WORKING PAPER

ITLS-WP-13-01

Confidence intervals of willingness-to-pay for random coefficient logit models.

By
Michiel C.J. Bliemer and John M. Rose

January 2013

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in
Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-13-01

TITLE: **Confidence intervals of willingness-to-pay for random coefficient logit models.**

ABSTRACT: Random coefficient logit (RCL) models containing random parameters are increasingly used for modelling travel choices. Willingness-to-pay (WTP) measures, such as the value of travel time savings (VTTS) are, in the case of such RCL models, ratios of random parameters. In this paper we apply the Delta method to compute the confidence intervals of such WTP measures, taking into account the variance-covariance matrix of the estimates of the distributional parameters. Compared to simulation methods such as proposed by Krinsky and Robb, the Delta method is able to avoid some of the simulations by deriving partly analytical expressions for the standard errors. Examples of such computations are shown for different combinations of random distributions.

KEY WORDS: ***Willingness-to-pay, confidence intervals, random coefficient logit, Delta method.***

AUTHORS: **Bliemer and Rose**

Acknowledgements: The authors are grateful for the private discussions on this topic with Bill Greene and Stephane Hess. Also, we would like to thank David Hensher for his comments on an early draft of this paper. Of course, only the authors can be blamed for possible surviving errors.

CONTACT: INSTITUTE of TRANSPORT and LOGISTICS STUDIES (C13)
The Australian Key Centre in Transport and Logistics Management

The University of Sydney NSW 2006 Australia

Telephone: +612 9351 0071
Facsimile: +612 9351 0088
E-mail: business.itlsinfo@sydney.edu.au
Internet: http://sydney.edu.au/business/itls

DATE: January 2013

1. Introduction

Discrete choice models are now widely applied to predict market shares, compute elasticities, or to derive willingness-to-pay (WTP) measures. To this end, stated preference or revealed preference data is collected and assuming random utility theory, parameters of the utility functions are estimated. While the traditional conditional logit model (or often referred to as the multinomial logit (MNL) model) proposed by McFadden (1974) is still widely used, in the last decade there has been a clear shift towards the more general mixed multinomial logit (MMNL) model, also commonly referred to as the random coefficients logit (RCL) model, that can handle more complex error component structures, can describe heterogeneous behaviour by means of random parameters, and can take panel effects into account (see McFadden and Train, 2000). Having random instead of fixed coefficients, the WTP is no longer a fixed value but rather represented by a random distribution as well.

In order to determine the reliability of the parameter estimates and resulting confidence intervals, standard errors play an important role. These standard errors are obtained from the (asymptotic) variance-covariance matrix, which is related to the second derivatives (curvature) of the estimated models log-likelihood function, and commonly reported in most standard estimation software. As the parameter estimates maximize the log-likelihood function, the higher the curvature of this function at the top, the more reliable these parameter estimates, hence the lower the standard errors on average. Since each parameter is not known with certainty but rather has some confidence interval, the WTP – which is typically defined as the ratio of two parameters – also has an associated confidence interval.

Several methods exist for computing the standard error of a function of parameter estimates. For example, Krinsky and Robb (1986, 1990) proposed a procedure for using the variance-covariance matrix in simulating a confidence interval for elasticities which has since been adapted for calculating the confidence intervals for WTP. This procedure involves the use of Monte Carlo simulation in at least two dimensions (see Haab and McConnell, 2003, for more detail). An analytical method for determining the standard error for the WTP ratio is the Delta method. Using the first derivatives of the ratio function, the standard error can be found without relying on simulation methods. The Krinsky and Robb and the Delta method have been applied to ratios of parameter estimates of the MNL model. Without referring to the Krinsky and Robb method, Ettema et al. (1997) proposed an identical simulation method, which was applied in Espino et al. (2006). The other main simulation approach for obtaining confidence intervals that has been applied in the literature is the use of bootstrapping (see e.g., Armstrong et al., 2001). Armstrong et al. (2001) also provides an alternative but more complex simulation approach.

Recently, Daly et al. (2012a) argued that under certain assumptions, for such a ratio the Delta method provides an exact expression for the standard error of WTP estimates. Simulation approaches merely offer an approximation of the confidence intervals.

Whilst most research effort has focused on obtaining the standard errors for the ratio of parameters within the MNL model framework, determining the standard errors for ratios of random parameters in the RCL model has become more important given that this model is now becoming more mainstream. Determining these standard errors and the resulting confidence intervals for the ratio of two distributions however is not trivial.

To illustrate the complication of determining the standard errors for WTP measures within the RCL model framework, consider the case of the ratio of the travel time parameter and the cost parameter, yielding a value that is often in the literature referred to as the value of travel time savings (VTTS), an important WTP measure in the transportation field. Suppose that the travel time parameter is normally distributed, and the cost parameter lognormally distributed such that it is always negative. According to Daly et al. (2012b), such a WTP would have a finite mean and variance. In estimation, we have to find values of the distributional parameters, namely the mean and standard deviation of the normally distributed travel time parameter, and the mean and standard deviation of the lognormally distributed cost parameter. Each of these four parameters has associated standard errors describing its uncertainty. Hence, there is uncertainty about the mean of the normal distribution, uncertainty about the standard deviation of the normal distribution, as well as uncertainty about the mean and standard deviation of the lognormal distribution. Furthermore, there are covariances describing the correlations between the parameter estimates. When we compute the WTP ratio between the two parameters, these uncertainties translate into uncertainty of the standard error of the WTP. Clearly, the entire variance-covariance matrix plays a role in determining this uncertainty. Therefore, it may be tempting to simply simulate the WTP ratio by drawing different values for each distribution of the coefficients, compute the ratio, and compute the interval in which 95 percent of the resulting values fall, as done for example in Campbell (2007). However, such a procedure does not take the variance-covariance matrix with the uncertainties in the parameter estimates into account, and is therefore not a valid procedure.

The Krinsky and Robb procedure could be applied by taking simulated draws for each of the estimated four structural parameters, which then results in a normal distribution and a lognormal distribution. From these two distributions, we could again take simulated draws and compute the ratios. Therefore, the equivalent Krinsky and Robb procedure for the WTP in an RCL model would involve a Monte Carlo simulation in six dimensions. Such a procedure is proposed in Hensher and Greene (2003) and applied in Sillano and Ortúzar (2005) and Michaud et al. (in press). The procedure in Armstrong et al. (2001) has been applied to random coefficient models by Amador et al. (2005). As far as we are aware, the Delta method has never been applied to obtain confidence intervals for the WTP in RCL models.

In this paper we propose to apply the Delta method for determining the standard error of the WTP ratio of two randomly distributed parameters, which can be used to compute confidence intervals. The main reason for preferring the Delta method over the Krinsky and Robb procedure is that the Delta method requires less simulation. To compare, the Krinsky and Robb procedure would require simulation of six random variables, while the Delta method would require simulation over only two random variables as will become clear later in this paper. Although the theorem in Daly et al. (2012a) is very powerful, the claim that the Delta method provides exact standard errors for the WTP in the case of fixed coefficients likely does not generalise to the ratio of any two random coefficients, as in general some simulation is needed (as the case in this paper), which may violate the assumption of an invertible function in the theorem. In the special case of a ratio of two lognormal distributions, the resulting distribution is again lognormal, such that the analytical results in Daly et al. (2012a) can be used.

The remainder of the paper is structured as follows. In Section 2 a brief introduction into discrete choice models and WTP is given, which mainly serves to introduce the necessary mathematical notation. Section 3 reviews how to compute the confidence intervals for WTPs in the MNL model using the Delta method. Section 4 presents the main contribution of the paper, namely applying the Delta method in case of the RCL model, both for independently and dependently randomly distributed parameters. To illustrate the method, four examples are given in Section 5. Section 6 concludes with a discussion.

