

Water Resources Research®

RESEARCH ARTICLE

10.1029/2023WR034803

Modeling Quality and Price Perception in the Choice of Drinking Water in France: A Hybrid Choice Model Approach



Key Points:

- Using survey data, we model water resources quality (mis)-perception and water price (mis)-perceptions by means of a hybrid choice model
- We show that people who are more likely to consider water quality as “very poor” or “poor” are less likely to drink tap water
- We find that people who fail to report the correct price difference between tap and bottled water are more likely to drink bottled water

Olivier Beaumais¹  and Romain Crastes dit Sourd² 

¹University of Rouen Normandy, LERN UR 4702, Rouen, France, ²Centre for Decision Research, Leeds University Business School, Management Department, Leeds, UK

Correspondence to:

O. Beaumais,
olivier.beaumais@univ-rouen.fr

Citation:

Beaumais, O., & Crastes dit Sourd, R. (2024). Modeling quality and price perception in the choice of drinking water in France: A hybrid choice model approach. *Water Resources Research*, 60, e2023WR034803. <https://doi.org/10.1029/2023WR034803>

Received 2 MAR 2023

Accepted 2 FEB 2024

Abstract The water resources literature usually discards the important price difference between bottled water and tap water as a predictor of drinking water choice. In France, bottled water is about 100 times more expensive than tap water. Using 4,003 survey responses, we model water resources quality (mis)-perception and water price (mis)-perceptions by means of a hybrid choice model. We show that respondents who are more likely to consider the quality of water resources as “very poor” or “poor” are less likely to drink tap water. Furthermore, we find that respondents who do not report the correct price difference between tap water and bottled water are more likely to drink bottled water, which is a novel finding, as significant price effects of this type have never been reported in the literature on drinking water choice.

1. Introduction

In this paper, we propose to explore the relationship between individual perception of water resources quality, water price, and drinking water choices (tap water, filtered water, bottled water) using a hybrid choice model. To this end, we use data from a unique survey conducted in France in 2013 involving 4,003 respondents. Respondents were asked to state their perception of the quality of water resources (surface water and groundwater) and to report their main choice of drinking water. The survey also featured questions on how people perceive the price difference between tap water and bottled water. Such data are usually not available to analysts and allow us to propose a more complete model than those usually found in the empirical literature on drinking water choices, integrating both *the perception of water resource quality* and *the perception of the price difference between bottled and tap water*.

In line with other findings, we find that people who are more likely to consider the quality of water resources in France to be “very poor” or “poor” are less likely to drink tap water and more likely to drink bottled water. Additionally, we find that respondents who are less likely to report the correct price difference between tap water and bottled water are also more likely to drink bottled water. This finding has never been reported in the literature to our knowledge, and indicates that price perception affects drinking water choices. We hypothesize that previous studies did not find such an effect because they did not account for “water price mis-perception” (Brent & Ward, 2019). In other words, consumers are typically unaware of how much they *actually* pay for tap water. However, in practice, as perception (or mis-perception) are latent constructs, they can only be approximated by a set of attitudinal (or perceptual) indicators derived from answers to attitudinal survey questions. These indicators, whether discrete or continuous, are not a direct measure of attitudes and perceptions but instead a function of these perceptions. Therefore, these indicators are at best an imperfect measure of the latent constructs they seek to capture. Including them in choice models as error-free variables raises two main issues. First, this type of approach does not take into account the fact that latent variables are never perfectly observed, and subject to measurement error. Second, endogeneity bias can occur when unobservable factors influence both respondents’ stated choices and their answers to attitudinal questions. Using a hybrid choice model allows to mitigate these issues, as widely acknowledged (Amaris et al., 2021; Buckell et al., 2021; Kim et al., 2014).

The article is organized as follows: in Section 1.1, we provide an overview of the literature on drinking water choices. Section 1.2 presents the data. The hybrid choice model is introduced in Section 1.3. The results are reported and discussed in Section 2. Section 3 concludes.

© 2024. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1.1. Literature Review

Notice that in what follows, we focus on drinking water choices in high-income countries, our data being taken from a survey conducted in France. However, there is also a rich and abundant literature for developing countries, which admittedly face specific issues regarding drinking water choice. For example, Pattanayak et al. (2005) explore the WTP for improved water supply in Kathmandu (Nepal), where people are confronted with a chronic shortage of water. By contrast, water shortages, when they occur in high-income countries, are more related to droughts than to a lack of investment in water infrastructure. Nevertheless, even if developing countries face specific water supply issues, consumers' perceptions of water quality (Cambodia, Orgill et al., 2013) and/or network water when compared to non-network water (Jordan, Orgill-Meyer et al., 2018) have been shown to play an important role in terms of drinking water choices, especially in terms of averting behaviors such as water storage, in-home treatment, etc. Such results and methodological contributions are partly related to the focus of our paper. However, we chose to restrict our literature review to high-income countries, selecting articles that explicitly address drinking water (tap water and filtered tap water vs. bottled water) using discrete choice econometrics, in line with what we intend to do in our paper. We start our literature review by recalling the role of non-price factors in drinking water choice models. We then continue by showing that there is actually a gap in the literature about the effect of water price (perceptions) in drinking water choice models. Finally, we end our literature review by pointing out the fact that including perceptions variables in drinking water choice models raises measurement error issues.

1.1.1. Non-Price Factors and Drinking Water Choice

The vast majority of the literature on drinking water choice in high-income countries focuses on the United States and Canada, presumably due to better data availability. The perception of health risks associated with tap water is at the heart of the first studies devoted to drinking water choices. As a result, risk perception variables are included as predictors in drinking water choice models, alongside sociodemographic variables or variables characterizing the good consumed, such as taste, appearance, or odor.

The seminal work of Abrahams et al. (2000) looks at drinking water choice as a means of assessing the averting costs associated with perceived risks of water contamination. Using data from a survey conducted in Georgia in 1995, the authors examine how drinking water choices (tap water, filtered tap water, bottled water) are associated with perceived risk, measured as a binary which takes the value 1 when the respondents stated that their tap water was unsafe (0 otherwise), with the perceived quality of the tap water, measured as a composite index reflecting the odor, taste, and appearance of the tap water, and finally with sociodemographic variables. It should be noted that the authors do not address the issue of measurement error in perception variables. The estimation of a simple multinomial logit model leads the authors to identify both the effects of sociodemographic variables (e.g., non-whites tend to consume more bottled water), and the effects of perception variables. Respondents who consider tap water unsafe tend to consume more bottled water. This is also true for respondents who are dissatisfied with the quality of tap water. Overall, the authors conclude that bottled water expenditures reflect not only risk averting costs, but also the effect of the perceived quality of tap water on drinking water choices.

From an Internet-based survey in Canada, Dupont et al. (2010) also show, by estimating a simple multinomial logit model, that the probability of being a bottled water drinker or filtered tap water drinker increases with income or for respondents who have had unpleasant past experiences with tap water taste. They find that two key perception variables influence the choice of bottled water and filtered tap water over tap water: the level of perceived health risk associated with tap water and the belief that bottled water is safer than tap water.

Other relevant studies include Viscusi et al. (2015), who show that risk perception and past experiences are major determinants of drinking water consumption choices, together with gender, income, and race. Similar results have been found by Doria (2006) and Javidi and Pierce (2018), who focus on whether respondents perceive tap water to be safe or not. Furthermore, Levêque and Burns (2017) report that the perception that consumers have regarding the risks related to drinking tap water, as well as the perceived quality of surface water, are predictors of bottled and filtered water consumption, respectively. Lastly, based on Canadian data, Lloyd-Smith et al. (2018) find, using a latent class model, that perceived risk negatively affects the probability of drinking tap water. However, their focus is on the issue of averting expenditures against health risks associated with tap water consumption.

