UNIVERSITY OF LEEDS

This is a repository copy of *Asymmetric triangular mixing densities for mixed logit models*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/105575/

Version: Accepted Version

Article:
Dekker, T orcid.org/0000-0003-2313-8419 (2016) Asymmetric triangular mixing densities for mixed logit models. Journal of Choice Modelling, 21. pp. 48-55. ISSN 1755-5345

https://doi.org/10.1016/j.jocm.2016.09.006

© 2016, Elsevier. This is an author produced version of a paper published in Journal of Choice Modelling. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Asymmetric triangular mixing densities for mixed logit models

October 5, 2016

Abstract

A novel method is proposed to estimate random parameter logit models using the asymmetric triangular distribution to describe unobserved preference heterogeneity in the population of interest. The asymmetric triangular mixing density has the potential to overcome behavioural limitations associated with the most frequently applied mixing densities like the normal and log-normal distribution. With only three parameters it remains parsimonious whilst its bounded support can easily be brought in line with behavioural intuitions. The triangular mixing density is not associated with an incredibly large upper (or lower) bound and it can accommodate varying degrees of skewness in unobserved preference heterogeneity. The proposed estimation procedure is based on the principle of mixture densities and circumvents additional simulation chatter arising when applying the inverse cumulative density function method to generate draws from the mixing density.

Keywords: Mixed logit, Mixing density, Asymmetry, Triangular Distribution

1 Introduction

The mixed logit model (MIXL), also referred to as the random parameters logit (RPL) model, represents one of the most popular econometric models to analyse discrete choice type data. Advantages of the MIXL model include i) the ability to model heterogeneity in the patterns of choices across respondents[1]; ii) non-constant error variances across alternatives via a relaxation of the independently and identically distributed error terms assumption; and iii) the potential accommodation of correlation in choices across repeated choice observations by the same respondent (e.g Hensher and Greene, 2003; Scarpa et al., 2005). Given its multi-functionality, Keane and Wasi (2013) acknowledge the MIXL model hosts an infinite number of alternative model specifications varying in the number and selection of alternative mixing densities.

The current paper adds the asymmetric triangular density to the set of potential mixing densities available to the analyst. By being able to control for skewness in the distribution of preferences over the population of interest, the asymmetric triangular density is more flexible than its symmetric counterpart. Its bounded support at both end of the distribution makes it a particularly attractive density relative to the more frequently used (log-)normal density. The asymmetric triangular density thereby answers the call of Hensher and Greene (2003) for the implementation of simple, but flexible distributional forms complying with behavioural expectations. So far this call mainly resulted in the adoption of the (constrained) symmetric triangular distribution (e.g. Brouwer et al., 2010; Hensher and Greene, 2003).

The focus of this paper is on the development of a maximum simulated likelihood (MSL) estimation method for the asymmetric triangular density. The proposed estimation method is based on the principle of mixing densities and recognizes that any triangular density can be described by means of two one-sided triangular densities with a common mode. A Bayesian estimation procedure was already developed by Dekker and Rose (2011).

The structure of the paper is as follows. Section 2 introduces the MIXL model and defines the triangular density. Section 3 then develops the MSL estimation framework. Section 4 presents a Monte Carlo simulation and Section 5 concludes the paper .

2 Model structure

2.1 The random coefficients multinomial logit model

Suppose individual n is presented with J alternatives in choice task $t = 1, \ldots, T$. The Random Utility Maximisation model postulates that the individual selects the alternative