2. Parameter estimation and WTP

Consider the usual utility function formulation in which the utility of alternative *j*, U_j, consists of a systematic utility part, V_j, and an unobserved part, ε_j,

$$U_j = V_j + \varepsilon_j, \tag{1}$$

where the systematic part is given by a (linear or nonlinear) function g_j of some known attribute levels for that alternative, x_j, and a vector of *K* unknown parameters, β,

$$V_j = g_j(x_j \mid \beta). \tag{2}$$

Often, a linear function is assumed, such that

$$V_j = \sum_k \beta_k x_{jk}. \tag{3}$$

We assume that the unobserved components ε_j are independently and identically extreme value type 1 (EV1) distributed, such that the probability of choosing a certain alternative is expressed by a logit type model.

The unknown parameters, also called coefficients (and used interchangeably in this paper), describing the preferences of agents under study, are to be estimated by observing (stated or revealed) choices of the agents in some choice situations. If the agents are assumed to be homogeneous, these parameters are constant over all agents, such that fixed parameters are estimated. In contrast, whenever agents are assumed to be heterogeneous (i.e., different preferences), then typically random parameters with distributions are estimated for the whole population of agents. For example, one could estimate a fixed parameter β_k which has a single value, or instead estimate a distribution in which the structural parameters are to be estimated (e.g., mean μ_k and standard deviation σ_k in case β_k is assumed to follow a normal distribution). In case of homogeneous agents, the MNL model is considered, while with agents with heterogeneous preferences, the RCL model results (either cross-sectional or panel).

Let $\hat{\beta}$ denote the vector of (maximum likelihood) estimates for the unknown parameters. According to McFadden (1974), these parameters will be asymptotically normally distributed with a mean corresponding to the true parameter values, β, and a variance-covariance matrix, Ω_β, equal to the negative inverse of the Fisher information matrix,

$$\hat{\beta} \xrightarrow{D} N(\beta, \Omega_\beta). \tag{4}$$

The asymptotic variance-covariance matrix, Ω_β, together with the parameter estimates, $\hat{\beta}$, is a typical side-product of estimation software such as Alogit, Biogeme, or Nlogit. The roots of the diagonal elements of this matrix denote the (asymptotic) standard errors. These standard errors denote how reliable the parameter estimates are, yielding *t*-ratios to test the null hypothesis of the parameter estimates.

Instead of the values of β themselves, one is often more interested in ratios of these parameters. WTP is a special case in which the denominator is the cost parameter. In case of a utility function that is linear in the parameters and linear in the attributes (as in Eqn. (3)), we define the WTP of attribute *k* as

$$w_k = \frac{\beta_k}{\beta_c}, \tag{5}$$

where β_k is the parameter for attribute *k* and β_c is the cost parameter. In the more general case of a nonlinear utility function, the WTP of attribute *k* is defined as

$$w_k = \frac{\partial g_j / \partial x_{jk}}{\partial g_j / \partial x_{jc}}. \tag{6}$$

The theory in this paper will be valid for the general nonlinear case, however, we will focus on the most widely assumed case of linear utility functions.

Since both β_k and β_c are both known but with uncertainty, there is also exists uncertainty about w_k. An interesting question then is, what is the standard error of w_k or alternatively, what is the confidence interval of w_k ? The next section first discusses how this has been solved in the literature for the case of fixed parameters in the MNL model. Then we show how to determine these confidence intervals in the case of random parameters in RCL models, which is the main contribution of this paper.

3. Confidence intervals for WTP in the MNL model

The Delta method can be applied to determine the variance of a ratio of parameter estimates. In fact, the Delta method can be applied in general for any function of the parameters. This method states that, if $\hat{\beta}$ is asymptotically distributed as mentioned in Eqn. (4), then a function $h(\hat{\beta})$ is asymptotically normally distributed with a mean of $h(\beta)$ and a variance of $\nabla_\beta h(\beta)^T \Omega_\beta \nabla_\beta h(\beta)$,

$$h(\hat{\beta}) \xrightarrow{D} N\left(\beta, \nabla_\beta h(\beta)^T \Omega_\beta \nabla_\beta h(\beta)\right), \tag{7}$$

where $\nabla_\beta h(\beta)$ denotes the Jacobian of $h(\beta)$. In case of $w_k = h(\beta_k, \beta_c) = \beta_k / \beta_c$, this yields

$$\hat{w}_k \xrightarrow{D} N\left(\beta, \begin{pmatrix} \dfrac{1}{\beta_c} \\ -\dfrac{\beta_k}{\beta_c^2} \end{pmatrix}^T \begin{pmatrix} \operatorname{var}(\beta_k) & \operatorname{cov}(\beta_k, \beta_c) \\ \operatorname{cov}(\beta_k, \beta_c) & \operatorname{var}(\beta_c) \end{pmatrix} \begin{pmatrix} \dfrac{1}{\beta_c} \\ -\dfrac{\beta_k}{\beta_c^2} \end{pmatrix} \right), \tag{8}$$

which simplifies to

$$\hat{w}_k \xrightarrow{D} N\left(\beta, \frac{1}{\beta_c^2}\left(\operatorname{var}(\beta_k) - 2w_k \operatorname{cov}(\beta_k, \beta_c) + w_k^2 \operatorname{var}(\beta_c)\right)\right). \tag{9}$$

Therefore, the asymptotic standard error of the WTP is

$$se(\hat{w}_k) = \frac{1}{\beta_c}\sqrt{\operatorname{var}(\beta_k) - 2w_k \operatorname{cov}(\beta_k, \beta_c) + w_k^2 \operatorname{var}(\beta_c)}, \tag{10}$$

which is the same formula as derived in for example Scarpa and Rose (2008) and Daly et al. (2012a). Using the parameter estimates $\hat{\beta}_k$ and $\hat{\beta}_c$ as the true parameters and using the corresponding elements in the asymptotic variance-covariance matrix (both provided by the estimation software), this standard error can be analytically computed.

This derivation holds in the case of fixed coefficients. With random coefficients estimated in a RCL model the same Delta method can be applied, realizing that β represents a probability distribution in which the distributional parameters are estimated with some uncertainty.

4. Confidence intervals for WTP in the RCL model

In case of random parameters, parameters β_k follow certain distributions. These distributions are described by parameters themselves, which have to be estimated and therefore have some degree of uncertainty. Hence, there is uncertainty about the exact shape of the distribution. Let us first of all assume that the distributions of these parameters are independent. Later we will consider the case in which β describes a vector of dependent normally distributed parameters.

4.1. Case I: Independently distributed random parameters

Let each β_k follow a probability distribution with (a vector of) parameters θ_k. Let us also assume that the cost parameter β_c can follow such a distribution. Estimating these distributional parameters in the RCL model will yield parameter estimates $\hat{\theta}_k$. The trick is to map the standard errors (and covariances) of the structural parameter estimates $\hat{\theta}_k$ to a standard error of β_k as well as to determine the standard error of β_k / β_c. The answer lies in rewriting the parameters β_k and β_c into functions of θ_k and θ_c using parameter-free distributions (such as the standard normal or the standard uniform distribution). These functions can then be used within the Delta method.

