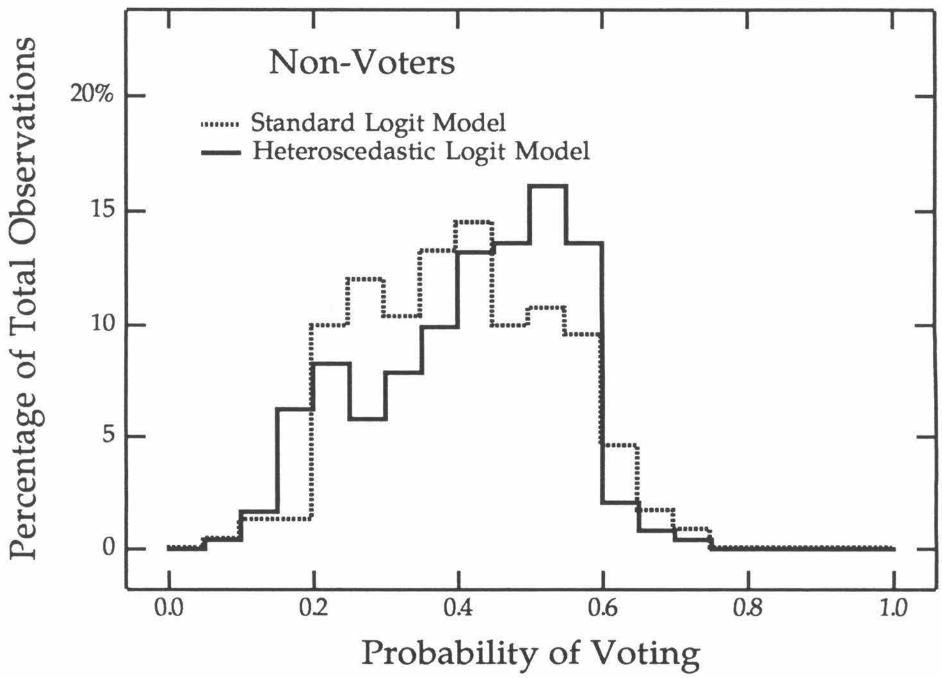


Figure 2.5

Performance of the Standard Logit and the Heteroscedastic Logit Voting Turnout Models: the 1980 Data





Chapter 3

Congressional Career Decisions

3.1 Introduction

This chapter studies choice behavior in congressional career decisions. Except the few congressmen who died in office or were expelled or were appointed to other offices, at the end of each term most members of the House face a career choice situation: they can choose to run for reelection, or to retire, or to seek other office (typically governor or U.S. senator) if an opportunity exists. A congressman's career choice decision reveals, of course, his motivation and direction of ambition, which "lie at the heart of politics" (Schlesinger, 1966, p.1). Moreover, it is how such choices are made that increasingly determines the overall level of House turnover and the composition of the House membership, which in turn "relate to partisan politics, to congressional reform, to basic indicators of congressional vitality and development stage, and to the nature of public policy and changes in that policy"

(Moore and Hibbing 1991, p.6). The choice among the three options "increasingly" determines the turnover level and membership composition because the House seat is becoming safer and the probability of defeat in reelection is decreasing (e.g., Bauer and Hibbing 1989). Figure 3.1 shows the percentage of members that were defeated, that retired, and that sought higher office (governor or senator) in each of the 20 congresses from 1947 to 1986. Indeed, while the rate of seeking higher office has been slightly increasing, and retirement rate has experienced great changes, the general trend of the defeat rate in reelections has been going down.

Many scholars recognize the importance of studying congressional career decisions. Rohde (1979) and Brace (1984) address progressive ambition of House members and study how they make the decision to run for higher office. A relatively large body of literature, partly triggered by the big shifts in retirement rate in the 1970s and the 1980s, concerns voluntary retirement from the House (e.g., Brace 1985; Frantzich 1978; Hibbing 1982; Moore and Hibbing 1991). A recent study by Schansberg (1992) attempts to model the decisions in both directions. This research identified many variables that are potentially responsible for congressional career decisions, and shed light on the substantive meanings of and roles played by these variables. It is, however, far from complete or satisfactory. With the exception of Schansberg (1992), previous research is mostly involved with the study of binary choices. Rohde and Brace study choices between seeking reelection and seeking higher office, with members who decide to retire excluded from the sample. The literature on retirement analyzes choices between seeking reelection and retiring, with members who decide to run for higher office excluded from the sample. In effect, these studies were not analyzing the unconditional probabilities of members choosing to retire, to seek reelection, or to run for higher office, but rather the conditional probability of choosing retirement over seeking reelection *given* that they have not otherwise chosen to run for higher office, or the conditional probability of seeking higher office versus remaining in the House *given* that the member has not otherwise chosen to retire. Conditional probabilities only show us an incomplete picture of the choice situation. We learn how a congressman chooses from a *subset* of all possible alternatives, but nothing about how the subset is chosen. Moreover, many authors are not aware of this fact. The subsample used is often taken as the whole sample, and variables affecting the conditional probabilities are interpreted as affecting the unconditional probabilities. In Rohde (1979), for example, members who "announce their retirement at the end of a term are not considered to have had an opportunity to run for higher office" (p.13). In fact about two thirds of them do have an opportunity, and they *choose* not to run for it. Simply excluding them is not justified, and may also lead to failure of identifying relevant variables that enter the choice calculation. For example, previous studies of voluntary retirement have not considered such variables as opportunity for higher offices as possible sources that affect the probability of retiring from the House. It may be the case that these variables do not affect the *conditional* probability of retiring versus seeking reelection, given that the option of seeking higher office has not been chosen. However, to the extent they play an important role in members' decision to seek higher office or not, they affect the probability of choosing the sub choice set { *retire, reelection*}, and can thus affect the probability of retiring.

