APPENDIX B

Conditional moments in the generalized extreme value family[1]

B.1. Introduction

The discrete/continuous econometric systems derived in Chapter 1 have the schematic form:

$$x = Z\beta^i + \eta \tag{1}$$

where Z is a vector of attributes of the alternatives and of the decision-maker, β^i is a vector of parameters that depend on the portfolio of appliances i, and η is an unobserved random variable whose density $f(\eta | i)$ depends in general on the observed choice i. The choice probabilities are functions of Z, the parameters β^i, and other parameters α:

$$P_i = \text{Prob(Portfolio } i \text{ is chosen)} = G^i(Z, \beta^1, \beta^2, ..., \beta^m, \alpha) \tag{2}$$

The likelihood of an observation (Z, i, x) is then:

$$G^i(Z, \beta^1, \beta^2, ..., \beta^m, \alpha) \cdot f(x - Z\beta^i | i) \tag{3}$$

Under standard regularity conditions, full information maximum likelihood (FIML) estimation of the parameters $\beta^1, ..., \beta^m, \alpha$ will yield consistent asymptotically efficient estimates, while maximum likelihood estimation of the discrete-choice model alone will yield consistent, but not usually efficient, estimates of the identifiable parameters.

[1]In the course of the exposition, several theorems related to the independent form of the generalized extreme value family, i.e., the multinomial logit model, are derived. Specifically, Corollary B.2.2 and Theorems B.3.3, B.4.1, B.4.2 and, B.4.3, which present the conditional moments for the multinomial logit model, are stated in Dubin and McFadden (1984). It should be further noted that Theorems B.3.3, B.4.1, and B.4.2 have been independently demonstrated by Hay (1980).

The system (1)-(3) is a variant of the "hybrid model with structural shift" analyzed in detail by Heckman (1978), and the estimators and properties he develops can be applied with straightforward modification. This system can also be interpreted as a "switching regression" with the structure analyzed by Lee (1981), Goldfeld and Quandt (1973, 1976), and Maddala and Nelson (1974). Note that (1) may be written:

$$x = Z\beta^i + E(\eta|i) + \nu \tag{4}$$

$$= \sum_{j=1}^{m} Z\, \delta_{ij}\, \beta^j + E(\eta|i) + \nu \tag{5}$$

$$= \sum_{j=1}^{m} Z\, P_j\, \beta^j + \xi \tag{6}$$

where:

$$\delta_{ij} = 1 \ \text{ iff } \ i=j\,, \quad E(\eta|i) = \int_{-\infty}^{+\infty} f(\eta|i)\, \eta\, d\eta$$

$$\nu = \eta - E(\eta|i)\,, \quad \xi = \eta + Z \sum_{j=1}^{m} (\delta_{ij} - P_j)\, \beta^j$$

Then $E(\eta|i) = 0$ and $E(\xi' Z P_j) = 0$ for $j = 1, 2, ..., m$. The choice probabilities P_i given by (2) and the conditional expectation $E(\eta|i)$ are nonlinear functions of the parameters of the problem. Under specific distributional assumptions, these functions may have computationally tractable forms. For example, if the discrete choice is binary and determined by a latent variable whose joint distribution with η is bivariate normal, then P_1 is a probit function and $E(\eta|i)$ is proportional to a Mill's ratio evaluated at the mean of the latent variables; see Heckman (1978).

Alternatively, suppose the utilities of different portfolios have independent extreme value distributions with a common variance. The choice probabilities are then multinomial logit; see McFadden (1973). The conditional expectation $E(\eta|i)$ is a simple function of the choice probabilities and their logs; see Dubin and McFadden (1984).

One method of estimating the parameters of (1) that is consistent under standard regularity conditions is to apply nonlinear least squares to (4) or (6). A second method is to replace $E(\eta | i)$ in (4) with a consistent estimate obtained by first estimating the parameters of the choice probabilities. Heckman (1979) shows that this procedure is also consistent under standard regularity conditions, although the asymptotic covariance matrices of the least squares coefficients obtained by this method differ from the standard formula due to the presence of estimated explanatory variables.[2]

The purpose of this appendix is to establish basic results on the conditional moments of generalized extreme value (GEV) random variables. This provides a useful generalization of the Heckman probit and the Dubin-McFadden logit selectivity corrections by including a class of dependent multinomial probability models. We introduce the GEV distribution and discuss its properties. We then derive the first, second, and cross-conditional moments for GEV variables given that a specific alternative has been selected. Finally, we allow the random variable η in equation (1) to have a linear conditional expectation in the space of GEV random variables and derive its properties. These results provide the distributional framework for consistent two-step estimation techniques.

B.2. Conditional moments in GEV

The following theorem due to McFadden (1978) introduces a general family of GEV choice models.

Theorem B.2.1. [McFadden]. Suppose $G(y_1, y_2, ..., y_J)$ is a nonnegative, homogeneous of degree one function of $(y_1, y_2, ..., y_J) \geqslant 0$. Suppose $\lim_{y_i \to +\infty} G(y_1, y_2, ..., y_J) = +\infty$ for $i = 1, 2, ..., J$. Suppose for any distinct $(i_1, i_2, ..., i_k)$ from $\{1, 2, ..., J\}$, $\partial^k G / \partial y_{i_1}, ..., \partial y_{i_k}$ is nonnegative if k is odd and nonpositive if k is even. Then:

$$P_i = e^{V_i} G_i(e^{V_1}, ..., e^{V_J}) / G(e^{V_1}, ..., e^{V_J}) \tag{7}$$

[2]A third estimation technique applies the method of instrumental variables to (1) using consistent estimates of the choice probabilities as instruments. Consistency and efficiency of the various procedures are considered in Chapter 5.

defines a choice model that is consistent with random utility maximization.

Proof. Theorem B.2.1 is proved in two steps. The first step demonstrates that the function:

$$F(\varepsilon_1, \varepsilon_2, ..., \varepsilon_J) = \exp[-G(e^{-\varepsilon_1}, e^{-\varepsilon_2}, ..., e^{-\varepsilon_J})] \tag{8}$$

is a multivariate extreme value distribution. The details may be found in McFadden (1978).

The second step assumes a population of individuals with utilities $u_i = V_i + \varepsilon_i$, where $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_J)$ is distributed F. Let ε denote the vector $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_J)$. Then:

$$P_i = \mathrm{Prob}[u_i \geqslant u_j, \forall i \mid i \neq j]$$

$$= \mathrm{Prob}[V_i + \varepsilon_i \geqslant V_j + \varepsilon_j, \forall i \mid i \neq j] \tag{9}$$

may be written:

$$\int_{\varepsilon_i=-\infty}^{+\infty} F_i(<V_i + \varepsilon_i - V_j>)\, d\varepsilon_i \tag{10}$$

where F_i denotes the derivative of F with respect to its ith argument, and $<a_j>$ denotes a vector with jth component a_j. From (10) and the definition of the GEV distribution, we have:

$$P_i = \int_{-\infty}^{+\infty} \exp\left[-G[<e^{-(V_i+\varepsilon_i-V_j)}>]\right] G_i[<e^{-(V_i+\varepsilon_i-V_j)}>]\, e^{-\varepsilon_i}\, d\varepsilon_i$$

$$= \int_{-\infty}^{+\infty} e^{-\varepsilon_i}\, G_i[<e^{V_j}>] \exp\left[-G[<e^{V_j}>] \cdot e^{-V_i}\, e^{-\varepsilon_i}\right] d\varepsilon_i$$

$$= \frac{G_i[<e^{V_j}>]}{G[<e^{V_j}>]}\, e^{V_i} \quad \text{Q.E.D.} \tag{11}$$

The second equality in equation (11) uses the fact that G is homogeneous of degree one and the implication that G_i is homogeneous of degree zero. The third equality makes use of the result:

$$\int_{-\infty}^{+\infty} e^{-\varepsilon} \exp[-\theta^{-1} e^{-\varepsilon}]\, d\varepsilon = \theta^{-1} \tag{12}$$

which follows from the substitution $u \rightarrow -\theta e^{-\varepsilon}$.

Corollary B.2.1. [Multinomial Logit Model]. Let:

$$G[y] = [\sum_{j=1}^{J} y_j^{1/\theta}]^{\theta}. \quad \text{Then:} \quad P_i = \frac{e^{V_i/\theta}}{\sum_{j=1}^{J} e^{V_j/\theta}}$$

Proof. This result is found in McFadden (1978). One need simply verify the linear homogeneity of G and apply equation (1). Q.E.D.

McFadden shows that if $\varepsilon_j \rightarrow +\infty\ \forall j \mid j \neq i$, then equation (2) implies, $F[\varepsilon_i] = \exp[-a_i\, e^{-\varepsilon}]$, where $a_i = G[0, ..., 0, 1, 0, ..., 0]$; with one in the ith coordinate. Under the assumptions of Corollary B.2.1, the marginal distribution, $F[\varepsilon_i]$, is $\exp[-e^{-\varepsilon_i}]$ (note $a_i = 1$); the cumulative distribution for an extreme value distributed random variable with variance $\pi^2/6$.

