Article:
Daly, A., Hess, S. and Ortúzar, J.D.D. (2023) Estimating willingness-to-pay from discrete choice models: Setting the record straight. Transportation Research Part A: Policy and Practice, 176. 103828. ISSN: 0965-8564

https://doi.org/10.1016/j.tra.2023.103828

Estimating willingness-to-pay from discrete choice models: setting the record straight

Andrew Daly[a], Stephane Hess[b] and Juan de Dios Ortúzar[c]

[a] Choice Modelling Centre & Institute for Transport Studies, University of Leeds, Leeds, UK; e-mail: andrew@alogit.com (ORCID 0000-0001-5319-2745)

[b] Choice Modelling Centre & Institute for Transport Studies, University of Leeds, Leeds, UK; e-mail: s.hess@leeds.ac.uk (ORCID: 0000-0002-3650-2518)

[c] Department of Transport Engineering and Logistics, Instituto Sistemas Complejos de Ingeniería (ISCI), BRT+ Centre of Excellence, Pontificia Universidad Católica de Chile, Santiago, Chile; e-mail: jos@ing.puc.cl (ORCID: 0000-0003-3452-3574)

Abstract

The estimation of indicators of willingness-to-pay (WTP) and the computation of associated measures of uncertainty have attracted much interest in the general area of choice modelling, but have also been characterised by a substantial amount of confusion and misguided discussions. We examine this problem in depth, both theoretically and empirically, in an effort to illustrate the most appropriate approaches to solving this problem, which has important repercussions for practice in our field. Our findings should be useful to analysts advising on transport policy and infrastructure project evaluation, in particular when requiring estimates of the value of travel time.

Keywords: willingness-to-pay; discrete choice models, Delta method, social evaluation

1. Introduction

Estimating willingness-to-pay (WTP) and computing the related measures of uncertainty from discrete choice models is one of the key applications of these models. In determining the most appropriate way to estimate the accuracy of WTP measures, it is useful to consider the reasons for requiring to know this accuracy. First, the standard errors of WTP measures may be needed to give a general indication of their accuracy, for statistical significance tests, and for comparison with other models and other studies. Second, confidence limits for WTP measures may be required, for example, to indicate the range of reliability of policy appraisals based on them.

Unfortunately, this topic has been plagued by confusion and misguided discussions, which we have partially tried to address in previous papers (Armstrong et al., 2001; Daly et al., 2012a, 2012b; Hess et al., 2005; Sillano & Ortúzar, 2005; but also Daly and Zachary, 1975). Notwithstanding these efforts, we have found that much confusion persists, especially outside our main discipline of transport research. For example, a recent paper by Carson & Czajkowski (2019), states that … "*a substantive problem exists with the widely-used ratio of coefficients approach to calculating willingness to pay (WTP) from discrete choice models*" and contains some misconceptions that motivated our discussion. We also provide additional detail and background with the hope of clarifying other ongoing confusions in several papers and conference presentations in the general choice modelling field, not losing sight of our initial interest in transport applications, particularly estimating the value of travel time.

In choice models using the paradigm of random utility maximisation, and with a particular focus on utility functions that are linear in attributes and coefficients, we define:

$$V = \beta_c . c + \beta_x . x + \cdots \quad (1)$$

where V is the systematic utility (we omit subscripts for decision makers, choice situations and alternatives), x is the attribute for which we wish to calculate the WTP[1], and c is the cost attribute. Assuming that the marginal utilities are the same across alternatives and choice situations, we have that, for a given decision maker (Beesley, 1965; Gaudry et al., 1989):

$$WTP = \frac{\delta V}{\delta x} / \frac{\delta V}{\delta c} = \frac{\beta_x}{\beta_c}, \quad (2)$$

that is, the WTP is simply calculated as the ratio of two marginal utility coefficients.

While some complications arise in situations where the marginal utilities $\frac{\delta V}{\delta x}$ and $\frac{\delta V}{\delta c}$ depend on the level of cost and/or the attribute in question (i.e. with non-linear utility functions) and/or on observable characteristics of the decision-maker (i.e. with deterministic heterogeneity), much greater issues arise of course when attributes are allowed to vary randomly across individuals, such as in the now extremely popular Mixed Multinomial Logit (MMNL) model (see for example Train, 2009). In that case, the analyst has the ability to make a choice of a population mixing distribution, and much work has looked at behavioural realism; for example, whether a Normal distribution would make sense for a cost coefficient as it implies accepting that this could have a positive value for some individuals (cf. Hess et al., 2005; Sillano & Ortúzar, 2005). An equally important issue in the MMNL case relates to the fact that we now have the possibility of the denominator in (2) not being bounded away from zero, that is, having a continuous distribution that includes zero. Daly et al. (2012b) provide extensive discussions in this context as well as a theorem showing which assumptions for the population-level distribution of the parameter β_c ensure that the moments of the distribution of WTP are defined.

A completely separate issue relates to the computation of measures of uncertainty for the components in (2), in the form of standard errors and/or confidence intervals. For the sake of simplicity, let us assume that we have a utility function that is linear in attributes and coefficients, meaning that the WTP is given by a ratio of two coefficients, say $WTP = \frac{\beta_x}{\beta_c}$. The question then arises of how to calculate an error measure for this ratio, and compute confidence intervals. This is an issue both with fixed coefficients models, which is the main focus of this paper, and in the presence of random heterogeneity, as discussed by Bliemer & Rose (2013).

We need to state categorically that these two issues, that is, the specification of random heterogeneity, and the computation of error measures, are quite separate. Unfortunately, however, findings in relation to the former (assumed population mixing distributions) have been used in discussions about the latter (computation of measures of uncertainty), often as a result of misunderstanding the asymptotic properties of maximum likelihood estimates.

The remainder of this paper is organised as follows. To motivate our discussion, we first discuss the utility specification proposed by Carson & Czajkowski (2019) and briefly show that this is ill-conceived. In section 3 we explain the differences between Normality and asymptotic Normality which are a typical source of confusion. Section 4 considers the calculation of

[1] Our focus here is on WTP, but the points apply to marginal rates of substitution more generally.

measures of uncertainty for ratios of marginal utilities, and Section 5 presents a brief empirical comparison. Finally, section 6 offers some conclusions

2. The Carson & Czajkowski 'new baseline' model

Carson & Czajkowski (2019) discuss the computation of WTP in the case of models without random heterogeneity. They apply a negative exponential to the coefficient estimated for the marginal cost to prevent it being positive, reformulating the model from (1) as follows:

$$V = -\exp(\gamma)\,.\,c + \beta_x.x + \cdots \quad (3)$$

The WTP from this specification would then be calculated as $\text{WTP} = \frac{\beta_x}{-\exp(\gamma)}$. While this restriction may appear reasonable in terms of economic theory, it is questionable whether it is beneficial in model building. Imagine a situation where due to either data problems or an endogenous price/quality relationship (e.g. Palma et al., 2016), the value for the cost coefficient is positive at the optimum. Such a finding is beneficial to the analyst in terms of then being able to improve the model or deal appropriately with any data issues. Therefore, preventing the estimation from highlighting a problem is not in anyone's interest. With the specification in Equation (3), γ would then tend to negative infinity, such that $-\exp(\gamma)$ would tend to zero, rather than the optimal value.

