DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

The Heterogenous Logit Model

Jeffrey A. Dubin

Langche Zeng

SOCIAL SCIENCE WORKING PAPER 759

February 1991

The Heterogenous Logit Model

Jeffrey A. Dubin Langche Zeng

Abstract

Probabilistic choice systems in the generalized extreme value (GEV) family embody two restrictions not shared by the covariance probit model. First, the unobserved components of random utility are homoscedastic across individuals and alternatives. Second, the degree of similarity among alternatives is also assumed to be constant across individuals. This paper considers extensions to models in the GEV class which relax these two restrictions. An empirical application concerning the demand for cameras is developed to demonstrate the potential significance of the heterogenous logit model.

The Heterogenous Logit Model

Jeffrey A. Dubin Langche Zeng*

1 Introduction

Of the existing models for qualitative decision making there is little question that the multinomial logit model has gained a certain popularity. Applications include such diverse subjects as labor force participation, occupational choice, travel mode choice, choice of residential location, and the purchase of consumer durable goods. As the use of the multinomial logit model has gained momentum, so has the reanalysis of its underlying assumptions.

One direction of inquiry has sought statistical tests for violation of the restrictive independence from irrelevant alternatives (IIA) property of the logit model.[1] Another direction of inquiry has looked for relaxation of the more rigid properties of the logit model while endeavoring to maintain computational flexibility. This research produced the nested logit model.[2] The nested logit model allows alternatives which are close substitutes to be grouped into subsets and identifies the degree of similarity within each subset. Consistency of the nested logit model with random utility maximization was demonstrated by showing that nested logit models are a special case of the generalized extreme value (GEV) family.[3]

The GEV family allows a relatively flexible pattern of correlation among the unobserved components of random utility. In this regard it bears some similarity to the covariance probit model (Hausman and Wise, 1978). The GEV class however, unlike the covariance probit model, embodies two restrictions. First, the unobserved components of utility are assumed to be homoscedastic across individuals and alternatives.[4] Second,

*The authors thank David M. Grether and Louis L. Wilde for helpful comments.

[1]See for example McFadden, Train, and Tye (1976), Hausman and McFadden (1984), Small and Hsaio (1985), McFadden (1987), and Wills (1987).

[2]The development of the nested logit model is due to Domencich and McFadden (1975), Williams (1977), McFadden (1978, 1981), and Daly and Zachary (1979).

[3]The GEV family was first examined in McFadden (1978) and Strauss (1979).

[4]Heterogeneity across alternatives can occur when different individuals have varying abilities to understand the choices which face them. See for example the discussion in Capon and Kuhn (1982).

the similarity coefficients are also assumed to be constant across individuals.[5] Neither of these restrictions is desirable. Moreover, when heterogeneity is present, maximum likelihood estimation of models within the GEV class is likely to be inconsistent.

The consequence of ignoring heterogeneity in the logit context has received some attention in the literature. Horowitz (1981) presents a numerical example which shows that the logit model is sensitive to the assumption of homoscedastic disturbances. Davidson and MacKinnon (1984) derive an LM test for heteroscedasticity in the binary logit model and present monte-carlo evidence on the properties of alternative test statistics. Dubin (1985), notes the heteroscedasticity problem in the context of a multinomial logit model for choice of residential space heating systems. Dubin considers simple normalizations of the explanatory variables to adjust for scale rather than the parametric approach given here. Hausman and Ruud (1987) develop extensions of the rank-ordered logit model in which the top ranked choices are selected more precisely by the individual than the lower ranked choices. Their model, termed the "heteroscedastic rank-ordered logit specification" introduces distinct variances in the lower and upper ranks. The scale factors (which depend only on the order of the choice set) differ "because unobserved characteristics play a greater role in selections among the poorest alternatives than among the best." Finally, Steckel and Vanhonacker (1988) present a model they term the "heterogeneous conditional logit model". The purpose of their study is to introduce an individual specific effect into the marginal extreme-value distribution. They assume that $F_t(\epsilon_j) = exp(-\tau_t \beta exp(-(\epsilon_j - \alpha)/\beta))$, where the individual specific effect, τ_t, is assumed to be gamma distributed. Under these assumptions, Steckel and Vanhonacker are able to find a closed-form expression for the choice probabilities which depends on the scale parameter β and the parameters which determine the distribution for τ_t. While their model avoids the IIA property of the multinomial logit model it does not, in fact, relax the homoscedasticity of the error terms.

The purpose of this paper is to provide a generalization of the generalized extreme value family which allows non-constancy of the similarity parameters in the nested logit model and which introduces non-homoscedasticity in the underlying random components of utility. While the heteroscedastic form of the logit model has appeared before in the literature, it has not been identified as a member of the GEV family.[6] Our development makes this connection and further extends the GEV family to include logit models with heteroscedasticity across alternatives.