McFadden's proof of Theorem B.2.1 may be modified to demonstrate that:

$$F[\varepsilon_1, \varepsilon_2, ..., \varepsilon_J] = \exp\left[-G[<e^{-\varepsilon_j/\theta}>]\right] \tag{13}$$

is a multivariate extreme value distribution. Application of (4) implies the probabilistic choice system:

$$P_i = e^{V_i/\theta}\, G_i[<e^{V_j/\theta}>] \,/\, G[<e^{V_j/\theta}>] \tag{14}$$

In this case, the marginal distribution for ε_i becomes $\exp[-a_i e^{-\varepsilon_i/\theta}]$, which is the cumulative distribution function for an extreme value distributed random variable with variance $(\pi^2/6)\,\theta^2$.[3] When:

$$G[<y_j>] = \sum_{j=1}^{J} y_j$$

equation (14) implies choice probabilities of the multinomial logit form. Furthermore, ε_i has mean $\gamma\theta$ and variance $(1/6)\,\pi^2\theta^2$. More generally, ε_i will have mean u and variance $(1/6)\,\pi^2\theta^2$ when:

$$G[y_1, y_2, \ldots, y_J] = \frac{\exp(u/\theta)}{\exp(\gamma)} \cdot \left[\sum_{j=1}^{J} y_j\right]$$

Let $\delta_j(\varepsilon)$ be an indicator random variable that is 1 when j is the chosen alternative, i.e., when $V_j + \varepsilon_j \geqslant V_i + \varepsilon_i$, $\forall i \mid i \neq j$, and 0 otherwise. We write δ_j as a function of ε to emphasize that it is a random variable whose outcome, conditioned on the V_j 's, depends on the realization of ε. Lemma B.2.1 presents the conditional moments.[4]

Lemma B.2.1. Let ε be GEV distributed with cumulative distribution function $F(\varepsilon)$ given by equation (13). Let $g(\cdot)$ be an arbitrary real-valued function. Then:

(a) $E(g(\varepsilon_1) \mid \delta_1(\varepsilon) = 1) = E(g(\varepsilon) \mid \varepsilon \sim EV[\theta\,(\ln G_1 - \ln P_1), \theta])$

where $EV[a,b]$ denotes an extreme-valued distributed random variable with location parameter a and scale parameter b.

(b) Let G be additively separable with $G(y) = G^A(y^A) + y_2$ where $y = (y^A, y_2)$ and where $G^A(\cdot)$ is homogeneous of degree one. Let ε have the corresponding partition, i.e., $\varepsilon = (\varepsilon^A, \varepsilon_2)$. Then:

[3] The random variable ε_i has the properties: $E[\varepsilon_i] = \theta(\gamma + \ln a_i)$ and $Var[\varepsilon_i] = (1/6)\,\pi^2\theta^2$, where $\gamma = .5772156649 \cdots$ is Euler's constant.

[4] Note that without loss of generality, it suffices to consider the expressions $E(\varepsilon_1 \mid \delta_1 = 1)$ and $E(\varepsilon_2 \mid \delta_1 = 1)$ rather than the more general expression $E(\varepsilon_i \mid \delta_j = 1)$ for $i = j$ and for $i \neq j$.

$$E(\varepsilon_2 | \delta_1(\varepsilon) = 1)$$

$$= \frac{G[<e^{V_j/\theta}>]}{G^A[<e^{V_j/\theta}>]} \left[E \left[g(\varepsilon_2) | \varepsilon_2 \sim EV[0,\theta] \right] \right.$$

$$\left. - P_2 \, E \left[g(\varepsilon_2) | \varepsilon_2 \sim EV[-\theta \, (\ln P_2), \theta] \right] \right]$$

Proof.

(a) We make use of the properties of conditional expectations. Recall:

$$\int_{-\infty}^{y} \int_{x \in A} f(x, y) \, dx \, dy$$

$$= \mathrm{Prob}[x \in A, \, Y \leqslant y]$$

$$= \mathrm{Prob}[Y \leqslant y | x \in A] \, \mathrm{Prob}[x \in A] \tag{15}$$

Thus:

$$\mathrm{Prob}[x \in A]^{-1} \int_{x \in A} f(x, y) dx = f(y | x \in A) \tag{16}$$

Equation (16) implies:

$$E(Y | x \in A) = \int_y y \, f(y | x \in A) \, dy$$

$$= \mathrm{Prob}[x \in A]^{-1} \int_y \int_{x \in A} y \, f(x, y) \, dx \, dy \tag{17}$$

As an application of (11) we find:

$$E(g(\varepsilon_1)|\delta_1(\varepsilon) = 1)$$

$$= \frac{1}{P_1} \int_{\varepsilon_1=-\infty}^{\infty} \int_{\varepsilon_2=-\infty}^{V_1-V_2+\varepsilon_1} \cdots \int_{\varepsilon_J=-\infty}^{V_1-V_J+\varepsilon_1} g(\varepsilon_1)\, dF(\varepsilon)$$

$$= \frac{1}{P_1} \int_{\varepsilon=-\infty}^{\infty} g(\varepsilon)\, F_1[<\varepsilon + V_1 - V_j>]\, d\varepsilon$$

$$= \frac{1}{P_1} \int_{\varepsilon=-\infty}^{\infty} g(\varepsilon)\, e^{-\varepsilon/\theta}\, G_1\left[< e^{-(\varepsilon+V_1-V_j)/\theta} > \right] \cdot$$

$$\exp\left[-G[< e^{-(\varepsilon+V_1-V_j)/\theta} >] \right] \frac{d\varepsilon}{\theta}$$

$$= \frac{1}{P_1} \int_{\varepsilon=-\infty}^{\infty} g(\varepsilon)\, e^{-\varepsilon/\theta}\, G_1\left[<e^{V_j/\theta}> \right] \cdot$$

$$\exp\left\{ -G[<e^{V_j/\theta}>]\, e^{-\varepsilon/\theta}\, e^{-V_1/\theta} \right\} \frac{d\varepsilon}{\theta}$$

Let $\theta_1 = G[<e^{V_j/\theta}>]\, e^{-V_1/\theta}$ and $\theta_2 = G_1[<e^{V_j/\theta}>]$. Then:

$$E(g(\varepsilon_1)|\delta_1(\varepsilon) = 1)$$

$$= \frac{\theta_2}{P_1} \cdot \int_{-\infty}^{\infty} g(\varepsilon)\, e^{-\varepsilon/\theta} \exp[-\theta_1\, e^{-\varepsilon/\theta}]\, \frac{d\varepsilon}{\theta}$$

$$= \frac{\theta_2}{P_1\, \theta_1} \cdot \int_{-\infty}^{\infty} g(\varepsilon)\, e^{-(\varepsilon-\theta\, k_1)/\theta} \exp[-e^{-(\varepsilon-\theta\, k_1)/\theta}]\, \frac{d\varepsilon}{\theta}$$

$$= E(g(\varepsilon)|\varepsilon \sim EV[\theta \ln \theta_1, \theta]) \tag{18}$$

where $k_1 = \ln \theta_1$ and $EV[a,b]$ denotes an extreme value distributed random variable with location parameter a and scale parameter b, i.e., $F_\varepsilon[t] = \exp[-e^{-(t-a)/b}]$.

From equation (14), $\theta_2/\theta_1 = G_1/\theta_1 = P_1$. Hence $\ln \theta_1 = (\ln G_1 - \ln P_1)$, so that substitution in the final equality of (18) proves the claim.

(*b*) $E(g(\varepsilon_2) | \delta_1(\varepsilon) = 1)$

$$= \frac{1}{P_1} \int_{\varepsilon_1=-\infty}^{+\infty} \int_{\varepsilon_2=-\infty}^{V_1-V_2+\varepsilon_1} \cdots \int_{\varepsilon_J=-\infty}^{V_1-V_J+\varepsilon_1} g(\varepsilon_2)\, dF(\varepsilon)$$

$$= \frac{1}{P_1} \int_{\varepsilon_1=-\infty}^{+\infty} \int_{\varepsilon_2=-\infty}^{V_1-V_2+\varepsilon_1} g(\varepsilon_2) \cdot$$

$$F_{12}[\varepsilon_1, \varepsilon_2, V_1-V_3+\varepsilon_1, \ldots, V_1-V_J+\varepsilon_1]\, d\varepsilon_2\, d\varepsilon_1$$

$$= \frac{1}{P_1} \int_{\varepsilon_2=-\infty}^{+\infty} \int_{\varepsilon_2+V_2-V_1}^{+\infty} g(\varepsilon_2) \cdot$$

$$F_{12}[\varepsilon_1, \varepsilon_2, V_1-V_3+\varepsilon_1, \ldots, V_1-V_J+\varepsilon_1]\, d\varepsilon_1\, d\varepsilon_2 \tag{19}$$

From equation (13):

$$F(\varepsilon) = \exp\Big[- G[<e^{-\varepsilon_j/\theta}>] \Big] \tag{20}$$

$$= \exp\Big[- G^A[<e^{-\varepsilon_j^A/\theta}>] \Big] \cdot \exp[-e^{-\varepsilon_2/\theta}]$$

so that:

$$F_{12}(\varepsilon) = \exp\Big[- G^A[<e^{-\varepsilon_j^A/\theta}>] \Big]\, G_1^A[<e^{-\varepsilon_j^A/\theta}>]\, e^{-\varepsilon_1/\theta}\, \frac{1}{\theta} \tag{21}$$

$$\cdot \exp[-e^{-\varepsilon_2/\theta}]\, e^{-\varepsilon_2/\theta}\, \frac{1}{\theta}$$

