APPENDIX B

Models for Heterogeneous Choices

Heteroskedastic Choice Models

In the empirical chapters of the printed book we are interested in testing two different types of propositions about the beliefs of Americans on many different aspects of public policy. First, we want to know how different predispositions and values influence the beliefs which people have about public opinion. For example, we are critically concerned in our analysis of the beliefs of white Americans about affirmative action to find whether the predisposition commonly called modern racism has a strong influence on what particular beliefs are held by individuals. In other words, in the case of affirmative action, we want to determine how much influence modern racism has on whether individual citizens want to see minority set-aside programs in government contracts continued, or racial and ethnic criteria utilized in public university admissions.

But, second, we are also trying to understand the variability in responses on particular issues across respondents. That is, we have theoretical reasons to believe that some respondents have greater underlying and unobserved variability in their responses to survey questions about public policy than other respondents. As we discuss in the early chapters of the book, the differences across respondents in the underlying variance of their survey responses can be due to uncertainty, equivocation, or ambivalence. What we need is a methodological tool which allows us to test both for the direct effects of predispositions on policy beliefs as well as for these three different effects on the variance of the survey response.

For our work in the book we use as our methodological tool an inferential statistical model to test for the direct effects of predispositions and for the possibility of uncertain, ambivalent, or equivocal survey responses. One of the first wrinkles we need to discuss, however, is the fact that the data we have on the beliefs of survey respondents about public policy are discrete, not continuous. Thus, we cannot use statistical modeling tools like ordinary least squares (as presented in appendix A) to make inferences about how predispositions matter or what drives response variability.

Instead, we use two different types of discrete choice models in our empirical analyses. Which type of model we use is determined by whether we have discrete survey responses which are binary or ordinal, which we discuss in appendix A. Recall that binary response data are usually in the form of “agree” and “disagree” survey responses about public policy. For binary response data we use the binary probit model, which is the appropriate technique to use for such response data. On the other hand, ordinal response data are those in which individuals are asked to use some type of scale to give their opinion about a particular policy issue. For example, an ordinal survey response would be one where individuals are asked whether a minority set-aside program for government contracting should be totally eliminated, partially eliminated, kept the same, increased somewhat, or increased greatly. The assumption is that there is an implicit ordering of these responses ranging from total elimination to a great increase; the statistical model we use for ordinal responses like these assumes such an implicit ordering. Our statistical model for such data is the ordinal probit model, which is the appropriate technique to use for ordinal response data. [1]

If we were not concerned about the problem of unequal survey response variability across individuals in our surveys, it would be relatively simple for us to examine the effects of predispositions on public policy preferences using our survey response data. Both the binary and ordinal probit choice models are well understood and are easy to implement in virtually all major statistics packages. However, if the process that causes the unequal survey response variance is not accounted for in an empirical model of the particular survey question, the model is likely to produce incorrect results. Moreover, the underlying variance in a respondent’s answers yields direct information about the degree of certainty that a respondent has in his or her opinions. When there is the possibility that uncertainty, ambivalence or equivocation are operative for some individuals, then these three different sources of response variability must be included in models of issue preferences.

Yet, the problem of unequal variance across observations is familiar to every analyst of regression models as heteroskedasticity. In the least squares regression model, if the errors are heteroskedastic, the estimator is unbiased and consistent but is inefficient; and the typical estimate of the parameter covariance matrix is incorrect. Unfortunately, unequal variance is a worse problem for both binary and ordinal choice models. In the specific case of the probit model, for example, heteroskedasticity makes the maximum likelihood estimates inconsistent and the estimate of the covariance matrix of the model estimates is incorrect (Yatchew and Griliches 1985). The same result holds for the ordinal choice model, as we will see below, because both the binary and ordinal choice models are nonlinear statistical models. Unlike the case of linear models (where heteroskedasticity only leads to inefficiency), in nonlinear choice models we will have both inconsistent and inefficient coefficient estimates because the heteroskedasticity will affect both the coefficient estimates and the estimated variances of the parameters. Therefore, if heteroskedasticity is suspected in a probit model, it must be tested for and modeled if we expect to obtain consistent estimates.

