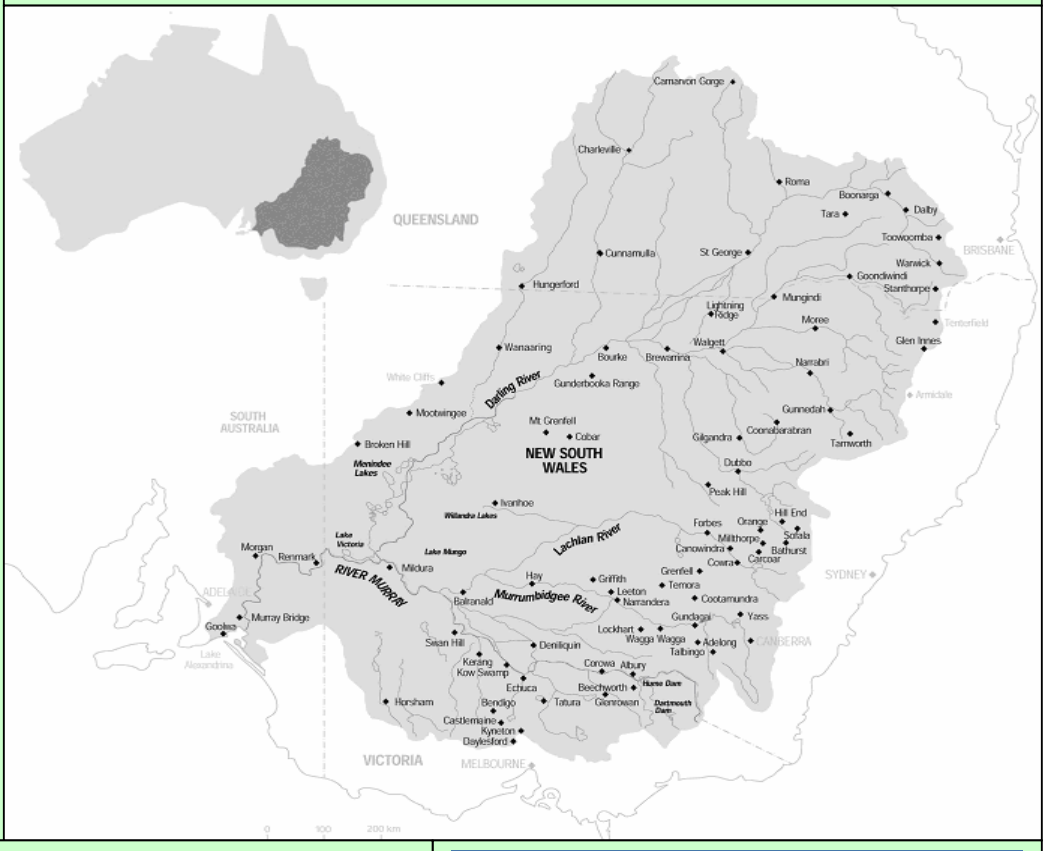
Risk & Sustainable Management Group

Murray Darling Program Working Paper: M07#1

Choice Modelling and laboratory experiments for non-market valuation: a framework

Giovanni B. Concu

Research supported by an Australian Research Council Federation Fellowship
http://www.arc.gov.au/grant_programs/discovery_federation.htm



Schools of Economics and Political Science
University of Queensland
Brisbane, 4072
rsmg@uq.edu.au

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Choice Modelling and laboratory experiments for non-market valuation: a framework.

Giovanni Concu
Risk and Sustainable Management Group
School of Economics
University of Queensland
Brisbane, 4072 QLD

Email: g.concu@uq.edu.au

Abstract
The willingness to pay (WTP)/willingness to accept (WTA) disparity raises serious questions about preference elicitation techniques based on the Hicksian model of decision-making. In this paper we investigate the possibility of incorporating the strategies suggested by Plott and Zeiler (2005) in a Pivot Process mechanism and in a Choice Modelling experiment to eliminate the WTP/WTA disparity. The goal is to improve the reliability of the benefit estimates by comparing and contrasting these two methodologies. No methodological and analytical obstacles prevent the use the Pivot Process to calibrate the Choice Modelling estimates. By combining the two methods, a calibration procedure can be usefully developed to validate good hypothetical surveys and correct bad ones.