Hence:

$$F_{12}(\varepsilon_1, \varepsilon_2, V_1-V_3+\varepsilon_1, \ldots, V_1-V_J+\varepsilon_1)$$

$$= \exp[\, - G^A[e^{-\varepsilon_1/\theta}, <e^{\frac{-V_1+V_j-\varepsilon_1}{\theta}}>]\,] \cdot$$

$$G_1^A\ [\, e^{-\varepsilon_1/\theta}, <e^{\frac{-V_1+V_j-\varepsilon_1}{\theta}}>\,]\, e^{-\varepsilon_1/\theta}\, \frac{1}{\theta}\, \exp\left[-e^{-\varepsilon_2/\theta}\right] e^{-\varepsilon_2/\theta}\, \frac{1}{\theta}$$

$$= \exp\left[-e^{-\varepsilon_1/\theta} e^{-V_1/\theta} \cdot G^A[<e^{V_j/\theta}>]\right] \cdot$$

$$G_1^A[<e^{V_j/\theta}>]\, e^{-\varepsilon_1/\theta}\, \frac{1}{\theta} \cdot \exp[-e^{-\varepsilon_2/\theta}]\, e^{-\varepsilon_2/\theta}\, \frac{1}{\theta} \quad \text{Thus} \tag{22}$$

$$E(g(\varepsilon_2) | \delta_1(\varepsilon) = 1)$$

$$= \frac{G_1^A[<e^{V_j/\theta}>]}{P_1} \cdot \int_{\varepsilon_2=-\infty}^{\infty} g(\varepsilon_2)\, e^{-\varepsilon_2/\theta} \exp[-e^{-\varepsilon_2/\theta}] \cdot$$

$$\int_{\varepsilon_2+V_2-V_1}^{\infty} \exp\left[-e^{-\varepsilon_1/\theta}\, e^{-V_1/\theta}\, G^A[<e^{V_j/\theta}>]\right] e^{-\varepsilon_1/\theta}\, \frac{d\varepsilon_1}{\theta}\, \frac{d\varepsilon_2}{\theta}$$

$$= \frac{G_1^A[<e^{V_j/\theta}>]}{P_1 \cdot \theta_1^A} \int_{\varepsilon_2=-\infty}^{\infty} g(\varepsilon_2)\, e^{-\varepsilon_2/\theta} \exp[-e^{-\varepsilon_2/\theta}] \cdot$$

$$\left[\, 1-\exp\left(-\theta_1^A\, e^{\frac{-\varepsilon_2-V_2+V_1}{\theta}}\right)\right] \frac{d\varepsilon_2}{\theta}$$

where $\theta_1^A = e^{-V_1/\theta}\, G^A[<e^{V_j/\theta}>]$. Thus:

$$E(g(\varepsilon_2) \mid \delta_1(\varepsilon) = 1)$$

$$= \frac{G_1^A[<e^{V_j/\theta}>]}{P_1 \cdot \theta_1^A}\, E(g(\varepsilon_2) \mid \varepsilon_2 \sim EV[0,\theta])$$

$$- \frac{G_1^A[<e^{V_j/\theta}>]}{P_1 \cdot \theta_1^A} \cdot \int_{\varepsilon_2=-\infty}^{\infty} g(\varepsilon_2)\, e^{-\varepsilon_2/\theta} \exp[-e^{-\varepsilon_2/\theta}] \cdot$$

$$\exp[-\theta_1^A\, e^{\frac{-\varepsilon_2 - V_2 + V_1}{\theta}}]\, \frac{d\varepsilon_2}{\theta} \tag{23}$$

But:

$$\int_{\varepsilon_2=-\infty}^{\infty} g(\varepsilon_2)\, e^{-\varepsilon_2/\theta} \exp[-e^{-\varepsilon_2/\theta} \cdot \theta_2]\, \frac{d\varepsilon_2}{\theta}$$

$$= \frac{1}{\theta_2}\, E\left[\, g(\varepsilon_2) \mid \varepsilon_2 \sim EV[\theta \ln \theta_2, \theta] \right]$$

where we define $\theta_2 = (1 + e^{(V_1 - V_2)/\theta} \cdot \theta_1^A)$. Hence:

$$E(g(\varepsilon_2) \mid \delta_1(\varepsilon) = 1)$$

$$= \frac{G_1^A[<e^{V_j/\theta}>]}{P_1 \cdot \theta_1^A} \cdot \left[E\left[g(\varepsilon_2) \mid \varepsilon_2 \sim EV[0, \theta]\right] \right.$$

$$\left. - \frac{1}{\theta_1} \cdot E\left[g(\varepsilon_2) \mid \varepsilon_2 \sim EV[\theta \ln \theta_2, \theta]\right] \right] \tag{24}$$

Note that $G_1^A[<e^{V_j/\theta}>] = G_1[<e^{V_j/\theta}>]$ implies:

$$\frac{G_1^A[<e^{V_j/\theta}>]}{P_1 \cdot \theta_1^A} = \frac{G_1[<e^{V_j/\theta}>]\, e^{V_1/\theta}}{G^A[<e^{V_j/\theta}>]\, P_1} = \frac{G[<e^{V_j/\theta}>]}{G^A[<e^{V_j/\theta}>]} \tag{25}$$

Also:

$$\theta_2 = \left(1 + e^{(V_1 - V_2)/\theta}\right) \theta_1^A = \left(1 + e^{-V_2/\theta}\, G^A[<e^{V_j/\theta}>]\right)$$

$$= e^{-V_2/\theta}\left[e^{V_2/\theta} + G^A[<e^{V_j\theta}>]\right] = e^{-V_2/\theta}\, G[<e^{V_j/\theta}>] \tag{26}$$

Furthermore, $G_2/\theta_2 = P_2$ and $G_2 \equiv 1$ imply:

$$\frac{e^{V_2/\theta}}{G[<e^{V_j/\theta}>]} = \frac{1}{\theta_2} = P_2 \tag{27}$$

Combining equations (25) and (27) with equation (24), we find:

$$E(g(\varepsilon_2)|\delta_1(\varepsilon) = 1)$$

$$= \frac{G[<e^{V_j/\theta}>]}{G^A[<e^{V_j/\theta}>]} \cdot \Bigg[E\left[g(\varepsilon_2)|\varepsilon_2 \sim EV[0, \theta]\right]$$

$$- P_2 \cdot E\left[g(\varepsilon_2)|\ \varepsilon_2 \sim EV[\theta \ln \theta_2, \theta]\right]\Bigg] \tag{28}$$

From equation (27) we have:

$$\ln \theta_2 = \ln G_2 - \ln P_2 = -\ln P_2 \tag{29}$$

Combining (28) and (29) with (27) proves the claim. Q.E.D.

As an application of Lemma B.2.1 we have:

Theorem B.2.2. Let ε be GEV distributed with cumulative distribution function $F(\varepsilon)$ given in (13). Then:

(a) $E(\varepsilon_1 | \delta_1(\varepsilon) = 1) = \theta\,(\gamma + \ln G_1 - \ln P_1)$

(b) $E(\varepsilon_1^2 | \delta_1(\varepsilon) = 1) = (\pi^2/6)\,\theta^2 + \theta^2\,(\gamma + \ln G_1 - \ln P_1)^2$

Let G be additively separable with $G(y) = G^A(y^A) + y_2$ where $y = (y^A, y_2)$ and where $G^A(\cdot)$ is homogeneous of degree one. Let ε have the corresponding partition, i.e., $\varepsilon = (\varepsilon^A, \varepsilon_2)$. Then:

(c) $$E(\varepsilon_2 | \delta_1(\varepsilon) = 1) = \frac{G[<e^{V_j/\theta}>]}{G^A[<e^{V_j/\theta}>]} \cdot \theta \cdot \Big[(1-P_2)\,\gamma + P_2 \ln P_2 \Big]$$

(d) $$E(\varepsilon_2^2 | \delta_1(\varepsilon) = 1) = \frac{G[<e^{V_j/\theta}>]}{G^A[<e^{V_j/\theta}>]} \cdot \theta^2 \cdot$$

$$\Big[\gamma^2 - P_2\,(\gamma - \ln P_2)^2 + (1-P_2)\,(\pi^2/6) \Big]$$

Proof.

(a) Using Lemma B.2.1 (a) with $g(\varepsilon) = \varepsilon$, we have:

$$E(\varepsilon_1 | \delta_1(\varepsilon) = 1) = E(\varepsilon | \varepsilon \sim EV[\theta\,(\ln G_1 - \ln P_1), \theta])$$

$$= \theta\,(\gamma + \ln G_1 - \ln P_1) \qquad (30)$$

(b) We take $g(\varepsilon) = \varepsilon^2$ so that:

$$E(\varepsilon_1^2 | \delta_1(\varepsilon) = 1)$$

$$= E\left[\varepsilon^2 | \varepsilon \sim EV[\theta\,(\ln G_1 - \ln P_1),\, \theta]\right]$$

$$= \left[E(\varepsilon | \varepsilon \sim EV[\theta\,(\ln G_1 - \ln P_1),\, \theta])\right]^2$$

$$+\ var\left[\varepsilon | \varepsilon \sim EV[\theta\,(\ln G_1 - \ln P_1),\, \theta]\right]$$

$$= \theta^2\ [\gamma + \ln G_1 - \ln P_1]^2 + (\pi^2/6)\,\theta^2 \tag{31}$$

(c) Using Lemma B.2.1 (b) with $g(\varepsilon) = \varepsilon$, we have:

$$E(\varepsilon_2 | \delta_1(\varepsilon) = 1)$$

$$= (\frac{G}{G^A}) \cdot \left(E\left[\varepsilon_2 | \varepsilon_2 \sim EV[0,\, \theta]\right] - \right.$$

$$\left. P_2\, E\left[\varepsilon_2 | \varepsilon_2 \sim EV[-\theta \ln P_2,\, \theta]\right] \right)$$

$$= (\frac{G}{G^A}) \cdot \left[\ \gamma\theta - P_2\,(\gamma\theta - \theta \ln P_2)\right]$$

$$= (\frac{G}{G^A} \cdot \theta) \cdot \left[\ (1-P_2)\,\gamma + P_2 \ln P_2\right] \tag{32}$$