In addition to the United States and Canada, there are similar studies for other high-income countries, such as France, which is the focus of our analysis.

To our knowledge, Bontemps and Nauges (2009) are the first to explore drinking water choices in France. More precisely, the authors seek to assess how the quality of raw water influences the choice of drinking tap water or not. One of the original features of the article comes from the way in which the quality of raw water is measured, through an index of poor quality of raw water, which in fact corresponds to the drinking water production cost where the respondents are located. Estimation of simple probit models shows that the poor quality of tap water influences whether respondents drink it or not.

Johnstone and Serret (2012) use survey data from 10 OECD countries, including France, to analyze drinking water choices. The authors estimate a simple multinomial logit model to identify the main predictors of the choice to consume bottled water and/or purified tap water, with tap water taken as the reference case. Beyond the effect of some sociodemographic variables, the results show that concerns about health risks and (bad) tap water taste have a positive impact on the consumption of bottled and/or filtered tap water. Again, variables that capture the level of concern about tap water quality are not considered to be prone to measurement error. Bontemps and Nauges (2016) use a subset of the same data set (focusing on Australia, Canada, and France) to compare several methods for controlling the endogeneity of a binary variable that encodes satisfaction or dissatisfaction, with the health impacts of tap water as a predictor of whether to drink tap water or not. Bontemps and Nauges (2016) argue that perception variables, such as the perception of the health risks related to drinking tap water, should be treated as endogenous to averting decisions (here, the decision to drink tap water or not). They compare two methods for controlling endogeneity: the recursive bivariate probit model and the special regressor approach. The results show that the marginal effect of being satisfied with tap water on the probability of drinking tap water is twice as large for the recursive bivariate model compared to what is found with the special regressor approach. As the latter is less demanding in terms of the underlying distributional assumptions, the authors suggest that it should be used whenever a special regressor is available.

Finally, Beaumais and Veyronnet (2017) use the same data we use for this paper to examine how drinking water choices (tap water, filtered water, bottled water) are associated with the degree of satisfaction with tap water, measured by a three-level categorical variable (not satisfied, satisfied, very satisfied). Taste heterogeneity across decision makers is modeled using a scaled multinomial logit model, which captures the influence of both observable and unobservable variables on the weights of the variables involved in the indirect utility function. The potential endogeneity of the satisfaction variable is controlled for by the two-stage residual inclusion method (2SRI; Terza et al., 2008). The results suggest that the categorical variable measuring satisfaction with tap water is endogenous. Specifically, comparing the results of models with and without endogeneity control shows significant differences between the two specifications: in both, satisfaction with tap water is a significant predictor of drinking water choices. However, when endogeneity is controlled for, it appears that individuals react in a binary way (not satisfied vs. satisfied) rather than in a nuanced way (not satisfied, satisfied, very satisfied). One of the salient points of all these articles, with the notable exception of Bontemps and Nauges (2016) and Lloyd-Smith et al. (2018), is that price is not considered as an attribute of drinking water choice.

Some authors (Abrahams et al., 2000; Dupont et al., 2010; Viscusi et al., 2015) recognize the potential role of price in drinking water choices, but face data limitations that prevent the use of water price in their modeling work. Others, such as Bontemps and Nauges (2009) argue that “consumer choice is not based on price comparison, as the price of bottled water is about 100 times higher than the price of tap water.” The idea is that this price difference for two close substitutes should lead consumers to turn away from bottled water. Beaumais and Veyronnet (2017) build similar arguments, and point that in their data, most respondents do not know the price of tap water and are unaware of the price difference between bottled water and tap water. In other studies, price is simply not considered a predictor of drinking water choices (Javidi & Pierce, 2018; Johnstone & Serret, 2012; Levêque & Burns, 2017).

1.1.2. Water Price and Drinking Water Choice: A Substantial Gap in the Literature

As mentioned above, the works of Bontemps and Nauges (2016) and Lloyd-Smith et al. (2018) are the only ones that we found, in which the authors include water price in a drinking water choice model. Although Bontemps and Nauges (2016) use data from a 2011 OECD survey to study the influence of perceived tap water quality on the choice of drinking bottled water instead of tap water, the authors also use the regional average price of water from

a previous 2008 OECD survey in their estimation. This modeling strategy is explained by the fact that their special regressor, introduced to control for the endogeneity of the perception variable measuring tap water quality, is precisely the average water price in 2008 (at the regional level). Furthermore, the authors “expect tap water to be a normal good.” In this vein, the data used by Lloyd-Smith et al. (2018) are sufficiently rich for them to include both the cost of bottled water and filtered tap water in their models, while the price of tap water is assumed to be zero. However, it should be noted that the latter authors do not address the measurement error that might be associated with these stated data on bottled water expenditure and filtered water expenditure. Their work focuses on ways of controlling for the potential endogeneity of perceived health risks associated with tap water, without questioning how the price of water is perceived.

As previously discussed in the Introduction, the term “water price mis-perception,” as used by Brent and Ward (2019), seems more appropriate to refer to how people perceive the price of water. Notice that cost mis-perceptions, for example, energy costs, are known to alter the decision-making process and can result in welfare losses, which can be mitigated by targeted policies (Allcott, 2013). Following Brent and Ward (2019), we therefore assume that using an objective tap water price measure to inform tap water consumption choices is inadequate. In fact, in France, only 20% of the consumers surveyed have been found to know the correct price of tap water, while 58% stated that they “do not know” (CGDD, 2014). It is then obvious that consumers do not use any objective water price variable in the utility maximization process underlying their drinking water choices.

In the literature on individual discrete choice behavior, be it in transport economics, energy economics or water economics, it is now commonplace to distinguish between decision utility and experienced utility (Allcott, 2013; Allcott & Sunstein, 2015; Brent & Wichman, 2022; Train, 2015). In a typical discrete choice model, each individual is confronted with a set of discrete alternatives (in this paper, tap water, filtered water and bottled water). Decision utility, also called anticipated utility (Train, 2015) or choice utility (Brent & Ward, 2019; Brent & Wichman, 2022), refers to the utility that an individual expects from choosing one alternative over other available options, at the decision step. Experienced utility, also called actual utility (Train, 2015), refers to the utility that an individual receives from experiencing (consuming) the chosen alternative (good). When the individual misperceives one or several attributes of the alternatives, at the decision step, experienced utility can be different from decision utility, thus inducing welfare losses in cases where experienced utility is lower than decision utility. Misperceptions can be due to imperfect information or internalities, that is, “costs we impose on ourselves by taking actions that are not in our own best interest.” (Allcott & Sunstein, 2015) due to behavioral biases. Notice that, as pointed out by Train (2015, Footnote 2), misperceptions induce a difference between choice utility and experienced utility only if the individual involved discovers ex post (at the experience step) that she misperceived one or several attributes of the chosen alternative (good). Nevertheless, beyond the assessment of welfare losses and from a policy recommendation perspective, identifying whether misperceived attributes actually influence discrete choice is of importance, as it can help design relevant interventions (see our conclusions).

To sum up, the papers on drinking water choice models that we have discussed in this literature review do not consider the role of the perceived price of water, or the perceived difference in water prices (bottled water, tap water) on drinking water choice. Our paper aims to fill this gap. However, in terms of econometric strategy, perception variables pose specific issues, which we now address.

1.1.3. Perception Variables as Predictors of Drinking Water Choice and Endogeneity

In the empirical literature on tap water consumption, many authors have used survey responses to attitudinal statements directly as covariates in a drinking water choice model (Bontemps & Nauges, 2016). However, it is now widely acknowledged that attitudes and perceptions are unobserved and that only manifestations or imperfect measurements of these attitudes can be observed (Hess & Stathopoulos, 2013). More precisely, the *perceived quality of water resources* is arguably not an exact measure of the way respondents perceive it, but a function thereof, and using it as an explanatory variable in a model of drinking water choice is hence likely to put the analyst at risk of measurement error, leading to endogeneity bias.