[1] Whilst it is common to interpret the random parameter coefficients as representing purely preference heterogeneity, the confoundment between scale and preference parameters in most discrete choice models implies that any modelled heterogeneity should more correctly be interpreted as representing a mixture of both preference and scale or error heterogeneity (Hess and Rose, 2012)

with the highest level of utility, i.e. $y_{nt} = i$ when $U_{nit} > U_{njt}, \forall j \neq i \in J$. Utility U_{nit} is decomposed in a structural part V_{nit} and a stochastic part ϵ_{nit}, where $U_{nit} = V_{nit} + \epsilon_{nit}$. After assuming that ϵ_{nit} follows a type-I extreme value distribution the choice probability for alternative i can be described by:

$$P_{nit} = \frac{exp(V_{nit})}{\sum_{j=1}^{J} exp(V_{njt})} \quad (1)$$

V_{nit} is characterised by a linear utility function $V_{nit} = X_{nit}\beta_n$. Let X_{nit} represent a set of exogenous variables and β_n defines the vector of marginal utility parameters. The subscript in β_n denotes marginal utility may vary across respondents. In most applications, an insufficient number of observations per respondent is available to estimate individual specific utility parameters. Hence, random coefficients are used to capture the heterogeneity in β_n across the population of interest. Let $f(\beta_n|\Omega)$ denote a mixing density function describing the distribution of marginal utility over the population of interest, where Ω is the vector of associated hyper-parameters. The expected choice probability of observing the sequence of choices y_n can then be described by the individual specific likelihood L_n:

$$L_n = \int_{\beta_n} \prod_{t=1}^{T} \frac{exp(X_{nit}\beta_n)}{\sum_{j=1}^{J} exp(X_{njt}\beta_n)} f(\beta_n|\Omega) d\beta_n \quad (2)$$

2.2 The triangular distribution

In this paper, $f(\beta_n|\Omega)$ is described by a triangular density. The density is a function of only three hyper-parameters being respectively the lower-bound a, the upper-bound b and the mode c. These three hyper-parameters define the density function:

$$f(\beta_n|a,b,c) = \begin{cases} \frac{2(\beta_n - a)}{(b-a)(c-a)} & \text{for } a \leq \beta_n \leq c \\ \frac{2(b - \beta_n)}{(b-a)(b-c)} & \text{for } c \leq \beta_n \leq b \end{cases} \quad (3)$$

When $\beta_n < a$ or $\beta_n > b$ the density $f(\cdot)$ will be zero. In short, the triangular distribution qualifies as a mixing density that is simple but flexible in shape and easily complies with behavioural expectations. Namely, the flexible mode of the triangular distribution allows for both positively- and negatively-skewed distributions, but also symmetry by setting $(c - a) = (b - c)$.[2] The support of the distribution can be constrained by fixing either the lower- or the upper-bound or both. Accordingly, the triangular distribution can accommodate non-negative (or non-positive) marginal utilities without inducing a fat upper-tail.

[2] Note that by drawing a straight line from the density at the mode to the zero density at the bounds the share of the population is decreasing at a constant rate when moving away from the mode.

3 Maximum Simulated Likelihood Estimation

3.1 The inverse cdf problem

The principles of MSL require that the simulated density can be obtained by means of rescaling and relocating a standard shape of the underlying distribution. For example, a normal distribution can be by simulated by taking draws from a standard normal distribution, which are subsequently relocated by the estimated mean and rescaled by the estimated standard deviation. For the asymmetric triangular density, draws from $f(\beta_n|\Omega)$ can be generated using an inverse cumulative density function (cdf) transformation approach for a given a, b, c (see (4)), where U_n^r represents a draw r from the standard uniform distribution defined over $[0, 1]$ for individual n.

$$\begin{aligned} \beta_n^r &= a + \sqrt{U_n^r(b-a)(c-a)} \text{ for } U_n^r < \frac{c-a}{b-a} \\ & \quad b - \sqrt{(1-U_n^r)(b-a)(b-c)} \text{ for } U_n^r \geq \frac{c-a}{b-a} \end{aligned} \tag{4}$$

The described inverse cdf approach, however, introduces additional chatter in the simulation. During each optimization iteration the values for a, b, and c adjust, implying that the number of draws assigned to the first (and second) part of (4) change. The accuracy by which the right- and left-hand side of the triangular density are approximated thereby varies at each iteration and potentially causes numerical difficulties.[3]

