Journal of Environmental Economics and Management 95 (2019) 57–61

Contents lists available at ScienceDirect

Journal of Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

# 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, La Jolla, CA 92093-0508, USA
[b] Department of Economics, University of Warsaw, Poland

ARTICLE INFO

*Article history:*
Received 2 April 2018
Received in revised form 3 March 2019
Accepted 8 March 2019
Available online 9 March 2019

*JEL codes:*
C01
C15
C18
Q0

*Keywords:*
Conditional logit
Confidence intervals
Contingent valuation delta method
Discrete choice experiment
Krinsky-Robb
Multinomial logit
Probit
Welfare measures

ABSTRACT

We show a substantive problem exists with the widely-used ratio of coefficients approach to calculating willingness to pay (WTP) from discrete choice models. The correctly calculated standard error for WTP using this approach is shown to be undefined. This occurs because the cost parameter's standard error implies some possibility the true parameter value is arbitrarily close to zero. We propose a simple yet elegant way to overcome this problem by reparameterizing the (negative) cost variable's coefficient using an exponential transformation to enforce the theoretically correct positive coefficient. With it the confidence interval for WTP is now finite and well behaved.



## 1. Introduction

Use of the ratio of an utility function attribute parameter to a cost parameter to measure willingness to pay (WTP) for the attribute in a discrete choice model is a long-standing widely-used practice in applied economics (Hensher et al., 2015). However, as a researcher usually deals with maximum likelihood (ML) estimates of the utility function parameters, there is uncertainty associated with them. Taking this uncertainty into account in numerical approximation of the standard error associated with WTP generally requires the assumption that the ML estimates of the utility function parameters have an asymptotically normal distribution, with means equal to their ML estimates and standard deviations equal to their standard errors (Bockstael and Strand, 1987). Calculating moments of a resulting *ratio distribution* (e.g., the empirical distribution of WTP) becomes problematic.

* Corresponding author.
*E-mail addresses:* rcarson@ucsd.edu (R.T. Carson), mc@uw.edu.pl (M. Czajkowski).

https://doi.org/10.1016/j.jeem.2019.03.003
0095-0696/

Non-existence of moments of a ratio distribution resulting from dividing two normally distributed variables has long been known in statistics. It 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).[1] As a result, the mean and standard deviation of the resulting WTP ratio distribution are undefined, and 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. This problem applies to the estimated standard error of WTP in all commonly used discrete choice logit and probit models with non-random monetary parameter (e.g., multinomial logit (MNL), nested logit, latent class logit, or some mixed logit (MXL) specifications).

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.

## 2. Problems with the delta method and Krinsky-Robb parametric bootstrapping method to estimate statistics of WTP distribution

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 method.[2,3]

The delta method provides a finite well-behaved estimate of the asymptotic variance. However, it does so even in cases when approximate linearity for all likely values of the random variable does not hold and can produce finite estimates of infinite quantities. Further, delta method-based estimates of the confidence interval are always symmetric, even though the distribution of the WTP ratio variable can be quite asymmetric, particularly if the sample size is not large. Finally, the resulting confidence interval is always finite, while following Gleser and Hwang (1987) and Dufour (1997) any valid 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.

Given these potential problems with the delta method, it is important to know when it can produce a reliable confidence interval. While the WTP ratio distribution is Cauchy, it is reasonably approximated by a normal if the denominator (the negative of the cost parameter $C$) 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). Hirschberg and Lye (2010) provide a review of the relevant 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. They point out that most Monte Carlo simulation studies (e.g., Dorfman et al., 1990; Hole, 2007), 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 will often perform reasonably well.

The KR approach involves simulating multiple draws from the distribution of structural parameters of the WTP ratio variable. The ratio of the simulated coefficients provides empirical distribution of WTP, which is used for calculating its mean, median, standard deviation or quantiles. Since, however, the Cauchy-like ratio (WTP) distribution has undefined moments the calculated mean and standard deviation (which are in fact infinite) are unstable and tend to 'explode' with increasing the number of random draws.[4] We illustrate this with KR simulation results, 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 the correlation coefficient equal to 0.5. These results are shown in Table 1 along with the analytical results, the delta and Fieller estimates.

Many empirical studies 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, this 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 result, according to the Gleser and Hwang (1987) theorem it must not have positive confidence.[5]

[1] Online Appendix A provides an extended discussion and formal mathematical treatment of this issue.

[2] Online Appendix B provides a detailed discussion of the main methods used for providing confidence intervals of the ratio, including Fieller bounds, which are not considered in main body of the paper.

[3] Applied researchers sometimes use a non-parametric rather than parametric bootstrap. Because the standard non-parametric bootstrap by construction cannot result in an infinite confidence interval it has the similar problems to those of the parametric bootstrap.

[4] 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.

[5] Note that a bootstrap procedure based on Fieller bounds does have positive confidence because it can return an interval with infinite length (Hwang, 1995).

**Table 1**
Behavior of delta, Krinsky Robb, and Fieller 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[a] | 0.51 | 1[a] | (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 | 1194.51 | 0.97 | (-0.52; 4.31) |
| | 100,000,000 | −0.05 | 12,769.87 | 0.97 | (-0.52; 4.29) |
| Fieller | – | 1[a] | – | 1[a] | $(-\infty, -2.18 \cdot 10^{13}) \cup (0.00, +\infty)$ |

[a] Calculated as ratio of coefficients.

## 3. An alternative specification for the discrete choice model

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:

$$U_i(Alternative = j) = U_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + \gamma z_{ij} + \varepsilon_{ij}, \tag{1}$$

where $\mathbf{x}_{ij}$ is a vector of alternative-specific attributes and $z_{ij}$ is the cost associated with it (with $\boldsymbol{\beta}$ and $\gamma$ being utility function parameters). We reformulate this utility function so that the coefficient of the monetary attribute only has support with respect to positive 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.[6] A simple specification that does this is given by:

$$U_i(Alternative = j) = U_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + \exp(\delta) z_{ij} + \varepsilon_{ij}. \tag{2}$$

Estimation of the parameters of the model follows in the usual way. However, now the parameter associated with the cost attribute ($\delta$), which has an asymptotic normal distribution, is exponentiated. Since $\exp(\delta)$ is strictly positive, the ratio distribution representing the empirical distribution of WTP will have well-defined moments including the mean and standard deviation.[7] 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 $\gamma$ equal to the exponentiation of the estimate of $\delta$. As such, nothing has been done to alter the data collected. Rather, the economic theory-supported assumption about the sign of the cost coefficient has been made.[8] Note that the z-statistic for δ will be different than for $\lambda$, typically but not always larger.[9]

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 log-normally distributed, with the standard deviation of the cost parameter constrained to 0. Online Appendix C provides pseudo code for implementing our proposed model in this fashion in LIMDEP/NLOGIT and STATA and a link to an implementation using MATLAB code.[10]

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. Online Appendix D shows how the associated formulas can be derived for several typical cases. It is important to note that this restriction does not keep the WTP estimate from being extremely large due to a lack of responsiveness to cost because the point estimate for this parameter can be arbitrary close to zero.

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

7 Daly et al. (2012) show that lognormal distribution approaches zero at a rate which assures the existence of finite moments.

8 The imposed 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) 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.

9 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 $\lambda$ or $\exp(\delta)$ 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).

10 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.

