

# Errors in functions of the parameters of statistically estimated models

Andrew Daly  
Gerard de Jong  
ITS, University of Leeds and RAND Europe

## 1. Introduction

In transport analysis, discrete choice models are often specified to contain parameters which must be estimated from observed behaviour, often using the maximum 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.

Once estimated, models can be used in a number of ways, such as:

- the calculation of simple functions of the parameters – in an important example the calculation of values of time as the ratio of time and cost coefficients;
- the prediction of future behaviour of populations, using aggregation techniques such as sample enumeration;
- the calculation of user benefit measures.

Outputs of all of these applications are, of course, functions of the estimated parameters and, as such, are 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.

Conventionally, two techniques have been used to derive these errors.

- Exploiting the asymptotically multivariate normal distribution of the estimates, text-book results concerning functions of normal variates have been used to derive distributions, and hence error measures, for simple functions of the parameters. This procedure has often been followed to derive error estimates for values of time.
- Sampling procedures have been applied to the estimated parameters, again treating them as multivariate normal distributions, and distributions of the relevant functions of the sampled values have been used to obtain information about the distribution of the function required.

In the paper it is shown that the first approach is inefficient and can better be replaced by the method set out in the paper. The second approach can also be replaced, with a substantial saving in convenience and time, when the functions concerned are not too complex. However, in very complicated situations or for very large models sampling from the parameter distribution remains the only practical approach (see de Jong *et al.*, 2005b).

The proposed approach is inspired by the well-known approximation for the variance of a function of random variables as a function of the covariance matrix of those variables and the first derivatives of the function with respect to those variables. In Section 2 of the paper it is shown that application of this formula gives the same results for a number of important cases, including the ratio of coefficients, as the *ad hoc* calculation from first principles given in text books. The formula is however more general and gives more insight into how the error depends on the structure of the problem. Further examples are given of calculations of error in forecasts, forecast changes and consumer surplus measures.

However, the status of these calculations is considerably greater than might be appreciated and this issue is explored in Section 3. First, because the asymptotic normality of the original

parameter estimates depends on a second-order approximation, while the calculation of the variance of the function is a first-order approximation, it is possible to prove (under reasonable conditions) that the function is also asymptotically normal. Second, because likelihood could just as well have been maximised with the derived function as one of the original arguments, the estimate of the function is itself a maximum likelihood estimate (again under reasonable conditions), with the properties of minimum variance and asymptotic lack of bias of all such estimates. We indicate the conditions under which the properties hold.

A final section uses these results to throw more light on the commonly-used ‘t ratio’ and some paradoxes concerning ‘t’ tests.

We conclude that these simple calculation methods can be used to obtain error measures for many commonly-needed statistics, that these error measures often have as much validity as the original parameter estimates and this approach gives considerable insight into a number of paradoxes concerning error calculations.

It is useful to make two notes on related subjects before moving on to the main discussion. First, the errors in the parameters discussed in this paper are of course not the only errors in model forecasts, as is discussed by de Jong *et al.* (2005b). In particular the impact of errors in input data can be calculated by similar methods to those presented here. Second, the discussion is presented here in terms of choice models but many of the issues and methods can be applied to other forms of model as well.

## 2. The error calculation procedure

In this section we set out a calculation procedure, which can be applied to the parameters of any statistically estimated model, to calculate the errors in functions of the estimated parameters. Typically, the criterion used to make the original estimation will be maximum likelihood.

### 2.1 The estimation context

In this section we review a few standard results which form the foundation for the ideas presented in the paper.

It is imagined that a choice model is estimated from a large number  $n$  of observations of revealed or stated preferences of consumers.<sup>1</sup> 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 a number of the properties of the parameter estimates.

The classical result which is widely used in this context is that, provided reasonable conditions are met and the model is correctly specified (not always a reasonable assumption!) then the expected score (first derivative of the likelihood function with respect to the model parameters) is zero and the maximum likelihood estimates  $\theta^+$  of the model parameters are distributed asymptotically normally around the true values  $\theta^*$ :

$$\sqrt{n} \cdot (\theta^+ - \theta^*) \rightarrow N(0, (-H)^{-1})$$

where  $H$  is the Hessian (second derivative) of the likelihood function with respect to the model parameters.