As it is well known, endogeneity can arise from three sources: measurement error, omitted variables and simultaneous determination. Regarding the perception of health risks related to tap water, Bontemps and Nauges (2016) seem to target endogeneity arising from simultaneity. They argue that “In typical theoretical averting-behavior models, both actual risk and perceived risk are endogenous since they depend on attitudes toward safety, and hence on the averting actions being taken.” Lloyd-Smith et al. (2018) explicitly add the source

of the omitted variable, given that “unobserved factors affecting water choices may be correlated with risk perceptions leading to omitted variable bias.” Notice that in their empirical work, Bontemps and Nauges (2016) actually control for the endogeneity of a binary variable that is equal to one when respondents stated to be satisfied with the health impact of tap water and zero otherwise. However, they also include the perceived local water quality in their models, without controlling for the potential endogeneity of this variable. The same holds for Lloyd-Smith et al. (2018), who focus exclusively on controlling the endogeneity of their risk perception variable, without controlling for the endogeneity of the water quality perception variables that they include in their model. On the contrary, Whitehead (2006), cited by both Bontemps and Nauges (2016) and Lloyd-Smith et al. (2018), acknowledges the potential endogeneity of water quality perception variables in a series of models designed to assess willingness to pay (WTP) for improved water quality. Interestingly, Whitehead (2006) considers the inclusion of perceived water quality in WTP models as a way to mitigate the omitted variable bias. From his point of view, endogeneity arises because unobservable variables could be correlated with both perceived quality and WTP. Likewise, Vásquez et al. (2015) include a variable measuring (tap water) quality perception in a series of probit models aimed at exploring various averting behaviors related to drinking water in Nicaragua, such as boiling water, treating water, or purchasing bottled water. They control for the potential endogeneity of the quality perception variable, which, they consider, stems from the fact that “in the context of developing countries where information of water quality is hardly available, household perceptions of water quality may be endogenous to averting behaviors as those perceptions can be based on several factors other than actual water quality.”

Although ignoring water resources quality and water price (mis-)perception might lead to endogeneity due to omitted variables, including these factors in drinking water choice models is also a potential source of endogeneity due to measurement error. This issue results from the fact that attitudes and perceptions are usually not direct measures of latent constructs, but rather their functions (Budziński & Czajkowski, 2022). More precisely, while it is essential to account for attitudes and perceptions, especially when individuals have a very poor or even erroneous knowledge of the objective reality, measurement error arises when directly incorporating indicators of attitudes and perceptions as error-free variables in standard choice models, as “attitudes can never be observed by an analyst” (Amaris et al., 2021). We argue, in line with Kim et al. (2014), Buckell et al. (2021), and Amaris et al. (2021), that an appropriate econometric strategy, given the modeling problem at hand, is to estimate a hybrid choice model. Such an approach allows us to introduce quality and price perceptions in a drinking water choice model. HCMs are made of three parts (Ben-Akiva et al., 2002; Kim et al., 2014): a choice model (commonly based on the random utility framework), a measurement component and a structural component. It is important to clarify that indicators do not directly enter the utility function. Instead, the indicators and the choice model are interacted by the means of latent variables. The measurement and structural components form what is called the latent variable model. The measurement component links the indicators to the latent variables, explicitly accounting for heterogeneous error terms across individuals, while the structural component links the latent variables to exogenous variables, such as sociodemographic factors. We introduce the data in the next section.

1.2. Data

Our research is based on data collected in 2013 from 4,003 respondents by IFOP (*Institut Français d'Opinion Publique*) as part of a survey commissioned by the *Commissariat général au développement durable*. Given the secondary nature of our data, we performed a series of checks to ensure the integrity of the data. A detailed data integrity statement is included in Appendix A, where limitations are also discussed.

In addition to the usual sociodemographic variables, respondents were asked about their perception of the quality of water resources, their choice of drinking water, their knowledge of the price of tap water, their knowledge of the difference between the price of tap water and the price of bottled water, etc. The full sample is representative of the French population (CGDD, 2014). The estimation sample is restricted to respondents for whom income is available, *that is*, 3,506 individuals without altering the representativeness of the sample (see Table 1). Additional variables on nitrate concentration in groundwater and rainfall level were added to the original database from external sources (from the Eider database of French Ministry of the Environment and from the public water data web portal, data.eaufrance.fr). We used the most detailed data for the nitrate levels that are not available at the catchment areas of the respondents but rather at the “arrondissement” level, which is an administrative subdivision of France (Regions > Departments > Arrondissement). At the time that we merged the survey data with external data on nitrate levels, France was divided into 332 arrondissements. It is worth noting that, in France, the

Table 1

Descriptive Statistics

Variables (Names of the variables used in the econometric model in parenthesis)	Mean	
	Full sample	Estimation sample
Bottled water	39%	39%
Filtered water	21%	21%
Tap water	40%	40%
Age (x_{age})	47 years old	47 years old
(standard deviation)	(16)	(16)
Gender (x_{female})	52%	51%
Education		
<i>No degree</i> (x_{edu1})	3%	3%
<i>Middle school degree</i> (x_{edu2})	26%	26%
<i>High school degree</i> (x_{edu3})	19%	19%
<i>Bachelor degree</i> (x_{edu4})	34%	34%
<i>Master and higher degree (reference)</i>	19%	19%
Rural location (x_{rural})	23%	23%
Average monthly income (x_{income})	–	2,759€
(standard deviation)	–	(1,658)
Number of children under 18 (x_{child})	0.521	0.526
Number of observations	4,003	3,506

Note. Authors' calculations.

vast majority (two-thirds) of drinking/tap water is produced from groundwater, which also explains why we chose to use nitrate concentration in groundwater as an objective measure of the quality of water resources.

In the survey, the question about drinking water choices was expressed as follows:

“At home, when you drink water, most often you drink...?”

(Only one possible answer)

- Mainly bottled water,
- Mainly filtered tap water,
- Mainly tap water.”

As is usual in the literature on drinking water choice, the wording of this question makes the choices mutually exclusive, so that discrete choice modeling techniques can be applied to the data. The main choice of drinking water in France is tap water (40%). Bottled water is the second most popular (39%). Finally, filtered water is third in the list and is consumed by 21% of respondents (see Table 1).

1.2.1. Choice of Drinking Water and Perceived Quality of Water Resources

Does the choice of drinking water depend on the perceived quality of water resources stated by the respondents? In response to the following question: “Apart from drinking water, would you say that the current quality of water resources in France (whether in rivers, lakes, groundwater) is...?”, the stated perceived quality of water resources can take one value among four: very poor (1), poor (2), good (3) and very good (4). Note that a separate question was asked about the perception of the expected quality of water resources in the future.

As a first step in our analysis, we report the choice of drinking water depending on the perceived quality of water resources (Table 2). We can observe that the share of tap water drinkers is 34.34% when the respondents consider that water resources are of very poor quality, while it increases to 53.90% when they consider them to be very

Table 2
Drinking Water Choice and Perceived Quality of Water Resources

	Bottled water	Filtered water	Tap water
Very poor	44.44%	21.21%	34.34%
Bad	45.36%	20.20%	34.45%
Good	35.61%	21.46%	42.93%
Very good	26.95%	19.15%	53.90%

Note. Authors' calculations.

good (an increase of 19.56% points). The dependence between the two variables, one nominal, the other ordered, is confirmed by a χ^2 test ($\chi^2(6) = 45.1451, p\text{-value} = 0.000$).

