Difficult Choices: An Evaluation of Heterogenous Choice Models *

Luke Keele
Department of Political Science
2140 Derby Hall
154 N. Oval Mall
Ohio State University
Columbus, OH 43210
Tele: 614-247-4256
Email: keele.4@polisci.osu.edu

David K. Park
Department of Political Science
Campus Box 1063
Washington University
One Brookings Drive
St Louis, MO 63130
Tele: 314-935-7276
Email: dpark@artsci.wustl.edu

March 3, 2006

*Prepared for the 2004 Meeting of the American Political Science Association, Chicago, IL Sept, 2-5. For comments and suggestions, we thank Chris Achen, Neal Beck, John Brehm, Andrew Gelman, Gary King, Andrew Martin, Walter Mebane, Jasjeet Sekhon, Curt Signorino, Jim Stimson, Marco Steenbergen, and the Statistics Working Group at Columbia University. Any errors that remain are our own.

Abstract

While the derivation and estimation of heterogeneous choice models appears straightforward, the properties of such models are not well understood. It is assumed that the properties of these models are identical to those of discrete choice models. We present analytical results that suggest the properties of these models are far more complex. Using a series of Monte Carlo experiments, we further analyze the properties of heteroskedastic probit and heteroskedastic ordered probit models. We test the relative efficiency of these models as well as how robust they are to specification and measurement error. We find that the estimates in heterogeneous choice models tend to be biased in all but ideal conditions and can often lead to incorrect inferences.

Unequal error variances, or heteroskedasticity, invariably causes problems for statistical inference. In the context of ordinary least squares, heteroskedasticity does not bias our parameter estimates, rather it either inflates or underestimates the standard errors. Heteroskedasticity, however, is more problematic in discrete choice models such as logit or probit and their ordered and multinomial variants. If we have nonconstant variances in the error term of a discrete choice model, not only are the standard errors incorrect, but the parameters are also biased and inconsistent (Yatchew and Griliches 1985).

As such, heteroskedasticity is usually treated like a disease, something to cure and then forget about. However, Alvarez and Brehm (1995) looked at the problem of heteroskedasticity and saw an opportunity to use unequal variances within our samples to offer powerful insights into empirical processes. They used heteroskedasticity as a means of exploring heterogeneity in choice situations. To do this, they developed a class of models they call heterogenous choice models, which include heteroskedastic probit and heteroskedastic ordered probit models.

The subsequent popularity of these models is testament to the interest in furthering our understanding of choice heterogeneity. These heteroskedastic models have been widely used to explore heterogenous choices and behaviors (Alvarez and Brehm 1997, 1998, 2002; Busch and Reinhardt 1999; Gabel 1998; Lee 2002; Krutz 2005). Routines for these models have become standard in statistical software such as Stata, Limdep, SAS, Eviews and Shazaam. These models regularly appear in working papers and are widely taught in graduate methodology courses at leading research institutions in political science.

What is the purpose of these models? In statistical terms, we typically estimate the following model: $Y \sim f(y|\theta, \alpha)$ and $\theta = g(X\beta)$ where θ and α are parameters to be estimated and $g(\cdot)$ is a link function. Normally, the theory is a test of some restriction on θ, while α represents an ancillary parameter such as σ^2, which tells us how much variability there is in the expected value of Y. As such, σ^2 is treated as a nuisance parameter assumed to be constant. But, at times, the variability in a choice may be more interesting than what determines the choice itself. Most people could easily identify the factors that structure attitudes on abortion, but understanding the choice as one fraught with ambivalence is much more interesting (Alvarez and Brehm 1995). There are many theoretical questions that should lead us more directly to a concern over the estimates of σ^2 than θ.

Such heterogeneity can arise from many sources and is more widespread than we often admit. For example, heterogeneity may be the product of different levels of information about a choice: certainty about a candidate's issue position might vary across levels of political sophistication (Alvarez and Franklin 1994). Heterogeneity might also be a by-product of the size of decision makers. Large firms or interest groups with greater resources might adopt a greater variety of lobbying strategies than smaller organizations. Or choice heterogeneity might result from income differences, as households with higher incomes might have greater variation in their purchasing decisions than lower income households. We might expect heterogenous choices due to socialization effects as those with higher amounts of socialization should exhibit less variation in their choices than those with little socialization. In international relations, more specifically in the democratic peace literature, we might expect choice heterogeneity in that some nations might have more variability in their propensity to engage in conflict: an autocratic regime might go to war under a variety of circumstances while democratic regimes may only go to war under limited conditions. Some have argued that heterogeneity may be the norm for international relations data (Lemke 2002).

At present, heteroskedastic probit and heteroskedastic ordered probit models are the tools of choice when investigating discrete heterogenous choices. The attraction of these models is the ability to test theories that relate directly to σ^2. The estimation of heterogeneous choice models is fairly straightforward. A researcher that suspects heterogeneity can select a set of covariates and model the heterogeneity. The ease with which we can estimate these models, however, belies the fact that the properties of these models are not well understood. While heterogenous choice models can be used for either "curing" probit models with unequal error variances or for testing hypotheses about heterogenous choices, there is little evidence, analytical or empirical, about how well these models perform at either task. We start by reviewing the derivation of these models as a first step toward an assessment of them.

1 Heterogenous Choice Models

To understand heteroskedastic probit and ordered probit models and their possible problems, we start with a review of their derivation. Consider a latent or unobserved, continuous variable

y_i^*:

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \tag{1}$$

where $\mathbf{x}_i$ is a vector of values for the ith observation, $\boldsymbol{\beta}$ is a vector of parameters, ε_i is the unobserved error. An observed realization of y_i^*, y_i, is related to y_i^* by the following mechanism:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > \tau \\ 0 & \text{if } y_i^* \leq \tau \end{cases} \tag{2}$$

where τ is a threshold parameter. Typically τ represents the threshold at which an individual makes a choice. For example, suppose that y_i^* represents the continuous unobserved propensity to vote, τ would then represent the threshold at which a nonvoter decides to vote. While in some settings τ is an estimated parameter, it is most often fixed at zero. To model y_i as a dichotomous choice, we must specify the systematic component of the probability of individual i adopting the choice y_i. If we set τ to zero, we can derive a parameterized functional form for a dichotomous choice:[1]

$$\begin{aligned} Pr(y_i = 1) &= Pr(y_i^* > 0) \\ &= Pr(\mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i > 0) \\ &= Pr(\varepsilon_i > -\mathbf{x}_i\boldsymbol{\beta}) \\ &= Pr(\varepsilon_i \leq \mathbf{x}_i\boldsymbol{\beta}). \end{aligned} \tag{3}$$

If we assume that ε_i follows some distribution, we can integrate over that distribution to estimate the probability that ε_i is less than or equal to $\mathbf{x}_i\boldsymbol{\beta}$. If we assume the errors of y_i^* are normally distributed, then we estimate a probit model typically written in the following way:

$$\Pr(y_i = 1) = \Phi(\mathbf{x}_i\boldsymbol{\beta}) \tag{4}$$

[1] This derivation is, of course, equivalent to the random utility derivation of dichotomous choice models. While there are other methods for deriving discrete choice models, we feel this method helps to clarify the nature of the error variance.

where Φ represents the cumulative normal distribution.

To estimate the probit model, however, we have to make an assumption about the error variance of the unobserved y_i^*. We must assume that the error term is homoskedastic or constant across individuals. This assumption is incorporated into the probit model by dividing both sides of the inequality in (3) by the standard deviation of ε_i:

$$Pr(y_i = 1) = Pr\left(\frac{\varepsilon_i}{\sigma} > -\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right). \tag{5}$$

This then gives us:

$$Pr(y_i = 1) = \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) \tag{6}$$

Normally, we assume that σ is a constant that equals 1, which, removes it from the equation, and we can estimate $\hat{\boldsymbol{\beta}}$.[2] In the context of heterogenous choices, σ is known or expected to vary systematically, such that $\sigma = \sigma_i$ with $i = 1 \ldots N$. For example, σ will be larger for respondents who experience value conflict or who are politically ignorant, but σ will be smaller for political sophisticates or people that are strongly committed to one side of the abortion debate. But if the errors are nonconstant (heteroskedastic), then $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}/\sigma$, and the parameter estimates will be biased, inconsistent, and inefficient (Yatchew and Griliches 1985).

