

Difficult Choices: An Evaluation of Heterogenous Choice Models *

Luke Keele

Department of Politics and International Relations

Nuffield College and Oxford University

Manor Rd, Oxford

OX1 3UQ UK

Tele: +44 1865 278716

Email: luke.keele@politics.ox.ac.uk

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 17, 2005

*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. We present analytical results that suggest these models are both inefficient and subject to larger than expected bias when misspecified or estimated with measurement error. 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 error 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 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.

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; Smith et al. 1999; 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.

Modeling heterogeneity takes a different view of which estimated parameters in a statistical model are substantively interesting. For example, let's assume that each response or behavior we observe is a random draw from an unobserved latent sampling distribution. This latent sampling distribution is assumed to have both first and second order moments, such that if this latent sampling distribution were observed we could calculate a mean and variance. If we observe two different choices, let's say one respondent supports abortion while another is opposed to it, it is assumed that the means of these respondents' latent sampling distributions are different. Typically statistical models in the social sciences seek to explain why one respondent made a different choice as compared to another respondent. Another way to think of this is that we only care to explain why the means differ across the sampling distributions, or that the mean is the only parameter of substantive interest.

If we could observe the sampling distributions for the respondents in our sample, it is entirely possible that not only will the means differ across respondents but the variances will differ as well. That is for two hypothetical respondents, if we were to calculate the variance, σ^2 , for the sampling distributions associated with their attitudes on abortion, we might find that they differ as well. Alvarez and Brehm (1995) offer two plausible explanations for why this might be the case in the context of opinions on abortion to demonstrate that σ^2 is more than a nuisance parameter but also substantively interesting. They argue that both differences in political sophistication and value conflict may account for increased variability in the responses we observe.

Public opinion on abortion is but one example of a heterogenous choice, which occur when some subset of our sample has greater unobserved variance in their choices than the rest of the sample. In statistical terms, we typically estimate the following model: $Y \sim \eta(\theta, \sigma^2)$, where θ and σ^2 are parameters to be estimated, with η being some link function. Normally, the theory is a test of some restriction on θ , as σ^2 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, as in the example above, the variability in the choice may be more interesting than what determines the choice. 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. 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.

Such heterogeneity may be the product of different levels of information about a choice. For example, 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). Diffusion studies is another place where we might observe heterogenous choices, in that states might have differing propensities to adopt policy innovations.

At present, heteroskedastic probit and heteroskedastic ordered probit models are the tools of choice when investigating heterogenous choices. The attraction of these models, beyond being a convenient cure for heteroskedasticity, 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. 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. We, first, define y_i as a discrete dependent variable with j categories. Here, we consider the special case where $j = 1$, however, the following derivations apply to higher values of j . Even if a researcher only considers a theory about the nature of σ^2 , he or she must model the choice or behavior represented by y_i . Since we only consider the special case where $j = 1$, to model the $\Pr(y_i = 1)$, we would use the following probit model:

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

The probability that y_i equals 1 is a function of a vector of explanatory variables represented by \mathbf{x}_i . $\boldsymbol{\beta}$ is the vector of parameters which describe the probability that $y_i = 1$ given a change

in \mathbf{x}_i . And Φ represents the cumulative normal distribution, the nonlinear link function, which ensures that our linear predictions fall between the interval 0-1.

To estimate a probit model, we must assume that the model error variances are constant (that is we must assume the errors are homoskedastic). We incorporate this assumption into the model by dividing the $\mathbf{x}_i\boldsymbol{\beta}$ matrix by σ , the standard deviation of the error distribution. Therefore, the probability that y_i equals one is actually:

$$\Pr(y_i = 1) = \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) \quad (2)$$

This assumption is critical for two reasons. First, if we do not assume that σ is constant the probit model will not be identified. Second, 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). Normally, we assume that σ equals 1, a convenient normalization which allows σ to drop out of the equation.

In the context of heterogenous choices, σ is known or expected to vary systematically. 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. Estimating a regular probit in this situation will yield biased and inconsistent estimates.

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 error variance:

$$\text{Var}(\varepsilon_i) = \sigma_i^2 = \exp(\mathbf{z}_i\boldsymbol{\gamma})^2 \quad (3)$$

where \mathbf{z}_i is a vector of covariates of the i th observation that define groups with different error variances in the underlying latent variable, and $\boldsymbol{\gamma}$ is a vector of parameters to be estimated. By taking the positive square root of (3), we have a model for the standard deviation of the error

distribution:

$$\sigma_i = \exp(\mathbf{z}_i\boldsymbol{\gamma}). \quad (4)$$

So, we now divide $\mathbf{x}_i\boldsymbol{\beta}$ by the right-hand side of Equation 4. The probability function for a particular observation now equals:

$$\Pr(y_i = 1) = \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\exp(\mathbf{z}_i\boldsymbol{\gamma})}\right) \quad (5)$$

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

$$\ln L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}|Y) = \sum_{y=1}^N \left(y_i \ln \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\exp(\mathbf{z}_i\boldsymbol{\gamma})}\right) + (1 - y_i) \ln \left[1 - \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\exp(\mathbf{z}_i\boldsymbol{\gamma})}\right) \right] \right) \quad (6)$$

Maximizing (6) 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.¹ 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 .²

This variance model, as it is called, serves two purposes. First, it explains the unequal variances and thus cures the heteroskedasticity in the model which should give us consistent estimates of the β 's. Second, the vector of $\boldsymbol{\gamma}$ parameters allows us to test hypotheses about the sources of choice heterogeneity. For example, if we have two measures thought to affect the variability in y_i . If the $\hat{\boldsymbol{\gamma}}$ parameter for one of the measures is positive and statistically significant, then we can say that measure is a significant predictor of the variance of y_i and for each increase in the level of that measure the variance in y_i will increase. If the second measure is negative and statistically significant then we can say the variance in y_i decreases as the level of the second measure increases. As such, we can test between two competing hypotheses about the variance in y_i .

The effect that \mathbf{x}_i has on the probability that y_i equals 1 is now conditional on the variance model. This is easily seen, in that if the value of the variance model is large, it will shrink the estimate of β . As the values of the variance model is smaller it decreases the effect of β by a

¹The 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.

²Just as ordered probit is a special case of probit, so to is heteroskedastic ordered probit a special case of heteroskedastic probit.

smaller amount.

1.1 Potential Problems with 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.³

We know, for example, that heterogenous choice models will be at best consistent but certainly not unbiased. The question, though, is whether the asymptotic properties of heterogeneous choice models differ from those of discrete choice models from which they are derived, and whether these differences may cause serious inferential problems.

First, we address the relative efficiency of the heteroskedastic probit model. It is useful, here, to use the latent variable derivation of the probit model. Consider a latent or unobserved, continuous variable y_i^* ,

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \tag{7}$$

where \mathbf{x}_i is a vector of values for the i th observation, $\boldsymbol{\beta}$ is a vector of parameters, ε_i is the unobserved error, $\varepsilon_i \sim N(0, \sigma^2)$.

Assume,

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases} \tag{8}$$

From this we can derive the standard probit model:⁴

$$\begin{aligned} Pr(y_i = 1) &= Pr(y_i^* > 0) \\ &= Pr(\mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i > 0) \\ &= Pr(\varepsilon_i > -\mathbf{x}_i\boldsymbol{\beta}) \\ &= \Phi(\mathbf{x}_i\boldsymbol{\beta}). \end{aligned} \tag{9}$$

³We 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.

⁴This derivation is, of course, equivalent to the random utility derivation of dichotomous choice models.

As Davidson and MacKinnon (1993) note, some efficiency must be lost as a binary y_i contains less information than the continuous y_i^* . Specifically, the loss of efficiency depends on the distribution of $\Phi(\mathbf{x}_i\boldsymbol{\beta})$. The probit model estimated with y_i will most closely match the efficiency of OLS estimated with y_i^* 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 for 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).

This implies that the efficiency of the heteroskedastic probit model should be less than the standard probit model. No information has been added to y_i , but some number of additional parameters have been added to the model. Therefore, we might expect that even in large samples that, unless distributional restrictions are imposed on $\Phi(\mathbf{x}_i\boldsymbol{\beta})$, the heteroskedastic probit model standard errors to be incorrect.

