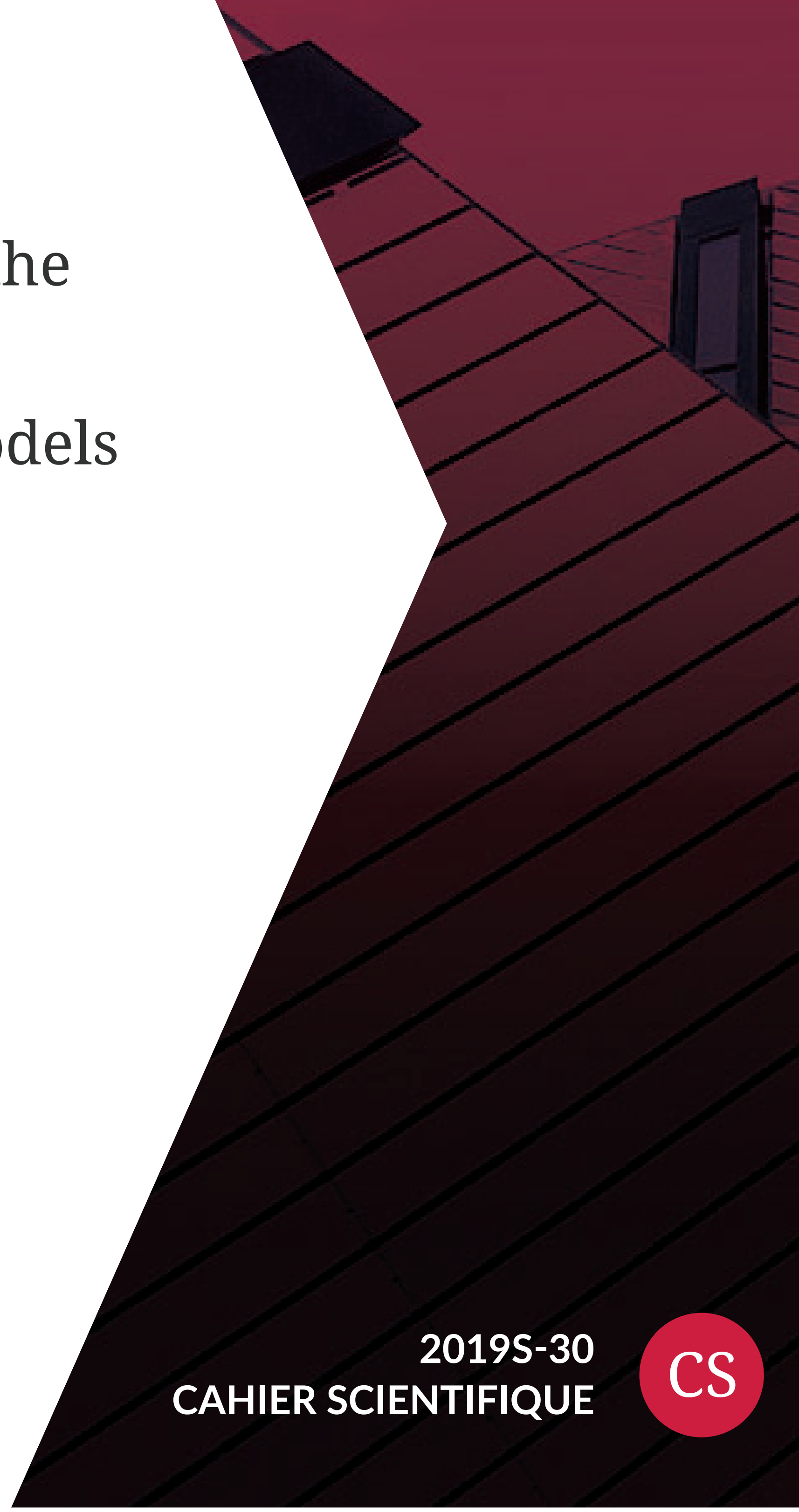
CIRANO
Allier savoir et décision

Factors Explaining the Hypothetical Bias: How to Improve Models for Meta-analyses

BAOUBADI **ATOZOU**
LOTA D. **TAMINI**
STÉPHANE **BERGERON**
MAURICE **DOYON**



2019S-30
CAHIER SCIENTIFIQUE
CS

2019s-30

Factors Explaining the Hypothetical Bias: How to Improve Models for Meta-analyses

Baoubadi Atozou, Lota D. Tamini, Stéphane Bergeron, Maurice Doyon

Série Scientifique
Scientific Series

Montréal
Décembre/December 2019

CIRANO
Allier savoir et décision

Centre interuniversitaire de recherche en analyse des organisations

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO

Partenaires corporatifs
Autorité des marchés financiers
Banque de développement du Canada
Banque du Canada
Banque Laurentienne
Banque Nationale du Canada
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du Québec
Canada Manuvie
Énergir
Hydro-Québec
Innovation, Sciences et Développement économique Canada
Intact Corporation Financière
Investissements PSP
Ministère de l'Économie, de la Science et de l'Innovation
Ministère des Finances du Québec
Mouvement Desjardins
Power Corporation du Canada
Rio Tinto
Ville de Montréal

Partenaires universitaires
École de technologie supérieure
École nationale d'administration publique
HEC Montréal
Institut national de la recherche scientifique
Polytechnique Montréal
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval
Université McGill

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

ISSN 2292-0838 (en ligne)

Factors Explaining the Hypothetical Bias: How to Improve Models for Meta-analyses

Baoubadi Atozou [*], *Lota D. Tamini* [†], *Stéphane Bergeron* [‡], *Maurice Doyon* [§]

Abstract/Résumé

Meta-analyses are getting more common in economics. However, little information is available regarding the choice of econometric models and its impact on results. Moreover, outlier data are common in meta-analyses and some authors have simply chosen to remove arbitrarily such data. We use the rich literature of meta-analysis on hypothetical bias (HB) related to contingent valuation methods to illustrate our point. More specifically, we review and update the meta-analyses of HB using a Meta-Regression Hierarchical Mixed Effect (MRHME) model and we apply a Bayesian Gibbs Sampling as classical results robustness check. A set of 462 observations from 87 economic valuation studies is used to this effect. The findings indicate that MRHME model is more efficient to explain HB. While respondents overstate their stated willingness-to-pay for a good by a factor of two, cheap talk, certainty correction, Ex Ante and Ex Post mitigation techniques significantly reduce the HB. Notwithstanding, mitigation techniques are more effective in private goods economic valuation.

Keywords/Mots-clés: Contingent Valuation, Hypothetical Bias, Meta-analysis, Willingness-To-Pay, Private Goods, Public Goods, Economic Valuation

JEL Codes/Codes JEL: C10, H40, Q00, Q20

[*] Corresponding author. Economist, Resilience Analysis and Policy, FAO. Email: baoubadi.atozou.1@ulaval.ca

[†] Université Laval and Center for Research on the Economics of the Environment, Agri-food, Transports and Energy (CREATE), Fellow CIRANO, Egg Industry Economic Research Chair

[‡] Université Laval, Egg Industry Economic Research Chair

[§] Université Laval and CREATE, Fellow CIRANO, Egg Industry Economic Research Chair

1 Introduction

Meta-analyses although widely used in health related science, were less common in economics until recently. Their use range from finance (Capon et al., 1990) to urban economics (Melo et al., 2009) to specific ecological goods (Loomis and White, 1996; Woodward and Wui, 2001). More recently, numerous studies using stated preference methods, such as contingent valuation, have emerged. A strong demand to assess public preferences for environmental and ecological goods' production programmes, ecosystems services, forest restauration services, and endangered species protection services motivate these studies (Carson et al., 1992; Johnston, 2006; Vossler and Evans, 2009; Murphy et al., 2010; Krawczyk, 2012; Lee and Hwang, 2016).

These methods have also been widely used for private goods, especially in the context of the development of new products, to assess consumers' willingness-to-pay (WTP) for these goods (Bergmo and Wangberg, 2007; Loomis et al., 2009; Moser et al., 2014; Doyon et al., 2015; Doyon and Bergeron, 2016). The literature emphasizes that stated preference methods potentially lead to a hypothetical bias (Bohm, 1972; NOAA, 1993; Cummings et al., 1995; Champ et al., 1997; Cummings and Taylor, 1999) defined as the difference between the hypothetical WTP measured by the declarative methods and the revealed or actual WTP (List et Gallet, 2001).

The large literature on HB has in turn generated numerous meta-analyses (List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005; Zawojska and Czajkowski, 2017; Foster and Burrows, 2017; Penn and Hu, 2018).

The findings of these meta-analyses are mixed and are sometimes not consistent in the signs and the significance of the effects of the explanatory factors of the hypothetical Bias (HB). The sensitivity of the results to the econometric models, the adequacy of these models with the structure of the meta-data and the fact that some unobservable characteristics and intra-study potential correlation and inter-study potential heteroscedasticities could explain the mixed results.

This article updates previous meta-analyses with emphasis on the use of a new sophisticated model, the Meta-Regression Hierarchical Mixed Effects model (MRHME), which corrects the effects of the unobservable characteristics and potential heteroscedasticity specific to each study (Moeltner et al., 2007; Dekker et al., 2011). To increase the robustness of our results, we estimate our MRHME models using the classical approach with the maximum likelihood method and the Bayesian approach with the Gibbs sampling process. The Bayesian approach corrects the effects of the outliers and low representativeness of some characteristics (Koop, 2003; Moeltner et al., 2007, Dekker et al., 2011) and allows us to test the robustness of our estimates. We also estimated the Log-linear benchmark models and used the Likelihood Ratio test to compare this model to our MRHME models.

The rest of the paper is as follows. Section 2 presents the literature review of previous meta-analyses. Section 3 describes the design of the database and the variables considered in the analysis. The econometric model and the empirical estimation approaches is the subject of section 4. Section 5 presents the results and Section 6 concludes.

2 Literature review of previous meta-analyses

Hypothetical bias is well known in the literature as a problem in contingent valuation methods (Bohm, 1972; Carson et al., 1992; NOAA, 1993; Penn and Hu, 2018). It represents the discrepancy between individuals' hypothetical and real WTPs.

Many studies using contingent valuation method (CVM) for economic valuation of goods highlighted that respondent-stated WTPs are significantly different from their real WTPs (Neill et al., 1994; Cummings et al., 1995; Loomis et al., 1996; Champ et al., 1997; Cummings and Taylor, 1999; Vossler et al., 2003; Brown et al., 2003; Murphy et al., 2005a; Blumenschein et al., 2007). It has since become important for researchers to understand and provide solutions to mitigate this bias in their stated preference valuations. Several studies analysed the HB in economic valuations of private and public goods with stated preference methods and ways to correct them (Champ et al., 1997; Cummings and Taylor, 1999; Vossler et al., 2003; Brown et al., 2003; Murphy et al., 2005b; Blumenschein et al., 2008). In addition, some CVM calibration techniques such as Certainty Correction (Champs et al., 1997), Cheap Talk (Cummings and Taylor, 1999), perceived consequentiality (Carson and Groves, 2007), honesty priming (de-Magistris et al., 2013), and religious priming (Stachtiaris et al., 2011) have been developed to eliminate or significantly reduce the HB. Despite the number of investigations conducted on this issue, there is no consensus in the literature on the determinants of HB, and the calibration techniques have mixed results.

Six main meta-analyses were conducted to summarize the empirical contributions of public and private economic evaluation studies in order to develop a theoretical basis for HB and to understand the factors that systematically drive it (List and Gallet, 2001; Little and

Berrens, 2004; Murphy et al., 2005; Little et al., 2012; Foster and Burrows, 2017; Penn and Hu, 2018). Table 1 presents the keys results and econometric models used for each of these previous meta-analyses.

Table 1: Selected Results from Previous Meta-Analyses on Hypothetical Bias

Study	List and Gallet (2001)	Little and Berrens (2004)	Murphy et al. (2005)	Little et al. (2012)	Foster and Burrows (2017)	Penn and Hu (2018)
Dependent Variable	Ln (Hypothetical WTP/ Real WTP)	Y = 1 if SS of hypothetical bias, 0 else	Ln (Actual WTP)	Y = 1 if SS of hypothetical bias, 0 else	Ln (Hypothetical WTP/ Real WTP)	Ln (Hypothetical WTP/ Real WTP)
Econometric Models	Log-linear	Probit	Log-linear	Probit	Log-linear and Log-linear Fixed Effects	Log-linear
Estimation approaches	Classical	Classical	Classical	Classical	Classical	Classical
Number of Studies (observations)	29 (58)	53 (85)	28 (77)	96 (220)	78 (432)	132 (908)
Private Good	SS, Less HB	Not SS, Less HB	SS, Less HB	-	SS, More HB	-
Public Good	-	-	-	-	-	SS, More HB
Student sample	-	-	SS, More HB	SS, More HB	Not SS, More HB	Not SS, Less HB
Within Respondent	Not SS, Less HB	Not SS, Less HB	SS, Less HB	Not SS, More HB	Not SS, More HB	-
Between-Respondent	-	-	-	-	-	Not SS, Less HB
WTP	SS, Less HB	-	-	-	-	-
WTA	-	Not SS, More HB	-	Not SS, More HB	-	SS, Less HB
Lab setting	Not SS, Less HB	Not SS, More HB	-	Not SS, Less HB	SS, Less HB	Not SS, More HB
HB mitigation approaches	-	-	SS, Less HB	SS, Less HB	-	-
Choice experiment	-	-	-	Not SS, Less HB	Not SS, Less HB	SS, Less HB
Induced Value	-	-	-	Not SS, Less HB		SS, Less HB

Study	List and Gallet (2001)	Little and Berrens (2004)	Murphy et al. (2005)	Little et al. (2012)	Foster and Burrows (2017)	Penn and Hu (2018)
Cheap Talk	-	-	-	-	SS, Less HB	SS, Less HB
Certainty follow-up	-	-	-	-	SS, Less HB	SS, Less HB
Consequentiality	-	-	-	-	-	SS, Less HB

[a]Updating the Table1 of Penn and Hu (2018) with adding their keys results. Notes: HB indicates "Hypothetical Bias", SS indicates "Statistical Significant", Not SS indicates "Not Statistical Significant" in 50% or more in the appropriate models. "-" indicates the variable was not included in the meta-analysis, WTA indicates "Willingness-To-Accept", WTP indicates "Willingness-To-Pay", Lab indicates "Laboratory", Less indicates "Negative Sign", More indicates "Positive Sign".