We begin this appendix by developing the binary choice model with heteroskedasticity. Then we extend the ordinal choice model to also include heteroskedasticity. The two sections that follow present two other statistical models with heteroskedasticity: the ordinal choice model for aggregated response data, and the negative binomial model with heteroskedasticity. The appendix then concludes with a technical summary of how to produce estimated marginal effects for these non-linear statistical models—estimates of the effects of each independent variable that are easier to understand and discuss than the actual parameter estimates.

Heteroskedastic Binary Choice Models

We begin by presenting our model for binary choices with heteroskedasticity, and then we present our model of ordinal choices with heteroskedasticity. In maximum likelihood terms, the idea behind modeling dichotomous choice is to specify the systematic component of some probability (π_i) of individual *i* adopting the choice (y_i). In conventional probit and logit estimations, the analyst assumes that the π_i were generated by a homogeneous process, or that the data are identically and independently distributed. This permits the analyst to write the likelihood function in a relatively simple form:

$$\log L(\pi|y) = \sum_{i=1}^{N} y_i \log \pi_i + (1 - y_i) * \log(1 - \pi_i) \tag{B.1}$$

(where π_i is reparameterized as a function, usually the normal distribution [Φ] of a set of explanatory variables). Our argument is that preferences for public policy choices are *not* identically distributed and that the process of generating responses to policy choices is heterogeneous: some respondents will be more uncertain, more ambivalent, or more equivocal than other respondents and this will cause them to have a wider underlying distribution of choices. This means that the standard probit (equation B.1) will yield inconsistent estimates (see Greene 1993: 649-50).

We can address this source of inconsistency by modeling the heterogeneity. A plausible choice for the functional form of the heterogeneity is a variation of Harvey's "multiplicative heteroskedasticity" approach (1976):

$$y_i^* = X_i\beta + \varepsilon_i \tag{B.2}$$

$$\text{var}(\varepsilon_i) \quad = \quad [exp(Z_i\gamma)]^2$$

where y_i^* is a binary response to the policy question, X_i and Z_i are matrices of independent variables, ε_i is an error term, and β and γ are coefficient vectors to estimate. The first equation is a model of choice, in which a person's policy beliefs are a linear combination of interests leading the respondent to opt for a particular choice. (In this equation, we will also add sets of control variables which allow us to obtain accurate estimates about the effects of the core beliefs and predispositions on preferences and to test alternative hypotheses about what determines particular policy preferences.) The second equation is a model for the error variance, where we introduce variables accounting for alternative explanations (the multiplicative heteroskedasticity idea). This means that the systematic component now describes an identically distributed process for π_i^*:

$$\pi_i^* = g(\frac{X_i\beta}{e^{Z_i\gamma}}) \tag{B.3}$$

where $g()$ is an appropriate link function bounded between zero and one such as the probit function ($\Phi()$). The only identifying assumption in this model is that the variance equation cannot have a constant.

This leads to a log-likelihood function very similar to the usual probit log-likelihood:

$$\log L = \sum_{i=1}^{n} \left(y_i \log \Phi\left(\frac{X_i\beta}{\exp^{Z_i\gamma}}\right) + (1-y_i)\log\left[1-\Phi\left(\frac{X_i\beta}{\exp^{Z_i\gamma}}\right)\right]\right) \tag{B.4}$$

The significant difference between the likelihood above and the conventional probit is the inclusion of the variance model in the denominator in equation B.4.

Since in the log-likelihood of equation B.4 we have the term $\frac{X_i\beta}{\exp^{Z_i\gamma}}$ it is easy to gain an intuition for why it is important that heteroskedasticity in discrete choice models must be explicitly dealt with. Given that we have the systemic component of the choice function $X_i\beta$ divided by the variance function $\exp^{Z_i\gamma}$ we clearly have a nonlinear model. That is, the estimated effect of each component of the choice function (each β) is conditional on the elements of the denominator of this fraction ($\exp^{Z_i\gamma}$). If the denominator takes a different value for each individual, then the parameters of the choice function will be incorrectly estimated unless this individual variation is taken into account in the estimation of the model parameters.

