ASSURING FINITE MOMENTS FOR WILLINGNESS TO PAY IN RANDOM COEFFICIENT MODELS

Andrew Daly, RAND Europe and ITS Leeds
Stephane Hess, ITS Leeds
Kenneth Train, University of California, Berkeley

Abstract

Random coefficient models such as mixed logit are increasingly being used to allow for random heterogeneity in willingness to pay (WTP) measures. In the most commonly used specifications, the distribution of WTP for an attribute is derived from the distribution of the ratio of individual coefficients. Since the cost coefficient enters the denominator, its distribution plays a major role in the distribution of the WTP. Depending on the choice of distribution for the cost coefficient, and its implied range, the distribution of the WTP may or may not have finite moments. In this paper, we identify a criterion to determine whether, with a given distribution for the cost coefficient, the distribution of WTP has finite moments. Using this criterion, we show that some popular distributions used for the cost coefficient in random coefficient models, including normal, truncated normal, uniform and triangular, imply infinite moments for the distribution of WTP, even if truncated or bounded at zero. We also point out that relying on simulation approaches to obtain moments of WTP from the estimated distribution of the cost and attribute coefficients can mask the issue by giving finite moments when the true ones are infinite.

Keywords: random coefficients; willingness-to-pay; mixed logit; discrete choice

1. Introduction

Discrete choice models with random coefficients are often used to estimate the distribution of willingness to pay (WTP) for attributes of goods or services. For example, mixed logit models (see e.g. McFadden & Train, 2000; Hensher & Greene, 2003; Sillano & Ortúzar, 2005), which are the most widely used form of random coefficient choice models, have been used extensively in transportation research to estimate various types of WTP, including travellers' WTP for changes in travel time. Under the standard approach, the analyst specifies the distribution of the cost and other attribute coefficients and estimates the parameters of this distribution. The distribution of WTP is then derived from the estimated distribution of the coefficients. This practice, while accurate in theory, can be problematic in practice because the cost coefficient enters the denominator of WTP. A value of the cost coefficient that is arbitrarily close to zero results in an arbitrarily large WTP. As a result, the moments of the WTP distribution might not exist for a given distribution of the cost coefficient. However, it is these moments of the WTP distribution, and especially the mean, that are of crucial interest, for example, in policy appraisal. The analyst must be especially careful, therefore, when specifying the distribution for the cost coefficient to assure that the resulting distribution of WTP is useful and meaningful.

In the field of choice modelling, the discussion of the *acceptability* of different distributions (primarily the normal) has generally focussed on the behavioural realism of allowing for *positive* values in the distribution of the cost coefficient (e.g. Hensher & Greene, 2003; Hess et al., 2005), rather than the possibility of non-existence of moments for WTP distributions. Although this possibility is known at least to some researchers, many authors continue to use the normal and other unsatisfactory distributions, and as a result produce misleading moments for the WTP distribution, for example using simulation (with or without censoring). In this paper, we not only show the inappropriateness of these methods, but also present a theorem that allows researchers to test whether the moments of the inverse of a distribution exist. Here, we show how some of the distributions used with a view to avoiding the above problems (e.g. Triangular bounded at zero) similarly do not have finite moments when inverted. The practical importance of this theorem should not be understated. It will allow authors to determine with certainty whether or not the moments of their WTP distribution are finite, hence avoiding the risk of producing misleading results. This will also give authors the confidence to allow for the important heterogeneity in the cost sensitivity, where they may otherwise have relied on a fixed cost coefficient, potentially leading to inferior model performance and bias in the heterogeneity retrieved for other sensitivities (due to confounding with the unexplained heterogeneity in the cost coefficient).

The central contribution of this paper is to provide a theorem that identifies, under certain conditions, whether or not the moments of WTP exist for any given distribution of the cost coefficient. Using the theorem, we show that, in addition to the normal distribution, many commonly used distributions for the cost coefficient, including truncated normal, uniform, and triangular (even if truncated or bounded at zero) imply that the distribution of WTP has undefined (i.e. infinite) moments. We point out, and illustrate with examples, that simulation of the WTP distribution from draws of the cost and attribute coefficients can mask the problem, providing (incorrect) finite moments even though the true moments are infinite. The problem is only masked further when relying on censoring of draws during simulation.

