University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 4/2018 (263)

A NEW BASELINE MODEL FOR ESTIMATING WILLINGNESS TO PAY FROM DISCRETE CHOICE MODELS

Richard T. Carson
Mikołaj Czajkowski

Warsaw 2018

WORKING PAPERS 4/2018 (263)

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

A New Baseline Model for Estimating Willingness to Pay from Discrete Choice Models

Richard T. Carson[a], **Mikołaj Czajkowski**[b*]

[a] *Department of Economics, University of California, San Diego, USA*
[b] *Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland*
[*] *Corresponding author: miq@wne.uw.edu.pl*

Abstract: We show a substantive problem exists with the widely-used ratio of coefficients approach to calculating willingness to pay (WTP) from choice models. The correctly calculated standard error for WTP using this approach is shown to always be infinity. A variant of this problem has long been recognized for mixed logit models. We show it occurs even in simple models like the conditional logit used as a baseline reference specification. It occurs because the standard error for the cost parameter implies some possibility that the true parameter value is arbitrarily close to zero. We propose a simple yet elegant way to overcome this problem by reparameterizing the coefficient of the (negative) cost variable to enforce the theoretically correct (and empirically almost always found) positive coefficient using an exponential transformation of the original parameter. This reparameterization enforces the desired restriction that no-part of the confidence region for original cost parameter spans zero. With it the confidence interval for WTP is now finite and well behaved. Our proposed model is straightforward to implement using readily available software. Its log-likelihood value is the same as the usual baseline discrete choice model and we recommend its use as the new standard baseline reference model.

Keywords: conditional logit, confidence intervals, contingent valuation delta method, discrete choice experiment, Krinsky-Robb, multinomial logit, probit, welfare measures

JEL codes: C01, C15, C18, Q0

Acknowledgments: The authors wish to thank Tirthankar Chakravarty, Andrew Daly, William Greene, Stephane Hess, James Powell, V. Kerry Smith, and Jeff Wooldridge and the participants of the European Association of Environmental and Resource Economists' Conference in Toulouse, and International Choice Modelling Conference in Sydney for valuable discussions and helpful comments. Any errors are authors' alone. Polish Ministry of Science and Higher Education and the Foundation for Polish Science support is gratefully acknowledged.

http://doi.org.10.26405/WP/WNE/2018/263/004

1. Introduction

The use of the ratio of an attribute parameter to a cost parameter to measure willingness to pay (WTP) for the attribute in a discrete choice model is a long-standing and widely used practice in applied economics. It appears to have first been used in Lisco's (1967) University of Chicago dissertation using a probit model to look at transportation mode choice and the value of time.[1] By the time of McFadden's seminal 1974 paper on conditional logit models, the ratio calculation of attributes divided by cost was made without substantive comment, and it is described in some detail in Ben-Akiva and Lerman's highly cited 1985 book. WTP as the ratio of parameters has subsequently been implemented in a vast number of papers in many fields of applied microeconomics, including agricultural economics, environmental economics, health policy, industrial organization, marketing, and transport (e.g., Haab and McConnell 2003; Anderson et al. 1992; Train 2009; Louviere et al. 2006). The standard reference text by Hensher, Rose and Greene (2005) succinctly sums up conventional wisdom: "In simple linear models, WTP measures are calculated as the ratio of two parameter estimates, holding all else constant. Provided at least one attribute is measured in monetary units, the ratio of the two parameters provides a financial indicator of WTP."[2] There are now thousands of studies and millions of WTP and consumer surplus estimates based on this approach.

In practice, the true parameters of the utility function are unknown. A researcher usually deals with maximum likelihood (ML) estimates of the utility function parameters, so there is uncertainty associated with them. Taking this uncertainty into account in estimating statistics related to the underlying WTP distribution, the ML estimates of the utility function parameters are considered to have an asymptotically normal distribution, with means equal to their ML estimates and standard deviations equal to their standard errors (Bockstael and Strand 1987). This in essence gives them an informal Bayesian interpretation (Geweke 1986). However, in this case, calculating moments of a resulting *ratio distribution* (e.g., the empirical distribution of WTP) becomes problematic, and we argue that it is generally performed incorrectly.[3]

---

[1] Lisco noted that his approach to obtaining a confidence interval for the value of time by using the confidence intervals of the parameters of his probit model was ad hoc and not strictly valid.
[2] A footnote to "linear models" notes: "As models of discrete choice are linear in the utility functions, the choice modeler is able to take advantage of this fact."
[3] This ratio is also often referred to as consumer surplus (subtracting price from WTP), an implicit price or the marginal rate of substitution. For simplicity, we will refer to all these ratio measures as WTP throughout this paper. Many other ratio measures in economics, other social sciences, biology and medicine have a similar structure. Here for concreteness we will assume that the denominator of this ratio is the parameter on some function of the

Non-existence of moments of a ratio distribution resulting from dividing two independent normally distributed variables has long been known in statistics (e.g., Fieller 1932; Geary 1930; Merrill 1928; Nicholson 1941). More recent, but by now quite old work (Marsaglia 1965; Hinkley 1969) in statistics deals with the case of the ratio of correlated normals which is the situation of interest here.[4] More generally the non-existence of moments issue confronts all ratio estimates that exhibit Cauchy-like tail behavior where that occurs because the distribution of the denominator spans zero (Lehmann and Shaffer 1988). We show that this problem exists with the estimated standard error of WTP in all commonly used discrete choice logit and probit models, with non-random parameters (e.g., multinomial logit (MNL), nested logit, latent class logit). We also show that the same problem also exists in models with random parameters such as the commonly used random parameters (mixed) logit (MXL) and newer generalized multinomial logit models that use a non-random monetary attribute coefficient as the denominator of WTP expressions to avoid the issues involved in having a distribution of cost parameters in the denominator of the expression for WTP.

In section 2, we look at problems that can occur with the delta method and the Krinsky-Robb parametric bootstrap, the two most commonly used ways of calculating standard errors and confidence intervals for WTP ratio estimators. In section 3, we examine where the problem of non-existent moments of a ratio distribution of two normally distributed random variables comes from, and analyze the probability density function (pdf) of the resulting ratio distribution. In section 4, we propose a new specification for selected parameters in discrete choice models which allows for the cost parameter to be constrained to strictly negative or positive values only. This alternative specification imposes a standard restriction from neoclassical economic theory and, at the same time, avoids problems associated with non-existent moments of the resulting WTP ratio distribution. Our proposed alternative specification can be easily implemented in available commercial software and used for any discrete choice model with a non-random cost coefficient. Section 5 considers more general discrete choice models where it is common practice to use a fixed cost coefficient to avoid having some type of random variation

---

monetary variable. The framework put forward here can be adapted to more complicated specifications involving functions of both cost and income variables.

[4] The non-existence of the expectation of the ratio of two independent standard normal occurs because the ratio has a Cauchy distribution. This result is a staple in books on counterexamples in probability (e.g., Romano and Siegel 1986). The Cauchy result can be obtained under considerably weaker conditions than two independent standard normals including allowing for non-zero expectations of the two random variables, and substituting the weaker condition of symmetry for normality coupled with a restriction of the relative size of the scale parameters (Stoyanov 1997).

in the cost coefficient. Section 6 briefly examines the use of estimation in WTP space and show that it causes a related problem when its estimated parameters are transformed in preference space. Section VII provides an illustration of how our proposed alternative specification works in practice using data from a well-known contingent valuation study valuing the prevention of oil spills off the central coast of California and, in a random-parameters setting, using an alternative-fuel vehicle discrete choice experiment study. Section VIII provides some concluding remarks and puts forward the proposition that our reformulation of the relevant likelihood become the new reference statistical model for WTP ratio estimates.

2. Delta Method and Krinsky-Robb Parametric Bootstrap

Two methods are commonly used in empirical applications for estimating statistics related to the WTP distribution: the delta method (Greene 2011) and the Krinsky and Robb (KR; 1986) parametric bootstrapping.[5] It will be useful to draw a clear distinction between (i) estimation of the variance of the WTP distribution which can be used to look at the risk of an estimator using a quadratic loss function like mean squared error and (ii) determining a confidence interval within which some fraction, such as 95%, of the estimates of WTP defined in terms of the ratio of ML parameters is expected to fall with repeated sampling. This distinction is often unimportant because confidence intervals are cast in terms of plus or minus the absolute value of the relevant z-statistic times the standard error.[6] However, it helps to explain why some estimators with unbounded risk appear to perform well when performance is judged in terms confidence interval coverage.

[5] The delta method has long been used for calculating the confidence interval for the ratio of normal parameters. Daly, Hess and de Jong (2012) put forward the case for using this approach for asymptotic WTP ratio estimates from choice models. For an early and highly cited application of the KR method (1986; 1991) for this purpose see Park et al. (1991). This application was motivated by the well-recognized potential problems with the delta method in this case. A third method, the Fieller confidence interval (Fieller 1954), is often used in the biometrics literature. Despite some proponents (Dufour 1997), it is used infrequently in econometrics (some notable examples include Valentine 1979; Staiger et al. 1997; Blomqvist 1973). The main reasons for this seem to be that it is based on using fiducial, rather than frequentist or Bayesian inference (Wallace 1980), and the less intuitive form of the interval which requires solving for roots of quadratic inequalities which may result in a finite interval, two disjoint semi-infinite intervals (a complement of a bounded interval) or even the whole real line. The Fieller method does not always result in finite confidence intervals is a result of a denominator which may have a distribution with significant mass around zero (Scheffe 1970; Zerbe 1978). Effectively, the Fieller confidence interval only requires the normality of numerator and denominator of the ratio rather than the ratio itself. Hirschberg and Lye (2010) provide a geometric comparison between the delta and Fieller confidence intervals as well as a frequentist interpretation. It is also possible to base confidence intervals on the likelihood ratio test statistic (Williams 1986) and non-parametric bootstrap approaches (e.g., Hole 2007). New approaches for providing confidence intervals in small samples continue to be proposed (e.g., Paige et al. 2011).

[6] Throughout the text we use z-stats for simplicity, although the equivalent t could be used for small samples.

The delta method in our case is based on a linear first-order Taylor series approximation of a non-linear function. The asymptotic variance of an estimator function g, which is assumed to be continuously differentiable, is given by:

$$\text{as.var}\left(g\left(b\right)\right)=\frac{\partial g\left(\beta\right)}{\partial\beta}\Sigma\left(\frac{\partial g\left(\beta\right)}{\partial\beta}\right)', \tag{1}$$

where b is the estimator of the parameters β and the asymptotic variance covariance matrix is Σ.

Applying this to the case of a ratio of two random variables following bivariate normal distribution $\left(B,C\right)$: $BVN\left(\mu_B,\mu_c;\sigma_B,\sigma_C;\rho\right)$, the asymptotic variance of their ratio becomes:

$$\text{as.var}\left(\frac{B}{C}\right)=\frac{\mu_B^2}{\mu_C^4}\sigma_C^2+\frac{1}{\mu_C^2}\sigma_B^2-\frac{2\mu_B}{\mu_C^3}\rho\sigma_B\sigma_C\ . \tag{2}$$

The quality of the delta approximation with respect to ratios of estimated coefficients for probit and logit models has long been questioned in the biometrics literature (Finney 1971). This has not deterred its use in applied economics work. Ruud (2000) for example, in his widely-used econometrics text, provides an example of using the delta method to look at the distribution of the ratio of two ML parameters. He cautions, though, that "sensible application of the delta method is limited to situations in which this approximate linearity holds for all likely values of the random variable". The difficulty with the ratio of two ML parameters, however, goes much deeper. Its moments do not exist and hence the delta estimate is inadmissible from a statistical perspective as it poses infinite risk (Zellner 1978).[7]

However, as can be seen in (1) above, the delta method provides a finite well-behaved estimate of the asymptotic variance. Most applied researchers are unaware that ML is often known to produce finite estimates of infinite quantities (Oehlert 1992).[8] Intuitively, the

---

[7] Interestingly, this problem was noted earlier in the context of welfare measurements in travel cost analysis using continuous variables which also involves a ratio estimate for WTP (Smith 1990). Smith's paper explicitly followed Zellner's (1978) view of ratio estimators. The impetus for treating the regression coefficients from travel cost models as random was Bockstael et al. (1987) seminal paper. Adamowicz et al. (1989) impose non-negativity on the consumer surplus estimate using Geweke's (1986) Bayesian oriented method of imposing an inequality constraint. Kling (1991) in a simulation study notes that while there were some differences between approaches to estimating "either the standard deviation or confidence intervals" that all "provided reasonable approximations". This seemed to cement the typical empirical practice of using either the delta method or the Krinsky-Robb approach.

[8] Some hint of the problem can be seen by noting that the higher order terms of the Taylor series expansion does not die out when the estimate of the price parameter is close to zero (Graham-Tomasi et al. 1990). Variants of the

continuity assumption is being violated. For any fixed value of B, the magnitude of the ratio variable B/C jumps dramatically when as C goes from being negative to being positive. Further, the assumption sometimes made that $C \neq 0$ does not solve the existence problem with respect to the delta estimate for a ratio variable confidence interval.