Alvarez and Brehm (1995) use a parametric model to both "cure" the biased caused by heteroskedasticity as well as to test theories about the nature of the variation in the error term of the model. Following Harvey (1976) and his model for heteroskedastic regression, Alvarez and Brehm (1995) adopt a multiplicative functional form for the variance of ε_i:

$$\mathrm{Var}(\varepsilon_i) = \sigma_i^2 = \exp(\mathbf{z}_i\boldsymbol{\gamma})^2 \tag{7}$$

where $\mathbf{z}_i$ is a vector of covariates of the ith observation that define groups with different error variances in the underlying latent variable, $\boldsymbol{\gamma}$ is a vector of parameters to be estimated, and ε_i is the error variance. By taking the positive square root of (7), we have a model for the standard deviation of the error distribution:

[2]The reader should note that this assumption is also required to identify the model.

$$\sigma_i = \exp(\mathbf{z}_i \boldsymbol{\gamma}). \tag{8}$$

So, we now divide $\mathbf{x}_i \boldsymbol{\beta}$ by the right-hand side of Equation 8. The probability function for a particular observation now equals:

$$\Pr(y_i = 1) = \Phi\left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{\exp(\mathbf{z}_i \boldsymbol{\gamma})}\right) \tag{9}$$

From this, we derive the heteroskedastic probit log-likelihood:

$$lnL(\hat{\boldsymbol{\beta}}, \hat{\gamma}|Y) = \sum_{y=1}^{N}\left(y_i \ln\Phi\left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{\exp(\mathbf{z}_i \boldsymbol{\gamma})}\right) + (1 - y_i)\ln\left[1 - \Phi\left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{\exp(\mathbf{z}_i \boldsymbol{\gamma})}\right)\right]\right) \tag{10}$$

Maximizing (10) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ given $\mathbf{x}_i$ and $\mathbf{z}_i$ is done as it would be for any maximum likelihood estimator.[3] This gives us estimates for both the effect of the $\mathbf{x}$ predictors on the discrete choice y_i, but also the effect of the $\mathbf{z}$ predictors on the variability of y_i.[4]

1.1 Properties of Heterogenous Choice Models

Despite the widespread use of heterogeneous choice models, there has been little examination of these models' statistical properties. Researchers implicitly assume that heterogeneous choice models share the same statistical properties as probit and ordinal probit models. As we demonstrate below, this is not true in several respects.[5]

First, we compare the efficiency of the heteroskedastic probit model to standard probit models. It is useful, here, to recall a few points from the derivation of the probit model. There exists a latent or unobserved, continuous variable y_i^*, but we only observe the binary y_i. Some efficiency must be lost by estimating a probit model with y_i as opposed to estimating a regression with y_i^*, as a binary y_i contains less information than the continuous y_i^* (Davidson and

[3]The reader should note that no constant is included in the variance model. The model is not identified if a constant is included in the variance model. If all the elements of $\boldsymbol{\gamma}$ are equal to 0, then $e^0 = 1$ and the model is identified just as it was in the standard probit model.

[4]Just as ordered probit is a special case of probit, so to is heteroskedastic ordered probit a special case of heteroskedastic probit.

[5]We are not the first to note that there may be problems with these models. Greene (1993) notes that this may be a difficult model to fit and may require many iterations to converge. And Achen (2002) notes that collinearity may make testing some types of hypotheses difficult.

MacKinnon 1993). Specifically, the loss of efficiency depends on the distribution of $\Phi(\mathbf{x}_i\boldsymbol{\beta})$. The probit model will most closely match the efficiency of OLS when a large share of the sample is such that $\Phi(\mathbf{x}_i\boldsymbol{\beta}) = 0.5$, but even when this is true, the probit model is still notably less efficient. Probit models where a large portion of the sample $\Phi(\mathbf{x}_i\boldsymbol{\beta})$ is near 0 or 1 will be particularly inefficient (Davidson and MacKinnon 1993).

The relative inefficiency of standard probit implies that the heteroskedastic probit model should be less efficient than the standard probit model. No information has been added to y_i, but a number of additional parameters have been added in the form of the variance model. Therefore, we expect that unless distributional restrictions are imposed on $\Phi(\mathbf{x}_i\boldsymbol{\beta})$, the heteroskedastic model should be less efficient than the standard probit model. Whether this inefficiency is something to be concerned about is an open question, one we need to explore through simulation.

We, next, more carefully consider the functional form of the heteroskedastic probit model. When we estimate a generalized linear model (GLM), the analyst must make two choices. First, the analyst must choose, $g(\cdot)$, the appropriate link function. Second, the analyst chooses the functional form of the regression function $f(X)$, which is almost always assumed to have the following linear form $f(\mathbf{x}_i\beta)$, hence a generalized *linear* model. But one could choose a functional form for $f(X)$ that was nonlinear in the following way:

$$\frac{\mathbf{x}_i\beta}{\exp(\mathbf{z}_i\gamma)} \tag{11}$$

while assuming that the data generating process is homoskedastic so that $\sigma = 1$. This model is, of course, identical to the heteroskedastic probit model:

$$Pr(y_i = 1) = \Phi\left[\frac{\frac{\mathbf{x}_i\beta}{\exp(\mathbf{z}_i\gamma)}}{1}\right] \tag{12}$$

The heteroskedastic model and the nonlinear model are exactly equivalent and there is no means for distinguishing between them (Achen 2002). Why is this true? When y_i is continuous, its first and second moments, the mean and the variance, are separately identified. But this is not the case when y_i is discrete, as the first and second moments are no longer separately identified. For a discrete y_i we cannot separately estimate β and σ (Ruud 2000). Maddalla (1983) puts it most succinctly, "It can be easily seen ... that we can only estimate $\frac{\beta}{\sigma}$ and not β and σ separately." Therefore, when estimating a heteroskedastic probit model one is not estimating a separate model for the variance, but instead estimating a nonlinear mean function for the $Pr(y_i = 1)$. In short, the heteroskedastic probit model falls within the class of generalized *nonlinear* models. The analytic properties of such models are generally intractable and are rarely considered, but McCullagh and Nelder (1989) consider the practical aspects of fitting generalized nonlinear models.

First, they suggest that such nonlinear models will have large sampling errors especially when they include exponential forms (McCullagh and Nelder 1989). This is of particular concern since we already suspect that such models should be less efficient than standard probit models.

Second, they suggest that generalized nonlinear models may be difficult to fit and have trouble converging. Moreover, all the generalized nonlinear models that they consider (none of which are fitted with binary data) are fit with quasi-likelihood methods utilized when there isn't enough information to form a likelihood. In general, they regard such models as feasible but difficult to fit (McCullagh and Nelder 1989).

In short, while heteroskedastic probit models may seem like simple adaptations of standard probit models, they contain an unusual form of nonlinearity that should complicate their estimation. The assumption in the statistics literature is that quasi-likelihood methods are necessary for such nonlinearity. Given the complications that appear to underlie such models, it would seem wise to perform at least some cursory examination of their properties.

For example, it would seem important to understand the effects of misspecification in these models. Considering the bias caused by heteroskedasticity in a probit model, it would seem that a failure to correct for heteroskedasticity via the correct specification of the variance model would cause for additional bias in the parameter estimates. In short, we may not only face bias from specification error, but specification error in the variance model could allow additional bias from heteroskedasticity to affect the model estimates. The same might be true of measurement error. Measurement error in variance model variables could allow for additional bias caused by our inability to correct for the heteroskedasticity we expect to be present. While the analytic foundations may be lacking for such models, we can turn to simulation to assess how these models perform under some basic conditions. We now turn to a set of Monte Carlo experiments to more fully assess the properties of heterogenous choice models.