The ratio of random terms has long been a topic in the statistics literature. For normally distributed variables, it has been known that their ratio does not have moments, though a direct proof is rare. For example, Geary (1930), Marsaglia (1965), and Phan-Gia et al. (2006) provide different ways of expressing the distribution of the ratio of two normally distributed variables, but do not discuss or derive its moments. Fieller (1932) shows that the moments of the ratio of two normal variables do not exist and uses this fact as motivation for restricting the support of the variables to a region of the positive quadrant such that moments exist. He does not examine normals that are truncated at zero, for which the support is an open region of the positive quadrant. Geary (1930) and Hinkley (1969) examine the ratio of two normals and a transformation of this ratio, which has become known as the Geary-Hinkley transformation. They consider situations in which the denominator is "very unlikely" to assume negative values -- a condition that Hayya et al. (1975) formalise in terms of the coefficient of variation in the denominator. Curtiss (1941) derives the density for the ratio of two random variables (without restricting to normality) under the assumption that the support for the denominator does not include zero, such that the density is always defined; however, he does not derive moments or discuss when they exist.

We build upon earlier work by providing a result that directly addresses the existence of moments for a ratio of two variables and is applicable for any distribution for the variables. The result implies the already-known fact that moments do not exist for the ratio of two normals, but does so in a way that is perhaps more transparent than the previous literature on ratios of normals. The result also implies that moments do not exist when the denominator is a truncated normal with truncation at zero, and that higher moments (beyond the mean) do not exist when the denominator has a triangular distribution with one end-point at zero. Some further examples are also given. To our knowledge, the implications for these distributions have not previously been shown. Most importantly, the result provides a mechanism for examining other distributions, which is helpful for researchers who are searching for alternative distributions that are supported by the data and yet realistic in their implications for willingness-to-pay.

The remainder of this paper is organised as follows. Section 2 gives the theorem and applies it to several well-known distributions. Section 3 discusses simulation of WTP moments and shows how simulation sometimes seems to obtain finite moments for distributions whose moments are known not to exist. Section 4 provides another caution, namely, that attempts to avoid problems by using the distribution of individual-level conditional means as an indication of the population distribution can mask undefined moments because it misapplies the relation between conditional and unconditional distributions. Section 5 provides a number of ways for the analyst to assure finite moments of WTP. Section 6 concludes.

2. Existence of moments of WTP distribution

In a typical model specification, the utility function for an alternative is a linear function of the various attributes of an alternative, such as travel time and travel cost, multiplied by their coefficients. For example, the utility obtained by a given person from alternative j may be represented as:

$$U_j = \theta\, c_j + \beta\, a_j + other\ terms + \varepsilon_j \tag{1}$$

where c_j is the cost of the alterative and a_j is a non-cost attribute. In the discussions in this paper, we focus on a specification that interacts the coefficients with the attributes in a purely linear fashion, but the issues highlighted here apply similarly to specifications incorporating non-linear interactions.

A random coefficients specification is obtained by considering θ and β to be random with a specified distribution in the population, whose parameters are estimated. The final error term ε is also considered random, most commonly independent and identically distributed extreme value, so that the model is a mixed logit.

The WTP for an attribute of alternative j is, by definition, the ratio of the marginal utility of the attribute to the marginal utility of its cost, which in the case of linear-in-parameters utility is the ratio of the attribute coefficient to the cost coefficient:

$$\text{WTP} = \frac{\partial U_j / \partial a_j}{\partial U_j / \partial c_j} = \frac{\beta}{\theta}$$

The distribution of β and θ in the population induces a distribution of WTP. The question that we are examining is whether the implied distribution of WTP has finite moments, which is certainly a desirable property and might even be considered a necessary property for a meaningfully specified model. For use in appraisal, the mean is clearly essential.

The main theorem of this paper can be applied directly when β and θ are independent, an assumption that is common in much current work using mixed logit models. It is however always possible to express β as

$$\beta = \beta^* + \alpha\theta$$

where β^* and θ are uncorrelated. Then WTP is $\beta/\theta = \beta^*/\theta + \alpha$, and the distribution of WTP depends on the distribution of two uncorrelated terms. A lack of correlation is only approximately the same as complete independence, however, unless the variables are normal. In Appendix 2 it is shown that equivalent results can be obtained for an extended class of 'linearly dependent' variables, so that by simple transformations the theorem can still be applied exactly. When there is dependence of another form, the theorem applies approximately.

