Heteroscedastic Qualitative Response Model*

Choon-Geol Moon
Rutgers University

A qualitative response model with heteroscedastic errors is studied. First, the heteroscedastic maximum likelihood estimator and its asymptotic properties are derived. Then, inconsistency of the standard (homoscedastic) maximum likelihood estimator is proved. Lastly, two empirical examples and numerical experiments follow to illustrate the feasibility of the heteroscedastic maximum likelihood estimator.

I. Introduction

When heteroscedasticity in the error terms for the latent (or tolerance) index variable is suspected in a qualitative response model, there are three general approaches to consistent estimation and inference. We may apply some transformation to correct violation to the homoscedasticity assumption and then estimate the resultant standard model. Egger (1979) and Goto *et al.* (1986) provide power transformations relevant to this and related models.

Alternatively, we may apply a semiparametric estimation method, which is robust to heteroscedasticity. The only known robust estimator in the qualitative response model is Manski's maximum score estimator[1] (Manski 1975, 1985). Strong consistency is proved under the independence assumption and the zero median assumption on the error terms for the latent index variable. That is, the normalized parameter vector $\beta^* = \beta / \| \beta \|$ is estimated consistently for the following model:

*This work was supported by National Science Foudation Grants DMS 82-19748 and SES 82-08180 at Stanford University. I would like to thank Takeshi Amemiya, Theodore W. Anderson, and Kimio Morimune for their very helpful advice and comments. Also, suggestions from A. S. Goldberger, Lawrence J. Lau and an anonymous referee are gratefully acknowledged. All remaining errors are mine.

[1]Cosslett's generalized maximum likelihood estimator requires the i.i.d. assumption, and hence it is not robust to heteroscedasticity. See Cosslett (1983).

$$y_i^* = x_i' \beta + \epsilon_i \text{ and } y_i = \begin{cases} 1 \text{ if } y_i^* > 0 \\ 0 \text{ if } y_i^* \leq 0, \end{cases}$$

where ϵ_i's are independent and $\text{med}(\epsilon_i) = 0$. From the consistent estimate of β^* we can predict the discrete choice decision (where $y_i = 0$ or $y_i = 1$) of the individual i, given his / her characteristics, x_i. However, we cannot obtain the success probability prediction, the estimate of $\Pr(y_i = 1 \mid x_i)$, since no particular parametric distribution is assumed from the start and, under the non-identical errors due to heteroscedasticity, nonparametric estimation of the error distributions is not possible. Besides that asymptotic distribution is not normal (Kim and Pollard 1987), computation time becomes burdensome when the dimension of $\beta^* = \beta / \| \beta \|$ is high or the data set is large (see Manski and Thompson 1986).

The other alternative method is to construct explicitly a heteroscedastic qualitative response model, which provides the consistent estimate of the success probability under heteroscedasticity. A qualitative response model is a probability model where the probability aspects of discrete choices play a key role. Hence, the most interesting questions[2] in a qualitative response model are i) the probability of choosing an alternative, ii) the change in the probability of choosing an alternative given a change in an explanatory variable, iii) the expected response, and iv) the change in the expected response given a change in an explanatory variable. The estimates of these magnitudes can be obtained by the last approach, but not by a semiparametric approach.

In this paper, an explicit heteroscedastic model is studied. First, a heteroscedastic model is delineated. Asymptotic properties of the heteroscedastic maximum likelihood estimator (MLE) are obtained. Then, inconsistency of the homoscedastic (or standard) MLE under heteroscedasticity is proved and discussed. Second, the two specifications are compared using the two real data sets. In the discussion of empirical results, large sample tests of the homoscedasticity hypothesis are mentioned. Third, a Monte Carlo study on the empirical size and power of a likelihood ratio test is performed, together with a numerical experiment for the evaluation of the magnitudes of inconsistency of the homoscedastic MLE under heteroscedasticity. A final section concludes the discussion.

[2]The asymptotic standard errors in a multinomial logit model for the efficient estimates of the four magnitudes listed below are the main topics in Fomby and Pearce (1986).

The literature on heteroscedasticity in a qualitative response model is very limited. Local approximation to the inconsistency of the homoscedastic MLE under the heteroscedasticity assumption is studied by Yatchew and Griliches (1979, 1985) and by Kiefer and Skoog (1984), using local asymptotic specification analysis. However, both derivations are not operational in approximating even locally the asymptotic bias in a realistic situation. This is due to the requirement that the common variance of the homoscedastic model prior to normalization be known. On the other hand, residual diagnostics based on the extended concepts of residuals for probit, Tobit and related models are developed by Gourieroux *et al.* (1985, 1987a, 1987b), and Chesher and Irish (1987). Specifically, the following definitions of residuals in probit, Tobit and related models are proposed: generalized residuals by Gourieroux *et al.* (1984, 1985, 1987a), simulated residuals by Gourieroux *et al.* (1987b), and standardized residuals by Chesher and Irish (1987).

II. Heteroscedastic Probit and Logit Models

Consider the following four models with y_i, $x'_i \equiv (1\ x_i^{*\prime})$, z_i being the observables, and y_i^* being the latent variable:

$$\text{Model } 0 : y_i = 1\ (y_i^* > 0)$$
$$y_i^* = x'_i \beta + \tilde{u}_i, \quad i = 1,\cdots,n,$$

where $\tilde{u}_i \sim N(0, 1)$ or a standard logistic distribution, independent over i.

$$\text{Model } 1 : y_i = 1\ (y_i^* > 0)$$
$$y_i^* = x'_i \beta + u_i, \quad i = 1,\cdots,n,$$

where $u_i \sim N(0, \sigma_i^2)$ or a logistic distribution with mean zero and variance σ_i^2, independent over i,

$$\sigma_i^2 = g(z'_i \alpha) \text{ with } g(\cdot) > 0$$

z_i should not include a constant term.

$$\text{Model 1A} : \text{Model 1 with } g(z'_i \alpha) = (1 + z'_i \alpha)^2.$$
$$\text{Model 1M} : \text{Model 1 with } g(z'_i \alpha) = \exp(2z'_i \alpha).$$

The standard probit or logit regression model corresponds to Model 0 above, while the additive heteroscedastic probit or logit model corresponds to Model 1A, and the multiplicative heteroscedastic probit or logit model corresponds to Model 1M. As for Model 1A and Model 1M, a normalization for parameter identification is incorporated into the heteroscedasticity structure by not allowing a constant term in z_i. In the main text below, only Model 1A is analyzed in comparison with Model 0, while derivations for Model 1M are briefly summarized in the Appendix B.

Alternatively, heteroscedasticity in the probit and logit models can be interpreted as a varying coefficients model where $y^*_i = x'_i \beta_i + \tilde{u}_i$. However, this alternative modeling gives rise to a serious problem that the number of parameters goes to infinity as the sample size goes to infinity.

From now on, we will use the following notation:

$$\Phi(t) = \int_{-\infty}^{t} (2\pi)^{-1/2} \exp(-v^2/2)dv$$

$$\phi(t) = (2\pi)^{-1/2} \exp(-t^2/2)$$

$$\Lambda(t:\lambda) = [1 + \exp(-\frac{\pi}{\sqrt{3}\lambda} t)]^{-1}$$

$$F_i \equiv F_i(x'_i \beta) = \Pr(y_i = 1) = \Pr(y^*_i > 0)$$

$$f_i \equiv f_i(x'_i \beta) = \frac{\partial F_i}{\partial (x'_i \beta)}.$$

A. Heteroscedastic Probit Model

Now, consider Model 1A with the probit specification. The log-likelihood function l is written as:

$$l = \sum_{i=1}^{n} y_i \ln F_i + \sum_{i=1}^{n} (1 - y_i) \ln(1 - F_i), \tag{1}$$

where

$$\begin{aligned} F_i &= \Pr(u_{1i} > -x'_i \beta) \\ &= \Pr(u_{1i} < x'_i \beta) \\ &= \int_{-\infty}^{x'_i \beta \sigma_i^{-1}} (2\pi)^{-1/2} \exp(-t^2/2)dt \\ &= \int_{-\infty}^{x'_i \beta} (2\pi \sigma_i^2)^{-1/2} \exp(-t^2/2\sigma_i^2)dt \\ &= \Phi(x'_i \beta / \sigma_i). \end{aligned} \tag{2}$$

