Journal of Agricultural and Resource Economics 45(2):376–395 ISSN: 1068-5502 (Print); 2327-8285 (Online)

doi: 10.22004/ag.econ.302460

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

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

Using a set of 462 observations from 87 public and private goods economic valuation studies, this study reviews and updates meta-analyses on hypothetical bias using a metaregression hierarchical mixed-effect (MRHME) model that corrects the effects of the unobservable characteristics, within-study error correlation, and potential heteroskedasticity specific to each study. The findings indicate that the MRHME model is more efficient than the log-linear models used in previous meta-analyses. Moreover, this modeling approach and the use of interaction variables by type of goods highlight significant differences relative to previous meta-analyses in the explanatory variables' effects, significance levels, magnitudes, and signs.

Key words: contingent valuation, economic valuation, private goods, public goods, willingness-to-pay

Introduction

Strong demand to assess the public's preferences for environmental and ecological goods' production, ecosystem services, forest restoration, and endangered species protection has spurred the demand for stated preference studies, including contingent valuations (Carson et al., 1992; Johnston, 2006; Vossler and Evans, 2009; Murphy, Stevens, and Yadav, 2010; Krawczyk, 2012; Lee and Hwang, 2016). Stated preferences have also been used for private goods to assess consumers' willingness-to-pay (WTP) for new products or innovative attributes (Bergmo and Wangberg, 2007; Loomis et al., 2009; Moser, Raffaelli, and Notaro, 2014; Doyon et al., 2015; Doyon and Bergeron, 2016). The literature emphasizes that stated preference methods potentially lead to hypothetical bias (HB) (Bohm, 1972; Arrow et al., 1993; Cummings, Harrison, and Rutström, 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.

The extensive literature on HB has in turn generated several meta-analyses, with mixed results regarding the factors contributing to HB (List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005; Zawojska and Czajkowski, 2017; Foster and Burrows, 2017; Penn and Hu, 2018). One hypothesis to explain the lack of consensus among these studies could be the adequacy of the type of meta-analysis models concerning the characteristics of the metadata. Specifically, previous models did not consider unobservable characteristics, intrastudy correlation, and interstudy heteroskedasticities.This article updates previous meta-analyses using the meta-regression hierarchical mixed-effect (MRHME) model, which corrects for the effects of unobservable characteristics, within-study error correlation, and potential heteroskedasticity specific to each study

Baoubadi Atozou (corresponding author) is an economist on the Resilience Analysis and Policy Team at the Food and Agriculture Organization of the United Nations. Lota D. Tamini is a professor in the Department of Agricultural Economics and Consumer Science at Laval University. Stéphane Bergeron is a senior researcher in the Department of Agricultural Economics and Consumer Science at Laval University. Maurice Doyon is a professor in the Department of Agricultural Economics and Consumer Science at Laval University.

Review coordinated by Richard T. Woodward.

(Moeltner, Boyle, and Paterson, 2007; Dekker et al., 2011). Comparing the log-linear benchmark models with the MRHME models using the likelihood ratio test reveals that the latter models are more efficient. In this paper, we discuss this modeling approach as well as how the inclusion of interaction variables by type of goods (private vs. public) provides new insights on the factors that most contribute to HB.

Previous Meta-Analyses

The presence of HB is a problem in contingent valuation estimates (Bohm, 1972; Carson et al., 1992; Arrow et al., 1993; Penn and Hu, 2018). Many studies using contingent valuation methods (CVM) for economic valuation of goods have highlighted that respondent-stated WTPs are significantly different from their real WTPs (Neill et al., 1994; Cummings, Harrison, and Rutström, 1995; Loomis et al., 1996; Champ et al., 1997; Cummings and Taylor, 1999; Vossler et al., 2003; Brown, Ajzen, and Hrubes, 2003; Murphy et al., 2005; Blumenschein et al., 2008). Several studies have devised ways to eliminate or reduce HB using calibration techniques such as certainty correction (Champ et al., 1997), cheap talk (Cummings and Taylor, 1999), perceived consequentiality (Carson and Groves, 2007; Vossler, Doyon, and Rondeau, 2012), honesty priming (de Magistris, Gracia, and Nayga, 2013), and religious priming (Stachtiaris et al., 2011). However, there is no consensus in the literature on the determinants of HB, the understanding of efficiency of calibration techniques, or the means by which these calibration techniques operate.

We have identified six main meta-analyses that 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, Broadbent, and Berrens (2012); Foster and Burrows (2017); and Penn and Hu (2018). Table 1 presents the key results and econometric models used for each of these studies.

List and Gallet (2001), Murphy et al. (2005), and Penn and Hu (2018) find that private goods reduce HB. By contrast, Little and Berrens (2004) reveal that the type of good has no significant effect on the probability of observing HB, while Foster and Burrows (2017) find that private goods increase HB. Previous results also diverge regarding the effects of the within-respondents, lab setting, and student variables on the magnitude of the HB.

The adequacy of the econometric models with the structure of metadata, some unobservable characteristics, and intrastudy potential heteroskedasticities may explain the mixed results. Several observations may come from the same study, and these observations may be correlated. Thus, according to Moeltner, Boyle, and Paterson (2007) and Dekker et al. (2011), it is likely that heteroskedascity will be observed due to this within-study correlation (Moeltner, 2019). Moreover, unobservable characteristics intrinsic to each study can also affect the results of the estimates (Rosenberger and Loomis, 2000; Johnston, Besedin, and Ranson, 2006; Moeltner, Boyle, and Paterson, 2007; Moeltner and Woodward, 2009; Johnston and Rosenberger, 2010; Moeltner, 2019). The omitted variables causes the endogeneity in ordinary least squares (OLS) regression if the omitted variables are correlated with other independent variables considered in the model (Kim and Frees, 2006), which will cause partial bias and inconsistency of the OLS estimators (Boardman and Murnane, 1979; Ehrenberg and Brewer, 1994; Kim and Frees, 2006). Simply ignoring within-study correlation, potential variable omission bias,[1] and heteroskedasticity would cast doubt on the reliability of standard errors for estimated coefficients and associated confidence intervals

[1] The omission of relevant predictor variables causes bias because it induces a correlation between the disturbance term and the explanatory variables (Wooldridge, 2002; Ebbes, Bockenholt, and Wedel, 2004; Kim and Frees, 2006; Cameron and Trivedi, 2010; Clark and Linzer, 2015).