The authors also suggest that their formulation … "*avoids problems associated with non-existent moments of the resulting WTP ratio distribution*". This point, unfortunately, mistakes asymptotic distributions of uncertainty for assumed population level distributions, a point we address in detail in the following section of the paper. They also state that, without this negative exponential restriction, "*the mean and standard deviation of the resulting WTP ratio distribution are undefined*". This is factually incorrect in the case of fixed coefficients models (i.e. without a population-level distribution), where the WTP ratio does not have a distribution; presumably they intended to refer to the distribution of the estimate of WTP. But again, this misunderstands the nature of asymptotic distributions. We show, below, that the results of specifications (1) and (3) are exactly equivalent across specifications, as are their statistical properties. The authors acknowledge the mathematical equivalence between the specifications, but motivate the use of the exponential transform on the grounds of small sample size properties. As we will explore in an empirical example later in the paper, these advantages do not in fact exist.

The use of the above specification also implies two practical problems. A key step in choice modelling involves testing the significance of individual model coefficients; in the case of a cost coefficient, testing whether the estimate is different from zero. With the proposed specification (3), the fact that we obtain an estimate of γ actually complicates matters. A t-ratio for γ against zero only tells us whether the cost coefficient is different from -1, which is not helpful. To establish whether the cost coefficient has a significant impact on the model would require either a likelihood ratio test against a (separately estimated) model excluding the cost coefficient, or the calculation of a standard error for $-\exp(\gamma)$, a point we return to in Section 3. Finally, and this was not considered by the authors, the use of an exponential transform in estimation can create more harm than good. A search for maximum likelihood estimates involves making changes to the parameter values, and changes of the same step size can have a much larger impact for those parameters to which an exponential transform is applied.

3. Population-level *vs* asymptotic distributions

It is well-known that the maximum likelihood estimates of utility function coefficients have an asymptotically Normal distribution. In this context, some authors have argued that calculating the moments of the distribution of the WTP estimate may be problematic, because of the non-existence of moments of a ratio distribution resulting from dividing two normally distributed variables (a fact long known in statistics). In that case, the mean and standard deviation of the resulting WTP ratio distribution would be undefined, and the resulting distribution would be not Normal, being typically skewed and potentially bimodal.

But this issue relates to the ratio of normally distributed parameters, and not to the ratio of estimates that have the lesser property of *asymptotic* Normality. Given this confusion, it seems necessary to explain the distinction between an assumption relating to the distribution of a parameter in the population and the uncertainty affecting the estimator of a coefficient.

Let us assume that our model contains a vector of unknown coefficients β which are estimated using the maximum likelihood criterion. We take a frequentist view that there exists a single true value β^*. Provided regularity conditions are met and the model is correctly specified, the *expected score* (first derivative of the log-likelihood function LL with respect to the model coefficients) is zero at the optimum and the maximum likelihood estimates (MLE) $\hat{\beta}$ tend, as the sample size N increases, to a Normal distribution around the true values β^*. In particular, we have that $\sqrt{N}(\hat{\beta} - \beta^*) \rightarrow \mathcal{N}(0, \Omega)$, where Ω is the covariance matrix, given by:

$$\Omega = \left(-E\left(\frac{\partial^2 LL(\beta)}{\partial \beta^2} \right) \right)^{-1} \tag{4}$$

Maximum likelihood estimates are asymptotically efficient, meaning that they achieve the Cramér-Rao lower bound in the limit. The standard errors, which give a measure of the uncertainty affecting an estimator of a coefficient in a model, are given by the square root of the diagonal of Ω.

Two important observations need to be made here.

First, given finite sample sizes, the key property of MLE is one of asymptotic Normality, rather than Normality, and this important distinction seems to be poorly understood. Asymptotic Normality can be applied to obtain point estimates like standard errors and 't' ratios, but what happens as we move further away from the optimum, e.g. in calculating confidence limits, is not defined by maximum likelihood theory. The estimates are not distributed exactly Normal, and any theoretical insights/claims based on an assumption of Normality may be misguided.

Second, asymptotic Normality is a property of MLE of model coefficients, unlike population mixing distributions (as in the case of, for example, MMNL models), where an analyst makes a decision on the assumed shape of a distribution for each marginal utility component. Asymptotic normality applies to estimates of fixed marginal utility components, as in the MNL, as well as estimates of the parameters explaining the distribution of coefficients varying randomly across individuals, as in MMNL. For the latter, the estimates of the parameters of the assumed distribution will have this property, and this is independent of the distributional assumptions made by the analyst.

The above two points highlight that asymptotic Normality of estimates is a fundamentally different property from a population level distribution. The similarity of form is beguiling but needs to be resisted. It is widely known that the inverse of a Normal distribution does not have defined moments, and this knowledge is at the heart of any guidance (e.g. in Daly et al., 2012b) to not assume Normal distributions over the population for a cost coefficient used as the denominator in (2). The confusion, which seems to persist in the literature, is the mistaken belief that this same finding also applies to the calculation of a ratio where the denominator is a parameter whose estimate follows an asymptotic Normal distribution. However, such theoretical insights on population level distributions of ratios based on distributed parameters must not be used to say anything about the statistical properties of these ratios, that is, their error measures, as we now discuss, drawing on Daly et al. (2012a).

Let us consider a re-parametrisation of the model by an invertible one-to-one vector function to obtain a vector η with the same dimension as β:

$$\eta = g(\beta) \quad \text{and} \quad \beta = g^{-1}(\eta) \tag{5}$$

With these conditions, and assuming continuity and differentiability, Cramer (1986) shows that the essential MLE properties are not affected by the transformation, so that $\hat{\eta} = g(\hat{\beta})$ is a maximum likelihood estimate of η. Greene (2008) similarly indicates that, if β is asymptotically Normally distributed and $g(\hat{\beta})$ is continuous, then $g(\hat{\beta})$ is also asymptotically Normally distributed.

This highlights the clear distinction between Normality and asymptotic Normality. If we have that a given parameter β_k follows a Normal distribution in a population, then the distribution of its inverse does not have finite moments. However, if $\hat{\beta}_k$ is an asymptotically Normal parameter estimate, then the estimate of its inverse maintains the property of asymptotic Normality.

In summary, because of the invertible relationship (5), we could just as easily have estimated a value for η and derived a value for β as vice versa. This point is fundamental to the understanding of the re-parametrisation of models and the properties of the MLE. The transformation does not affect the properties of minimum variance, consistency and asymptotic Normality that apply to MLE.

Let us now consider the implications of this discussion for a utility specification such as (3). Its use will yield estimates for γ, β_x and any other parameters included in the model. Using the notation from (5), we then have that $\eta = -\exp(\gamma)$, that is, the function g is the negative exponential. Assuming that the optimum estimate of η is negative, we can then state a number of facts:

- With g being a differentiable function, and with $\hat{\gamma}$ being a maximum likelihood estimate, the property of asymptotic Normality also applies to $\eta = -\exp(\gamma)$, such that $\hat{\eta} = -\exp(\hat{\gamma})$ is a MLE of η, with the associated MLE properties of minimum variance and consistency.
- The estimate we obtain for γ (i.e. $\hat{\gamma}$) is exactly equal to $\log(-\hat{\beta}_c)$, with $\hat{\beta}_c$ being the estimate obtained when using a model of the form in (1).
- The estimate and standard errors for any other untransformed coefficients (e.g. β_x) remain the same as those obtained when using a model of the form in (1).

- The estimate for the WTP (i.e. $\frac{\hat{\beta}_x}{-\exp(\hat{\gamma})}$) is the same as that obtained using a model of the form in (1), which gives $\frac{\hat{\beta}_x}{\hat{\beta}_c}$.