This can also be seen in the derivatives of the log-likelihood function given in equation B.4 for the two sets of coefficients. First, the derivative of the log-likelihood function with respect to β is:

$$\frac{\partial \log L}{\partial \beta} \quad = \quad \sum_{i=1}^{n}\left[\frac{\phi_i(y_i-\Phi_i)}{\Phi_i(1-\Phi_i)}\right]\exp^{-Z_i\gamma}X_i \tag{B.5}$$

And the derivative of the log-likelihood function with respect to γ is:

$$\frac{\partial \log L}{\partial \beta} \quad = \quad \sum_{i=1}^{n} \left[\frac{\phi_i (y_i - \Phi_i)}{\Phi_i (1 - \Phi_i)} \right] \exp^{-Z_i \gamma} Z_i (-X_i \beta) \tag{B.6}$$

Thus, from the derivatives in equation B.5, it is easy to see that the estimates of β depend on the variance function. This implies that if there is heteroskedasticity in the data which is ignored, then the coefficient estimates are biased.

Fortunately our prediction of heterogenous responses to public policy questions can be formulated as a statistical test; so we test for the presence of heteroskedasticity in our models of public policy preferences using a simple likelihood ratio test (Davidson and MacKinnon 1984; Engle 1984).[2] This test compares an unrestricted model (with a fully specified variance model, equation B.4) to a restricted model (in which homoskedasticity is assumed, equation B.1). The null hypothesis is that the error variances are homoskedastic (i.e., that $\gamma = 0$), indicating that an ordinary probit will suffice. The alternative is that at least one γ is not zero. Let L_0 be the log likelihood for the restricted (homoskedastic) probit, L_H be the log likelihood for the unrestricted (heteroskedastic) probit, and k be the number of γ_i coefficients in the variance portion of the model. Then the likelihood ratio

$$LR = 2 \times (L_H - L_0) \tag{B.7}$$

is distributed as a χ^2 with k degrees of freedom.[3] If we cannot reject the null hypothesis that the error variances are homoskedastic (i.e., that $\gamma = 0$), then an ordinary probit will suffice.[4]

Heteroskedastic Ordinal Choice Models

The likelihood function for our ordinal heteroskedastic probit model is also relatively easy to derive. In fact, the ordinal heteroskedastic choice model is a simple extension of the binary heteroskedastic choice model derived in the previous section. We begin by assuming that there is a continuous underlying process Y_i such that:

$$Y_i \sim F(y_i \mid \pi_i) \tag{B.8}$$

where the systemic component is:

$$\pi_i = F(X_i \beta) \tag{B.9}$$

Next we denote our threshold parameters by μ_j, where $j = 1, \ldots, m$ and $\mu_1 = -\infty$ and $\mu_m = \infty$. We constrain the thresholds so that the probabilities are always positive:

$$\mu_{j-1} < \mu_j < \ldots < \mu_m \tag{B.10}$$

We know from the data which category y_i belongs to, so we can write that y_i belongs to category j if the following expression holds:

$$\mu_{j-1} < y_i \leq \mu_j \tag{B.11}$$

To make the exposition easier, we assume that y_i is a series of j binary variables (instead of being coded as one ordinal variable) such that:

$$y_{ij} = \begin{cases} 1 & \text{if } \mu_{j-1} < y_i \leq \mu_j \\ 0 & \text{otherwise} \end{cases} \tag{B.12}$$

We next write the probability that y_i is in j as:

$$P(y_i = j) = P(\mu_{j-1} < y_i \leq \mu_j) \tag{B.13}$$

$$= F\left(\frac{\mu_j - \beta' X_i}{\sigma_i}\right) - F\left(\frac{\mu_{j-1} - \beta' X_i}{\sigma_i}\right) \tag{B.14}$$

Usual derivations of this likelihood at this point assume that $\sigma_i = 1$. As we argue in the text, we wish to assume that choice is heterogeneous, so we assume instead that

$$\sigma_i = \exp(\gamma' Z_i) \tag{B.15}$$

where Z_i are variables which we believe measure the heterogeneity in choices across individuals and γ are coefficients.