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
No. 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
Cheap talk	-	-	-	-	SS, Less HB	SS, Less HB
Certainty follow-up	-	-	-	-	SS, Less HB	SS, Less HB
Consequentiality	-	-	-	-	-	SS, Less HB

Notes: This table updates (**?**, Table 1) by adding their keys results. HB indicates hypothetical bias, SS indicates statistically Significant, Not SS indicates not statistically significant in 50% or more in the appropriate models. "-" indicates that 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.

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

HBF	Mean	Median	SD	CV	N
Full Sample	2.11	1.41	2.44	0.86	462
Calibrationa	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 subsample of observations using calibration techniques.

(Moeltner and Woodward, 2009).[2] Blakely and Woodward (2000) and Bell, Fairbrother, and Jones (2019) indicate that statistical analysis that ignores the multilevel nature of the dataset may bias the standard error. Therefore, to consistently estimate the standard error, separate random error terms may be specified for each level of analysis. Random error terms may also be included for the individual-level coefficients. Previous meta-analyses on HB did not control these aspects nor the problem of low frequency of certain characteristics in the database that may affect the results in the estimation of their model. The MRHME model controls for those aspects to produce consistent results (Moeltner, Boyle, and Paterson, 2007; Moeltner and Woodward, 2009).

Finally, 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.

Data Description

Selection Criteria

We adopt two inclusion criteria for studies in our meta-analysis. First, we include studies that reported both average hypothetical WTP (WTP_h) and revealed WTP (WTP_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 WTP surveys. Following Murphy et al. (2005), we excluded willingness-to-accept studies, which are rarely used. However, unlike Murphy et al. (2005), we did not exclude studies that estimated the hypothetical and real WTPs using different survey mechanisms. Instead, we created a variable (*Same Mechanism*) to detect the potential effect of using different elicitation survey effects on HB.

We included the papers that were used in the first four meta-analyses (List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005; Little, Broadbent, and Berrens, 2012). In addition, following the suggestions of Littell, Corcoran, and Pillai (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 to include the studies not considered in these previous meta-analyses. We obtained 87 studies, including 44 studies about private goods and 43 about public goods. Appendix Table A1 summarizes the selected studies.

Dependent Variable: Hypothetical Bias Factor (HBF)

We use the average hypothetical WTP (WTPh) and revealed WTP (WTPr) to obtain our dependent variable, the hypothetical bias factor $\left(HBF = \frac{MeanWTP_h}{MeanWTP_r}\right)$. If HBF is equal to 1, then HB is 0.

Table 2 presents the descriptive statistics of the HBF. The average of 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 obtained by using mitigation techniques is 32.61% of the full sample (Table 3). The results obtained with these mitigation techniques seem to be on average more accurate than those obtained without them. In fact,

[2] Bell, Fairbrother, and Jones (2019) argue that estimates from models that incorrectly assume within-effect homogeneity will suffer from bias; this bias can be more easily avoided by explicitly modeling such heterogeneity, with the inclusion of random slopes (Western, 1998).

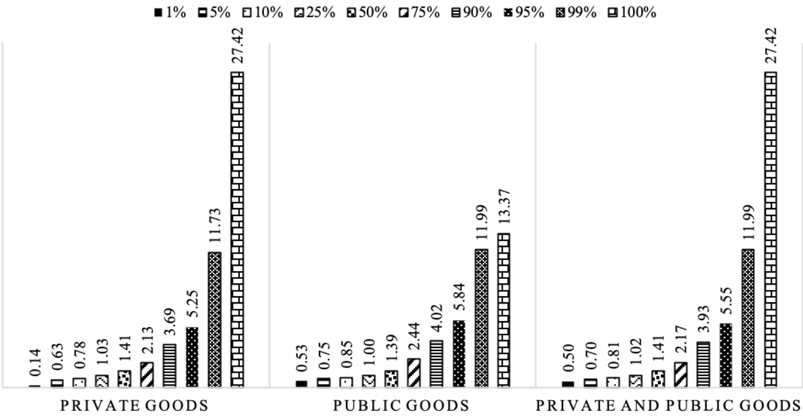

Figure 1. Percentile Distribution of Hypothetical Bias Factor (HBF)

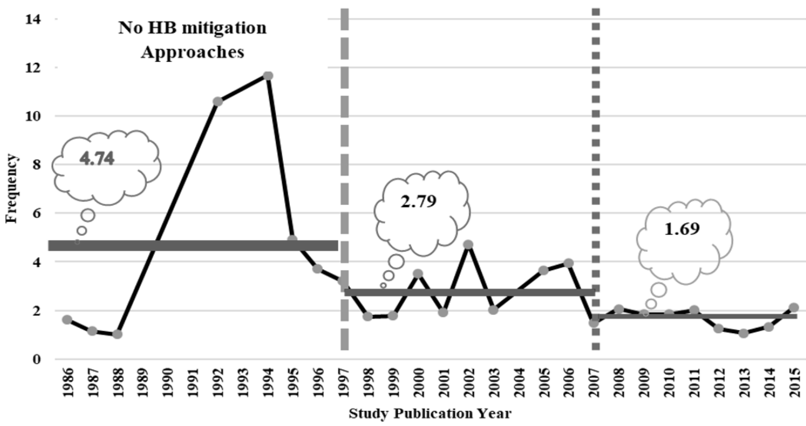

Figure 2. Dynamic of the Average of Bypothetical Bias Factor (HBF) by Publication Year and Different Periods of Key HB Mitigation

average HBF for the mitigation techniques subsample is 1.42 compared with 2.52 in the subsample without HB mitigation 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 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. The timeline of improvement corresponds to the development and use of mitigation techniques such as cheap talk, certainty correction, and perceived consequentiality.

