

# Simulation Methods for Nested Logit Models

A. Colin Cameron and Noah (Yong-eok) Kim  
Department of Economics, University of California, Davis, CA 95616

email: [accameron@ucdavis.edu](mailto:accameron@ucdavis.edu)

June 16, 2001

## **Abstract**

The development of simulation estimation methods has led to increased application of the multinomial probit and mixed logit models for multinomial choice data, such as data on choice of recreational site, transportation mode or occupation.

Simulation methods have been rarely applied to the other leading multinomial choice model, the nested logit or generalized extreme value (GEV) model. Notable exceptions are papers by McFadden (1996) and Herriges and Kling (1999) who estimate compensated variation in nested logit or GEV models that are nonlinear in income using simulation methods. These papers use a Markov chain Monte Carlo method to obtain draws from a GEV distribution that is computationally very intensive.

In this paper we propose alternative methods for drawing GEV errors that lead to quicker calculation of compensating variation in nested logit models, though results at this stage are preliminary.

## 1. Introduction

We consider models for unordered categorical data, such as choice of recreational site, transportation mode or occupation. The multinomial logit (MNL) model is particularly simple to estimate and is provided in many statistical packages. Economists emphasize the use of models based on rational consumer choice, and restrict their attention to models based on random utility models (RUM). The MNL model is a RUM. It is viewed as too restrictive, however, due to its assumption of independence of irrelevant alternatives.

This limitation of MNL has led to a considerable literature on more general multinomial models. The multinomial probit (MNP) model is an alternative model that is viewed as potentially the most general model, but its estimation is difficult because of the need to evaluate an integral of increasing dimension as the number of choices increases. The random parameters logit model or mixed logit model or mixed multinomial logit model (MMNL) is a generalization of the MNL that overcomes the independence of irrelevant alternatives and is potentially quite flexible, but again requires evaluation of a high dimensional integral, albeit one whose dimension does not increase with the number of choices. The nested logit model or the generalized extreme value (GEV) model is also a generalization of the MNL model. It has the attraction of being estimated with little more difficulty than the MNL, but is more restrictive than MNP and MMNL and requires imposition of a nesting structure or ordering of decision-making that may not be tenable in some applications.

The development of simulation-based estimation techniques, McFadden (1989) and Pakes and Pollard (1989), has led to an increasing number of empirical applications of the MNP and MMNL models. These simulation methods entail draws from the multivariate normal, which is straightforward to do by transformation to independent normals, though the computational burden is high.

Simulation methods have been only rarely applied to GEV models, because there is no need to use simulation methods to estimate the model and because it is difficult to draw from the multivariate GEV distribution. This gap has been partly filled in a recent paper, McFadden (1995), which used simulation methods to conduct welfare analysis using GEV models. For example, one might wish to place a dollar value, using compensated variation, on the impact of a policy change that changes the characteristics of a recreational site. If income enters the RUM nonlinearly it becomes necessary to use simulation or other numerical methods to compute the compensated variation, even in the easily estimated MNL

and GEV models. McFadden showed how to draw from the GEV distribution using Markov chain Monte Carlo (MCMC) methods, and applied the method in a simple example with artificial data. The method is computationally very intensive. Herriges and Kling (1999) applied this method to a four-choice model with two regressors and 1,182 observations. Computation of a point estimate of mean compensating variation took 15 minutes on a 200 MHz Pentium and computation of confidence intervals took 48 hours.

In this paper we propose alternative methods to obtain draws from the GEV distribution that permit much faster computation of compensating variation. These methods can also be applied to estimation of the GEV model, allowing comparison of simulation methods with conventional estimation methods.

The calculation of compensating variation in nested logit or GEV models is presented in section 2. Methods to generate GEV random variables, including the new method of this paper, are detailed in section 3. In section 4 we present simulations that demonstrate the usefulness of the proposed method. Section 5 gives an application and section 6 concludes.

## 2. Nested Logit Models and Compensating Variation

### 2.1. Nested Logit Models

Standard references include Maddala (1983) and Greene (2000). The consumer chooses one of  $m$  mutually exclusive choices, and we observe the discrete outcome  $y = k$  if the consumer chooses the  $k^{th}$  alternative. This choice will be the one that gives the highest utility to the consumer.