Let us first write the parameters in terms of the distributional parameters and a parameter-free distribution:

$$\beta_k = \beta_k(z_k \mid \theta_k), \tag{11}$$

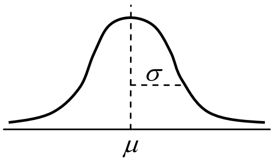
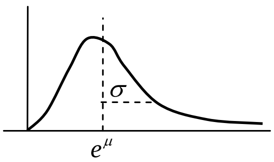
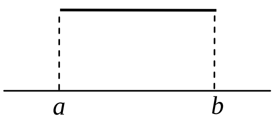
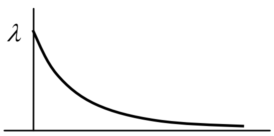
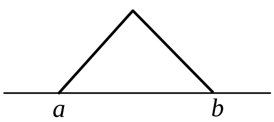
where z_k is a standard probability distribution (or in some cases, a vector of standard probability distributions). For example, in order to describe a normal distribution, $N(\mu_k, \sigma_k^2)$, we can write $\beta_k = \mu_k + \sigma_k z_k$, where z_k follows a standard normal distribution, $N(0,1)$. Table 1 provides a (non-exhaustive) list of other probability distributions which can be derived from a standard normal or uniform distribution. The table also contains the first derivatives (Jacobians) of β to the distributional parameter(s) and standard distributed random variable(s) *z*, which we will need later in applying the Delta method.

The WTP can be written as

$$w_k(z_k, z_c \mid \theta_k, \theta_c) = \frac{\beta_k(z_k \mid \theta_k)}{\beta_c(z_c \mid \theta_c)}. \tag{12}$$

Note that since the parameters are distributions, the WTP will also be a distribution.

Table 1: Different parameter distributions

Distribution	$\beta(z\mid\theta)$	Standard distributio n	$\nabla_\theta \beta$	$\nabla_z \beta$
Normal 	$\beta(z\mid\mu,\sigma)=\mu+\sigma z$	$z \propto N(0,1)$	$\begin{pmatrix}1\\ z\end{pmatrix}$	σ
Lognormal 	$\beta(z\mid\mu,\sigma)=\exp(\mu+\sigma z)$	$z \propto N(0,1)$	$\begin{pmatrix}\beta\\ z\beta\end{pmatrix}$	$\sigma\beta$
Uniform 	$\beta(z\mid a,b)=a+(b-a)z$	$z \propto U(0,1)$	$\begin{pmatrix}1-z\\ z\end{pmatrix}$	$b-a$
Exponential 	$\beta(z\mid\lambda)=-\dfrac{\ln z}{\lambda}$	$z \propto U(0,1)$	$-\dfrac{\beta}{\lambda}$	$-\dfrac{1}{\lambda z}$
Triangular 	$\beta(z\mid a,b)=a+(b-a)\dfrac{z_1+z_2}{2}$	$z_1 \propto U(0,1)$ $z_2 \propto U(0,1)$	$\begin{pmatrix}1-\dfrac{z_1+z_2}{2}\\ \dfrac{z_1+z_2}{2}\end{pmatrix}$	$\begin{pmatrix}\dfrac{b-a}{2}\\ \dfrac{b-a}{2}\end{pmatrix}$

First, let us focus on the WTP for a specific value (draw) of (z_k, z_c). Suppose that θ_k and θ_c have p_k and p_c elements, and that z_k and z_c have s_k and s_c elements, respectively. Applying the Delta method, we arrive at

$$\hat{w}_k(z_k,z_c)\xrightarrow{D} N\left(w_k, \begin{pmatrix}\nabla_{\theta_k} w_k\\ \nabla_{\theta_c} w_k\\ \nabla_{z_k} w_k\\ \nabla_{z_c} w_k\end{pmatrix}^T \begin{pmatrix}\Omega_{\theta_{kc}} & \mathbf{0}\\ \mathbf{0} & \mathrm{diag}(1,\ldots,1)\end{pmatrix} \begin{pmatrix}\nabla_{\theta_k} w_k\\ \nabla_{\theta_c} w_k\\ \nabla_{z_k} w_k\\ \nabla_{z_c} w_k\end{pmatrix}\right), \tag{13}$$

in which $\nabla_{\theta_k} w_k \in R^{p_k}$ and $\nabla_{\theta_c} w_k \in R^{p_c}$ are the Jacobians of the WTP to θ_k and θ_c, respectively, evaluated in the true values of the parameters, $\nabla_{z_k} w_k \in R^{s_k}$ and $\nabla_{z_c} w_k \in R^{s_c}$ are the Jacobian of the WTP to z_k and z_c, respectively, $\Omega_{\theta_{kc}}$ is the submatrix of the variances and covariances of distributional parameters θ_k and θ_c, $\mathbf{0} \in R^{(p_k+p_c)\times(s_k+s_c)}$ is a matrix with zeros, and $\mathrm{diag}(1,\ldots,1) \in R^{(s_k+s_c)\times(s_k+s_c)}$ is a diagonal matrix with ones. The zeros and diagonal matrix follow from the fact that all standard distributions are independently distributed without any correlations with the other parameters.

The Jacobians can be calculated as

$$\begin{aligned}
\nabla_{\theta_k} w_k &= \nabla_{\theta_k}\left(\frac{\beta_k(z_k \mid \theta_k)}{\beta_c(z_c \mid \theta_c)}\right) = \frac{1}{\beta_c}\nabla_{\theta_k}\beta_k, \\
\nabla_{\theta_c} w_k &= \nabla_{\theta_c}\left(\frac{\beta_k(z_k \mid \theta_k)}{\beta_c(z_c \mid \theta_c)}\right) = -\frac{\beta_k}{\beta_c^2}\nabla_{\theta_c}\beta_c = -\frac{w_k}{\beta_c}\nabla_{\theta_c}\beta_c, \\
\nabla_{z_k} w_k &= \nabla_{\theta_c}\left(\frac{\beta_k(z_k \mid \theta_k)}{\beta_c(z_c \mid \theta_c)}\right) = \frac{1}{\beta_c}\nabla_{z_k}\beta_k, \\
\nabla_{z_c} w_k &= \nabla_{\theta_c}\left(\frac{\beta_k(z_k \mid \theta_k)}{\beta_c(z_c \mid \theta_c)}\right) = -\frac{\beta_k}{\beta_c^2}\nabla_{z_c}\beta_c = -\frac{w_k}{\beta_c}\nabla_{z_c}\beta_c.
\end{aligned} \tag{14}$$

In other words, we can rewrite formula (13) as

$$\hat{w}_k(z_k, z_c) \xrightarrow{D} N\left(w_k, \frac{1}{\beta_c^2}\begin{pmatrix} \nabla_{\theta_k}\beta_k \\ -w_k\nabla_{\theta_c}\beta_c \\ \nabla_{z_k}\beta_k \\ -w_k\nabla_{z_c}\beta_c \end{pmatrix}^T \begin{pmatrix} \Omega_{\theta_{kc}} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(1,\ldots,1) \end{pmatrix} \begin{pmatrix} \nabla_{\theta_k}\beta_k \\ -w_k\nabla_{\theta_c}\beta_c \\ \nabla_{z_k}\beta_k \\ -w_k\nabla_{z_c}\beta_c \end{pmatrix} \right). \tag{15}$$

In the special case of having both fixed (non-random) coefficients, $\nabla_{\theta_k}\beta_k = \nabla_{\theta_c}\beta_c = 1$ and $\nabla_{z_k}\beta_k = \nabla_{z_c}\beta_c = 0$, such that the variance simplifies to the Eqn. (8). The asymptotic distribution in (15) is for the conditional parameter estimate $\hat{w}_k(z_k, z_c)$. The (unconditional) expected WTP estimate, denoted by $\hat{w}_k$ (without being conditional to specific draws (z_k, z_c)), is defined as

$$\hat{w}_k = \int_{z_k}\int_{z_c} \hat{w}_k(z_k, z_c)\, dF_k(z_k)\, dF_c(z_c), \tag{16}$$

where $F_k(z_k)$ and $F_c(z_c)$ are the (possibly multivariate) cumulative distribution functions of the standard distributed z_k and z_k. Since the integrals are linear operators, the resulting asymptotic distribution of $\hat{w}_k$ is also normally distributed, where the expectation and the variance are integrals over $\hat{w}_k(z_k, z_c)$ defined in Eqn. (15). Theoretically, this leads to a problem, as for unbounded distributions (like the normal distribution that is defined on the complete domain of $(-\infty,+\infty)$) these integrals will not be defined at $\beta_c = 0.$ That the moments of the distribution are undefined, does not mean that the distribution does not exist. Daly et al. (2012b) show that the probability of observing $\beta_c = 0$ should be zero in order for the moments to be finite. Hence, they suggest that the cost parameter should not follow a normal distribution or a distribution truncated at zero, but rather a lognormal distribution or another distribution with no probability mass at $\beta_c = 0.$ Alternatively, one could use the median to replace the mean. In the remainder of this paper we will assume that the probability distribution of the cost parameter has no mass at $\beta_c = 0$ or that the median replaces the mean in case there is a positive mass around zero.

