

Calculating errors for measures derived from choice modelling estimates

Andrew Daly* Stephane Hess† Gerard de Jong‡

July 31, 2011

Abstract

The calibration of choice models produces a set of parameter estimates and an associated covariance matrix, usually based on maximum likelihood estimation. However, in many cases, the values of interest to analysts are in fact functions of these parameters rather than the parameters themselves. It is thus also crucial to have a measure of variance for these derived quantities and it is preferable that this can be guaranteed to have the maximum likelihood properties, such as minimum variance. While the calculation of standard errors using the Delta method has been described for a number of such measures in the literature, including the ratio of two parameters, these results are often seen to be approximate calculations and do not claim maximum likelihood properties. In this paper, we show that many measures commonly used in transport studies and elsewhere are themselves maximum likelihood estimates and that the standard errors are thus exact, a point we illustrate for a substantial number of commonly used functions. We also discuss less appropriate methods, notably highlighting the issues with using simulation for obtaining the variance of a function of estimates.

Keywords: discrete choice models; standard errors; parameter significance; delta method; maximum likelihood

1 Introduction

The use of discrete choice models entails the estimation of values for the various parameters used in the model specification, often on the basis of the maximum

*Institute for Transport Studies, University of Leeds, and RAND Europe, daly@rand.org

†Institute for Transport Studies, University of Leeds, s.hess@its.leeds.ac.uk, Tel: +44 (0)113 34 36611

‡Institute for Transport Studies, University of Leeds, Significance and CTS Stockholm, G.C.Dejong@its.leeds.ac.uk

likelihood criterion. This criterion has the advantages that it yields minimum-variance, asymptotically unbiased and asymptotically multivariate-normal estimates and also gives asymptotic estimates of the errors associated with those estimates. These error estimates allow analysts to assess the success of their estimation, using techniques such as t -ratios or (for non-linear models) asymptotic t -ratios.

Independently of the model structure and utility specification used in a discrete choice analysis, the final estimates thus consist of a vector of parameters β and a covariance matrix Ω . However, in many cases, the values of interest to analysts are in fact not the elements of the vector β itself, but other model outputs based on these estimates.

The calculation of these derived measures as functions of the estimated parameters β is well documented in the existing literature. However, by being functions of β , these measures are also subject to the errors associated with the estimation of the parameters. It is important therefore to be able to assess the error associated with statistics derived from the estimated parameters. This has received relatively little attention, and many studies still report derived measures without the associated standard errors.

In many cases, the required statistics are also functions of the data, which may well have measurement error. Calculating the error in statistics caused by data error requires assumptions concerning the data error distribution, which is of a different nature and may be substantially greater than the error caused by estimation error. Consideration of both types of error has been discussed in a small literature; for a review in the context of travel demand see [de Jong et al. \(2007\)](#). Data error is beyond the scope of this paper, which focusses on estimation error.

Where analysts have made the extra effort to produce standard errors for functions of their parameter estimates, they have often relied on two different approaches, namely simulation of the standard errors, or calculation on the basis of the so-called Delta method. While its use may be attractive in the case of complex functions, simulation of the standard errors can in fact be seriously misleading, as highlighted once more in numerical examples in the present paper. The Delta method on the other hand is typically regarded as providing an *approximation* to the true standard errors, the quality of which is not usually discussed. The approach of [Armstrong et al. \(2001\)](#) is not widely used, because of its computational complexity, but is a valid alternative.

In this paper, we refer to work by [Cramer \(1986\)](#) to show that the estimates given by many functions of model parameters estimated by maximum likelihood are in fact themselves maximum likelihood estimates and the standard errors obtained using the Delta method are correct estimates of the accuracy of these

functions with full maximum likelihood properties.

The remainder of this paper is organised as follows. The following section discusses the nature of standard errors calculated by the Delta method. This is followed by a discussion of a number of regularly used measures in Section 3, showing in each case that the calculated standard errors are indeed exact rather than approximations. Section 4 presents an empirical example for the ratio of coefficients that contrasts the Delta method with simple simulation of the errors; it is indicated that the variance of the simulation estimator does not exist, in line with discussions in [Daly et al. \(2011\)](#). Finally, Section 5 summarises the findings of the paper.

2 General methodology

This section begins by recalling the basic properties of parameters estimated by maximum likelihood methods and the error measures associated with them. It goes on to discuss the application of functions to obtain derived measures from parameter measures and the Delta method for obtaining estimates of error in derived measures. The second subsection discusses the estimation of errors by sampling, while the third subsection presents the main theorem of the paper, which shows that the transformed estimates are themselves maximum likelihood estimates under quite general conditions and that the Delta method errors are the Cramér-Rao lower bound of errors. Finally, a general method is described for developing error estimates for new parameter transformations in a multivariate context.

2.1 Parameter estimates and error estimates

It is imagined that our choice model is estimated from a large number N of observations of revealed or stated preferences of individual decision makers¹. We further assume that this model contains a number of unknown parameters which are estimated using the maximum likelihood criterion. Because of this context it is possible in fairly general terms to state the key properties of the parameter estimates.