3.2 Using a mixture of densities

To work around this issue, a simulation approach comparable to the mixtures of normal densities (e.g. Fosgerau and Hess, 2009) is proposed. It is easily recognized that the asymmetric triangular distribution can be constructed by means of two one-sided triangular densities with a common mode c. The first one-sided density has its lower bound at a and its upper bound is equivalent to its mode c. For the second one-sided triangular density, c describes the mode and the lower bound while b describes its upper bound. Draws from both distributions can be generated independently using (5). Where U_{1n}^r and U_{2n}^r represent draws from two independent standard uniform distributions and β_{1n}^r and β_{2n}^r the associated transformations following from the inverse cdf method.

$$\begin{aligned} \beta_{1n}^r &= a + (c-a)\sqrt{U_{1n}^r} \\ \beta_{2n}^r &= b - (b-c)\sqrt{U_{2n}^r} \end{aligned} \tag{5}$$

Since both sides of the distribution are generated independently, the problem of additional simulation chatter no longer prevails. The two independent densities, however, still

[3]Additional chatter is not arising with the symmetric triangular density. 50% of its mass is always situated on each side of the mode.

need combining into a single density for which the mass is normalized to unity. This is accomplished by respectively assigning the weights $\frac{c-a}{b-a}$ and $\frac{b-c}{b-a}$, i.e. the share of mass assigned to the left and right-hand side of the mode. The simulated likelihood function is accordingly described by (6).

$$L_n = \left(\frac{c-a}{b-a}\right)\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}\frac{exp(X_{nit}\beta_{1n}^r)}{\sum_{j=1}^{J} exp(X_{njt}\beta_{1n}^r)} + \left(\frac{b-c}{b-a}\right)\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}\frac{exp(X_{nit}\beta_{2n}^r)}{\sum_{j=1}^{J} exp(X_{njt}\beta_{2n}^r)} \quad (6)$$

A step by step description of the simulation procedure is provided below:

1. For individual n generate two independent sets of R draws from a standard uniform distribution. Label these sets of draws respectively as U_{1n} and U_{2n}.
2. Transform U_{1n} and U_{2n} into draws for respectively β_{1n} and β_{2n} using (5).
3. Evaluate the multinomial logit choice probability for all choices $t = 1, \ldots, T$ made by individual n at each draw β_{1n}^r.
4. Multiply the outcomes of step 3 across the T choices made by individual n and subsequently average across the R draws.
5. Repeat steps 3 and 4 for β_{2n}.
6. L_n is then a weighted average of steps 4 and 5 using $\frac{c-a}{b-a}$ and $\frac{b-c}{b-a}$ as weights.
7. Repeat steps 1 to 6 for each individual.
8. Take the logarithm of each L_n and sum over all respondents for the log-likelihood.

3.3 Alternative parameterisations

Estimations are conducted in Ox (Doornik and Ooms, 2006). Codes are available upon request and easily transferable to other software packages. To avoid restrictions on the parameters of interest during estimation, it is common practice for the symmetric triangular density to estimate the mode c and the log of the spread such that $exp(s) = (c-a) = (b-c)$. A natural extension is the estimation of two spread parameters for the asymmetric triangular density. Empirical exploration revealed that directly estimating i) the mode and upper and lower bounds, or ii) the mode, mean μ and standard deviation σ of the triangular density[4] may reduce the correlation patterns between parameter estimates, but more often results in failure of the estimation routine and higher standard errors on the bounds of the triangular density . This paper therefore uses the parameterisation based on the mode and spreads parameters. Analytical gradients are provided in A.

[4]The lower and upper bound are then defined by $a = \frac{(c-3\mu)-\sqrt{-3(c^2+\mu^2)+6\mu c+24\sigma^2}}{2}$ and $b = 3\mu - a - c$.