This discussion highlights misconceptions in the need for a specification such as (3) and already, in itself, removes the need for such a solution.

4. Calculating measures of uncertainty for WTP

We look now at the key question of how to calculate measures of uncertainty for WTP. We start by looking at the computation of standard errors before discussing the computation of confidence intervals.

A simple method for calculating the standard deviation of a function of random parameters is the Delta method, which goes back at least to Bessel (1838). It is generally presented in our field as an approximation, based on a Taylor series expansion (see Ortúzar & Willumsen, 2011, Chapter 9). Statistical textbooks show that, given appropriate conditions, a first-order approximation to the error in the estimate of $\eta = g(\beta)$ induced by the error in the estimate of β is given by (e.g. Greene, 2008):

$$cov(\eta) = g'^{T} \Omega g' \tag{6}$$

where g' is the matrix of first derivatives of the function g with respect to β, and Ω is the covariance matrix of the estimates of β.

However, in the case of MLE, these formulae can be given a different interpretation. Cramer (1986) shows that, in the context of (5), the covariance of η as given in (6) is the Cramér-Rao lower bound, so that the estimate of η has the full MLE properties. Daly et al. (2012a) used this result to show that a likelihood function which has been maximised with respect to one set of parameters can also be considered to have been maximised with respect to parameters derived by transformations of the first set. The optimum values of the new coefficients are the transformed values of the old ones, and their estimation errors are given by the formulae of the Delta method, which in this context, and contrary to prevailing views in the literature, is not an approximation. Therefore, a model can be estimated with a specification that is convenient and then transformed to a different specification as required, without losing the maximum likelihood properties of the estimates. A full statement of this central result is given in Appendix 1.

In the context of WTP estimation based on a model without random heterogeneity, such as the MNL[2], it may be convenient to estimate the model with a utility function of the form shown in (1). This would yield optimal estimates $\hat{\beta}_c$ (which we can assume to be negative) and $\hat{\beta}_x$, along with estimation of asymptotic standard errors σ_c and σ_x and covariance $\sigma_{c,x}$ of the estimates. Equation (2) is used to calculate the WTP from this model. We then have that $g = \frac{\beta_x}{\beta_c}$, and applying (6), we can calculate the standard error of the WTP as:

$$\sigma_{WTP} = \left(\hat{\beta}_x / \hat{\beta}_c\right) \sqrt{\left(\frac{\sigma_x^2}{\hat{\beta}_x^{\,2}} + \frac{\sigma_c^2}{\hat{\beta}_c^{\,2}} - \frac{2\sigma_{c,x}}{\hat{\beta}_x \hat{\beta}_c}\right)} \tag{7}$$

[2] This includes models with non-linear functions and/or with systematic taste variations.

This is the correct asymptotic error for the WTP rather than an approximation. Readers can easily convince themselves of this using a simple empirical proof, by estimating a model in WTP space, as we do in Section 5. It is well known (Cameron, 1988; Train & Weeks, 2005; AHCG, 1996) that the model in (1) can be rewritten as:

$$V = \beta_c.(c + v_x.x + \cdots) \tag{8}$$

in which case we would have obtained optimal estimates $\hat{\beta}_c$ (as before) and the WTP $\hat{v}_x$, along with their standard errors and covariance. Train & Weeks (2005) show that (8) is consistent with (1), and leads to the same model fit in the specific case of fixed coefficient models with linear utility functions. Similarly, it can then easily be seen that $\hat{v}_x = \hat{\beta}_x/\hat{\beta}_c$, that is, the directly estimated WTP from a model using (8) is in line with that based on the estimates of a model in preference space (2). What the results in Daly et al. (2012a) show, is that the standard error obtained for $\hat{v}_x$ is identical to the standard error calculated using (7) for a model estimated in preference space, that is, the value that would be given by the Delta method. The model can be estimated using (1) or equivalently using (8), as is convenient.

The same reasoning applies to a specification such as (3). We have already discussed how this specification yields the same WTP as (1) and hence also (8). We can now also state that the Delta method can be used to calculate the standard error for $\hat{\eta}$ as $\sigma_\eta = \exp(\hat{\gamma})\,\sigma_\gamma$, where $\hat{\gamma}$ is the MLE, with standard error σ_γ. The resulting standard error σ_η is the same as that obtained for $\hat{\beta}_c$ using a model of the form in (1). The asymptotic distribution of the estimate for the parameter without a restriction ($\hat{\beta}_c$) is thus the same as the asymptotic distribution of the restricted parameter (i.e. $-\exp(\hat{\gamma})$). How can it be that an estimate to which we apply the negative of an exponential, which is by definition negative, still has an asymptotic Normal distribution? The key lies in the word *asymptotic* – Normality is derived as an asymptotic property and applies only in a neighbourhood around the MLE, where the second-order approximations for these two functions are equal.

Further, applying the Delta method, we have that $\sigma_{WTP(3)} = \sqrt{\frac{\sigma_x^2 + \hat{\beta}_x{}^2\sigma_c^2 - 2\hat{\beta}_x\sigma_{x,\gamma}}{\exp(2\hat{\gamma})}}$, which is the same as the standard error for the WTP for models using (1), and the same as σ_{v_x} for models using (8).

It should also be noted that the use of the Delta method does not, in any way, imply that the cost coefficient needs to be fixed. If β_c and possibly also β_x are distributed across individuals, the analyst needs to first use the results in Daly et al. (2012b) to determine the existence of moments for the distribution of WTP. If these moments can be expressed as a function of estimated parameters, then the analyst can again use the Delta method for the moments of the distributed WTP. We present a simple illustration of this in Appendix 2.

In addition to the Delta method, the work of Krinsky and Robb (1986, 1991) is often cited in environmental economics and marketing. Their simulation procedure is lengthy but correct in principle for the cases they present in 1991, where they found close agreement with the Delta method. The Krinsky and Robb approach is based on drawing from the asymptotically Normal distribution of the estimates, so that it yields asymptotically Normal estimates of functions of the estimates, providing conditions of boundedness are met. The approach can be used to calculate standard errors for functions only in those cases where the functions do not involve ratios. In cases where ratios are involved, such as in WTP, the Krinsky and Robb approach

must not be used to compute standard errors, though it can still be used for the computation of empirical confidence intervals.[3]

The above discussion has explained how standard errors can be calculated accurately for WTP measures. Our attention now turns to confidence intervals around the estimates. Care is again required given the asymptotic nature of the distribution of coefficient estimates. The natural tendency to calculate a C% confidence interval using $\hat{\beta} \pm z^* \sigma_\beta$, where z^* is the upper $\frac{1-C}{2}$ critical value for a $N(0,1)$ distribution, relies on the assumption of asymptotic Normality. In the case of coefficients (or functions thereof, such as WTP) with small standard errors relative to their estimates (i.e. high t-ratios), the above calculation can be acceptable. However, it is far from clear what level of statistical significance is required to make the $\hat{\beta} \pm z^* \sigma_\beta$ calculation acceptable. Other approaches for calculating confidence intervals thus deserve attention too.

A commonly used method in environmental economics is that of Fieller (1954), which calculates the confidence limits as:

$$V_{min}, V_{max} = \frac{S_{cx} - \sqrt{S_{cx}^2 - S_{cc}S_{xx}}}{S_{cc}}, \frac{S_{cx} + \sqrt{S_{cx}^2 - S_{cc}S_{xx}}}{S_{cc}} \qquad (9)$$

where $S_{cc} = \hat{\beta}_c^2 - t^2\sigma_c^2$; $S_{xx} = \hat{\beta}_x^2 - t^2\sigma_x^2$; $S_{cx} = \hat{\beta}_c\hat{\beta}_t - t^2\sigma_{cx}$, and where t is the critical t value. What is often not recognised is that this assumes exact Normality rather than asymptotic Normality. This calculation is also offered as one approach by Armstrong et al. (2001); note that although their formula looks rather different it is in fact the same.