The unconditional mean WTP estimate $\hat{w}_k$ can be approximated by Monte Carlo simulation,

$$\hat{w}_k \approx \frac{1}{R}\sum_{r=1}^{R} \hat{w}_k(z_k^{(r)}, z_c^{(r)}), \tag{17}$$

where $(z_k^{(r)}, z_c^{(r)}),\ r = 1,\ldots,R,$ are pseudo-random or quasi-random draws from the distributions defined by $F_k(z_k)$ and $F_c(z_c).$ The larger R is, the more accurate the approximation will be. Since $\hat{w}_k(z_k^{(r)}, z_c^{(r)})$ is asymptotically normally distributed, $\hat{w}_k$ will also be normally distributed in the limit, with the following simulated variance:

$$\mathrm{var}(\hat{w}_k) \approx \frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{\left(\beta_c^{(r)}\right)^2} \begin{pmatrix} \nabla_{\theta_k}\beta_k^{(r)} \\ -w_k^{(r)}\nabla_{\theta_c}\beta_c^{(r)} \\ \nabla_{z_k}\beta_k^{(r)} \\ -w_k^{(r)}\nabla_{z_c}\beta_c^{(r)} \end{pmatrix}^T \begin{pmatrix} \Omega_{\theta_{kc}} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(1,\ldots,1) \end{pmatrix} \begin{pmatrix} \nabla_{\theta_k}\beta_k^{(r)} \\ -w_k^{(r)}\nabla_{\theta_c}\beta_c^{(r)} \\ \nabla_{z_k}\beta_k^{(r)} \\ -w_k^{(r)}\nabla_{z_c}\beta_c^{(r)} \end{pmatrix} \right), \tag{18}$$

where $\beta_k^{(r)} = \beta_k(z_k^{(r)}, z_c^{(r)}),\ \beta_c^{(r)} = \beta_c(z_k^{(r)}, z_c^{(r)}),$ and $w_k^{(r)} = \beta_k^{(r)} / \beta_c^{(r)}.$ The draws $(z_k^{(r)}, z_c^{(r)})$ can be obtained using for example Halton draws (e.g., Bhat, 2001) or other quasi-random draws (e.g., Sándor and Train, 2004; Bliemer et al., 2008) $(\omega_k^{(r)}, \omega_c^{(r)})$ such that the $z_k^{(r)} = F_k^{-1}(\omega_k^{(r)})$ and $z_c^{(r)} = F_c^{-1}(\omega_c^{(r)}).$

Once the asymptotic variance in Equation (18) has been calculated, the $(1-\alpha)$ confidence interval of the expected WTP estimate can be determined as

$$\left(\hat{w}_k - t_{1-\alpha/2}\sqrt{\mathrm{var}(\hat{w}_k)},\ \hat{w}_k + t_{1-\alpha/2}\sqrt{\mathrm{var}(\hat{w}_k)}\right), \tag{19}$$

where t_α is the *t*-value corresponding to a level of significance of $\alpha.$ For example, for a 95 percent confidence interval, $t_{0.975} = 1.96.$

It is important to realize that the variance of the unconditional WTP computed directly from the conditional WTP's, considering only simulated values of $w_k(z_k^{(r)}, z_c^{(r)}),$ is incorrect as it ignores the uncertainty (expressed in the variance-covariance matrix) of the distributional parameter estimates, $\hat{\theta},$ while we explicitly take this into account in Eqn. (18).

4.2. Case II: Dependently distributed random parameters

Estimation of dependent random parameters is typically limited to the multivariate normal distribution, as for other distributions or mixtures of distributions the multivariate distribution are not easy to estimate or can only be approximated. In this section, we restrict ourselves to the ratio of two dependent normally or lognormally distributed parameters.

Consider a vector of parameters, β, which are assumed normally distributed with a vector of means, μ, and a matrix of (co)variances, Σ,

$$\beta \propto N(\mu,\Sigma). \tag{20}$$

Non-zero covariances mean that the parameters are correlated (dependent). In order to estimate these variances and covariances, a Cholesky decomposition can be used in which the vector of dependent normally distributed parameters, β, is written a linear combination of a vector of independent standard normally distributed parameters, z,

$$\beta = \mu + Az, \tag{21}$$

where A is a lower triangular (Cholesky) matrix such that $AA^T = \Sigma$ (see e.g., Greene, 2008). The values in the A matrix are then estimated, and using these values the matrix with estimated (co)variances can be obtained. Writing Eqn. (21) in extensive form, this becomes

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_K \end{pmatrix} + \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \cdots & a_{KK} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_K \end{pmatrix} = \begin{pmatrix} \mu_1 + a_{11}z_1 \\ \mu_2 + a_{21}z_1 + a_{22}z_2 \\ \vdots \\ \mu_K + a_{K1}z_1 + \ldots + a_{KK}z_K \end{pmatrix}. \tag{22}$$

This means that each single parameter can be written as

$$\beta_k = \mu_k + \sum_{i=1}^{k} a_{ki} z_i. \tag{23}$$

An important difference with the case of independent normally distributed parameters is, that β_k no longer just depends on only z_k, but on $z_1,\ldots,z_k$. The K^{th} random parameter, β_K, depends on $K+1$ distributional parameters, $\theta_K = (\mu_K, a_{K1},\ldots,a_{KK})$. Hence, the vector of all parameters that need to be estimated can be denoted by $\theta = (\theta_k)_{k=1,\ldots,K}$. Estimation of this vector produces not only the parameter estimates, $\hat{\theta}$, but also yields an asymptotic variance-covariance matrix, Ω_θ. As indicated before, the square roots of the diagonal elements of this matrix denote the standard errors describing the uncertainty of each element in the vector of parameter estimates.

Computing the variance of the unconditional WTP requires again simulation. It is clear that the ranking order in which the parameters are represented in the Cholesky matrix determines the number of distributions that needs to be drawn from when computing the WTP. If $w_k = \beta_k / \beta_c$ is of main interest, then it is best to use β_k and β_c as the first two parameters, requiring only two standard normal distributions to be drawn from. If such an ordering is not made in advance, then in theory one may need to draw from all K standard normal distributions. Larger numbers of draws, R, are then required to obtain a good approximation of the expected WTP and its asymptotic variance.