Schansberg (1992) avoids the above problems and considers the full range of choices. Unfortunately, this study runs into another problem of sample treatment: it treats *all* members as having all three options, and models how a member chooses to seek higher office even if he does not have an opportunity to do so. This causes serious problems in the properties of the estimators and the substantive interpretation of the results, because more than one third of the members do not have an opportunity for higher office. A figurative summary of the problems in sample treatment by previous research is presented in Figure 3.2.

In this chapter, we seek to overcome the weaknesses in previous research and to study congressional career decisions in a rigorous fashion. Using data from the twenty congresses spanning 1947 to 1986, we examine three possible career options: retiring (including leaving for unimportant offices such as small city mayor), seeking reelection, and seeking higher office (governor or senator). All members of the 20 congresses are included in the sample except those died in office, expelled, or appointed to another office — those who do not really have a choice.[1] The chapter unfolds as follows: Section 2 concerns the theoretical aspects

[1] **Also excluded are the few who run for offices other than governor or Senator *that are of certain* significance (e.g., vice president, state-level supreme court judge). Decisions to run for such offices have their own peculiar nature and the sample is too small for systematic study.**

of the analysis. We discuss rational choice behavior in the context of congressional career decisions, setting up a framework in which variables that enter the decision making can be identified. We then proceed to the discussion of the choice model and estimation strategy. Section 3 presents the empirical analysis. We begin with the specification of relevant variables based on previous research. We then present the estimation results of the choice model and discuss their substantive meanings and implications. Finally, Section 4 offers concluding remarks, and identifies problems deserving of future research.

3.2 A Model of Congressional Career Decisions

We begin by assuming that congressmen are rational in the sense of expected utility maximization. Consider the choice situation a congressman faces at the end of each term. If he has an opportunity to run for a higher office, his choice set is $J = \{a_1, a_2, a_3\}$, where $a_1 =$ retire, $a_2 =$ run for reelection to the house, and $a_3 =$ run for a higher office.[2] If he does not have an opportunity for higher office, he chooses between a_1 and a_2. Under the rationality assumption, a representative will choose option a_j if and only if the expected utility of a_j is greater than that of all other available options. The following analysis considers members with all three options. The analysis for the other members can be carried out in the same

[2]As in Rohde (1979), a House member is considered to have an opportunity for higher office if an election is held in his state for a Senate seat or the governorship, and there is not an own party incumbent seeking reelection to the office in question. Also, odd-year governorships are excluded because it is not of an exclusive nature.

way. A graphical illustration of the choice situation is presented in Figure 3.3. If the congressman chooses to retire, he retires with probability $P_1 = 1$, receiving the utility of retirement U_1; if he chooses to run for reelection, he wins the reelection with probability P_2, receiving utility of the House seat U_2 minus the cost of running for the election C_2, and loses (hence retires) with probability $1-P_2$, receiving U_1-C_2; if he chooses to run for higher office, he wins with probability P_3, receiving the utility of the higher office seat U_3 minus the cost of seeking the seat C_3, and loses (hence retires) with probability $1-P_3$, receiving U_1-C_3. Let E_j denote the expected utility of choosing option a_j, then,

$$\begin{aligned}
E_1 &= U_1 \\
E_2 &= P_2U_2 + (1-P_2)U_1 - C_2 \\
&= U_1 + P_2(U_2 - U_1) - C_2 \qquad (3.1) \\
E_3 &= P_3U_3 + (1-P_3)U_1 - C_3 \\
&= U_1 + P_3(U_3 - U_1) - C_3.
\end{aligned}$$

As any affine transformation of an expected utility function is also an expected utility function (Varian 1984), subtracting U_1 from all E_j, we can rewrite (3.1) as:

$$\begin{aligned}
E_1 &= 0 \\
E_2 &= P_2(U_2 - U_1) - C_2 \qquad (3.2)
\end{aligned}$$

$$E_3 = P_3(U_3 - U_1) - C_3.$$

Expressions (3.2) states that if we normalize the expected utility of "retire" as zero, then the expected utility of "run for reelection" depends on the probability of winning the reelection, the value of the House seat *relative* to the value of retirement, and the cost of seeking reelection; and the expected utility of seeking higher office depends on the probability of winning the higher office, the value of the higher office *relative* to the value of retirement, and the cost of seeking the higher office. An important difference between our specification and that in Rohde (1979) is that in ours both the expected utilities of seeking reelection and seeking higher office depend on the value of retirement, while in Rohde (1979) they do not. Because we do not have specific data on the value of retirement, U_1 is implicitly expressed in terms of the expected utilities of the House seat and higher office seat. As a consequence, E_2 and E_3 are correlated because they both depend on the value of retirement. Operationally, this implies that no variable that enters E_2 should be excluded *a priori* from E_3, or vice versa. We will discuss the identification of variables entering E's later.

Given the expected utilities of the options, the congressman then chooses the option with the greatest E. Had we known the deterministic values of E_j's, we could easily predict with certainty which option the congressman will choose. Unfortunately, in empirical research we can never know the exact values of E's, which are functions of many variables only part of which are observable or are measurable. The empirical modeling of E's hence always includes errors, and the prediction of the choice behavior can only be in probability terms. Therefore our next task is to select a choice model which gives the probabilities of choosing to retire, to seek reelection, or to seek higher office based on the empirical values of E's.

It may seem likely that the career choices have a two-level tree structure. For example, the congressman may first decide whether to run for higher office or not (the first level), if not, then he chooses between seeking reelection and retiring (second level). If "seeking reelection" and "retiring" are considered to be "similar" alternatives, then a model that specifically allows for correlation between the error terms of the two alternatives, such as a nested logit model, may be appropriate. We estimated such models, and found, through hypothesis testings, that the tree structure with {retire, run for reelection} as the second-level sub-choice set is more likely the "true" structure than the other two possibilities. However, we do not get meaningful estimates for the similarity parameter in this model (it is not in the admissible range (0,1] for the model to be proper probability choice model). There are two possibilities that could lead to this situation. First, it may be the case that congressional career decisions cannot be described by a tree structure. Second, it may be the case that the decisions do have a tree structure, but the marginal probabilities in the sub-choice set (the probability of seeking reelection and the probability of retiring) are extremely skewed, causing the similarity parameter to fall out of the admissible range (Börsch-Supan 1990).