1.2.2. Perception of Water Resources Quality

Furthermore, Table 3 reports the concentration of nitrates in milligrams per liter (mg/l) in groundwater used to supply drinking water. Although it is not relevant for our analysis to know *where* the highest concentrations of nitrate are found in mainland France, it is important to point out that there are large differences across the territory. The South of France are the regions where

people drink tap water the most (Auvergne-Rhône-Alpes: 56.91%; Provence-Alpes-Côte d'Azur and Corsica: 51.27%; Occitanie: 50.88%), and are also regions where average nitrate concentrations in raw water are low. On the other hand, Hauts-de-France (17.83%), Centre-Val de Loire (28.85%) and Normandy (31.43%) are the regions where individuals drink tap water the least and average nitrate concentrations relatively high. The last column of Table 3 shows the average perceived quality by region. As perceived quality is an ordered categorical variable that ranges from 1 (very poor) to 4 (very good), a higher average perceived quality roughly corresponds to a better perceived quality. The region where the average nitrate concentration is lowest (Provence-Alpes-Côte d'Azur and Corsica) is also the region where the perceived quality is the highest on average. The regions where the average nitrate concentration is the highest (Normandy, Hauts-de-France) exhibit a low perceived quality on average, Brittany (Bretagne) being an exception, which illustrates that subjective perceptions can be far from objective measures of water quality (Artell et al., 2013).

1.2.3. Perceived Water Price

Finally, Table 4 shows that 58% of the respondents do not know the price of tap water; only 20% can state a value in line with the price actually charged (about €3.4 per m³; the mean value given by the respondents for how much a cubic meter of tap water costs is €5.97 and the standard deviation €8.89). Furthermore, 56% of the respondents consider tap water to be somewhat expensive, about 30% consider it to be somewhat affordable, while 14% state that they do not know whether tap water is expensive or affordable. These findings are in line with other sources, as further discussed in Appendix A. Additionally, the price difference between tap and bottled water is underestimated by approximately half the population, with about 25% unable to comment on the difference: on average, at the time of the survey, bottled water was about 100 times more expensive than tap water according to CGDD (2014). Notice that the same figure is used by Bontemps and Nauges (2009) based on bottled water

Table 3
Nitrate Concentration in Raw Water, Tap Water Consumption and Perceived Water Resources Quality by Geographical Location

	Average nitrate concentration	Tap water consumption	Average perceived quality
Provence-Alpes-Côte d'Azur + Corse	8.13	51.27%	2.76
Île-de-France	10.20	39.15%	2.66
Nouvelle-Aquitaine	12.63	42.90%	2.71
Auvergne-Rhône-Alpes	13.56	56.91%	2.75
Bourgogne-Franche-Comté	14.57	44.44%	2.60
Occitanie	14.82	50.88%	2.63
Grand Est	16.53	33.01%	2.69
Centre-Val de Loire	19.93	28.85%	2.62
Pays de la Loire	20.89	40.21%	2.62
Hauts-de-France	23.58	17.83%	2.55
Normandie	26.13	31.43%	2.56
Bretagne	28.70	32.97%	2.65

Note. Authors' calculations.

Table 4
Price and Quality Perception—Descriptive Statistics

Perception question	Possible answers	Freq. (%)
<i>How much does one cubic meter of tap water cost?</i>	<i>I don't know</i>	58.27
	Respondent gives an amount	41.73
<i>How much more expensive do you think bottled water is than tap water?</i>	0 (<i>I don't know</i>)	24.76
	1 (<i>bottled water is as expensive as tap water</i>)	6.42
	2 (<i>bottled water is 2 times more expensive</i>)	17
	3 (<i>bottled water is 5 times more expensive</i>)	16.94
	4 (<i>bottled water is 10 times more expensive</i>)	13.98
	5 (<i>bottled water is 100 times more expensive</i>)	20.91
<i>How expensive do you feel tap water is?</i>	0 (<i>I don't know</i>)	14.15
	1 (<i>somewhat affordable</i>)	29.66
	2 (<i>somewhat expensive</i>)	56.19
<i>Apart from drinking water, would you say that the current quality of water resources in France (whether in rivers, lakes, groundwater) is...?</i>	1 (<i>very poor</i>)	2.82
	2 (<i>poor</i>)	31.63
	3 (<i>good</i>)	61.52
	4 (<i>very good</i>)	4.02
<i>For the years to come, do you think the quality of water resources...?</i>	1 (<i>will decrease</i>)	49.94
	2 (<i>will remain the same</i>)	24.07
	3 (<i>will improve</i>)	25.98

Note. Authors' calculations.

purchase data from a sample of French households. The actual price difference at the time of the CGDD survey was that tap water is 92 times more expensive than bottled water. We hypothesize that the correct answer has been rounded up to 100 to reduce cognitive burden and avoid making it too salient compared to other response options. Therefore, it is likely that whether an individual correctly perceives the price difference between tap water and bottled water influences whether she chooses to consume bottled water. Hence, our econometric strategy consists of modeling the perception of the price of tap water with respect to bottled water as a latent variable, where the stated perception of the price difference between bottled water and tap water is used as an indicator jointly estimated together with the choice model.

1.3. Econometric Strategy: Latent Variables and Hybrid Choice Model

Typical HCMs are composed of three parts: a measurement component, a structural component, and a discrete choice model component (Ben-Akiva et al., 2002; Budziński & Czajkowski, 2022; Groothuis et al., 2021; Kim et al., 2014). In the following, we present each part in detail to end with the log-likelihood of the whole system of equations. A detailed graph summarizing how the different parts are connected is available in Appendix A (Figure A1). The measurement component of the HCM is composed of a series of indicators measuring two underlying latent variables related to the *perception of water quality* and the *perception of tap water price*, labeled $\alpha_{quality}$ and α_{price} respectively.

1.3.1. Measurement Component: Perception of Water Quality

The latent variable $\alpha_{quality}$ is informed by two indicators related to the perception of water resources quality, today and in the future. The measurement models for these indicators are ordered logit model structures. The probability $PI_{quality}$ of observing a given outcome n where ($n = 1-4$) for the indicator $I_{quality}$ related to the perception of water quality today is given by an ordered logit (OL) model which is specified as follows:

$$P(I_{quality} = n) =$$

$$\left(\frac{1}{1 + e^{(\zeta_{quality} \cdot \alpha_{quality} - \tau_{quality_{n+1}})}} \right) - \left(\frac{1}{1 + e^{(\zeta_{quality} \cdot \alpha_{quality} - \tau_{quality_n})}} \right) \quad (1)$$

where the parameter $\zeta_{quality}$ thus corresponds to the effect of a variation in the latent variable $\alpha_{quality}$ on the perception of water quality. The parameters labeled as $\tau_{quality_n}$ are latent thresholds estimated by the model and normalized so that $\tau_{quality_1} = -\infty$ and $\tau_{quality_5} = +\infty$. The indicator related to the expected quality in the future is labeled $I_{qualityf}$, and is also modeled via an ordered logit structure (where $\zeta_{quality}$ is instead replaced by $\zeta_{qualityf}$ and where another set of thresholds labeled $\tau_{qualityf}$ is estimated.