Gleser and Hwang (1987) in an important theorem for a class of problems, which includes the ratio of two normal, show that it is impossible to construct confidence intervals for key parameters which have both positive confidence and finite expected length.[9] The underlying difficulty is that there is part of the parameter space for which identification is tenuous. This issue has been explored at some length in the econometrics literature by Dufour (1997) who shows that the Gleser and Hwang theorem holds when near some value (e.g., $C = 0$) the function of interest is *locally almost unidentified*. Effectively, if the interval on which a potential confidence interval is defined contains a locally almost unidentified region, then the method used to develop the confidence interval must be capable of producing an infinite interval if too much of the density is close to zero, which the delta method is incapable of doing in this case. Serving to underscore that the nature of the problem is not at the single point $C = 0$, Lai et al. (2004) study the inverse of a "punctured" normal looking at how large the fraction of the density trimmed off on each side of zero needs to be in order to obtain an estimate of the ratio variable with finite first and second moments.

There are other symptoms of the problem with the delta method in forming confidence intervals for ratio variables. It always produces a symmetric confidence interval, even though the distribution of the WTP ratio variable can be quite asymmetric, particularly if the sample size is not large. The delta confidence interval also can diverge considerably from the Fieller confidence interval which is valid under more general circumstances. The underlying reason for both problems is that the delta method can be a poor approximation if C is not sufficiently far from zero and the sample size not large.

---

ratio estimator problem appear in simultaneous equation models (Bergstrom 1962), distributed lagged models (Lianos and Rausser 1972) and the reduce rank regression used in tests of cointegration (Phillips 1994). More recently, the ratio estimation problem has been shown to lie behind the notion of weak instruments in econometric models (Woglom 2001).

[9] Gleser and Hwang (1987) show this result holds for a number of important statistical problems. Koschat (1987) independently derives it for the ratio of two normal. Franz (2007) provides a useful discussion of the importance of the Gleser and Hwang (1987) theorem to a set of long standing issues with ratio estimators.

Given these potential problems with the delta method, there are two related questions for applied work. Why does the delta method appear to work well in many simulation studies and is it possible to identify conditions where the delta confidence interval is likely to work well? Hirschberg and Lye (2010) provide a review of the previous analytical and simulation studies starting with Finney (1971) who argued that the delta method is only adequate if the t-statistic on the denominator of the ratio was above 8.75, a condition not typically met in applied economic work.[10] They point out that most Monte Carlo simulation studies (e.g., Hole 2007; Dorfman et al. 1990), have assumed that the denominator of the ratio is highly significant and the numerator less so, and thus examine a situation where the delta method should perform reasonably well.[11]

From a technical perspective, to see when the delta method will produce a reliable confidence interval, it is useful to first note that both the normal and the Cauchy are both members of the symmetric stable family of distributions. The ML location parameter for the Cauchy distribution is the median and the scale parameter is the half-width, which is half the distance between the 25th and 75th quantiles. The half-width is deterministically linked to the standard deviation in the normal case. What is important to keep in mind is that the closer to zero C gets the larger the Cauchy scale parameter gets because it is the density near zero that is generating the extreme Cauchy tails. While the WTP ratio distribution is still Cauchy, it is reasonably approximated by a normal if C is far enough away from zero in a statistical sense (i.e., a function of the actual distance, the half-width scale parameter, and sample size) if one does not go too far out into the tails. The reason the delta method has good performance in this situation is that among the class of asymptotically unbiased median estimators, no estimator has a higher probability than the ML estimate of being in a specified interval around the true value of the ratio (Zaman 1981; Fiebig 1985). Thus, while the delta method's estimate of the variance

---

[10] Much of this literature is cast in terms of the coefficient of variation, which is the inverse of the t-statistic, and assuming that the correlation coefficient is equal zero. Marsaglia (2006) provides a tighter bound for the ratio variable having an approximate normal distribution that requires the t-statistic on the denominator, t_C , be greater than 4 and $(t_B - \rho t_C)\big/(1-\rho^2)^{.5}$ be less than 2.256, where t_B is the t-statistic for the numerator. This condition is also frequently not met in empirical studies. Further, letting $n \to \infty$ does not guarantee that this condition is met. If the correlation coefficient ρ is equal to zero, meeting this condition involves the numerator being substantially less significant than a very significant denominator. Hirschberg and Lye (2010) show that when ρ is not equal to zero, the case where the ratio variable differs in sign from ρ is particularly problematic in terms of the delta method providing erroneous results.

[11] Hole (2007) is particularly careful to point out that good performance of the delta method is dependent on having a highly significant cost parameter.

may be useful in forming percentile-based confidence intervals it is not a valid estimate of the variance.

The KR approach assumes joint asymptotic normality of the individual estimated parameters. It is viewed as avoiding some of the potential problems with the delta approximation and often results in somewhat larger confidence intervals than the delta method. The KR approach is parametric bootstrap procedure and involves simulating multiple draws from the distribution of structural parameters of the WTP ratio variable.[12] The function of the draws (in our case, the ratio of simulated coefficients) provides empirical distribution of WTP which is used for calculating its mean, median, standard deviation or quantiles. Since, however, in this case, the ratio (WTP) distribution has undefined moments (i.e., mean and standard deviation) the KR simulation method is used incorrectly, as calculated mean and standard deviation (which are in fact infinite) are unstable and tend to 'explode' with increasing the number of random draws.[13] We illustrate this with simulation results presented in Table 1, where two normally distributed variables were assumed to have means equal to 1, standard deviations corresponding to p-value of 0.05 (i.e., $\sigma = 0.5102$; the minimum p-value reported in many empirical studies), and set the correlation coefficient equal to 0.5.[14]

[12] The multivariate normal distribution is characterized by a coefficient vector (as a vector of means) and an asymptotic variance-covariance matrix.

[13] The common application of the KR method to calculate means and standard deviations of ratios of normally distributed parameters seems to be a misinterpretation of the method. The original application, calculating elasticities, for the KR approach did not involve ratios of parameters (Krinsky and Robb 1991; Krinsky and Robb 1986).

[14] The Krinsky-Robb method for calculating median or other quantiles (e.g. 0.025 and 0.975 quantiles used to report 'confidence interval' of WTP) is robust, since the quantiles of the ratio distribution of two normals are well-defined. The popular Poe et al. (1994); (1997) convolutions test for whether two WTP distributions are different is often treated as a test of the difference in mean WTP between the two samples. However, this test (as its authors note) allows the two distributions to vary on extreme quantiles so it is possible to accept that two WTP distributions are statistically equivalent when they have radically different means. Ignoring divergences in extreme quantiles where there is little information may be a desirable property in comparing two distributions, but the test should not be treated as looking at the difference in means. The overlapping confidence interval approach to testing the difference between two ratio estimates of WTP (Park et al. 1991) is invalid because the confidence intervals do not exist. Classical tests of the difference in the means of two ratio WTP estimates that rely on having well-defined standard deviations likewise fail.

Table 1: Behavior of delta and Krinsky Robb estimates for ratio-based WTP

Method	Draws	Mean	Standard Deviation	Median	95% c.i. / quantile range
Analytical	–	Undefined	Undefined	0.97	(0.63;4.29)
Delta	–	1*	0.51	1*	(0.00;2.00)
Krinsky and Robb	100	0.89	0.82	0.84	(-1.00;2.61)
	1000	0.36	24.19	0.97	(-0.68;3.78)
	10,000	1.20	12.70	0.97	(-0.45;4.12)
	100,000	1.12	21.06	0.97	(-0.51;4.32)
	1,000,000	0.76	302.93	0.97	(-0.51;4.33)
	10,000,000	0.54	1,194.51	0.97	(-0.52;4.31)
	100,000,000	-0.05	12,769.87	0.97	(-0.52;4.29)
Fieller	–	1*	–	1*	$\left(-\infty,-2.18\cdot10^{13}\right)\cup\left(0.00,+\infty\right)$

* calculated as ratio of coefficients

Many empirical works use a KR simulation to derive means, standard deviations, and confidence intervals of the WTP distribution with 100 to 10,000 draws being common. As illustrated with an example provided in Table 1 and explained in more detail below, this method leads to erroneous results. Like the delta method, the KR approach always produces a confidence interval of finite length even though one can observe sizeable increases in the standard deviation in Table 1 as n increases. As a consequence, it and non-parametric bootstrap procedures must not have positive confidence according to the Gleser and Hwang (1987) theorem. A bootstrap procedure based on Fieller bounds does have positive confidence because it can return an interval with infinite length (Hwang 1995).

3. The Algebra of Ratio Variables

Typical algebraic operations on multivariate random variables result in sum distributions, difference distributions, product distributions, and ratio (or quotient) distributions (Springer 1979). Drawing inferences from the ratio of coefficients is elemental in many statistical applications. Formally, let W be a random variable defined as $W = B/C$, where B and C are random variables following some known distributions, with joint distribution function $f(b,c)$. The pdf of W can then be calculated from (Curtiss 1941):

$$h(w) = \int_{-\infty}^{+\infty} |q| f(wq,q)\,dq . \tag{3}$$

In the case of discrete choice models estimated with ML techniques, the parameter estimates are known with uncertainty and are almost always treated as being asymptotically normally distributed random variables with known means (estimated parameters) and standard deviations (equal to estimated standard errors). A researcher interested in deriving the marginal rate of substitution of one choice attribute for another wants to be able to estimate key statistics related to its empirical (ratio) distribution.

Assume B and C are normally distributed, so that (B,C) has bivariate normal density:

$$w_{BC}(b,c;\mu_B,\mu_C;\sigma_B,\sigma_C;\rho)= \\ =\frac{1}{2\pi\sigma_B\sigma_C\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}\frac{1}{1-\rho^2}\left(\left(\frac{b-\mu_B}{\sigma_B}\right)^2-2\rho\left(\frac{b-\mu_B}{\sigma_B}\right)\left(\frac{c-\mu_C}{\sigma_C}\right)+\left(\frac{c-\mu_C}{\sigma_C}\right)^2\right)\right). \quad (4)$$

Solving (3) for a closed form solution is troublesome, as the integral becomes (Fieller 1932):

$$h(w) = \frac{\sigma_C\sigma_B\sqrt{1-\rho^2}}{\pi\left(\sigma_B^2-2\rho\sigma_B\sigma_C w+\sigma_C^2w^2\right)}\exp\left(-\frac{1}{2}\frac{1}{1-\rho^2}\left(\frac{\mu_B^2}{\sigma_B^2}-2\rho\frac{\mu_B\mu_C}{\sigma_B\sigma_C}+\frac{\mu_C^2}{\sigma_C^2}\right)\right)+ \\ \exp\left(-\frac{1}{2}\frac{\left(\mu_B-\mu_C w\right)^2}{\sigma_B^2-2\rho\sigma_B\sigma_C q+\sigma_C^2w^2}\right)\frac{\sigma_B\left(\rho\mu_B\sigma_C-\mu_C\sigma_B\right)+\sigma_C\left(\rho\mu_C\sigma_B-\mu_B\sigma_C\right)w}{\pi\left(\sigma_B^2-2\rho\sigma_B\sigma_C w+\sigma_C^2w^2\right)^{\frac{3}{2}}}\times. \\ \int_0^{\frac{\sigma_B\left(\rho\mu_B\sigma_C-\mu_C\sigma_B\right)+\sigma_C\left(\rho\mu_C\sigma_B-\mu_B\sigma_C\right)w}{\sigma_B\sigma_C\left(\left(1-\rho^2\right)\left(\sigma_B^2-2\rho\sigma_B\sigma_C w+\sigma_C^2w^2\right)\right)^{\frac{1}{2}}}}\exp\left(-\frac{1}{2}q^2\right)dq$$

(5)

If $\mu_B=\mu_C=0$, (5) simplifies and $h(w)$ becomes the pdf of a Cauchy distribution. In a more general case, expression (5) does not have closed-form solution in terms of elementary functions, as the densities of B and C are not negligible at $w\le 0$. However, (5) can be expressed in terms of the standard normal CDF (Φ) (Hinkley 1969):

$$h(w)=\frac{K_3K_4}{\sqrt{2\pi}\sigma_B\sigma_C K_1^3}\left(\Phi\left(\frac{K_3}{\sqrt{1-\rho^2}K_1}\right)-\Phi\left(\frac{-K_3}{\sqrt{1-\rho^2}K_1}\right)\right)+\frac{\sqrt{1-\rho^2}}{\pi\sigma_B\sigma_C K_1^2}\exp\left(\frac{-K_2}{2\left(1-\rho^2\right)}\right),$$

where: (6)

$$K_1=\left(\frac{w^2}{\sigma_B^2}-\frac{2\rho w}{\sigma_B\sigma_C}+\frac{1}{\sigma_C^2}\right)^{\frac{1}{2}}, \qquad K_3=\frac{\mu_B w}{\sigma_B^2}-\rho\frac{\mu_B+\mu_C w}{\sigma_B\sigma_C}+\frac{\mu_C}{\sigma_C^2},$$

$$K_2=\frac{\mu_B^2}{\sigma_B^2}-\frac{2\rho\mu_B\mu_C}{\sigma_B\sigma_C}+\frac{\mu_C^2}{\sigma_C^2}, \qquad K_4=\exp\left(\frac{1}{2\left(1-\rho^2\right)}\left(\frac{K_3^2}{K_1^2}-K_2\right)\right),$$

or in terms of the Kummer's confluent hypergeometric (Hermite) function $\left({}_1F_1\right)$ (Pham-Gia et al. 2006).[15]

Daly, Hess and Train (2012) have recently shown in the context of random parameter choice models that if the distribution of the cost parameter has positive density at zero, then the resulting ratio distribution for WTP does not have finite moments. The Daly, Hess and Train result is applicable to a much wider range of contexts.[16] For normally distributed B and C, this follows directly from (5), where non-zero density of C at zero creates the 'Cauchy component', causes the integrals to diverge and, as a result, moments to be infinite. Only if $\Pr(C \leq 0) = 0$ is the resulting ratio distribution 'well-behaved'. Even if the mean and the standard deviation of the ratio distribution are undefined, (6) can still be used to derive its quantiles, such as the median or 0.025 and 0.975 percentiles, which can serve as a measure of spread or a substitute for confidence intervals.