2 A Monte Carlo Analysis

To better understand the nature of heterogenous choice models, we developed a number of Monte Carlo experiments. We performed separate Monte Carlo analyses for heteroskedastic probit and heteroskedastic ordered probit. In the ordered case, the outcome variable has four categories. We fit both models with two predictors in the choice model and two predictors in the variance model. So the data generating process has the following form:

$$\begin{aligned} y_i &= \Phi(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i) \\ \mathrm{Var}(\varepsilon_i) &= \exp(\gamma_1 z_{1i} + \gamma_2 z_{2i})^2 \end{aligned}$$

We set the parameters for the choice model to the following values: $\beta_0 = 1$, $\beta_1 = 2$ and $\beta_2 = 3$ and for the variance model $\gamma_1 = 1$ and $\gamma_2 = 2$.

We performed two different Monte Carlo experiments. In the first experiment, we use a perfectly specified model to assess how efficient these models are. In the second experiment, we examine the effects of specification and measurement error. In these experiments, we either omitted a relevant variable or substituted a variable that was correlated with one of the true variables into the estimating equations. For example, in one case we substituted a variable that was correlated at 0.90 with one of the variables from the data generating process into the estimated model.

For each experiment, we recorded a variety of statistics. We track the Root Mean Squared Error (RMSE), the bias in the estimates (which we often report as a percentage), the mean absolute error (MAE), the coverage rate, or the percentage of times the 95% confidence interval contains the true parameter value, and a measure of overconfidence. With the measure of overconfidence, the quality of the ML estimates of variability are assessed by calculating the ratio between the root mean square average of the 1000 estimated errors and the corresponding standard deviations of the 1000 estimates.[6] If the measure of overconfidence is above 100%, let's say 150%, then the true sampling variability is, on average, one and a half times the

[6]The measure is more precisely: $100\frac{\sqrt{\sum_{l=1}^{1000}(\beta-\hat{\beta})^2}}{\sqrt{\sum_{l=1}^{1000}(s.e.(\hat{\beta}))^2}}$.

reported estimate of variability, in other words the true standard errors are 1.5 times larger than the estimated standard errors. For each experiment, we report whichever statistics are most appropriate.

We purposely leave the design of the Monte Carlo experiments fairly uncomplicated. First, all the exogenous variables are orthogonal to each other. This will, of course, minimize the effects of misspecification. Moreover, all the exogenous variables are continuous and normally distributed. We wouldn't expect either of these conditions to hold in most applied work. Both these conditions should work in favor of observing better performance by the heteroskedastic models.[7]

Before presenting the results, we stop briefly to consider a benchmark for the general performance of the heterogenous models. The benchmark we use for the estimates of $\boldsymbol{\beta}$ are the estimates from standard probit and ordered probit models used with homoskedastic data. We assume that heterogenous models should work as well with heteroskedatic data as normal models work with homoskedastic data. So for each experiment, we construct a baseline with regular probit models using homoskedastic data generating processes. There are some limits to this baseline, however, for the γ parameters, no such obvious baseline exists. For later experiments, we construct an alternative baseline. In the first experiment, we examine the efficiency of the heteroskedastic probit models.

2.1 Efficiency

First, we present the results for when the models are perfectly specified with a sample size of 1000. We report the coverage rates and level of confidence. We are most interested in the later two measures, since it is the relative efficiency of the models that we are testing. We also compare the results to standard probit models. Table 1 contains the MAE, levels of overconfidence and the coverage rates, for both heteroskedastic and homoskedastic models under ideal circumstances.

The heteroskedastic ordered probit model can be given a clean bill of health, as both the level of overconfidence and coverage rates are close to ideal. Its performance mirrors that of

[7]We also use a random X as opposed to a fixed X design. Binkley and Abbott (1987) argue that random X designs are in general superior to fixed X designs particularly when simulating the effects of misspecification.

normal ordered probit models. The heteroskedastic probit model, however, fares less well.

The heteroskedastic model probit performs noticeably less well than its homoskedastic counterpart. The regular probit performs as expected; in fact, we would not expect a regression on the latent y_i to be too much more efficient. But there is noticeable inefficiency in the heteroskedastic probit model. For all the parameters in the model, the estimated sampling variability tends to be over 20% too large and the coverage rates are lower well below the value of 95 we would expect. While larger samples should improve the performance, here, this is not ideal given the sample sizes normally used with these models in applied work. We might find better estimates of the standard errors using a more restrictive distribution for the model, but we assume it is unrealistic to expect such a distribution in applied data unless one happens to be fortunate. As expected, the heteroskedastic probit model exhibits inefficiency even under ideal circumstances. This inefficiency is also troubling in that it makes an applied researcher more likely to declare a parameter statistically significant than would otherwise be the case. We, next, see how the models fare under less ideal conditions.

Table 1: Relative Efficiency of Heterogenous Choice Models

SPECIFICATION		OVERCONFIDENCE[a] (%)	LEVEL[b]
Het. Probit	β_1	122	82
	β_2	116	86
	γ_1	127	83
	γ_2	123	83
Probit	β_1	104	92
	β_2	102	95
Het. Ordered Probit	β_1	92	99
	β_2	95	96
	γ_1	105	95
	γ_2	95	97
Ordered Probit	β_1	104	94
	β_2	102	94

Results are based on 1000 Monte Carlo replications with a sample size of 1000.

[a]Overconfidence $= 100\frac{\sqrt{\sum_{l=1}^{1000}(\beta-\hat{\beta})^2}}{\sqrt{\sum_{l=1}^{1000}(s.e.(\hat{\beta}))^2}}$

[b]Percentage of 95% confidence intervals that contain $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$.

2.2 Misspecification and Measurement Error

The second set of Monte Carlo experiments test how well the two models perform when either the choice or variance model is misspecified or measurement error is present. To examine the effect of misspecification, we either omit a relevant variable from either the choice or variance model. While measurement error can arise in many forms, we induce it here by using a variable in one of the estimating equations that is correlated with the true variable in the DGP. We estimated a series of models where one of the included variables is correlated with the true variable at either 0.90 or 0.70. We purposefully set the correlations between the true and proxy variables to be quite high. For example, correlations in cross-sectional data sets above 0.50 are fairly rare. With such high correlations, the tests should be friendly to the heterogenous choice models. We manipulate the specification condition across both the choice and variance model, thus giving us four different experimental conditions for the choice model and four different conditions for the variance model. In these experiments, we report the coverage rate and level of overconfidence as well as the percentage of bias.

We start by constructing our baseline with a standard probit model. This will allow us to understand whether misspecification of or measurement error in an $\mathbf{x}_i$ variable in a heterogenous choice model has consequences beyond those of misspecification or measurement error in standard probit models. To construct the baseline, we conducted a Monte Carlo analysis using a standard probit with two uncorrelated variables under the experimental conditions described above. The results are in Table 2.

The results range from serious to fairly benign. When, we included a variable that is highly correlated with the true variable, the estimates of the correctly specified variable were off by 8% while the estimates of the measurement error variable were around 19% too small. Also notice, here, that the standard errors are unaffected by the measurement error, as levels of confidence are close to ideal. Omitted variable bias is less than for larger amounts of measurement error but larger than small amounts of measurement error as the coefficient for the included variable is now attenuated by around 30%. Notice, however, that there is bias, despite the fact that X_1 and X_2 are independent of each other.

For these experiments, we also construct a second baseline. We compare the results to

Table 2: Misspecification of the Probit Model

SPECIFICATION[a]		BIAS (%)	OVERCONFIDENCE[b] (%)	LEVEL[c]
.90	β_1	-8	99	74
	β_2	-19	90	16
.70	β_1	-19	94	10
	β_2	-44	90	0
Omitted X_2	β_1	-30	137	0

Results are based on 1000 Monte Carlo replications with a sample size of 1000.