4 Monte Carlo simulations

A simple way to verify the proposed estimation procedure, whilst avoiding additional simulation chatter, is to estimate a model using the symmetric triangular density and contrast it against the inverse cdf method, which is known to work in this context. A simulated dataset of $N = 1,000$ respondents each making $T = 10$ choices over $J = 3$ randomly generated alternatives is generated. The scale parameter of the additive error term is normalised to one. The attribute values for a single explanatory variable are drawn from a standard normal distribution. The corresponding preference parameter β_n follows a symmetric triangular density with $c = 0$ and $s = ln(4)$. Table 1 reveals the two methods provide nearly identical results. The negligible differences are numerical and caused by the additional set of Halton draws applied in the mixing approach.[5]

Table 1: Verification of the proposed estimation procedure using simulated data

	Inv. Cdf		Mixing approach	
	Estimate	t-ratio	Estimate	t-ratio
mode	0.0172	0.297	0.0171	0.295
ln(spread)	1.3666	41.90	1.3667	41.90
LL	-9,012.38		-9,012.36	
obs	10,000		10,000	
n	1,000		1,000	

A full set of Monte Carlo simulations, based on the same data structure, is then performed contrasting the performance of the inverse cdf and mixing approach. For each of in total seven model specifications, 50 datasets are generated. Each dataset takes a unique set of random draws from the error term. The hyper parameters of the triangular density structurally vary across the model specifications controlling the degree of skewness.

Table 2 describes the results for the symmetric density as defined above. The table includes estimation results for the symmetric triangular density to act as a point of reference. As expected, the model fit is highly comparable, but on average slightly better when using the asymmetric triangular density. This is a direct result of including an additional parameter in the model. The asymmetric estimation procedure is more data intensive as reflected by the increase in the standard errors on all parameters, in particular the mode. This also causes a slight bias and more variation in the actual parameter estimates across the 50 datasets for the asymmetric triangular density. On average, the size of the bias is limited and comparable between the inverse cdf and the mixing approach.

The subsequent six model specifications all keep the mode at zero whilst shifting both bounds by one unit at a time to the left or right. As such, the [*lower*,*upper*] bounds range between [-7,1] and [1,7]. Tables 3 and 4 reflect that estimation of the mode remains somewhat of an issue. Especially when one of the bounds moves close to the mode, the mode tends to be drawn somewhat to that bound and the bias of the mode and that

[5] Sets of 1,000 Halton draws are used in all estimations for each mixing component.

Table 2: Monte Carlo simulations for 50 datasets using the symmetric triangular density

		Symmetric		Inverse CDF		Mixing approach	
		value	st. error*	value	st. error*	value	st. error*
LL	average	-8975.72		-8975.47		-8975.47	
	5%	-9063.37		-9062.86		-9062.84	
	95%	-8886.73		-8886.58		-8886.57	
mode	average	-0.003	0.059	0.0316	0.2410	0.0306	0.2406
	5%	-0.033	0.057	-0.1938	0.1859	-0.1939	0.1862
	95%	0.040	0.061	0.2857	0.3137	0.2847	0.3139
lower	average	-4.014	0.021	-4.0313	0.0401	-4.0310	0.0401
	5%	-4.146	0.019	-4.2428	0.0325	-4.2425	0.0325
	95%	-3.870	0.023	-3.8222	0.0495	-3.8219	0.0496
upper	average	4.009	0.020	3.9855	0.0408	3.9859	0.0407
	5%	3.844	0.018	3.7620	0.0330	3.7627	0.0330
	95%	4.167	0.023	4.2191	0.0490	4.2191	0.0492

* These are summary statistics for the st. errors, not st. errors of the reported values.

bound slightly increase. Estimation of the bounds, however, remains reasonably accurate although the standard errors tend to increase slightly with the degree of asymmetry in the distribution. Despite this minor issue, these Monte Carlo simulations illustrate the asymmetric triangular distribution can be used as an alternative mixing density for the mixed logit model.