To avoid making any ungrounded assumptions about the nature of the errors in E's, then, we will adopt a "mother logit" model, which is shown by McFadden (1975) to be extremely general and can represent *any* choice probabilities under the correct specification of the utility functions. As formulated by Train (1986), let $P^*_{in} = f(z_{in}; z_{jn}, \forall j \neq i; s_n)$ be the "true" probability that decision-maker n chooses alternative i, where z_{in} is the observed data relating to alternative i as faced by decision-maker n, z_{jn} is the observed data relating to alternatives other than i, and s_n is a vector of characteristics associated with the decision-maker. There is no restriction on the functional form of f so this specification is completely general. All choice models are special cases of this model, including the standard logit and probit models.

Now define $W_{in} = \log P^*_{in}$ and evaluate the logit probabilities based on W_{in}:

$$P_{in} = \frac{e^{W_{in}}}{\sum_j e^{W_{jn}}} = \frac{e^{\log P^*_{in}}}{\sum_j e^{\log P^*_{jn}}} = P^*_{in}, \tag{3.3}$$

where the last equality is due to the fact that choice probabilities necessarily sum to one.

Equation (3.3) shows that *any* choice probability P^*_{in} can be expressed in the logit probability form, with W_{in} being the deterministic utility of alternative i to decision-maker n.[3] Note that in general $W_{in} = \log P^*_{in}$ depends not only on observed data

[3]Recall that in a logit specification like (3.3) the underlying (expected) utility for alternative i is

relating to alternative i, but also on data relating to all other alternatives. Hence the mother logit model is not subject to the IIA (Independence from Irrelevant Alternatives) property as a standard logit model does. The IIA property holds if the ratio of the choice probabilities of any two alternatives is not affected by the presence of other alternatives, which is unlikely the case in our congressional career choice situation.

The next question is, of course, the appropriate specification of W_i,[4] the utility functions of model (3.3). Once we obtain W_i we can estimate a logit-form model straightforwardly. As noted above, $W_{in} = \log P^*_{in}$ depends on (expected) utilities of all alternatives. Hence all variables that enter the utility function of any alternative may enter each W_{in}. From the expected utility functions (3.2), we have

$$W_i = g_i(U_1, U_2, U_3, P_2, P_3, C_2, C_3).$$

Let X denote the vector of variables that appear in g_i, and assume, for simplicity, that g_i's are linear, then

$$W_i = \beta_i X,$$

where β_i is a vector of coefficients. The choice probabilities are therefore given by

$$P_i = \frac{e^{\beta_i X}}{\sum_j e^{\beta_j X}}, \tag{3.4}$$

assumed to be $W_{in} + \epsilon_{in}$, where ϵ_{in} is the stochastic part of the total utility of alternative i.

[4]For simplicity of notations, we hereafter drop the subscript n in W_{in}.

where the index j runs from 1 to 2 if the choice set contains two alternatives, and from 1 to 3 if the choice set contains three alternatives.[5]

Let $c_{in} = 1$ if individual n chooses alternative i, and $c_{in} = 0$ otherwise, then the log-likelihood of observation n is:

$$l_n(\beta) = \sum_{i=1}^{I} c_{in} \ln p_{in},$$

where β denotes coefficients in p_{in} to be estimated; $I = 2$ if individual n has two alternatives, and $I = 3$ if individual n has three alternatives; p_{in} is given by (3.4). The total sample log-likelihood is then given by

$$L(\beta) = \sum_{n=1}^{N} l_n(\beta).$$

Maximum likelihood estimates of β can be obtained by maximizing $L(\beta)$ with respect to β. The computation can be carried out using a standard statistical package.

As only utility *differences* matter, from (3.4) we can write (taking the case of three choices as an example):

$$P_1 = \frac{e^{(\beta_1-\beta_2)X}}{e^{(\beta_1-\beta_2)X} + 1 + e^{(\beta_3-\beta_2)X}}$$

[5]The reader can see that we assume the functional form of W_1 and W_2 are the same for members with three choices and those with two choices. We have estimated a model in which the utility functions are allowed to have different variables and coefficients for the two groups of House members, but tests show that the two models are statistically indistinguishable.

or

$$P_1 = \frac{1}{1 + e^{(\beta_2-\beta_1)X} + e^{(\beta_3-\beta_1)X}}.$$

P_2 and P_3 can be expressed similarly. Estimation of the model will give $\beta_2 - \beta_1$ and $\beta_3 - \beta_1$, if the first alternative is taken as the base of normalization; or it gives $\beta_1 - \beta_2$ and $\beta_3 - \beta_2$, if the second alternative is taken as the base of normalization. The case is similar if the third alternative serves as the base. To normalize the utilities using different bases can often facilitate the interpreting of estimation results, as $\beta_i - \beta_j$ affects the conditional probability of choosing alternative i over j given the third alternative is not chosen, or the unconditional probabilities if there are only two alternatives. Take the probability of choosing the first alternative as an example. Suppose we use the second alternative as the base of normalization (which we do in this study in order to facilitate comparison of empirical results with previous research), if there are two alternatives,

$$P_1 = \frac{e^{\beta_1 X}}{\sum_{j=1}^{2} e^{\beta_j X}} = \frac{e^{(\beta_1-\beta_2)X}}{e^{(\beta_1-\beta_2)X} + 1}$$

and it is easy to verify that

$$\frac{\partial P_1}{\partial X} = P_1 P_2(\beta_1 - \beta_2) \tag{3.5}$$

which are signed by $\beta_1 - \beta_2$. Hence we can infer the direction of the effects on P_1 by an independent variable x directly from the sign of the estimated coefficient of

x. If there are three alternatives, then

$$P_1 = \frac{e^{\beta_1 X}}{\sum_{j=1}^{3} e^{\beta_j X}} = \frac{e^{\beta_1 X}}{\sum_{j=1}^{2} e^{\beta_j X}} \frac{\sum_{j=1}^{2} e^{\beta_j X}}{\sum_{j=1}^{3} e^{\beta_j X}},$$

where the second fraction is the probability of choosing the sub choice set $\{1,2\}$, and the first fraction is the conditional probability of choosing alternative 1 over alternative 2 given that the subset $\{1,2\}$ is chosen. This conditional probability has the same functional form as the unconditional probability in the 2-choice case, and a similar analysis will show that its partial derivatives are signed by $(\beta_1 - \beta_2)$. In the empirical analysis below we will report the estimates of $(\beta_1 - \beta_2)$, $(\beta_3 - \beta_2)$ and $(\beta_3 - \beta_1)$ to facilitate pairwise comparisons.