(d) Using Lemma B.2.1 (b) with $g(\varepsilon) = \varepsilon^2$, we have:

$$E(\varepsilon_2^2 | \delta_1(\varepsilon) = 1)$$

$$= (\frac{G}{G^A}) \cdot \left[E \left[\varepsilon_2^2 | \varepsilon_2 \sim EV[0, \theta]\right] \right.$$

$$\left. - P_2 \cdot E \left[\varepsilon_2^2 | \varepsilon_2 \sim EV[-\theta \ln P_2, \theta]\right] \right]$$

$$= (\frac{G}{G^A}) \cdot \left(\left[(\gamma\theta)^2 + (\pi^2/6)\, \theta^2\right] - P_2\left[\theta^2\, (\gamma - \ln P_2 + (\pi^2/6)\, \theta^2\right] \right)$$

$$= (\frac{G}{G^A}) \cdot \left[(\gamma\theta)^2 - P_2\, \theta^2\, (\gamma - \ln P_2)^2 + (1 - P_2)\, (\pi^2/6)\, \theta^2 \right] \qquad (33)$$

$$= (\frac{G}{G^A} \cdot \theta^2) \cdot \left[\gamma^2 - P_2\, (\gamma - \ln P_2)^2 + (1 - P_2)\, (\pi^2/6) \right] \quad \text{Q.E.D.}$$

Comments: Theorem B.2.2 imposes strong separability in the functional form for G to obtain a closed form conditional expectation. When G has the additive from $G[y] = G^A[y_2^A] + y_2$, ε_2 is independent from ε^A. If we do not impose strong separability, then $F_{12}(\varepsilon)$ becomes:

$$F_{12}(\varepsilon) = \exp\left[-G[<e^{-\varepsilon_j/\theta}>] \right] e^{-\varepsilon_1/\theta}\, e^{-\varepsilon_2/\theta}\, \frac{1}{\theta^2} \cdot$$

$$\left[G_1[<e^{-\varepsilon_j/\theta}>]\, G_2[<e^{-\varepsilon_j/\theta}>] - G_{12}[<e^{-\varepsilon_j/\theta}>] \right] \qquad (34)$$

Following the proof of Lemma B.2.1 (b), we see that the analogue of (22) for equation (34) does not permit an easy integration in (17). It is possible, however, to extend the results of Theorems B.2.2 (c) and (d) for $G[y] = G^A[y^A] + \alpha y_2$. We present the results in Corollary B.2.2.

Corollary B.2.2 Let ε be GEV distributed with cumulative distribution function $F(\varepsilon)$ given in (13). Let G be additively separable with $G(y) = G^A(y^A) + \alpha y_2$ where $y = (Y^A, y_2)$ and where $G^A(\cdot)$ is homogeneous of degree one. Define $\alpha^* = \theta \ln \alpha$. Then:

$$\text{(a)}\quad E(\varepsilon_2 | \delta_1(\varepsilon) = 1) = \frac{G[<e^{V_j/\theta}>]}{G^A[<e^{V_j/\theta}>]} \cdot \Big[(\gamma\theta + \alpha^*)\,(1-P_2) + \theta\, P_2 \ln P_2 \Big]$$

$$\text{(b)}\quad E(\varepsilon_2^2 | \delta_1(\varepsilon) = 1) = \frac{G[<e^{V_j/\theta}>]}{G^A[<e^{V_j/\theta}>]} \cdot$$

$$\Big[(\pi^2/6)\,\theta^2\,(1 - P_2) + (\gamma\theta + \alpha^*)^2\,(1 - P_2)$$

$$+\, 2\,\theta\,(\gamma\theta + \alpha^*)\,P_2 \ln P_2 - \theta^2\,P_2(\ln P_2)^2 \Big]$$

Proof. The proof of Corollary B.2.2 requires minor modifications in the arguments that demonstrate Lemma B.2.1 (b) and Theorems B.2.2 (c) and (d). It is therefore omitted. Q.E.D.

As an illustration of Theorem B.2.2 and its corollary, we derive the conditional moments for the multinomial and nested logit models.

Example B.2.1. [Conditional Moments in the Multinomial Logit Model].

Let: $G[y] = \alpha\,[\sum_{j=1}^{J} y_j]$ and $\alpha^* = \theta \ln \alpha$. Then:

$$\text{(a)}\quad E(\varepsilon_1 | \delta_1(\varepsilon) = 1) = (\alpha^* + \gamma\theta) - \theta \ln P_1$$

$$\text{(b)}\quad E(\varepsilon_1^2 | \delta_1(\varepsilon) = 1) = \pi^2\,\theta^2/6 + (\alpha^* + \gamma\theta)^2 + \theta^2\,(\ln P_1)^2$$

$$-\, 2\,(\alpha^* + \gamma\theta) \cdot \theta\,(\ln P_1)$$

(c) $E(\varepsilon_2 | \delta_1(\varepsilon) = 1) = (\alpha^* + \gamma\theta) + \theta\, P_2\, (\ln P_2)/(1-P_2)$

(d) $E(\varepsilon_2^2 | \delta_1(\varepsilon) = 1) = \pi^2\theta^2/6 + (\alpha^* + \gamma\theta)^2$

$- P_2\, \theta^2\, (\ln P_2)^2/(1-P_2) + 2\, (\alpha^* + \gamma\theta)\, (\theta \ln P_2)\, P_2/(1-P_2)$

Proof.

(a) $G_1 = \alpha$ and $\theta \ln G_1 = \theta \ln \alpha = \alpha^*$. Apply Theorem B.2.2 (a).

(b) Use Theorem B.2.2 (b) with $G_1 = \alpha$. Then:

$$E(\varepsilon_1^2 | \delta_1(\varepsilon) = 1) = \frac{\pi^2}{6}\, \theta^2 + \theta^2\, (\gamma + \ln \alpha - \ln P_1)^2$$

$$= \frac{\pi^2}{6}\, \theta^2 + \theta^2\, (\gamma + \ln \alpha)^2 - 2\, \theta^2\, (\gamma + \ln \alpha)\, (\ln P_1) + \theta^2\, (\ln P_1)^2$$

$$= \frac{\pi^2}{6} \theta^2 + (\gamma\theta + \alpha^*)^2 - 2\, (\gamma\theta + \alpha^*)\, \theta\, (\ln P_1) + \theta^2\, (\ln P_1)^2$$

(c) Apply Corollary B.2.2 (a) with $G^A[y^A] = \alpha\, (\sum_{j \neq 2} y_j\,)$. Then:

$$E(\varepsilon_2 | \delta_1(\varepsilon) = 1) = \frac{G[<e^{V_j/\theta}>]}{G^A[<e^{V_j/\theta}>]}\ \cdot$$

$$(\gamma\theta + \alpha^*)\, (1-P_2) + \theta\, P_2 \ln P_2$$

From equation (14):

$$\frac{G[<e^{V_j/\theta}>]}{G^A[<e^{V_j/\theta}>]} = \frac{\alpha \sum_{j=1}^{J} e^{V_j/\theta}}{\alpha \sum_{j \neq 2}^{J} e^{V_j/\theta}} = 1/(1-P_2) \qquad (35)$$

Therefore:

$$E(\varepsilon_2|\delta_1(\varepsilon) = 1) = (\gamma\theta + \alpha^*) + \theta\, P_2\, (\ln P_2)/(1-P_2)$$

(d) Apply Corollary B.2.2 (b) with $G^A[y^A] = \alpha \sum_{j \neq 2}^{J} y_j$ and equation (35).

Then:

$$E(\varepsilon_2^2|\delta_1(\varepsilon) = 1) = \frac{\pi^2}{6}\,\theta^2 + (\gamma\theta + \alpha^*)^2 + 2\theta\,(\gamma\theta + \alpha^*)\cdot$$

$$P_2\,(\ln P_2)\,/\,(1-P_2) - \theta^2\, P_2\,(\ln P_2)^2\,/\,(1-P_2) \quad \text{Q.E.D.}$$

As a second illustration of Theorem B.2.2 and Corollary B.2.2, we derive the conditional moments for the nested logit model.

Example B.2.2. Consider a two-level nested logit model with three alternatives:

$$G[y_1, y_2, y_3] = \left[\, y_1^{1/(1-\sigma)} + y_3^{1/(1-\sigma)} \,\right]^{(1-\sigma)} + y_2 \tag{36}$$