Differentiating (1) with respect to $\gamma \equiv \binom{\beta}{\alpha}$, we get the gradient vector and the Hessian matrix of l:

$$\frac{\partial l}{\partial \gamma} = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \\ \frac{\partial l}{\partial \alpha} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} f_i \{\frac{y_i}{F_i} - \frac{(1-y_i)}{(1-F_i)}\} x_i \\ \\ -\sum_{i=1}^{n} (\text{sgn}_i)(\frac{x'_i \beta f_i}{\sigma_i}) \{\frac{y_i}{F_i} - \frac{(1-y_i)}{(1-F_i)}\} z_i \end{bmatrix} \tag{3}$$

where $\text{sgn}_i = \text{sign}(1 + z'_i \alpha)$, and

$$\frac{\partial^2 l}{\partial \gamma \partial \gamma'} = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta \partial \beta'} & \frac{\partial^2 l}{\partial \beta \partial \alpha'} \\ \\ \frac{\partial^2 l}{\partial \alpha \partial \beta'} & \frac{\partial^2 l}{\partial \alpha \partial \alpha'} \end{bmatrix} = \begin{bmatrix} A & C \\ \\ C' & B \end{bmatrix} \tag{4}$$

where

$$A = -\sum_{i=1}^{n} f_i [\{\frac{y_i}{F_i^2} + \frac{(1-y_i)}{(1-F_i)^2}\} f_i$$

$$+ \{\frac{y_i}{F_i} - \frac{(1-y_i)}{(1-F_i)}\} \frac{x'_i \beta}{\sigma_i^2}] x_i x'_i$$

$$B = -\sum_{i=1}^{n} y_i x'_i \beta [\frac{f_i^2 x'_i \beta}{F_i^2 \sigma_i^2} + \frac{f_i(-2\sigma_i^2 + (x_i' \beta)^2)}{F_i \sigma_i^4}] z_i z'_i$$

$$+ \sum_{i=1}^{n} (1-y_i) x'_i \beta [-\frac{f_i^2 x'_i \beta}{(1-F_i)^2 \sigma_i^2}$$

$$+ \frac{f_i(-2\sigma_i^2 + (x'_i \beta)^2)}{(1-F_i)\sigma_i^4}] z_i z'_i$$

$$C = \sum_{i=1}^{n} (\text{sgn}_i) y_i [\frac{f_i \{-\sigma_i^2 + (x'_i \beta)^2\}}{F_i \sigma_i^3} + \frac{f_i^2 x'_i \beta}{F_i^2 \sigma_i}] x_i z'_i$$

$$-\sum_{i=1}^{n} (\text{sgn}_i)(1-y_i) [\frac{f_i \{-\sigma_i^2 + (x'_i \beta)^2\}}{(1-F_i)\sigma_i^3}$$

$$- \frac{f_i^2 x'_i \beta}{(1-F_i)^2 \sigma_i}] x_i z'_i$$

and

$$f_i = (2\pi \sigma_i^2)^{-1/2} \exp\{-(2\sigma_i^2)^{-1}(x'_i \beta)^2\}$$
$$= (1/\sigma_i) \cdot \phi(x'_i \beta / \sigma_i).$$

Using the above notation, the homoscedastic probit model gives rise to the gradient vector $\frac{\partial l}{\partial \beta}$ and the Hessian matrix A after setting $\sigma_i = \sigma = 1$. It is known that in the homoscedastic probit model, $(-A)$ is positive definite and hence the loglikelihood function is globally concave. In contrast to this, we have the following negative result in the heteroscedastic probit model.

Theorem 1
The loglikelihood function (1) of the heteroscedastic probit model is not globally concave.

Proof: It is easy to see that $(-A)$ is still positive definite in the heteroscedastic case, but the following simple counter-example to the positive definiteness shows that $(-\frac{\partial^2 l}{\partial \gamma \partial \gamma'})$ is not necessarily positive definite.

We consider the following simple heteroscedastic probit model:

$$y^*_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \tag{5}$$

where x_{2i} and x_{3i} are scalars, $u_i \sim N(0, \sigma_i^2)$ independent over i and $\sigma_i^2 = (1 + \alpha x_{2i})^2$. The data for y_i, x_{2i} and x_{3i} consist of

$$y' = (0\ 0\ 1\ 0\ 1\ 0\ 1\ 0)$$

$$x'_2 = (2.875\ 4.5\ 5.625\ 5.625\ 6.25\ 6.25\ 6.75\ 6.75)$$

and

$$x'_3 = (1\ 0\ 1\ 0\ 1\ 0\ 1\ 0).$$

Then one of the diagonal elements of $-\frac{\partial^2 l}{\partial \gamma \partial \gamma'}$, $-\frac{\partial^2 l}{\partial \alpha^2}$ evaluated at $\beta_1 = -0.005$, $\beta_2 = -0.003$, $\beta_3 = 0.004$ and $\alpha = 0.008$, has the negative value -1.1086.

Q.E.D.

Since the loglikelihood function of the heteroscedastic probit model is, in general, not globally concave, one has to try out several different sets of starting values in search of the global maximum.

The heteroscedastic MLE is defined as

$$\hat{\gamma} = \operatorname*{argmax}_{\gamma \in \Gamma \equiv B \times A} l.$$

Then, we need the following assumptions to prove consistency and asymptotic normality of $\hat{\gamma}$.

Assumption 1
The parameter space Γ is open bounded subset of the Euclidean

K-space.

Assumption 2
$\{x_i\}$ are uniformly bounded in i and $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} x_i x'_i$ is a finite nonsingular matrix. Furthermore, the empirical distribution function of $\{x_i\}$ converges to a distribution function.

Assumption 3
$\{z_i\}$ are uniformly bounded in i and $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} z_i z'_i$ is a finite nonsingular matrix. Furthermore, $(1 + z'_i\alpha)$ is bounded away from zero for every i.

Remark
The first two assumptions are the same as the Assumptions 9.2.1-9.2.3 in Amemiya (1985) and Assumption 3 is a simple adaptation of the assumptions in Amemiya (1977). Assumption 3 should be replaced by the following Assumption 3M when we consider the multiplicative heteroscedastic model [Model 1M] in the Appendix B. Also note that $\{z_i\}$ are not necessarily subsets of $\{x_i\}$.

Assumption 3M
$\{z_i\}$ are uniformly bounded in i and $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} z_i z'_i$ is a finite nonsingular matrix.

Theorem 2
Under the Assumptions 1-3, the heteroscedastic probit MLE $\hat{\gamma}$ of γ is consistent and asymptotically normal with the asymptotic variance-covariance matrix $[-\mathcal{E}\frac{\partial^2 l}{\partial\gamma\partial\gamma'}]^{-1}\Big|_{(\beta'_0\alpha'_0)'}$, where

$$-\mathcal{E}\frac{\partial^2 l}{\partial\gamma\partial\gamma'} =$$

$$\begin{bmatrix} \sum_{i=1}^{n}\frac{f_i^2}{F_i(1-F_i)}x_i x'_i & \sum_{i=1}^{n}(\text{sgn}_i)\frac{f_i^2 x'_i\beta}{\sigma_i F_i(1-F_i)}x_i z'_i \\ & \sum_{i=1}^{n}\frac{f_i^2(x'_i\beta)^2}{\sigma_i^2 F_i(1-F_i)}z_i z'_i \end{bmatrix} \qquad (6)$$

Proof: Consistency is immediately shown from Theorems 4.1.2, 4.2.2 and 4.2.3 in Amemiya (1985). Asymptotic normality is based on Theorem 4.2.4 in Amemiya (1985).

Q.E.D.

B. Heteroscedastic Logit Model

Basically the same results as those stated for the probit specification hold under the logit specification, but they are re-stated briefly since the variance term containing a set of parameters α appears differently in the distribution functions.