The most commonly-used version of the **random utility model** (RUM) specifies the utility of the  $j^{th}$  choice to be given by

$$U_j = V_j + \varepsilon_j, \quad j = 1, 2, \dots, m, \quad (2.1)$$

where  $V_j = V_j(\mathbf{z}, \boldsymbol{\beta})$  denotes the deterministic part of utility, which is a function of regressors  $\mathbf{z}$  and parameters  $\boldsymbol{\beta}$ , and  $\varepsilon_j$  denotes the random component of utility which is drawn from a distribution of known functional form, though may also be dependent on unknown parameters. The consumer makes the choice that maximizes utility, so that

$$\Pr[y = k] = \Pr[U_k \geq U_j, \text{ for all } j]. \quad (2.2)$$

Different assumptions on the distribution of the errors lead to different specifications of  $\Pr[y = k]$  and hence to different multinomial models. Estimation is by maximum likelihood.

**Estimation** of multinomial models is theoretically straightforward. The density for an observation can be written compactly as

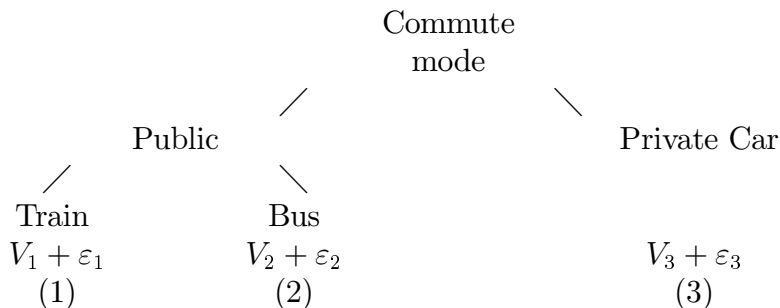
$$f(y) = \prod_{j=1}^m p_j^{d_j}, \quad (2.3)$$

where  $d_j$  is an indicator variable that equals 1 if  $y = j$  and equals 0 otherwise, and  $p_j = \Pr[y = j]$  is of known functional form obtained from (2.2) that depends on regressors and parameters to be estimated. Given a sample of size  $n$  with  $i^{\text{th}}$  observation  $(y_i, \mathbf{z}_i)$  the ML estimator maximizes the log-likelihood

$$\ln L = \sum_{i=1}^n \sum_{j=1}^3 d_{ij} \ln p_{ij}. \quad (2.4)$$

Implementation requires knowledge of the functional forms for the probabilities  $p_j$ .

A leading example for this paper is a **three-choice nested logit model**. An example is transportation mode choice between train, bus and car. These clearly depend on observables such as individual income, direct cost and time cost of travel, but will also depend on characteristics that are not observed. One might believe that the unobserved components  $\varepsilon_1, \varepsilon_2$  and  $\varepsilon_3$  are correlated between train (choice 1) and bus (choice 2), which are both public transit modes, but uncorrelated with car (choice 3) which is a private transit mode. This can be represented in the following decision tree



The nested logit model in this example specifies the correlated errors  $(\varepsilon_1, \varepsilon_2)$  to be bivariate extreme value distributed and the independent error  $\varepsilon_3$  to be extreme value distributed. Thus  $m = 3$  and the errors have cumulative distribution

function

$$\begin{aligned} F(\varepsilon_1, \varepsilon_2) &= \exp(-[e^{-\varepsilon_1 s} + e^{-\varepsilon_2 s}]^{1/s}), \quad s \geq 1, \\ F(\varepsilon_3) &= \exp(-e^{-\varepsilon_3}). \end{aligned} \quad (2.5)$$

The marginal densities of  $\varepsilon_1$  and  $\varepsilon_2$  can be shown to be extreme value, with mean  $\gamma = \Gamma'(1) = .57722$  and variance  $\pi^2/6 = 1.64493$ . The parameter  $s$  will need to be estimated, and is related to the correlation between  $\varepsilon_1$  and  $\varepsilon_2$  according to

$$\text{Corr}[\varepsilon_1, \varepsilon_2] = 1 - (1/s)^2. \quad (2.6)$$

One can show that this three-choice model leads to the following probabilities

$$\begin{aligned} p_1 = \Pr[y = 1] &= \frac{e^{sV_1}}{(e^{sV_1} + e^{sV_2})} \times \Pr[y = 3] \\ p_2 = \Pr[y = 2] &= \frac{e^{sV_2}}{(e^{sV_1} + e^{sV_2})} \times \Pr[y = 3] \\ p_3 = \Pr[y = 3] &= \frac{e^{V_3}}{(e^{sV_1} + e^{sV_2})^{1/s} + e^{V_3}} \end{aligned} \quad (2.7)$$

If  $s = 1$  the model reduces to a three-choice multinomial logit model.