Another approach is the 'likelihood ratio' (LR) method also proposed by Armstrong et al. (2001), who draw a parallel between the test that a statistic is outside the confidence limit and the χ^2 test that estimating a free parameter gives a significant improvement from fixing the parameter at the confidence limit. This approach is ingenious and avoids the issue caused by the denominator approaching zero, but relies heavily on knowledge of the likelihood function. For example, if the function contains local optima, it is quite possible that spurious results can be obtained. The LR approach is, as acknowledged by Armstrong et al. (2001), also computationally intensive. They find that the confidence limits given by the Fieller approach are similar to those of the LR approach, at least for the data they tested. The Fieller formula is based only on the curvature of the likelihood function at the optimum, while the LR approach looks at likelihood values away from the optimum, when the curvature of the function may change. The Fieller formula is simply not applicable in this case, because the distribution of the estimates is asymptotically Normal, not Normal.

An approach that has received less attention than one would expect in this context is bootstrapping[4]. The Bootstrap operates by sampling N observations from the original sample, *with replacement*, repeated a number of times as chosen by the analyst, say leading to S samples $D_1, \ldots, D_S$. Individual models are then estimated yielding S sets of coefficient values (e.g. the vector $\hat{\beta}^{(s)}$ in run s). The concept on which the Bootstrap is based is that, if the original data

[3] A referee points out that unbounded results can be avoided in the Krinsky & Robb approach by truncating the distribution of the denominator at or close to zero. However, this changes the mean and standard deviation of the distribution and can only be acceptable when an extremely small fraction of the distribution is censored. In calculating confidence limits, however, the extreme values are automatically censored as they fall outside a reasonable confidence range.

[4] An alternative resampling technique that could also be used is the Jackknife (Shao & Tu, 1995).

is a representative sample from the population being studied, then the Bootstrap samples also resemble samples that might be drawn if the sampling were done again. For that reason, they give the sampling variance that may be expected.

In the context of data with multiple observations per individual, it is necessary to sample at the level of individuals rather than observations. The covariance matrix of the coefficient estimates $cov(\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(S)})$ is calculated as their covariance over the Bootstrap samples and an empirical confidence interval for WTP can be obtained from the distribution of $WTP^{(s)} = \frac{\hat{\beta}_x^{(s)}}{\hat{\beta}_c^{(s)}}, \forall s$. This process is, of course, computationally expensive with a large value for S, especially with complex models, but does not rely on any assumptions about Normality, asymptotic or otherwise.

5. Empirical example

We now present a brief empirical example using a stated preference (SP) dataset collected by Axhausen et al. (2008) in the context of value of time calculations. A set of 388 people faced nine choices each between two public transport routes, both using train. The alternatives were described by travel time (tt), travel cost (tc), headway (hw, time between successive trains) and the number of interchanges (ch). We estimate three binary logit models[5] using specification (1), in preference space, specification (8), in WTP space, and specification (3) in preference space. All models were estimated using Apollo (Hess & Palma, 2019).

The results of the model estimation on the full sample are shown in Table 1. For the conventional preference space specification and the Equation (3) specification we also show the calculated WTP values, along with the standard errors computed using the Delta method. For the Equation (3) specification we also do this for the cost coefficient, which is obtained as $-\exp(\gamma)$. We show both 'Classical' errors, based on the inverse Hessian matrix, and 'Robust' errors, based on the 'sandwich' matrix (cf., Train, 2009; Huber, 1967). The 'robust' errors allow for certain types of specification error in the model. Appendix 1 shows that the Delta method can also be applied to robust error measures.

As expected, all three models produce exactly the same log-likelihood (LL), but the specifications where the utility is not linear in coefficients require more iterations to reach convergence. Any untransformed coefficients are the same across models, such as the non-cost coefficients for the first and third model, and the cost coefficient for the first and second model.

In addition, the implied WTP are the same in the first and third model, and equal to the estimated WTP measures in the second model (up to numerical precision). More importantly, for both classical and robust errors, the standard errors calculated using the Delta method for the WTP measures in the first model are equal (within the accuracy of the computation) to the standard errors for the directly estimated WTP in the second model. Finally, the same applies for the calculated standard errors for WTP when using the Equation (3) specification.

We have already pointed out that asymptotic Normality is a large sample property. Now, we also investigate small sample implications. For this, we estimated 96 models for each

[5] Note that this specification would not be acceptable for a proper estimation of WTP, as the MNL does not consider the 'pseudo panel' effect implicit in SP data. However, here it is convenient to illustrate the issues to which the present paper is addressed.

specification, going from the full sample to retaining only the first 20 respondents, removing around 1% of respondents from the end of the data at each step.

Table 1: Full sample estimation results on Swiss data

	Preference space (Equation 1)			**WTP space (Equation 8)**			**Equation (3) specification**		
Estimated parameters	4			4			4		
Iterations	15			21			18		
LL(final)	-1665.699			-1665.699			-1665.699		
Adj. $\rho^2(0)$	0.3102			0.3102			0.3102		
	estimate	**classical s.e.**	**robust s.e.**	**estimate**	**classical s.e.**	**robust s.e.**	**estimate**	**classical s.e.**	**robust s.e.**
β_{tt} (travel time)	-0.0598	0.0043	0.0067	-		-	-0.0598	0.0043	0.0067
β_{tc} (travel cost)	-0.1318	0.0135	0.0236	-0.1318	0.0135	0.0236	-		-
β_{hw} (headway)	-0.0375	0.0018	0.0023	-		-	-0.0375	0.0018	0.0023
β_{ch} (changes)	-1.1521	0.0434	0.0613	-		-	-1.1521	0.0434	0.0613
v_{tt}	-		-	0.4534	0.0285	0.0555	-		-
v_{hw}	-		-	0.2841	0.0301	0.0518	-		-
v_{ch}	-		-	8.7393	0.8993	1.5323	-		-
γ (travel cost, exp form)	-		-	-		-	-2.0264	0.1025	0.1791
Implied values									
β_{tc} (travel cost)	-		-	-		-	-0.1318	0.0135	0.0236
v_{tt}	0.4534	0.0283	0.0555	-		-	0.4534	0.0285	0.0555
v_{hw}	0.2841	0.0302	0.0518	-		-	0.2841	0.0302	0.0518
v_{ch}	8.7400	0.8996	1.5329	-		-	8.7400	0.8996	1.5329

Given the wealth of results, we present an overview in graphical format in Figure 1, where we rely on four key metrics, showing the evolution of the log-likelihood per observation, the estimate of one of the three WTP measures, namely the value of travel time (VTT), and finally the classical and robust standard error of the VTT, multiplied by the square root of the sample size used.