We consider Model 1A with $u_i \sim$ an independent logistic distribution. To accommodate the heteroscedastic variance structure [note that $\sigma_i^2 = (1 + z'_i \alpha)^2$ is assumed], we write

$$F_i = \Pr(y_i = 1) = \Pr(u_i > -x'_i \beta)$$
$$= \Pr(u_i < x'_i \beta) = \Lambda(x'_i \beta : \sigma_i).$$

Since $f_i = \frac{\partial F_i}{\partial (x'_i \beta)} = \frac{F_i(1 - F_i)}{\sigma_i}$, the gradient vector and the Hessian matrix of l are derived as:

$$\frac{\partial l}{\partial \gamma} = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \\ \frac{\partial l}{\partial \alpha} \end{bmatrix} = \left(\frac{\pi}{\sqrt{3}}\right) \begin{bmatrix} \sum_{i=1}^{n} \frac{(y_i - F_i)}{\sigma_i} x_i \\ \\ -\sum_{i=1}^{n} \frac{(\mathrm{sgn}_i)(y_i - F_i) x'_i \beta}{\sigma_i^2} z_i \end{bmatrix} \tag{7}$$

and

$$\frac{\partial^2 l}{\partial \gamma \partial \gamma'} = \begin{bmatrix} A & C \\ C' & B \end{bmatrix}, \tag{8}$$

where

$$A = -\left(\frac{\pi^2}{3}\right) \sum_{i=1}^{n} \frac{F_i(1 - F_i)}{\sigma_i^2} x_i x'_i$$

$$B = -\left(\frac{\pi^2}{3}\right) \sum_{i=1}^{n} \frac{F_i(1 - F_i)(x'_i \beta)^2}{\sigma_i^4} z_i z'_i$$
$$+ \left(\frac{2\pi}{\sqrt{3}}\right) \sum_{i=1}^{n} \frac{(y_i - F_i)(x'_i \beta)}{\sigma_i^3} z_i z'_i$$

$$C = \left(\frac{\pi^2}{3}\right) \sum_{i=1}^{n} \frac{F_i(1 - F_i)(x'_i \beta)}{\sigma_i^3} x_i z'_i$$
$$- \left(\frac{\pi}{\sqrt{3}}\right) \sum_{i=1}^{n} \frac{(\mathrm{sgn}_i)(y_i - F_i)}{\sigma_i^2} x_i z'_i.$$

As in the probit case, $(-A)$ is p.d., but $-\frac{\partial^2 l}{\partial \gamma \partial \gamma'}$ is not necessarily p.d. And the value of the counter-example in the previous subsection is -1.2305, with the same set of data and at the same parameter values.

Finally, under the Assumptions 1-3, the heteroscedastic logit MLE is consistent and asymptotically normal with the asymptotic variance-covariance matrix $(-\mathcal{E}\frac{\partial^2 l}{\partial \gamma \partial \gamma'})^{-1}\Big|_{(\beta'_0\alpha'_0)'}$, where

$$-\mathcal{E}\frac{\partial^2 l}{\partial \gamma \partial \gamma'} = (\frac{\pi^2}{3})$$

$$\times \begin{bmatrix} \sum_{i=1}^{n} \frac{F_i(1-F_i)}{\sigma_i^2} x_i x'_i & -\sum_{i=1}^{n} (\text{sgn}_i) \frac{F_i(1-F_i)(x'_i\beta)}{\sigma_i^3} x_i z'_i \\ & \sum_{i=1}^{n} \frac{F_i(1-F_i)(x'_i\beta)^2}{\sigma_i^4} z_i z'_i \end{bmatrix}.$$

III. Inconsistency of the Homoscedastic MLE under Heteroscedasticity

Consistency of the homoscedastic MLE under regularity conditions, including the homoscedasticity assumption, is proved by Amemiya (1985) and by Gourieroux and Monfort (1981). Gourieroux and Monfort prove consistency of the homoscedastic logit MLE under a less restrictive setting. In this section, we study the asymptotic property of the homoscedastic MLE when the true model is heteroscedastic. Note that, before the usual normalization, the true argument of F_i is $x'_i\beta/\sigma_i$, but the homoscedastic MLE uses $x'_i\beta/\sigma$ incorrectly.

Under a very restrictive structure—the same number of observations for each distinct σ_i and the same replications of covariates for each distinct σ_i, Yatchew and Griliches (1979) derive from their local approximation formula for inconsistency the following locally approximated condition among σ and σ_i's, which results in no asymptotic bias of the homoscedastic MLE:

$$\frac{1}{K}\sum_{j=1}^{K}\frac{1}{\sigma_j} = \frac{1}{\sigma} \text{ (where } K \text{ is the number of distinct } \sigma_i\text{'s).}$$

However, generally, with the probit specification, in a nonlocal or global case the following condition is necessary even under the same restrictive structure:

$$\frac{1}{K}\sum_{j=1}^{K}\Phi(\frac{x_l'\beta}{\sigma_j}) = \Phi(\frac{x_l'\beta}{\sigma}), \text{ for every } l = 1,\cdots,N,$$

where N is the number of covariate vectors for each distinct σ_i. [N is the same across $j = 1,\cdots,K$.]

Let us consider the following homoscedastic probit model:

$$y^*_i = x'_i\beta + \tilde{u}_i, \quad \tilde{u}_i \sim IIN(0, \sigma^2). \tag{9}$$

Then, the homoscedastic probit MLE $\hat{b}$ *of* $b = \frac{\beta}{\sigma}$ maximizes the loglikelihood function:

$$l^* = \sum_{i=1}^{n} y_i \ln\Phi(x'_ib) + \sum_{i=1}^{n}(1 - y_i)\ln(1 - \Phi(x'_ib)).$$

Theorem 3

Under Model 1 with normal errors, the homoscedastic probit MLE $\hat{b}$ is in general inconsistent.

Proof: Under Model 1 with normal errors, $\hat{b}$ is a consistent estimator of $b^* \in B \equiv \{b \mid \frac{\partial Q^*(b)}{\partial b} = 0\}$, where

$$Q^*(b) = \lim_{n\to\infty}\frac{1}{n}[\sum_{i=1}^{n}\Phi(x'_i\beta_0/\sigma_0)\ln\Phi(x'_ib)$$

$$+ \sum_{i=1}^{n}\{1 - \Phi(x'_i\beta_0/\sigma_0)\}\ln\{1 - \Phi(x'_ib)\}]$$

and

$$\frac{\partial Q^*(b)}{\partial b} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\phi(x'_ib)x_i\{\Phi(x'_i\beta_0/\sigma_0)$$

$$-\Phi(x'_ib)\}[\Phi(x'_ib)\{1 - \Phi(x'_ib)\}]^{-1}.$$

However, in general, $\frac{\partial Q^*(b)}{\partial b}$ does not vanish unless $\sigma_{i0} = \sigma_0$ for every i. (It corresponds to $\alpha_0 = 0$ under Model 1A or 1M after normalization.) Hence, the homoscedastic probit MLE $\hat{b}$ for b is generally inconsistent when the true specification is heteroscedastic. When $\beta_0 = 0$, Model 0 and Model 1A or Model 1M are indistinguishable, and we can easily see that the homoscedastic MLE $\hat{b}$ is consistent for $b^* = 0$.

Q.E.D.

IV. Empirical Examples

In this section, we report the estimation results of both homoscedastic and heteroscedastic specifications using two real data sets. However, in this and following sections, we impose the restriction that $z_i = x_i^*$.