In Section 2 we consider some approaches to introducing heterogeneity into logit models and present our own approach to this problem. In Section 3 we consider issues

[5]The nested logit model is the best known member of the GEV class. A second member is the "ordered generalized extreme value" model (Small (1987)). In each case, the degree of similarity among alternatives is assumed to be constant across individuals.

[6]In fact there appears to be some minor confusion on the theoretical underpinnings of the heteroscedastic logit model. Davidson and MacKinnon (1984, p. 247), for example, remark that "since this logit specification involves no latent variable, [it] cannot properly be called a specification of heteroscedasticity." As we demonstrate in Section 2, the heteroscedastic logit model is a member of the generalized GEV family and therefore is fully consistent with random utility maximization.

of specification and derive tests for the presence of heterogeneity in GEV models. In Section 4 we present an empirical example based on the choice of a consumer product both to the methods and to validate the potential significance of heterogeneity in discrete choice contexts. Section 5 presents our conclusions.

2 The Heterogeneous Logit Model

We begin with a population of T decision makers, each choosing among I discrete alternatives. Alternative i is assumed to provide utility u_{it} to individual t which consists of a deterministic component v_{it} and an unobserved random component ϵ_{it} with $u_{it} = v_{it} + \epsilon_{it}$. Under the assumption of random utility maximization (RUM), individual t chooses alternative i whenever $u_{it} \geq u_{jt}, \ \forall j \neq i$. The probability of this event is:

$$\begin{aligned} p_{it} &= \text{Prob } (v_{it} + \epsilon_{it} > v_{jt} + \epsilon_{jt}, \ \forall j \neq i) && (1) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{v_{it}+\epsilon_{it}-v_{1t}} \cdots \int_{-\infty}^{v_{it}+\epsilon_{it}-v_{It}} f(\epsilon_{1t}, \ldots, \epsilon_{It}) d\epsilon_{It} \ldots d\epsilon_{1t} d\epsilon_{it} && (2) \end{aligned}$$

where $f(\epsilon_{1t}, \ldots, \epsilon_{It})$ is the joint density of the ϵ_{it}. Equation (2) implies that

$$p_{it} = \int_{-\infty}^{\infty} F_i(v_{it} + \epsilon_{it} - v_{1t}, \ldots, \epsilon_{it}, \ldots, v_{it} + \epsilon_{it} - v_{It}) d\epsilon_{it} \tag{3}$$

where $F(\epsilon_{1t}, \ldots, \epsilon_{It})$ is the cumulative distribution of the ϵ_{it} and where F_i denotes the partial derivative of F with respect to its ith argument.

The generalized extreme value (GEV) probabilistic choice family is obtained by assuming that the distribution function F takes the form

$$F(\epsilon_{1t}, \ldots, \epsilon_{It}) = exp(-G(e^{-\epsilon_{1t}}, \ldots, e^{-\epsilon_{It}})) \tag{4}$$

with $G(Y_1, \ldots, Y_I)$ a nonnegative, homogeneous of degree one function on $\Re_I^+$ with the properties that $\lim_{Y_i \to \infty} G = +\infty$ and that the lth order cross-partial derivatives of G with respect to distinct Y_i's is nonnegative if l is odd and nonpositive if l is even.[7] Under these conditions, McFadden (1978) shows that equation (3) defines a multivariate extreme value distribution with

[7]Ben-Akiva and Lerman (1985) note that replacing the condition of linear homogeneity with μ homogeneity does not alter the result that $F(\cdot)$ is a proper cumulative distribution function. It does leave a free parameter in the probabilistic choice model which to date has not been exploited in empirical applications. Our results can be extended to G functions with μ homogeneity as well.

$$p_{it} = e^{v_{it}} G_i(e^{v_{1t}}, \ldots, e^{v_{It}}) / G(e^{v_{1t}}, \ldots, e^{v_{It}}). \tag{5}$$

The GEV class contains the multinomial and nested logit models as special cases. For example, when $G(\cdot) = \sum_{i=1}^{I} Y_i$, equation (5) is the multinomial logit model with $p_{it} = e^{v_{it}} / \sum e^{v_{jt}}$. The nested logit model groups the I alternatives into J subsets with each subset containing alternatives that are similar. For example, the two-level nested logit model results from $G = \sum_{k=1}^{J} (\sum_{i \in C_k} Y_i^{1/\sigma_k})^{\sigma_k}$, where $C_k \subset \{1, 2, \ldots, I\}$, $\cup_{k=1}^{J} C_k = \{1, 2, \ldots, I\}$, and $0 < \sigma_k \leq 1$. Here, σ_k measures the similarity of the random components of utility within subset C_k. For this G function, the probabilities in (5) can be written as the product of two probabilities each in standard logit form,

$$p_j = \frac{e^{v_j/\sigma_k}}{\sum_{i \in C_k} e^{v_i/\sigma_k}} \cdot \frac{e^{\sigma_k I_k}}{\sum_{k=1}^{J} e^{\sigma_k I_k}}, \quad j \in C_k$$

where $I_k = \log \sum_{i \epsilon C_k} e^{v_i/\sigma_k}$ defines the "inclusive value" of alternatives in subset C_k.