The results on the effects of factors that systematically affect the HB are mixed. The findings of List and Gallet (2001), Murphy et al. (2005) and Penn and Hu (2018) indicate that private goods significantly reduce HB. By contrast, Little and Berrens' (2004) results reveal that the type of good has no significant effect on the probability of observing a HB, while Foster and Burrows (2017) find that private goods significantly increase the HB. The results of the previous meta-analyses also diverge regarding the effects of the "within respondents", "laboratory" and "students" variables on the magnitude of the HB. Indeed, List and Gallet (2001) and Little and Berrens (2004) find that Within-Respondents reduces the HB, but this reduction is not statistically significant. Murphy et al. (2005) results suggest that Within-Respondents significantly reduces the magnitude of the HB, while Foster and Burrows (2017) and Penn and Hu (2018) find that Within-Respondents increases the HB, but this increase is not statistically significant. Furthermore, List and Gallet (2001) and Murphy et al. (2005) show that "Students" significantly increases the HB, and Foster and Burrows (2017) and Penn and Hu (2018) found that "Students" insignificantly increase and decrease the HB, respectively. Foster and Burrows (2017) findings indicate that the “Lab setting” significantly reduces the HB. However, although the “Lab setting” effect is not statistically significant in the results of List and Gallet (2001), Little and Berrens (2004), Little et al. (2012), and Penn and Hu (2018), it is negative for List and Gallet (2001) and Little et al. (2012) and is positive for Little and Berrens (2004) and Penn et Hu (2018).

The sensitivity of the results to the choice of econometric models, the adequacy of these models with the structure of meta-data and the fact that some unobservable characteristics and intra-study potential heteroscedasticities may explain the mixed results. List and Gallet (2001), Murphy et al. (2005), and Penn and Hu (2018) used Log-linear model in their meta-analyses of HB, Little and Berrens (2004), and Little et al. (2012) used a Probit model, while Foster and Burrows (2017) used Log-linear and Log-linear Fixed Effect model. The presence of outliers introduces potential bias in the estimates of linear models (Cook, 1977; Andrews and Pregibon, 1978; West, 1984; Chetterjee and Hadi, 1986; Anderson and Legendre, 1999; Wisnowski et al., 2001; Adnan et al., 2003; Zuur et al., 2010). Another issue is that several observations may come from the same study (List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005; Little et al., 2012; Foster and Burrows, 2017; Penn and Hu, 2018), and there may be a correlation between these observations. According to Moeltner et al. (2007) and Dekker et al. (2011), it is thus likely to observe a heteroscedascity due to this potential correlation, but also unobservable characteristics intrinsic to each study that can also affect the results of the estimates. The previous meta-analyses on HB did not control these aspects as well as the problem of low frequency of certain characteristics in the database that may affect the results in the estimation of their model. Finally, the estimates provide average effects of the factors. A variable that has a non-significant effect on HB does not indicate that this effect is statistically zero for private and the public goods. Introducing interaction variables between the key explanatory factors and the type of good could substantially improve understanding of HB and highlight the adequacy of factors with the type of property to reduce or eliminate HB in WTP valuation. Previous meta-analyses did not take into account this relevant aspect.

3 Data Description

3.1 Selection criteria

We adopt two inclusion criteria for relevant studies in our meta-analysis. First, we include studies that reported the average hypothetical WTP (WTP_h) and stated WTP (WTP_r). These values are used to obtain our dependent variable, the Hypothetical Bias Factor $(HBF) = \frac{MeanWTP_h}{MeanWTP_r}$. Second, we include studies that clearly and accurately described their experimental designs, the target population, and the good for both the hypothetical and real WTPs survey treatments. Following Murphy et al. (2005), we excluded the Willingness-to-Accept (WTA) studies because they are rarely used. We also included in our meta-analysis the studies that estimated the hypothetical and real WTPs using different survey mechanisms, and we used the "same mechanism" as the explanatory variable to detect the different elicitation survey effects on the HB. First, according to the criteria of inclusions, we selected the papers that were used in the previous meta-analysis (List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005; Little et al., 2012). Second, following the protocol of the meta-analysis implementation described by Little et al. (2008) and Stanley et al. (2013), we searched for keywords and their combinations through electronic databases such as *Google Scholar, Econlit, Web of*

Science, Business Source Complete, CAB Abstracts, Academic Search, and Cairn, and the studies that were not considered in these previous meta-analyses that matched our inclusion criteria. We obtained 87 studies, including 44 studies about private goods and 43 studies about public goods. Table A2 (see Appendix) summarized the selected studies.

Table 2 presents the descriptive statistics of the HBF. The average HBF is 2.11 with a standard deviation of 2.44 and a median of 1.41. The proportion of the observations of the HB mitigation techniques is 32.61% of the full sample (Table 3). The results obtained with these calibration techniques seem to be on average more accurate than those obtained without a calibration technique. In fact, the average HBF for the mitigation techniques sub-sample is 1.42 compared with 2.52 in the sub-sample without the HB mitigation techniques.

Table 2: Descriptive Statistics of Hypothetical Bias Factor (HBF)

HBF	Mean	Median	SD	CV	Observations.
Full Sample	2.11	1.41	2.44	0.86	462
Calibration[a]	1.42	1.08	0.94	0.66	171
Without Calibration	2.52	1.58	2.91	1.15	291

Notes: Standard Deviation (SD), Coefficient of Variation (CV), [a] Sub-sample of observations using calibration techniques.

Figure 1 shows the percentile values of the HBF according to the type of good and the total sample. The median of the hypothetical bias factor is 1.41 for private goods versus 1.39 for public goods (Figure 1). The figure shows that 90% of HBF observations are below 3.69 for private goods and 4.02 for public goods.

Figure 1 : Percentiles distribution of hypothetical bias factor (HBF)

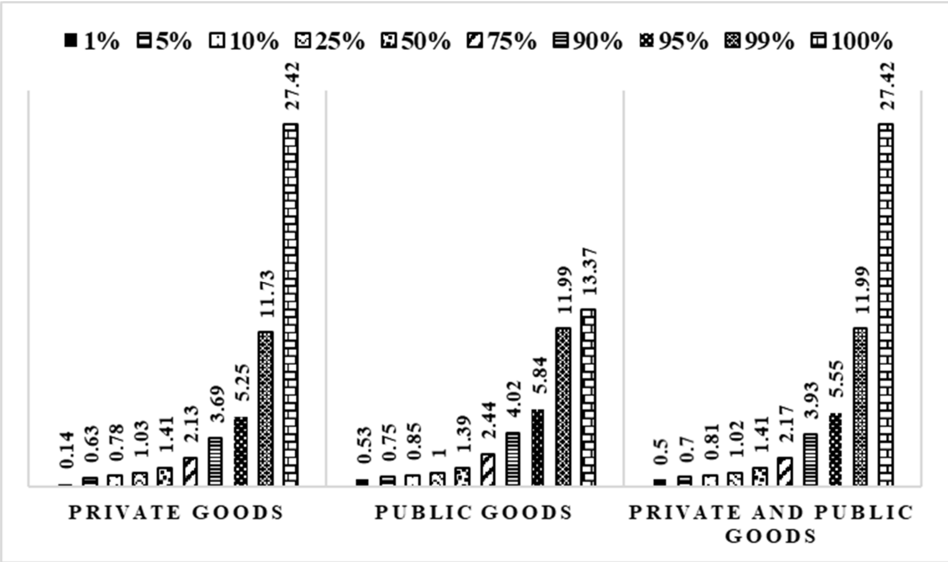
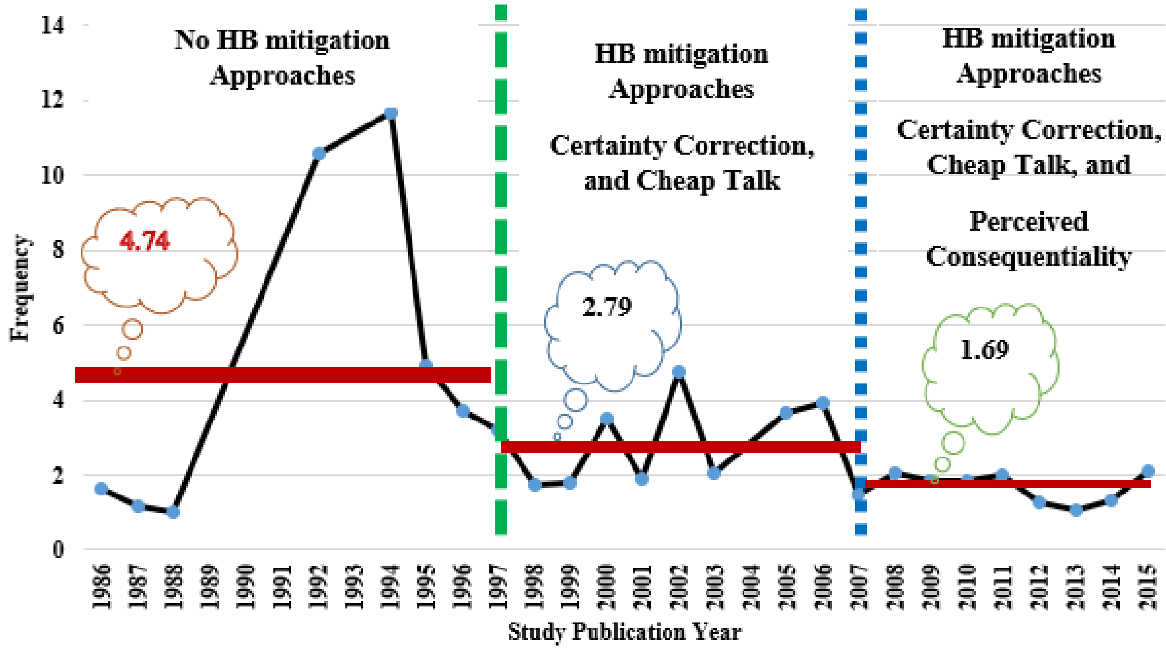


Figure 2 shows the change in the mean of the HBF by the year of publication. This figure shows a gradual improvement in the estimates of subjects' preferences with the stated preferences methods, especially with the use of cheap talk and certainty correction. Moreover, with the introduction of perceived consequentiality by Carson and Groves (2007) as the HB mitigation approach since 2007, in addition to cheap talk and certainty correction, the WTP predictions with CVMs have continued to improve.

Figure 2 : Dynamic of the average of hypothetical bias factor (HBF) by Publication year and different periods of key HB mitigation approaches



3.2 Dependent variable: hypothetical bias factor (HBF)

The use of the discrepancy between the hypothetical and real WTPs as the unit factor could introduce some issues because the studies are carried out in various countries and the WTPs are measured in different units. Therefore, our dependent variable is the hypothetical bias, which is the ratio of the hypothetical WTP and actual WTP and we called hypothetical bias factor (HBF). If the HBF is equal to one, then the hypothetical bias is zero.

3.3 Explanatory factors

We define the WTP elicitation mechanisms, survey respondents' characteristics, mitigation techniques, and type of good as independents variables (List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2012).

3.3.1 *Calibration techniques*

The authors developed techniques for calibrating contingent valuation methods to eliminate or reduce the HB and improve the reliability of the results obtained with the stated preference methods (CVMs) (Champs et al., 1997; Cummings and Taylor, 1999; Carson and Groves, 2007; Vossler et al., 2012; Loomis, 2011, 2014) as follows: *Cheap talk script* (Cummings and Taylor, 1999) and *certainty correction* (Champ et al., 1997). Several authors have used these techniques and mixed results have emerged in the literature (Poe et al., 2002; Brown, 2003; Aker et al., 2008; Bedate et al., 2009; Moser et al., 2014; Doyon et al., 2015). Nevertheless, previous meta-analyses have shown that both techniques significantly reduce the HB (Little and Berrens, 2004; Murphy et al., 2005; Foster and Burrows, 2017; Penn and Hu, 2018). Since 2005, several other techniques have been created, such as *perceived consequentiality* (Carson and Groves, 2007), *honesty priming* (de-Magistris et al., 2013), and *religious priming* (Stachtiaris et al., 2011), and they have been used to reduce HB. We aggregate these mitigation techniques into a single *calibration* variable that is set to one if a mitigation approach is used in the study, and it is set to zero otherwise (Murphy et al. 2005; Little et al., 2012). We consider two other variables. First, an *ex-ante* calibration technique variable is set to one if cheap talk, honesty, or religious priming is used as the mitigation technique, and it is set to zero otherwise. Second, an *ex post* calibration technique is incorporated that takes the value of 1 if certainty correction or explicit perceived consequentiality is used to calibrate the stated preference methods, and it takes the value of 0 otherwise. *Cheap talk* and *certainty correction* are also the binary explanatory factors. The *cheap talk* indicator takes the value of one if this calibration technique is applied in the hypothetical survey treatment and 0 otherwise. The *certainty correction* binary variable is set to one if it is used to calibrate the CVM and 0 otherwise.