Figure 1 approximately here

For the LL per observation, we see that, in the plot, it is not possible to distinguish the results of the standard preference space model (Equation 1) and the WTP space model (Equation 8), where we observed a maximum difference of less than 0.001% for the WTP space model compared to the preference space model. For the model using the Equation (3) specification, two larger differences arise in the models removing 88% and 95% of the sample, respectively. For these two segments, that model converges to an inferior solution than the other two models. When estimating a model on just 12% of the sample, we obtain a LL per observation of -0.4511 compared to the base model LL of -0.4368, i.e. a drop of 3.2%, while, when estimating a model on just 5% of the sample, the difference is even larger, with the specification using Equation 3 giving a LL per observation of -0.4574, compared to -0.4312 for the base model, i.e. a drop of 5.7%. However, far from this being a result of some benefits of the specification, it seems an

illustration of the numerical issues that can arise in estimation due to the exponential transform, meaning that the model based on Equation 3 converges to an inferior solution.

The findings for the estimates of the WTP are in line with theory and the findings from the LL plot. The WTP space results never differ by more than 0.57% compared to preference space (with an average of 0.047% deviation). The same is true for the Equation (3) specification with the exception of very large deviations in the models removing 88% and 95% of the sample, respectively. For the first of these, we obtain a VTT of CHF 516.24/hr, compared to the base model result of CHF 34.21/hr (i.e. an overestimation by more than 1,400%). For the estimation on our smallest sample size, the difference is even larger, with an estimate of CHF 2767.92/hr, compared to the base model result of CHF 24.84/hr (i.e. an overestimation by more than 11,000%). Removing these two outliers gives us differences of up to 0.51% compared to the preference space model using Equation 1, with an average of 0.049%.

Finally, for the standard errors for WTP, we see maximum differences of 2.37% (classical) and 2.85% (robust) when comparing WTP space with preference space, with average differences of 0.19% (classical) and 0.22% (robust). For the Equation (3) specification we again see the impact of the two runs with poor convergence, with infinite classical standard errors and very large robust standard errors in those two cases. Removing these, we see maximum differences of 1.96% (classical) and 2.07% (robust) compared to the preference space model, with averages of 0.18% for both classical and robust.

This analysis of progressively smaller samples highlights that the three model specifications, except for two outliers, remain entirely consistent with each other, in line with our theoretical points. In addition, not only are the differences negligible, but there is no indication that they are anything other than white noise, and do not become larger at smaller sample sizes

We finally look at the use of bootstrapping for confidence intervals, contrasting the findings for the specification in equation (1) with the equation (3) specification. We look at the full data (388 individuals), and reduced samples containing the first 100 individuals, and the first 50 individuals, respectively. This latter is a reasonable lower bound in terms of sample size for any credible study. We use 200 Bootstrap samples in each case. The results of this process are summarised in Table 2.

Table 2: Bootstrap results

		N=388		N=100		N=50	
		Base Eq. (1)	Eq. (3)	Base Eq. (1)	Eq. (3)	Base Eq. (1)	Eq. (3)
MLE	$\widehat{WTP}$	27.21	27.21	23.43	23.43	34.24	34.25
	σ_{WTP} (robust s.e.)	3.33	3.33	5.32	5.32	13.88	13.90
Bootstrap results	Mean	27.84	27.85	24.79	43.20	34.65	636.33
	Median	27.57	27.57	23.85	23.85	32.97	33.44
	Standard deviation	3.26	3.26	7.08	117.30	38.67	6,655.84
	2.5th percentile	22.40	22.39	15.54	15.54	17.50	19.76
	97.5th percentile	34.75	34.75	40.47	441.14	95.00	3,176.13
	Skewness	0.37	0.37	2.16	6.87	-5.52	13.83
	Kurtosis	2.99	2.99	12.70	52.87	55.17	194.12
	$mean\left(\frac{LL_{C\&C} - LL_{base}}{LL_{base}}\right)$	-	0.00%	-	0.02%	-	0.11%

$max\left(\frac{LL_{C\&C} - LL_{base}}{LL_{base}}\right)$	-	0.00%	-	1.25%	-	8.41%

For the full data case, we see that the Bootstrap means are close to the MLE, and the Bootstrap standard deviations are very similar to the robust standard errors. In addition, there is only a little skewness in the distribution, and no excess kurtosis (with 3 corresponding to Normal). Thus, for the full sample the distribution remains largely symmetrical, allowing an analyst to compute confidence intervals on the basis of the standard errors (i.e. the asymptotic results). We also see no difference between both specifications in model fit.

Once we move to smaller samples (N = 100 and N = 50), we see evidence of asymmetry in the distribution of Bootstrap results for the base model specification (1), along with standard deviations that quickly outstrip the asymptotic standard errors, indicating the limited applicability of the asymptotic assumption. We also see substantial excess kurtosis and greater skewness. The negative value for skewness for the base specification (1) model with N = 50 deserves attention. This is caused by a small number of the samples (4 out of 200) yielding a positive cost coefficient, which, by being close to zero, gives large negative WTP. The Equation (3) specification prohibits such positive cost coefficients – leading to a loss in fit for these four samples (by up to 8.41%) given that the optimum is at a value not allowed by the specification. The further undesirable side effect of this is that for these samples the cost coefficient of the alternative specification becomes arbitrarily close to zero, leading to an explosion in WTP, and making the standard deviation and confidence intervals unusable (exactly the problem that the specification was trying to avoid).

Problems for the Equation (3) specification also arise with N = 100. Even though none of the Bootstrap samples led to positive cost coefficients with the base specification, the alternative model yields some outlying values, most likely due to slightly inferior solutions, a problem we attribute to the use of the exponential transform and the resulting numerical problems.

These findings suggest that, in those cases where the specification of Equation (3) is in line with the standard model, it offers no benefits. But in those cases where its results differ from the standard specification, the model fit in repeated sampling is worse than for the base specification (1), and the implied confidence intervals are not useful.

6. Conclusions

We have set out, in general, how modellers can deal in practice with the issues that arise in estimating WTP and its accuracy.

The fundamental issue is that many authors seemingly fail to appreciate the difference between asymptotic Normality and Normality. The lesser property of maximum likelihood estimates, that they are distributed asymptotically Normal, cannot justify imposing a full Normal distribution. It is the assumption of a full Normal distribution that causes the problem that the denominator of the WTP ratio seems to be close to zero.

In contrast, the result in Daly et al. (2012a) shows that the same model could have been estimated with a different specification, in which the WTP is estimated directly and therefore its estimate has an asymptotic Normal distribution. This apparent paradox is caused by the fact that the estimates are not distributed Normal, but only asymptotically Normal. Once again, the errors can be calculated exactly using the Delta method, giving the same standard errors as for preference and WTP space.

Given these theoretical points, and our empirical results in Section 5, we therefore conclude that using, for example, an exponential transform to try to solve the issues discussed, is not only technically incorrect, but also unnecessary and misleading. It is unnecessary because simpler formulations, less prone to practical difficulties, can give the same result and both the estimates and the estimation errors can be transformed as required. It is misleading because it suggests that there is an issue with the ratio calculation. Furthermore, it does not have any of the promised small sample size advantages put forward by the authors advocating its use.

In determining the most appropriate way to estimate the accuracy of WTP measures, it is useful to consider the reasons for requiring to know the accuracy.

- To test whether WTP is significantly different from zero, going beyond a test for the numerator alone, and thus taking into account positive or negative correlation between the estimates.
- The standard error of WTP may be needed to give a general indication of its accuracy, for comparison with other models and other studies, for example; in this case the Delta method gives a simple and suitable calculation.
- It may be required to indicate confidence limits for WTP, for example to give ranges for which a specific policy might be appropriate; in this case the Delta method can also be used, but it must be recognised that this gives only asymptotic results. When the estimation accuracy is low, alternative methods may be needed.