Models in the GEV class embody two important restrictions. First, the marginal distribution of each random disturbance has constant variance. To see this note that the marginal distribution for ϵ_{it} is given by

$$F^i(\epsilon_{it}) = F(\infty, \ldots, \epsilon_{it}, \ldots, \infty) = \exp(-G(0, \ldots, e^{-\epsilon_{it}}, \ldots, 0)) = \exp(-a_i e^{\epsilon_{it}}) \tag{6}$$

where $a_i = G(\delta_{i1}, \ldots, \delta_{iI})$ and where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.[8] But $F^i(\epsilon_{it})$, so defined, is the cumulative distribution function of an extreme value random variable with location parameter lna_i and variance $\pi^2/6$.[9] Therefore the variance of the marginal distributions is independent of the functional form of the GEV generating function. Second, the similarity coefficients (in the special case of nested logit models) are assumed to be constant across individuals, i.e., decision makers are assumed to have identical perceptions about the degree of similarity among alternatives.[10] Since these conditions are unduly restrictive we now present some simple alternatives.

[8]The last equality in equation (6) uses the linear homogeneity of the G function.

[9]The extreme-value distribution has cumulative distribution functions $F(\epsilon) = exp(-e^{-(\epsilon-\xi)/\theta})$ with $E(\epsilon) = \xi + \theta\gamma$ and $\text{var}(\epsilon) = (\pi^2/6)\theta^2$ where γ is Euler's constant.

[10]The problem of constant similarity of alternatives has been studied in the marketing literature and is reviewed in Huber and Lerman (1983). Our treatment of individual level heterogeneity in the nested logit model allows for "differential substitutability" which implies "the effect of a new item on the choice probabilities of other alternatives depends both on the ***similarity*** to other items as well as on its systematic ***utility***" (Huber and Lerman, p. 13).

2.1 Heteroscedastic Errors in the GEV Class

As we have just demonstrated, it is not possible to introduce heteroscedasticity in the marginal distributions for ϵ_{it} by choice of generating function G. Instead we introduce heteroscedasticity through a scale parameter which depends on the individual and the alternative. Following Dubin (1985, p. 221) observe that if $G(Y_1, \ldots, Y_I)$ is a proper GEV function, and if $\theta_{it} > 0$ then

$$\tilde{F}(\epsilon_{1t}, \ldots, \epsilon_{It}) = exp(-G(\langle e^{-\epsilon_{jt}\theta_{jt}}\rangle)) \tag{7}$$

is multivariate extreme-value distributed.[11] The marginal distributions corresponding to equation (7) are of the form

$$\tilde{F}^i = \tilde{F}(\infty, \ldots, \epsilon_{it}, \ldots, \infty) = exp(-G(0, \ldots, e^{-\epsilon_{it}\theta_{it}}, \ldots, 0)) = exp(-a_i e^{-\epsilon_{it}\theta_{it}}),$$

where we have again used the linear homogeneity of G. Now, however, $\tilde{F}^i$ is extreme-value distributed with variance $(\pi^2/6)(1/\theta_{it}^2)$. In principal, this can vary both by alternative and by individual. Moreover, as equation (7) nests the standard GEV distribution when θ_{jt} is constant, it is possible to derive simple tests of specification as we illustrate in Section 3.

Under RUM, the choice probabilities corresponding to (7) satisfy equation (2). But in this case, $\tilde{F}_i = \theta_{it} e^{-\epsilon_{it}\theta_{it}} G_i(\cdot) exp(-G(\langle e^{-\epsilon_{it}\theta_{it}}\rangle))$, implies

$$p_{it} = \int_{-\infty}^{\infty} \theta_{it} e^{-\epsilon\theta_{it}} G_i(e^{-(v_{it}+\epsilon-v_{1t})\theta_{1t}}, \ldots, e^{-\epsilon\theta_{it}}, \ldots, e^{-(v_{it}+\epsilon-v_{It})\theta_{It}}) e^{-G(\cdot)} d\epsilon. \tag{8}$$

When heteroscedasticity is present across individuals, but not alternatives, $\theta_{it} = \theta_t$ and equation (8) becomes

$$\begin{aligned} p_{it} &= \int_{-\infty}^{\infty} \theta_t e^{-\epsilon\theta_t} G_i(\langle e^{v_{jt}\theta_t}\rangle) exp\{-e^{-(v_{it}+\epsilon)\theta_t} G(\langle e^{v_{jt}\theta_t}\rangle)\} d\epsilon \\ &= \theta_t G_i(\cdot) \int_{-\infty}^{\infty} e^{-\epsilon\theta_t} exp\{-G(\cdot) e^{-v_{it}\theta_t} e^{-\epsilon\theta_t}\} d\epsilon \end{aligned}$$