3.3.2 Other variables

Among the stated preference mechanisms used in the identified studies, we can mention the following, although it is not an exhaustive list: the Vickrey Auction, Open-ended, Dichotomous Choice, Referendum, the Becker-DeGroot-Marschak (BDM) procedure, and the Nth Prize Auction. List and Gallet (2001), Little and Berrens (2004) and Murphy et al. (2005) have shown that these elicitation mechanisms have potentially different effects on the HB. Each of these elicitation mechanisms is set to one if it is used in the survey treatment, and it is set to 0 otherwise. The variable "*same mechanism*" takes the value of one if the same survey treatment is used for the hypothetical and real WTP elicitations, and it takes the value of zero otherwise.

The preferences of an economic agent may differ depending on whether the appraised good is a public or private good. The previous meta-analyses find mixed effects of the type of good on the HBF (List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005; Little et al., 2012; Foster and Burrows, 2017; Penn and Hu, 2018). The binary variable *Private* is set to one if the economic valuation study is for a private good and zero otherwise. Carson and Groves (2007) suggest that when the treatment survey experiment is consequential, using the incentive compatible mechanisms allows for obtaining the respondents' real WTPs in a hypothetical treatment. The *incentive compatible mechanism* indicator (ICM) variable takes the value of 1 if the WTP is estimated using Dichotomous

Choice, Referendum, the Vickrey Auction, the Nth Price Auction, or the BDM Procedure (Carson and Groves, 2007; Carson et al., 2014), and 0 otherwise.

Student samples are broadly used as subjects in economic valuation studies using experimental economics, whether for public goods (Carlsson and Martinson, 2001; Murphy et al., 2005b; Mozumder and Berrens, 2007; Vossler and Evans, 2009; Lee and Hwang, 2016) or for private goods (List and Shorgen, 1998; Johannesson et al., 1999; Balisteri et al., 2001; Ehmke et al., 2008). The student-respondents variable is considered as the explanatory factor in previous meta-analyses (Murphy et al., 2005; Little et al., 2012; Foster and Burrows, 2017; Penn and Hu, 2018). The variable *Student* take the value of one if the study's subjects are entirely students, and it is set to zero otherwise.

Table 3 provides the descriptions and statistics of the variables.

Table 3: Variable Description

Variable	Description	Obs.	Prop.
	Good Characteristics		
Private	1 if the evaluated good is a private good and 0 otherwise	236	50.97
	Type of Experimental Survey		
Laboratory	1 if experimental survey was performed in a laboratory, 0 otherwise	225	48.6
Field Survey	1 if experimental survey is a field survey, 0 otherwise	168	36.29
Mail Survey	1 if experimental survey is a mail survey, 0 otherwise	66	14.25
Phone Survey	1 if experimental survey is a phone survey, 0 otherwise	4	0.86
	Type of Survey Respondents		
Student	1 if subjects used in the experiment survey are students, and 0 otherwise	212	45.79
	Type of Comparison		
Between-Respondents	1 if respondents are different in hypothetical and real WTP valuation, and 0 otherwise	397	85.75
	Contingent Valuation Methods		
Open-ended	1 if WTP elicitation mechanism is Open-Ended, 0 otherwise	66	14.25
Vickrey Auction	1 if WTP elicitation mechanism is Vickrey Auction, 0 otherwise	35	7.56
Nth Price Auction	1 if WTP elicitation mechanism is Nth Price Auction, 0 otherwise	29	6.26
IACA	1 if WTP elicitation mechanism is IACA, 0 otherwise	4	0.86

Variable	Description	Obs.	Prop.
BDM	1 if WTP elicitation mechanism is BDM, 0 otherwise	7	1.51
Referendum BDM	2 if WTP elicitation mechanism is a referendum BDM, 0 otherwise	10	2.16
Dichotomous Choice	1 if WTP elicitation mechanism is dichotomous choice, 0 otherwise	141	30.45
MDC	1 if WTP elicitation mechanism is MDC, 0 otherwise	88	19.01
Referendum	1 if WTP elicitation mechanism is a referendum, 0 otherwise	70	15.12
SDCE	1 if WTP elicitation mechanism is SDCE, 0 otherwise	9	1.94
Same Mechanism	1 si le mécanisme de l'expérience réelle est le même que celui de l'expérience hypothétique et 0 sinon	395	85.31
	Calibration Techniques		
Cheap Talk	1 if HB mitigation technique is cheap talk, and 0 otherwise	79	17.06
Certainty Correction	1 if HB mitigation technique is certainty correction, and 0 otherwise	39	8.42
Honesty	1 if HB mitigation technique is honesty, and 0 otherwise	4	0.86
Own Money	1 if HB mitigation technique is own money, and 0 otherwise	6	1.30
Explicit Consequentiality	1 if explicit consequentiality question is asked, and 0 otherwise	17	3.67
Calibrate (Aggregated)	1 if a calibration technique is used and 0 otherwise	145	31.31
	Hypothetical Bias Factor		
Facteur du biais hypothétique (FBH)	Ratio of Hypothetical and Real WTP (WTPh/WTPr)	463	NA
Observations	Total of Observations	463	NA

4 Econometric Model

We adopt the meta-regression hierarchical mixed effect (MRHME) model used by Moeltner et al. (2007) and Dekker et al. (2011) for two main reasons. The MRHME model (i) addresses the study-specific heteroskedasticity by random parameter specifications (Moeltner et al., 2007) and (ii) controls the effects of unobservable characteristics. We assign fixed coefficients to the explanatory factors that do not have sufficient inter-study variability to allow for random coefficient specifications, as suggested by Moeltner et al. (2007). These variables include all the explanatory variables that are generally invariant between the observations of a given study, such as the study's authors.

Let y_{ijs} be the calibration factor that is estimated in study s with hypothetical experience i and actual experience j. For the same study, the characteristics of the experimental design and the stated preference methods influence the HBFs. The unobservable characteristics associated with the authors also have influences. Therefore, we take into account the intra-study variability of the HBF related to the experimental design, the WTP assessment methods and the inter-study variability of the HBF related to the unobservable factors related to each study. These factors may lead to heteroscedasticity that is related to the methodological features (Koop, 2003, Chapter 6). To solve this problem, Moeltner et al. (2007) proposed making the effects of these explanatory factors random and considering the effects of the other variables that do not generate this internal variability of the HBF as fixed. Thus, our model relies on that of Moeltner et al. (2007) and Dekker et al. (2011):

$$\begin{aligned} & Y_{ijs} \left| (.) \right. = \exp\left(M'_{r,ijs}\beta_{rs} + B'_{f,ijs}\beta_{f,x} + \varepsilon_{ijs} \right) \exp\left(E'_{f,ijs}\beta_{f,e} \right) \\ & \quad with\ \beta_{rs} \sim mvn\left(b,\Sigma\right)\ and\ \varepsilon_{ijs} \sim n\left(0,\sigma^2\right) \end{aligned} \tag{1}$$

where *mvn* and *n* represent the multivariate and univariate normal distributions, respectively. The vectors $M_{r,ijs}$ and $B_{f,ijs}$ are the methodological characteristics and those of the evaluated good, respectively. The parameters β_{rs} associated with the methodological characteristics are random coefficients. The matrix of regressors $E_{f,ijs}$ refers to the matrix of the characteristics of the sample of WTP treatment survey-respondents. The parameters associated with the type of good, the type of WTP treatment survey-respondents and the author level are fixed coefficients. The vectors of coefficients β_{rs}, $\beta_{f,ijs}$, and $\beta_{f,e}$ are the sub-vectors of the vector of coefficients that are respectively associated with the explanatory regressors of the following vectors: $M_{r,ijs}$, $B_{f,ijs}$ and $E_{f,e}$. The vector of random coefficients follows a multivariate normal distribution of mean *b* and the variance-covariance matrix Σ. The stochastic error term also follows according to equation (1) as a normal distribution with zero mean and variance σ^2. The logarithmic transformation of equation (1) gives the following expression of the meta-regression model:

$$\begin{aligned} \ln\left(Y_{ijs} \left| X_{r,ijs}, Z_{ijs} \right. \right) &= M'_{r,ijs}\beta_{r,ijs} + B'_{f,ijs}\beta_{f,ijs} + E'_{f,ijs}\beta_{f,e} + \varepsilon_{ijs} \\ &= X'_{r,ijs}\beta_{r,ijs} + Z'_{ijs}\beta_{f} + \varepsilon_{ijs} \end{aligned} \tag{2}$$

where X_{ijs} is the matrix of random coefficient regressors $\left(M_{ijs}\right)$, and Z_{ijs} is the matrix of explanatory variables with fixed effects $\left(B_{f,ijs}, E_{f,ijs}\right)$. The hypothesis of the normality of the random coefficients $\left(\beta_{rs}\right)$ and the stochastic error term $\left(\varepsilon\right)$ implies that the HBF vector of the study, which is noted as $\left(\ln\left(Y_{ijs}\left|X_{r,ijs}, Z_{ijs}\right.\right)\right)$, follows a multivariate normal distribution. Thus, the statistical inference of our variable of interest is estimated by the following equations (Dekker et al., 2011):

$$\ln\left(Y_s\left|X_{rs}, Z_{fs}\right.\right) = X_{rs}\beta_{rs} + Z_{fs}\beta_{fs} + \varepsilon_s \text{ with,}$$

$$E\left[\ln\left(Y_s\left|X_{rs}, Z_{fs}\right.\right)\right] = X_{rs}b + Z_{fs}\beta_{fs} \text{ and} \tag{3}$$

$$E\left[\ln\left(Y_s\right)\ln\left(Y_t\right)'\right] = \begin{cases} X_{rs}\Sigma X'_{rs} + \sigma^2 I_{n_s}, & s = t \\ 0, & \text{if not} \end{cases}$$

The dimension of the vectors $\ln\left(Y_s\left|X_{rs}, Z_{fs}\right.\right)$, X_{rs} and Z_{fs} are all equal to the number of observations n_s reported by study *s*, and I_n is a square matrix of dimension $\left(n_s * n_s\right)$. Since the matrix of random effects variables X_{rs} is included in the variance-covariance matrix of the dependent variable, the model specification captures the observed and study-specific heteroscedasticity (Moeltner et al., 2007; Dekker et al., 2011). According to Moeltner et al. (2007) and Swamy (1970), the estimation of the MRHME model under the normality hypothesis with random coefficients has good, desirable properties. First, it corrects for heteroscedasticity. Second, as indicated in equation (3) and specifically in expression of $E\left[\ln\left(Y_s\right)\ln\left(Y_t\right)'\right]$, the random coefficient specification introduces correlation across intra-study observations, both via the regressors included in matrix X_{rs} and via the unobserved elements common to all observations for a given study through the random intercept. This specification of the MRHME model increases the efficiency of the model and avoids the erroneous estimation of the standard error compared to the simple model that treats all variables as independent (Moeltner et al., 2007). Newman et al. (2010) show that MRHME models are most appropriate when variables are nested and intra-class or intra-study observations (Moeltner et al., 2007; Dekker et al., 2011) correlation. Field (2009); Kreft (1996); Morris, (1995); Mundfrom and Schultz (2002); Raudenbush and Bryk (2002); and Tabachnick and Fidell, (2007) have indicated that hierarchical model is superior to ordinary least squares (OLS) because it theoretically produces appropriate error terms that control for potential dependency due to nesting effects, while OLS does not. We estimate the MRHME and conduct the Likelihood ratio test to check its efficiency versus OLS. The specific expression of the meta-regression hierarchical mixed effects model is as follows:

$$\ln HBF = \ln\left(\frac{Hypothetical\ WTP}{Real\ WTP}\right) =$$

$$f\left(\begin{array}{l} Private\ Good,\ Student,\ Between\text{-}Respondent,\ Fiel\ Survey,\ Vickery\ Auction \\ MDC,\ Dichotomous\ Choice,\ Open\text{-}ended,\ Referendum,\ Same\ Mechamism, \\ ICM,\ Calibrate,\ Calibrate\ Ex\ Ante,\ Calibrate\ Ex\ Post,\ Cheap\ Talk,\ Certainty \\ Correction,\ Calibrate\ \times Private,\ Calibrate\ Ex\ Ante\ \times Private,\ Calibrate\ Ex\ Post \\ \times Private,\ Same\ Mechanim\ \times Private,\ Field\ Survey\ \times Private\ ICM\ \times Private \\ \left|\ \underbrace{Authors,\ Calibration\ Techniques,\ Mechanism}_{Ramdom\ levels\ control\ variables}\right. \end{array}\right)$$

Given the fact that the design of mitigation approach depends on the choice and sometime subjective with certainty correction approach (Blumenschein et al., 2007; Blomquist et al., 2009, Champ et al. 2009, Morrison and Brown 2009; Broadbent 2014) and the length of cheap talk script, we introduce in our model the random slopes for the calibration techniques (Moeltner et al., 2007; Dekker et al., 2011).