Explanatory Factors

Table 3 describes the variables and reports their statistics.

Table 3. Variable Descriptions

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
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 if the mechanism of real experience is the same as that of the hypothetical experiment, and 0 otherwise	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			
Hypothetical Bias Factor (HBF)	Ratio of hypothetical and real WTP (WTP_h/WTP_r)	463	n/a
No. of obs.	Total observations	463	n/a

Calibration Techniques

We use *Cheap Talk* and *Certainty Correction* as binary explanatory variables. These take a value of 1 if applied in the hypothetical survey treatment and 0 otherwise. Perceived consequentiality (Carson and Groves, 2007; Vossler, Doyon, and Rondeau, 2012), honesty priming (de Magistris, Gracia, and Nayga, 2013), and religious priming (Stachtiaris et al., 2011) have also been used to reduce HB. For different models, we aggregate these mitigation techniques as well as the previously described into a single binary *Calibration* variable, which is equal to 1 if a mitigation technique is used, and 0 otherwise.

In different instances, the calibration variables are aggregated into two variables. First, an *Ex Ante Calibration* technique variable is set to 1 if cheap talk, honesty, or religious priming is used as the mitigation technique, and 0 otherwise. Second, an *Ex Post Calibration* variable takes a value of 1 if certainty correction or perceived consequentiality is used to calibrate the stated preference methods, and 0 otherwise.

Other Variables

List and Gallet (2001) have shown that different elicitation mechanisms have potentially different effects on HB. Binary variables are used for elicitation mechanisms. The variable *Same Mechanism* takes a value of 1 if the same survey treatment is used for the hypothetical and real WTP, and 0 otherwise.

The preferences of an economic agent may differ depending on whether the good is a public or private good. The binary variable *Private* is set to 1 if the economic valuation study is for a private good, and 0 otherwise. Carson and Groves (2007) suggest that respondents reveal their real WTPs when the treatment survey experiment is consequential, even in a hypothetical treatment. The incentive-compatible mechanism indicator variable (*ICM*) takes a value of 1 if the WTP is estimated using an ICM such as dichotomous choice, referendum, Vickrey auction, *n*th price auction, or Becker–DeGroot–Marschak method (BDM). As previous meta-analysis, we introduce *Open-Ended* as explanatory variable that takes a value of 1 if an open-ended mechanism is used and 0 otherwise.

Students are broadly used as subjects in economic valuation studies in experimental economics (see, e.g., Ehmke, Lusk, and List, 2008; Vossler and Evans, 2009; Lee and Hwang, 2016). The variable *Student* takes a value of 1 if the study's subjects are entirely student subjects, and 0 otherwise.

Econometric Model

We adopt the MRHME model used by Moeltner, Boyle, and Paterson (2007) and Dekker et al. (2011) because it (i) addresses the study-specific heteroskedasticity by random parameter specifications (Moeltner, Boyle, and Paterson, 2007) and (ii) controls for the effects of unobservable characteristics. We assign fixed coefficients to the explanatory factors that do not have sufficient interstudy variability to allow for random coefficient specifications (Moeltner, Boyle, and Paterson, 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 intrastudy variability of the HBF related to the experimental design, the WTP assessment methods, and the interstudy variability of the HBF related to the unobservable factors related to each study. These factors may lead to heteroskedasticity, which is related to the methodological features (Koop, 2003, ch. 6). To solve this problem, Moeltner, Boyle, and Paterson (2007) propose 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:

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

where *mvn* and *n* represent the multivariate and univariate normal distributions, respectively. The vectors $M_{r,ijs}$ and $B_{f,ijs}$ are the characteristics of the method and 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 subvectors of the vector of coefficients that are respectively associated with the explanatory regressors of 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 variance–covariance matrix σ. The stochastic error term also follows according to equation (1) as a normal distribution with 0 mean and variance σ^2. The logarithmic transformation of equation (1) gives the following expression of the metaregression 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 (M_{ijs}) 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 (β_{rs}) and the stochastic error term (ε) implies that the HBF vector of the study, $\left(\ln\left(Y_{ijs} \left| X_{r,ijs}, Z_{ijs}\right.\right)\right)$, follows a multivariate normal distribution. Thus, the statistical inference of our variable of interest is estimated by the following equations (Dekker et al., 2011):

$$\begin{aligned} &\ln\left(Y_s \left| X_{rs}, Z_{fs}\right.\right) = X_{rs}\beta_{rs} + Z_{fs}\beta_{fs} + \varepsilon_s \text{ with,} \\ &E\left[\ln\left(Y_s \left| X_{rs}, Z_{fs}\right.\right)\right] = X_{rs}b + Z_{fs}\beta_{fs} \text{ and} \\ &E\left[\ln\left(Y_s\right)\ln\left(Y_t\right)'\right] = \begin{cases} lcX_{rs}\Sigma X'_{rs} + \sigma^2 I_{n_s} & \text{s = t} \\ 0 & \text{if not} \end{cases}. \end{aligned} \tag{3}$$

The dimensions 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 $(n_s \times n_s)$. 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 heteroskedasticity (Moeltner, Boyle, and Paterson, 2007; Dekker et al., 2011). According to Swamy (1970) and Moeltner, Boyle, and Paterson (2007), the estimation of the MRHME model under the normality hypothesis with random coefficients has desirable properties. First, it corrects for heteroskedasticity. 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 intrastudy observations, both via the regressors included in the matrix X_{rs} and via the unobserved elements common to all observations for a given study through the random intercept. The MRHME model increases the efficiency of the model and avoids erroneous estimations of the standard error compared to models that treat all variables as independent (Moeltner, Boyle, and Paterson, 2007). Newman, Newman, and Salzman (2010) show that MRHME models are more appropriate when variables are nested and have intraclass or intrastudy observation correlation (Moeltner, Boyle, and Paterson, 2007; Dekker et al., 2011; Moeltner, 2019). In addition, Tabachnick and Fidell (2007) and Field (2009), among others, have indicated that the hierarchical model is superior to OLS because it theoretically produces appropriate error terms that control for potential dependency due to nesting effects. We estimate the MRHME and conduct the likelihood ratio test to check its efficiency versus OLS. Given that the choice of a mitigation approach can be subjective, we introduce random slopes for the calibration techniques.