It is clear, and generally agreed, that when the relevant coefficients are estimated with reasonable accuracy, most of the methods proposed will give reasonable results in good agreement with the other methods[6]. Errors implying t ratios around 6 or 8 might be considered acceptable in this context[7]. In such cases, the Delta method can generally be used, as it is the simplest of the methods.

To enhance the applicability of the Delta method, we present a theorem (see Appendix 1), which shows that Delta calculations can be applied to 'robust' error measures derived from the 'sandwich' matrix, as well as to classical errors derived from the inverted Hessian matrix. These robust error measures have a wider applicability than classical errors, in terms of the assumptions needed.

When the parameters are less accurately estimated it is more difficult to choose an appropriate approach. Asymptotic approaches will be less appropriate, as the confidence limits will lie further from the optimum and the assumption of a quadratic log-likelihood function will be less satisfactory. In these circumstances it may be useful to use the likelihood ratio approach of Armstrong et al. (2001) or a resampling approach such as bootstrapping, as in Section 5. While these approaches require considerable calculation, they are not asymptotic. The Armstrong et al. (2001) approach depends only on the correctness of the likelihood function, while bootstrapping is even more general.

We should also note that in some cases the accuracy of estimation of WTP may be inadequate for the purpose. Even a t ratio of 4, which may well be thought to show that the WTP is

[6] However, methods involving sampling from a Normal distribution will ultimately return values close to or beyond zero and therefore cannot be recommended.

[7] It should be noted that when the coefficients are positively correlated, then the WTP may be more accurately estimated than either of the components in the ratio. The opposite could apply in the case of negative correlation.

significantly different from zero, implies two-sided 95% confidence limits of roughly ±50% and may well be inadequate as a basis for important investment or policy decisions.

For transport applications, these findings should help analysts to report the accuracy of their results more easily, since we show that the simple Delta method can be applied in most circumstances. The Delta method can also be applied to 'robust' errors.

Reporting the accuracy of results, in particular for the crucial estimates of the value of travel time, is an important contribution to professional transport policy analysis.

Acknowledgements

Stephane Hess was supported by the European Research Council through the consolidator grant 615596-DECISIONS and advanced Grant 101020940-SYNERGY. Juan de Dios Ortúzar wishes to acknowledge the support of Instituto Sistemas Complejos de Ingeniería (ANID PIA/BASAL AFB180003) and the BRT+ Centre of Excellence funded by the Volvo Research and Educational Foundations (www.brt.cl). We are grateful for the useful and insightful comments of three anonymous referees that helped us to improve the paper significantly.

References

AHCG (1996). The Value of Travel Time on UK Roads. Report to UK Department of Transport, Accent and Hague Consulting Group, London.

Armstrong, P.M., Garrido, R.A. & Ortúzar, J. de D. (2001) Confidence intervals to bound the value of time. *Transportation Research Part E: Logistics and Transportation Review* **37**, 143-161.

Axhausen, K.W., Hess, S., König, A., Abay, G., Bates, J.J. & Bierlaire, M. (2008). State of the art estimates of the Swiss value of travel time savings. *Transport Policy* **15**, 173-185.

Beesley, M.E. (1965). The value of time spent in travelling: some new evidence. *Economica* **32**, 174–185.

Berndt, E., Hall, B., Hall, R. & Hausman, J. (1974). Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement* **3**, 653–665.

Bessel, F.W. (1838). Untersuchungen uber die Wahrscheinlichkeit der Beobachtungsfehler. *Astronomische Nachrichten* **15**, 369-404 (in German).

Bliemer, M.C.J. & Rose, J.M. (2013). Confidence intervals of willingness-to-pay for random coefficient logit models. *Transportation Research Part B*: Methodological **58**, 199-214.

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

Carson, R.T. & Czajkowski, M. (2019). A new baseline model for estimating willingness-to-pay from discrete choice models. *Journal of Environmental Economics and Management* **95**, 57-61.

Cramer, J.S. (1986). *Econometric Applications of Maximum Likelihood Methods*. Cambridge University Press, Cambridge.

Daly, A. & Zachary, S. (1975). Commuters' values of time. *LGORU Report T55*, Local Government Operational Research Unit, Reading. Available at: (http://www.alogit.com/papers/CommutersValues.pdf.)

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

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

Fieller, E.C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society*, *Series B* **16**, 175-185.

Gaudry, M.J.I., Jara-Díaz, S.R. & Ortúzar, J. de D. (1989). Value of time sensitivity to model specification. *Transportation Research Part B: Methodological* **23**, 151-158.

Greene, W.H. (2008). *Econometric Analysis*. Pearson Education Inc, Upper Saddle River, NJ.

Hess, S., Bierlaire, M. & Polak, J.W. (2005). Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A: Policy and Practice* **39**, 221-236.

Hess, S. & Palma, D. (2019). Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling* **32**, 100170.

Huber, P.J. (1967). The behaviour of maximum likelihood estimates under nonstandard conditions. *Proceedings Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, 221–233.

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

Krinsky, I. & Robb, A. (1991). On approximating the statistical properties of elasticities: a correction. *Review of Economics and Statistics* **72**, 189–190.

Ortúzar, J. de D. & Willumsen, L.G. (2011). *Modelling Transport.* John Wiley & Sons, Chichester.

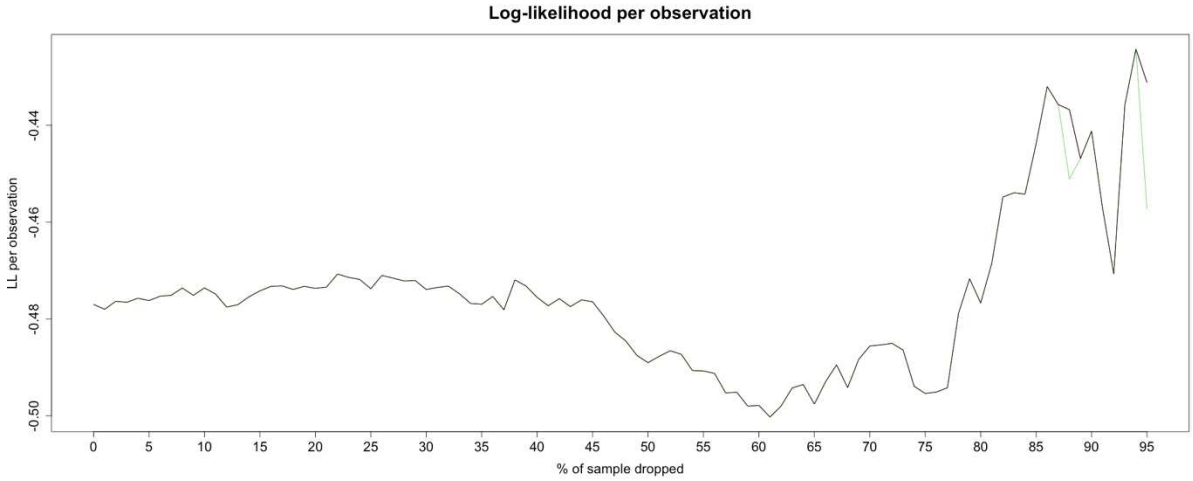
Palma, D., Ortúzar, J. de D., Rizzi, L.I., Guevara, C.A., Casaubon, G. & Ma, H. (2016). Modelling choice when price is a cue for quality: a case study with Chinese consumers. *Journal of Choice Modelling* **19**, 24-39.

