ELASTICITY, MODEL SCALE AND ERROR

Andrew Daly
RAND Europe and Institute for Transport Studies, University of Leeds

1. INTRODUCTION

Elasticity is often used as a convenient means of summarising the sensitivity of a population to changes in transport policy variables, such as the prices of alternative modes or routes. In developing models, analysts are often required to demonstrate that the elasticities implied by their models are consistent with generally accepted notions of what elasticities ought to be. Most travel demand models used in practical forecasting studies are formulated with a single measure of utility (or its negative, generalised cost) for each alternative mode, route (or destination etc.) and in such models it is clear that the elasticity they yield is directly proportional to the scale that multiplies the utility of the alternatives, if utility is measured in fixed units, e.g. of cost or time. Obtaining an acceptable value for elasticity is then a matter of obtaining a suitable value for that scale, conditional on the structure of the model.

However, it is generally known that when statistical estimates are made of the coefficients of a model the value of the model scale is inversely proportional to the error with which utility is measured. There is therefore an apparent paradox that an improvement to the model, i.e. a reduction in error, hence an increase in the model scale parameter, seems to indicate an increase in elasticity. It cannot be the case that the quality of the model affects the behaviour of the population!

The paper presents a resolution of this apparent paradox, which is analysed in terms of the distribution of the utility function among the population. When a model is improved, although the error is reduced, the variance of the *measured* part of the utility is increased and it is shown that, *ceteris paribus*, this implies a reduction in elasticity that balances the increase caused by the increase in model scale. The basic mechanism is quite simple, though the details can be complicated for particular model types. The functioning of the effect is illustrated by models estimated and applied on simulated data.

After a brief sketch of the background the central issue is presented and some previous work is discussed. The problem is then analysed in Section 3 through simple models based on simulated data. Section 4 is concerned with giving an explanation of the effect that is described. Conclusions are then drawn.

2. BACKGROUND

In this section of the paper we present the context of choice modelling, taking the opportunity to define the notation used subsequently, discuss how measures of sensitivity can be defined for choice models, introduce the specific problem to which the paper is addressed and review briefly previous work in the area.

2.1 The context of choice modelling

Transport demand arises because of the choices made by people to travel and how, when and where to travel. It is then natural to set an analysis of issues concerning travel demand in the context of choice modelling and in the context of Random Utility Models (RUM) which are used as the basis for almost all advanced work in choice modelling and which can also be identified as the basis for most of the transport demand models used in practice.

The RUM framework postulates that choices are made as the result of utility maximisation by individuals and that the utility U_{ij} of an individual i for a choice alternative j can be approximated by

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$

where V is the analyst's best approximation of the utility and
ε is the error in the analyst's approximation.

In the analysis, ε is treated as a random number (hence the concept of *random* utility) and the initial focus is on specifying an appropriate form for its distribution. Given a distribution of ε, the probability p_{ij} that individual i will choose alternative j in the total choice set C can in principle be calculated

$$\Pr\{i\, chooses\, j\} = p_{ij} = \Pr\{V_{ij} + \varepsilon_{ij} \geq V_{ik} + \varepsilon_{ik}, for\, all\, k \in C\}$$

This is in principle a more-or-less mechanical calculation, but depending on the distribution specified for ε, the actual process can be more or less difficult, sometimes involving Monte Carlo procedures. The simplest calculations arise in models of the logit family, which are used (for this reason) in most research and practical applications, and for which ε is specified to have an extreme-value or Gumbel distribution with the same variance across all of the alternatives.

For the present paper, we need to be careful how we specify the scale of the model. Of course, the scale of the utility U is entirely arbitrary but the scales of V and ε need to be compatible so that the probability calculations can be made appropriately. The way in which this is achieved in practice for logit models is that ε is specified to have a *standard* Gumbel distribution:

$$\Pr\{\varepsilon \leq t\} = \exp\left(-e^{-t}\right)$$

This distribution has a mean which is equal to Euler's constant (0.5772..) but this is of no consequence since all alternatives have the same distribution, i.e. the same mean. More importantly, the Gumbel distribution has variance $\pi^2/6$, and it is this variance that defines the scale of the model. Thus the scale in which V is measured has to be adjusted to match the assumption that the standard error in estimating the utilities is $\pi/\sqrt{6}$.

In practice, V is generally specified as a function linear in unknown parameters β

$$V_{ij} = \sum \beta_r x_{rij}$$

where x_{rij} represents measured data items, different items (e.g. cost, time etc.) indexed by $r = 1,...$, relating to alternative j for individual i.