Other models such as the multinomial probit model have a similar structure, except that they usually specify a distribution for the errors  $\varepsilon_j$  that does not lead to an explicit solution for the probabilities  $p_j$  such as (2.7). Instead one is led, in the three-choice case, to solving expressions such as

$$\begin{aligned} \Pr[y = 3] &= \Pr[V_3 + \varepsilon_3 > V_1 + \varepsilon_1, \quad V_3 + \varepsilon_3 > V_2 + \varepsilon_2] \\ &= \Pr[\varepsilon_3 - \varepsilon_1 > V_1 - V_3, \quad \varepsilon_3 - \varepsilon_2 > V_2 - V_3] \end{aligned} \quad (2.8)$$

which is a bivariate integral that needs to be evaluated using numerical or simulation methods.

Nested logit models such as the preceding three-choice model are special cases of the more general class of **generalized extreme value (GEV) models**. This class specifies the joint cumulative distribution of the errors in an  $m$ -choice model to be

$$F(\boldsymbol{\varepsilon}) = \exp[-H(e^{-\varepsilon_1}, \dots, e^{-\varepsilon_m})] \quad (2.9)$$

where the function  $H(Y_1, Y_2, \dots, Y_m)$  satisfies a number of assumptions including nonnegativity, homogeneity of degree one, mixed partial derivatives that are continuous and nonpositive for even order and nonnegative for odd order, and  $\lim_{Y_j \rightarrow \infty} H(Y_1, Y_2, \dots, Y_m) = \infty$ . Analytical expressions for the choice probabilities are possible for this more general class. The preceding three choice nested logit model specifies

$$H(Y_1, Y_2, Y_3) = [Y_1^s + Y_2^s]^{1/s} + Y_3. \quad (2.10)$$

## 2.2. Compensating Variation

Herriges and Kling (1999) give an excellent summary.

We wish to put a dollar value on the effect of changing one or more of the characteristics in the deterministic component of utility, such as price or time cost of travel in the transportation mode choice example. We first need to specify a deterministic part of utility. We specify the indirect utility form

$$V_j = V(y - p_j, \mathbf{x}_j) \quad (2.11)$$

where  $y$  is income,  $p_j$  is the price of the  $j^{\text{th}}$  alternative and  $\mathbf{x}_j$  are characteristics associated with the  $j^{\text{th}}$  alternative<sup>1</sup>. For notational simplicity the unknown regression parameters  $\beta$  are suppressed. Then utility of alternative  $j$  is

$$U_j = U(y - p_j, \mathbf{x}_j, \varepsilon_j) = V(y - p_j, \mathbf{x}_j) + \varepsilon_j. \quad (2.12)$$

Suppose we change the characteristics from  $\mathbf{x}'_j$  to  $\mathbf{x}''_j$ . Then compensating variation  $CV$  is the change in income needed to hold utility at its initial level, so that the highest utility level attainable with income  $y$  and characteristics  $\mathbf{x}'_j$  must equal the highest level attainable with income  $(y - CV)$  and characteristics  $\mathbf{x}''_j$ . Thus **compensating variation**  $CV$  is implicitly defined as the solution to

$$\max_{j=1, \dots, m} U(y - p_j, \mathbf{x}'_j, \varepsilon_j) = \max_{j=1, \dots, m} U(y - CV - p_j, \mathbf{x}''_j, \varepsilon_j). \quad (2.13)$$

As an example, consider a two-choice model where  $U_j = y + x_j + \varepsilon_j$ ,  $j = 1, 2$  and the scalar  $x_j$  changes from  $x'_j$  to  $x''_j$ . Then there are four possibilities. If alternative 1 is chosen before and after then  $CV = (x''_1 - x'_1)$ , since then  $U''_1 = y - CV + x''_1 + \varepsilon_1 = y + x'_1 + \varepsilon_1 = U'_1$ . Similarly if alternative 2 is chosen before and after then  $CV = (x''_2 - x'_2)$ . If switch occurs from alternative 1 to alternative 2 then  $U''_2 = U'_1$  implies  $y - CV + x''_2 + \varepsilon_2 = y + x'_1 + \varepsilon_1$  which implies  $CV = x''_2 - x'_1 + \varepsilon_2 - \varepsilon_1$ . Similarly if switch occurs from alternative 2 to alternative 1 then  $CV = x''_1 - x'_2 + \varepsilon_1 - \varepsilon_2$ . More generally, for  $m$ -choice linear models such as this, the compensating variation is  $CV_{jk} = V''_k - V'_j + \varepsilon_k - \varepsilon_j$  if the change in  $\mathbf{x}$  leads to a change from alternative  $j$  to alternative  $k$ .