We now write the likelihood for a given set of parameters as:

$$L = \prod_{i=1}^{n} \prod_{j=1}^{m} \left[F\left(\frac{\mu_j - \beta' X_i}{\sigma_i}\right) - F\left(\frac{\mu_{j-1} - \beta' X_i}{\sigma_i}\right)\right]^{y_{ij}} \tag{B.16}$$

We take logs to produce the log-likelihood function:

$$\ln L = \sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij} \ln\left[F\left(\frac{\mu_j - \beta' X_i}{\sigma_i}\right) - F\left(\frac{\mu_{j-1} - \beta' X_i}{\sigma_i}\right)\right] \tag{B.17}$$

where we assume that F represents the standard cumulative normal distribution.

The first derivatives of the model are easy to present (Alvarez and Brehm 1998; Greene 1997). Using the same notation, except referencing one of the ordinal categories as k from the set j and noting that f gives the normal density ϕ, we obtain the following derivative:

$$\frac{\partial \ln L_i}{\partial \beta} = \frac{f[(\mu_{j-1,k} - \beta' X_i)/\sigma_i] - f[(\mu_{j,k} - \beta' X_i)/\sigma_i]}{F[(\mu_{j,k} - \beta' X_i)/\sigma_i] - F[(\mu_{j-1,k} - \beta' X_i)/\sigma_i]} X_i/\sigma_i \tag{B.18}$$

Define:

$$f_{j,k} = f[(\mu_{j,k} - \beta' X_i)/\sigma_i] \tag{B.19}$$

and

$$F_{j,k} = F[(\mu_{j,k} - \beta' X_i)/\sigma_i] \tag{B.20}$$

This allows us to easily write the remaining first derivatives:

$$\frac{\partial \ln L_i}{\partial \mu_{j,k}} = [f_{j,k}/(F_{j,k} - F_{j-1,k})]/\sigma_i \tag{B.21}$$

$$\frac{\partial \ln L_i}{\partial \mu_{j-1,k}} = -[f_{j-1,k}/(F_{j,k} - F_{j-1,k})]/\sigma_i \tag{B.22}$$

$$\frac{\partial \ln L_i}{\partial \gamma} = -([1/(F_{j,k} - F_{j-1,k})]/\sigma_i)[(\mu_{j,k} - \beta' X_i) - (\mu_{j-1,k} - \beta' X_i) f_{j-1,k}] Z_i \tag{B.23}$$

Aggregate Heteroskedastic Ordinal Choice Models

The next step is to extend the heteroskedastic ordered choice model to aggregates. The dependent variable now measures the number of people who respond in each of the three or four categories. If we make the strong assumption that each observation within an aggregate is uncorrelated with any other and generated by the same process (i.e., distributed independently and identically), then the extension is simple:

$$\begin{aligned} N(y_i = 1) &= \binom{\sum N_i}{N_1} P(y_i = 1)^{N_1} \\ N(y_i = 2) &= \binom{\sum N_i}{N_2} P(y_i = 2)^{N_2} \\ &\vdots \end{aligned} \tag{B.24}$$

$$N(y_i = 4) \quad = \quad \binom{\sum N_i}{N_4} P(y_i = 4)^{N_4} \tag{B.25}$$

where N_i represents the observed number who respond in category i.[5]

Substituting in the formula for the probabilities of each choice (B.14), we have:

$$\begin{aligned} N(y_i = 1) &= \binom{\sum N_i}{N_1} F\left(\frac{-X_i\beta}{\exp(Z_i\gamma)^2}\right)^{N_1} \\ N(y_i = 2) &= \binom{\sum N_i}{N_1} \left[F\left(\frac{\mu_1 - X_i\beta}{\exp(Z_i\gamma)^2}\right) - F\left(\frac{-X_i\beta}{\exp(Z_i\gamma)^2}\right)\right]^{N_2} \\ &\vdots \end{aligned} \tag{B.26}$$

$$N(y_i = 4) \quad = \quad \binom{\sum N_i}{N_1} \left[1 - F\left(\frac{\mu_3 - X_i\beta}{\exp(Z_i\gamma)^2}\right)\right]^{N_4} \tag{B.27}$$

What is important to notice is that, as a consequence of the strong independent and identical distribution assumption above, the parameters of interest, β and γ, refer to exactly the same thing across the two levels of analysis: the effect of a unit change upon the probability of individual choice and the variability of individual choice, respectively. In (B.27) we explicitly use aggregate information to draw inferences about individual choice.