In order to determine the asymptotic distribution of the estimated conditional WTP, $\hat{w}_k(z_k, z_c),$ again Eqn. (13) can be used. Assuming that β_k and β_c as the first two parameters, the Jacobians for the WTP are given by

$$\nabla_{\theta_k} w_k = \frac{1}{\beta_c}\begin{pmatrix} 1 \\ z_k \end{pmatrix}, \nabla_{\theta_c} w_c = -\frac{w_k}{\beta_c}\begin{pmatrix} 1 \\ z_k \\ z_c \end{pmatrix}, \nabla_{z_k} w_k = \frac{1}{\beta_c}(a_{11} - a_{21}w_k), \text{ and} \qquad (24)$$
$$\nabla_{z_c} w_c = -\frac{a_{22}w_k}{\beta_c}.$$

Now assume that all coefficients are lognormally distributed in which the underlying normal distribution has a vector of means, μ, and a matrix of (co)variances, Σ. Eqn. (23) then becomes

$$\beta_k = \exp\left(\mu_k + \sum_{i=1}^{k} a_{ki} z_i \right), \qquad (25)$$

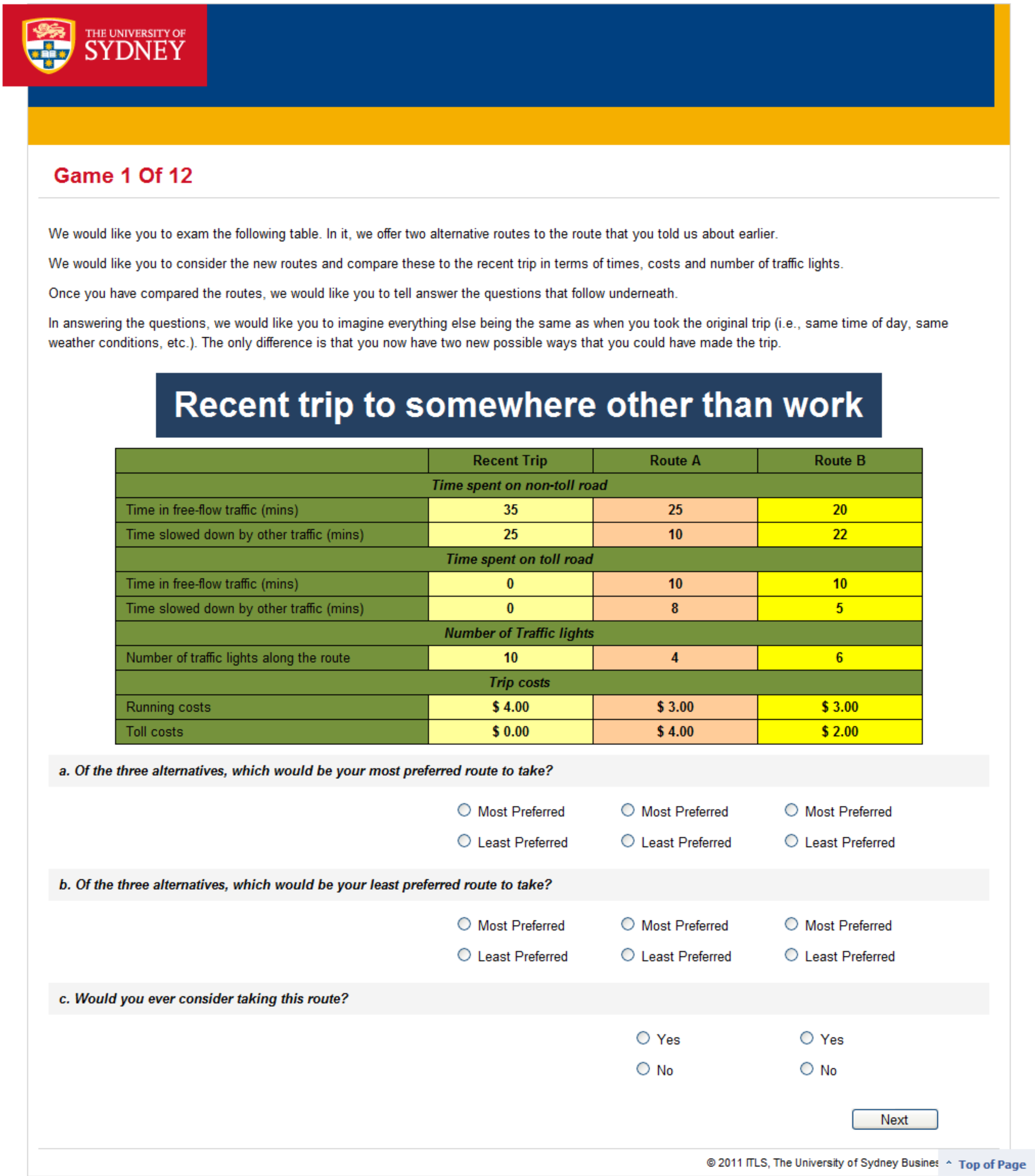
such that the Jacobians are

$$\nabla_{\theta_k} w_k = w_k \begin{pmatrix} 1 \\ z_k \end{pmatrix}, \nabla_{\theta_c} w_c = -w_k \begin{pmatrix} 1 \\ z_k \\ z_c \end{pmatrix}, \nabla_{z_k} w_k = (a_{11} - a_{21})w_k, \text{ and } \nabla_{z_c} w_c = -a_{22}w_k. \qquad (26)$$

It is possible to mix normal and lognormal distributions, hence β_k can be normally distributed and β_c lognormally with a joint matrix of (co)variances of the underlying normal distribution, Σ.

5. Examples

In this section we will provide a few numerical examples, illustrating the computation of the confidence intervals of the willingness-to-pay under different distributional assumptions of the parameter estimates. We use an empirical data set collected in a simple route choice experiment, where respondents had to choose between their current route, and two hypothetical route alternatives that included a tolled route. The routes were identified by four travel times described as the time spent in free flow and congested travel conditions travelling on non-tolled road during the trip, and free flow and congested travel conditions travelling on a toll road during the trip, as well as the toll and petrol costs, and the number of traffic lights (see Figure 1). For the current study, we combine the free flow and time spent in congested traffic conditions for both road types to form combined travel times non-tolled, and tolled roads, and use only the toll cost and the number of traffic lights.


THE UNIVERSITY OF SYDNEY

Game 1 Of 12

We would like you to exam the following table. In it, we offer two alternative routes to the route that you told us about earlier.

We would like you to consider the new routes and compare these to the recent trip in terms of times, costs and number of traffic lights.

Once you have compared the routes, we would like you to tell answer the questions that follow underneath.

In answering the questions, we would like you to imagine everything else being the same as when you took the original trip (i.e., same time of day, same weather conditions, etc.). The only difference is that you now have two new possible ways that you could have made the trip.

Recent trip to somewhere other than work

	Recent Trip	Route A	Route B
Time spent on non-toll road			
Time in free-flow traffic (mins)	35	25	20
Time slowed down by other traffic (mins)	25	10	22
Time spent on toll road			
Time in free-flow traffic (mins)	0	10	10
Time slowed down by other traffic (mins)	0	8	5
Number of Traffic lights			
Number of traffic lights along the route	10	4	6
Trip costs			
Running costs	$ 4.00	$ 3.00	$ 3.00
Toll costs	$ 0.00	$ 4.00	$ 2.00

a. Of the three alternatives, which would be your most preferred route to take?

Most Preferred / Least Preferred | Most Preferred / Least Preferred | Most Preferred / Least Preferred

b. Of the three alternatives, which would be your least preferred route to take?

Most Preferred / Least Preferred | Most Preferred / Least Preferred | Most Preferred / Least Preferred

c. Would you ever consider taking this route?

Yes / No | Yes / No

Next

© 2011 ITLS, The University of Sydney Busines Top of Page

Figure 1: Screen capture of stated choice task

Data were collected in October 2011 from 148 respondents, each of whom completed 12 choice tasks each. As such, the data consists of 1776 choice observations. Respondents were recruited from an internet panel (http://www.pureprofile.com/) and were required to have recently taken a commuting trip in Sydney Australia.

For illustration purposes, we will focus on the willingness-to-pay for a reduction in the travel time of the non-tolled route (TUR), i.e., $w_{TUR} = \beta_{TUR} / \beta_{TC}$. In four different examples, we will assume the following combinations of distributions for the two parameters: (i) normally distributed travel time parameter, fixed cost parameter, (ii) both parameters independently normally distributed, (iii) both parameters dependently normally distributed, and (iv) fixed travel time parameter, lognormally distributed cost parameter. Table 2 summarizes the different parameter estimates that are used to illustrate these willingness-to-pay computations for RCL models. The parameter estimates used in the four examples are shaded in grey. Note that the standard deviations of the random parameters are fairly large, such that we would expect relatively wide confidence intervals for the WTP estimates.