3.3 Empirical Analysis

We now turn to the empirical analysis of congressional career decisions in the framework of the model developed above. Our first task, of course, is to identify variables that enter X, the vector involving $U_1, U_2, U_3, P_2, P_3, C_2$, and C_3. That is, we need to specify variables that reflect the values of retirement, of the House seat, and of the higher office seat; the probability of winning the House reelection, and of winning the higher office seat; and the cost of seeking reelection or seeking higher office. Based on theoretical considerations and on results from previous research, we include the following variables:

(1) Age. Age plays an important role in members' evaluation of the career options. The elderly will derive less utility from serving in the House or higher office, for the jobs are mentally, emotionally and physically demanding. It is therefore not surprising that previous studies, without exception, observe that the rate of voluntary retirement increases with age. In comparing the House seat with higher office seat, however, we would expect age to be inversely related to the propensity to seek the latter option, because "the older a politician is the less chance he has for promotion and the less likely he is to harbor ambitions to advance" (Hain 1974, p.265).

(2) Republican. A dummy variable that registers whether the member belongs to the minority party, which, for the period covered by this study, has almost always meant the Republican. Many scholars realize the disutility the minority party status brings to the member. The commonly cited reasons include the ideological frustration felt by the Republicans as they watch the Democratic majority preside over the growth of the federal government, and the fact that Republicans are blocked from committee chairmanships (Hibbing 1982). Being a Republican thus is expected to encourage a member to leave the House in both ways: retiring or seeking higher office (particularly seeking higher office, according to both Gilmour and Rothstein 1991 and Schansberg 1992).

(3) Chair/Leader. A dummy variable that takes the value 1 if the member is the chair of a standing committee or a majority party leader, 0 otherwise. This variable

is meant to reflect that the House seat is particularly valuable for the member, and should therefore encourage him to stay. Previous researchers often use the seniority rank of the member in the House to indicate the value of his House seat. There are two problems associated with this. First, the seniority rank is very highly correlated with age, causing difficulties in interpreting the estimated parameters. Second, it is not really seniority rank *per se* but formal positions held that captures the value of the House seat, and the correlation between the two is decreasing in the modern Congress.

(4) Ideol.Pos.Dem and Ideol.Pos.Rep. A pair of variables that denote the member's ideological position within his party caucus, calculated separately for Democrat and Republican members. The value of the variable is the member's percentile rank on an underlying liberal-conservative dimension derived by scaling their roll-call votes with Poole and Rosenthal (1985)'s NOMINATE procedure. The higher the percentile ranking, the more conservative the member and the more likely he is expected to retire (Frantzich 1978, Moore and Hibbing 1991).

(5) Institutional Reform. A dummy variable taking the value 1 if the member is elected before the 91st congress and serves in the four subsequent ones, and 0 otherwise. This variable is meant to reflect the effects of the institutional changes in or around 1970 that are believed by many authors to diminish the relative value of the House seat. Such changes include improved retirement pensions, requirement of financial disclosure, declined seniority system, to name a few (Brace 1985,

Hibbing 1982). The effects should be small for members entering the congress after the changes, and should decrease over time, as partly evidenced by the decreased aggregate level of retirement in the 1980's. We expect that members seriously affected by the changes (having value 1 for the dummy variable) will more likely to either retire from the House or to seek higher office.

(6) Previous Margin. The vote percentage received by the member in the previous election, which should serve as an indicator of the member's safeness in the next election, and which is therefore expected to enter into P_2 and C_2. However, as observed by many scholars, the House seat is becoming safer and safer, and defeat rate is steadily decreasing over the years. Hence this variable may not be playing a significant role as it seems on the surface. This is indeed the case in our findings.

(7) Scandal. A dummy variable taking the value 1 if the member is charged with scandals, 0 otherwise. The reason to include this variable is obvious. Charge of scandal reduces the probability of the incumbent winning the election, and increases the cost of campaign (Peters 1980). Furthermore, in the face of decreasing electoral competition, it is one of the only few sources that determine the member's fate in the election. While in our data no member charged with scandals ever tries to run for higher office, we expect such members are also discouraged for reelection and are therefore more likely to retire.

(8) Redistricting. A dummy variable taking the value 1 if the member is seriously hurt by redistricting — if he has to run against another incumbent or if he has

to appeal to a completely new constituency — and 0 otherwise. The effects of redistricting are widely recognized (Bullock 1972, Fiorina et al. 1975, Frantzich 1978), and Brace (1984, 1985) finds empirical evidence that members hurt by redistricting are more likely to retire or to seek higher office.

(9) Constituency Overlap. This variable is the reciprocal of the number of districts in the member's state. Rohde (1979), Brace (1984) and Schansberg (1992) all find that members from smaller states are more likely to run for higher office, because the overlap of the constituencies makes campaign easier and the race less risky. In estimating the choice model we use the logarithm of this variable due to the expectation that the effects of this variable are increasingly strong when the size of the state decreases.

(10) Senate Election. A dummy variable taking the value 1 if there is a Senate election, 0 otherwise, is meant to reflect the value of the higher office seat. Rohde (1979) and Brace (1984) find that a House member values a Senate seat more than a governor seat, because the former has longer terms and is electorally safer. So the opportunity for the Senate would encourage a member to seek higher office.