We estimate the model using the classical maximum likelihood method and the Bayesian approach to test the robustness of our results. Indeed, the efficiency of the classical model strongly depends on the number of observations and the number of parameters to be estimated. An alternative approach for solving this shortcoming of classical models is the Bayesian approach (Moeltner et al., 2007; Moeltner and Woodward, 2009) that has several advantages over a conventional approach. First, the theory of large samples is not necessary. Second, a single additional parameter allows for hierarchically modelling of the heteroscedasticity of the error terms. Finally, the specification of the prior distributions allows for consideration of the missing relevant information in the metadata database. The details information on our Bayesian model specification is in the Appendix.

5 Empirical Results

We estimate four models. The first model (model 1) includes the following as explanatory variables: *Private*, the experimental design characteristics and the calibration variable summarizing the HB mitigation techniques. Model 2 investigates the effect of the calibration techniques (*Calibrate*) regarding the type of good (*Private*). In this model, we also investigate the interaction between the experimental characteristics and the type of good. In model 3, we go further and test the effectiveness of *ex ante* calibration techniques (Cheap talk and honesty) and *ex post* calibration techniques (Certainty and explicit perceived consequence) in reducing the HB according the type of good (*Private*). Several authors show that the *Dichotomous Choice* (Carson and Groves, 2007; NOAA 1993), *Vickrey auction, Nth Price Auction, BDM* procedure, and *Referendum* mechanisms (Lusk et al., 2007; Carson and Groves, 2007; Carson et al., 2014) are consistent with subjects' real preference revelations in the WTP CVM. Therefore, model 4 includes the incentive compatible mechanism (*ICM*) variable that takes the value of one if one of the compatible incentive mechanisms (dichotomous choice, Referendum, the Vickrey Auction, the Nth price auction and the BDM procedure) is used in the WTP estimation, and it takes the value of 0 otherwise. In addition, *Cheap Talk* (Cummings and Taylor, 1999) and *Certainty Correction* (Champ et al., 1997) are introduced in this second model as explanatory variables.

5.1 Log linear models versus MRHME Models

We estimate the hierarchical mixed-effect meta-regression models using the maximum likelihood method. Table 4 presents the results. The overall significance test of the model (Wald test) shows that all the models are significant and valid at the 1% level - model 1 (Chi2 (11) 56.27, P-value <0.01), model 2 (Chi2 (12) 71.98, P-value <0.01), model 3 (Chi2 (14) 87.25, P-value <0.01) and model 4 (Chi2 (8) 74.43, P-value <0.01).

The results of the likelihood ratio test show that the four MRHME models explain better the HB than the log-linear models (Cameron and Trivedi, 2010). The LR-test results for the four models (Table 4) are respectively (LR-stat 162.09, P-value <0.001), (LR-stat 187.31, P-value <0.01), (LR-stat 153.82, P-value <0.01), and (LR-stat 178.56, P-value <0.01), and indicate that the unobservable characteristics and heteroscedasticity have significant effects on the estimated parameters. Therefore, the use of the Log-linear regression leads to biased results. The MRHME model provides a significant and substantial improvement for the explanation of the HBF.

Table 4: Classical Estimation Results of MRHME Models

Variables	Model 1		Model 2		Model 3		Model 4	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Constant	0.709***	0.224	0.523*	0.295	0.692***	0.218	0.741***	0.182
Private	-0.153	0.134	0.393	0.356	-0.128	0.132	-0.040	0.132
Field Survey	0.262*	0.145	0.151	0.270	0.255*	0.141	0.233	0.147
Students	0.092	0.124	0.039	0.134	0.081	0.120	0.042	0.124
Between-Respondents	-0.250**	0.119	0.163	0.202	-0.264**	0.116	-0.326***	0.118
Vickrey Auction	0.193	0.149			0.183	0.146		
MDC	-0.177	0.187			-0.185	0.182		
DC	0.042	0.161			0.076	0.156		
Open-Ended	0.133	0.164			0.147	0.160		
Referendum	-0.358*	0.221			-0.375*	0.213		
Same Mechanism	0.108	0.123	0.143	0.245	0.115	0.120	0.058	0.115
ICM			-0.308**	0.149			-0.004	0.094
Calibrate	-0.331***	0.058	-0.261***	0.075				
Calibrate Ex Ante					-0.220***	0.091		
Calibrate Ex Post					-0.339***	0.109		
Cheap Talk							-0.285***	0.066
Certainty Correction							-0.644***	0.090
Calibrate × Private			-0.193**	0.108				
Calibrate Ex Ante × Private					-0.133	0.127		
Calibrate Ex Post × Private					-0.456***	0.173		
Same Mechanism × Private			-0.229	0.283				
ICM × Private			0.499**	0.197				
Between-Respondent × Private			-0.692***	0.250				
Field Survey × Private			0.147	0.315				

Variables	Model 1		Model 2		Model 3		Model 4	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Random Effects								
Sd(calibrate)	0.146	0.077	0.099	0.098				
Sd(calibrate Ex ante)					0.133	0.122		
Sd(calibration tech.)					0.008	0.019		
Sd(certainty correction)							6.3e-7	5.42e-6
Sd(Cheap Talk)							0.050	213
Sd(_cons)	0.497	0.051	0.537	0.052	0.478	0.056	0.514	0.050
Sd(Residual)	0.423	0.016	0.415	0.016	0.412	0.016	0.422	0.016
Observations	460		460		460		460	
Wald Test	ddl (11)		dd(12)		ddl(14)		ddl(8)	
Chi2 (ddl)	56.27		71.98		87.25		74.43	
P-value	<0.0001		<0.0001		<0.0001		<0.0001	
Likelihood Ratio Test (LR-Test)								
Likelihood LL	-341.012		-337.194		-326.97		-338.707	
Likelihood LL C	-422.054		-430.848		-403.88		-427.987	
Chi2 test ddl	2		2		3		3	
Chi2 stat. (LR test)	162.09		187.31		153.82		178.56	
P-value	<0.0001		<0.0001		<0.0001		<0.0001	

Note: *** (1 %), ** (5 %), and * (10 %), Standard Error (SE), Dichotomous Choice (DC), Multiple Discrete Choice (MDC), Incentive Compatible Mechanism (ICM)

5.2 Classical Estimation Results of MRHME Models

5.2.1 Calibration techniques and hypothetical bias

The results of Model 1 (Table 4) show that the calibration technique has statistically significant negative effects at the 1% level. This confirms the results of Little and Berrens

(2004) and Murphy et al. (2005). Moreover, Ex Ante calibration techniques (*Cheap Talk, Honesty priming,* and *Religious priming*) and Ex Post (*Certainty Correction and perceived consequentiality*) calibration techniques reduce the HB, as indicated by the estimates of model 3. In models 2 and 3, we also investigated the effect of the interaction between calibration techniques and the type of good (*Private*). Model 2 shows that compared to public goods, calibration techniques are more effective in reducing the HB for private goods. The *Calibrate* × *Private* good interaction variable has a significant negative effect on the HBF. In addition, the results of model 3 indicate that the interaction variable between the *Ex Post* calibration techniques and the private good (*Ex Post Calibrate* × *Private*) has a negative and significant effect on the HBF. However, there is no interaction impact between the ex-ante calibration techniques and the type of goods (*Ex Ante Calibrate* × *Private*).

In model 4, more specifically, we introduced *Cheap Talk* (Cummings and Taylor, 1999) and *Certainty Correction* (Champ et al., 1997) as the explanatory variables. Our results show that the *Cheap Talk* and *Certainty Correction* calibration techniques are effective in eliminating or reducing the HB. The *Cheap Talk* approach and *Certainty Correction* have negative and statistically significant coefficients on the HBF, respectively, at the 1% level. Furthermore, the effect of the *Certainty Correction* technique (-0.645) remains higher than that of *Cheap Talk* (-0.286).

5.2.2 Contingent valuation methods

Several authors show that the *Dichotomous Choice* (Carson and Groves, 2007; NOAA 1993), *Vickrey auction*, *Nth Price Auction*, *BDM* procedure, and *Referendum* mechanisms

(Lusk et al., 2007b; Carson and Groves, 2007; Carson et al., 2014) are consistent with subjects' real preference revelations with the CVM. We estimated a model (model 4) that includes the incentive compatible mechanism (*ICM*) variable that takes the value of 1 if one of the compatible incentive mechanisms (dichotomous choice, Referendum, the Vickrey Auction, the Nth price auction, and the BDM procedure) is used in the WTP estimation, and it takes the value of 0 otherwise. Model 4's findings reveal that incentive compatible mechanisms (ICM) have a negative effect on the HBF, as predicted by Carson and Groves (2007), but it is not statistically significant.

To take into account the different possible impacts regarding the type of goods, we created an interaction variable between *ICM* and *Private*. Our estimated results that are presented in model 2 indicate a negative effect that is statistically significant for public goods (-0.257) while the impact is positive and statistically significant for private goods (the sum of the coefficients of ICM and ICM*Private is statistically equal to 0, Chi2 = 2.170, P-value = 0.140).

5.2.3 Type of goods

The results of model 1 show that the type of good (private) has no statistically significant direct effects on the HBF. This result contrasts the results of List and Gallet (2001), Murphy et al. (2005), and Penn and Hu (2018) that concluded that the HB increases when the evaluated good is a public good. Foster and Burrows (2017) find that private goods significantly increase the HB. In models 2-4, we further investigated the impact of the type of good by introducing interaction variables in the estimated models. The results show that the type of good has an impact when using the Between-Respondents design and mitigation techniques (*Between-Respondents* × *Private*, *Calibrate* × *Private*). It reduces the HBF, and their effects are statistically significant at the 1% and 5% levels, respectively. In contrast, the use of incentive compatible mechanisms (ICM) in the valuation of private goods appears to be ineffective in reducing the HB. In fact, the results reveal a positive and significant coefficient (p-value < 5%) of *ICM* × *Private* with respect to the HBF. In addition, the use of the same elicitation mechanism for the hypothetical and real WTPs and the field survey in the private goods valuation (*Same Mechanism* × *Private*, and *Field Survey*× *Private*) have no statistically significant effects on the HBF.

5.2.4 Other variables impacting the hypothetical bias

Type of comparison

Our findings suggest that the Between-Respondents experimental design significantly reduces the HB (Models 1). This result contradicts the findings by Murphy et al. (2005) and List and Gallet's (2001) that the between-respondents comparison has no significant effect on the HBF. In contrast, List and Gallet (2001) suggested that within-respondents experimental designs are appropriate because they allow researchers to control the large individual specific-effects in the statistical analysis. In model 2, we investigate the impact regarding the type of good. The results indicate a non-statistically significant positive impact when analysing the HB of public goods (0.161) while the impact is negative and not statistically significant in the case of private goods (the sum of the coefficients of Between-Respondents and Between-Respondents*Private is negative and statistically significant, Chi2(1) = 12.33, P-value < 0.001).

Type of experimental survey

The analysis reveals that all stated preference methods do not significantly affect the HBF. Nevertheless, the referendum-type mechanism has a significant effect at the 10% level on the HBF. In addition, we found that the use of the same elicitation mechanism in the hypothetical and actual WTP treatments has no significant effect on the HBF. This result indicates that there is no statistically significant gap between the HB obtained using the same or different elicitation mechanisms for the hypothetical and real WTP treatments, and thus it empirically rejects the hypothesis made by Murphy et al. (2005) in their meta-analysis. Indeed, in their meta-analysis, these authors imposed that the valuation mechanisms of the hypothetical and real WTPs need to be identical to avoid any confusion regarding any effects due to the different mechanisms. However, our results clearly show that the use of the same mechanism in the hypothetical and actual WTP survey-treatments has no significant effect on the HB. The incentive compatible mechanisms (ICMs) insignificantly reduce the HB. This result confirms the predictions of Carson and Groves (2007). However, the interaction variable *ICM* × *Private* good has a positive and statistically significant effect on the HBF, which suggest that incentive compatible mechanisms are effective in reducing the HB in the public good valuation study.

Field surveys increase the HB compared to laboratory experiments. Indeed, the field survey has a positive and statistically significant coefficient on the HBF (Model 1), but at the 10% level. The sign of this coefficient remains positive, but it is not significant in models 2 and 3 (Table 4). List and Gallet (2001) find that laboratory experiments do not have a statistically significant effect on the HBF. On the other hand, the results of Murphy et al.

(2005) indicated that performing the laboratory treatment has a positive and significant effect on the HBF.

Type of survey respondent

The results show that using only *Student* respondents do not have statistically significant effects on the HBF. This contradicts the results of Murphy et al. (2005). These authors found that using only students as the study's participants can be a source of HB.