An important stage in most studies is to make an estimate of the β's on the basis of the revealed or stated preferences of a sample of individuals. The ratios of the β's can be interpreted as 'trade-off' ratios indicating the *relative* importance of marginal changes in x variables. While these ratios are of primary importance in many studies, for the present work the interest is in the *absolute* value of β.

The x variables include concrete measures of the attributes of the alternatives, such as the number of minutes or Euros that are spent when that alternative is chosen[1]. Because of the relationship between the scale of V and the measurement error in the utilities, the β_t attached to (say) a time attribute gives the relationship between minutes and the error measure in the model. Specifically, we can say that the utility standard error, measured in minutes, is $\left(\pi/\sqrt{6}\right)/\beta_t$.

In model estimation, what is happening with respect to the absolute values of the β's is therefore that they are adjusted to make the measured utility compatible with the standard error. The β's are therefore determined by the analyst's success in approximating the true utility U by the measured utility V, and not by the responsiveness of the population to changes in x. Specifically, if the analyst is more successful, the error will be reduced and β will increase.

2.2 Elasticity and Demand Sensitivity

The concept of elasticity is widely used in economics to measure demand sensitivity to variables such as price and has an attraction in that context because it is dimensionless. The standard formulation of the elasticity η_{jr} of the demand Q_j for alternative j with respect to a variable x is

for 'arc' elasticity $\eta_{jr} = \frac{x_r}{Q_j} . \frac{\Delta Q_j}{\Delta x_r}$ or for point elasticity $\eta_{jr} = \frac{x_r}{Q_j} . \frac{\partial Q_j}{\partial x_r}$

where ΔQ and Δx_r represent respectively the change in demand and the change in the variable that causes it. The point elasticity form, using differentials rather than finite differences, is often more convenient for theoretical discussions.

---

[1] The possibility of non-linear transformations of x exists, of course, but dealing with this possibility would complicated the discussion unnecessarily at this point.

For application to predict aggregate demand, a choice model is implemented to give the total expected demand for alternative j by

$$Q_j = \sum_i w_i p_{ij}$$

where w_i is a weight or expansion factor, effectively giving the number of individuals of type i being considered.

Providing we allow i to include the specification of the home location of individuals, this simple formulation covers almost all forecasting models used in research and practice, including both what is normally considered as 'sample enumeration' and aggregate models. The formulation allows us to express elasticity for a choice model by

$$\eta_{jr} = \frac{\overline{x_r}}{Q_j} . \frac{\partial Q_j}{\partial x_r} = \frac{\overline{x_r}}{Q_j} . \sum_i w_i \frac{\partial p_{ij}}{\partial x_r}$$

A difficulty in applying this formulation in a choice modelling context is that the value of x_r often varies in the population so that the mean value $\overline{x_r}$ has to be included in the equation and this may not give a good representation of the population, for example when x_r is zero for some observations, or when a non-linear transformation is applied. The presence of the total demand Q_j in the equation is also a confusing factor. An alternative measure of model sensitivity is suggested by Cramer (2007): the Average Sample Effect (ASE). Generalising his measure slightly (allowing for the presence of *w*) this can be given as

$$ASE_{rj} = \frac{1}{W} \frac{\partial Q_j}{\partial x_r} = \sum_i w_i \frac{\partial p_{ij}}{\partial x_r} \Big/ W \qquad (1)$$

where $W = \sum_i w_i$, the total amount of demand.

This definition applies only for a binary model, but a more general measure can be obtained if we utilise the fact that x_r has its impact in the model as part of the utility of an alternative, so the demand sensitivity can be defined

$$\begin{aligned} \zeta_{ik} &= (1/\beta_r) ASE_{rj} \\ &= (1/\beta_r) \sum_i w_i \frac{\partial V_k}{\partial x_r} \frac{\partial p_{ij}}{\partial V_k} \Big/ W \\ &= \sum_i w_i \frac{\partial p_{ij}}{\partial V_k} \Big/ W \end{aligned}$$

where β_r is the coefficient of x_r in the utility function of alternative k .

The function ζ can be applied to any form of model; it is dimensionless, independent of r and depends only on model structure and on the measured utilities for each individual, so that it can be considered to be a fundamental property of the model and the distribution of the x's in the population.