The compensating variation depends on observables  $(y, p_j$  and  $\mathbf{x}_j)$ , parameters that can be estimated and unobservable errors  $\varepsilon_j$ . We need to eliminate the unobservables and compute the expected compensating variation  $E[CV]$  which

---

<sup>1</sup>Note that  $y$  here refers to income and should not be confused with the discrete outcome measure  $y$  used in the previous section.

involves integrating over  $\varepsilon_j$ . From the preceding example it should be clear that this integration is quite difficult. Remarkably a closed form solution is possible **if the errors are GEV distributed and income appears linearly**, so that  $V(y - p_j, \mathbf{x}_j) = \alpha(y - p_j) + f(\mathbf{x}_j)$ . This result is given in Section 4.

When income appears **nonlinearly**, however, an exact solution is not possible and one instead needs to numerically integrate over  $\varepsilon_j$  the function for  $CV$  defined in (2.13). This can be simulated in the following way:

1. At iteration  $t$  draw  $\varepsilon^t$  from the distribution of  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_m)$ .
2. Calculate  $CV^t$  from  $\max_{j=1, \dots, m} U(y - p_j, \mathbf{x}'_j, \varepsilon_j^t) = \max_{j=1, \dots, m} U(y - CV^t - p_j, \mathbf{x}''_j, \varepsilon_j^t)$ .
3. Repeat  $T$  times.
4. Estimate  $E[CV]$  by  $\frac{1}{T} \sum_{t=1}^T CV^t$ .

This method is used to obtain  $E[CV]$  for each individual in the sample. Averaging, possibly with weighting, provides a population estimate.

### 3. Methods to Obtain Draws from the GEV

The problem that is at the heart of the McFadden (1995) and Herriges and Kling (1999) papers is that there is no simple way to draw from the GEV distribution for  $\varepsilon$ , except in the special case of product of univariate extreme values which is the special case of MNL. In this section we consider ways to draw from the GEV.

#### 3.1. GEV Distribution

The GEV distribution is defined in (2.9). Different choices of  $H()$  lead to different GEV models, with nested logit models being special cases.

In the special case that  $H(Y_1, Y_2, \dots, Y_m) = Y_1 + Y_2 + \dots + Y_m$  the joint density is just the product of  $m$  independent extreme value random variables with c.d.f.  $F(\varepsilon) = \exp(-e^{-\varepsilon})$  and density  $f(\varepsilon) = e^{-\varepsilon} \exp(-e^{-\varepsilon})$ . This is easily generated using the transformation method - draw  $u$  from the uniform on  $(0, 1)$  and then setting  $\varepsilon = -\ln(-\ln(u))$ .

But for any departure from this, even one as simple as bivariate extreme value, there is no simple way to draw random variables from the GEV. McFadden (1995) proposed using Markov chain Monte Carlo simulation methods.

### 3.2. Markov Chain Monte Carlo

Let  $f(\boldsymbol{\varepsilon})$  denote the joint density of the random variable of interest, and let  $g(\boldsymbol{\varepsilon})$  denote a **sampling density**, distinct from  $f(\boldsymbol{\varepsilon})$ . A **Markov chain Monte Carlo** (MCMC) is used to generate draws from  $f(\boldsymbol{\varepsilon})$ , based on draws from  $g(\boldsymbol{\varepsilon})$ , in the following way (Metropolis-Hastings).

1. At iteration  $t$  draw  $\tilde{\boldsymbol{\varepsilon}}^t$  from the sampling density  $f(\boldsymbol{\varepsilon})$ , and also draw a uniform random variable  $\eta^t$ .
2. Compute

$$\hat{\boldsymbol{\varepsilon}}^t = \begin{cases} \tilde{\boldsymbol{\varepsilon}}^t & \text{if } \eta^t \leq \frac{f(\tilde{\boldsymbol{\varepsilon}}^t)/g(\tilde{\boldsymbol{\varepsilon}}^t)}{f(\hat{\boldsymbol{\varepsilon}}^{t-1})/g(\hat{\boldsymbol{\varepsilon}}^{t-1})} \\ \hat{\boldsymbol{\varepsilon}}^{t-1} & \text{otherwise} \end{cases} \quad (3.1)$$

3. Repeat  $T_0 + T$  times.
4. Discard the first  $T_0$  realizations of  $\hat{\boldsymbol{\varepsilon}}^t$  (burn-in) and use the remaining  $T$  estimates.