Keywords: incentive-compatible, stated preferences, pivot process, WTP/WTA disparity

Introduction

Theory and experiments have been widely used to analyse the disparity between willingness to pay (WTP) and willingness to accept (WTA). The gap is often interpreted as an "endowment effect" where the initial endowment affects the rate of exchange between goods. Prospect theory states that endowment effects and loss aversion—people commonly value losses much more than commensurate gains—describe a fundamental feature of human preferences (Tversky and Kahnemann 1991). The neoclassical model of decision making, on the contrary, assumes that preferences do not depend on the current assets and that the initial entitlement does not change the final allocation. The WTP/WTA disparity is then taken as evidence that individuals do not have Hicksian preferences. Many experimenters have observed significant differences between WTA and WTP measures (see Horowitz and McConnell 2002); many others have not (see Plott and Zeiler 2005). Hence, the observed disparity could be the result of weak methods for preference elicitations. That is, in some experimental settings individuals may not reveal their true preferences.

The existence and interpretations of the WTP/WTA disparity have important implications for the design and implementation of environmental policies. WTA and WTP imply different property right regimes. If the two measures diverge, environmental policies could be substantially altered by the assignment of property rights. For instance, assume the public's WTP for native vegetation conservation is half the WTA compensations for loss of native plants. Also assume developers would willingly buy or sell land at the prevailing market prices. The amount of native vegetation protected when rights are assigned to developers and had to be purchased *by* the public would be less than the amount protected if the developers had to purchase rights *from* the public (Horowitz and McConnell 2003).

The WTP/WTA anomaly raises some serious questions about preference elicitation techniques based on the neoclassical model of preference. An alternative theory—the "Discovered Preferences Hypothesis" (see Braga and Starmer 2005)— claims some anomalies are errors in stated preference that would disappear in environments that allow learning. Recent laboratory experiments have indeed showed how a set of

design strategies can control for "subjects' misconceptions" and eliminate the WTA/WTP gap (Plott and Zeiler 2005). This result questions the interpretation of observed gaps as evidence of loss aversion. Hence, the Discovered Preference Hypothesis offers hopes in another important direction. If laboratory experiments of the type performed by Plott and Zeiler (2005) are anomaly-free, they could provide an important empirical point of reference to assess benefit estimates obtained with Stated Preference (SP) methods. In other words, they may offer a constructive basis for mitigating or calibrating WTA/WTP disparity, or to assess innovations in elicitation methods that aim to correct for errors.

In order to assess the use of laboratory experiments to calibrate SP estimates, this paper compares and contrasts a SP method and a "direct revelation mechanism" for preference elicitation. The aim is to discuss the theoretical and analytical properties of the Choice Modelling (CM) technique and the Pivot Process (PP) mechanism. CM is a SP technique that generates data from observations of behaviour in hypothetical markets. PP also uses hypothetical markets but in conjunction with a mechanism that turns the economic commitments into real economic consequences. Comparing the two elicitation methods highlights analytical differences that could cause estimate to differ. It also provides for the differences to be resolved either through design strategies or econometric procedures. The paper is organised as follow. Section 1 contains a review of anomalies in SP and the way they have been addressed. In section 2, the properties of the PP are discussed. The CM method is introduced in section 3. Section 4 contains a set of methodological proposals for combining PP and CM. Section 5 concludes.