Heterogeneous Negative Binomial Models

Our inferential method for estimating the determinants of variance in the heterogeneous negative binomial model that follows is similiar to our heterogenous discrete choice models. The two parameters in the negative binomial model are the mean event count rate (λ_i) and the dispersion (α). The event count rate must be greater than or equal to zero, while the dispersion must be greater than one.

We begin by assuming that each dependent variable ($Y_i = 0, 1, 2, 3, \ldots$ and i indexes respondents) has the following distribution:[6]

$$Y_i \sim f_{nb}(y_i|\lambda_i, \alpha) \tag{B.28}$$

where $E(Y_i) = \lambda_i$. This produces a probability distribution for $P(Y_i = y_i|\lambda_i, \alpha_i)$:

$$P(y_i|\lambda_i, \alpha_i) = \prod_{i=1}^{N} \frac{\Gamma(1/\alpha_i + Y_i)}{\Gamma(Y_i + 1)\Gamma(1/\alpha_i)} \left(\frac{1}{1 + \alpha_i\lambda_i}\right)^{1/\alpha_i} \left(1 - \frac{1}{1 + \alpha_i\lambda_i}\right)^{Y} \tag{B.29}$$

The log-likelihood for the negative binomial model is given by:

$$\ln L(\lambda,\alpha|Y) = \sum_{i=1}^{N} \frac{\Gamma(1/\alpha_i + Y_i)}{\Gamma(Y_i+1)\Gamma(1/\alpha_i)}(\frac{1}{1+\alpha_i\lambda_i})^{1/\alpha_i}(1-\frac{1}{1+\alpha_i\lambda_i})^{Y} \tag{B.30}$$

Now we reparameterize the event count rate to be a function of the explanatory variables in the choice component of the model (X) and coefficients (β):

$$\lambda_i = \exp(X_i\beta) \tag{B.31}$$

and similarly for the variance model variables (Z) and coefficients (γ):

$$\alpha_i = \exp(Z_i\gamma) \tag{B.32}$$

So after parameterizing for the rate (λ) and dispersion (α), this produces a log-likelihood function slightly different from the standard one presented above:

$$\begin{aligned}\ln L(\beta,\gamma|Y_i,X_i,Z_i) \quad = \quad & \sum_{i=1}^{N} \ln\Gamma(1/\exp(Z_i\gamma)+Y_i) - \ln\Gamma(Y_i+1) - \ln\Gamma(1/\exp(Z_i\gamma)) \\ & - \ln\Gamma(1/\exp(Z_i\gamma)) - \ln((1+\exp(X_i\beta))/\exp(Z_i\gamma)) \\ & + Y_i \ln(\exp(X_i\beta)/(1+\exp(X_i\beta)))\end{aligned} \tag{B.33}$$

Again, we simply will estimate the choice function parameters using this log-likelihood function (i.e., the β parameters) in the choice function and we parameterize the error variance as in our ordinal heteroskedastic choice model with a set of variables measuring the potential sources of heterogeneity (the Z_i variables) and their associated parameters (the γ parameters).

Estimation of Heteroskedastic Choice Models

These heteroskedastic choice models are widely used in this book. Fortunately, they are also relatively easy statistical models to estimate. In the previous sections, we presented all of the information needed to program both heteroskedastic choice models in any type of statistical software package which can perform maximum likelihood analysis, like GAUSS. Armed with the log likelihood and the first derivatives, it is relatively simple to program either model in GAUSS; also, the heteroskedastic choice models do not take an extremely long time to reach convergence.

We have used GAUSS, SHAZAM, and STATA to maximize our log-likelihood functions for the heteroskedastic choice models presented in this book. Also, LIMDEP can estimate both models without programming the log-likelihood function, making it a very simple and effective way to estimate this class of discrete choice models; we also use LIMDEP to estimate the models we present in this book. The most recent release of STATA has included a command to estimate heteroskedastic binary probit models. Computer code using each of these statistical software packages is available from the authors at http://hardchoices.caltech.edu.