[11] Observe that $\tilde{F}(\epsilon_1, \ldots, \epsilon_I) = F(\epsilon_1\theta_1, \ldots, \epsilon_I\theta_I)$ where $F(\cdot)$ is the cumulative distribution function given in equation (3). Therefore $\tilde{F}$ is also a proper cumulative distribution function. This may also be verified by directly checking the conditions given in McFadden (1978) or in Smith (1984). Dubin considered the distribution $\tilde{F}(\epsilon_1, \ldots, \epsilon_I) = F(\epsilon_1\theta, \ldots, \epsilon_I\theta)$ which introduces individual level (but not alternative specific) heterogeneity.

$$
\begin{aligned}
&= \theta_t G_i(\cdot) \int_{-\infty}^{\infty} \frac{1}{\theta_t e^{v_{it}\theta_t} G(\cdot)} d\left(exp\{-G(\cdot)e^{-v_{it}\theta_t} e^{-\epsilon\theta_t}\}\right) \\
&= e^{v_{it}\theta_t} G_i(\langle e^{v_{it}\theta_t}\rangle)/G(\langle e^{v_{it}\theta_t}\rangle). \qquad (9)
\end{aligned}
$$

In the special case, $G(\cdot) = \sum_{i=1}^{I} Y_i$, equation (9) is a multinomial logit model with heteroscedasticity across individuals; i.e.,

$$
p_{it} = e^{v_{it}\theta_t} / \sum_{j=1}^{I} e^{v_{jt}\theta_t}. \qquad (10)
$$

Davidson and MacKinnon's binary logit model with heteroscedasticity is a special case of (10) wherein $v_{1t} - v_{2t} = x_t'\beta$ and θ_t is parameterized by a function $e^{-z_t'\gamma}$. Dubin and McFadden (1984) employ choice probabilities given by equation (10) in their study of the demand for energy consuming durable goods. In their model $\ln(\theta_t)$ is linear in the price of electricity.[12] These examples, as well as the general equation (10), are therefore members of the GEV class with heteroscedasticity.

As equation (9) demonstrates, when heteroscedasticity is present across individuals, but not alternatives a relatively simple form for the choice probabilities exists.[13] It is also possible to simplify equation (8) by choosing simple forms for $G(\cdot)$. Suppose for example that $G = Y_1 + Y_2$ which, as we have remarked corresponds to the binary logit model specification. From equation (8) we have

$$
p_{1t} = \int_{-\infty}^{\infty} \theta_{1t} e^{-\epsilon\theta_{1t}} exp\{-(e^{-\epsilon\theta_{1t}} + e^{-(v_{1t}+\epsilon-v_{2t})\theta_{2t}})\} d\epsilon. \qquad (11)
$$

If we make a change in variables with $\mu = \epsilon\theta_{1t}$, then

$$
p_{1t} = \int_{-\infty}^{\infty} e^{-\mu} exp\{-e^{-\mu} - e^{-(v_{1t}+\mu/\theta_{1t}-v_{2t})\theta_{2t}}\} d\mu.
$$

Now let $\alpha = \theta_{2t}/\theta_{1t}$ and let $K = e^{-\theta_{2t}(v_{1t}-v_{2t})}$. Then

$$
p_{1t} = \int_{-\infty}^{\infty} e^{-\mu} exp\{-e^{-\mu} - e^{-\alpha\mu} \cdot K\} d\mu. \qquad (12)
$$

[12]In Dubin and McFadden, the scale factor θ_t is required as an integrating factor to solve the partial differential equation which relates indirect utility and linear demand via Roy's identity.

[13]While the general form of equation (8) is somewhat complicated, it does have an integrand which exists in closed-form for commonly specified G functions, and therefore should be amenable to numerical integration. The system may therefore prove to be as flexible yet computationally more attractive than the covariance probit model.

To simplify further we make a change in variables with $t = e^{-u}$. Then $dt = -e^{-u}du$ and

$$p_{1t} = \int_{\infty}^{0} -exp(-t - t^{\alpha}K)dt = \int_{0}^{\infty} e^{-t}e^{-Kt^{\alpha}}dt. \tag{13}$$

But (13) is the Laplace transform of $e^{-Kt^{\alpha}}$

$$F(s) = \int_{0}^{\infty} e^{-st}e^{-Kt^{\alpha}}dt = L\{e^{-Kt^{\alpha}}\} \tag{14}$$

evaluated at $s = 1$. Using the series expansion for e^{x} we have

$$\begin{aligned} p_{1t} &= L\{e^{-Kt^{\alpha}}\} \\ &= L\{\sum_{n=0}^{\infty}(-Kt^{\alpha})^{n}/n!\} \\ &= \sum_{n=0}^{\infty} L\{t^{\alpha n}\}(-1)^{n}K^{n}/n!. \end{aligned} \tag{15}$$