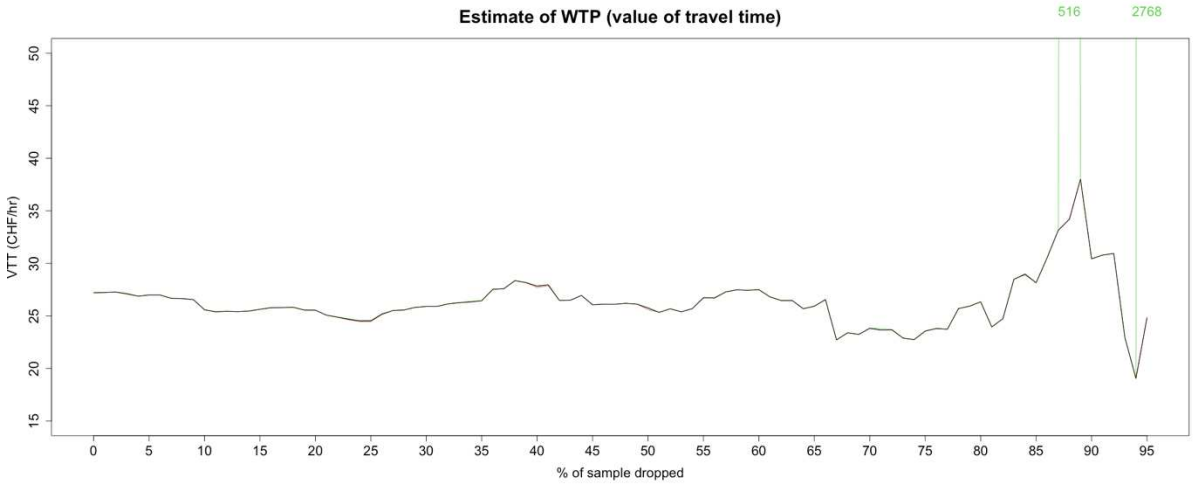
Shao, J. & Tu, D. (1995). *The Jackknife and Bootstrap.* Springer-Verlag, New York.

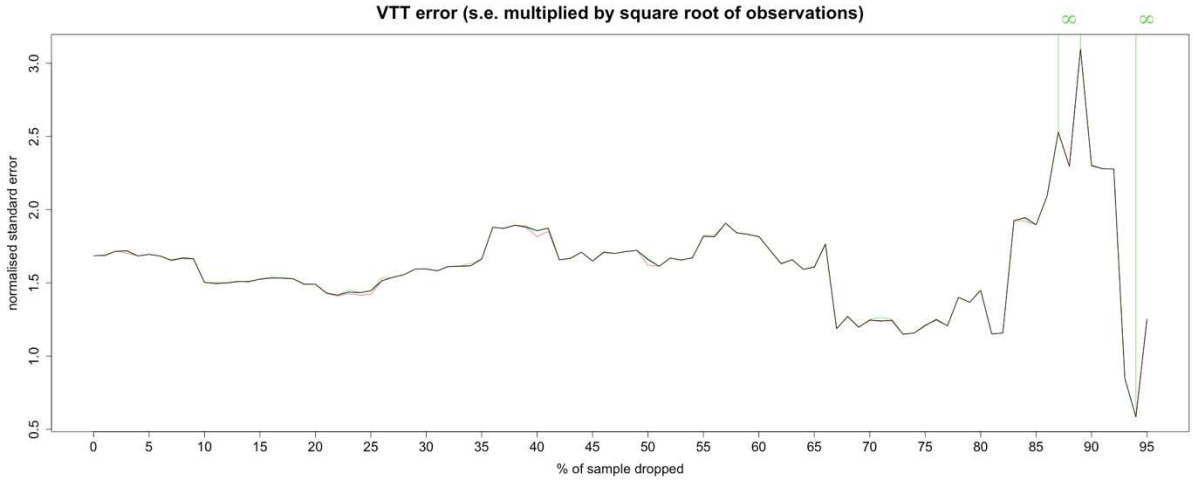
Sillano, M. & Ortúzar, J. de D. (2005). Willingness-to-pay estimation with mixed logit models: some new evidence. *Environment and Planning Part A: Economy and Space* **37**, 525-550.

Train, K.E. (2009). *Discrete Choice Models with Simulation.* Cambridge University Press, Cambridge.

Train, K. & Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space. In R. Scarpa & A. Alberini (eds), *Application of Simulation Methods in Environmental and Resource Economics*, 1-16. Springer Publisher, Dordrecht.


Log-likelihood per observation
LL per observation
-0.44
-0.46
-0.48
-0.50
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
% of sample dropped


Estimate of WTP (value of travel time)
516
2768
VTT (CHF/hr)
50 45 40 35 30 25 20 15
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
% of sample dropped


VTT error (s.e. multiplied by square root of observations)
∞
∞
normalised standard error
3.0 2.5 2.0 1.5 1.0 0.5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
% of sample dropped

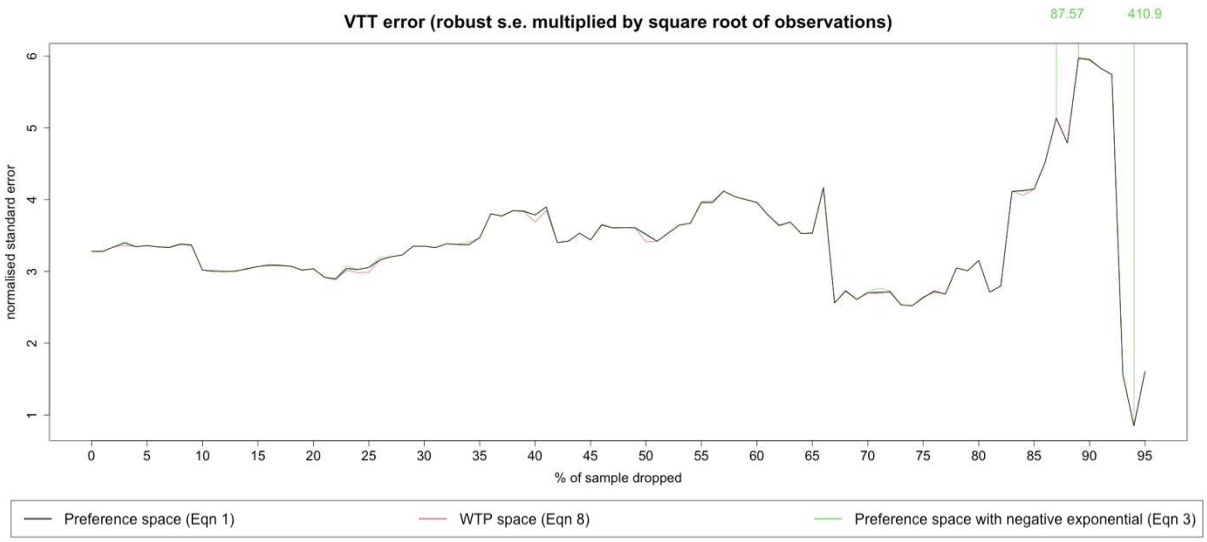


Figure 1: Small sample properties

Appendix 1

The theorems given in this Appendix justify the use of the Delta method with both classical and robust errors.