Moreover, the additional parameters in the heteroskedastic probit model should cause the estimates to converge to their true values at slower rates than for regular probit models, which implies larger sample sizes are needed to consistently estimate a heteroskedastic probit model. Whether this is true for heteroskedastic ordered probit model remains an open question. While an ordinal y_i contain less information than for a continuous y_i^* , the ordinal y_i does contain considerably more information than a binary y_i does.

Next, misspecification error in heterogenous choice models should be worse than in standard probit and ordinal probit models. Yatchew and Griliches (1985) derived the formula for the asymptotic bias in standard probit models, and we rely on their work to understand the effect of misspecification in heterogenous choice models. Consider a basic probit model:

$$y_i = \Phi(\alpha + \beta\mathbf{x}_i + \delta\mathbf{w}_i + \varepsilon_i) \tag{10}$$

Suppose that the distribution of \mathbf{w}_i is conditional on \mathbf{x}_i in the following fashion:

$$\mathbf{w}_i = \phi_0 + \phi_1\mathbf{x}_i + v_i \tag{11}$$

where:

$$\begin{aligned} E(v_i|\mathbf{x}_i) &= 0 \\ \text{var}(v_i|\mathbf{x}_i) &= \sigma_v^2 \end{aligned} \tag{12}$$

If \mathbf{w}_i is normally distributed given \mathbf{x}_i and is omitted from the estimating equation, then the Maximum Likelihood estimate of β will converge to:

$$\frac{\beta + \delta\phi_1}{\sqrt{\delta^2\sigma_v^2 + \sigma_\varepsilon^2}} \tag{13}$$

Unlike in a regression context, the estimate of β will be biased even if $\phi_1 = 0$, that is if the two variables are uncorrelated. The amount of bias will depend on the distribution of \mathbf{w}_i as well as the magnitude of δ . With the identity that follows, we can derive the asymptotic bias for heterogenous choice models.

The literature on heterogenous choices implicitly assumes that the analyst is estimating two separate models: a choice model and a variance model. Such an assumption implies that the first and second moments of y_i , the mean and the variance, are separately identified as would be true for a continuous distribution. But 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.” The inability to separately estimate β and σ has important implications for deriving the asymptotic bias for heteroskedastic models.

Assume we have a model for the mean of y_i that is non-linear and takes the following functional form:

$$\mathbf{x}_i\beta = \frac{\mathbf{g}_i\beta}{\exp(\mathbf{z}_i\gamma)} \tag{14}$$

but the data generating process (DGP) is homoskedastic so that $\sigma = 1$. If this is the case, then

$$\Pr(y_i = 1) = \Phi\left[\frac{\mathbf{x}_i\beta}{\exp(\mathbf{z}_i\gamma)}\right] \tag{15}$$

This model is, of course, identical to the heteroskedastic model, and the heteroskedastic probit

model and cannot be distinguished from it. This identity implies that the asymptotic bias for a misspecified heteroskedastic probit model will be as follows:

$$\frac{\frac{\beta}{\exp(\mathbf{z}_i\gamma)} + \delta\phi_1}{\sqrt{\delta^2\sigma_v^2 + \sigma_\varepsilon^2}} \quad (16)$$

Clearly the bias from misspecification will be greater than that found in a probit model. Calculating the exact magnitude is difficult since it depends on the distribution of the variables as well as the magnitude of $\exp(\mathbf{z}_i\gamma)$.

Measurement error also presents a singular problem for heterogenous choice models. Yatchew and Griliches (1985) derive the asymptotic bias given measurement error, which takes the same functional form as that of the bias. Importantly, their formula implies that if measurement error is found in the \mathbf{z} -variables, we should expect a compounding of the bias similar to that of misspecification. Therefore, heterogenous choice models should exhibit larger bias under specification error than regular probit and ordinal probit models, and prone to larger amounts of bias if the \mathbf{z} -variables are measured with error. We may also suspect some level of inefficiency, but this is a lesser concern if the estimates are biased.

One may argue that while these biases may be present, the general effect of the bias is to attenuate the true effects. While this is generally true, we are rarely interested in the actual estimates of β or γ . Instead, we care about the substantive interpretation of these parameters in terms of predicted probabilities or marginal effects. Given the non-linear transformations required to calculate predicted probabilities and marginal effects, the bias may not appear as an attenuation, but could, in some circumstances, lead to incorrectly signed predicted probabilities and marginal effects, particularly in the ordered model.

So in short, we expect that the heterogenous choice models should converge at a slower rate than normal probit models. The heteroskedastic probit model should also be less efficient with same amount of data. Moreover, the amount of bias introduced by misspecification and measurement error may be substantially larger than for standard probit models.

2 A Monte Carlo Analysis

To better understand the implications of the analytical results we also use Monte Carlo simulations to study the properties of heterogenous choice models. 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 general model has the following form:

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

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 three different Monte Carlo experiment for both models. In experiment 1, we use a perfectly specified variance model and then varied the sample size, doing 1000 experiments with each sample size. This analysis allows us to both assess the rate of convergence for the estimates and to observe how efficient these models are. In experiment 2, we added normal random errors to the variance model to simulate measurement error in the variance model and again varied the sample sizes. In the final set of experiments, we examine the effects of specification 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 either the variance or choice model. 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.⁵ 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 reported estimate of variability. For each experiment, we report whichever statistics are most appropriate.

2.1 Efficiency and Asymptotic Consistency

Before presenting the results, we stop briefly to consider a benchmark for the general performance of the models. The benchmark we use for the estimates of β are the estimates from standard probit and ordered probit models used with homoskedastic data. That is, we compare whether the estimates of β for the heteroskedastic models converge and are as efficient as standard probit and ordered probit models. Unfortunately, for γ no such obvious baseline exists.

First, we present the results for when the models are perfectly specified but we vary the sample size for each experiment. To construct a baseline for this experiment, we also performed two sets of Monte Carlo experiments where we estimated standard probit and ordered probit models with a homoskedastic DGP across the same range of sample sizes. Figure 1 plots the RMSE for both heteroskedastic and homoskedastic probit and ordered probit models for sample sizes of 50, 100, 250, 500, 750 and 1000.

The importance of sample size is immediately evident; in the smaller sample sizes the RMSE for the heteroskedastic model is much higher than that of the standard probit model. For the heteroskedastic probit model estimated with 50 cases, the RMSE is 90.43 while it is 33.1 for the standard probit model. There is a large reduction in error as the sample size increases to 100 cases, as the RMSE falls from 90.43 to 30.36 for the heteroskedastic probit model. But the RMSE for the probit model with 1000 cases is a mere 0.8. Once the sample size increases to more than 250 cases, however, the RMSE values for the two models start to converge.

The heteroskedastic ordered probit model performs substantially better than the heteroskedastic probit model. The RMSE starts out at a much lower level for 50 cases, 8.97, but is still much larger than the value for a standard ordered probit model. The difference between the two

⁵The measure is more precisely: $100 \frac{\sqrt{\sum_{i=1}^{1000} (\beta - \hat{\beta}_i)^2}}{\sqrt{\sum_{i=1}^{1000} (s.e.(\hat{\beta}_i))^2}}$.