The classical result that is widely used in this context is that, provided reasonable conditions are met and the model is correctly specified, then the expected score (first derivative of the likelihood function with respect to the model parameters) is zero and the maximum likelihood estimates β of the model parameters

¹Because we have assumed that there is a large number of observations, we shall not investigate small-sample properties of the estimations.

converge, as N increases, to a normal distribution around the true values β^* :

$$\sqrt{N}(\beta - \beta^*) \rightarrow n(0, \Omega) \tag{1}$$

where Ω is a symmetric positive definite matrix.

If the model is correctly specified and the optimum is well defined, then Ω is minus the inverse of the Hessian (matrix of second derivatives) of the likelihood function with respect to the model parameters and forms the Cramér-Rao lower bound, so these estimates are minimum variance.

If the model is not correctly specified but the expected score of the likelihood function is zero at β^* , or when it is inconvenient to calculate the true second derivative matrix, other matrices can be substituted, as indicated in the textbooks (e.g. [Train, 2009](#)). However, in these cases we no longer have maximum likelihood estimates, although the Delta method can still be used as an approximation.

Then let β be a correctly-specified maximum likelihood estimator of a vector β^* of dimension L and let Ω be the covariance matrix of β around β^* . Let $\Phi = \Phi(\beta): \mathbf{R}_L \rightarrow \mathbf{R}_L$ be a differentiable function. The interest now lies in computing standard errors for Φ .

The first point is that if β^* is the true value of β , then $\Phi^* = \Phi(\beta^*)$ is the true value of Φ . This depends only on Φ being an ordinary single-valued function (e.g. not a square root, which would leave Φ^* being defined ambiguously).

The Slutsky Theorem states that continuous functions of consistent estimators are consistent estimators of the functions. That is, making calculations of functions of model parameters gives results that have at least the reasonable property of consistency. For the Slutsky theorem, β does not have to be a maximum likelihood estimator, but if it is, and under certain other conditions, these results can have substantially more status and correspondingly better properties.

It is shown by simple calculus in statistical textbooks that a first-order approximation to the error in Φ induced by the error in β is given by

$$cov(\Phi) = \Phi'{}^T \Omega \Phi' \tag{2}$$

where Φ' is the vector first derivative of the function Φ with respect to β and Ω is the covariance matrix of the estimates of β . Moreover, [Greene \(2008, pp. 1055-1056\)](#) indicates that if β is asymptotically normally distributed, and Φ is continuous (and of course differentiable, though this is not stated by [Greene 2008](#)) then Φ is also asymptotically normally distributed.

This approximation, widely known as the Delta Method, can be used in a wide range of circumstances to estimate the error in functions of estimated parameters. It may be noted that the first-derivative calculation does not depend on the method that was used to derive the parameter estimates - we need only consistency.

2.2 Estimation of errors by sampling

In the context of maximum likelihood estimates of β , Equation 2 can be seen as a two-stage calculation of the variance of the function Φ . First, the matrix Ω is an asymptotic approximation to the true covariance matrix of the estimates β . Second, the distribution of Φ has commonly been seen as some complicated function derived from the asymptotic normality of the distribution of β . In the case of the estimate of the ratio of two parameters, we might naturally call on literature which describes the distribution of the ratio of two normally distributed random variables: skewed, with complications arising when the denominator gets close to 0.

For example, [Armstrong et al. \(2001\)](#) present an approach to calculating asymptotic t -ratios for parameter ratios based on the asymptotic normality of the numerator and denominator. They also give an alternative approach, calculating likelihood ratios relative to restricted versions of the model, an approach with greater sophistication but not always feasible and, as they concede, often “tedious” in practice.

The approach of making calculations based on the assumption that the numerator and denominator of the parameter ratio are normally distributed is also widespread in other literatures such as health economics (see e.g. [Hole, 2007](#)). The difficulties of this rigid approach, requiring an exact normal distribution of the estimated parameters (which is only an approximation) and therefore losing sight of the fact that their ratio is equally justifiably normally distributed (see the [Greene 2008](#) result mentioned above) leads to a range of complicated and unsatisfactory approaches, such as sampling from the ‘normal’ distributions of the original parameters and using the samples as arguments for the function, an approach sometimes attributed to [Krinsky and Robb \(1986, 1991\)](#), in cases where these complications are not necessary and sometimes incorrect.

The Krinsky-Robb papers appear to have been misinterpreted. First, it is important to read the correction paper ([Krinsky and Robb, 1991](#)), which finds that the Delta method and the sampling approach give very similar results, always, they find, less than 5% different, in contradiction to the first paper where results were distorted by a computer error. Second, in both papers Krinsky and Robb considered random parameters which were estimated by linear models, and could therefore be considered as exactly normally distributed, while the elasticity function of which they were arguments did not involve a ratio. Their results should not be interpreted as showing that the Delta and sampling approaches give very different results, nor should they be interpreted as giving information about ratio estimators.

2.3 Status of transformed estimates

A deeper understanding of Equation 2 can be obtained along the lines set out by Cramer (1986, Section 3.1). In Section 2.1 we established that providing Φ is differentiable, then the distribution of Φ converges asymptotically to a normal distribution around the true value Φ^* :

$$\sqrt{N}(\Phi - \Phi^*) \rightarrow n(0, \Phi'^T \Omega \Phi') \quad (3)$$

where Ω is the covariance matrix of β , i.e. Φ is asymptotically equivalent to an MLE of Φ^* , as it has the same asymptotic distribution.