If β and θ are independent, then the moments of WTP are the product of the moments of β and $1/\theta$. In particular, the k^{th} moment of WTP is

$$E[(\beta/\theta)^k] = E(\beta^k) * E(1/\theta^k)$$

The analyst in specifying the distribution of β can directly assure that $E(\beta^k)$ exists. The relevant question, then, is whether the inverse moments $E(1/\theta^k)$ exist for a specified distribution of θ. The following theorem provides the necessary guidance.

> If a random variable θ has an absolutely continuous probability density $f(\theta)$, then for any positive integer k the inverse moment $E(1/\theta^k)$ exists if and only if $\lim_{\theta\to 0} \dfrac{f(\theta)}{\theta^h}$ exists for some $h > k-1$.

The proof is given in Appendix 1. The theorem has the following Corollary, also proved in Appendix 1.

> If a random variable θ has an absolutely continuous probability density $f(\theta)$ defined on the positive half-line and $\lim_{\theta\to 0} f(\theta) > 0$, then none of the inverse moments $E(1/\theta^k)$ exists.

The Corollary has the following applications to some commonly used distributions.

- For a uniform distribution bounded at 0, all inverse moments are non-existent.
- For a normal distribution truncated at 0, as with the uniform distribution, all the inverse moments are non-existent.

The main theorem can be used to investigate other commonly used distributions.

- For a triangular distribution bounded at 0, used quite commonly in practice, $f(\theta) = \lambda\theta$ for θ less than the mean. Then $f(\theta)/\theta^1 = \lambda$ so that the limit is finite for $h = 1$ and the mean of $1/\theta$ exists. However, for $h > 1$, $f(\theta)/\theta^h = \lambda/\theta^{h-1}$ and the limit does not exist. The variance of $1/\theta$ and its higher moments do not exist.
- Both lognormal and Johnson's Sb distribution have all inverse moments in their basic specifications (i.e. with the domains between 0 and infinity for the lognormal, and 0 and 1, or any other positive number, for the Johnson Sb). Of course, this result is already known for the lognormal, but the theorem can also be used to show it: since the lognormal density approaches zero faster than any power function, the limit exists for all h and so all inverse moments exist.

The Johnson's Sb distribution tends to a constant times the lognormal as it approaches zero, which implies that its inverse moments also exist.

- For the gamma distribution with shape parameter ν, inverse moments exist for $k < \nu$ but no higher. This is because the gamma distribution tends to a constant times θ^{ν} as its argument θ approaches zero. The negative exponential distribution is a gamma distribution with $\nu = 0$, and so it has no inverse moments.

- For the Weibull distribution with frequency function $f(\theta) = \frac{r}{\lambda}\left(\frac{\theta}{\lambda}\right)^{r-1} e^{-(\theta/\lambda)^r}$, with support on the non-negative half-line, the inverse moment k exists if and only if $k < r$.

Further, no inverse moments exist for distributions with strictly positive density at zero. This result applies to the commonly used normal distribution, as well as to bounded distributions (e.g., uniform, triangular, Johnson's Sb, and lognormal) if they are offset such that they straddle 0.

The theorem also implies that if $f(\theta) = 0$ in an interval around zero, then all inverse moments exist. This result means that all inverse moments can be assured to exist by setting bounds on the distribution such that the density is zero within an interval around zero. These bounds can be set a priori (if some means of determining them is available) or estimated. For example, the uniform and triangular distributions can be used and still assure finite inverse moments by shifting the distributions away from zero, through an extra parameter that is estimated or specified *a priori*. However, pre-imposing such bounds might be viewed as arbitrary.

In summary, suppose the support of θ is the continuous set (a, b). Then if a<0<b, no moments exist, while if a<b<0, all moments exist. In the critical (and common) case that b=0 the result depends on the form of the distribution and the above theorem needs to be used. Table 1 presents a summary of the above results for common choices of distributions for the cost coefficient in mixed logit studies.

Table 1: Summary of existence of WTP moments for common choices of distribution for cost coefficient

Distribution for cost coefficient		**Existence of moments**
Uniform	Bounded between a and b, with a<0	• b<0: all WTP moments exist • b≥0: no WTP moments exist
Normal	Unbounded	• No WTP moments exist
	Truncated, with a domain between negative infinity and b	• b<0: all WTP moments exist • b≥0: no WTP moments exist
Triangular	Bounded between a and b, with a<0	• b<0: all WTP moments exist • b=0: mean WTP exists, but no other WTP moments exist • b>0: no WTP moments exist
Lognormal (with sign change)	Bounded between negative infinity and b, where b is either estimated or is zero by default	• b≤0: all WTP moments exist • b>0: no WTP moments exist
Johnson SB	Bounded between a and b (estimated or fixed), with a<0	• b≤0: all WTP moments exist • b>0: no WTP moments exist