It is instructive to analyze the shape of the ratio distribution resulting from dividing two normally distributed random variables. Appendix 1 provides illustrative examples of ratio distributions resulting from dividing normally distributed variables characterized by different coefficients of variation (i.e., the ratios of standard deviations to the means of the distributions $(c.v. = \sigma/\mu)$, which correspond to p-values of the estimated coefficients) and different correlation coefficients.[17] The resulting ratio distribution is clearly not normal and often not symmetrical. The distribution can even be bimodal (Marsaglia 1965). Thus, the two standard location parameters, mean and median, of the resulting ratio distribution are likely to differ. The WTP ratio estimate is not a consistent estimate of either location statistic.

To illustrate this divergence further, Appendix 2 contains examples of how taking the ratio of coefficients (in our case normalized to 1) differs from the 'pseudo-mean', and the median of the ratio distribution.[18] The true mean and standard deviation of this distribution are

---

[15] As an aside, we found that the popular statistical packages (e.g., MATLAB) evaluate (6) more quickly than the frequently used Kummer's confluent hypergeometric (Hermite) function representation.

[16] See Piegorsch and Casella (1985) and Khuri and Casella (2002) for general discussions and proofs regarding conditions required for the existence of negative moments of random variable (e.g., $1/C$) moments.

[17] The mean (μ) is normalized to 1. The standard deviation (σ) is selected ensure the specified p-value.

[18] The pseudo-means and pseudo-standard deviations were simulated using 10^8 draws from multivariate normal distribution of B and C. The results are calculated for a few 'illustrative' cases (p-value of B and C equal to 0.01, 0.05, 0.1; correlation coefficient equal to -0.9, -0.5, 0, 0.5, 0.9).

undefined. The simulation masks this problem by not taking enough draws, especially in the case when means of B and C are relatively far from 0.[19]

Appendix 2 reports the 'pseudo-standard deviation', and 2.5'th and 97.5'th percentiles of the distribution along with the delta, KR, and Fieller confidence intervals. The delta interval mistakenly suggested the process is well behaved. The KR-based interval is sometimes substantially different from the delta one but it too appears well behaved. The admissible Fieller bounds show that 95% confidence intervals frequently include $\pm\infty$.

To summarize, we have shown, under the usual specification of a discrete choice model, that (a) the mean and standard deviation of the resulting WTP ratio distribution are undefined, and (b) the resulting distribution is not normal, being typically skewed and potentially bimodal. The ratio formed by the point estimate of the coefficients is also different from the median of the ratio distribution. The usual implementations of the delta and KR approaches to obtaining confidence intervals help to mask the nature of the problem. It is still possible to report median WTP calculated using (6) or (7) and reporting of extreme quantiles such as those used here (i.e., 0.025 and 0.975) can help to illustrate the spread of the distribution.

If one is prepared to rule out $C=0$ as a point estimate, then the question can be raised, is it is any less likely that the true parameter value is negative? Once the possibility of the cost parameter has a zero or the negative true parameter value is ruled out,[20] the estimation procedure used should be revised to take account of this restriction. We delve into this issue in the next section and show that implementing this restriction is straightforward and solves several issues related to the estimation of WTP.

---

[19] Generally, the ratio of coefficients is closer to the median of the distribution than to its pseudo-mean; however, only when p-values of B and C are relatively small and correlation coefficient is quite large does the *median* of the resulting ratio distribution become reasonably close to the ratio of coefficients.

[20] That this is often the case and the restriction is required by standard economic theory (Varian 1992). Only goods not conforming to the law of demand (Giffen and Veblen goods) have positive price coefficients. We note that the same procedure can be applied to other coefficients that economic theory clearly signs. Gelman (2011), a prominent statistician, has pointed out that serious inference problems exist for any ratio variable in which the denominator of the ratio can plausibly take on either a positive or negative value. Marsaglia (2006) notes that the normality assumption for a ratio variable is much less tenuous if one is prepared to assume that the denominator is always positive.

4. An Alternative Specification for the Discrete Choice Model

In this section, we propose a different approach that directly ensures existence of moments of the empirical distribution of WTP. Usually all parameters enter the utility function linearly so individual i's utility associated with choosing alternative j is given by:

$$U_i\left(Alternative=j\right)=U_{ij}=\boldsymbol{\beta}'\mathbf{x}_{ij}+\gamma z_{ij}+\varepsilon_{ij}, \quad (7)$$

where $\mathbf{x}_{ij}$ is a vector of alternative-specific attributes and z_{ij} is the cost associated with it (with $\boldsymbol{\beta}$ and γ being utility function parameters). We reformulate this utility function so that the coefficient of the monetary attribute has no support in negative values. When cost is used as the monetary variable, as is typically the case, it will be necessary to redefine z_{ij} to be the negative of the cost variable.[21] A simple specification that does this is given by:

$$U_i\left(Alternative=j\right)=U_{ij}=\boldsymbol{\beta}'\mathbf{x}_{ij}+\exp\left(\delta\right)z_{ij}+\varepsilon_{ij} \ . \quad (8)$$

Estimation of the parameters of the model follows in the usual way. However, now the parameter associated with the cost attribute $\left(\delta\right)$, which has an asymptotic normal distribution, is exponentiated. Since $\exp\left(\delta\right)$ is strictly positive, the ratio distribution representing the empirical distribution of WTP will have well-defined moments including the mean and standard deviation.[22] If the cost parameter originally had the expected sign, the log-likelihood (LL) of the model will be unchanged since maximization of the LL occurs at the same place with the estimate of λ equal to the exponentiation of the estimate of δ. Note that the z-statistic for δ will be different than for λ, typically but not always larger.[23]

Utilizing a lognormal transformation to impose bounds on the resulting transformation is a standard practice in econometric applications. For example in a discrete choice context, Train and Sonnier (2005) use lognormal transformations in order to introduce bounds on distributions of WTP for changes in attribute levels in a random parameter logit model with correlated

---

[21] Specifications using some function of income and cost will generally fulfill this requirement without having to change the sign of the variable.

[22] Daly, Hess and Train (2012) show that lognormal distribution approaches zero at a rate which assures the existence of finite moments.

[23] The invariance principle of ML does not extend to standard errors so that rescaling the parameters with a non-linear transformation alters significance levels. The likelihood ratio test is invariant to whether the model is specified in terms of λ or $\exp\left(\delta\right)$ in terms of the inclusion of the cost variable in the model. Our transformation of one of the parameters in a ML model is similar to what is frequently done with the variance parameter to enforce the known sign restriction on the variance parameter (Ruud 2000).

parameters,[24] while Fiebig et al. (2010) use lognormal transformations to assure positive distribution of scale in generalized-MNL model.

On a practical note, implementation of this approach does not require programing the MNL model from scratch. Even though the MNL model included in many popular statistical packages assumes that all the parameters enter utility function linearly, it is often possible to estimate this model as a constrained version of a Random Parameters Logit model. The negative of the cost parameter is specified as random and lognormally distributed. Then a constraint is specified that the standard deviation on the cost parameter is equal to 0. Appendix 3 provides pseudo code for implementing our proposed model in this fashion in LIMDEP/NLOGIT and STATA.[25]

Introducing the above reparameterization of the model assures finite moments of the ratio (WTP) distribution. It is therefore possible to utilize the delta, KR or the Fieller approaches for deriving the standard error or the confidence interval of WTP. Appendix 4 presents how the associated formulas can be derived for a few typical cases.[26]

5. More General Discrete Choice Models

Our proposed restriction on the cost parameter can be imposed in other more general discrete choice models when the cost parameter is specified a fixed rather than random parameter. For instance, nested logit or latent class models, both have fixed parameters that are used in WTP calculation, even though these parameters are often allowed to differ across nests or classes. Our approach can also be used in the case of specifications involving random parameters, if the cost parameter is not modeled as random, which is a common practice in applied work.[27] Several motivations have been advocating for using a fixed cost parameter. The first is the recognition that having a normally distributed distribution of cost parameters parameter causes problems with respect to the existence of the moments of the WTP distribution (Meijer and Rouwendal 2006). Daly, Hess and Train (2012) show in an MXL model that if the parameter

---

[24] They also consider other distributions such as the censored normal and Johnson's (1949) S_B to restrict the sign of the cost coefficient.

[25] In the online supplement to this paper we also show how to use the DCE estimation package developed in Matlab, available from github.com/czaj/DCE under Creative Commons BY 4.0 license, to implement the specification we propose.

[26] In the online supplement to this paper we make available a collection of Matlab functions useful for applying delta and Fieller methods for the calculating confidence intervals / sets.

[27] Train and Weeks (2005) provide an extensive list of empirical examples that use a fixed cost parameter in RPL models. Their discussion is motivated in terms of directly specifying the distribution of WTP, an issue we consider below.

associated with cost has non-zero probability at zero, then the ratio distribution of WTP has undefined moments. Other reasons why researchers use a fixed cost parameter in an MXL model include wanting to restrict the monetary parameter to have the same sign for all individuals,[28] ensuring that the distribution of the WTP has the theoretically expected sign, making the calculation of WTP (and the associated confidence intervals) less burdensome, and making the identification easier, in particular in models which allow for correlated parameters, and in datasets with few observed choices per individual (Hess and Train 2011; Revelt and Train 1998).

6. Discrete Choice Models in WTP-Space

A solution that is sometimes proposed for potential problems with WTP defined as the ratio of two parameters is to reparametrize WTP so that it is a direct argument in the LL function (Cameron 1988; Train and Weeks 2005). The standard formulation assumes that WTP follows normal distribution, although other distributional assumptions are possible.[29] While mixed results have been obtained in terms of goodness of fit and out-of-sample prediction when comparing discrete choice models specified in preference and WTP space, specification in WTP space overcomes the problems we have been discussing. However, there is a direct (asymptotic) translation between parameters in models estimated in probability space and WTP space (Scarpa et al. 2008). We show that fixing the WTP problem creates a different problem if one wants to go from the WTP space parameters to probability space parameter.

The nature of this new problem stems from the key parameter of interest in preference space generally being one of the attribute parameters and the reason that a normal distribution was originally selected for it was the belief that people might be indifferent to changes in the level of this attribute. Thus, the hypothesis of interest is that a specific $\beta_k = 0$. However, the WTP space representation rules out this possibility. The reason is that going from the WTP space representation to the preference space representation involves the product of two normally distributed parameters rather than the ratio of such parameters. The resulting distribution is

[28] Alternatively, the cost parameter can be modeled as random, following a bounded distribution, such as lognormal, censored normal or Johnson SB (Train and Sonnier 2005). Using some of these distributions has been shown, however, to entail numerical difficulties (especially in the case of correlated parameters modeled in the classical framework) or to produce behaviorally implausible results (Greene et al. 2005).

[29] Some of these such as the log-normal often result in implausibly large mean WTP estimates because of the distributions long right tail, which is often not well pinned down due to range of observed data.

known as a product normal distribution, which has rather peculiar properties – its pdf is not defined at zero.