We estimate four models. Model 1 includes the 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 regarding the type of good and 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). Model 4 includes *ICM* (dichotomous choice, referendum, Vickrey auction, *n*th price auction or a BDM procedure), *Cheap Talk*, and *Certainty Correction* as explanatory variables.

Log-Linear Models versus MRHME Models

We estimate the MRHME models using the maximum likelihood method (Table 4). The overall significance test of the model (Wald test) shows that all the models are significant and valid at the 1% level: model 1 (χ^2 (11) 56.27, p-value <0.01), model 2 (χ^2 (12) 71.98, p-value <0.01), model 3 (χ^2 (14) 87.25, p-value <0.01) and model 4 (χ^2 (8) 74.43, p-value <0.01).

The results of the likelihood ratio test show that the four MRHME models explain the HB better than the log-linear models do (Cameron and Trivedi, 2010).[3] 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. These results indicate that the unobservable characteristics and heteroskedasticity have significant effects on the estimated parameters. Therefore, the use of the log-linear regression leads to potentially biased results. The MRHME model provides improvement for the explanation of the HBF.

Estimation Results of MRHME Models

Calibration Techniques and Hypothetical Bias

The results of model 1 (Table 4) show that calibration technique has statistically significant negative effects at the 1% level, confirming 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* calibration techniques (*Certainty Correction* and *Perceived Consequentiality*) reduce the HB, as indicated by the estimates of model 3. In models 2 and 3, we investigated the effect of the interaction between calibration techniques and the type of good (*Calibrate* × *Private*).[4] Compared to public goods, calibration techniques are more effective in reducing the HB for private goods (see model 2). The coefficient of the variable *Calibrate* × *Private* is negative and statistically significant. In addition, the results of model 3 indicate that the interaction variable between the *ex post* calibration techniques and the private good (*Ex Post Calibration* × *Private*) has a negative and significant effect on the HBF.

In model 4, we introduced *Cheap Talk* and *Certainty Correction* as explanatory variables. Our results show that the *Cheap Talk* and *Certainty Correction* calibration techniques are effective in reducing the HB. The effect of the *Certainty Correction* technique (−0.644) is larger in magnitude than the effect of *Cheap Talk* (−0.285).

[3] Appendix Table A2 reports the results of the log-linear models.
[4] *Calibrate* variable represents the *Calibration* variable.

Table 4. Classical Estimation Results of MRHME Models (*N* = 460)

	Model 1		Model 2		Model 3		Model 4	
Variables	**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**
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				
Ex Ante Calibrate					−0.220***	0.091		
Ex Post Calibrate					−0.339***	0.109		
Cheap Talk							−0.285***	0.066
Certainty Correction							−0.644***	0.090
Calibrate × *Private*			−0.193**	0.108				
Ex Ante Calibrate × *Private*					−0.133	0.127		
Ex Post Calibrate × *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				
Random Effects								
Sd(*Calibrate*)	0.146	0.077	0.099	0.098				
Sd(*Ex Ante Calibrate*)					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
Wald test	ddl (11)		dd(12)		ddl(14)		ddl(8)	
χ^2 (ddl)	56.27		71.98		87.25		74.43	
p-value	<0.001		<0.001		<0.001		<0.001	
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	
χ^2 test ddl	2		2		3		3	
χ^2 statistic (LR test)	162.09		187.31		153.82		178.56	
p-value	<0.001		<0.001		<0.001		<0.001	
Convergence tolerance (1e-10)	Yes		Yes		Yes		Yes	

Notes: Single, double, and triple asterisks (*, **, ***) indicate significance at the 10%, 5%, and 1% levels, respectively. Dichotomous choice (DC), multiple discrete choice (MDC), incentive compatible mechanism (ICM).

Contingent Valuation Methods: Incentive-Compatibility Mechanisms

Model 4 reveals that *ICM* does not seem to be a factor on HB. To account for the different possible impacts regarding the type of goods, investigated the effect of the interaction between ICM and private good (*ICM*× *Private*). Our estimated results (model 2) indicate a negative effect that is statistically significant for public goods (−0.308), while the impact is positive and statistically significant for private goods.

Type of Goods

The result of model 1 shows that the type of good (*Private*) has no statistically significant direct effect on the HBF. This result contrasts the result of Foster and Burrows (2017), who found that private goods significantly increase the HB, and those of List and Gallet (2001), Murphy et al. (2005), and Penn and Hu (2018), who concluded that the HB increases when the evaluated good is a public good.

The results in models 2–4 show that using the between-respondents design and mitigation techniques (*Between-Respondents* × *Private*, *Calibrate* × *Private*) reduce HB for private goods. In contrast, the use of ICM in the valuation of private goods (*ICM* × *Private*) appears to be ineffective in reducing the HB. In addition, the use of the same elicitation mechanism for the hypothetical and real WTPs and the field survey in the private good valuation (*Same Mechanism* × *Private* and *Field Survey* × *Private*) have no statistically significant effects on the HBF.