The Laplace transform (evaluated at $s = 1$) of $t^{\alpha n}$, is $\Gamma(\alpha n + 1)$, where $\Gamma(x)$ is the standard gamma function.[14] Therefore

$$\begin{aligned} p_{1t} &= \sum_{n=0}^{\infty}(-1)^{n}K^{n}/n! \cdot \Gamma(\alpha n + 1) \\ &= \sum_{n=0}^{\infty}(-1)^{n}K^{n}\Gamma(\alpha n + 1)/\Gamma(n + 1), \end{aligned} \tag{16}$$

where convergence holds for $K < 1$.[15] For the case $K > 1$, we have

$$\begin{aligned} p_{1t} &= \int_{0}^{\infty} \sum_{n=0}^{\infty}(-t)^{n}/n! \cdot e^{-Kt^{\alpha}}dt \\ &= \sum_{i=0}^{\infty}(-1)^{n}/n! \cdot \int_{0}^{\infty} t^{n}e^{-Kt^{\alpha}}dt \end{aligned}$$

[14]See "Handbook of Mathematical Functions", Milton Abramowitz and Irene Stegun (eds.), National Bureau of Standards, Applied Mathematics Series 55, 1964. (29.3.7).

[15]The condition $K < 1$ occurs when the strict utilities obey $v_{1t} > v_{2t}$.

$$= \sum_{i=0}^{\infty}(-1)^n/n! \cdot \frac{\Gamma((n+1)/\alpha)}{\alpha K^{(n+1)/\alpha}}$$

$$= \frac{1}{\alpha\psi}\sum_{n=0}^{\infty}(\frac{-1}{\psi})^n\frac{\Gamma((n+1)/\alpha)}{\Gamma(n+1)} \tag{17}$$

where $\psi = K^{\frac{1}{\alpha}}$.[16]

When $\theta_{1t} = \theta_{2t}$, $p_{1t} = \sum_{n=0}^{\infty}(-1)^n K^n = e^{v_{1t}\theta_{1t}}/(e^{v_{1t}\theta_{1t}} + e^{v_{2t}\theta_{2t}})$ is a binary logit model with heteroscedasticity across individuals. When θ_{1t} and θ_{2t} differ, the binary logit model exhibits heteroscedasticity across individuals and alternatives. Maximum likelihood estimation of the parameters in this system via gradient methods requires the derivative of the gamma function. This derivative is given by the digamma function and can be computed using a series expansion (Bernardo, 1976).

2.2 Heterogeneous Nested Logit Models

We now relax the constant similarity coefficient restriction in nested logit models. Our approach is to parameterize the similarity coefficients as functions of the underlying explanatory variables much as is often done for heteroscedasticity in the linear model. In the case of the two-level nested logit model we employ a variant of the usual G function in which the similarity parameters σ_k are individual specific:

$$G(Y_1, \ldots, Y_I) = \sum_{k=1}^{J}(\sum_{i\epsilon I_k} Y_i^{1/\sigma_{kt}})^{\sigma_{kt}}. \tag{18}$$

Here it is useful to impose $0 < \sigma_{kt} \leq 1$ so that G has the necessary properties for $F(\cdot)$ to be a proper cumulative distribution function.[17] The choice probabilities are given by

$$p_j = \frac{e^{v_j/\sigma_{kt}}}{\sum_{i\in C_k} e^{v_i/\sigma_{kt}}} \cdot \frac{e^{\sigma_{kt}I_k}}{\sum_{k=1}^{J} e^{\sigma_{kt}I_k}}$$

where $I_k = \log \sum_{i\in C_k} e^{v_i/\sigma_{kt}}$.

Two-stage estimation of this model is complicated by the presence of the scaling factors σ_{kt}. In the first stage, the conditional choice probabilities (choice of alternative

[16]See Standard Mathematical Tables, Nineteenth Edition, Samuel M. Selby (ed.), The Chemical Rubber Co., Cleveland, Ohio. Formula 662, Page 449.

[17]Börsch-Supan (1990) shows that the condition $0 < \sigma_{kt} \leq 1$ is in fact sufficient, but not necessary to guarantee that $F(\cdot)$ is a proper cumulative distribution function.

j from subset C_k) can be estimated using the heteroscedastic multinomial logit model. Then the inclusive values I_k can be calculated and used for the calibration of the marginal choice probabilities. Sequential estimation, in this context, is probably less useful than full information methods since it may be difficult to insure that the scale effects σ_{kt} are identical at each stage.