Gamma (with sign change)	With shape parameter ν, and with domain between negative infinity and b, where b is either estimated or is zero by default	• b<0: all WTP moments exist • b=0: k^{th} WTP moment exists if and only if $k < \nu$ • b>0: no WTP moments exist
Negative exponential (with sign change)	Gamma with shape parameter set to zero, and with domain between negative infinity and b, where b is either estimated or is zero by default	• b<0: all WTP moments exist • b≥0: no WTP moments exist
Weibull (with sign change)	With shape parameter *r*, and with domain between negative infinity and b, where b is either estimated or is zero by default	• b<0: all WTP moments exist • b=0: k^{th} WTP moment exists if and only if $k < r$ • b>0: no WTP moments exist

3. Simulation of WTP ratios

The most common practice when dealing with induced distributions is to estimate their moments through simulation (see e.g. Hensher & Greene, 2003, for an in-depth discussion). In the case of WTP, numerous draws are taken from the distribution of β and θ, and the ratio of β to θ is calculated for each draw. The ratios are draws from the distribution of WTP, and the mean and variance of the draws of the ratios are used as estimates of the mean and variance of WTP in the population. For distributions with finite moments, simulation of the moments is an appropriate and useful procedure. However, when the moments do not exist, simulation can serve to mask their non-existence, providing finite simulated moments when the true moments are infinite. Indeed, in simulation, it is highly unlikely that a draw of the denominator will be obtained that provides a value of the ratio that is larger than the computer can recognise. In fact, a particular set of draws of the denominator might not result in any unreasonably large ratios, so that the mean ratio seems not only finite but reasonable. Of course, if the simulation is repeated, quite different results will probably be obtained in different simulations, and this variance over simulations might alert the analyst to the problem. However, simulation is usually performed once rather than repeatedly, and an analyst obtaining a finite, and perhaps even reasonable, simulated mean WTP would not have any reason to suspect that the true moments for the specified model are infinite.

Even though the point (that simulation gives finite moments even when the true moments are infinite) is fairly obvious, it is perhaps useful to provide some illustration of the phenomenon, in order to emphasise its importance. For these illustrations, we simulate the value of time (VOT, aka, the WTP for time reductions) with time measured in minutes and cost in £. To focus the situation, we specify the travel time coefficient to be fixed, with a value of -0.05. The cost coefficient is given the following distributions:

- Uniform, distributed between 0 and -1, i.e. with mean of -0.5
- Triangular, distributed between 0 and -1, i.e. with mean and mode of -0.5
- Normal, distributed with a mean of -0.5 and a standard deviation of 0.3, so that 95% of the mass is on the negative side of zero.

In each simulation, we use 10^7 (ten million) draws of the cost coefficient for each distribution. We calculate the mean and variance of the simulated draws of VOT. Note that 10^7 is a far larger number of draws than is used in most simulations; we chose such a large number of draws in order to show that *finite* moments are wrongly obtained even with an extremely large number of draws.

We repeated the simulations 10 times, with a different set of 10^7 random draws each time. Table 2 gives the mean and standard deviation over the ten simulations of the simulated mean and variance.

Table 2: Simulation results

		Mean	Variance
Uniform	**Mean (across ten runs)**	60.23	$2.05 \cdot 10^9$
	Std. dev. (across ten runs)	13.92	$3.03 \cdot 10^9$
Triangular	**Mean (across ten runs)**	8.32	252.02
	Std. dev. (across ten runs)	0.01	38.53
Normal	**Mean (across ten runs)**	1.88	$5.73 \cdot 10^8$
	Std. dev. (across ten runs)	6.57	$8.51 \cdot 10^8$

The uniform distribution produces very large values of simulated mean VOT, while the normal distribution produced very low values, even though in both cases the mean is actually undefined given the above theorem. Of course, in both cases, the standard deviation over simulations is large, especially given the huge number of draws in each simulation. For the triangular distribution, the true mean VOT is actually finite, such that it is amenable to simulation. As expected, the standard deviation of the mean over simulations is exceedingly small (0.01) for the triangular distribution, since the true mean exists and each simulation uses a very large number of draws.