Other Variables Impacting the Hypothetical Bias

Our findings suggest that the between-respondents experimental design significantly reduces the HB (model 1). This result contradicts the findings by Murphy et al. (2005) and List and Gallet (2001) that the between-respondent comparison has no significant effect on the HBF. In model 2, we investigate the impact regarding the type of good (*Between-Respondents* × *Private*). The results indicate a nonstatistically significant positive impact for public goods (0.161), while the impact is negative and not statistically significant in the case of private goods.

The analysis reveals that the type of elicitation technique does not significantly affect the HBF. Nevertheless, the referendum-type mechanism has a significant effect on the HBF at the 10% level. In addition, we find 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. Thus, we empirically reject the hypothesis made by Murphy et al. (2005) 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.

As in List and Gallet (2001), the impact of field surveys is not statistically significant at the 5% level and less (models 1–3) when compared to laboratory experiments. In contrast, Murphy et al. (2005) indicated that performing the laboratory treatment has a positive and significant influence on the HBF.

The results show that using only *Student* respondents has no statistically significant effect 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.

We summarize the methodological approaches, the results of the previous meta-analyses and our main results in Table 5.

As indicated in the previous section, *Private*, *Student*, *Vickrey Auction*, *Open-Ended*, and *Same Mechanism* have no significant effects on HB. Similarly, our results suggest that compared to public goods, mitigation techniques (*Calibrate*) and between-respondents design (*Between-Respondents*) significantly reduce the HBF in the WTP valuation for private goods. The interaction variables of the *ex post* calibration techniques and the private good and the incentive-compatible mechanisms have statistically significant effects on the HBF. It should be noted that the effects of these variables are not statistically significant with log-linear models (Appendix Table A1). Moreover, the results indicate that the log-linear models overestimate the magnitude of the effects of the calibration technique

Table 5. Key Empirical Evidence 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 otherwise	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)
Econometric Models	Log-linear	Probit	Log-linear	Probit	Log-linear and Fixed Effects	Log-linear	Log-linear	MRHME
Estimation approach	Classical	Classical	Classical	Classical	Classical	Classical	Classical	Classical
No. of Studies (observations)	29 (58)	53 (85)	28 (77)	96 (220)	78 (432)	132 (908)	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
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
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
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	-	-	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	-	-
Cheap Talk	-	-	-	-	SS, Less HB	SS, Less HB	SS, Less HB	SS, Less HB
Certainty follow-up	-	-	-	-	SS, Less HB	SS, Less HB	SS, Less HB	SS, Less HB
Consequentiality	-	-	-	-	-	SS, Less HB	-	-
Ex Ante Calibration	-	-	-	-	-	-	SS, Less HB	SS, Less HB
Ex Post Calibration	-	-	-	-	-	-	SS, Less HB	SS, Less HB
Field Survey	-	-	-	-	-	-	Not SS, More HB	SS, More HB
Same Mechanism	-	-	-	-	-	-	Not SS, More HB	Not SS, More HB
ICM	-	-	-	-	-	-	Not SS, Less HB	SS, Less HB
Calibrate Private	-	-	-	-	-	-	Not SS, Less HB	SS, Less HB
Ex Ante Calibrate Private	-	-	-	-	-	-	Not SS, Less HB	Not SS, Less HB
Ex Post Calibrate Private	-	-	-	-	-	-	Not SS, Less HB	SS, Less HB
ICM Private	-	-	-	-	-	-	Not SS, More HB	SS, More HB
Same Mechanism Private	-	-	-	-	-	-	SS, Less HB	Not SS, Less HB
Field Survey Private	-	-	-	-	-	-	Not SS, Less HB	Not SS, More HB
Between-Respondent Private	-	-	-	-	-	-	Not SS, Less HB	SS, Less HB

variables compared to the MRHME models (see, e.g., the magnitudes of the effects of the *Calibrate*, *Cheap Talk*, *Certainty Correction*, *Ex Ante Calibrate*, and *Ex Post Calibrate* variables in Table 4 and Appendix Table A1).

Conclusion

Contingent valuation methods (CVM) are widely used in the economic valuation of environmental goods and services and for private goods with new attributes. However, CVM are likely to suffer from hypothetical bias (HB), creating unreliable estimates on which to build environmental, public health, or business policy decisions. Using meta-analysis of the HB literature can help identify factors that can improve CVM estimates and stated preference methodologies. We estimated an MRHME model using the classical approach. In contrast to earlier meta-analysis models, this hierarchical model controls for the unobservable effects, within-study error correlation, and heteroskedasticity specific to each study. 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 explains the HB than the log-linear models and indicate that the unobservable characteristics and heteroskedasticity have significant effects on the estimated parameters. Therefore, the use of the log-linear regression leads to potentially biased results. The MRHME model provides a significant improvement for the explanation of the HBF.

Results related to HB show that the average of the HBF is 2.11 and its median is 1.41 for the total sample. The results generally indicate that the use of calibration techniques, the between-respondents design, the referendum mechanism, and—in some instances—incentive-compatible mechanisms significantly reduce the HB in WTP estimates with declarative methods. Conversely, the use of the same mechanism in hypothetical and real treatment surveys was not found to affect HB. However, between-respondents design and calibration techniques significantly reduce HB in the case of private good valuations with stated preference methods. An unexpected result is that the use of incentive-compatible (IC) mechanisms for private goods significantly increases HBF. A possible explanation is that using an IC auction in a hypothetical setting would create cognitive confusion

Foster and Burrows (2017) and Penn and Hu (2018) use log-linear and fixed effects models and ignore other important variable in their estimation. While Foster and Burrows find that private goods significantly increase HB, Penn and Hu (2018) find that public goods significantly increase HB. Our study highlights that the type of good does not have significant effect on HB. Moreover, relative to our results, Foster and Burrows (2017) seem to overestimate the effect of the cheap talk and certainty correction variables, while Penn and Hu (2018) seem to underestimate the effect of cheap talk and overestimate the effect of certainty correction on HB. In addition, the introduction of the interactions variables in this study produce relevant information on HB, which could eventually help reduce the size of HB in economic evaluation for ecological goods and services and new private good attributes using stated preferences.