It may be tempting to claim that Φ can be regarded as an MLE with no further ado, that it represents the maximisation of likelihood over a space induced by the transformation Φ . Cramer (1986), however, prefers the more widely accepted view, which is to consider the reparametrisation of the model by an invertible vector function to obtain a vector η with the same dimension as β :

$$\eta = g(\beta) \text{ and } \beta = g^{-1}(\eta) \quad (4)$$

For the transformation g to be invertible it must be one-to-one², as well as differentiable. While these conditions may be restrictive from a mathematical point of view, in practice many important functions can be shown to have the required properties. With these conditions, Cramer shows that the properties of the dependent variable are not affected by the transformation, so that $\eta = g(\beta)$ is an MLE of η^* . Cramer then goes on to derive the covariance of Φ around Φ^* as:

$$\text{cov}(\Phi) = g'^T \Omega g' \quad (5)$$

where g' is the derivative matrix (Jacobian) of g with respect to β . This is exactly the result from Equation 2, but now, because we know η to be an MLE, we also know that $\text{cov}(\Phi)$ is the Cramér-Rao lower bound of minimum variance for the estimator.

The approach of Cramer can be summarised in the following theorem.

Theorem: Let β be a correctly-specified maximum likelihood estimator of a vector β^* of dimension L and let Ω be the covariance matrix of β around β^* . Let $\Phi: \mathbf{R}_L \rightarrow \mathbf{R}_L$ be a differentiable and invertible function. Then:

1. $\Phi^* = \Phi(\beta^*)$ is the true value of $\Phi(\beta^*)$;
2. $\Phi = \Phi(\beta)$ is a maximum likelihood estimator of Φ^* ; and

²This also implies g and g^{-1} are non-singular, within the space of interest.

3. the covariance matrix of Φ around Φ^* attains the Cramér-Rao lower bound and is given by:

$$\text{cov}(\Phi) = \Phi'^T \Omega \Phi', \quad (6)$$

where Φ' is the first derivative matrix of Φ .

We are now in a position to reassess the results derived by the Delta Method. Instead of seeing this approach as being a general way to develop useful *approximations* for the error in functions of parameters, we can now see that the functions of parameters can themselves be interpreted as true maximum likelihood estimates, while the Delta Method is exactly what is required to obtain the *true* Cramér-Rao covariance of the transformed parameter estimates around the true values.

The Theorem gives us two principal benefits. First, it generalises a number of results that have been derived as approximate calculations. Second, it gives a new status to the transformed parameters $\Phi(\beta)$.

Consider the important example of estimating the ratio of two estimated parameters. By considering the parameters to be maximum likelihood estimators and therefore asymptotically normally distributed about the true value, a Taylor expansion can be used to approximate the distribution and thus derive the variance. Given the theorem above, however, it is not necessary to make approximate calculations based on Taylor expansions but a direct statement of the result can be given, based on the first derivatives. Furthermore, the result indicates that the parameter ratio is itself a maximum likelihood estimate and is therefore itself consistent, unbiased and normally distributed around the true ratio. The variance has the same value as given by the textbook formula, but the result shows that this is in fact the true Cramér-Rao lower bound, rather than an approximate value.

It may also be noted from [Cramer \(1986\)](#) that if the theorem does not hold because Φ is not invertible, then Φ is still asymptotically equivalent to a maximum likelihood estimator. The use of Φ is unlikely to lead to false conclusions.

2.4 A general approach

It is useful to note that Φ can be defined to transform as many or as few elements of β as convenient. In a later remark in Section 3.1, [Cramer \(1986\)](#) notes that “a single derived parameter can be regarded as part of a larger transformation”. That is, a transformation can be applied to a single parameter which is a component of a parameter vector and this transformed parameter can be viewed along with the untransformed parameters as being the maximum likelihood estimates.

Similar reasoning can be applied for any subset of the estimates. Of course, the transformations being applied must remain invertible and differentiable.

Thus, Φ in the above notation may contain a large number of functions of β while we may only be interested in a single function. Here, we could have that Φ_k gives a single function of a specific subset of parameters of β , say β_{Φ_k} with covariance matrix Ω_{Φ_k} . Providing that the function defined by changing β_k to this Φ_k , leaving all the other β values unchanged, satisfies the conditions of the theorem, in particular that it is invertible, then we can make the claims of the Theorem for Φ_k . In the remainder of this paper, we will assume that Φ relates to a single function of a subset of the parameters and will drop subscripts accordingly. We are then interested in the variances of an individual function Φ rather than the covariance matrix of a set of functions.

The transformations that satisfy the requirements of the Theorem are then those that convert a single variable to another single variable by a differentiable and invertible function. All of the functions we shall consider are differentiable and these functions are invertible if the derivative does not change sign or become zero, because they are continuous and strictly monotonic. That is, we can apply the Theorem if the derivative of Φ with respect to one β parameter does not change sign as β changes.

Before proceeding, we will introduce a formulation that facilitates the calculation when dealing with a single Φ and a large number of parameters β . As shown above, we have that $var(\Phi) = \Phi'^T \Omega \Phi'$. Here, Φ' has L elements, namely the derivative of Φ with respect to each of the L elements in β . Furthermore, Ω has L^2 elements, of which only $\frac{L^2+L}{2}$ are unique, since the matrix is symmetric. Denoting the individual elements in Φ' and Ω as ϕ'_l and ω_{lm} respectively, we have that:

$$var(\Phi) = \Phi'^T \Omega \Phi' = \sum_{l=1}^L \phi'_l \left(\sum_{m=1}^L \phi'_m \omega_{ml} \right), \quad (7)$$

where it can be further seen that this is equal to:

$$var(\Phi) = \sum_{l=1}^L \phi'^2_l \omega_{ll} + 2 \sum_{l=2}^L \sum_{m=1}^{l-1} \phi'_l \phi'_m \omega_{lm}, \quad (8)$$

where the final simplification is possible by imposing the restriction on the double summation.

