Deriving welfare measures from discrete choice experiments: a response to Ryan and Santos Silva

Emily Lancsar\textsuperscript{a,∗} and Elizabeth Savage\textsuperscript{b}
\textsuperscript{a}Business School (Economics), Centre for Health Services Research, University of Newcastle upon Tyne, UK
\textsuperscript{b}CHERE, University of Technology, Sydney, Australia

Summary

In this response we start by highlighting the key area of agreement between the commentaries and our original paper: if there is uncertainty regarding which alternative will be chosen, in a DCE or in the real world, then the compensating variation as modified for discrete data by Small and Rosen is the appropriate method of deriving welfare measures from DCEs. Both commentators point out circumstances in which the method traditionally used in the health economics arena may be consistent with the compensating variation. We show that these circumstances require a number of potentially unrealistic and ad hoc assumptions, and argue that using the traditional method could produce erroneous welfare estimates if these assumptions fail to hold in practice. We show that the compensating variation method can accommodate each of the special cases raised by the commentators and therefore is the more general and appropriate approach to deriving welfare measures from DCEs. We also respond to issues raised regarding the estimation of DCEs in general and our application to asthma medication in particular. Copyright © 2004 John Wiley & Sons, Ltd.

Keywords welfare measurement; compensating variation; willingness to pay; discrete choice experiments; cost benefit analysis

Introduction

In this rejoinder, we start by highlighting the key area of agreement between Lancsar and Savage [Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. Health Econ, this issue], Ryan [Deriving welfare measures in discrete choice experiments: a comment to Lancsar and Savage (1). Health Econ, this issue], and Santos Silva [Deriving welfare measures from discrete choice experiments: a comment to Lancsar and Savage (2). Health Econ, this issue]. Namely, if there is uncertainty regarding which alternative will be chosen, in the experiment or in the real world, then the compensating variation (CV) as modified for discrete data by Small and Rosen [1] is the appropriate method for deriving welfare measures from discrete choice experiments (DCEs).

The commentators raise a number of cases in which the method traditionally used in the health economics literature to derive measures of willingness to pay (WTP) from DCEs can be forced to be consistent with the Small and Rosen method. We review each case highlighting the restrictive assumptions that must be made for the former to be consistent with the latter and question how realistic these assumptions are in practice. We also
show that the CV method can accommodate each of the special circumstances raised by the commentors and as such is the more general approach to deriving welfare measures from DCEs.

One of the key cases highlighted by both commentators is when we know with certainty which alternative will be chosen. We agree that in this case the traditional method may be consistent with the CV. Where we disagree is with regard to how relevant this is in practice. In this context, we draw a distinction between constructed and real markets, raising the important issue, overlooked by both commentators, that even if there is only one alternative on offer in the real world, individuals can choose not to consume it. Hence, there is uncertainty regarding what will be chosen in the experiment and the real market even if this relates only to whether or not individuals will consume the product on offer.

We draw a distinction between marginal and total welfare measures and advocate the latter if the purpose of such welfare measures is as an estimate of benefit in cost benefit analyses (CBAs). Finally, we comment on the estimation of DCEs and offer some concluding comments.

Taking account of uncertainty

The statements made in both commentaries suggest they agree with our key conclusion that if there is uncertainty regarding which alternative will be chosen then the probability of choosing each alternative must be taken into account when calculating welfare measures from DCEs. Indeed Ryan (this issue) states: ‘if there are multiple options, estimation of aggregate welfare must involve the expected value arising from each option [where]... the expected value takes account of the change in utility, weighted by the probability of choosing that option’ (Ryan, this issue, p. 1). The CV formula does this. Similarly, Santos Silva (this issue) states: ‘in a heterogeneous population, the expected compensating variation required by a change in $X_{ij}$ depends on the proportion of individuals that choose each alternative before and after the change in this characteristic’ (Santos Silva, this issue, p. 5) which is also taken into account by the CV formula. In contrast, the method traditionally used in the health economics literature does not do this and by implication is not necessarily consistent with welfare and random utility theory (RUT) in the contexts in which it has been applied.