Our proposed restriction on the cost parameter can be imposed in all discrete choice models when the cost parameter is specified as fixed rather than random parameter (e.g., MNL, nested logit, latent class logit, or MXL with non-random monetary attribute coefficient, which is a common practice in applied work[11]).

Two empirical examples of how the proposed method performs are included in online Appendix E. The first uses data from a binary discrete choice question about preventing oil spills off California's central coast (Carson et al., 2004) and the second uses data from a discrete choice experiment involving attributes that can influence alternative-fuel vehicle choice (Train and Sonnier, 2005). To briefly summarize, the prior information on the sign of the cost parameter our proposed method uses has little influence on estimated WTP measures and their confidence intervals with the large sample sizes characterizing the original studies. However, with a much smaller random sample drawn from the original data, our proposed method produces a substantially smaller confidence interval.[12] Further, restricting the sign of the numerator to also be positive can help stabilize estimates and produce useful confidence intervals in smaller samples.

## 4. 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 asymptotic sampling distribution of the ratio of the two ML parameter estimates is not, as often assumed, normal. Computing the ratio at the point estimates does not produce the expected value of the asymptotic distribution of the WTP nor its median. The expected value of WTP and its standard deviation are both undefined. The two workhorses the delta method and the KR approach used for estimating WTP confidence intervals do not produce valid estimates. For the delta method, the issue is producing a misleading finite and often reasonable appearing estimate of the standard error when the true quantity is undefined. The KR approach will always eventually show the degenerate nature of the WTP ratio with large enough sample sizes. Unfortunately, it too often produces plausible statistics for the number of replications typically used in applied work.

Another way that is sometimes suggested for this problem is to estimate the model in WTP space with WTP as direct argument in the log-likelihood function (Cameron, 1988; Train and Weeks, 2005). This solves the problem in the sense of estimating WTP under the assumption that it is normally distributed. However, this solution can mask the converse problem. Recovering the preference-space equivalents of key parameters of interest (e.g., to simulate market shares) involves simulating 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 (Ware and Lad, 2003). Its use raises a set of issues involving quantities like elasticities typically associated with the preference space that have not been explored. 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.[13]

We are clearly 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 because the delta method and the KR approach typically produce plausible confidence intervals.

We provide a simple, easy to implement solution to this problem by making selected preference parameters enter utility function exponentially. 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 in a wide variety of choice models. When the absolute value of the z-statistic on the cost parameter is very large (>8.75) (Finney, 1971), the confidence interval from the delta method and the one we propose will be virtually identical and the risk of a draw from the Krinsky-Robb estimator that causes the ratio WTP estimate to blow up are de minimis, so that main advantage from using our approach in this case is to ensure the statistical admissibility of the estimator. In the more typical case, imposition of the restriction on cost and other parameters will tighten, often substantially, the confidence intervals around expected WTP, suggesting the restriction(s) brings 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.

[11] Constraining the cost parameter to be non-random may be more problematic than addressing the confidence interval issue. However, there are cases when it may be reasonable. Train and Weeks (2005) provide an extensive list of empirical MXL examples using a fixed cost parameter. Motivations for this restriction include problems with the existence of the moments of the WTP distribution (Meijer and Rouwendal, 2006), wanting to restrict the monetary parameter to have the same sign for all individuals, ensuring that the distribution of the WTP has the theoretically expected sign, making the calculation of WTP (and associated confidence intervals) less burdensome, and making the identification easier, particularly in models allowing for correlated parameters, and in datasets with few observed choices per individual (Revelt and Train, 1998; Hess and Train, 2011).

[12] Equivalent result can be expected in the case there were fewer choice tasks per respondent, resulting in larger standard errors and hence larger bias of the WTP derived from the linear specification.

[13] Online Appendix F provides a more detailed discussion of some of the issues associated with using WTP-space approach, when other statistics that are now ratio variables, such as market shares, need to be calculated.

## Conflict of interest/financial interest

The authors declare that they have no conflict of interest with respect to the material contained in this manuscript nor any financial interest in the results.

## Acknowledgements

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. Mikolaj Czajkowski gratefully acknowledges the support of the National Science Centre of Poland (Opus 2017/25/B/HS4/01076).

## Appendix A. Supplementary materials

Supplementary materials to this article can be found online at https://doi.org/10.1016/j.jeem.2019.03.003.

## References

Bockstael, N.E., Strand Jr., I.E., 1987. The effect of common sources of regression error on benefit estimates. Land Econ. 63 (1), 11–20.
Cameron, T.A., 1988. A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression. J. Econ. Manag. 15 (3), 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.
Daly, A., Hess, S., Train, K., 2012. Assuring finite moments for willingness to pay in random coefficient models. Transportation 39 (1), 19–31.
Dorfman, J.H., Kling, C.L., Sexton, R.J., 1990. Confidence intervals for elasticities and flexibilities: reevaluating the ratios of normals case. Am. J. Agric. Econ. 72 (4), 1006–1017.
Dufour, J.-M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. Econometrica 65 (6), 1365–1387.
Finney, D.J., 1971. Probit Analysis, 3 Ed. Cambridge University Press, London.
Gelman, A., 2011. Fundamental difficulty of inference for a ratio when the denominator could be positive or negative. In: Statistical Modeling, Causal Inference, and Social Science blog available at. http://andrewgelman.com/2011/06/inference_for_a/.
Gleser, L.J., Hwang, J.T., 1987. The nonexistence of 100(1 - $\alpha$)% confidence sets of finite expected diameter in errors-in-variables and related models. Ann. Stat. 15 (4), 1351–1362.
Greene, W.H., 2011. Econometric Analysis, 7 Ed. Prentice Hall, Upper Saddle River, NJ.
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., 2015. Applied Choice Analysis, 2 Ed. Cambridge University Press, Cambridge, UK.
Hess, S., Train, K., 2011. Recovery of inter- and intra-personal heterogeneity using mixed logit models. Transp. Res. Part B Methodol. 45 (7), 973–990.
Hirschberg, J., Lye, J., 2010. A geometric comparison of the delta and Fieller confidence intervals. Am. Statistician 64 (3), 234–241.
Hole, A.R., 2007. A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Econ. 16 (8), 827–840.
Hwang, J.T., 1995. Fieller's problem and resampling techniques. Stat. Sin. 5, 161–171.
Krinsky, I., Robb, A.L., 1986. On approximating the statistical properties of elasticities. Rev. Econ. Stat. 68 (4), 715–719.
Lehmann, E.L., Shaffer, J.P., 1988. Inverted distributions. Am. Statistician 42 (3), 191–194.
Marsaglia, G., 2006. Ratios of normal variables. J. Stat. Softw. 16 (4), 1–10.
Meijer, E., Rouwendal, J., 2006. Measuring welfare effects in models with random coefficients. J. Appl. Econom. 21 (2), 227–244.
Revelt, D., Train, K., 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. Rev. Econ. Stat. 80 (4), 647–657.
Ruud, P.A., 2000. An Introduction to Classical Econometric Theory. Oxford University Press, New York, USA.
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. Springer Netherlands, pp. 117–134.
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. Springer, Dordrecht, pp. 1–16.
Varian, H.R., 1992. Microeconomic Analysis, 3 Ed. W. W. Norton, New York.
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.