3 Standard errors for commonly used measures

This section presents formulae for the variances of a number of measures commonly used in transport analysis and elsewhere, obtained using the approach described in the previous section. Specifically, we look at different specifications of $\Omega(\beta)$, where β is a vector grouping together all estimated parameters, with Ω being the covariance matrix of the vector β , with individual elements defined as ω_{kl} . Each time, we discuss how the obtained variances can be shown to have the same MLE properties as the coefficients from which they are obtained.

Table 1 summarises the calculations needed to obtain the variances for four basic but widely used measures, namely the sum and difference of two parameters, and the ratio and product of two parameters. We also show the variance for the reciprocal of β_1 , which implies that the t -ratio of the reciprocal of β_1 with respect to zero is exactly the same as the t -ratio of β_1 with respect to zero, namely $\frac{\beta_1}{\sqrt{\omega_{11}}}$, i.e. β_1 divided by its standard error. Alongside the variance for the product of two estimators, we also see that the variance of the square of an estimator is given by the variance of the estimator multiplied by four times the square of the estimator; the t -ratio for the square of a parameter is thus half the t -ratio for the parameter itself, a result that is useful when moving between standard deviations and variances for random coefficients. Finally, Table 1 also shows the variances for ratios of parameters (e.g. willingness-to-pay indicators, WTP) in the case of non-linear utility functions, in particular when a Box-Cox transform is used.

We now turn to the issue of invertibility, and by extension the maximum likelihood properties of these error measures. As can be seen from the fifth column in Table 1, the signs of the partial derivatives of Φ are either fixed or independent of the relevant parameter (i.e. the parameter which we differentiate against), with the exception of the square, i.e. $\Phi = \beta_1^2$. Here, if the sign of β is known, as will generally be the case, the derivative has a fixed sign. As a conclusion, we can state the standard errors obtained using the Delta method for the various measures in Table 1 are exact³.

We now move to a discussion of model-specific quantities of interest. Specifically, in advanced models, the behaviour is quantified by a number of parameters that impose a certain structure on the error term. This includes the correlation between error terms in Generalised Extreme Value (GEV) models (cf. [McFadden, 1978](#)) as well as the parameters used for the distributions of preferences in random coefficients models. However, the actual measures of interest are often again a function of these parameters rather than the parameters themselves, lead-

³A further case that deserves special attention is with the WTP in the presence of a Box-Cox transform, when one of the concerned attributes takes on a value of 1. In this case, the resulting WTP does not depend on the corresponding λ value.

Table 1: Standard errors for commonly used measures

Function	Φ	β^*	$\phi'_k = \frac{\partial \Omega}{\partial \beta^*}$	Sign of ϕ'_k	$var(\Phi)$	Limitations
Sum	$\beta_1 + \beta_2$	β_1	$\phi'_1 = 1$	positive	$\omega_{11} + \omega_{22} + 2\omega_{12}$	None
		β_2	$\phi'_2 = 1$	positive		
Difference	$\beta_1 - \beta_2$	β_1	$\phi'_1 = 1$	positive	$\omega_{11} + \omega_{22} - 2\omega_{12}$	None
		β_2	$\phi'_2 = -1$	negative		
Ratio	$\frac{\beta_1}{\beta_2}$	β_1	$\phi'_1 = \frac{1}{\beta_2}$	sign of β_2	$\left(\frac{\beta_1}{\beta_2}\right)^2 \left(\frac{\omega_{11}}{\beta_1^2} + \frac{\omega_{22}}{\beta_2^2} - 2\frac{\omega_{12}}{\beta_1\beta_2}\right)$	$\beta_2 \neq 0$
		β_2	$\phi'_2 = -\frac{\beta_1}{\beta_2^2}$	minus sign of β_1		
Inverse	$\frac{1}{\beta_1}$	β_1	$\phi'_1 = -\frac{1}{\beta_1^2}$	negative	$\frac{\omega_{11}}{\beta_1^4}$	$\beta_1 \neq 0$
		β_2	$\phi'_2 = \beta_2$	sign of β_2	$\beta_2^2\omega_{11} + \beta_1^2\omega_{22} + 2\beta_1\beta_2\omega_{12}$	None
Product	$\beta_1\beta_2$	β_1	$\phi'_1 = \beta_2$	sign of β_2		
		β_2	$\phi'_2 = \beta_1$	sign of β_1		
Square	β_1^2	β_1	$\phi'_1 = 2\beta_1$	sign of β_1		None
		β_2	$\phi'_2 = \frac{1}{\beta_2} \frac{x_1^{\lambda_1-1}}{x_2^{\lambda_2-1}}$	sign of β_2		
WTP with Box-Cox transform	$\frac{\beta_1}{\beta_2} \left(\frac{x_1^{\lambda_1-1}}{x_2^{\lambda_2-1}}\right)$	β_1	$\phi'_1 = \frac{1}{\beta_2} \frac{x_1^{\lambda_1-1}}{x_2^{\lambda_2-1}}$	sign of β_2	use Equation 8	$x_1 > 0, x_2 > 0$
		λ_1	$\phi'_2 = \frac{\beta_1}{\beta_2} \frac{x_1^{\lambda_1-1} \ln(x_1)}{x_2^{\lambda_2-1}}$	sign of $\frac{\beta_1 \ln(x_1)}{\beta_2}$		
		β_2	$\phi'_3 = -\frac{\beta_1}{\beta_2^2} \frac{x_1^{\lambda_1-1}}{x_2^{\lambda_2-1}}$	minus sign of β_1		
		λ_2	$\phi'_4 = -\frac{\beta_1}{\beta_2} \frac{x_1^{\lambda_1-1} x_2^{\lambda_2-1} \ln(x_2)}{x_2^{2\lambda_2-2}}$	minus sign of $\frac{\beta_1 \ln(x_2)}{\beta_2}$		