Revisiting the case of consistency between welfare measures

As explained in Lancsar and Savage (this issue), the traditional method is consistent with the Small and Rosen CV when: (1) there are only two alternatives on offer and the approximation in Equation (8) in Lancsar and Savage holds and (2) if we know with certainty which alternative will be chosen. To this list Santos Silva adds (3) the case of an infinitesimal change in the probability of choosing a given alternative and (4) the case in which the attributes of the choice alternatives change in such a way as to leave the choice probabilities unchanged.

(1) Binary case plus an approximation

The CV formula and the WTP formula traditionally used in the health economics literature can be forced to be consistent in the binary choice case if we make the following approximation:

$$\ln(1 + e^{V_i^j}) \approx V_i^j$$

(1)

See Lancsar [2] for the relevant proof. This approximation does not hold if there are more than two choice alternatives. Furthermore, as we demonstrated in the empirical application in our original paper, it may not hold even in the binary case and the two methods can produce vastly different results.

Santos Silva appears to be confused regarding our explanation of when the CV and the standard WTP equations can be made to be consistent, stating incorrectly that we claim this is when there is a binary choice and the probability of choosing a given alternative is equal to 1. The approximation in Equation (1), which replicates Equation (8) in our original paper, is not equivalent to setting the probability of choosing a given alternative equal to 1. Assuming that the probability is equal to 1 is equivalent to simply not putting the other $J - 1$ alternatives into the CV equation.
The certainty case: constructed compared to hypothetical markets and choosing not to consume

Both commentaries highlight an important point raised in Lancsar and Savage (this issue), and in more detail in Lancsar [2], that if we know a priori which alternative will be chosen or there is only one option and it is consumed, then the WTP method is consistent with, or a special case of, the CV method. Whether this holds in practice is open to debate. Indeed it is with regard to how relevant this certainty case is in real world applications that we disagree with the commentators.

In thinking about the relevance of this certainty case it is important to draw a distinction between markets constructed in DCEs and real world markets. A DCE essentially constructs a hypothetical market in which respondents are presented with a series of choice sets containing two or more alternatives and are asked to choose their preferred alternative from each set. A priori there is always uncertainty about which alternative will be chosen; we do not know which will be chosen, we only know the probability with which each is chosen which is why we model this choice process in a RUT framework. It is never the case in DCEs that respondents are presented with only one option; the choice is between different treatments or at a minimum between treatment and non-treatment, that is, allowing respondents to choose not to consume. Indeed, there would be no point in modelling ‘choice’ where there is no choice. As such, the welfare measures derived from DCEs must take account of the inherent uncertainty in such constructed markets by taking account of the probability of choosing each alternative in the choice set.

When we come to apply the model from the constructed to the real market, the next issue to address is whether there is uncertainty about which alternative will be chosen in the real market or, more specifically, whether more than one option is relevant in the real world? If more than one alternative is available in the real world then the welfare measures derived from the parameter results of DCEs should be calculated using the Small and Rosen CV.

For example, the empirical application outlined in our paper calculated measures of welfare for preventive asthma medication (both inhalers and tablets). In the real world several preventive medications are available from which asthmatics generally choose one. This implies there is uncertainty even in the real market regarding which asthma medication will be chosen and again this uncertainty must be taken into account when eliciting welfare measures. As such, the CV method is appropriate in both the experiment and the real world context.

Ryan (this issue) suggests that state of the world models in which only one alternative is on offer are the norm when patients make real world choices about health care but offers no evidence to support this. The choice of asthma medication, the example in our paper, is clearly not a state of the world model. Similarly, all 14 studies which derived MWTP or WTP in the health economics literature reviewed in Lancsar and Savage (this issue) can be considered to have uncertainty surrounding which alternative would be chosen in the real world context, either because there is more than one alternative on offer in the real world or because even when there is only one alternative available, it must be acknowledged that individuals can choose not to consume the good or service in the real market.