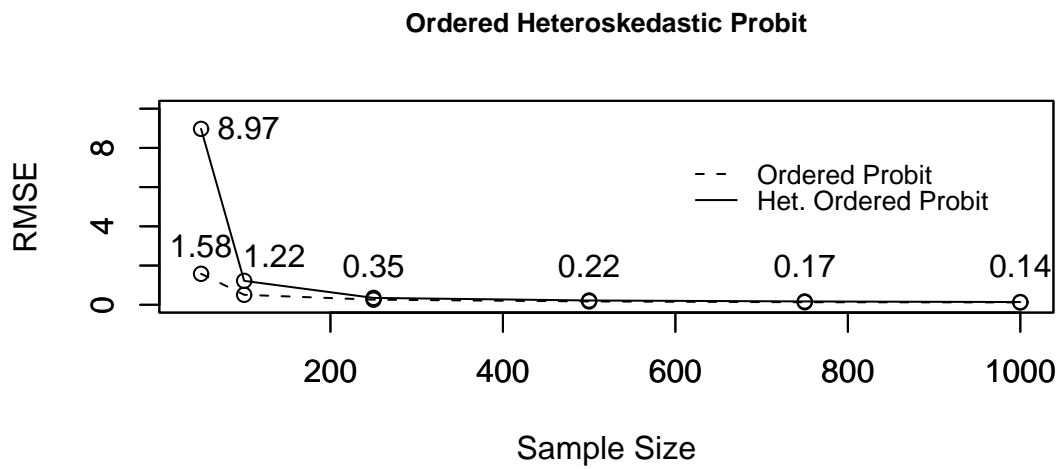
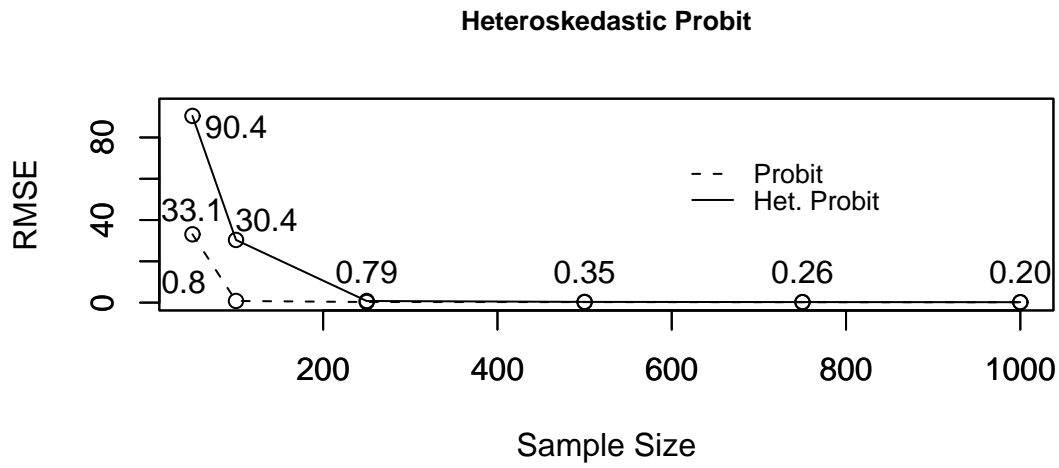


Figure 1: General Properties of Heterogenous Choice Models-Root Mean Squared Error

models is much smaller for a 100 cases and the RMSE values converge at 250 cases.

Figure 2 plots the average MAE over the different sample sizes for the two $\hat{\beta}$ parameters, which provides a better sense of the models' properties over differing case sizes. In the case of the heteroskedastic probit model with 50 cases, the $\hat{\beta}_1$ parameter tended to be off by about 51 points and the error for $\hat{\beta}_2$ was, on average, around 77 points; the estimates, then, are incorrect by around 2500%! Obviously, the model tends to provide poor estimates of the β parameters. Again, the error is much smaller in the model with 100 cases but is still unacceptable considering the average error for the $\hat{\beta}_1$ parameter was nearly 8 points and just under 12 points for the $\hat{\beta}_2$ parameter. Once we move to sample sizes of 250 cases and greater, the error starts to level off to more acceptable levels. The error is reduced substantially for each increment in the sample size until it is quite small for 1000 cases with an average error of only 0.09.

The error in the estimates of γ , while not small, is always smaller than that of the β 's. At worst, the estimates of γ are incorrect by around 160-170%, which, while not good, is much better than that of the β 's. Again, the situation improves considerably once we reach 250 cases. Here, the estimates of the γ 's are off by around 14%, and this drops to around 6% for 500 cases.

For the ordered probit model, the error in $\hat{\beta}$ is smaller, but not acceptable. For a model with 50 cases, the estimates are off by around 230%. The effect of larger sample sizes takes hold much earlier. For 250 cases, the error is fairly small (around 8%), improving to 1-2% for 1000 cases.⁶

From perspective of curing heteroskedasticity, the empirical evidence presents a provocative result. For small sample sizes, the variance model cure appears to do little good. Clearly, heterogenous choice models should not be used with small sample sizes. Five hundred cases for a heteroskedastic probit model and 250 for heteroskedastic ordered probit are rough guidelines for minimum sample sizes. While these are minimum levels, it is also clear that the difference in both cases between a 500 case model and a 1000 case model is still substantial.

We now examine the efficiency of the two models when estimated with 1000 cases. Table 1 contains the percentage of bias, level of overconfidence and the coverage rate, for the models under ideal circumstances. The heteroskedastic ordered probit model clearly can be given a

⁶We should also note that for both models when the case size falls below 250, the hessian matrix would frequently fail to invert. The failure to produce standard errors was the same across both the ordered and regular heteroskedastic models.

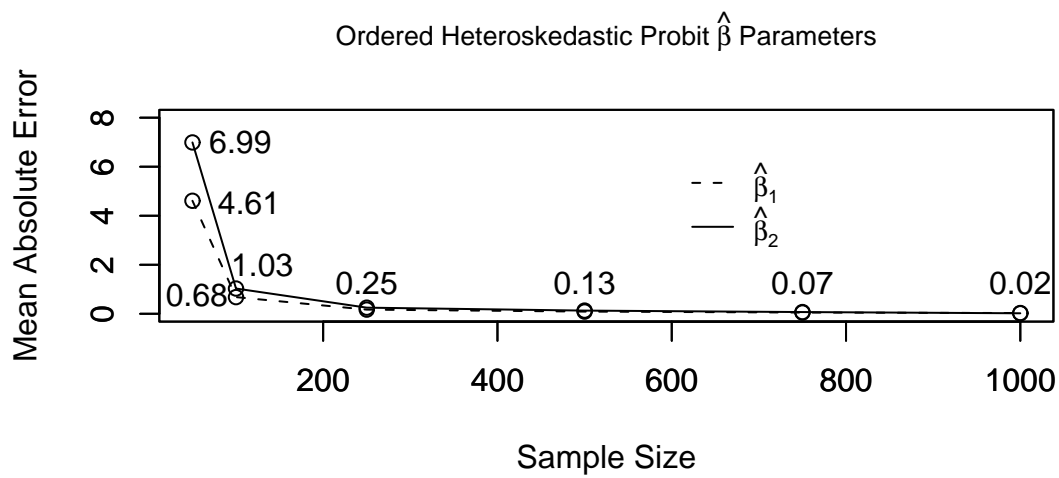
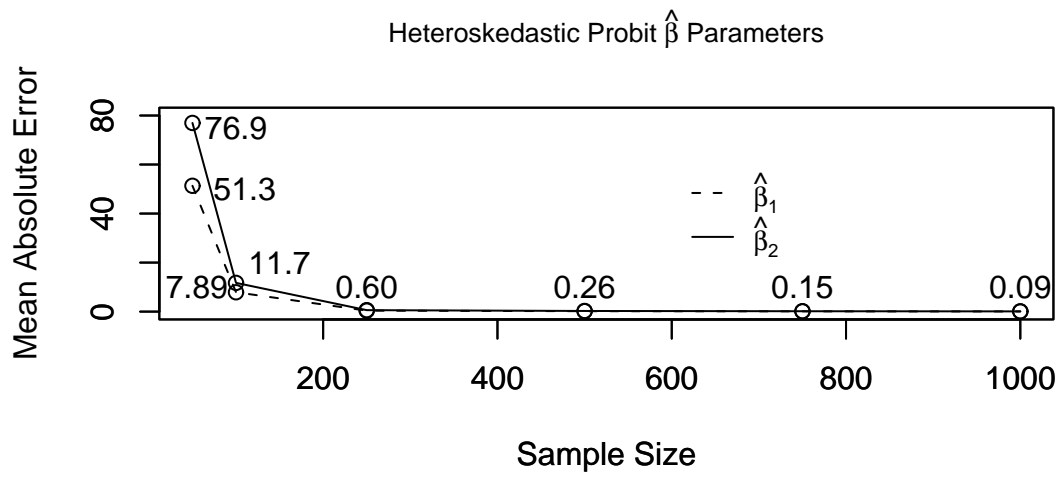


Figure 2: General Properties of Heterogenous Choice Models-Average Bias

clean bill of health, as both the level of overconfidence and coverage rates are close to ideal. The heteroskedastic probit model, however, fares less well. For all the parameters in the model, the estimated sampling variability tends to be over 20% too large and the coverage rates are lower than they should be. 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 under a more restrictive distribution of 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 than ideal conditions.