ing to a requirement for the calculation of appropriate error measures. Table 2 summarises the calculation of variances for a number of such measures.

We first look at the correlation between error terms in a two-level Nested Logit (NL) model as a function of the structural parameter λ , and an approximation to the correlation between the errors for two alternatives i and j in a two-level Cross-Nested Logit (CNL) model, with K nests, where $\alpha_{j,k}$ is the allocation parameter for alternative j and nest k (cf. Papola, 2004; Marzano and Papola, 2008). For both measures, the fifth column shows no problems with invertibility as the signs of the partial derivatives are fixed, hence ensuring that the standard errors obtained with the Delta method are exact.

Researchers and practitioners are increasingly relying on the Mixed Multinomial Logit (MMNL) model to represent random variations in the marginal utilities across respondents. From an interpretation point of view, we are interested in the moments of the estimated distribution, but, in contrast to the commonly used Normal distribution, these moments in other distributions do not generally equate to the estimated parameters. Table 2 shows the calculations for the variances for the means and standard deviations for two commonly used alternatives to the Normal, namely the Triangular distribution⁴ and the Lognormal distribution⁵. Corresponding variances for the estimated variance can be obtained with the help of the formula for squares in Table 1. Invertibility and hence exactness of the standard errors obtained with the Delta method is clearly ensured for all of these measures except for the standard deviation of the Triangular distribution, where problems arise for the mode, i.e. c . If, as is commonly the case, a symmetrical Triangular is used, we simply have that $c = \frac{a+b}{2}$, so that no problem arises. If an asymmetrical Triangular is used, invertibility is guaranteed (i.e. the variances are still exact) if we assume that the sign of the skewness is known.

While most MMNL applications still make use of independently distributed random coefficients, it has been recognised that it is important to also allow for correlation between the coefficients. In the rare cases where a multivariate distribution is actually used, this is almost exclusively based on a multivariate Normal (MVN) distribution. The estimation of a multivariate Normal produces estimates of the mean coefficient values along with the Cholesky matrix Λ , on the basis of which it is possible to calculate covariance and correlation measures. The outputs of interest in this context are the lower triangular Cholesky matrix

⁴With a , b and c giving the lower boundary, upper boundary and mode of the Triangular distribution.

⁵With the mean and standard deviation of the underlying Normal distribution being given by μ_N and σ_N .