5.3 Results of Bayesian estimates: MRHME according to Gibbs sampling

We conducted our estimations according to the Bayesian approach, while all previous models were estimated using the classical approach. The prior distributions of parameters and hyper-parameters are set so that they are non-informative. This specification makes it possible to prevent these initial distributions from having an effect on the final or posteriori distributions of the explanatory effects on the HBF (Chib and Carlin, 1999; Martin et al., 2016). The average mean vectors of the effects of the random effects and fixed effects variables are set to zero $\left(b = 0;\ \mu_f = 0\right)$ and the matrices of their variances-covariances are equal to the matrix identities of the *kr* and *kf* $\left(V_r = I_{kr} \text{ et } V_f = I_{kf}\right)$ vector dimensions of the random effects and fixed effects variables, respectively. The parameters of the inverse-gamma distribution (the scale parameter and the slope) for the hyper-parameters are set to 10^{-2}, i.e., $\eta = \kappa = \alpha_{ii} = \gamma_{ii} = 10^{-2}$ (Martin et al., 2016). The number of iterations of the Gibbs sample is eleven thousand iterations. We suppressed the first thousand simulations in order to eliminate the effects of prior distributions and used the last ten thousand iterations to derive the posterior effect distributions and the average effects of the explanatory variables on the HBF. Therefore, the results obtained are the posterior effect distributions and the numerical results such as the average effect, and its standard deviation and its standard error.

These results (Table A1 in Appendix) generally corroborate those obtained with the classical methods. Nevertheless, while the classical estimates find that *Ex Ante* × *Private* and *Same Mechanism* × *Private* do not have statistically significant effects on the HBF, the Bayesian results suggest that the posterior effect of these interaction variables are entirely and negatively distributed. This suggests that, compared to public goods, the use of *Ex Ante calibration* techniques and the *Same Mechanism* design reduce the HB more in private goods' economic valuations (Figure 3).

Figure 3 : Posterior distribution of the effects of HB mitigation approaches and experimental design characteristics

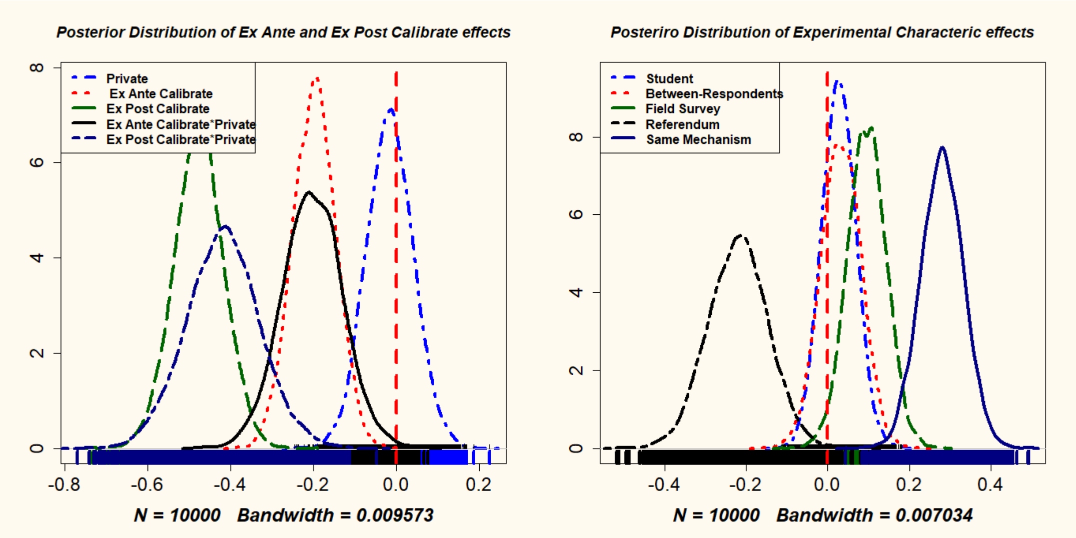
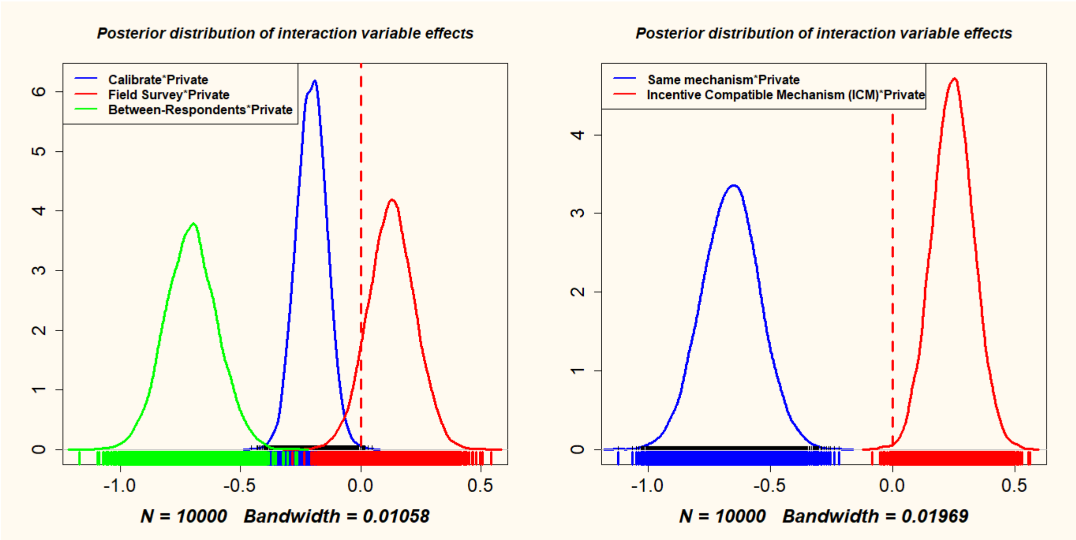


Figure 4 : Posterior distribution of the effects of interaction variables between experimental design characteristics and calibration techniques and private good



The classical estimates indicated that *Same Mechanism* does not have a statistically significant effect on the HBF and *Between-Respondents* significantly reduces the HB. Conversely, the Bayesian results reveal that the posterior effects of *Same Mechanism* and those of *Between-Respondents* are entirely positively distributed and not entirely negatively distributed (Figure 4). These results suggest that the outliers could affect the estimates when using classical estimation and combining classical and Bayesian approaches maybe a good alternative to deal with outliers and ensure results robustness.

We summarize the methodological approaches and the results of the previous meta-analyses and those of the study at hand in Table 5.

Table 5 : Keys Empirical Evidences to Reduce Hypothetical Bias in Economic Valuation with CVM

Study	List and Gallet (2001)	Little and Berrens (2004)	Murphy et al. (2005)	Little et al. (2012)	Foster and Burrows (2017)	Penn and Hu (2018)	Our Empirical Results		
Dependent Variable	Ln (Hypothetical WTP/ Real WTP)	Y = 1 if SS of hypothetical bias, 0 else	Ln (Actual WTP)	Y = 1 if SS of hypothetical bias, 0 else	Ln (Hypothetical WTP/ Real WTP)	Ln (Hypothetical WTP/ Real WTP)	Ln (Hypothetical WTP/ Real WTP)	Ln (Hypothetical WTP/ Real WTP)	Ln (Hypothetical WTP/ Real WTP)
Econometric Models	Log-linear	Probit	Log-linear	Probit	Log-linear and Fixed Effects	Log-linear	Log-linear	MRHME	MRHME
Estimation approach	Classical	Classical	Classical	Classical	Classical	Classical	Classical	Classical	Bayesian
Number of Studies (observations)	29 (58)	53 (85)	28 (77)	96 (220)	78 (432)	132 (908)	87 (462)	87 (462)	87 (462)
Private Good	SS, Less HB	Not SS, Less HB	SS, Less HB	-	SS, More HB	-	Not SS, Less HB	Not SS, Less HB	Not END, Less HB
Public Good	-	-	-	-	-	SS, More HB	-	-	-
Student sample	-	-	SS, More HB	SS, More HB	Not SS, More HB	Not SS, Less HB	SS, Less HB	Not SS, More HB	Not END, Less HB
Within Respondent	Not SS, Less HB	Not SS, Less HB	SS, Less HB	Not SS, More HB	Not SS, More HB	-	-	-	-
Between-Respondent	-	-	-	-	-	Not SS, Less HB	SS, Less HB	SS, Less HB	Not END, Less HB
WTP	SS, Less HB	-	-	-	-	-	-	-	-
WTA	-	Not SS, More HB	-	Not SS, More HB	-	SS, Less HB	-	-	-
Lab setting	Not SS, Less HB	Not SS, More HB	-	Not SS, Less HB	SS, Less HB	Not SS, More HB	-	-	-

Study	List and Gallet (2001)	Little and Berrens (2004)	Murphy et al. (2005)	Little et al. (2012)	Foster and Burrows (2017)	Penn and Hu (2018)	Our Empirical Results		
HB mitigation approaches	-	-	SS, Less HB	SS, Less HB	-	-	SS, Less HB	SS, Less HB	END, Less HB
Choice experiment	-	-	-	Not SS, Less HB	Not SS, Less HB	SS, Less HB	-	-	-
Induced Value	-	-	-	Not SS, Less HB		SS, Less HB	-	-	-
Cheap Talk	-	-	-	-	SS, Less HB	SS, Less HB	SS, Less HB	SS, Less HB	END, Less HB
Certainty follow-up	-	-	-	-	SS, Less HB	SS, Less HB	SS, Less HB	SS, Less HB	END, Less HB
Consequentiality	-	-	-	-	-	SS, Less HB	-	-	-
Ex Ante Calibration	-	-	-	-	-	-	SS, Less HB	SS, Less HB	END, Less HB
Ex Post Calibration	-	-	-	-	-	-	SS, Less HB	SS, Less HB	END, Less HB
Field Survey	-	-	-	-	-	-	Not SS, More HB	SS, More HB	Not END, More HB
Same Mechanism	-	-	-	-	-	-	Not SS, More HB	Not SS, More HB	EPD, More HB
ICM	-	-	-	-	-	-	Not SS, Less HB	SS, Less HB	END, Less HB
Calibrate Private	-	-	-	-	-	-	Not SS, Less HB	SS, Less HB	END, Less HB
Ex Ante Calibrate Private	-	-	-	-	-	-	Not SS, Less HB	Not SS, Less HB	END, Less HB
Ex Post Calibrate Private	-	-	-	-	-	-	Not SS, Less HB	SS, Less HB	END, Less HB

Study	List and Gallet (2001)	Little and Berrens (2004)	Murphy et al. (2005)	Little et al. (2012)	Foster and Burrows (2017)	Penn and Hu (2018)	Our Empirical Results		
ICM Private	-	-	-	-	-	-	Not SS, More HB	SS, More HB	EPD, More HB
Same Mechanism Private	-	-	-	-	-	-	SS, Less HB	Not SS, Less HB	END, Less HB
Field Survey Private	-	-	-	-	-	-	Not SS, Less HB	Not SS, More HB	Not END, More HB
Between-Respondent Private	-	-	-	-	-	-	Not SS, Less HB	SS, Less HB	END, Less HB

Notes: HB indicates "Hypothetical Bias", SS indicates "Statistical Significant", Not SS indicates "Not Statistical Significant" in 50% or more in the appropriate models. "-" indicates the variable was not included in the meta-analysis, WTA indicates "Willingness-To-Accept", WTP indicates "Willingness-To-Pay", Lab indicates "Laboratory", Less indicates "negative Sign", More indicates "Positive Sign", END indicates "Entirely Negation Distribution" of posterior effect, Not END Less indicates "Not Entirely Negative Distribution" of posterior effect, but the posterior distribution is dominated by negative value, Not END, More indicates "Not Entirely Negative Distribution" of posterior effect, but the posterior distribution is dominated by positive value; EPD indicates " Entirely Positive Distribution" of posterior effect; Less indicates "Reduce HB"; and More indicates "Increase HB".

6 Conclusion

This article updates previous meta-analyses with emphasis on the use of new sophisticated models. We estimated a Meta-Regression Hierarchical Mixed Effect (MRHME) model using the classical and Bayesian approaches. In contrast to earlier meta-analysis models, this hierarchical model controls the unobservable effects and heteroscedasticity specific to each study. Several observations may come from the same study, and there may be a correlation between these observations. Therefore, this potential correlation between the intra-study observations and the unobservable and study-specific characteristics that can also affect the results of the estimates introduce heteroscedasticity. We use Bayesian Gibbs sampling procedure approach to estimate the posterior distributions of the effects of the explanatory variables on the HBF. It solves the problems related to the effects of outliers and the low representativeness of certain variables in the sample and tests the robustness of the classically estimated results. The previous meta-analyses did not control for the potential effect of these factors in their empirical estimations.

The results of the likelihood ratio tests show that the use of MRHME models better explain the HB than the log-linear models and indicate that the unobservable characteristics and heteroscedasticity have significant effects on the estimated parameters. Therefore, the use of the Log-linear regression leads to biased results. The MRHME model provides a significant and substantial improvement for the explanation of the HBF.

Results related to HB show that the average of the HBF is 2.112 and its median is 1.41 for the total sample. The econometric estimate results generally indicate that the use of calibration techniques (*Aggregated calibration, Cheap Talk*, *Certainty Correction*, *Ex Ante*

Calibration, and *Ex Post Calibration*), the Between-Respondents design, the Referendum mechanism and incentive compatible mechanisms significantly reduce the HB in WTP estimates with declarative methods. Conversely, *Field Survey* seems to be a source of bias in evaluation studies with declarative methods. The use of the same mechanism in hypothetical and real treatment surveys has no effect on the hypothetical bias. However, compared to public goods, the same mechanism design in the hypothetical and actual WTP treatments, the *Between-Respondents* design and the calibration techniques are more effective and significantly reduce HB in the case of private good valuations with stated preference methods. In addition, the results highlighted that the *incentive compatible mechanisms* is biased in the case of the private goods' economic valuation.