1.3.2. Measurement Component: Water Price Perception

The latent variable α_{price} is informed by three indicators related to whether respondents *feel* that bottled water is expensive compared to tap water and whether they *know* the *actual* price of tap water as well as the *actual* price difference between tap water and bottled water. For each indicator, we first model by the means of a binary logit whether each respondent has answered the question or simply stated “I don’t know” and, if the respondent has provided an answer, what her answer is (using different model structures as further described in Table 5 below). As an illustration, the probability of observing the set of answers for question 2, “How much more expensive do you think bottled water is than tap water?” is labeled PI_{price2} and corresponds to:

$$PI_{price2} = \left(1 - \left(\frac{1}{1 + e^{(\zeta_{price2_NA} \cdot \alpha_{price} - \tau_{price2_NA})}} \right) \right)^{(I_{price2_NA} == 1)} \left(\left(\frac{1}{1 + e^{(\zeta_{price2} \cdot \alpha_{price} - \tau_{price2_{n+1}})}} \right) - \left(\frac{1}{1 + e^{(\zeta_{price2} \cdot \alpha_{price} - \tau_{price2_n})}} \right) \right) \quad (2)$$

The equation for PI_{price2} combines a binary logit (BL) model for whether the respondent has given a response to the question or not as well as an ordered logit model which is only estimated if the respondent has answered the question (this is the purpose of the indicator variable $I_{price2NA}$ which takes the value 1 if the respondent has answered and 0 else). The ordered logit part of the equation (from which “do not know” is excluded) is coded in such a way that a higher value for the dependent variable corresponds to an answer that is closer or equal to the right statement which is that “bottled water is 100 times more expensive [than tap water]”.

The other two measurement equations enter the model in a similar fashion, with PI_{price1} combining a logit model and a Linear Regression (LR) for modeling “how much does one cubic meter of water cost” and PI_{price3} combining two logit models for modeling whether the respondent feels that tap water is “somewhat expensive” or “somewhat affordable” conditionally on whether the respondent has provided a response for this question. The entire set of indicators as well as the latent variables they are interacted with are reported in Table 5 below.

1.3.3. Structural Component

The structural component links the latent variables to exogenous factors, including sociodemographic variables. More formally, $\alpha_{quality}$ and α_{price} are defined as follows for respondent i :

$$\begin{aligned} \alpha_{quality,i} = & \alpha_{quality,nitrates} \cdot x_{nitrates} + \alpha_{quality,rainfall} \cdot x_{rainfall} + \alpha_{quality,age} \cdot x_{age} + \alpha_{quality,female} \cdot x_{female} \\ & + \alpha_{quality,child} \cdot x_{child} + \alpha_{quality,rural} \cdot x_{rural} + \alpha_{quality,income} \cdot x_{income} + \alpha_{quality,edu1} \cdot x_{edu1} \\ & + \alpha_{quality,edu2} \cdot x_{edu2} + \alpha_{quality,edu3} \cdot x_{edu3} + \alpha_{quality,edu4} \cdot x_{edu4} + \sigma_{1,i} \end{aligned} \quad (3)$$

Table 5
Latent Variables and Related Measurement Equations

Latent variable	Indicator	Label	Coding	Structure
$\alpha_{quality}$	Perception of water resources quality (nowadays)	<i>quality</i>	1 (<i>very poor</i>) 2 (<i>poor</i>) 3 (<i>good</i>) 4 (<i>very good</i>)	OL
	Perception of water resources quality (in the future)	<i>quality, f</i>	1 (<i>it will decrease</i>) 2 (<i>it will remain the same</i>) 3 (<i>it will improve</i>)	OL
α_{price}	<i>How much does one cubic meter of tap water cost?</i>	<i>Price1</i>	Open-ended	LR
	<i>How expensive is bottled water compared to tap water?</i>	<i>Price2</i>	1 (<i>bottled water is as expensive as tap water</i>) 2 (<i>bottled water is 2 times more expensive</i>) 3 (<i>bottled water is 5 times more expensive</i>) 4 (<i>bottled water is 10 times more expensive</i>) 5 (<i>bottled water is 100 times more expensive</i>)	OL
	<i>How expensive do you feel tap water is?</i>	<i>Price3</i>	1 (<i>somewhat affordable</i>) 2 (<i>somewhat expensive</i>)	BL
	Response is “ <i>I don't know</i> ” for indicator <i>price1</i>	<i>priceNA1</i>	0 (<i>I don't know</i>) 1 (respondent provided an answer)	BL
	Response is “ <i>I don't know</i> ” for indicator <i>price2</i>	<i>priceNA2</i>	0 (<i>I don't know</i>) 1 (respondent provided an answer)	BL
	Response is “ <i>I don't know</i> ” for indicator <i>price3</i>	<i>priceNA3</i>	0 (<i>I don't know</i>) 1 (respondent provided an answer)	BL

$$\begin{aligned} \alpha_{price,i} = & \alpha_{price,nitrates} \cdot x_{nitrates} + \alpha_{price,rainfall} \cdot x_{rainfall} + \alpha_{price,age} \cdot x_{age} + \alpha_{price,female} \cdot x_{female} \\ & + \alpha_{price,child} \cdot x_{child} + \alpha_{price,rural} \cdot x_{rural} + \alpha_{price,income} \cdot x_{income} + \alpha_{price,edu1} \cdot x_{edu1} \\ & + \alpha_{price,edu2} \cdot x_{edu2} + \alpha_{price,edu3} \cdot x_{edu3} + \alpha_{price,edu4} \cdot x_{edu4} + \sigma_{2,i} \end{aligned} \quad (4)$$

The continuous variables are recoded so that they are now centered on 0. The binary variable “*rural*” takes the value 1 for respondents living in a rural area and 0 else. The variables “*edu1*,” “*edu2*,” “*edu3*,” and “*edu4*” are binary variables capturing different levels of education and where “*Master and higher degree*” is the base. The variable $x_{nitrates}$ is the observed nitrate concentration and $x_{rainfall}$ is the local level of rainfall. Other variables have been defined in Table 1. Finally, σ_1 and σ_2 are random disturbances specified as such: $\sigma_1 \sim N(0,1)$ and $\sigma_2 \sim N(0,1)$.

1.3.4. Choice Model

The choice model corresponds to the well-known multinomial logit model. The utility that an individual, noted i , derives from choosing alternative j from a set of J alternatives is:

$$U_{ij} = V_{ij} + \varepsilon_{ij} = \beta x_{ij} + \varepsilon_{ij} \quad (5)$$

where x_{ij} is a vector of observed characteristics of alternative j and observed characteristics of individual i and β a vector of coefficients, which are interpreted as weights in the utility function. $V_{ij} = \beta x_{ij}$ is often referred to as the deterministic component of the utility U_{ij} . The error term, ε_{ij} , captures the effect of variables that influence the utility U_{ij} but are not included in βx_{ij} . Individual i chooses alternative k if and only if $U_{ik} > U_{ij}, \forall j \neq k$. Therefore, the probability that individual i chooses alternative k can be written as follows:

$$\begin{aligned}
 P_{ik} &= \Pr(U_{ik} > U_{ij}, \forall j \neq k) \\
 &= \Pr(\beta x_{ik} + \varepsilon_{ik} > \beta x_{ij} + \varepsilon_{ij}, \forall j \neq k) \\
 &= \Pr(\varepsilon_{ij} - \varepsilon_{ik} < \beta x_{ik} - \beta x_{ij}, \forall j \neq k)
 \end{aligned} \tag{6}$$