3 Specification and Tests of Specification

Implementation of either the heteroscedastic GEV model or the heterogeneous nested logit model requires explicit parameterization of θ_{it} and σ_{kt}. For the heteroscedastic GEV model we require that θ_{it} be positive and therefore specify θ_{it} as $(1 + \sigma_1' z_i + \sigma_2' z_t + \sigma_3' z_{it})^2$ or as $e^{(\sigma_1' z_i + \sigma_2' z_t + \sigma_3' z_{it})}$.[18] Each of these forms nests the homoscedastic case when σ_i is identically zero. For the heterogenous nested logit model we impose the additional restriction $\sigma_{kt} \leq 1$. In this regard, any cumulative distribution function will provide an acceptable parameterization for σ_{kt}. For example a logistic distribution could be used with $\sigma_{kt} = 1/(1 + e^{-\sigma_k' z_t})$. Estimation can then be accomplished using maximum likelihood methods.[19]

As an example we demonstrate how the scale factors can be specified to depend on the size of the choice set. The example is motivated by empirical applications in which individuals face choice sets of varying size. We adapt the specification of Kiefer and Skoog (1984) and Yatchew and Griliches (1985) and let $\theta_t = 1 + \alpha_j d_{tj}$ where $d_{tj} = 1$ if $j = J_t$ and zero otherwise. In this case, the variance of random utility for individual t is inversely proportional to α_j where j indexes the number of alternatives, J_t, in the choice set. If larger choice sets make it more likely that individuals make choices less carefully than when they face fewer choices then we would expect α_j to decline as j increases.

Following Davidson and MacKinnon (1984) it is also possible to derive tests for the heterogenous forms of the nested logit and GEV models against the homoscedastic forms using the LM statistic. As an example we consider a test for heteroscedasticity within the multinomial logit model. We adopt a linear-in-parameters specification for strict utility, $v_{it} = \beta' x_{it}$ and choose an exponential parameterization for the scale effect, $\theta_t = e^{\alpha' z_t}$.[20] The log-likelihood for the sample is $L = \sum_{t=1}^{T} \sum_{i=1}^{I} c_{it} ln p_{it}$ where $c_{it} = 1$ if alternative i is chosen and 0 otherwise, and where $p_{it} = e^{v_{it}\theta_t} / \sum_{j=1}^{I} e^{v_{jt}\theta_t}$. To construct the LM statistic we require the derivatives:

[18]Similar specifications for heteroscedasticity appear in Jarque (1981), Bera, Jarque, and Lee (1984), and Lee and Maddala (1985).

[19]The regularity conditions for consistency and asymptotic normality of maximum likelihood estimates for models in the GEV class are given in McFadden (1984). Our extensions continue to satisfy regularity since the underlying probabilities are continuous and differentiable in the scale parameters.

[20]The Lagrange multiplier statistic is invariant to the parameterization of θ_t provided that $\partial\theta_t/\partial\alpha = K \cdot z_t$ where K is a scalar under the null hypothesis. Our specification of θ_t has this property but alternative specifications could serve equally well. For a discussion of such locally equivalent alternatives see Engle (1984) and Lee and Maddala (1985).

$$
\begin{aligned}
L_\beta &= \sum_t \sum_i c_{it}\theta_t(x_{it} - x_t) \quad \text{and} \\
L_\alpha &= \sum_t \sum_i c_{it}\theta_t z_t \beta'(x_{it} - x_t)
\end{aligned}
$$

and elements of the information matrix,

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

where $x_t = \sum_i p_{it} x_{it}$.

Under the null hypothesis $\alpha = 0$ and $\theta_t = 1$. Therefore we can apply standard multinomial logit techniques to obtain estimates of the parameters β and form the LM statistic using $L'_\alpha[V_{\alpha\alpha} - V_{\alpha\beta}V_{\beta\beta}^{-1}V_{\beta\alpha}]^{-1}L_\alpha$ (Breush and Pagan, 1980). This statistic has a limiting chi-squared distribution with degrees of freedom equal to the number of elements in the vector α.

4 An Application of the Heterogeneous Logit Model to Consumer Camera Purchases

The empirical example we present is based on the demand for cameras. We examine the selections made from among three broad types of cameras: 35mm, instant, and other still cameras. The data are taken from surveys of households conducted in 1982 and 1983 by National Family Opinion, Inc. (NFO). We select the sample for analysis based on three criteria (1) the household is interviewed in both 1982 and 1983, (2) the household buys some type of camera in 1983, and (3) complete data is available for all household characteristics in the 1982 base year. The first criteria allows us to characterize the choices made in 1983 as functions of the stock of cameras (and their types) previously selected by the household. The second criteria conditions our analysis on the decision to buy some type of camera. Including the alternative of not buying a camera in a given year among the alternatives which represent types of camera purchased would be a clear violation of IIA. The third criteria is of course required to undertake the estimation. The estimation sample consists of 875 households.