Using the notation of (7), the vector of WTPs for the choice attributes $\mathbf{x}$ is $\mathbf{w}=\boldsymbol{\beta}/\gamma$ and that the utility function can be expressed in WTP space as:

$$U_i\left(Alternative=j\right)=U_{ij}=\left(\gamma_i\mathbf{w}_i\right)'\mathbf{x}_{ij}+\gamma_i z_{ij}+\varepsilon_{ij}\ . \tag{9}$$

These expressions of utility function are behaviorally equivalent. Note, however, that any distribution of parameters in preference space implies some distributions in WTP space, and *vice versa*. Therefore, when moving from a model estimated in WTP space to probability space, and the WTP vector estimate is assumed to have a normal distribution, the key implied parameters will be the product of two normals $\left(\beta=\gamma w\right)$. Unfortunately, just as the ratio of two normal is not normal, the product of two normal is not normal.[30]

The product normal distribution is a rather unusual distribution whose pdf expressed in terms of the parameters in (9) is given by:

$$h\left(b\right)=\int\frac{1}{|q|}f\left(\frac{b}{q},q\right)dq \tag{10}$$

While this distribution has a finite expectation and variance, it is not defined at $b=0$, and further, tends to have sharply spiked exponential-like shoulders around its expectation (Ware and Lad 2003). This is not the type of distribution one would typically associate with a preference parameter. Its use raises a set of issues involving quantities like elasticities and market shares typically associated with the preference space that have not been explored.

7. Two Empirical Examples

To provide an illustration of how our approach may be used, we provide two empirical examples using data from two publicly available stated preference experiments.[31] The first is a large contingent valuation (CV) survey designed to value WTP to prevent oil spills along California's

---

[30] The closed form pdf of the product normal was derived by Craig (1936) and Rohatgi (1976). It is, however, inconvenient to use as it is expressed in the form of the difference between two integrals. Several approximations to this distribution were proposed (e.g., Aroian 1947), as well as series expansions for purposes of numerical computation which rely mostly on the Mellin and Laplace transformation techniques (e.g., Cornwell et al. 1978; Glen et al. 2004).,

[31] The use of stated preference data here avoids the issue of potential endogeneity of the cost variable that often characterizes revealed preference data. Note that may influence how the cost parameter is estimated from revealed data but not the way that the WTP estimate is generally formed as the ratio of parameters.

central coast (Carson et al. 2004) and the second is a discrete choice experiment (DCE) study of an alternative-fuel vehicle choice (Train and Sonnier 2005).[32] [33]

A. WTP Estimates from California Oil Spill Prevention CV Study

In the oil spill prevention study, a single binary choice question elicitation format was used. Random assignment of cost to respondents allows us to focus on the unconditional expected WTP estimated using a simple conditional logit model. There are only two variables, an alternative specific constant (ASC) associated with implementing the new prevent scenario (B), versus the status quo and the cost (C) of the prevention program if implemented.[34] Table 2 provides estimation results with all parameters entering linearly (panel 1), the cost parameter entering exponentially (panel 2), and both cost and the ASC parameters entering exponentially (panel 3). This last case is equivalent to assuming it is implausible for a consumer to have negative utility associated with both money and introducing the program.

Table 2: Results from Typical and Two Alternative Specifications (Complete Dataset)[35]

	MNL – typical specification (cost enters linearly)	MNL – alternative specification 1 (cost enters exponentially)	MNL – alternative specification 2 (ASC and cost enter exponentially)
B – ASC associated with introducing the scenario	0.5602*** (0.0934)	0.5602*** (0.0934)	-0.5794*** (0.1667)
C – cost associated with introducing the scenario	0.7152*** (0.0845)	-0.3351*** (0.1181)	-0.3351*** (0.1181)
Ratio of coefficients[36]	78.33	78.33	78.33
Median WTP – KR	78.33	77.79	78.33
E(WTP) – KR	78.28 (undefined)	77.73	78.83
Std. err. E(WTP) – delta	8.81 (undefined)	8.81	8.81
Std. err. E(WTP) – KR	9.01 (undefined)	8.98	8.90
95% c.i. E(WTP) – delta	(61.05;95.60)	(61.05;95.60)	(61.05;95.60)
95% c.i. E(WTP) – KR (quantile range)	(60.39;95.90)	(59.83;95.25)	(62.82;97.66)

[32] We use the 1484 choice observations from 100 respondents included in the dataset available at Kenneth Train's website: http://elsa.berkeley.edu/~train/.
[33] The code and data for estimating the models presented in this paper are available from http://czaj.org/research/supplementary-materials.
[34] In estimation, we use negative of the actual cost divided by 100.
[35] ***, **, * – Significance at 1%, 5%, 10% level; standard errors in parentheses.
[36] B/C, $B/\exp(C)$ or $\exp(B)/\exp(C)$, respectively; WTP results are scaled back to $ (from the cost parameter specified in $100).

95% c.s. E(WTP) – Fieller	(60.39;95.90)	(60.39;95.90)	(60.38;95.90)
Log-likelihood	-712.7737	-712.774	-712.7737
AIC/*n*	1.3176	1.3176	1.3176
n (observations)	1,085	1,085	1,085

The first aspect of Table 2 to note is that the alternative specifications do not differ in terms of model fit, while the parameters of the attributes entering exponentially are equal to the natural logarithm of parameters entering linearly.[37] Since probability of $C \le 0$ is now 0 in specifications 1 and 2, the moments of the ratio distribution exist and can be easily calculated. We use the KR approach to numerically simulate draws from a bivariate normal and then calculate the mean and standard deviation of the resulting ratio distribution.[38]

The dataset used for this illustration is about as well-behaved as possible, considering p-values of B and C in linear MNL specification are smaller than 10^{-8}. In datasets in which p-values of the parameters are not so low, one can expect larger differences between ratio of maximum likelihood coefficients B/C, mean (i.e., expected value of WTP, if defined) and median of the resulting ratio distribution. To illustrate this, we estimated the model again for a sub-sample of 100 respondents that resulted in standard errors of the coefficients being substantially larger. The results are provided in Table 3.

Once again, the alternative specifications provide the same fit in terms of the LL, but allow us to calculate mean and standard deviation of the ratio distribution. This time, with many fewer observations in the sample, there is now considerable uncertainty with respect to the true value of parameters B and C as illustrated by much larger standard errors. Thus, the spread of the empirical distribution of WTP is larger. This can be best seen by noting that as the extreme 0.025 and 0.975 quantiles are much further away from each other than in Table 2. The commonly used ratio of the two ML coefficients for the estimate of WTP ($48.35) is now substantially different from either the expected value of the WTP (estimated mean, $34.09) or median ($43.12) of the ratio distribution in the case of the cost parameter entering exponentially. When both ASC and cost parameters enter exponentially, the mean is $64.72, and the median ($48.35) effectively becomes identical to the B/C ratio.

---

[37] This is a result of the invariance principle which states that the maximum likelihood estimator of a function is a function of the maximum likelihood estimator.

[38] We used 10^8 draws from a multivariate normal distribution of parameters B and C to derive the empirical distribution of WTP using KR method.

Table 3: Results from Typical and Two Alternative Specification (N=100 Subsample of Data)

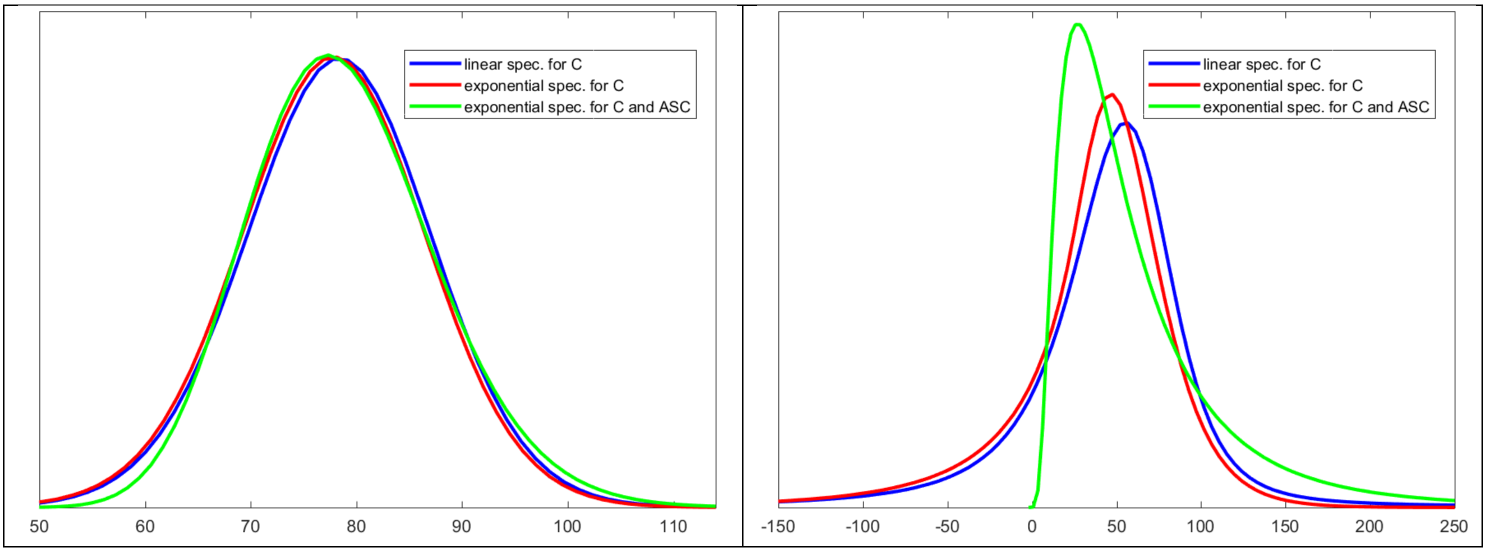
	MNL – typical specification (cost enters linearly)	MNL – alternative specification 1 (cost enters exponentially)	MNL – alternative specification 2 (ASC and cost enter exponentially)
B – ASC associated with introducing the scenario	0.2843 (0.2964)	0.2843 (0.2964)	-1.2578 (1.0431)
C – cost associated with introducing the scenario	0.5879** (0.2954)	-0.5312 (0.5026)	-0.5312 (0.5027)
Ratio of coefficients[39]	48.35	48.35	48.35
Median WTP – KR	49.60	43.12	48.35
E(WTP) – KR	60.03 (undefined)	34.09	64.72
Std. err. E(WTP) – delta	36.90 (undefined)	36.90	36.92
Std. err. E(WTP) – KR	$39.59 \cdot 10^4$ (undefined)	57.26	57.57
95% c.i. E(WTP) – delta	(-23.98;120.68)	(-23.98;120.69)	(-24.00;120.71)
95% c.i. E(WTP) – KR (quantile range)	(-153.64;168.62)	(-108.47;114.86)	(10.83;215.96)
95% c.s. E(WTP) – Fieller	(-1,610.35;153.12)	(-1,640.96;153.19)	(-1,658.72;153.19)
Log-likelihood	-66.8654	-66.8654	-66.8654
AIC/n	1.3773	1.3773	1.3773
n (observations)	100	100	100

Figure 1 shows the empirical distribution of WTP for the alternative specifications for the full sample and the much smaller sub-sample of the California oil spill data. For the full sample used in the left panel of Figure 1, the estimated ratio distributions have very similar shapes, indicating that using the proposed specification does not alter the results much, while at the same time allowing for calculating correct (bounded) moments of the WTP distribution. This similarity in the estimated ratio distributions does not carry over to the smaller sample used in the right panel of Figure 1. The linear specification results in a spread-out distribution for WTP with a substantial fraction estimated to hold negative WTP values. The alternative specification constraining the cost parameter to be strictly positive allows calculation of the moments. This concentrates the empirical distribution somewhat but does not constrain the distribution of WTP to be positive. There is substantial uncertainty exists with respect to ASC, so a non-trivial fraction of the sample is still estimated to hold negative WTP estimates. If it is justifiable to assume that the utility associated with implementing a new prevention program at zero cost cannot be negative, then the second alternative specification that further constraints the

[39] B/C, $B/\exp(C)$ or $\exp(B)/\exp(C)$, respectively; WTP results are scaled back to $ (from the cost parameter specified in $100).

empirical distribution of WTP to be positive should be used. It results in a much more asymmetric distribution and the inference one would draw about mean WTP from the sample of 100 observations is similar to that from the original sample of 1,000 observations.

Figure 1. Probability Density Function of Empirical Distribution of E(WTP) for Full Sample ($n = 1085$) in the Left Panel and for Subsample ($n = 100$) in the Right Panel



B. Alternative Fuels Vehicle DCE Study

In the case of the vehicle choice study, the elicitation format was a sequential multinomial choice. Each respondent was presented with 10-15 choice tasks consisting of 3 alternatives. The choice attributes in this study included: range (for non-gas fueled cars; $range$), engine type (dummy coded as $electric$ or $hybrid$, with gas as a reference level), performance (dummy coded as p_medium or p_high) and cost, in terms of purchase price ($c\_purchase$) and monthly operating cost ($c\_operate$).