## Appendix A. 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 :

$$h(w) = \int_{-\infty}^{+\infty} |q| f(wq,q)\,dq\,. \tag{1}$$

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:

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

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

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

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 $(\Phi)$ :

$$h(w) = \frac{K_3 K_4}{\sqrt{2\pi}\sigma_B\sigma_C K_1^3}\left(\Phi\left(\frac{K_3}{\sqrt{1-\rho^2}K_1}\right) - \Phi\left(\frac{-K_3}{\sqrt{1-\rho^2}K_1}\right)\right) + \frac{\sqrt{1-\rho^2}}{\pi\sigma_B\sigma_C K_1^2}\exp\left(\frac{-K_2}{2\left(1-\rho^2\right)}\right),$$

where: (4)

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

or in terms of the Kummer's confluent hypergeometric (Hermite) function $\left({}_1F_1\right)$ (Pham-Gia, Turkkan and Marchand, 2006).[1]

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

[1] 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.

result is applicable to a much wider range of contexts.[2] 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 $\Pr(C \le 0) = 0$ 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. Table A1 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.[3] The resulting ratio distribution is clearly not normal and often not symmetrical. The distribution can even be bimodal . 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, Table A2 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.[4] The true mean and standard deviation of this distribution are undefined.

---

[2] 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.

[3] The mean $(\mu)$ is normalized to 1. The standard deviation $(\sigma)$ is selected ensure the specified p-value.

[4] 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).

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.[5]

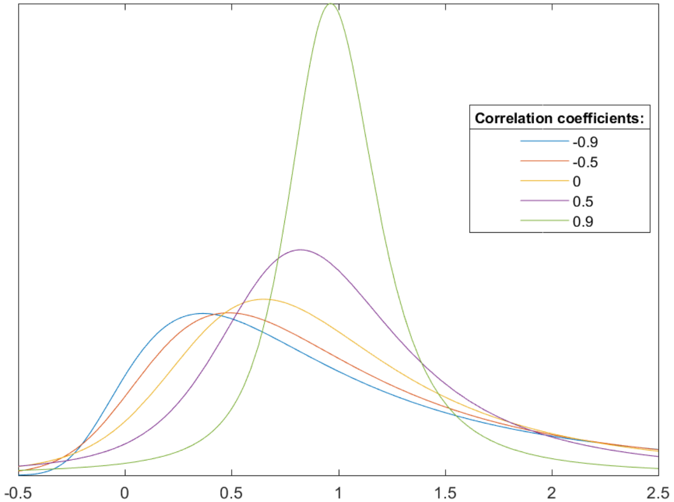
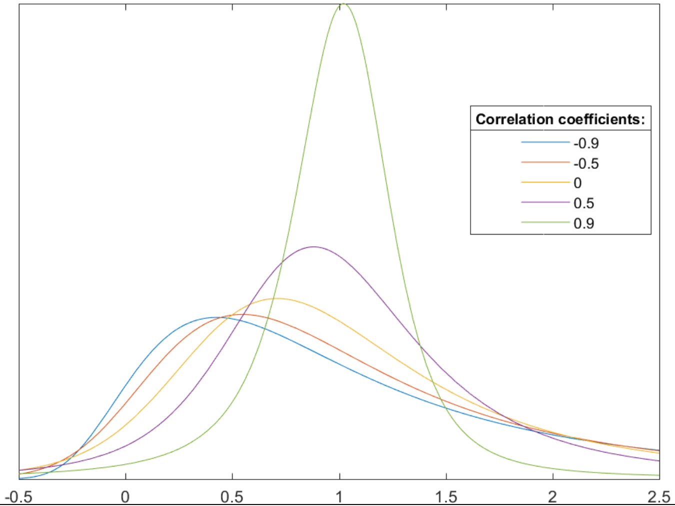
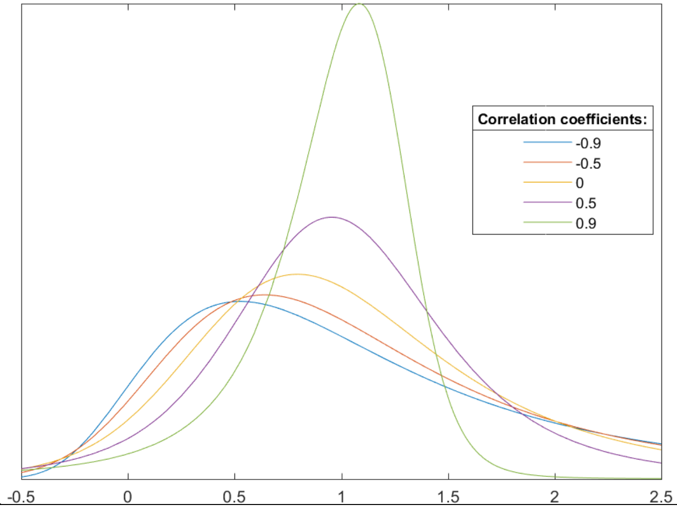
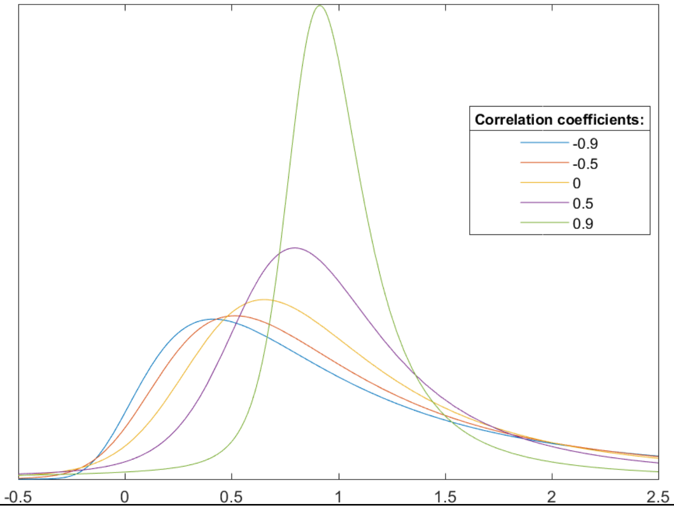
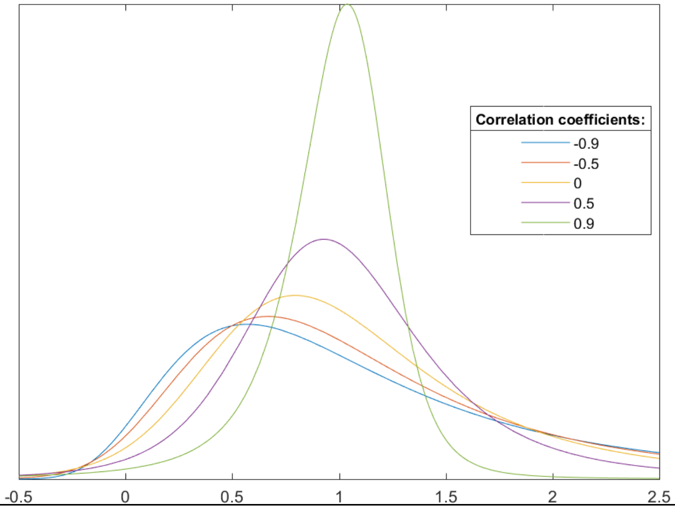
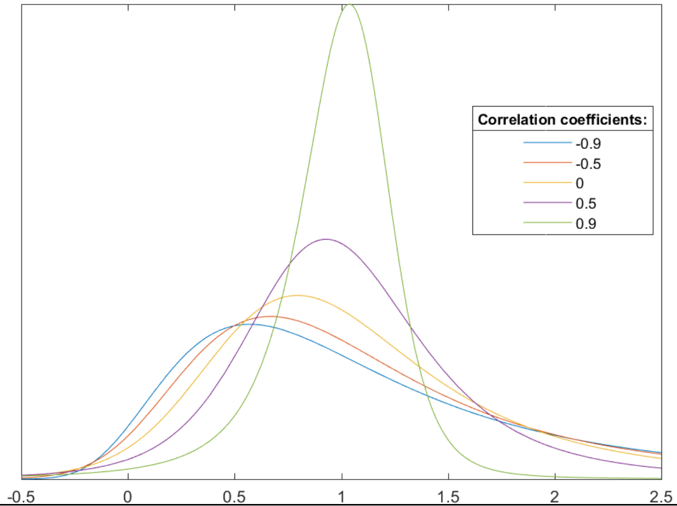
Table A2 reports the 'pseudo-standard deviation', and 2.5'th and 97.5'th percentiles of the distribution along with the delta, KR, and Fieller confidence intervals. The delta interval mistakenly suggests the process is well behaved. The KR-based interval is sometimes substantially different from the delta one, but it too appears well behaved. The admissible Fieller bounds show that 95% confidence intervals frequently include $\pm\infty$.