---

<sup>1</sup> In the transportation context, ‘consumers’ may be travellers, households (e.g. in car ownership decisions), freight shippers, etc..

If the optimum of the likelihood is well defined, then  $H$  will be negative definite and there is no problem about inverting  $(-H)$  which is a positive definite matrix, to derive  $(-H)^{-1}$  as another positive definite matrix. We also note that  $(-H)^{-1}$  is 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  $\theta^*$ , then the ‘sandwich estimator’ can be used:

$$\sqrt{n} \cdot (\hat{\theta}^+ - \theta^*) \rightarrow N(0, (-H)^{-1} \cdot V \cdot (-H)^{-1})$$

where  $V$  is the variance-covariance matrix of the score vector<sup>2</sup> and is of course positive definite unless the data is linearly dependent, so that the product defining the covariance of the normal distribution is positive definite.

Sometimes, e.g. when it is inconvenient to calculate the true second derivative matrix,  $V$  is used as an approximation to  $-H$ , and  $V^{-1}$  as an approximation to the covariance matrix of errors in  $\theta$ , a method popularised by Berndt *et al.* (1974); for that reason  $V$  is often referred to as the BHHH matrix. In any of these three cases, we have a matrix which we can use to describe the covariance of the estimates  $\hat{\theta}^+$ , so we can write

$$\sqrt{n} \cdot (\hat{\theta}^+ - \theta^*) \rightarrow N(0, \Theta)$$

where  $\Theta$  is  $(-H)^{-1}$ ,  $(-H)^{-1} \cdot V \cdot (-H)^{-1}$  or  $V^{-1}$  as appropriate.

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

All the estimators are asymptotically unbiased, i.e. distributed with a mean which is equal to the true value, and consistent, i.e. converge to the true value as the amount of data increases (this is the role of  $\sqrt{n}$  in the formulae above). Further, the distribution of those estimates can be taken as normal, with a covariance matrix which can be calculated without great difficulty.

## 2.2 First-order approximation of error in functions of parameters

Suppose we can express a required output as a differentiable function of the parameters, e.g.

$$\phi = \phi(\theta).$$

The Slutsky Theorem<sup>3</sup> 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 reasonable properties. For this theorem  $\theta$  does not have to be a maximum likelihood estimator, but we shall see later that if  $\theta$  is a maximum likelihood estimator, and under certain other conditions, these results can have substantially more status and correspondingly better properties.

Then it is shown by simple calculus in statistical textbooks that a first-order approximation to the error in  $\phi$  induced by the error in  $\theta$  is given by

<sup>2</sup> This result is quoted by Train (2003), p. 228.

<sup>3</sup> Quoted by Ben-Akiva and Lerman (1985), p. 19. Ben-Akiva and Lerman give the example of the ratio estimator for  $\theta_1/\theta_2$ , failing to notice that this is not continuous at  $\theta_2=0$ ! However, replacing  $\phi$  by  $\phi^*=\theta_1 \cdot \theta_2^*$ , with  $\theta_2^*=1/\theta_2$  if  $|\theta_2|>\delta>0$  and  $\theta_2^*=\theta_2/\delta^2$  otherwise produces a continuous estimator; then letting  $\delta \rightarrow 0$  covers as much of the range as we wish.

$$\text{var}(\phi) \cong \phi'^T \Theta \phi' \quad (1)$$

where  $\phi'$  is the vector first derivative of the function  $\phi$  with respect to  $\theta$  and  $\Theta$  is the covariance matrix of the estimates of  $\theta$ .

This approximation can be used in a wide range of circumstances to estimate the error in functions of estimated parameters.