Using this dataset, we estimated a random parameters multinomial logit model. $Range$ is assumed to be distributed log-normally, p_medium, p_high, $electric$ and $hybrid$ normally distributed. The two cost coefficients were assumed to be fixed (non-random) parameters. Assuming a log-normally distributed $range$ allows us to impose a behavioral restriction that says that marginal utility associated with this attribute cannot be negative. When the cost parameter is assumed fixed, the researcher can let the parameter multiplying the negative of cost enter the utility function in the same exponential form. Table 4 provides estimation results for the case where the two (non-random) cost parameters enter linearly (panel

1) and exponentially (panel 2) as well as WTP for *range* expressed as the marginal rate of substitution for monthly operating cost ($c\_operate$).[40]

Table 4. Results for Range from Typical and Alternative Specification [41]

	MXL – typical specification (cost parameters enter linearly)		MXL – alternative specification (cost parameters enter exponentially)	
	Means	Standard deviations	Means	Standard deviations
range (log-normally distributed)	-0.7328 (0.4636)	0.5915** (0.2980)	-0.7328 (0.4519)	0.5915** (0.2917)
electric (normally distributed)	-1.7908*** (0.3426)	1.2658*** (0.2156)	-1.7908*** (0.3380)	1.2658*** (0.2153)
hybrid (normally distributed)	0.4395*** (0.1694)	0.9745*** (0.1369)	0.4395*** (0.1693)	0.9745*** (0.1368)
p_medium (normally distributed)	0.5310*** (0.1134)	0.5696*** (0.1420)	0.5310*** (0.1133)	0.5696*** (0.1420)
p_high (normally distributed)	0.0770 (0.0999)	0.3514** (0.1675)	0.0770 (0.1000)	0.3514** (0.1677)
$c\_purchase$ (fixed)	0.4748*** (0.0380)	–	-0.7448*** (0.0799)	–
$c\_operate$ (fixed)	0.0136*** (0.0039)	–	-4.3006*** (0.2902)	–
Ratio of coefficients[42]	42.21		42.21	
Median WTP – KR	41.16		40.86	
E(WTP) – K&R	49.43 (undefined)		48.69	
Std. err. E(WTP) – delta	18.74 (undefined)		18.58	
Std. err. E(WTP) – KR	$30.04 \cdot 10^3$ (undefined)		23.42	
95% c.i. E(WTP) – delta	(5.47;78.95)		(5.80;78.63)	
95% c.i. E(WTP) – KR (quantile range)	(19.89;129.53)		(18.79;107.49)	
95% c.s. E(WTP) – Fieller	(12.99;110.54)		(13.43;110.33)	
Log-likelihood	-1,347.0764		-1,347.0764	
AIC/*n*	1.8316		1.8316	
n (observations)	1,484		1,484	

Our results demonstrate that the alternative specifications result in the same LL, while the parameters of the attributes entering exponentially are equal to the natural logarithm of

---

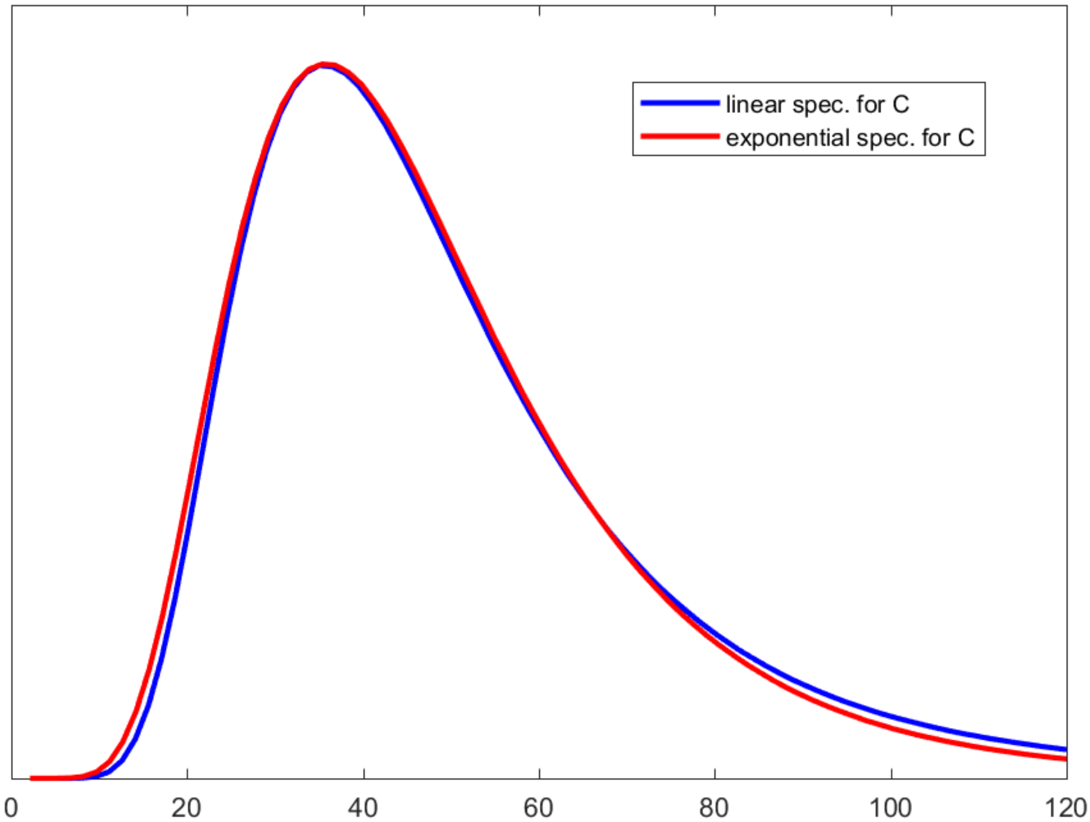
[40] We used 10^8 draws from a multivariate normal distribution of parameters μ_{range}, σ_{range} and $c\_operate$ to derive the empirical distribution of WTP using KR method.

[41] ***, **, * – Significance at 1%, 5%, 10% level; standard errors in parentheses. For log-normally distributed parameters, estimates of the mean and standard deviation of the underlying normal distribution are provided.

[42] $\exp\left(\mu_{range} + 0.5\sigma_{range}^2\right) / c\_operate$ or $\exp\left(\mu_{range} + 0.5\sigma_{range}^2\right) / \exp\left(c\_operate\right)$, respectively.

parameters entering linearly.[43] Even though the medians of WTP distributions are similar, the alternative specification assures the existence of finite moments of WTP. Restricting the cost parameter to be strictly positive allows for the moments (mean, standard deviation) of the ratio distribution to be well defined. Figure 2 illustrates these findings with kernel densities of empirical distributions of mean WTP under alternative specifications. As can be seen, the approach we propose does not 'tamper' with the results – it merely assures that the simulated ratio (WTP) distribution has finite moments.

Figure 2. Probability Density Function of Empirical Distribution of E(WTP) $(range/c\_operate)$



8. Concluding Remarks

We show that the usual practice of calculating the moments and confidence intervals of WTP and other similar economic quantities estimated using the ratio of parameters from a discrete choice model is seriously flawed. The ratio of the two maximum likelihood parameters is not, as often assumed, normally distributed and is not equal to expected WTP nor median WTP. The expected value of WTP and its standard deviation are both undefined. The workhorse delta method and the KR approach used for estimating WTP confidence intervals do not work. For

[43] There are small differences in results due to numerical approximations.

the delta method, the issue is producing a misleading finite and often reasonable estimate of the standard error when the true quantity is undefined and the potential poor quality of the delta method approximated percentile-based confidence interval unless the cost parameter is highly significant. The KR approach eventually shows the degenerate nature of the WTP ratio at extremely large sample sizes. Unfortunately, it typically produces plausible statistics for the number of replications used in applied work.

We are not the first to note problems with ratio-based WTP estimators. The problem is well-known in statistics and some of the key issues have been long pointed out in the literature on estimating WTP (Hanemann and Kanninen 2002). This has not, however, stopped the practice of estimating statistics related to WTP and similar economic quantities using the ratio of coefficients from discrete choice models. Part of the reason for this lies with convenience and following tradition. Part of it stems from the mistaken belief that mean and median WTP are equal and both equal to the ratio of ML parameters. Empirical practice has been reinforced by the delta method and the KR approach usually producing plausible statistics.

The problem is perceived to be much more acute in random parameter models. Allowing cost to be random usually results in enough density near zero and in the negative range that the problem cannot be ignored, as the implied distribution of WTP is completely implausible. Two ways of dealing with the issue have emerged. The first way is make the denominator fixed. We argue that this is only assumes all individuals are characterized by the same parameter and it does not eliminate uncertainty over the fixed parameter estimate which leads to non-existent moments of WTP. The second is to estimate the model in WTP space with WTP as direct argument LL function. This solves the problem in the sense of estimating WTP under the assumption that it is normally distributed. However, this solution hides the converse problem. In preference space, key parameters of interest such as market shares are simulated as the product of two normal distributions. The resulting distribution is known as a product-normal distribution and has a very peculiar pdf which is not defined for the central tendency and has very steep double exponential like shoulders. Lying at the heart of the problem is implausibility of the assumptions associated with using ratios or products of normals as the estimator for many quantities of economic interest.

We provide a different, simple and elegant solution to this problem by making selected preference parameters enter utility function exponentially, whenever no support in negative values is justified by economic theory and logic. Our approach will almost never change the overall model goodness of fit because key parameters such as cost almost always have the correct sign. Yet, our approach assures the existence of moments and finite confidence intervals of WTP and allows for a technically correct way of calculating mean WTP and its standard deviation. It is easily implemented using commonly used software for estimating random parameters choice models and is generalizable in a straightforward manner to many models beyond the simple conditional logit model if the parameter in the denominator of the ratio expression for WTP is fixed. Further, imposition of the restriction on cost and other parameters will often substantially tighten the confidence intervals around expected WTP, suggesting the restriction(s) brings considerable useful information. When both an attribute and the cost parameter are constrained, our approach will result in empirical estimates of the WTP distribution that have support only with respect to positive values, which makes behavioral sense in many if not most cases.

References

Adamowicz, W.L., Fletcher, J.J., Graham-Tomasi, T. 1989. Functional form and the statistical properties of welfare measures. American Journal of Agricultural Economics 71, 414−421.

Anderson, S.P., Palma, d.d. A., Thisse, J.-F. 1992. Discrete Choice Theory of Product Differentiation. MIT Press, Cambridge, MA.

Aroian, L.A. 1947. The probability function of the product of two normally distributed variables. The Annals of Mathematical Statistics 18, 265−271.

Ben-Akiva, M., Lerman, S.R. 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, MA.

Bergstrom, A.R. 1962. The exact sampling distributions of least squares and maximum likelihood estimators of the marginal propensity to consume. Econometrica 30, 480−490.

Blomqvist, A.G. 1973. Hypothesis tests and confidence intervals for steady-state coefficients in models with lagged dependent variables: some notes on Fieller's Method. Oxford Bulletin of Economics and Statistics 35, 69−74.

Bockstael, N.E., Hanemann, W.M., Kling, C.L. 1987. Estimating the value of water quality improvements in a recreational demand framework. Water Resources Research 23, 951−960.

Bockstael, N.E., Strand, Jr. I.E. 1987. The effect of common sources of regression error on benefit estimates. Land Economics 63, 11−20.

Cameron, T.A. 1988. A New paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression. Journal of Economics and Management 15, 355−379.

Carson, R.T., Conaway, M. B., Hanemann, M.W., Krosnick, J.A., Mitchell, R.C., Presser, S. 2004. Valuing Oil Spill Prevention: A Case Study of California's Central Coast. Springer, Dordrecht, The Netherlands.

Cornwell, L.W., Aroian, L.A., Taneja, V.S. 1978. Numerical evaluation of the distribution of the product of two normal variables. Journal of Statistical Computation and Simulation 7, 123−131.

Craig, C.C., 1936. On the frequency function of xy. The Annals of Mathematical Statistics 7 , 1−15.

Curtiss, J. H., 1941. On the distribution of the quotient of two chance variables. The Annals of Mathematical Statistics 12, 409−421.

Daly, A., Hess, S., de Jong, G. 2012. Calculating errors for measures derived from choice modelling estimates. Transportation Research Part B: Methodological 46, 333−341.

Daly, A., Hess, S., Train, K. 2012. Assuring finite moments for willingness to pay in random coefficient models. Transportation 39, 19−31.

Dorfman, J.H., Kling, C.L., Sexton, R.J. 1990. Confidence intervals for elasticities and flexibilities: reevaluating the ratios of normals case. American Journal of Agricultural Economics 72, 1006−1017.

Dufour, J.-M. 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. Econometrica 65, 1365−1387.

Fiebig, D.G., 1985. Evaluating estimators without moments. The Review of Economics and Statistics 67, 529−534.

Fiebig, D.G., Keane, M.P. Louviere, J. Wasi, N. 2010. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. Marketing Science 29, 393−421.

Fieller, E.C. 1932. The distribution of the index in a normal bivariate population. Biometrika 24, 428−440.

———. 1954. Some problems in interval estimation. Journal of the Royal Statistical Society. Series B (Methodological) 16, 175−185.

Finney, D.J. 1971. Probit Analysis. Cambridge University Press, London.

Franz, V.H. 2007. Ratios: A Short Guide to Confidence Limits and Proper Use. Discussion paper, Department of Psychology, Universität Gießen.

Geary, R.C. 1930. The frequency distribution of the quotient of two normal variates. Journal of the Royal Statistical Society 93, 442−446.