To summarize, under the usual specification of a discrete choice model: (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.[6]

[5] 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.

[6] 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.

**Table A1. Ratio Distribution for Different Correlation Coefficients of Normally Distributed Random Variables (Expressed in Terms of p-values)**

<table>
<tr><th></th><th>p-value of C = 0.1</th><th>p-value of C = 0.05</th><th>p-value of C = 0.01</th></tr>
<tr><td>p-value of B = 0.1</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>p-value of B = 0.05</td><td>
</td><td>
</td><td>
</td></tr>
</table>

p-value of B = 0.01

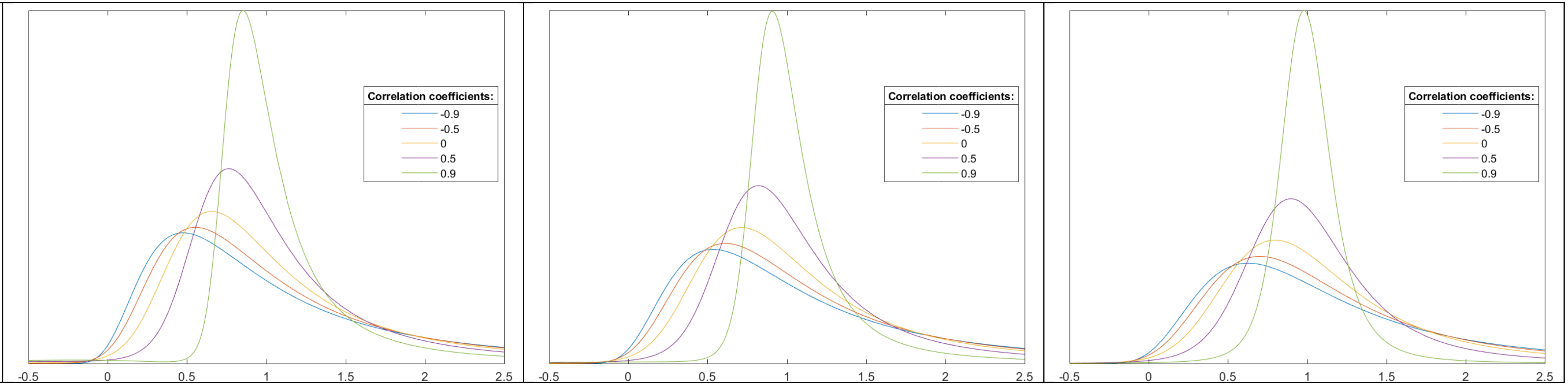


**Table A2. Pseudo-Moments and Quantiles of the Ratio Distribution for a Range of Correlation Coefficients and Coefficients of Variation of Normally Distributed Random variables (Expressed in Terms of p-values)**

| p-value of B | p-value of C | corr. coef. | $\mu_B/\mu_C$ | 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 | $1.89 \cdot 10^{4}$ | 0.57 | (-2.48;5.64) | (-0.11;2.11) | $(-\infty;-1.93)\cup(-5.67 \cdot 10^{-14};+\infty)$ |
| 0.05 | 0.10 | 0.90 | 1.00 | 1.01 | 0.97 | $7.76 \cdot 10^{2}$ | 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 | $1.42 \cdot 10^{3}$ | 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 | $1.45 \cdot 10^{3}$ | 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 | $1.71 \cdot 10^{3}$ | 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 | $6.66 \cdot 10^{2}$ | 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 | $2.39 \cdot 10^{4}$ | 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 | $6.51 \cdot 10^{3}$ | 1.09 | (-0.56;11.88) | (-1.14;3.14) | $(-\infty;-9.02 \cdot 10^{13})\cup(-0.10;+\infty)$ |
| 0.10 | 0.05 | -0.50 | 1.00 | 1.13 | 0.94 | $9.56 \cdot 10^{3}$ | 0.97 | (-1.16;9.60) | (-0.90;2.90) | $(-\infty;-6.94 \cdot 10^{13})\cup(-0.13;+\infty)$ |
| 0.10 | 0.05 | 0.00 | 1.00 | 1.39 | 0.95 | $1.59 \cdot 10^{3}$ | 0.79 | (-1.13;6.84) | (-0.56;2.56) | $(-\infty;-4.35 \cdot 10^{13})\cup(-0.21;+\infty)$ |
| 0.10 | 0.05 | 0.50 | 1.00 | 1.05 | 0.98 | $1.64 \cdot 10^{3}$ | 0.57 | (-0.87;4.25) | (-0.11;2.11) | $(-\infty;-1.76 \cdot 10^{13})\cup(-0.52;+\infty)$ |
| 0.10 | 0.05 | 0.90 | 1.00 | 0.93 | 1.00 | $3.49 \cdot 10^{2}$ | 0.27 | (-0.28;2.04) | (0.48;1.52) | $(-\infty;2.92)\cup(3.15 \cdot 10^{12};+\infty)$ |
| 0.10 | 0.10 | -0.90 | 1.00 | 0.57 | 0.86 | $8.31 \cdot 10^{3}$ | 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 | $2.76 \cdot 10^{4}$ | 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 | $5.65 \cdot 10^{3}$ | 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 | $2.12 \cdot 10^{5}$ | 0.61 | (-2.36;5.53) | (-0.19;2.19) | $(-\infty;+\infty)$ |
| 0.10 | 0.10 | 0.90 | 1.00 | 0.60 | 0.99 | $3.42 \cdot 10^{3}$ | 0.27 | (-0.44;2.71) | (0.47;1.53) | $(-\infty;+\infty)$ |

## Appendix B. Approaches for providing confidence intervals for a ratio

The delta method (Greene, 2011) 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. 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:

$$\text{as.var}\left(g(b)\right)=\frac{\partial g(\beta)}{\partial \beta}\Sigma\left(\frac{\partial g(\beta)}{\partial \beta}\right)', \tag{5}$$

where $b$ is the estimator of the parameters $\beta$ and the asymptotic variance covariance matrix is $\Sigma$.