This study contributes to the literature on meta-analyses in economics by demonstrating potential biases associated with the common use of Log-linear regression models. We demonstrated that the use of MRHME model is more appropriate and that Bayesian Gibbs sampling procedure approach could solve the problems related to the effects of outliers. This study also update and highlight the relevant hypothetical bias reduction factors when using stated preference approaches.

References

Adnan, R., Mohamad, M. N., Setan, H. (2003). Multiple outliers detection procedures in linear regression. Matematika, 19, 29-45.

Akter, S., Bennett, J., Akhter, S. (2008). Preference uncertainty in contingent valuation. Ecological Economics, 67(3), 345-351.

Anderson, M. J., Legendre, P. (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. Journal of statistical computation and simulation, 62(3), 271-303.

Andrews, D. F., Pregibon, D. (1978). Finding the outliers that matter. Journal of the Royal Statistical Society. Series B (Methodological), 85-93.

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

Bedate, A. M., Herrero, L. C., Sanz, J. Á. (2009). Economic valuation of a contemporary art museum: correction of hypothetical bias using a certainty question. Journal of Cultural Economics, 33(3), 185-199.

Bergmo, T. S., Wangberg, S. C. (2007). Patients' willingness to pay for electronic communication with their general practitioner. The European Journal of Health Economics, 8(2), 105.

Blomquist, G. C., Blumenschein, K., & Johannesson, M. (2009). Eliciting willingness to pay without bias using follow-up certainty statements: comparisons between probably/definitely and a 10-point certainty scale. Environmental and Resource Economics, 43(4), 473-502.

Blumenschein, K., Blomquist, G. C., Johannesson, M., Horn, N., Freeman, P. (2007). Eliciting willingness to pay without bias: evidence from a field experiment. The Economic Journal, 118(525), 114-137.

Bohm, P. (1972). Estimating Demand for Public Goods: An Experiment. European Economic Review, 3(2), 111–30.

Broadbent, C. D. 2014. "Evaluating Mitigation and Calibration Techniques for Hypothetical Bias in Choice Experiments." Journal of Environmental Planning and Management 57 (12): 1831–48.

Brown, T.C., Ajzen, I., Hrubes, D. (2003). Further tests of entreaties to avoid hypothetical bias in referendum contingent valuation. Journal of environmental Economics and Management, 46(2), 353-361.

Cameron, A. C., Trivedi, P. K. (2010). *Microeconometrics using stata* (Vol. 2). College Station, TX: Stata press.

Capon, N., Farley, J. U., Hoenig, S. (1990). Determinants of financial performance: a meta-analysis. Management science, 36(10), 1143-1159.

Carlsson, F., Martinsson, P. (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments?: Application to the valuation of the environment. Journal of Environmental Economics and Management, 41(2), 179-192.

Carson, R. T., Mitchell, R. C., Hanemann, W. M., Kopp, R. J., Presser, S., Ruud, P. A. (1992). A contingent valuation study of lost passive use values resulting from the Exxon Valdez oil spill (No. 6984). University Library of Munich, Germany.

Carson, R. T., Groves, T. (2007). Incentive and Informational Properties of Preference Questions. Environmental and Resource Economics, 37 (1), 181–210.

Carson, R. T., Groves, T., List, J. A. (2014). Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice. Journal of the Association of Environmental and Resource Economists, 1 (1/2), 171–207.

Champ, P. A., Bishop, R. C., Brown, T. C., McCollum, D. W. (1997). Using Donation Mechanisms to Value Nonuse Benefits from Public Goods. Journal of Environmental Economics and Management, 33(2), 151–62.

Champ, P. A., Moore, R., & Bishop, R. C. (2009). A comparison of approaches to mitigate hypothetical bias. Agricultural and Resource Economics Review, 38(2), 166-180.

Chatterjee, S., Hadi, A.S. (1986). Influential observations, high leverage points, and outliers in linear regression. Statistical Science, 379-393.

Chib, S., Carlin, B. P. (1999). On MCMC sampling in hierarchical longitudinal models. Statistics and Computing, 9(1), 17-26.

Cook, R.D. (1977). Detection of influential observation in linear regression. Technometrics, 19(1), 15-18.

Cummings, R. G., Taylor, L. O. (1999). Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method. The American Economic Review, 89(3), 649–65.

Cummings, R. G., Harrison, G. W., Rutström, E. E. (1995). Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible? The American Economic Review, 85(1), 260–66.

De-Magistris, T., Gracia, A., Nayga Jr, R.M. (2013). On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments. American Journal of Agricultural Economics, 95(5), 1136-1154.

Dekker, T., Brouwer, R., Hofkes, M., Moeltner, K. (2011). The Effect of Risk Context on the Value of a Statistical Life: A Bayesian Meta-Model. Environmental and Resource Economics, 49(4), 597–624.

Doyon, M., Bergeron, S. (2016). Understanding Strategic Behavior and Its Contribution to Hypothetical Bias When Eliciting Values for a Private Good. Canadian Journal of Agricultural Economics, 64(4), 653-666.

Doyon, M., Saulais, L., Ruffieux, B., Bweli D. (2015). Hypothetical Bias for Private

Goods: Does Cheap Talk Make a Difference? Theoretical Economics Letters, 5(6), 749–56.

Ehmke, M.D., Lusk, J.L., List, J.A. (2008). Is hypothetical bias a universal phenomenon? A multinational investigation. Land economics, 84(3), 489-500.

Field, A. (2009). Discovering statistics using SPSS. Thousand Oaks, CA: Sage Publication.

Foster, H., Burrows, J. (2017). 10. Hypothetical bias: a new meta-analysis. Contingent Valuation of Environmental Goods: A Comprehensive Critique, 270.

Johannesson, M., Blomquist, G.C., Blumenschein, K., Johansson, P.O., Liljas, B., O'conor, R.M. (1999). Calibrating hypothetical willingness to pay responses. Journal of Risk and Uncertainty, 18(1), 21-32.

Johnston, R.J. (2006). Is hypothetical bias universal? Validating contingent valuation responses using a binding public referendum. Journal of Environmental Economics and Management, 52(1), 469-481.

Koop, G. (2003). Bayesian Econometrics. Wiley, New York

Krawczyk, M. (2012). Testing for hypothetical bias in willingness to support a reforestation program. Journal of forest economics, 18(4), 282-289.

Lee, J., Hwang, U. (2016). Hypothetical Bias in Risk Preferences as a Driver of Hypothetical Bias in Willingness to Pay: Experimental Evidence. Environmental and Resource Economics, 65(4), 789-811.

List, J. A., Gallet, C. A. (2001). What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values? Environmental and Resource Economics. 20(3), 241–54.

List, J. A., Shogren, J. F. (1998). The deadweight loss of Christmas: comment. The American Economic Review, 88(5), 1350-1355.

Little, J., Berrens, R. (2004). Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta-Analysis. Economics Bulletin, 3(6), 1–13.

Little, J., Broadbent, C. D., Berrens, R. P. (2012). Meta-analysis of the probability of disparity between actual and hypothetical valuation responses: Extension and preliminary new results. In Western Economic Forum, 11(1), 1-12.

Loomis, J. (2011). What's to know about hypothetical bias in stated preference valuation studies?. Journal of Economic Surveys, 25(2), 363-370.

Loomis, J.B. (2014). 2013 WAEA keynote address: Strategies for overcoming hypothetical bias in stated preference surveys. Journal of Agricultural and Resource Economics, 39(1), 34-46.

Loomis, J., Brown, T., Lucero, B., Peterson, G. (1996). Improving Validity Experiments of Contingent Valuation Methods: Results of Efforts to Reduce the Disparity of Hypothetical and Actual Willingness to Pay. Land Economics, 72(4), 450–61.

Loomis, J., Bell, P., Cooney, H., Asmus, C. (2009). A comparison of actual and hypothetical willingness to pay of parents and non-parents for protecting infant health: the case of nitrates in drinking water. Journal of Agricultural and Applied Economics, 41(3), 697-712.

Loomis, J. B., White, D. S. (1996). Economic benefits of rare and endangered species: summary and meta-analysis. Ecological Economics, 18(3), 197-206.

Lusk, J.L., McLaughlin, L., Jaeger, S.R. (2007). Strategy and response to purchase intention questions. Marketing letters, 18(1-2), 31-44.

Martin, A. D., Quinn K. M., Park J. H., and Park, M. J. H. (2016)."Package ‘MCMCpack’."

Melo, P. C., Graham, D. J., Noland, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. Regional science and urban Economics, 39(3), 332-342.

Moeltner, K., Boyle, K. J., Paterson, R. W. (2007). Meta-Analysis and Benefit Transfer for Resource Valuation-Addressing Classical Challenges with Bayesian Modeling. Journal of Environmental Economics and Management, 53(2), 250–69.

Moeltner, K., Woodward, R. (2009). Meta-functional benefit transfer for wetland valuation: making the most of small samples. Environmental and Resource Economics, 42(1), 89-108.

Morris, C. (1995). Hierarchical models for educational data- an overview. Journal of Educational and Behavior Statistics, 20, 190-200.

Morrison, M., & Brown, T. C. (2009). Testing the effectiveness of certainty scales, cheap talk, and dissonance-minimization in reducing hypothetical bias in contingent valuation studies. Environmental and Resource Economics, 44(3), 307-326.

Moser, R., Raffaelli, R., Notaro, S. (2014). Testing Hypothetical Bias with a Real Choice Experiment Using Respondents’ Own Money. European Review of Agricultural Economics, 41(1), 25–46.

Mozumder, P., Berrens, R.P. (2007). Investigating hypothetical bias: induced-value tests of the referendum voting mechanism with uncertainty. Applied Economics Letters, 14(10), 705-709.

Mundfrom, D. J., & Schults, M. R. (2002). A monte carlo simulation comparing parameter estimates from multiple linear regression and hierarchical linear modeling. Multiple Regression Viewpoints, 28, 18-21.

Murphy, J. J., Allen, P. G., Stevens, T. H., Weatherhead, D. (2005a). A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation. Environmental and Resource Economics, 30(3), 313–25.

Murphy, J. J., Stevens, T., Weatherhead, D. (2005b). Is Cheap Talk Effective at Eliminating Hypothetical Bias in a Provision Point Mechanism? Environmental and Resource Economics, 30(3), 327–43.

Murphy, J.J., Stevens, T.H., Yadav, L. (2010). A comparison of induced value and home-grown value experiments to test for hypothetical bias in contingent valuation. Environmental and Resource Economics, 47(1), 111-123.

Neill, H. R., Cummings, R. G., Ganderton, P. T., Harrison, G. W., McGuckin, T. (1994). Hypothetical Surveys and Real Economic Commitments. Land Economics, 70(2),145–54.

Newman, Newman, and Salzman (2010). Comparing OLS and HLM Models and the Questions They Answer: Potential Concerns for Type VI Errors. Multiple Linear Regression Viewpoints, 2010, Vol. 36(1)

NOAA Panel (Arrow, Kenneth, Robert Solow, and others). (1993). Report of the NOAA Panel on Contingent Valuation.

Penn, J. M., Hu, W. (2018). Understanding hypothetical bias: An enhanced meta-analysis. American Journal of Agricultural Economics, 100(4), 1186–1206.

Poe, G. L., Clark, J. E., Rondeau, D., Schulze W. D. (2002). Provision Point Mechanisms and Field Validity Tests of Contingent Valuation. Environmental and Resource Economics, 23(1), 105–31.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage Publications.

Stachtiaris, S., Drichoutis, A., Nayga, R. and Klonaris, S. (2011). Can religious priming induce truthful preference revelation?.

Stanley, T.D., Doucouliagos, H., Giles, M., Heckemeyer, J.H., Johnston, R.J., Laroche, P., Nelson, J.P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R.S. (2013). Meta-analysis of economics research reporting guidelines. Journal of Economic Surveys, 27(2), 390-394.

Swamy, P. (1970). Efficient inference in a random coefficient regression model. Econometrica: Journal of the Econometric Society, 38(2), 311-323.

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th ed.). Boston: Allyn & Bacon.

Vossler, C. A., Ethier, R. G., Poe, G. L., Welsh, M. P. (2003). Payment Certainty in Discrete Choice Contingent Valuation Responses: Results from a Field Validity Test. Southern Economic Journal, 69(4), 886–902.

Vossler, C. A., Evans, M. F. (2009). Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality. Journal of Environmental Economics and Management, 58(3), 338-345.

Vossler, C.A., Doyon, M., Rondeau, D. (2012). Truth in consequentiality: theory and field evidence on discrete choice experiments. American Economic Journal: Microeconomics, 4(4), 145-71.

West, M. (1984). Outlier models and prior distributions in Bayesian linear regression. Journal of the Royal Statistical Society. Series B (Methodological), 46(3), 431-439.

Wisnowski, J.W., Montgomery, D.C., Simpson, J.R. (2001). A comparative analysis of multiple outlier detection procedures in the linear regression model. Computational Statistics Data Analysis, 36(3), 351-382.