Our model of camera selection specifies the utility of a given camera type as a function of four socio-economic characteristics: education, income, presence of children, and marital status and a variable which measures the stock of cameras held by households in the base year by type of camera. The explanatory variables are defined as follows: EDUC, the number of years of education for the head of household; INC, household income in thousands of dollars, HKIDS, one if the household has children, MARITAL, one if the head of household is married, and HELD, the number of cameras of each type (35mm, instant, or other still) present in the household in the base year. Variable definitions and summary statistics are given in Table 1. In this sample, roughly equal numbers of households choose the 35mm and other still formats (41 and 43 percent respectively) while the remainder choose instant photography (16 percent). Nearly 90 percent of the households had some camera in use in the base year.

The results of the estimation for the standard multinomial logit model are presented in Table 2.[21] The alternative specific constant terms show significant preference for the other still format as compared with either 35mm or instant type cameras. Households with higher income are more likely to choose 35mm cameras over other still and instant formats. This is consistant with the fact that 35mm cameras were relatively more expensive than other types of cameras in the early 1980's. Households with children are significantly less likely to purchase instant cameras and instead prefer other still formats. Higher education increases the probability of purchasing 35mm cameras. This is constant with the commonly held view that 35mm cameras were more complicated to use than other types of cameras.[22] Households in which the head is married do not reveal any significant preference among the camera types. Finally, the presence of cameras of a given type in the base year significantly increases the probability of repurchasing that type in the year of camera acquisition. Households therefore reveal significant "type" loyalty.

To test whether heteroscedasticity is present in the standard logit model we employ the specification test derived in Section 3. We considered several specifications for possible scale effects and found that the size of the family (SIZE) is a significant determinant of the variance of random utility. The value of the LM statistic, in this case, is 5.92 (significant at the five percent level). We therefore estimate the heteroscedastic logit model using the specification $\log(\theta_t) = \alpha \cdot \text{SIZE}_t$.[23] The results of this estimation are also presented in Table 2.

Comparing the standard and heteroscedastic logit estimates we observe relative stability in the estimated coefficients and only modest changes in the levels of statistical significance. The most prominently affected coefficient is that of HKIDS. This might

[21] Without loss of generality, we have used the "other still" alternative as the basis for normalization.

[22] Recall that in 1982 few, if any, 35mm cameras provided automatic focusing, automatic shutter control, or self-winding film.

[23] The heteroscedastic logit model was estimated in just under one hour while the standard logit model was estimated in approximately 13 seconds. Both estimations were completed on a SUN 4/280 using Statistical Software Tools (Dubin/Rivers Research).

have been expected since the scale effect is modeled in terms of family size and family size increases when children are present.[24] The implication of this finding is that larger families have a greater diversity of camera needs arising from their larger and potentially more diverse constituency.

5 Conclusion

In the linear model the detection and subsequent correction of heteroscedastic disturbances is motivated by issues of efficiency. But for models in the GEV family the detection and correction of heterogeneity is fundamental to achieving even consistent parameter estimates. And while in the linear model it is possible to consider the issues of consistency and efficiency separately, this is not possible in the maximum likelihood setting. As we have seen, the specification of the form of heterogeneity becomes an integral part of the model to be estimated. In our development of the heterogeneous logit model we have relaxed two fundamental restrictions of the GEV family. Our approach, which introduces non-constant scale parameters into the functional forms for the standard and nested logit models, leads to new probabilistic choice systems.

The new models have substantial intuitive appeal. In the heteroscedastic GEV model, for example, as the variance of random utility increases relative to the difference in strict utilities, standardized utility differences are attenuated toward zero. The model then gives more equal weight to each probability. On the other hand, observations which have less disperse random utility components tend to accentuate observed utility differences which leads to more pronounced differences in the predicted choice probabilities. Our extensions of the nested logit model have a similar character. When the random components of utility are large relative to the observed components, the degree of similarity of alternatives diminishes.

Application of the heterogeneous logit model in empirical settings will in fact reveal its usefulness. Our study of the demand for camera types suggests that certain forms of population heterogeneity can extend beyond the factors one normally uses to characterize alternatives.

[24]Similar results were obtained in an alternative specification where the explanatory variable HKIDS was replaced by SIZE. The value of the LM statistic in this case was 5.58.

Table 1
Variable Definitions and Summary Statistics

Variable	Description	Mean	Stdev
EDUC	head of household's education in years	13.59	3.49
INC	household's income in thousands	21.54	9.39
HKIDS	one if household has children	0.58	0.49
MARITAL	one if head of household is married	0.69	0.46
HELD 35mm	number of 35mm cameras in use in base year	0.47	0.81
HELD instant	number of instant cameras in use in base year	0.24	0.49
HELD other	number of other still cameras in use in base year	0.80	0.76
SIZE	number of people in household	2.87	1.49

Table 2
Standard and Heteroscedastic Multinomial Logit
Models for Camera Type Choice
(asymptotic t-values in parenthesis)