In our work with the heteroskedastic choice models we have experienced no unusual problems in the actual estimation of the models. For example, the ordinal heteroskedastic probit models we use in chapter 6 (on attitudes about race and affirmative action) converge rapidly and in just thirty to thirty-five iterations. The only problem which researchers must be aware of when using these models is one which is actually more an aspect of the ordinal choice model in general; in some cases the maximum-likelihood estimation of any ordinal choice model (with or without heteroskedasticity) might produce estimates of the underlying thresholds which are not strictly ordered. If this occurs, the estimation routine is likely either to not converge or to produce estimates which are clearly incorrect. But we have not encountered this problem in any of the applications which we report in the book.

Interpretation of Heteroskedastic Choice Models

The primary problem with using these heteroskedastic discrete choice models is that the estimates can be difficult to interpret. Since both the binary and the ordinal heteroskedastic choice models are highly nonlinear, that means that any particular coefficient estimate depends on both the other coefficient estimates and the values of the independent variables. The coefficient estimates themselves can only reveal the sign of the estimates relationship and whether it is statistically different from zero.

Thus, we largely refrain from presenting coefficient estimates in the text of the book. Instead we present the coefficient estimates in the appendices posted here; we also provide discussion in these same appendices of how we code variables, how we generate many of the scales used in our analysis, and some general discussion of the fit of each particular model. Readers who are interested in the details of model estimation can examine the results directly for themselves in the appropriate appendices.

When we present results from our heteroskedastic discrete choice models in the printed chapters of the book we present secondary analyses. That is, we use the coefficient estimates we obtain to produce results which we can more easily interpret and discuss in the context of each chapter. To produce our secondary estimates we rely upon three different approaches.

First, we often make use of the *estimated marginal effects* of variables in the choice function of the model. Once we have coefficient estimates in hand, it is relatively easy to compute the estimated marginal effects of each right-hand side variable. These marginal effect estimates are simply the estimates of the change in probability of choice which we expect conditioned on a change in the value of the particular independent variable. In other words, the marginal effect estimates for the effects of variables in the choice function of our heteroskedastic discrete choice models are simply the partial derivatives of the probability of a certain choice being made with respect to a change in one of the independent variables.

For the binary choice model with heteroskedasticity, we write the marginal effects as:

$$\frac{\partial P(Y=1)}{\partial X_i} = \left[\frac{\phi X_i \beta}{\exp^{Z_i \gamma}}\right] \frac{\beta}{\exp^{Z_i \gamma}} \tag{B.34}$$

The expression given in equation B.34 holds for each particular independent variable (that is, each X) and the respective coefficient estimate (the corresponding estimated value of β).

We write the marginal effect estimates for the ordinal choice model with heteroskedasticity in a very similar way:

$$\frac{\partial P(k)}{\partial X_i} = [f(\mu_{j-1} - X_i\beta)/\exp(Z_i\gamma) - f(\mu_j - X_i\beta)/\exp(Z_i\gamma)]\beta/\exp(Z_i\gamma) \tag{B.35}$$

Since we present the marginal of each right-hand side variable on the probabilities of choosing the low (L) and high (H) categories, these expressions are given as:

$$\frac{\partial P(L)}{\partial X_i} = -\frac{\phi(X_i\beta)}{\exp(\gamma' Z_i)} \frac{\beta}{\exp(\gamma' Z_i)} \tag{B.36}$$

and

$$\frac{\partial P(H)}{\partial X_i} = \frac{\phi(\mu_4 - X_i\beta)}{\exp(\gamma' Z_i)} \frac{\beta}{\exp(\gamma' Z_i)} \tag{B.37}$$

The second way we present our secondary analyses is by using *differencing*.[7] We use the differencing approach primarily for our presentation of the effects of the variables in the variance function. The essence of the differencing approach begins with the estimated values of the coefficients in the variance function (γ) and with the sample distribution information of each variable in the variance function (the mean, the standard deviation, and the sample minimum and maximum of each variance function variable). We set all of the variables in the variance function to their mean values; multiplying each of these mean values by the coefficient estimates, and then summing these products gives us an estimate of the magnitude of the error variance. Thus, we begin with the estimated magnitude of the error variance at the sample means of the variables in the variance function.