Table 2: Standard errors for model specific measures

Function	Φ	β^*	$\phi'_k = \frac{\partial \Omega}{\partial \beta^k}$	Sign of ϕ'_k	$var(\Phi)$	Limitations
NL correlation	$1 - \lambda^2$	λ	$\phi'_1 = -2\lambda$	negative	$4\lambda^2 \omega^2$	$0 < \lambda \leq 1$
CNL correlation	$\sum_{k=1}^K \frac{1}{\alpha_{i,k}} \alpha_{j,k}^{\frac{1}{2}} (1 - \lambda_k^2)$	$\lambda_k, k = 1, \dots, K$ $\alpha_{i,k}, k = 1, \dots, K$ $\alpha_{j,k}, k = 1, \dots, K$	$\phi'_1 = -2\lambda_i \alpha_{i,l}^{\frac{1}{2}} \alpha_{j,l}^{\frac{1}{2}}, 1 \leq l \leq K$ $\phi'_2 = \frac{1}{2} \alpha_{i,l}^{\frac{1}{2}} \alpha_{j,l}^{\frac{1}{2}} (1 - \lambda_l^2), K < l \leq 2K$ $\phi'_3 = \frac{1}{2} \alpha_{j,l}^{\frac{1}{2}} \alpha_{i,l}^{\frac{1}{2}} (1 - \lambda_l^2), 2K < l \leq 3K$	negative positive positive	use Eq. 8	$0 \leq \lambda_k \leq 1, \forall k$ $0 \leq \alpha_{i,k} \leq 1, \forall k, \sum_k \alpha_{i,k}$ $0 \leq \alpha_{j,k} \leq 1, \forall k, \sum_k \alpha_{j,k}$
Mean of Triangular	$\mu_T = \frac{a+b+c}{3}$	a b c	$\phi'_1 = \frac{1}{3}$ $\phi'_2 = \frac{1}{3}$ $\phi'_3 = \frac{1}{3}$	positive positive positive	$\frac{\omega_{11} + \omega_{22} + \omega_{33} + 2\omega_{12} + 2\omega_{13} + 2\omega_{23}}{9}$	$a \leq c \leq b$
SD of Triangular	$\sigma_T = \sqrt{\frac{a^2 + b^2 + c^2 - ab - ac - bc}{18}}$	a b c	$\phi'_1 = \frac{2a-b-c}{18\sigma_T}$ $\phi'_2 = \frac{2b-a-c}{18\sigma_T}$ $\phi'_3 = \frac{2c-a-b}{18\sigma_T}$	negative positive sign of $2c - a - b$	$\frac{(2a-b-c)((2a-b-c)\omega_{11} + (2b-a-c)\omega_{12} + (2c-a-b)\omega_{13})}{18^2 \sigma_T^2}$ $+$ $\frac{(2b-a-c)((2b-a-c)\omega_{22} + (2c-a-b)\omega_{23})}{18^2 \sigma_T^2}$ $+$ $\frac{(2c-a-b)^2 \omega_{33}}{18^2 \sigma_T^2}$	$a \leq c \leq b$
Mean of Lognormal	$\mu_{LN} = e^{\mu_N + \frac{\sigma_N^2}{2}}$	μ_N σ_N	$\phi'_1 = e^{\mu_N + \frac{\sigma_N^2}{2}} = \mu_{LN}$ $\phi'_2 = \sigma_N e^{\mu_N + \frac{\sigma_N^2}{2}} = \sigma_N \mu_{LN}$	positive (μ_{LN}) positive (μ_{LN})	$\mu_{LN}^2 (\omega_{11} + \sigma_N^2 \omega_{22} + 2\omega_{12} \sigma_N)$	$\sigma_N > 0$
SD of Lognormal	$\sigma_{LN} = \sqrt{e^{2\mu_N + 2\sigma_N^2} - e^{2\mu_N + \sigma_N^2}}$	μ_N σ_N	$\phi'_1 = \frac{2e^{2\mu_N + 2\sigma_N^2} - 2e^{2\mu_N + \sigma_N^2}}{2\sqrt{e^{2\mu_N + 2\sigma_N^2} - e^{2\mu_N + \sigma_N^2}}} = \sigma_{LN}$ $\phi'_2 = \sigma_N \left(\frac{e^{2\mu_N + 2\sigma_N^2}}{\sigma_{LN}} + \sigma_{LN} \right)$	positive positive	$\sigma_{LN}^2 \omega_{11} + 2\sigma_N (e^{2\mu_N + 2\sigma_N^2} + \sigma_{LN}^2) \omega_{12}$ $+$ $\sigma_N^2 \left(\frac{e^{2\mu_N + 2\sigma_N^2}}{\sigma_{LN}} + \sigma_{LN} \right)^2 \omega_{22}$	$\sigma_N \geq 0$
Variance in MVN	$\Phi_A(a) = \sum_{k=1}^a (s_{a,k})^2$	$s_{a,k}, a \geq k$	$\phi'_{A,a,k} = 2s_{a,k}$	see text	use Eq. 8	$a \geq k$
Covariance in MVN	$\Phi_B(a, b) = \sum_{k=1}^a s_{a,k} s_{b,k}$	$s_{a,k}, a \geq k$ $s_{b,k}, b \geq k$	$\phi'_{B,a,k} = s_{b,k}$ $\phi'_{B,b,k} = s_{a,k}$	see text	use Eq. 8	$k \leq a < b$
Correlation in MVN	$\Phi_C(a, b) = \frac{\Phi_B(a, b)}{\sqrt{\Phi_A(a)\Phi_A(b)}}$	$s_{j,k}, j \geq k$	$\phi'_{C,j,k} = \frac{\phi_{B,j,k}(a, b)}{\sqrt{\Phi_A(a)\Phi_A(b)}}$ $\Phi_B(a, b) (\phi'_{A,j,k}(a) \Phi_A(b) + \phi'_{A,j,k}(b) \Phi_A(a))$ $2(\Phi_A(a)\Phi_A(b))^{\frac{3}{2}}$	see text	use Eq. 8	$j \geq k, a < b$

Λ , given by:

$$\Lambda = \begin{pmatrix} s_{1,1} & 0 & 0 & \dots & 0 \\ s_{2,1} & s_{2,2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{K,1} & s_{K,2} & s_{K,3} & \dots & s_{K,K} \end{pmatrix}, \quad (9)$$

with corresponding covariance matrix of the estimates Ω_Λ where this is a subset of the covariance matrix for all estimated parameters Ω , and where Λ can be written in vector form as $\Lambda = (s_{1,1}, s_{2,1}, s_{2,2}, \dots, s_{K,K})$. We have that the covariance matrix of the multivariate Normal distribution is now given by $\Lambda^T \Lambda$. Table 2 shows the calculations required to obtain variances for three main measures that can be obtained from his, namely the variance of individual coefficients and the covariance and correlation between pairs of coefficients. For the actual calculation of the standard errors, it can be seen that Λ has $\sum_{k=1}^K k$ elements, with the same being the case for $\phi'_{A,j,k}(a)$, $\phi'_{B,j,k}(a, b)$, and $\phi'_{C,j,k}(a, b)$. The covariance matrix of Λ , Ω_Λ has $\left(\sum_{k=1}^K k\right)^2$ elements ($\sum_{k=1}^K k$ by $\sum_{k=1}^K k$), and the variances of the measures can be obtained straightforwardly by applying Equation 8 to the partial derivatives.