Even if there is only one alternative on offer in the real world we cannot assume, as Ryan states, that ‘individuals therefore take up the service/drug with certainty’ (Ryan, this issue, p. 1). This ignores the fact that individuals can choose to delay consumption, substitute to an imperfect alternative or simply not consume at all. The decision made by parents regarding whether to have their children immunised provides a case in point. Even if there is only one vaccine available for a particular disease, parents can, and in some cases do, choose not to immunise their children. At the extreme end of the spectrum there are even examples of individuals choosing not to have potentially life saving treatment [3], the point being that we cannot simply assume that individuals will consume a good or service just because it is the only option on offer.

Thus Ryan’s (this issue) statement that the ‘current methods for estimating WTP are consistent with random utility theory and welfare theory’ should be followed by the caveat ‘in the case where the individual consumes the product or program with certainty’ and even then it is difficult to know for certain what individuals would choose in the real world without access to revealed preference (RP) data.
(3) Infinitesimal change

Santos Silva suggests that if there is an infinitesimal change in one of the attributes the choice probability can essentially be treated as constant and the traditional formula is appropriate. We agree that in this case the traditional method will approximate the CV. Indeed, this point is outlined in Lancsar [2,4]. However, it raises two questions: how relevant are infinitesimal changes in practice and why use an approximation when the CV formula is so tractable?

(4) Constant choice probabilities

Santos Silva suggests that the product of the change in the level of an attribute times the MWTP (which when summed over all changes in the attributes is the traditional method of calculating WTP from DCEs in the health economics literature) will be consistent with the Small and Rosen method when variations in the attributes leave the choice probabilities unchanged. While this is a theoretical possibility, it is far from realistic to assume that the choice probabilities remain unchanged following a change in the attributes of a good or service. Changing the levels of any of the attributes can change the level of utility derived from each alternative which, by definition, changes the probability with which each option will be chosen, both of which impact on the resulting change in welfare.

For example, consider a consequence of taking a particular asthma medication that causes deterioration of a physical activity attribute from being able to participate in all sporting/strenuous activities to being able to participate in only a restrictive range. It is highly unlikely that this would have no effect on the probability of choosing this particular medication. Instead, it may induce patients to substitute to an alternative medication in the experiment or in practice.

Santos Silva also suggests that the probability of choosing each of the alternatives will remain unchanged if a single attribute changes by the same amount across all the choice alternatives. This means if, for example, the attribute ‘ability to participate in sporting/strenuous activities’ changed by the same amount across all asthma medications, then the traditional method of calculating WTP from DCEs in the health economics literature would be consistent with the CV. While this may be correct if attributes interact in the same way across alternatives, it is a restrictive assumption which begs the question: why would this attribute change by the same amount across all asthma medications? Again, the Small and Rosen method is less restrictive and more appropriate.

Total rather than marginal welfare measures

Santos Silva (this issue) focused on MWTP for a single attribute to the exclusion of WTP for an entire good or service and suggested that the former is related to the Small and Rosen method. While it may be related, by his own admission Santos Silva states that ‘the MWTP is not appropriate to evaluate the total welfare change resulting from discrete changes in the characteristics of the alternatives’ (Santos Silva, this issue, p. 4). Furthermore, if the purpose of welfare measurement is to use such measures of benefit in a CBA of entire products or programs, then the total, rather than the marginal, change in welfare is most relevant.

Estimation issues

As Santos Silva (this issue) correctly points out, the empirical application is not the main focus of our paper. The DCE data used in the paper were simply a vehicle to demonstrate the two methods and to elucidate the point that the WTP method is generally not consistent with the CV and they give different results in practice. We could equally have used simulated data because the purpose of the empirical application was neither to discuss the estimation of DCEs per se nor to draw policy conclusions with regard to asthma medication.