A weakness of this method is that it is computationally intensive, as illustrated below.

### 3.3. One-Factor Multivariate EV

We focus on the case where  $\varepsilon_1$  and  $\varepsilon_2$  are bivariate extreme value with joint c.d.f.  $F(\varepsilon_1, \varepsilon_2) = \exp(-[e^{-\varepsilon_1 s} + e^{-\varepsilon_2 s}]^{1/s})$ .

We propose generating  $\varepsilon_1$  and  $\varepsilon_2$  using the following **one-factor multivariate EV model**

$$\begin{aligned} \varepsilon_1 &= a\xi + bv_1 + c \\ \varepsilon_2 &= a\xi + bv_2 + c \end{aligned} \quad (3.2)$$

where  $\xi$ ,  $\nu_1$  and  $\nu_2$  are independent univariate extreme value and  $a$ ,  $b$  and  $c$  are weights. These weights are chosen to reproduce the same first two moments and correlation as for the bivariate extreme value. Thus where  $\gamma = \Gamma'(1) \simeq 0.57722$  we have

$$\begin{aligned} \mathbb{E}[\varepsilon_j] &= \gamma = a\gamma + b\gamma + c \\ \text{Var}[\varepsilon_j] &= \frac{\pi^2}{6} = a^2 \frac{\pi^2}{6} + b^2 \frac{\pi^2}{6} \\ \text{Corr}[\varepsilon_1, \varepsilon_2] &= [1 - (1/s)^2] = \frac{a^2}{a^2 + b^2}. \end{aligned} \quad (3.3)$$

Table 3.1: One-factor extreme value model approximation to the bivariate extreme value. Results of 10,000 draws when  $s=2$ . Sample moments from the simulation are compared to theoretical moments.

	Sample value	Theoretical Value	Difference
$E[\varepsilon_1]$	0.57885	0.57722	0.00163
$E[\varepsilon_2]$	0.58471	0.57722	0.00749
$V[\varepsilon_1]$	1.59151	1.64493	-0.05342
$V[\varepsilon_2]$	1.61917	1.64493	-0.02576
$\text{Cov}[\varepsilon_1, \varepsilon_2]$	0.74160	0.75000	0.00840

Solving (2.9) yields

$$\begin{aligned}
 a &= \sqrt{1 - (1/s)^2} \\
 b &= \sqrt{1 - a^2} \\
 c &= (1 - a - b)\gamma
 \end{aligned}
 \tag{3.4}$$

As an illustration of this method let  $s = 2$ , in which case the correlation between  $\varepsilon_1$  and  $\varepsilon_2$  equals 0.75. Then using we have  $\varepsilon_1 = 0.8660\xi + 0.5\nu_1 - 0.2113$ , and similarly for  $\varepsilon_2$ . A simulation of 10,000 draws of  $\xi$ ,  $\nu_1$  and  $\nu_2$  from the extreme value (using  $-\ln(-\ln(u))$  where  $u$  is uniform) yielded 10,000 draws of  $\varepsilon_1$  and  $\varepsilon_2$  using the one-factor model. From Table 3.1 the bivariate extreme value distribution is approximated quite well. The sample means are within two standard deviations of the theoretical means (as the estimated standard error of the sample means is 0.01260).

The one-factor approximation to the bivariate GEV is not perfect, however. One way to see this is to note that the bivariate GEV has a univariate extreme value marginal whereas the one-factor approximation is not. Figure 3.1 shows the univariate marginal density of  $\varepsilon_1$  that arises from the one-factor approximation with a univariate extreme value density, using the same 10,000 draws as in Table 3.1. The approximation is very good though is not perfect.

The one-factor method generalizes immediately to multivariate extreme densities with  $H(Y_1, Y_2, \dots, Y_m)$  in (2.9) equal to  $[Y_1^s + Y_2^s + \dots + Y_m^s]^{1/s}$ . Thus it can be applied to any two-level nested logit model. Suppose the  $m$  alternatives can be nested into  $L$  subgroups or branches, where the  $l^{\text{th}}$  subgroup has  $m_l$  alternatives