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

$$\text{as.var}\left(\frac{B}{C}\right)=\frac{\mu_B^2}{\mu_C^4}\sigma_C^2+\frac{1}{\mu_C^2}\sigma_B^2-\frac{2\mu_B}{\mu_C^3}\rho\sigma_B\sigma_C \ . \tag{6}$$

For an early and highly cited application of the Krinsky and Robb (KR; 1986) parametric bootstrapping method for calculating the confidence interval for the ratio of parameters see . This application was motivated by the well-recognized potential problems with the delta method in this case.

A third method, the Fieller confidence interval , is often used in the biometrics literature. Despite some proponents (Dufour, 1997), it is used infrequently in econometrics (some notable examples include Blomqvist, 1973; Valentine, 1979; Staiger, Stock and Watson, 1997). 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, while Lye and Hirschberg (2018) offer additional guidance regarding the use of Fieller statistics and when these diverge from the delta method.

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). Other solutions that make the problem less pronounced include making marginal utility of money income-dependent (Morey, Sharma and Karlstrom, 2003; Giergiczny et al., 2012). New approaches for providing confidence intervals in small samples continue to be proposed (e.g., Paige, Chapman and Butler, 2011).

**Discussion**

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]

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 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

[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 the seminal Bockstael, Hanemann and Kling (1987) paper. Adamowicz, Fletcher and Graham-Tomasi (1989) impose non-negativity on the consumer surplus estimate using Geweke's 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, Adamowicz and Fletcher, 1990). Variants of the 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. 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.

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, Wood and Qiao (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.

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., Dorfman, Kling and Sexton, 1990; Hole, 2007), 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

[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 $(t_B - \rho t_C)/(1-\rho^2)^{.5}$ 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 $\rho$ 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 $\rho$ is not equal to zero, the case where the ratio variable differs in sign from $\rho$ 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.

the true value of the ratio (Zaman, 1981; Fiebig, 1985). Thus, while the delta method's estimate of the variance 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]

[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 KR application, calculating elasticities, did not involve ratios of parameters (Krinsky and Robb, 1986).

**Appendix C. Implementing alternative model specification in LIMDEP/NLOGIT, MATLAB 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.

<table>
<tr><th>Limdep/Nlogit</th><th>Stata</th></tr>
<tr><td colspan="2">Typical (MNL) specification</td></tr>
<tr><td><pre>nlogit
    ; lhs = Y
    ; choices = ...
    ; model: U(*) = B*X1 + … + C*Xc
    ; …$</pre></td><td><pre>clogit Y X1 … Xc, …</pre></td></tr>
<tr><td colspan="2">The alternative specification with the parameter of variable Xc entering exponentially</td></tr>
<tr><td><pre>nlogit
    ; lhs = Y
    ; choices = ...
    ; model: U(*) = B*X1 + … + C*Xc
    ; rpl
    ; fcn: C(l)
    ; sdv = 0
    ; …$</pre></td><td><pre>constraint define 1 [SD]Xc = 0
mixlogit Y X1 … Xc, rand(Xc)
constraint(1) ln(1) …</pre></td></tr>
</table>

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.

The DCE estimation package developed in Matlab, available from github.com/czaj/DCE under Creative Commons BY 4.0 license, also has the ability to implement this specification we propose as well as the delta and Fieller methods for the calculating confidence intervals / sets.

## Appendix D. 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:

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

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 D1.[14]

**Table D1. 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 $g$:

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

where $\sigma_Z^2$ 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

[14] 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 (2) in the main text), and the alternative specification in which both $B$ and $C$ are constrained to be strictly positive, respectively.

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

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 $\Pr\left(z^2 < z_0^2\right) = 0.95$ . 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.[15] 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 D1 and D2, 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 log-normally distributed, such that:

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

where $\mu_B$ and $s_B$ correspond to the mean and standard deviation of the underlying normal distribution of log-normally 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

[15] 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.

asymptotic variance using the delta method and Fieller confidence sets are presented in Table D3.

**Table D2. Fieller confidence sets under different model specifications**

| $g(B,C)$ | Fieller confidence set – $g$ such that: | Fieller bounds[16] |
|---|---|---|
| $\frac{B}{C}$ | $\left(z_0^2\sigma_C^2 - c^2\right)g^2 + \left(2bc - 2z_0^2\sigma_{BC}\right)g + z_0^2\sigma_B^2 - b^2 > 0$ | $\frac{z_0^2\sigma_{BC} - bc \pm z_0\sqrt{\left(\sigma_B^2\left(c^2 - z_0^2\sigma_C^2\right) + z_0^2\sigma_{BC}^2 - 2bc\sigma_{BC} + b^2\sigma_C^2\right)}}{z_0^2\sigma_C^2 - c^2}$ |
| $\frac{B}{\exp(C)}$ | $\left(\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)\right)g^2 + \left(2\exp(c)\left(\mathrm{b} - z_0^2\sigma_{BC}\right)\right)g + z_0^2\sigma_B^2 - b^2 > 0$ | $\frac{\exp(c)\left(z_0^2\sigma_{BC} - b\right) \pm z_0\sqrt{\exp(2c)\left(\sigma_B^2\left(1 - z_0^2\sigma_C^2\right) + z_0^2\sigma_{BC}^2 - 2b\sigma_{BC} + b^2\sigma_C^2\right)}}{\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)}$ |
| $\frac{\exp(B)}{\exp(C)}$ | $\left(\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)\right)g^2 + \left(2\exp(b+c)\left(1 - z_0^2\sigma_{BC}\right)\right)g + \exp(2b)\left(z_0^2\sigma_B^2 - 1\right) > 0$ | $\frac{\exp(b+c)\left(z_0^2\sigma_{BC} - 1\right) \pm z_0\sqrt{\exp\left(2(b+c)\right)\left(\sigma_B^2\left(1 - z_0^2\sigma_C^2\right) + z_0^2\sigma_{BC}^2 - 2\sigma_{BC} + \sigma_C^2\right)}}{\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)}$ |

[16] 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 D3. Asymptotic variance using the delta method and Fieller confidence sets under two different model specifications for a random (lognormally distributed) numerator of WTP**