Table 2: Parameter estimates of four examples

		Example 1		Example 2		Example 3		Example 4	
Attribute		β	(*t-ratio*)	β	(*t-ratio*)	β	(*t-ratio*)	β	(*t-ratio*)
Con1		-0.86424	(-9.87)	-2.07409	(-13.33)	-2.16449	(-12.83)	-0.73291	(-11.30)
Con2		0.30547	(4.48)	0.25094	(3.28)	0.23754	(3.12)	0.23771	(4.07)
Travel time UR	Mu	-0.04694	(-4.76)	-0.02895	(-3.27)	-0.02586	(-2.54)	-0.03470	(-10.77)
	Sigma	0.06611	(5.52)	0.05111	(4.33)	--	(4.37)	--	--
Travel time TR	Mu	-0.25723	(-7.47)	-0.01952	(-1.58)	-0.05938	(-2.79)	-0.01284	(-2.49)
	Sigma	0.86313	(48.02)	0.07005	(2.62)	--	(2.38)	--	--
Traffic Lights	Mu	-0.20925	(-7.87)	-0.07167	(-2.38)	0.02272	(0.49)	-0.12124	(-5.18)
	Sigma	0.05230	(0.85)	0.14364	(2.89)	--	(3.44)	--	--
Toll Costs	Mu	-0.50606	(-24.38)	-0.95054	(-9.51)	-0.81403	(-9.94)	-0.99440	(-7.06)
	Sigma	--	--	0.91316	(10.46)	--	(7.76)	1.22291	(26.35)
Cholesky Sigma	TUR:TUR	--	--	--	--	0.05780	(4.37)	--	--
	TTR:TTR	--	--	--	--	0.06466	(1.71)	--	--
	TL:TL	--	--	--	--	0.09822	(1.52)	--	--
	TC:TC	--	--	--	--	0.74988	(7.66)	--	--
	TTR:TUR	--	--	--	--	0.01751	(0.36)	--	--
	TL:TUR	--	--	--	--	-0.11592	(-1.92)	--	--
	TL:TTR	--	--	--	--	-0.01686	(-0.23)	--	--
	TC:TUR	--	--	--	--	0.01824	(0.14)	--	--
	TC:TTR	--	--	--	--	0.06003	(2.26)	--	--
	TC:TL	--	--	--	--	-0.15613	(-2.59)	--	--
Log-likelihood		-1377		-1295		-1285		-1319	
Adj. ρ^2		0.29		0.33		0.34		0.32	

5.1. Normal divided by fixed

Assume that β_k is a random parameter following a normal distribution, hence $\beta_k = \mu_k + \sigma_k z_k$ with $z_k \propto N(0,1)$, and that β_c is a fixed parameter. The Jacobians are $\nabla_{(\mu_k,\sigma_k)}\beta_k = (1, z_k)^T$, $\nabla_{\beta_c}\beta_c = 1$, and $\nabla_{z_k}\beta_k = \sigma_k$, respectively, such that the conditional WTP is asymptotically distributed as

$$\hat{w}_k(z_k) \xrightarrow{D} N\left(w_k(z_k), \frac{1}{\beta_c^2} \begin{pmatrix} 1 \\ z_k \\ -w_k(z_k) \\ \sigma_k \end{pmatrix}^T \begin{pmatrix} \Omega_{(\mu_k,\sigma_k,\beta_c)} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ z_k \\ -w_k(z_k) \\ \sigma_k \end{pmatrix} \right), \tag{27}$$

with the subset of the covariance matrix taken directly from the estimation software,

$$\Omega_{(\mu_k,\sigma_k,\beta_c)} = \begin{pmatrix} \mathrm{var}(\mu_k) & \mathrm{cov}(\mu_k,\sigma_k) & \mathrm{cov}(\mu_k,\beta_c) \\ \mathrm{cov}(\mu_k,\sigma_k) & \mathrm{var}(\sigma_k) & \mathrm{cov}(\sigma_k,\beta_c) \\ \mathrm{cov}(\mu_k,\beta_c) & \mathrm{cov}(\sigma_k,\beta_c) & \mathrm{var}(\beta_c) \end{pmatrix} = \begin{pmatrix} 0.00010 & 0.00000 & 0.00005 \\ 0.00000 & 0.00014 & -0.0001 \\ 0.00005 & -0.00011 & 0.00043 \end{pmatrix}$$

Using 25,000 Halton draws for simulating the standard normally distributed variable z_k, the average WTP can be computed as 0.0928, and the average variance is 0.0179, such that the average standard error is 0.1138. Hence, the 95 percent confidence interval is (-0.1694, 0.3550).

To graphically illustrate, with each Halton draw we obtain a WTP value and a variance of the WTP. Hence, each draw represents a normal distribution of the WTP. In Figure 2 we have plotted (in blue) 50 normal distributions obtained from the first 50 Halton draws. The sampling distribution is then determined by taking the mean WTP and the mean variance, represented by the thick solid line (in red) in Figure 1.

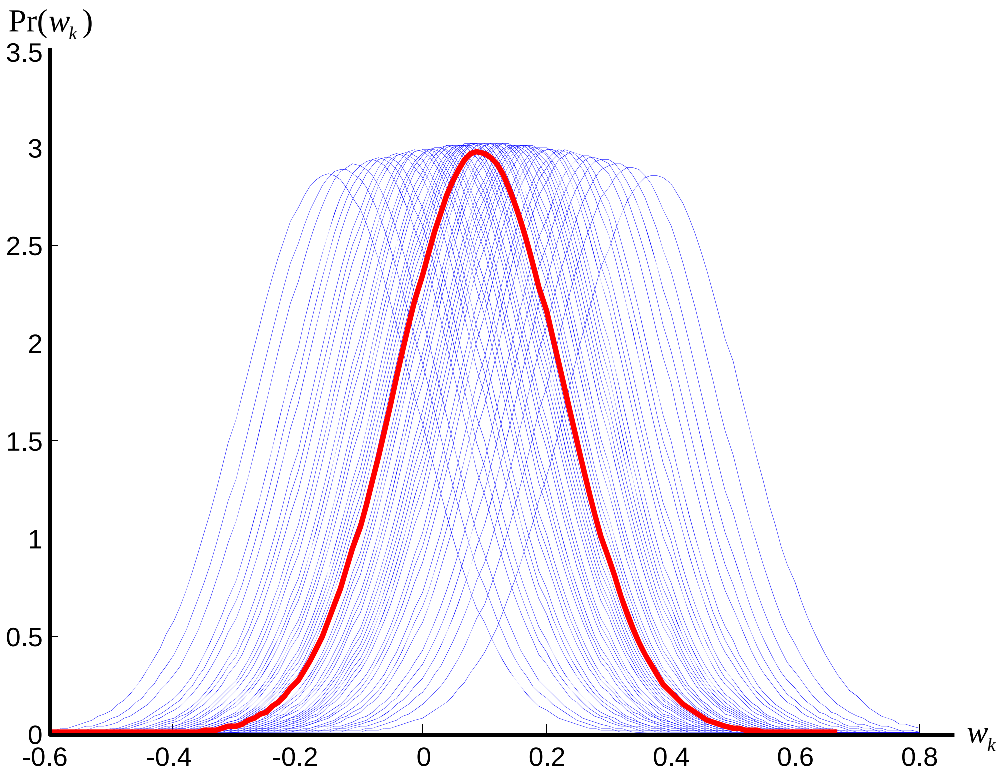


Figure 2: Simulated normal distributions and the sampling distribution (Normal / Fixed)

We also compare the results using the (Krinsky and Robb) simulation procedure for random coefficients logit models, as outlined by Hensher and Greene (2003). Obtaining the lower triangular Cholesky matrix from the covariance matrix, which is then used to simulate μ_k, σ_k, and β_c using 25,000 Halton draws, and obtaining β_k by simulating z_k (hence, a simulation over four dimensions in total), we find a mean WTP of 0.0943 and a variance of 0.0175. Taking the 0.025 and 0.975 percentiles results in a confidence interval of (-0.1674, 0.3587). Hence, the Delta method reproduces the confidence intervals found by applying the Krinsky and Robb method, but instead requiring integration over only a single random variables instead of over four dimensions.