Note that for given values of $\overline{x_r}$ and Q_j we can always obtain the elasticity from the demand sensitivity ζ :

$$\eta_{jr} = \beta_r \frac{\overline{x_r}}{\overline{p_j}} . \zeta_{jk(r)}$$

where $\overline{p_j} = Q_j / W$, the average probability of choosing alternative j and $k(r)$ is the alternative to which x_r applies.

In this form we see the apparent direct connection between elasticity and coefficient value. If β increases, then the rest of the equation is apparently unchanged, since ζ seems to be a fundamental property of the model, so the elasticity increases. But in the previous section we saw that β depended on the accuracy of the model, not on the responsiveness of the population: this is the basic paradox.

2.3 The work of Cramer (2007)

The paradox discussed in the two previous sections has been addressed by the evergreen Cramer in a recent paper. The context of that paper is that the omission of a relevant variable (orthogonal to the other variables) from a model of binary response, whether logit or probit, shifts the values of the coefficients towards zero. This is in contrast with a linear model, where the omission of an orthogonal explanatory variable increases the error in the model but *does not* change the values of the coefficients. The reason for this difference is that in a linear model the scale of the coefficients is defined by the scale of the response, usually denoted by y, which is not affected by the set of variables x chosen as regressors; in a binary response model, i.e. a choice model, the scale of the coefficients is defined by the error in the utility approximation, as described in Section 2.1.

Cramer describes work by Wooldridge (2002) that analyses the effect of omitting orthogonal regressors in a probit model. When the distribution of the omitted regressor is assumed to be normal in the data, it simply adds to the variance of the error (also, in a probit model, assumed to be normal) and a new probit model is obtained with a greater error variance than the model with all regressors included. In a logit model, the issue is not so simple, because Gumbel distributions do not combine well – logit is not always more straightforward than probit! – and Cramer resorts to simulation to investigate the issue.

To quantify the results, Wooldridge defines the Average Partial Effect (this is the expectation, whereas the ASE in equation (1) is the sample effect, i.e. the estimate of the expectation derived from the sample). The APE can be shown to be invariant in probit models to the removal of orthogonal explanatory variables. However, for the logit model, the APE is not immediately apparent, so Cramer derives the ASE as in equation (1), but omitting the weights:

$$ASE_{rj} = (1/N)\sum_i \frac{\partial p_{ij}}{\partial x_r} = (1/N)\beta_r \sum_i p_{ij}(1 - p_{ij})$$

where N is the number of elements in the sample, but he finds that the expected value of this expression is not easy to analyse. However, he does state that if a variable is removed, then, because of the increased error, β will move towards zero, the probabilities will all move towards 0.5 and the function $p(1-p)$ will move towards its maximum value of 0.25.

In contrast to the unclear analytical situation, the results of Cramer's simulation tests are very clear. Using simulations very similar to those described in the following section, he shows that the changes in $p(1-p)$ are *almost exactly* sufficient to counteract the change in β. That is, although β changes, the ASE is scarcely affected by the removal of an orthogonal regressor. In other simulations, Cramer shows that this remarkable result is also not noticeably affected by the distribution in the data of the variable removed: it may be normal, logistic, skewed or even a (0, 1) dummy. In this respect Cramer goes considerably beyond the Wooldridge result which applies only for the removal of a regressor with a specific distribution.

Thus Cramer gives the basis for a belief that the resolution of the paradox can be obtained by looking at the way in which the predicted probabilities change when the model is reduced in quality by the removal of a relevant variable. A strong point of his paper is that he shows that the APE remains stable, regardless of the nature of the variable that is removed, i.e. the form of its distribution in the data. However, Cramer's simulations are restricted in specific ways, which make further investigation necessary.

- The models in Cramer's simulations are binary and it is useful to investigate whether the introduction of further alternatives changes the issue. In particular, when the model has a nesting structure different results may be expected, because the nesting structure is intended to accommodate different response scales.
- Cramer studied only the removal of *orthogonal* regressors. it is common, however, for regressors to be correlated.

More important, the explanation of the mechanism by which adjustments to the demand sensitivity take place has not yet been made intuitive: what is the precise nature of the changes to p which bring this about? In particular, the explanation Cramer offers for the adjustment to the ASE by the probabilities moving towards 0.5 cannot apply in every case, since in any model containing alternative-specific constants the average probabilities will remain constant, whatever the quality of the model. An intuitive grasp of these mechanisms would be of great help to modellers trying to understand and explain the way models work in practice. We begin by undertaking some very simple experiments.