Assuming that ε_{ij} is distributed iid extreme value, the probability of choosing a given alternative (e.g., alternative 1 among three) is given by:

$$P_1 = \frac{e^{V_1}}{e^{V_1} + e^{V_2} + e^{V_3}} \tag{7}$$

The two latent variables $\alpha_{quality}$ and α_{price} directly enter the utility functions for the multinomial logit model, which are specified as follows for each individual i (index i has been omitted again for variables x):

$$\begin{aligned}
 V_{tap,i} &= \beta_{tap,constant} + \beta_{tap,nitrates} \cdot x_{nitrates} + \beta_{tap,rainfall} \cdot x_{rainfall} + \beta_{tap,age} \cdot x_{age} + \beta_{tap,female} \cdot x_{female} \\
 &\quad + \beta_{tap,child} \cdot x_{child} + \beta_{tap,rural} \cdot x_{rural} + \beta_{tap,income} \cdot x_{income} + \beta_{tap,edu1} \cdot x_{edu1} + \beta_{tap,edu2} \cdot x_{edu2} \\
 &\quad + \beta_{tap,edu3} \cdot x_{edu3} + \beta_{tap,edu4} \cdot x_{edu4} + \theta_{tap,quality} \alpha_{quality,i} + \theta_{tap,price} \alpha_{price,i}
 \end{aligned} \tag{8}$$

$$\begin{aligned}
 V_{filtered,i} &= \beta_{filtered,constant} + \beta_{filtered,nitrates} \cdot x_{nitrates} + \beta_{filtered,rainfall} \cdot x_{rainfall} + \beta_{filtered,age} \cdot x_{age} \\
 &\quad + \beta_{filtered,female} \cdot x_{female} + \beta_{filtered,child} \cdot x_{child} + \beta_{filtered,rural} \cdot x_{rural} + \beta_{filtered,income} \cdot x_{income} \\
 &\quad + \beta_{filtered,edu1} \cdot x_{edu1} + \beta_{filtered,edu2} \cdot x_{edu2} + \beta_{filtered,edu3} \cdot x_{edu3} + \beta_{filtered,edu4} \cdot x_{edu4} \\
 &\quad + \theta_{filtered,quality} \alpha_{quality,i} + \theta_{filtered,price} \alpha_{price,i}
 \end{aligned} \tag{9}$$

$$V_{bottled,i} = 0 \tag{10}$$

where the parameters $\theta_{tap, quality}$, $\theta_{tap, price}$, $\theta_{filtered, quality}$, and $\theta_{filtered, price}$ correspond to the effect of the latent variables $\alpha_{quality}$ and α_{price} on the probability of choosing tap water and filtered water with respect to bottled water, which is the reference alternative of the model. It is worth noting that the same variables enter V_{tap} , $V_{filtered}$ as well as $\alpha_{quality}$ and α_{price} , in line with the recommendations of Vij and Walker (2016) on how to adequately specify HCMs. Notice that the HCM model results were compared with “naïve” multinomial logit (MNL) and nested logit models where the indicators enter the models in the same way as other covariates labeled as β . The HCM model was found to be more efficient (as defined and following the same measurement method proposed by Abou-Zeid and Ben-Akiva (2014)) than its MNL counterpart, while nested logit structures were not found to substantially improve fit and behavioral insights. Efficiency measures comparing a naïve MNL and the HCM are reported in Appendix A (Table A2).

The model hence jointly explains the values of 6 dependent variables including the main drinking water choice question, the answers to the two water quality perception questions as well as the three water price perception questions. Therefore, the log-likelihood corresponds to the joint probability of observing the tap water consumption choice for each individual, as well as the likelihood of observing the responses related to each one of the indicators described above. The presence of random parameters in the model allows to relax the Independence of Irrelevant Alternative (IIA) property of the basic multinomial logit model (without mixing):

$$LL(\Omega_U, \Omega_\alpha, \Omega_\theta) = \sum_{i=1}^I \ln \int \int_{\sigma_1 \sigma_2} P_i \cdot PI_{quality,i} \cdot PI_{quality,f,i} \cdot PI_{price1,i} \cdot PI_{price2,i} \cdot PI_{price3,i} g(\sigma_1) m(\sigma_2) d\sigma_1 d\sigma_2 \tag{11}$$

2. Results

The final model was estimated using 25,000 Halton draws. We used the Apollo package (Hess & Palma, 2019) running on the Myria HPC system (CRIANN, Normandy, France). Results are described in Table 6 (robust standard errors and robust T-ratios are also reported).

2.1. Quality Perception

The results should be interpreted as follows: an increase in the (normally distributed and unitless) latent variable $\alpha_{quality}$ decreases the probability of drinking tap water compared to bottled water as $\theta_{tap, quality}$ is negative (-0.4015) and significant at the 1% level. An increase in this same latent variable also decreases, to a lesser extent, the probability of consuming filtered water given that $\theta_{filtered, quality}$ is also negative (-0.2327) and significant at the 1% level. An increase in the latent variable $\alpha_{quality}$ decreases the probability of stating a better perception of water quality. Indeed, the parameter $\zeta_{quality}$ is found to be significant and equal to -0.9402 . This indicates that respondents who are less likely to state that the current quality of water resources is good are also less likely to drink tap water, in line with expectations. The sign of $\zeta_{quality}$ shows that respondents who are more likely to consider that the current quality of water resources in France is good are more likely to consider that this quality is going to decrease in the future. Overall, our results suggest that respondents are more likely to drink tap water or filtered tap water than bottled water when they are more likely to consider that the quality of water resources is high.

Surprisingly, the latent variable $\alpha_{quality}$ is not influenced by nitrate and rainfall levels. This result is driven by multiple factors. First, when estimating the ordered logit models related to *quality* and *quality_f* as separate models rather than as indicators in a hybrid choice model structure, we find that the variables *nitrates* and *rainfall* have a significant effect on *quality* but not on *quality_f*, which might affect the significance of the parameter when used as part of a latent construct informed by both indicators. Second, the means of both *nitrates* and *rainfall* are significantly higher for respondents living in a rural area according to a series of *t*-tests significant at the 1% level ($t = -5.5827$ and -4.5784 respectively). Hence, it might be that the effect of *nitrates* and *rainfall* on $\alpha_{quality}$ is also captured by *rural*, which is found to be negative and significant. At the same time, there are reasons for keeping all three variables in the model as the distributions of *nitrates* for rural and nonrural respondents (unreported) largely overlap (and the same argument can apply to *rainfall*). This does not mean that respondents do not adjust their behavior according to objective measures of water quality, given that $\beta_{tap, nitrates}$ is negative and significant at the 1% level in the choice model component of the hybrid choice model. Other results indicate that $\alpha_{quality}$ is lower for older individuals (-0.0171) and higher for women (0.4591). The latent variable also varies by level of education and income level, where respondents who reported a higher income exhibit a lower value for $\alpha_{quality}$ all else being equal.

2.2. Price Perception

We now investigate the relationship between the indicators related to the perception of the price of tap water (with respect to bottled water) and the choice of water consumption. We find that an increase in the (normally distributed and unitless) latent variable related to the perception of the price of bottled water, α_{price} , increases the probability of consuming tap water and filtered tap water with respect to bottled water given that $\theta_{tap, price}$ and $\theta_{filtered, price}$ are both positive and significant at the 1% level (the parameter values are 0.3718 and 0.3565 respectively). At the same time, an increase of α_{price} decreases the probability of reporting a high value for a cubic meter of tap water given that ζ_{price1} is negative (-0.5209) and significant at the 1% level. Moreover, an increase of α_{price} also increases the probability of reporting the correct price of bottled water with respect to tap water given that ζ_{price2} is positive (1.7140) and also significant at the 1% level. In other words, respondents who are less likely to know the correct price of bottled water with respect to tap water are also more likely to consume it. This important result has not been observed elsewhere in the literature, to the best of our knowledge. Also, an increase in α_{price} increases the probability of stating that tap water is somewhat affordable given the sign and value of ζ_{price3} , which is significant at the 5% level. Finally, the sign of $\zeta_{price1na}$, $\zeta_{price2na}$, and $\zeta_{price3na}$ indicate that respondents who are more likely to give an answer other than “I don’t know” to the questions related to water price perception (*Price1*, *Price2*, *Price3*) are also more likely to consume tap and filtered water compared to bottled water. Overall, these results suggest that there is a significant price effect when consumers choose between bottled water, tap water, and filtered tap water.