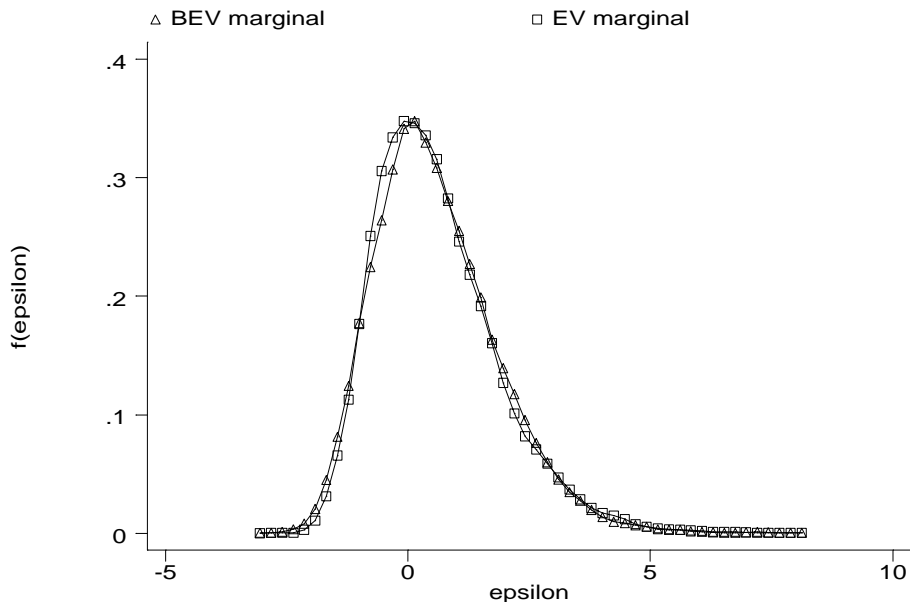


Figure 3.1: Comparison of univariate marginal density of the one-factor approximation to bivariate extreme value (with correlation 0.75) with the univariate extreme value density. Kernel density estimates based on 10,000 simulated draws.

with errors  $\varepsilon_{l1}, \dots, \varepsilon_{lm_l}$  that have GEV density with

$$H(\varepsilon_{l1}, \dots, \varepsilon_{lm_l}) = \exp \left( - \left[ \sum_{j=1}^{m_l} e^{-\varepsilon_{lj} s_l} \right]^{1/s_l} \right), \quad (3.5)$$

so that the errors are pair-wise correlated with correlation  $1 - (1/s_l)^2$ . Then one uses

$$\varepsilon_j = a_l \lambda + b_l v_j + c_l,$$

where  $a_l$ ,  $b_l$  and  $c_l$  are computed using (3.4) with  $s$  replaced by  $s_l$ .

## 4. Simulations

We consider application to an example by McFadden. The motivation for simulation methods is for GEV models nonlinear in income. The example studied

is linear in income, however, since then an exact measure of expected compensated variation is available, permitting comparison of simulation estimates with a known value.

#### 4.1. McFadden's Example

McFadden (1995) considers the specific three-choice example

$$U_j = V_j + \varepsilon_j = (y + x_j) + \varepsilon_j, \quad j = 1, 2, 3, \quad (4.1)$$

where  $x_j$  is scalar,  $\varepsilon_1$  and  $\varepsilon_2$  are bivariate EV and  $\varepsilon_3$  is EV.

We consider the welfare effect of a change in  $\mathbf{x} = (x_1, x_2, x_3)$  from  $\mathbf{x}' = (x'_1, x'_2, x'_3) = (0, 0, 0)$  to  $\mathbf{x}'' = (x''_1, x''_2, x''_3) = (0.2, 0, 0)$ .

For this design the only possible improvements in utility are to switch to (or remain with) alternative 1 since this is the only alternative whose utility has been changed (and it has increased). If alternative 1 is preferred beforehand it will remain preferred afterwards and  $CV = 0.2$ . If alternative 2 is preferred beforehand and one switches to alternative 1 then  $CV = 0.2 + (\varepsilon_1 - \varepsilon_2)$  since then  $y + \varepsilon_2 = y + 0.2 - CV + \varepsilon_1$ . Similarly a switch from alternative 3 to alternative 1 has  $CV = 0.2 + (\varepsilon_1 - \varepsilon_3)$ .

It is shown in the appendix that in this case an exact expression for expected compensating variation is possible, with

$$E[CV] = \ln \left[ (e^{0.2s} + 1)^{1/s} + 1 \right] - \ln [2^{1/s} + 1]. \quad (4.2)$$

#### 4.2. Simulation Estimates

Table 4.1 reports results given in Tables 1 and 2 of McFadden (1995).

The first five columns summarize the experimental design. The bivariate EV parameter  $s$  is set to 1, 2 and 10 corresponding to correlations between  $\varepsilon_1$  and  $\varepsilon_2$  of 0.0, 0.75 and 0.99. The exact probabilities before and after are given in the third and fourth columns. The exact  $E[CV]$  computed by (4.2) is given in the fifth column. Note that the GEV model parameters are set to their true values rather than estimated.