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 demonstrate that the use of MRHME mode is more appropriate for meta-analysis with observations coming from the same research teams. One should note that numerous other specifications for the MRHME are possible. Variants could be a different hierarchical distribution, a different mix of interaction terms, higher-order interaction, or different functional forms, to name a few. Finally, this study also contributes to the understanding of the relevant HB reduction factors when using stated preference methodologies.

[First submitted March 2019; accepted for publication September 2019.]

References

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 58(1993):4602–4614.

Bell, A., M. Fairbrother, and K. Jones. "Fixed and Random Effects Models: Making an Informed Choice." *Quality & Quantity* 53(2019):1051–1074. doi: 10.1007/s11135-018-0802-x.

Bergmo, T. S., and S. C. Wangberg. "Patients' Willingness to Pay for Electronic Communication with Their General Practitioner." *European Journal of Health Economics* 8(2007):105–110. doi: 10.1007/s10198-006-0014-5.

Blakely, T. A., and A. J. Woodward. "Ecological Effects in Multi-Level Studies." *Journal of Epidemiology & Community Health* 54(2000):367–374. doi: 10.1136/jech.54.5.367.

Blumenschein, K., G. C. Blomquist, M. Johannesson, N. Horn, and P. Freeman. "Eliciting Willingness to Pay without Bias: Evidence from a Field Experiment." *Economic Journal* 118(2008):114–137. doi: 10.1111/j.1468-0297.2007.02106.x.

Boardman, A. E., and R. J. Murnane. "Using Panel Data to Improve Estimates of the Determinants of Educational Achievement." *Sociology of Education* 52(1979):113–121. doi: 10.2307/2112449.

Bohm, P. "Estimating Demand for Public Goods: An Experiment." *European Economic Review* 3(1972):111–130. doi: 10.1016/0014-2921(72)90001-3.

Brown, T. C., I. Ajzen, and D. Hrubes. "Further Tests of Entreaties to Avoid Hypothetical Bias in Referendum Contingent Valuation." *Journal of Environmental Economics and Management* 46(2003):353–361. doi: 10.1016/S0095-0696(02)00041-4.

Cameron, A. C., and P. K. Trivedi. *Microeconometrics Using Stata*, vol. 2. College Station, TX: Stata Press, 2010.

Carson, R. T., and T. Groves. "Incentive and Informational Properties of Preference Questions." *Environmental and Resource Economics* 37(2007):181–210. doi: 10.1007/s10640-007-9124-5.

Carson, R. T., R. C. Mitchell, W. M. Hanemann, R. J. Kopp, S. Presser, and P. A. Ruud. "A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill." MPRA Paper 6984, University Library of Munich, Munich, Germany, 1992.

Champ, P. A., R. C. Bishop, T. C. Brown, and D. W. McCollum. "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods." *Journal of Environmental Economics and Management* 33(1997):151–162. doi: 10.1006/jeem.1997.0988.

Clark, T. S., and D. A. Linzer. "Should I Use Fixed or Random Effects?" *Political Science Research and Methods* 3(2015):399–408. doi: 10.1017/psrm.2014.32.

Cummings, R. G., G. W. Harrison, and E. E. Rutström. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?" *American Economic Review* 85(1995):260–266.

Cummings, R. G., and L. O. Taylor. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review* 89(1999):649–665. doi: 10.1257/aer.89.3.649.

de Magistris, T., A. Gracia, and R. M. Nayga. "On the Use of Honesty Priming Tasks to Mitigate Hypothetical Bias in Choice Experiments." *American Journal of Agricultural Economics* 95(2013):1136–1154. doi: 10.1093/ajae/aat052.

Dekker, T., R. Brouwer, M. Hofkes, and K. Moeltner. "The Effect of Risk Context on the Value of a Statistical Life: a Bayesian Meta-model." *Environmental and Resource Economics* 49(2011):597–624. doi: 10.1007/s10640-011-9456-z.

Doyon, M., and S. Bergeron. "Understanding Strategic Behavior and Its Contribution to Hypothetical Bias When Eliciting Values for a Private Good." *Canadian Journal of Agricultural Economics* 64(2016):653–666 . doi: 10.1111/cjag.12114.

Doyon, M., L. Saulais, B. Ruffieux, and D. Bweli. "Hypothetical Bias for Private Goods: Does Cheap Talk Make a Difference?" *Theoretical Economics Letters* 05(2015):749–756.

Ebbes, P., U. Bockenholt, and M. Wedel. "Regressor and Random-Effects Dependencies in Multilevel Models." *Statistica Neerlandica* 58(2004):161–178. doi: 10.1046/j.0039-0402.2003.00254.x.

Ehmke, M. D., J. L. Lusk, and J. A. List. "Is Hypothetical Bias a Universal Phenomenon? A Multinational Investigation." *Land Economics* 84(2008):489–500. doi: 10.3368/le.84.3.489.

Ehrenberg, R. G., and D. J. Brewer. "Do School and Teacher Characteristics Matter? Evidence from High School and beyond." *Economics of Education Review* 13(1994):1–17.

Field, A. *Discovering Statistics Using SPSS*. Thousand Oaks, CA: SAGE, 2009, 3rd ed.

Foster, H., and J. Burrows. "Hypothetical Bias: A New Meta-Analysis." In D. McFadden and K. Train, eds., *Contingent Valuation of Environmental Goods: A Conprehensive Critique,* Cheltenham, UK: Edward Elgar, 2017, 270–291.

Johnston, R. J. "Is Hypothetical Bias Universal? Validating Contingent Valuation Responses Using a Binding Public Referendum." *Journal of Environmental Economics and Management* 52(2006):469–481. doi: 10.1016/j.jeem.2005.12.003.