Gelman, A. 2011. Fundamental difficulty of inference for a ratio when the denominator could be positive or negative. Statistical Modeling, Causal Inference, and Social Science' blog available at http://andrewgelman.com/2011/06/inference_for_a/.

Geweke, J. 1986. Exact inference in the inequality constrained normal linear regression model. Journal of Applied Econometrics 1, 127−141.

Glen, A.G., Leemis, L.M., Drew, J.H. 2004. Computing the distribution of the product of two continuous random variables. Computational Statistics & Data Analysis 44, 451−464.

Gleser, L.J., Hwang, J.T. 1987. The nonexistence of 100(1 - α)% confidence sets of finite expected diameter in errors-in-variables and related models. The Annals of Statistics 15 , 1351−1362.

Graham-Tomasi, T., Adamowicz, W.L., Fletcher, J.J. 1990. Errors of truncation in approximations to expected consumer surplus. Land Economics 66, 50.

Greene, W.H. 2011. Econometric Analysis. Prentice Hall, Upper Saddle River, NJ.

Greene, W.H., Hensher, D.A., Rose, J.M. 2005. Using Classical Simulation-Based Estimators to Estimate Individual WTP Values. A Mixed Logit Case Study of Commuters, in: Scarpa, R., Alberini, A. (eds), Applications of Simulation Methods in Environmental and Resource Economics. Dordrecht, Springer, pp. 17−33.

Haab, T., McConnell, K. 2003. The Econometrics Of Non-Market Valuation. Edward Elgar, Northampton, MA.

Hanemann, M.W., Kanninen, B. 2002. The Statistical Analysis of Discrete-Response CV Data, in: Bateman, I.J., Willis, K.G. (eds.), Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries. Oxford University Press, New York, pp. 302−441

Hensher, D. A., Rose, J. M., Greene, W.H. 2005. Applied Choice Analysis: A Primer. Cambridge, University Press Cambridge.

Hess, S., Train, K. 2011. Recovery of inter- and intra-personal heterogeneity using mixed logit models. Transportation Research Part B: Methodological 45, 973−990.

Hinkley, D.V. 1969. On the ratio of two correlated normal random variables. Biometrika 56, 635−639.

Hirschberg, J., Lye, J. 2010. A geometric comparison of the delta and Fieller confidence intervals. The American Statistician 64, 234−241.

Hole, A.R. 2007. A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Economics 16, 827−840.

Hwang, J.T. 1995. Fieller's problem and resampling techniques. Statistica Sinica 5, 161−171.

Johnson, N.L. 1949. Bivariate distributions based on simple translation systems. Biometrika 36, 297−304.

Khuri, A., Casella, G. 2002. The Existence of the First Negative Moment Revisited. The American Statistician 56, 44−47.

Kling, C.L. 1991. Estimating the precision of welfare measures. Journal of Environmental Economics and Management 21, 244−259.

Koschat, M.A. 1987. A characterization of the Fieller solution. The Annals of Statistics 15, 462−468.

Krinsky, I., Robb, A.L. 1986. On approximating the statistical properties of elasticities. The Review of Economics and Statistics 68, 715−719.

Krinsky, I., Robb, A.L. 1991. Three methods for calculating the statistical properties of elasticities: A comparison. Empirical Economics 16, 199−209.

Lai, C.D., Wood, G.R., Qiao, C.G. 2004. The mean of the inverse of a punctured normal distribution and its application. Biometrical Journal 46, 420−429.

Lehmann, E.L., Shaffer, J.P. 1988. Inverted distributions. The American Statistician 42, 191−194.

Lianos, T.P., Rausser, G.C. 1972. Approximate distribution of parameters in a distributed lag model. Journal of the American Statistical Association 67, 64−67.

Lisco, T. 1967. Value of Computer's Time: A Study in Urban Transportation. Ph.D. dissertation, Department of Economics, University of Chicago.

Louviere, J.J., Hensher, D.A., Swait, J.D. 2006. Stated Choice Methods: Analysis and Applications. Cambridge University Press, Cambridge.

Marsaglia, G., 1965. Ratios of normal variables and ratios of sums of uniform variables. Journal of the American Statistical Association 60, 193−204.

———. 2006. Ratios of normal variables. Journal of Statistical Software 16 (1), 10.

McFadden, D. 1974. Conditional Logit Analysis of Qualititative Choice Behaviour, in: Zarembka, P. (ed.), Frontiers in Econometrics, Academic Press, New York, NY, pp. 105−142.

Meijer, E., Rouwendal, J. 2006. Measuring welfare effects in models with random coefficients," Journal of Applied Econometrics 21, 227−244.

Merrill, A.S. 1928. Frequency distribution of an index when both the components follow the normal law. Biometrika 20A, 53−63.

Nicholson, C. 1941. A geometrical Analysis of the frequency distribution of the ratio between two variables. Biometrika 32, 16−28.

Oehlert, G.W. 1992. A note on the delta method. The American Statistician 46, 27−29.

Paige, R.L., Chapman, P.L., Butler, R.W. 2011. Small sample LD50 confidence intervals using saddlepoint approximations. Journal of the American Statistical Association 106, 334−344.

Park, T., Loomis, J.B., Creel, M. 1991. Confidence intervals for evaluating benefits estimates from dichotomous choice contingent valuation studies. Land Economics 67, 64.

Pham-Gia, T., Turkkan, N., Marchand, E. 2006. Density of the ratio of two normal random variables and applications. Communications in Statistics: Theory & Methods 35, 1569−1591.

Phillips, P.C.B., 1994. Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models. Econometrica 62, 73−93.

Piegorsch, W.W., Casella, G. 1985. The existence of the first negative moment. The American Statistician 39, 60−62.

Poe, G.L., Severance-Lossin, E.K., Welsh, M.P. 1994. Measuring the difference (X - Y) of simulated distributions: a convolutions approach. American Journal of Agricultural Economics 76, 904−915.

Poe, G.L., Welsh, M.P., Champ, P.A. 1997. Measuring the difference in mean willingness to pay when dichotomous choice contingent valuation responses are not independent. Land Economics 73, 255−267.

Revelt, D., Train, K. 1998. Mixed Logit with repeated choices: households' choices of appliance efficiency level. Review of Economics and Statistics 80, 647−657.

Rohatgi, V.K. 1976. An Introduction to Probability Theory and Mathematical Statistics. Wiley and Sons, New York.

Romano, J.P., Siegel, A.F. 1986. Counterexamples in Probability and Statistics. Wadsworth Press, Belmont, CA:.

Ruud, P.A. 2000. An Introduction to Classical Econometric Theory. Oxfrod University Press, New York, USA.

Scarpa, R., Thiene, M., Train, K. 2008. Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps. American Journal of Agricultural Economics 90, 994−1010.

Scheffe, H. 1970. Multiple testing versus multiple estimation. improper confidence sets. estimation of directions and ratios. The Annals of Mathematical Statistics 41, 1−29.

Smith, V.K. 1990. Estimating recreation demand using the properties of the implied consumer surplus. Land Economics 66, 111−120.

Springer, M. D., The Algebra of Random Variables (John Wiley & Sons, 1979).

Staiger, D., Stock, J.H., Watson, M.W. 1997. The NAIRU, unemployment and monetary policy. The Journal of Economic Perspectives 11, 33−49.

Stoyanov, J.M. 1997. Counterexamples in Probability. John Wiley & Sons, New York.

Train, K., Sonnier, G. 2005. Mixed Logit with Bounded Distributions of Correlated Partworths, in: Scarpa, R., Alberini, A. (eds), Applications of Simulation Methods in Environmental and Resource Economics. Dordredcht, Springer, pp. 117−134.

Train, K. E., Discrete Choice Methods with Simulation (New York: Cambridge University Press, 2009).

Train, K.E., Weeks, M. 2005. Discrete Choice Models in Preference Space and Willingness-to-pay Space, in: Scarpa, R., Alberini, A. (eds), Applications of Simulation Methods in Environmental and Resource Economics. Dordredcht, Springer.

Valentine, T.J. 1979. Hypothesis tests and confidence intervals for mean elasticities calculated from linear regression equations. Economics Letters 4, 363−367.

Varian, H.R 1992. Microeconomic Analysis. W. W. Norton, New York.

Wallace, D.L. 1980. The Behrens-Fisher and Fieller-Creasy Problems, in: Fienberg, S.E., Hinkley, D.V. (eds.), R.A. Fisher: An Appreciation. Springer, Berlin: Springer.

Ware, R., Lad, F. 2003. Approximating the Distribution for Sums of Products of Normal Variables. Department of Mathematics and Statistics, University of Canterbury, New Zealand: working paper available at: http://www.math.canterbury.ac.nz/research/ucdms2003n15.pdf.

Williams, D.A. 1986. Interval estimation of the median lethal dose. Biometrics 42, 641−645.

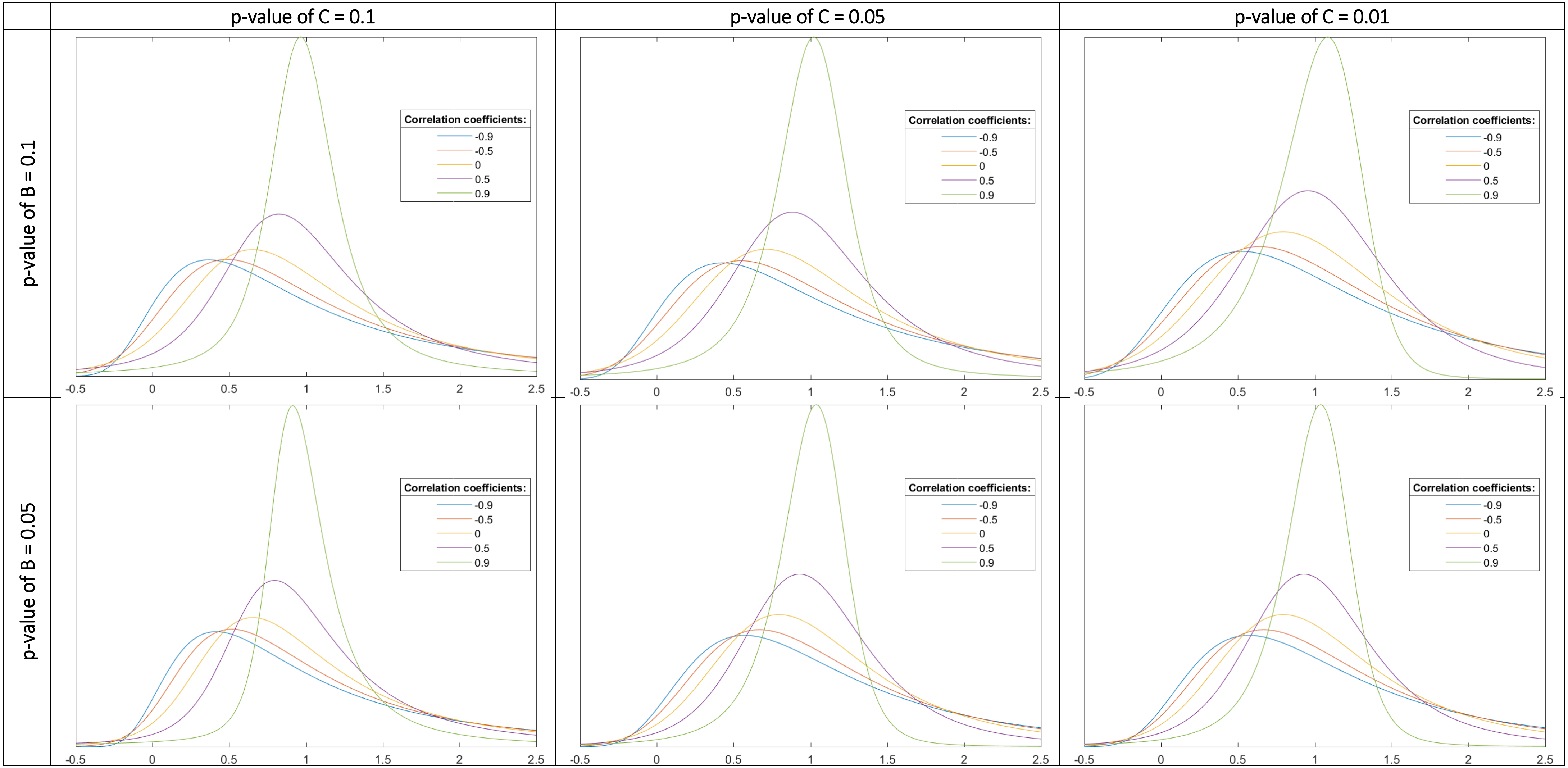
Woglom, G. 2001. More results on the exact small sample properties of the instrumental variable estimator. Econometrica 69, 1381−1389.