3. SIMPLE EXPERIMENTS

When the analytical situation is unclear, simulation can offer a way to gain understanding of the effects operating in a model. In this case, we have conducted a number of very simple simulation experiments to illustrate how model scale and demand sensitivity change in a range of circumstances chosen to illuminate the issues being discussed.

3.1 Binary model, independent variables

Closely following Cramer (2007), the following simple experiment has been made. Suppose choice is made on the basis of utility functions:

$$U_0 = 0$$

$$U_1 = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

where β are coefficients whose true value is 1;

ε has the standard logistic distribution.

The standard logistic distribution is the distribution of the difference of two standard independent Gumbel-distributed variables, and has the cumulative distribution function $F(t) = 1/(1+\exp(-t))$ with mean zero and variance $\pi^2/3$.

In the first experiment, the x's were distributed in the data as independent standard logistic variables. Hence their averages are very close to zero and they each have variance very similar to that of ε.

Table 1 shows the results of models estimated from a simulation in which 1,000,000 choices were simulated according to the model above. The large number of choices was selected to reduce problems with errors in estimation and this appears to have been successful, judging by the standard errors obtained. Two models were estimated from the data generated in this experiment, M2 where both β coefficients were estimated and M1 where β_1 was estimated but β_2 was artificially held to zero.

Table 1: First Simulation Results

Variable	M2	M1	Ratio
β_1	0.9954	0.6782	1.468
β_1 standard error	0.0023	0.0017	1.35
β_2	1.0000[2]	0	n/a
β_2 standard error	0.0023	n/a	n/a
kurtosis of ε	1.197	0.613	1.95

In the first model, the error in the measured utility is simply the variance of the logistic distribution of ε, i.e. $\pi^2/3$. In the second model, there is a further error due to the omission from the model of x_2, which also has variance $\pi^2/3$.

---

[2] The correlation of the estimates of β_1 and β_2 is 0.499.

Standard modelling theory would indicate the ratio of the coefficients to be $\sqrt{2}$ ≈ 1.414, whereas we find the ratio for β_1 is about 1.468 as shown in the table. The cause of this small difference is probably the changed kurtosis. Although the value for M1 is not exactly half the M2 value (which it should be, theoretically[3]) it is clear that the double error distribution in M1 is considerably less leptokurtic than the single one in M2. That is, the single distribution has fatter tails and relatively more outliers, given its variance, and we would therefore expect a slightly smaller value of β in M1 than indicated solely by the variances, as indeed is the case.

Forecast tests were then made by adding 0.5 to x_1 in the models, with the results shown in Table 2.

Table 2: Sensitivity Results from First Simulation

Choice of alternative 1	M2	M1	Ratio
Simulated[4]	499608	499608	1
Predicted base[5]	500525	500392	1.0003
Predicted in test	566562	566536	1.0001
Proportional change	0.13194	0.13218	0.9982

Clearly, the fact that the coefficient in model M1 is 46.8% larger than the corresponding coefficient in M2 has not had a direct impact on the sensitivity of the model. Essentially the changes in the pattern of $p(1-p)$ have more-or-less exactly compensated for the substantial decrease in the parameter.

However, both in M1 and M2 the average probability p is 0.5, so it is clear that the compensation for the scale change is *not* due to a change in the average. The variation of the p values around 0.5 is different in the two cases, however, and this may begin to explain the effect.

3.2 Binary model, independent variables, with constant

To investigate further the extent to which the scale correction effect depends on the average values of the probabilities, a further experiment was set up, very similar to the first, but including a constant. Specifically, the model was

$$U_0 = 0$$

$$U_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

---

3 Wikipedia (http://en.wikipedia.org/wiki/Gumbel_distribution) gives an expected value for the Gumbel distribution of +2.4. Adding or subtracting two identically-distributed variables halves the expected kurtosis (http://en.wikipedia.org/wiki/Kurtosis).

4 With reference to the simulated choices, we note that the standard error in the number choosing each alternative can be obtained from the binomial formula $\sqrt{np(1-p)}$, which in this case with $n=10^6$ and $p=½$ is 500, so that the observed discrepancy of 392 is entirely reasonable. The same pseudo-random numbers were used in both models, so it is expected that the same chosen numbers are found.

5 With reference to the base application of the model, it should not be expected that this would reproduce the simulated shares exactly, as there is no alternative-specific constant in the model.

where all the β's, including β_0, have the true value 1.

The results of estimating a model, once more based on one million simulated choices, are shown in Table 3 with β_2 estimated in model P2 and constrained to zero in model P1.