Table 3: Monte Carlo simulations for 50 datasets using positive skewed triangular densities

		Bounds [-3,5]				Bounds [-2,6]				Bounds [-1,7]			
		Inverse cdf		Mixing approach		Inverse cdf		Mixing approach		Inverse cdf		Mixing approach	
		value	st. error	value	st. error	value	st. error	value	st. error	value	st. error	value	st. error
LL	average	-8836.374		-8836.35		-8416.39		-8416.35		-7653.25		-7653.24	
	5%	-8909.503		-8909.49		-8506.42		-8506.41		-7726.82		-7726.81	
	95%	-8755.417		-8755.44		-8326.74		-8326.75		-7573.95		-7573.87	
mode	average	-0.042	0.252	-0.042	0.252	-0.059	0.232	-0.059	0.231	-0.092	0.231	-0.086	0.223
	5%	-0.286	0.197	-0.286	0.198	-0.355	0.173	-0.355	0.172	-0.488	0.047	-0.480	0.060
	95%	0.240	0.333	0.239	0.332	0.153	0.314	0.151	0.312	0.126	0.357	0.130	0.304
lower	average	-2.982	0.033	-2.982	0.033	-1.959	0.028	-1.960	0.027	-0.893	0.040	-0.898	0.033
	5%	-3.220	0.027	-3.221	0.026	-2.179	0.021	-2.178	0.021	-1.124	0.002	-1.124	0.004
	95%	-2.770	0.046	-2.769	0.046	-1.741	0.047	-1.741	0.044	-0.515	0.085	-0.533	0.065
upper	average	5.024	0.052	5.024	0.052	6.012	0.062	6.012	0.062	6.984	0.074	6.983	0.074
	5%	4.797	0.045	4.797	0.045	5.704	0.052	5.704	0.052	6.721	0.065	6.719	0.064
	95%	5.357	0.063	5.358	0.063	6.286	0.075	6.284	0.074	7.315	0.086	7.317	0.085

∞

Table 4: Monte Carlo simulations for 50 datasets using negative skewed triangular densities

		Bounds [-5,3]				Bounds [-6,2]				Bounds [-7,1]			
		Inverse cdf		Mixing approach		Inverse cdf		Mixing approach		Inverse cdf		Mixing approach	
		value	st. error	value	st. error	value	st. error	value	st. error	value	st. error	value	st. error
LL	average	-8827.68		-8827.69		-8390.17		-8390.19		-7616.27		-7616.27	
	5%	-8899.51		-8899.52		-8458.76		-8458.75		-7707.00		-7707.05	
	95%	-8765.68		-8765.71		-8294.21		-8294.20		-7537.79		-7537.79	
mode	average	0.028	0.238	0.028	0.239	-0.021	0.234	-0.020	0.234	0.086	0.292	0.088	0.282
	5%	-0.204	0.190	-0.204	0.190	-0.361	0.187	-0.362	0.187	-0.318	0.102	-0.317	0.047
	95%	0.254	0.308	0.255	0.309	0.331	0.297	0.333	0.300	0.505	0.469	0.516	0.494
lower	average	-5.035	0.050	-5.035	0.051	-5.970	0.063	-5.970	0.063	-7.019	0.078	-7.019	0.078
	5%	-5.241	0.044	-5.241	0.044	-6.263	0.053	-6.263	0.053	-7.371	0.069	-7.371	0.069
	95%	-4.831	0.058	-4.830	0.058	-5.618	0.074	-5.618	0.074	-6.663	0.089	-6.663	0.088
upper	average	2.980	0.032	2.980	0.032	1.999	0.026	1.999	0.026	0.923	0.091	0.920	0.071
	5%	2.779	0.025	2.779	0.025	1.767	0.020	1.767	0.020	0.552	0.008	0.539	0.002
	95%	3.187	0.043	3.187	0.043	2.211	0.038	2.211	0.038	1.153	0.168	1.153	0.198