We know that invertibility is ensured for the covariance in multivariate Normals as the Cholesky decomposition is unique, so that as a result, the standard errors for variances and covariances obtained with the Delta method are exact. By the same reasoning, the correlation transformation is also invertible, given the standard errors.

4 Empirical example

As a final step, we present a simple comparison between the Delta method and simulation, with a view to highlighting the inadequacy of simulation in this context. Here, we limit ourselves to the most basic meaningful example, namely the ratio between two coefficients, β and γ . Extensions to more complex situations are straightforward. Our example uses values of -0.05 for β and -0.1 for γ , where we use variances of 0.0001 and 0.0009 respectively, with a covariance of -0.0001 . This gives t -ratios of -5 and -3.33 for β and γ , with a correlation of -0.33 between the two coefficients. We thus have that $\frac{\beta}{\gamma} = 0.5$ and $\frac{\gamma}{\beta} = 2$. Using the appropriate formulae from Table 1, we obtain standard errors for these two ratios of 0.2062 and 0.8246 respectively, giving t -ratios of 2.43 for both $\frac{\beta}{\gamma}$ and $\frac{\gamma}{\beta}$; from Table 1, it can be seen that the t -ratio against zero for the inverse of a function is the same as that for the function itself.

Table 3: Empirical example: standard errors for a ratio of two coefficients, statistics over 50 runs, each with 10^7 draws

	β/γ	γ/β
<i>true</i> s.e.	0.2062	0.8246
<i>true</i> t -ratio	2.43	2.43
min simulated s.e.	11.07	0.96
mean simulated s.e.	88.56	1.24
max simulated s.e.	831.96	5.26
min simulated t -ratio	0.0006	0.3805
mean simulated t -ratio	0.0136	1.8382
max simulated t -ratio	0.0452	2.0815

As an alternative to using the exact formulae, some analysts may rely on simulation, especially in the case of complex measures where formulae have not been discussed in detail prior to this paper. To illustrate the shortcomings of this approach, simulation was used to try to produce standard errors for $\frac{\beta}{\gamma}$ and $\frac{\gamma}{\beta}$ in this example. In fact, for a ratio of random normal variables, the variance does not exist, as is shown by [Daly et al. \(2011\)](#). Although that paper deals with the different context of distribution in the population rather than errors, the mathematics of the ratio of two normally distributed variables is the same, and the literature reviewed and results of the paper show that none of the moments of these ratios exist. However, the simulation approach is sometimes used in practice and it is illuminating to try an extensive simulation to show how that works out. Specifically, we used fifty runs with 10^7 (ten million) draws for β and γ ⁶.

The results of this application are summarised in [Table 3](#), giving statistics for the simulated standard errors and the resulting t -ratios for the ratios $\frac{\beta}{\gamma}$ and $\frac{\gamma}{\beta}$ ⁷. Several observations can be made. Firstly, there is a lack of stability across the fifty runs, again highlighting an issue with simulation which depends on the draws used in a specific run, even when making use of 10^7 draws! We can further see that the simulation overestimates the true standard error for $\frac{\beta}{\gamma}$ by over 5,000% in the *best* run, and by over 400,000% in the *worst* run. For $\frac{\gamma}{\beta}$, the problems are slightly less pronounced, which is a result of a higher t -ratio for the denominator,

⁶With appropriate consideration of the covariance.

⁷These t -ratios are based on the estimates rather than simulated values for the ratios $\frac{\beta}{\gamma}$ and $\frac{\gamma}{\beta}$, along with the simulated standard errors, as would be common practice when simulating standard errors. The simulated mean values would also be biased, unlike the ratio of the best-estimates, which is a maximum likelihood estimator and therefore unbiased.

reducing the chances of values close to zero in the denominator for the second ratio. But even here, the *best* run overestimates the standard error by over 15%, while the *worst* run overestimates it by over 500%. With these problems, it should come as no surprise that the t -ratios are also severely biased, and are not the same for the two ratios as they should be.

It is important to be clear that the simulation results for this ratio are always incorrect. The discussions in [Daly et al. \(2011\)](#) show that all simulations of this ratio give undefined results. In this sense, the simulations of $\frac{\beta}{\gamma}$ are less concerning, as the risk of accepting them as valid is small. For $\frac{\gamma}{\beta}$, however, there is a clear danger that results in this range could be accepted as valid.

Simulation can be used, of course, in cases when the result that is required actually exists. For a ratio, percentiles may be used. However, in most simple cases it is easier and, as we have shown, more exact, to use the Delta calculation. In more complicated situations (cf. [de Jong et al., 2007](#)), simulation is still needed.

5 Discussion and conclusions

5.1 Discussion: t ratios

We close the paper by looking at the consequences that the discussions above have for our interpretation of estimates and their significance. In particular the interpretation of t -ratios and confidence limits requires a little thought.

Earlier, we showed that the t -ratio of the estimate of a reciprocal of an MLE parameter was equal to the t -ratio of the parameter itself. We also showed that the estimate of the reciprocal has just as much status as the initial estimate. Does this mean that a test that a parameter is significantly different from 0 is exactly the same as a test that it is significantly different from infinity? Moreover, how can it be that a parameter estimate and an estimate of the reciprocal of the parameter are both distributed asymptotically normal?