Table 3: Second Simulation Results

Variable	P2	P1	Ratio
β_0	0.9959	0.6790	1.467
β_0 standard error	0.0031	0.0024	1.29
β_1	0.9962	0.6813	1.462
β_1 standard error	0.0024	0.0017	1.41
β_2	1.0002	0	n/a
β_2 standard error	0.0024	n/a	n/a

The difference in error variance between these models is the same as that between the models in the first experiment and the scale change in the β's is very nearly the same as the 1.468 observed in that case.

Sensitivity tests were conducted on these models by adding 0.5 to x_1, as in the first case, and the results are reported in Table 4.

Table 4: Sensitivity Results from Second Simulation

Choice of alternative 1	P2	P1	Ratio
Simulated	630643	630643	1
Predicted base[6]	630643	630643	1
Predicted in test	690965	691018	0.9999
Proportional change	0.09565	0.09574	0.9991

Clearly, the addition of the constant has not materially affected the stability of model sensitivity. We may note that, to preserve the average p value between P2 and P1, as many p values have moved away from 0.5 as have moved towards it. The explanation that stability is caused by shift towards 0.5 is not valid and an alternative explanation must be found.

3.3 Binary model, correlated variables[7]

In a further experiment, models with 2 and 1 coefficients were estimated from simulated data drawn in the same way as in the first experiment (i.e. without a constant), but with *positively correlated* x values, constructed so that the correlation between the variables was $1/\sqrt{2}$ (= 0.707..). The results from the estimation are shown in Table 3.

6 Because of the constant in this model, these totals are 'predicted' exactly.
7 I am grateful to Stephane Hess for suggesting this test.

Table 5: Third Simulation Results

Variable	C2	C1	Ratio
β_1	1.0005	1.3174	0.7591
β_1 standard error	0.0036	0.0027	1.33
β_2	0.9955	0	n/a
β_2 standard error	0.0029	n/a	n/a

The key point here is that the coefficient in the reduced model is *larger* than that in the better model, despite the increase in error. That is, reducing the explanation given by the model does not always reduce the coefficient. This is quite an intuitive result, as x_1 takes on part of the explanation given in the better model by x_2, an effect which in this case outweighs the scale correction.

We may also note that the standard errors of estimation in this model are larger than those in the first model, which follows from the correlation of the variables, reducing the amount of information in the data concerning the values of the coefficients β. The correlation between the coefficient estimates is -0.215, not positive as it was in the uncorrelated first simulation.

The results of a sensitivity test, again increasing x_1 by 0.5, are shown in Table 6.

Table 6: Sensitivity Results from Third Simulation

Choice of alternative 1	C2	C1	Ratio
Predicted base	500594	500466	1.0002
Predicted in test	562934	593701	0.9482
Proportional change	0.1245	0.1863	0.6683

Here we see that in these circumstances the model sensitivity is by no means stable to the removal of an explanatory variable. The sensitivity in C1 increases by *more* than the change in the ratio of the relevant coefficient would indicate. It seems that the mechanism that usually operates to increase the sensitivity of the simplified model is operating here to increase the bias.

3.4 Tree logit model

In a final experiment, the effect of correlation between the error terms ε, rather than between the data x was investigated. This was done by setting up a model of the following form.

$$U_1 = \beta x_1 + \varepsilon_1$$
$$U_2 = \beta x_2 + \varepsilon_2$$
$$U_3 = \beta x_3 + \varepsilon_3$$

In this model, the x variables were independently distributed, again with standard logistic distribution; β had the true value 1. However, ε was distributed standard Gumbel (i.e. with variance $\pi^2/6$) and ε_1 and ε_2 had correlation 0.25, so that estimating a tree (nested) logit model would give a nesting coefficient θ of value 0.5 (the square root of the correlation). This model was estimated based on one million simulated choices, as before, model T1 taking account of ignoring the correlation in the ε's and model T0 ignoring it. The results are shown in Table 7.

Table 7: Fourth Simulation Results

Variable	T1	T0	Ratio
β	0.9981	0.8159	1.2233
β standard error	0.0017	0.0013	1..31
θ	0.5015	1	n/a
θ standard error	0.0015	n/a	n/a

The estimation results show that, as in the previous cases with uncorrelated data where effects were ignored, the coefficient β is reduced. However, in this case the cause and effect of this change are rather different, as can be seen from the model sensitivity, illustrated in Table 8. Again, 0.5 has been added to x_1 to simulate a forecast.