Following McFadden (1978), one may verify that (36) satisfies the conditions of Theorem B.2.1. Therefore, application of equation (14) yields:

$$P[2|1, 2, 3] = \frac{e^{V_2/\theta}}{\left[\, e^{V_1/\theta(1-\sigma)} + e^{V_3/\theta(1-\sigma)} \,\right]^{(1-\sigma)} + e^{V_2/\theta}} \tag{37}$$

$$P[1|1, 2, 3] = \frac{\left[\, e^{V_1/\theta(1-\sigma)} + e^{V_3/\theta(1-\sigma)} \,\right]^{(1-\sigma)}}{\left[\, e^{V_1/\theta(1-\sigma)} + e^{V_3/\theta(1-\sigma)} \,\right]^{(1-\sigma)} + e^{V_2/\theta}} \cdot$$

$$\frac{e^{V_1/\theta(1-\sigma)}}{\left[\, e^{V_1/\theta(1-\sigma)} + e^{V_3/\theta(1-\sigma)} \,\right]} \tag{38}$$

$$= P[(1, 3)|(1, 2, 3)] \cdot P[1|(1, 3)]$$

where $P(i|A)$ denotes the probability that i is chosen from the set A. From equation (36) we find:

$$G_1 = \left[e^{V_1/\theta(1-\sigma)} + e^{V_3/\theta(1-\sigma)} \right]^{-\sigma} \cdot e^{\sigma V_1/\theta(1-\sigma)}$$

$$= P[1|1, 3]^{\sigma} \tag{39}$$

We define $G^A[y_1, y_2, y_3] = [y_1^{1/(1-\sigma)} + y_3^{1/(1-\sigma)}]^{(1-\sigma)}$. It follows that:

$$(\frac{G}{G^A})\ [<e^{V_i/\theta}>] = \frac{\left[e^{V_1/\theta(1-\sigma)} + e^{V_3/\theta(1-\sigma)} \right]^{(1-\sigma)}}{\left[e^{V_1/\theta(1-\sigma)} + e^{V_3/\theta(1-\sigma)} \right]^{(1-\sigma)}}$$

$$+ \frac{e^{V_2/\theta}}{\left[e^{V_1/\theta(1-\sigma)} + e^{V_3/\theta(1-\sigma)} \right]^{(1-\sigma)}}$$

$$= 1 + \frac{P[2|(1, 2, 3)]}{P[(1, 3)|(1, 2, 3)]}$$

$$= 1/(1 - P[2|(1, 2, 3)]) \tag{40}$$

For G given in equation (36), Theorem B.2.2 implies:

$$E(\varepsilon_1|\delta_1(\varepsilon) = 1) = \theta\left[\gamma + \sigma \ln P(1|1, 3) - \ln P(1|1, 2, 3)\right] \tag{41}$$

$$E(\varepsilon_1^2|\delta_1(\varepsilon) = 1) = \frac{\pi^2}{6}\theta^2 + \theta^2\left[\gamma + \sigma \ln P(1|1, 3) - \ln P(1|1, 2, 3)\right]^2 \tag{42}$$

Application of Theorems B.2.2 (c) and (d) to equation (34) implies:[5]

$$E(\varepsilon_2|\delta_1(\varepsilon) = 1) = \theta\ [\gamma + P_2\ (\ln P_2)/(1-P_2)] \tag{43}$$

and:

$$E(\varepsilon_2^2|\delta_1(\varepsilon) = 1) = \theta^2\left[\frac{\pi^2}{6} + \left[\gamma^2 - P_2\ (\gamma - \ln P_2)^2\right]/(1-P_2)\right] \tag{44}$$

In equations (41) and (42), one observes that the nested logit model implies a closed-form expression in the conditional probabilities of reaching alternative one from various nodes of the tree. The conditional expectations in (41) and (42) differ from their counterparts derived in the multinomial logit example by the term $\sigma \ln P(1|(1, 3))$. As σ tends to zero in the limit, the nested logit model converges to the multinomial logit model and the term $\sigma \ln P(1|(1, 3))$ vanishes.

Comparison of (37), (38), and the corresponding expressions in the multinomial logit example reveals equal conditional expectations. The essence of the separability assumption is that the variable ε_2 behaves as a multinomial rather than nested logit random variable.

The calculations involved in equations (41)—(44) are easily modified to trees of any depth or size. As an illustration, consider a two-level nested logit model with M alternatives:

$$G(y) = \sum_{m=1}^{M} a_m \left[\sum_{i \in B_m} y_i^{1/(1-\sigma_m)}\right]^{1-\sigma_m} \tag{45}$$

[5] Expressions for the conditional expectation of ε_3 are analogous to those presented for ε_1.

where $B_m \subseteq \{1, 2, ..., J\}$, $\bigcup_{m-1}^{M} B_m = \{1, 2, ..., J\}$, $a_m > 0$, and $0 \leqslant \sigma_m < 1$. McFadden (1978) derives the choice probabilities for equation (45) and shows that they satisfy:

$$P_i = \frac{\sum_{m \mid i \in B_m}^{M} e^{V_i/(1-\sigma_m)} \, a_m \left[\sum_{j \in B_m} e^{V_j/(1-\sigma_m)} \right]^{-\sigma_m}}{\sum_{n-1}^{M} a_n \left[\sum_{k \in B_n} e^{V_k/(1-\sigma_n)} \right]^{(1-\sigma_n)}}$$

$$= \sum_{m \mid i \in B_m} P[i \mid B_m] \cdot P[B_m] \tag{46}$$

where:

$$P[i \mid B_m] = \begin{cases} e^{V_i/(1-\sigma_m)} / \sum_{j \in B_m} e^{V_j/(1-\sigma_m)} & \text{if } i \in B_m \\ 0 & \text{otherwise} \end{cases} \tag{47}$$

and:

$$P[B_m] = \frac{a_m \left[\sum_{j \in B_m} e^{V_j/(1-\sigma_m)} \right]^{(1-\sigma_m)}}{\sum_{n-1}^{M} a_n \left[\sum_{k \in B_n} e^{V_k/(1-\sigma_n)} \right]^{(1-\sigma_n)}} \tag{48}$$

From equation (45) we have:

$$G_i(y) = \sum_{m \mid i \in B_m}^{M} a_m \left[\sum_{j \in B_m} y_j^{1/(1-\sigma_m)} \right]^{-\sigma_m} \cdot y_i^{\sigma_m/(1-\sigma_m)} \tag{49}$$

so that:

$$G_i(<e^{V_j}>) = \sum_{m=1}^{M} a_m \, P[i \,|\, B_m]^{\sigma_m} \tag{50}$$

The form of the derivative in (50) generalizes in higher order trees. As an example consider the three-level tree structure:

$$G = \sum_a \left[\sum_d \left[\sum_m y_{mda}^{1/(1-\sigma)} \right]^{(1-\sigma)/(1-\delta)} \right]^{(1-\delta)} \tag{51}$$

In this case one may show:

$$G_{mda}[<e^{V_j}>] = \sum_a \sum_d P[d \,|\, a]^{\delta} \cdot P[m \,|\, da]^{\sigma} \tag{52}$$

where G_{mda} denotes the derivative of G in (51) with respect to y_{mda}. Further, equation (40) generalizes to cases in which G exhibits strong separability in some of its arguments. For example, suppose:

$$G = G^A + a_{M+1}\, y_{M+1} \,. \text{ Then: } P_{M+1} = a_{M+1}\, e^{V_{M+1}/\theta}/G$$

and:

$$((G-G^A)/G)\,(<e^{V_j/\theta}>) = P_{M+1}$$

Thus:

$$(G/G^A)\,(<e^{V_j/\theta}>) = (1 - P_{M+1})^{-1}$$

as in equation (40).

B.3. Conditional covariance in the GEV family

We now consider the conditional moment of the product of two GEV random variables. Rather than calculate $E[\varepsilon_1\varepsilon_2|\delta_1(\varepsilon) = 1]$, we will alternatively find $E[(\varepsilon_2 - \varepsilon_1)^2|\delta_1(\varepsilon) = 1]$ and use the relation $(\varepsilon_2 - \varepsilon_1)^2 = \varepsilon_2^2 - 2\varepsilon_1\varepsilon_2 + \varepsilon_1^2$ in conjunction with Theorem B.2.2. It is well known that the difference $(\varepsilon_2 - \varepsilon_1)$ has a logistic distribution when ε_1 and ε_2 are independent identically extreme value distributed. Our next result finds the joint distribution function for $(Y_2, Y_3, ..., Y_J) = (\varepsilon_2 - \varepsilon_1, \varepsilon_3 - \varepsilon_1, ..., \varepsilon_J - \varepsilon_1)$ when ε has a GEV distribution.

Theorem B.3.1. [Generalized Logistic Distribution]. Let $Y_j = \varepsilon_j - \varepsilon_1$ for $j = 2, 3, ..., J$ where ε has a GEV distribution given by equation (13). Then:

$$H[w_2, w_3, ..., w_J] = Prob[Y_2 \leqslant w_2, Y_3 \leqslant w_3, ..., Y_J \leqslant w_J]$$

$$= G_1[<e^{-w_j/\theta}>]/G[<e^{-w_j/\theta}>]$$

with $w_1 \equiv 0$.

Proof.

$$H = Prob[Y_2 \leqslant w_2, ..., Y_J \leqslant w_J]$$

$$= \int_{\varepsilon_1=-\infty}^{\infty} \int_{\varepsilon_2=-\infty}^{\varepsilon_1+w_2} \cdots \int_{\varepsilon_J=-\infty}^{\varepsilon_1+w_J} dF(\varepsilon)$$

$$= \int_{\varepsilon_1=-\infty}^{\infty} F_1[\varepsilon, \varepsilon+w_2, ..., \varepsilon+w_J]\, d\varepsilon$$

$$= \int_{\varepsilon=-\infty}^{\infty} \exp\Big[-G[<e^{-(\varepsilon+w_j)/\theta}>]\Big] G_1[<e^{(-\varepsilon-w_j)/\theta}>]e^{-\varepsilon/\theta}\, \frac{d\varepsilon}{\theta}$$

$$= \int_{\varepsilon=-\infty}^{\infty} \exp\Big[-e^{-\varepsilon/\theta}G[<e^{-w_j/\theta}>]\Big] G_1[<e^{-w_j/\theta}>]e^{-\varepsilon/\theta}\, \frac{d\varepsilon}{\theta}$$

$$= \frac{G_1[<e^{-w_j/\theta}>]}{G[<e^{-w_j/\theta}>]} \quad \text{Q.E.D.}$$

Two familiar results follow immediately from Theorem B.3.1.

Corollary B.3.1.

(a) $H[V_1-V_2, V_1-V_3, ..., V_1-V_J] = P_1$

(b) $(Y_2, Y_3, ..., Y_J)$ is logistically distributed when $G[y] = \sum_{j=1}^{J} y_j$.