The first data is a grouped data extracted from the *Survey of Consumer Finances* (1969) by Amemiya and Nold (1975). The data and variable definitions are reproduced in Table A1 in the Appendix A. Table A2 reports the maximum likelihood estimation results. Since there are many observations per cell, we may apply a similar procedure to minimum-χ^2 in the estimation of the heteroscedastic model. However, the extension of minimum-χ^2 procedure to the heteroscedastic model results in no computational simplicity since we end up with a nonlinear weighted least squares procedure which requires iterations. Hence, we simply report the maximum likelihood estimation results. Also note that in the logit expression $\frac{\pi}{\sqrt{3}}$ is explicitly factored out so that comparison of parameters from probit and logit models is immediate without rescaling. Using the notation defined in Table A1, the latent variable is described as

$$y^*_i = b_1 + b_2 x_{2i} + b_3 x_{3i} + \tilde{u}_i$$

$\tilde{u}_i \sim$ independent normal or logistic with mean 0 and variance 1

under the homoscedastic specification, and

$$y^*_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

$u_i \sim$ independent normal or logistic with mean 0 and variance σ_i^2

$$\sigma_i^2 = (1 + \alpha_2 x_{2i} + \alpha_3 x_{3i})^2$$

under the heteroscedastic specification.

The second data is an individual ungrouped data from the *Census of Population and Housing, 1980: Public-Use Microdata Samples* ("A" Sample, nationwide file 1/1000). For the modification and reduction of the data, see Table A3. For the definition of the variables and their summary statistics, see Table A4. We estimate a conventional model for female labor force decision [e.g., Bowen and

Finegan (1969), and Smith (1980)] by the maximum likelihood estimation and the results are tabulated in Table A5. Note that the format of the table is the same as that of Table A2. Among many possible heteroscedastic specifications, the reported ones are the best in terms of Akaike information criterion (AIC).

For the test of the homoscedasticity assumption on the error term versus the heteroscedasticity assumption, first of all, we can apply the trilogy of the asymptotic tests — likelihood ratio test, Lagrange multiplier test, and Wald test. Davidson and MacKinnon (1984) propose several Lagrange multiplier tests for various forms of model inadequacy — omitted variables or heteroscedasticity of known forms — in the logit and probit models. Their experiment suggests that except for the test on the coefficients appearing in the heteroscedastic variance structure, one of their proposed LM test statistics, LM_2, seems to have a small-sample distribution which is remarkably close to its asymptotic one. For the test on the coefficients in the heteroscedastic structure, which is of interest to us, LM_2 tends to reject the null less often than it should. Also known is that LM tests for homoscedasticity are invariant to different specifications for the alternatives such as multiplicative heteroscedasticity and additive heteroscedasticity, as long as the heteroscedastic structures share the same variables. For this matter, consult Bera and Jarque (1981), and Bera and McKenzie (1986). Our empirical model estimation provides the following numerical results when we compare two logit specifications. For the first data set, the likelihood ratio test for additive heteroscedasticity is 7.448 and that for multiplicative heteroscedasticity [which is described in the Appendix B], 4.226, while LM_2 is 2.694. For the second data set, the likelihood ratio tests are 27.486 and 28.568 for additive and multiplicative heteroscedasticity respectively, while LM_2 is 24.650. Indeed we find that LM_2 is smaller in magnitude than the likelihood ratio test. Based on Davidson and MacKinnon's finding of the poor performance of LM_2 in testing for the coefficients in the heteroscedastic structure and our later Monte Carlo finding of the reliable performance of the likelihood ratio test, we conclude that both data sets support the heteroscedastic specification over the homoscedastic specification based on the likelihood ratio tests.

Also applicable to our model choice between the homoscedastic and the heteroscedastic models are AIC, the sum of squared residuals weighted by the estimated probabilities, and the sum of squared residuals weighted by observed frequencies. For these, see

Amemiya (1981).

To find whether the two model specifications in fact different in some practical aspects, we first compare the closeness of the estimated success probabilities to the observed frequencies, and then we will assess the estimates of the changes in the success probabilities given a change in an explanatory variable. For these, we will use only the logit results from the first data set. Table A6 tabulates the estimated probabilities together with the observed frequencies. We find that out of 32 evaluation points the heteroscedastic success probability estimates are closer to the observed frequencies at 18 points. At one point, the difference in the success probability estimates is 0.2.

To further assess deviation of the homoscedastic model from the heteroscedastic model, we look at the change in the probability of $\{ y_i = 1 \mid x_i \}$, or the change in the expected response given a change in an explanatory variable. First, we have the following differentiation of $\Pr(y_i = 1 \mid x_i) = \mathcal{E}(y_i \mid x_i)$ from the homoscedastic logit model and the heteroscedastic logit model:

$$\frac{\partial \Lambda(x'_i b : 1)}{\partial x_{2i}} = \left(\frac{\pi}{\sqrt{3}}\right) \exp\{(\frac{\pi}{\sqrt{3}}) x'_i b\} \cdot [1 + \exp\{(\frac{\pi}{\sqrt{3}}) x'_i b\}]^{-2} \cdot b_2 \tag{10}$$

$$\frac{\partial \Lambda(x'_i \beta : \sigma_i)}{\partial x_{2i}} = \left(\frac{\pi}{\sqrt{3}\sigma_i}\right) \exp\{(\frac{\pi}{\sqrt{3}\sigma_i}) x'_i \beta\} \cdot [1 + \exp\{(\frac{\pi}{\sqrt{3}\sigma_i}) x'_i \beta\}]^{-2} \cdot \{\beta_2 - (\mathrm{sgn}_i)(x'_i \beta)\, \sigma_i^{-1}\, \alpha_2\}. \tag{11}$$

(11) shows that the incremental effect of explanatory variables on $\Pr(y_i = 1 \mid x_i)$ is not restricted to being the same when evaluated at $\Pr(y_i = 1 \mid x_i) = p$ and $\Pr(y_i = 1 \mid x_i) = 1 - p$, while (10) shows that the incremental effect is restricted to being the same. We evaluate (10) and (11) at the mean value of x_{2i}, $\bar{x}_2 = 9.8887$ using the estimates in Table A2:

$$HO0 = \frac{\partial \Lambda(\cdot:1)}{\partial x_{2i}} \Bigg|_{(x_{1i}, x_{2i}, x_{3i})=(1,\bar{x}_2,0)} = 0.01783$$

$$HE0 = \frac{\partial \Lambda(\cdot:\sigma_i)}{\partial x_{2i}} \Bigg|_{(x_{1i}, x_{2i}, x_{3i})=(1,\bar{x}_2,0)} = 0.01643$$

$$HO1 = \frac{\partial \Lambda(\cdot:1)}{\partial x_{2i}} \Bigg|_{(x_{1i}, x_{2i}, x_{3i})=(1,\bar{x}_2,1)} = 0.01508$$

$$HE1 = \frac{\partial \Lambda(\cdot:\sigma_i)}{\partial x_{2i}} \Bigg|_{(x_{1i}, x_{2i}, x_{3i})=(1,\bar{x}_2,1)} = 0.01190$$

We note that the difference between $HO0$ and $HO1$ from the homoscedastic model is the smaller than that between $HE0$ and $HE1$ from the heteroscedastic model. Hence, the two specifications seem to differ sufficiently both in probability predictions and in predicted changes in success probability.

Finally, we assess the implied degree of inconsistency of the homoscedastic model in terms of estimated probabilities. First, the probability limit b^* of the homoscedastic MLE for b is obtained under the assumption that the true model specification is heteroscedastic with $\beta_{01} = \hat{\beta}_1$, $\beta_{02} = \hat{\beta}_2$, $\beta_{03} = \hat{\beta}_3$, $\alpha_{02} = \hat{\alpha}_2$, and $\alpha_{03} = \hat{\alpha}_3$ from Table A2. By solving the following optimization problem:

$$\max_b \sum_{i=1}^{n} \mathcal{E}(y_i) \cdot \ln \Lambda(x'_i b : 1) + \sum_{i=1}^{n} \{1 - \mathcal{E}(y_i)\} \cdot \ln \{1 - \Lambda(x'_i b : 1)\}, \tag{12}$$

where $\mathcal{E}$ is taken over the true (assumed) parameter values, we get the following probability limits:[3] $\text{plim}\hat{b}_1 = -0.3619$, $\text{plim}\hat{b}_2 = 0.03828$, and $\text{plim}\hat{b}_3 = 0.4419$. The last column of Table A6 reports the estimated probabilities evaluated at the above probability limits. To measure the degree of inconsistency in terms of success probability predictions, we compare the columns 5 and 6. In general, the implied asymptotic bias in terms of the discrepancy between the estimated probabilities and the true assumed success probabilities

[3]The probability limits of the heteroscedastic logit MLE under the homoscedasticity assumption with $b_{01} = \hat{b}_1$, $b_{02} = \hat{b}_2$, and $b_{03} = \hat{b}_3$ from Table A2 are as follows: $\text{plim}\,\hat{\beta}_{01} = -0.3718$, $\text{plim}\,\hat{\beta}_{02} = 0.03934$, $\text{plim}\,\hat{\beta}_{03} = 0.4408$, $\text{plim}\,\hat{\alpha}_{02} = 0.00000$, and $\text{plim}\,\hat{\alpha}_{03} = -0.00000$.

seems relatively small except at one case where the bias is as large as 0.204.