Then we start with one of the variables in the variance function. We change the value of this one variable by some predetermined amount (either by one standard deviation or by changing the value to the sample minimum or maximum). We then recompute the magnitude of the error variance at this second value of the particular variable; the difference between this estimated error variance and the first gives us an estimate of the effect of the particular variable in the variance function on the error variance. We repeat this procedure for all of the variables in the error variance function.

This differencing procedure also can be used to present the estimated effect of each variable in the choice function. All this involves is the identical approach to that which we just discussed; we set all of the choice function variables to their sample means and there we compute the probability of a particular policy choice. We then vary one of the choice function variables by some predetermined amount (by one standard deviation, or by changing it to the sample minimum or maximum), and then we recompute the probability of policy choice. The difference between these two probability estimates gives us an estimate of the effect of the particular choice function variable on the probability that a representative individual might make a certain policy choice.

The third approach we use for presenting estimated effects of variables in both the choice and variance function is a *graphical approach*. Here, we just generalize the differencing approach for both the choice and variance function variables. For example, if we were interested in measuring the impact of a particular variable from the choice function we would begin by setting all of the variables (but not the variable of interest) to their sample mean values. We would set the variable of interest to the sample minimum value, and we could compute the probability of policy choice. Then, we would increase the variable of interest by some small amount, and again recompute the probability of policy choice. We would continue to increase the variable of interest by incremental amounts, each time recomputing the probability of policy choice, until we reached the sample maximum value for the variable of interest. Last, we would graph each of these probability estimates for each value of the variable of interest.

The graphical approach also works for the variance function estimates. The only important difference is that for each value of the variable of interest from the variance function we would compute the magnitude of the error variance, holding the other variables in the variance function at their sample means. We would compute the magnitude of error variance for each incremental value of the variable of interest; again, we would graph these two sets of points for presentation purposes.

Notes

1.For both binary and ordinal response data another widely used statistical model is called the logit binary or categorical choice model (these are sometimes called the logistic choice models in the literature). This is essentially identical to the probit model; the only major distinction between the two models is that the former assumes that the response data have slightly different statistical distributions (the logistic rather than the normal). The logit function can be substituted here for a similar model (Dubin and Zeng 1991; Gerber and Lupia 1993). Another interesting application of the heteroskedastic probit model is given by Knapp and Seaks (1992).

2.There are two other tests for heteroskedasticity in the binary choice framework—the Lagrange multiplier and Wald test statistics. We use the likelihood ratio test here since it is most familiar to political scientists, since we have a theoretical specification for the variance function and since we have a strong interest in estimating the parameters in γ. Also, they are asymptotically equivalent tests (Engle 1984).

3.Davidson and MacKinnon (1984) demonstrate through Monte Carlo simulations that this LR test falsely rejects the null in less than 1 percent of the replications at the $p < .01$ level, with only five hundred observations in each replication. With the greater number of observations in our sample, Davidson and MacKinnon's findings suggest even greater power in our application of the LR test.

4.All tests for heteroskedasticity are sensitive to model misspecification (Davidson and MacKinnon 1984). In fact, an alternative approach to heteroskedasticity is to regard it as a problem of misspecified functional form, and to incorporate a series of interactive terms into the model. For testing our model, though, the variance function is of intrinsic interest, and is a function of understandable parameters. Estimating the variance function, moreover, is a direct test of our argument.

5.Assuming that observations are independently and identically distributed is very clearly a strong assumption. This implies, for example, that the beliefs about racial policies in one congressional district in a state are uncorrelated with those same beliefs in the adjacent district. Our next task is to determine how sensitive this empirical model is to this assumption. Then we will have to weaken this assumption, and to develop an alternative estimation model which does not require that observations be independently and identically distributed.

6.This follows the explication in Long (1997).

7.This is similar to the *first difference* approach discussed by King (1989).