(11) Open Seat–Senate. A dummy variable taking the value 1 if there is a Senate election and the seat is not sought by an incumbent, 0 otherwise. This variable obviously affects the probability of winning the Senate seat. Without exception, previous research reports that an open seat opportunity greatly encourages a House member's propensity to run for higher office (Rohde 1979, Brace 1984, Schansberg

1992).

(12) Open Seat–Governor. Similar to Open Seat–Senate.

Table 3.1 reports the mean values of the variables for members retiring, seeking reelection, and seeking higher office, respectively. The entries provide a rough idea of how these variables affect the career decisions. They may be misleading however, because the effects of other variables are not controlled. Keep this in mind and we are safe for an initial investigation. Age, as expected, encourages retirement but discourages higher office seeking. Minority party members are more likely to retire or seek higher office (the latter turns out to be spurious). Standing committee chairs or majority party leaders are more likely to retire — this is obviously due to the effect of age, as such members are normally old. Conservative members of the Democratic party tend to retire more than the liberals, while the latter are more likely to seek higher office (the effects are less obvious for the Republicans). Institutional reform apparently does diminish the value of the House seat. The previous vote margin of a House member is typically large, but it does not seem to figure much into the career decisions. The effects of scandal charge is in the expected direction, as are the effects of the redistricting variable. The last four variables, as expected, all encourage progressive ambition, although they do not seem to affect the choice between staying in the House and retiring.

Table 3.2 reports the estimation results of the choice model, from which we can learn the independent effects of the variables. Recall from the discussion in Section

2 that we can infer the following from the estimated coefficients: for members without an opportunity for higher office, the directions of impacts on choice probabilities by independent variables; for members with an option to run for higher office, the directions of impacts on *conditional* probabilities by independent variables. The first two columns of the table give $\beta_1 - \beta_2$ and their t-statistics, from which we see that in choosing between seeking reelection and retiring, the effects of age, minority party status, ideological location for members of the Democratic party, institutional reform, scandal charge, and redistricting are all statistically significant and in the expected direction. We should point out that the effects of the ideological location of the Democrats are partly generated by the fact that Democratic members from the South are more conservative than the rest and have a greater propensity to retire. When the model is reestimated with Southern Democrats excluded, the ideology variable for the Democrats is no longer statistically significant at the standard significance level (the t-statistic drops to 1.76). The coefficient of all other variables in the model are not statistically different from zero. While this may not be surprising for the variables that mainly reflect the value of the higher office seat (Senate Election) or the probability of winning the seat (Const. Overlap, Open Seat–Senate, and Open Seat–Governor), the finding that a standing committee chair or a majority leader is no less likely to retire than other members, everything else equal, is not at all expected. This implies that formal positions held will not deter a member from retreating from politics, if other factors amount to favoring the

option. Nor do they deter members from running for higher office (column 3 and 4). Previous vote margin does not figure into the career decisions, either. This confirms Bauer and Hibbing (1989)'s conclusion that the level of competition in congressional elections has declined. Another variable that has no effect on career decisions is the ideological location of Republican members, given the effect of the minority party status and other variables controlled.

Column 3 and 4 of Table 3.2 report $\beta_3 - \beta_2$ and their t statistics. The effects of age, institutional reform, degree of constituency overlap, opportunity for Senate seat and probability of winning higher office seat are as expected in choosing between seeking reelection and seeking higher office. There are two findings, however, that are quite novel. One is that being a Republican *per se* does not encourage higher office seeking, contrary to Gilmour and Rothstein (1991) and Schansberg (1992)'s statements. This is probably due to the fact that while Schansberg's results are problematic due to the incorrect sample treatment, Gilmour and Rothstein reach the conclusion by examining aggregate data, without controlling for other variables like age and the opportunity for higher office seat. Indeed, our data show that the higher aggregate level of higher office seeking by Republicans, observed by Gilmour and Rothstein for the period of 1972-1988 in their study, is due to the historical fact that during this period Republicans were on average younger, had better opportunities, were from smaller states, and were hurt more by redistricting. Another finding that has not been reported in previous research is that liberal members of the Democratic party have a greater propensity of seeking higher office seat. To the extent that the House serves as a stepping stone to the Senate, this helps to explain the ideological polarization, i.e., the greater disparity in the voting records of Senate Democrats and Republicans, described in Poole and Rosenthal (1984).

To complete the pairwise comparisons, the last two columns of Table 3.2 report $\beta_3 - \beta_1$ and their t-statistics, obtained from estimating the model using the first alternative "retiring" as the base of normalization. Age, as expected, strongly discourages higher office seeking in comparison with retirement. Minority party members are more likely to retire than seeking higher office — again contrary to conclusions from previous research. Conservative members of the Democratic party (e.g., the Southern Democrats) are also more likely to retire, compared with either seeking reelection or seeking higher office. The last four variables, reflecting the value of the higher office seat or the probability/cost of winning it, all strongly encourage higher office seeking. All other variables, however, do not significantly affect the choice between retiring and seeking higher office.[6]

From Table 3.2 we can only infer to which direction the independent variables affect choice probabilities in pair-wise comparisons. To measure the extent of the effects or, for members with three choices, to infer even the directions of the effects on the unconditional probabilities will require other measures. For our purpose we

[6]Scandal charge do not have significant coefficients when comparing seeking higher office with other alternatives, due to the fact that no members charged with scandals have tried to run for higher office, therefore the variances of the coefficients blow up.

report in Table 3.3 the average elasticities of the choice probabilities with respect to continuous independent variables, $\partial \log P / \partial \log x = (\partial P/\partial x) \ / \ (P/x)$, and the average percentage change in choice probabilities with respect to the change in a dummy variable from zero to one, holding other variables constant.[7] Calculation of entries in Table 3.3 is based on re-estimation of the model with insignificant variables excluded (such that the entries are statistically non-zero). From Table 3.3 we see that the effects on the probability of retiring are greatest by scandal charge and age, next are minority party status, redistricting, ideological position of the Democratics, and institutional reform.[8] We also see that for members with opportunity for higher office, the set of variables that primarily affect the value of higher office seat or the probability of winning it (*Constit. Overlap, Senate Election, Open Seat-Senate, and Open Seat-Governor*) all decrease the probability of retiring. To see why, recall that

$$P_1 = \frac{e^{(\beta_1-\beta_2)X}}{e^{(\beta_1-\beta_2)X} + 1 + e^{(\beta_3-\beta_2)X}}. \tag{3.6}$$

Therefore if a variable affects either the utility difference between retiring and seeking reelection or the utility difference between seeking higher office and seeking reelection, it will affect the unconditional probability of retiring from the House.