Johnston, R. J., E. Y. Besedin, and M. H. Ranson. "Characterizing the Effects of Valuation Methodology in Function-Based Benefits Transfer." *Ecological Economics* 60(2006):407–419. doi: 10.1016/j.ecolecon.2006.03.020.

Johnston, R. J., and R. S. Rosenberger. "Methods, Trends and Controversies in Contemporary Benefit Transfer." *Journal of Economic Surveys* 24(2010):479–510. doi: 10.1111/j.1467-6419.2009.00592.x.

Kim, J.-S., and E. W. Frees. "Omitted Variables in Multilevel Models." *Psychometrika* 71(2006):659–690. doi: 10.1007/s11336-005-1283-0.

Koop, G. *Bayesian Econometrics*. New York, NY: Wiley, 2003.

Krawczyk, M. "Testing for Hypothetical Bias in Willingness to Support a Reforestation Program." *Journal of Forest Economics* 18(2012):282–289. doi: 10.1016/j.jfe.2012.07.003.

Lee, J., and U. Hwang. "Hypothetical Bias in Risk Preferences as a Driver of Hypothetical Bias in Willingness to Pay: Experimental Evidence." *Environmental and Resource Economics* 65(2016):789–811. doi: 10.1007/s10640-015-9926-9.

List, J. A., and C. A. Gallet. "What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values?" *Environmental \& Resource Economics* 20(2001):241–254. doi: 10.1023/A:1012791822804.

Littell, J., J. Corcoran, and V. Pillai. *Systematic Reviews and Meta-Analysis*. Oxford, UK: Oxford University Press, 2008.

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

Little, J., C. D. Broadbent, and R. P. Berrens. "Meta-Analysis of the Probability of Disparity between Actual and Hypothetical Valuation Responses: Extension and Preliminary New Results." *Western Economics Forum* 11(2012):1–12. doi: 10.22004/ag.econ.176522.

Loomis, J., P. Bell, H. Cooney, and C. Asmus. "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(2009):697–712.

Loomis, J., T. Brown, B. Lucero, and G. Peterson. "Improving Validity Experiments of Contingent Valuation Methods: Results of Efforts to Reduce the Disparity of Hypothetical and Actual Willingness to Pay." *Land Economics* 72(1996):450–461. doi: 10.2307/3146908.

Moeltner, K. "Bayesian Nonlinear Meta Regression for Benefit Transfer." *Journal of Environmental Economics and Management* 93(2019):44–62. doi: 10.1016/j.jeem.2018.10.008.

Moeltner, K., K. J. Boyle, and R. W. Paterson. "Meta-Analysis and Benefit Transfer for Resource Valuation-Addressing Classical Challenges with Bayesian Modeling." *Journal of Environmental Economics and Management* 53(2007):250–269. doi: 10.1016/j.jeem.2006.08.004.

Moeltner, K., and R. Woodward. "Meta-Functional Benefit Transfer for Wetland Valuation: Making the Most of Small Samples." *Environmental and Resource Economics* 42(2009):89–108. doi: 10.1007/s10640-008-9205-0.

Moser, R., R. Raffaelli, and S. Notaro. "Testing Hypothetical Bias with a Real Choice Experiment Using Respondents' Own Money." *European Review of Agricultural Economics* 41(2014):25–46. doi: 10.1093/erae/jbt016.

Murphy, J. J., P. G. Allen, T. H. Stevens, and D. Weatherhead. "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation." *Environmental \& Resource Economics* 30(2005):313–325. doi: 10.1007/s10640-004-3332-z.

Murphy, J. J., T. H. Stevens, and L. Yadav. "A Comparison of Induced Value and Home-Grown Value Experiments to Test for Hypothetical Bias in Contingent Valuation." *Environmental and Resource Economics* 47(2010):111–123. doi: 10.1007/s10640-010-9367-4.

Neill, H. R., R. G. Cummings, P. T. Ganderton, G. W. Harrison, and T. McGuckin. "Hypothetical Surveys and Real Economic Commitments." *Land Economics* 70(1994):145–154. doi: 10.2307/3146318.

Newman, D., I. Newman, and J. Salzman. "Comparing OLS and HLM Models and the Questions They Answer: Potential Concerns for Type VI Errors." *Multiple Linear Regression Viewpoints* 36(2010):1–8.

Penn, J. M., and W. Hu. "Understanding Hypothetical Bias: An Enhanced Meta-Analysis." *American Journal of Agricultural Economics* 100(2018):1186–1206. doi: 10.1093/ajae/aay021.

Rosenberger, R. S., and J. B. Loomis. "Panel Stratification in Meta-Analysis of Economic Studies: An Investigation of Its Effects in the Recreation Valuation Literature." *Journal of Agricultural and Applied Economics* 32(2000):459–470. doi: 10.1017/S1074070800020563.

Stachtiaris, S., A. Drichoutis, R. Nayga, and S. Klonaris. "Can Religious Priming Induce Truthful Preference Revelation?" MPRA Paper 34433, University Library of Munich, Munich, Germany, 2011.

Stanley, T., H. Doucouliagos, M. Giles, J. H. Heckemeyer, R. J. Johnston, P. Laroche, J. P. Nelson, M. Paldam, J. Poot, G. Pugh, R. S. Rosenberger, and K. Rost. "Meta-Analysis of Economics Research Reporting Guidelines." *Journal of Economic Surveys* 27(2013):390–394. doi: 10.1111/joes.12008.

Swamy, P. A. V. B. "Efficient Inference in a Random Coefficient Regression Model." *Econometrica* 38(1970):311–323. doi: 10.2307/1913012.

Tabachnick, B. G., and L. S. Fidell. *Using Multivariate Statistics*. Boston, MA: Allyn & Bacon, 2007, 5th ed.

Vossler, C. A., M. Doyon, and D. Rondeau. "Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments." *American Economic Journal: Microeconomics* 4(2012):145–171. doi: 10.1257/mic.4.4.145.