The amount of simulation chatter appears negligible between the inverse cdf and mixing approach. To investigate this further, the number of draws is systematically decreased for the [-7,1] dataset whilst comparing the performance of both estimators.[6] Table 5 highlights that at lower numbers of draws the MSL procedure becomes less accurate and differences start to arise from the log-likelihood values at 1,000 draws when using the mixing approach. The Root Means Square Difference (RMSD) for the log-likelihoods shows more rapidly increasing degrees of simulation chatter for the inverse cdf relative to the mixing approach. Not surprisingly, this chatter mainly affects the estimation of the upper bound and the mode, i.e. the short-end of the asymmetric triangular density. The bias and RMSD for the lower bound, relative to the true model parameter of -7, stay fairly constant, also at a lower number of draws in both approaches. The bias and RMSD on the mode and upper bound, however, increase more rapidly for the inverse cdf.

5 Conclusions

The Monte Carlo simulations illustrate that the asymmetric triangular density can be added to the toolbox of the discrete choice modeller. The recommended mixing approach relies on a mixture of densities. Its application will therefore have an impact on estimation time. Namely, when z random parameters are assumed to follow an asymmetric triangular density then 2^z 'classes' of respondents can possibly formed. This rapidly increases the number of times the likelihood function needs to be evaluated. The Monte Carlo simulations, however, reveal that the amount of simulation chatter associated with the more traditional and quicker inverse cdf approach is limited, even at a reasonably low number of draws. More draws may, however, be required when the underlying density is heavily skewed or when complex models are estimated with multiple random parameters.

The data requirements of the asymmetric triangular density are higher than those for its symmetric counterpart, or other two-parameter densities such as the normal and log-normal density. Increasing degrees of correlation between the parameter estimates and limited empirical identification are not uncommon when estimating more complex densities such as the Johnson-SB density or off-set parameters for the log-normal density(e.g. Train and Sonnier, 2005). In empirical applications it may therefore be useful to fix one of the bounds. This is not controversial since in many empirical applications we which to restrict β_n to a particular domain. The asymmetric triangular density might, however, not be the best choice when modelling heterogeneity in cost sensitivities. Daly et al. (2012) point out that no moments of the willingness-to-pay density exist when the upper bound is strictly positive. Moreover, when fixing the upper bound at zero only the mean will be defined, but not any higher moments.

The contributions of this paper hopefully spur the empirical application of what has extensively been discussed as a potentially attractive distribution for mixed logit models.

[6] The number of draws used should not be used as a benchmark. Empirical datasets may require more draws to ensure stability of the likelihood function. This is especially the case when using multiple random parameters or more complex model specifications.

Table 5: Simulation chatter at lower numbers of draws

		Mixing approach	Inverse CDF approach				Mixing approach			
		1,000 draws	500 draws	250 draws	200 draws	100 draws	500 draws	250 draws	200 draws	100 draws
LL	average	-7616.27	-7616.25	-7616.21	-7616.46	-7616.75	-7616.24	-7616.24	-7616.25	-7616.31
	average diff		0.0192	0.0627	-0.1827	-0.4773	0.0320	0.0372	0.0248	-0.0353
	RMSD		0.0977	0.1978	0.2541	0.6266	0.0486	0.1314	0.0896	0.1899
mode	bias	0.088	0.0857	0.0863	0.0999	0.1216	0.0837	0.0933	0.0782	0.0691
	RMSD	0.265	0.2653	0.2713	0.2781	0.2938	0.2586	0.2696	0.2579	0.2443
lb	bias	0.019	0.0183	0.0185	0.0217	0.0272	0.0188	0.0211	0.0163	0.0150
	RMSD	0.224	0.2238	0.2242	0.2249	0.2261	0.2235	0.2254	0.2207	0.2166
ub	bias	0.080	0.0780	0.0786	0.0877	0.1026	0.0756	0.0833	0.0722	0.0643
	RMSD	0.204	0.2046	0.2106	0.2154	0.2268	0.1976	0.2060	0.1988	0.1873

References

Brouwer, R., Dekker, T., Rolfe, J., and Windle, J. (2010). Choice certainty and consistency in repeated choice experiments. *Environmental and Resource Economics*, 46(1):93–109.