| | |
|---|---|
| $g\left(\mu_B, s_B, \mathrm{C}\right)$ | $\dfrac{\exp\left(\mu_B + 0.5 s_B^2\right)}{C}$ |
| Asymptotic variance (delta method) | $\dfrac{\exp\left(s_b^2 + 2\mu_b\right)\left(c^2\sigma_B^2 + \sigma_C^2 + c\left(-2\sigma_{C\mu_B} + c s_b^2 \sigma_{s_B}^2 - 2 s_b\left(\sigma_{Cs_B} - c\sigma_{s_B\mu_B}\right)\right)\right)}{c^4}$ |
| Fieller confidence set − $g$ such that: | $\left(\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)\right)g^2 + \left(2\exp\left(c + \dfrac{s_b^2}{2} + \mu_b\right)\left(1 - s_b z_0^2 \sigma_{Cs_B} - z_0^2\sigma_{C\mu_B}\right)\right)g + \exp\left(s_b^2 + 2\mu_b\right)\left(s_b^2 z_0^2 \sigma_{s_B}^2 + 2 s_b z_0^2 \sigma_{s_B\mu_B} + z_0^2\sigma_B^2 - 1\right) > 0$ |
| Fieller bounds[17] | $\exp\left(c + \dfrac{s_b^2}{2} + \mu_b\right)\left(s_b z_0^2 \sigma_{Cs_B} + z_0^2 \sigma_{C\mu_B} - 1\right) \pm \dfrac{z_0\sqrt{\exp\left(s_b^2 + 2\left(c + \mu_b\right)\right)\left(\begin{array}{l}\sigma_B^2\left(1 - z_0^2\sigma_C^2\right) - 2 s_b \sigma_{Cs_B} + s_b^2 z_0^2 \sigma_{Cs_B}^2 - 2\sigma_{C\mu_B} + 2 s_b z_0^2 \sigma_{Cs_B}\sigma_{C\mu_B} + \\ z_0^2\sigma_{C\mu_B}^2 + s_b^2\sigma_{s_B}^2 + 2 s_b \sigma_{s_B\mu_B} - \sigma_C^2\left(-1 + s_b^2 z_0^2 \sigma_{s_B}^2 + 2 s_b z_0^2 \sigma_{s_B\mu_B}\right)\end{array}\right)}}{\exp(2c)\left(-1 + z_0^2\sigma_C^2\right)}$ |
| $g\left(\mu_B, s_B, \mathrm{C}\right)$ | $\dfrac{\exp\left(\mu_B + 0.5 s_B^2\right)}{\exp\left(C\right)}$ |
| Asymptotic variance (delta method) | $\dfrac{\sigma_B^2 + \sigma_C^2 - 2 s_b \sigma_{Cs_B} - 2\sigma_{C\mu_B} + s_b^2 \sigma_{s_B}^2 + 2 s_b \sigma_{s_B\mu_B}}{\exp\left(2c - s_b^2 - 2\mu_b\right)}$ |
| Fieller confidence set − $g$ such that: | $\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)g^2 + \left(2\exp\left(c + 0.5 s_b^2 + \mu_b\right)\left(1 - s_b z_0^2 \sigma_{Cs_B} - z_0^2\sigma_{C\mu_B}\right)\right)g + \exp\left(s_b^2 + 2\mu_b\right)\left(s_b^2 z_0^2 \sigma_{s_B}^2 + 2 s_b z_0^2 \sigma_{s_B\mu_B} + z_0^2\sigma_B^2 - 1\right) > 0$ |
| Fieller bounds[48] | $\exp\left(c + \dfrac{s_b^2}{2} + \mu_b\right)\left(s_b z_0^2 \sigma_{Cs_B} + z_0^2 \sigma_{C\mu_B} - 1\right) \pm \dfrac{z_0\sqrt{\exp\left(s_b^2 + 2\left(c + \mu_b\right)\right)\left(\begin{array}{l}\sigma_B^2\left(1 - z_0^2\sigma_C^2\right) - 2 s_b \sigma_{Cs_B} + s_b^2 z_0^2 \sigma_{Cs_B}^2 - 2\sigma_{C\mu_B} + 2 s_b z_0^2 \sigma_{Cs_B}\sigma_{C\mu_B} + \\ z_0^2\sigma_{C\mu_B}^2 + s_b^2\sigma_{s_B}^2 + 2 s_b \sigma_{s_B\mu_B} - \sigma_C^2\left(-1 + s_b^2 z_0^2 \sigma_{s_B}^2 + 2 s_b z_0^2 \sigma_{s_B\mu_B}\right)\end{array}\right)}}{\exp(2c)\left(z_0^2\sigma_C^2 - 1\right)}$ |

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

## Appendix E. 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.[18] The first is a large contingent valuation (CV) survey designed to value WTP to prevent oil spills along California's 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).[19] [20]

### 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.[21] Table E1 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).[22] 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.

---

[18] 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.

[19] We use the 1484 choice observations from 100 respondents included in the dataset available at Kenneth Train's website: http://elsa.berkeley.edu/~train/.

[20] The code and data for estimating the models presented in this paper are available from http://czaj.org/research/supplementary-materials.

[21] In estimation, we use negative of the actual cost divided by 100.

[22] Note that while restricting the sign of the cost parameter is clearly supported by theory, restricting the sign of an attribute (or the ASC) is more problematic. Theory does not necessarily provide guidance for whether individuals have positive or negative preferences for many attributes. While we include such a case as a demonstration, we caution against imposing restrictions that are invalid, unwarranted, and may result in a loss of information.

Table E1. Results from Typical and Two Alternative Specifications (Complete Dataset)[23]

| | 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*** | 0.5602*** | -0.5794*** |
| | (0.0934) | (0.0934) | (0.1667) |
| $C$ – cost associated with introducing the scenario | 0.7152*** | -0.3351*** | -0.3351*** |
| | (0.0845) | (0.1181) | (0.1181) |
| Ratio of coefficients[24] | 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) |
| 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.[25] Since probability of $C \leq 0$ 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.[26]

[23] ***, **, * – Significance at 1%, 5%, 10% level; standard errors in parentheses.

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

[25] 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.

[26] 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.

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 E2.

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.

Table E2. Results from Typical and Two Alternative Specification ($n$=100 Subsample of Data)

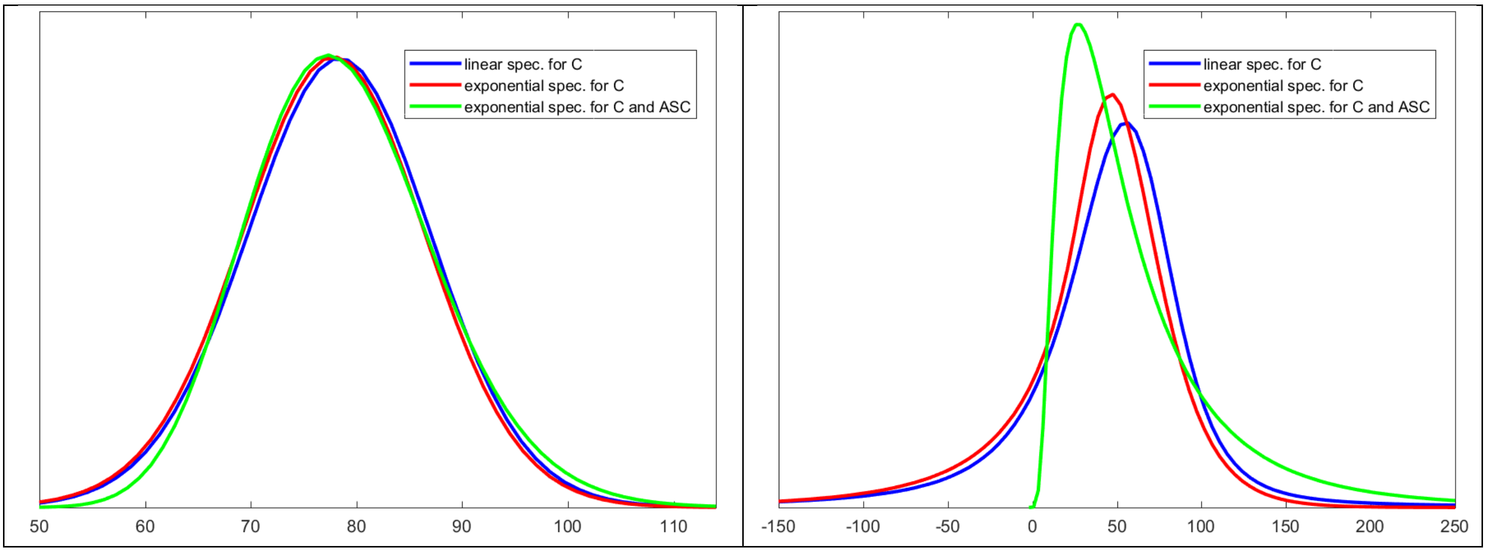
| | MNL – typical specification (cost enters linearly) | MNL – alternative specification 1 (cost enters exponentially) | MNL – alternative specification 2 (ASC and cost enter exponentially) |
|---|---|---|---|
| $B$ – ASC associated with introducing the scenario | 0.2843<br>(0.2964) | 0.2843<br>(0.2964) | -1.2578<br>(1.0431) |
| $C$ – cost associated with introducing the scenario | 0.5879**<br>(0.2954) | -0.5312<br>(0.5026) | -0.5312<br>(0.5027) |
| Ratio of coefficients[27] | 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 E1 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 E1, 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 E1. 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