The remaining three columns give the results of computation of  $E[CV]$  using the simulation method of section 2.2. The draws from  $f(\boldsymbol{\varepsilon})$  are obtained using the MCMC method in section 3.2 with sampling density being the product of three independent EV random variables. For one simulation the Markov chain (after

Table 4.1: McFadden (1995) Example. Three choices and linear in income :  $V_j = y + x_j + e_j$ . First two errors are bivariate extreme value and third is extreme value.  $x$  changes from  $(0, 0, 0)$  to  $(0.2, 0, 0)$ . MCMC estimates of  $E[CV]$  are from 1,000 replications where each replication has chain of length 5,000. The sample mean, percentage deviation of this sample mean from the true WTP, and root mean square error for these 1,000 replications are given in the last three columns.

s	$\rho$	Before			After			Exact $E[CV]$	MCMC Estimates of $E[CV]$		
		$p1'$ , $p2'$ , $p3'$	$p1''$ , $p2''$ , $p3''$			T=5000	% Error		RMSE		
1	0.0	0.333, 0.333, 0.333	0.379, 0.310, 0.310		0.0712046	0.07120	0.00%	0.0014			
2	0.75	0.293, 0.293, 0.414	0.367, 0.246, 0.388		0.0658542	0.06579	-0.11%	0.0028			
10	0.99	0.483, 0.259, 0.259	0.487, 0.066, 0.477		0.0767327	0.07640	-0.43%	0.0075			

burn-in) is of length  $T = 5,000$  leading to 5,000 draws of  $\varepsilon$  and 5,000 values of  $CV^t$ .<sup>2</sup> Then  $E[CV]$  is estimated by  $\frac{1}{T} \sum_{t=1}^{5000} CV^t$ . This procedure is repeated 1,000 times giving 1,000 estimates of  $E[CV]$ . The last three columns of Table 4.1 give the sample mean, the percentage deviation (% error) of this from the true  $E[CV]$ , and the empirical root mean square error for these 1,000 estimates. The MCMC approximation does well on average, though is by no means exact for  $s \neq 0$  considering that the average is based on 5 million computations of  $E[CV]$ . The RMSE is relatively large, being 1.94%, 4.28% and 9.84% of exact  $E[CV]$  for  $s$  equal to 1, 2 and 10.

An alternative method is to use the same simulation methods as given in section 2.2, except draw the bivariate EV errors from the one-factor EV model. Thus  $\varepsilon_1 = a\lambda + bv_1 + c$ ,  $\varepsilon_2 = a\lambda + bv_2 + c$ , where  $\lambda$ ,  $v_1$  and  $v_2$  are draws from EV, and  $\varepsilon_3$  is a draw from EV. Results are given in the last three columns of Table 4.2. For one simulation there were 3,000 draws of  $\varepsilon$ , somewhat less than the 5,000 using the MCMC method as here the vector  $\varepsilon$  is independent whereas in the MCMC method the vector  $\varepsilon$  is correlated. There were only 100 simulations, which may effect the closeness of the overall estimated average to the true  $E[CV]$  but will have little effect on the RMSE. Clearly the method has less noise (lower RMSE) than the MCMC method. And the method is on average remarkably accurate for large  $s$ .

<sup>2</sup>McFadden (1995) also considered  $T = 500$ ,  $T = 1000$  and  $T = 10000$ .

Table 4.2: McFadden (1995) Example: One-factor GEV estimates of  $E[CV]$ . Same set up as Table 4.1 Columns 4 to 6 reproduce MCMC results in Table 4.1. Columns 7 to 9 give results when instead approximate using one-factor GEV.

s	$\rho$	Exact $E[CV]$	MCMC Estimates (T=5000)			One-Factor Estimates		
			Average	% Error	RMSE	Average	% Error	RMSE
1	0.0	0.07120	0.07116	-0.06%	0.0014	0.07138	0.19%	0.0019
2	0.75	0.06585	0.06592	0.10%	0.0028	0.06610	0.38%	0.0015
10	0.99	0.07673	0.07689	0.21%	0.0075	0.07673	-0.01%	0.0013

## 5. Application

Herriges and Kling (1999) studied choice between four types of fishing sites - beach, pier, charter boat and private boat - using nested logit models. There were 1,182 observations with regressors income ( $y$ ), prices ( $p_j$ ) and catch rates ( $x_j$ ). They estimated a translog model with

$$U_j = \beta_{10} \ln(y - p_j) + \beta_{20} \ln(x_j) + \beta_{11} (\ln(y - p_j))^2 + \beta_{22} (\ln x_j)^2 + \beta_{12} \ln(y - p_j) \ln x_j + \varepsilon_j, \quad j = 1, \dots, 4,$$

as well as a linear model and a generalized Leontief model. Welfare effects were calculated for doubling the price of each alternative fishing mode, doubling the catch rate at all sites, and eliminating beach and pier fishing.