Example (1): suppose we wish to calculate a simple ratio of parameters (assuming the denominator is never zero)

$$v = \theta_1 / \theta_2$$

we can differentiate  $v$  to obtain

$$\begin{aligned} v_1 &= 1 / \theta_2 \\ v_2 &= -\theta_1 / \theta_2^2 \\ v_j &= 0 \text{ for any other components in } \theta. \end{aligned}$$

Application of equation (1) for the variance of  $v$  gives

$$\text{var}(v) \cong \phi'^T \Theta \phi' = \Theta_{11} v_1^2 + \Theta_{22} v_2^2 + 2 \Theta_{12} v_1 v_2$$

writing out the matrix multiplication in full. Then, substituting the derivative values we get

$$\begin{aligned} \text{var}(v) &\cong \Theta_{11} / \theta_2^2 + \Theta_{22} \cdot \theta_1^2 / \theta_2^4 - 2 \Theta_{12} \theta_1 / \theta_2^3 \\ &\cong v^2 (\Theta_{11} / \theta_1^2 + \Theta_{22} / \theta_2^2 - 2 \Theta_{12} / \theta_1 \theta_2) \end{aligned}$$

which is the well-known formula for the approximate variance of a ratio of random variables and is often applied to calculate the value of time and similar ratios in travel demand analysis.

This formula is usually derived by more *ad hoc* methods specific to the case of a ratio, which conceal the way in which the first derivative operates. Moreover, for each new function we usually need to go back to first principles. The generality of the first derivative approach, i.e. equation (1), will be illustrated by application in more complex contexts below.

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. The deeper analyses presented later require that maximum likelihood methods are used, however.

### 2.3 Generalisation and further examples

An immediate generalisation of equation (1) is to a vector function  $\phi$ :

$$\text{covar}(\phi) \cong \phi'^T \Theta \phi'$$

In this equation  $\phi'$  is to be seen as the matrix of partial derivatives whose elements are

$$\phi'_{ij} = \partial \phi_i / \partial \theta_j$$

The specific advantage of this generalisation is that it allows us to calculate the correlations of the errors in functions of estimated parameters.

### Example 1a: Reciprocal

Suppose we estimate a parameter  $\theta$  with an error variance  $\Theta$  and we wish to know the error in its reciprocal. This is actually a special case of the ratio estimator derived above, with the numerator being constant. The formulae above simplify considerably and we can calculate

$$\text{var}(1/\theta) \cong \Theta / \theta^4$$

It is interesting in this case to note that if we calculate the t ratio

$$t(\theta) = \theta / \sqrt{\Theta}$$

but

$$t(1/\theta) = (1/\theta) / \sqrt{\text{var}(1/\theta)} = \theta / \sqrt{\Theta}$$

That is, the 't' value for the reciprocal is *exactly* the same as the 't' value for the original parameter, although the meaning of the test is of course radically different.<sup>4</sup> This is the first of a number of paradoxes concerning 't' ratios that are illustrated by this approach.

### Example 2: Product

The calculations for a product are straightforward, since  $v_1 = \theta_2$  and  $v_2 = \theta_1$ , we get

$$\text{var}(v) \cong \Theta_{11} \theta_2^2 + \Theta_{22} \theta_1^2 + 2 \Theta_{12} \theta_1 \theta_2$$

which again is a well-known result that can be found in textbooks., but then as the outcome of an *ad hoc* and complicated calculation procedure, whereas here it is derived from a simple formula.

### Example 3: Predicted demand

Suppose we calculate demand  $Q_j$  for alternative  $j$  using a simple sample enumeration procedure over each member  $s$  of a sample

$$Q_j = \sum_s w_s \cdot p_{js}$$

where  $w$  is the expansion factor for each member of the sample and  $p$  is the choice probability.