Explanatory Variable	Model	
	Standard	*Heteroscedastic*
35mm Alternative		
CONSTANT	-1.53	-2.44
	(-4.26)	(-3.56)
INC	0.044	0.063
	(4.54)	(3.58)
HKIDS	-0.275	-0.345
	(-1.68)	(-1.38)
EDUC	0.066	0.111
	(2.67)	(2.67)
MARITAL	-0.22	-0.27
	(-1.18)	(-1.02)
Instant Alternative		
CONSTANT	-0.88	-1.10
	(-2.10)	(-1.79)
INC	0.014	0.021
	(1.15)	(1.12)
HKIDS	-0.462	-1.04
	(-2.24)	(-2.16)
EDUC	0.017	0.029
	(0.57)	(0.65)
MARITAL	-0.31	-0.66
	(-1.37)	(-1.76)
Alternative Specific Variables		
HELD	0.201	0.333
	(3.32)	(2.94)
Heteroscedasticity Factors		
SIZE	-	-0.151
		(-2.05)
Log-likelihood at convergence	-860.9	-858.2

References

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

Bera, A., Jarque, M. and Lee, L. (1984): "Testing the Normality Assumption in Limited Dependent Variable Models". *International Economic Review* 25, 536-578.

Bernardo, J. (1976): "Psi (Digamma) Function", *Applied Statistics* 25, 315-317.

Börsch-Supan, A. (1990): "On the Compatibility of Nested Logit Models With Utility Maximization". *Journal of Econometrics* 43, 373-388.

Breusch, T. and Pagan, A. (1990): "The Lagrange Multiplier Test And Its Applications To Model Specification in Econometrics". *Review Of Economic Studies* 47, 239-254.

Capon, N. and Kuhn, D. (1982): "Can Consumers Calculate Best Buys?", *Journal of Consumer Research* 8, 449-453.

Daly A. and Zachary, S. (1979): "Improved Multiple Choice Models", D. Hensher and Q. Dalvi, eds., *Identifying and Measuring the Determinants of Mode Choice*, London: Teakfield 335-357.

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

Domencich, T. and McFadden, D. (1975): *Urban Travel Demand: A Behavioral Analysis*, Amsterdam: North-Holland.

Dubin, J. and McFadden, D. (1984): "An Econometric Analysis of Residential Electric Appliance Holdings And Consumption". *Econometrica* 52, 345-362.

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

Engle, R. (1984): "Wald, LR, and LM Tests in Econometrics", in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics* Volume II, Amsterdam: North-Holland.

Hausman, J. and Ruud, P. (1987): "Specifying and Testing Econometric Models for Rank-Ordered Data", *Journal of Econometrics* 34, 83-104.

Horowitz, J. (1981): "Identification and Diagnosis Of Specification Errors in the Multinomial Logit Model". *Transportation Research B* 15B, 345-360.

Hausman, J. and Wise, D. (1978): "A Conditional Probit Model For Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences". *Econometrica* 46, 403-426.

Hausman, J. and McFadden, D. (1984): "Specification Tests For The Multinomial Logit Model". *Econometrica* 52, 1219-1240.

Huber, J. and Lerman, S.(1983): "Adjusting Choice Models for Similarity Effects", mimeo, Massachusetts Institute of Technology.

Jarque, C. (1981): "A Test for Heteroscedasticity in a Limited Dependent Variable Model". *Australian Journal of Statistics* 23, 159-163.

Kiefer, N. and Skoog, G. (1984): "Local Asymptotic Specification Error Analysis". *Econometrics* 52, 837-875.

Lee, L. and Maddala, G. (1985): "The Common Structure of Tests for Selectivity Bias, Serial Correlation, Heteroscedasticity and Nonnormality in the Tobit Model". *International Economic Review* 26, 1-20.

McFadden, D., Train, K. and Tye, W. (1976): " An Application of Diagnostic Tests For The Independence From Irrelevant Alternatives Property Of The Multinomial Logit Model". *Transportation Research Record* 637, 39-45.

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

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

McFadden, D. (1984): "Econometric Analysis of Qualitative Respond Models". in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics, Volume II.* Amsterdam: North-Holland.

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

Small, K. (1987): "A Discrete Choice Model for Ordered Alternatives". *Econometrica* 55, 407-424.

Small, K. and Hsiao, C. (1985): "Multinomial Logit Specification Tests". *International Economic Review* 26, 619-627.

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

Steckel, J. and Vanhonacker, W. (1988): "A Heterogeneous Conditional Logit Model Of Choice". *Journal Of Business and Economic Statistics* 6, 391-398.

Strauss, D. (1979): "Some Results on Random Utility Models". *Journal of Mathematical Psychology* 20, 35-52.

Williams, H. (1977): "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit", *Environment Planning* A9, 285-344.

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

Yatchew, A. and Griliches, Z. (1985): "Specification Error in Probit Models". *Review of Economics and Statistics* 18, 134-139.