[7]For *Senate Election*, the change is from *Senate Election* = 0 and *Open Seat-Senate* = 0 to *Senate Election* = 1 and *Open Seat-Senate* = real sample value; For *Open Seat-Senate*, the change is from *Senate Election* = 1 and *Open Seat-Senate* = 0 to *Senate Election* = 1 and *Open Seat-Senate* = 1.

[8]Bear in mind, however, that the entries in Table 3.3 are samples of random variables and hence the conclusions are not in the exact sense.

The opportunity variables do not significantly affect the utility differences between retirement and remaining in the House, but they do all act to increase the utility difference between seeking higher office and remaining in the House, thus increasing the denominator of (3.6) and reducing the probability of choosing retirement.

Other entries in Table 3.3 can be interpreted in a similar fashion. We see that redistricting, the existence of a Senate election, age, and a Senate open seat all have great impact on the probability of choosing to run for higher office. The effects of other variables are relatively smaller. For the probability of choosing to run for reelection, the effect of age is smaller for members with the choice of seeking higher office. The reason is that while age encourages retirement, it discourages higher office seeking, hence reduces the net effect on the probability of remaining. The effect of redistricting is stronger for such members because it encourages leaving the House in both ways.

In the introduction section we indicated the importance of individual career decision making in determining the long term membership composition of the House. Indeed, the effects are cumulating. Figure 3.4 shows that as a result of the Republicans being more likely to retire, and as a whole also more likely to seek higher office due to reasons discussed earlier, the percentage of Republicans in the House is decreasing with terms served. Similarly, because members from smaller states (greater constituency overlap) are more likely to leave the House for Higher office,[9] such members are few among seniors. This means that districts in small states or that have elected a Republican are more likely to be represented by junior members lacking experience and connections. That fewer Republicans ever build up their seniority also explains, from another angle, the Democratic dominance of the House.

3.4 Conclusion

This chapter formulates and estimates an integrated model of congressional career decisions in a rigorous fashion, and examines the impact of factors that enter the decision making on both the conditional and unconditional choice probabilities. While some of our findings confirm previous research or conform to expectations, others offer fresh insight into the nature of the decision making process. Notably, we find that formal positions held and previous vote margins do not figure into House members' career decisions, and being a Republican *per se* does not encourage progressive ambition. We also see that a number of factors previously identified as predisposing House members to seek higher office also affect retirement decision. Our analysis also reveals, for the first time, the direction and extent of the effects of independent variables on the unconditional probabilities of choosing the congressional career options. Even more important, however, is what such an analysis tells us about those who choose to stay. Given that it has become rare for House in-

[9]They are also less likely to retire, but the effects is much smaller compared with the propensity to seek high office, so that the net effect on the decision to stay is negative. Refer to Table 3.3.

cumbents to fail in their reelection bids, it is the joint effect of progressive ambition and voluntary retirement that has come to increasingly determine the composition of the House of Representatives, which in turn bears on the policy outcomes of the Congress.

In closing, we shall point out a potential problem deserving of future research. Because our data are panel data — each member is repeatedly sampled until he leaves the House — auto-correlation of errors in the expected utility functions is likely present. The fact that the cross-sectional sample size is far bigger than the number of repetitions (the data set includes nearly 2,000 different individuals who on average serve only about 5 terms) may lighten the harm of the problem, but may not eliminate it. Although in theory the mother logit model employed in this study applies to error structures of general patterns, we expect that the performance of the model can be improved if the information on the error structure is explicitly incorporated. We leave the search of a suitable model that has good theoretical property and is computationally feasible for such data to future research.

References

Bauer, M. and Hibbing, J. 1989. "Which Incumbents Lose in House Elections: A Response to Jacobson's "The Margins Never Vanished." *American Journal of Political Science*, 33(1):263-271.

Bösch-Supan,A.(1990). "On the Compatibility of NMNL Models With Utility Maximization." *Journal of Econometrics.* 43, 373-388. North-Holland.

Brace, P. 1984. "Progressive Ambition in the House: A Probabilistic Approach." *Journal of Politics*, 46:556-69.

Brace, P. 1985. "A Probabilistic Approach to Retirement from the U.S.Congress." *Legislative Studies Quarterly*, X, 1:107-123.

Bullock, C. 1972. "House Careerist: Changing Patterns of Longevity and Attrition." *American Political Science Review*, 66:1295-1300.

Fiorina, M. et al. 1975. "Historical Change in House Turnover." In *Congress in Change: Evolution & Reform*, Norman J. Ornstein, ed. New York: Praeger, 1975.

Frantzich, S. 1978. "Opting Out: Retirement from the House of Representatives." *American Political Science Quarterly*, 6:251-73.

Gilmour, J. and Rothstein, P. 1991. "Early Republican Retirement: A Cause of Democratic Dominance in the House of Representatives." Political Economy Working Paper, Washington University.

Hain, P. 1974. "Age, Ambitions, and Political Careers: The Middle-Age Crisis." *The Western Political Quarterly*, 27(2):265-274.

Hibbing, J. 1982. *Choosing to Leave: Voluntary Retirement from the U.S. House of Representatives.* Washington, D.C.: University Press of America.

Hibbing, J. 1991. "Contours of the Modern Congressional Career." *American Political Science Review,* 85:405-28.