The information on the likelihood function comes out of the estimation process. At the optimum, we know that the first derivative of the function is zero and we have an estimate (from one or other matrix) of the second derivatives. All of the usual information on errors, t -ratios and confidence limits comes from this matrix of second derivatives.

By assuming that the likelihood function is quadratic, i.e. that the second derivative at the optimum applies everywhere, we can obtain an approximate view of how the value of the function declines away from the optimum. It is very rare to make systematic tests of the true value of the likelihood function at points away from the optimum; [Daly and Zachary \(1975\)](#) and [Armstrong et al. \(2001\)](#)

are exceptions. It is therefore not surprising that we find paradoxes such as the t -ratio of a parameter and its inverse being equal. At the optimum value, we know that the estimate, the first and second derivatives are all consistent between the parameter and its inverse. But as soon as we move away from the optimum, there is no guarantee at all, and it is clear that the likelihood function defined in terms of one or both of the formulations (parameter or inverse) must fail to be quadratic. It is frequently stated that the t -ratios given for non-linear models are approximate. The extent of this approximation is perhaps often underestimated.

Is there a better approach? The conventional calculations made for models estimated on the maximum likelihood criterion appear to make best use of the information available at the optimum likelihood value. To get better information, it would be necessary to investigate the true variation of the likelihood function as we move away from the optimum. For example, to obtain 95% confidence limits for a parameter, it would be useful to find the upper and lower values beyond which 2.5% of the likelihood lies, the approach of [Armstrong et al. \(2001\)](#). These would not necessarily be symmetric around the optimum values of the parameter, but their inverses would represent the upper and lower confidence limits of the inverse of the parameter. One could then test whether any particular value, e.g. 0, lay within the 95% confidence bands. Making these calculations would be time-consuming, as specialised software does not appear to exist and in most cases, therefore, it is necessary to continue to use t -ratios.

5.2 Conclusions

In this paper, we have discussed the issue of computing standard errors for measures that are functions of parameters estimated from discrete choice models. This is an issue of crucial importance as, in choice modelling, the values of interest to analysts are in fact often functions of these parameters rather than the parameters themselves. The paper has shown how the simple Delta formula can be used to derive such standard errors while maintaining desirable properties for the standard errors, where, previously, analysts have seen this as an approximate approach.

The paper then presents formulae for the standard errors of a number of regularly used measures, going beyond what is currently available in the choice modelling literature⁸, and illustrates how they are indeed exact calculations.

The simple example in Section 4 highlights the benefit of this approach over simple simulation; [Daly et al. \(2011\)](#) show that simple simulation must ultimately fail for the important example of a coefficient ratio.

⁸To allow readers to exploit the formulae developed in this paper, freeware software is being made available online, at <http://www.stephanehess.me.uk>

Future work should extend the set of functions developed in this paper. In this context, given the results of the simulation and the non-existence proof for the variance of the ratio estimator, and the fact that simulation may in the short term remain necessary for some of the most complicated functions, further investigation should be undertaken to determine whether the common practice of estimating errors by simulation is likely to yield serious misinterpretations in other cases as well. The analytical results are clearly more reliable and hence preferable if they can be calculated without undue complication.

Acknowledgements

The authors are grateful to Chandra Bhat and three anonymous referees for extensive comments on earlier versions of the paper. The second author acknowledges the financial support of the Leverhulme Trust in the form of a *Leverhulme Early Career Fellowship*.

References

- Armstrong, P., Garrido, R. A., Ortúzar, J. de D., 2001. Confidence interval to bound the value of time. *Transportation Research Part E* 37 (1), 143–161.
- Cramer, J. S., 1986. *Econometric applications of Maximum Likelihood methods*. Cambridge University Press, Cambridge.
- Daly, A., Hess, S., Train, K., 2011. Assuring finite moments for willingness to pay estimates from random coefficients models. *Transportation*, forthcoming.
- Daly, A., Zachary, S., 1975. Commuters' values of time. Local Government O.R. Unit Report T55, Reading, UK.
- de Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R., Hofman, F., 2007. Uncertainty in traffic forecasts: literature review and new results for the netherlands. *Transportation* 34 (4), 375–395.
- Greene, W. H., 2008. *Econometric Analysis (Sixth Edition)*. Pearson Education Inc, Upper Saddle River, NJ.
- Hole, A. R., 2007. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics* 16 (8), 827–840.
- Krinsky, I., Robb, A., 1986. On approximating the statistical properties of elasticities. *Review of Economics and Statistics* 68 (4), 715–719.

- Krinsky, I., Robb, A., 1991. On approximating the statistical properties of elasticities: a correction. *Review of Economics and Statistics* 72 (1), 189–190.
- Marzano, V., Papola, A., 2008. On the covariance structure of the cross-nested logit model. *Transportation Research Part B* 42 (2), 83–98.
- McFadden, D., 1978. Modelling the choice of residential location. In: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. W. (Eds.), *Spatial Interaction Theory and Planning Models*. North Holland, Amsterdam, Ch. 25, pp. 75–96.
- Papola, A., 2004. Some developments on the cross-nested Logit model. *Transportation Research Part B* 38 (9), 833–854.
- Train, K., 2009. *Discrete Choice Methods with Simulation*, 2nd Edition. Cambridge University Press, Cambridge, MA.