Table 8: Sensitivity Results from Fourth Simulation

Predictions..	T1	T0	Ratio
Alt. 2 base	298426	333671	0.8944
Alt. 2 predicted in test	265196	302797	0.8778
Alt. 2 change	-0.1114	-0.0925	0.8303
Alt. 3 base	403394	333338	1.2102
Alt. 3 predicted in test	383369	302472	1.2675
Alt. 3 change	-0.0496	-0.0926	0.5356

Here we see the typical effects that would be expected from the omission of a tree structure that is present in the data. In model T0, the impact of the test is to reduce equally the numbers choosing alternatives 2 and 3, as is typical in a multinomial logit model. However, when the tree structure is correctly included, as in model T1, the reduction of demand for alternative 3 is nearly halved, corresponding well to the true value of the tree coefficient, 0.5; alternative 2 loses more, as would be expected because the coefficient β is larger in this model. The results are slightly confused by the differences in the base situation, which results from the simplicity of the model in having no constant to get the base shares correct.

3.5 Summary of experimental results

The first two experiments demonstrate clearly the existence of the model quality paradox: removing an orthogonal variable from the model reduces its quality and therefore reduces the value of β. However this reduction does not materially affect the sensitivity of the model. The presence of a constant in experiment 2 is also not material, showing that it is *not* a movement of all the p's towards 0.5 that brings about the result observed.

In contrast, there is nothing paradoxical about the third and fourth experiments.

The third experiment shows that omitting a variable that is correlated with another variable in the model will bias the coefficient of the second variable and hence the sensitivity of the model. If the correlation of the variables is positive the coefficient of the other variable may increase, because the scale reduction given by loss of quality may be insufficient to overcome the bias. When the coefficient is biased upwards, the quality scale effect operates to reinforce the bias when model sensitivity is calculated.

The fourth experiment shows that reducing model quality by omitting structural effects is *not* corrected by analogous effects to those that operate on sensitivity when variables are omitted.

4. EXPLANATION OF THE PARADOX

Among the simple experiments described in the previous section, the first experiment demonstrates the paradox clearly and simply and so gives a suitable context for trying to explain it.

4.1 Variance of choice probabilities

In 2.2 we defined the sensitivity function ζ_{jk}, for an alternative j with respect to the utility of another alternative k. The expectation of this function can be estimated from the sample as

$$E(\zeta_{jk}) = \sum_i w_i \frac{\partial p_{ij}}{\partial V_k} \bigg/ W = E\left(\frac{\partial p_j}{\partial V_k}\right)$$

For a binary logit model, and focussing on the own-effect, the calculation can be worked out further, as was done by Cramer (2007),

$$E(\zeta) = E\left(\frac{\partial p}{\partial V}\right) = E(p(1-p))$$

and this can be worked out further

$$E(\zeta) = \overline{p}(1-\overline{p}) - \mathrm{var}(p)$$

where $\overline{p}$ is the mean value of p.

The sensitivity of the model applied to a population is less than the sensitivity of individuals at the mean value of p. This was a well-known effect in early discrete choice modelling (e.g. Ben-Akiva and Lerman, 1985, Ch. 6), but its application in the present context seems to have been overlooked. Moreover, the magnitude of the effect does not seem to have been estimated previously.

Table 9 presents a number of statistics relevant to the calculation of ζ in experiment 1 and taken from the data generated for that experiment. The first three rows present calculations of the components of $E(\zeta)$. In model M2, the variance of p is about twice as much as that in M1, so that $E(\zeta)$ is substantially lower. However, in the fourth and fifth rows we calculate the expected response to a change in x_1, finding that the difference caused by the coefficient values and the difference caused by $Var(p)$ almost exactly cancel out. This is the explanation of the fact that M2 and M1 showed almost identical proportional change in the experiment, as repeated in row 6 of the table; row 6 also shows that ζ gives a reasonable representation of the change in demand to be expected from a change of 0.5 in x_1.

Table 9: Results on sensitivity from experiment 1

Variable		M2	M1	Ratio
$\bar{p}(1-\bar{p})$	1	0.25	0.25	1
sample variance of p	2	0.1164	0.0537	2.168
$\zeta = \bar{p}(1-\bar{p}) - Var(p)$	3	0.1336	0.1963	0.681
β_1	4*	0.9954	0.6782	1.468
$0.5 * \beta_1 * \zeta$	5	0.0665	0.0665	0.999
Demand change	6*	0.0661	0.0661	0.998
$Var(V)$	7	6.546	2.231	2.934
$Var(V)/Var(p)$	8	0.0178	0.0240	0.742

* Information copied from Table 2.