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

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 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 E1. 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




**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 E3 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 ($c\_operate$).[28]

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

Table E3. Results for Range from Typical and Alternative Specification [29]

| | MXL – typical specification (cost parameters enter linearly) | | MXL – alternative specification (cost parameters enter exponentially) | |
|---|---|---|---|---|
| | Means | Standard deviations | Means | Standard deviations |
| *range* | -0.7328 | 0.5915** | -0.7328 | 0.5915** |
| (log-normally distributed) | (0.4636) | (0.2980) | (0.4519) | (0.2917) |
| *electric* | -1.7908*** | 1.2658*** | -1.7908*** | 1.2658*** |
| (normally distributed) | (0.3426) | (0.2156) | (0.3380) | (0.2153) |
| *hybrid* | 0.4395*** | 0.9745*** | 0.4395*** | 0.9745*** |
| (normally distributed) | (0.1694) | (0.1369) | (0.1693) | (0.1368) |
| $p\_medium$ | 0.5310*** | 0.5696*** | 0.5310*** | 0.5696*** |
| (normally distributed) | (0.1134) | (0.1420) | (0.1133) | (0.1420) |
| $p\_high$ | 0.0770 | 0.3514** | 0.0770 | 0.3514** |
| (normally distributed) | (0.0999) | (0.1675) | (0.1000) | (0.1677) |
| $c\_purchase$ | 0.4748*** | – | -0.7448*** | – |
| (fixed) | (0.0380) | | (0.0799) | |
| $c\_operate$ | 0.0136*** | – | -4.3006*** | – |
| (fixed) | (0.0039) | | (0.2902) | |
| Ratio of coefficients[30] | 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 parameters entering linearly.[31] Even though the medians of WTP distributions are similar, the alternative specification assures the existence of finite moments of WTP. Restricting the cost

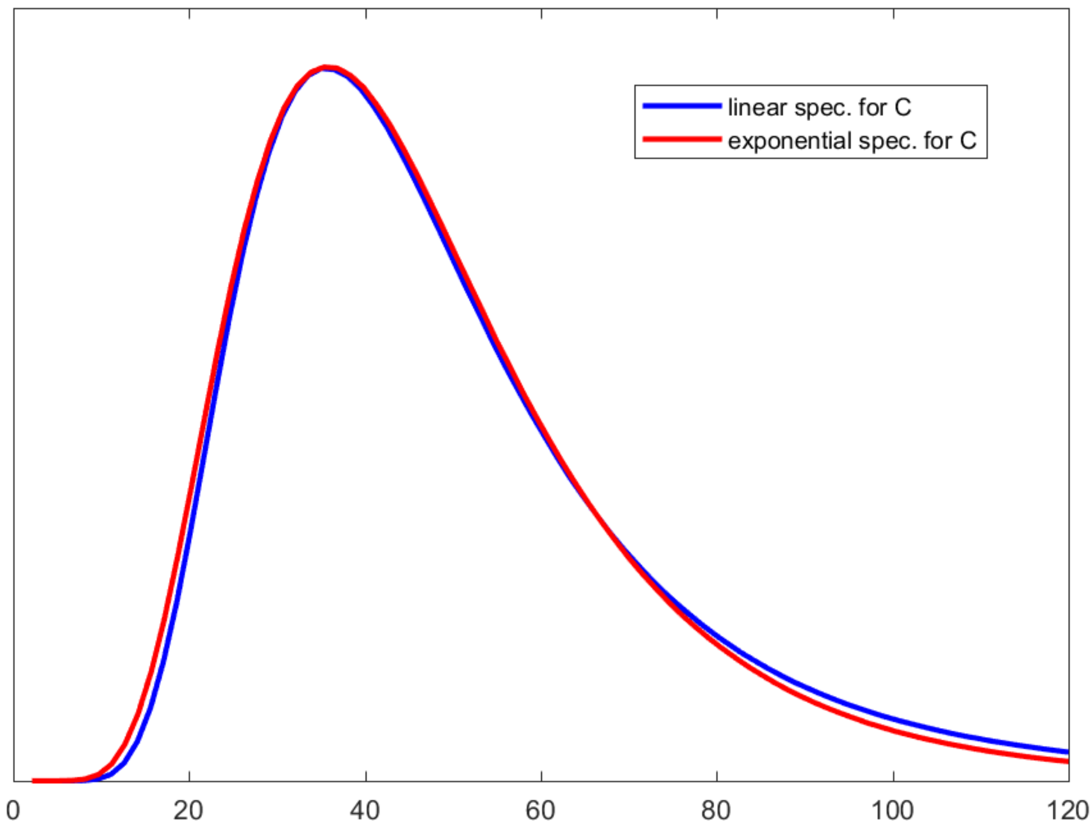
[29] ***, **, * – Significance at 1%, 5%, 10% level; standard errors in parentheses. For log-normally distributed parameters, estimates of the mean and standard deviation of the underlying normal distribution are provided.

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

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

parameter to be strictly positive allows for the moments (mean, standard deviation) of the ratio distribution to be well defined. Figure E2 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 E2. Probability Density Function of Empirical Distribution of E(WTP) ( $range/c\_operate$ )



**Appendix F. WTP-space approach to deal with problems associated with calculating WTP as a ratio**

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.[32] 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, Thiene and Train, 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

[32] 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. It is worth noting here that the use of ln($C$) as a fixed regressor in a logit or probit model does avoid the issue with WTP being undefined but only by implicitly assuming that WTP has a log-normal distribution.

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

Using the notation of (1), 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:

$$U_i\left(Alternative = j\right) = U_{ij} = \left(\gamma_i \mathbf{w}_i\right)' \mathbf{x}_{ij} + \gamma_i z_{ij} + \varepsilon_{ij} \ . \tag{7}$$

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.[33]

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

$$h\left(b\right) = \int \frac{1}{|q|} f\left(\frac{b}{q}, q\right) dq \tag{8}$$

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.

[33] 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, Aroian and Taneja, 1978; Glen, Leemis and Drew, 2004).

## Appendix G. References used in appendices


Adamowicz, W. L., Fletcher, J. J., and Graham-Tomasi, T., 1989. Functional Form and the Statistical Properties of Welfare Measures. *American Journal of Agricultural Economics*, 71(2):414-421.

Aroian, L. A. (1947). The Probability function of the product of two normally distributed variables. *Annals of Mathematical Statistics*, 18, 265-271.