The simulated variances are very large for the uniform and normal distributions, which should provide a clue to the analyst that something is amiss. However, for the triangular distribution, the simulated variance of VOT, while large, is not nearly as great as for the other two distributions, even though the true variance is infinite under all three of them. As with the simulated means, the simulated variances do not provide a reliable guide as to whether the true variances are defined, and, indeed, one would not expect simulation to be useful for this purpose.

Analysts have occasionally advocated the use of censoring in order to account for the fact that the simulation results are unduly influenced by a very small number of extreme draws. Using the example of the normal distribution, we simulated the mean and variance of VOT using different degrees of censoring of the distribution, ranging between one percent and ten percent with symmetrical censoring. The results are shown in Table 3, where once again, we present summary results across the ten different sets of simulation draws. The results for the full sample (i.e. 0% censoring) are given in the first column.

Table 3: Impacts of censoring

		0%	1%	2%	5%	10%
Mean	**Mean (across ten runs)**	1.88	7.24	7.24	7.23	7.22
	Std. err (across ten runs)	6.57	0.02	0.01	0.01	0.01
Variance	**Mean (across ten runs)**	$5.73 \cdot 10^8$	1,955.10	961.61	368.39	171.11
	Std. err (across ten runs)	$8.51 \cdot 10^8$	14.18	4.47	0.86	0.30

The results suggest that a small amount of censoring leads to a high degree of stability across runs, so that the variation across runs is indeed just a result of a few extreme values. From this perspective, censoring may appear to be a very desirable solution. However, this conclusion neglects the fact that the true mean is in fact infinite: the censoring, by creating stability over simulations, serves to further mask the reality of the situation. This fact is shown more vividly with the simulated variance. With each additional degree of censoring, there is a reduction in the simulated variance of

VOT. The analyst essentially selects the variance of VOT by selecting a degree of censoring, rather than estimating the variance of VOT from the data. The basic problem is that the model, as specified, implies infinite mean and variance of VOT, and so the solution is to re-specify the model rather than censor under the existing specification.

Another observation can be made at this point. While the moments of a distribution may not be defined, the percentiles always exist. As an illustration, Table 4 shows the simulated lower and upper 1%, 5% and 10% points for the ten sets of draws for the three choices of distributions above. As can be seen from these results, the percentile estimates are very stable across the ten runs. This finding may be useful in scenarios where the median WTP can be used (£6/hr for uniform and triangular in this case, and £5.60/hr for normal), or where we are interested in knowing what share of the sample population have a WTP above some threshold value of interest. However, while the temptation may exist to consider the mean WTP to be approximated by the midway point of the 5-percentile and 95 percentile values (or some similarly symmetric percentiles), this concept is not correct. In fact, we observed scenarios with the normal distribution where the *simulated* mean exceeded the 99% percentile. And of course, the actual mean is infinite.

Table 4: Simulating percentiles of WTP distribution

		1%	5%	10%	50%	90%	95%	99%
Uniform	**Mean**	3.03	3.16	3.33	6.00	29.99	59.99	299.67
	Std. Err	0.00	0.00	0.00	0.00	0.02	0.07	0.90
Triangular	**Mean**	3.23	3.56	3.86	6.00	13.42	18.97	42.42
	Std. Err	0.00	0.00	0.00	0.00	0.01	0.01	0.07
Normal	**Mean**	-90.63	2.21	3.04	5.60	16.11	26.86	107.44
	Std. Err	0.27	0.00	0.00	0.00	0.01	0.02	0.31

4. On the use of conditional distributions

Train (2003) provides a procedure for estimating the distribution of coefficients for each individual in a dataset conditional on the choices that the person made. It may be tempting to use these conditional distributions to provide alternative distributions of WTP when the estimated models imply unreasonable distributions. In particular, it is possible to calculate the conditional mean coefficients for each sampled individual, take the ratio of these conditional means, and interpret the mean and variance of these ratios over individuals as the mean and variance of WTP in the population. Generally, the variance of the ratio of conditional means is smaller than the variance of WTP calculated from the unconditional distribution of coefficients, and so the former may be considered (erroneously) an improvement in estimation when the later are unreasonably large.

There are two problems in this approach. First, the ratio of conditional means is not the mean of the conditional distribution of the ratio, since $E(\beta/\theta) \neq E(\beta)/E(\theta)$, even if β and θ are independent. Second, the procedure mistakenly ignores the fact that conditional distributions are derived from, and must aggregate to, the unconditional distribution. In particular, the (unconditional) population variance is equal to the variance of the conditional means PLUS the variance of the conditional distribution around the conditional means. The variance of conditional means is less than the unconditional variance not because it is a more reasonable estimate, but rather because it incorrectly excludes the variance around the conditional means.