Thus we conclude that overall model sensitivity depends on the variance of the predicted probability.

In turn, the variance in p depends on the variance in V, which is shown in row 7 of the table. As a first-order approximation, the variance of p would be directly proportional to the variance of V, with a proportionality equal to $(p(1-p))^2$, i.e. 0.0625. However, because of the large range of V in M2, this relationship does not hold, as shown in the final row of the table, and even with the smaller variance in M1 the approximation is poor. Tests with more limited variance in V have shown the relationship to be accurate for those values, however. This dependence of $Var(p)$ on $Var(V)$ is important, because the variance of V will always increase when additional significant variables are included in the model.

For models other than binary logit, the calculations for ζ are of course more complicated but the principles are the same. When the formula for $\partial p_j / \partial V_k$ contains terms in higher powers of p, we may need to refer to the skewness or other moments of p, so that the calculation can become quite complicated.

However, the principles of the mechanism are established:

- model sensitivity ζ depends on the distribution of p in the population;
- for logit models the calculations are quite straightforward and explain completely the results observed in the first experiment;

- the relationship between variance in p and variance in the underlying measured utility V is not clear when the variance is large, but it is obvious that variance in p increases when variance in V increases.

Thus, when a model is reduced in quality by the elimination of a variable, *providing that the change does not bias the coefficients of the remaining variables*, no change in sensitivity is expected. Conversely, improving the model would not affect sensitivity either. A similar effect would also be expected when models are changed in quality by increasing or reducing the accuracy with which variables are measured, *again assuming that these changes introduce no biases*.

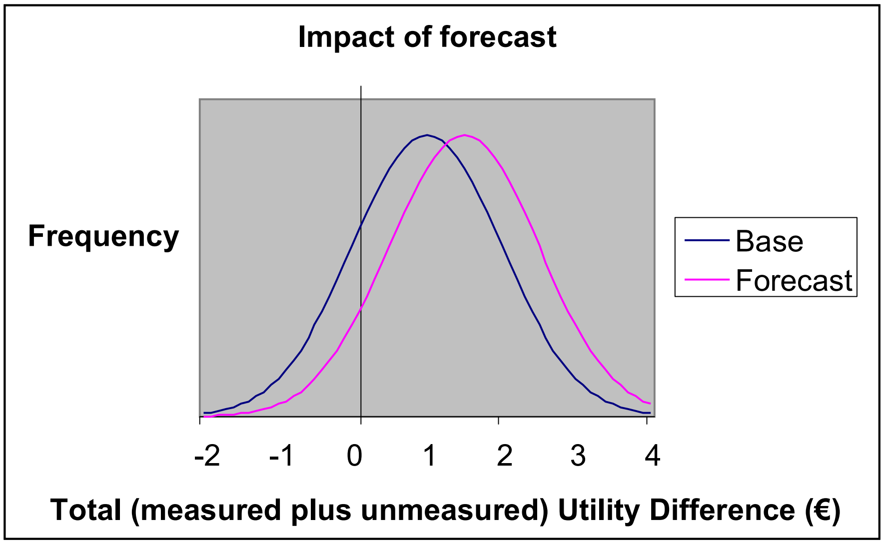
An example of change in model quality of this type would be to change the level of spatial disaggregation at which the model operates (cf. Daly and Ortúzar, 1990). In consequence, when the level of aggregation of the model is changed, β values need to be adjusted to match the quality of the model, increasing when more detailed measurements are made (e.g. smaller zones) or when additional variables are added, *so that* model sensitivity is not affected. When a model is re-estimated with the more detailed data, this adjustment will be made automatically.

A further point is that model sensitivity *will* change when applied with unadjusted β's in circumstances where the distribution of p is different. When the variance of p is high, model sensitivity will be reduced. This should not be seen as a failing, rather a model whose sensitivity changes in that way is responding correctly to predict the behaviour of a population whose circumstances vary. This finding may also be seen as further justification of the practice of re-estimating the scale, as well as the constants, when a model is transferred from one area to another, where the variance of p may well be different. It cannot be expected that coefficients that are applicable in one area can simply be transferred unadjusted to another area.

4.2 Intuitive explanation

The previous section gives an explanation of the paradox in quantitative terms and in the context of the first experiment. A further interpretation can also be given in a more intuitive way.