Then we can calculate the variance in this demand by

$$\text{var}(Q_j) = Q_j'^T \Theta Q_j'$$

where  $\Theta$  is the covariance matrix of the model parameters and  $Q_j'$  is a vector whose  $k^{\text{th}}$  element is

$$Q_{jk}' = \sum_s w_s \cdot \partial p_{js} / \partial \theta_k$$

If the choice model is a linear multinomial logit and  $x$  are the characteristics of the alternatives as faced by each member of the sample:

---

<sup>4</sup> This property has been observed in practice when different software (e.g. ALOGIT and Biogeme) formulates structural parameters in different ways, such that a parameter in one program is the inverse of the corresponding parameter in the other.

$$p_{js} = \exp V_{js} / \sum_i \exp V_{is}$$

$$V_{is} = \sum_k \theta_k X_{isk}$$

then

$$Q_{jk}' = \sum_s w_s \cdot p_{js} \cdot (1-p_{js}) \cdot X_{jsk}$$

This calculation appears simple enough, but when the model is no longer multinomial logit and/or the sample being expanded is large, with complicated calculations for  $w$ , the amount of calculation involved here can be prohibitive and a sampling approach can be necessary (see de Jong *et al.*, 2005b).

For policy issues, a crucial aspect of the calculation is not so much the error in predicted demand as the error in the *difference* in demand for  $j$  between two scenarios. To obtain this we could calculate the difference in demand as arising from the change in choice probabilities between the two scenarios (i.e. assuming  $w$  does not change)

$$\Delta Q_j = \sum_s w_s \cdot \Delta p_{js}$$

and make the calculations on  $\Delta Q$  as was done for  $Q$  above. The problem with this approach is that the variance of  $\Delta p$  is complicated to calculate, so it might be preferable to calculate

$$\text{var}(\Delta Q) = \text{var}(Q^1 - Q^0) = \text{var} Q^0 + \text{var} Q^1 - 2 \text{covar}(Q^0, Q^1)$$

where the 0 and 1 superscripts refer to base and policy scenarios respectively. This approach takes advantage of the generalisation of  $\phi$  to a vector function indicated at the beginning of this section. Reading out from the formula, we can calculate

$$\text{covar}(Q^0, Q^1) = Q^{0T} \Theta Q^1 = Q^{1T} \Theta Q^0$$

where the  $Q'$  are differentials of the respective  $Q$ 's for base and policy scenarios. Since the calculation of these differentials is straightforward, the whole calculation becomes quite simple. The covariance is quite important in this calculation, because often demand estimates will be highly positively correlated and omitting this correlation from the calculation would seriously distort the result.

An important application of this difference calculation is in 'pivoting' applications for making forecasts. Here an estimate is made of the change in demand between a base case and a forecast scenario, for which the difference error calculation can be applied as above. Then, because of the positive correlation between the model outputs, *and if an accurate estimate of base traffic is known*, a much more accurate estimate of demand under the forecast scenario can be obtained.

#### Example 4: Logsum

A final example of this type of calculation is with the logsum, often used to calculate consumer surplus (de Jong *et al.*, 2005a, Kohli and Daly, 2006). The surplus  $S$  and logsum  $V$  are defined for an individual by

$$S = (1/\theta_c)V = (1/\theta_c) \log \sum_k \exp V_k$$

where  $\theta_c$  is the coefficient of cost in the model, assumed for simplicity to be constant. For further simplicity we shall assume that the utilities  $V_k$  of the alternatives  $k$  are associated with a linear logit choice model as given in the equations above. We then obtain

$$\text{var}(S) = S'^T \cdot \Theta \cdot S'$$

and

$$S'_j = (1/\theta_c) \{ \sum_k p_k x_{kj} - \delta_{jc} S \}$$

where  $\delta_{jc}$  is 1 when  $j$  refers to the cost coefficient and 0 otherwise.

This calculation is also very simple, even though it takes into account the correlation with the estimate of the cost coefficient, and can be extended, as in the case of the demand calculations, to give the surplus for the difference between base and policy scenarios.

To calculate the consumer surplus for a population the separate values for individuals would normally be added up (subject to suitable economic reservations!) and error can then be calculated using a summed value of  $S'_j$  which would take proper account of positive correlation between the surplus of different individuals.

### 3. The status of transformed estimates

In the previous section of the paper a useful calculation procedure was presented *as an approximation* to obtain estimates of the estimation error associated with the parameters of statistical models. The status of these apparent approximations is now investigated further, under the explicit assumption that the estimation is made by maximum likelihood methods.