5.2. Normal divided by normal

Now assume that both β_k and β_c are random parameters following a normal distribution, hence $\beta_k = \mu_k + \sigma_k z_k$ with $z_k \propto N(0,1)$, and $\beta_c = \mu_c + \sigma_c z_c$, with $z_c \propto N(0,1)$. Then the Jacobians are $\nabla_{(\mu_k,\sigma_k)}\beta_k = (1, z_k)^T$, $\nabla_{(\mu_c,\sigma_c)}\beta_c = (1, z_c)^T$, $\nabla_{z_k}\beta_k = \sigma_k$, and $\nabla_{z_c}\beta_c = \sigma_c$, respectively, such that the conditional WTP is asymptotically distributed as

$$\hat{w}_k(z_k, z_c) \xrightarrow{D} N\left(w_k(z_k, z_c), \frac{1}{\beta_c^2(z_c)} \begin{pmatrix} 1 \\ z_k \\ -w_k(z_k, z_c) \\ -w_k(z_k, z_c)z_c \\ \sigma_k \\ -w_k(z_k, z_c)\sigma_c \end{pmatrix}^T \begin{pmatrix} \Omega_{(\mu_k,\sigma_k,\mu_c,\sigma_c)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & 0 \\ \mathbf{0} & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ z_k \\ -w_k(z_k, \\ -w_k(z_k, z \\ \sigma_k \\ -w_k(z_k, z_(\end{pmatrix} \right. \quad (28)$$

with

$$\Omega_{(\mu_k,\sigma_k,\mu_c,\sigma_c)} = \begin{pmatrix} 0.00008 & -0.00002 & 0.00007 & 0.00001 \\ -0.00002 & 0.00014 & 0.00001 & -0.00007 \\ 0.00007 & 0.00001 & 0.00999 & 0.00463 \\ 0.00001 & -0.00007 & 0.00463 & 0.00762 \end{pmatrix}.$$

First of all, we note that having the cost parameter normally distributed is problematic, as stated in Daly et al. (2012b), as a normal distribution has a positive probability mass at zero and therefore draws close to zero lead to very large WTP values. Hence, theoretically the mean and variance of the WTP are undefined. Again using 25,000 Halton draws, we therefore take the median of the WTPs and the median of the variances of the WTP. The median WTP is 0.0190 and the median variance is 0.0029, such that the median standard error is 0.0542. Hence, the 95 percent confidence interval based on the median values is (-0.0873, 0.1253). For completeness, we also computed the average WTP and average variance, yielding -0.0312 and 1656835850.8956, respectively, leading to a not very meaningful 95 percent confidence interval of (-79778.8924, 79778.8500).

5.3. Normal divided by normal (dependent)

Similar to the previous example, we assume that both β_k and β_c are random parameters following a normal distribution, however, this time we assume they are dependent such that the covariances are also estimated. We describe the parameters as a function of the first three elements in the Cholesky matrix, namely (a_{11}, a_{21}, a_{22}), which yields $\beta_k = \mu_k + a_{11}z_k$ and $\beta_c = \mu_c + a_{21}z_k + a_{22}z_c$, with $z_k \propto N(0,1)$ and $z_c \propto N(0,1)$. Hence, the conditional WTP is asymptotically distributed as

$$\hat{w}_k(z_k,z_c)\xrightarrow{D}N\left(w_k(z_k,z_c),\frac{1}{\beta_c^2(z_k,z_c)}\begin{pmatrix}1\\ z_k\\ -w_k(z_k,z_c)\\ -w_k(z_k,z_c)z_k\\ -w_k(z_k,z_c)z_c\\ a_{11}-a_{21}w_k(z_k,z_c)\\ -a_{22}w_k(z_k,z_c)\end{pmatrix}^T\right.\cdot \tag{29}$$

$$\left.\begin{pmatrix}\Omega_{(\mu_k,a_{11},\mu_c,a_{21},a_{22})} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & 1 & 0\\ \mathbf{0} & 0 & 1\end{pmatrix}\begin{pmatrix}1\\ z_k\\ -w_k(z_k,z_c)\\ -w_k(z_k,z_c)z_k\\ -w_k(z_k,z_c)z_c\\ a_{11}-a_{21}w_k(z_k,z_c)\\ -a_{22}w_k(z_k,z_c)\end{pmatrix}\right),$$

with

$$\Omega_{(\mu_k,a_{11},\mu_c,a_{21},a_{22})}=\begin{pmatrix}0.00010 & 0.00000 & 0.00011 & 0.00058 & -0.00026\\ 0.00000 & 0.00018 & 0.00011 & -0.00013 & 0.00011\\ 0.00011 & 0.00011 & 0.00671 & -0.00008 & -0.00381\\ 0.00058 & -0.00013 & -0.00008 & 0.01626 & -0.00459\\ -0.00026 & 0.00011 & -0.00381 & -0.00459 & 0.00957\end{pmatrix}.$$

As in the previous example, a normally distributed parameter in the denominator is problematic and the WTP moments are undefined. Hence, the only meaningful results we can state are based on the median values. The median WTP is 0.0211, and the median variance is 0.0081, such that the median standard error is 0.0898 and the 95 percent confidence interval is (-0.1549, 0.1972).

5.4. Fixed divided by lognormal

In the fourth and final example, we consider a fixed parameter for β_k and a randomly distributed cost parameter following a lognormal distribution, hence $\beta_c=\exp(\mu_c+\sigma_c z_c)$ with $z_c\propto N(0,1)$. The Jacobians are $\nabla_{(\beta_k)}\beta_k=1$, $\nabla_{(\mu_c,\sigma_c)}\beta_c=(\beta_c,z_c\beta_c)^T$, and $\nabla_{z_c}\beta_c=\sigma_c\beta_c$, respectively, such that the conditional WTP is asymptotically distributed as

$$\hat{w}_k(z_c)\xrightarrow{D}N\left(w_k(z_k,z_c),\frac{1}{\beta_c^2(z_c)}\begin{pmatrix}1\\ -w_k(z_k,z_c)\beta_c(z_c)\\ -w_k(z_k,z_c)z_c\beta_c(z_c)\\ -\sigma_c\beta_k(z_k)\end{pmatrix}^T\begin{pmatrix}\Omega_{(\beta_k,\mu_c,\sigma_c)} & \mathbf{0}\\ \mathbf{0} & 1\end{pmatrix}\begin{pmatrix}1\\ -w_k(z_k,z_c)\beta_c(z_c)\\ -w_k(z_k,z_c)z_c\beta_c(z_c)\\ -\sigma_c\beta_k(z_k)\end{pmatrix}\right). \tag{30}$$

with

$$\Omega_{(\beta_k,\mu_c,\sigma_c)} = \begin{pmatrix} 0.00001 & -0.00002 & 0.00001 \\ -0.00002 & 0.01985 & -0.00652 \\ 0.00001 & -0.00652 & 0.00215 \end{pmatrix}.$$

The lognormal produces always negative values for the cost parameter, hence the mean and variance of the WTP are computable and meaningful. The mean WTP is 0.1959 with a mean variance of 0.2213, such that the mean standard error is 0.4704, yielding a 95 percent confidence interval of (-0.7261, 1.1179). If we would again take the median instead of the mean, we would obtain a median WTP of 0.0941 with a median standard error of 0.1161, resulting in a confidence interval of (-0.1335, 0.3216), which is more in line with the findings from the first example with a normally distributed random coefficient divided by a fixed coefficient. The difference between the confidence intervals obtained through the mean and the median are quite different, which is also illustrated in Figure 3. The red line indicates the sampling distribution using the mean, while the green line represents the sampling distribution using the median. Since dividing by the lognormal distribution results in some cases to rather large values for the WTP (since values close to zero are likely to occur, although values equal to zero cannot occur), the mean variance is large. Using the median, extreme values do not have a large impact.