Jacobson, G. 1987. "The Margins Never Vanished: Incumbency and Competition in elections to the U.S. House of Representatives, 1952-1982." *American Journal of Political Science,* 31:126-41.

McFadden, D. 1975."On Independence, Structure, And Simultaneity in Transportation Demand Analysis." Working Paper 7511, Urban Travel Demand Forecasting Project, Institute of Transportation and Traffic Engineering, University of California, Berkeley.

Moore, M. and Hibbing, J. 1991. "Is Serving in Congress Fun Again? Voluntary Retirements and Congressional Careers Through the 1980's." Manuscript, University of Nebraska.

Peters, J. 1980. "the Effects of Charges of Corruption on Voting Behavior in Congressional Elections." *American Political Science Review.* 74:697-708.

Poole, K. and Rosenthal, H. 1984. "The Polarization of American Politics."*Journal of American Politics,*46:1061-79.

Poole, K. and Rosenthal, H. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science,* 29:357-84.

Rohde, D. 1979. "Risk-Bearing and Progressive Ambition: The Case of Members of the United State House of Representatives." *American Journal of Political Science*, 23:1-26.

Schansberg, E. 1992. "Moving Out of the House: An Analysis of Congressional Quits." Paper presented at the Annual Meeting of the Public Choice Society, New Orleans, LA.

Schlesinger, J. A. 1966. *Ambition and Politics: Political Career in the United States.* Chicago, Rand McNally.

Train, K. 1986. *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand.* Cambridge, MA: The MIT Press.

Varian, H. 1984. *Microeconomic Analysis.* New York: W.W. Norton and Company.

Table 3.1

Mean Values of Factors Affecting Congressional Career Decisions: A Comparison of Members Retiring, Running for Reelection, or Seeking Higher Office

Variable	Retiring	Reelection	Higher Office
Age	59.6	51.4	46.6
Republican	0.47	0.41	0.50
Chair/Leader	0.09	0.05	0.01
Ideol. Pos., Dem.	0.61	0.50	0.43
Ideol. Pos., Rep.	0.54	0.50	0.49
Institutional Reform	0.21	0.13	0.16
Previous Margin	0.71	0.70	0.67
Scandal	0.03	0.01	0.00
Redistricting	0.03	0.01	0.04
Senate Election	0.31	0.32	0.76
Constit. Overlap	0.11	0.12	0.27
Open Seat — Senate	0.11	0.10	0.35
Open Seat — Governor	0.31	0.30	0.43

N = 8353

Table 3.2

Analysis of Congressional Career Decisions, 1947-1986

Estimation of the Choice Model

Variable	Retire vs. Run for Reelection		Seek Higher Office vs. Run for Reelection		Seek Higher Office vs. Retire	
	($\beta_1 - \beta_2$)	t-stat.	($\beta_3 - \beta_2$)	t-stat.	($\beta_3 - \beta_1$)	t-stat.
constant	-8.79	-14.8**	1.18	1.8	9.98	11.2**
Age	0.07	13.5**	-0.04	-5.6**	-0.12	-12.3**
Republican	0.94	3.9**	-0.48	-1.7	-1.43	-3.9**
Chair/Leader	-0.20	-1.0	-0.86	-1.2	-0.67	-0.9
Ideol. Pos., Dem.	1.46	5.1**	-1.22	-3.1**	-2.69	-5.6**
Ideol. Pos., Rep.	0.20	0.8	0.18	0.5	-0.02	-0.1
Institutional Reform	0.29	2.2*	0.41	2.1*	0.12	0.5
Previous Margin	-0.004	-1.1	-0.000	-0.1	0.003	0.6
Scandal	1.94	5.4**	—	—	—	—
Redistricting	0.87	2.6**	1.77	4.7**	0.93	1.9
Constit. Overlap	-0.06	-0.8	0.94	13.1**	1.00	10.2**
Senate Election	-0.07	-0.5	1.10	5.6**	1.17	5.0**
Open Seat — Senate	0.22	1.1	0.82	5.0**	0.60	2.4*
Open Seat — Governor	-0.07	-0.7	0.35	2.4*	0.43	2.3*

Log-Likelihood	initial:	-7785.6
	at converg.:	-2185.4
Number of Observations:		8353

** Significant at 0.01 level
* Significant at 0.05 level

Table 3.3
The Effects of Independent Variables on Choice Probabilities

Variable	With Three Choices			With Two Choices	
	P_1	P_2	P_3	P_1	P_2
Age	3.57	-0.08	-2.51	3.47	-0.21
Republican	1.71	-0.05	-0.05	1.69	-0.06
Ideol. Pos. Dem.	0.33	-0.01	-0.20	0.42	-0.02
Institutional Reform	0.30	-0.04	0.45	0.33	-0.02
Scandal	4.64	-0.18	-0.18	4.55	-0.20
Redistricting	0.88	-0.20	3.85	1.20	-0.06
Constit. Overlap	-0.04	-0.04	0.89	—	—
Senate Election	-0.05	-0.05	2.54	—	—
Open Seat — Senate	-0.06	-0.06	1.11	—	—
Open Seat — Governor	-0.01	-0.01	0.39	—	—
Number of Observations		4922		3431	