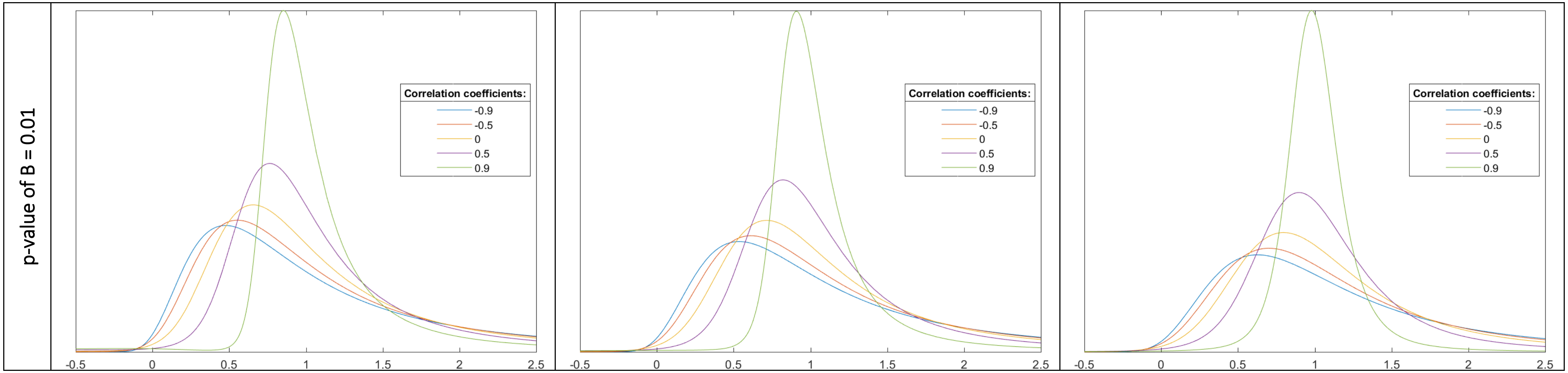
Zaman, A. 1981. Estimators without moments: The case of the reciprocal of a normal mean. Journal of Econometrics 15, 289−298.

Zellner, A. 1978. Estimation of functions of population means and regression coefficients including structural coefficients: A minimum expected loss (MELO) approach. Journal of Econometrics 8, 127−158.

Zerbe, G.O. 1978. On Fieller's theorem and the general linear model. The American Statistician 32, 103−105.

Appendix 1. Ratio Distribution for Different Correlation Coefficients of Normally Distributed Random Variables (Expressed in Terms of p-values)




p-value of B = 0.01
Correlation coefficients:
-0.9
-0.5
0
0.5
0.9
-0.5
0
0.5
1
1.5
2
2.5

Appendix 2. Pseudo-Moments and Quantiles of the Ratio Distribution for a Range of Correlation Coefficients and Coefficients of Variation of Normally Distributed Random variables (Expressed in Terms of p-values)

p-value of B	p-value of C	corr. coef.	μ_B/μ_C	pseudo-mean (KR)	median (KR)	pseudo-st.dev. (KR)	pseudo-st.dev. (delta)	95% c.i. of mean – KR (quantile range)	95% c.i. of mean – delta	95% c.i. of mean – Fieller
0.01	0.01	-0.90	1.00	1.36	0.99	$1.52\cdot10^{3}$	0.76	(0.12;6.18)	(-0.48;2.48)	(0.14;7.08)
0.01	0.01	-0.50	1.00	1.26	0.99	$1.06\cdot10^{3}$	0.67	(0.14;5.23)	(-0.32;2.32)	(0.17;5.96)
0.01	0.01	0.00	1.00	0.19	0.99	$1.21\cdot10^{4}$	0.55	(0.19;4.04)	(-0.08;2.08)	(0.22;4.53)
0.01	0.01	0.50	1.00	1.18	1.00	$5.00\cdot10^{2}$	0.39	(0.30;2.79)	(0.24;1.76)	(0.33;3.05)
0.01	0.01	0.90	1.00	1.03	1.00	$1.35\cdot10^{2}$	0.17	(0.59;1.63)	(0.66;1.34)	(0.60;1.68)
0.01	0.05	-0.90	1.00	1.50	0.95	$2.11\cdot10^{3}$	0.88	(-1.99;9.82)	(-0.72;2.72)	$(-\infty;-7.33\cdot10^{12})\cup(0.13;+\infty)$
0.01	0.05	-0.50	1.00	0.95	0.95	$3.99\cdot10^{3}$	0.78	(-0.45;8.34)	(-0.53;2.53)	$(-\infty;-6.01\cdot10^{13})\cup(0.15;+\infty)$
0.01	0.05	0.00	1.00	0.78	0.96	$4.19\cdot10^{3}$	0.64	(-0.50;6.46)	(-0.26;2.26)	$(-\infty;-4.35\cdot10^{13})\cup(0.21;+\infty)$
0.01	0.05	0.50	1.00	1.24	0.97	$2.16\cdot10^{3}$	0.46	(-0.19;4.53)	(0.10;1.90)	$(-\infty;-2.70\cdot10^{13})\cup(0.34;+\infty)$
0.01	0.05	0.90	1.00	1.23	0.99	$2.53\cdot10^{3}$	0.23	(0.40;2.79)	(0.54;1.46)	$(-\infty;-1.37\cdot10^{13})\cup(0.67;+\infty)$
0.01	0.10	-0.90	1.00	-1.26	0.89	$1.47\cdot10^{4}$	0.97	(-8.76;12.06)	(-0.91;2.91)	$(-\infty;-8.77)\cup(0.11;+\infty)$
0.01	0.10	-0.50	1.00	0.63	0.90	$5.79\cdot10^{3}$	0.87	(-7.08;10.36)	(-0.70;2.70)	$(-\infty;-7.07)\cup(0.14;+\infty)$
0.01	0.10	0.00	1.00	1.14	0.92	$3.88\cdot10^{3}$	0.72	(-4.98;8.18)	(-0.41;2.41)	$(-\infty;-4.97)\cup(0.20;+\infty)$
0.01	0.10	0.50	1.00	0.85	0.94	$1.81\cdot10^{3}$	0.53	(-3.00;5.99)	(-0.05;2.05)	$(-\infty;-2.94)\cup(0.34;+\infty)$
0.01	0.10	0.90	1.00	1.04	0.96	$8.91\cdot10^{2}$	0.31	(-1.54;4.06)	(0.39;1.61)	$(-\infty;-1.53)\cup(0.65;+\infty)$
0.05	0.01	-0.90	1.00	1.54	0.99	$1.73\cdot10^{3}$	0.88	(-0.03;6.96)	(-0.72;2.72)	$(-1.37\cdot10^{-14};8.00)$
0.05	0.01	-0.50	1.00	1.90	0.99	$5.30\cdot10^{3}$	0.78	(-0.03;5.75)	(-0.53;2.53)	$(-1.69\cdot10^{-14};6.56)$
0.05	0.01	0.00	1.00	1.37	0.99	$1.10\cdot10^{3}$	0.64	(-0.05;4.25)	(-0.26;2.26)	$(-2.27\cdot10^{-14};4.75)$
0.05	0.01	0.50	1.00	1.07	1.00	$3.58\cdot10^{2}$	0.46	(-0.03;2.76)	(0.10;1.90)	$(-3.69\cdot10^{-14};2.94)$
0.05	0.01	0.90	1.00	0.96	1.00	$7.09\cdot10^{1}$	0.23	(0.10;1.51)	(0.54;1.46)	$(-7.30\cdot10^{-14};1.50)$
0.05	0.05	-0.90	1.00	2.44	0.94	$1.21\cdot10^{4}$	0.99	(-0.42;10.96)	(-0.95;2.95)	$(-\infty;-8.27\cdot10^{13})\cup(0.00; ;+\infty)$
0.05	0.05	-0.50	1.00	2.27	0.95	$1.09\cdot10^{4}$	0.88	(-1.06;9.04)	(-0.73;2.73)	$(-\infty;-6.53\cdot10^{13})\cup(-4.83\cdot10^{-3};+\infty)$
0.05	0.05	0.00	1.00	1.10	0.96	$2.29\cdot10^{3}$	0.72	(-0.84;6.64)	(-0.41;2.41)	$(-\infty;-4.35\cdot10^{13})\cup(0.00;+\infty)$
0.05	0.05	0.50	1.00	0.98	0.97	$4.05\cdot10^{3}$	0.51	(-0.52;4.29)	(0.00;2.00)	$(-\infty;-2.18\cdot10^{13})\cup(1.21\cdot10^{-3};+\infty)$
0.05	0.05	0.90	1.00	102.36	0.99	$1.01\cdot10^{6}$	0.23	(0.16;2.17)	(0.55;1.45)	$(-\infty;-4.35\cdot10^{12})\cup(6.64\cdot10^{-3};+\infty)$
0.05	0.10	-0.90	1.00	0.21	0.87	$9.11\cdot10^{3}$	1.09	(-9.87;13.34)	(-1.14;3.14)	$(-\infty;-9.87)\cup(-1.06\cdot10^{-14};+\infty)$
0.05	0.10	-0.50	1.00	2.10	0.89	$7.54\cdot10^{3}$	0.97	(-7.60;11.12)	(-0.90;2.90)	$(-\infty;-7.60)\cup(-1.38\cdot10^{-14};+\infty)$
0.05	0.10	0.00	1.00	0.53	0.91	$5.85\cdot10^{3}$	0.79	(-4.87;8.36)	(-0.56;2.56)	$(-\infty;-4.76)\cup(-2.30\cdot10^{-14};+\infty)$

0.05	0.10	0.50	1.00	3.63	0.94	$1.89\cdot10^4$	0.57	(-2.48;5.64)	(-0.11;2.11)	$(-\infty;-1.93)\cup(-5.67\cdot10^{-14};+\infty)$
0.05	0.10	0.90	1.00	1.01	0.97	$7.76\cdot10^2$	0.27	(-0.54;3.08)	(0.48;1.52)	$(-\infty;0.00)\cup(0.34;+\infty)$
0.10	0.01	-0.90	1.00	1.79	0.99	$1.42\cdot10^3$	0.97	(-0.15;7.59)	(-0.91;2.91)	(-0.11;8.74)
0.10	0.01	-0.50	1.00	1.41	0.99	$1.45\cdot10^3$	0.87	(-0.18;6.18)	(-0.70;2.70)	(-0.14;7.05)
0.10	0.01	0.00	1.00	1.50	0.99	$1.71\cdot10^3$	0.72	(-0.26;4.45)	(-0.41;2.41)	(-0.20;4.95)
0.10	0.01	0.50	1.00	1.08	1.00	$6.66\cdot10^2$	0.53	(-0.35;2.81)	(-0.05;2.05)	(-0.34;2.94)
0.10	0.01	0.90	1.00	3.34	1.00	$2.39\cdot10^4$	0.31	(-0.45;1.55)	(0.39;1.61)	(-0.65;1.53)
0.10	0.05	-0.90	1.00	1.55	0.93	$6.51\cdot10^3$	1.09	(-0.56;11.88)	(-1.14;3.14)	$(-\infty;-9.02\cdot10^{13})\cup(-0.10;+\infty)$
0.10	0.05	-0.50	1.00	1.13	0.94	$9.56\cdot10^3$	0.97	(-1.16;9.60)	(-0.90;2.90)	$(-\infty;-6.94\cdot10^{13})\cup(-0.13;+\infty)$
0.10	0.05	0.00	1.00	1.39	0.95	$1.59\cdot10^3$	0.79	(-1.13;6.84)	(-0.56;2.56)	$(-\infty;-4.35\cdot10^{13})\cup(-0.21;+\infty)$
0.10	0.05	0.50	1.00	1.05	0.98	$1.64\cdot10^3$	0.57	(-0.87;4.25)	(-0.11;2.11)	$(-\infty;-1.76\cdot10^{13})\cup(-0.52;+\infty)$
0.10	0.05	0.90	1.00	0.93	1.00	$3.49\cdot10^2$	0.27	(-0.28;2.04)	(0.48;1.52)	$(-\infty;2.92)\cup(3.15\cdot10^{12};+\infty)$
0.10	0.10	-0.90	1.00	0.57	0.86	$8.31\cdot10^3$	1.19	(-10.76;14.36)	(-1.32;3.32)	$(-\infty;-10.76)\cup(-0.09;+\infty)$
0.10	0.10	-0.50	1.00	-0.69	0.88	$2.76\cdot10^4$	1.05	(-8.03;11.74)	(-1.06;3.06)	$(-\infty;-8.02)\cup(-0.12;+\infty)$
0.10	0.10	0.00	1.00	0.57	0.90	$5.65\cdot10^3$	0.86	(-4.86;8.57)	(-0.69;2.69)	$(-\infty;-4.54)\cup(-0.22;+\infty)$
0.10	0.10	0.50	1.00	22.06	0.94	$2.12\cdot10^5$	0.61	(-2.36;5.53)	(-0.19;2.19)	$(-\infty;+\infty)$
0.10	0.10	0.90	1.00	0.60	0.99	$3.42\cdot10^3$	0.27	(-0.44;2.71)	(0.47;1.53)	$(-\infty;+\infty)$

Appendix 3. Implementation of the Alternative Model Specification in IMDEP/NLOGIT and STATA

Our proposed model can be implemented without special programming in many statistical packages that can estimate mixed logit models. This can be done by declaring the variable that will serve as the denominator of the WTP ratio to have a random parameter that follows a log-normal distribution and then constraining the standard deviation of that parameter to be zero. Below we provide pseudo code for two commonly used statistical packages LIMPEP/NLOGIT and Stata. In these examples, Y represents the dependent variable while X1, …, Xc are choice attributes, of which Xc is the monetary attribute which in the alternative specification enters with an exponentiated parameter.