Daly, A., Hess, S., and Train, K. (2012). Assuring finite moments for willingness to pay in random coefficient models. *Transportation*, 39(1):19–31.

Dekker, T. and Rose, J. (2011). Shape shifters: simple asymmetric mixing densities for mixed logit models. VU University Amterdam. IVM Working Paper: IVM 11/01.

Doornik, J. and Ooms, M. (2006). *Introduction to Ox*. Timberlake Consultants.

Fosgerau, M. and Hess, S. (2009). A comparison of methods for representing random taste heterogeneity in discrete choice models. *European Transport*, 42(1):1–25.

Hensher, D. and Greene, W. (2003). The mixed logit model: The state of practice. *Transportation*, 30(2):133–176.

Hess, S. and Rose, J. (2012). Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation*, 39(6):1225–1239.

Keane, M. and Wasi, N. (2013). Comparing alternative models of heterogeneity in consumer choice behavior. *Journal of Applied Econometrics*, 28(6):1018–1045.

Scarpa, R., Ferrini, S., and Willis, K. (2005). Performance of error component models for status-quo effects in choice experiments. In Scarpa, R. and Alberini, A., editors, *Applications of Simulation Methods in Environmental and Resource Economics*, volume 6 of *The Economics of Non-Market Goods and Resources*, pages 247–273. Springer Netherlands.

Train, K. and Sonnier, G. (2005). Mixed logit with bounded distributions of correlated partworths. In Scarpa, R. and Alberini, A., editors, *Applications of Simulation Methods in Environmental and Resource Economics*, volume 6 of *The Economics of Non-Market Goods and Resources*, pages 117–134. Springer Netherlands.

A Gradients

This appendix derives the gradients for the proposed mixture of one-sided triangular densities. The first thing to note is that the weights assigned to the one-sided triangular density are independent of the mode c and can be rewritten respectively to $\frac{exp(s_1)}{exp(s_1)+exp(s_2)}$ and $\frac{exp(s_2)}{exp(s_1)+exp(s_2)}$. This results in the following partial derivatives:

$$\frac{\partial \frac{exp(s_i)}{exp(s_i)+exp(s_j)}}{\partial s_i} = \frac{exp(s_i+s_j)}{(exp(s_i)+exp(s_j))^2} \tag{7}$$

$$\frac{\partial \frac{exp(s_i)}{exp(s_i)+exp(s_j)}}{\partial s_j} = -\frac{exp(s_i+s_j)}{(exp(s_i)+exp(s_j))^2} \tag{8}$$

The first order derivatives of the two one-sided triangular probability density functions give rise to (9)-(12). Recognizing that the first term is always the original pdf, from which it is easy to take draws, makes writing the simulated equivalent of the gradient convenient. Before we do that define $E(P_n^1)$ and $E(P_n^2)$ as the expected choice probability for individual n based on either the first (or second) one-sided triangular density for notational convenience.