The latent variable α_{price} increases with age (0.0116) and is found to be lower for women (-0.6281), which means that women are less likely to know the actual price of bottled price with respect to tap water. This is a surprising result, but it is in line with what is found in the data. Indeed, only 27.71% of the male respondents and 14.37% of the female respondents stated that the price of bottled water is about 100 times higher than the price of tap water. This difference between genders is significant at the 1% level according to a χ^2 test ($\chi^2(1) = 94.1904$, p -

Table 6
Estimation Results

Param.		Coeff.	Rob. T.	Param.		Coeff.	Rob. T.	
β_{tap} (Choice model)	<i>const</i>	1.4184	4.9	$\beta_{filtered}$ (Choice model)	<i>const</i>	0.2175	0.65	
	<i>nitrates</i>	-0.0283	-7.37		<i>nitrates</i>	-0.001	-0.23	
	<i>rainfall</i>	0.4168	2.24		<i>rainfall</i>	-0.2079	-0.92	
	<i>age</i>	-0.0171	-4.87		<i>age</i>	-0.0101	-2.6	
	<i>female</i>	0.3712	3.71		<i>female</i>	0.2178	1.92	
	<i>child</i>	0.0545	1.12		<i>child</i>	0.1375	2.54	
	<i>income</i>	-0.1066	-3.51		<i>income</i>	-0.0199	-0.63	
	<i>edu1</i>	-0.0886	-0.31		<i>edu1</i>	-0.0851	-0.27	
	<i>edu2</i>	-0.1315	-0.93		<i>edu2</i>	0.1148	0.72	
	<i>edu3</i>	-0.1149	-0.78		<i>edu3</i>	-0.0073	-0.04	
<i>edu4</i>	-0.0912	-0.72	<i>edu4</i>	-0.1261	-0.86			
<i>rural</i>	0.1183	1.18	<i>rural</i>	0.1429	1.25			
θ_{tap} (Choice model)	<i>price</i>	0.3718	5.55	$\theta_{filtered}$ (Choice model)	<i>price</i>	0.3565	5.53	
	<i>quality</i>	-0.4015	-5.07		<i>quality</i>	-0.2327	-2.75	
$\alpha_{quality}$ (Structural model)	<i>nitrates</i>	0.0037	1.19	α_{price} (Structural model)	<i>nitrates</i>	-0.0012	-0.56	
	<i>rainfall</i>	-0.1361	-0.99		<i>rainfall</i>	0.4483	3.94	
	<i>age</i>	-0.0214	-9.84		<i>age</i>	0.0116	5.45	
	<i>female</i>	0.4591	7.55		<i>female</i>	-0.6281	-10.03	
	<i>child</i>	-0.0084	-0.25		<i>child</i>	0.0291	0.94	
	<i>income</i>	-0.0606	-3		<i>income</i>	0.1045	6.02	
	<i>edu1</i>	0.0314	0.15		<i>edu1</i>	-0.542	-3.54	
	<i>edu2</i>	0.2876	3.07		<i>edu2</i>	-0.4236	-4.62	
<i>edu3</i>	0.2137	2.15	<i>edu3</i>	-0.2742	-3.09			
<i>edu4</i>	0.1478	1.71	<i>edu4</i>	-0.0985	-1.29			
<i>rural</i>	-0.1477	-2.13	<i>rural</i>	0.1875	2.9			
ζ (Measurement model)	<i>price1</i>	-0.5209	-1.98	τ (Measurement model)	<i>price na1</i>	0.097	1.2	
	<i>price2</i>	1.714	5.21		<i>price na2</i>	2.1759	9.87	
	<i>price3</i>	0.2644	3.96		<i>price na3</i>	2.4762	18.84	
	<i>price1na</i>	1.0592	9.17		<i>price2 1</i>	-3.6669	-9.98	
	<i>price2na</i>	1.3111	6.93		<i>price2 2</i>	-1.5717	-7.14	
	<i>price3na</i>	0.8185	6.88		<i>price2 3</i>	-0.1669	-1.26	
	<i>quality</i>	-0.9402	-10.72		<i>price2 4</i>	1.0783	8.07	
	<i>quality f</i>	1.2481	8.31		<i>price3</i>	-0.5555	-11.18	
	σ (LR)	<i>price1</i>	8.87		12.55	<i>Quality2</i>	-4.3866	-25.48
	Goodness-of-fit					<i>Quality3</i>	-1.1295	-11.58
Log-likelihood		-26707.99		<i>Quality4</i>	3.3091	26.63		
AIC		53,559.97		<i>quality f2</i>	-0.9654	-9.04		
BIC		54,003.65		<i>quality f3</i>	0.4526	4.23		

value = 0.000). Moreover, α_{price} increases with income. We also find that different levels of education have different effects on the value of the latent variable.

2.3. Other Findings From the Discrete Choice Model

The discrete choice model results also show that an increase in nitrates negatively affects the probability of drinking unfiltered tap water, while *rainfall* increases it. Older people are less likely to drink tap water and filtered

water compared to bottled water. Furthermore, women are less likely to drink bottled water, all other things being equal, and the probability of drinking tap water decreases as income increases. The parameters $\beta_{tap, drur}$ and $\beta_{filtered, drur}$ do not affect water consumption choice, although this doesn't mean that geographical location has no effect on water consumption choice given that geographical location has an impact on both $\alpha_{quality}$ and α_{price} as previously stated. This raises the issue of how to further interpret model output when the same variables are introduced in multiple model components. In what follows, we adapt the sample enumeration procedure suggested by Hess et al. (2018) to test different scenarios related to the reduction of nitrates in groundwater and changes in the level of education of the population.

2.4. Analysis of Scenarios

The process consists of predicting the likelihood that each of the 3,506 respondents included in the model chooses a given drinking water alternative in different scenarios. We test how nitrate levels found in groundwater affect choice probabilities by making predictions for when the level of nitrates found in groundwater is the highest of the sample for all respondents (62.81), the third quartile (23.74), the median level (15.07), the first quartile (6.91) and the minimum level found in the sample (0). The other variables are kept at their respective values in the sample to test these different scenarios *all else being equal*. We follow the same approach for testing how variations in education level affect choice probabilities. More precisely, we test five scenarios and make predictions for when all respondents have no degree, a middle school education, a high school education, hold a bachelor, and finally hold a master's degree or above, again *all else being equal*.

The scenarios considered were chosen because they are related to variables that can be acted upon using policy instruments. For example, decision makers can decide to increase nitrate depollution efforts or better inform respondents about water pricing. For each scenario and each respondent, 200 predictions are made by taking draws from a multivariate normal distribution where the means correspond to the model parameter values at convergence and the variance-covariance corresponds to the robust variance-covariance matrix of the model at convergence. This protocol is analogous to the Krinsky and Robb approach (Krinsky & Robb, 1986, 1990) which is commonly used to calculate standard errors for willingness to pay estimates derived from discrete choice models, only in this instance we calculate standard errors for the choice probabilities derived from the different scenarios considered. The results are reported in Figure 1 below (detailed results are provided in Appendix A (Table A1)).

Figure 1 shows that large changes in trends can be achieved with depollution efforts, although whether these efforts should be implemented in practice depends on cost-benefit analysis data, which are neither available nor the focus of this study. Moreover, we observe that the main drinking water choice for respondents with a BSc or a MSc degree or higher is tap water, while it is bottled water for the rest of the population. Of course, a higher education level does not necessarily suggest a higher knowledge of water pricing. Future research could investigate where and how different categories of the population gather information about water pricing, to further improve knowledge dissemination.

3. Conclusion

In this paper, we propose to use a hybrid choice model to assess the role of perception variables in drinking water choices, which has not previously been done in the received literature. In doing so, we address the issues posed by the use of perception variables in empirical works on pro-environmental and/or averting behaviors. We argue that such an approach will continue to prove useful not only in the field of environmental economics but also in all empirical work that seeks to associate individual economic choices with perception variables.