1. Anomalies in SP methods.

Stated preference techniques such as Contingent Valuation (CV) and Choice Modelling (CM) are the only available methods for the estimation of 'passive use' or 'non-use values'. These techniques are based on the creation of *hypothetical* markets for the exchange of public goods. The transactions in hypothetical markets generate a set of data which is then used to estimate the benefits of environmental changes. SP techniques are often regarded with scepticism because they are troubled with *anomalies*–i.e. systematic inconsistencies between the theory used to organise the data collection and interpretation, and the pattern of individuals' responses. The disparity

between WTP and WTA is the first anomaly. While the Hicksian theory does not imply WTP=WTA (Hanemann 1991), Sudgen (1999) convincingly argues that the difference should be small. For an individual with Hicksian preferences, if the WTP is only a small fraction of income—as is expected for environmental goods— the gap between WTP and WTA should be in the order of few percentage points (Sudgen 1999, p.159). Horowitz and McConnell (2002) review 45 WTP/WTA studies—for a total of 201 experiments—and find that WTA is about seven times higher than WTP. Theory and estimates are clearly at odds.

A second anomaly in SP studies is the insufficient sensitivity of individual's valuations to changes in the quantity of the goods. A third common anomaly is the influence of irrelevant clues on respondents. Sudgen (2005) interprets these two inconsistencies as the consequences of respondents' use of mental processes that do not take proper account of quantities but seek for clues to come out with an answer. There is indeed an extensive literature on the use of heuristics or rule of thumbs in decision making (Gigerenzer and Selten 1999).

In order to improve the reliability of elicitation methods, hypothetical data have been enriched with observations obtained with revealed preference methods. This approach combines observations from real but surrogate markets and hypothetical markets. Adamowicz et al. (1994) pioneered this approach in the field of environmental valuation. They found sufficient evidence to claim that revealed and SP data contain a similar preference structure. They also cannot reject the hypothesis that the underlying variance of the two datasets is different. In other words, while the parameter point estimates are not statistically different, stated preference estimates have larger variance.

Controlled, theory-driven experiments in laboratory settings have also been designed to assess the magnitude of the hypothetical bias. Experiments are typically set up as direct revelation mechanisms in which individuals have a dominant strategy to reveal their true valuation of the good. In SP methods, individual responses to valuation questions do not entail a real economic commitment or real economic consequences. On the contrary, in experimental valuation, the revelation mechanism entails a real

economic commitment, or consequence, or both. In three simple Dichotomous Choice (DC)—or take-it-or-leave-it—experiments, Cumming et al. (1995) found that the hypothetical DC experiments did not generate the same responses as the real DC experiments. List and Gallet (2001) reviewed 29 field and laboratory studies. They found that, on average, respondents overstated their preference by a factor of 3 in hypothetical settings. In Balistreri et al. (2001), DC and Open-Ended (OE) question formats were used to estimate WTP for insurance against an environmental hazard with known probability. WTP estimates were then compared with auction values. Hypothetical values systematically overestimated auction values. The OE format had a bias smaller than the DC format. Also Veisten and Navrud (2006) compared hypothetical DC and Open-Ended (OE) hypothetical experiments to Actual Payment (AP) for the provision of an environmental good. Respondents facing DC and AP questions stated lower willingness to pay than respondents confronted with just DC question or OE and AP questions. The gap between stated and actual WTP was lower when respondents face mechanisms that induce truth telling.

Two important developments in experimental economics may offer alternative strategies to investigate anomalies in SP. First, Healy (2006) assessed which mechanism for public good provision generates the most efficient outcome, i.e. the Pareto optimal allocation. Healy compared five mechanisms—voluntary contribution, proportional tax, Groves-Ledyard mechanism, Walker mechanism, and a Vickery-Clarke-Groves (VCG) mechanism. This last one was found to be the most efficient and stable. In repeated rounds the VCG mechanism converged close to the efficient equilibrium with over half the subjects truthfully revealing their preference.