Proof.

(a) $H[V_1-V_2, ..., V_1-V_J] = \dfrac{G_1[<e^{-(V_1-V_j)/\theta}>]}{G[<e^{-(V_1-V_j)/\theta}>]}$

$$= e^{V_1/\theta}\, G_1[<e^{V_j/\theta}>]\, G[<e^{V_j/\theta}>] = P_1$$

where the first equality uses the result of Theorem B.3.1, the second equality uses the homogeneity property of G, and the third equality uses equation (14).

(b) Because $G[y] = \sum_{j=1}^{J} y_j$, $G_1[y] = 1$.

Theorem B.3.1 implies:

$$H[w_2, ..., w_J] = \left[\sum_{j=1}^{J} e^{-w_j/\theta} \right]^{-1}$$

which is a multivariate logistic distribution. Q.E.D.

We now make the assumption that ε_1 and ε_2 are independent from each other and from ε^A.

Theorem B.3.2. Let ε be GEV distributed with $G[y] = \alpha\, y_1 + \alpha\, y_2 + \alpha\, G^A[<y_j^A>]$ where G^A is homogeneous of degree one and where $y = (y_1, y_2, y^A)$. Then:

$$E((\varepsilon_2 - \varepsilon_1)^2 | \delta_1(\varepsilon) = 1)$$

$$= \theta^2\, [\ln ((1-P_2)/P_1)]^2 - 2\theta^2 \ln ((1-P_2)/P_1) \cdot$$

$$\Big[P_2 \ln P_2/(1-P_2) + \ln (1-P_2) \Big] + \theta^2/(1-P_2) \cdot$$

$$\int_{-\infty}^{\ln [(1-P_2)/P_2]} h(z)\, dz \quad \text{with} \quad h(z) = \frac{z^2\, e^{-z}}{(1+e^{-z})^2}$$

Proof.

$$E((\varepsilon_2 - \varepsilon_1)^2 | \delta_1(\varepsilon) = 1)$$

$$= \frac{1}{P_1} \int_{\varepsilon_1=-\infty}^{\infty} \int_{\varepsilon_2=-\infty}^{V_1-V_2+\varepsilon_1} (\varepsilon_2 - \varepsilon_1)^2 \cdot$$

$$F_{12}[\varepsilon_1, \varepsilon_2, V_1-V_3+\varepsilon_1, \ldots, V_1-V_J+\varepsilon_1]\, d\varepsilon_2\, d\varepsilon_1$$

We make the logistic transformation: $z_1 = \varepsilon_1$, $z_2 = \varepsilon_2 - \varepsilon_1$. It is easily verified that this transformation has unit Jacobian. Thus:

$$E((\varepsilon_2 - \varepsilon_1)^2 | \delta_1(\varepsilon) = 1)$$

$$= \frac{1}{P_1} \int_{z_1=-\infty}^{\infty} \int_{z_2=-\infty}^{V_1-V_2} z_2^2 \cdot$$

$$F_{12}[z_1, z_1 + z_2, V_1 - V_3 + z_1, ..., V_1 - V_J + z_1]\, dz_2\, dz_1$$

$$= \frac{1}{P_1} \int_{z_2=-\infty}^{V_1-V_2} z_2^2 \int_{z_1=-\infty}^{\infty}$$

$$F_{12}[z_1, z_1 + z_2, V_1 - V_3 + z_1, ..., V_1 - V_J + z_1]\, dz_1\, dz_2$$

Define:

$$H[w_2, ..., w_J] = \int_{\varepsilon=-\infty}^{\infty} F_1[\varepsilon, \varepsilon + w_2, ..., \varepsilon + w_J]\, d\varepsilon$$

Then:

$$E((\varepsilon_2 - \varepsilon_1)^2 | \delta_1(\varepsilon) = 1)$$

$$= \int_{z_2=-\infty}^{V_1-V_2} z_2^2 \cdot H_2[z_2, V_1 - V_3, ..., V_1 - V_J]\, dz_2$$

Recall that $G[y_1, y_2, ..., y_J] = \alpha y_1 + \alpha y_2 + \alpha G^A[<y_j^A>]$. Thus $G_1 = \alpha$, and by Theorem B.3.1:

$$H[w_2, ..., w_J] = \alpha \left[\alpha + \alpha e^{-w_2/\theta} + \alpha G^A[<e^{-w_j/\theta}>]\right]^{-1}$$

from which follows:

$$H_2[w_2, ..., w_J] = e^{-w_2/\theta} \left[\, 1 + G^A[<e^{-w_j/\theta}>] + e^{-w_2/\theta} \right]^{-2}$$

and:

$$E((\varepsilon_2 - \varepsilon_1)^2 | \delta_1 = 1)$$

$$= \frac{\theta^2}{P_1} \int_{-\infty}^{(V_1-V_2)/\theta} \frac{y^2 e^{-y}}{(A + e^{-y})^2}\, dy$$

$$= \frac{\theta^2}{P_1\, A^2} \cdot \int_{-\infty}^{(V_1-V_2)/\theta} \frac{y^2\, e^{-y}}{(1 + e^{-\ln A - y})^2}\, dy$$

where $A = 1 + G^A[<e^{-(V_1-V_j)/\theta}>]$ and we have used the transformation $y = z_2/\theta$. Note that:

$$\frac{(1-P_2)}{P_1} = \frac{G-\alpha\, e^{V_2/\theta}}{\alpha\, e^{V_1/\theta}} = \frac{\alpha\, e^{V_1/\theta} + \alpha\, G^A}{\alpha\, e^{V_1/\theta}}$$

$$= 1 + e^{-V_1/\theta} \cdot G^A[<e^{V_j/\theta}>] = A$$

Define $z = y + \ln A$. Then:

$$E(Y_2^2 | \delta_1 = 1) = \frac{\theta^2}{P_1\, A^2} \cdot \int_{-\infty}^{((V_1-V_2)/\theta) + \ln A} \frac{(z - \ln A)^2\, A\, e^{-z}}{(1+e^{-z})^2}\, dz$$

$$= \frac{\theta^2}{P_1\, A} \cdot \int_{-\infty}^{((V_1-V_2)/\theta) + \ln A} \frac{(z^2 - 2\, z \ln A + (\ln A)^2)\, e^{-z}}{(1+e^{-z})^2}\, dz$$

Because:

$$(V_1-V_2)/\theta = \ln (P_1/P_2) \;\text{ and }\; A = (1-P_2)/P_1$$

It follows that:

$$(V_1-V_2)/\theta + \ln A = \ln (P_1/P_2) + \ln [(1-P_2)/P_1] = \ln [(1-P_2)/P_2]$$

We let $x = \ln [(1-P_2)/P_2]$, so that:

$$E(Y_2^2 | \delta_1 = 1) = \frac{\theta^2}{P_1 A} \int_{-\infty}^{x} \frac{z^2 e^{-z}}{(1 + e^{-z})^2} dz +$$

$$\frac{\theta^2}{P_1 A} \int_{-\infty}^{x} \frac{e^{-z}}{(1 + e^{-z})^2} dz \cdot (\ln A)^2 -$$

$$\frac{2 (\ln A) \theta^2}{P_1 A} \int_{-\infty}^{x} \frac{e^{z}}{(1 + e^{-z})^2} dz$$

$$= \frac{\theta^2}{(1-P_2)} \int_{-\infty}^{x} \frac{z^2 e^{-z}}{(1 + e^{-z})^2} dz + \theta^2 (\ln ((1-P_2)/P_1))^2$$

$$\frac{-2 (\ln A) \theta^2}{(1-P_2)} \int_{-\infty}^{x} \frac{z e^{-z}}{(1 + e^{-z})^2} dz$$

Using integration by parts, we have:

$$\int_{t=-\infty}^{x} \frac{t e^{-t} dt}{(1 + e^{-t})^2} = \frac{x}{1 + e^{-x}} - \ln (1 + e^{x})$$

from which follows:

$$\int_{t=-\infty}^{\ln [(1-P_2)/P_2]} \frac{t e^{-t} dt}{(1 + e^{-t})^2} = \frac{\ln [(1-P_2)/P_2]}{(1-P_2)^{-1}} + \ln P_2$$

$$= \left[P_2 \ln P_2 + (1-P_2) \ln(1-P_2) \right]$$

Hence:

$$E(Y_2^2 | \delta_1(\varepsilon) = 1) = \theta^2 (\ln ((1-P_2)/P_1))^2$$

$$- \frac{2\,\theta^2}{(1-P_2)} \ln\,((1-P_2)/P_1)\Big[\,P_2 \ln P_2 + (1-P_2) \ln\,(1-P_2)\Big]$$

$$+ \frac{\theta^2}{(1-P_2)} \cdot \int_{-\infty}^{\ln\,[(1-P_2)/P_1]} h(z)\,dz$$

$$= \theta^2\,[\ln\,((1-P_2)/P_1)]^2$$

$$-\,2\,\theta^2 \ln\,((1-P_2)/P_1)\Big[\,P_2 \ln P_2/(1-P_2) \;+\; \ln\,(1-P_2)\Big]$$

$$+ \frac{\theta^2}{(1-P_2)} \cdot \int_{-\infty}^{\ln\,[(1-P_2)/P_2]} h(z)\,dz \quad \text{Q.E.D.}$$

The integral $\int_{-\infty}^{x} h(z)\,dz$ where $h(z) = (z^2\,e^{-z})/(1 + e^{-z})^2$ is in fact related to $E[y^2 | y < x]$ where y has a univariate logistic distribution. A closed-form solution for this conditional expectation does not exist. Hay (1980) and Lee (1981) determine the expectation in terms of a series expansion involving the incomplete gamma distribution. Using an alternative series expansion, we derive a computationally simple form for the integral.

Theorem B.3.3. For $0 < \lambda < 1$:

$$\int_0^{\ln \lambda^{-1}} \frac{u^2 e^{-u}}{(1 + e^{-u})^2} du = \frac{\pi^2}{6} - \frac{\lambda (\ln \lambda)^2}{(1 + \lambda)} - 2 (\ln \lambda) (\ln (1 + \lambda))$$

$$+ 2 \sum_{i=0}^{\infty} (-1)^i \frac{\lambda^{i+1}}{(i+1)^2}$$

Proof. From the formula for the sum of a geometric series we have:

$$(1 + x)^{-1} = \sum_{i=0}^{\infty} (-1)^i x^i \quad \text{for} \quad |x| < 1$$

Differentiating and integrating, term by term, provides two useful relations:

$$\frac{1}{(1+x)^2} = \sum_{i=1}^{\infty} (-1)^{i+1} i\, x^{i-1} = \sum_{i=0}^{\infty} (-1)^i (i+1)\, x^i$$

$$\ln (1+x) = \sum_{i=0}^{\infty} \frac{(-1)^i}{(i+1)} x^{i+1} \text{ for } |x| < 1$$

For $x = e^{-u}$ with $u > 0$:

$$(1 + e^{-u})^{-2} = \sum_{i=0}^{\infty} (-1)^i (i+1) e^{-ui} \quad \text{and}$$

$$\int_0^{\ln \lambda^{-1}} \frac{u^2 e^{-u}}{(1 + e^{-u})^2} du = \int_0^{\ln \lambda^{-1}} u^2 \sum_{i=0}^{\infty} (-1)^i (i+1)\, e^{-u(i+1)} du$$

$$= \sum_{i=0}^{\infty} (-1)^i (i+1) \int_0^{\ln \lambda^{-1}} u^2 e^{-u(i+1)} du$$

Next we use the fact that:

$$\int y^2\, e^{-iy}\, dy = \frac{-1}{i}\left[y^2 + \frac{2}{i}\, y + \frac{2}{i^2} \right] e^{-iy}$$

Then:

$$\int_0^{\ln \lambda^{-1}} \frac{u^2\, e^{-u}}{(1+e^{-u})^2}\, du$$

$$= \sum_{i=0}^{\infty} (-1)^i\, (i+1) \left[\frac{-1}{(i+1)}\, y^2 + \frac{2}{(i+1)}\, y + \frac{2}{(i+1)^2} \right] e^{-(i+1)y} \Bigg|_0^{-\ln \lambda}$$

$$= \sum_{i=0}^{\infty} (-1)^{i+1} \left[\left\{ (\ln \lambda^{-1})^2 + \frac{2}{(i+1)} \ln \lambda^{-1} + \frac{2}{(i+1)^2} \right\} \lambda^{i+1} - \frac{2}{(i+1)^2} \right]$$

$$= -2 \sum_{i=0}^{\infty} \frac{(-1)^{i+1}}{(i+1)^2} + (\ln \lambda^{-1})^2 \cdot \sum_{i=0}^{\infty} (-1)^{i+1}\, \lambda^{i+1}$$

$$+ 2\, (\ln \lambda^{-1}) \cdot \sum_{i=0}^{\infty} \frac{(-1)^{i+1}}{(i+1)} \lambda^{i+1} + 2 \sum_{i=0}^{\infty} (-1)^{i+1}\, \lambda^{i+1}/(i+1)^2$$

$$= \frac{\pi^2}{6} - \left[\frac{\lambda (\ln \lambda)^2}{(1+\lambda)} - 2\, (\ln \lambda)\, (\ln\, (1+\lambda)) + 2 \sum_{i=0}^{\infty} (-1)^i\, \lambda^{i+1}/(i+1)^2 \right]$$

where we have used the fact that:

$$\sum_{i=0}^{\infty} (-1)^i/(i+1)^2 = \pi^2/12 \quad \text{Q.E.D.}$$

For reference below, we let:

$$G(\lambda) = \left[\frac{\lambda\,(\ln \lambda)^2}{(1+\lambda)} - 2\,(\ln \lambda) \ln (1+\lambda) + 2 \sum_{i=0}^{\infty} (-1)^i \lambda^{i+1}/(i+1)^2 \right]$$

Application of Theorem B.3.2 in the case of binary alternatives yields:

Corollary B.3.2.

$$E(Y_2^2 | \delta_1 = 1) = \begin{cases} (\theta^2/P_1) \cdot (\pi^2/3 - G(P_2/P_1)) & \text{for } P_1 > P_2 \\ (\theta^2/P_1) \cdot (G(P_1/P_2)) & \text{for } P_1 < P_2 \\ (\theta^2/P_1) \cdot (\pi^2/6) & \text{for } P_1 = P_2 \end{cases}$$

Proof. Using Theorem B.3.2:

$$E(Y_2^2 | \delta_1 = 1) = \frac{\theta^2}{P_1} \int_{-\infty}^{\ln (P_1/P_2)} h(z)\, dz$$

where we have imposed the restriction $P_1 + P_2 = 1$ implied in the case of binary alternatives. For $P_1 > P_2$:

$$E(Y_2^2 | \delta_1 = 1) = \frac{\theta^2}{P_1} \int_{-\infty}^{0} h(z)\, dz + \frac{\theta^2}{P_1} \int_{0}^{\ln (P_1/P_2)} h(z)\, dz$$

Now make the substitution $\lambda^{-1} = P_1/P_2$. Then Theorem B.3.3 implies:

$$E(Y_2^2 | \delta_1 = 1) = (\theta^2/P_1)(\pi^2/6) + (\theta^2/P_1) \left[(\pi^2/6) - G(P_2/P_1) \right]$$

For $P_1 < P_2$:

$$E(Y_2^2 | \delta_1 = 1) = \frac{\theta^2}{P_1} \int_{-\infty}^{0} h(z)\, dz - \frac{\theta^2}{P_1} \int_{\ln (P_1/P_2)}^{0} h(z)\, dz$$

$$= \frac{\theta^2}{P_1}\frac{\pi^2}{6} - \frac{\theta^2}{P_1}\left[\pi^2/6 - G(P_1/P_2)\right] = \frac{\theta^2}{P_1} G(P_1/P_2)$$

Finally, when $P_1 = P_2$, $G(1) = \pi^2/6$, which implies continuity for $E(Y_2^2 | \delta = 1)$ Q.E.D.

B.4. Continuous/discrete econometric systems

We now introduce a random variable η and suppose that conditioned on ε, η has mean $(\sqrt{6}\sigma/\pi\theta)\sum_{i=1}^m R_i\varepsilon_i$ and variance $\sigma^2(1 - \sum_{i=1}^m R_i^2)$ with $\sum_{i=1}^m R_i = 0$ and $\sum_{i=1}^m R_i^2 < 1$.

For the present, we assume that $<\varepsilon_i>$ are independently, identically extreme value distributed with $E(\varepsilon_i) = 0$. This is accomplished by assuming that location parameter $\alpha = -\gamma\theta$. Note that $(\sqrt{6}\sigma/\pi\theta) = (\sigma/\sigma_\varepsilon)$ where σ_ε is the standard deviation of ε_i. The next theorem presents the unconditional moments of η.

Theorem B.4.1. [Dubin and McFadden].

(a) $E(\eta) = 0$

(b) $E(\eta)^2 = \sigma^2$

(c) $Correl\ (\eta, \varepsilon_i) = R_i$

Proof.

(a) $E(\eta) = E[E(\eta|\varepsilon)] = E\left[\frac{\sigma}{\sigma_\varepsilon}\sum_{i=1}^m R_i\,\varepsilon_i\right] = 0$

(b) $E(\eta^2|\varepsilon) = var\ (\eta|\varepsilon) + (E(\eta|\varepsilon))^2$

$$E(\eta^2) = E\left[\sigma^2(1 - \sum_{i=1}^m R_i^2) + \left[\frac{\sigma}{\sigma_\varepsilon}\sum_{i=1}^m R_i\,\varepsilon_i\right]^2\right]$$

$$= \sigma^2 \, (1 - \sum_{i=1}^{m} R_i^2) + \frac{\sigma^2}{\sigma_\varepsilon^2} \sum_{i=1}^{m} R_i^2 \, \sigma_\varepsilon^2 = \sigma^2$$

(c) $E(\eta \, \varepsilon_i) = E \left[E(\eta \varepsilon_i | \varepsilon) \right] = E \left[\varepsilon_i \, E(\eta | \varepsilon) \right]$

$$= E \left[\varepsilon_i \, \frac{\sigma}{\sigma_\varepsilon} \sum_{i=1}^{m} R_i \, \varepsilon_i \right] = \frac{\sigma}{\sigma_\varepsilon} R_i \, \sigma_\varepsilon^2 = \sigma \, R_i \, \sigma_\varepsilon$$

Therefore, *Correl* $(\eta, \varepsilon_i) = E(\eta \, \varepsilon_i)/\sigma \, \sigma_\varepsilon = R_i$ Q.E.D.

We now derive the expectation of η conditioned on the event that a particular alternative is chosen.