In the context of maximum likelihood estimates of  $\theta$ , equation (1) can be seen as a two-stage calculation of the variance of the function  $\phi$ .

- First, the matrix  $\Theta$  is an asymptotic approximation to the true covariance matrix of the estimates  $\theta^+$ .
- Second, the distribution of  $\phi$  would naturally be seen as some complicated function derived from the asymptotic normality of the distribution of  $\theta^+$ . For example, in the case of the estimate  $v$  of the ratio of two parameters, we might naturally call on literature which describes the distribution of the ratio of two normal variates: skewed, with complications arising when the denominator gets close to 0.

Altogether, the distribution of  $v$  described in this way is unappealing. However, this 'two-stage' interpretation of equation (1) lacks insight. A much deeper understanding can be obtained along the lines set out by Cramer (1986, Section 3.1).

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

Then we recall the Slutsky theorem, mentioned above, that if  $\theta^+$  is a consistent estimator (e.g. a maximum likelihood estimator, MLE) of  $\theta$ , then if  $\phi$  is continuous  $\phi^+ = \phi(\theta^+)$  is a consistent estimator of  $\phi$ .

Moreover, if  $\phi$  is differentiable then a much stronger property can be established, that is that the distribution of  $\phi$  converges asymptotically to a normal distribution

$$\sqrt{n} \cdot (\phi^+ - \phi^*) \rightarrow N(0, \phi'^T \Theta \phi')$$

where  $\Theta$  is the covariance of  $\theta^+$ ; i.e.  $\phi^+$  is asymptotically equivalent to an MLE of  $\phi^*$ , as it has the same asymptotic distribution.

Some authors apparently claim that  $\phi^+$  can be regarded as an MLE with no further ado, that it represents the maximisation of likelihood over a space induced by the transformation  $\phi$ . Cramer, 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  $\eta$  with the same dimension as  $\theta$

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

For the transformation  $g$  to be invertible it must be one-to-one<sup>5</sup>, as well as differentiable. While these conditions may be restrictive from a mathematical point of view, in practice they cause little difficulty. With these conditions, Cramer shows that the distribution of the dependent variable is not affected by the transformation, so that  $\eta^+ = g(\theta^+)$  is an MLE of  $\eta^*$ . Cramer then goes on to derive the covariance  $\Phi$  of  $\phi$  around  $\phi^*$

$$\Phi = g'^T \Theta g'$$

where  $g'$  is the derivative matrix (Jacobian) of  $g$  with respect to  $\theta$ . This is exactly the result of section 2 again but now, because we know  $\eta^+$  to be an MLE, we also know that  $\Phi$  is the Cramér-Rao lower bound<sup>6</sup> of minimum variance for the estimator.

We are now in a position to reassess the results derived in section 2. Instead of seeing the first-derivative 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 first-derivative approach is exactly what is required to obtain the covariance of the transformed parameter estimates around the true values.

In a later remark in the same section, Cramer 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 one-to-one and differentiable.

In an important example, we recall Example 1 of section 2, calculating the ratio of parameter estimates, for example in estimating the value of time. Applying the reasoning of the present section, we now see that the ratio, i.e. the value of time, is itself a maximum likelihood estimate. Thus it has the properties of consistency, unbiasedness, minimum variance and asymptotic normality, that we would wish.

How the ratio of asymptotically normal estimates can itself be asymptotically normal is the subject of the following section.

---

<sup>5</sup> This also implies  $g$  and  $g^{-1}$  are non-singular, within the space of interest.

<sup>6</sup> The accent in Cramér is important in this context!



#### 4. Interpretation of errors of estimates

The issue discussed in this section of the paper is 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.

In section 2 of the paper, the example was given that the t ratio of the estimate of a reciprocal of an MLE parameter was equal to the t ratio of the parameter itself. In section 3 it was shown that the estimate of the reciprocal had 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 which the estimates and the error estimates are based comes from the likelihood function. The maximum point of this function is clearly defined, provided there are no numerical problems, and presents no problem when we make transformations. The issue arises in defining the distribution around the maximum value.

The information on which the error estimates are based comes from the description of the likelihood function that 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.