Woodward, R. T., Wui, Y. S. (2001). The economic value of wetland services: a meta-analysis. Ecological economics, 37(2), 257-270.

Zawojska, E., Czajkowski, M. (2017). Re-examining empirical evidence on stated preferences: importance of incentive compatibility. Journal of Environmental Economics and Policy, 6(4), 374-403.

Zuur, A.F., Ieno, E.N., Elphick, C.S. (2010). A protocol for data exploration to avoid common statistical problems Methods in Ecology and Evolution, 1(1), 3-14.

7 Appendix

Bayesian Model Specification and Estimation Procedure

According to equation (3), $\ln\left(Y_s \middle| X_{rs}, Z_{fs}\right)$ follows a multivariate normal distribution. We obtain the following likelihood function:

$$p\left(\ln(Y)\middle|X, Z, \beta_r, \beta_f, \sigma^2\right) = \prod_{s=1}^{S} \frac{1}{\left(2\pi\sigma^2\right)^{n_s/2}} \exp\left(-\frac{1}{\sigma^2}\left(\ln(Y_s) - X_{rs}\beta_{rs} - Z_{fs}\beta_{fs}\right)'\left(\ln(Y_s) - X_{rs}\beta_{rs} - Z_{fs}\beta_{fs}\right)\right) \quad (4)$$

The prior distributions of this model concern the parameters and the hyperparameters to be estimated:

$$\begin{aligned}
&(a)\ p\left(\beta_{rs}\middle|b,\Sigma\right) = mvn(b,\Sigma),\\
&(b)\ p(b) = mvn(\mu, V),\\
&(c)\ p\left(\beta_f\right) = mvn\left(\mu_{f,}V_f\right),\\
&(d)\ p\left(\Sigma_{ii}\right) = ig\left(\alpha_{ii}, \gamma_{ii}\right),\\
&(e)\ p\left(\sigma^2\right) = ig(\eta, \kappa).
\end{aligned} \quad (5)$$

The likelihood function (equation (4)) and the prior distributions of parameters (system of equations (5)) define the Bayesian Hierarchical Model of Meta-regression with mixed coefficients. The random coefficient β_{rs} (equation (a)) and fixed coefficient β_f (equation (c)) vectors follow both multivariate Gaussian distributions of respective mean vectors b and μ_f, and respective variance-covariance matrices Σ, and V_f. The theoretical mean vector of the random parameters b is generated by a multivariate Gaussian distribution process of mean μ and variance-covariance matrix V (equation (b)). The variances Σ_{ii} of the random coefficients β_{rs} respectively follow an inverse-Gamma distribution of shape parameter α_{ii}, and scale parameter γ_{ii}. The prior distribution of the variance σ^2 of the error term is an inverse-Gamma distribution of shape parameter η, and scale parameter κ. The choice of this inverse-gamma distribution makes it possible to ensure the positive sign of the terms of variances.

The likelihood function (equation (4)) and the set of prior distributions of the parameters β_{rs}, b, β_f, $\left(\Sigma_{ii}\right)_{i=1,\dots,K}$ and σ^2 (equation (5)) are associated to derive the posterior distributions of the parameters of the mixed-effect meta-regression hierarchical model conditionally to the observation matrix $\left(Y, X_{rs}, Z_{fs}\right)$. We follow the procedure of Moeltner et al. (2007) and Dekker et al. (2011) for the empirical estimation of prior distributions. This procedure of Gibbs sampling described by Koop (2003) is used to empirically derive the posterior distributions of hierarchical Bayesian model parameters. Given the prior distributions (equation (5)), the Gibbs sample procedure allows to draw iteratively $r_1 = 1, 2, ..., R_1$ sets of the parameters of the posterior distributions according to the following steps:

$$\begin{aligned}
&p\left(\beta_{f,r_1} \middle| b_{r_1}, \Sigma_{r_1}, \sigma_{r_1}^2, Y, X_{rs}, Z_{fs}\right),\\
&p\left(\beta_{rs,r_1} \middle| \beta_{f,r_1}, b_{r_1}, \Sigma_{r_1}, \sigma_{r_1}^2, Y, X_{rs}, Z_{fs}\right),\\
&p\left(b_{r_1} \middle| \beta_{rs,r_1}, \Sigma_{r_1}\right),\\
&p\left(\Sigma_{ii,r_1} \middle| \beta_{rs,r_1}, b_{r_1}\right),\ i = 1, 2, ..., K,\\
&p\left(\sigma_{r_1}^2 \middle| \beta_{f,r_1}, b_{r_1}, Y, X_{rs}, Z_{fs}\right)
\end{aligned} \tag{6}$$

This iterative procedure thus makes it possible to generate a set of R_1 sets of parameters of the posterior distributions. By tending R_1 towards infinity, that is to say with a large number of iteration (weak law of large numbers (Koop, 2003)), the conditionally drawn parameters $\left(\beta_f, b, \Sigma, \sigma^2\right)$ will converge towards the posterior distribution $p\left(\beta_f, b, \Sigma, \sigma^2 \middle| Y, X_r, Z\right)$, whose processes generating the series of observations obtained for each parameter at the end of the Gibbs sample simulation are posterior marginal distributions of each of these parameters (Koop 2003, Moeltner et al., 2007, and Dekker et al., 2011):

$$\begin{aligned}
&p\left(\beta_f \middle| Y, X_{rs}, Z_{fs}\right),\\
&p\left(b \middle| Y, X_{rs}, Z_{fs}\right),\\
&p\left(\Sigma \middle| Y, X_{rs}, Z_{fs}\right),\\
&p\left(\sigma^2 \middle| Y, X_{rs}, Z_{fs}\right)
\end{aligned} \tag{7}$$

We use the Markov Chain Monte Carlo (MCMC) method to verify the convergence of the Gibbs sample (Koop 2003, Albert 2009).

Table A 1 : Results of Bayesian estimates of MRHME Models

Variables	Model 1		Model 2		Model 3		Model 4	
	Coef. Mean	SE	Coef. Mean	SE	Coef. Mean	SE	Coef. Mean	SE
Constant	0.176	0.0009	0.403	0.0008	-0.033	0.0011	0.145	0.0009
Private	-0.062	0.0006	0.099	0.0005	1.048	0.0015	-0.015	0.0005
Field Survey	0.083	0.0005	0.037	0.0004	0.022	0.0007	0.095	0.0004
Student	0.036	0.0004	-0.042	0.0004	-0.145	0.0004	0.027	0.0004
Between-Respondents	0.078	0.0005	-0.069	0.0005	0.333	0.0007	0.032	0.0004
Vickrey Auction	0.442	0.0007					0.426	0.0007
MDC	0.163	0.0006					0.141	0.0006
DC	0.194	0.0006					0.274	0.0005
Open-Ended	0.287	0.0006					0.308	0.0006
Referendum	-0.232	0.0007					-0.219	0.0007
Same Mechanism	0.271	0.0005	0.311	0.0005	0.652	0.0008	0.280	0.0005
ICM			-0.11	0.0003	-0.239	0.0005		
Calibrate	-0.385	0.0003			-0.326	0.0004		
Calibrate Ex Ante							-0.198	0.0005
Calibrate Ex Post							-0.477	0.0006
Cheap Talk			-0.291	0.0003				
Certainty Correction			-0.734	0.0005				
Calibrate × Private					-0.204	0.0006		
Calibrate Ex Ante × Private							-0.202	0.0007
Calibrate Ex Post × Private							-0.418	0.0008
Same Mechanism × Private					-0.662	0.0011		
ICM × Private					0.245	0.0008		
Between-Respondents × Private					-0.705	0.001		
Field Survey × Private					0.129	0.0009		
Sigma2	0.259	0.0001	0.262	0.0001	0.252	0.0001	0.240	0.0001

Table A 2: Meta-studies description

Authors	Publication year	Respondents	Type of experience	Calibration techniques	FBH (Min-Meam-Max)
Alfnes and al. (2010)	2010	University staff	Laboratory	Cheap talk, Real talk	1.28 - 1.69 - 2.72
Arana and Leon (2013)	2013	consumers	Laboratory		0.73 - 1.01 - 1.20
Balistreri and al. (2001)	2001	Students	Laboratory		1.25 - 1.25 - 1.25
Bergmo and Wangberg (2007)	2007	Patients	Field survey		1.50 - 1.50 -1.50
Bhatia and Fox-Rushby (2010)	2010	Households	Field survey		0.94 - 0.94 - 0.94
Blomquist and al. (2009)	2009	Patients	Field survey	Certainty correction	0.47 - 1.47 - 3.68
Blumenschein and al. (1997)	1997	Students	Laboratory		3.69 - 7.71 -11.74
Blumenschein and al. (2008)	2008	Patients	Field survey	Certainty correction; Cheap talk	0.77 - 1.53 - 4.10
Burchardi and al. (2005)	2005	Consumers	Field survey		1.21 - 1.33 - 1.44
Burton and al. (2007)	2007	Students	Laboratory		1.14 - 1.31 - 1.51
Camacho-Cuena and al. (2004)	2004	Consumers	Laboratory		1.04 - 1.04 - 1.04
Chowdhury and al. (2011)	2011	Consumers	Field survey	Cheap talk	1.03 - 2.25 - 4.72
Cummings and al. (1995)	1995	Students, Non-students	Laboratory		2.56 - 4.93 - 10.50
De-Magistris and al. (2013)	2013	Consumers	Field survey	Cheap talk, Honesty	0.75 - 1.14 - 1.50
Dicky and al. (1987)	1987	Households	Field survey		1.15 - 1.15 - 1.15
Doyon and al. (2015)	2015	Consumers	Laboratory	Cheap talk	1.40 - 1.41 - 1.43
Fox and al. (1998)	1998	Households	Phone survey		0.86 - 0.96 - 1.05
Frykblom (1997)	1997	Students	Laboratory		1.50 - 1.60 - 1.71
Frykblom (2010)	2010	Students	Laboratory		1.32 - 1.73 - 2.13

Authors	**Publication year**	**Respondents**	**Type of experience**	**Calibration techniques**	**FBH (Min-Meam-Max)**
Grebitus and al. (2013)	2013	consumers	Laboratory		1.13 - 1.55 - 1.97
Heberlein and Bishop (1986)	1986	Hunters	Survey by mail		1.24 - 1.61 - 2.26
Johannesson (1997)	1996	Students	Laboratory		1.63 - 1.63 - 1.63
Johannesson and al. (1997)	1997	Students	Laboratory		1.02 - 1.02 - 1.02
Johannesson and al. (1999)	1999	Students	Laboratory	Certainty correction	0.81 - 2.04 - 8.50
Johennesson and al. (2010)	2010	Students	Laboratory	Certainty correction	0.52 - 1.73 - 8.01
Kealy and al. (1988)	1988	Students	Field survey		1.01 - 1.13 - 1.41
List (2001)	2001	Merchants; Non- merchants	Field survey	Cheap talk	1.02 - 1.67 - 1.95
List (2003)	2003	merchants; Non- merchants	Field survey	Cheap talk	0.75 - 1.96 - 3.15
List and Shorgren (1998) 1995	1998	Consumers; Retailers	Field survey		2.18 - 2.73 - 3.47
List and Shorgren (1998) 1998	1998	Student	Laboratory		0.61 - 0.80 - 1.00
Loomis and al. (1997)	1996	University staff	Laboratory		1.95 - 2.80 - 3.64
Loomis and al. (1997)	1997	University staff	Laboratory		1.86 - 2.20 - 2.55
Loomis and al. (2009)	2009	Households	Mixed survey (mail and field)		7.05 - 7.06 - 7.07
Morkbar and al. (2014)	2014	Consumers	Field survey	Cheap talk	0.59 - 0.76 - 1.15
Moser and al. (2014)	2014	Consumers	Field survey	Cheap talk; Own money	0.14 - 1.85 - 2.88
Murphy and al. (2010)	2010	Students	Laboratory		0.99 - 1.39 - 2.13
Neill et al. (1994)	1994	Students	Laboratory		3.10 - 10.27 - 27.42
Paradiso and Trisorio (2001)	2001	Students	Laboratory		2.79 - 3.13 - 3.46
Silva et al. (2007)	2007	Adult buyers	Field survey		1.08 - 1.21 - 1.40

Authors	**Publication year**	**Respondents**	**Type of experience**	**Calibration techniques**	**FBH (Min-Meam-Max)**
Silva et al. (2011)	2011	Consumers	Field survey	Cheap talk	0.93 - 1.08 - 1.26
Silva et al. (2012)	2012	Adult Buyers	Field survey	Cheap talk	0.89 - 1.05 - 1.21
Stachtiaris et al. (2011)	2011	Students	Laboratory	Religion prime	1.04 - 1.19 - 1.41
Stefani and Scarpa (2009)	2009	Students	Laboratory		0.76 - 1.43 - 2.45
Taylor et al. (2010)	2010	Students	Field survey		4.98 - 5.05 - 5.11
Volinskiy et al. (2011)	2011	Consumers	Laboratory		0.70 - 2.33 - 4.16
Alpizar et al. (2008)	2008	tourists	Field survey		1.94 - 3.10 - 5.25
Barrage and Lee (2010)	2010	General	Laboratory	Cheap talk, Explicit consequence	0.53 - 1.54 - 2.59
Botelho and Pinto (2002)	2002	Students	Laboratory		11.51 - 11.51 - 11.51
Broadbent (2013)	2013	Students	Laboratory	Certainty correction, Cheap talk	0.49 - 0.78 - 1.06
Broadbent et al. (2010)	2010	Students	Laboratory	Explicit consequence	1.01 - 1.22 - 1.47
Brown et al. (1996)	1996	Households	Survey by mail		1.50 - 3.94 - 8.25
Brown et al. (2003)	2003	Students	Laboratory	Cheap talk	0.78 - 1.52 - 2.86
Caplan et al. (2010)	2010	Students	Laboratory		1.17 - 1.61 - 2.14
Carlson et Martinsson (2001)	2001	Students	Laboratory		1.13 - 1.13 - 1.13
Champ et Bishop (2009)	2009	Residents	Survey by mail	Certainty correction, Cheap talk	0.50 - 1.36 - 3.24