Classical Theorem

Let $\hat{\beta}$ be a correctly-specified maximum likelihood estimator of a vector with true value β^* of dimension L and let Ω be the covariance matrix of $\hat{\beta}$ around β^*, given by the inverse of the negative of the Hessian. Let $\Phi: R_L \rightarrow R_L$ be a differentiable and invertible function. Then:

1. $\Phi^* = \Phi(\beta^*)$ is the true value of $\Phi(\beta^*)$;
2. $\widehat{\Phi} = \Phi(\hat{\beta})$ is a maximum likelihood estimator of Φ^*; and
3. the covariance matrix of $\widehat{\Phi}$ around Φ^* attains the Cramér-Rao lower bound and is given by $cov(\Phi) = \Psi = \Phi'^T \Omega \Phi'$, where Φ' is the first derivative matrix of Φ.

Proof: outlined in Daly et al. (2012a) and given more completely in Cramer (1986).

The theorem above applies only when the model is correctly specified. However, provided the likelihood scores are zero at the true values of the parameters, we can use an alternative error estimate, usually termed the 'robust' error matrix (for a discussion see, e.g., Train, 2009). Using the same notation as in the Classical Theorem, the robust matrix is given by the 'sandwich' (sw) formula

$$cov^{(sw)}(\widehat{\beta}) = \Omega^{(sw)} = \Omega\, B\, \Omega, \qquad [A1]$$

where B is the 'BHHH' matrix (Berndt et al., 1974), given by $b_{jk} = \sum_n b_{jk}^n = \sum_n \frac{\partial LL^n}{\partial \beta_j} \frac{\partial LL^n}{\partial \beta_k}$, that is, the sample covariance of the scores. We use the notation of lower-case letters to indicate elements of vectors or matrices indicated by the corresponding upper-case symbols.

The following theorem allows us to apply the Delta method to the robust matrix.

Robust Theorem

In the context defined above for the Classical Theorem, $\Phi'^{T}\Omega^{(sw)}\,\Phi'$ is the robust error matrix $cov^{(sw)}(\widehat{\Phi})$ for maximum likelihood estimation over Φ.

Proof: The robust error matrix with respect to Φ is given by

$$\Psi^{(sw)} = XMX \qquad [A2]$$

where X and M are respectively the classical error matrix and the BHHH matrix, in each case defined with respect to Φ. From the classical theorem we know that the transformation of the classical matrix from β to Φ gives $X = \Psi$ as defined above. The required BHHH matrix M is given by the sample covariance of the scores with respect to Φ

$$M = [m_{jk}] = \left[\sum_n m_{jk}^n\right] = \left[\sum_n \frac{\partial L^n}{\partial \phi_j}\frac{\partial L^n}{\partial \phi_k}\right], \qquad [A3]$$

using the notation that $[x_{jk}]$ is the matrix whose jk^{th} element is x_{jk} and n to index the sample. Now, by the chain rule,

$$\frac{\partial L^n}{\partial \beta_r} = \sum_i \frac{\partial L^n}{\partial \phi_i}.\phi'_{ri} \qquad [A4]$$

Hence, the β-based BHHH matrix can be expressed in terms of ϕ derivatives:

$$b_{rs}^n = \frac{\partial L^n}{\partial \beta_r}\frac{\partial L^n}{\partial \beta_s} = \left(\sum_i \frac{\partial L^n}{\partial \phi_i}.\phi'_{ri}\right)\left(\sum_i \frac{\partial L^n}{\partial \phi_i}.\phi'_{si}\right) = \sum_j \sum_k \phi'_{rj}\left(\frac{\partial L^n}{\partial \phi_j}\frac{\partial L^n}{\partial \phi_k}\right)\phi'_{sk} \qquad [A5]$$

Noting that Φ does not vary with n, terms can be summed and rearranged

$$b_{rs} = \sum_n b_{rs}^n = \sum_n \sum_j \sum_k \phi'_{rj}\left(\frac{\partial L^n}{\partial \phi_j}\frac{\partial L^n}{\partial \phi_k}\right)\phi'_{sk} = \sum_j \sum_k \phi'_{rj} m_{jk}\,\phi'_{sk} \qquad [A6]$$

B can therefore be expressed as a matrix multiplication[8]

$$B = \Phi'^{T} M\,\Phi', \text{ i.e. } M = (\Phi'^{T})^{-1} B (\Phi')^{-1} \qquad [A7]$$

So, we can now write the required matrix

$$\Psi^{(sw)} = XMX = \Psi\left(\Phi'^{T}\right)^{-1} B\left(\Phi'\right)^{-1}\Psi = \Phi'^{T}\Omega\Phi'\left(\Phi'\right)^{-1} B\left(\Phi'^{T}\right)^{-1}\Phi'^{T}\Omega\Phi'$$

$$= \Phi'^{T}\Omega B\Omega\Phi' = \Phi'^{T}\Omega^{(sw)}\Phi' \qquad [A8]$$

which is the Φ transformation of the original robust error matrix as required.

[8] Note that in the matrix multiplication $A = B^T CD$, elements of A are given by $a_{ij} = \sum_k \sum_l b_{ki} c_{kl} d_{lj}$.

Appendix 2

To illustrate the use of the Delta method in the case of random coefficient models, let us look at the simple case of using two independent negative Lognormal distributions, with:

$$\log(-\beta_x) \sim N\left(\mu_{log-\beta_x}, \sigma_{log-\beta_x}\right) \ \& \ \log(-\beta_c) \sim N\left(\mu_{log-\beta_c}, \sigma_{log-\beta_c}\right), \qquad [A9]$$

that is, the logarithms of the negatives of the two coefficients follow a Normal distribution.

The ratio of two Lognormal distributions is itself a Lognormal distribution, such that:

$$\log(WTP) \sim N\left(\mu_{logWTP}, \sigma_{logWTP}\right), \qquad [A10]$$

with $\mu_{logWTP} = \mu_{log-\beta_x} - \mu_{log-\beta_c}$ and $\sigma_{logWTP=}\sqrt{\sigma^2_{log-\beta_x} + \sigma^2_{log-\beta_c}}$. With maximum likelihood estimates $\hat{\mu}_{log-\beta_x}$, $\hat{\sigma}_{log-\beta_x}$, $\hat{\mu}_{log-\beta_c}$ and $\hat{\sigma}_{log-\beta_c}$, we then now that $\hat{\mu}_{logWTP}$ and $\hat{\sigma}_{logWTP}$ have these same properties, and using the Delta method, we have that:

$$s.e.\left(\hat{\mu}_{logWTP}\right) = \sqrt{var(\hat{\mu}_{log-\beta_x}) + var(\hat{\mu}_{log-\beta_c}) - 2cov(\hat{\mu}_{log-\beta_x}, \hat{\mu}_{log-\beta_c})} \qquad [A11]$$

and

$$s.e.(\hat{\sigma}_{logWTP}) = \sqrt{\frac{\hat{\sigma}_{log-\beta_x}{}^2}{\hat{\sigma}_{log-\beta_x}{}^2+\hat{\sigma}_{log-\beta_c}{}^2} var(\hat{\sigma}_{log-\beta_x}) + \frac{\hat{\sigma}_{log-\beta_c}{}^2}{\hat{\sigma}_{log-\beta_x}{}^2+\hat{\sigma}_{log-\beta_c}{}^2} var(\hat{\sigma}_{log-\beta_c}) + 2\frac{\hat{\sigma}_{log-\beta_x}\cdot\hat{\sigma}_{log-\beta_c}}{\hat{\sigma}_{log-\beta_x}+\hat{\sigma}_{log-\beta_c}} cov(\hat{\sigma}_{log-\beta_x}, \hat{\sigma}_{log-\beta_c})} \qquad [A12]$$

where var and cov relate to the variance (i.e. square of the standard error) and covariance of the maximum likelihood estimates, rather than to the moments of the population level distribution.