Vossler, C. A., R. G. Ethier, G. L. Poe, and M. P. Welsh. "Payment Certainty in Discrete Choice Contingent Valuation Responses: Results from a Field Validity Test." *Southern Economic Journal* 69(2003):886. doi: 10.2307/1061656.

Vossler, C. A., and M. F. Evans. "Bridging the Gap Between the Field and the Lab: Environmental Goods, Policy Maker Input, and Consequentiality." *Journal of Environmental Economics and Management* 58(2009):338–345. doi: 10.1016/j.jeem.2009.04.007.

Western, B. "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach." *American Journal of Political Science* 42(1998):1233–1259. doi: 10.2307/2991856.

Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2002.

Zawojska, E., and M. Czajkowski. "Re-Examining Empirical Evidence on Stated Preferences: Importance of Incentive Compatibility." *Journal of Environmental Economics and Policy* 6(2017):374–403. doi: 10.1080/21606544.2017.1322537.

Appendix

Table A1. Description of Meta-Studies

Authors	Pub. Year	Respondents	Type of Experience	Calibration Techniques	FBH (Min-Mean-Max)
Alfnes et 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 et 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 et al. (2009)	2009	Patients	Field survey	Certainty correction	0.47 - 1.47 - 3.68
Blumenschein et al. (1997)	1997	Students	Laboratory		3.69 - 7.71 -11.74
Blumenschein et al. (2008)	2008	Patients	Field survey	Certainty correction; Cheap talk	0.77 - 1.53 - 4.10
Burchardi et al. (2005)	2005	Consumers	Field survey		1.21 - 1.33 - 1.44
Burton et al. (2007)	2007	Students	Laboratory		1.14 - 1.31 - 1.51
Camacho-Cuena et al. (2004)	2004	Consumers	Laboratory		1.04 - 1.04 - 1.04
Chowdhury et al. (2011)	2011	Consumers	Field survey	Cheap talk	1.03 - 2.25 - 4.72
Cummings et al. (1995)	1995	Students, Non-students	Laboratory		2.56 - 4.93 - 10.50
De-Magistris et al. (2013)	2013	Consumers	Field survey	Cheap talk, Honesty	0.75 - 1.14 - 1.50
Dicky et al. (1987)	1987	Households	Field survey		1.15 - 1.15 - 1.15
Doyon et al. (2015)	2015	Consumers	Laboratory	Cheap talk	1.40 - 1.41 - 1.43
Fox et 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
Grebitus et 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 et al. (1997)	1997	Students	Laboratory		1.02 - 1.02 - 1.02
Johannesson et al. (1999)	1999	Students	Laboratory	Certainty correction	0.81 - 2.04 - 8.50
Johennesson et al. (2010)	2010	Students	Laboratory	Certainty correction	0.52 - 1.73 - 8.01
Kealy et 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 et al. (1997)	1996	University staff	Laboratory		1.95 - 2.80 - 3.64
Loomis et al. (1997)	1997	University staff	Laboratory		1.86 - 2.20 - 2.55

Continued on next page...

Table A1. – continued from previous page

Authors	Pub. Year	Respondents	Type of Experience	Calibration Techniques	FBH (Min-Mean-Max)
Loomis et al. (2009)	2009	Households	Mixed survey (mail and field)		7.05 - 7.06 - 7.07
Morkbar et al. (2014)	2014	Consumers	Field survey	Cheap talk	0.59 - 0.76 - 1.15
Moser et al. (2014)	2014	Consumers	Field survey	Cheap talk; Own money	0.14 - 1.85 - 2.88
Murphy et 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
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
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

Continued on next page...

Table A1. – continued from previous page

Authors	Pub. Year	Respondents	Type of Experience	Calibration Techniques	FBH (Min-Mean-Max)
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
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

Table A2. Results of Log-Linear Models

	Model 1		Model 2		Model 3		Model 4	
Variables	**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**
Constant	0.471***	0.152	0.666***	0.127	0.498***	0.184	0.450***	0.148
Private	−0.113*	0.066	0.027	0.63	0.391	0.072	−0.064	0.071
Field Survey	−0.011	0.073	0.016	0.074	0.100	0.128	0.020	0.073
Student	−0.057	0.068	−0.092	0.065	−0.167**	0.072	−0.045	0.068
Between-Respondent	−0.065	0.090	−0.230***	0.085	−0.085	0.118	−0.106	0.088
Vickrey Auction	0.405***	0.131					0.392***	0.127
MDC	0.096	0.109					0.090	0.106
DC	0.111	0.095					0.255**	0.096
Open-Ended	0.228*	0.118					0.243**	0.113
Referendum	−0.229*	0.121					−0.195*	0.116
Same Mechanism	0.230**	0.093	0.186**	0.086	0.355***	0.137	0.202**	0.091
ICM			−0.029	0.068	−0.082	0.113		
Calibrate	−0.355***	0.063			−0.354***	0.092		
Calibrate Ex Ante							−0.225**	0.112
Calibrate Ex Post							−0.588***	0.124
Cheap Talk			−0.314***	0.079				
Certainty Correction			−0.741***	0.112				
Calibrate × Private					−0.046	0.127		
Calibrate Ex Ante × Private							−0.174	0.156
Calibrate Ex Post × Private							−0.283	0.133
Same Mechanism × Private					−0.374**	0.183		
ICM × Private					0.108	0.148		
Between-Respondent × Private					−0.133	0.180		
Field Survey × Private					−0.147	0.159		
No. of obs.	460		460		460		460	
Adjusted R^2	15.2%		13.07%		11.98%		19.26%	
F-statistics	(7.35, *p*-value<0.001)		(8.48, *p*-value<0.001)		(5.07, *p*-value <0.001)		(8.82, P<0.001)	

Notes: Single, double, and triple asterisks (*, **, ***) indicate significance at the 10%, 5%, and 1% levels, respectively. Dichotomous choice (DC), multiple discrete choice (MDC), incentive compatible mechanism (ICM).