Figure 3.1
Sources of House Membership Turnover

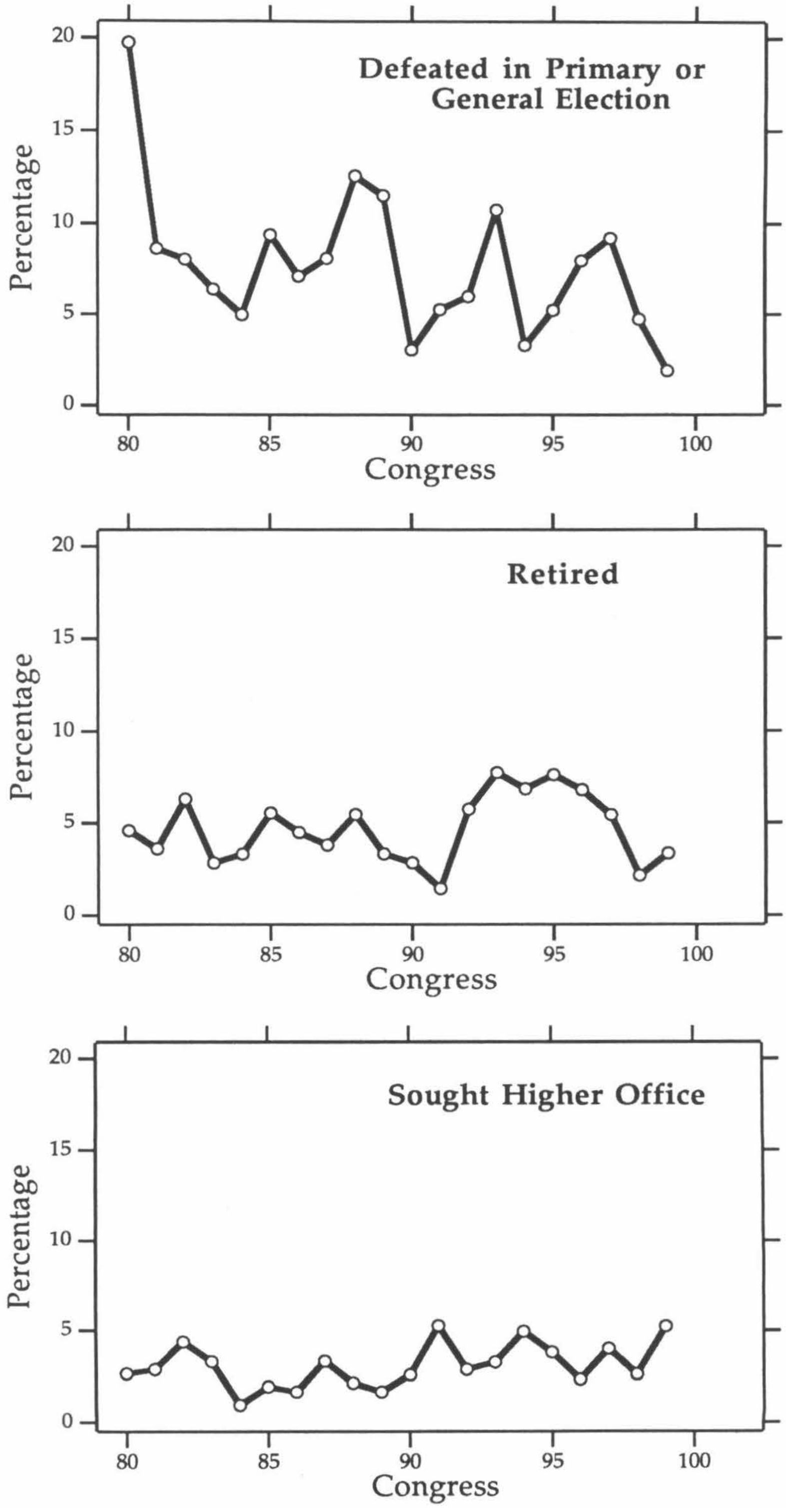


Figure 3.2

Previous Research: Problems in Sample Treatment

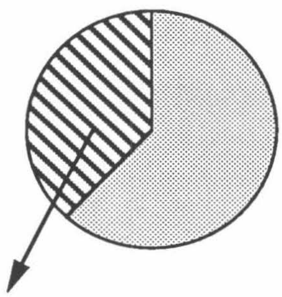
1. Voluntary retirement

2. Progressive ambition

- Run for re-election (92.3 %)
- Retire (4.6 %)
- Run for higher office (3.1 %)

- The conditional choice probabilities are interpreted as unconditional.
- No knowledge about the intercorrelation of all available choices.
- May fail to identify relevant variables.

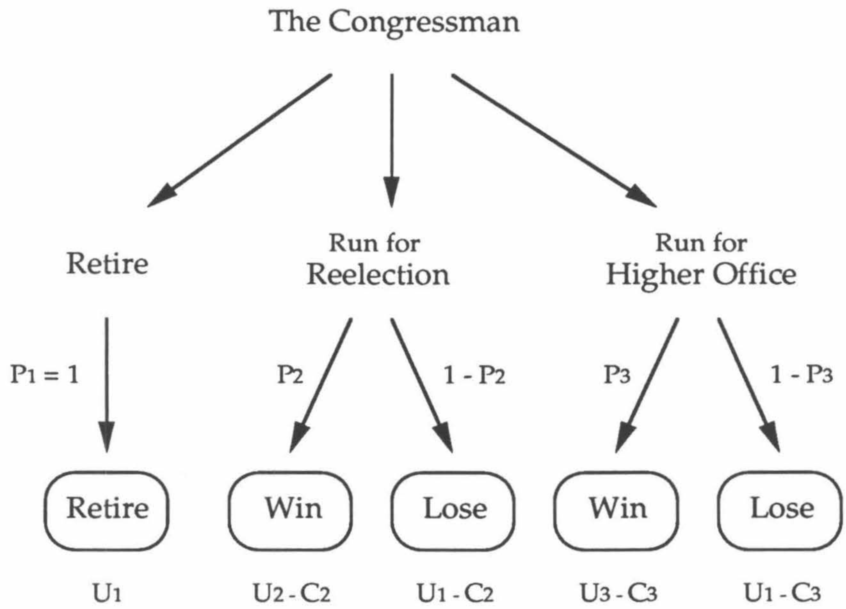
3. Schansburg (1992)



- 3 choices, have an opportunity for higher office (62%)
- 2 choices, no opportunity for higher office (38%)

Treated as having 3 choices in Schansburg (1992).

Figure 3.3
Congressional Career Choices



$$E_1 = U_1$$

$$E_2 = U_1 + P_2\,(U_2 - U_1) - C_2$$

$$E_3 = U_1 + P_3\,(U_3 - U_1) - C_3$$

Figure 3.4

Individual Career Choice and House Membership Composition