Theorem B.4.2. [Dubin and McFadden].

$$E(\eta | \delta_i(\varepsilon) = 1) = \frac{\sqrt{6}\sigma}{\pi} \left[\sum_{j=1}^{m} \frac{R_j \, P_j}{(1-P_j)} \ln P_j - R_i \, \frac{\ln P_i}{(1-P_i)} \right]$$

Proof. Define $A_i \equiv \{\varepsilon \mid \delta_i(\varepsilon) = 1\}$. Then:

$$E(\eta | \delta_i = 1) = \frac{1}{P_i} \int_{A_i} E(\eta | \varepsilon) \prod_{j=1}^{m} f(\varepsilon_j) \, d\varepsilon$$

$$= \frac{1}{P_i} \int_{A_i} \left(\frac{\sigma}{\sigma_\varepsilon} \sum_{j=1}^{m} R_j \, \varepsilon_j \right) \prod_{j=1}^{m} f(\varepsilon_j) \, d\varepsilon$$

$$= \frac{\sigma}{\sigma_\varepsilon} \sum_{j=1}^{m} \frac{R_j}{P_i} \int_{A_i} \varepsilon_j \prod_{j=1}^{m} f(\varepsilon_j) \, d\varepsilon$$

$$= \frac{\sigma}{\sigma_\varepsilon} \sum_{j=1}^{m} E[\varepsilon_j | \delta_i(\varepsilon) = 1] \cdot R_j$$

$$= \frac{\sigma}{\sigma_\varepsilon} \sum_{j \mid j \neq i}^{m} E[\varepsilon_j \mid \delta_i(\varepsilon) = 1]\, R_j + \frac{\sigma}{\sigma_\varepsilon} E[\varepsilon_i \mid \delta_i(\varepsilon) = 1]\, R_i$$

Using Example B.2.1, we find:

$$E(\eta \mid \delta_i(\varepsilon) = 1) = \frac{\sigma}{\sigma_\varepsilon} \sum_{j \mid j \neq i}^{m} \frac{\theta\, R_j\, P_j \ln P_j}{(1-P_j)} - \frac{\sigma}{\sigma_\varepsilon} R_i\, \theta \ln P_i$$

where we have imposed $\alpha = -\gamma\theta$. Noting that $\sigma_\varepsilon = (\pi\, \theta/\sqrt{6})$, we have:

$$E(\eta \mid \delta_i(\varepsilon) = 1) = \frac{\sqrt{6}\sigma}{\pi} \left[\left(\sum_{j \mid j \neq i}^{m} \frac{R_j\, P_j \ln P_j}{(1-P_j)} \right) - R_i \ln P_i \right]$$

$$= \frac{\sqrt{6}\sigma}{\pi} \left[\left(\sum_{j=1}^{m} \frac{R_j\, P_j \ln P_j}{(1-P_j)} \right) - \frac{R_i \ln P_i}{(1-P_i)} \right] \quad \text{Q.E.D.}$$

Let $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Then we may rewrite the result of Theorem B.4.2 as:

$$E(\eta \mid \delta_i(\varepsilon) = 1) = \frac{\sqrt{6}\sigma}{\pi} \left[\left(\sum_{j \mid j \neq i}^{m} \frac{R_j\, P_j \ln P_j}{(1-P_j)} \right) + \frac{R_i \ln P_i\, (P_i - 1)}{(1-P_i)} \right]$$

$$= \frac{\sqrt{6}\sigma}{\pi} \left[\sum_{j=1}^{m} \frac{R_j \ln P_j}{(1-P_j)} (P_j - \delta_{ij}) \right]$$

We now consider the conditional variance of η for the binary case $m = 2$. Recall that:

$$E(\eta^2 \mid \delta_i = 1) = \frac{1}{P_i} \int_{A_i} E(\eta^2 \mid (\varepsilon))\, f(\varepsilon)\, d\varepsilon$$

where $f(\varepsilon) = \prod_{i=1}^{m} f(\varepsilon_i)$. We use the relation:

$$E(\eta^2|\varepsilon) = var(\eta|\varepsilon) + (E(\eta|\varepsilon))^2$$

$$= \sigma^2 \,(1 - \sum_{i=1}^{2} R_i^2) + \frac{\sigma^2}{\sigma_\varepsilon^2} \,(\sum_{i=1}^{2} R_i \,\varepsilon_i)^2$$

to obtain:

$$E(\eta^2|\delta_i = 1) = \sigma^2 \,(1 - \sum_{i=1}^{2} R_i^2) + \frac{\sigma^2}{\sigma_\varepsilon^2} \sum_{i=1}^{2} R_i^2 \, E(\varepsilon_i^2|\delta_i = 1)$$

$$+ \frac{2\,\sigma^2}{\sigma_\varepsilon^2} R_1 \, R_2 \, E(\varepsilon_1 \, \varepsilon_2|\delta_i = 1)$$

We collect results in the next theorem:

Theorem B.4.3. [Dubin and McFadden].

$E(\eta^2|\delta_1) = \sigma^2 + 2\,\sigma^2 R_2^2 \, J(P_1,\delta_1)$ where $J(P_1, \delta_1) =$

$$\begin{cases} 1/P_1 - 1 - (3/\pi^2)\,(1/P_1)\cdot G\left[\dfrac{(1-P_1)}{P_1}\right] & \text{for } \delta_1 = 1,\ P_1 > 1/2 \\ -1 + (3/\pi^2)\,(1/P_1)\cdot G\left[\dfrac{P_1}{(1-P_1)}\right] & \text{for } \delta_1 = 1,\ P_1 \leqslant 1/2 \\ -1 + (3/\pi^2)\,[1/(1-P_1)]\cdot G\left[\dfrac{(1-P_1)}{P_1}\right] & \text{for } \delta_1 = 0,\ P_1 > 1/2 \\ P_1/(1-P_1) - (3/\pi^2)\,[1/(1-P_1)]\cdot G\left[\dfrac{P_1}{(1-P_1)}\right] & \text{for } \delta_1 = 0,\ P_1 \leqslant 1/2 \end{cases}$$

Proof.

$$E[\eta^2|\delta_1 = 1] = \sigma^2 (1 - (R_1^2 + R_2^2)) + \frac{\sigma^2}{\sigma_\varepsilon^2} \Big[R_1^2 \, E(\varepsilon_1^2|\delta_1 = 1)$$

$$+ R_2^2 \, E(\varepsilon_2^2|\delta_1 = 1) + 2 \, R_1 \, R_2 \, E(\varepsilon_1\varepsilon_2|\delta_1 = 1)\Big]$$

For the binary case: $P_1 + P_2 = 1$ and $R_1 + R_2 = 0$. Example B.2.1 and Corollary B.3.2 then imply:

$$E(\eta^2|\delta_1 = 1) = \sigma^2 (1 - 2 \, R_2^2)$$

$$+ (\sigma_2^2/\sigma_\varepsilon^2) \, R_2^2 \begin{cases} \dfrac{\theta^2}{P_1}\left[(\pi^2/3) - G\left(\dfrac{(1-P_1)}{P_1}\right)\right] & \text{for } P_1 > 1/2 \\[2ex] \dfrac{\theta^2}{P_1} \, G\left(\dfrac{P_1}{(1-P_1)}\right) & \text{for } P_1 \leqslant 1/2 \end{cases}$$

Using $\sigma_\varepsilon^2 = \pi^2 \, \theta^2/6$ yields the first two parts of the theorem. It is then simple to derive the expression for $E(\eta^2|\delta_1 = 0)$ from:

$$E(\eta^2|\delta_1 = 1) \, P_1 + E(\eta^2|\delta_1 = 0) \, (1-P_1) = E(\eta^2) = \sigma^2 \quad \text{Q.E.D.}$$

If we relax the assumption that $<\varepsilon_i>$ are independently, identically extreme value distributed, then conditional moments of η are not easily derived. Indeed, the strong separability used for the function G in Corollary B.2.2 and Theorem B.3.2 if applied symmetrically to all components of G would imply the simple multinomial logit specification.

The sequential form of the GEV family does, however, provide a tractable alternative. Rather than assume that η has a linear conditional expectation in $<\varepsilon_i>$, we instead assume that η has a linear conditional expectation in the space of the "induced" independent extreme value random variables that generate the conditional branch probabilities. This assumption is motivated by the consideration that the simple multinomial logit probability form is implied by but does not itself imply an independent extreme value error structure. This point is usefully illustrated in the bivariate extreme value distribution:

$$G(y) = \left[y_1^{1/(1-\sigma)} + y_2^{1/(1-\sigma)} \right]^{(1-\sigma)} \tag{53}$$

Equation (53) implies a probability choice system:

$$P_1 = \frac{e^{V_1/\theta(1-\sigma)}}{e^{V_1/\theta(1-\sigma)} + e^{V_2/\theta(1-\sigma)}} \tag{54}$$

Alternatively, consider the independent form of the GEV:

$$G[y] = y_1 + y_2 \tag{55}$$

which implies the multinomial logit probability choice system:

$$P_1 = \frac{e^{V_1/\theta}}{e^{V_1/\theta} + e^{V_2/\theta}} \tag{56}$$

As the scale parameters $\theta(1-\sigma)$ and θ are *not* identified in (54) and (56), the resulting models are observationally equivalent.

Furthermore, in the sequential or nested logit model, we may view the second-level conditional probabilities in equation (47) as being generated by the independent extreme value random variables $<\varepsilon_j^{B_m}>$ with variance $(\pi^2/6)(1-\sigma_m)^2$. Specifically:

$$P[i \mid B_m] = Prob\ [V_i + \varepsilon_i^{B_m} \geqslant V_j + \varepsilon_j^{B_m},\ \forall j \mid j \in B_m \text{ and } j \neq i] \tag{57}$$

Finally, the error structure $<\varepsilon_j^{B_m}>$ may be analyzed through Theorems B.4.1, B.4.2, and B.4.3.