Table 1: General Properties of Heterogenous Choice Models

SPECIFICATION		M.A.E.	OVERCONFIDENCE ^a	
		(%)	(%)	LEVEL ^b
Probit	β_1	4	122	82
	β_2	4	116	86
	γ_1	7	127	83
	γ_2	5	123	83
Ordered	β_1	1	92	99
Probit	β_2	1	95	96
	γ_1	0	105	95
	γ_2	2	95	97

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

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

^bPercentage of 95% confidence intervals that contain β or γ .

2.2 Measurement Error

The second Monte Carlo analysis examines the effect of measurement error in the variance model. For both probit and ordered probit models, normal random errors were added to the variables in the variance model. Figure 3 plots the average bias in the two $\hat{\gamma}$ parameters for both models. The bias in the $\hat{\gamma}$ parameters is quite high as both of the γ parameters are biased

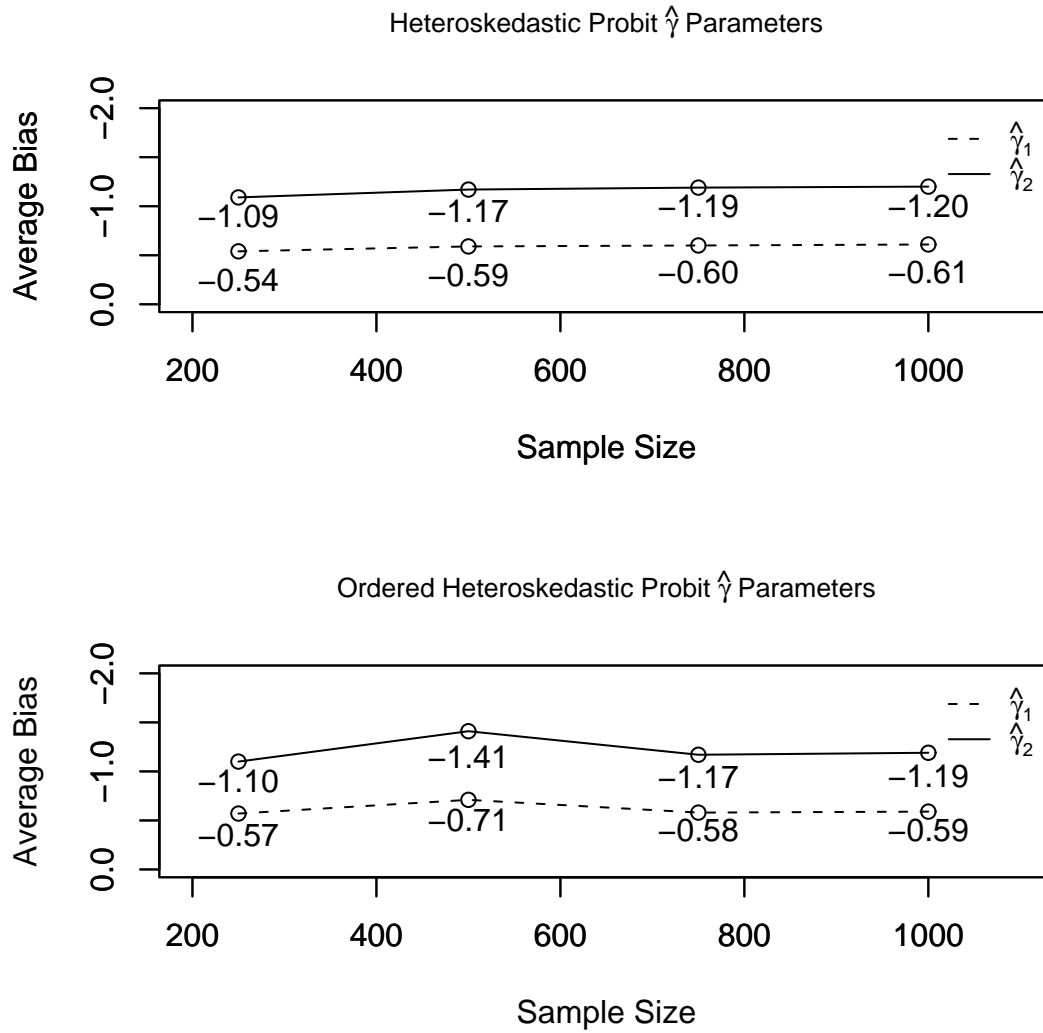


Figure 3: Measurement Error in Heterogenous Choice Models-Average Bias

by between 50% to 60%. The estimates of the β 's are also affected by the noise in the variance model. While the bias is not as large as for the γ 's, it is still a substantial 25-35%.

In the case of ordered probit model, we see the same pattern. Other than the unusual bump at 500 cases, again we see inconsistent estimates of the γ parameters. This is also true for the estimates of the β 's, as the noise in the variance model generates similar amounts of bias, with the estimates being off by 30 to 55%.

The empirical evidence, here, highlights a second problem with heteroskedastic probit models. Not only does this model have poor small sample properties and less than ideal standard errors, but the bias induced by measurement error in the variance model is substantial as the

analytic results suggest. While the heteroskedastic ordered probit model fared well in the first analysis, it, too, has substantial problems with measurement error in the variance model.

2.3 Specification Error in Heterogenous Choice Models

The last set of Monte Carlo experiments test how well the two models perform when either the choice or variance model is misspecified. The asymptotic bias formula for heterogenous choice models implies that the bias caused by misspecification could be substantial. Our analysis, here, attempts to quantify whether this bias is, in fact, large or perhaps more trivial.

We examine the effect of misspecification in two ways, one blunt and the other subtle. The blunt form of misspecification is that of omitted variables. That is, one of the estimating equations does not include a variable from the DGP. But we can also commit a more subtle form of misspecification. A variable we use in one of the estimating equations may only be correlated with the true variable in the DGP. To capture this more subtle form of misspecification, we estimated a series of models where one of the included variables is correlated with the true variable at three different levels: 0.90, 0.70 and 0.50. We purposefully set the correlations between the true and proxy variables to be quite high. For example, correlations between variables 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 misspecifications for the choice model and four different misspecifications for the variance model.

In these experiments, we report the coverage rate and level of overconfidence as well as the percentage of bias. When assessing the amount of bias in the estimates of β , we use a misspecified standard probit model as a baseline. This will allow us to understand whether misspecifying an \mathbf{x}_i variable in a heterogenous choice model has consequences beyond those of misspecifying an \mathbf{x}_i variable in standard probit models. 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.

2.3.1 Heteroskedastic Probit Models

We start with the consequences of misspecifying a standard probit model. We know that the estimates of all the variables in the model will be biased (regardless of whether they are correlated or not), but by how much isn't clear. To find out, we estimate a Monte Carlo analysis using a standard probit with two uncorrelated variables, one of which was correlated with the true variable in the DGP at 0.90. The effects of this misspecification are fairly benign as seen in Table 2.

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
.50	β_1	-25	92	0
	β_2	-60	92	0
Omitted X_2	β_1	-30	137	0

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

^aCorrelation between true variable in DGP and variable in the estimated model.

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

^cPercentage of 95% confidence intervals that contain β or γ .

The estimates of the correctly specified variable were off by 8% while the estimates of the misspecified variable were around 19% too small. How does this compare to a similar misspecification of an \mathbf{x}_i variable in a heteroskedastic probit? Here, the effect of the same misspecification is dramatic, as seen in Table 3. Now, the estimates of *every* parameter in the model are biased by 60-70%. Moreover, the 95% confidence interval never contains the true parameter value and the standard errors are too confident in some cases by over 200%. The heteroskedastic probit model, then, performs significantly worse than a standard probit under a similar misspecification.