5. Available solutions

There are several paths that analysts can take – and some have taken – to assure that their models have distributions of WTP with finite moments. A few of the most prominent are:

- In selecting a distribution for the cost coefficient, the analyst can use the theorem above to determine whether the implied distribution of WTP has finite moments. As highlighted above, the existence of inverse moments may, for some distributions, depend on the actual estimated parameters of the distribution.
- The analyst can require that the distribution of the cost coefficient be bounded away from zero. Inverse moments exist for any distribution that does not have support arbitrarily close to zero. The bound can be set by the analyst (which entails a degree of arbitrariness that may be deemed unacceptable, cf. Hess et al., 2005) or it can be treated as an additional parameter to estimate.
- Another possible solution is the use of a finite mixture models, including latent class models (see e.g. Hess et al., 2007) and the non-parametric estimation procedures suggested by Train (2008). In these specifications, the continuous distribution for coefficients discussed above is replaced with a finite distribution, i.e., a distribution that has mass at a finite number of coefficient values. If all the points for the cost coefficient are either estimated or constrained to be away from zero, the implied WTP distribution is also finite with defined values.
- Finally, the model can be re-parameterised in WTP space, as suggested by Train and Weeks (2005). This solution is perhaps the most straightforward, since it avoids the need to consider the distribution of inverse coefficients. Utility in equation 1 can be re-written as

$$U_j = \theta\, c_j + \theta\, \lambda\, a_j + other\ terms + \varepsilon_j \qquad (2)$$

where $\lambda = \beta/\theta$ is the WTP for the attribute. Instead of specifying distributions for θ and β, the analyst specifies distributions for θ and λ. In estimation, the coefficient of the attribute is calculated as the product of θ and λ rather than as one coefficient in itself. Any model specified as in equation (1) can be re-expressed in the form of equation (2), and vice versa, and so the re-expression is simply a re-parameterisation rather than a new model.

6. Conclusions

The majority of discrete choice models estimated by academics and a growing share of models estimated by non-academic practitioners now utilise random coefficients for cost and non-cost attributes. Many of these studies have as their objective the computation of WTP measures, where the distribution of WTP for an attribute (which is the coefficient of that attribute divided by the cost coefficient) is derived from the distribution of the coefficients. In particular, mean WTP is central to transport policy appraisal.

In this paper, we explore the impact that the distributional assumptions for the cost coefficient have on the distribution of WTP. In particular, we focus on the moments of the WTP distribution. While it is known that such moments do not exist in the case of a normally distributed cost coefficient, there continue to be examples of studies that mistakenly compute the mean and/or variance of the WTP from such specifications. While there is growing use of alternative distributions, it is important to select these distributions not only with a focus on behavioural realism and computational convenience, but also considering the implications for the WTP distribution.

The core contribution of this paper is to identify a criterion to determine whether the distribution of WTP has finite moments. Using this criterion, we show that some popular distributions used for the cost coefficient in random coefficient models, including normal, truncated normal, uniform and triangular, imply infinite moments for the distribution of WTP. We also point out that relying on simulation approaches to obtain moments of WTP from the estimated distributions of the cost and attribute coefficients can mask the problem by giving finite moments when the true ones are infinite. Similarly, using conditional distributions is inappropriate, and percentiles cannot be used to infer any information on the moments (since the moments don't exist).

The theorem presented in this paper provides analysts with a reliable way of establishing whether their chosen distributional assumptions permit them to compute a WTP distribution with finite moments. At the same time, we realise that there will be cases in which it is not straightforward to arrive at a distribution that meets behavioural and computational requirements while also leading to finite WTP moments. For this reason, we also briefly discuss a number of alternative ways of obtaining meaningful WTP results, namely using finite mixture such as latent class models and a form of non-parametrics; and by re-parameterising the model in WTP space so that the distribution of WTP is estimated directly rather than derived.

Acknowledgements

The second author acknowledges the financial support of the Leverhulme Trust in the form of a "Leverhulme Early Career Fellowship".

References

Curtiss, J. H. (1941), 'On the Distribution of the Quotient of Two Chance Variables', The Annals of Mathematical Statistics 12 (4): 409-421.

Geary, R. C. (1930), 'The Frequency Distribution of the Quotient of Two Normal Variates', Journal of the Royal Statistical Society 93 (3): 442.446.