Expected compensated variation was calculated for each observation using the MCMC simulation with chain length  $T = 1000$  and then averaged over the sample. Computational time for each model and change was fifteen minutes on a Pentium 100. Unlike the McFadden (1995) example the parameters here were first estimated. To account for the sampling error in these estimates ( $\beta$  and  $s$ ) five hundred coefficient vectors were drawn from the asymptotic distribution of  $(\hat{\beta}, \hat{s})$  and  $E[CV]$  was then computed 500 times. The middle 95% of these 500 estimates provides a 95% confidence interval. Total computational time was forty-eight hours. It is hoped that the one-factor EV model may permit much faster computation.

## **6. Conclusions**

At this stage the one-factor model appears to give a remarkably good approximation to expected compensating variation. The closeness of the approximation needs to be tested further over a wider range of GEV models. But it is expected that the approximation will continue to perform well, certainly much better than other approximations such as using a representative consumer approach or using bounds presented in McFadden (1995) and Herriges and Kling (1999).

## 7. Appendix: Derivation of (4.2)

For GEV models that are linear in income,  $V(y - p_j, \mathbf{x}_j) = \alpha(y - p_j) + f(\mathbf{x}_j)$ , McFadden (1995) and earlier papers show there is an explicit solution

$$E[CV] = \frac{1}{\alpha} \left( \ln H(e^{V_1''}, \dots, e^{V_m''}) - \ln H(e^{V_1'}, \dots, e^{V_m'}) \right), \quad (7.1)$$

where the function  $H()$  from the GEV distribution is defined in (2.10) and  $V_j'$  and  $V_j''$  are the before and after values of the deterministic component of utility.

For the current example (4.1)

$$\begin{aligned} \ln H(e^{V_1'}, e^{V_2}, e^{V_3'}) &= \ln \left[ \left( (e^{V_1'})^s + (e^{V_2})^s \right)^{1/s} + e^{V_3'} \right] \\ &= \ln \left[ \left( e^{s(y+x_1)} + e^{s(y+x_2)} \right)^{1/s} + e^{s(y+x_3)} \right] \\ &= \ln \left[ e^y \left\{ (e^{sx_1} + e^{sx_2})^{1/s} + e^{sx_3} \right\} \right] \\ &= y + \ln \left[ (e^{sx_1} + e^{sx_2})^{1/s} + e^{sx_3} \right]. \end{aligned}$$

Then using (7.1) yields

$$\begin{aligned} E[CV] &= \frac{1}{\alpha} \left( \ln H(e^{V_1'}, \dots, e^{V_m'}) - \ln H(e^{V_1''}, \dots, e^{V_m''}) \right) \\ &= \ln \left[ \left( e^{sx_1''} + e^{sx_2''} \right)^{1/s} + e^{x_3''} \right] - \ln \left[ \left( e^{sx_1'} + e^{sx_2'} \right)^{1/s} + e^{x_3'} \right]. \end{aligned}$$

Setting  $(x_1', x_2', x_3') = (0, 0, 0)$  and  $(x_1'', x_2'', x_3'') = (0.2, 0, 0)$  yields (4.2).

## 8. References

- Greene, W.H. (2000), *Econometric Analysis*, 4th edition, New York, MacMillan.
- Herriges, J.A. and C.L. Kling (1999), "Nonlinear Income Effects in Random Utility Models," *Review of Economics and Statistics*, 81(1), 62-72.
- Johnson, N.L. and S. Kotz (1972), *Continuous Multivariate Distributions*, New York, John Wiley.
- Maddala, G.S. (1983), *Limited-Dependent and Qualitative Variables in Economics*, Cambridge, Cambridge University Press.
- McFadden, D. (1995), "Computing Willingness-to-Pay in Random Utility Models," Department of Economics, University of California - Berkeley, December.
- McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica*, 57, 995-1026.
- Pakes, A.S. and Pollard, D. (1989), "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57, 1027-1058.
- Robert, C.P. and G. Casella (1999), *Monte Carlo Statistical Methods*, New York, Springer-Verlag.