Authors	Publication year	Respondents	Type of experience	Calibration techniques	FBH (Min-Meam-Max)
Christie (2007)	2007	visitors	Field survey		1.28 - 2.34 - 3.40
Commigs and Taylor (1999)	1999	Students	Laboratory	Cheap talk	0.88 - 1.25 - 1.68
Elmke et al. (2008)	2008	Students	Laboratory		0.55 - 1.11 - 1.56
Getzner (2000)	2000	Students	Laboratory		2.67 - 3.50 - 4.33
Jacquemet et al. (2011)	2011	Students	Laboratory	Cheap talk	3.12 - 4.17 - 5.85
Jacquemet et al. (2013)	2013	Students	Laboratory	Honesty	0.98 - 0.98 - 0.98
Johansson-Stenman and Svedsader (2008)	2008	Students	Laboratory		1.08 - 2.45 - 3.82
Johnston (2006)	2006	Households	Survey by mail	Explicit consequence	1.06 - 1.06 - 1.06
Krawczyk (2012)	2012	Mixed	Laboratory		1.37 - 1.45 - 1.52
Lee and Hwang (2015)	2015	Students	Laboratory	Cheap talk	1.74 - 2.59 - 3.30
Letry and List (2007)	2007	Students	Field survey	Cheap talk, Explicit consequence	0.97 - 1.91 - 3.95
List et al. (2006)	2006	Residents	Survey by mail	Cheap talk	0.65 - 1.54 - 3.23
Mitani and Flores (2009)	2009	Mixed	Laboratory		0.98 - 0.98 - 0.98
Morrison and Brown (2009)	2009	Students	Laboratory	Certainty correction, Cheap talk	0.61 - 0.98 - 1.51
Mozumder and Berrens (2007)	2007	Students	Laboratory	Cheap talk	0.97 - 1.03 - 1.17
Murphy et al. (2003)	2003	Students	Laboratory	Cheap talk	4.77 - 6.17 - 7.57
Murphy et al. (2005)	2005	Students	Laboratory	Cheap talk	2.44 - 4.80 - 7.20
Murphy et al. (2010)	2010	Students	Laboratory		0.95 - 1.21 - 1.63
Poe et al.(2002)	2002	Households	Phone survey		1.19 - 1.34 - 1.50
Ready et al. (2010)	2010	Students	Laboratory		3.15 - 3.15 - 3.15
Seip and Stret (1992)	1992	Adults	Field survey		10.61 - 10.61 - 10.61

Authors	Publication year	Respondents	Type of experience	Calibration techniques	FBH (Min-Meam-Max)
Sinden (1988)	1988	Students	Field survey		0.76 - 0.94 - 1.14
Spencer et al. (1998)	1998	Students	Laboratory		0.77 - 2.53 - 4.67
Stefani and Scarpa (2009)	2009	Students	Laboratory		0.72 - 0.93 - 1.07
Stevens et al. (2013)	2013	Students	Laboratory	Honesty	0.96 - 1.08 - 1.19
Swardh (2008)	2008	Students	Laboratory	Certainty correction	0.75 - 1.85 - 3.50
Taylor (1998)	1998	Students	Laboratory		1.44 - 1.44 - 1.44
Taylor et al. (2010)	2010	Students	Field survey		1.55 - 2.17 - 4.12
Veisten and Narvud (2006)	2006	Residents	Survey by mail		1.78 - 5.79 - 13.38
Vossler and Evans (2009)	2009	Students	Laboratory	Explicit consequence	0.86 - 1.24 - 1.65
Vossler and Kerkvliet (2003)	2003	Adult Residents	Survey by mail		1.010 - 1.01 - 1.013
Vossler and Watson (2013)	2013	Registered voters	Survey by mail	Explicit consequence	0.79 - 0.98 - 1.16

Figure A1 : Posterior Distribution of the Effects of Good and Experimental Design Characteristics

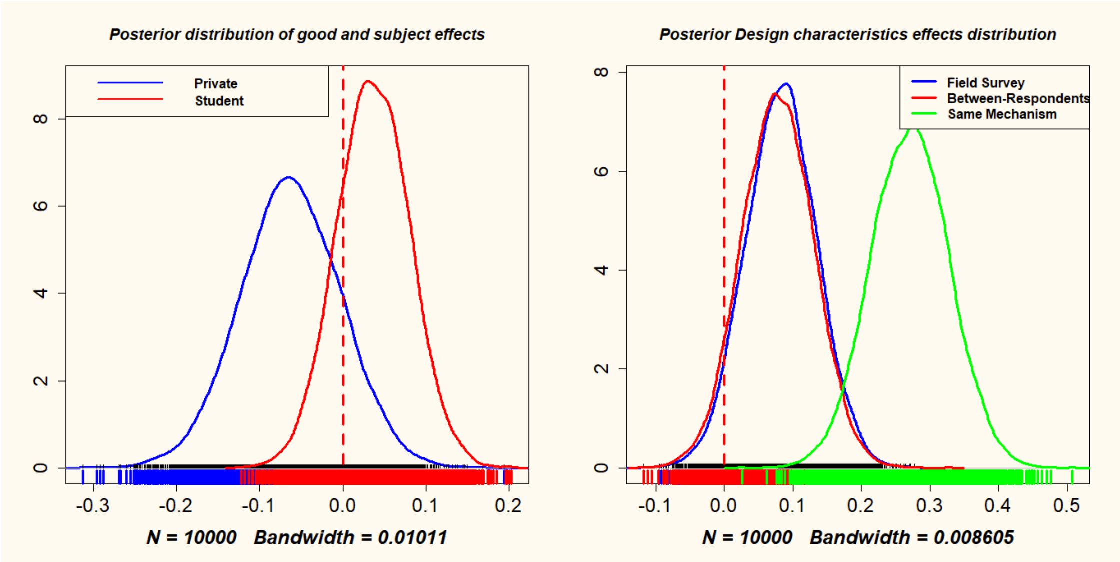


Figure A2 : Posterior Distribution of the Effects of Stated Preferences Methods and HB Calibration Techniques

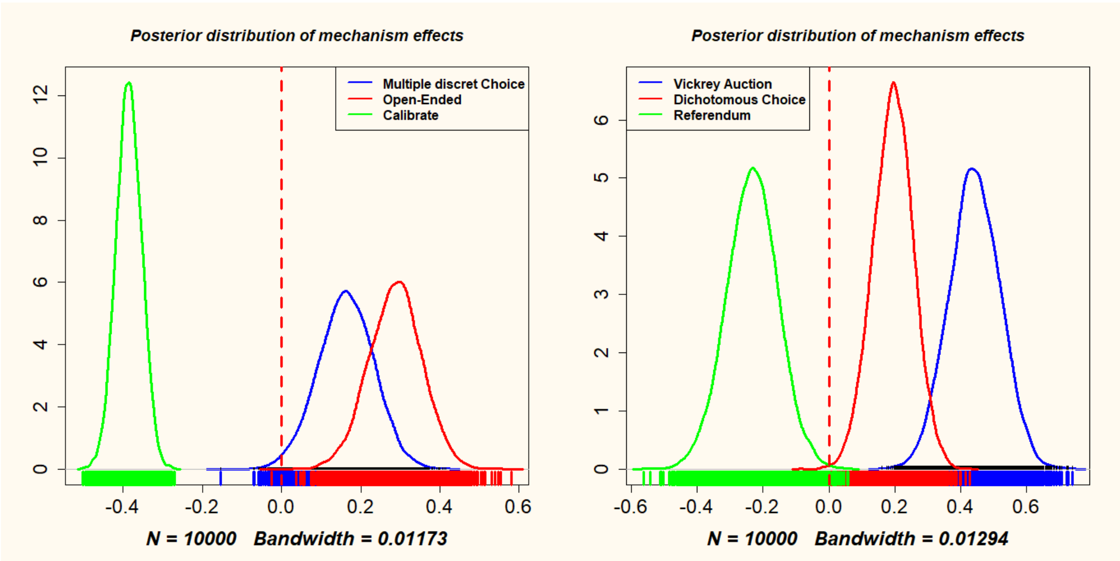


Figure A3 : Posterior Distribution of the Effects of Incentive Compatible Mechanism (ICM) and HB Mitigation Approaches

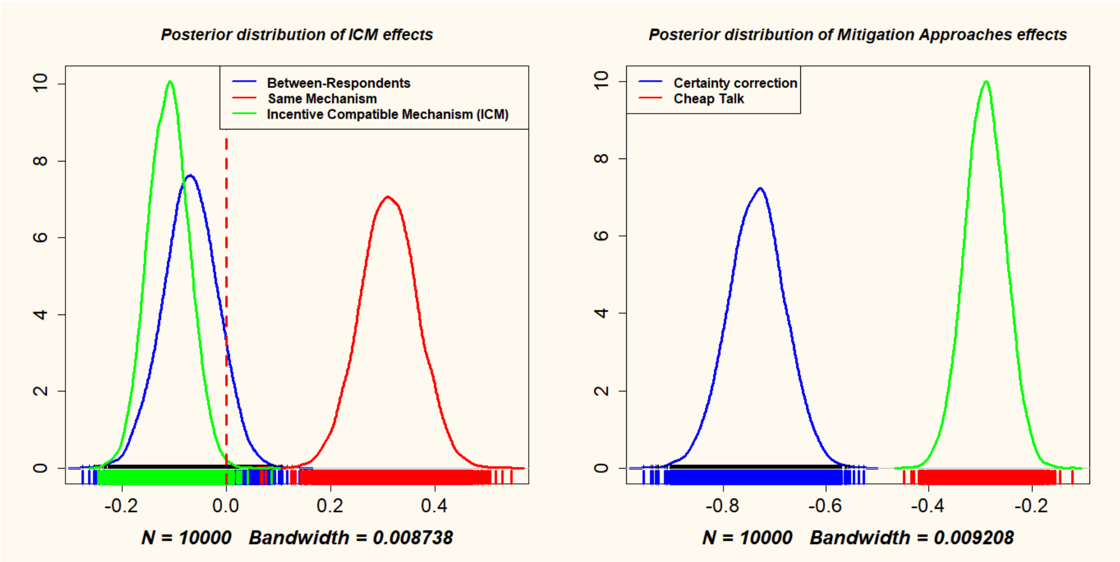


Figure A4 : Posterior Distribution of the Effects of Interaction Variables Between Experimental Design Characteristics and Calibration Techniques and Private Good

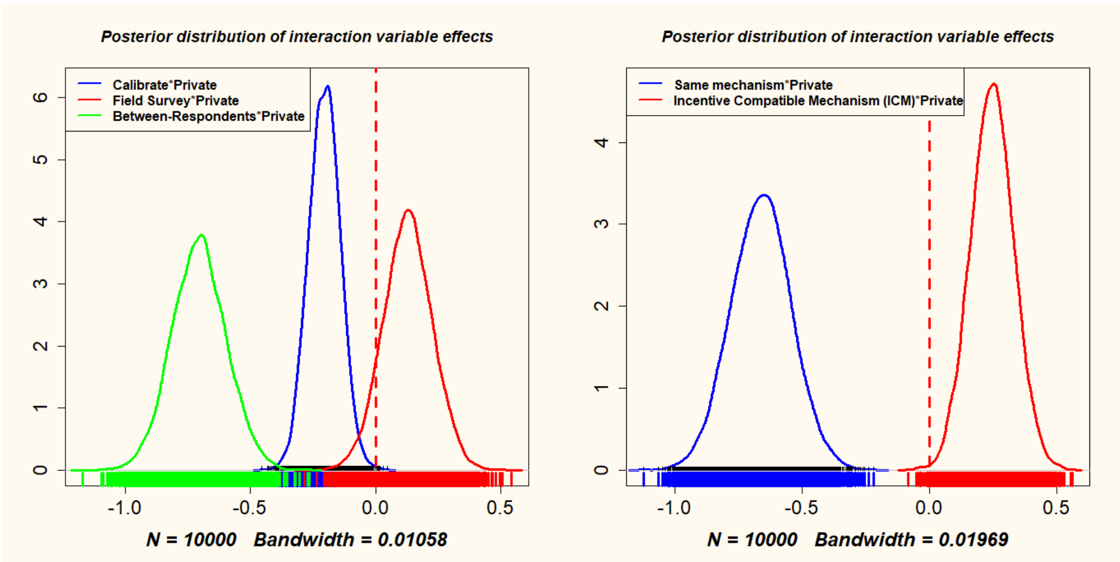


Figure A5: Posterior Distribution of the Effects of HB Mitigation Approaches (Ex Ante and Ex Post)