The figure shows the distribution of the total utility difference (in Euro, for example) between two alternatives before and after the application of a policy which improves the utility of one alternative by €0.50. The fraction of the population with a positive utility difference, i.e. choosing the better alternative, increases because the fraction of the distribution that is positive increases.



The graph is not related to a choice model but simply describes the distribution of population preferences. To set up a choice model, two steps are required:

- some aspects of the utility experienced by travellers are measured;
- a model is estimated, which means
    a. determining the relative values of the measured aspects and
    b. scaling the coefficients of the measured aspects so that the error between the measured and total utility difference has the required variance, i.e. for a binary logit model, $\pi^2/3$.

Providing there is no bias (that is, that the measured and unmeasured variables are not correlated), and assuming that the cost of the alternatives was measured, the choice modeller could then draw a graph like the one above and predict the increase in demand for one alternative when its utility advantage over the other alternative was increased by €0.50. The graph could be drawn with a scale either in Euro as in the figure or with the scale given by the logit standard deviation. The prediction of change in demand would not be different.

If the model were improved, e.g. by the addition of a measure of reliability, then, *providing this did not introduce bias*, the relative values of the other aspects would be unchanged but the error (measured in Euro) would be reduced so that the model scale would have to increase to get back to the $\pi/\sqrt{3}$ standard deviation required. However, the graphs would still look the same, whether drawn in Euro or in the new model scale, and the predicted demand change would be the same. Of course, improving the model will mean that more reliable estimates can be made of the coefficients, and that forecasts for various segments of the market will be different, as well as giving the model the ability to respond to changes in the additional variable.

In summary, the key issue is that there is an overall variance in utility among the population, and how much of this the model captures makes no difference to the overall model predictions (*providing bias is avoided*). The objective of model improvement is then to allow the incorporation of further effects, to reduce bias and to improve the reliability of the estimates.

5. CONCLUSIONS

The objective of this paper is to investigate and explain the paradox that changes in model scale can take place without affecting the sensitivity of the model. Without this effect, it would appear that model sensitivity would be increased by the improvement of a model, which (unless bias exists before and/or after the improvement) always increases the model scale.

A measure ζ for model sensitivity was defined that is dimensionless, independent of the specific variable considered and avoids some of the difficulties arising in the use of elasticity.

The issue of the dependence on model quality of model scale but not model sensitivity was discussed and illustrated in a recent paper by Cramer (2007). This work shows that the result does not depend on the distributional form of variables included or omitted. However, Cramer does not offer an explanation of the effect.

Experiments were conducted to illustrate and quantify the effect. A very simple experiment showed that the omission or inclusion of a variable in a model did indeed change the scale when the model was estimated, but that this had no impact on the sensitivity of the model. A second experiment showed that the inclusion of a constant in the model, so that a scale reduction does not move all the probabilities towards 0.5, did not affect the stability of sensitivity. However, a third experiment illustrated that, with correlated variables, the omission of one variable may bias the coefficient of another, to the extent that its coefficient increases, and this bias then does allow the model sensitivity to change. A final experiment showed that simplifying a model by ignoring correlation in the errors can change the model sensitivity.

The explanation of the paradox offered in this paper is that model sensitivity depends on the variance in the predicted probabilities among the population. If that variance increases, as it will when more variables are measured, then model sensitivity is reduced. The impact of this dependence was shown to explain the effect observed in the first experiment. In the absence of bias, the magnitudes of the effects of variance change and coefficient change are almost exactly equal and opposite . An intuitive explanation was offered with the idea that the variation of utility in the population exists independently of choice modelling, so that a change in utility will bring about a change in demand, which an unbiased choice model of any quality will expect to predict correctly.

These findings are important for the adjustment and transfer of models in a number of contexts. For example, model transfer between areas or over time needs to take account of the variation of predicted probabilities. It is not possible to prescribe a single model scale that will apply in all circumstances. Similarly, the adjustment of a model developed from (say) Stated Preference data to a context of Revealed Preferences will require a scale change, which may have the effect of maintaining rather than changing the sensitivity of the model.

6. REFERENCES

Ben-Akiva, M. and Lerman, S. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Mass..

Cramer, J. S. (2007) Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-specified Disturbances, Oxford Bulletin of Economics and Statistics 69 (4) , 545–555.

Daly, A. J. and Ortúzar, J. de D. (1990) Forecasting and Data Aggregation: Theory and Practice, *Traffic Engineering and Control*, Vol. 31, No. 12, December.

Wooldridge, J. M. (2002) *Econometric analysis of cross section and panel data*, MIT Press, Cambridge, Massachusetts.