V. Numerical Experiment

In this section, we conduct two separate numerical experiments. The first one is designed to evaluate further the degree of inconsistency of the homoscedastic MLE when the true model is heteroscedastic. The second one is to examine the empirical performance of the likelihood ratio test to show that it is a reliable test to distinguish the heteroscedastic model from the homoscedastic model.

The first experiment assumes 6 sets of the true parameters, which provide us with varying degrees of heteroscedasticity. We consider the logit model based on the first data set investigated in the previous section, but with the following sets of assumed true parameter values:

	β_{01}	β_{02}	β_{03}	α_{02}	α_{03}
I	−0.5918512	0.06794184	0.3376301	0.00078318	0.0
II	−0.5918512	0.06794184	0.3376301	0.0078318	0.0
III	−0.5918512	0.06794184	0.3376301	0.078318	0.0
IV	−0.5918512	0.06794184	0.3376301	0.78318	0.0
V	−0.5918512	0.06794184	0.3376301	7.8318	0.0
VI	−0.5918512	0.06794184	0.3376301	78.318	0.0

In the experiment, the original variance structure is further normalized by the mean of the variance terms $(1 + z'_i \alpha)^2$ where $i = 1,\cdots,n$, to investigate the pure effect of increasing heteroscedasticity. That is, the experiments from No. I to No. VI are designed to increase the variability of the variance terms across observations, with the mean magnitude of the variance terms held fixed[4] in all 6 experimental set-ups. Hence, the variance structure adopted in this numerical study is

$$\sigma_i^2 = (1 + z'_i \alpha)^2 / [\frac{1}{n} \sum_{k=1}^{n} (1 + z'_k \alpha)^2]. \qquad (10)$$

The probability limits of the homoscedastic MLE, the deviations of

[4]If we do not apply this normalization, probability deviations first increase with heteroscedasticity and then they start to decrease since dominant magnitudes of the variance terms implied by increasing unnormalized heteroscedasticity eventually drive all the success probabilities close to 0.5.

probabilities between estimated probabilities and true assumed probabilities for 6 different sets of assumed parameter values are compiled in Table A7. The following expected relationship

$$\min_i(\beta_{0k}/\sigma_i) \leq \text{plim}\, \hat{b}_k \leq \max_i(\beta_{0k}/\sigma_i), \quad k = 1,2,3,$$

is observed, but σ_k^*'s, such that $\text{plim}\, \hat{b}_k = \beta_{0k}/\sigma_k^*$ $(k = 1,2,3)$, are not the same. In fact, $\frac{\sigma_2^*}{\sigma_3^*}$ and $\frac{\sigma_2^*}{\sigma_1^*}$ are increasing as the heteroscedasticity through α_{02} increases.[5] It is also observed that inconsistency increases with heteroscedasticity, in terms of the deviations of the estimated probabilities from the true assumed probabilities.

The second experiment is a Monte Carlo study for a homoscedastic probit model and a heteroscedastic probit model with an exponential variance structure (described in the Appendix B) to examine the empirical performance of the likelihood ratio test in distinguishing the two models. Here, the data for the explanatory variables are the same as those used in the first empirical examples, but the disturbance terms for the latent index function are generated according to $N(0, \sigma_i^2)$ where $\sigma_i = \exp(z'_i \alpha)$ for the experiment under heteroscedasticity, and according to $N(0, 1)$ for the experiment under homoscedasticity. The true models used in this Monte Carlo study are the estimated homoscedastic probit model and the estimated heteroscedastic probit model with exponential heteroscedastic structure, whose estimated coefficients are reported in Tables A2 and A10. For both cases, 100 replications with 1523 samples each are made. Table A8 summarizes the Monte Carlo results. Included are the sample means and medians of the 100 homoscedastic MLE's and the 100 heteroscedastic MLE's, together with the sample means and medians of the maximized loglikelihoods times (−2). The table shows that when the true model is heteroscedastic, the likelihood ratio test is 7.453 at the sample mean of the 100 replications and 8.006 at the sample median. On the other hand, when the true model is homoscedastic, the likelihood ratio test is 1.816 at the sample mean and 1.101 at the sample median of the 100 replications. Table A9 tabulates the percentage of 100 replications that reject the homoscedastic model in favor of the heteroscedastic model at the varying significance level. The first column lists the

[5] Hence, the common variance corresponding to a homoscedastic model before normalization does not have an easily identifiable relationship with the variances in a heteroscedastic model.

significance level chosen, the second column provides the empirical power of the likelihood ratio test when the asymptotic critical points based on χ^2_2 is applied, and the last column corresponds to the empirical size of the test. Even with this small number of replications the closeness of the theoretical significance level and the level of empirical Type I error for the likelihood ratio test is noted. For example, if we set the significance level at 5%, then the empirical Type I error is around 3%, while the power of the test is around 55%. And, if we apply *AIC* in choosing the appropriate model, then we choose the heteroscedastic model correctly with the 55% success rate when true is the heteroscedastic probit model, and we choose the homoscedastic model correctly with the 89% success rate when true is the homoscedastic probit model. From Tables A8 and A9, we conclude that the heteroscedastic model is satisfactorily distinguishable from the homoscedastic model by using the likelihood ratio test.

VI. Conclusion

The qualitative response model with an explicit heteroscedastic structure is studied. Then, the heteroscedastic MLE is shown to be consistent and asymptotically normal, while the standard MLE is inconsistent under heteroscedasticity. From the estimation results from the two real data sets, a heteroscedastic structure is shown to be supported. It is also illustrated that the degree of inconsistency from the homoscedastic ML estimation, in terms of probability predictions, can be very large as heteroscedasticity increases. A heteroscedastic model is successfully distinguishable from a homoscedastic model by the likelihood ratio test. However, computation time increases substantially due to nonconcavity of the loglikelihood function and the added nonlinearity through the variance structure.

Appendix A

TABLE A1
DATA ON SUBSAMPLE OF 1523 HOUSEHOLDS SELECTED FROM THE *SURVEY OF CONSUMER FINANCES* (1969)

Household Disposable Income (mid-range in $1,000)	Those who moved in 1967 or 1968		Those who did not move	
	n_t	r_t	n_t	r_t
2.875	32	13	43	18
4.500	36	15	69	24
5.625	35	24	45	21
6.250	24	14	34	16
6.750	45	35	45	18
7.250	34	29	48	22
7.750	26	19	56	27
8.250	27	20	41	24
8.750	32	19	51	23
9.375	31	26	73	38
10.075	53	34	51	26
11.000	39	24	89	50
12.000	43	30	86	41
13.500	35	28	105	69
16.000	32	26	78	41
21.250	25	20	60	41

Note: 1. n_t : number of families in the t-th category (I_t)

2. r_t : number of families in the t-th category who bought consumer durables in 1968

3. x_{1t} : variable identically equal to 1

4. x_{2t} : the mid-range of the household disposable income used to define the t-th category—that is, in the estimation I used $x_{2t} = x_{2i}$ for every i in I_t, where I_t is the t-th category index set

5. x_{3t} : dummy variable equal to 1 if the household moved in 1967 or in 1968, and 0 otherwise. $x_{3t} = x_{3i}$ for every i in I_t.