Second, Plott and Zeiler (2005) designed a set of controls to reduce or eliminate "subject misconceptions"—that can loosely define as "confusion"—under the assumption that it is a source of the WTP/WTA disparity. The first control was the use of an incentive compatible mechanism to induce subjects to reveal their true preference for private goods. By telling the truth, subjects increase the probability of gaining the maximum amount possible. Training was the second control strategy. It provides subjects with a basic understanding of the elicitation mechanism. Many incentive-compatible mechanisms are indeed unfamiliar to subjects, even if the task

may appear to be a simple buying or selling task. The third control was a set of practice rounds to allow subject to experience the instrument while gaining familiarity with its properties. In particular, Plott and Zeiler (2005) used *paid* practice rounds so that subjects were immediately exposed to the consequences of their decisions. And finally, the experiments assured anonymity. If decisions are not made anonymously, subjects may be concerned with how others view their bids. The main finding of Plott and Zeiler's experiments is that no WTP/WTA disparity is observed. Their primary conclusion is that the observed WTP/WTA gaps do not reflect a fundamental feature of human preferences.

These results suggest an avenue for further research. Is it possible to eliminate the WTP/WTA in an environment for the provision of public goods designed following the results of Healy (2006) and Plott and Zeiler (2005)? How do the WTP and WTA estimates in the laboratory compare with estimates from SP techniques for the same environmental good? Is it possible to calibrate SP estimates using laboratory experiments? However, a cautionary note is needed. It is tempting to consider laboratory experiments as an alternative to SP institutions for the valuation of environmental changes: couldn't researchers use laboratory experiments to obtain *true* values instead of estimating values using expensive surveys? However, direct revelation mechanisms applied to public goods are still a step removed from real markets. They entail real economic commitment and real economic consequences, but do not lead to the provision of a public good. If one were to describe the degree of realism of each evaluation technique along a continuum from hypothetical to real markets, direct revelation mechanisms applied to public goods would sit exactly between state preferences and revealed preference techniques (fig 1). Direct revelation mechanisms have an advantage over revealed preference techniques. The latter can only provide information from actual markets. The former–as well as SP techniques–can provide information on institutions that are fundamentally different from the existing ones. It must also be stressed that eliciting true values in experimental settings is clearly dependent on the definition of the public good, the source of incentive payment, the size of the interview, and so forth. It is also unclear how to extrapolate experiment results to population and other dimensions of the public goods.

Combining the newly developed strategies in laboratory experiments and in SP could provide a better understanding of preference formation and elicitation. The simultaneous use of the experimental and stated preference methods aim to:

a) estimate WTP and WTA benefit measures for the same environmental changes in two institutional contexts. The laboratory experiment and the survey-based data generation mechanism should be designed to offer the same hypothetical market;
b) determine the difference in WTP and WTA estimates within and across institutional environments;
c) assess the possibility of calibrating benefit estimates whenever the WTP and WTA responses do not satisfy Hicksian conditions.

The following sections contain a description of the properties of the PP mechanism and the CM method. The aim is to determine if they can be used to evaluate preference for the same environmental goods, and understand if there are theoretical and methodological obstacles to comparing their respective estimates.

2. Features of the Pivot Process.

Demand revealing mechanisms for public good provision are usually associated with the work of Vickery, Clarke and Groves (see Attiyeh et al. 2000). They were originally developed to overcome undesirable properties such as free-riding commonly occurring in public good decisions. Demand revealing mechanisms —or dominant strategy mechanisms— are incentive compatible in the sense that it is always individually rational to behave truthfully.

There several variants of the demand revealing mechanism. The essence of these institutions/tools is that each person pays (or is paid for) the benefit of her actions, but no one is charged (or credited) as required by budget balance. In these environments, participants are motivated to reveal their preferences through a tax mechanism which rewards truthfully presentation and penalises concealment. Preferences are usually assumed to be quasilinear:

$$u_i(y,x_i)=v(y)+x_i \qquad (1)$$

where $v(y)$ is a strictly concave function of the level of public good provided y and x_i is the bundle of private goods or a "composite commodity" or a "divisible numeraire" (Ledyard, 1994, Healy 2006). Two features of this formulation are important to note. First, the additive structure implies substitutability of the private and public goods. This is necessary to interpersonal comparison of utility. Since everyone has a common currency, the numeraire, and since the numeraire x and the valuation v are summed to form the overall utility, the valuation v must be expressed in terms of the units of the numeraire. This means that everyone's utility units are expressed in the same currency and are therefore comparable. Second, the function is entirely deterministic. Direct revelation mechanisms provide the monetary measure of the utility change associated with a change in the provision of the public good y.

Among the class of demand revealing mechanisms, the pivot process is a version of the Vickery-Clarke-Groves mechanism with a binary public good (Tideman and Tullock 1976). It has been found to be the most stable and efficient among the public good mechanisms (Healy 2006). It also resembles a hypothetical WTP/WTA elicitation procedure. In the pivot mechanism, a public good of fixed description and size will either be provided or will not. Each individual transmits a private bid that could be either positive or negative valuation of changing from the status quo. Bids are all summed up, and the public good is provided if the positive bids are larger than the negative one. However, the pivot process contains a tax system: each individual whose bid changed the outcome (relative to what would have happened without her bid) is required to pay a tax equal to the amount of her bid which was required to switch the outcome. Suppose bids are collected among individual j and k for policy proposal A. They amount to -$20—the policy is rejected. Suppose now that individual i bids $25, so that the total of the bids is $5. The policy is implemented. According to the rules of the pivot mechanism, individual i pays a tax—called Clarke Tax—equal to the amount of her bid used to change the outcome of the policy decision that is, $20. A similar tax is calculated for every other individual. Those whose bids do not change the outcome do not pay tax—technically they are not "pivot" players.

Note that had player i understated her preference by an amount less than $5, she would have paid exactly the same tax. If she has understated her preference by more

than $5, policy A would have been rejected. But player *i* would rather have A at the price of $20—and gain at least $5 ($25-$20)—than renounce A and gain nothing. Similarly, if a player overstates her preference, she either makes no difference in what is selected or what she pays, or she changes the result and pays more for her choice than it is worth to her. The pivot rule give players the choice of leaving the outcome unchanged or changing it at price equal to the reported net loss to the other players. If the value of the alternative is less than the net value to the other players, than it is rational to bid truthfully and leave the outcome unchanged. But it is also rational to bid truthfully when the player's value for the policy is larger than the net value to the other players. Understating values, players may pass the opportunity to obtain the desired outcome at an attractive price.

The Clarke tax is indeed a peculiar tax mechanism. It is important, then, to instruct players of the properties of the game and provide training and time for the subjects to learn the dominant-strategy bidding features of the mechanism. Unless the objective of the game is to understand the propensity of participants to the use of decision rules, instructing respondents about the demand-strategy properties would not change the preference structure. It could also reduce the noise associated with misconception and random bidding (Harrison, 2006, Plott and Zeiler 2005).

3. Features of the Choice Modelling method.

The Choice Modelling (CM) technique has been increasingly applied in environmental valuation (Adamowicz 2004). It is a technique belonging to Conjoint Analysis, a set of experimental tools designed in the early 1960s by mathematical psychologists (McFadden 1986, Mackenzie 1993). CM combines Lancaster's approach to consumer theory with Random Utility Theory (Louviere et al. 2000). Individuals are assumed to choose the alternative that yields the highest utility. Each alternative's utility is represented by a utility function U_i that contains an observable (deterministic) element V_i and a stochastic element ε_i:

$$U_i = V_i + \varepsilon_i \qquad (2)$$

The alternative's characteristics–or attributes–enter the deterministic element of the utility function. An individual will choose alternative *i* if $U_i > U_j$ for all $i \neq j$. Since the stochastic elements are not observed, the analyst can only describe the probability of

choosing *i* as:

$$\Pr[i \quad is \quad chosen] = \Pr[(V_i + \varepsilon_i) > (V_j + \varepsilon_j)] \qquad \forall j \in C \qquad (3)$$

where *C* is the set of all possible alternatives. Probabilities of choice can be computed from (3) once the distribution of the error terms is specified. The deterministic component is usually specified a linear, additive function of the choice attributes (Louviere et al. 2000):

$$V_i = \Sigma_i \beta_{ik} X_{ik} \qquad (4)$$

where β_{ik} is a parameter vector conditional on a matrix of *k* alternative's attributes. In a CM experiment, subjects are presented with several alternatives usually partitioned in choice sets of two or three. What researchers observe in this experimental setting is a series of yes/no answers that indicate which alternative provide the maximum utility. β_{ik} is estimated as the set of parameters that maximise the utility of the chosen alternatives.

The alternatives presented to the subjects are selected from the universe of possible alternatives by a mechanism called design of experiment (Louviere et al. 2000). Consultation with experts, focus group and pilot studies are usually set up with the purposes of identifying the attributes and their levels. Variables that are expected to affect the utility of any alternative but that do not vary across alternatives, such as socio-economic characteristics and distance, have to be interacted with choice specific attributes. The great advantage of the CM technique is the possibility of breaking down the observable element of utility function into explanatory variables that can strategically varied by the researcher. It allows estimating marginal values for each single attribute that enter V_i, testing its significance and evaluating the welfare impacts of policies as different bundles of attributes.

4. Overcoming differences in CM and PP.

The main differences between CM and PP derive from their respective behavioural models and the data they generate. The PP provides a deterministic monetary measure of the utility associated with alternative *i*. The CM gives an estimates of utility of alternative *i* conditional on the structure of the error terms ε and the matrix *X* describing the alternative and individual characteristics. Differences could be minimised by:

(a) Lancasterizing the PP model. This means specifying the same functional form and arguments of the *v(g)* term in PP and *V(X)* in CM. Apart from preserving strictly concavity and quasilinearity, there are no other technical reasons that prevent *v(g)* to use a description of *a policy as a bundle of attributes*. This is the Lancaster's approach used in CM. Because usually the set of attributes in CM contains a monetary attribute, it works as the numeraire as the composite good in the PP utility model. Using the same public good both in CM and PP means that in both institutions there is no actual provision of public good.

(b) Randomising the utility function in PP. Introducing an error term in PP could be justified by the hypothetical nature of the market, i.e. the public goods do not get delivered. This error could be interpreted as idiosyncrasies in preferences associated with the hypothetical nature of the experiment and the real economic incentive. This approach has the advantage of producing estimates of utility parameters that can be compared to those estimated in CM. It would be possible then to test if the underlying preference structures are the same by comparing estimated utility parameters. This approach relies on the availability of a sufficient number of observations in PP for a reliable regression analysis.

(c) Impose the same functional forms on utility functions. The PP use a strictly concave quasi linear function. The CM uses a linear in parameters function. A strictly concave form in CM would require interaction terms–and hence a more complex experimental design that allows second or higher order interactions.

(d) Organise practice rounds and training. The CM is itself organised as a set of choice tasks. Some of these choice tasks should be interpreted as training rounds and not used in parameter estimates. Indeed, Swait and Adamowicz (2001) note that variance of estimates decreases as participants in CM move through the choice tasks. Following Plott and Zeiler (2005) practice rounds and training should be part of both PP and CM.

Points (a), (c) and (d) do not pose exceptional challenges. More difficult–and probably controversial–is the proposal in point (b). In support of the randomisation of

the PP utility function is the extensive evidence that, even in the best of conditions, only around half of the subjects choose the dominant strategy–i.e. truth-telling–in PP games (Attiyeh et al. 2000, Kawagoe and Mori 2001, Healy 2006). There is clearly something that researchers–and possibly the subjects themselves–are disregarding or misinterpreting. The error terms would capture the unobservable elements of the decision process in the PP environment.

5. Conclusion.

Relatively few applications have made use of laboratory elicitation procedures with incentive compatible mechanisms to replicate hypothetical markets created with stated preference methods. The complementary use of the two methodologies has the potential of providing useful insight both on the hypothetical bias created by stated preference methods, and on the dreaded WTP/WTA disparity so often recorded in the literature.

While the both the methodologies seem to be flexible enough to accommodate specific experimental needs, only a field trial would probably dispels hopes and make problems evident. There are always problems in mixing and matching different institutions. Strictly adherence to best practice both in state preference application and laboratory experiments would assure study replicability and comparability.

The promise of combining hypothetical survey and responses that entail real economic commitment in the context incentive compatible institutions is the possibility of a calibration procedure. Calibration would validate a good hypothetical survey and correct bad one. Will the promise be kept?

REFERENCE

Adamowicz, W. L., Louviere J., William M. (1994), Combining revealed and stated preference methods for valuing environmental amenities, *Journal of Environmental Economics and Management* 26(3): 271-92.

Adamowicz, W. L. (2004), What's It Worth? An examination of historical trends and future directions in environmental valuation, *Australian Journal of Agricultural and Resource Economics*, 48(3): 419-43.

Attiyeh, G., R. Franciosi, Isaac R.M. (2000), Experiments with the pivot process for providing public goods, *Public Choice*, 102(1): 93-112.

Balistreri, E., McClelland G., Poe G., Schulze W. (2001), Can hypothetical questions reveal true values? A laboratory comparison of dichotomous choice and open-ended contingent values with auction values, *Environmental and Resource Economics*, 18(3): 275-292.

Braga, J. And C. Starmer (2005), Preference anomalies, preference elicitation and the Discovered Preference Hypothesis, *Environmental and Resource Economics,* 32(1): 55-89.

Cummings, R. G., G. W. Harrison, Rustrom E.E. (1995), Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible?, *American Economic Review,* 85(1): 260.

Gigerenzer G., Selten R. (eds) (2002), *Bounded rationality: the adaptive toolbox*, MIT Press.

Hanemann, W. M. (1991), Willingness to pay and willingness to accept: how much can they differ?, *American Economic Review*, 81(3): 635-647.

Harrison, G. W. (2006), Experimental evidence on alternative environmental valuation methods, *Environmental and Resource Economics*, 34(1): 125-162.

Healy P. J.(2006), Learning dynamics for mechanism design: an experimental comparison of public goods mechanisms, *Journal of Economic Theory*, 129(1): 114-149.

Horowitz, J. K. And K. E. McConnell (2002), A review of WTA/WTP studies, *Journal of Environmental Economics and Management*, 44(3): 426-447.

Horowitz J.K., McConnell K. E. (2003), Willingness to accept, willingness to pay and the income effect, *Journal of Economic Behavior and Organization*, 51(4): 537-545.

Ledyard J.O. (1994), Public goods: a survey of experimental research, in Kagel J. H. And Roth A. E. (eds), *Handbook of experimental economics*, Princeton University

Press.

List, J. A., Gallet C. A. (2001),What experimental protocol influence disparities between actual and hypothetical stated values?, *Environmental and Resource Economics*, 20(3): 241-254.

Louviere J.J., Hensher D. A., Swait J. (2000), *Stated choice methods*, Cambridge University Press, New York.

Mackenzie J. (1993), A comparison of contingent preference models, *American Journal of Agriculture Economics*, 75: 593-603.

Mcfadden D. (1986), The choice theory approach to market research, *Marketing Science*, 5(4): 275-297.

Plott, C. R., Zeiler K. (2005), The willingness to pay-willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations, *American Economic Review*, 95(3): 530-545.

Sugden, R. (2005), Anomalies and stated preference techniques: a framework for a discussion of coping strategies, *Environmental and Resource Economics*, 32(1): 1-12.

Sugden R. (1999), Alternatives to the neoclassical theory of choice, in: I. Bateman, K. G. Willis (Eds.), *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, Oxford University Press, Oxford.

Swait, J., Adamowicz W. (2001), Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice, *Organizational Behavior and Human Decision Processes*, 86(2): 141-167.

Tversky, A., Kahneman D. (1991), Loss aversion in riskless choice: a reference-dependent model, *The Quarterly Journal of Economics*, 106(4): 1039-1061.

Figure 1. Market and hypothetical data and the elicitation techniques.