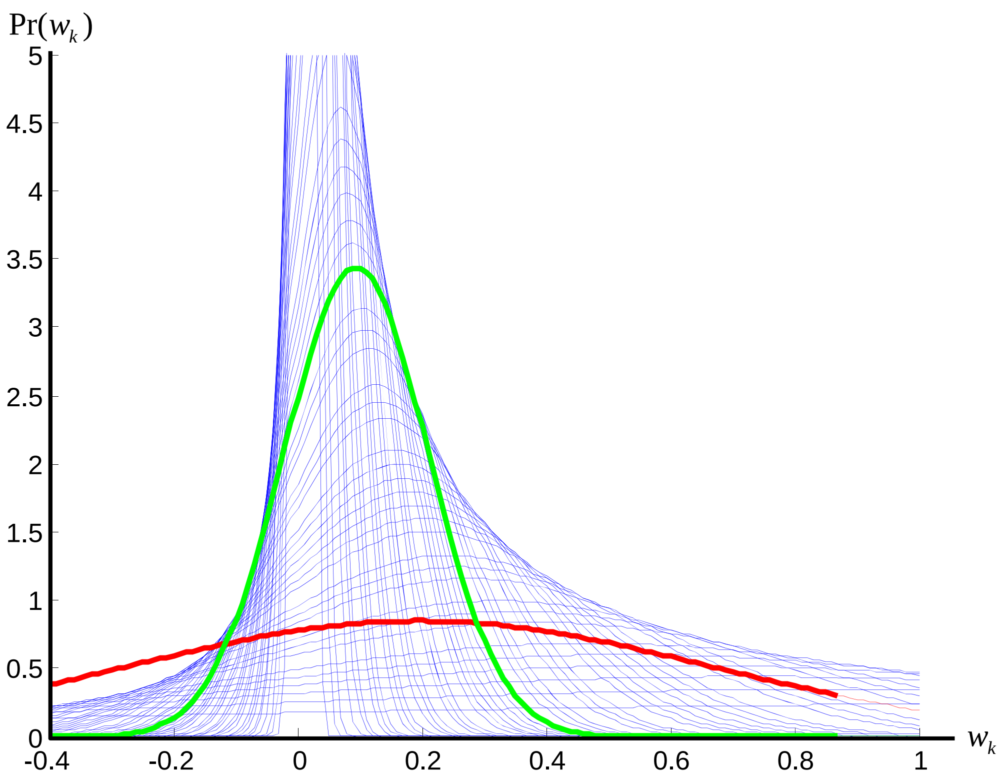


Figure 3: Simulated normal distributions and the sampling distribution (Fixed / Lognormal)

6. Discussion

In this paper we have presented a method to determine the confidence intervals of WTP measures taken from a RCL model in which one or more of the parameters following a random distribution. The method works by first reformulating the WTP as a function of the distributional parameters and some parameter-free standard distributions and then applying the Delta method. Hence, the method can be applied for any combination of normal distributions, lognormal distributions, uniform distributions, exponential distributions, triangular distributions, and more. We have also shown that correlations between (log)normally distributed parameters can be taken into account. The method takes the variance-covariance matrix of the respective model parameter estimates into consideration, translating the uncertainties in the estimation of the distributional parameters into uncertainty in the WTP measure as presented in confidence intervals.

As Daly et al. (2012b) points out, one has to be careful that the random parameter in the denominator (typically the cost parameter) does not go through zero. Hence, the probability mass of this distribution should be nil at zero, such as in the lognormal distribution. Otherwise, the mean and variance of the WTP are theoretically not defined. The parameter estimates in Table 2 illustrate that in this example using a fixed coefficient or a lognormally distributed coefficient for the cost parameter results in a worse model fit, although these are the only two models presented in the table that are able to produce theoretically defined WTPs. If one would like to select the model with the best model fit, the pragmatic way out would be to take the median instead of the mean. The analyst is therefore confronted with a dilemma, which deserves a closer look at the matter.

As mentioned, it is not necessary to assume linear utility functions, the methodology proposed in this paper can also handle nonlinearities in the parameters and/or in the attributes. In that case, the derivatives in Eqn. (6) will not be a simple ratio of β_k and β_c, but rather a more general function of these parameters and possibly the attribute levels, *x*. Furthermore, the Jacobians $\nabla_{\theta_k}\beta_k$ need to be replaced by a more general Jacobian $\nabla_{\theta_k} h_k(\beta_k)$, where $h_k(\beta_k) = \partial g_j / \partial x_{jk}$. The algebra may become a bit more tedious, but the equations and the main principle remain the same.

References

Amador, F.J., R.M. González, and J. de D. Ortúzar (2005). Preference heterogeneity and willingness to pay for travel time savings. *Transportation*, **32**, 627-647.

Armstrong, P., R.A. Garrido, J. de D. Ortúzar (2001). Confidence interval to bound the value of time. *Transportation Research Part E*, **37**(1), 143-161.

Bhat, C. (2001) Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, **35**, 677-693.

Bliemer, M.C.J., J.M. Rose, and S. Hess (2008) Approximation of Bayesian efficiency in experimental choice designs. *Journal of Choice Modelling*, **1**, pp. 98-127.

Campbell, D. (2007). Willingness to pay for rural landscape improvements: combining mixed logit and random-effects models. *Journal of Agricultural Economics*, **58**(3), 467-483.

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

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

Espino, R., J. de D. Ortúzar, and C. Román (2006). Confidence intervals for willingness to pay measures in mode choice models. *Networks and Spatial Economics*, **6**, 81-96.

Ettema, D., H. Gunn, G. de Jong, and K. Lindveld (1997). A simulation method for determining the confidence interval of a weighted group average value of time. *Proceedings of the European Transport Conference*, 101-112.

Greene, W.H. (2008) *Econometric analysis*. Prentice Hall, 7th ediction.

Haab, T.C. and K.E. McConnell (2003) *Valuing environmental and natural resources: The econometrics of non-market valuation*, Edward Elger Publishing.

Hensher, D.A., and W.H. Greene (2003). The mixed logit model: the state of the practice. *Transportation*, **30**, 133-176.

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

Krinsky, I. and Robb, A.L. (1990). On approximating the statistical properties of elasticities: a correction. *Review of Economics and Statistics*, **72**(1), 189–190.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P. (ed.). *Frontiers in Econometrics*, Academic Press, New York, 105-142.

McFadden, D. and Train, K. (2000). Mixed MNL Models for discrete response. *Journal of Applied Econometrics*, **15**(5), 447-470.

Michaud, C., D. Llerena, and I. Joly (in press). Willingness to pay for environmental attributes of non-food products: a real choice experiment. *European Review of Agricultural Economics*.

Sándor, Z., and K. Train (2004) Quasi-random simulation of discrete choice models. *Transportation Research Part B*, **38**, pp. 313-327.

Scarpa, R. and J.M. Rose (2008). Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *Australian Journal of Agricultural and Resource Economics*, **52**(3), 253-282.

Sillano, M., and J. de D. Ortúzar (2005). Willingness-to-pay estimation with mixed logit models: some new evidence. *Environment and Planning A*, **37**, 525-550.