However, Santos Silva (this issue) raised a number of issues with regard to the estimation of DCEs in general and our application to asthma medication in particular. First is the charge that welfare measures cannot be calculated from conditional fixed effects logit models because such models do not allow the estimation of the probability that an individual will choose a given alternative. This is not entirely correct. In fixed
effects models differences in predicted probabilities across individuals arise from two sources, differences in attribute levels and differences in individual covariates. Santos Silva suggests that the fixed effect estimator cannot be used to identify parameters associated with regressors that do not vary across alternatives. This relates to individual covariates such as income, age, gender, etc. However, this has been addressed in the DCE literature by interacting such variables with dummy variables created for the choice alternatives or with individual attributes [5]. In the current example, we did not include individual covariates because our focus was not on predicting individual probabilities, but rather on determining the probability with which each alternative will be chosen for the representative respondent in the population in our experiment.

Santos Silva (this issue) raised the issue of the five individuals who always chose the current medication. Having a set of individuals who always choose one alternative is not uncommon in DCEs. For example, a third of the sample chose the same alternative in every scenario in Hall et al [6] who suggest this may be an accurate reflection of patient preferences or respondents may simply not be willing to trade over the attribute range presented. Furthermore, as discussed in the review of the use of DCEs in the health economics literature by Viney et al. [7], RUT allows such apparently 'irrational' responses due to the random component which captures, among other things, 'errors in the perception and optimization by the consumer' [8]. Deleting such observations may yield precise estimates of effects for attributes for respondents who are willing to trade, but would impact on the orthogonality of the design and the results would suffer from sample selection bias.

The issue of recalibration of the results from DCEs is an important one which to date, to the authors’ knowledge, has neither been undertaken in the DCEs reported in the health economics literature nor is common in the wider DCE literature. Recalibration of ASCs using RP data is suggested when the models are used to predict out of sample or for forecasting [5,9]. The purpose is to account for the confounding of the parameter estimates with the variance of the error term. If the MRS is used, the error variances cancel out and rescaling is not required. When using the Small and Rosen method rescaling should be considered if RP data are available. However, when such data are not available, rescaling is an ad hoc procedure with no empirical basis and is therefore not to be recommended and more research is needed on the appropriate rescaling techniques for such situations.

It is important to end this section by noting that the welfare measures estimated in the empirical example in Lancsar and Savage (this issue) were not sensitive to changing the estimation of the discrete choice model from a conditional fixed effects logit to a straight logit which did not take account of the panel nature of the data, in which case the resulting WTP was still 13 times larger than the CV.

Conclusion

Prior to Lancsar and Savage (this issue), the topic of our paper, namely how to measure welfare changes from DCEs in a theoretically consistent manner, had not been raised in the health economics literature, nor had the Small and Rosen method been demonstrated in an empirical application of a DCE in a health setting. While the topic has clearly caused debate, we are glad we have stimulated such interest in this area.

As the preceding discussion has hopefully made clear, the Small and Rosen CV is the appropriate method for deriving welfare measures from DCEs because it: (a) is derived using microeconomic welfare theory and is consistent with RUT; (b) is more general than the current method used in the health economics literature and can accommodate each of the special circumstances raised by the commentators; and (c) is not difficult to compute in practice.

The traditional method is generally not consistent with the Small and Rosen method; however the commentators have gone to great lengths to point out circumstances in which it may be consistent and may therefore be appropriate. However, these circumstances require a number of potentially unrealistic and ad hoc assumptions. Using the traditional method is likely to produce erroneous welfare estimates if the restrictive assumptions discussed here fail to hold in practice.

Acknowledgements

Emily Lancsar is funded by the Health Foundation. The authors are grateful for the advice and valuable
comments provided by Jordan Louviere and Cam Donaldson.

References