$$\frac{\partial \frac{2(\beta_n-c+exp(s_1))}{exp(2s_1)}}{\partial c} = \frac{2(\beta_n-c+exp(s_1))}{exp(2s_1)}\left[\frac{-1}{\beta_n-c+exp(s_1)}\right] \tag{9}$$

$$\frac{\partial \frac{2(\beta_n-c+exp(s_1))}{exp(2s_1)}}{\partial s_1} = \frac{2(\beta_n-c+exp(s_1))}{exp(2s_1)}\left[\frac{exp(s_1)}{\beta_n-c+exp(s_1)}-2\right] \tag{10}$$

$$\frac{\partial \frac{2(c+exp(s_2)-\beta_n)}{exp(2s_2)}}{\partial c} = \frac{2(c+exp(s_2)-\beta_n)}{exp(2s_2)}\left[\frac{1}{c+exp(s_2)-\beta_n}\right] \tag{11}$$

$$\frac{\partial \frac{2(c+exp(s_2)-\beta_n)}{exp(2s_2)}}{\partial s_2} = \frac{2(c+exp(s_2)-\beta_n)}{exp(2s_2)}\left[\frac{exp(s_2)}{c+exp(s_2)-\beta_n}-2\right] \tag{12}$$

When (13) defines the likelihood function of interest for individual n

$$\begin{aligned} L_n &= \frac{exp(s_1)}{exp(s_1)+exp(s_2)}\int_{\beta_{1n}}\prod_{t=1}^{T}\frac{exp(X_{nit}\beta_{1n})}{\sum_{j=1}^{J}exp(X_{njt}\beta_{1n})}\frac{\beta_{1n}-c+exp(s_1)}{exp(2s_1)}d\beta_{1n} \\ &+ \frac{exp(s_2)}{exp(s_1)+exp(s_2)}\int_{\beta_{2n}}\prod_{t=1}^{T}\frac{exp(X_{nit}\beta_{2n})}{\sum_{j=1}^{J}exp(X_{njt}\beta_{2n})}\frac{c+exp(s_2)-\beta_{2n}}{exp(2s_2)}d\beta_{2n} \\ &= \frac{exp(s_1)}{exp(s_1)+exp(s_2)}E(P_n^1)+\frac{exp(s_2)}{exp(s_1)+exp(s_2)}E(P_n^2) \end{aligned} \tag{13}$$

Then the simulated gradients are provided by (14)-(16). Note that the same draws as used in the main estimation procedure can be used to evaluate the gradient.

$$
\begin{aligned}
\frac{\partial L_n}{\partial c} &= \frac{exp(s_1)}{exp(s_1)+exp(s_2)} \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \frac{exp(X_{nit}\beta_{1n}^r)}{\sum_{j=1}^{J} exp(X_{njt}\beta_{1n}^r)} \left[\frac{-1}{\beta_{1n}^r - c + exp(s_1)} \right] \qquad (14) \\
&+ \frac{exp(s_2)}{exp(s_1)+exp(s_2)} \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \frac{exp(X_{nit}\beta_{2n}^r)}{\sum_{j=1}^{J} exp(X_{njt}\beta_{2n}^r)} \left[\frac{1}{c + exp(s_2) - \beta_{2n}^r} \right] \\
\frac{\partial L_n}{\partial s_1} &= \frac{exp(s_1+s_2)}{(exp(s_1)+exp(s_2))^2} \left(E(P_n^1) - E(P_n^2) \right) \qquad (15) \\
&+ \frac{exp(s_1)}{exp(s_1)+exp(s_2)} \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \frac{exp(X_{nit}\beta_{1n}^r)}{\sum_{j=1}^{J} exp(X_{njt}\beta_{1n}^r)} \left[\frac{exp(s_1)}{\beta_{1n}^r - c + exp(s_1)} - 2 \right] \\
\frac{\partial L_n}{\partial s_2} &= \frac{exp(s_1+s_2)}{(exp(s_1)+exp(s_2))^2} \left(E(P_n^2) - E(P_n^1) \right) \qquad (16) \\
&+ \frac{exp(s_2)}{exp(s_1)+exp(s_2)} \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \frac{exp(X_{nit}\beta_{2n}^r)}{\sum_{j=1}^{J} exp(X_{njt}\beta_{2n}^r)} \left[\frac{exp(s_2)}{c + exp(s_2) - \beta_{2n}^r} - 2 \right]
\end{aligned}
$$