Fieller, E. C. (1932) 'The Distribution of the Index in a Normal Bivariate Population', Biometrika 24 (3/4): 428-440.

Hayya, J., Armstrong, D. and Gressis, N. (1975), 'A Note on the Ratio of Two Normally Distributed Variables', Management Science 21 (11): 1338-1341.

Hensher, D. & Greene, W. (2003), 'The Mixed Logit Model: The State of Practice,' *Transportation* 30(2), 133–176.

Hess, S., Bierlaire, M. & Polak, J.W. (2005), 'Estimation of Value of Travel-Time Savings using Mixed Logit Models,' *Transportation Research Part A*, 39(2-3), pp. 221-236.

Hess, S., Bierlaire, M. & Polak, J.W. (2007), 'A Systematic Comparison of Continuous and Discrete Mixture Models,' *European Transport* 37, pp. 35-61.

D. V. Hinkley (1969), 'On the Ratio of Two Correlated Normal Random Variables', Biometrika 56 (3): 635-639.

Marsaglia, G. (1965), 'Ratios of Normal Variables and Ratios of Sums of Uniform Variables', Journal of the American Statistical Association 60 (309): 193-204.

McFadden, D. (1998) 'Measuring Willingness-to-Pay for Transportation Improvements,' in T Gärling, T Laitila, K Westin (eds.), Theoretical Foundation for Travel Choice Modelling, Pergamon.

McFadden, D. & Train, K. (2000), 'Mixed MNL Models for Discrete Response,' *Journal of Applied Econometrics* 15, 447–470.

T. Pham-Gia, N. Turkkan and E. Marchand (2006), 'Density of the Ratio of Two Normal Random Variables and Applications', Communications in Statistics: Theory and Methods (Taylor & Francis) 35 (9): 1569-1591.

Sillano, M. and Ortúzar, J. de D. (2005), 'WTP estimation with mixed logit models: some new evidence', Environment and Planning A 37 (2005)

Train, K. E. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, Massachusetts.

Train, K.E. (2008), 'EM Algorithms for Nonparametric Estimation of Mixing Distributions,' *Journal of Choice Modelling*, 1(1), pp. 40-69.

Train, K. and Weeks, M. (2005), 'Discrete Choice Models in Preference Space and Willingness-to-Pay Space,' in R. Scarpa and A. Alberini (eds), Application of Simulation Methods in Environmental and Resource Economics, Springer Publisher, Dordrecht, The Netherlands, chapter 1, pp. 1-16.

APPENDIX 1: Proof of theorem in Section 2

> If a random variable θ has an absolutely continuous probability density $f(\theta)$, then for any positive integer k the inverse moment $E(1/\theta^k)$ exists if and only if $\lim_{\theta\to 0}\frac{f(\theta)}{\theta^h}$ exists for some $h > k-1$.

We first prove the following Lemma.

> Let $f(\theta)$ be the probability density function of a random variable θ, where $f(\theta)$:
>
> 1. is absolutely continuous
> 2. has support only on the positive half-line
> 3. is monotonic (either non-decreasing or non-increasing) in an interval $(0,r)$ for some r
>
> Then for a positive integer k, $E(1/\theta^k)$ exists if and only if $\lim_{\theta\to 0}\frac{f(\theta)}{\theta^h}$ exists for some $h > k-1$.

Suppose $\lim_{\theta\to 0}\frac{f(\theta)}{\theta^h} = H$ for some non-negative value of h. Certainly the limit exists for $h = 0$ because of the continuity of f.

Define $S_n^k(r) = \int_{r/(n+1)}^{r/n} \theta^{-k} f(\theta)d\theta$. Then, because of the monotonicity property, we know that $S_n^k(r)$ is contained in a closed interval

$$S_n^k(r) \in \left[\left(\frac{r}{n}-\frac{r}{n+1}\right)\left(\frac{r}{n+1}\right)^{-k} f\left(\frac{r}{n+1}\right), \left(\frac{r}{n}-\frac{r}{n+1}\right)\left(\frac{r}{n}\right)^{-k} f\left(\frac{r}{n}\right)\right]$$

i.e.