Limdep/Nlogit	**Stata**
Typical (MNL) specification	

`nlogit`
`     ; lhs = Y`
`     ; choices = ...`
`     ; model: U(*) = B*X1 + … + C*Xc`
`     ; …$`
	`clogit Y X1 … Xc, …`

The alternative specification with the parameter of variable Xc entering exponentially	

`nlogit`
`     ; lhs = Y`
`     ; choices = ...`
`     ; model: U(*) = B*X1 + … + C*Xc`
`     ; rpl`
`     ; fcn: C(l)`
`     ; sdv = 0`
`     ; …$`
	`constraint define 1 [SD]Xc = 0`
	`mixlogit Y X1 … Xc, rand(Xc)`
	`constraint(1) ln(1) …`

Note that this specification may, in some cases, require providing better starting values than those obtained from the MNL model. Specifically, the parameter estimate of logarithm of monetary attribute from a MNL model is usually a good starting value.

Appendix 4. Deriving formulas for standard errors and confidence sets of WTP under alternative model specifications

Let b and c be normally distributed estimates of B and C, such that:

$$\begin{bmatrix} b \\ c \end{bmatrix} : N\left\{\begin{bmatrix} B \\ C \end{bmatrix}, \begin{bmatrix} \sigma_B^2 & \sigma_{CB} \\ \sigma_{BC} & \sigma_C^2 \end{bmatrix}\right\}.$$

The estimate of the function $g(B,C)$ can be defined as $g(\mathrm{b},c)$, and approximated using the first-order Taylor series around (B,C). The asymptotic variance of g can thus be derived using the delta method (see formula (1) in the main text). The results for the cases of interest for this paper are provided in Table 5.[44]

Table 5. Asymptotic variance using the delta method

$g(B,C)$	Asymptotic variance
$\frac{B}{C}$	$\frac{1}{c^2}\sigma_B^2 - \frac{2b}{c^3}\sigma_{BC} + \frac{b^2}{c^4}\sigma_C^2$
$\frac{B}{\exp(C)}$	$\frac{\sigma_B^2 - 2b\sigma_{BC} + b^2\sigma_C^2}{\exp(2c)}$
$\frac{\exp(B)}{\exp(C)}$	$\exp(2b-2c)\left(\sigma_B^2 - 2\sigma_{BC} + \sigma_C^2\right)$

The Fieller bounds are derived by specifying the z-statistic. For illustration, consider the first case of interest here, i.e. $g(B,C) = B/C$. Whatever the true value of g:

$$Z = b - gc : N\left(0, \sigma_Z^2\right),$$

[44] The functions one may be particularly interested in are $g(B,C) = B/C$, $g(B,C) = B/\exp(C)$ and $g(B,C) = \exp(B)/\exp(C)$ which correspond to the usual (linear) model specification, the alternative specification with C constrained to be strictly positive (see equation (8) in the main text), and the alternative specification in which both B and C are constrained to be strictly positive, respectively.

where σ_Z^2 can be derived using formula (1), i.e. $\sigma_Z^2 = \begin{bmatrix} \frac{\partial Z}{\partial b} & \frac{\partial Z}{\partial c} \end{bmatrix} \times \begin{bmatrix} \sigma_B^2 & \sigma_{CB} \\ \sigma_{BC} & \sigma_C^2 \end{bmatrix} \times \begin{bmatrix} \frac{\partial Z}{\partial b} & \frac{\partial Z}{\partial c} \end{bmatrix}'$. In our first case, since $g(B,C) = B/C$ then $\sigma_Z^2 = \sigma_C^2 g^2 - 2\sigma_{BC} g + \sigma_B^2$. The z-statistic of interest is thus

$$z = \frac{b - gc}{\sigma_Z^2} .$$

Setting the desired confidence level requires using the appropriate value of z-statistics z_0 (e.g., for the 95% confidence interval $z_0 \approx 1.6449$) so that $\Pr(z^2 < z_0^2) = 0.95$. This yields the following inequality:

$$\left(z_0^2\sigma_C^2 - c^2\right)g^2 + \left(2bc - 2z_0^2\sigma_{BC}\right)g + z_0^2\sigma_B^2 - b^2 > 0 .$$

Solving it for g results in the confidence set.[45] The same algorithm can be applied to derive Fieller bounds for the other cases of interest for this paper. The results are provided in Table 6.

Finally, we note that in the case of random parameter models, if B is normally distributed one can still use formulas provided in Table 5 and 6, while substituting b for the estimate of the mean of B, and using the appropriate submatrix of the full asymptotic variance-covariance matrix of the model. If, however, B is assumed to be lognormally distributed, such that:

$$\begin{bmatrix} \mu_b \\ s_b \\ c \end{bmatrix} : N\left\{\begin{bmatrix} \mu_B \\ s_B \\ C \end{bmatrix}, \begin{bmatrix} \sigma_B^2 & \sigma_{s_B\mu_B} & \sigma_{C\mu_B} \\ \sigma_{\mu_B s_B} & \sigma_{s_B}^2 & \sigma_{Cs_B} \\ \sigma_{\mu_B C} & \sigma_{s_B C} & \sigma_C^2 \end{bmatrix}\right\} ,$$

where μ_B and s_B correspond to the mean and standard deviation of the underlying normal distribution of lognormally distributed B , the expected value of b is equal to $\exp\left(\mu_b + 0.5s_b^2\right)$. Applying the same procedures as described above, the formulas for deriving asymptotic variance using the delta method and Fieller confidence sets are presented in Table 7.

---

[45] In addition to the ‘typical’ bounded interval case, it is possible the set will include entire real line, or will be unbounded. This first case occurs if $z_0\sigma_B^2 - b^2 < 0$, i.e., if z-test of the hypothesis $c = 0$ is significant at the specified level. See Appendix 2 for an illustration.

Table 6. Fieller confidence sets under different model specifications

$g(B,C)$	Fieller confidence set – g such that:	Fieller bounds[46]
$\frac{B}{C}$	$\left(z_0^2\sigma_C^2 - c^2\right)g^2 + \left(2bc - 2z_0^2\sigma_{BC}\right)g + z_0^2\sigma_B^2 - b^2 > 0$	$\frac{z_0^2\sigma_{BC} - bc \pm z_0\sqrt{\left(\sigma_B^2\left(c^2 - z_0^2\sigma_C^2\right) + z_0^2\sigma_{BC}^2 - 2bc\sigma_{BC} + b^2\sigma_C^2\right)}}{z_0^2\sigma_C^2 - c^2}$
$\frac{B}{\exp(C)}$	$\left(\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)\right)g^2 + \left(2\exp(c)\left(\mathrm{b} - z_0^2\sigma_{BC}\right)\right)g + z_0^2\sigma_B^2 - b^2 > 0$	$\frac{\exp(c)\left(z_0^2\sigma_{BC} - b\right) \pm z_0\sqrt{\exp(2c)\left(\sigma_B^2\left(1 - z_0^2\sigma_C^2\right) + z_0^2\sigma_{BC}^2 - 2b\sigma_{BC} + b^2\sigma_C^2\right)}}{\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)}$
$\frac{\exp(B)}{\exp(C)}$	$\left(\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)\right)g^2 + \left(2\exp(b+c)\left(1 - z_0^2\sigma_{BC}\right)\right)g + \exp(2b)\left(z_0^2\sigma_B^2 - 1\right) > 0$	$\frac{\exp(b+c)\left(z_0^2\sigma_{BC} - 1\right) \pm z_0\sqrt{\exp\left(2(b+c)\right)\left(\sigma_B^2\left(1 - z_0^2\sigma_C^2\right) + z_0^2\sigma_{BC}^2 - 2\sigma_{BC} + \sigma_C^2\right)}}{\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)}$

[46] The interval is bounded if $z_0^2\sigma_C^2 - c^2 < 0$ or $z_0^2\sigma_C^2 - 1 < 0$, for the usual (B/C) or for the alternative ($B/\exp(C)$, $\exp(B)/\exp(C)$) specifications, respectively.

Table 7. Asymptotic variance using the delta method and Fieller confidence sets under two different model specifications for a random (lognormally distributed) numerator of WTP

$g(\mu_B, s_B, \mathrm{C})$	$\frac{\exp(\mu_B + 0.5 s_B^2)}{C}$
Asymptotic variance (delta method)	$\frac{\exp(s_b^2 + 2\mu_b)\left(c^2\sigma_B^2 + \sigma_C^2 + c\left(-2\sigma_{C\mu_B} + c s_b^2 \sigma_{s_B}^2 - 2 s_b\left(\sigma_{Cs_B} - c\sigma_{s_B\mu_B}\right)\right)\right)}{c^4}$
Fieller confidence set – g such that:	$\left(\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)\right)g^2 + \left(2\exp\left(c + \frac{s_b^2}{2} + \mu_b\right)\left(1 - s_b z_0^2 \sigma_{Cs_B} - z_0^2\sigma_{C\mu_B}\right)\right)g + \exp(s_b^2 + 2\mu_b)\left(s_b^2 z_0^2 \sigma_{s_B}^2 + 2 s_b z_0^2 \sigma_{s_B\mu_B} + z_0^2\sigma_B^2 - 1\right) > 0$
Fieller bounds[47]	$\frac{\exp\left(c + \frac{s_b^2}{2} + \mu_b\right)\left(s_b z_0^2 \sigma_{Cs_B} + z_0^2 \sigma_{C\mu_B} - 1\right) \pm z_0 \sqrt{\exp\left(s_b^2 + 2(c + \mu_b)\right)\left(\sigma_B^2\left(1 - z_0^2\sigma_C^2\right) - 2 s_b \sigma_{Cs_B} + s_b^2 z_0^2 \sigma_{Cs_B}^2 - 2\sigma_{C\mu_B} + 2 s_b z_0^2 \sigma_{Cs_B}\sigma_{C\mu_B} + z_0^2\sigma_{C\mu_B}^2 + s_b^2\sigma_{s_B}^2 + 2 s_b \sigma_{s_B\mu_B} - \sigma_C^2\left(-1 + s_b^2 z_0^2 \sigma_{s_B}^2 + 2 s_b z_0^2 \sigma_{s_B\mu_B}\right)\right)}}{\exp(2c)\left(-1 + z_0^2\sigma_C^2\right)}$
$g(\mu_B, s_B, \mathrm{C})$	$\frac{\exp(\mu_B + 0.5 s_B^2)}{\exp(C)}$
Asymptotic variance (delta method)	$\frac{\sigma_B^2 + \sigma_C^2 - 2 s_b \sigma_{Cs_B} - 2\sigma_{C\mu_B} + s_b^2\sigma_{s_B}^2 + 2 s_b \sigma_{s_B\mu_B}}{\exp\left(2c - s_b^2 - 2\mu_b\right)}$
Fieller confidence set – g such that:	$\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)g^2 + \left(2\exp\left(c + 0.5 s_b^2 + \mu_b\right)\left(1 - s_b z_0^2\sigma_{Cs_B} - z_0^2\sigma_{C\mu_B}\right)\right)g + \exp(s_b^2 + 2\mu_b)\left(s_b^2 z_0^2\sigma_{s_B}^2 + 2 s_b z_0^2 \sigma_{s_B\mu_B} + z_0^2\sigma_B^2 - 1\right) > 0$
Fieller bounds[48]	$\frac{\exp\left(c + \frac{s_b^2}{2} + \mu_b\right)\left(s_b z_0^2 \sigma_{Cs_B} + z_0^2 \sigma_{C\mu_B} - 1\right) \pm z_0 \sqrt{\exp\left(s_b^2 + 2(c + \mu_b)\right)\left(\sigma_B^2\left(1 - z_0^2\sigma_C^2\right) - 2 s_b \sigma_{Cs_B} + s_b^2 z_0^2 \sigma_{Cs_B}^2 - 2\sigma_{C\mu_B} + 2 s_b z_0^2 \sigma_{Cs_B}\sigma_{C\mu_B} + z_0^2\sigma_{C\mu_B}^2 + s_b^2\sigma_{s_B}^2 + 2 s_b \sigma_{s_B\mu_B} - \sigma_C^2\left(-1 + s_b^2 z_0^2 \sigma_{s_B}^2 + 2 s_b z_0^2 \sigma_{s_B\mu_B}\right)\right)}}{\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)}$

[47] The interval is bounded if $z_0^2\sigma_C^2 - c^2 < 0$.

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl