Methodological issues in non-market valuation

Victor Champonnois

Thèse de Doctorat d'Aix-Marseille Université en Sciences Économiques

Soutenue publiquement le 11 décembre 2018 devant le jury composé de :

Bengt KRISTRÖM, Professeur, Department of Forest Economics, Swedish University of Agricultural Sciences (*Rapporteur*)
Jürgen MEYERHOFF, Professeur, Technische Universität Berlin, Institute for Landscape Architecture and Environmental Planning, Berlin (*Rapporteur*)
Dominique BUREAU, Délégué général du Conseil Économique pour le Développement Durable (*Examinateur*)
Katrin ERDLENBRUCH, Directrice de recherche, Irstea, Montpellier (*Examinateur*)
Emmanuel FLACHAIRE, Professeur, Aix-Marseille Université, AMSE (*Examinateur*)
Pierre-Alexandre MAHIEU, Maître de conférences, University of Nantes, Lemna (*Examinateur*)
Olivier CHANEL, Directeur de recherche, CNRS, AMSE (*Directeur de thèse*)

Remerciements

Tout d'abord je remercie mon directeur de thèse, Olivier Chanel. Il a toujours été disponible, réactif et très généreux dans l'aide qu'il m'a apportée. Je pense avoir profité d'un environnement exceptionnel et avoir beaucoup appris grâce à lui.

I would like to thank Bengt Kriström and Jürgen Meyerhoff for accepting to review this thesis. I also thank them, as well as Ivana Logar because they kindly hosted my visits in their respective research center, in Umeå, Berlin and Zurich. These stays were highlights of this PhD, where I have made beautiful encounters, both professionally and personally.

Je remercie Pierre-Alexandre Mahieu pour ses conseils et pour l'organisation du fameux WONV, qui a été un rendez-vous annuel dans ma thèse et m'a permis de rencontrer des chercheurs spécialisés de mon domaine. Merci à Katrin Erdlenbruch et Emmanuel Flachaire pour leurs remarques avisées et leurs conseils.

Je remercie également Bénédicte Meurisse pour la supervision de mon séjour au ministère en charge de l'environnement. Grâce à elle j'ai pu prendre du recul sur mes recherches et mieux comprendre le point de vue du décideur. Merci à Dominique Bureau pour m'avoir donné l'opportunité de ce séjour.

Merci à tous mes amis et collègues de l'AMSE, pour les discussions, les projets, les déjeuners, les séminaires et tous les bons moments passés ensemble. Merci particulièrement à mes fabuleux collègues du bureau Y15.

Enfin, je remercie ma famille et mes amis, qui m'ont soutenu et grâce à qui j'ai pu m'épanouir pendant ces trois années.

Abstract

In this thesis I explore different methodological issues arising in non-market valuation. In the first part of the thesis, I try to provide solutions to some problems of preference elicitation. In particular, I analyze the performance of a new elicitation format to reduce anchoring bias in multiple willingness to pay (WTP) elicitation (Chapter 1) and I propose a new strategy to identify the effect of trust in institution on protesting behaviors (Chapter 2). The second part of the thesis is devoted to the statistical analysis of WTP. I compare quantile regression models with standard models to assess their respective ability to account for recurrent issues in WTP data (Chapter 3), I also propose a test for a new type of publication bias (Chapter 4). In the last part of the thesis, I investigate the equity issues in WTP aggregation as a measure of benefits, and the role of subsistence needs (Chapter 5).

Résumé

Cette thèse explore différents problèmes méthodologiques associés à l'évaluation non-marchande. Dans la première partie, je m'intéresse à certaines difficultés posées par l'élicitation des préférences. En particulier, j'analyse les performances d'un nouveau format d'élicitation pour réduire le biais d'ancrage des consentements à payer (Chapitre 1). J'étudie aussi l'effet de la confiance dans les institutions sur les comportements de protestation dans les questionnaires d'évaluation (Chapitre 2). La seconde partie de la thèse est dédiée à l'analyse statistique des consentements à payer. Je compare les modèles de régressions quantiles avec les modèles standards pour mesurer leur capacité à prendre en compte des caractéristiques récurrentes des données de consentement à payer (Chapitre 3). Je propose aussi un test pour un nouveau type de biais de publication (Chapitre 4). Dans la dernière partie, je m'intéresse aux problèmes d'équité liés à l'agrégation des consentements à payer pour mesurer les bénéfices d'un projet, et le rôle joué par les besoins de subsistance (Chapitre 5).

Contents

Chapter 1

General introduction

1.1 Introducing non-market valuation

1.1.1 Cost benefit analysis

In France, before the 19th century, infrastructure projects were not evaluated based on our actual definition of rentability. One of the criteria, explained by the engineer Havre-Lapeyre, considers that a project should be evaluated based on its cost divided by the life-time of the infrastructure: “let’s build a bridge [...] with one bloc of granite that will last as long as the world [...] no matter the cost, being divided by eternity, this fraction will be infinitely small”. With the technological progress in the 19th century emerged the necessity to justify public expenditures. As a consequence, new methods of economic calculations were developed, including the expected benefits for trade in the analysis (Picon (1992)). These methods, which basically confront the costs and the benefits of the projects, were further developed in the US at the beginning of the 20th century, mainly to evaluate water resources projects. Because of competitions among regions, group of users and bureaucratic jurisdiction, there was a need for an objective and standardized manner.

Cost-benefit analyses (CBA) are a simple comparison of benefits and costs of a projects, expressed in monetary terms, that gives a measure of the rentability of this project. For appraisal of public investments and public policy, CBA is based on social costs and social benefits, which are defined as increase or decrease of the well-being of individuals which are part of the society. While cost computation may appear a simple accounting exercise, it can convey large uncertainties, and some aspects like the impact of projects on competitiveness and employments need to be accounted for (see Pearce et al. (2006) for more details). However, because the focus of this thesis is on computation of the benefits, I will not develop further the cost component.

Benefits were initially measured by simply taking the net income to productive factor (quantity times price), ignoring the benefits for the consumer (Spencer Banzhaf (2010)). In 1932, Hotelling, proposed the consumer surplus as a measure of benefits (Hotelling (1932)), which has now become standard practice. The notion of consumer surplus is key in the implementation of CBA, so it deserves a conceptual detour.

1.1.2 Consumer surplus

What does the price of a product represent? Most of us would answer "the value of this product". The idea that a price represents the value of a good or a service is intuitive, but it is not that simple. Adam Smith in a passage in the Wealth of Nations, presents the paradox of value explaining that price represents only a certain type of value.

> "The word value [...] has two different meanings, and sometimes expresses the utility of some particular object, and sometimes the power of purchasing other goods which the possession of that object conveys. The one may be called value in use; the other, value in exchange. The things which have the greatest value in use have frequently little or no value in exchange; and, on the contrary, those which have the greatest value in exchange have frequently little or no value in use. Nothing is more useful than water; but it will purchase scarce anything; scarce anything can be had in exchange for it. A diamond, on the contrary, has scarce any value in use; but a very great quantity of other goods may frequently be had in exchange for it" (book I, chapter IV).

This distinction leaves one question unsolved. "Why would something like water, which has so much value, to humans have a so low price?" In other words, why doesn't a price reflect value in use?

The proper answer came from Dupuit (1844), by introducing an important distinction that has been overlooked by Smith, the difference between average value and marginal value. The marginal value can be defined as the value of one additional item. This notion of marginal value is of great importance because it is this value that determines the purchase of a good: one is willing to buy an additional item only if the marginal value for this item is greater than its cost (the price). In the end, if the consumer is free to choose the quantity she wants, the marginal value will be equal to the price. Therefore, the price is a measure of the marginal value of the last unit of consumption of the good, which is not equal to the average value of the good. In fact, marginal value is very often decreasing with the quantity consumed. For instance, I put a lot of value on one glass of water when I am thirsty, but a little less on the second one, even less on the third one. This means that the total value is greater than the total payment. The *profit* made by the consumer, the difference between total value of consumption and total payment, has been called by Marshall (1890) the consumer surplus. Stated in Adam Smith's words, this surplus is the difference between the value in use and the value in exchange.

Dupuit (1844) gave a simple definition of consumer surplus as "maximum sacrifice expressed in money which each consumer would be willing to make in order to acquire an object".[1] In other words, the value of a good is the result of a trade-off with another good (or money): I am willing to sacrifice ten units of money to obtain this good, so this good is worth 10 times one unit of money. Therefore, benefits are quantified in monetary terms as the total Willingness to pay (WTP) for these benefits.

When prices are available, WTP function (demand function) can be estimated directly. However, some goods or services are not exchanged in markets. Therefore they have no market price. An absence of price does not mean an absence of WTP : if an individual could exchange some money against better air quality or a nicer landscape, they would reveal their WTP for these goods. So new methods were developed in order to measure these unobserved WTP.

1.1.3 Eliciting preferences

Two types of methods exist to assess the economic value of non market goods, revealed preference methods and stated preference methods.

Revealed preferences

The theory of revealed preferences introduced by Samuelson (1938), postulate that the consumer's preferences can be revealed by their purchasing habit. Applied to the non market goods, the revealed preference methods allow to reveal consumer's preference for non market goods based on their observed consumption in other markets that are related to the non-market good.

The first example of a measure of WTP for non market good (outdoor recreational activities) was thought of by Hotelling: "If we assume that the benefits are the same no matter what the distance, we have, for those living near the park, a consumers' surplus

[1] Dupuit (1844) stated that the "maximum sacrifice expressed in money that each consumer would be willing to make in order to acquire an object" provides "the measure of the object's utility".

consisting of the differences in transportation costs." (Spencer Banzhaf (2010)). This intuition will serve as the base of the famous travel cost model still used today. In other words the value of recreational activities in a given site can be estimated by the transportation cost faced by the visitors (including the opportunity cost of time). The intuition laid out by Hotelling was later formalized by Clawson and Knetsch (2013).

Another example of revealed preference method, this time to estimate the impact of air pollution, is to use the price difference between a house in a low pollution area and a house with similar characteristics but located in a high pollution area. This "hedonic price" approach was first proposed by Rosen (1974) based on the Lancaster's new consumer theory, which considers that consumers value the characteristics of the good, not the good in itself (Lancaster (1966)) .

Although these methods are very useful, they have two main drawbacks. First their inability to estimate non-use values, that is values that do not depend on the actual use of the good by the consumers themselves (e.g. altruistic, existence, or option values) (Champ et al. (2003)). Second, they rely on several strong assumptions, for instance the hedonic price method assumes that consumer are perfectly informed, and the market analyzed in pure and perfect competition.

Stated preferences

The second stream of methods is not based on what individuals do, but on what they say they would do. For practitioners, surveys are a very flexible tool to elicit values, because they do not need to rest on any real market and can embark the respondents in purely hypothetical scenarios, to ask them what they would do in this situation.

It is Davis (1963) who first used a survey to measure the value of non-market good, with a method that will later be called contingent valuation. Increasingly used in the 80's, it was first used in court to assess damages from the Exxon Valdez oil spill in 1989 and ask monetary compensation. After this episode, a panel of scientists published a report, known as the National Oceanic and Atmospheric Administration (NOAA) report, giving important guidelines for contingent valuation.

The idea of the method is to propose a fictional market in which the respondents could "buy" the non-market good. In this situation, the respondent states the maximum amount she would be willing to pay for the good.

The first important aspect of stated preference studies is to determine the population that should be surveyed. The answer depends on the type of value looked for: for use values the users of the good or service is the appropriate population, but for non-use values, a more global population is needed. Once the type of population is chosen, the next choice is the survey mode (face to face, web surveys, mail, phone). Face to face interviews have certain qualities such as a possibility to interact with the interviewer to be sure the questions are properly understood, while being able to show figures to help the understanding and capture attention, which is not possible with phone interviews. However, face to face interviews are expensive, and as a consequence web surveys have been increasingly used in the past years, due to their low cost, their ability to show multimedia contents (videos) and to interact automatically on the respondent's clicking behavior. In spite of these advantages, there is a concern of a non-representativeness of the sample in Web surveys (Lindhjem (2011)). When the choices of the population and the survey mode have been made, the good or service of interest and their attributes are described to the respondents and the WTP scenario is presented. Although fictional, this scenario should be realistic: the payment vehicle, institutional context and planned action should be credible.

An important element is the actual WTP questions, i.e. the elicitation format. An important distinction should be made between contingent valuation and choice experiment.

- Contingent valuation is used to value one good (or a few goods with multiple elicitation questions). Among contingent valuation, several elicitation format exist. The main ones are dichotomous choice (single – the referendum - or repeated – double bounded dichotomous choice or bidding games), open ended question and payment cards. Single dichotomous choice has been established as being the most incentive compatible Carson and Groves (2007), it also appear less cognitively demanding for the respondent, although as a consequence, less information are drawn from the responses.

- Choice experiments are used to value attributes of the goods. It is based on Lancaster (1966)'s framework, the respondents makes repeated choice between alternatives with differing attributes (including a cost attribute). Therefore, the respondent implicitly states how she makes trade-off between attributes. The respondent makes several successive choices among different alternatives presented in choice cards.

Both methods give different measures, one valuing the overall value of a good, the other valuing a given set of attributes, so the main determinant of the choice between the two depends on the value of interest and the type of good that is evaluated.

1.1.4 Aggregating preferences

The standard criterion that justifies the CBA is the Kaldor Hicks criterion: if the net benefits from a project is positive, then the overall welfare is increased by the project, and the winners can compensate the losers with a monetary transfer. Therefore in most applied CBA, individual WTP are simply summed up, and the overall gain from the project is measured by the net present value (net benefits with a time discounting). However, this transfer between winners and losers is not made in practice, which raises the question of the equity of CBA. According to Broberg (2010) "a disadvantage of the Kaldor-Hicks criterion is that it ignores the possible importance of distributional issues. It can lead to decisions which favor the rich at the expense of the poor because the rich have a greater ability to pay in support of any given strength of preference."

Theoretically, simply summing up (i.e. applying equal weights) means that the marginal social utility of income is the same for everyone. This is only the case if the fiscal system is considered "optimal", which is of course never the case in the real world. This issue calls for the use of equity weights, which for each individual, should be equal to the individual's marginal utility of income times the weight of this individual in the social welfare function. Some solutions exist to measure marginal utility of income (Fleurbaey and Abi-Rafeh (2016)), but are not yet implemented in practice. More practical solutions are based on the level of income only, not the preferences. One such rule is to adjust WTP measures by the ratio of individual income, Y_i, to average income, $\overline{Y}$ (Fankhauser et al. (1997)). Regarding the weights in the social welfare function, there is no consensus on the proper function to use, so this issue is ignored in practice, which means that everyone is considered to have the same weights (Pearce et al. (2006)).

1.2 Issues in non-market valuation

Although some issues are related to the monetary valuation of non market goods, I will mostly focus on those that are specific to stated preference methods, since it is the main focus of my thesis.

1.2.1 Debates in stated preference surveys

In reaction to the rise of contingent valuation in the scientific literature, virulent criticism have been formulated towards the methodology.

Debates among economists

One of the most famous critics to contingent valuation method (CVM), by Diamond and Hausman (1994), casts a strong doubt on the validity of CVM when evaluating non-use values. They claim that respondents do not have preferences for the goods or services we ask them to evaluate. Another strong critic on CVM (not only focusing on non-use values) was made by Kahneman and Knetsch (2005), who argue that CVM measures the moral satisfaction people get from contributing to public goods, not their economic value. Both of these critics are based on the same empirical fact: the monetary insensitivity to the scale or the scope of the change of good or service. Almost 20 years later, Hausman (2012) persists and sharpens his critic, focusing also on the divergence between WTP and willingness to accept (which in theory should not be too far apart), and the hypothetical bias (an upward bias of WTP compared to what individuals would actually pay in a revealed preference framework). Several responses to these critics (for instance Haab et al. (2013)) offer a more positive view, and highlight the progress that have been made in the past decades on these dimensions (Carson (2012)).

Another issue of importance in stated preference surveys is the so-called "protest zeros". In stated preference surveys, some respondents refuse to state their true preferences and give a zero amount instead, for various reasons. For instance, they think someone else should pay for the good, the choice of payment vehicle is not adequate or the scenario is not credible enough. Even if the practitioner is able to detect these protest responses and remove them from the sample, it affects the representativeness of the sample.

In recent years, the choice experiment approach has become predominant in stated preference studies (Mahieu et al. (2017)). Although it is often claimed that it performs better than CVM with respect to the main issues faced by the method, results from the literature are not clear. Regarding sensitivity to scope, the evidence of a better sensitivity to scope has been recently put under question (Andersson et al. (2016)), the evidence of a better incentive compatibility are mixed (Hoyos (2010)) and DCM does not seems to lead to significantly fewer protest responses (Meyerhoff and Liebe (2010, 2008)).

The ultimate argument for the defender of stated preference surveys is that there is no alternative to account for non-use values, and therefore some numbers should be better than zero. This is true only in the framework of CBA. As I will show below, institutions can be creative and find other ways to account for non-market values by stepping out of a unidimensional and fully monetary evaluation. Nonetheless, there are important issues arising with multidimensional evaluations. For instance, there is no weighting of dimensions, so either all dimensions are weighted equally, either the weighting is at the discretion of the decision maker, both solutions are ad-hoc and not based on a theoretical ground. Besides such analysis neglects the possibility of co-benefit (benefits arising in different dimensions) (Bureau (2018)). So as far as reasonable, improving stated preference methods is still the best way towards a comprehensive, precise and clear evaluation of projects.

Some ethical aspects of non-market valuation

Giving a monetary value to non-market goods can raise ethical concerns. For instance, measuring the benefits of projects that have effects on health and safety requires to put a value on a decrease in the probability of death, implicitly putting a value on human life. Economists have a very careful terminology and speak of value of a statistical life or value of a prevented fatality. The important aspect of this value is that it is not the value of an identified life that is saved for sure, but the value of an anonymous individual whose risk of death is decreased by a small probability. Despite these precautions, there is still a controversy around putting a value on life (Viscusi (2005)). Against this reluctance, some economists like Tirole (2017) reply with a consequentialist argument: if we refuse to set an explicit value on life to use in policy evaluation, decision maker will anyway put an implicit value on life through their choice of policy. So the question is then, do we prefer to have a collective reflexion on this value or do we let this value be at the discretion of the deciders?

In any case, even if institutions (in France for instance, see next section) accept this trade-off between life and money, an elicitation of the WTP for a decrease in the risk of death through stated preference surveys requires the respondents to be willing to make this trade-off. In other words, respondents have to be utilitarian in order for their WTP to exist. Sagoff (2007) claim that people think about the environment as citizens and not as consumers, which means that respondents who hold right based beliefs as citizens cannot put themselves in the "environmental consumer's" shoes to state a valid WTP for the valuation survey. In theory, this means that the whole cost benefit method and the Kaldor Hicks criterion falls apart if some individuals have non-utilitarian position. Indeed, this individual cannot make an explicit trade-off between money and life risk, so she cannot be compensated.

Valuation of biodiversity raises other concerns, because it requires values to be anthropocentric. In other words the value of any "good" is defined with respect to human well-being. This value can come from the use (e.g. a safari) or non use (e.g. altruism towards animals) of the good. This implies that in CBA there is no intrinsic values of environmental goods, which means that non-human living species have no absolute rights that should be protected no matter the cost (see for instance Abson and Termansen (2011) and O'Neill (1992) for discussions on the ethical position of anthropocentrism in environmental valuation). Thus, for economic valuation to properly measure anthropocentric values, all the population must share an anthropocentric view on environmental values in order for them to make biodiversity-money trade-offs. This is not likely to be the case (Veisten et al. (2006)).

To conclude, the question of ethical beliefs of respondents in non-market valuation is an important dimension that is often overlooked by economists.

1.2.2 At the institutional level: the case of France

Since the Loi de programmation pluriannuelle des finances publiques in 2012, an economic evaluation is required for all investment projects in France. The Commissariat général à la stratégie et à la prospective (2013) provides the guidelines and reference values for these evaluations of projects. Transportation is the first sector for which the CBA became the main tool for project evaluation. Stated preference surveys are used to give reference values to some dimensions of the benefits. For instance the reference value of a statistical life in France is 3 million euros, and follows the OECD recommendation based on a meta-analysis of VSL, estimates from stated preference surveys (OECD (2012)). However, some reference values are still arbitrary or based on an insufficient literature. As an example, mild and severe injuries have respectively values of 2% and 15% of the VSL, which are originally only based on one study with a production function approach (ignoring all intangible effects). Thus, there is still some work needed to properly account for all the dimensions of CBA evaluation of transport infrastructures.

For flood prevention projects of more than 5 million euros, a Multicriteria analysis (MCA) is required. MCA can be defined as an extended CBA: beside a traditional CBA, elements that cannot be monetized in the actual state of knowledge are evaluated using non-monetary measures. The project is evaluated for these dimensions using efficiency criterion such as Cost-effectiveness analysis. For instance, in flood prevention projects, impacts on the safety of the population are evaluated using the number of persons living in a risky area before and after the project. This measure is divided by the cost of the project to obtain a measure of "people put in safety" per euro spent, which is easy to compare between projects. This type of analysis present several issues, such as implicit weights that are put on the non-monetary dimensions, or the inability to account for co-benefits (Bureau (2018)). However, it serves as a second-best solution in the absence of robust methods to estimate a monetary value for these dimensions.

The French national ecosystem assessment, called Évaluation Française des Écosystèmes et Services Écosystémiques (EFESE), is a program led by the French ministry of environment. It is in charge of proposing ways to evaluate the contribution of ecosystem to human populations (CGDD (2016)). Among the different values assessed, an important one is the notion of natural heritage ("patrimoine naturel") : elements of biodiversity that have an identity dimension and a particular status due to a notable feature. This natural heritage has a non-use value, and in some cases, a non-utilitarian value. Therefore it is very difficult to attach a monetary or even a quantitative value to this aspect of biodiversity. Some of the propositions of the program are based on a description of the way these natural heritages are acknowledged, or on an inventory of the elements of this heritage. For instance, protection status (natural parks) can reveal our disposition to protect an ecosystem, labels like "Grand site de France" qualify notable natural or cultural aspects of territories, and the symbolic aspect of some species can be found in artistic works (paintings, literature, ...), in which they are source of inspirations and build individual and collective identity.

Although there is an upward trend, project evaluation is still insufficiently made by decision makers. While methodological issues are important, they are not the only issues when it comes to concrete implementation. At least in France, an additional barrier seems to stand at the institutional level (Bureau (2018)).

1.3 Objective of this thesis

The aim of this thesis is to contribute to the literature on methodological issues in non-market valuation. The process of measuring the non-market benefits from a project can be decomposed in three stages. First, studies are conducted to elicit the WTP of the population for a non-market good or service. Then these WTP data are analyzed using statistical techniques, for instance to study the determinants of the WTP or to test for theoretical validity. Finally, WTP are aggregated to provide a measure of the overall benefit. This thesis aims to provide insights on some of the issues arising in all three stages and concrete solutions to improve the valuation process.

1.3.1 Issues in willingness to pay elicitation

Elicitation of valid WTP is known to be a difficult task. One important aspect is the cognitive process followed by the respondents to give a WTP. Anchoring belongs to the larger class of implied value cues. The respondent tends to anchor his/her stated WTP on the bid(s) offered during the elicitation step, being influenced by starting values (starting point bias; Tversky and Kahneman (1974)) and the range and the centering of bids (Covey et al. (2007)). When WTP answers include an anchoring effect, "true" unobserved WTP differs from stated WTP in a non-random way, making it unsound to base any decision on such values. An additional type of anchoring is involved where multiple WTPs are elicited within a single CV survey. Although multiple eliciting has advantages, like explicitly offering potential substitutes for the good valued (see Luchini et al. (2003), respondents may be influenced by their prior answers (Payne et al. (2000)). **The first chapter** of this thesis assesses how a recently proposed variant of PC – the CPC - performs with respect to anchoring in multiple elicitation questions, using the results from two empirical multiple elicitation CV surveys.

In stated preference surveys, some respondents refuse to state their true preferences and give a zero amount instead, for various reasons. For instance, they think that someone else should pay for the good, that the choice of payment vehicle is not adequate or that the scenario is not credible enough. Usually, these respondents must justify their zero amount by answering follow-up questions, which enable the practitioner to detect these so called "protest respondents". Even if the practitioner is able to detect these protest responses and remove them from the sample, the distribution of protest respondents is very unlikely to be random. Then, the samples on which the aggregate WTP is computed are no more representative. For this reason, it is important to understand the motivations of protest behaviors and to find ways to mitigate them. Many studies explore the determinants of protest responses. An important determinant is the respondent's perception of the managing authority. However, the effect of trust in institutions on the protest responses is not clear yet, and could be better understood by tackling two issues. First, the effect differs depending on how we look at the institutional determinant of the protest responses. Second, since most studies are conducted in one place at one time, the institutional context that surrounds the respondents does not change. Therefore, these studies cannot tell anything about the effect of the institutional factors that motivate the protest answers. In **the second chapter** I tackle these issues by relying on meta-data on environmental valuation studies, merged with institutional variables. By using intra-country variations in these institutional variables, I am able to capture the effect of trust institutions on the protest rate, wiping out the effect of each studies' specificities.

1.3.2 Statistical analysis of willingness to pay

Econometric models help inform private or public decision-makers by clarifying and predicting preferences; to do so, they seek WTP determinants amongst respondent and survey characteristics. Effective modeling, however, needs to take several econometric issues into account. First, there is the treatment of zero WTP, which introduces censoring from below. Second, there can be outliers and/or extremely large values, due either to the hypothetical nature of the CV exercise or to the difficulty of the valuation task. Third, the impact of respondents' characteristics on WTP is potentially heterogeneous, which may bias estimates. Quantile regression (QR) and censored quantile regression (CQR) can help tackle these issues by estimating the impact of explanatory variables on any conditional WTP quantiles chosen, instead of simply on mean WTP. Yet despite the importance of CV in environmental studies, and its increasing use of quantile-based methods, the field lacks a systematic analysis of the performance of (C)QR. **The third chapter** proposes to fill this gap by comparing the performances of (C)QR models and standard models (Ordinary Least Square and Tobit).

Meta-analyses have been increasingly used in various areas during the past decades. This research technique allows to average key estimates from different studies to obtain a unique and more precise measure to use for policy makers. It also enables identifying determinants of the heterogeneity in estimates between studies, which can be used to provide methodological guidelines. One of the main issues in meta-analyses is the selection of estimates, both by the researcher and the journals. If not accounted for, this selection creates a publication bias, which can lead to a misrepresentation of the aggregate estimate and inappropriate recommendations Stanley et al. (2017). In two recent meta-analyses on the value of a statistical life (VSL), Viscusi (2015); Viscusi and Masterman (2017) suggest a previously unexplored type of publication bias. This bias arises "because of the efforts by researchers to provide estimates in line with the previous literature". In **the fourth chapter**, I propose two tests for the anchoring of estimates on prior studies. Using a meta- analysis on the VSL, I show that this publication selection process exists and that it affects the mean and the bias of the VSL estimates.

1.3.3 The plutocracy of Cost-benefit analysis

Once individual WTP are properly elicited and statistically analyzed, their aggregation as a measure of benefits appears very democratic, directly feeding the preferences into support decision-making. Nevertheless, it hides two methodological issues when preferences are elicited through willingness to exchange money (or a composite market good) for non-market goods or service provision. First, there is the budget constraint effect, already empirically investigated and for which solutions have been proposed to accurately reflect the preferences of low-income individuals. Second, subsistence needs limit the realm of possibility when expressing preferences, and therefore WTP, which affects the poorest more than the richest. Thus, subsistence needs exacerbate the problem of inequity in CBA through their effect on the marginal utility of income. **The fifth and last chapter** of this thesis investigates how subsistence needs distort the WTP-based expression of preferences, insidiously turning CBA into a plutocratic process. Surprisingly, to our knowledge, their consequences on preference and WTP elicitations have not previously been explored theoretically. We propose to fill this gap by comparing the standard framework with one that accounts for subsistence needs. First, we show how preference elicitation for a non-market good or service is affected. Second, we show how the non-market implicit (or shadow) price is under-estimated w.r.t. the standard framework. Finally, we provide both numerical and empirical illustrations of this under-estimation.

Bibliography

Abson, D. J. and Termansen, M. (2011). Valuing ecosystem services in terms of ecological risks and returns. *Conservation Biology*, 25(2):250–258.

Andersson, H., Hole, A. R., and Svensson, M. (2016). Valuation of small and multiple health risks: A critical analysis of SP data applied to food and water safety. *Journal of Environmental Economics and Management*, 75(C):41–53.

Broberg, T. (2010). Income Treatment Effects in Contingent Valuation: The Case of the Swedish Predator Policy. *Environmental and Resource Economics*, 46(1):1–17.

Bureau, D. (2018). Feasibility of costs-benefits analysis for environmental health choices. *Environnement Risques Santé*, (4):368–372.

Carson, R. T. (2012). Contingent Valuation: A Practical Alternative when Prices Aren't Available. *Journal of Economic Perspectives*, 26(4):27–42.

Carson, R. T. and Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, 37(1):181–210.

CGDD (2016). EFESE L'essentiel du cadre conceptuel.

Champ, P. A., Boyle, K. J., Brown, T. C., and Bateman, I. J., editors (2003). *A Primer on Nonmarket Valuation*, volume 3 of *The Economics of Non-Market Goods and Resources*. Springer Netherlands, Dordrecht.

Clawson, M. and Knetsch, J. L. (2013). *Economics of outdoor recreation*. RFF Press.

Commissariat général à la stratégie et à la prospective (2013). L'évaluation socioéconomique des investissements publics.

Covey, J., Loomes, G., and Bateman, I. J. (2007). Valuing risk reductions: Testing for range biases in payment card and random card sorting methods. *Journal of Environmental Planning and Management*, 50(4):467–482.

Davis, R. K. (1963). Recreation planning as an economic problem. *Nat. Resources J.*, 3:239.

Diamond, P. A. and Hausman, J. A. (1994). Contingent Valuation: Is Some Number Better than No Number? *Journal of Economic Perspectives*, 8(4):45–64.

Dupuit, J. (1844). De la Mesure de l'Utilite des Travaux Publiques. *Annales des Ponts et Chaussees*.

Fankhauser, S., Tol, R. S. J., and Pearce, D. W. (1997). The Aggregation of Climate Change Damages: a Welfare Theoretic Approach. *Environmental and Resource Economics*, 10(3):249–266.

Fleurbaey, M. and Abi-Rafeh, R. (2016). The Use of Distributional Weights in Benefit–Cost Analysis: Insights from Welfare Economics. *Review of Environmental Economics and Policy*, 10(2):286–307.

Haab, T. C., Interis, M. G., Petrolia, D. R., and Whitehead, J. C. (2013). From Hopeless to Curious? Thoughts on Hausman's "Dubious to Hopeless" Critique of Contingent Valuation. *Applied Economic Perspectives and Policy*, 35(4):593–612.

Hausman, J. (2012). Contingent Valuation: From Dubious to Hopeless. *Journal of Economic Perspectives*, 26(4):43–56.

Hotelling, H. (1932). Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions. *Journal of Political Economy*, 40(5):577–616.

Hoyos, D. (2010). The state of the art of environmental valuation with discrete choice experiments. *Ecological Economics*, 69(8):1595–1603.

Kahneman, D. and Knetsch, J. L. (2005). Valuing public goods: the purchase of moral satisfaction. *The Earthscan reader in environmental values*, page 231.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2):132–157.

Lindhjem, H. (2011). Using Internet in Stated Preference Surveys: A Review and Comparison of Survey Modes. *International Review of Environmental and Resource Economics*, 5.

Luchini, S., Protière, C., and Moatti, J.-P. (2003). Eliciting several willingness to pay in a single contingent valuation survey: application to health care. *Health Economics*, 12(1):51–64.

Mahieu, P.-A., Andersson, H., Beaumais, O., Crastes dit Sourd, R., Hess, S., and Wolff, F.-C. (2017). Stated preferences: a unique database composed of 1657 recent published articles in journals related to agriculture, environment, or health. *Review of Agricultural, Food and Environmental Studies*, 98(3):201–220.

Marshall, A. (1890). *Principles of economics*. OCLC: 1045114844.

Meyerhoff, J. and Liebe, U. (2008). Do protest responses to a contingent valuation question and a choice experiment differ? *Environmental and Resource Economics*, 39(4):433–446.

Meyerhoff, J. and Liebe, U. (2010). Determinants of protest responses in environmental valuation: A meta-study. *Ecological Economics*, 70(2):366–374.

OECD (2012). Mortality Risk Valuation in Environment, Health and Transport Policies - OECD.

O'Neill, J. (1992). The varieties of intrinsic value. *The Monist*, 75(2):119–137.

Payne, J. W., Schkade, D. A., Desvousges, W. H., and Aultman, C. (2000). Valuation of multiple environmental programs. *Journal of Risk and Uncertainty*, 21(1):95–115.

Pearce, D. W., Atkinson, G., and Mourato, S. (2006). *Cost-benefit analysis and the environment: recent developments*. Organisation for Economic Co-operation and Development, Paris. OCLC: ocm64672331.

Picon, A. (1992). De l'utilité des travaux publics en France au XIXe siècle. *Culture Technique*, (26):122–127.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1):34–55.

Sagoff, M. (2007). *The Economy of the Earth: Philosophy, Law, and the Environment*. Cambridge University Press, Cambridge, 2 edition.

Samuelson, P. A. (1938). A Note on the Pure Theory of Consumer's Behaviour. *Economica*, 5(17):61.

Spencer Banzhaf, H. (2010). Consumer Surplus with Apology: A Historical Perspective on Nonmarket Valuation and Recreation Demand. *Annual Review of Resource Economics*, 2(1):183–207.

Stanley, T. D., Doucouliagos, H., and Ioannidis, J. P. A. (2017). Finding the power to reduce publication bias. *Statistics in Medicine*, 36(10):1580–1598.

Tirole, J. (2017). *Economics for the common good*. Princeton University Press.

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131.

Veisten, K., Navrud, S., and Valen, J. S. Y. (2006). Lexicographic preference in biodiversity valuation: Tests of inconsistencies and willingness to pay. *Journal of Environmental Planning and Management*, 49(2):167–180.

Viscusi, W. K. (2005). The Value of Life. SSRN Scholarly Paper ID 827205, Social Science Research Network, Rochester, NY.

Viscusi, W. K. (2015). The role of publication selection bias in estimates of the value of a statistical life. *American Journal of Health Economics*, 1(1):27–52.

Viscusi, W. K. and Masterman, C. (2017). Anchoring biases in international estimates of the value of a statistical life. *Journal of Risk and Uncertainty*, 54(2):103–128.

Part I

Issues in willingness to pay elicitation

Chapter 2

Reducing the anchoring bias in multiple question CV surveys[1]

Abstract: The elicitation format is a crucial aspect of Contingent Valuation (CV) surveys and can impact their reliability. This paper contributes to the extensive debate on WTP (Willingness To Pay) elicitation formats by assessing whether the Circular Payment Card (CPC) can reduce anchoring on respondents' previous answers under multiple elicitation questions. This new format uses a visual pie-chart representation without start or end points: respondents spin the circular card in any direction until they find the section that best matches their WTP. We used a CV survey based on two ways of reducing risks associated with flooding, each randomly presented first to half of the respondents, to test the absolute performance of CPC. We presented a second survey on two social insurance schemes for subjects currently uninsured to respondents randomly split into three subgroups. Each group's WTP was elicited using one of three formats: Open-Ended (OE), standard Payment Card (PC) and the new CPC. The two insurance schemes were always proposed in the same order, and we assessed the relative performance of CPC by comparing anchoring across respondents. Our results provide evidence that CPC is likely to reduce anchoring in multiple elicitation questions and that respondents may rely on different heuristic decisions when giving WTP in the OE and in the two PC formats.

[1]This paper is a joint work with Olivier Chanel and Khaled Makhloufi. It has been published in the Journal of Choice Modelling.

2.1 Introduction

Stated preference methods are increasingly used to inform public policies or company strategies when market prices cannot be observed directly. Examples include surveys dealing with non market values (in health, environment, education or transport) or with products not yet available on marketplaces (in marketing, finance or consumer research). Values are obtained either by choice modelling, i.e. analysing choices from several sets of alternatives, or by contingent valuation (CV), i.e. eliciting maximum willingness to pay (WTP) for a given level of good proposed in a hypothetical scenario.

Since respondents to CV surveys rely solely on a hypothetical scenario and an elicitation format when making their decision, the design of these components may affect stated WTP. The format needs to be appropriate, to ensure that CV surveys reveal respondents' "true" unobserved WTP values. In terms of the hypothetical scenario, this can generally be achieved by following practitioner guidelines and by doing careful pre-tests. However, although widely discussed in the literature, none of the existing elicitation formats appears clearly to outperform the others (see Mitchell and Carson, 1989; Carson and Groves, 2011). Choice of elicitation format is likely to impact the quality and quantity of WTP information collected, as well as to introduce potential errors/biases.

In particular, elicitation formats that rely on closed-ended answers may encourage the anchoring of respondents' WTP on the bids offered (Boyle et al., 1997; Herriges and Shogren, 1996). A variant of the Payment Card (PC) - referred to as the Circular PC (CPC) – was recently proposed (Chanel et al., 2013) to limit such anchoring. By using a visual pie-chart representation without start or end points (see Figure 1), it helps eliminate starting-bid and middle-card biases, and strongly reduces the range effect. This new format has been compared favourably to two of the most common elicitation formats: the Open Ended (OE) and PC (Chanel et al., 2017).

This paper seeks to contribute to the methodological literature on WTP elicitation formats by testing CPC's ability to limit anchoring on the respondent's previous answers under multiple CV elicitation questions. We use the results of two CV surveys to assess both the absolute and the relative ability of the CPC format to cope with such anchoring. The first survey proposes two successive scenarios for flood-related risk reduction and exploits the random assignment of scenario order over the sample of respondents to test the format's absolute ability. We expect to find no evidence of anchoring on the first elicited WTP. The second survey successively elicits WTP for first, health and second, pension insurance schemes over the whole sample. Keeping the order unchanged, we randomly use a different elicitation format (OE, PC or CPC) for each of three sub-groups to test the relative ability of the CPC format w.r.t. the two others. We expect to find less anchoring on first WTP with the CPC format than with the two other formats.

The remainder of the paper is structured as follows. Section 2 describes the anchoring issues in CV and details the new format proposed, while Section 3 explains the empirical strategy. Section 4 tests for anchoring in a survey dealing with two ways of reducing risks associated with flooding, while Section 5 tests for anchoring in a survey dealing with two social insurance schemes for respondents currently uninsured. Section 6 discusses and concludes.

2.2 Anchoring and the elicitation format

2.2.1 Anchoring issues in CV surveys

Anchoring belongs to the larger class of implied value cues that directly lead to respondents' answers being sensitive to framing effects in ascending vs. descending bidding games (DeShazo, 2002), to a greater tendency to "yea/nay" saying in dichotomous choice (DC) questions (Kanninen, 1995; Chien et al., 2005) or to incentive incompatibility in multiple DC questions (Whitehead, 2002). A first type of anchoring arises when the respondent tends to anchor his/her stated WTP on the bid(s) offered during the elicitation step, being influenced by starting values (the starting point bias first evoked by Tversky and Kahneman, 1974; in psychology), follow-up values (Araña and León, 2007) and the range and the centring of bids (Covey et al., 2007). When WTP answers include an anchoring effect, "true" unobserved WTP differs from stated WTP in a non-random way, making it unsound to base any decision on such values.

It is worth noting that the anchoring tendency is partly unconscious. It has been observed in psychological studies eliciting objective quantities (e.g. the length of the Amazon, the yearly average mileage travelled by car or the number of physicians in the local yellow pages) even when respondents know that the numbers proposed are random and unrelated to the good (O'Conor et al., 1999; Tversky and Kahneman, 1974; Wilson et al., 1996). Anchoring is especially prevalent when preferences regarding the good are uncertain, due to poor definition or too limited knowledge. Flachaire and Hollard (2007), for instance, develop a model in which the bids proposed during the survey help respondents reduce uncertainty, resulting in a stronger effect on the initial bid than on subsequent bids. Overall, this type of anchoring arises when the bid proposed (or any number proposed) during the elicitation process is considered as providing an indication of the value of the good (i.e., a cognitive anchor).

A second type of anchoring arises when multiple WTPs are elicited within a single CV survey implicitly assuming that respondents' answers are unaffected (see the question order issue in Boyle et al., 1993). This is sometimes done when the choice modelling approach is not methodologically feasible, or as a way of reducing survey costs or exploring various alternatives within the same scenario. Although multiple eliciting has advantages, like explicitly offering potential substitutes for the good valued (see Luchini et al., 2003), respondents may be influenced by their prior answers (Payne et al., 2000; Longo et al., 2015). They may also have difficulties with the alternative-specific cognitive assessment for each valuation question, i.e. be unable to construct the appropriate reference framework for each question (Selart, 1996). The consequence is a deliberate or unconscious tendency for respondents to self-anchor on their own previously stated WTP. This type of anchoring is different from the position effect in a top-down or a bottom-up design, in which a composite good is valued either directly or as a package built from a smaller subset or extracted from a larger multipack good (Powe and Bateman, 2003; Veisten et al., 2004). It is also different from a nested effect, either true (when one good is a subset of another, Carson and Mitchell, 1995) or perceived by the respondents (when one good is an improved version of another, De Ridder and De Graeve, 2005).

In this article, we focus on the second type of anchoring: a tendency to rely on one's own previous answers in multiple elicitation questions, without fully considering what makes the goods assessed differ across questions. We assess how a recently proposed variant of the PC – the CPC – performs with respect to this type of anchoring, using the results from two empirical multiple elicitation CV surveys.

2.2.2 A new variant of the PC: the CPC

Despite recent improvements (Wang, 1997; Cook et al., 2012; Mahieu et al., 2017) the PC format still presents disadvantages, chiefly the risk of implied value cues from the range of the bid interval, the starting values and the position of the bids. Chanel et al. (2013) recently introduced a variant of the PC: the CPC. It uses a visual representation of a circular card with no predetermined start or end points, no top or bottom, no left or right (see Figures 1 and 2 for two examples). The interviewer asks the respondent to think about his/her maximum WTP, and then presents the CPC in a random position to help her/him in the elicitation process. Respondents are asked to spin it until they find the section that best corresponds to their WTP to benefit from the improvement proposed in the scenario. The text containing the WTP values is curved around the circle to allow easy handling and spinning, with no predetermined direction of rotation. The respondent is then asked which section corresponds to his/her WTP; there may also be an OE follow-up question to elicit a more precise WTP.

Chanel et al. (2017) establish the advantages of the CPC over the standard PC on a single elicitation question: it helps eliminate starting-bid bias (because each section is equally likely to be seen at first glance) and middle-card bias (by construction), and strongly reduces the range effect associated with the bids chosen (the succession of bid ranges mimics a continuous distribution). Since the respondent has to spin the circular card to reach the section corresponding to his/her WTP, both cognitive and (a small) physical effort are involved. This extends the cognitive process from the pure reflection needed to choose a value to the motor skills needed to spin the CPC. These 'efforts', repeated for each elicitation question, may increase the respondent's engagement in the elicitation process. Because the interviewer presents the CPC in a new random position for every question, respondents are assumed to be less likely to anchor on their own previous WTP when answering a subsequent WTP question in multiple elicitation surveys. To test this, we used two empirical CV studies.

2.3 Empirical strategy

We used the results of two CV surveys – on flooding and on social insurance - to assess both the absolute and the relative ability of the CPC format to cope with the anchoring issue under multiple elicitation questions.

2.4 Testing the CPC format's absolute anchoring reduction potential

In the first CV survey, residents of France's Provence Alpes Côte d'Azur (PACA) region with differing degrees of flood-risk exposure were presented with two successive scenarios for flood-related risk reduction (Protective Devices or Insurance). Each scenario was presented in first position to a randomly chosen half of the sample, and all respondents' WTPs were elicited using the CPC (see Figure 1). Consequently, through random assignment of scenario ordering over the sample, this survey offers a test of the absolute ability of CPC to cope with anchoring, by examining the relationship between WTP elicited and scenario ordering. Let us denote the j^{th} elicited WTP for scenario k by

WTP_{kj} with j=1,2, k=A,B, and omit the respondent's index i to lighten notations.

First, we perform unconditional analyses. We start by checking whether WTPs differ according to respondent group (WTP_{A1} and WTP_{B2} pooled vs. WTP_{B1} and WTP_{A2} pooled) and scenario (WTP_{A1} and WTP_{A2} pooled vs. WTP_{B1} and WTP_{B2} pooled). We then check for the presence of multiple anchoring by looking at whether, for a given scenario, WTPs are the same when stated in first position (no anchoring) and in second position (after potential anchoring): $\mathrm{WTP_{A1}}$ vs. $\mathrm{WTP_{A2}}$ and $\mathrm{WTP_{B1}}$ vs. $\mathrm{WTP_{B2}}$.

We use four different equality tests to compare WTPs, paired or unpaired depending on the pooling applied. We perform a Welch (1947)'s t-test assuming that the mean WTPs are the same. Then we test whether the median WTPs are the same using a Chi-square non-parametric test of independence. Finally we use two other non-parametric tests of the equality of the distribution: the Wilcoxon rank-sum and the Kolmogorov-Smirnov tests.

Second, we check whether the unconditional results hold when overall potential heterogeneity of respondents is taken into account by estimating the anchoring effect with the following model for both scenarios (Models 1 and 2):

$$WTP_{k,i} = \alpha Order_i + x_i'\beta + u_i \qquad k = 1,2; i = 1,...,n \qquad (1)$$

where x_i is the vector of individual covariates, β is a vector of coefficients, $Order_i$ is a dummy accounting for the order of the scenarios, α is the associated coefficient and u_i is an error term. Testing for the null hypothesis $\alpha = 0$ allows us to detect a potential anchoring effect, while controlling for observed characteristics.

2.4.1 Testing the CPC format's relative anchoring reduction potential

The second CV survey successively asked uninsured Tunisians their WTP for health and pension insurance schemes that could be made available to them. In this survey, although the health insurance scheme was always proposed first (scheme A), each of three sub-groups of equal size was randomly assigned to one elicitation format: OE, PC or CPC. Consequently, this survey offers the opportunity to test the relative ability of the CPC format to cope with multiple anchoring w.r.t. the two standard formats.

We analyse differences in WTP between respondents, to explore the tendency to anchor - both specific to each elicitation format and linked to the first elicited WTP - while controlling for WTP determinants. We focus below on the anchoring of WTP for the pension scheme (B) on WTP for the health scheme (A), accounting for the impact of the elicitation formats *l=OE, PC, CPC* (see for instance Luchini et al., 2003; or Protière et al., 2004; for a more complete treatment of interdependencies involving information delivered). We proceed in two ways. First, we introduce WTP for scheme A (WTP_{Ai}) and dummy-specific terms for the elicitation format ($Elicitation_{li}$) (Model 3):

$$WTP_{Bi} = x_i'\beta + \delta WTP_{Ai} + \epsilon Elicitation_i + u_i \qquad (2)$$

Then, we replace the WTP_{Ai} variable by specific interaction terms WTP_{Ai} *x* $Elicitation_{li}$ to account for an elicitation format-specific anchoring effect on scheme A (Model 4):

$$WTP_{Bi} = x_i'\beta + \psi WTP_{Ai} * Elicitation_{li} + \gamma Elicitation_i + u_i \qquad (3)$$

2.4.2 Methodological issues common to both surveys

Because we are concerned here with the ability of the CPC format to reduce anchoring under multiple elicitation questions, we restrict our analysis to respondents actually exposed to the elicitation format. Consequently, in both surveys, respondents answering "No" to a prior willingness-to-join question (protest and true null responses) are removed from the samples. In addition, we restrict the samples to respondents who provide valid WTP for each multiple elicitation question.

Although it is common in the literature to use increasingly spaced bid amounts with the standard PC, we choose equally spaced amounts for our CPC, both based on pre-tests and as inherent to the design (better mimicking a continuous distribution). Finally, regressions use the middle of the bid-range elicited for the PC and CPC formats (as in Cameron and Huppert, 1989; Yang et al., 2012). In the social insurance survey, this enables a single model to be used to elicit WTP via three formats. Note that interval regression models (Wooldridge, 2002; Anderson and Mellor, 2009) that account for differences in the type of WTP elicited have also been estimated. They yield very similar results, whether they use point estimate both in the OE format and the OE follow-up question of the CPC format and interval with two specified thresholds for the PC format, or point estimate for the OE format and interval for the PC and CPC formats (details upon request).

2.5 Evidence on the CPC's absolute anchoring reduction potential

2.5.1 Survey design of the CV survey on flooding

The first CV survey was conducted through face-to-face interviews in April-June 2012 with inhabitants living in four municipalities of Southeastern France. Two scenarios for flood risk reduction were presented to each respondent, in the spirit of Deronzier and Terra (2006) (see Appendix A for details). The first one (Protective Devices) proposed a collective action consisting in funding municipality-level protective devices; the WTP stated by the respondents covered both the material and the psychological costs of flooding. The second scenario (Insurance) proposed a payment for insurance against the financial risk related to flooding; the WTP covered only material damage. The sample was randomly divided into two groups differing in scenario presentation order. This enabled us to disentangle the impact of the scenario (Insurance vs. Protective Devices) from the impact of the order of presentation (first vs. second).

The initial sample was composed of 599 adults representative of the PACA population in terms of three stratification variables (age, gender, and profession) and differing with respect to the flood risk inherent in their place of residence. Of the municipalities, Miramas had never been flooded and was not in a flood plain, Berre l'Etang had never been flooded but was located in a potentially risky area, Vaison-la-Romaine was flooded in 1992 (41 dead or missing), Draguignan was flooded in 2010 (25 dead or missing). All respondents from previously flooded municipalities had to be living there when the flood occurred. Questions covered socio-demographic variables, preferences regarding time and risk, flood risk perception, information and behaviours regarding flood risk (see Appendix B for a description of the variables). 200 respondents answered "No" to the willingness-to-join question for both scenarios and 132 for at least one scenario, leaving us with 277 respondents who encountered the CPC twice.

2.5.2 Results

We first conduct the four equality tests presented in section 3.1 on pooled WTPs, to check whether the WTPs differ significantly depending on respondent's group and scenario (see Table I). Whatever the test of equality, the null assumption of equality of WTP by group cannot be rejected (lowest p-value=.231), but the null assumption by scenario is always rejected: WTPs for Insurance are higher than for Protective Devices. The

	Group effect	Scenario effect	Anchoring effect
Test of equality	WTP_{A1} and WTP_{B2} vs. WTP_{B1} and WTP_{A2}	WTP_{A1} and WTP_{A2} vs. WTP_{B1} and WTP_{B2}	WTP_{A1} vs. WTP_{A2}
Mean (t-test)	.9011	.0076	.9088
Median (non parametric test)[a]	.231	.0725	.595
Distribution (Wilcoxon rank-sum test)	.7394	.0298	.9447
Distribution (Kolmogorov Smirnov test)[b]	.787	-	.742

Table I: P-values of the equality tests

absence of any group effect rules out a spurious association with anchoring effect through presentation order, while a significant difference across scenarios leaves room for potential anchoring. Because WTP for Insurance is significantly higher than for Protective Devices, in presence of multiple anchoring we expect the second WTP elicited to be anchored on the first and hence both a lower WTP for Insurance and a higher WTP for Protective Devices when elicited second. The last column in Table I presents the results of the test of multiple question anchoring. Whatever the equality test, we never reject the null assumption of equality (lowest p-value=.445).

[a] WTPs equal to the median are equally split between the two groups. [b] No standard KS test is available for paired data.

Then, to more accurately account for individual characteristics, we estimate Eq. (1) by regressing the WTPs on a set of covariates, plus a dummy controlling for scenario ordering. First, we test for independence of covariates across groups (see the p-values in Appendix B) and reject independence only for preference for the Protective Devices scenario (*PrefProtective*), despite random assignment across respondents. Consequently, we force this variable into the parsimonious model presented in Table II, which keeps only variables significant at 10% threshold. Then, we use cluster-robust standard errors at the municipality level to account for potential correlation due to survey design (Bhattacharya, 2005) and unobservable characteristics. Moreover, we test for endogeneity between the perceived likelihood of being flooded in the next 10 years (*ProbaFlood*) and each of the WTPs.[2] We cannot reject endogeneity (p-values $< .05$), so we use instrumental variable (IV) estimators to obtain unbiased estimates.

The results in Table II confirm that scenario order is never significant: whatever the

[2] We thank a reviewer for drawing our attention to this issue.

Variables	Model 1 WTP for Insurance	Model 2 WTP for Protective Devices
Order (=1)	4.004 (.829)	-7.757 (.575)
Socio demographic		
Income	0.029 (<.001)	0.031 (.017)
Flood-related risk		
Inform (=1)	59.366 (.043)	51.78 (<.001)
NbrInfo	20.91 (<.001)	18.725 (<.001)
ProbaFlood	8.152 (<.001)	6.361 (<.001))
PrefProtective (=1)	21.62 (.590)	17. 40 (.464)
Attitudinal		
Impatience	-9.036 (.084)	-7.054 (.025)
RiskTolerance	51.74 (<.001)	29.67 (<.001)
Constant	-106.6 (<.001)	-114.5 (<.001)
Observations	263	263
Adjusted R^2	0.176	0.291

Table II: Model estimates for the CV survey on Flooding

scenario order, we cannot reject the absence of anchoring (p-value greater than .575). This also holds in strictly parsimonious models without forcing *PrefProtective* (not shown, p-value greater than .626) as well as in all models estimated with different sets of control variables used as robustness checks (see Appendix C, p-value greater than .299).

The effects of individual determinants have the expected sign and are of comparable significance across models. *Income* has a positive effect of comparable magnitude in explaining both WTPs. Other standard socio-demographic variables (*Male*, *Age*, *Couple*, *Education*, *Child*) are not significant at the 10% threshold and are hence not shown in Table II, meaning they are weaker determinants of WTPs than flood-related variables. Indeed, respondents' information about flood risk (*Inform* and *NbrInfo)* is a significant and positive determinant of their level of WTP, as is the perceived likelihood of being flooded in the next 10 years (*ProbaFlood*). *Impatience* has an intuitive negative effect. Finally, *RiskTolerance* has a counter-intuitive positive effect, which suggests that risk aversion elicited through lotteries is a poor predictor of risk behaviours elicited in real-life situations.

P-values in parentheses are computed with cluster-robust standard errors by municipality. In each model, the variables used to explain *ProbaFlood* in the first stage (not shown) are all exogenous variables plus the following instruments: municipality dummies and *HousingRisk.*

Overall, both the unconditional and the conditional tests provide evidence of absence of multiple question anchoring. However, because this CV survey on Flooding exclusively uses the CPC format, we cannot with certainty attribute the absence of anchoring to the use of CPC. It may be due to other aspects of the survey (no huge difference in WTP across scenarios, enough time given to respondents to set their WTP at each question, etc.). We thus turn to the CV survey on Social Insurance to isolate the effect of CPC on anchoring by comparing it with two other elicitation formats within a multiple question framework.

2.6 Evidence on the CPC's relative anchoring reduction potential

2.6.1 Survey design of the CV survey on social insurance

The second CV survey was conducted between August and September 2013, on Tunisian citizens not covered by – nor benefiting from - any social insurance scheme, i.e. more likely to be young, unemployed or informal workers (see Abu-Zaineh et al., 2013, 2014; Makhloufi et al., 2015). Two sampling locations in eight Tunisian governorates were consequently chosen so as to target these citizens: the "Souk", characterized by the high presence of informal activities, and the public squares where many peaceful demonstrations involving unemployed youths were organized after the so-called "*Arab Spring*".

Among the initial sample of 456 respondents surveyed using face-to-face interviews, 30 refused to answer the CV module. The remaining 426 were randomly split into three mutually exclusive and equal groups differing in the WTP elicitation format used (OE, PC and CPC, see Figure 2 for the latter). All respondents answered the same questionnaire (pre-tested for wording and choice of the number, range and values of the bids).

The valuation task started by asking respondents their willingness-to-join and their quarterly WTP for a Voluntary pre-payment Health Insurance Scheme (VHIS) made available to them. This was similar to the existing mandatory health insurance scheme currently run by the National Health Insurance Fund (known as "*Caisse Nationale d'Assurance Maladie*", CNAM, see Appendix D).

Then, all respondents were asked their willingness-to-join and their quarterly WTP for a Voluntary Pension Insurance Scheme (VPIS) available in addition to the VHIS. This also mimicked the existing mandatory retirement scheme for the self-employed entitled to the National Social Security Fund (known as "*Caisse Nationale de Sécurité Sociale*", CNSS, see Appendix D). 26 respondents answered "No" to the willingness-to-join question for both schemes, 197 for at least one scheme, leaving us with 203 respondents exposed twice to the elicitation question.

The lack of random assignment in the scenario prevented us from using the approach used in the Flooding survey, but the random assignment of the elicitation formats enabled us to test relative CPC performance regarding multiple anchoring.

2.6.2 Results

We explicitly take into account the potential interdependencies / anchoring between elicitations for the two social insurance schemes. Appendix E provides descriptive statistics on respondents' WTPs, socio-economic, socio-demographic and health characteristics (see Chanel et al., 2017 for a detailed presentation of the survey).

Results of the parsimonious (i.e. keeping only variables significant at 10% threshold) OLS models are reported in Table III. We use cluster-robust standard errors at interviewer level to account for potential correlation of unobservable characteristics specific to the interviewer or to the geographical area the interviewer was assigned to.[3] Model 3 estimates Eq. (2) and shows a significant and positive anchoring effect on the previously elicited WTP with a coefficient of about 0.526 (p-value=.003), which means that, on average, respondents anchor at 53% on their WTP previously elicited for VHIS. The OE and PC formats do not lead to WTPs that are significantly different from those elicited using the CPC format.

Cluster-robust standard errors are used at interviewer level. P-values in parentheses

In Model 4, we replace the overall anchoring term by three elicitation format-specific anchoring terms, and we find a significant and positive effect for each of them, larger for OE (0.708, p-value=.021) than for CPC (0.337, p-value<.0001) and PC (0.474, p-value=.012). The difference in anchoring is significant between OE and PC estimates (p-value=.0607) and OE and CPC (p-value=.0966) but not between CPC and PC (p-value=.1949).

Regarding the determinants, we find a positive effect of household income on WTP in Model 4, and evidence of interviewer effects in both models. Respondent's age has a significantly positive quadratic effect (with a maximum at 39 and 40, depending on the model), while education has a positive effect: *NoSchool* decreases WTP. Finally, living in a rural governorate negatively affects WTP, and the effect of having at least one outpatient consultation appears positive. Sensitivity analyses are provided in Appendix

[3] Only five interviewers covered the eight sample locations, generating strongly imbalanced distributions of the *Rural*, *Disadvantaged governorate* and *Sample point* variables by interviewer. To avoid high collinearity, we do not use these three spatially related variables in the same model.

Variables	Model 3 WTP for VPIS	Model 4 WTP for VPIS
Elicitation_OE (=1)	0.533 (.887)	-14.885 (.054)
Elicitation_PC (=1)	-5. 123 (.187)	-11.024 (.078)
Elicitation_CPC (=1)	(ref)	(ref)
WTP-VHIS	0.526 (.003)	-
WTP-VHIS x Elicitation_OE	-	0.708 (.021)
WTP-VHIS x Elicitation_PC	-	0.474 (.012)
WTP-VHIS x Elicitation_CPC	-	0.337 (<.001)
Survey		
Interviewer #2 (=1)	-21.213 (.002)	-19.531 (.005)
Interviewer #5 (=1)	-	8.528 (.001)
Socio demographic		
Income	-	0.0083 (.036)
Age	1.667 (.085)	1.524 (.059)
(Age)^2	-0.021 (.084)	-0.0190 (.066)
Rural (=1)	-8.646 (.033)	-7.921 (.058)
NoSchool (=1)	-22.280 (.005)	-17.477 (.016)
Health respondent		
Outpatient (=1)	9.924 (.027)	7.402 (.060)
Constant	-4.436 (.770)	-0.667 (.955)
Observations	203	203
Adjusted R^2	.5045	.5314

Table III: Model estimates for the CV survey on Social Insurance

F, with models that consecutively use different sets of control variables for Models 3 and 4: survey-specific, socio-demographic, specific to respondent's health, specific to the health of respondent's family members and others (respondent's risk aversion and reasons for not yet having a health insurance scheme). They confirm the results obtained with the parsimonious models.

2.7 Discussion

In the CV survey on Flooding, which uses the CPC format alone but randomly changes the order of scenario presentation, we found no evidence of anchoring on the first elicited WTP. In the CV survey on Social Insurance, which randomly uses three elicitation formats, we found greater anchoring on the first WTP with the OE format than with the two PC-type formats. This suggests that respondents may rely on different heuristic decisions when stating WTP in the OE and in the two PC formats (Hanemann, 1996; Welsh and Poe, 1998; Frör, 2008). A possible explanation is that answering an OE question is not typical of purchasing decisions, because the respondent has to set the price. S/he thus needs to reflect deeply before giving an amount. Faced with a second question related to the first, s/he may therefore rely on his/her first answer, giving an amount related to the previous one rather than again reflecting deeply. The PC formats, by providing the amounts, along with a visual aid, mimic real-life decisions (i.e. as price-taker), thereby facilitating the construction of the evaluation. This is also consistent with Van Exel et al. (2006): "*the anchoring and adjustment process often involves a great deal of inertia. People tend to hold on to their anchor and adjustment is typically insufficient, so that the final estimate is pulled toward the anchor* (p. 841)". Thus, when there is a second question, it is easier to change the amount by assessing how much better (or worse) the second good proposed is than the first one. The lesson here is that PC formats need to include provisions to limit anchoring effects from multiple successive elicitation questions. The CPC, being presented in a new random position to each respondent at every question, is a first attempt to limit anchoring. Longo et al. (2015) suggest offering respondents the opportunity to revise their WTP at the end of the multiple elicitation sequence, to help them better account for differences across the goods assessed.

A survey specifically designed to test whether the CPC reduces anchoring bias in multiple question CV surveys would use several elicitation formats, including several versions of the CPC (i.e. various settings of bid amounts, numbers, spacing (constant or increasing), ordering (ordered vs. non ordered)), with several scenarios successively proposed in varying order. This would however require a large respondent sample to obtain sufficient statistical accuracy, and would be expensive. Incidentally, a larger sample size would allow multilevel models to be used, thus better accounting for clustering within a community of respondents (by city in the flooding survey, or by governorate in the social insurance survey). The aim here was more modest: to assess, by exploiting the results of two already existing surveys, whether the CPC helps reduce anchoring. This paper provides evidence that it does.

Finally, it should be noted that in both applications, our analyses focus exclusively on differences across elicitation formats with respect to anchoring. Because respondents answered a willingness-to-join question before the elicitation format was used, we purposely only consider those who did not answer 'No'. However, while this is consistent with our intentions, respondents answering 'No' should be accounted for in any modelling aimed at predicting WTP, in particular by disentangling protest WTP from true null WTP based on (closed-ended) debriefing questions on the reasons for refusing to join.

Glossary

CNAM Caisse Nationale d'Assurance Maladie

CNSS Caisse Nationale de Sécurité Sociale

CPC Circular Payment Card

CV Contingent valuation

DC dichotomous choice

IV instrumental variables

OE Open-Ended

PACA Provence Alpes Côte d'Azur

PC Payment Card

TND Tunisian Dinar

VHIS Voluntary Health Insurance Scheme

VPIS Voluntary Pension Insurance Scheme

WTP Willingness to Pay

Bibliography

Abu-Zaineh M, Romdhane HB, Ventelou B, Moatti J-P, Chokri A. 2013. Appraising financial protection in health: the case of Tunisia. *Int J Health Care Finance Econ* **13(1)**: 73–93.

Abu-Zaineh M, Arfa C, Ventelou B, Romdhane HB, Moatti J-P. 2014. Fairness in healthcare finance and delivery: What about Tunisia? *Health Policy Planning* **29(4)**: 433–442.

Anderson L R, Mellor JM. 2009. Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. *J Risk Uncertain* **39**: 137–160.

Araña JE, León, CJ 2007. Repeated dichotomous formats for eliciting Willingness to Pay: Simultaneous estimation and anchoring effects, *Environ Resour Econ*, **36(4)**, 75-497.

Barsky RB, Juster FT, Kimball MS, Shapiro MD. 1997. Preference parameters and behavioral heterogeneity: an experimental approach in the health and retirement study. *Quarterly J Econ* **112(2)**: 537-579.

Bhattacharya D. 2005. Asymptotic inference from multi-stage samples. *J Econometrics* **126**: 145-171.

Boyle K, Johnson FR, McCollum D. 1997. Anchor and adjustment in single-bounded dichotomous-choice questions. *Amer J Agric Econ* **79(5)**: 1495-1500.

Boyle K, Welsh, M, Bishop R. 1993. The role of question order and respondent experience in contingent-valuation studies. *J Environ Econ Manag* **25**: 80-99.

Cameron TA, Huppert DD. 1989. OLS versus ML estimation of non-market resource values with payment card interval data. *J Environ Econ Manag* **17**: 230-246.

Carson, RT, Mitchell RC. 1995. Sequencing and Nesting in Contingent Valuation Surveys. *J Environ Econ Manag* **28 (2)**: 155–73.

Carson RT, Groves T. 2011. Incentive and information properties of preference questions commentary and extensions. In International Handbook on Non-market Valuation, Bennett J (ed.), Edward Elgar: Northampton.

Chanel O, Chichilnisky G, Massoni S, Vergnaud J-C, Vincent Lyk-Jensen S. 2013. Décision en présence d'incertitude et d'émotions face à des risques de catastrophes naturelles, Final Report ANR-08-RISKNAT-007-01, Greqam: Marseille.

Chanel O, Makhloufi K, Abu-Zaineh M. 2017. Can a circular payment card format effectively elicit preferences? Evidence from a survey on mandatory health insurance scheme in Tunisia. *Appl Health Econ Health Policy* 15(3), 385-398.

Chien, YL, Huang C, Shaw D. 2005. A general model of starting point bias in double-bounded dichotomous contingent valuation surveys. *J Environ Econ Manag* **50**: 362–77.

Cook J, Jeuland M, Maskery B, Whittington D. 2012. Giving stated preference respondents "time to think": results from four countries. *Environ Resour Econ* **51 (4)**: 473-496.

Covey J, Loomes G, Bateman I. 2007. Valuing risk reductions: testing for range biases in payment card and random card sorting methods. *J Environ Planning Manag* **50(4)**: 467-482.

De Ridder A, De Graeve, D. 2005. Order bias in estimates of willingness to pay for drugs to treat attention-deficit/hyperactivity disorder. *Eur J Health Econ* **6**: 146–151.

Deronzier P, Terra S. 2006. Bénéfices économiques de la protection contre le risque d'inondation. Document de travail D4E, Série Etudes 06-E05, MEDE: Paris.

DeShazo, J. 2002. Designing transactions without framing effects in iterative question formats. *J Environ Econ Manag* **43**: 360–85.

Flachaire E, Hollard G (2007) Starting-point bias and respondent uncertainty in dichotomous choice contingent valuation surveys. *Resour Energy Econ* 29:183–194

Frör O. 2008. Bounded rationality in contingent valuation: empirical evidence using cognitive psychology. *Ecol Econ* **68**: 570-581.

Hanemann WM. 1996. Theory versus data in the contingent valuation debate, In

The contingent valuation of environmental resources: methodological issues and research needs. Bjornstad J, Kahn JR (eds.), Elgar: Brookfield.

Herriges J, Shogren, J. 1996. Starting point bias in dichotomous choice valuation with follow-up questioning. *J Environ Econ Manag.* **30**: 112-131.

Kanninen B. 1995. Bias in discrete response contingent valuation. *J Environ Econ Manag* **28**: 114-125.

Longo A, Hoyos D, Markandya, A. 2015. Sequence Effects in the Valuation of Multiple Environmental Programs Using the Contingent Valuation Method. *Land Econ* **91(1)**: 20–35

Luchini S, Protière C, Moatti J-P. 2003. Eliciting several willingness to pay in a single contingent valuation survey: application to health care. *Health Econ* **12**: 51–64

Mahieu P-A, Wolff FC, Shogren J, Gastineau P. 2017. Interval bidding in a distribution elicitation format, *Appl Econ* 49(51), 5200-5211.

Makhloufi K, Ventelou B, Abu-Zaineh M. 2015. Have health insurance reforms in Tunisia attained their intended objectives? *Int J Health Econ Manag* **15(1)**: 29-51.

Mitchell RC, Carson RT. 1989. *Using surveys to value public goods: the contingent valuation method*. Johns Hopkins University Press: Baltimore.

O'Conor, RM, Johannesson M, Johansson P-O. 1999. Stated preferences, real behaviour and anchoring: Some empirical evidence. *Environ Resour Econ* **13**: 235–48.

Payne, J.W., Schkade, D.A., Desvousges, W.H. and Aultman, C., 2000. Valuation of multiple environmental programs. *J Risk Uncertain* **21(1)**: 95-115.

Powe NA Bateman I.J. 2003. Ordering effects in nested 'top-down' and 'bottom-up' contingent valuation designs *Ecol Econ* **45** 255-270.

Protière C, Donaldson C, Luchini S, Moatti J-P, Shackley P. 2004. The impact of information on non-health attributes on willingness to pay for multiple health care programmes. *Soc Sci Med* **58(7)**: 1257-1269.

Selart M. 1996. Structure compatibility and restructuring in judgment and choice. *Organ Behav Hum Process* **65(2)**: 106–116.

Tversky A, Kahneman D. 1974. Judgement under Uncertainty: Heuristics and biases. *Science* **185**: 1124–1131.

Van Exel NJA, Brouwer WBF, van den Berg B, Koopmanschap MA. 2006. With a little help from an anchor. Discussion and evidence of anchoring effects in contingent valuation. *J Socio-Econ* **35**: 836–853.

Veisten K., Hoen H., Starand J. 2004. Sequencing and the Adding-up Property in Contingent Valuation of Endangered Species: Are Contingent Non-Use Values Economic

Values? *Environ Resour Econ* **29**: 419–433.

Wang H. 1997. *Contingent Valuation of Environmental Resources: A Stochastic Perspective.* Ph.D. Dissertation, University of N. Carolina at Chapel Hill., School of Public Health.

Welch, B. L. 1947. The generalization of "Student's" problem when several different population variances are involved. *Biometrika.* 34 (1–2): 28–35.

Welsh MP. and Poe GL. 1998. Elicitation effects in contingent valuation: comparisons to a multiple bounded discrete choice approach. *J Environ Econ Manag.* **36**: 170-185.

Whitehead JC. 2002. Incentive incompatibility and starting-point bias in iterative valuation questions. *Land Econ* **78**: 285–97.

Wilson T, Houston C, Etling K, Brekke N. 1996. A new look at anchoring effects: Basic anchoring and its antecedents. *J Exp Psychol: General* **125**: 387–402.

Wooldridge JM. 2002. *Econometric analysis of cross section and panel data* (2^{nd} ed.). MIT Press: Cambridge.

Yang S-H, Hu W, Mupandawana M, Liu Y. 2012. Consumer willingness to pay for fair trade coffee: a Chinese case study. *J Agri Appl Econ* **44(1)**: 21–34.

2.A Hypothetical scenarios of the CV survey on Flooding

Introduction by Interviewer:

You are going to be the main actor in our fictitious scenarios. You will have to take the best decision regarding your housing. Only your opinion matters, there is no wrong or right answer. Not everyone is fully aware of the way the flood insurance system works, so we present it briefly. In France, every third-party liability insurance policy regarding fire or damage includes a mandatory contribution known as CatNat. To benefit from this type of compensation in the event of flood, the flood event must have been declared a 'natural catastrophe' by joint ministerial decree and the goods (property and belongings) must be insured. Compensation will be subject to a 380 euros deductible. Personal injuries are not covered by the CatNat system. They are covered either by a personal insurance policy, or by the national government if a civil servant (administrative or elected) can be held responsible for the occurrence of the flood event.

Protective devices (randomly proposed first to half the sample)

Let us imagine that the CatNat insurance still covers flood-related events. Your current insurance contract still covers all other types of events, and your premium remains unchanged. Imagine that the national government creates a Flood Management Fund to finance protective devices against flood. Building dikes, water retention ponds or improving rainwater evacuation networks would reduce the height and speed of water and would completely eliminate the risk of flooding in your commune. This work will only be carried out if the population involved contributes to the Flood Management Fund. We would like to know how much maximum you would be willing to pay per year to this Fund.

Insurance scenario (randomly proposed first to the other half sample)

Let us imagine that the CatNat insurance no longer covers flood-related events. Your current insurance contract still covers all other types of events, and your premium remains unchanged. Imagine that the national government creates a Flood Management Fund that is now the only flood-related damage compensation system. It allows you to be compensated in case of personal, property or material damage. You can freely choose to contribute or not to this Flood Management Fund, but if you do not contribute, you will not be compensated in case of flood-related damage. We would like to know how much maximum you would be willing to pay per year to this Fund."

2.B Table B1 Summary statistics of the CV survey on Flooding (N=277)

Variable	Description	**Mean**	**Std. Dev.**	**Min.**	**Max.**	P-value*
WTP-Insurance	WTP for Insurance scenario (yearly, in euros)	112.76	143.29	0	1300	.523
WTP-Protective	WTP for Protective Devices scenario (yearly, in euros)	98.306	145.54	0	1500	.405
Respondents characteristics						
Male	Gender (Male=1)	47.3%		0	1	.771
Age	Age (in years)	50.79[a]	17	16	94	.914
Income	Monthly income of the respondent (in euros)	1389.2	944.87	0	8000	.753
Couple	In a relationship (=1)	56.9%	-	0	1	.232
Child	Has at least one child (=1)	36.1%	-	0	1	.105
JuniorEd	Junior high school Ed. (=1)	63.2%	-	0	1	.104
SeniorEd	Senior high school Ed. (=1)	12.6%	-	0	1	.690
VocatEd	Vocational Education (=1)	11.2%	-	0	1	.105
FurtherEd	Further Education (=1)	13%	-	0	1	.228
Flood-related risk variables						
HousingRisk	Living on the ground floor or in a house (=1)	57.8%	-	0	1	.306
Inform	Looked for information about flood risk (=1)	15.5%	-	0	1	.368
NbrInfo	Number of media known for information about flood risk (integer)	2.47	1.42	1	7	.271
ProbaFlood	Perceived likelihood of being flooded in the next 10 years (in %)	7.20	12.71	0	100	.650
PrefInsurance	Preference for the Insurance scenario (=1)	24.2%	-	0	1	.326
PrefProtective	Preference for the Protective devices scenario (=1)	39.0%	-	0	1	.013
Attitudinal variables						
Impatience	Preference for the present score (1-7 score)	2.95	2.76	0	7	.254
RiskTolerance	Risk tolerance score (1-4 score)	1.65	0.92	1	4	.872
Happy	Declared subjective well-being (0-10 score)	6.77	1.98	0	10	.124
Survey specific variables						
Order	WTP for the Insurance scenario is elicited first (=1)	42.6%	-	0	1	
Municipality#1-4	Dummy variables for each of the 4 municipalities					

P-values of the test of independence of each variable with the order of the scenario. Continuous variables have been discretized in 4 groups based on quantiles.

[a] Because respondents in Vaison-la-Romaine had to be living there when the flood occurred in 1992, the sample mean age is no longer representative of the PACA population age.

2.C Robustness checks for the CV survey on Flooding

	Regression models on WTP for Insurance					Regression models on WTP for Protective devices				
Models	(1a)	(1b)	(1c)	(1d)[a]	(1e)[a]	(2a)	(2b)	(2c)	(2d)[a]	(2e)[a]
Variables										
Order (=1) *(p-value)*	2.474 (.851)	14.57 (.299)	9.827 (.572)	2.131 (.931)	10.070 (.626)	-6.042 (.572)	6.225 (.497)	3.217 (.801)	-5.222 (.738)	0.109 (.992)
Survey var.	-	YES	YES	-	-	-	YES	YES	-	-
Socio var.	-	-	YES	YES	YES	-	-	YES	YES	YES
Flood-related risk variables	-	-	-	YES	YES	-	-	-	YES	YES
Attitudinal variables	-	-	-	-	YES	-	-	-	-	YES
# of variables	1	4	13	15	18	1	4	13	15	18
Adjusted R^2	-.004	.063	.127	.080	.175	-.003	.073	.167	.216	.300
Observations	277	277	273	273	263	277	277	273	273	263

Cluster-robust standard errors are used at municipality level. [a] Instrumental variable estimation, endogeneity first stage equations for *Probaflood* not shown, municipality dummies and *HousingRisk* removed from the explanatory variables in the WTP structural equations.

2.D Hypothetical scenarios of the CV survey on Health and Pension schemes

(translated from Arabic)

Introduction by Interviewer:

No one is safe from injury or illness. Valuation of WTP is very important, allowing the implementation of a new Voluntary Health Insurance Scheme (VHIS). The VHIS covers only healthcare benefits and is not conditional on the exercise of a professional activity (employed or self-employed). It covers the healthcare expenditures of the insured and his/her household members. It offers a package of healthcare services identical to those offered by the public scheme currently run by 'CNAM'.

We will now ask you questions on the amount that you are willing-to-pay to join this new voluntary scheme. The amount that you are willing to pay represents the importance that you attach to the health insurance scheme and to healthcare services in general. Please note that this amount will reduce your expenditure on other items.

Note to the interviewer: [Please give the interviewee the blue list that describes the scheme under consideration and ask her/him to take time to reply to all the questions]

Pension insurance scheme (VPIS scenario, translated from Arabic)

Introduction by Interviewer:

In addition to the Health Insurance, a voluntary Pension Insurance Scheme (VPIS) might also be available independently. Before answering the following questions, please consider your potential need for a monthly income when you become elderly and are unable to work.

2.E Table E1 Descriptive statistics of the CV survey on Social Insurance (n= 203)

Variable	Description	Mean	Std. Dev.	Min.	Max.
WTP-VHIS	WTP for VHIS (quarterly, in TND)	41.867	26.351	5	170
WTP-VPIS	WTP for VPIS (quarterly, in TND)	44.532	28.798	5	180
Respondent characteristics					
Male	Male (=1)	67.4%	-	0	1
Age	Age (in years)	35.763	9.247	20	70
Household size	Number of household members (integer)	2.852	2.131	1	9
Child	Living with at least one child (=1)	18.7%	-	0	1
Elderly	Living with one person over 65 years (=1)	6.8%	-	0	1
Married	Married (=1)	46.3%	-	0	1
NoSchool	No schooling (=1)	0.9%	-	0	1
Elementary	Primary school (=1)	24.1%	-	0	1
Secondary	Secondary education (=1)	55.1%	-	0	1
Higher Ed.	Higher education (=1)	19.7%	-	0	1
Income	Monthly household income (in TND)	484.354	370.212	50	3000
Work	Employed /self-employed (=1)	79.8%	-	0	1
Rural	Living in rural area (=1)	18.2%	-	0	1
DisadvGov	Living in disadvantaged governorate[a] (=1)	45.3%	0.499	0	1
Other variables					
NonDeclared	Uninsured due to no declared work (=1)	51.7%	-	0	1
Administration	Uninsured due to administrative procedures (=1)	33.4%	-	0	1
RiskAversion	Risk-averse [b] (1-6 score)	5.233	1.345	1	6
Respondent-specific health variables					
HealthStatus	Good self-reported health status (=1)	78.8%	-	0	1
Outpatient	At least one outpatient care during the last 3 months (=1)	39.9%	-	0	1
Inpatient	At least one hospitalization during the last 8 months (=1)	11.8%	-	0	1
Chronic	Reports a chronic condition (=1)	14.7%	-	0	1
Afford	Can afford health services (=1)	33.9%	-	0	1
Smoking	Consuming tobacco products (=1)	47.7%	-	0	1
Health variables specific to the family members of the resp.					
OutpatientHous	At least one outpatient care in household during the last 3 months (=1)	53.6%	-	0	1
InpatientHous	At least one hospitalization in household during the last 8 months (=1)	17.7%	-	0	1
OtherConsHous	At least one other medical consultation in household during the last 3 months (=1)	94.1%	-	0	1
ChronicHous	One household member reports a chronic condition (=1)	21.6%	-	0	1
Survey specific variables					
PublicSquare	Sample point is public square (=1), vs. informal market (=0)	41.3%	-	0	1
Interviewer#1-5	Dummy variables for each of the 5 interviewers	-	-	-	-

* At the time of the survey, 1 Tunisian Dinar (TND) = € 0.455 = $ 0.605.
[a] According to decree n° 2008-387 of February 11, 2008.
[b] Six modalities were generated according to the method of Barsky *et al.*, (1997).

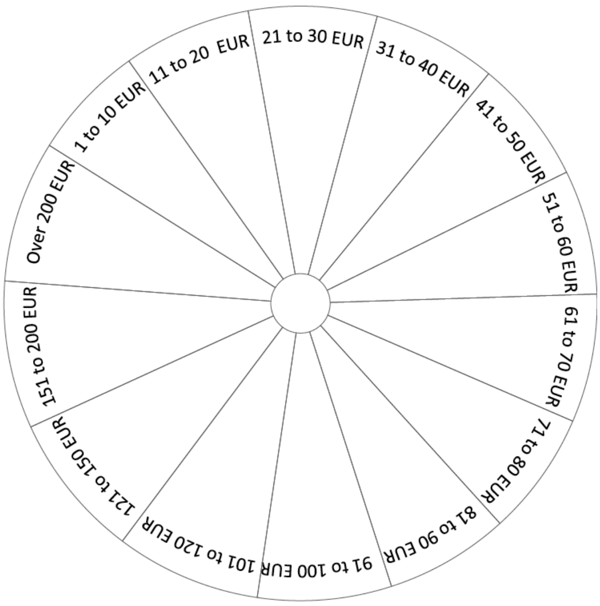
2.F Table F1 Robustness checks for the CV survey on Social Insurance

	Regression models on WTP for VPIS									
Models	(3a)	(3b)	(3c)	(3d)	(3e)	(4a)	(4b)	(4c)	(4d)	(4e)
Variables										
Elicitation_OE (=1)	-.874	.961	1.094	4.332	3.788	-16.559*	-17.863	-16.570	-12.848	-12.516
Elicitation_PC (=1)	-6.083	-5.413	-5.668	-3.731	-3.315	-10.093	-13.887*	-14.189*	-13.037	-10.674
WTP-VHIS	.595***	.568***	.575***	.578***	.580***	-	-	-	-	-
WTP-VHIS x *Elicitation_OE*	-	-	-	-	-	.824***	.800**	.790**	.777**	.782**
WTP-VHIS x *Elicitation_PC*	-	-	-	-	-	.532***	.555***	.581**	.608***	.577***
WTP-VHIS x *Elicitation_CPC*	-	-	-	-	-	.463***	.374***	.392***	.397***	.414***
Survey var.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Socio var.		YES	YES	YES	YES		YES	YES	YES	YES
Health (resp.)			YES	YES	YES			YES	YES	YES
Health (family)				YES	YES				YES	YES
Other var.					YES					YES
# of variables	8	21	26	30	37	10	23	28	32	39
Adjusted R^2	.4933	.5014	.5086	.5437	.5428	.5094	.5241	.5274	.5603	.5579
Observations	203	203	203	203	201	203	203	203	203	201

Cluster-robust standard errors are used at interviewer level. * if p<0.05, ** if p<0.01, *** if p<0.001

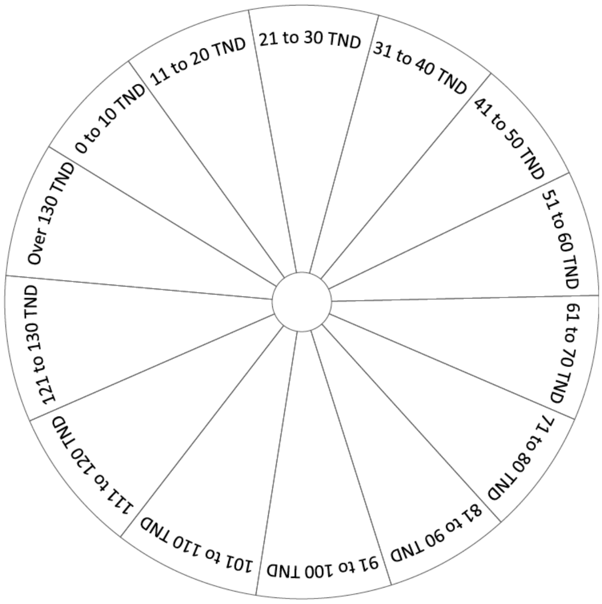
2.G Figure 1 Circular payment card used in the Flooding survey

(English translation)



2.H Figure 2 Circular payment card used in the Social Ins. survey

(English translation)


21 to 30 TND
31 to 40 TND
41 to 50 TND
51 to 60 TND
61 to 70 TND
71 to 80 TND
81 to 90 TND
91 to 100 TND
101 to 110 TND
111 to 120 TND
121 to 130 TND
Over 130 TND
0 to 10 TND
11 to 20 TND

Chapter 3

How does trust in institutions affect protest responses in environmental valuation surveys?

Abstract: In environmental valuation surveys, some respondents state a zero willingness to pay that does not reflect their preferences. Among the rationale for such a value, I focus on mistrust in institutions. The results in the existing literature depends on whether the effect is identified by respondents' statements or by a random assignment of the managing institution. This paper tackle this issue by using a new identification strategy. By merging country data on perception of institutions with meta-data from environmental valuation surveys, I am able to estimate the effect of trust in institutions on the surveys' protest rates, wiping out the effect of each studies' specificities. Results show that trust in institutions is not a significant determinants of protest responses.

3.1 Introduction

Stated preference studies use surveys to elicit preferences for non market goods. In these surveys, the respondents state how much they would be willing to pay (WTP) for the non market good provision. However, some of them refuse to state their true preferences and give a zero amount instead, for various reasons. For instance, they think someone else should pay for the good, the choice of payment vehicle is not adequate or the scenario is not credible enough. Usually, these respondents must justify their zero amount by answering follow-up questions, which enable the practitioner to detect these so called "protest respondents".

Protest zeros are then different from "valid zeros", i.e. respondents who state a zero WTP which is the result of a decision that depends solely on their preferences and their income (either their marginal utility for the environmental good is zero, either they face a budget constraint preventing them to have a positive WTP).[1] Besides, protest responses should also be distinguished from missing responses. A missing response arises when a respondent refuse explicitly to answer the WTP question.

Even if the practitioner is able to detect these protest responses and remove them from the sample, the distribution of protest respondents is very unlikely to be random. Then, the samples on which the aggregate WTP is computed are no more representative. This poses a threat to some of the fundamental hypothesis of cost-benefit analysis. For this reason, it is important to understand the motivations of protest behaviors and to find ways to mitigate them.

A lot of studies explore the determinants of protest responses. Some determinants are related to survey characteristics (see Meyerhoff and Liebe (2010) for a meta-study). Others are related to individual characteristics of the respondents. While some of these individual determinants come from personal preferences and attitudes ("I can't put a price on nature"), others depend on the respondent's perception of the managing authority ("I don't trust the actor providing the good"). However, the effect of trust in institutions on the protest responses is not clear yet, and could be better understood by tackling two issues.

First, the effect differs depending on how we look at the institutional determinant of the protest responses. On the one hand, several studies in developing countries find that mistrust in the government is often the reason stated by the respondents for their protest responses (Hadker et al. (1997); Cunha-e Sá et al. (2012)). On the other hand, two recent articles (Oehlmann and Meyerhoff (2017) and Remoundou et al. (2012)) look at the impact of a change in the authority in charge of the project on stated WTP, finding no significant effect. Remoundou et al. (2012) also test for an impact on the protest rate, but they don't find any. Overall, we face a paradox: in some studies respondents justify their protest responses with respect to the institutions, but other studies that look at the effect of changing the managing institution in the survey don't find any impact.

Second, since most studies are conducted in one place at one time, the institutional context that surrounds the respondents doesn't change. Therefore, these studies cannot

[1]Sometimes, the respondent is first asked whether or not he wants to be part of the market presented in the fictional scenario (screening mechanism). In this case, refusing to enter the marjet is considered a zero WTP that still needs to be classified as valid or protest. Although the use of a screening mechanism might affect the respondents behavior, no information about their use is included in the meta-data.

tell anything about the effect of the institutional factors that motivate the protest. The two studies mentioned above (Remoundou et al. (2012) and Oehlmann and Meyerhoff (2017)) propose a way to overcome this issue by changing the managing authority in the survey. However, as discussed by Remoundou et al. (2012), the effect of the managing institution is then study- and good-specific, allowing no inference on a more general effect of institutional context on protest behavior. Besides, a random assignment of managing institution might not be able to significantly affect a lack of trust in institutions by this survey design. Indeed, mistrust in the institutions could lead to a systematic defiance towards all public entities. Besides, a distrustful respondent may not believe that the stated managing authority in the survey is the true authority in charge. For these reasons, an identification of the effect of institutional factors through survey design should not be considered as a perfect substitute for an actual variation in the institutional context. Such a variation is used by Schläpfer and Bräuer (2007), who conduct two identical contingent valuation studies in Switzerland and Germany to account for the variations in perception of the framing. They find significant differences in the results between the two locations, which they suspect are coming from differences in countries institutions (habits of voting for local policies in Switzerland). However, their study does not investigate the effect of specific institutional characteristics on the protest rate.

In this paper I try to tackle these issues. I rely on meta-data on environmental valuation studies, merged with institutional variables. By using intra-country variations in these institutional variables, I am able to capture the effect of trust institutions on the protest rate, wiping out the effect of each studies' specificities.

The remainder of the paper proceeds as follows. Section 3.2 describes of the data, the choice of variables and the identification of the effect of the institutional variables. Section 3.3 presents the descriptive statistics and the regression results. Section 3.4 concludes.

3.2 Methods and Data

I use meta-data on environmental valuation studies merged with institutional variables to identify the effect of the institutional context on protest rate. In this section I describe the meta-data, the institutional data and the other variables used as controls.

The dataset was collected by Meyerhoff and Liebe (2010). It has observations for 254 independent samples from 157 different stated preference studies from 1988 to 2010 across 34 countries.[2] It contains information about elicitation methods, payment vehicles, survey methods, type of goods, protest rates, year and country of collection. The data was collected using the Google Scholar search engine and the web-pages of journals in the field of environmental economics. Only studies with reported number of protest responses and sufficient information about survey characteristics were used (see Meyerhoff and Liebe (2010) for more details).[3]

I merge the dataset by year and country with trust in institutions variables.

I measure trust in the institutions, using variables from the World Value Survey (WVS (2015)) and European Value Survey (EVS (2011)). For these two surveys, respondents from a representative sample of each country state whether they trust various institutions, particularly the government and civil services (from "A great deal" to "Not at all"). There are several waves of surveys for each country, which are not more distant than 10 years. Thus I match each stated preference survey with a WVS wave which is closer than 5 years at most. I compute the mean by country and year to get two global measures of trust in institutions. I then rescale the variables to be in a range from 0 to 100, where 100 stand for "A great deal" of trust in the government.

I use additional variables as control. I account for the main survey characteristics using the variables collected by Meyerhoff and Liebe (2010): payment vehicle, elicitation format and survey method. I also account for the nature of the environmental good. I add country fixed effects to control for any country specificity (e.g. cultural aspects). Consequently, I remove 10 surveys that were the only one in their country. The addition of country fixed effects means that the only variation left in the dependent variable comes from the characteristics of the surveys and the variability of context variables across time.[4]

I control for tax revenue, because respondents may feel like they give already too much money for the collectivity, thereby affecting their probability to protest. To do this, I use OECD data of Total Tax revenue as a percentage of GDP (OECD (2017)). I account for GDP per capita, since a correlation between income and the probability to protest has been repeatedly observed in stated preference studies.

The model can be written as follows:

$$p_{ijct} = x'_{ijct}\beta_1 + y'_{tc}\beta_2 + \alpha_c + e_{ijct}$$

[2] Several studies have split samples.

[3] The protest responses are detected based on answers of follow-up questions after a respondent stated a zero WTP.

[4] Intra-country variations could also have an impact, but are impossible to capture without losing almost all of the variability.

p_{ijct} and x_{ijct} are respectively the protest rate and the survey characteristics of the survey i in study j that was conducted in country c at year t. y_{tc} is a set of time varying country characteristics, including trust in institutions variables, α_c is the set of country fixed effects and e_{ijct} is the error term.

I estimate this model with three different methods. I use OLS regression with the log of the protest rates, firstly without weights and then weighting each observation by the inverse of the number of survey in the study, in order to give the same weight to all studies. In the third model I use a fractional logit model on the (untransformed) protest rates to account for the fractional nature of the dependent variable (Papke and Wooldridge (1996)). In all specifications, I use cluster standard errors by study.

3.3 Results

In this section, I first provide descriptive statistics of the data set. Then I present the main results.

3.3.1 Descriptive Statistics

Out of the full sample of 255 observations, I delete 15, either because of missing values or because the survey was the only one conducted in a given country. Figure 3.1 shows the number of surveys per country. US and UK are the most represented countries, with more than 30 surveys each. Western Europe countries are also present, but there are very few developing countries. Since a large share of stated preference studies are conducted in developed countries, the sample is biased towards them. This could be an issue in terms of external validity, but I should be able to provide valid findings for developed countries.

The first part of Table I shows the frequency of each survey characteristics, and the second part provides statistics for other variables. One can see that the tax payment vehicle (PV) is the most used. As explained above, the quality of the institutions may affect reactions to tax payments.

Regarding survey design (SD), phone interviews are the most frequent, followed by face to face and mail. There does not seem to be a consensus in the literature on the effect of particular survey design on the protest rate. Likewise, there is no clear evidence on the impact of a specific elicitation format (EF) on the protest rate.

On average, 20% of the stated preferences are not valid WTP. This rate can go up to 60%. This shows that protest responses should be a major concern for stated preference

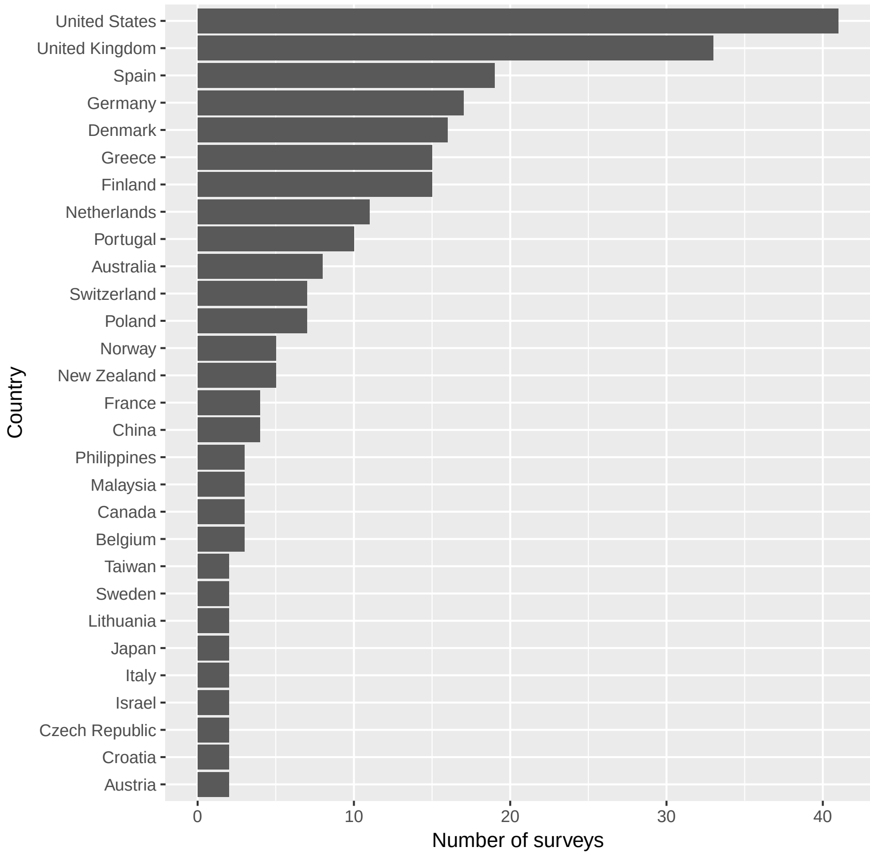


Figure 3.1: Number of observations by country

studies, and that it can greatly bias the aggregate WTP. Note that the values of the institutional variables are by no mean representative. This is only the values merged with the meta-data, so only for the countries and years corresponding to a valuation study. The mean of the trust in the government index is 69. This reflect the fact that most countries are developed countries, with a relatively high amount of trust in institutions. The values span from a quite low range (from 42 to 78) which also may come from the similarity of the countries in the sample size. Note that the fact that countries are similar in my sample is not necessarily an issue for identification, since I only use time variation in trust, not geographical variations. Table II shows the distribution of the type of goods in the meta-data. Nature and biodiversity protection, Species protection and Wetlands and Recreation are the most represented areas.

Variable	Mean	SD	Min	Max	N
proshare	0.173	0.113	0	0.593	240
Elicitation Method					
Choice Experiment	0.138	0.345	0	1	240
Dichotomous Choice	0.417	0.494	0	1	240
Open Ended	0.192	0.394	0	1	240
Payment Card	0.237	0.426	0	1	240
Other	0.017	0.128	0	1	240
Survey Method					
Other	0.067	0.25	0	1	240
Face to face	0.354	0.479	0	1	240
Mail	0.358	0.481	0	1	240
On site	0.096	0.295	0	1	240
Phone	0.071	0.257	0	1	240
Web	0.054	0.227	0	1	240
Other	0.067	0.25	0	1	240
Payment Vehicle					
Bill	0.204	0.404	0	1	240
Donation	0.092	0.289	0	1	240
Entrance fee	0.096	0.295	0	1	240
Fund	0.121	0.327	0	1	240
Other	0.125	0.331	0	1	240
Tax	0.363	0.482	0	1	240
Country variables					
Gouv	69.155	6.732	42.28	78.951	207
Civil_services	64.817	5.211	47.066	80.381	235
tax_revenue	33.73	6.282	23.017	48.984	227
gdp_capita	32.35	9.231	5.821	62.434	233

Table I: Summary statistics

Type of good	Freq.	Percent
Species protection	44	17.25
Wetland along rivers	38	14.90
Biodiversity	35	13.73
Recreation	28	10.98
Other	24	9.41
Landscape	20	7.84
Wetland (lakes, coastal)	16	6.27
Forest resources	15	5.88
Air quality	11	4.31
Water quality and groundwater	11	4.31
Agriculture	8	3.14
Tropical resources	5	1.96
Total	255	100

Table II: Types of goods in the sample

3.3.2 Regression Analysis

The different specifications are reported in Tables III for the trust in the government and Table IV for the trust in civil services. Model 1 is an OLS model with Country Fixed effects. In Model 2 I add dummies for the type of good. Model 3 is a Weighted least square models, that weight each survey based on the sample size. Finally, model 4 is fractional model accounting for the bounded nature of the dependent variable.[5] Finally I account for cluster effects between surveys coming from the same studies in the standard errors computations.

Regarding the effect of survey characteristics, the findings are roughly consistent with Meyerhoff and Liebe (2010). Sign and significance of coefficients are quite stable across specifications. The choice of payment vehicle seems to be the most important survey determinant on the probability to protest. The reference alternative is the bill, and it leads to the least protest responses. Fund and entrance fee have the biggest impact on protest rates, closely followed by tax. Surprisingly, donation leads to fewer protests than the other Payment vehicle (except bill). Elicitation formats do not have significant effects on the protest rate, in the first three models. However the WLS model detects a positive effect of Open Ended format and Payment cards. On-site survey tends to produce significantly less protest responses than face to face (at the respondent's residence) surveys, according to the WLS model.

The effect of trust in institutions is not significantly different from zero in all models,

[5]Because of the difference in scale of the dependent variable the estimated coefficients are much lower in magnitude than the other models.

for both government and civil services. This result suggest that the level of trust in institutions itself is not a determinant of the protest rate. In this regard, it is in line with the results from Remoundou et al. (2012).

Due to the small number of observations, the use of country fixed effect could cause identification problems due to a potential lack of remaining variability in the covariates. In order to test this issue, I estimate models keeping only countries with 5 or more and 10 or more surveys, both using the WLS model.

Results are reported in Table V. The coefficient of trust in institutions results are larger in magnitude compare to the standard models, but are still non-significant.

3.4 Discussion

This paper investigates the relationship between institutional variables and protest rate in environmental valuation studies. It provides insights to practitioners on how the protest rate can be affected depending on the country where a survey is conducted. Using meta-data merged with institutional variables and exploiting intra-country variations, I find that trust in the institutions is not a significant determinant of the protest behaviors. This result is in line with the results from Remoundou et al. (2012). In their discussion, they used caution when interpreting the results of absence of the effect of institutional context on the protest rate, because it might be case-study specific. The present paper reinforces their findings by wiping out the study specificities and the good specificity.

Using two different datasets, each one based on a different population might present some issues. First, the level of trust is measured based on a representative population from the whole country, while the population targeted by the valuation studies are often from more specific places. The level of trust in the institutions in the whole country might be different than at the place of study. Since precise location information is not reported in the meta-data, due to the difficulty to collect this data, it is not possible to match them with the specific population from the WVS. Therefore this leads to some errors in the measure of trust in institutions. Second, the institutions managing authority could also differ from the institutions for which the trust is measured in the World Value Survey. For instance, the managing institution could be a local authority. Since there is only a few institutions accounted for in the WVS study, that information about the institutions in charge is not collected in the meta-analysis, it is not possible to match the correct institution type. Again, this introduce some noise in the measure of trust in institutions. To conclude, the best way to improve the identification strategy proposed

VARIABLES	WLS	Fractional	WLS	Fractional
Payment vehicle				
Donation	0.801**	0.0619**	0.701**	0.0404
	(0.364)	(0.0273)	(0.290)	(0.0267)
Entrance fee	1.082**	0.101	1.142***	0.0994
	(0.419)	(0.0691)	(0.392)	(0.0620)
Fund	1.283***	0.0893**	1.309***	0.101***
	(0.237)	(0.0349)	(0.228)	(0.0324)
Tax	1.078***	0.0620**	0.964***	0.0571**
	(0.245)	(0.0258)	(0.231)	(0.0240)
Other	1.169***	0.103**	1.034***	0.102**
	(0.373)	(0.0504)	(0.334)	(0.0399)
Elicitation method				
Dichotomous Choice	0.292	0.0404	0.211	0.0286
	(0.214)	(0.0318)	(0.184)	(0.0262)
Open ended	0.343	0.0439	0.445*	0.0566
	(0.347)	(0.0456)	(0.253)	(0.0348)
Payment card	0.291	0.0192	0.296	0.0198
	(0.198)	(0.0264)	(0.178)	(0.0236)
Other	-0.108	-0.0374	-0.243	-0.0434
	(0.928)	(0.0670)	(0.914)	(0.0642)
Survey method				
Mail	0.155	0.0339	0.169	0.0426
	(0.253)	(0.0372)	(0.201)	(0.0268)
On site	-0.569*	-0.0375	-0.479*	-0.0244
	(0.322)	(0.0332)	(0.272)	(0.0315)
Phone	0.122	0.0113	0.139	0.0321
	(0.295)	(0.0377)	(0.253)	(0.0404)
Web	-0.109	-0.0166	-0.0860	0.00845
	(0.251)	(0.0364)	(0.221)	(0.0355)
Other	-0.413	-0.0401	-0.455	-0.0485
	(0.474)	(0.0462)	(0.417)	(0.0351)
Country variables				
GDP/capita	0.00344	-0.00176	0.00197	-0.000141
	(0.0242)	(0.00233)	(0.0154)	(0.00147)
tax_revenue	0.0301	-0.00525	0.0369	-0.000350
	(0.0641)	(0.00794)	(0.0587)	(0.00745)
Gouv	-0.0217	-0.00217		
	(0.0297)	(0.00351)		
Civil_services			-0.0151	-0.000378
			(0.0326)	(0.00381)
Constant	2.007		1.510	
	(3.500)		(2.688)	
Country FE	X	X	X	X
Type of good FE	X	X	X	X
Observations	190	196	216	222
R^2	0.568		0.535	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) WLS	(2) Fractional	(3) WLS	(4) Fractional
2.Payment_vehicle	0.701**	0.0404	0.701**	0.0404
	(0.290)	(0.0267)	(0.290)	(0.0267)
3.Payment_vehicle	1.142***	0.0994	1.142***	0.0994
	(0.392)	(0.0620)	(0.392)	(0.0620)
4.Payment_vehicle	1.309***	0.101***	1.309***	0.101***
	(0.228)	(0.0324)	(0.228)	(0.0324)
5.Payment_vehicle	1.034***	0.102**	1.034***	0.102**
	(0.334)	(0.0399)	(0.334)	(0.0399)
6.Payment_vehicle	0.964***	0.0571**	0.964***	0.0571**
	(0.231)	(0.0240)	(0.231)	(0.0240)
2.Elicitation_method	0.211	0.0286	0.211	0.0286
	(0.184)	(0.0262)	(0.184)	(0.0262)
3.Elicitation_method	0.445*	0.0566	0.445*	0.0566
	(0.253)	(0.0348)	(0.253)	(0.0348)
4.Elicitation_method	0.296	0.0198	0.296	0.0198
	(0.178)	(0.0236)	(0.178)	(0.0236)
5.Elicitation_method	-0.243	-0.0434	-0.243	-0.0434
	(0.914)	(0.0642)	(0.914)	(0.0642)
2.Survey_method	0.169	0.0426	0.169	0.0426
	(0.201)	(0.0268)	(0.201)	(0.0268)
3.Survey_method	-0.479*	-0.0244	-0.479*	-0.0244
	(0.272)	(0.0315)	(0.272)	(0.0315)
4.Survey_method	0.139	0.0321	0.139	0.0321
	(0.253)	(0.0404)	(0.253)	(0.0404)
5.Survey_method	-0.455	-0.0485	-0.455	-0.0485
	(0.417)	(0.0351)	(0.417)	(0.0351)
6.Survey_method	-0.0860	0.00845	-0.0860	0.00845
	(0.221)	(0.0355)	(0.221)	(0.0355)
gdp_capita	0.00197	-0.000141	0.00197	-0.000141
	(0.0154)	(0.00147)	(0.0154)	(0.00147)
tax_revenue	0.0369	-0.000350	0.0369	-0.000350
	(0.0587)	(0.00745)	(0.0587)	(0.00745)
Civil_services	-0.0151	-0.000378	-0.0151	-0.000378
	(0.0326)	(0.00381)	(0.0326)	(0.00381)
Constant	1.510		1.510	
	(2.688)		(2.688)	
Observations	216	222	216	222
R^2	0.535		0.535	

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table IV: Results for Trust in Civil Services

VARIABLES	(1) Gouv. (5+)	(2) Gouv. (10+)	(3) Civil (5+)	(4) Civil (10+)
Payment Vehicle				
Donation	0.722*	0.803**	0.734**	0.833**
	(0.385)	(0.387)	(0.312)	(0.339)
Entrance Fee	1.065***	1.265***	1.123***	1.271***
	(0.405)	(0.377)	(0.385)	(0.383)
Fund	1.302***	1.447***	1.360***	1.539***
	(0.284)	(0.299)	(0.263)	(0.245)
Tax	1.009***	1.199***	0.950***	1.134***
	(0.255)	(0.258)	(0.242)	(0.248)
Other	0.977**	1.330***	0.934**	1.278***
	(0.411)	(0.443)	(0.374)	(0.385)
Elicitation Method				
Dichotomous choice	0.219	0.208	0.179	0.260
	(0.217)	(0.244)	(0.183)	(0.206)
Open Ended	0.348	0.319	0.430	0.460
	(0.339)	(0.337)	(0.264)	(0.282)
Payment Card	0.203	0.208	0.234	0.275
	(0.220)	(0.251)	(0.190)	(0.236)
Other	-0.264	-0.261	-0.343	-0.286
	(0.939)	(1.016)	(0.929)	(0.964)
Survey Method				
Mail	0.0499	-0.349	0.118	-0.0507
	(0.275)	(0.328)	(0.204)	(0.210)
On Site	-0.601*	-0.832**	-0.510*	-0.769***
	(0.323)	(0.336)	(0.271)	(0.281)
Phone	0.324	0.217	0.212	0.105
	(0.330)	(0.373)	(0.282)	(0.283)
Web	-0.191	-0.619	-0.116	-0.279
	(0.285)	(0.390)	(0.242)	(0.313)
Other	-0.392	-0.763	-0.440	-0.805
	(0.466)	(0.642)	(0.408)	(0.621)
Country Variables				
gdp_capita	0.00909	0.0269	0.00294	0.00800
	(0.0247)	(0.0311)	(0.0158)	(0.0209)
tax_revenue	0.0380	0.0448	0.0362	0.0480
	(0.0680)	(0.0795)	(0.0596)	(0.0642)
Gouv	-0.0237	-0.0293		
	(0.0293)	(0.0342)		
Civil_services			-0.0162	-0.0376
			(0.0320)	(0.0483)
Constant	1.820	0.111	1.592	0.801
	(3.538)	(4.906)	(2.675)	(3.885)
Observations	175	144	198	166
R-squared	0.545	0.561	0.497	0.533

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table V: Robustness checks

in this paper is through the collection of more precise geographic and institutional data in the valuation studies.

Bibliography

Cunha-e Sá, M. A., Madureira, L., Nunes, L. C., and Otrachshenko, V. (2012). Protesting and Justifying: A Latent Class Model for Contingent Valuation with Attitudinal Data. *Environmental and Resource Economics*, 52(4):531–548.

EVS (2011). European Values Study 1981-2008, Longitudinal Data File. GESIS Data Archive, Cologne, Germany, ZA4804 Data File Version 2.0.0 (2011-12-30).

Hadker, N., Sharma, S., David, A., and Muraleedharan, T. R. (1997). Willingness-to-pay for Borivli National Park: evidence from a Contingent Valuation. *Ecological Economics*, 21(2):105–122.

Meyerhoff, J. and Liebe, U. (2010). Determinants of protest responses in environmental valuation: A meta-study. *Ecological Economics*, 70(2):366–374.

OECD (2017). Tax revenue (indicator). (Accessed on 13 March 2017).

Oehlmann, M. and Meyerhoff, J. (2017). Stated preferences towards renewable energy alternatives in Germany – do the consequentiality of the survey and trust in institutions matter? *Journal of Environmental Economics and Policy*, 6(1):1–16.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–632.

Remoundou, K., Kountouris, Y., and Koundouri, P. (2012). Is the value of an environmental public good sensitive to the providing institution? *Resource and Energy Economics*, 34(3):381–395.

Schläpfer, F. and Bräuer, I. (2007). Theoretical incentive properties of contingent valuation questions: Do they matter in the field? *Ecological Economics*, 62(3-4):451–460.

WVS (2015). World Value Survey 1981-2014 Longitudinal Aggregate v.20150418, 2015. World Values Survey Association (www.worldvaluessurvey.org). Aggregate File Producer: JDSystems Data Archive, Madrid, Spain.

Part II

Statistical analysis of willingness to pay

Chapter 4

How useful are (censored) quantile regressions for analyzing willingness to pay data?[1]

Abstract: Recurring econometric issues, such as censoring and heteroskedasticity, often impact the analysis of willingness to pay (WTP) data. We investigate the potential advantages of models based on quantile regression (QR) and censored quantile regression (CQR) for addressing these issues. First, we provide analytical arguments showing how (C)QR can tackle these issues. Second, we show by means of a Monte Carlo experiment how (C)QR performs compared to standard (linear and censored) models. Third, we apply these four models to a French Contingent Valuation survey dealing with flood risk and compare performance. Our findings confirm the usefulness of (C)QR for analyzing WTP data, especially CQR in presence of heteroskedasticity or censored data.

4.1 Introduction

Non-market valuation methods have been increasingly used in the past decades to elicit the population's willingness to pay (WTP) for various non-market goods and services, through either revealed or stated preferences. These WTP data are then analyzed using econometric models to help inform private or public decision-makers by clarifying and predicting preferences. This involves identifying the WTP determinants among individual characteristics.

[1]This paper is a joint work with Olivier Chanel. A new version is under revision.

Effective modeling, however, needs to take several econometric issues into account. First, there is the treatment of zero WTP, which introduces censoring from below. Second, the impact of respondents' characteristics on WTP is potentially heterogeneous, which may bias estimates. Third, the WTP variance is likely to be non-independent of some WTP determinants, such as income, which generates heteroskedasticity. Fourth, in CV studies particularly, there can be outliers and/or extremely large values, due either to the hypothetical nature of the CV exercise or to the difficulty of the valuation task. Quantile regression (QR) and censored quantile regression (CQR) can help tackle these issues by estimating the impact of explanatory variables on any conditional WTP quantiles chosen, instead of simply on mean WTP. Yet despite the importance non-market valuation studies, and its increasing use of quantile-based methods, the field lacks a systematic analysis of the performance of (C)QR.

We propose to fill this gap by comparing the performances of (C)QR models and standard models (Ordinary Least Square, OLS, and Tobit). First, we provide analytical arguments showing how (C)QR can tackle the above-mentioned econometric issues associated with WTP data. Second, we carry out a Monte Carlo experiment especially designed to include heterogeneity and censoring, in order to compare the statistical performances of the four models in a controlled framework. Finally, we apply these methods to real data from a French CV survey on reducing risks associated with flooding.

Our results confirm the advantages of (C)QR models w.r.t. standard conditional mean estimates for analyzing WTP data. The Monte Carlo experiment shows that (C)QR are less impacted by heteroskedasticity than OLS and Tobit, and that the Tobit and CQR models perform better than their linear counterparts in presence of censored WTP. Applying QR models to the flood risk survey confirms their superiority over standard approaches, although improvements from CQR estimates over QR estimates are more limited.

We make two primary contributions to the existing literature. First, our analysis contributes to the methodological literature on non-market valuation. To date, despite their numerous advantages over most of the standard methods, QR applications that use WTP as dependent variable are still rare, although on the increase. One article alone (Krishnamurthy and Kriström 2016) applies CQR to CV data with zero WTP although CQR has been used in various empirical applications for years to account for null data either observed (like vegetable demand, see Gustavsen and Rickertsen 2006; local precipitation, see Friederichs and Hense 2007; agricultural surfaces, see Motamed et al. 2016) or revealed (like food safety, see Lagerkvist and Okello 2016; demand for opera Laamanen 2013). Overall, all find heterogeneity in the relationships between WTP and some of the explanatory variables, confirming the superiority of QR-type approaches over standard approaches. However, we are not aware of any study (stated preference or not) that compares standard OLS-like and Tobit-like models with QR and CQR methods. Our study helps fill these gaps by comparing the properties of the four models (OLS, Tobit, QR and CQR), first through a Monte Carlo experiment and then on real WTP.

Second, our analysis contributes to the CV empirical literature on flooding impacts, especially regarding two aspects: the nature of beneficiaries and the nature of flood-related effects. In the literature, the nature of beneficiaries is dealt with by two types of scenario: an individual action scenario that evaluates the WTP for a decrease in the respondent's own consequences of a flood or a collective action scenario that evaluates the WTP for a decrease in flood risk. The nature of the flood-related effects covers both tangible effects (through compensation for monetary losses) and intangible effects (like emotions or psychological aspects related to a flood event and its aftermath). Some studies propose a scenario for an individual action involving tangible effects (Hung 2005; Abbas et al. 2014); intangible effects (Defra 2005; Joseph et al. 2015; Owusu et al. 2015) or both (Kuo 2016), while others propose a collective action that decreases both flood-related effects (Shabman et al. 1998; Novotny et al. 2001; Grelot 2004; Zhai et al. 2006; Glenk and Fischer 2010). Only three propose two scenarios, one for each action (Deronzier and Terra 2006; Chanel et al. 2013a; Ghanbarpour et al. 2014). Comparisons of these studies are difficult, due to differences in the risk reduction proposed, in purchasing power across countries, in the elicitation format or in the two aspects of flooding impact mentioned above (see Appendix 4.A for a summary with mean WTP values). Our CV study sheds light on the impact of beneficiaries by using both an individual and a collective action on the same sample.

The remainder of the article proceeds as follows. In the following section we describe the advantages of quantile regressions for WTP data and the relevant econometric models. We then present the Monte Carlo experiment and the empirical application. In the last section we discuss and conclude.

4.2 Quantile regressions and WTP data

The reasons why (C)QR may perform better than standard models in tackling issues arising from WTP data are given below, followed by the presentation of the four models used in the experiment and the CV application.

4.2.1 Advantages of QR and CQR for WTP data

First, the treatment of zero WTP requires care. OLS regressions are known to be biased when the dependent variable is censored, calling for the use of the Tobit model to properly account for censoring. QR are also biased when there is censoring, as noted by Kowalski (2016), “since quantile regression uses information from the entire sample to generate the estimate at each quantile, if some observations on [the dependent variable] are censored, the quantile regression lines can be biased toward zero at all quantiles.” CQR should thus be used to properly deal with censoring (see Powell 1986). In non-market valuation, whether negative WTPs should be allowed in the modeling is an issue per se (Carson and Hanemann 2005): if yes, zero WTPs may correspond to negative WTPs censored at zero, if no, they are strictly null WTP, i.e.‘corner solutions’. Following standard practice in the literature (see Krishnamurthy and Kriström 2016, footnote 3 for instance), we use ‘censoring’ and ‘corner solution’ as perfect substitutes in the following even though they are not strictly equivalent when negative WTP truly represents a negative utility change.

Second, (C)QR is only one of several econometric methods (like nonparametric estimations, latent class models, random parameter models) that allow for heterogeneity in the coefficients. Indeed, each coefficient of a (C)QR corresponds to the coefficient of a regression in which an explanatory variable interacts with an unobserved latent variable that influences the position of respondents in the conditional distribution of the dependent variable. It therefore offers a more comprehensive view of the relationship between the dependent variable and the covariates, since the covariates are allowed to have a different impact at each quantile of the conditional distribution of the dependent variable, not only at the mean. This can be useful when analyzing WTP data: for instance, in their CV study, it is only through QR that Furno et al. (2016) manage to detect the effect of hypothetical bias on the tails of the distribution. This feature should also have advantages for policy makers, offering a picture of how the effects of WTP determinants are distributed across the population, not only on their conditional mean, which can clearly be misleading. Although (C)QR accounts for heterogeneity, it is not a total substitute for latent class or random parameter models, which account for preference heterogeneity (Nahuelhual et al. 2004; Boxall and Adamowicz 2002). The source of the heterogeneity that (C)QR accounts for is, by definition, unobserved, and various sources may be at play.

Third, the relationships between WTP and its determinants are usually heteroskedastic, especially regarding the income variable. As income increases, WTP is less and less constrained and the heterogeneity of preferences leads to greater dispersion of WTP values. This kind of feature has already been observed, see for instance the Engel Curves between income and food expenditure in Koenker (2005). (C)QR is able to capture these scale shifts and reveals how they affect the marginal effects at a given conditional quantile of the WTP. Therefore, (C)QR allows us to interpret heteroskedasticity as a special case of heterogeneity. For instance, the scale effect of income (which we model in the Monte Carlo experiment) leads to heterogeneity of the income coefficient along the conditional distribution of WTP.

Fourth, WTP from CV studies may contain many low and/or very high WTPs because no actual out-of-pocket payment is required for the provision of the hypothetical good or service and because the valuation task is difficult. This will hinder the estimation of conditional mean WTP, because strong influence from the upper tail of the WTP distribution potentially leads to mean and median WTPs that significantly differ from each other. QR are more robust than OLS regressions to the presence of outliers, fat tails and to non-normal errors (Powell 1986; O'Garra and Mourato 2006; Huang and Chen 2015). Moreover, although OLS is more efficient than QR when the errors are homoskedastic and normally distributed (according to the Gauss Markov theorem), empirical evidence suggests that QR tends to provide more efficient estimates if these assumptions are not met (Deaton 1997; Hung et al. 2010).

The main difficulty in applying (C)QR is interpreting the coefficients. (C)QR results express how the coefficients vary, not along the marginal distribution, but rather along the conditional distribution. Hence, the estimates capture the marginal effect of each observed characteristic on the specific quantile of the WTP conditional on these characteristics: it is not heterogeneity across the distribution of the WTP that is accounted for, but heterogeneity across the distribution of the unobserved determinants of the WTP. Some interpretations (e.g. in O'Garra and Mourato 2006; and Viscusi et al. 2012) consider, for instance, that respondents with high WTP are more likely to take into account their resources, attitudes and behaviors, whereas respondents with low WTP may be more affected by budget constraints than by individual preferences. In our opinion, the heterogeneity of the marginal effects should not be interpreted as heterogeneity across income groups or preference groups. Resources and preferences are observed (although imperfectly for preferences), while the heterogeneity revealed by (C)QR is heterogeneity across the unobserved determinants of WTP.

Special features of WTP-type data

Stated preference methods used to elicit WTP rely on hypothetical surveys that make them prone to several long-standing problems extensively debated in the literature (see Diamond and Hausman 1994; or Hausman 2012). Some challenge the validity of using WTP as reliable economic values (hypothetical bias, the fact that preferences must be well-defined *ex ante* or that non-economic components might be embedded in WTP), while others raise the question of whether elicited WTP is a reliable expression of individuals' genuine WTP. Although our contribution is more concerned with the modeling of observational-type data than with the nature of WTP, incentive compatibility and potential negativity deserve a brief discussion.

As stated by Carson and Groves (2011), consequentiality is a prerequisite of incentive compatibility: individuals must care about the issue at stake and believe their responses can influence the final decision. These authors consider that a carefully-constructed survey can plausibly guarantee adequate consequentiality (see however Vossler and Watson 2013). Then, the elicitation mechanism needs to be incentive compatible for responses to reflect individuals' genuine WTPs. This is more tricky according to Carson and Groves (2011), who conclude that, of the various existing elicitation formats, none stands out as having better statistical and practical properties. We ourselves know of no elicitation format that can be considered both generally theoretically incentive compatible (i.e. the truthful preference revelation is an optimal (and the dominant) strategy for the respondent) and procedurally invariant in the CV setting (answers should not depend on the psychological or psychometric properties of the elicitation format). Some authors, fortunately, consider that "for many practical purposes, then, the question of the supposed incentive compatibility of any particular elicitation procedure may be very much a second order issue" (Bateman et al. 2002, p. 381). Overall, the researcher must pay careful attention to the choice of elicitation format *ex ante* to ensure proper incentive properties, and then choose the appropriate econometric model to properly model the decision. Note that although we deal with open-ended data in the following, the econometric models we propose can be applied to referendum-type data, with a bivariate Probit or a binary quantile regression with selection for instance (see also Strazzera et al. 2003 for a double-bounded dichotomous choice model with selection). They can also match payment card formats (in which one bid-range is chosen from several proposed), as the middle of the bid-range elicited is a reliable approximation of the genuine WTP (Cameron and Huppert 1989; Yang et al. 2012a).

Finally, there is the question of whether negative WTP should be allowed in the econometric modeling (Carson and Hanemann 2005). The answer depends on the revelation mechanism. If yes, zero WTPs stand as negative WTPs that truly represent a negative utility change but are censored at zero due to an observability issue, and the underlying distributional assumptions of the Tobit model are fulfilled (see Sigelman and Zeng 1999). If no, zero WTPs are strictly null WTP, i.e. 'corner solutions' under an economic rationale (a budget constraint or change in utility associated with the good) that could be modeled separately. Although it is standard practice in the literature to consider 'censoring' and 'corner solution' as strictly equivalent (see Krishnamurthy and Kriström 2016, footnote 3 for instance), the use of a Tobit model is debatable in the latter case (see section 2.1.1).

4.2.2 Econometric models

We present the models used both in the MC experiment and the CV application, relying on conditional means and on conditional quantiles both for the linear specification and the specification accounting for censoring.

Linear models not accounting for censored WTP

The **conditional mean model** for WTP can be written as:

$$WTP_i = x_i'\beta + u_i \tag{4.1}$$

where $u_i \sim N(0, \sigma^2)$ is a random term, x_i is a matrix of explanatory variables and β a vector of parameters.

Linear models estimated by Ordinary Least Squares (OLS) are simple and have an interesting feature: the point estimates are more robust to non-normality than Maximum

Likelihood-based methods because no distributional assumptions are necessary (although they are required for inference). However, the strong assumptions about model specification (linearity, homogeneity, homoskedasticity) often violated by WTP data mean that the OLS model only provides a first, simple benchmark.

Following Koenker (2005)'s presentation of the **conditional Quantile Regression (QR) model**, the conditional distribution of a random variable WTP is denoted $F_{WTP|X}(WTP|x)$, where X is a set of random explanatory variables. The conditional quantile Q_τ is defined as:

$$Q_\tau(WTP|x) = \inf(u : F_{WTP|X}(u|x) \geq \tau) = F^{-1}(\tau|x) \tag{4.2}$$

If we assume that the conditional τ-quantile function is linear in x we can write:

$$Q_\tau(WTP|x) = x'\beta_\tau \tag{4.3}$$

where β_τ is a vector of k parameters associated with the τ-quantile. The QR estimator of β_τ for a random sample $(WTP_i, x_i)_{i=1,\dots,n}$ is obtained by solving:

$$\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau(WTP_i - x_i'\beta_\tau) \tag{4.4}$$

where ρ_τ is the check function defined by:

$$\rho_\tau(u) = u(\tau - I(u < 0)) \tag{4.5}$$

where $I(.)$ is an indicator variable. We interpret the coefficient $\beta_{\tau,k}$ as the change in the quantile of order τ of the conditional distribution for a marginal change in variables x_k.

Although QR have several advantages over OLS regressions, especially for CV data, neither of them takes into account valid zero WTPs, which induces non-linearities in the relationship between the dependent variable and the explanatory variables, and biases the estimates (Amemiya 1984).

Models accounting for censored WTP

The **Tobit model** accounts for censoring of the dependent variable, and can be written as:

$$\begin{cases} WTP_i = WTP_i^* & \text{if} \quad WTP_i^* > 0 \\ WTP_i = 0 & \text{if} \quad WTP_i^* \leq 0 \end{cases} \tag{4.6}$$

where WTP_i is the observed WTP and WTP_i^* is a latent variable corresponding to the true WTP. Under the parametric assumption $WTP_i^* \sim N(x_i'\beta, \sigma^2)$, the likelihood function of this model is:

$$L(\beta, \sigma; WTP_i, x_i) = \prod_{i=1}^n \left(\frac{1}{\sigma} \phi \left(\frac{WTP_i - x_i'\beta}{\sigma} \right) \right)^{I(WTP_i > 0)} \left(1 - \Phi \left(\frac{WTP_i - x_i'\beta}{\sigma} \right) \right)^{I(WTP_i = 0)} \tag{4.7}$$

with $\phi(.)$ the probability density function of the standard Normal distribution and $\Phi(.)$ the cumulative density function (cdf).

Although this model is simple to implement and relatively easy to interpret, it is sensitive to incorrect assumptions regarding the error term distribution. Another drawback is that it assumes that the latent variable corresponding to the true WTP can be negative, which is rarely the case in CV studies. The Tobit model is still appropriate as a statistical method but not fully as an economic framework (Wooldridge 2001).

The **Censored Quantile Regression (CQR) model**, based on the QR model and first proposed by Powell (1986), assumes that the conditional τ-quantile function in x is $Q_\tau(WTP_i|x_i) = \max(0, x_i'\beta_\tau)$. Thus the estimator is found by solving:

$$\min_{\beta_\tau} \sum_{i=1}^n \rho_\tau(WTP_i - \max(0, x_i'\beta_\tau)) \tag{4.8}$$

The CQR model allows for coefficients' heterogeneity, and Powell (1986) shows that, under some regularity conditions, it is consistent and asymptotically normal whatever the error distribution, which is not true for the Tobit.

4.3 Monte Carlo experiment

To characterize the empirical properties of the four models on WTP data, we carry out a Monte Carlo experiment especially designed to include censoring and heterogeneity, modeled via a heteroskedastic error term. Specifications that explicitly express the heterogeneity of the coefficients as a function of the quantile (for instance $\beta(\tau) = exp(\tau)$) are possible (Hoshino 2013), but their interpretation would be less intuitive.

4.3.1 Design of the Monte Carlo experiment

The data-generating process (DGP) is a linear specification with a censored dependent variable WTP_i:

$$WTP_i = \max(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + e_i, 0) \tag{4.9}$$

where:

- $x_{i,1}$ is a standard log-normal continuous variable $lnN(0,1)$. It stands for the income variable, with a location shift effect on WTP: respondents with high incomes are more likely to have higher WTPs. A scale effect is also accounted for in the error term e_i below: respondents with higher incomes are also likely to have higher variance in WTP.[2]

- $x_{i,2}$ is a standard normal variable $N(0,1)$.

- e_i is an error term that covers three different heteroskedasticity intensities j ($j = 0, 1, 2$):

$$e_i = (1 + \gamma_j x_{i,1})v_i \tag{4.10}$$

  where v_i is i.i.d standard normal, $\gamma_0 = 0$, $\gamma_1 = 0.4$ and $\gamma_2 = 0.8$.
  The first case corresponds to homoskedasticity, while the two others produce linear heteroskedasticity in $x_{i,1}$, mimicking the scale effect of income, with increasing strength from γ_1 to γ_2.

The heteroskedasticity of the error term leads to heterogeneity in the relationship between the quantiles of the conditional WTP distribution and the covariates. The marginal effect of a covariate on the quantile covariates affects both the location and the scale of the dependent variable:

$$\frac{\partial Q_\tau(WTP|x_1)}{\partial x_1} = \beta_1 + \gamma_j F_u^{-1}(\tau) \tag{4.11}$$

[2]Although we consider an income effect on WTP here, this specification covers any shift and scale effects an explanatory variable is assumed to have on a dependent variable.

where $F_u^{-1}(\tau)$ is the inverse cdf (i.e. quantile function) of the error term distribution.

We use several specifications for the DGP, varying the sample size n (50, 300 and 1000) and the censoring rate c (0% when β_0=7, 20% when β_0=0 and 40% when β_0=-1.7). We finally set β_1=2 (to mimic the positive relationship between income and WTP), β_2=-2 (to mimic a negative relationship with WTP), and simulate 10000 samples for each of the 3 sample sizes x 3 heteroskedasticity intensities x 3 censoring rates = 27 specifications (see details in Appendix 4.B for reviewers' use only).

4.3.2 Results of the Monte Carlo experiment

Because the experiment provides a large amount of information, we focus on the Mean Bias and the Root-Mean-Square Error (RMSE). Table I show these statistics for the slope coefficients, 2 censor rates (0% and 40%), 2 heterogeneity intensities ($\gamma = 0$ and $\gamma = 0.8$) and $n = 1000$, while tables VI to IX (see Appendix 4.C for reviewers' use only) show all the specifications. For the Mean Bias, the p-values of equality tests with the corresponding true parameter are given in parentheses for each model and each DGP. To make results clearer, cells are highlighted in gray when the Mean Bias differs by less than 10% from the true parameter. For the RMSE, cells are highlighted in gray when the RMSE is less than .1, as a rule-of-thumb measure of the error of each model (see Ferrini and Scarpa 2007 for similar choices).

For Mean Bias, the linear models are systematically biased towards zero when the censoring rate is positive, consistent with expectations and empirical findings (Chen and Kashiwagi 2017): negative for β_1 that is positive and positive for β_2 that is negative. The QR estimates tend to be more biased than OLS estimates for heteroskedastic DGP ($j \neq 0$), a bias that increases with the censoring rate. Conversely, the models accounting for censoring are less (or not at all) biased. A notable exception is the Tobit model for heteroskedastic DGP, because of inconsistency introduced by the violation of the homoskedasticity assumption. Only the CQR model systematically has a Mean Bias close to zero. We can conclude that QR when modeling WTP data might be problematic because it will be very biased if the WTP are censored. In this regard, CQR is a better choice.

Regarding RMSE, we consistently find that it increases with censoring rate. We find lower RMSE for the models accounting for censoring, particularly striking for β_2. This can be explained by the increase in bias described above. Finally, we observe that the

β_1

DGP	Statistics	OLS	Tobit	QR 25%	QR 50%	QR 75%	CQR 25%	CQR50%	CQR 75%
c=0%, γ =0, n=1000	MB	-0.000	0.000	0.000	0.000	-0.000	0.000	0.000	-0.000
c=0%, γ =0, n=1000	p-val	0.944	0.958	0.150	0.793	0.055	0.135	0.754	0.056
c=0%, γ =0, n=1000	RMSE	0.008	0.008	0.010	0.009	0.010	0.010	0.009	0.010
c=0%, γ =0.8, n=1000	MB	-0.001	-0.001	0.003	0.001	-0.004	0.003	0.001	-0.004
c=0%, γ =0.8, n=1000	p-val	0.741	0.763	0.040	0.612	0.001	0.035	0.595	0.001
c=0%, γ =0.8, n=1000	RMSE	0.101	0.101	0.087	0.058	0.050	0.087	0.058	0.050
c=40%, γ =0, n=1000	MB	-0.178	0.000	-0.342	-0.223	-0.142	0.002	0.000	-0.001
c=40%, γ =0, n=1000	p-val	0.000	0.412	0.000	0.000	0.000	0.000	0.114	0.000
c=40%, γ =0, n=1000	RMSE	0.091	0.008	0.173	0.113	0.073	0.012	0.011	0.011
c=40%, γ =0.8, n=1000	MB	-0.170	0.170	-0.587	-0.395	-0.306	0.022	0.004	-0.005
c=40%, γ =0.8, n=1000	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.002
c=40%, γ =0.8, n=1000	RMSE	0.132	0.137	0.412	0.206	0.130	0.138	0.076	0.059

β_2

DGP	Statistics	OLS	Tobit	QR 25%	QR 50%	QR 75%	CQR 25%	CQR50%	CQR 75%
c=0%, γ =0, n=1000	MB	-0.000	-0.000	0.000	-0.000	-0.001	0.000	-0.000	-0.001
c=0%, γ =0, n=1000	p-val	0.586	0.216	0.409	0.745	0.132	0.721	0.562	0.109
c=0%, γ =0, n=1000	RMSE	0.016	0.016	0.022	0.020	0.021	0.022	0.020	0.021
c=0%, γ =0.8, n=1000	MB	0.000	-0.002	-0.000	-0.000	0.000	-0.001	-0.000	0.000
c=0%, γ =0.8, n=1000	p-val	0.918	0.049	0.843	0.997	0.692	0.327	0.761	0.779
c=0%, γ =0.8, n=1000	RMSE	0.045	0.045	0.040	0.036	0.039	0.040	0.036	0.039
c=40%, γ =0, n=1000	MB	0.791	0.000	0.784	0.822	0.882	-0.002	-0.000	0.001
c=40%, γ =0, n=1000	p-val	0.000	0.832	0.000	0.000	0.000	0.017	0.543	0.234
c=40%, γ =0, n=1000	RMSE	0.396	0.022	0.394	0.412	0.442	0.035	0.031	0.031
c=40%, γ =0.8, n=1000	MB	0.850	-0.448	1.408	1.094	0.944	-0.041	-0.011	-0.004
c=40%, γ =0.8, n=1000	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008
c=40%, γ =0.8, n=1000	RMSE	0.427	0.241	0.705	0.548	0.474	0.154	0.088	0.072

Note: Cells are highlighted in gray when the Mean Bias differs by less than 10% from the true parameter. For the RMSE, cells are highlighted in gray when the RMSE is less than .1.

Table I: Mean Bias and Root-Mean-Square Error for n=1000

RMSE for QR and CQR is constant across quantiles when there is homoskedasticity, but decreases along the quantiles when there is heteroskedasticity. This could be because, due to censoring, we have less information at the bottom of the conditional distribution, which causes a loss in efficiency.

Looking at the effect of sample size (see Appendix 4.C), only CQR shows a consistent decrease in bias for both β_1 and β_2 when sample size increases, whatever the specifications. With the exception of the lowest quantile for the smallest sample size (n=50), CQR tends to outperform the other models in terms of bias. RMSE decreases with sample size for all models and specifications. Again, CQR does not perform well for a small sample size at the 25% quantile, but previous results for the larger sample sizes still hold. Therefore, one should be careful when using CQR in a study with a very small sample size, conditional mean models should be more robust in this case.

Since the statistics of interest for policy-makers are mean and median WTP, we compare the ability of the different models to predict the theoretical values. Table II shows the expected WTP, the mean of the medians, and the mean of the predicted WTP for the models (conditional mean for OLS and Tobit, conditional median for QR and CQR).

The number of replications is large enough for the mean of the median across the simulations to provide an accurate estimate of the theoretical median. The p-values of the equality test of the mean of predictions and the true mean are given in parentheses next to the mean of predictions, and the 95% empirical confidence intervals (CI) for the mean of predictions are given in parentheses below. It is worth noting that the test of equality of the mean of the conditional median to the true mean for QR and CQR is statistically meaningless, and only reproduced to enable comparisons with OLS and Tobit.[3]

We observe that the sample median WTP is always lower than the theoretical mean, especially for censored DGP, due to the skewness to the right of the simulated WTP. Predicted mean WTPs are very close whatever the models without censoring, but differ when there is censoring. The predicted mean WTPs of the four models are systematically downwardly biased with respect to the theoretical mean WTP, especially for the censored

[3]We could have compared the median of the QR and CQR predictions with the unconditional median, but the two are not directly linked, since there is no law of iterated expectation for the median. So we chose the mean for consistency.

DGP	Theo. Mean	Sample Median	OLS	Tobit	QR	CQR
c=0%, γ =0, n=50	10.323	9.537	10.302 (0.912)	10.302 (0.912)	10.303 (0.915)	10.302 (0.918)
c=0%, γ =0, n=50	10.323	(9.032,10.016)	(9.112,11.773)	(9.112,11.773)	(9.085,11.76)	(9.096,11.789)
c=0%, γ =0, n=300	10.323	9.389	10.298 (0.906)	10.297 (0.906)	10.298 (0.905)	10.297 (0.902)
c=0%, γ =0, n=300	10.323	(9.143,9.587)	(9.77,10.869)	(9.77,10.869)	(9.769,10.877)	(9.766,10.879)
c=0%, γ =0, n=1000	10.323	9.557	10.296 (0.843)	10.296 (0.843)	10.296 (0.851)	10.296 (0.843)
c=0%, γ =0, n=1000	10.323	(9.378,9.762)	(10.006,10.61)	(10.006,10.61)	(10.002,10.613)	(10.001,10.613)
c=0%, γ =0.8, n=50	10.408	9.587	10.302 (0.834)	10.302 (0.833)	10.301 (0.84)	10.303 (0.838)
c=0%, γ =0.8, n=50	10.408	(8.303,10.369)	(8.976,11.971)	(8.975,11.971)	(8.853,12.121)	(8.891,12.118)
c=0%, γ =0.8, n=300	10.408	9.326	10.3 (0.708)	10.299 (0.706)	10.3 (0.733)	10.3 (0.721)
c=0%, γ =0.8, n=300	10.408	(8.996,9.627)	(9.699,10.976)	(9.699,10.976)	(9.65,11.018)	(9.65,11.019)
c=0%, γ =0.8, n=1000	10.408	9.332	10.3 (0.523)	10.299 (0.522)	10.3 (0.558)	10.299 (0.549)
c=0%, γ =0.8, n=1000	10.408	(9.195,9.457)	(9.969,10.657)	(9.968,10.657)	(9.942,10.689)	(9.944,10.688)
c=40%, γ =0, n=50	2.805	0.743	2.369 (0.42)	2.304 (0.372)	2.327 (0.399)	2.437 (0.475)
c=40%, γ =0, n=50	2.805	(0,1.763)	(1.395,3.711)	(1.318,3.652)	(1.287,3.732)	(1.46,3.793)
c=40%, γ =0, n=300	2.805	0.841	2.368 (0.09)	2.301 (0.051)	2.319 (0.069)	2.436 (0.139)
c=40%, γ =0, n=300	2.805	(0.448,1.278)	(1.918,2.873)	(1.847,2.81)	(1.847,2.854)	(1.983,2.948)
c=40%, γ =0, n=1000	2.805	0.795	2.363 (0.003)	2.295 (0.001)	2.312 (0.002)	2.43 (0.011)
c=40%, γ =0, n=1000	2.805	(0.673,0.94)	(2.112,2.647)	(2.044,2.582)	(2.047,2.608)	(2.179,2.716)
c=40%, γ =0.8, n=50	3.297	0.806	2.519 (0.247)	2.179 (0.14)	2.328 (0.223)	2.782 (0.425)
c=40%, γ =0.8, n=50	3.297	(0.043,1.381)	(1.412,4.121)	(1.058,3.814)	(1.017,4.193)	(1.599,4.49)
c=40%, γ =0.8, n=300	3.297	0.69	2.523 (0.015)	2.111 (0.002)	2.304 (0.007)	2.753 (0.094)
c=40%, γ =0.8, n=300	3.297	(0.437,1.017)	(2.019,3.13)	(1.587,2.738)	(1.718,2.988)	(2.216,3.402)
c=40%, γ =0.8, n=1000	3.297	0.598	2.522 (0)	2.089 (0)	2.302 (0)	2.749 (0.005)
c=40%, γ =0.8, n=1000	3.297	(0.311,0.759)	(2.23,2.847)	(1.773,2.433)	(1.965,2.666)	(2.441,3.087)

Table II: Comparison of Predicted WTP

DGP, although the best predictions are obtained with CQR. As expected, increasing the sample size results in tighter confidence intervals. As for the coefficients, increasing the intensity of heteroskedasticity enlarges confidence intervals, especially for the OLS and Tobit models.

Overall, accounting for censoring in the models has an undeniable impact: the Tobit and CQR models perform better than their linear counterparts. In addition, as expected, (C)QR are less impacted by heteroskedasticity than OLS and Tobit, making CQR the model of choice when the sample size is large enough. These results obviously hold with any type of data supporting censoring and heteroscedasticity, but the next section compares the four models to actual data from a CV study on flood risk (see Appendix 4.D for a summary of existing applications of QR Models to CV Studies).

4.4 Empirical application

Decreasing the impact of floods is a major public concern. Of the world's ten most costly disasters between 1970 and 2015, five involved flooding, two hurricanes, two earthquakes, and the last was the terrorist attacks on 11 September 2001, the only non-natural disaster (Sigma 2016). The intensity and frequency of flooding and hurricanes are likely to increase with climate change in the 21st century (Intergovernmental Panel on Climate Change 2013). The highest ever number of natural catastrophes was observed in 2015 (198), leading to a rise in their annual worldwide cost (USD 150 billion per year on average over the 2004-2014 period).

In France, a quarter of the population is at risk of flooding (MEDE, 2012), and catastrophic river risings or flash floods regularly hit the front pages. Since 1999, some 200 people have died in five major flood events, with twofold consequences. First, for the population directly involved, the physical (deaths and injuries), psychological (Post-Traumatic Stress Disorder, PTSD) and financial consequences of catastrophic flooding have long-lasting effects. Second, for the insurance sector (and indirectly the whole population through insurance premiums), flooding represents the major hazard in terms of number of claims paid and in terms of cost for the French insurance regime providing reimbursement for damage due to natural disasters (Cat Nat regime). With €4.7 billion paid out between 1995 and 2006 under the natural disaster warranty (10% for individuals, 90% for firms), flooding accounts for 57% of overall Cat Nat expenditure (CEPRI, 2013). In spite of the importance of the issue at stake, comprehensive assessment of the damages from floods are still difficult to make. This type of assessment should include both the tangible aspect of floods, and the impacts on morbidity and mortality (Johansson and Kriström (2015)). While evaluation of the costs of disasters often manage to account for loss in private consumption and (to a lesser extent) the value of lives lost, the loss of consumer surplus is more difficult to assess. There is also a lack of data on the impact of floods on consumer surplus, that calls for more stated preference studies on flood risk protection.

4.4.1 Method and data

The survey was administered in Southeastern France between 26 April and 30 June 2012 via individual face-to-face interviews with respondents both having and not having experience of floods.

Study design

The questionnaire included eight modules (housing, risk perception, hypothetical monetary choices, personality, PTSD, flood-specific issues, socio-demographic factors and contingent valuation scenario). However, we only present in detail the findings relevant to this article: respondents' WTP to reduce their vulnerability and exposure to flooding (see Chanel et al. 2013b for additional results and the full questionnaire).

Two scenarios were proposed to each respondent, in the spirit of Deronzier and Terra (2006), to determine respondents' willingness to participate in actions aimed at reducing risks and, if so, the corresponding WTP (see Appendix 4.E for the exact wording). One scenario (randomly proposed first to half of the sample) is collective and assesses intangible effects, proposing a contribution to the funding of city-level protective devices. The respondent gives a WTP that reflects his/her utility for a decrease in flood risk, while maintaining existing insurance. This WTP covers both the tangible and intangible / psychological gains from prevention. The other scenario is individual and restricted to tangible effects, proposing a contribution to insurance against flood risk that will only reduce vulnerability. The respondent gives a WTP that reflects his/her utility for full insurance on flood risk, i.e. a decrease in the financial risk linked to flooding, without any change in flood risk and the related psychological effects.

For ease of comparison between the two scenarios, and to limit possible framing effects, we use a fictitious Flood Management Fund to manage both risk reduction (protective devices) and consequences (individual insurance) of a flood. The payment vehicle in both scenarios is a voluntary contribution to this Fund, and the same elicitation format is used.

Of the various existing elicitation formats, none stands out as having better statistical and practical properties (see Carson and Groves 2011). Here, we used the circular payment card (CPC) which, unlike the standard payment card (PC), relies on a circular visual representation with no predetermined start or end points, no top or bottom, no left or right. The respondent is asked to think about her/his WTP, and is then given the CPC in a random position to help her/him in the elicitation process. The respondent is then asked "How much maximum would you be willing to pay per year?". In addition to the advantages of the standard PC format (low rate of non-response and a visual aid to facilitate WTP elicitation), the circular version eliminates starting-bid bias (because each section is equally likely to be seen at first glance), middle-card bias (by construction), and helps strongly reduce the range effect associated with the bids chosen (as the succession of bid ranges mimics a continuous distribution, see Carson and Groves 2011). A between-respondent analysis comparing the WTPs elicited using CPC, Open-Ended (OE) and PC formats within the same format can be found in Chanel et al. (2017). Note finally that the use of the same elicitation format for both scenarios eases their comparison, unlike Deronzier and Terra (2006).

Data

The empirical analysis is based on a sample of 599 respondents interviewed at home face-to-face by a specialized survey institute. Four municipalities in Southeastern France, within a 65 km radius, were chosen for their varying degrees of exposure to flood risk. Two municipalities had never been flooded: Miramas (25,300 inhabitants), at no risk of flooding, and Berre-l'Etang (13,800 inhabitants), located in an area with a potential risk of flooding due to torrential rivers and dam failure. Two municipalities had experienced flash floods in the past twenty years: Vaison-la-Romaine (6,200 inhabitants) in September 1992 (20 years before the survey) by the Ouvèze river rising, with 37 deaths and four missing, and Draguignan (36,600 inhabitants) in June 2010 (two years before the survey) by the Nartuby river rising, with 23 deaths (12 in Draguignan itself) and two missing.

The respondents interviewed had to meet the following inclusion criteria: be older than 18 at the time of the survey, live in one of the four municipalities and, for the two flooded cities, have been physically present and over 18 when flooding occurred. A pre-test of 20 respondents was used to fine-tune the wording as well as to choose the range and centering of bids. The various modules of the survey captured a large set of potential determinants of WTP. In addition to standard socio-demographic variables, several questions were aimed at capturing risk and loss aversion, time preference, risk perception, information about housing, insurance, personality traits, flood-related knowledge and behaviors, as well as the material and psychological impacts of previous flood(s) where applicable.

Table III presents the summary statistics. The mean WTPs for the individual and the collective action scenarios are close, as was the case in the three other studies that proposed both scenarios (Deronzier and Terra 2006; Chanel et al. 2013a; Ghanbarpour et al. 2014). This may appear surprising, since a larger WTP could be expected from the latter scenario, which entails intangible effects and involves many potential beneficiaries outside the household, hence an altruistic dimension going beyond the family. There are three possible explanations for this result: a lack of faith in the effectiveness of the city-level protective devices, a kind of anchoring on the WTP elicited in first position, or the fact that because the collective scenario is about public good provision, respondents may have less incentive to reveal their true preferences (underestimation due to free-rider behavior). The average age of the sample is 51.3 years; 55.1% are female; 36.2% have at least one child at home; 41.8% have at least a high school certificate; monthly mean respondent income is €1,422 ; monthly mean household income is €2,106 and 47.6% are homeowners .

In standard PC format, the WTP bid range elicited should be used in preference

Variable	Label	Mean	Std. Dev.	Min.	Max.	N
WTP-I	Willingness to pay (Insurance scenario)	100.48	143.29	0	1300	341
Gender	Gender (Male=1)	.449	.40	0	1	599
Age	Age (in years)	51.293	17.003	16	94	593
AgeSquare	Square of the age (in years)	2919.621	1805.891	256	8836	593
Child	Has at least one child (=1)	0.362	0.481	0	1	599
Education	Education (ordinal variable)	1.853	1.14	1	4	599
Income	Monthly income of the respondent (in euros)	1423.478	904.531	0	8000	575
Owner	Is the owner of the housing (=1)	0.476	0.5	0	1	599
HousingRisk	Living on the ground floor or in a house (=1)	0.605	0.489	0	1	593
Inform	Looked for information about flood risk (=1)	0.14	0.347	0	1	593
NbrInfo	Number of media known for information about flood risk (integer)	2.526	1.422	0	8	593
PastExperience	Already experienced a flood (=1)	0.521	0.5	0	1	593
ProbaFlood	Perceived likelihood of being flooded in the next 10 years (in %)	9.353	14.958	0	100	593
Impatience	Preference for the present score (1-7 score)	2.974	2.756	0	7	568
RiskTolerance	Loss lover score (1-4 score)	1.56	0.86	1	4	593
Happy	Declared subjective well-being (0-10 score)	6.772	2.043	0	10	593

Table III: Summary Statistics (n=599)

to the middle of the bid range, although empirical studies do not seem to find major differences between point- and interval-based estimates (Cameron and Huppert 1989; Yang et al. 2012b). This is not an issue here, because we know respondents' exact WTP (the CPC only being a visual aid); however, we need to account for left-censoring. There is actually a noticeable proportion of zero WTP (28.27% for the collective action scenario and 30.86% for the insurance scenario) among respondents giving a WTP, as shown in figure 4.1.

4.4.2 Results

Because estimates from both scenarios are very similar, we only provide results for the insurance scenario (the results for the collective scenario are available upon request).[4] Our choice of WTP determinants is based on the main variables found across the CV studies in Appendix 4.A, and we choose to limit unobserved heterogeneity by keeping all explanatory variables - even non-significant ones - in the four models (see Table IV).

Variables	OLS	Tobit	QR 25%	QR 50%	QR 75%	CQR 25%	CQR 50%	CQR 75%
Intercept	-64.308	-147.205	-30.399	-47.535	23.65	-70.824	-83.216	-90.036
(p-values)	(0.046)	(0.001)	(0.077)	(0.029)	(0.497)	(0.113)	(0.004)	(0.012)
revenu	0.019	0.024	0.005	0.007	0.018	0.009	0.013	0.009
(p-values)	(0.019)	(0.021)	(0.177)	(0.283)	(0.17)	(0.207)	(0.046)	(0.469)
SeRenseigne	80.095	69.092	8.71	25.374	61.741	26.273	33.02	44.949
(p-values)	(0)	(0.006)	(0.489)	(0.312)	(0.222)	(0.214)	(0.163)	(0.317)
RISKVECUDUM	-27.789	-70.139	-15.696	-28.289	-23.019	-30.62	-35.243	-19.976
(p-values)	(0.051)	(0)	(0.033)	(0.001)	(0.12)	(0.13)	(0.005)	(0.163)
NbreMoyenInfo	11.76	13.898	4.732	8.878	5.555	8.858	8.276	13.371
(p-values)	(0.021)	(0.031)	(0.069)	(0.02)	(0.257)	(0.086)	(0.043)	(0.021)
PROBSUBDIX	2.014	2.742	0.814	1.28	3.066	0.739	1.316	3.244
(p-values)	(0)	(0)	(0.046)	(0.067)	(0)	(0.174)	(0.073)	(0)
PREFPRESENT	-7.413	-12.59	-2.477	-5.888	-9.92	-3.116	-9.175	-11.357
(p-values)	(0.003)	(0)	(0.008)	(0.001)	(0)	(0.157)	(0)	(0)
LOSSLOV	18.693	26.118	2.52	18.253	25.229	4.534	18.741	18.87
(p-values)	(0.02)	(0.01)	(0.603)	(0.012)	(0.05)	(0.552)	(0.003)	(0.136)
HAPPY	8.006	14.91	4.7	7.543	1.555	7.761	12.689	15.564
(p-values)	(0.028)	(0.002)	(0.004)	(0.014)	(0.71)	(0.014)	(0)	(0.001)
PATRIMOINE	0.088	0.108	0.015	0.086	0.099	0.032	0.083	0.152
(p-values)	(0.011)	(0.013)	(0.434)	(0.023)	(0.161)	(0.353)	(0.069)	(0.021)
PTSDDUM	40.894	50.934	20.74	29.446	93.797	29.366	35.611	121.391
(p-values)	(0.082)	(0.081)	(0.252)	(0.401)	(0.061)	(0.207)	(0.325)	(0.014)

Table IV: Results by model, Collective Action Scenario

Linear models not accounting for censored WTP

Regarding the **OLS** results, socio-demographic variables seem less important than variables characterizing preferences and psychological variables. Living in a house or on the ground flood (*HousingRisk*) does not seem to have an impact (although Novotny

[4] A dummy variable controls for the order of the scenario in the regressions but is never significantly different from zero.

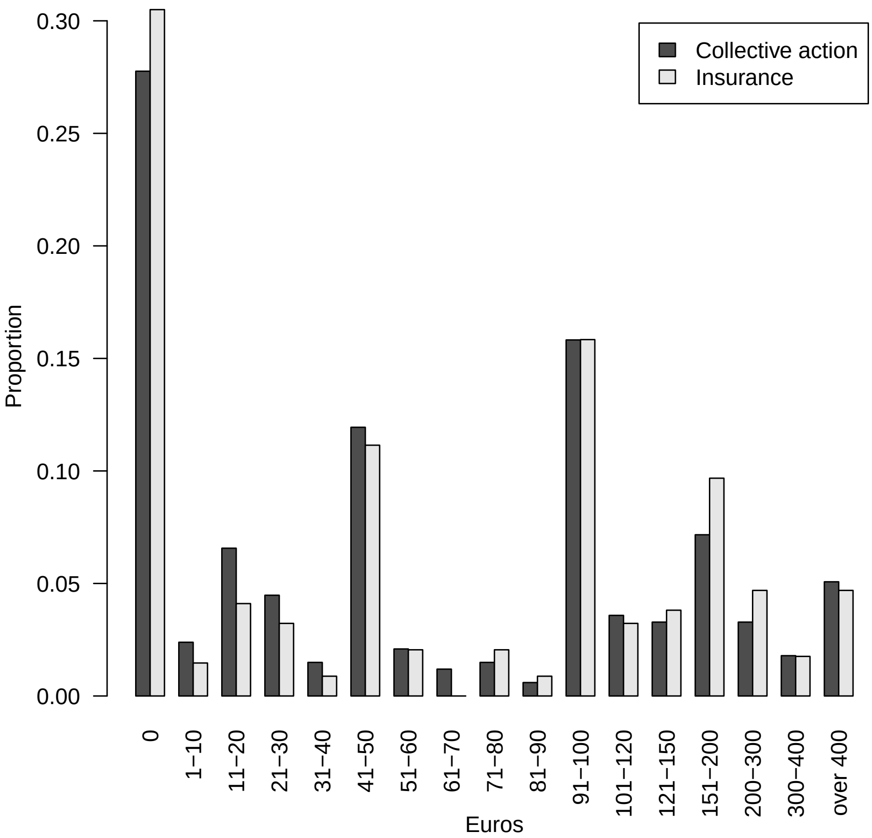


Figure 4.1: WTP distribution among respondents giving a WTP

et al. 2001; Hung 2005; and Deronzier and Terra 2006 find a positive impact), which suggests that the objective risk of flooding does not influence WTP. *Age* is not significant, undoubtedly because the scenario would affect all of a respondent's household members rather than simply the respondents themselves, as suggested in Konishi and Adachi (2011). Its impact is positive in Joseph et al. (2015), Defra (2005), Deronzier and Terra (2006) or Owusu et al. (2015); but negative in Abbas et al. (2014).

Income is positively related to WTP, which is intuitive and argues for the validity of the CV surveys (which Bishop and Woodward 1995 defined as *theoretical construct validity*). This positive impact is consistent with all flood CV studies (with an inverted U-shape in Joseph et al. 2015 and Deronzier and Terra 2006).

The amount of information (*NbrInfo*) and having looked for information about flood risk when moving to the current place of residence (*Inform*) have a positive impact on WTP. This is consistent with previous findings that individual preparedness for flooding (Zhai et al. 2006) or being insured against flood (Shabman et al. 1998) have a positive effect on WTP.

Individual preferences and psychological variables seem important. WTP is increasing with the subjective probability of a future flood (*ProbaFlood*), consistent with the findings that flood fear — either expressed in terms of flood-related stress, worrying about housing value reductions or future flooding — positively and significantly influence WTP in Novotny et al. (2001), Defra (2005) and Joseph et al. (2015). WTP is also increasing with subjective well-being (*Happy*), and decreasing with preference for the present (*Impatience*). A counterintuitive result is the negative relationship between WTP and loss loving score (*LossLover*), which may indicate that preferences elicited from purely monetary gains and losses in loteries are bad predictors of those elicited from hypothetical situations mimicking real life.

Note that the OLS estimates are inefficient because residuals are non-normally distributed (Jarque-Bera p-value ≤ 0.0001) and heteroskedastic (Breusch-Pagan p-value ≤ 0.0001).

Figure 4.2 provides a representation of the **QR** estimates. For each variable, the horizontal axis represents the conditional quantiles in the range 10%-90%[5] and the vertical axis represents the values of the coefficient. The black lines represent the QR estimates (along with the 95% CI in gray) and the OLS estimates are represented by the dashed

[5] Unlike Krishnamurthy and Kriström (2016) we did not face computational issues when estimating the conditional quantiles below 25%.

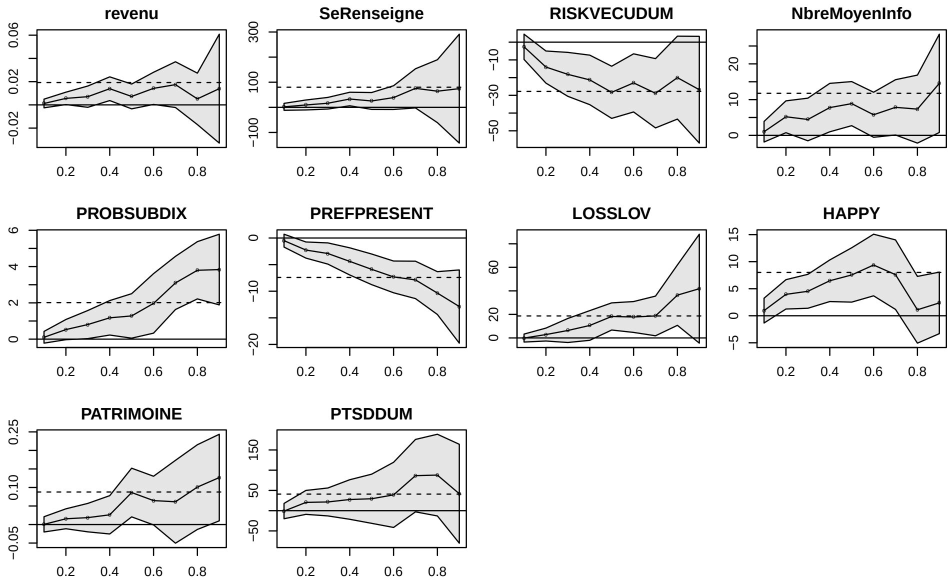
lines. We find significant heterogeneity along the conditional WTP distribution, which was obviously not accounted for by the OLS model. For instance, the marginal effect on the conditional WTP of being a loss lover (*LossLover*) increases along the conditional distribution, i.e. has a greater impact at the top of the conditional distribution than at the bottom, whereas preferring the present (*Impatience*) decreases along the conditional distribution.

The effect of income on WTP is positive but only significant in the middle of the conditional distribution, as in Krishnamurthy and Kriström (2016). The left-censored part of the conditional distribution may represent a fraction of respondents whose stated WTP does not correspond to their true WTP, and thus cannot be predicted by their income. This could also explain why the marginal effect of most of the variables is non-significant at the lowest conditional quantiles. The WTP of the observations at the top of the conditional distribution also seems not to be affected by income. This could be due to the fact that respondents may not consider their income when stating WTP, leading to an upward bias in WTP. These respondents are likely to be located at the top of the conditional distribution, and the marginal effect of income on their WTP is almost zero, since income is not accounted for in the decision.

We also observe that WTP is more sensitive to individual preferences in the highest conditional quantiles. Following the same reasoning as above, some respondents put more weight on their preferences than on their resources, so they give a high WTP. As a result, the WTP at the top of the conditional distribution is greatly influenced by preferences and not by income, while the WTP at the center of the conditional distribution seems to be affected by both. Thus QR could reveal heterogeneity in respondents' attitude to the survey: for instance, how seriously they consider the income constraint when stating their WTP.

Models accounting for censored WTP

Tobit's marginal effects on latent WTP are comparable in sign and significance with those of the OLS model, although the latter are, as expected, biased towards zero compared to Tobit. The only exception is the impact of having already experienced a flood (*PastExperience*), which becomes significant. Note however that, because of the rejection of the homoskedasticity assumption (see above), Tobit estimates are likely to be biased according to the results of the Monte Carlo simulations.



Note: The black lines represent the QR estimates (along with the 95% CI in gray) and the OLS estimates are represented by the dashed lines.

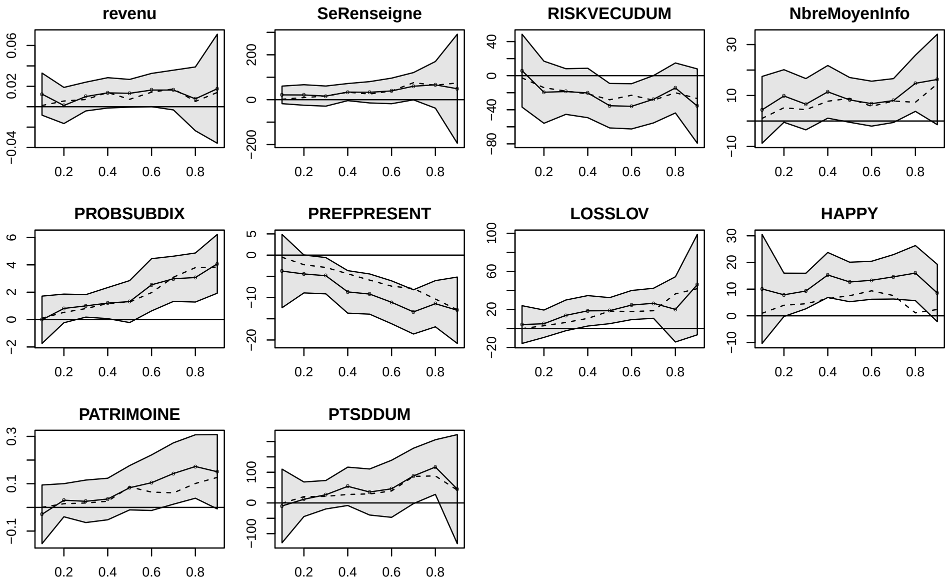
Figure 4.2: QR estimates, Collective Action scenario

The **CQR** results in figure 4.3 do not show major differences from the QR results. The black lines represent the CQR estimates (along with 95% CI in gray) and the dashed lines the QR estimates. Even though differences between the two estimates are small, the QR estimates tend to be biased towards zero compared to the CQR estimates, particularly for *Happy*, *NbrInfo*, and *LossLover*, which is consistent with the findings in the Monte Carlo experiment.

Overall, we find clear signs of heterogeneity among the respondents, differing according to the unobserved determinants of the WTP (i.e. to their rank in the conditional WTP distribution). Although it is difficult to determine what exactly is embedded in these unobserved components (attitude to the survey, differences in sensitivity to hypothetical bias, etc.), (C)QR shows how they can affect the relation between WTP and observed characteristics.

4.5 Conclusion

This article confirms the advantages of (C)QR models w.r.t. standard conditional mean estimates for analyzing WTP data, first through a Monte Carlo experiment and second, by applying them to a CV study on flood risk protection. Moreover, the CQR model outperforms the other three models due to its ability to account both for censoring and



Note: The black lines represent the CQR estimates (along with 95% CI in gray) and the dashed lines the QR estimates.

Figure 4.3: CQR and QR estimates, Collective Action scenario

heteroskedasticity, two common problems in WTP data. Using CQR appears relevant with any data that entail this type of problem (like number of occurrences of an event or quantities consumed) in many fields such as agriculture, energy, climate, environment and health. This also applies to economic data either directly observed on markets (prices, rates, taxes), indirectly revealed (shadow prices), or stated in surveys (WTP). Nor did we find significant differences in WTP elicited depending on the beneficiaries in the two scenarios tested.

Two important issues remain for further research. First, despite the fact that many models can account for heterogeneity in the coefficients, they have not been comprehensively compared to date. A study comparing these models and defining the kind of heterogeneity accounted for by each would be useful when choosing the proper model to analyze data from non-market valuation studies. Second, from a practical perspective, although the use of a circular payment card rules out elicitation-specific anchoring, the respondent may have relied on his/her answer from the first scenario when answering the second. As a result, the WTP elicited in second position does not necessarily represent the true WTP for this scenario, but may include some component of the WTP elicited in first position, although no significant effect of scenario order has been found so far. Consequently, future work could consider a joint analysis of both WTPs, by explicitly modeling potential anchoring effects, for instance.

Bibliography

Abbas, A., Amjath-Babu, T., Kächele, H., and Müller, K. (2014). Non-structural flood risk mitigation under developing country conditions: an analysis on the determinants of willingness to pay for flood insurance in rural pakistan. *Natural Hazards*, 75(3):2119–2135.

Akter, S., Brouwer, R., Choudhury, S., and Aziz, S. (2009). Is there a commercially viable market for crop insurance in rural Bangladesh? *Mitigation and Adaptation Strategies for Global Change*, 14(3):215–229.

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, 24:3–61.

Bateman, I., Carson, R., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., and Loomes, G. (2002). *Economic Valuation with Stated Preference Techniques.* Edward Elgar Publishing.

Belluzzo, W. (2004). Semiparametric approaches to welfare evaluations in binary response models. *Journal of Business & Economic Statistics*, 22(3):322–330.

Bishop, R. and Woodward, R. (1995). Valuation of environmental amenities under certainty. In Bromley, D., editor, *Handbook of Environmental Economics*, pages 543–667. Blackwell Publishers, Oxford.

Botzen, W. and van den Bergh, J. (2012). Risk attitudes to low-probability climate change risks: Wtp for flood insurance. *Journal of Economic Behavior and Organization*, 82(1):151–166.

Boxall, P. C. and Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: a latent class approach. *Environmental and Resource Economics*, 23(4):421–446.

Brummett, R., Nayga, R., and Wu, X. (2007). On the use of cheap talk in new product valuation. *Economics Bulletin*, 2(1):1–9.

Buchinsky, M. and Hahn, J. (1998). An Alternative Estimator for the Censored Quantile Regression Model. *Econometrica*, 66(3):653–671.

Cameron, T. and Huppert, D. (1989). OLS versus ML estimation of non-market resource values with payment card interval data. *Journal of Environmental Economics and Management*, 17:230–246.

Carson, R. and Groves, T. (2011). Incentive and information properties of preference questions commentary and extensions. In Bennett, J., editor, *International Handbook on Non-market Valuation*. Edward Elgar, Northampton.

Carson, R. T. and Hanemann, W. M. (2005). *Valuing Environmental Changes*, volume 2 of *Handbook of Environmental Economics*. Elsevier.

Centre Européen de Prévention du Risque Inondation (2013). Indemnisation des victimes. "http://www.cepri.net/indemnisation-des-victimes.html, acceded on 28/03/13".

Chanel, O., Chichilnisky, G., Massoni, S., Vergnaud, J.-C., and Lyk-Jensen, S. V. (2013a). Décision en présence d'incertitude et d'émotions face á des risques de catastrophes naturelles (RISKEMOTION), Final Report ANR-08-RISKNAT-007-01, Greqam, Marseille, France.

Chanel, O., Chichilnisky, G., Massoni, S., Vergnaud, J.-C., and Vincent Lyk-Jensen, S. (2013b). Décision en présence d'incertitude et d'émotions face á des risques de catastrophes naturelles (riskemotion), final report anr-08-risknat-007-01, greqam, marseille, france.

Chanel, O., Makhloufi, K., and Abu-Zaineh, M. (2017). Can a circular payment card format effectively elicit preferences? Evidence from a survey on a mandatory health insurance scheme in Tunisia. *Applied Health Economics and Health Policy*, 15(3):385–398.

Chen, J.-e. and Kashiwagi, M. (2017). The japanese taylor rule estimated using censored quantile regressions. *Empirical Economics*, 52(1):357–371.

Chernozhukov, V., Fernández-Val, I., and Kowalski, A. (2015). Quantile regression with censoring and endogeneity. *Journal of Econometrics*, 186(1):201–221.

Deaton, A. (1997). *The analysis of household surveys: a microeconometric approach to development policy*. World Bank Publications.

Defra (2005). The appraisal of human-related intangible impacts of flooding, R&D Technical Report FD2005/TR, 352 p. Defra Environment Agency, Flood and Coastal Defence R&D Programme.

Deronzier and Terra (2006). Bénéfices économiques de la protection contre le risque d'inondation. *Document de travail D4E, Série Etudes 06-E05.*

Diamond, P. A. and Hausman, J. A. (1994). Contingent valuation: Is some number better than no number? *Journal of Economic Perspectives*, 8(4):45–64.

Dodd, M. (2014). Intertemporal discounting as a risk factor for high BMI: Evidence from Australia, 2008. *Economics and Human Biology*, 12(1):83–97.

Ferrini, S. and Scarpa, R. (2007). Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management*, 53:342–363.

Fitzenberger, B. (1994). A note on estimating censored quantile regressions. Technical report, Discussion Paper, Center for International Labor Economics (CILE), University of Konstanz.

Friederichs, P. and Hense, A. (2007). Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly Weather Review*, 135(6):2365–2378.

Furno, M., Verneau, F., and Sannino, G. (2016). Assessing hypothetical bias: An analysis beyond the mean of functional food. *Food Quality and Preference*, 50:15–26.

Ghanbarpour, M., Saravi, M., and Salimi, S. (2014). Floodplain inundation analysis combined with contingent valuation: Implications for sustainable flood risk management. *Water Resources Management*, 28(9):2491–2505.

Glenk, K. and Fischer, A. (2010). Insurance, prevention or just wait and see? Public preferences for water management strategies in the context of climate change. *Ecological Economics*, 69(11):2279–2291.

Grelot, F. (2004). *Gestion collective des inondations. Peut-on tenir compte de l'avis de la population dans la phase d'évaluation économique a priori?* PhD thesis, Ecole nationale supérieure d'arts et métiers-ENSAM.

Gustavsen, G. W. and Rickertsen, K. (2006). A censored quantile regression analysis of vegetable demand: The effects of changes in prices and total expenditure. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 54(4):631–645.

Hanley, N., Kriström, B., and Shogren, J. (2009). Coherent arbitrariness: On value uncertainty for environmental goods. *Land Economics*, 85(1):41–50.

Hausman, J. (2012). Contingent valuation: From dubious to hopeless. *Journal of Economic Perspectives*, 26(4):43–56.

Hoshino, T. (2013). Estimation of the preference heterogeneity within stated choice data using semiparametric varying-coefficient methods. *Empirical Economics*, 45(3):1129–1148.

Huang, H. and Chen, Z. (2015). Bayesian composite quantile regression. *Journal of Statistical Computation and Simulation*, 85(18):3744–3754.

Hung, H.-C. (2005). The attitude towards flood insurance purchase when respondents' preferences are uncertain: a fuzzy approach. *Journal of Risk Research*, 12(1):239–258.

Hung, W.-T., Shang, J.-K., and Wang, F.-C. (2010). Pricing determinants in the hotel industry: Quantile regression analysis. *International Journal of Hospitality Management*, 29(3):378–384.

Intergovernmental Panel on Climate Change (2013). Summary for policymakers. In Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Jackman, M. and Lorde, T. (2013). Why buy when we can pirate? The role of intentions and willingness to pay in predicting piracy behavior. *International Journal of Social Economics*, 41(9):801–819.

Jerome, G., Alavi, R., Daumit, G., Wang, N.-Y., Durkin, N., Yeh, H.-C., Clark, J., Dalcin, A., Coughlin, J., Charleston, J., Louis, T., and Appel, L. (2015). Willingness to pay for continued delivery of a lifestyle-based weight loss program: The hopkins power trial. *Obesity*, 23(2):282–285.

Johansson, P.-O. and Kriström, B. (2015). On the Social Cost of Water-Related Disasters. *Water Economics and Policy*, 01(03):1550015.

Johnston, R., Swallow, S., and Weaver, T. (1999). Estimating willingness to pay and resource tradeoffs with different payment mechanisms: An evaluation of a funding guarantee for watershed management. *Journal of Environmental Economics and Management*, 38(1):97–120.

Joseph, R., Proverbs, D., and Lamond, J. (2015). Assessing the value of intangible benefits of property level flood risk adaptation (PLFRA) measures. *Natural Hazards*, 79(2):1275–1297.

Kleiber, C. and Zeileis, A. (2008). *Applied Econometrics with R*. Springer-Verlag, New York. ISBN 978-0-387-77316-2.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press, Cambridge; New York.

Koenker, R. (2015). *quantreg: Quantile Regression*. R package version 5.19.

Konishi, Y. and Adachi, K. (2011). A framework for estimating willingness-to-pay to avoid endogenous environmental risks. *Resource and Energy Economics*, 33(1):130–154.

Kowalski, A. E. (2016). Censored Quantile instrumental variable estimates of the price elasticity of expenditure on medical care. *Journal of Business and Economic Statistics*, 34(1):107–117.

Kreibich, H., Seifert, I., Thieken, A., Lindquist, E., Wagner, K., and Merz, B. (2011). Recent changes in flood preparedness of private households and businesses in Germany. *Regional Environmental Change*, 11(1):59–71.

Krishnamurthy, C. K. B. and Kriström, B. (2016). Determinants of the price-premium for green energy: Evidence from an oecd cross-section. *Environmental and Resource Economics*, 64:173–204.

Kuo, Y.-L. (2016). Is there a trade-off between households' precautions, mitigations and public protection for flood risk? *Environmental Hazards*, 15(4):311–326.

Laamanen, J.-P. (2013). Estimating demand for opera using sales system data: the case of finnish national opera. *Journal of Cultural Economics*, 37(4):417–432.

Lagerkvist, C. J. and Okello, J. (2016). Using the integrative model of behavioral prediction and censored quantile regression to explain consumers' revealed preferences for food safety: Evidence from a field experiment in kenya. *Food Quality and Preference*, 49:75 – 86.

Landry, C., Hindsley, P., Bin, O., Kruse, J., Whitehead, J., and Wilson, K. (2011). Weathering the storm: Measuring household willingness-to-Pay for risk-reduction in post-Katrina New Orleans. *Southern Economic Journal*, 77(4):991–1013.

Last, A.-K. (2007). The monetary value of cultural goods: A contingent valuation study of the municipal supply of cultural goods in Lueneburg, Germany. Working paper series in economics 63, University of Lueneburg.

Lavín, F. V., Flores, R., and Ibarnegaray, V. (2017). A Bayesian quantile binary regression approach to estimate payments for environmental services. *Environment and Development Economics*, 22(2):156–176.

Lusk, J., Traill, W., House, L., Valli, C., Jaeger, S., Moore, M., and Morrow, B. (2006). Comparative advantage in demand: Experimental evidence of preferences for genetically modified food in the United States and European Union. *Journal of Agricultural Economics*, 57(1):1–21.

Meleddu, M., Pulina, M., and Ladu, M.-G. (2013). Evaluating the demand for cultural goods: Just income and tastes do matter? *Economia Della Cultura*, XXIII(2):203–215.

Ministère de l'écologie du développement durable et de l'énergie (2012). Première évaluation nationale des risques d'inondation, principaux résultats. EPRI 2011, mimeo.

Motamed, M., McPhail, L., and Williams, R. (2016). Corn area response to local ethanol markets in the united states: A grid cell level analysis. *American Journal of Agricultural Economics*, 98(3):726–743.

Nahuelhual, L., Loureiro, M. L., and Loomis, J. (2004). Using random parameters to account for heterogeneous preferences in contingent valuation of public open space. *Journal of Agricultural and Resource Economics*, 29(3):537–552.

Navrud, S., Tuan, T., and Tinh, B. (2012). Estimating the welfare loss to households from natural disasters in developing countries: A contingent valuation study of flooding in Vietnam. *Global Health Action*, 5(1).

Notaro, S. and De Salvo, M. (2010). Estimating the economic benefits of the landscape function of ornamental trees in a sub-mediterranean area. *Urban forestry & urban greening*, 9(2):71–81.

Novotny, V., Clark, D. F., Griffin, R. J., Bartosova, A., Booth, D., and Anderson, R. (2001). Risk based urban watershed management-integration of water quality and flood control objectives. usepa/nsf/usda star watershed program final report. Technical report, www.coe.neu.edu/environment/UrbanWatershed.

O'Garra, T. and Mourato, S. (2006). Public Preferences for Hydrogen Buses: Comparing Interval Data, OLS and Quantile Regression Approaches. *Environmental and Resource Economics*, 36(4):389–411.

Owusu, S., Wright, G., and Arthur, S. (2015). Public attitudes towards flooding and property-level flood protection measures. *Natural Hazards*, 77(3):1963–1978.

Paarsch, H. J. (1984). A Monte Carlo comparison of estimators for censored regression models. *Journal of Econometrics*, 24(1):197–213.

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, 32(1):143–155.

Ren, J. and Wang, H. H. (2016). Rural Homeowners' Willingness to Buy Flood Insurance. *Emerging Markets Finance and Trade*, 52(5):1156–1166.

Reynaud, A. and Nguyen, M.-H. (2016). Valuing Flood Risk Reductions. *Environmental Modeling and Assessment*, 21(5):603–617.

Seymour, J., McNamee, P., Scott, A., and Tinelli, M. (2010). Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. *Health Economics*, 19(6):683–696.

Shabman, L., Stephenson, K., Thunberg, E., Dietz, B., Driscoll, P., and O'Grady, K. (1998). Comparing benefit estimation techniques: residential flood hazard reduction benefits in roanoke, virginia. Technical report, DTIC Document.

Sigelman, L. and Zeng, L. (1999). Analyzing censored and sample-selected data with tobit and heckit models. *Political Analysis*, 8(02):167–182.

Sigma (2016). Natural catastrophes and man-made disasters in 2015.

Simpson, K. and Hanley, N. (2016). Managed Realignment for Flood Risk Reductions: What are the Drivers of Public Willingness to Pay? Working papers, University of St. Andrews, Department of Geography and Sustainable Development.

Strazzera, E., Scarpa, R., Calia, P., Carrot, G., and Willis, K. (2003). Modelling zero values and protest responses in contingent valuation surveys. *Applied Economics*, 44(3):1165 – 1192.

Su, Q. (2012). A quantile regression analysis of the rebound effect: Evidence from the 2009 national household transportation survey in the United States. *Energy Policy*, 45:368–377.

Thunberg, E. (1988). Willingness to pay for property and non-property flood hazard reduction benefits: An experiment using the contingent value survey method. Ph.D dissertation, Department of Agricultural Economics, Virginia Tech VPI&SU.

Tinelli, M., Ryan, M., Bond, C., and Scott, A. (2013). Valuing benefits to inform a clinical trial in pharmacy: Do differences in utility measures at baseline affect the effectiveness of the intervention? *PharmacoEconomics*, 31(2):163–171.

Trent, M., Lehmann, H., Qian, Q., Thompson, C., Ellen, J., and Frick, K. (2011). Adolescent and parental utilities for the health states associated with pelvic inflammatory disease. *Sexually Transmitted Infections*, 87(7):583–587.

Viscusi, W. K., Huber, J., and Bell, J. (2012). Heterogeneity in values of morbidity risks from drinking water. *Environmental and Resource Economics*, 52(1):23–48.

Vossler, C. and Holladay, J. (2016). Alternative Value Elicitation Formats in Contingent Valuation: A New Hope. Working Papers 2016-02, University of Tennessee, Department of Economics.

Vossler, C. and Watson, S. B. (2013). Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. *Journal of Economic Behavior & Organization*, 86(C):137–147.

Werritty, A., Houston, D., Ball, T., Tavendale, A., and Black, A. (2007). Exploring the social impacts of flooding and flood risk in Scotland. Scottish Executive Social Research, Report, 137 p.

Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data*, volume 1 of *MIT Press Books*. The MIT Press.

Yang, S.-H., Hu, W., Mupandawana, M., and Liu, Y. (2012a). Consumer willingness to pay for fair trade coffee: A chinese case study. *Journal of Agricultural and Applied Economics*, 44(1):21–34.

Yang, S.-H., Hu, W., Mupandawana, M., and Liu, Y. (2012b). Consumer willingness to pay for fair trade coffee: A Chinese case study. *Journal of Agricultural and Applied Economics*, 44(1):21–34.

Yao, X.-L., Liu, Y., and Yan, X. (2014). A quantile approach to assess the effectiveness of the subsidy policy for energy-efficient home appliances: Evidence from Rizhao, China. *Energy Policy*, 73:512–518.

Zhai, G., Sato, T., Fukuzono, T., Ikeda, S., and Yoshida, K. (2006). Willingness to pay for flood risk reduction and its determinants in Japan. *JAWRA Journal of the American Water Resources Association*, 42(4):927–940.

4.A Literature review of CV studies on flood risk

We looked for studies providing estimations of WTP to decrease flood risk, via a structured literature search with keywords in the Scopus database and on Google Scholar.[6] We found 26 different surveys (published in some 35 papers from 1988 to 2017) that use stated preference methods to explore respondents' WTP to reduce the risk of flooding in their place of residence.

Nine studies were discarded for various reasons: unsuccessful attempts to get the original document (Thunberg 1988; a Ph.D thesis dealing with 142 US respondents; or Kreibich et al. 2011; a study on 310 Germans); WTP not in monetary terms but in person-days (Navrud et al. 2012, in Vietnam) or in-kind (Akter et al. 2009, in Bengladesh) contributions; no CV data included (Werritty et al. 2007 with a sociology-oriented study; or Johnston et al. 1999; Landry et al. 2011; Botzen and van den Bergh 2012; Reynaud and Nguyen 2016, with discrete choice experiment surveys).

Table V summarizes the 17 remaining studies with their mean WTP (expressed in € 2012). However, comparisons are difficult, due to differences in the risk reduction proposed, in purchasing power across countries, and in three factors mentioned in the table and detailed below: elicitation format, beneficiaries and nature of effects assessed.

The elicitation format affects the nature of the stated WTP - binary for the single-bounded format, discrete for the payment card, descending bid for the double-bounded format, and continuous for the open-ended format - and consequently, the econometric analysis. It also affects the quality of the elicited WTP due to the biases or errors inherent to the different formats (Carson and Hanemann 2005).

The beneficiaries of the scenario are involved in two types of action. An individual action scenario evaluates the WTP for a decrease in the respondent's own consequences from a flood, typically by purchasing insurance or adapting the property against flood risk. A collective action scenario evaluates the WTP for a decrease in flood risk, e.g. by financial participation in collective protection measures. Seven studies propose a scenario for individual action only (Defra 2005; Hung 2005; Abbas et al. 2014; Joseph et al. 2015; Owusu et al. 2015; Ren and Wang 2016; Kuo 2016), seven studies for collective action only (Shabman et al. 1998; Novotny et al. 2001; Grelot 2004; Zhai et al. 2006; Glenk and Fischer 2010; Simpson and Hanley 2016; Vossler and Holladay 2016), and three propose two scenarios, one for each type of action (Deronzier and Terra 2006; Chanel et al. 2013a; Ghanbarpour et al. 2014).

[6] The keywords used are "contingent valuation" and "flood", "stated preference" and "flood", "discrete choice experiment" and "flood", and "conjoint analysis" and "flood".

The third factor is the nature of the flood-related effects to be impacted by the measure proposed in the scenario. These effects may be purely tangible, with the scenario offering an insurance policy that fully compensates for monetary losses; purely intangible, with a scenario stipulating that only emotional or psychological effects are to be considered; or both, with a scenario proposing to avoid flooding thanks to (individual or collective) measures against flood risk.

Author	Country	N	Elicit. format	Beneficiaries	Effects	Mean annual WTP (€)
Defra (2005)	England	1510	PC	Individual	Intangible	236-314
Hung (2005)	Taiwan	405	SB	Individual	Tangible	108-145
Abbas et al. (2014)	Pakistan	250	DB	Individual	Tangible	7.3
Joseph et al. (2015)	England	243	OE	Individual	Intangible	850
Owusu et al. (2015)	Scotland	256	OE	Individual	Intangible	1037
Ren and Wang (2016)	China	1322	DBDC	Individual	Tangible	3-9
Kuo (2016)	Taiwan	600	SB	Individual	Tangible /Both	47/ NA
Shabman et al. (1998)	USA	74	PC	Collectivity	Both	47-140
Novotny et al. (2001)	USA	1000	SB	Collectivity	Both	97
Grelot (2004)	France	213	SB	Collectivity	Both	46-58
Zhai et al. (2006)	Japan	428	PC	Collectivity	Both	24-41
Glenk and Fischer (2010)	Scotland	1033	PC	Collectivity	Both	67
Simpson and Hanley (2016)	Scotland	593	PC	Collectivity	Both	71
Vossler and Holladay (2016)	USA	1719	OE/PC/SB	Collectivity	Both	7-18
Deronzier and Terra (2006)	France	500	DB/OE	Individual/Collectivity	Tangible/Both	41.1/40.8
Chanel et al. (2013a)	France	599	PC	Individual/Collectivity	Tangible/Both	107/103
Ghanbarpour et al. (2014)	Iran	83	OE	Individual/Collectivity	Tangible/Both	36/45

Note: Elicitation format (DB: descending bid, DBDC: double-bounded dichotomous choice, OE: open-ended, PC: payment card, SBDC: single-bounded dichotomous choice), NA: Non available, N: Sample size

Table V: Summary of 17 CV Studies on Flood Risk

4.B Technical note

Computations are performed with R and the quantreg (Koenker 2015) and AER (Kleiber and Zeileis 2008) packages (codes available upon request). The CQR estimation uses the BRCENS algorithm of Fitzenberger (1994), which is based on the Barrodale-Roberts-Algorithm for standard QR.

Two standard criteria used in the Monte Carlo and QR literature to compare performance of different models are Mean Bias and Root-Mean-Square Error (RMSE). They both measure the magnitude of the deviations of Monte Carlo estimates from the true estimate (Paarsch 1984; Buchinsky and Hahn 1998; Chernozhukov et al. 2015). For a given specification, the Mean Bias is defined as:

$$\frac{1}{R}\sum_{r=1}^{R}(\hat{b}_r - b) \tag{4.12}$$

and the RMSE is defined as:

$$\sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{b}_r - b}{b}\right)^2} \tag{4.13}$$

where b is the true value of the marginal effect of x on WTP and $\hat{b}_r$ the estimation of the marginal effect of x_r on WTP_r for the r^{th} of the R Monte Carlo replications.

Note that this marginal effect is not equal to β for QR and CQR when the quantile is different from 0.5. For the homoskedastic case, it is equal to β. For the heteroskedastic case, it is equal to $\beta + \gamma_{j=1,2}F_u^{-1}(\tau)$, with $F_u^{-1}(\tau) = 0$ if and only if $\tau = 0.5$ for a symmetric, zero-centered distribution (as the standard normal).

Moreover, for symmetric (zero-centered) distributions and for conditional mean models, the mean is equal to the median (and equals zero for zero-centered distributions) and $b = \beta$.

4.C Additional Monte-Carlo results

DGP	Statistics	OLS	Tobit	QR 25%	QR 50%	QR 75%	CQR 25%	CQR50%	CQR 75%
c=0%, γ =0, n=50	MB	0.000	0.001	0.008	0.000	-0.006	0.008	0.000	-0.006
c=0%, γ =0, n=50	p-val	0.578	0.558	0.000	0.717	0.000	0.000	0.709	0.000
c=0%, γ =0, n=300	MB	-0.000	-0.000	0.001	-0.000	-0.002	0.001	-0.000	-0.002
c=0%, γ =0, n=300	p-val	0.314	0.351	0.012	0.203	0.000	0.010	0.209	0.000
c=0%, γ =0, n=1000	MB	-0.000	0.000	0.000	0.000	-0.000	0.000	0.000	-0.000
c=0%, γ =0, n=1000	p-val	0.944	0.958	0.150	0.793	0.055	0.135	0.754	0.056
c=0%, γ =0.4, n=50	MB	0.002	0.002	0.043	0.003	-0.036	0.043	0.003	-0.036
c=0%, γ =0.4, n=50	p-val	0.587	0.573	0.000	0.383	0.000	0.000	0.375	0.000
c=0%, γ =0.4, n=300	MB	0.001	0.001	0.008	0.002	-0.004	0.008	0.002	-0.004
c=0%, γ =0.4, n=300	p-val	0.533	0.513	0.000	0.206	0.002	0.000	0.202	0.002
c=0%, γ =0.4, n=1000	MB	-0.001	-0.001	0.001	-0.001	-0.003	0.001	-0.001	-0.003
c=0%, γ =0.4, n=1000	p-val	0.307	0.330	0.125	0.358	0.000	0.109	0.376	0.000
c=0%, γ =0.8, n=50	MB	0.001	0.001	0.066	0.007	-0.064	0.066	0.007	-0.064
c=0%, γ =0.8, n=50	p-val	0.828	0.823	0.000	0.184	0.000	0.000	0.180	0.000
c=0%, γ =0.8, n=300	MB	0.000	0.000	0.014	-0.001	-0.012	0.014	-0.001	-0.012
c=0%, γ =0.8, n=300	p-val	0.897	0.884	0.000	0.703	0.000	0.000	0.715	0.000
c=0%, γ =0.8, n=1000	MB	-0.001	-0.001	0.003	0.001	-0.004	0.003	0.001	-0.004
c=0%, γ =0.8, n=1000	p-val	0.741	0.763	0.040	0.612	0.001	0.035	0.595	0.001
c=20%, γ =0, n=50	MB	-0.094	0.001	-0.105	-0.085	-0.079	0.013	0.002	-0.020
c=20%, γ =0, n=50	p-val	0.000	0.135	0.000	0.000	0.000	0.000	0.048	0.000
c=20%, γ =0, n=300	MB	-0.074	-0.000	-0.094	-0.070	-0.057	0.002	0.000	-0.005
c=20%, γ =0, n=300	p-val	0.000	0.717	0.000	0.000	0.000	0.000	0.833	0.000
c=20%, γ =0, n=1000	MB	-0.069	-0.000	-0.092	-0.067	-0.053	0.001	-0.000	-0.001
c=20%, γ =0, n=1000	p-val	0.000	0.180	0.000	0.000	0.000	0.003	0.266	0.000
c=20%, γ =0.4, n=50	MB	-0.096	0.038	-0.130	-0.123	-0.141	0.088	0.015	-0.059
c=20%, γ =0.4, n=50	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
c=20%, γ =0.4, n=300	MB	-0.076	0.033	-0.160	-0.123	-0.106	0.015	0.003	-0.012
c=20%, γ =0.4, n=300	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.000
c=20%, γ =0.4, n=1000	MB	-0.072	0.031	-0.166	-0.124	-0.105	0.004	0.001	-0.003
c=20%, γ =0.4, n=1000	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.083	0.001
c=20%, γ =0.8, n=50	MB	-0.086	0.092	-0.136	-0.139	-0.171	0.245	0.044	-0.077
c=20%, γ =0.8, n=50	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
c=20%, γ =0.8, n=300	MB	-0.076	0.076	-0.200	-0.154	-0.132	0.026	0.001	-0.024
c=20%, γ =0.8, n=300	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.720	0.000
c=20%, γ =0.8, n=1000	MB	-0.066	0.079	-0.204	-0.149	-0.123	0.010	0.002	-0.005
c=20%, γ =0.8, n=1000	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.065	0.000
c=40%, γ =0, n=50	MB	-0.249	-0.002	-0.384	-0.264	-0.194	0.020	0.002	-0.029
c=40%, γ =0, n=50	p-val	0.000	0.117	0.000	0.000	0.000	0.000	0.160	0.000
c=40%, γ =0, n=300	MB	-0.192	-0.001	-0.348	-0.228	-0.149	0.004	0.000	-0.006
c=40%, γ =0, n=300	p-val	0.000	0.110	0.000	0.000	0.000	0.000	0.965	0.000
c=40%, γ =0, n=1000	MB	-0.178	0.000	-0.342	-0.223	-0.142	0.002	0.000	-0.001
c=40%, γ =0, n=1000	p-val	0.000	0.412	0.000	0.000	0.000	0.000	0.114	0.000
c=40%, γ =0.4, n=50	MB	-0.245	0.081	-0.471	-0.347	-0.298	0.226	0.042	-0.076
c=40%, γ =0.4, n=50	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
c=40%, γ =0.4, n=300	MB	-0.193	0.075	-0.480	-0.335	-0.263	0.031	0.008	-0.016
c=40%, γ =0.4, n=300	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
c=40%, γ =0.4, n=1000	MB	-0.180	0.072	-0.481	-0.334	-0.258	0.009	0.001	-0.004
c=40%, γ =0.4, n=1000	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.224	0.000
c=40%, γ =0.8, n=50	MB	-0.238	0.177	-0.522	-0.404	-0.368	1.263	0.148	-0.093
c=40%, γ =0.8, n=50	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
c=40%, γ =0.8, n=300	MB	-0.181	0.173	-0.578	-0.399	-0.312	0.091	0.012	-0.023
c=40%, γ =0.8, n=300	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
c=40%, γ =0.8, n=1000	MB	-0.170	0.170	-0.587	-0.395	-0.306	0.022	0.004	-0.005
c=40%, γ =0.8, n=1000	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.002

Table VI: MB for $\beta 1$

DGP	Statistics	OLS	Tobit	QR 25%	QR 50%	QR 75%	CQR 25%	CQR50%	CQR 75%
c=0%, γ =0, n=50	MB	-0.000	-0.001	0.002	0.000	-0.000	0.002	0.000	-0.001
c=0%, γ =0, n=50	p-val	0.831	0.692	0.311	0.908	0.807	0.356	0.945	0.799
c=0%, γ =0, n=300	MB	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
c=0%, γ =0, n=300	p-val	0.115	0.243	0.067	0.372	0.235	0.109	0.469	0.272
c=0%, γ =0, n=1000	MB	-0.000	-0.000	0.000	-0.000	-0.001	0.000	-0.000	-0.001
c=0%, γ =0, n=1000	p-val	0.586	0.216	0.409	0.745	0.132	0.721	0.562	0.109
c=0%, γ =0.4, n=50	MB	0.002	0.001	-0.000	0.001	0.003	-0.001	0.001	0.003
c=0%, γ =0.4, n=50	p-val	0.446	0.580	0.896	0.727	0.394	0.715	0.770	0.398
c=0%, γ =0.4, n=300	MB	0.002	0.002	0.001	0.001	0.003	0.001	0.001	0.002
c=0%, γ =0.4, n=300	p-val	0.027	0.098	0.422	0.183	0.031	0.636	0.236	0.035
c=0%, γ =0.4, n=1000	MB	0.001	0.001	-0.000	0.000	0.000	-0.000	0.000	0.000
c=0%, γ =0.4, n=1000	p-val	0.047	0.344	0.907	0.514	0.543	0.478	0.728	0.627
c=0%, γ =0.8, n=50	MB	0.000	-0.001	-0.003	0.004	0.000	-0.005	0.004	0.000
c=0%, γ =0.8, n=50	p-val	0.924	0.737	0.412	0.240	0.958	0.224	0.291	0.977
c=0%, γ =0.8, n=300	MB	0.003	0.001	0.003	0.002	0.001	0.002	0.002	0.001
c=0%, γ =0.8, n=300	p-val	0.077	0.499	0.043	0.127	0.621	0.119	0.179	0.655
c=0%, γ =0.8, n=1000	MB	0.000	-0.002	-0.000	-0.000	0.000	-0.001	-0.000	0.000
c=0%, γ =0.8, n=1000	p-val	0.918	0.049	0.843	0.997	0.692	0.327	0.761	0.779
c=20%, γ =0, n=50	MB	0.384	-0.003	0.365	0.361	0.380	-0.005	-0.004	0.027
c=20%, γ =0, n=50	p-val	0.000	0.115	0.000	0.000	0.000	0.068	0.127	0.000
c=20%, γ =0, n=300	MB	0.383	-0.000	0.355	0.356	0.378	-0.002	-0.003	0.001
c=20%, γ =0, n=300	p-val	0.000	0.652	0.000	0.000	0.000	0.055	0.001	0.283
c=20%, γ =0, n=1000	MB	0.382	-0.001	0.353	0.354	0.378	-0.002	-0.001	-0.001
c=20%, γ =0, n=1000	p-val	0.000	0.001	0.000	0.000	0.000	0.000	0.045	0.021
c=20%, γ =0.4, n=50	MB	0.410	-0.107	0.559	0.467	0.421	-0.078	-0.036	0.023
c=20%, γ =0.4, n=50	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
c=20%, γ =0.4, n=300	MB	0.416	-0.114	0.573	0.476	0.435	-0.016	-0.005	0.000
c=20%, γ =0.4, n=300	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.881
c=20%, γ =0.4, n=1000	MB	0.415	-0.119	0.574	0.475	0.434	-0.004	-0.002	-0.001
c=20%, γ =0.4, n=1000	p-val	0.000	0.000	0.000	0.000	0.000	0.001	0.031	0.313
c=20%, γ =0.8, n=50	MB	0.462	-0.226	0.732	0.537	0.439	-0.283	-0.087	0.024
c=20%, γ =0.8, n=50	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
c=20%, γ =0.8, n=300	MB	0.454	-0.272	0.748	0.544	0.449	-0.042	-0.019	-0.006
c=20%, γ =0.8, n=300	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004
c=20%, γ =0.8, n=1000	MB	0.458	-0.277	0.753	0.546	0.454	-0.012	-0.005	-0.001
c=20%, γ =0.8, n=1000	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.261
c=40%, γ =0, n=50	MB	0.799	-0.007	0.815	0.838	0.872	-0.022	-0.015	0.018
c=40%, γ =0, n=50	p-val	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.000
c=40%, γ =0, n=300	MB	0.792	-0.001	0.788	0.822	0.881	-0.006	-0.002	0.003
c=40%, γ =0, n=300	p-val	0.000	0.421	0.000	0.000	0.000	0.000	0.097	0.002
c=40%, γ =0, n=1000	MB	0.791	0.000	0.784	0.822	0.882	-0.002	-0.000	0.001
c=40%, γ =0, n=1000	p-val	0.000	0.832	0.000	0.000	0.000	0.017	0.543	0.234
c=40%, γ =0.4, n=50	MB	0.831	-0.162	1.138	1.000	0.921	-0.238	-0.098	0.020
c=40%, γ =0.4, n=50	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
c=40%, γ =0.4, n=300	MB	0.818	-0.186	1.137	0.995	0.928	-0.042	-0.019	-0.003
c=40%, γ =0.4, n=300	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.204
c=40%, γ =0.4, n=1000	MB	0.816	-0.194	1.138	0.995	0.929	-0.012	-0.004	-0.000
c=40%, γ =0.4, n=1000	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.648
c=40%, γ =0.8, n=50	MB	0.865	-0.356	1.374	1.081	0.933	-2.226	-0.278	-0.030
c=40%, γ =0.8, n=50	p-val	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000
c=40%, γ =0.8, n=300	MB	0.851	-0.423	1.403	1.092	0.943	-0.158	-0.035	-0.010
c=40%, γ =0.8, n=300	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
c=40%, γ =0.8, n=1000	MB	0.850	-0.448	1.408	1.094	0.944	-0.041	-0.011	-0.004
c=40%, γ =0.8, n=1000	p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008

Table VII: MB for $\beta 2$

DGP	Statistics	OLS	Tobit	QR 25%	QR 50%	QR 75%	CQR 25%	CQR50%	CQR 75%
c=0%, γ =0, n=50	RMSE	0.043	0.043	0.059	0.053	0.059	0.059	0.053	0.059
c=0%, γ =0, n=300	RMSE	0.015	0.015	0.020	0.018	0.020	0.020	0.018	0.020
c=0%, γ =0, n=1000	RMSE	0.008	0.008	0.010	0.009	0.010	0.010	0.009	0.010
c=0%, γ =0.4, n=50	RMSE	0.150	0.150	0.194	0.154	0.148	0.194	0.154	0.148
c=0%, γ =0.4, n=300	RMSE	0.082	0.082	0.083	0.066	0.063	0.083	0.066	0.063
c=0%, γ =0.4, n=1000	RMSE	0.054	0.054	0.046	0.037	0.035	0.046	0.037	0.035
c=0%, γ =0.8, n=50	RMSE	0.267	0.267	0.373	0.251	0.211	0.373	0.251	0.211
c=0%, γ =0.8, n=300	RMSE	0.153	0.153	0.158	0.107	0.092	0.158	0.107	0.092
c=0%, γ =0.8, n=1000	RMSE	0.101	0.101	0.087	0.058	0.050	0.087	0.058	0.050
c=20%, γ =0, n=50	RMSE	0.069	0.046	0.090	0.075	0.074	0.068	0.059	0.063
c=20%, γ =0, n=300	RMSE	0.041	0.015	0.054	0.041	0.036	0.022	0.019	0.021
c=20%, γ =0, n=1000	RMSE	0.036	0.008	0.048	0.036	0.029	0.011	0.010	0.011
c=20%, γ =0.4, n=50	RMSE	0.163	0.161	0.207	0.170	0.162	0.240	0.174	0.159
c=20%, γ =0.4, n=300	RMSE	0.090	0.085	0.124	0.090	0.079	0.096	0.072	0.066
c=20%, γ =0.4, n=1000	RMSE	0.065	0.057	0.106	0.072	0.058	0.053	0.040	0.037
c=20%, γ =0.8, n=50	RMSE	0.266	0.277	0.365	0.257	0.221	0.686	0.282	0.223
c=20%, γ =0.8, n=300	RMSE	0.156	0.160	0.203	0.130	0.104	0.190	0.117	0.096
c=20%, γ =0.8, n=1000	RMSE	0.105	0.110	0.162	0.094	0.069	0.104	0.064	0.052
c=40%, γ =0, n=50	RMSE	0.146	0.051	0.233	0.161	0.127	0.084	0.069	0.072
c=40%, γ =0, n=300	RMSE	0.101	0.016	0.182	0.120	0.081	0.025	0.021	0.022
c=40%, γ =0, n=1000	RMSE	0.091	0.008	0.173	0.113	0.073	0.012	0.011	0.011
c=40%, γ =0.4, n=50	RMSE	0.205	0.175	0.357	0.242	0.203	0.561	0.219	0.178
c=40%, γ =0.4, n=300	RMSE	0.130	0.097	0.294	0.182	0.133	0.124	0.085	0.075
c=40%, γ =0.4, n=1000	RMSE	0.106	0.068	0.283	0.172	0.119	0.067	0.046	0.040
c=40%, γ =0.8, n=50	RMSE	0.291	0.303	0.510	0.323	0.256	12.730	0.725	0.251
c=40%, γ =0.8, n=300	RMSE	0.177	0.183	0.426	0.225	0.153	0.295	0.142	0.109
c=40%, γ =0.8, n=1000	RMSE	0.132	0.137	0.412	0.206	0.130	0.138	0.076	0.059

Table VIII: RMSE for $\beta 1$

DGP	Statistics	OLS	Tobit	QR 25%	QR 50%	QR 75%	CQR 25%	CQR50%	CQR 75%
c=0%, γ =0, n=50	RMSE	0.074	0.074	0.100	0.091	0.101	0.100	0.091	0.101
c=0%, γ =0, n=300	RMSE	0.029	0.029	0.040	0.036	0.039	0.040	0.036	0.039
c=0%, γ =0, n=1000	RMSE	0.016	0.016	0.022	0.020	0.021	0.022	0.020	0.021
c=0%, γ =0.4, n=50	RMSE	0.129	0.129	0.150	0.138	0.149	0.150	0.138	0.149
c=0%, γ =0.4, n=300	RMSE	0.054	0.054	0.058	0.054	0.058	0.058	0.054	0.058
c=0%, γ =0.4, n=1000	RMSE	0.029	0.029	0.031	0.029	0.032	0.031	0.029	0.032
c=0%, γ =0.8, n=50	RMSE	0.192	0.193	0.189	0.173	0.190	0.190	0.173	0.190
c=0%, γ =0.8, n=300	RMSE	0.081	0.081	0.073	0.067	0.072	0.073	0.067	0.072
c=0%, γ =0.8, n=1000	RMSE	0.045	0.045	0.040	0.036	0.039	0.040	0.036	0.039
c=20%, γ =0, n=50	RMSE	0.212	0.088	0.220	0.210	0.221	0.142	0.118	0.121
c=20%, γ =0, n=300	RMSE	0.195	0.034	0.184	0.183	0.194	0.054	0.047	0.047
c=20%, γ =0, n=1000	RMSE	0.192	0.019	0.179	0.179	0.190	0.029	0.025	0.026
c=20%, γ =0.4, n=50	RMSE	0.244	0.172	0.326	0.274	0.255	0.317	0.210	0.197
c=20%, γ =0.4, n=300	RMSE	0.215	0.091	0.294	0.245	0.225	0.106	0.080	0.076
c=20%, γ =0.4, n=1000	RMSE	0.209	0.071	0.289	0.239	0.219	0.056	0.043	0.041
c=20%, γ =0.8, n=50	RMSE	0.294	0.273	0.415	0.316	0.278	0.854	0.307	0.262
c=20%, γ =0.8, n=300	RMSE	0.240	0.176	0.383	0.279	0.234	0.162	0.109	0.096
c=20%, γ =0.8, n=1000	RMSE	0.233	0.153	0.379	0.275	0.230	0.086	0.058	0.052
c=40%, γ =0, n=50	RMSE	0.412	0.127	0.437	0.441	0.456	0.188	0.152	0.151
c=40%, γ =0, n=300	RMSE	0.398	0.044	0.399	0.415	0.444	0.067	0.057	0.056
c=40%, γ =0, n=1000	RMSE	0.396	0.022	0.394	0.412	0.442	0.035	0.031	0.031
c=40%, γ =0.4, n=50	RMSE	0.437	0.229	0.595	0.523	0.485	1.307	0.366	0.269
c=40%, γ =0.4, n=300	RMSE	0.413	0.131	0.573	0.501	0.468	0.169	0.116	0.101
c=40%, γ =0.4, n=1000	RMSE	0.409	0.110	0.571	0.498	0.466	0.088	0.061	0.055
c=40%, γ =0.8, n=50	RMSE	0.469	0.372	0.712	0.566	0.496	38.337	1.199	0.396
c=40%, γ =0.8, n=300	RMSE	0.433	0.257	0.706	0.550	0.477	0.396	0.171	0.132
c=40%, γ =0.8, n=1000	RMSE	0.427	0.241	0.705	0.548	0.474	0.154	0.088	0.072

Table IX: RMSE for $\beta 2$

4.D Applications of (C)QR models to CV studies

A structured literature search and a systematic review of the abstract and the full text of the two hundred or so existing publications yielded 20 applications of QR to CV studies.[7] In seven of them, however, the dependent variable is not WTP: in the health field, it is Body Mass Index (Dodd 2014), Time Trade Off (Trent et al. 2011), or health-related quality-of-life indicators (Tinelli et al. 2013; Seymour et al. 2010), while in the energy field, it is vehicle miles traveled (Su 2012) and electricity consumption (Yao et al. 2014), and in the cultural field, visits to museums(Meleddu et al. 2013). Overall, 13 applications actually use WTP as the dependent variable (see table X for a summary), and all find heterogeneity in the relationships between WTP and some of the explanatory variables. For binary QR, Belluzzo (2004) and Lavín et al. (2017) succeed in capturing heterogeneity that the logit model misses. For QR, O'Garra and Mourato (2006) observe differences in the determinants of WTP at the two tails of the WTP distribution and Viscusi et al. (2012) confirm the 'richer picture" of QR w.r.t. interval regression. In the only article that applies CQR, Krishnamurthy and Kriström (2016) find significant effects (in particular for income) on WTP, whereas Tobit-like models indicate non-significant effects.

4.E Hypothetical scenarios

A translation of the questions and the scenarios presented to respondents and relevant to this study is reproduced below. Sentences in italics are for the reader and were not read to respondents.

Introduction by Interviewer

"You are going to be the main actor in our fictitious scenarios. You will have to take the best decision regarding your housing. Only your opinion matters, there is no wrong or right answer. Not everyone is fully aware of the way the flood insurance system works, so we present it briefly. In France, every third-party liability insurance policy regarding fire or damage includes a mandatory contribution known as CatNat. To benefit from this type of compensation in the event of flood, the flood event must have been declared a

[7] The keywords used in the Scopus database and on Google Scholar were "contingent valuation" and "quantile regression", "stated preference" and "quantile regression", "discrete choice experiment" and "quantile regression" and "conjoint analysis" and "quantile regression". We exclude studies that only use the median in quantile regressions.

Author	Method	Elicitation format	Dependent variable	N
Belluzzo (2004)	Binary QR	Referendum	WTP for the improvement of water resources	1026
Lusk et al. (2006)	QR	Open ended	WTA to consume genetically modified chocolate chip cookies	346
O'Garra and Mourato (2006)	QR	Payment card	WTP for the air and noise pollution reductions associated with the introduction of hydrogen buses	531
Brummett et al. (2007)	QR	Payment card	WTP for irradiated mango	304
Last (2007)	QR	Payment card	WTP for municipal cultural provision	1062
Hanley et al. (2009)	QR	Payment card	WTP for an improvement in coastal water quality	800
Notaro and De Salvo (2010)	QR	Payment card	WTP for research expenditure and treatments to preserve cypresses	308
Viscusi et al. (2012)	QR	Bidding game	WTP for reducing the morbidity risks from drinking water	3585
Jackman and Lorde (2013)	QR	Open ended	WTP for digital products	390
Jerome et al. (2015)	QR	Bidding game	WTP for the continuation of a weight loss program	234
Krishnamurthy and Kriström (2016)	Censored QR	Open ended	WTP for a completely green residential electricity system	8229
Furno et al. (2016)	QR	Auctions	WTP for canned crushed tomatoes enriched with lycopene	190
Lavín et al. (2017)	Binary QR	Referendum	WTP for hydrological ecosystem services provided by forest	501

Note: QR: Quantile regression

Table X: Summary of the Applications of QR Models to CV Studies (N: Sample size)

'natural catastrophe' by joint ministerial decree and the goods (property and belongings) must be insured. Compensation will be subject to a €380 deductible. Personal injuries are not covered by the CatNat system. They are covered either by a personal insurance policy, or by the national government if a civil servant (administrative or elected) can be held responsible for the occurrence of the flood event."

Protective devices scenario (randomly proposed first to half the sample)

"Let us imagine that the CatNat insurance still covers the flood-related events. Your current insurance contract still covers all other types of events, and your premium remains unchanged. Imagine that the national government creates a Flood Management Fund to finance protective devices against flood. Building dikes, water retention ponds or improving rain water evacuation networks would reduce the height and speed of water and would completely eliminate the risk of flood in your commune. This work will only be realized if the population involved contributes to the Flood Management Fund. We would like to know how much maximum you would be willing to pay per year to this Fund."

Note to the interviewer: *If the respondent asks for details on the level of protection, the cost of the protective devices or the way they would be funded, please give the following answer:*

"This survey is part of a research project that involves several communes. What we are considering here is a fictitious situation, so that the exact way the protective devices would be implemented is not yet decided. When answering, however, imagine that everybody covered by this protective devices pays, like the household waste removal tax, for instance."

Insurance scenario (randomly proposed first to the other half sample)

"Let us imagine that the CatNat insurance no longer covers the flood-related events. Your current insurance contract still covers all other types of events, and your premium remains unchanged. Imagine that the national government creates a Flood Management Fund that is now the only flood-related damage compensation system. It allows you to be compensated in case of personal, property or material damage. You can freely choose to contribute or not to this Flood Management Fund, but if you do not contribute, you will not be compensated in case of flood-related damage. We would like to know how much maximum you would be willing to pay per year to this Fund."

Note to the interviewer: *If the respondent asks for details on the way the insurance would be implemented, please give the following answer (depending on the question):*

a) "This survey presents a fictitious situation that assumes that the insurance premium remains unchanged despite the fact that it no longer covers flood-related damage. There may be two reasons for this. First, it still covers all the other risks, including other natural hazard risks, which represent 95% of the compensation paid. Second, the CatNat system currently suffers from financial imbalance, because of the increase in natural hazard-related risks, so that the national government had to contribute as much as 50% of the premiums paid in recent years (Centre Européen de Prévention du Risque Inondation, 2013)."

b) "This survey presents a fictitious situation that assumes that the insurance premium remains unchanged in the event of flood. This means that the compensation will be subject to a €380 deductible, that an obsolescence coefficient is applicable, that the housing will be rebuilt as the original and that personal injuries will be covered under the same provisions as for personal insurance policies."

Note to the interviewer: *Repeat the following after each scenario*

"We remind you that you have previously declared that the probability of being flooded during the coming year is ...." *(remind the respondent of his/her previous answer to question L16-1).*

WTP Question 1. "Would you be willing to contribute to the Flood Management Fund to finance protective devices against flood / to be fully compensated in the event of flood ?" *(depending on the scenario).*

Note to the interviewer:
If the answer to question WTP Question 1 is "No", then ask for the reasons.
If the answer to WTP Question 1 is "Yes", then ask the following:

WTP Question 2. “How much maximum would you be willing to pay per year? To help you, here is a card with several amounts.”

Note to the interviewer: *[Present the circular payment card].*

“Do not forget that this money will be drawn from your household’s budget! You will therefore have less money at the end of the month for consumption or savings.”

Chapter 5

Testing anchoring on previous studies in meta-analyses: An application to VSL

Abstract: In this paper, I propose a formal test for a publication bias suggested by Viscusi and Masterman (2017). This bias arises from the selection of estimates based on their distance from the usual range in the literature. It implies that, in a given research area, there is a progressive concentration of estimates around this range as the literature grows. I test this hypothesis using OECD meta-data on VSL elicited via stated preferences. Using conditional variance regressions and quantile regressions, I am able to test for the effect of the number of prior studies on the dispersion of estimates, controlling for confounding factors. Consistently with the hypothesis, my results show a decrease in the dispersion of VSL estimates, which mostly happens for the lower part of the conditional distribution. These results imply that there is a systematic underestimation of the variance of mean estimates in meta-analyses of VSL. It might also affect meta-analyses in other research areas.

5.1 Introduction

Meta-analyses have been increasingly used in various areas during the past decades. This research technique allows to average key estimates from different studies to obtain a unique and more precise measure. It also enables identifying determinants of the heterogeneity in estimates between studies, which can be used to provide methodological guidelines. One of the main issue in meta-analyses is the selection of estimates, both by the researcher and the journals. If not accounted for, this selection creates a publication bias, which can lead to a misrepresentation of the aggregate estimate and inappropriate recommendations (Stanley et al. (2017)).

To my knowledge, two main types of publication bias have been identified so far. The first one is based on the significance of the results. This publication bias has been acknowledged for a long time and there have been many attempts to correct for it (Doucouliagos et al. (2012, 2014); Viscusi (2015)). The second type of publication bias comes from the preference towards estimates that are consistent with theory (see for example Card and Krueger (1995) on the effect of minimum wage or Doucouliagos et al. (2014) on positive income elasticities.

In two recent meta-analyses on the value of a statistical life (VSL), Viscusi (2015); Viscusi and Masterman (2017) suggest a third type of publication bias. This bias arises "because of the efforts by researchers to provide estimates in line with the previous literature". Besides researchers, "journal editors and reviewers likewise may be more likely to favor publication of result in the usual range". This hypothesis is based on their findings of a large difference in magnitude of publication bias between US and non-US studies. This bias would be due to non-US researchers anchoring their results on US studies. However, these authors do not provide a general way to test for this effect, so it has not been formally identified.

In this paper, I propose a test for a broader form of "publication anchoring bias". Instead of relying on the distinction between US v.s. non-US studies, I base my approach on the size of the literature at the time of publication. The hypothesis (denoted "Hypothesis 1") is the following. If researchers and reviewers actually judge the quality of an estimate based on its distance from the central tendency in the literature, they would prefer an estimate close to the central range to one far from it. So if a literature is well developed, there exists a range of values seen as acceptable and the selected estimates will tend to lie in this predetermined range. This selection pattern will then lead to a concentration of estimates around the central tendency in the literature. However, when a literature emerges there is no consensus on what the magnitude of the estimates should be, so it is not be possible to discriminate estimates based on this pattern. It is only when the body of works grows that a central tendency appears and that selection can happen. Thus, when the number of studies increases there should be a decrease in the dispersion of estimates. Figure 1 illustrates this pattern using simulated data. Estimates are represented in the y-axis and the number of prior studies in the x-axis. Black dots represent selected estimates, and red dots rejected estimates. When the literature is

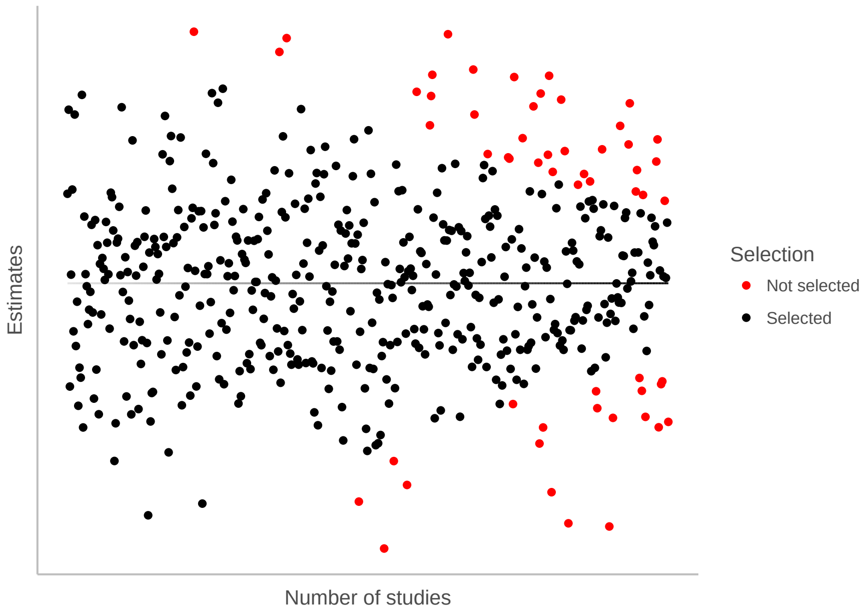


Figure 5.1: Estimates w.r.t. number of previous studies in the literature

scarce (i.e. small number of studies) there is no selection because of the uncertainty on what the appropriate range should be, but as the size of the literature increases, a central tendency emerges (the black line in the graph) and estimates start to be selected based on their distance from this central tendency. This selection becomes more strict as the central trend is more apparent.

Under hypothesis 1, if the selection is symmetric, the concentration of estimates towards the central tendency would not be strictly speaking a bias. The selection based on the distance from this central tendency would lead to an underestimation of the variance of the mean estimates. This would only affect the accuracy, not the mean estimate. The selection could be asymmetric though, i.e. there could be a selection more strict on large estimates than on small estimates. Such a pattern would lead to a bias in the mean estimate, in addition to an underestimation of the variance. Unlike in Viscusi (2015); Viscusi and Masterman (2017) I propose a general way to test for the anchoring of estimates on prior studies.

Given the importance of mortality effects in health-related (public) decision making, a large share of meta-analyses have been performed on the value of a statistical life (Ashenfelter and Greenstone (2004); Bellavance et al. (2009); Lindhjem et al. (2011); Robinson and Hammitt (2015)) and attempted to account for the publication bias. The hypothesis from Viscusi (2015); Viscusi and Masterman (2017) is also based on a VSL meta-analysis. I therefore will also apply the tests to meta-data on VSL.

In this paper, I propose two tests for the anchoring of estimates on prior studies. Using a meta-analysis on the VSL, I show that this publication selection process exists and that it affects the mean and the bias of the VSL estimates. Instead of using meta-data from revealed preference studies as in Viscusi and Masterman (2017), I use OECD meta-data (OECD (2012)) on stated preference studies. The stated preference literature on VSL is very different from the revealed preference literature.

I try to test whether Viscusi and Masterman (2017)'s hypothesis of the influence of prior U.S. studies holds for stated preference studies, and whether the anchoring bias affects differently US and non US studies.

The remainder of the paper proceeds as follows. Section 2 presents the design of the test, Section 3 describes the meta-data on the VSL, Section 4 presents and discusses the results and Section 5 concludes.

5.2 Methodology

5.2.1 Econometric modeling

In this section, I present a general methodology to detect the anchoring of estimates on prior studies. I propose two approaches to test for a decrease of the conditional dispersion of estimates when the number of studies increases.

The first approach is based on the Breush Pagan test for heteroskedasticity. It proceeds in two steps. In the first step I estimate the conditional mean of the estimates, for a set of covariates including the number of prior studies.[1]

$$y_i = \alpha_1 + \gamma_1 Nstudies_i + x_i' \beta_1 + \epsilon_i$$

Where y_i is the estimate of study i, x_i is a set of controls associated with study i, $Nstudies_i$ is the number of prior studies, ϵ_i is the error term. This model is estimated with OLS.

From this regression, I compute the residuals and take their squares, obtaining a measure of the conditional variance of the estimates. I can now perform a second regression of the squared of the residuals on the same set of covariates.

$$\hat{\epsilon_i}^2 = \alpha_2 + \gamma_2 Nstudies_i + x_i' \beta_2 + u_i$$

[1]In this section I call estimates the values of estimates in the meta-data, and coefficient the estimates of the econometric model.

γ_2 is the meta-estimate of the effect of the number of prior studies on the conditional variance of the estimates. Therefore I can test if γ_2 is negative to verify the hypothesis of a decrease of the variance as the number of studies increases.

Using this model, I can correct the unconditional variance for this publication bias. The relationship between unconditional variance and conditional variance is defined in the law of total variance: $V(y) = V(E(y|x)) + E(V(y|x))$. In the least squares framework, the first term is equal to the variance of the fitted values of the OLS model and the second term is the mean of the conditional variance, i.e. mean of the the squared residuals. So I correct the conditional variance by removing the effect of prior studies, that is substracting $\gamma_2 Nstudies_i$ to $\hat{\epsilon_i}^2$ for each observation i. Formally I compute the corrected squared residuals $\tilde{\epsilon_i}^2$ with:

$$\tilde{\epsilon_i}^2 = \hat{\epsilon_i}^2 - \hat{\gamma}_2 Nstudies_i$$

Then I obtain the corrected unconditional variance using the law of total variance.

The main issue with this approach based on variance is that it is sensitive to outliers, which might often be present in meta-analyses. Therefore I propose a second approach based on quantile regressions, which are less impacted by this problem. (Koenker (2005)).

Quantile regressions are mainly used to model the heteroskedasticity by looking at how the marginal effect of a variable varies along the conditional distribution of the dependent variable. If the marginal effects are the same along the conditional distribution, then the variance does not vary with the covariate, i.e. the variable is homoskedastic. However if the marginal effect increases/decreases, the variance increases/decreases with the covariate, the variable is then heteroskedastic (Koenker (2005)). Thus, I can check how the number of prior studies affects the variance of the estimates by looking at the coefficients associated with this covariate for a set of quantiles of the conditional distribution. The model is written:

$$Q_\tau(y_i|x_i, Nstudies_i) = \alpha_\tau + \gamma_\tau Nstudies_i + x_i'\beta_\tau + \epsilon_{\tau,i}$$

Where γ_τ is the meta-estimates of the effect of the number of prior studies on the quantile τ of the conditional distribution of estimates. Then, if the γ_τ decreases with the quantiles τ, it verifies the hypothesis of a concentration of estimates as the literature grows. Koenker and Bassett (1982) proposed a formal test of heteroskedasticity based on this idea. If the quantile regression lines are parallel, i.e. if the γ_τ are equal for all τ, then the conditional variance is constant. On the other hand, if the coefficient associated with a covariate increases/decreases along the conditional distribution, the conditional variance is increasing/decreasing with this covariate. One should note that this method does not allow to have a corrected variance, because there is no direct relationship between variance and interquantile ranges, unless making strong parametric assumptions.

5.2.2 Identification

In order to choose the right set of controls, one should know which factors could have an effect on the concentration of published estimates. Three important factors should be relevant for most meta-data. The first one is convergence in methods: as the literature grows, a consensus often emerges on which methods are the most appropriate. Therefore there might be an homogenization of the methods, which could lead to a concentration of published estimates. If a study uses a method that is considered unsuitable by the majority, it should be rightly disregarded. Therefore, one should control for the set of methods, which includes data collection methods and statistical methods. These variables need to be reported in the meta-analysis if one wants to test for this type of publication selection.

The second factor is the increase in sample sizes through time. It is well known that larger datasets have become available in the past years. An increase in sample sizes implies an increase in the precision of the estimates. The estimates of the most recent studies should then be more concentrated.

The last aspect is a positive correlation between the number of prior studies and the number of estimates in a study. It might be the case that old studies reported fewer estimates than more recent ones. A straightforward explanation could be the increase of econometric models available. If this correlation exists, it might lead to a spurious effect between the number of prior studies and the variability of estimates, because estimates from the same study come from the same sample or population and thus should be less varying than estimates from different studies. It is also common practice to give less weights to estimates that come from a study with a lot of reported estimates. I will thus use the inverse of estimates per studies as a weight in the econometric models.

These three factors should be the main causes of concern for the validity of the test. However, some other aspects specific to the field of study and the nature of the estimates may have an impact as well. Regarding the application on the VSL, I will discuss which other controls should be included in the next section.

5.3 Meta-data on VSL

I use a meta-data on stated preference surveys for mortality risk valuation. Most of the VSL meta-study mentioned in the introduction are based on revealed preferences. Both approaches have strengths and drawbacks. Stated preference studies only elicit hypothetical preferences because they are based on surveys, unlike revealed preference which are based on market behavior. However revealed preference studies can only estimate VSL regarding a specific risk for a specific population. Besides, it is possible to obtain negative VSL using revealed preference studies, although they are not economically meaningful Doucouliagos et al. (2012), while stated preference, because the WTP for a risk reduction are (almost) always positive, the implied VSL are also always positive. This means that one can't expect the funnel plot to be symmetric, even without publication bias.

The dataset was collected by the OECD (OECD (2012)). All the surveys elicit VSL from the stated WTP for a reduction in mortality risk related to environmental, health and transport policies.[2] The first study published is dated from 1973 and the most recent ones are from 2009. Figure 2 represents the VSL estimates as a function of the number of prior studies.[3] One can indeed see that there is a convergence in the estimates. The first forty estimates have a very large variance. Then there is a sharp decrease in the variance, followed by another smaller decrease after eighty studies. As mentioned before, many factors may explain this evolution, the most important ones are likely to be the change in methodological guidelines and the increase in sample size. The estimates published in journals are represented in red and other studies in black. Interestingly, the variance seems to decrease faster for published studies than for unpublished ones. This could be a sign that the homogenization of the methods is faster for published studies than for unpublished ones. Another possibility is a selection of studies based on the distance from the central tendency.

For each study, the OECD meta-data contains information about the methodology, some characteristics of the affected population and the characteristics of the risk (Lindhjem et al. (2011)).

Variables characterizing the survey design and the statistical method are summarized in Table 1. Four variables characterize the survey design. The first one is the elicitation method, i.e. the mechanism by which the WTP was elicited. The majority of studies

[2] VSL are converted in USD 2005 PPP adjusted based on the AIC (actual individual consumption).

[3] This variable is created using the year of Publication. Among studies published in the same year, I could not identify which one came first.

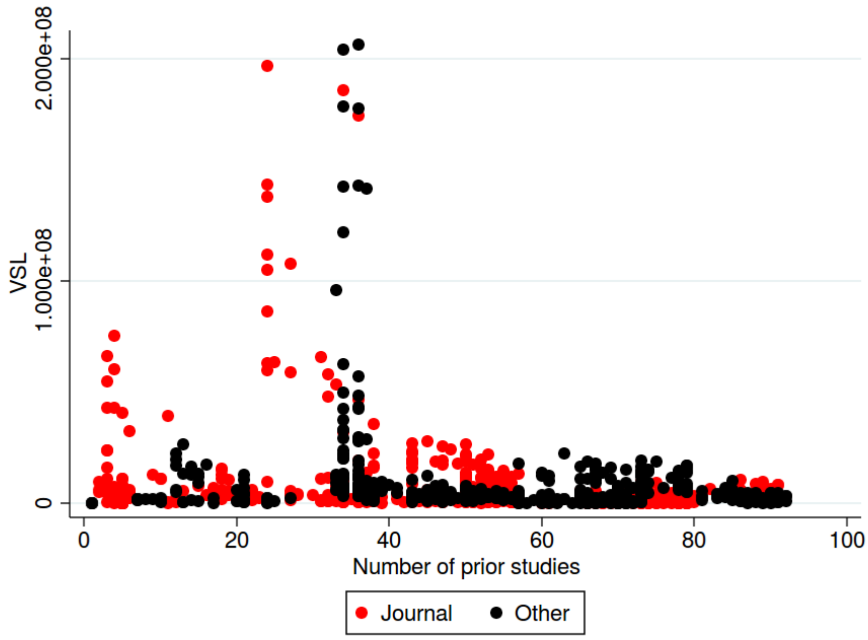


Figure 5.2: VSL estimates and quantity of studies in the literature

are contingent valuation studies, using dichotomous choice, as it was prescribed by the NOAA panel. Open-ended questions and payment cards are also widely used. Conjoint analyses, which also seem to include discrete choice experiments, represent a minor part of the studies. The second aspect is the survey approach. Most surveys are conducted face-to-face or are self-administrated (this include emails). Web-based surveys also represent a significant part of the studies. Regarding the payment vehicle, most studies use the price of a product. Other frequently used options are cost of living and a tax. The last aspect is whether the risk change was explained in a way that makes it more intuitive to the respondents, as we see most surveys did.

Regarding the statistical method, I use a classification that depends on the flexibility of the specification. I distinguish between fully parametric models (mostly models estimated by maximum likelihood), semi-parametric models (which do not need distributional assumptions for consistent estimations, mostly OLS) and non-parametric (descriptive statistics and Turnbull model). Within these categories, the choice of models is mostly determined by the nature of the dependent variables (continuous, interval, binary, etc.), which depends on the elicitation methods. Therefore, more precision on the methods would be redundant with the elicitation method information.

Other relevant variables are summarized in Table 2. The average VSL is 8.7 million, the 1009 estimates range from 5 thousand to 200 million.

Some variables characterize the magnitude and nature of the risk change. As noted

Elicitation Methods	CV - cards	CV - dicho	CV - open	Conjoint analysis	Other	Total
Freq	145	448	229	118	226	1166
Survey Approach	Face-to-face	Self-administrated	Telephone	Web-based	Other	Total
Freq	482	403	65	148	54	1152
Payment Vehicle	Cost of living	Donation	Price of product	Tax	Other	Total
Freq	146	27	730	171	75	1149
Risk Explanation	Yes	No				Total
Freq	935	186				1121
Statistical Method	Parametric	Semi-parametric	Non-Parametric			Total
Freq	815	37	256			1,108

Table I: Summary of methods, N=1167

Variable	Description	Mean	Std. Dev.	N
vslaic	Sample mean VSL in PPP-adjusted USD 2005	8728300	21686387	1009
riskchange	Change in mortality risk on an annual basis per 1000	0.003	0.016	757
cancerrisk	=1 if reference to cancer risk in survey; 0 if not	0.147	0.354	1167
public	=1 if public good; 0 if private	0.274	0.446	1167
envir	=1 if environment-related risk change; 0 if not	0.264	0.441	1167
health	=1 if health-related risk change; 0 if not	0.422	0.494	1167
traffic	=1 if traffic-related risk change; 0 if not	0.314	0.465	1167
samplesize	Number of valid response used to estimate VSL	846.1	2099.2	848
gdpcapita	Country GDP/capita in PPP-adjusted USD 2005	26616	12215	1155
gdpgrowth	Country growth in GDP at the collection year	3.434	3.032	1165
lifeexp	Country life expectancy at the collection year	76.542	3.514	1155
num_est	Number of VSL estimates reported in the study	33.456	28.429	1167
num_stud	Number of studies in the literature at year of publication	49.458	24.284	1167

Table II: Summary statistics, N=1167

by Lindhjem et al. (2011), the magnitude of the change is a significant determinant of the VSL. Besides, there might be an evolution of the risk change through time, because surveys conducted in more developed countries tend to value smaller risk changes. There is a reasonable balance between studies on environment, health and transportation policies, with a majority of health policies. Whether there is a cancer risk may also be important, since it could be perceived as more dreadful.

Regarding the sample size, the meta-data report both the total sample size and the subsample size (of the sample that was actually used to compute the VSL). I choose to include the subsample size, because it is this variable that affects the precision of the estimates. As expected, it is positively correlated with the year of data collection (corr = 0.26, p-val < 0.0001). As suspected in the previous section, I find that the number of estimates per study is increasing with the number of prior studies (correlation = 0.066, p-value = 0.024): recent studies tend to report more estimates. Therefore, I use the inverse of the number of estimates per studies as a weight in the econometric models.

Some other factors might influence how VSL estimates evolve as the literature grows. In their meta-analysis on VSL from hedonic wage method, Bellavance et al. (2009) observe an increase of VSL through time. They advance several hypotheses to explain this phenomenon. While the first two are related to the hedonic wage methodology, the third is more general: an increase in the life expectancy through the year could lead to a higher VSL. Thus I control for life expectancy in the country at the collection year. Another plausible hypothesis that they do not mention is the income growth since 1970, which could also affect the VSL. Besides, if the estimates are collected in different countries, a convergence in GDP between countries may also explain a convergence in VSL. Therefore I control for the growth rate of GDP per capita of the country during the collection year of the survey and the level of GDP in 2005. Finally, I take into account the characteristics that are found to have an impact in Lindhjem et al. (2011) (the type of policy, the magnitude of risk change, whether there is a risk of cancer and whether the good is public). Following Lindhjem et al. (2011), I use a log-log model: all continuous variables are transformed by taking their log. Using a log-form reduces the number of outliers, which is an important issue in VSL data.

5.4 Results

As explained in Section 2, I use two approach to identify the anchoring effect. I first use conditional variance regressions, then quantile regressions.

5.4.1 Conditional variance

Results are reported in Table 3. The first two columns report the unweighted mean and variance models. As in Lindhjem et al. (2011), whether the risk was explained with a visual tool has a significant positive effect on the mean VSL and the magnitude of risk change has a significant negative effect on the mean. Life expectation has a positive effect on the mean, but not GDP (both in growth or in level). Unlike what is expected, the number of prior studies has a significant negative effect on the mean VSL. This is unexpected because there is no apparent reason why the number of prior studies should affect the mean. Two reasons could explain this effect. The first is an asymmetric selection: if high estimates (compared to the literature) are selected in a different way than the low estimates, then beside a scale shift, there will be a location shift in the VSL distribution as the literature grows. For instance if high estimates are more strictly selected than low estimates, then the mean VSL will increase.

Another explanation is the aforementioned positive correlation between the number of prior studies and the number of estimates in a study. Since estimates from the same study should be less varying than estimates from different studies because they come from the same sample or population, this might lead to a spurious correlation.

Regarding the results for the variance, GDP growth has a negative and significant effect. Variance is also negatively affected by the magnitude of the risk. The effect of the number of studies is negative and significant, as expected from my hypothesis.

In order to account for the potential bias due to estimates coming from the same studies, I present the results of a weighted least square model, using the inverse of the number of estimates in the study as weights. Results tend to be smaller in magnitude. One can see that the effect of the number of studies on the mean VSL is no longer statistically significant, suggesting that not using weight was causing the issue discussed above. In the variance regression however it stays significant, although with a lower magnitudes. Overall the results are in accordance with my hypothesis.

Using the law of total variance, I compute the corrected standard deviations (SD). The ratio of corrected SD on non-corrected SD is equal to 1.297 for the unweighted model and 1.632 for the weighted models. Note that this computations are based on untransformed VSL (not the log transformed model).

These results confirm the necessity to account for this anchoring effect, because it can lead to severe overestimation of the precision of VSL estimates.

In the next section, I investigate further with a more robust statistical method.

5.4.2 Quantile regressions

In this section I apply the robust test for heteroskedasticity proposed by Koenker and Bassett (1982). This test rely on the property that if the coefficient of a covariate in a quantile regression increases/decreases along the conditional distribution, the conditional variance is increasing/decreasing with this covariate. Because it is based on quantile regressions, it is possible to condition for other covariates. Therefore, I use the same specification as the previous model. I first provide visual insights by reporting graphs of the coefficients of the log of the number of prior studies as a function of conditional VSL quantiles in Figure 3.

In accordance with the conditional variance results, the curve tends to decrease with VSL conditional quantiles, which means that the dispersion decreases as the number of studies increases. More specifically there is a clear decrease below the 0.6 quantile, with relatively narrow 95% confidence intervals. However, coefficients above the 0.6 quantile are very imprecise though, there is a widening of the confidence interval, and no clear downward or upward trend. From this graph, one can suspect an asymmetric selection, since only the VSL dispersion on the lower half of the conditional distribution seems to

VARIABLES	(1) Mean	(2) Variance	(3) Mean (WLS)	(4) Variance (WLS)
log(num_stud)	-0.996***	-0.695**	-0.414	-0.371*
	(0.234)	(0.307)	(0.252)	(0.193)
log(gdp)	-0.363	-0.147	0.342	-0.253
	(0.466)	(0.521)	(0.385)	(0.380)
log(riskchange)	-0.517***	-0.220**	-0.425***	-0.0703
	(0.0790)	(0.106)	(0.0907)	(0.0975)
cancerrisk	0.766	0.882	1.623*	0.770
	(0.648)	(0.688)	(0.826)	(0.501)
public	-1.124	0.550	-0.737	-0.364
	(0.752)	(1.130)	(0.455)	(0.385)
envir	0.597	-0.212	-0.141	0.556
	(0.766)	(1.192)	(0.586)	(0.455)
traffic	-0.458	-1.249	-0.0586	-0.528
	(0.559)	(0.871)	(0.394)	(0.325)
noexplan	1.089***	0.247	0.698**	0.0925
	(0.328)	(0.411)	(0.287)	(0.240)
gdpgrowth	-0.0639	-0.158***	-0.0216	-0.00172
	(0.104)	(0.0472)	(0.0850)	(0.130)
log(lifeexp)	10.88*	-13.35*	9.635**	-0.900
	(6.461)	(7.524)	(4.626)	(4.821)
samplesize	8.18e-05	3.28e-05	0.000106	-0.000185
	(0.000143)	(0.000169)	(0.000232)	(0.000186)
Constant	-30.61	60.98**	-33.65*	8.001
	(26.16)	(27.87)	(17.72)	(21.91)
Observations	500	500	500	500
R-squared	0.748	0.360	0.812	0.249

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table III: OLS Results

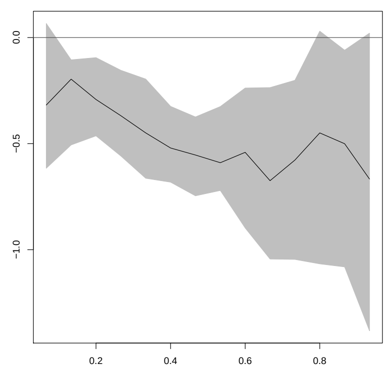


Figure 5.3: Weighted Quantile regressions, graphical results

Difference in coefficients (p-value)	0.5-0.1	0.75-0.25	0.9-0.5
Effect of the number all prior studies	-0.349 (0.058)	-0.166 (0.612)	-0.032 (0.942)

Table IV: Results of the Wald test

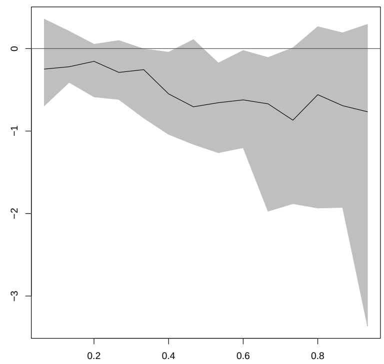


Figure 5.4: Quantile regressions, effect of prior US studies

be affected by the number of prior studies. This would imply an upward bias on the estimates. However, regarding the effect of the number of prior estimates on location of VSL (the coefficient for the median quantile, at 0.5 on Figure 3), it is negative, which contradicts the results on the variance. This result might come from an omitted variable bias in the mean regression. Therefore it is not possible to conclude on the existence and direction of a “location bias”.

Nevertheless, it is still possible to test formally for the presence of anchoring effect on the variance. I use a Wald test of equality of coefficients between quantiles (Koenker and Bassett (1982)). Due to the asymmetry observed in the quantile regressions coefficients, I test for equality of coefficients between the 0.25 quantile and the 0.75 quantile, but also between the 0.1 quantile and the 0.5 quantile and between the 0.5 quantile and the 0.9 quantile. Results are reported in Table IV. The difference of coefficients between the 0.25 quantile and the 0.75 quantile is negative as we saw in the graph, but it is not significantly different from zero. Nor is the difference between the 0.5 quantile and the 0.9 quantile. However, there is a significant difference at 10% between the 0.1 quantile and the 0.5 quantile. As a consequences we reject the null of no anchoring effect for the lower part of the conditional distribution, the higher part seems unaffected.

5.4.3 Specificity of US studies

In this section I perform an analysis of two Viscusi and Masterman (2017)’s hypotheses that US studies serve as anchor for non-US studies, and that only non-US studies are

Difference in coefficients (p-value)	0.5-0.1	0.75-0.25	0.9-0.5
Effect of the number of prior US studies	-0.589 (0.114)	-0.514 (0.174)	-0.024 (0.970)
Effect on Non-US studies	-0.378 (0.353)	-0.270 (0.551)	-0.015 (0.975)

Table V: Results of the Wald test, US vs non-US distinction

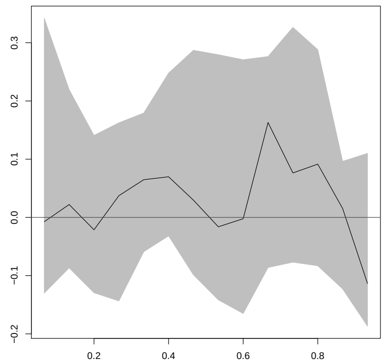


Figure 5.5: Quantile regressions, interaction effect on US studies

affected by anchoring and US studies are not. In Figure 5.4, I report the coefficients of two regressions: in the first regression, the size of the literature is based only on US studies, and in the second regression I add the interaction effect between the number of prior studies and a dummy equal to one when the study is from the US.

Figure 5.4 shows the coefficient associated with the model where I use the number of US studies as measure of the literature size. The effect seems much weaker than when using all studies. This is confirmed by the tests reported in the first line of Table V: I do not reject the null of equality of coefficients for all three pairs of quantiles. Therefore there is no evidence of a particular anchoring on US studies, which is not in line with the result from Viscusi and Masterman (2017).

Then I look at the effect of the number of prior studies on non-US studies by using an interaction term between number of studies and a US dummy. Looking at the second line of Table V, one can see that the baseline effect of the number of prior studies on non-US estimates is not significant for all quantiles. The interaction term is not statistically different from zero for all quantiles (Figure 5.5), which suggests that US studies and international studies are affected with the same intensity by anchoring bias in stated preference studies.

Note that the distinction between US and non US studies might only be a proxy for another characteristic of the study, such as its quality. Therefore one should be cautious when interpreting these results.

5.5 Conclusion

In this paper, I test the hypothesis of a selection of estimates based on their distance from the usual range in the literature. I show that this selection happens for estimates in the lower part of the conditional distribution. No statistically significant effect is found for the upper part of the conditional distribution. These results imply that there is a systematic underestimation of the variance of mean estimates in meta-analyses of VSL. The correct standard deviation of VSL estimates would be around 60% larger than the non-corrected standard deviation. I am not able to conclude for an anchoring effect on the mean estimates.

Unlike what is suspected in Viscusi and Masterman (2017), I find no significant anchoring effect when only US studies are considered as anchors. Besides I do not find that non-US studies are more anchored than US studies. Caution should be used when interpreting these last results because they might be non-causal.

It could be interesting to focus on the "best estimates" in the meta-data, rather than the whole set of estimates. Indeed, regarding publication bias in the VSL literature, Viscusi (2018) finds significantly a starker selection process (in terms of "traditional publication bias") for best-estimates than for the whole dataset. The type of selection might also differ.

Although in theory the two tests proposed in this paper are applicable to other fields, some criteria must be satisfied. For instance, because the identification strategy is based on a variation trough time, the literature should cover a reasonably long period. Also, as for all meta-analysis, the key estimate must be the same throughout all the period, otherwise a change in the reported values could be attributed to a change in definition. Finally, as it was done in the present application, a careful reflection must be undertaken to account for all the possible confounding factor in order to properly identify the effect of the number of previous studies.

Bibliography

Ashenfelter, O. and Greenstone, M. (2004). Estimating the Value of a Statistical Life: The Importance of Omitted Variables and Publication Bias. *The American Economic Review*, 94(2):454–460.

Bellavance, F., Dionne, G., and Lebeau, M. (2009). The value of a statistical life: A meta-analysis with a mixed effects regression model. *Journal of Health Economics*, 28(2):444–464.

Card, D. and Krueger, A. B. (1995). Time-Series Minimum-Wage Studies: A Meta-analysis. *The American Economic Review*, 85(2):238–243.

Doucouliagos, C., Stanley, T. D., and Giles, M. (2012). Are estimates of the value of a statistical life exaggerated? *Journal of Health Economics*, 31(1):197–206.

Doucouliagos, H., Stanley, T., and Viscusi, W. K. (2014). Publication selection and the income elasticity of the value of a statistical life. *Journal of Health Economics*, 33:67–75.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press, Cambridge ; New York.

Koenker, R. and Bassett, G. (1982). Robust Tests for Heteroscedasticity Based on Regression Quantiles. *Econometrica*, 50(1):43–61.

Lindhjem, H., Navrud, S., Braathen, N. A., and Biausque, V. (2011). Valuing mortality risk reductions from environmental, transport, and health policies: a global meta-analysis of stated preference studies. *Risk Analysis*, 31(9):1381–1407.

OECD (2012). Mortality Risk Valuation in Environment, Health and Transport Policies - OECD.

Robinson, L. A. and Hammitt, J. K. (2015). Research Synthesis and the Value per Statistical Life: Research Synthesis and the Value per Statistical Life. *Risk Analysis*, 35(6):1086–1100.

Stanley, T. D., Doucouliagos, H., and Ioannidis, J. P. A. (2017). Finding the power to reduce publication bias. *Statistics in Medicine*, 36(10):1580–1598.

Viscusi, W. K. (2015). The role of publication selection bias in estimates of the value of a statistical life. *American Journal of Health Economics*, 1(1):27–52.

Viscusi, W. K. (2018). Best Estimate Selection Bias in the Value of a Statistical Life. *Journal of Benefit-Cost Analysis*, 9(02):205–246.

Viscusi, W. K. and Masterman, C. (2017). Anchoring biases in international estimates of the value of a statistical life. *Journal of Risk and Uncertainty*, 54(2):103–128.

Part III

Aggregating preferences

Chapter 6

On the plutocracy of cost benefit analysis[1]

Abstract: Revealed and stated preference techniques are widely used to assess individuals' non-market preferences, in particular in cost benefit analysis (CBA). First, however, individuals have to satisfy subsistence needs through market good consumption, which affects their budget constraint. The impact of subsistence needs on preference elicitation for, and pricing of, a non-market good or service have not been extensively explored. In this paper, we first provide a methodological framework showing how both preferences and pricing depend on level of subsistence needs and income. We then quantify these impacts by comparing this framework with the standard framework from a theoretical, a numerical and an empirical perspective. In particular, we consider individual preferences for the non-market good that differ according to level of income. Our findings confirm the relevance of accounting for subsistence needs when relying on non-market vdraluation methods, especially in CBA.

6.1 Introduction

Cost-benefit analysis (CBA) has increasingly been used in all economic sectors to support public and private decision-making since the last century. The past 50 years have seen the introduction of non-market components into CBA, with considerations like improved recreation, visual amenities, odours, noise, loss of biodiversity, psychological aspects, or the valuation of premature deaths. Because there is no marketplace to set economic prices for these components, they require specific methods of valuation based on stated

[1]This paper is a joint work with Olivier Chanel.

or revealed preferences. These methods elicit - the former directly through surveys, the latter indirectly through data collection - individuals' preferences for a given non-market good or service. They then derive willingness to pay (WTP) for the corresponding welfare change and, after statistical treatments, this WTP is used by private and public decision-makers in CBA.

This process appears very democratic, directly feeding the preferences of the whole population into decision-making Pearce et al. (2006). Nevertheless, it hides two methodological issues when preferences are elicited through willingness to exchange money (or a composite market good) for non-market goods or service provision. First, there is the budget constraint effect, already empirically investigated and for which solutions have been proposed to accurately reflect the preferences of low-income individuals: either contributions in kind or in work (Brouwer et al. 2009; Abramson et al. 2011, Hossack and An 2015) or an "equity-adjusted WTP" (Breffle et al. 2015). Second, subsistence needs limit the realm of possibility when expressing preferences, and therefore WTP, which affects the poorest more than the richest. Thus, subsistence needs exacerbate the problem of inequity in CBA through their effect on the marginal utility of income.

The standard criterion that justifies the CBA is the Kaldor Hicks criterion: if the net benefits from a project is positive, then the overall welfare is increased by the project, and the winners can compensate the losers with a monetary transfer (Johansson and Kriström (2018)). Therefore in most applied cost–benefit analyses, individual WTP are simply summed up, and the overall gain from the project is measured by the net present value (net benefits with a time discounting). However, this transfer between winners and losers is not made in practice, which raises the question of the equity of CBA. According to Broberg (2010) "A disadvantage of the Kaldor-Hicks criterion is that it ignores the possible importance of distributional issues. It can lead to decisions which favour the rich at the expense of the poor because the rich have a greater ability to pay in support of any given strength of preference."

Theoretically, simply summing up (i.e. applying equal weights) means that the marginal social utility of income is the same for everyone. This is only the case if the fiscal system is considered "optimal", which is of course never the case in the real world. This issue calls for the use of equity weights, which for each individual, should be equal to the individual's marginal utility of income times the weight of this individual in the social welfare function (Mäler, 1974; Kanninen and Kriström, 1992). The question of the impact of subsistence needs on this equity issue, and how it can be reflected in equity weighting of WTP has been overlooked in the CBA literature.

This paper investigates how subsistence needs distort the WTP-based expression of preferences, insidiously turning CBA into a plutocratic process. Surprisingly, to our knowledge, their consequences on preference and WTP elicitations have not previously been explored theoretically. We propose to fill this gap by comparing the standard framework with one that accounts for subsistence needs. First, we show how preference elicitation for a non-market good or service is affected. Second, we show how the non-market implicit (or shadow) price is under-estimated w.r.t. the standard framework. Finally, we provide both numerical and empirical illustrations of this under-estimation.

We find that the preferences of individuals tend towards less substitutability between market and non-market goods when subsistence needs are accounted for. This effect decreases as income increases or subsistence needs decrease. We also find that the WTP of the richest individuals is less impacted by subsistence needs, and exceeds the WTP of the poorest even when they show similar preferences for the non-market good. Finally, where preferences are heterogeneous preferences, it is the richest whose views are more likely to impact decision-making based on CBA.

Our principal contribution to the literature is twofold.
First, our analysis contributes to the methodological literature on preference revelation. We propose a framework for assessing the impact of subsistence needs on both individual preferences and WTP elicitation for non-market goods. This impact depends on "true" unobserved preferences, income and level of subsistence needs. It should be noted that this contribution differs from assessments of the impacts of income distribution on CBA, a long-acknowledged issue (see Ebert 1986; or Adler 2016 for a recent overview). Because WTP is the ratio of marginal utility for the good valued over marginal utility of income, heterogeneous preferences over the income distribution increase the likelihood of non-democratic outcomes in CBA. This problem can be overcome by assigning distributional weights. Nor does our contribution belong to the debate on the democratic / plutocratic antagonism in consumer price indexes for households weighted according to their total expenditures (see Deaton and Zaidi 2002; Ley 2002), or in health when priorities depend on income or age (Olsen and Donaldson 1998). In a recent paper, Baumgärtner et al. (2017a) study the impact of income inequality on mean WTP and show that, depending

on whether the non-market good is a substitute or a complement for the manufactured goods, increased inequality can either decrease or increase mean WTP.

Second, our analysis contributes to the empirical literature on the "budget saturation effect". To date, most stated and revealed preference studies consider a positive and significant relationship between income and WTP as indicating that individuals are behaving as they would on actual marketplaces, where level of income drives level of consumption. However, this interpretation can be questioned. Breffle et al. (2015) found that "when more than one program is necessary in order to provide a complete cleanup or improvements are required at one or more sites, then lower-income individuals simply run out of expressed WTP even though problems that extend beyond that WTP still impair them and their uses". Along the same lines, Smith (2005) was interested in the role of the budget constraint in the scale-sensitivity of WTP. He found "an increasing 'relevance' of the budget constraint as the value of the good (relative to income) increases: as the benefit increases, WTP for that benefit rises and consequently the budget constraint becomes an increasingly significant determinant of WTP". Budget saturation also implies that, due to the narrowing of their set of choices, the WTP of the poorest individuals will show little or no variation, irrespective of the strength of their preferences. Therefore, an implicit ranking of different programs based on their WTP would misleadingly suggest indifference between programs for these individuals. Olsen et al. (2005) compared the implicit ranking inferred from the ordinal differences in WTP values with the explicit ranking of the same programs. They showed that a large share of individuals stating indifference through their WTP stated a clear ranking based on the preference elicitation exercise, but unfortunately the study did not link this to level of income.

The remainder of the paper proceeds as follows. Section 2 describes the relevant models and Section 3 analyses the impacts of introducing subsistence needs. Section 4 presents numerical illustrations and an empirical analysis is presented in Section 5. We finally discuss and conclude in Section 6.

6.2 Models

We present the standard model and the model with subsistence needs before comparing them in terms of elicitation of preferences and shadow prices.

6.2.1 Standard model

Consider an individual whose preference relation is continuous, monotonic and convex. Let us consider that this preference relation is represented by a two-good utility function $u(x, q)$, where x represents the quantity of a composite market good and q the quantity of a non-market good. $u(x, q)$ is assumed to be twice continuously differentiable, strictly increasing and strictly quasi concave in x and q, with $x, q \geq 0$.

The preferences of the individual for the non-market good depend on the composite market good's substitutability for this non-market good, which is measured via the elasticity of substitution $\sigma(x, q)$. We choose Hicks (1932)'s original definition of the elasticity of substitution for two factors:

$$\sigma(x, q) = \frac{\partial ln(\frac{x}{q})}{\partial ln(MRS)}$$

where MRS is the marginal rate of substitution, i.e. $\frac{\partial u(.)}{\partial q} / \frac{\partial u(.)}{\partial x}$.

We are also interested in the relation between the WTP for the non-market good, this elasticity of substitution and the individual's income. We follow the main trend in the literature on non-market valuation (Hanemann 1991; Lankford 1988; Ebert 2003) by defining the marginal WTP as the shadow price for q:

$$\pi(p, q, y) = \frac{\partial V(p, q, y)/\partial q}{\partial V(p, q, y)/\partial y} \tag{6.1}$$

where $V(p, q, y)$ is the indirect utility function which is obtained from the following maximization problem:

$$\underset{x}{\text{Max}}\, u(x, q) \;\text{ subject to }\; px = y \text{ and } q \text{ fixed} \tag{6.2}$$

In this problem, the individual pays for the quantity of the non-market good q at the shadow price π and has the shadow income $y + \pi q$. The individual's income is compensated so that all his real income is spent on market goods. We then have the following equivalence with the inverse Hicksian demand: $\pi(p, q, y) = \pi(p, q, v(p, q, y))$. It is clear from Eq. (6.1) that the shadow price of the non-market good only depends on the individual's preferences represented by the parameters characterising V(.), the quantity of non-market goods q and the income y.

6.2.2 Model with subsistence needs

Let us consider that individuals face minimum subsistence needs defined by the level of consumption x_s of the composite good. We adapt Baumgärtner et al. (2017b)'s and Drupp (2016)'s approaches, which address a subsistence requirement in terms of environmental services, by introducing these subsistence needs into u as follows:

$$\begin{cases} u(x,q) = u^l(x) & \text{for} \quad x \leq x_s \\ u(x,q) = u^h(x,q) & \text{else} \end{cases} \tag{6.3}$$

We assume that u^l is strictly increasing in x and u^h has the same properties as u (see above). We also adopt Baumgärtner et al. (2017b)'s assumption that individuals always prefer to be in the domain where subsistence needs are satisfied, i.e.:

$$\inf_{x>x_s,\ q\geq 0} u^h(x,q) > \sup_{x_s \geq x \geq 0} u^l(x) \tag{6.4}$$

Hence, below the minimum subsistence needs, nobody is willing to trade the composite good for the non-market good. The elasticity of substitution between x and q is then obviously set to $\sigma(x,q) \equiv 0$ when $x_s \geq x$ because q does not enter u^l, and π is consequently also set to 0. We will no longer consider this case in what follows. Above the minimum subsistence needs, $\sigma(x,q)$ and π are defined as in the standard model but based on the $u^h(x,q)$ function that accounts for these subsistence needs.

6.2.3 Defining the two models under a CES utility function

In order to be able to determine explicit relationships between the standard and the subsistence needs frameworks, we need to start from a functional form as flexible as possible regarding preferences over x and q. A relevant, easy to interpret, tractable and well-known specification of $u(x,q)$ is the Constant Elasticity of Substitution (CES) function first proposed by Arrow et al. (1961):

$$u(x,q) = [\alpha x^\theta + (1-\alpha) q^\theta]^{\frac{1}{\theta}} \text{ for } \theta \in]-\infty; 1] \ ; 0 < \alpha < 1 \tag{6.5}$$

Note that the elasticity of substitution $\sigma(x,q) \equiv \frac{1}{1-\theta}$, and that the CES function covers a range from perfect complement ($\theta \to -\infty$; $\sigma(x,q) \to 0$) to perfect substitute

($\theta \to 1$; $\sigma(x,q) \to \infty$), as well as the Cobb Douglas function ($\theta \to 0$; $\sigma(x,q) \to 1$).[2] α represents the preference for the market good x relative to the preference for the non-market good q.

At equilibrium, the shadow price for q in the standard CES model is (from Eq. (6.1) and Eq. (6.5)):

$$\pi = \frac{(1-\alpha)q^{\theta-1}}{(\alpha/p)(\frac{y}{p})^{\theta-1}} \tag{6.6}$$

Following the modification proposed by Geary (1950) and Stone (1954) to account for minimum levels of consumption and known as Stone-Geary function, and its extension to CES proposed by Baumgärtner et al. (2017b) or Drupp (2016), we define the extended CES above the minimum subsistence level as follows:

$$u^h(x,q) = [\alpha(x-x_s)^\theta + (1-\alpha)q^\theta]^{\frac{1}{\theta}} \quad \text{for} \quad x > x_s; \quad \theta \in]-\infty; 1]; \ 0 < \alpha < 1 \tag{6.7}$$

When $x > x_s > 0$, the elasticity function of the CES is no longer $\sigma(x,q) \equiv \frac{1}{1-\theta}$ but (Baumgärtner et al. 2017b; proposition 1):

$$\sigma_s(x,q) \equiv \frac{1}{1-\theta}[1 - F(x_s, x, q; \alpha, \theta)] \tag{6.8}$$

$$\text{where} \quad F(.) = \frac{(1-\alpha)\frac{x_s}{x}}{\alpha[\frac{x-x_s}{q}]^\theta + (1-\alpha)}$$

The shadow price for q in the CES model with subsistence needs is (from Eq.(6.1) and Eq.(6.7)):

$$\pi_s = \frac{(1-\alpha)q^{\theta-1}}{(\alpha/p)(\frac{y}{p} - x_s)^{\theta-1}} \tag{6.9}$$

6.3 Impacts of the introduction of subsistence needs

6.3.1 Impact on the elicitation of preferences

Baumgärtner et al. (2017b) proved that $F(.) > 0$ for all x and q in Eq. (6.8). Hence, the elasticity of substitution is always lower with minimum subsistence needs than without, which induces a shift towards complementarity in the relationship between the composite market good and the non-market good. Whatever their true unobserved preferences, individuals are less inclined to trade market goods for a non-market good when they face subsistence needs, hence expressing lower preference for the non-market good.

[2]Indeed, $u(x,q) \to x^\alpha q^{(1-\alpha)}$ *when* $\theta \to 0$.

Baumgärtner et al. (2017b) also proved that $F(.) \to 0$ when $x_s \to 0$ (absence of subsistence needs) and/or $x \to \infty$ (increase in consumption of the composite market good). This means that the magnitude of misrepresentation of preference for the non-market good decreases with income. Moreover, they proved that the elasticity of substitution $\sigma_s(x, q)$ increases with income, although non-monotonically for negative θ. This implies that the preferences of the poorest are more inelastic w.r.t. the composite market good.

Finally, we can determine how the misrepresentation evolves depending on the preference for the market good, α. The derivative of F(.) w.r.t. α is:

$$\frac{\partial F(.)}{\partial \alpha} = \frac{\partial}{\partial \alpha} \frac{(1-\alpha)\frac{x_s}{x}}{\alpha[\frac{x-x_s}{q}]^\theta + (1-\alpha)} = -\frac{x_s[\frac{x-x_s}{q}]^\theta}{x\left(\left([\frac{x-x_s}{q}]^\theta - 1\right)\alpha + 1\right)^2}$$

For $x > x_s$, and for all $\theta < 1$, it is negative since the numerator and the denominator are always positive. This means that the shift towards complementarity monotonically increases with the preference for the non-market good. *Ceteris paribus*, the higher the preferences for the non-market good, the less likely the trade-off of market goods for the non-market good in the model with subsistence needs (compared to the standard model). This magnifies the magnitude of the under-estimation of preferences towards the non-market good.

Overall, the poorer an individual, the higher the under-estimation of his/er true unobserved preferences for the non-market good.

6.3.2 Impact on the relationship between preferences and shadow prices

We examine the derivative of π with respect to θ. From Eq.(6.6), we have:

$$\frac{\partial \pi}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{(1-\alpha)}{(\alpha/p)} \left(\frac{q}{y/p}\right)^{\theta-1} = \frac{(1-\alpha)}{(\alpha/p)} \frac{\partial}{\partial \theta} exp\left(log\left(\frac{q}{y/p}\right)\right)^{\theta-1}$$

$$= \frac{(1-\alpha)}{(\alpha/p)} exp\left(log\left((\theta-1)\frac{q}{y/p}\right)\right) log\left(\frac{q}{y/p}\right)$$

$$= \frac{(1-\alpha)}{(\alpha/p)} \left(\frac{q}{y/p}\right)^{\theta-1} log\left(\frac{q}{y/p}\right)$$

Because the first two terms are always positive, the sign of the expression is determined by the third term. From Eq. (6.9), we find the derivative of π_s with respect to θ

to be similar, except that y/p is replaced by $(y/p - x_s)$ in the expression above. For π (resp. π_s), it is positive when $q > y/p$ (resp. $q > y/p - x_s$) and negative when $q < y/p$ (resp. $q < y/p - x_s$). Consequently, it depends on the relative quantity of market and non-market goods. An increase in substitutability will thus have an effect on the shadow price of the non-market good that depends on this relative quantity, but for given y and p, will be negative for smaller values of q in presence of subsistence needs.

Then, we examine the derivative of π with respect to α:

$$\frac{\partial \pi}{\partial \alpha} = \frac{\partial}{\partial \alpha} \frac{(1-\alpha)}{(\alpha/p)} \left(\frac{q}{y/p} \right)^{\theta - 1} = - \frac{p}{(\alpha^2)} \left(\frac{q}{y/p} \right)^{\theta - 1}$$

As above, the derivative of π_s with respect to θ is obtained by replacing y/p by $(y/p - x_s)$. Both derivatives are always negative, as expected: for a given y and p, the lower the preferences for the market good x (i.e. α), the higher the shadow price for the non-market good. Consequently, the effect of α on π is smaller in presence of subsistence needs, *ceteris paribus*.

6.3.3 Impact on shadow prices

We are interested in the spread between π_s (defined in Eq. (6.6)) and π (defined in Eq. (6.9)). It is always strictly negative for $p, q > 0$, and null only for $x_s = 0$. This means that the true shadow price of the non-market good is under-estimated when subsistence needs are ignored. Moreover, *ceteris paribus*, the spread is increasing with income when preferences express complementarity ($\theta > 0$), decreasing when preferences express substitutability ($\theta < 0$) and constant for $\theta = 0$. However the ratio π_s/π is always decreasing with income, even for negative θ. Moreover, it can be shown that this ratio is independent of α, p, q.

We can show that the use of distributional weights (see Adler 2016; or Fleurbaey and Abi-Rafeh 2016) based on the marginal value of the individual's income, cannot properly correct the shadow prices to account for the effect of subsistence needs. The marginal utility of income is equal to:

$$\frac{\partial v(y, q, x_s)}{\partial y} = \frac{\partial [\alpha(\frac{y}{p} - x_s)^\theta + (1-\alpha) q^\theta]^{\frac{1}{\theta}}}{\partial y}$$

$$= [\alpha(\frac{y}{p} - x_s)^\theta + (1-\alpha) q^\theta]^{\frac{1-\theta}{\theta}} \frac{\alpha}{p} (\frac{y}{p} - x_s)^{\theta - 1}$$

If we weight the shadow prices in the model with subsistence needs (Eq. (6.9)) by the marginal utility of income, we obtain:

$$\pi_{w_s} = [\alpha(\frac{y}{p} - x_s)^\theta + (1-\alpha)q^\theta]^{\frac{1-\theta}{\theta}}(1-\alpha)(q)^{\theta-1}$$

The resulting weighted shadow price π_{w_s} still depends on the level of subsistence needs x_s, which means that accounting for the marginal utility of income might not fully correct for under-estimation. This is because the subsistence needs not only affect the marginal utility of income, but also the marginal utility of the non-market good.

Theoretically, it would be possible to apply a "social weight", denoted SW, and obtain the same shadow price that would be obtained without subsistence needs but after weighting by the marginal utility of income and a utilitarian social welfare function. This social weight would be equal to:[3]

$$SW = \frac{[\alpha(\frac{y}{p})^\theta + (1-\alpha)q^\theta]^{\frac{1-\theta}{\theta}}}{[\alpha(\frac{y}{p} - x_s)^\theta + (1-\alpha)q^\theta]^{\frac{1-\theta}{\theta}}}$$

The shadow price weighted by the marginal utility of income and by SW would then be equal to:

$$\pi_{w_s,SW} = [\alpha(\frac{y}{p})^\theta + (1-\alpha)q^\theta]^{\frac{1-\theta}{\theta}}(1-\alpha)(q)^{\theta-1}$$

In practice, the SW is both individual and non-market good-specific, and would be almost impossible to elicit.

Overall, regarding decision-making based on CBA, the under-estimation of shadow prices of any non-market good assessed via shadow prices - derived from revealed or stated preference data - affects the assessment of benefits. It not only decreases the overall desirability of a project at a given cost, but also has consequences that differ depending on whether the preferences for the non-market good in the population (measured by $\sigma(x,q)$ and α in Eq. (6.5)) are homogeneous or heterogeneous, depending on income. This will be explored in the next section.

[3]The implied social welfare function is defined by the primitive of this social weight with respect to π.

6.4 Numerical illustrations

We provide numerical illustrations to show how the introduction of subsistence needs impacts both the substitutability and the shadow prices of non-market goods. Since we consider that an individual spends all his/her income on market goods (see maximization problem (6.2)), we set $x \equiv y/p$ and $p \equiv 1$ in the following. Thus, the composite market good x can be seen as the numéraire and represents the income of the individual.

6.4.1 Elasticity of substitution

We first look at the way the introduction of subsistence needs affects the elasticity of substitution σ_s, which is known to be no longer constant. From Eq. (6.8), we represent in Figure 6.1 the relationship between the elasticity of substitution and income expressed in terms of $\frac{x}{x_s}$, for $\theta \geq 0$ and $\theta < 0$.[4]

We see that elasticity of substitution σ_s is always lower than in the standard case, but converges towards the standard case value $\sigma \equiv \frac{1}{1-\theta}$ as income increases. Three points are worthy of note.

First, when $\theta \geq 0$, the convergence towards σ is monotonic as income increases. However, we observe that despite the fact that x and q are substitute goods, σ_s is lower than 1 when subsistence needs represent a large share of income, hence exhibiting complementarity. This means that subsistence needs push preferences toward complementarity.

Second, when $\theta < 0$, σ_s always exhibits complementarity, but the convergence is no longer monotonic. When income is just above subsistence needs, complementarity increases before converging towards σ when income grows.

Third, for low incomes, the ordering of the elasticity of substitution measured by σ_s does not match that measured by σ. This can be formally proven by scrutinizing the derivative of σ_s w.r.t. σ (details upon request) but is clearly seen from Figure 6.1. Close to the level of subsistence needs, for θ respectively equal to -0.5, 0 and 0.5 for instance, σ decreases (2, 1 and .6667) whereas σ_s increases (.2, .5 and .6). This means that in presence of subsistence needs and for very low incomes, preferences for substitutability ($\theta \geq 0$) tend more towards complementarity than preferences for complementarity ($\theta < 0$), making them even more inelastic w.r.t. the composite market good.

[4]The other parameters are set to q=1, α=0.5, x_s=2.

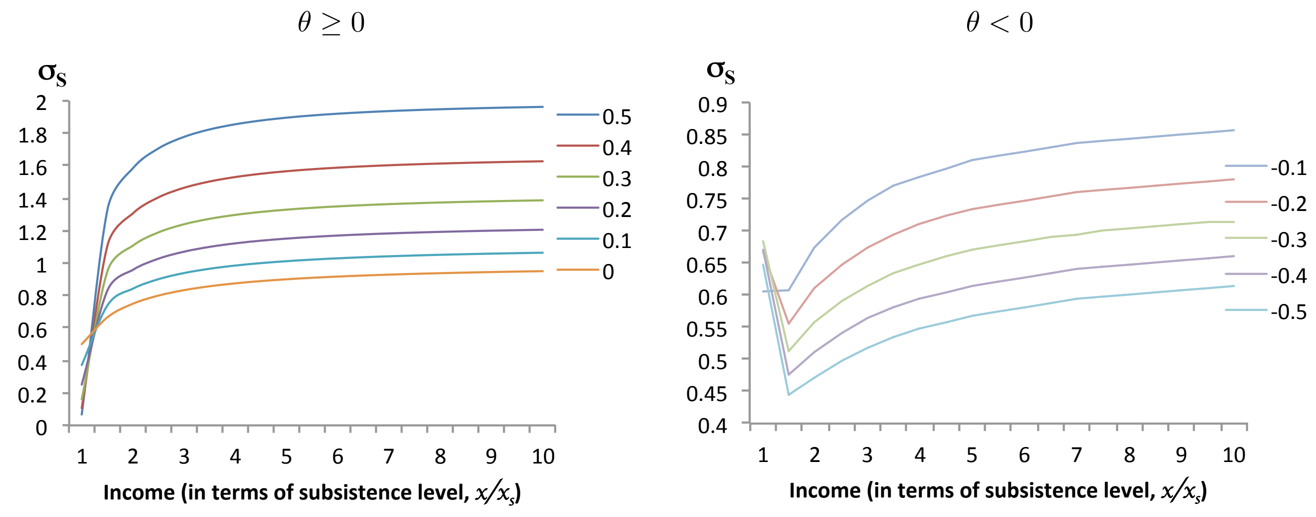


Figure 6.1: σ_s vs. Income for various values of θ

6.4.2 Shadow Price

We are interested in how introducing subsistence needs impacts the shadow price of the non-market good. Consequently, based on Eq. (6.6) and (6.9), we consider how the ratio π_s/π evolves when income changes. The use of a ratio removes the price level effect: the closer to 1, the smaller the under-estimation due to the presence of subsistence needs. We consider first preferences that are homogeneous over individuals (i.e. same θ and α) and then heterogeneous preferences depending on income.

Homogenous preferences

If preferences are homogeneous over the population, ignoring the minimum subsistence level will lead to under-estimating the benefits measured with shadow prices, which decreases the overall desirability of a project for a given cost. Figure 6.2 illustrates the magnitude of the under-estimation by computing the ratio for different θ and income combinations for $x_s = 2$ (the ratio does not depend on p, q nor α). We see that the lower the income, the smaller the ratio for any given value of θ, which is due to the fact that π_s tends towards zero when income tends towards x_s. For an income seven times larger than x_s, for instance, the ratio is about 0.8 (i.e. 20% under-estimation) when θ exhibits complementarity ($\theta = -0.5$), but about 0.93 (i.e. 7% under-estimation) for substitutability ($\theta = 0.5$). Note that, based on the actual distribution of income in the population, we can compute for any θ the "subsistence needs-corrected" aggregated shadow price to properly assess non-market benefits in CBA (see an application in section 5).

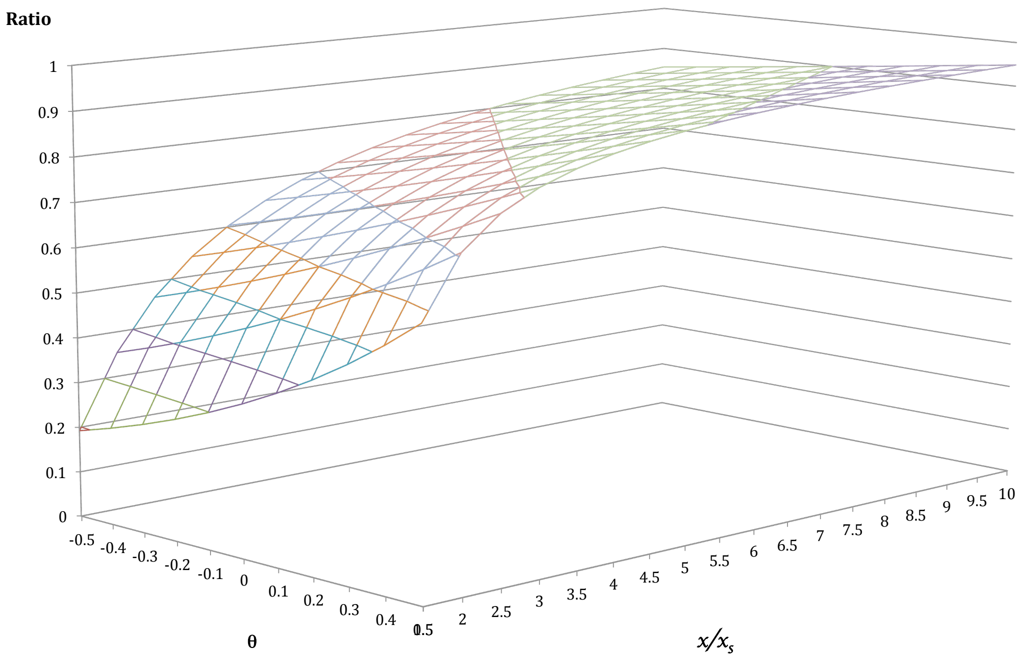


Figure 6.2: Ratio between shadow prices accounting for and not accounting for subsistence needs (by θ and relative income)

To illustrate the fact that accounting for the marginal utility of income does not fully correct for under-estimation, we consider the ratio of the weighted shadow price with subsistence needs to the shadow price without (π_{w_s}/π). Figure 6.3 shows that, although the ratio converges faster to 1 than in the unweighted case (see Figure 6.2), it is not always equal to 1 and is smaller for low incomes.

Heterogeneous preferences

If preferences are heterogeneous (i.e. income-dependent) regarding the non-market good, the overall desirability of a project is still downwardly biased, but the relative desirability is also biased in favour of the preferences of the richest fraction of the population. This is because under-estimation of the preferences for the non-market good is greater for the poorest fraction of the population than for the richest.

As an illustration, imagine (for the sake of simplicity) a bi-modal income distribution: one half of the population has a low income (1.5 times the subsistence needs), and the other half has a high income (10 times the subsistence needs). Let us consider that the preferences of each of the two subpopulations for non-market good are measured

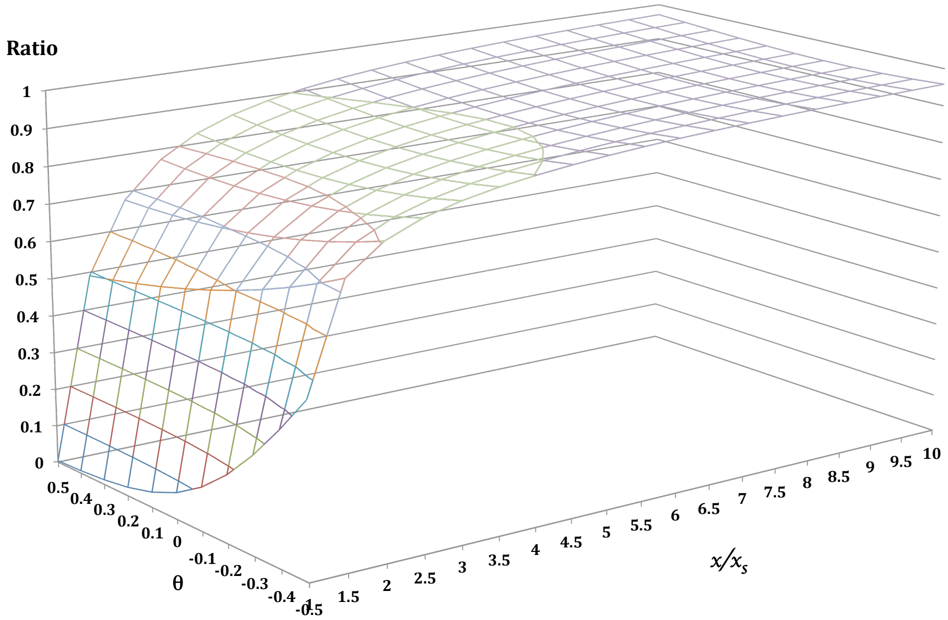


Figure 6.3: Ratio between weighted shadow prices accounting for and not accounting for subsistence needs (by θ and relative income)

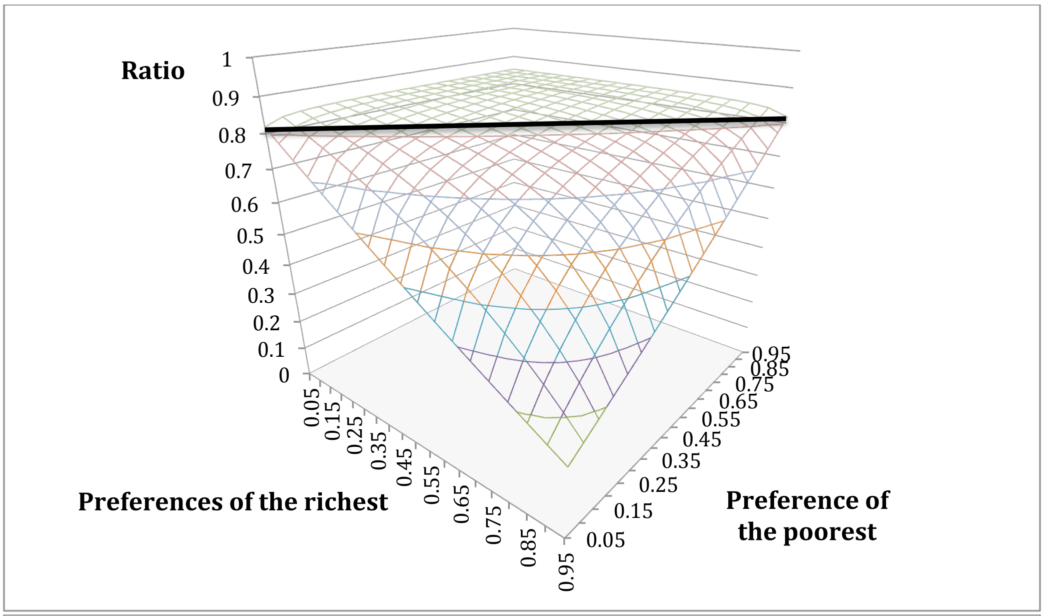
by α, varying by stepsize .05, from .05 (strong preference for the non-market good) to .95 (strong preference for the market good). For each combination of preferences, we compute the ratio of the average elicited shadow price with minimum subsistence needs to the average true unobserved shadow price, i.e. $E(\pi_s)/E(\pi)$. This tells us how far the valuation of non-market benefits in a CBA based on elicited shadow prices would be from the true unobserved shadow price. The closer to one the ratio, the truer the representation of the non-market preferences of the population. We compute this ratio for three values of elasticity of substitution ($\theta = 0.5, 0, -0.5$), and give the results in Figures 6.4.

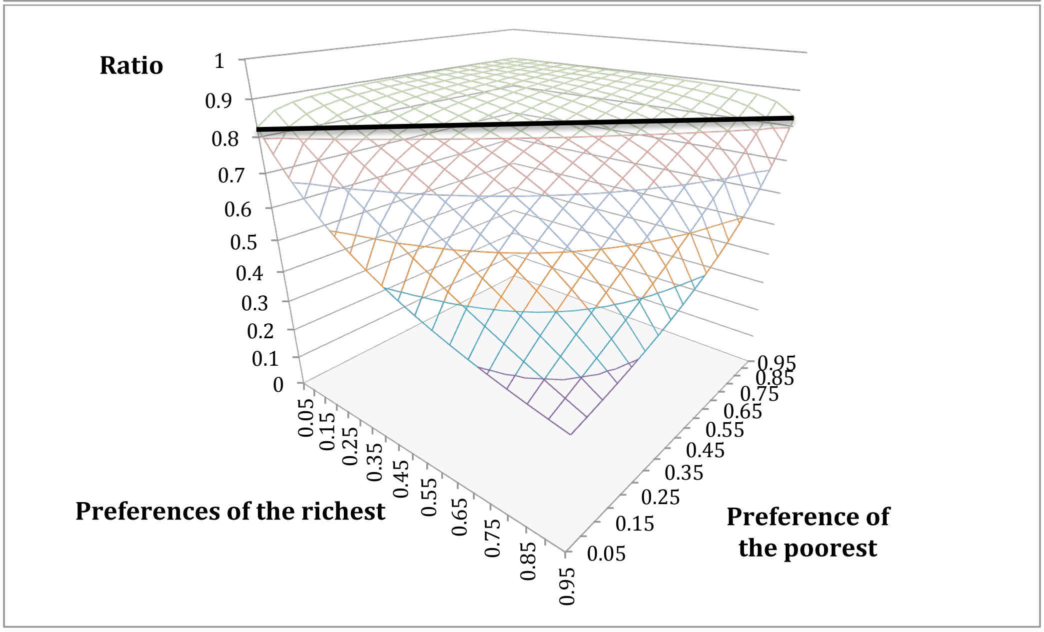
Whatever the substitutability value, we have the following results. For homogeneous preferences (represented by the horizontal black line segment on the three figures), the ratio is clearly constant and shows a spread lower than 15% w.r.t. the average true unobserved shadow price. Provided the preferences of the richest are more non-market oriented than those of the poorest, the spread is lower than in the homogeneous case (background of each figure), hence favouring the preferences of the richest. Provided the preferences of the richest are less non-market oriented than those of the poorest, the spread is higher than in the homogeneous case (foreground of each figure), hence again favouring the preferences of the richest. Overall, whatever the preferences of the richest, they are always better represented in CBA than those of the poorest, which shows a plutocratic bias. In particular, the non-market preferences of the poorest appear never to be properly accounted for unless shared by the richest, whereas the non-market preferences of the richest appear always to be better accounted for, their shadow prices not only being higher but also less under-estimated.

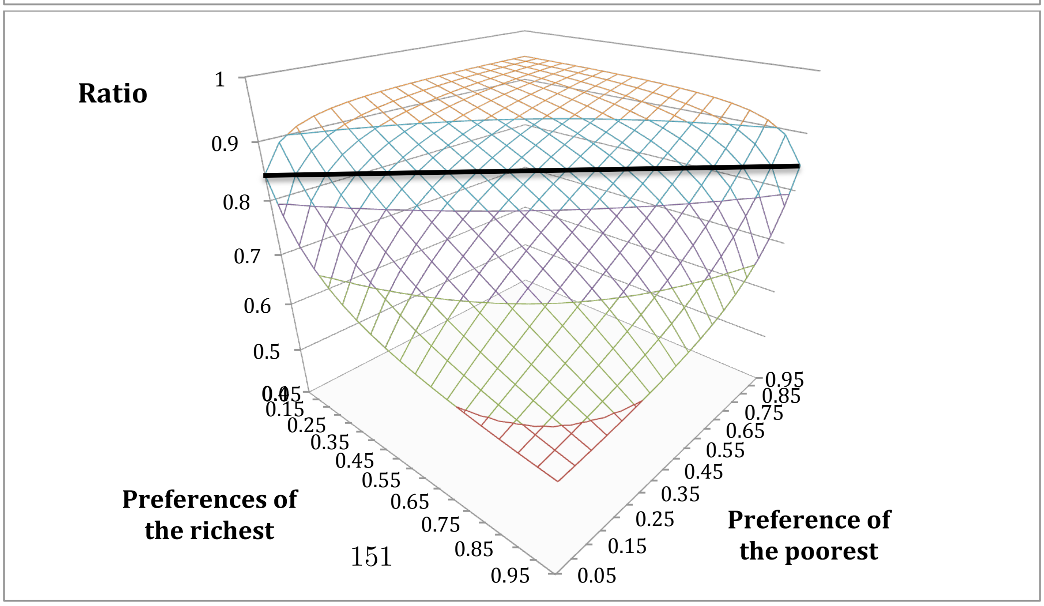
6.5 Empirical analysis on French income data

We explore the magnitude of the distortion of WTP-based expression of preferences when subsistence needs are accounted for. We rely both on French income distribution data and on empirical studies eliciting elasticities of substitution for various non-market goods. This allows us to scrutinise the extent to which CBA might be affected by the subsistence needs issue.

Regarding the distribution of income, we first need to set a value for subsistence needs x_s. This is different from absolute monetary poverty, defined by the World Bank based on a minimum number of calories (costing about US$ 1.9 in 2015). It is also different from relative monetary poverty, which takes into account the distribution of income in a given society. Organisation for Economic Co-operation and Development (OECD) (and Institut National de la Statistique et des Etudes Economiques in France) consider for instance that the share of the population with an income below the 50^{th} percentile (median) is below the poverty threshold, whereas Eurostat considers the threshold to be at the 60^{th} percentile. We are actually interested in the minimum amount required to live in a given country, including minimum expenditure on food, water, energy, housing, clothes, transportation, etc. A French survey estimates this to be about € 600 per month for one adult in 2016 (Carrefour des Solidarités 2011). This amount is slightly lower than the median standard of living for a person considered poor (€ 705 per month in 2015, Argouarc'h and Cazenave-Lacrouts 2017) and slightly higher than the active solidarity revenue (RSA) paid by the French government to individuals with no resources (€ 545 per month in 2017). We therefore adopt it as a reasonable benchmark. Based on this subsistence income per individual of x_s= € 600, we use the French annual income distribution by individual (Institut National de la Statistique et des Etudes Economiques


Ratio
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85 0.95
Preferences of the richest
0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85 0.95
Preference of the poorest


Ratio
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85 0.95
Preferences of the richest
0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85 0.95
Preference of the poorest


Ratio
1
0.9
0.8
0.7
0.6
0.5
0.4
0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85 0.95
Preferences of the richest
0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85 0.95
Preference of the poorest

2017) to express the French income distribution over the population in terms of x/x_s values.

Regarding estimates of the substitutability of market goods for non-market goods, Drupp (2016) presents various estimations of θ for non-market services around the world (air or water quality improvements, forest or marine services, landscape or recreational amenities, biodiversity, etc.). He obtains a mean empirical estimate of $\theta = 0.57$ with a mean empirical error range of $(0.28 - 0.86)$. This means that, on average, individuals exhibit substitutability between market goods and non-market goods or services.

We assume that preferences regarding the non-market good are homogeneous (i.e. not income-dependent) due to lack of relevant data on the actual distribution of preferences w.r.t. income in the population. Figure 6.5 represents the ratio of the mean shadow price accounting for income distribution with minimum subsistence needs over the shadow price of the standard model, for various values of θ. Remember from section 5 that this ratio does not depend on p, q or α. In particular, when $\theta = 0.57$ (see dotted line segment), the difference is about 20%. This means that, from a CBA perspective, the non-market benefits of a project (assessed through survey-based shadow prices) will be under-estimated by 20% on average compared to the "true" (unobserved) benefits when the effect of subsistence needs is not accounted for. Consequently, because of the distortion of the benefit-cost test (benefits need to be 20% greater than costs for a project to pass), this rules out a fraction of socially desirable projects. In addition, heterogeneous preferences depending on income would add a plutocratic bias, as shown in Figures 6.4, favouring the preferences of the richest. We cannot currently assess the extent of this bias due to lack of specific data on the population's distribution of preferences for the non-market good w.r.t. income.

6.6 Discussion

Our findings have important implications for both theoretical and empirical research in economics, as well as major policy implications.

From a theoretical perspective, these findings contribute to the equity vs. equality debate. Should we increase the WTP elicited by the poorest when preferences for a non-market good are known to be income-dependent? By doing so, we are favouring

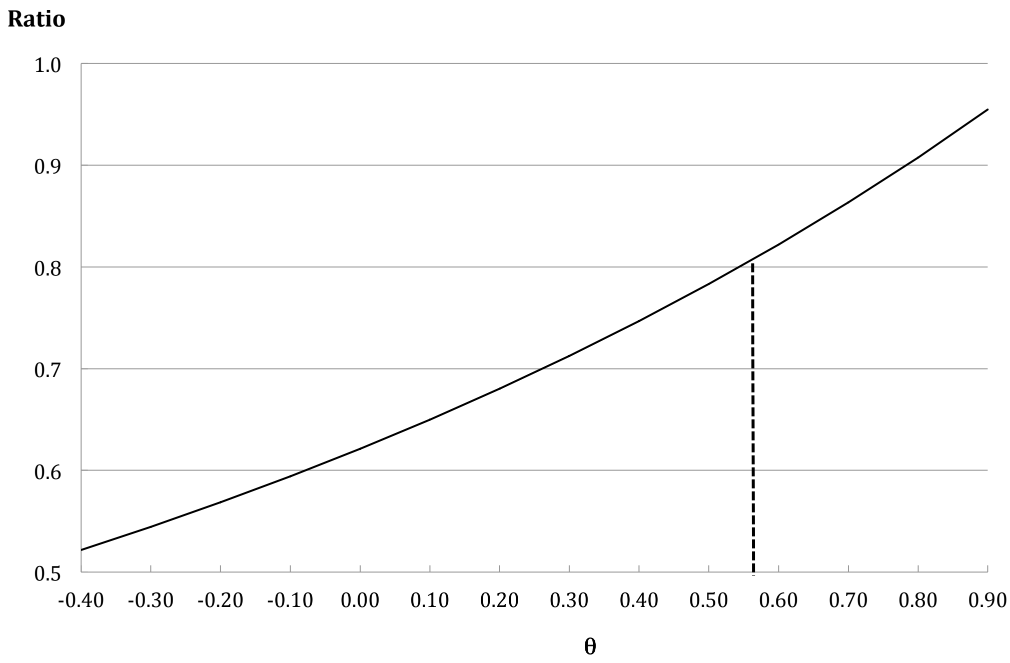


Figure 6.5: Mean ratio between shadow prices accounting and not accounting for subsistence needs, based on French income distribution, by θ

equity (an as-fair-as-possible representation of preferences) over equality (a common representation of preferences). The introduction of subsistence needs may also be worth considering when studying the scope / scale effect in contingent valuation. Findings that WTP is not proportional to the scope or the scale of an improvement proposed in a survey may partly be due to the constraint that subsistence needs impose on WTP. This constraint is not only more binding for the poorest individuals than for the richest, but is also heightened when the scale / scope valuation is large.

From an empirical perspective, our findings challenge the use of stated and revealed preference techniques in CBA involving preferences for a non-market good and measured through WTP. When preferences are income-independent, the overall desirability of a project is under-estimated but the average preference of the population is accounted for. When preferences are income-dependent, decision-makers should be aware that the preferences of the poorest individuals may be under-represented and those of the richest over-represented. Using other methods to account for preferences may help identify the likelihood of such a distorsion. Such methods include using rankings among alternatives instead of WTP, or applying normalised scenarios that elicit preferences under a given (hypothetical) income assumed equal for everyone.

In terms of policy implications, our findings raise important issues. During its one-century-long history, CBA has been used by national governments (via their various agencies), supra-national organisations and private firms to assess the effectiveness of policies and prioritise them. Use of CBA gradually extends to all economic sectors (Swenson 2015): beginning with navigation in the 1900s, the CBA covered agricultural and land issues during the New Deal, then was extended to public urban and transportation infrastructures after World War II, to social, educational and health issues in the 1960s, to occupational and environmental issues in the 1970s. CBA was used to assess the value of regulation / deregulation and of central government interventions in various economic sectors during the 1980s and 1990s, to help compute various public profitability / efficiency ratios in the 2000s and from then on has been widely used in all sectors to support public and private decision-making. Once non-market values are involved in a CBA, we need to be on the lookout for any potential plutocratic bias.

Bibliography

Abramson, A., Becker, N., Garb, Y., and Lazarovitch, N. (2011). Willingness to pay, borrow, and work for rural water service improvements in developing countries. *Water Resources Research*, 47(11):n/a–n/a. W11512.

Adler, M. D. (2016). Benefit–Cost Analysis and Distributional Weights: An Overview. *Review of Environmental Economics and Policy*, 10(2):264–285.

Argouarc'h, J. and Cazenave-Lacrouts, M.-C. (2017). Les niveaux de vie en 2015. *INSEE Première*, 1665(September).

Arrow, K. J., Chenery, H. B., Minhas, B. S., and Solow, R. M. (1961). Capital-labor substitution and economic efficiency. *The Review of Economics and Statistics*, 43(3):225–250.

Baumgärtner, S., Drupp, M. A., Meya, J. N., Munz, J. M., and Quaas, M. F. (2017a). Income inequality and willingness to pay for public environmental goods. *Journal of Environmental Economics and Management*.

Baumgärtner, S., Drupp, M. A., and Quaas, M. F. (2017b). Subsistence, Substitutability and Sustainability in Consumption. *Environmental and Resource Economics*, 67(1):47–66.

Breffle, W. S., Eiswerth, M. E., Muralidharan, D., and Thornton, J. (2015). Understanding how income influences willingness to pay for joint programs: A more equitable value measure for the less wealthy. *Ecological Economics*, 109:17–25.

Broberg, T. (2010). Income Treatment Effects in Contingent Valuation: The Case of the Swedish Predator Policy. *Environmental and Resource Economics*, 46(1):1–17.

Brouwer, R., Akter, S., Brander, L., and Haque, E. (2009). Economic valuation of flood risk exposure and reduction in a severely flood prone developing country. *Environment and Development Economics*, 14(3):397?417.

Carrefour des Solidarités (2011). Pour un Revenu Minimum de Survie.

Deaton, A. and Zaidi, S. (2002). *Guidelines for Constructing Consumption Aggregates for Welfare Analysis*. Washington DC: World Bank.

Drupp, M. A. (2016). Limits to substitution between ecosystem services and manufactured goods and implications for social discounting. *Environmental and Resource Economics*, pages 1–24.

Ebert, U. (1986). Equity and distribution in cost-benefit analysis. *Journal of Economics*, 46(1):67–78.

Ebert, U. (2003). Environmental Goods and the Distribution of Income. *Environmental and Resource Economics*, 25(4):435–459.

Fleurbaey, M. and Abi-Rafeh, R. (2016). The use of distributional weights in benefit–cost analysis: Insights from welfare economics 1. *Review of Environmental Economics and Policy*, 10(2):286–307.

Geary, R. C. (1950). A note on "a constant-utility index of the cost of living". *The Review of Economic Studies*, 18(1):65–66.

Hanemann, W. M. (1991). Willingness to Pay and Willingness to Accept: How Much Can They Differ? *The American Economic Review*, 81(3):635–647.

Hicks, J. R. (1932). *The Theory of Wages*. Macmillan: London.

Hossack, F. and An, H. (2015). Does payment type affect willingness-to-pay? valuing new seed varieties in india. *Environment and Development Economics*, 20(3):407?423.

Institut National de la Statistique et des Etudes Economiques (2017). Revenu, niveau de vie et pauvreté en 2014 Enquête Revenus fiscaux et sociaux (ERFS) - Insee Résultats.

Johansson, P.-O. and Kriström, B. (2018). *Cost–Benefit Analysis*. Cambridge University Press, 1 edition.

Kanninen, B. J. and Kriström, B. (1992). Welfare Benefit Estimation and the Income Distribution. *Beijer Discussion Paper Series No 20. Stockholm: Stockholm School of Economics.*

Lankford, R. H. (1988). Measuring welfare changes in settings with imposed quantities. *Journal of Environmental Economics and Management*, 15(1):45–63.

Ley, E. (2002). On plutocratic and democratic cpis. *Economics Bulletin*, 4(3):1–5.

Mäler, K.-G. (21974). *Environmental economics: a theoretical inquiry*. RFF Press.

Olsen, J. A. and Donaldson, C. (1998). Helicopters, hearts and hips: Using willingness to pay to set priorities for public sector health care programmes. *Social Science & Medicine*, 46(1):1–12.

Olsen, J. A., Donaldson, C., and Shackley, P. (2005). Implicit versus explicit ranking: On inferring ordinal preferences for health care programmes based on differences in willingness-to-pay. *Journal of Health Economics*, 24(5):990–996.

Pearce, D. W., Atkinson, G., and Mourato, S. (2006). *Cost-benefit analysis and the environment: recent developments*. Organisation for Economic Co-operation and Development, Paris. OCLC: ocm64672331.

Smith, R. D. (2005). Sensitivity to scale in contingent valuation: The importance of the budget constraint. *Journal of Health Economics*, 24(3):515–529.

Stone, R. (1954). Linear expenditure systems and demand analysis: An application to the pattern of british demand. *The Economic Journal*, 64(255):511–527.

Swenson, D. (2015). Foundations and methods for evaluating public policy effectiveness: Benefits and costs measures.

Chapter 7

General conclusion

The aim of this thesis was to contribute to the literature on methodological issues in non-market valuation. As explained previously, these issues arise at several stages of the non market valuation process: willingness to pay (WTP) elicitation, statistical analysis of WTP and WTP aggregation. In this thesis I have provided insights on some of the issues arising in all three stages, and concrete solutions to improve the valuation process.

Regarding the anchoring bias in multiple contingent valuation (CV) questions studies in the first chapter, I found no evidence of anchoring when a circular payment card is used in the Flood survey. In the CV survey on Social Insurance, which randomly uses three elicitation formats, we found greater anchoring on the first WTP with the open ended format than with the two payment card-type formats. This suggests that respondents may rely on different heuristic decisions when stating WTP in the open ended and in the two payment card formats.

The second chapter investigated the relationship between institutional variables and protest rate in environmental valuation studies. It provided insights to practitioners on how the protest rate can be affected depending on the country where a survey is conducted. Using meta-data merged with institutional variables and exploiting intra-country variations, I found that trust in the institutions is not a significant determinant of the protest behaviors. This result is in line with the results from previous literature. This chapter reinforces existing results by wiping out the study specificities and the good specificity.

Turning now to the second part on statistical analysis of WTP, the third chapter confirmed the advantages of (censored) quantile regression ((C)QR) models w.r.t. standard conditional mean estimates for analyzing WTP data, first through a Monte Carlo experiment and second, by applying them to a CV study on flood risk protection.

Moreover, the CQR model outperformed the other three models due to its ability to account both for censoring and heteroskedasticity, two common problems in WTP data. Using CQR appears relevant with any data that entail this type of problem (like number of occurrences of an event or quantities consumed) in many fields such as agriculture, energy, climate, environment and health. This also applies to economic data either directly observed on markets (prices, rates, taxes), indirectly revealed (shadow prices), or stated in surveys (WTP). Nor did we find significant differences in WTP elicited depending on the beneficiaries in the two scenarios tested.

In the fourth chapter, I tested the hypothesis of a selection of estimates based on their distance from the usual range in the literature. I showed that this selection happens for estimates in the lower part of the conditional distribution. No statistically significant effect was found for the upper part of the conditional distribution. These results imply that there is a systematic underestimation of the variance of mean estimates in meta-analyses of the value of a statistical life. The correct standard deviation of the value of a statistical life estimates would be around 60% larger than the non-corrected standard deviation. I was not able to conclude for an anchoring effect on the mean estimates.

The last chapter on WTP aggregation have important implications for both theoretical and empirical research in economics, as well as major policy implications. From a theoretical perspective, these findings contribute to the equity vs. equality debate. Should we increase the WTP elicited by the poorest when preferences for a non-market good are known to be income-dependent? By doing so, we are favouring equity (an as-fair-as-possible representation of preferences) over equality (a common representation of preferences). From an empirical perspective, our findings challenge the use of stated and revealed preference techniques in cost-benefit analysis (CBA) involving preferences for a non-market good and measured through WTP. When preferences are income-independent, the overall desirability of a project is under-estimated but the average preference of the population is accounted for. When preferences are income-dependent, decision-makers should be aware that the preferences of the poorest individuals may be under-represented and those of the richest over-represented. In terms of policy implications, we need to be on the lookout for any potential plutocratic bias when non-market values are involved in a CBA.

In conclusion, non-market valuation present many challenges. This thesis has only attempted to bring answers to some of the many questions that these methods raise, and a lot of work remains to be done. These questions involves various fields of research such as psychology and survey methodology (for preference elicitation), statistics and econometrics (for WTP analysis), and welfare economics (for WTP agregation). There is no doubt that the future work of non-market valuation methodology lies at the intersection of these different fields of research.