$$S_n^k(r) \in \left[\left(\frac{1}{n}\right)\left(\frac{n+1}{r}\right)^{k-1} f\left(\frac{r}{n+1}\right), \left(\frac{1}{n+1}\right)\left(\frac{n}{r}\right)^{k-1} f\left(\frac{r}{n}\right)\right] = \left[L_n^k(r), U_n^k(r)\right], \text{ say}$$

and

$$\lim_{n\to\infty} L_n^k(r) = \left(\frac{1}{n}\right)\left(\frac{n+1}{r}\right)^{k-1} H\left(\frac{r}{n+1}\right)^{h} = H\left(\frac{1}{n}\right)\left(\frac{n+1}{r}\right)^{k-h-1}$$

$$\lim_{n\to\infty} U_n^k(r) = \left(\frac{1}{n+1}\right)\left(\frac{n}{r}\right)^{k-1} H\left(\frac{r}{n}\right)^{h} = H\left(\frac{1}{n+1}\right)\left(\frac{n}{r}\right)^{k-h-1}$$

Then, providing $k < h+2$, this interval shrinks to zero and, because of the continuity and limit properties of f,

$$\lim_{n\to\infty} S_n^k(r) = \left(\frac{H}{r^{k-h-1}}\right) n^{k-h-2}$$

Then define $T_N^k = \sum_{n=1}^{N} S_n^k$. If the limit exists, $\lim_{N\to\infty} T_N^k = \int_0^r \theta^{-k} f(\theta) d\theta$, and the k^th inverse moment also exists. Conversely, if the limit does not exist, the k^th inverse moment does not exist.

It is a classical result that $\sum n^{\alpha}$ converges if and only if $\alpha < -1$, so that the series T converges if and only if $(k-h-2) < -1$, i.e. $k-1 < h$, proving the Lemma.

The Lemma has the following Corollary

> If a random variable θ has an absolutely continuous probability density $f(\theta)$ defined on the positive half-line and $\lim_{\theta\to 0} f(\theta) > 0$, then none of the inverse moments $E(1/\theta^k)$ exists.

The corollary follows immediately by noting that the limit in the Lemma fails to exist for any $h > 0$ and so the moments do not exist for any $k \geq 1$.

We can now conclude that the condition of monotonicity in the Lemma is not required. If $\lim_{\theta\to 0} f(\theta) > 0$, then the inverse moments and the limit fail to exist, as in the Corollary. However, if $\lim_{\theta\to 0} f(\theta) = 0$, then because of continuity the function must be non-decreasing in a neighbourhood of 0 (it must remain non-negative). That is, any function for which inverse moments or the limit might exist must be monotonic close to zero.

Suppose the function f is defined over the negative half line. For negative θ, the limit is defined for integer h only but with that reservation the same existence result then applies to the negative half-line as for the positive half-line.

Finally, if f is defined over the whole line then the k th moment exists for the whole line if and only if the moments over both positive and negative half-lines exist, completing the proof of the theorem.

Note that in the case when $\lim_{\theta\to 0-} \frac{f(\theta)}{\theta^h} \neq \lim_{\theta\to 0+} \frac{f(\theta)}{\theta^h}$, i.e. there is some sort of 'kink', then if both limits exist the k th moment exists for $k < h-1$; but if one does not exist then the k th moment does not exist.

APPENDIX 2: On the ratio of correlated variables

Suppose we are interested in the ratio of two correlated random variables, A and B. In the main text it is indicated that we can always define a variable

$$A^* = A - \alpha B$$

which allows us the express investigate the distribution of the ratio by setting

$$\frac{A}{B} = \alpha + \frac{A^*}{B}$$

with A^* and B uncorrelated, *irrespective* of the distribution of the variables A and B. In the case of jointly normal variables A and B this implies that A^* and B are independent, but in the case of variables with other distributions independence does not follow from lack of correlation, though it might be considered to hold approximately, except in special cases.

If the variables A and B can be related by a *linear dependence*, which we can define by

$$A = f_1(B) + f_2(B)A^*$$

with A^* and B truly independent, then, if $f_1 \neq 0$, we can calculate the ratio

$$\frac{A}{B} = \frac{1}{B_1^*} + \frac{A^*}{B_2^*}, \text{ with } B_i^* = \frac{B}{f_i(B)}.$$

This ratio exists only if both components exist and these can be tested by the theorem in the usual way, since A^* is independent of B_2^*. If $f_1 = 0$, then we just have the second term and that can be tested as usual. For the joint normal distribution, f_1 is a constant times B and f_2 is a constant, so that the tests can be made directly on B.

The concept of linear dependence thus defines a fairly wide class of joint distributions for which the existence of ratio inverse moments can be tested using the theorem presented in this paper.