[a]Correlation between true variable in DGP and variable in the estimated model.

[b]Overconfidence $= 100\frac{\sqrt{\sum_{l=1}^{1000}(\beta-\bar{\beta})^2}}{\sqrt{\sum_{l=1}^{1000}(s.e.(\hat{\beta}))^2}}$

[c]Percentage of 95% confidence intervals that contain $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$.

standard probit models that were estimated with the same heteroskedastic DGP. Here, we ask what would have happened if we had ignored the heteroskedasticity. Would we be worse off? Of course, this comparison is of limited usefulness since often analysts care about the γ estimates which are obviously lacking in a standard probit model. We present the result from this baseline in Table 3

Table 3: Ordered Probit and Probit Model With Heteroskedastic Data

SPECIFICATION		BIAS (%)	OVERCONFIDENCE[a] (%)	LEVEL[b]
Probit	β_1	-50	190	0
	β_2	-50	210	0
Ordered	β_1	-48	145	0
Probit	β_2	-48	162	0

Results are based on 1000 Monte Carlo replications with a sample size of 1000.

[a]Overconfidence $= 100\frac{\sqrt{\sum_{l=1}^{1000}(\beta-\bar{\beta})^2}}{\sqrt{\sum_{l=1}^{1000}(s.e.(\hat{\beta}))^2}}$

[b]Percentage of 95% confidence intervals that contain $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$.

As is evident in Table 3, neither model performs particularly well. The coefficients are about half their true size and the standard errors are highly inflated. The question as we review the results is whether the bias caused by misspecification or measuremente error is any larger than the bias we see here. We now turn to the results from the Monte Carlo analyses. Since there will be four sets of results for each of the probit and ordered probit models, we present the results from each separately starting with the probit models. How does this compare to a similar amount of measurement error in an $\mathbf{x}_i$ variable in a heteroskedastic probit?

2.2.1 Heteroskedastic Probit Models

Table 4 contains the results from our Monte Carlo analysis of misspecification and measurement error in the heteroskedastic probit model. We, first, consider misspecification and measurement error in the choice model. Here, the same amount of measurement as we used in the baseline models produces much larger biases. Now, the estimates of *every* parameter in the model are biased by 60-70%. Moveover, the 95% confidence interval never contains the true parameter value and the standard errors are too confident in some cases by over 200%. So unlike in the standard probit, where the standard errors were close to correct, here, they are poorly estimated.

The heteroskedastic probit model is clearly more sensitive to the specification and measurement error than the standard probit upon which it is based. But with a standard probit model, we cannot incorrectly specify the variance model as we can with the heteroskedastic probit model. That said, the addition of a variance model allows for the possibility of providing a cure for heteroskedasticity. We, now, focus on the quality of the estimates when the variance model is misspecified.

The consequences of misspecification and measurement error in the variance model are noticeably different than what we saw with the choice model. Here, as seen in Table 4, the effect of measurement error and omitted variables is more pronounced on our ability to make inferences. To be sure, misspecifying the variance model does induce bias. For example, in the case where the estimating equation of the variance model contains a variable correlated with the true variable at 0.90, both the $\boldsymbol{\beta}$'s in the choice model are off by 36%, and the $\boldsymbol{\gamma}$'s are biased by 19% and 28% respectively. But the reduction in the bias comes at the expense of the estimates of the sampling variability. Here, for every parameter, the standard errors are overconfident by over 200%, and the 95% confidence intervals rarely contain the true parameter value.

Decreasing the correlation between the estimating variable and the true variable to 0.70 has

Table 4: Misspecification and Measurement Error in the Heteroskedastic Probit Model

SPECIFICATION[a]		BIAS (%)	OVERCONFIDENCE[b] (%)	LEVEL[c]
		Choice Model		
.90	β_1	-61	167	0
	β_2	-65	145	0
	γ_1	-70	164	0
	γ_2	-70	219	0
.70	β_1	-70	154	0
	β_2	-79	138	0
	γ_1	-80	151	0
	γ_2	-80	175	0
Omitted X_2	β_1	-76	137	0
	γ_1	-85	138	0
	γ_2	-86	137	0
		Variance Model		
.90	β_1	-36	216	7
	β_2	-36	221	7
	γ_1	-19	215	37
	γ_2	-28	200	8
.70	β_1	-56	207	0
	β_2	-56	219	0
	γ_1	-42	246	5
	γ_2	-60	223	0
Omitted Z_2	β_1	-66	189	0
	β_2	-65	203	0
	γ_1	-61	203	0

Results are based on 1000 Monte Carlo replications with a sample size of 1000.

[a]Correlation between true variable in DGP and variable in the estimated model.

[b]Overconfidence $= 100\frac{\sqrt{\sum_{l=1}^{1000}(\beta-\bar{\beta})^2}}{\sqrt{\sum_{l=1}^{1000}(s.e.(\hat{\beta}))^2}}$

[c]Percentage of 95% confidence intervals that contain $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$.

a pronounced effect on the amount of bias in the model. The bias increases by 20% for both the estimates of the $\boldsymbol{\beta}$'s and $\boldsymbol{\gamma}$'s. The estimates of the sampling variability continue to worsen but by a small margin, and the coverage rates remain poor. We also find that in general ignoring the heteroskedasticity is better than using a variance model. Omitting a relevant variable from the variance model induces bias worse than that found when we ignored the unequal error variances. Again this matters little if one is interested in the estimates of γ, but suggests that for estimates of β ignoring the heteroskedasticity is a reasonable solution.

Bias and Response Probabilities Given the nonlinear link function used in heteroskedastic probit models, bias in the parameter estimates does not directly translate into biased estimates of how a change in $\mathbf{x}_i$ affects the probability that $y_i = 1$. Wooldridge (2001) recommends comparisons of the response probabilities across various values of $\mathbf{x}_i$.[8]

That is, even though the true size of β's are always underestimated, we might, in fact, overestimate the change in the probability of $y_i = 1$ for a change in $\mathbf{x}_i$. To understand how the bias affects the probability that $y_i = 1$ given changes in one of the x_i variables, we calculated the predicted probabilities of y_i given changes in the x_2 variable for the model where one of the variables in the estimated choice model is correlated with the true variable at 0.90.

In Figure 1, we plot the predicted probability that $y_i = 1$ as x_2 changes from -1 to 1 with the rest of the variables in the model held constant at their means. In the figure, we plot the predicted probabilities from the true model against the predicted probabilities against an estimated model. The coefficient values for the estimated model are the average of the 1000 estimates from the Monte Carlo analysis.

As the reader can see, the effect of the bias depends on the value of x_2. The differences in predicted probabilities can be quite large for negative values of x_2 as the estimated predicted probabilities are too large by about 0.50. While the two predicted probabilities are identical for one value of x_2, for larger values of x_2 the predicted probabilities are too small. So despite the fact that the estimated β's are too small, depending on the values of x_2 the predicted probability for y_i may be either too small or too large.

[8]One can also compare marginal effects across various values of $\mathbf{x}_i$. We computed the marginal effects across the models, and while not directly comparable to the predicted probabilities, we didn't find the discrepancies between the true and estimated marginal effects to be any smaller or larger than those in the predicted probabilities.

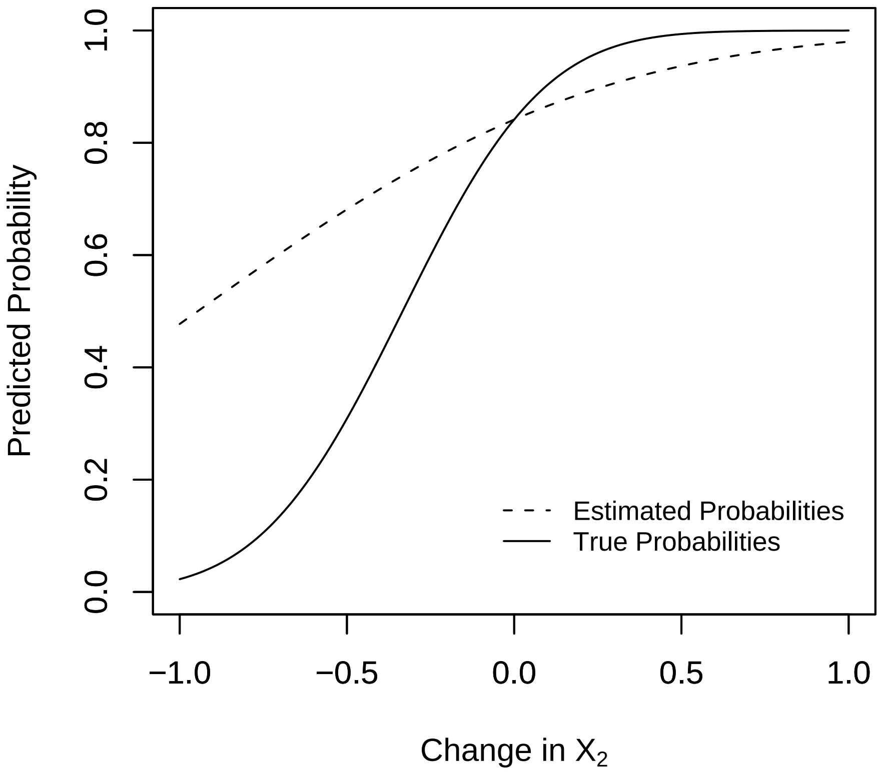

Figure 1: True Versus Estimated Probability that $y = 1$

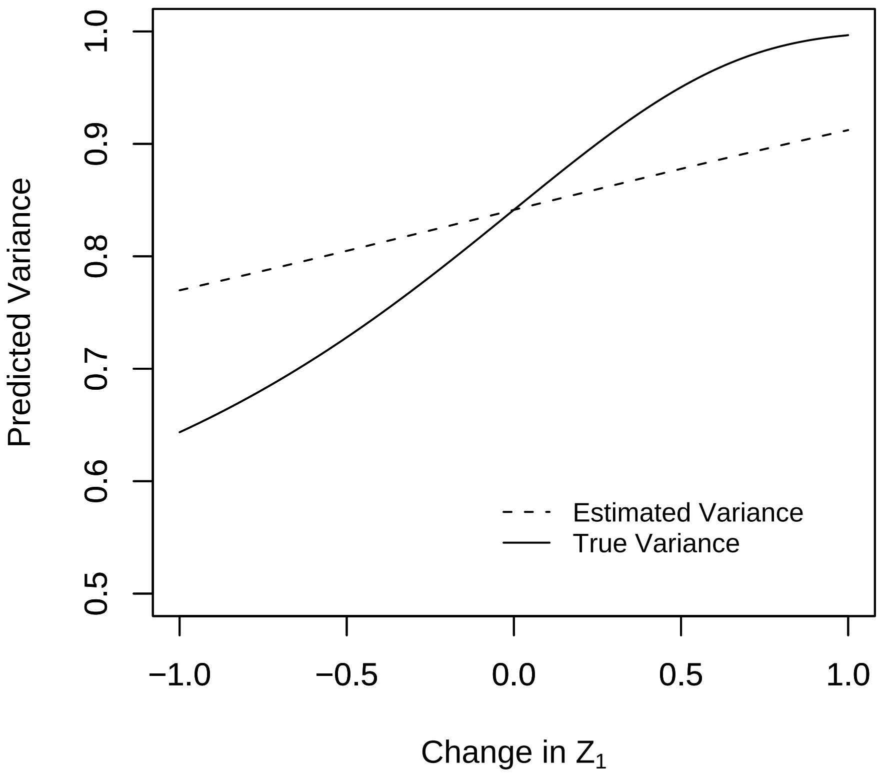

Figure 2: True Versus Estimated Predicted Variance of y_i

We, next, plot the predicted variance of y_i for both the estimated and true models. Here, we plot how the variance for y_i changes as we change z_1 from -1 to 1 with the rest of the variables in the model held constant. In Figure 2, we see a similar pattern where for negative values of $\mathbf{z}_1$ we under-predict the variance of y_i, but for positive values of z_1 we over-predict the variance of y_i. Both figures emphasize the incorrect inferences that are possible in the face of minor specifications of the heteroskedastic probit model. While the size of the coefficients are generally attenuated, due to the non-linear nature of the model, one can both over and underestimate the quantity of substantive interest.

In sum, it is not only the point estimates that are affected but also the resulting probabilities from these models. Moreover, despite the fact that the bias attenuates the parameters, the predicted probabilities may be too large. We, next, turn to the results for the heteroskedastic ordered probit model, which, thus far, has performed somewhat better than the heteroskedastic probit model.

2.2.2 Heteroskedastic Ordered Probit Models

In the first analysis, the ordered probit model performed better than the probit model. But, here, that is not the case, as the differences between the ordered probit model and the probit model when misspecified or measurement error is present are minimal. Table 5 contains the results from when the heteroskedastic ordered probit model is misspecified or contains measurement error.

When the choice model estimating equation contains a variable correlated with the true variable at 0.90, all the parameters are biased by at least 60%. For a similar amount of measurement error in a standard ordered probit model, the bias would be 20 percentage points lower. As the level of correlation between the estimating variable and the true variable decreases, the bias increases.

In Table 5, we also see the results of when the variance model is either misspecified or contains measurement error, while the amount of bias is half that of when the choice model has measurement error or is misspecified, the estimated sampling variability tends to be twice as large. As before, when the probit variance model contained errors, not only does it bias the parameters (albeit at a lower level) but causes the estimated sampling variance to be substantially larger than the true sampling variance. Moreover, if one only cared about the estimates of the β's it would be better to estimate a standard probit model and ignore the heteroskedasticity than leave a relevant variable out of the variance model.

Bias and Response Probabilities Again to gain a better sense of how the bias induced by misspecification and measurement error affects the inferences in the heteroskedastic ordered probit model, we plot the predicted probabilities for the model with the least amount of measurement error. Since the dependent variable we use in the analysis has four different categories there are four different predicted probabilities that we can plot. Here, we plot the results for the probability that y_i equals 2 and 4.

In Figure 3, we plot the predicted probability that $y_i = 2$. The heteroskedastic ordered

Table 5: Misspecification and Measurement Error in the Heteroskedastic Ordered Probit Model

SPECIFICATION[a]		BIAS (%)	OVERCONFIDENCE[b] (%)	LEVEL[c]
		Choice Model		
.90	β_1	-61	115	0
	β_2	-65	106	0
	γ_1	-72	127	0
	γ_2	-72	168	0
.70	β_1	-71	106	0
	β_2	-79	101	0
	γ_1	-80	108	0
	γ_2	-80	128	0
Omitted X_2	β_1	-77	85	0
	γ_1	-84	109	0
	γ_2	-84	87	0
		Variance Model		
.90	β_1	-33	126	35
	β_2	-33	120	32
	γ_1	-19	136	83
	γ_2	-26	130	43
.70	β_1	-54	171	0
	β_2	-54	174	0
	γ_1	-39	203	3
	γ_2	-58	199	0
Omitted Z_2	β_1	-65	144	3
	γ_1	-64	162	0
	γ_2	-57	173	0

Results are based on 1000 Monte Carlo replications with a sample size of 1000.

[a]Correlation between true variable in DGP and variable in the estimated model.

[b]Overconfidence $= 100\frac{\sqrt{\sum_{l=1}^{1000}(\beta-\bar{\beta})^2}}{\sqrt{\sum_{l=1}^{1000}(s.e.(\hat{\beta}))^2}}$

[c]Percentage of 95% confidence intervals that contain β or $\boldsymbol{\gamma}$.

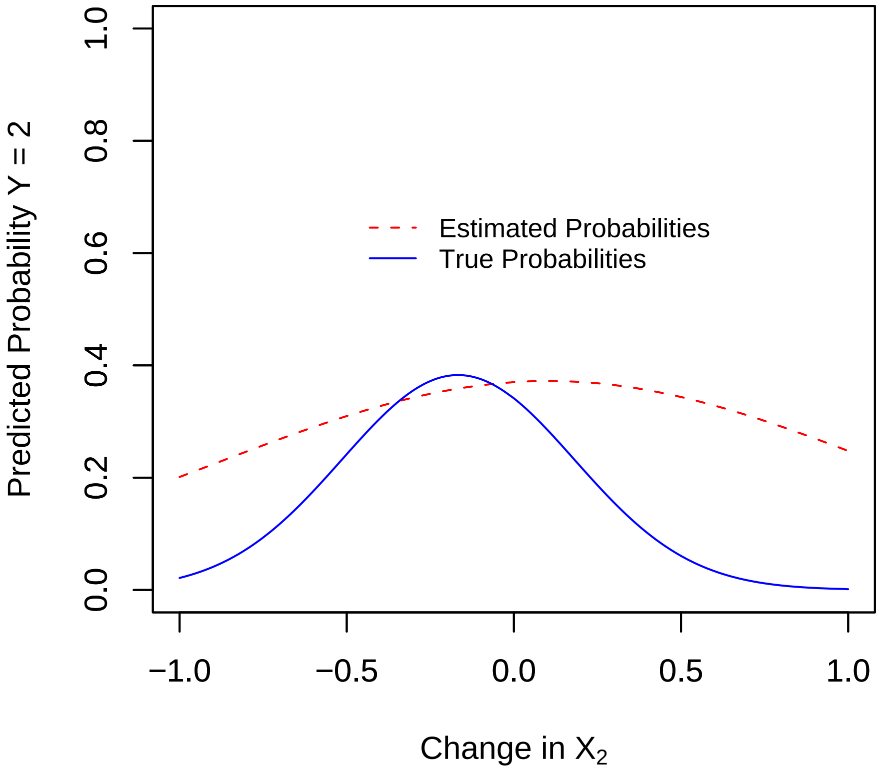

Figure 3: True Versus Estimated Predicted Probability that $y = 2$

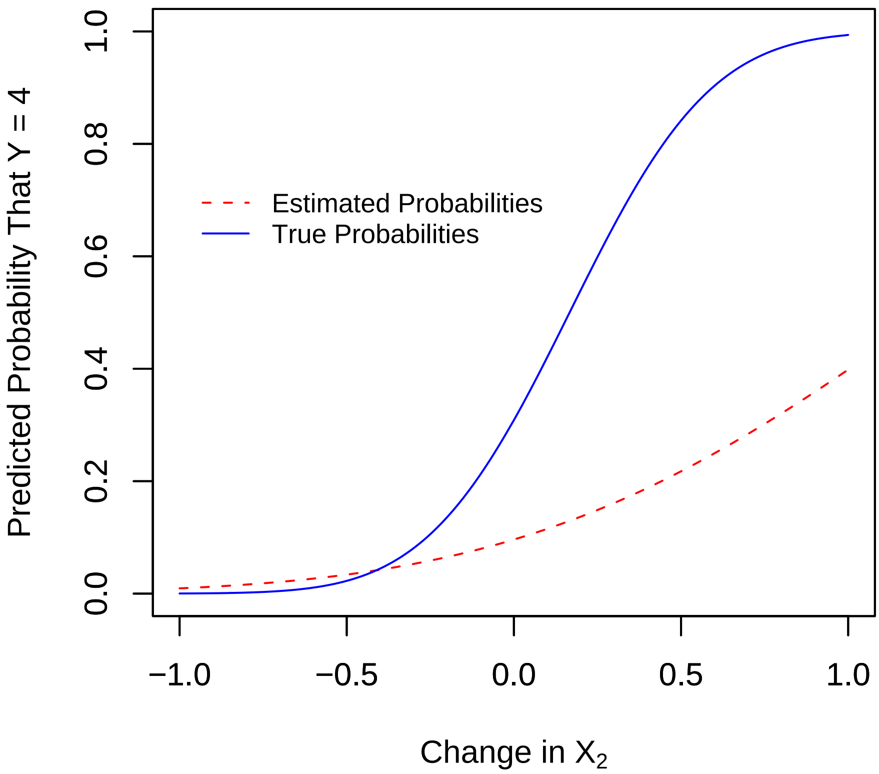

Figure 4: True Versus Estimated Predicted Probability that $y = 4$

probit model with measurement error provides obviously incorrect estimates of the predicted probability that $y_i = 2$. Except for a narrow range of values, the estimated predicted probabilities are typically much higher than the true values. So, here, despite estimated values for the β's that are too small, we tend to estimate predicted probabilities that are too high. In Figure 4, we plot the predicted probability that $y_i = 4$. Here, we see the opposite problem, in that for most values of $\mathbf{x}_i$, the predicted probabilities are too low. In fact, for some value of $\mathbf{x}_i$, the difference is as large as 0.6. The figures provided further evidence that a minor amounts of measurement error in the model can lead to incorrect inferences about the effects of the independent variables on y_i.

3 Implications For Empirical Models

Finally, we use the results from the Monte Carlo experiments to shed light on work published with heterogenous choice models. We have shown that heterogenous choice models are extremely sensitive to misspecification and measurement error and that the heteroskedastic probit model is fairly inefficient even under ideal conditions. While we cannot know published heteroskedastic models are misspecfied or contain measurement error, we can assume that the models have poorly estimated standard errors, and that researchers need to correct their standard errors. To illustrate, we use the data from the seminal article which introduced these models to the political science literature (Alvarez and Brehm 1995).

The authors test whether value conflict and firmness of opinion affects the variance in opinions on different aspects of abortion. In particular, the authors look for signs that value conflict causes increased variance in the opinions on abortion. They found that value conflict did prove to be an important consideration in choices on abortion. This was the first demonstration that the public struggles with policy choices due to ambivalence. And one could argue that this paper helped launch research on the study of ambivalence in public opinion, a small but burgeoning area of current research.

To find evidence of ambivalence, they estimate heteroskedastic models for seven different questions about abortion. The focus of the statistical models is the number of statistically significant variables in the variance model. Given the evidence from the Monte Carlo experiments, we would expect the standard error estimates in these models to be overconfident, making it more likely that we observe a statistically significant effect. This may be true to a greater extent than we observed in our analyses since many of the variables in this analysis are correlated and categorical, while in the Monte Carlo analysis we used independent continuous covariates. In short, given what we now know about these models, it should be too easy to find statistical evidence of ambivalence. We can, however, use bootstrapping to re-estimate the standard errors for these models.

Bootstrapping is a useful technique for estimating standard errors without relying on the distributional properties of a statistical model (Efron and Tibshirani 1993). We can use bootstrapping to recalculate both the standard errors and the confidence intervals and test whether the underestimation of the standard errors might affect the inferences made from this set of models. We used a nonparametric random-X bootstrap with 1000 bootstrap resamples to calculate the standard errors and confidence intervals.[9] With bootstrapping, we repeatedly resample from the data with replacement and estimate the model with the resampled dat set each time to create a sampling distribution for the heteroskedastic probit model parameters. Using the bootstrapped standard errors, we recalculated the t-statistics for the published models.[10] While we recalculated the standard errors for all the variables in the model, we focus the analysis on the variables in the variance model, since the authors only have predictions from theory about the variables in the variance model. This analysis allows to understand whether the inferences made with the original models were overly optimistic.

Table 6 contains the number of times a variable was statistically significant across the seven models estimated with both the original standard errors and the bootstrapped standard errors. In the original analysis, the seven variables were significant a total of 22 times; with the bootstrapped model, that is reduced to 13 times. While the average reduction in the t-statistics was around 0.40 some dropped by as much as 1.20. In particular, the theoretically important value conflict measure is now significant 3 out of 7 times instead of 5 out of 7 times. It would appear that value conflict driven ambivalence is substantially less pervasive than first thought. Clearly the poor standard errors found in heteroskedastic probit models can have a substantial effect on our ability to make inferences.

[9] We could have also used what is called a fixed-X bootstrap and resampled residuals. However, as Efron and Tibshirani (1993) demonstrate, random-X bootstrapping is less sensitive to assumptions. Moreover, fixed-X bootstrapping is mostly useful when we care about capturing some aspect of the error term, which, here, we are not.

[10] To calculate whether a variable was statistically significant, we simply divided the coefficient estimated by Alvarez and Brehm by the new bootstrapped standard error and compared the t-statistics. This process assumes that the data are normally distributed. Given the sample sizes for these models, this is not an unreasonable assumption. A better method is to compare the bootstrapped confidence intervals to the estimated confidence intervals. We found the confidence interval comparison to be more conservative than using the bootstrapped standard errors as there were a greater number of bootstrapped confidence intervals that contained zero than original estimated confidence intervals. We compared both bias corrected (BC_a) and normal percentile confidence intervals to the estimated confidence intervals.

Table 6: Replication of Alvarez and Brehm: Bootstrapping the Standard Errors

Independent Variable	Original Model	Bootstrapped Model
Pro Count	3	2
Con Count	7	4
Pro Count × Con Count	5	3
Importance	0	0
Information	2	2
Firmness of Opinion	5	3

Nonparametric bootstrap with random-X resampling.
Cell entries represent number of times variable
is significant at 0.05 level
1000 bootstrap resamples.

4 Conclusion

To understand the variation in choices is often as important as understanding the choice itself. Understanding heterogeneity in choices and behavior has become a fertile area of research. But given our findings, we offer several recommendations.

First, the poor performance of the heteroskedastic probit also implies that one should generally refrain from taking an ordered measure and dichotomizing it for use in a heteroskedastic probit instead of running a heteroskedastic ordered probit model. Why would anyone do this? Well, for one, heteroskedastic probit is implemented in Stata but heteroskedastic ordered probit is only available in Limdep or requires coding and optimizing your own likelihood.

Second, the standard errors from heteroskedastic probit models should not be relied upon. The standard errors from these models are overly optimistic and can lead to incorrect infernces. As we demonstrated in our reanalysis of Alazarez and Brehm, much of the evidence for ambivalence toward abortion vanishes once the standard errors are corrected. Bootstrapping provides a simple and convenient correction for this problem. Bootstrapping procedures are easy to implement in standard software packages such as Stata allowing for a straightforward corrective.

The results, here, emphasize the need to pay attention to the specification of *both* the so-called choice and variance models. This is important given how these models are typically used. Most often, these models are used to test theories about what causes variation in a choice setting. The focus on the variance model can cause the choice model to be of secondary interest. But as the evidence, here, demonstrates, even if the choice model is of secondary interest its specification cannot be. Misspecification or measurement error in the choice model was just as serious or worse than the same errors in the variance model. Moreover, whatever specification is adopted, analysts should demonstrate through rigorous sensitivity testing that the results are not sensitive to specification.

Researchers should also take great care with the measurement of the variables used in the analysis. While this is good advice for all models, we have seen that fairly benign measurement error has serious consequences. Given that these models are frequently used with survey data, scales should be used instead of single survey items wherever possible to minimize measurement error.

If researchers are only interested in the parameters from the choice model, but suspect heterskedasticity, these models may not be the best alternative. If the error variance differs across well defined groups, specification of the variance model should be relatively easy. But if the source of the heteroskedasticity is less clear and harder to specify, it is better to estimate a standard probit and ignore the heteroskedasticity than poorly specify a heteroskedastic model.

The findings, here, must also revise our substantive knowledge about politics. Alvarez and Brehm's models of abortion are widely cited as evidence that the public experiences value conflict. However, our reanalysis suggests that the inferences drawn from the model were based on incorrect standard errors. Of the measures that we found to be no longer be statistically significant, most were measures of value conflict. Future work should devise an alternative way to model heterogenous choices, but until then analysts must use care with heterogenous choice models.

References

Achen, Christopher H. 2002. "Toward a New Political Methodology: Microfoundations and ART." *Annual Review of Political Science* 5:423–450.

Alvarez, R. Michael and Charles Franklin. 1994. "Uncertainty and Political Perceptions." *Journal of Politics* 56:671–689.

Alvarez, R. Michael and John Brehm. 1995. "American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values." *American Journal of Political Science* 39:1055–1082.

Alvarez, R. Michael and John Brehm. 1997. "Are Americans Ambivalent Towards Racial Policies." *American Journal of Political Science* 41:345–374.

Alvarez, R. Michael and John Brehm. 1998. "Speaking In Two Voices: American Equivocation About the Internal Revenue Service." *American Journal of Political Science* 42:418–452.

Alvarez, R. Michael and John Brehm. 2002. *Hard Choices, Easy Answers.* Princeton, NJ: Princeton university Press.

Binkley, James K. and Philip C. Abbott. 1987. "The Fixed X Assumption in Econometrics: Can The Textbooks Be Trusted?" *The American Statistician* 41:206–214.

Busch, Marc L. and Eric Reinhardt. 1999. "Industrial Location and Protection: The Political and Economic Geography of U.S. Nontariff Barriers." *American Journal of Political Science* 43:1028–1050.

Davidson, Russell and James G. MacKinnon. 1993. *Estimation and Inference in Econometrics.* New York: Oxford University Press.

Efron, Bradley and Robert J. Tibshirani. 1993. *An Introduction To The Bootstrap.* Boca Raton: Chapman & Hall/CRC.

Gabel, M. J. 1998. "Economic Integration and Mass Politics: Market Liberalization and Public Attitudes in the European Union." *American Journal of Political Science* 42:936–953.

Greene, William H. 1993. *Econometric Analysis.* New York: Macmillan.

Harvey, Andrew. 1976. "Estimating Regression Models with Multiplicative Heteroskedasticity." *Econometrica* 44:461–465.

Krutz, Glen S. 2005. "Issues and Institutions: Winnowing in the U.S. Congress." *American Journal of Political Science* 49:436–447.

Lee, Taeku. 2002. *Mobilizing Public Opinion: Black Insurgency and Racial Attitudes in the Civil Rights Era.* Chicago: University of Chicago Press.

Lemke, Douglas. 2002. *Regions of War and Peace.* Cambridge: Cambridge University Press.

Maddalla, G.S. 1983. "Limited-Dependent and Qualitative Variables in Econometrics." *Cambridge: Cambridge University Press* .

McCullagh, P. and J.A. Nelder. 1989. *Generalized Linear Models*. 2 ed. Boca Raton: Chapman & Hall/CRC.

Ruud, Paul, A. 2000. *An Introduction to Classical Econometric Theory*. Oxford: Oxford University Press.

Wooldridge, Jeffrey M. 2001. *Econometric Analysis of Cross Section and Panel Data*. Boston: The MIT Press.

Yatchew, A. and Z. Griliches. 1985. "Specification Error In Probit Models." *Review of Economics and Statistics* 18:134–139.

A Additional Analyses

We also studied the effects of several other possible misspecifications of the heteroskedastic model. We first tested the effect of having the same variable in both the choice and variance model. Here, the model performed well. However, the specifications in the analysis is much simpler than what would be found in applied work, and there are, of course, no guarantees that this might not be problematic with more complicated specifications.

We also examined the effect of having extraneous variables in either the choice or variance model. Here, the standard errors were slightly inflated, but not at a level any higher than we would expect for extraneous variables.