Bergstrom, A. R., 1962. The Exact Sampling Distributions of Least Squares and Maximum Likelihood Estimators of the Marginal Propensity to Consume. *Econometrica*, 30(3):480-490.

Blomqvist, A. G., 1973. Hypothesis Tests aAnd 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(1):69-74.

Bockstael, N. E., Hanemann, W. M., and Kling, C. L., 1987. Estimating the value of water quality improvements in a recreational demand framework. *Water Resour. Res.*, 23(5):951-960.

Bockstael, N. E., and Strand, I. E., Jr., 1987. The Effect of Common Sources of Regression Error on Benefit Estimates. *Land Economics*, 63(1):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(3):355-379.

Carson, R. T., Conaway, M. B., Hanemann, M. W., Krosnick, J. A., Mitchell, R. C., and 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., and Taneja, V. S., 1978. Numerical evaluation of the distribution of the product of two normal variables. *Journal of Statistical Computation and Simulation*, 7(2):123-131.

Craig, C. C. (1936). On the frequency function of xy. Annals of Mathematical Statistics, 7, 1-15.

Daly, A., Hess, S., and de Jong, G., 2012. Calculating errors for measures derived from choice modelling estimates. *Transportation Research Part B: Methodological*, 46(2):333-341.

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

Dorfman, J. H., Kling, C. L., and Sexton, R. J., 1990. Confidence Intervals for Elasticities and Flexibilities: Reevaluating the Ratios of Normals Case. *American Journal of Agricultural Economics*, 72(4):1006-1017.

Dufour, J.-M., 1997. Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models. *Econometrica*, 65(6):1365-1387.

Fiebig, D. G., 1985. Evaluating Estimators without Moments. *The Review of Economics and Statistics*, 67(3):529-534.

Finney, D. J., 1971. Probit Analysis. 3 Ed., 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.

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/.

Giergiczny, M., Valasiuk, S., Czajkowski, M., De Salvo, M., and Signorello, G., 2012. Including cost income ratio into utility function as a way of dealing with 'exploding' implicit prices in mixed logit models. *Journal of Forest Economics*, 18(4):370-380.

Glen, A. G., Leemis, L. M., and Drew, J. H., 2004. Computing the distribution of the product of two continuous random variables. *Computational Statistics & Data Analysis*, 44(3):451-464.

Gleser, L. J., and 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(4):1351-1362.

Graham-Tomasi, T., Adamowicz, W. L., and Fletcher, J. J., 1990. Errors of Truncation in Approximations to Expected Consumer Surplus. *Land Economics*, 66(1):50.

Greene, W. H., 2011. Econometric Analysis. 7 Ed., Prentice Hall, Upper Saddle River, NJ.

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

Hensher, D. A., Rose, J. M., and Greene, W. H., 2015. Applied Choice Analysis. 2 Ed., Cambridge University Press, Cambridge, UK.

Hess, S., and Train, K., 2011. Recovery of inter- and intra-personal heterogeneity using mixed logit models. *Transportation Research Part B: Methodological*, 45(7):973-990.

Hirschberg, J., and Lye, J., 2010. A Geometric Comparison of the Delta and Fieller Confidence Intervals. *The American Statistician*, 64(3):234-241.

Hole, A. R., 2007. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics*, 16(8):827-840.

Hwang, J. T., 1995. Fieller's Problem and Resampling Techniques. *Statistica Sinica*, 5:161-171.

Khuri, A., and Casella, G., 2002. The Existence of the First Negative Moment Revisited. *The American Statistician*, 56(1):44-47.

Kling, C. L., 1991. Estimating the precision of welfare measures. *Journal of Environmental Economics and Management*, 21(3):244-259.

Krinsky, I., and Robb, A. L., 1986. On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, 68(4):715-719.

Lai, C. D., Wood, G. R., and Qiao, C. G., 2004. The Mean of the Inverse of a Punctured Normal Distribution and Its Application. *Biometrical Journal*, 46(4):420-429.

Lehmann, E. L., and Shaffer, J. P., 1988. Inverted Distributions. *The American Statistician*, 42(3):191-194.

Lianos, T. P., and Rausser, G. C., 1972. Approximate Distribution of Parameters in a Distributed Lag Model. *Journal of the American Statistical Association*, 67(337):64-67.

Lye, J., and Hirschberg, J., 2018. Confidence Intervals for Ratios: Econometric Examples with Stata. Department of Economics - Working Papers Series 2037, The University of Melbourne.

Marsaglia, G., 2006. Ratios of Normal Variables. *Journal of Statistical Software*, 16(4):1:10.

Meijer, E., and Rouwendal, J., 2006. Measuring welfare effects in models with random coefficients. *Journal of Applied Econometrics*, 21(2):227-244.

Morey, E. R., Sharma, V. R., and Karlstrom, A., 2003. A Simple Method of Incorporating Income Effects into Logit and Nested-Logit Models: Theory and Application. *American Journal of Agricultural Economics*, 85(1):248-253.

Oehlert, G. W., 1992. A Note on the Delta Method. *The American Statistician*, 46(1):27-29.

Paige, R. L., Chapman, P. L., and Butler, R. W., 2011. Small Sample LD50 Confidence Intervals Using Saddlepoint Approximations. *Journal of the American Statistical Association*, 106(493):334-344.

Pham-Gia, T., Turkkan, N., and Marchand, E., 2006. Density of the Ratio of Two Normal Random Variables and Applications. *Communications in Statistics: Theory & Methods*, 35(9):1569-1591.

Phillips, P. C. B., 1994. Some Exact Distribution Theory for Maximum Likelihood Estimators of Cointegrating Coefficients in Error Correction Models. *Econometrica*, 62(1):73-93.

Piegorsch, W. W., and Casella, G., 1985. The Existence of the First Negative Moment. *The American Statistician*, 39(1):60-62.

Revelt, D., and Train, K., 1998. Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. *Review of Economics and Statistics*, 80(4):647-657.

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

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

Scarpa, R., Thiene, M., and 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(4):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):1-29.

Smith, V. K., 1990. Estimating Recreation Demand Using the Properties of the Implied Consumer Surplus. *Land Economics*, 66(2):111-120.

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

Staiger, D., Stock, J. H., and Watson, M. W., 1997. The NAIRU, Unemployment and Monetary Policy. *The Journal of Economic Perspectives*, 11(1):33-49.

Train, K., and Sonnier, G., 2005. Mixed Logit with Bounded Distributions of Correlated Partworths. In: *Applications of Simulation Methods in Environmental and Resource Economics*, R. Scarpa and A. Alberini, eds., Springer Netherlands, 117-134.

Train, K. E., and Weeks, M., 2005. Discrete Choice Models in Preference Space and Willingness-to-pay Space. In: *Applications of Simulation Methods in Environmental and Resource Economics*, R. Scarpa and A. Alberini, eds., Springer, Dordrecht, 1-16.

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

Varian, H. R., 1992. Microeconomic Analysis. 3 Ed., W. W. Norton, New York.

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

Ware, R., and 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(3):641-645.

Woglom, G., 2001. More Results on the Exact Small Sample Properties of the Instrumental Variable Estimator. *Econometrica*, 69(5):1381-1389.

Zaman, A., 1981. Estimators without moments: The case of the reciprocal of a normal mean. *Journal of Econometrics*, 15(2):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(2):127-158.

Zerbe, G. O., 1978. On Fieller's Theorem and the General Linear Model. *The American Statistician*, 32(3):103-105.