What if the misspecification is worse? Under the next experimental condition, the \mathbf{x}_i variable included in the estimating equation is correlated with the variable in the DGP at 0.70. In the

Table 3: Misspecification of the Heteroskedastic Probit Choice Model

SPECIFICATION ^a		BIAS	OVERCONFIDENCE ^b	LEVEL ^c
		(%)	(%)	
.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
.50	β_1	-74	150	0
	β_2	-87	139	0
	γ_1	-83	141	0
	γ_2	-83	157	0
Omitted X_2	β_1	-76	137	0
	γ_1	-85	138	0
	γ_2	-86	137	0

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

^aCorrelation between true variable in DGP and variable in the estimated model.

$${}^b\text{Overconfidence} = 100 \frac{\sqrt{\sum_{i=1}^{1000} (\beta - \hat{\beta})^2}}{\sqrt{\sum_{i=1}^{1000} (s.e.(\hat{\beta}))^2}}$$

^cPercentage of 95% confidence intervals that contain β or γ .

baseline probit model, the effect is not trivial. The parameter for the correctly specified variable in the model will be biased by 16%, and the estimated effect of the misspecified variable will be 45% too small. But in the heteroskedastic probit, the parameters for all the variables in the model will be biased by at least 70% if not closer to 80%. And, again, the 95% confidence interval does not contain the true parameter in any of the Monte Carlo trials. In the two other Monte Carlo experiments where the specification of the choice model worsens, the performance of the model continues to deteriorate.

The heteroskedastic probit model is clearly more sensitive to the specification of the \mathbf{x}_i variables 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 misspecifying the variance model are noticeably different from misspecifying the choice model. Here, as seen in Table 4, the effect of misspecification 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 β 's in the choice model are off by 36%, and the γ '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 a pronounced effect on the amount of bias in the model. The bias increases by 20% for both the estimates of the β 's and γ 's. The estimates of the sampling variability continue to worsen but by a small margin, and the coverage rates remain poor.

Bias and Response Probabilities Given the nonlinear link function used in heteroskedastic probit models, bias in the parameter estimates does not directly translate into how the level of bias affects the estimate of the effect \mathbf{x}_i on the probability that $y_i = 1$. For example, Wooldridge

Table 4: Misspecification of the Heteroskedastic Probit Variance Model

SPECIFICATION ^a		BIAS	OVERCONFIDENCE ^b	LEVEL ^c
		(%)	(%)	
.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
.50	β_1	-62	189	0
	β_2	-62	206	0
	γ_1	-51	238	0
	γ_2	-76	213	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.

^aCorrelation between true variable in DGP and variable in the estimated model.

$${}^b\text{Overconfidence} = 100 \frac{\sqrt{\sum_{i=1}^{1000} (\beta - \hat{\beta})^2}}{\sqrt{\sum_{i=1}^{1000} (s.e.(\hat{\beta}))^2}}$$

^cPercentage of 95% confidence intervals that contain β or γ .

(2001) recommends comparisons of the response probabilities across various values of \mathbf{x}_i .⁷

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 \mathbf{x}_i variables, we calculated the predicted probabilities of y_i given changes in the \mathbf{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 4, we plot the predicted probability that $y_i = 1$ as \mathbf{x}_2 changes from its minimum to maximum value 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 \mathbf{x}_2 . The differences in predicted probabilities can be quite large for negative values of \mathbf{x}_2 as the estimated predicted probabilities are much too large by about 0.50. While the two predicted probabilities are identical for one value of \mathbf{x}_2 , for larger values of \mathbf{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 \mathbf{x}_2 the predicted probability for y_i may be either too small or too large.

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 \mathbf{z}_1 with the rest of the variables in the model held constant. In Figure 5, 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 \mathbf{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, the empirical evidence demonstrates that the estimates from heteroskedastic probit models with even minor misspecifications are poor. The estimates of the parameters are biased and the standard errors are incorrect as well. Clearly, the model is substantially less robust

⁷One 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.

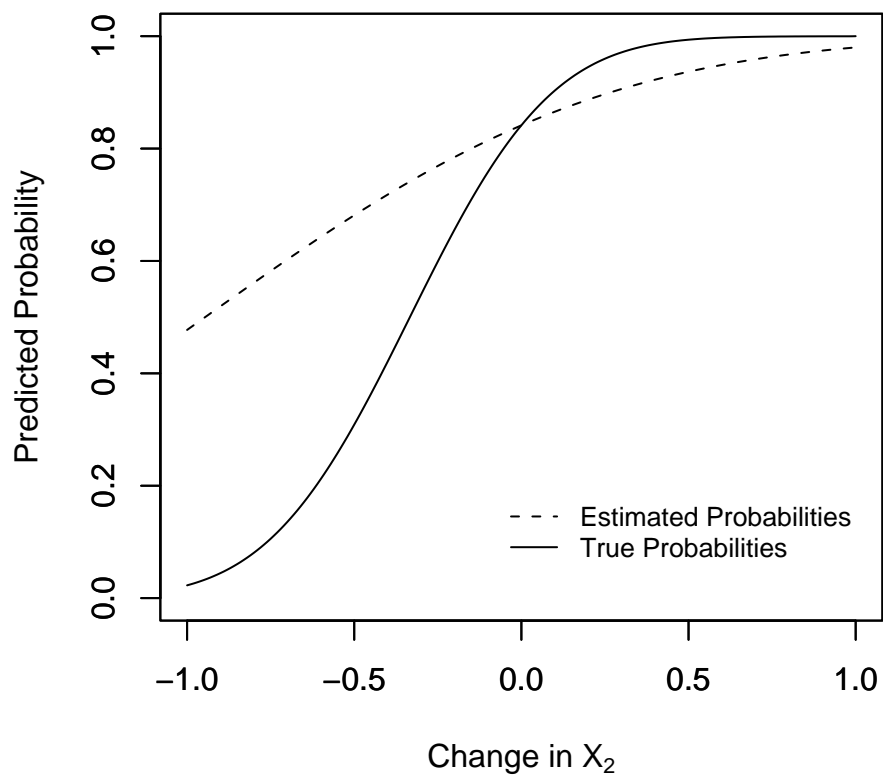


Figure 4: True Versus Estimated Probability that $y = 1$

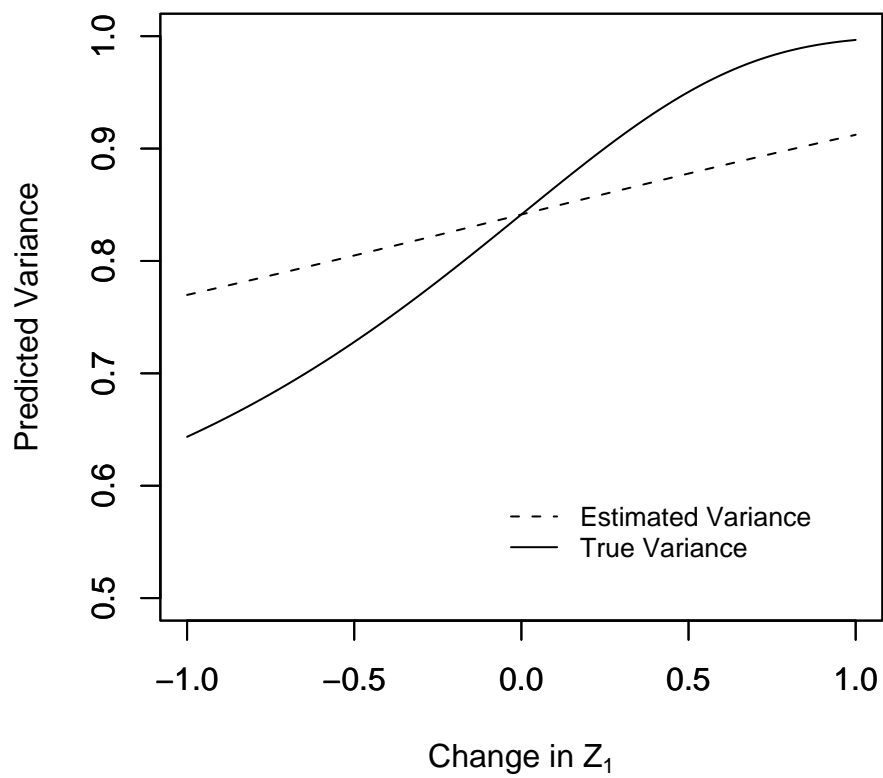


Figure 5: True Versus Estimated Predicted Variance of y_i

than the standard probit model upon which it is based. We, next, turn to the results for the heteroskedastic ordered probit model, which, thus far, has performed somewhat better than the heteroskedastic probit model.

2.3.2 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 are minimal. Table 5 contains the results from when the heteroskedastic ordered probit choice model is misspecified.

Table 5: Misspecification of the Heteroskedastic Ordered Probit Choice Model

SPECIFICATION ^a		BIAS (%)	OVERCONFIDENCE ^b (%)	LEVEL ^c
.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
.50	β_1	-74	103	0
	β_2	-87	100	0
	γ_1	-83	100	0
	γ_2	-83	113	0
Omitted X_2	β_1	-77	85	0
	γ_1	-84	109	0
	γ_2	-84	87	0

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

^aCorrelation between true variable in DGP and variable in the estimated model.

$$^b\text{Overconfidence} = 100 \frac{\sqrt{\sum_{i=1}^{1000} (\beta - \hat{\beta})^2}}{\sqrt{\sum_{i=1}^{1000} (s.e.(\hat{\beta}))^2}}$$

^cPercentage of 95% confidence intervals that contain β or γ .

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 misspecification 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. Here, too, when the choice model is misspecified, the sampling variance is fairly well estimated.

In Table 6, the variance model is misspecified, while the amount of bias is half that of when the choice model is misspecified, the estimated sampling variability tends to be twice as large. As before, when we misspecified the probit variance model, 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.

Table 6: Misspecification of the Heteroskedastic Ordered Probit Variance Model

SPECIFICATION ^a		BIAS (%)	OVERCONFIDENCE ^b (%)	LEVEL ^c
.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
.50	β_1	-60	156	0
	β_2	-60	188	0
	γ_1	-49	206	0
	γ_2	-73	201	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.

^aCorrelation between true variable in DGP and variable in the estimated model.

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

^cPercentage of 95% confidence intervals that contain β or γ .

Bias and Response Probabilities Again to gain a better sense of how the bias induced by misspecification affects the inferences in the heteroskedastic ordered probit model, we plot the predicted probabilities for the model with the least amount of misspecification. Since the

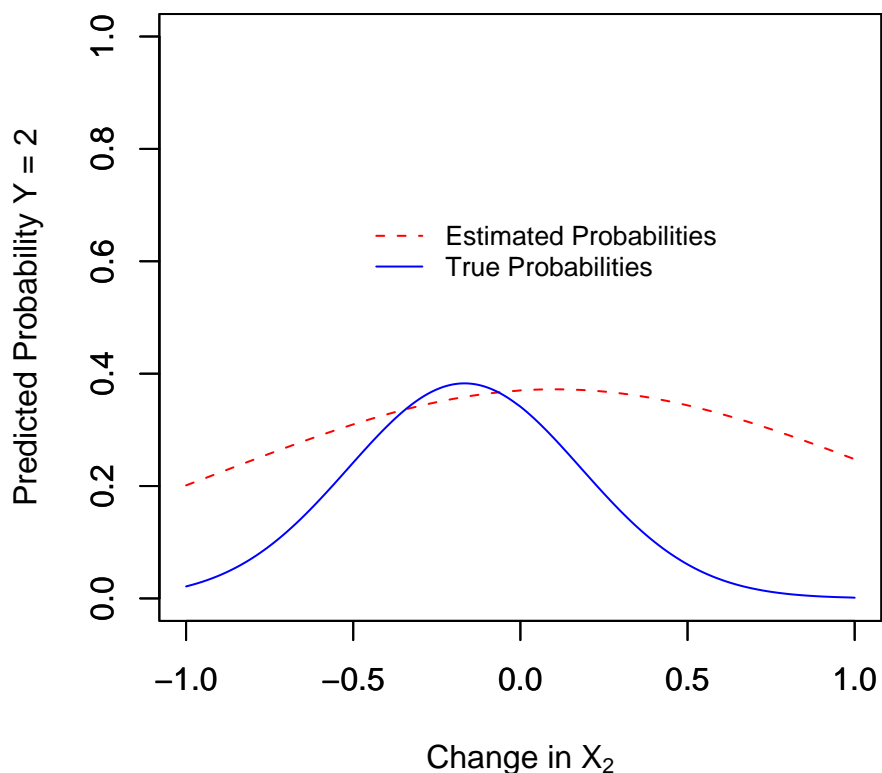


Figure 6: True Versus Estimated Predicted Probability that $y = 2$

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 6, we plot the predicted probability that $y_i = 2$. The misspecified heteroskedastic ordered probit model 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 7, 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 misspecification of the model

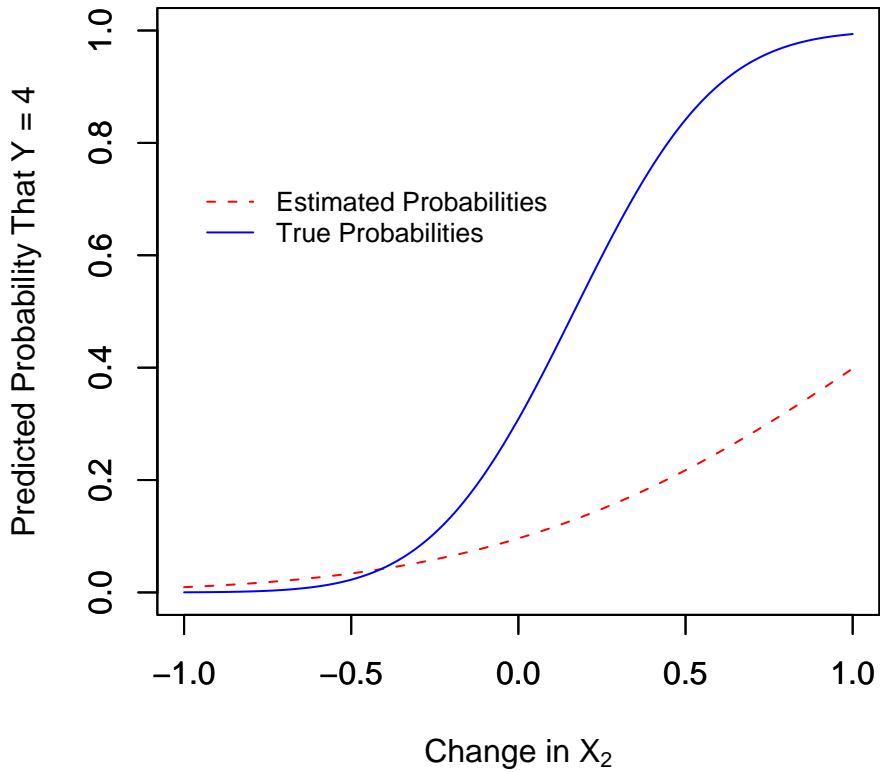


Figure 7: True Versus Estimated Predicted Probability that $y = 4$

can lead to incorrect inferences about the effects of the independent variables on y_i .

To give the reader a better idea of how poorly the γ parameters are estimated, we plot the sampling distribution of the γ parameters from the models where the choice model had the smallest amount of misspecification for both the probit and ordered probit models. In Figure 8, the sampling distribution for the two γ parameters from the heteroskedastic probit model are plotted in a histogram. In both cases, the true values are not even on the scale of the x-axis. In fact for both γ 's, the true value of the parameter is not within the estimated sampling distribution at all.

Figure 9 plots the sampling density for the γ parameters from the heteroskedastic ordered probit model. Once again, the true parameters values do not fall with the estimated sampling distribution. In effect, the bias is large enough that the estimate will almost never coincide with

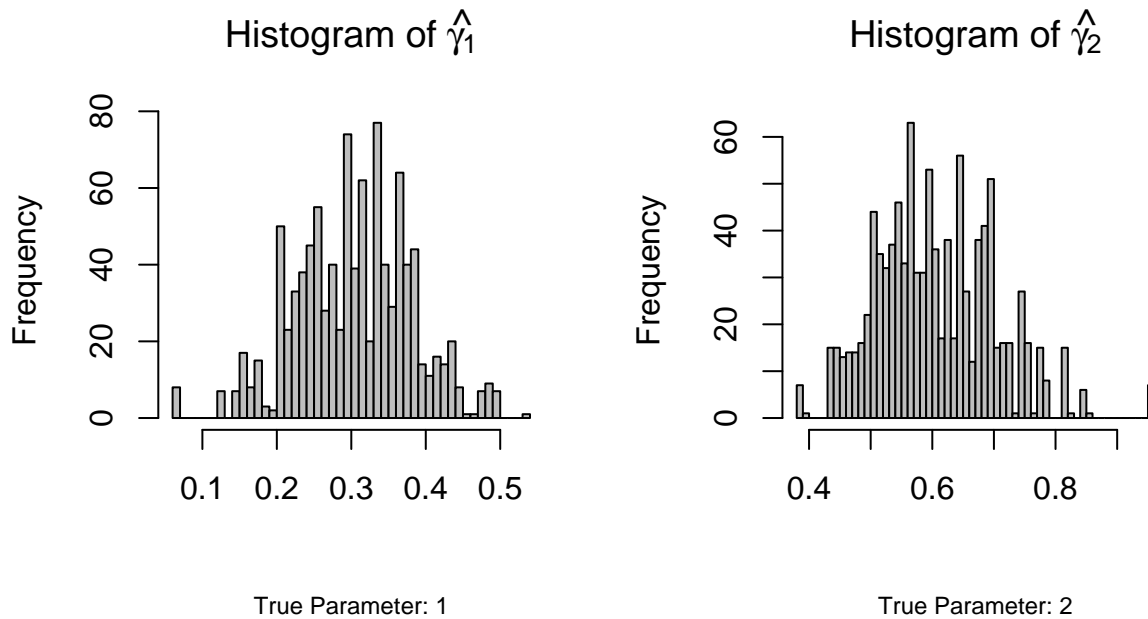


Figure 8: Histograms of Estimates for Heteroskedastic Probit Model

the true value. We, next, examine the consequences of our analysis for published results that rely on heterogenous choice models.

3 Implications For Empirical Models

Researchers typically spend little time judging model fit, even when such checks can result in discoveries that can greatly improve their model (Bafumi et al. 2004). We use simulations to judge the model fit of some published heteroskedastic probit estimates. To do this, we compare the posterior density (sampling distribution) of the model parameter estimates from actual data with a posterior density that utilizes information from simulated data. We do this with a parametric simulation approach developed by King, Tomz and Wittenberg (2000) that is similar to bootstrap methods.

We use the following procedure: we estimate the variance model parameters, γ . We, then, generate two posterior densities that are multivariate normal with mean vector γ . For the first posterior density, σ_1 , the variance-covariance matrix, $\mathbf{\Omega}_1$, is that of the γ 's from the published model. The second posterior density, σ_2 , uses $\mathbf{\Omega}_2$, a variance-covariance matrix from simulated

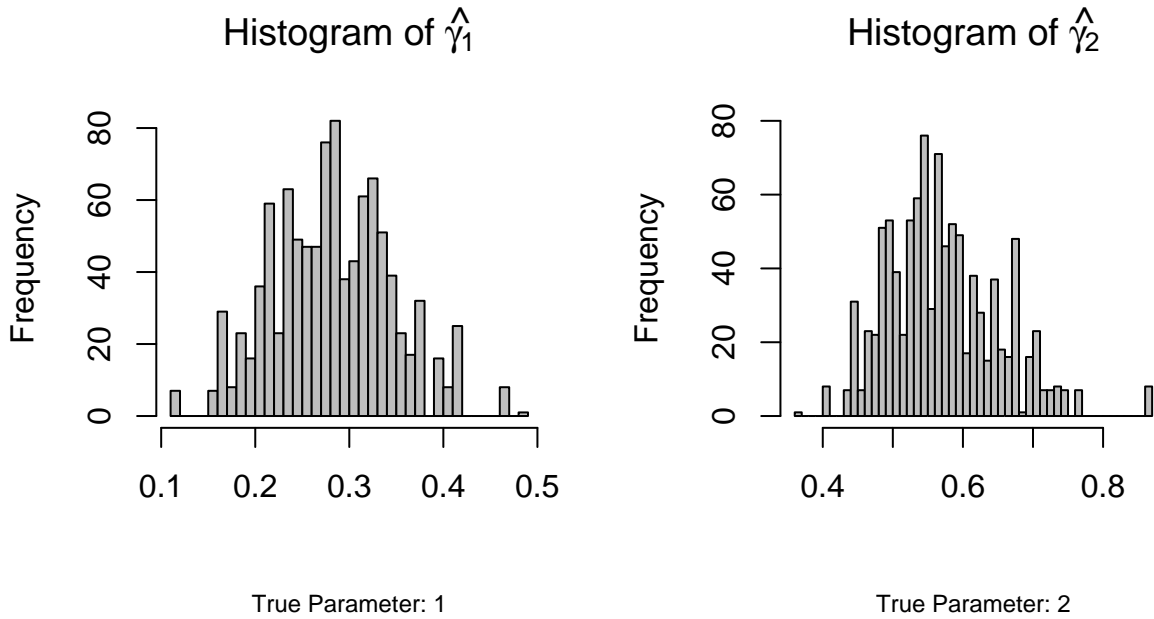


Figure 9: Histograms of Estimates for Heteroskedastic Ordered Probit Model

data.

To illustrate, we use the data from the seminal article which introduced these models to the political science literature (Alvarez and Brehm 1995). First, we replicate one of the models in the article to estimate the parameters γ and the variance-covariance matrix, $\mathbf{\Omega}_1$ for the variance model. We, then, generate the posterior density of the error distribution for the actual data, σ_1 , by generating 1000 draws from a multivariate normal distribution with mean γ and variance-covariance matrix $\mathbf{\Omega}_1$ and multiply that by the vector, $\bar{\mathbf{z}}$, which contains the means of the explanatory variables from the variance model, such that:

$$\sigma_1 = \exp(\bar{\mathbf{z}}\gamma_1) \text{ where } \gamma_1 \sim N_m(\mu_\gamma, \mathbf{\Omega}_1).$$

This forms a posterior density that summarizes the magnitude of the error variance when the variance model is held at its mean. The variation in this density summarizes all the knowledge about the parameters given the statistical model. The mean of this distribution represents the key quantity of interest: the average magnitude of the error variance, but as a distribution it also represents our uncertainty about the quantity of interest. If we knew the values for the γ parameters perfectly, then the distribution would be a spike, however, the larger the variances of the γ parameters are, the more the draws will differ. Now let's say we thought there was a minor

specification error in the model, one where a variable correlated with one in the choice model was included instead of the true variable. We can simulate the effect of such a specification error using the results from the Monte Carlo analysis.

To do this, we simulate a data set where one of the variables in the estimated choice models is correlated with the true variable at 0.70. We, then, generate a new posterior density of the variance model for the simulated data, σ_2 . Here, we again generate 1000 draws from a multivariate normal distribution using the same mean vector, γ from the Alvarez and Brehm model but with variance-covariance matrix Ω_2 where Ω_2 is estimated from the simulated data. We, then, multiply that with the same vector \bar{z} , such that:

$$\sigma_2 = \exp(\bar{z}\gamma_2) \text{ where } \gamma_2 \sim N_m(\mu_\gamma, \Omega_2).$$

This new posterior density simulates how a minor specification error effects our uncertainty about the magnitude of the error variance. We, next, plot the two posterior densities of the variance model in Fig. 10. The posterior density from the published model summarizes the estimates of the variance model with bounds of uncertainty. Given that this density has an extremely tight distribution, it indicates that there is very little model based uncertainty in the estimates. But, the model uncertainty, here, should be understated given that we use normal random error to approximate the model error, when, in fact, there might be considerably more systematic model error present.

The second posterior density should capture other systematic model specific error that may be present. This density has a much wider distribution that is significantly skewed right. Figure 10 demonstrates how a small specification error in the choice model can radically increase the amount of model specific uncertainty, as there is much more variability in the density of the model with the simulated variance-covariance matrix than in the published model.

It's even easier to see if we compare the 95% confidence intervals for the two distributions. The mean of $\sigma_1 = 0.93$ and $\sigma_2 = 1.00$, with the 95% interquartile range for $\sigma_1 = (0.87, 0.99)$ and for $\sigma_2 = (0.50, 1.88)$. While we cannot know whether the authors committed such a specification error in their model, Figure 10 indicates just how much uncertainty may be present in the estimates due to systematic model error.

Finally, let's assume that the abortion models are correctly specified. Given that these are

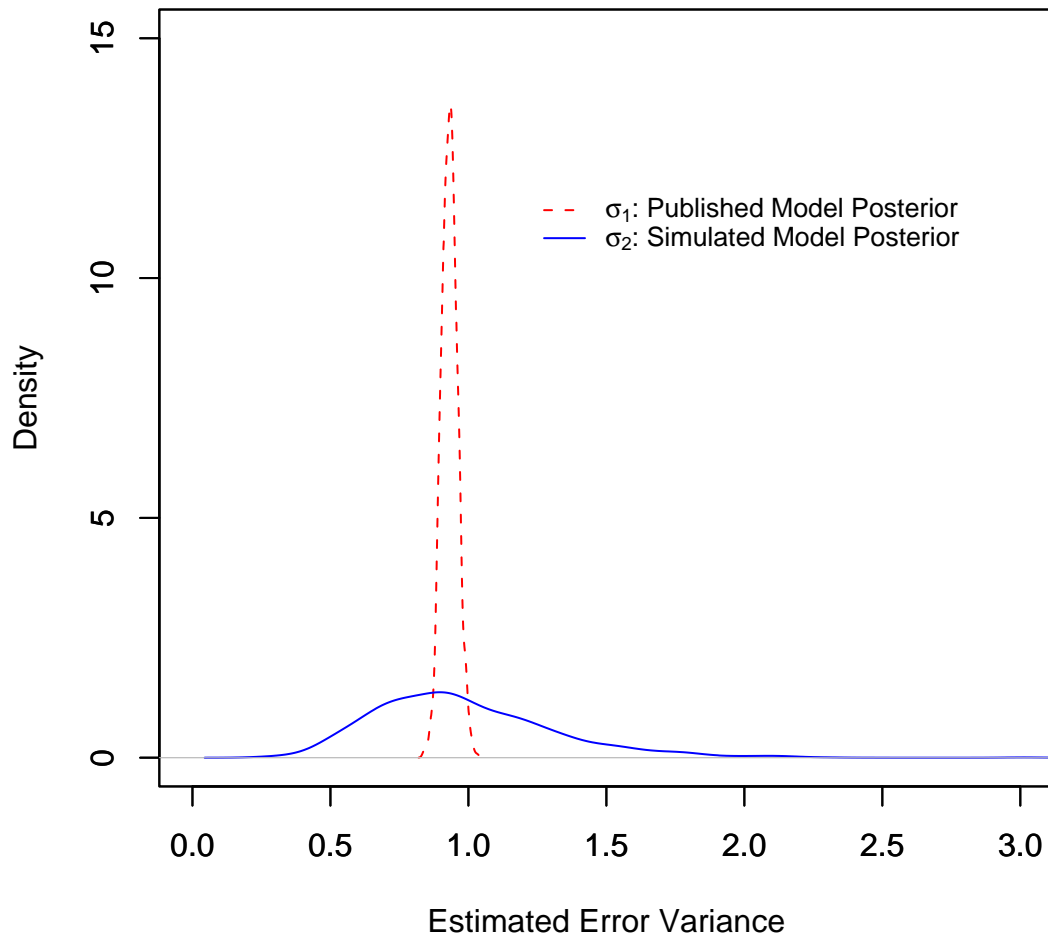


Figure 10: Posterior Density of the Variance Model: Brehm Alvarez Model (σ_1) versus Monte Carlo Simulation (σ_2)

heteroskedastic probit models, it is quite likely that the standard errors are understated as we found in the initial Monte Carlo analysis. We bootstrapped the standard errors on the published models to test how the underestimation of the standard errors might affect our inferences.⁸ We found that of the twenty-two coefficients that would be judged statistically significant by conventional standards nine were no longer statistically significant. While the average reduction in the t-statistics was around 0.40 some dropped by as much as 1.20. Clearly the poor standard errors found in heteroskedastic probit models can have a substantial effect on our ability to make inferences.

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, the research in this field must be altered in two ways.

The first change must be in the current use of heteroskedastic modeling techniques. The empirical evidence that we present demonstrates that the heteroskedastic probit model is too unsound for use in research applications. The estimated sampling variability and coverage rates were less than ideal even under a perfect specification. Measurement error in the variance model induced significant amounts of bias, and almost any specification error causes the estimates of both the choice and variance model to be completely unreliable. Moreover, the results, here, occur in the much simpler world of Monte Carlo experiments. In the real world, we should probably expect some combination of not having a large enough sample size, measurement error and an incorrect specification.

The poor performance of the heteroskedastic probit also implies that one should never take an ordered measure and dichotomize 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. A more thorny question is whether the heteroskedastic ordered probit model is safe for use in applied work. While the ordered model

⁸We used a 1000 bootstrap samples to recalculate the standard errors.

performed better under ideal conditions, it proved to be as sensitive to misspecification and measurement error as the probit model. A researcher with a *strong* theoretical specification and ample sample size might be comfortable fitting these models. The results, here, emphasize the need to pay attention to the specification of both models. Often the choice model is given little thought since the researcher is only interested in making inferences about the variability in choices. But as we have shown, misspecification of either model will badly bias the estimates. Therefore, without a very strong theoretical specification for *both* the choice and variance model, we feel that the estimates of even the ordered model should be treated with caution.

The findings here must also revise our substantive knowledge about politics. Given the empirical evidence, here, any findings that rely solely on heterogenous choice models must be cast into doubt. 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. Any finding based on these models must be treated with caution.

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.
- Bafumi, Joseph, Andrew Gelman, David K. Park and Noah Kaplan. 2004. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation." Unpublished Manuscript.
- 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.
- Greene, William H. 1993. *Econometric Analysis*. New York: Macmillan.
- Harvey, Andrew. 1976. "Estimating Regression Models with Multiplicative Heteroskedasticity." *Econometrica* 44:461–465.
- King, Gary, Michael Tomz and Jason Wittenberg. 2000. "'Making the Most of Statistical Analyses: Improving Interpretation and Presentation'" *American Journal of Political Science* 44:341–355.
- 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* .
- Ruud, Paul, A. 2000. *An Introduction to Classical Econometric Theory*. Oxford: Oxford University Press.

- Smith, Charles E., Robert D. Brown, John M. Bruce and L. Martin Overby. 1999. "Party Balancing and Voting for Congress in the 1996 National Election." *American Journal of Political Science* 43:737–764.
- 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.

Appendix

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.

We also performed some Monte Carlo experiments where we misspecified regular probit and ordered probit models, but estimated these models with heteroskedastic data. We found that the bias in the models was slightly worse than when the choice model of the heterogenous choice models was misspecified. As such, it would appear that the heterogenous choice models buy the analyst little when they are misspecified.

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(iffelse(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 Specified 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 Counter And Start Loop
j <- 1
for(j in 1 : i){
set.seed(round(runif(1, min=1, max=30000)))

#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")

```