TABLE A2
MAXIMUM LIKELIHOOD ESTIMATION (FIRST EXAMPLE)

Variable	Probit		Logit	
	Homo.	Hetero.	Homo.	Hetero.
		Main Equation		
X_1	−0.413(0.0876)	−0.678(0.2873)	−0.372(0.0790)	−0.592(0.2499)
X_2	0.044(0.0076)	0.078(0.0342)	0.039(0.0070)	0.068(0.0296)
X_3	0.494(0.0698)	0.387(0.1757)	0.441(0.0631)	0.338(0.1531)
		Variance Equation		
X_2		0.082(0.0488)		0.078(0.0475)
X_3		−1.011(0.1764)		−1.004(0.1745)
$-2l_{max}$	2001.360	1993.902	2001.280	1993.832

Note: Standard errors are in parentheses.

TABLE A3
DATA SET MODIFICATION AND REDUCTION (ORDER OF PRECEDENCE)

Number of households on the original tape	94,205
Number of households vacant at the time of the census	(−) 8,021
Number of households with female (who is either head or spouse of male head) not married or married but separated	(−)36,990
Number of households with female (who is either head or spouse of male head) not in the specified age range (25–55, inclusive) at the time of census	(−)15,976
Number of households with female (who is either head or spouse of male head) neither black nor white	(−) 1,549
Number of households with female (who is either head or spouse of male head) having farm income in 1979	(−) 203
Number of households with female (who is either head or spouse of male head) having bad (i.e., allocated) labor data or truncated income	(−) 9,593
Number of households with female (who is either head or spouse of male head) having nonfarm self-employment income	(−) 334
Number of households with male (who is either head or spouse of female head) having farm or nonfarm self-employment income	(−) 3,009
Number of households with male (who is either head or spouse of female head) having bad (i.e., allocated) labor data or truncated income	(−) 3,493

TABLE A3 (Continued)

Number of households with male (who is either head or spouse of female head) not present in the household	(—) 99
Number of households with female (who is either head or spouse of male head) not white	(—) 1,002
Total Eligible Households	13,486
Number of households picked up every 10th observation	1,348

TABLE A4

SUMMARY ON VARIABLES AND SAMPLE MEANS

Variable	Description	Sample Mean
Y	Dummy on wife's annual labor force participation in 1979	0.6454
X_1	Constant term	
X_2	Family income other than wife's wage income ($1,000) in 1979	22.5677
X_3	Wife's age (as of census date)	38.5475
X_4	Wife's completed grade (as of 1979)	12.4696
X_5	Number of young children (0–6 yrs. of age in 1979, or 1–7 of age in 1980)	0.4303
X_6	Number of older children (7–16 yrs. of age in 1979, or 8–17 yrs. of age in 1980)	0.8954
X_7	Number of adults in the household other than wife and husband (17 and over in 1979, or 18 and over in 1980)	0.3197
X_8	Dummy on SMSA's (urban dummy)	0.8190
X_9	Southern residence dummy	0.3131
X_{10}	Health disability dummy	0.0653
X_{11}	Dummy of family living on social security income	0.0438

TABLE A5

MAXIMUM LIKELIHOOD ESTIMATION (SECOND EXAMPLE)

Variable	Probit		Logit	
	Homo.	Hetero.	Homo.	Hetero.
		Main Equation		
X_1	1.482(0.3277)	1.588(0.2732)	1.353(0.3103)	1.533(0.2987)
X_2	−0.015(0.0028)	−0.010(0.0029)	−0.014(0.0028)	−0.010(0.0030)
X_3	−0.033(0.0053)	−0.030(0.0048)	−0.031(0.0049)	−0.029(0.0052)
X_4	0.084(0.0157)	0.035(0.0120)	0.078(0.0152)	0.035(0.0128)
X_5	−0.631(0.0611)	−0.503(0.1059)	−0.580(0.0602)	−0.490(0.1122)
X_6	−0.092(0.0350)	−0.089(0.0259)	−0.087(0.0316)	−0.087(0.0265)
X_7	0.107(0.0598)	0.157(0.0768)	0.100(0.0546)	0.155(0.0771)
X_8	−0.101(0.1009)	−0.093(0.0641)	−0.093(0.0914)	−0.090(0.0628)
X_9	−0.090(0.0809)	−0.055(0.0500)	−0.084(0.0730)	−0.054(0.0490)
X_{10}	−0.450(0.1385)	−0.315(0.1086)	−0.409(0.1318)	−0.309(0.1158)
X_{11}	−0.367(0.1925)	−0.235(0.1112)	−0.332(0.1635)	−0.229(0.1094)
		Variance Equation		
X_3		−0.013(0.0022)		−0.012(0.0026)
X_5		0.615(0.3403)		0.711(0.3839)
X_7		0.244(0.1382)		0.271(0.1613)
$-2l_{max}$	1559.056	1532.335	1560.081	1532.335

Note: Standard errors are in parentheses.

TABLE A6
COMPARISON OF THE ESTIMATED PROBABILITIES

x_{2i}	x_{3i}	$\hat{P}_t$	$\hat{F}_t$(homo.)	$\hat{F}_t$(hete.)	$\hat{F}_t$(plim homo.)
2.875	1	0.4063	0.5818	0.3818	0.5853
	0	0.4186	0.3848	0.3573	0.3877
4.500	1	0.4167	0.6097	0.5665	0.6124
	0	0.3478	0.4126	0.4052	0.4148
5.625	1	0.6857	0.6287	0.6297	0.6308
	0	0.4667	0.4322	0.4344	0.4339
6.250	1	0.5833	0.6390	0.6538	0.6409
	0	0.4706	0.4431	0.4493	0.4446
6.750	1	0.7778	0.6472	0.6695	0.6488
	0	0.4000	0.4520	0.4606	0.4532
7.250	1	0.8529	0.6553	0.6827	0.6567
	0	0.4583	0.4608	0.4713	0.4618
7.750	1	0.7308	0.6633	0.6939	0.6645
	0	0.4821	0.4697	0.4816	0.4704
8.250	1	0.7407	0.6712	0.7036	0.6721
	0	0.5854	0.4786	0.4914	0.4790
8.750	1	0.5938	0.6791	0.7120	0.6798
	0	0.4510	0.4875	0.5007	0.4878
9.375	1	0.8387	0.6887	0.7211	0.6891
	0	0.5205	0.4986	0.5118	0.4986
10.075	1	0.6415	0.6993	0.7298	0.6994
	0	0.5098	0.5111	0.5235	0.5108
11.000	1	0.6154	0.7130	0.7393	0.7128
	0	0.5618	0.5276	0.5378	0.5268
12.000	1	0.6977	0.7274	0.7478	0.7268
	0	0.4767	0.5453	0.5520	0.5441
13.500	1	0.8000	0.7481	0.7579	0.7470
	0	0.6571	0.5717	0.5712	0.5698
16.000	1	0.8125	0.7802	0.7701	0.7783
	0	0.5256	0.6147	0.5984	0.6117
21.250	1	0.8000	0.8377	0.7857	0.8349
	0	0.6833	0.6989	0.6411	0.6940

Note: $\hat{P}_t$ denotes the observed frequencies from the raw data. $\hat{F}_t$(homo.) and $\hat{F}_t$(hete.) denote the estimates for $\Pr(y_i = 1 \mid x_{2i}, x_{3i})$ from the homoscedastic model estimation and the heteroscedastic model estimation, respectively. $\hat{F}_t$(plim homo.) is the success probability estimates from the homosecdastic model but evaluated at the plim of the parameter estimate assuming that the true model is the estimated heteroscedastic model.

TABLE A7

NUMERICAL STUDY ON HOMOSCEDASTIC MLE UNDER HETEROSCEDASTICITY

No.	plim $\hat{b}_1$	plim $\hat{b}_2$	plim $\hat{b}_3$	$\frac{\text{plim } \hat{b}_2}{\text{plim } \hat{b}_1}$	$\frac{\text{plim } \hat{b}_3}{\text{plim } \hat{b}_1}$
I	−0.5902	0.06769	0.3378	−0.1147	−0.5724
II	−0.5779	0.06573	0.3398	−0.1137	−0.5879
III	−0.5503	0.05917	0.3628	−0.1075	−0.6593
IV	−0.6147	0.06068	0.4283	−0.09828	−0.6938
V	−0.6497	0.06249	0.4497	−0.09618	−0.6922
VI	−0.6537	0.06273	0.4522	−0.09595	−0.6918

No.	Max. dev. of est. prob. from true	Average deviation	Est. prob. at$(1, \bar{x}_2, 0)$	Est. prob. at$(1, \bar{x}_2, 1)$
I	0.001229	0.000351	0.5358	0.6805
II	0.01157	0.003225	0.5326	0.6785
III	0.07693	0.01836	0.5158	0.6729
IV	0.1978	0.04074	0.4921	0.6782
V	0.2307	0.04749	0.4856	0.6809
VI	0.2342	0.04830	0.4848	0.6813

Note: The variance is normalized as in (10). The true assumed parameter values are $\beta_{01} = -0.5918512$; $\beta_{02} = 0.06794184$; $\beta_{03} = 0.3376301$; $\alpha_{02} =$ 0.00078318 (No. I), 0.0078318 (No. II), 0.078318 (No. III), 0.78318 (No. IV), 7.8318 (No. V), 78.318 (No. VI); and $\alpha_{03} = 0.0$. And the true probabilities at $(1, \bar{x}_2, 0)$ and $(1, \bar{x}_2, 1)$ are 0.536214 and 0.680811, respectively.

TABLE A8
MONTE CARLO RESULTS

	I(a)	I(b)	II(a)	II(b)
$-2l^{*}_{max}$	2000.552	2002.487	1998.223	1999.917
$-2l_{max}$	1993.099	1994.481	1996.407	1997.816
$\hat{b}_1$	−0.392	−0.393	−0.424	−0.426
$\hat{b}_2$	0.0412	0.0416	0.0447	0.0437
$\hat{b}_3$	0.502	0.493	0.495	0.491
$\hat{\beta}_{01}$	−0.709	−0.690	−0.428	−0.422
$\hat{\beta}_{02}$	0.0835	0.0797	0.0456	0.0449
$\hat{\beta}_{03}$	0.444	0.427	0.588	0.492
$\hat{\alpha}_{02}$	0.0592	0.0592	−0.00127	0.00247
$\hat{\alpha}_{03}$	−0.702	−0.670	0.0506	0.0246

Note: The true model for I(a) and I(b) is the heteroscedastic probit model with $\sigma_i = \exp(z'_i\alpha)$ and the parameter values $\beta_{01} = -0.726261$, $\beta_{02} = 0.085520$, $\beta_{03} = 0.440229$, $\alpha_{02} = 0.060630$, and $\alpha_{03} = -0.663690$. And the true model for II(a) and II(b) is the homoscedastic model with the parameter values $b_{01} = -0.413360$, $b_{02} = 0.043738$, and $b_{03} = 0.493866$. The above assumed parameter values are from the estimation results (for the first empirical data set) of the models described in the Appendix B. The numbers recorded under I(a) and II(a) are sample means from 100 replications, while those recorded under I(b) and II(b) are sample medians.

TABLE A9

PERCENTAGE OF REPLICATIONS REJECTING A HOMOSCEDASTIC MODEL IN FAVOR OF A HETEROSCEDASTIC MODEL AT THE SIGNIFICANCE LEVEL α

$$\Pr\{LRT = -2(l^*_{\max} - l_{\max}) \geq \chi^2_{(2:\alpha)}\} = \alpha$$

Sig. level(α)	When true is a hete. model	When true is a homo. model
0.001	10	0
0.005	30	1
0.010	34	1
0.025	45	2
0.050	55	3
0.075	61	5
0.100	66	7
0.200	80	14
0.300	85	25
0.400	91	44
0.500	93	48
0.600	95	58
0.700	95	65
0.800	97	81
0.900	99	90

Note:The percentage rejecting a homoscedastic model in favor of a heteroscedastic model based on Akaike information criterion is as follows:

1) When the true model is heteroscedastic — 55%.
2) When the true model is homoscedastic — 11%.

TABLE A10

HETEROSCEDASTIC LOGIT AND PROBIT MLE UNDER MULTIPLICATIVE HETEROSCEDASTICITY

Variable	Probit	Logit
	First Data Set	
	Main Equation	
X_1	−0.726(0.2399)	−0.643(0.2135)
X_2	0.086(0.0301)	0.076(0.0270)
X_3	0.440(0.1494)	0.385(0.1325)
	Variance Equation	
X_2	0.061(0.0250)	0.061(0.0255)
X_3	−0.664(0.2985)	−0.690(0.2986)
$-2l_{max}$	1997.191	1997.014
	Second Data Set	
	Main Equation	
X_1	1.164(0.3828)	1.170(0.4248)
X_2	−0.007(0.0032)	−0.007(0.0036)
X_3	−0.022(0.0072)	−0.022(0.0081)
X_4	0.026(0.0125)	0.027(0.0139)
X_5	−0.367(0.1387)	−0.374(0.1543)
X_6	−0.065(0.0266)	−0.066(0.0296)
X_7	0.124(0.0709)	0.126(0.0752)
X_8	−0.068(0.0509)	−0.068(0.0539)
X_9	−0.039(0.0385)	−0.040(0.0403)
X_{10}	−0.230(0.1132)	−0.233(0.1217)
X_{11}	−0.162(0.0978)	−0.165(0.1035)
	Variance Equation	
X_3	−0.026(0.0097)	−0.024(0.0105)
X_5	0.619(0.2209)	0.665(0.2281)
X_7	0.438(0.1718)	0.446(0.1771)
$-2l_{max}$	1531.253	1531.513

Note: Standard errors are in parentheses.

Appendix B

Here, probit and logit models with the exponential heteroscedastic structure are considered. By the exponential heteroscedastic structure or multiplicative heteroscedasticity, we mean

$$\sigma_i^2 = \exp(2z'_i\alpha) \text{ or } \sigma_i = \exp(z'_i\alpha). \tag{B1}$$

Then, we arrive at the same expression for the loglikelihood function as in (1) once (B1) is noted.

First, we consider the heteroscedastic probit model and then the heteroscedastic logit model. Table A10 tabulates the estimation results from both grouped and ungrouped data sets.

B1. Heteroscedastic Probit Model

With the exponential heteroscedasticity structure, we obtain the following gradient vector, Hessian matrix and the inverse of the asymptotic variance-covariance matrix from the heteroscedastic probit specification:

$$\frac{\partial l}{\partial \gamma} = \begin{bmatrix} \sum_{i=1}^{n} (\frac{y_i}{F_i} - \frac{1-y_i}{1-F_i}) f_i x_i \\ -\sum_{i=1}^{n} (\frac{y_i}{F_i} - \frac{1-y_i}{1-F_i}) f_i x'_i \beta z_i \end{bmatrix}$$

$$\frac{\partial^2 l}{\partial \gamma \partial \gamma'} = \begin{bmatrix} A & C \\ C' & B \end{bmatrix}$$

where

$$A = -\sum_{i=1}^{n} (\frac{y_i}{F_i} - \frac{1-y_i}{1-F_i}) f_i (\frac{x'_i\beta}{\sigma_i}) x_i x'_i$$

$$-\sum_{i=1}^{n} (\frac{y_i}{F_i^2} + \frac{1-y_i}{(1-F_i)^2}) f_i^2 x_i x'_i$$

$$B = -\sum_{i=1}^{n} (\frac{y_i}{F_i} - \frac{1-y_i}{1-F_i}) \{-1 + \frac{(x'_i\beta)^2}{\sigma_i^2}\} f_i x_i'\beta z_i z'_i$$

$$-\sum_{i=1}^{n} (\frac{y_i}{F_i^2} + \frac{1-y_i}{(1-F_i)^2}) f_i^2 (x_i'\beta)^2 z_i z_i'$$

and

$$C = \sum_{i=1}^{n} (\frac{y_i}{F_i} - \frac{1-y_i}{1-F_i}) \{-1 + \frac{(x_i'\beta)^2}{\sigma_i^2}\} f_i x_i z_i'$$

$$+ \sum_{i=1}^{n} (\frac{y_i}{F_i^2} + \frac{1-y_i}{(1-F_i)^2}) f_i^2 x_i' \beta x_i z_i'.$$

$$-\varepsilon \frac{\partial^2 l}{\partial \gamma \partial \gamma'} =$$

$$\begin{bmatrix} \sum_{i=1}^{n} (\frac{1}{F_i} + \frac{1}{1-F_i}) f_i^2 x_i x_i' & -\sum_{i=1}^{n} (\frac{1}{F_i} + \frac{1}{1-F_i}) f_i^2 x_i' \beta x_i z_i' \\ & \sum_{i=1}^{n} (\frac{1}{F_i} + \frac{1}{1-F_i}) f_i^2 (x_i'\beta)^2 z_i z_i' \end{bmatrix}$$

B2. Heteroscedastic Logit Model

With the exponential heteroscedastic structure, we obtain the following gradient vector, Hessian matrix and the inverse of the asymptotic variance-covariance matrix from the heteroscedastic logit specification:

$$\frac{\partial l}{\partial \gamma} = \frac{\pi}{\sqrt{3}} \begin{bmatrix} \sum_{i=1}^{n} (y_i - F_i)(\frac{1}{\sigma_i}) x_i \\ -\sum_{i=1}^{n} (y_i - F_i)(\frac{x_i'\beta}{\sigma_i}) z_i \end{bmatrix}$$

$$\frac{\partial^2 l}{\partial \gamma \partial \gamma'} = \begin{bmatrix} A & C \\ C' & B \end{bmatrix},$$

where

$$A = -\sum_{i=1}^{n} (\frac{\pi}{\sqrt{3}\sigma_i})^2 F_i (1 - F_i) x_i x_i'$$

$$B = -\sum_{i=1}^{n} (\frac{\pi}{\sqrt{3}\sigma_i})^2 F_i (1 - F_i) (x_i'\beta)^2 z_i z_i'$$

$$+ \sum_{i=1}^{n} (\frac{\pi}{\sqrt{3}\sigma_i})(y_i - F_i)x_i'\beta z_i z_i'$$

$$C = \sum_{i=1}^{n} (\frac{\pi}{\sqrt{3}\sigma_i})^2 F_i(1 - F_i)x_i'\beta x_i z_i'$$

$$- \sum_{i=1}^{n} (\frac{\pi}{\sqrt{3}\sigma_i})(y_i - F_i)x_i z_i'.$$

$$-\varepsilon \frac{\partial^2 l}{\partial \gamma \partial \gamma'} = \frac{\pi}{\sqrt{3}}$$

$$\times \begin{bmatrix} \sum_{i=1}^{n} F_i(1 - F_i)(\frac{1}{\sigma_i^2})x_i x_i' & -\sum_{i=1}^{n} F_i(1 - F_i)(\frac{x_i'\beta}{\sigma_i^2})x_i z_i' \\ & \sum_{i=1}^{n} F_i(1 - F_i)(\frac{x_i'\beta}{\sigma_i})^2 z_i z_i' \end{bmatrix}$$

The formulae to evaluate the change in the expected response or the success probability due to change in one explanatory variable (x_{2i}) become

$$\frac{\partial \Lambda_i}{\partial x_{2i}} = (\frac{\pi}{\sqrt{3}\sigma_i})\Lambda_i(1 - \Lambda_i)\{\beta_2 - (x_i'\beta)\alpha_2\}$$

$$\frac{\partial \Lambda^*_i}{\partial x_{2i}} = (\frac{\pi}{\sqrt{3}})\Lambda^*_i(1 - \Lambda^*_i)b_2$$

where $\Lambda_i = \Lambda(x_i'\beta : \sigma_i)$ and $\Lambda^*_i = \Lambda(x_i'b : 1)$.

References

Amemiya, T. "A Note on a Heteroscedastic Model." *Journal of Econometrics* 6 (1977): 365-70.

__________. "Qualitative Response Models: A Survey." *Journal of Economic Literature* 19 (1981): 1483-536.

__________. *Advanced Econometrics.* Harvard University Press, 1985.

Amemiya, T. and Nold, F. "A Modified Logit Model." *Review of Economics and Statistics* 57 (1975): 255-57.

Bera, Anil K. and Jarque, Carlos M. "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals: Monte Carlo Evidence." *Economics Letters* 7 (1981): 313-18.

Bera, Anil K. and McKenzie, Colin R. "Alternative Forms and Properties of the Score Test." *Journal of Applied Statistics* 13 (1986) : 13-25.

Bowen, W. G. and Finegan, T. A. *The Economics of Labor Force Participation.* Princeton University Press, 1969.

Bureau of the Census. *Census of Population and Housing, 1980 : Public-Use Microdata Sample (A Sample).* Washington, 1983.

Chesher, Andrew and Irish, Margaret "Residual Analysis in the Grouped and Censored Normal Linear Model." *Journal of Econometrics* 34 (1987) : 33-61.

Cosslett, S. R. "Distribution-free Maximum Likelihood Estimator of the Binary Choice Model." *Econometrica* 51 (1983) : 765-82.

Davidson, Russell and MacKinnon, James G. "Convenient Specification Tests for Logit and Probit Models." *Journal of Econometrics* 25 (1984) : 241-62.

Egger, M. J. "Power Transformations to Achieve Symmetry in Quantal Bioassay." Technical Report No. 47, Division of Biostatistics, Stanford University, 1979.

Fomby, T. B. and Pearce, J. E. "Standard Errors in the Multinomial Logit Models." *Commun. Statist.-Theor. Meth.* 15 (1986) : 2555-68.

Goto, M., Inoue, T., and Tsuchiya, Y. "Power-transformation for Ordered Categorical Data." *Behaviormetrika* 19 (1986) : 17-39.

Gourieroux, C., and Monfort, A. "Asymptotic Properties of the Maximum Likelihood Estimator in Dichotomous Logit Models." *Journal of Econometrics* 17 (1981) : 83-97.

Gourieroux, C., Monfort, A., Renault E., and Trognon, A. "Generalized Residuals." *Journal of Econometrics* 34 (1987a) : 5-32.

____________. "Simulated Residuals." *Journal of Econometrics* 34 (1987b) : 201-52.

Gourieroux, C,. Monfort A., and Trognon, A. "Estimation and Test in Probit Models with Serial Correlation." In J. P. Florence, *et al.* (eds.), *Alternative Approaches to Time Series Analysis* (Proceedings of the 3rd Franco-Belgian Meeting of Statisticians, Nov. 1982), 1984.

____________. "A General Approach to Serial Correlation." *Econometric Theory* 1 (1985) : 315-40.

Kiefer, Nicholas M. and Skoog, Gary R. "Local Asymptotic Specification Analysis." *Econometrica* 52 (1984) : 873-85.

Kim, Jeankyung and Pollard, David. "Cube Root Asymptotics." Working Paper, Department of Statistics, Yale University, 1987.

Manski, Charles F. "Maximum Score Estimation of the Stochastic Utility Model of Choice." *Journal of Econometrics* 3 (1975) : 205-28.

____________. "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator." *Journal of Econometrics* 27 (1985) : 313-33.

Manski, Charles F. and Thompson, T. S. "Operational Charateristics of Maximum Score Estimation." *Journal of Econometrics* 32 (1986) : 85-108.

Smith, J. P. (ed.) *Female Labor Supply : Theory and Estimation.* Princeton

University Press, 1980.

Survey Research Center. *Survey of Consumer Finances.* Michigan: University of Michigan, 1969.

Yatchew, Adonis and Griliches, Zvi. "Specification Error in Probit and Logit Models." Discussion Paper No. 717. Harvard Institute of Economic Research, Harvard University, 1979.

___________. "Specification Error in Probit Models." *Review of Economics and Statistics* 67 (1985): 134-39.