The information we have is thus that the optimum value of the likelihood function is at a certain point, the first derivative is zero and we also know the second derivative (or we have a good estimate) at the optimum value. *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.

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. If we want to make conventional significance tests and we are looking at 't ratios' around 2,<sup>7</sup> we are in fact already some distance from the optimum and it would not be surprising to find inconsistencies.

It is frequently stated that the 't ratios' given for non-linear models are approximate and the reasons for this are clarified by the results obtained earlier in this paper. 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 2½% of the likelihood lies. These would not necessarily be symmetric around the optimum values of the parameter, but their inverses would represent the upper and

---

<sup>7</sup> Conventionally, 95% confidence limits are often taken for t values of  $\pm 1.96$ .

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. In most cases, therefore, it is necessary to continue to use 't ratios', but with an enhanced understanding that attributing any great precision to these values would be incorrect.

T ratios should therefore be taken as general indications of the importance and accuracy of estimation of a parameter. A value of 5 indicates that the parameter should probably be included in the model (providing it has the right sign!); a value of 0.2 indicates that it has almost no effect. Given an intermediate value, we need to examine our prior beliefs about the importance of the variable in question for the model.

For example, if we estimate a cost parameter in a choice model, after our best efforts to improve the model, with a t ratio of  $-1.3$ , then:

- our prior beliefs would usually be strong that cost had a negative impact on the choice of an alternative;
- the data says the value estimated is better than zero, but the data is insufficient to make a very good estimate.

In the absence of any other information, one would usually conclude that the model would be better with this parameter taking the estimated value, rather than omitting cost from the model, i.e. taking zero to be a better value despite the prior beliefs *and* evidence of the data, albeit weak. However, if the parameter represented a preference for bus use for people aged 25-40 compared with the rest of the population, one would conclude that this was probably a random effect and there was no reason to include the variable in the model.

Intelligent judgement is, as always, decisive and tests such as the t ratio should be taken only as indicative.

## 5. Conclusions

The paper sets out to investigate the way in which error measures can be derived for models based on maximum likelihood estimation. These methods yield estimates with many attractive properties, with error estimates that can be estimated in one of three ways, depending on the circumstances of the estimation.

Given error estimates for the parameters, a method is presented to determine errors for functions of those parameters, based on the first derivative of the function in question. This method is applied to give formulae for errors in the

- ratio of parameters,
- and, as a special case of the ratio, the inverse of a parameter,
- forecast demand from a choice model and
- consumer surplus estimated from a choice model.

These calculations are simple in principle and can be applied to numerous other cases, avoiding the need for *ad hoc* calculation in each case. They are very useful but appear to give approximate results.

However, following the explanation of Cramer (1986) it becomes apparent that, under fairly weak conditions which would generally apply in transport analysis, functions of maximum likelihood estimates are themselves maximum likelihood estimates and the first-derivative method yields the true (minimum variance) error measures appropriate to the transformed parameters.

A final section discusses some of the apparent paradoxes revealed by these methods.

## References

Ben-Akiva, M. and Lerman, S.R. (1985) *Discrete Choice Models*, MIT Press, Cambridge, Mass..

Berndt, E.U., Hall, B.H., Hall, R.E. and Hausman, J.A. (1974) Estimation and inference in non-linear structural models, *Annals of Economic and Social Measurement*, **3/4**, pp. 653-665.

Cramer, J.S. (1986) *Econometric applications of Maximum Likelihood methods*; Cambridge University Press, Cambridge.

Jong, G. de, Daly, A., Pieters, M., Hoorn, A. van der (2005a) The logsum as an evaluation measure: review of the literature and new results, presented at conference of European Regional Science Association.

de Jong, G. de, Daly, A., Pieters, M., Miller, S., Plasmeijer, R. and Hofman, F. (2005b) Uncertainty in traffic forecasts: literature review and new results for The Netherlands, presented to ETC, Strasbourg.

Kohli, S. and Daly, A. (2006) Use of logsums in welfare estimation: PRISM application, presented to European Transport Conference, Strasbourg.

Train, K. (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press.