Yet another possibility we explored was the effect of estimating heteroskedastic models with homoskedastic data. Since the coefficients on the variables in the variance model should be zero, the model should be equivalent to a standard probit model. However, while the estimates were not biased, the standard errors were noticeably overconfident. For the probit model, the measure of overconfidence was around 165% for the β's and around 140% for the γ's. Again, however, the ordered model had no trouble when fitted with an extraneous variance model. Researchers should use a likelihood ratio test to detect that the heteroskedastic model is not an improvement over a standard model since using the heteroskedastic model on normal data is not without consequences.

We also replicated the analysis using categorical variables instead of the the continuous variables used in the analyses. Specifically, we made $\mathbf{x}_1$ and $\mathbf{z}_1$ categorical and $\mathbf{x}_2$ and $\mathbf{z}_2$ continuous (both $\mathbf{x}_1$ and $\mathbf{z}_1$ had four categories) for both the probit and ordered probit models. While the RMSE for the probit model with continuous variables fell to 0.20 with 1000 cases, the model with categorical variables remained high at 0.65. The ordered probit model produced similar results, where the model with continuous variables fell to 0.14 with 1000 cases, the categorical model remained at 0.52.

Finally, we also studied the effect of sample size and additive measurement error. We found that sample sizes of at least 500 cases were need to get consistent estimates. The effect of additive measurement error was similar the measurement error model used in the reported analyses.

B Monte Carlo Code

```
##Heteroscedastic Probit for Monte Carlo With Misspecification Error
##In The Choice Model

#Define Likelihood
lik.hetprobit <-function(par, X, Y, Z){

#Pull Out Parameters
Y <- as.matrix(y)
X <- as.matrix(x)
Z <- as.matrix(z)
K <-ncol(X)
M <-ncol(Z)
```

```
b <- as.matrix(par[1:K])
gamma <- as.matrix(par[K+1:M])
mu <- (X%*%b)
sd <- exp(Z%*%gamma)
mu.sd <-(mu/sd)

#Form Likelihood
   log.phi <- pnorm(ifelse(Y == 0, -1, 1) * mu.sd, log.p = TRUE)
   2 * sum(log.phi)
}

##########################Monte Carlo Starts Here#######################
# Number of Experiments
i <- 1000
# Number of Cases
n <- 1000
#Correlation Btw True and Specifed Variables
r <- .90

#Define Empty Matrices
beta.na <- matrix(NA,i,5)
stder.na <- matrix(NA,i,5)
convg.na <- matrix(NA,i,1)

set.seed(736451
)
#Set Counter And Start
Loop j <- 1 for(j in 1 : i){


#Simulate Data
z1 <- rnorm(n)
z2 <- rnorm(n)
x4 <- rnorm(n)

#Create Correlated Variables
d <- chol(matrix(c(1,r,r,1), 2,2))
x1 <- rnorm(n)
x2 <- rnorm(n)
x3 <- r*x2 + d[2,2]*x1
latentvar <- exp(1.0 * z1 + 2.0 * z2)
latentz <- 1.0 + 2.0 * x4 + 3.0 * x2 + rnorm(n, mean=0, sd=latentvar)
y <- latentz
y[latentz < 1] <- 0
y[latentz >=1] <- 1
x <- cbind(1, x4, x3)
z <- cbind(z1, z2)

#Fit Model
```

```
hetresult <- optim(c(1,0,0,0,0), lik.hetprobit, Y=y, X=x, Z=z, method="BFGS",
                   control=list(maxit=2000, fnscale=-1),hessian=T)
beta.na[j,] <-hetresult$par
stder.na[j,] <-sqrt(diag(solve(-hetresult$hessian)))
convg.na[j,] <-hetresult$convergence

j <- j+1
}
#Analyze Results
s1 <- matrix(0,i,1)
s2 <- matrix(2,i,1)
s3 <- matrix(3,i,1)
s4 <- matrix(1,i,1)
s5 <- matrix(2,i,1)
se <- cbind(s1,s2,s3,s4,s5)
#RMSE
rmse <- sqrt(mean((beta.na-se)^2))
cat("RMSE: ", rmse, "\n")

#Calculate Bias In Parameters
bias.b1 <- (mean(beta.na[,2] - 2))
bias.b2 <- (mean(beta.na[,3] - 3))
bias.z1 <- (mean(beta.na[,4] - 1))
bias.z2 <- (mean(beta.na[,5] - 2))

cat("Average Bias in Beta1: ", bias.b1, "\n")
cat("Average Bias in Beta2: ", bias.b2, "\n")
cat("Average Bias in Gamma1: ", bias.z1, "\n")
cat("Average Bias in Gamma2: ", bias.z2, "\n")
cat("Mean of Beta_1 Hat: ", mean(beta.na[,2]), "\n")
cat("Mean of Beta_2 Hat: ", mean(beta.na[,3]), "\n")
cat("Mean of Gamma_1 Hat: ", mean(beta.na[,4]), "\n")
cat("Mean of Gamma_2 Hat: ", mean(beta.na[,5]), "\n")

#Calculate Percentage of Bias
b1 <- 2
b2 <- 3
g1 <- 1
g2 <- 2
Beta1.Probit.Pct <- (1-((b1-bias.b1)/b1))*100
Beta2.Probit.Pct <- (1-((b2-bias.b2)/b2))*100
Gamma1.Probit.Pct <- (1-((g1-bias.z1)/g1))*100
Gamma2.Probit.Pct <- (1-((g2-bias.z2)/g2))*100
cat("% of Bias in Beta_1 Hat: ", Beta1.Probit.Pct, "\n")
cat("% of Bias in Beta_2 Hat: ", Beta2.Probit.Pct, "\n")
cat("% of Bias in Gamma_1 Hat: ", Gamma1.Probit.Pct, "\n")
cat("% of Bias in Gamma_2 Hat: ", Gamma2.Probit.Pct, "\n")

#Convergence
```

```
cat("Number Of Times Failed to Converge:", sum(convg.na), "\n")

#Coverage Rate
ci.plus <- beta.na[,2] + 1.96*stder.na[,2]
ci.neg <- beta.na[,2] - 1.96*stder.na[,2]
covg <- as.numeric(ci.neg < s2 & s2 < ci.plus)
cat("Coverage Rate Beta_1: ", mean(covg),"\n")
ci.plus <- beta.na[,3] + 1.96*stder.na[,3]
ci.neg <- beta.na[,3] - 1.96*stder.na[,3]
covg <- as.numeric(ci.neg < s3 & s3 < ci.plus)
cat("Coverage Rate Beta_2: ", mean(covg),"\n")
ci.plus <- beta.na[,4] + 1.96*stder.na[,4]
ci.neg <- beta.na[,4] - 1.96*stder.na[,4]
covg <- as.numeric(ci.neg < s4 & s4 < ci.plus)
cat("Coverage Rate Gamma_1: ", mean(covg),"\n")
ci.plus <- beta.na[,5] + 1.96*stder.na[,5]
ci.neg <- beta.na[,5] - 1.96*stder.na[,5]
covg <- as.numeric(ci.neg < s5 & s5 < ci.plus)
cat("Coverage Rate Gamma_2: ", mean(covg),"\n")

#Confidence Measure
con.beta.1 <- 100 * sqrt(mean((beta.na[,2]-mean(beta.na[,2]))^2))
/sqrt(mean(stder.na[,2]^2))
con.beta.2 <- 100 * sqrt(mean((beta.na[,3]-mean(beta.na[,3]))^2))
/sqrt(mean(stder.na[,3]^2))
con.gamma.1 <- 100 * sqrt(mean((beta.na[,4]-mean(beta.na[,4]))^2))
/sqrt(mean(stder.na[,4]^2))
con.gamma.2 <- 100 * sqrt(mean((beta.na[,5]-mean(beta.na[,5]))^2))
/sqrt(mean(stder.na[,5]^2))
cat("Confidence of Beta_1: ", con.beta.1, "\n")
cat("Confidence of Beta_2: ", con.beta.2, "\n")
cat("Confidence of Gamma_1: ", con.gamma.1, "\n")
cat("Confidence of Gamma_2: ", con.gamma.2, "\n")
```