From an empirical point of view, we show that the perception of a poor quality of water resources encourages the consumption of bottled water. Although this result is not new, as a known predictor of drinking water choice, quality perception must be incorporated in our modeling work in order to control for its specific effect. Although the perception of water resources quality is mostly informed by sociodemographic variables and nitrate concentration in groundwater, other pollutants could have been considered, such as pesticides. However, available data on pesticides showed poorer granularity, and pesticide concentration was highly correlated with nitrate concentration. We further acknowledge some of the limitations of the data used in this study in a "Data integrity statement and limitations" section (see Appendix A).

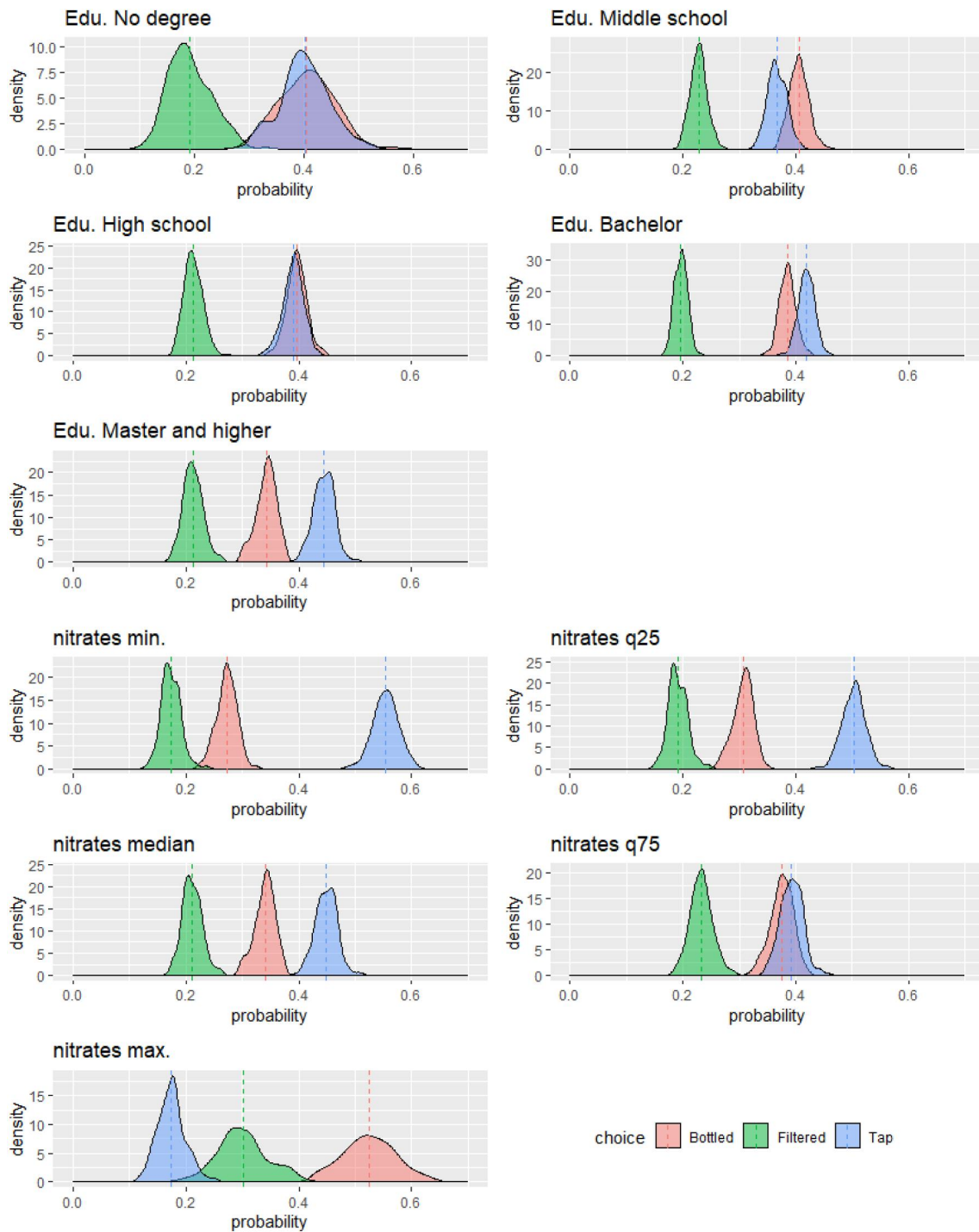


Figure 1. Sample enumeration results.

Regardless, this modeling strategy allows us to isolate the main effect of interest in this paper, which is the link between water price perception and drinking water choices. In fact, our econometric strategy makes it possible to show that people who are less likely to report the correct price difference between tap water and bottled water are less likely to consume tap water, which has not been shown previously. Taken together, our results suggest that people form perceptions, perhaps even beliefs, about tap water based on the information they have at their disposal, which influence their choice of drinking water at the decision step. However, these perceptions may be

far from the objective reality. Individuals underestimate the difference between bottled water and tap water, which raises “internalities” (Allcott & Sunstein, 2015). One source of these “internalities” is precisely cost (price) mis-perception (see Allcott, 2011, for the case of energy costs).

In the case of water price mis-perception, the associated welfare losses have to be assessed, and we argue that our modeling approach could help, as we are able to show how drinking water choices change in response to price perception. In 2012, in France, bottled water expenditures amount to approximately 3.5 billion euros (CGDD, 2014), which can be compared to the total billing of water for domestic use at the same time, which is about 7 billion euros (BIPE, 2019). Considering that only 1% of the latter is for drinking and that only 20% of the population knows the actual price difference between bottled water and tap water, welfare losses due to water price mis-perception are probably significant, and a promising area for future research. In addition, from a policy recommendation perspective, regulating these “internalities” could be done on the basis of the principles discussed by Allcott and Sunstein (2015).

Appendix A: Data Integrity Statement and Limitations

A1. Questionnaire Design

The questionnaire was designed by researchers from the *Commissariat général au développement durable* (CGDD), in collaboration with a major survey company (IFOP). Annex 1 of CGDD (2014) indicates that the final design of the survey was based on an iterative process informed by two pilot surveys involving 11 and 104 respondents respectively. Furthermore, CGDD (2014) indicates that this process ensured that questions were properly understood, which was also informed by analyzing response time data. We hypothesize that the first pilot survey corresponds to a pretest, while the second pilot survey is in line with the recommendations found in the literature on survey design. For example, Ruel et al. (2016, Chapter 6: Pretesting and Pilot Testing) state that: “A general rule of thumb is to pilot test the survey on 30 to 100 pilot participants.” We are thus confident that the wording of each question included in the final questionnaire is adequate for academic research purposes. However, we cannot completely rule out other potential flaws in the design of the questionnaire, which could have been addressed by pretesting the survey using cognitive interviews, as recommended by Presser et al. (2004). This would have contributed to inform whether the Likert scales used for measuring water quality perception need a mid-point response option, for example, Another potential improvement would have been to allow respondents to state that the price of tap water is acceptable for question *Price3* (*How expensive do you feel tap water is*). As a result and given that the knowledge of the public on how drinking water is produced and priced is often imperfect, we have sought to corroborate our data using other sources whenever possible, as further described below.

A2. Data Corroboration

Table 4 in the main text reports that 58% of the population does not know how much tap water costs; only 20% of the population is able to state a value in line with the actual price charged at the time of the survey (about €3.4 per m³; while the mean value given by the respondents regarding how much a cubic meter of tap water cost is €5.97, with a standard deviation of €8.89). Moreover, 56% of the respondents consider tap water to be “somewhat expensive,” while about 30% consider it to be “somewhat affordable” and 14% state that they do not know whether tap water is expensive or affordable. Other data sources can corroborate these findings. Indeed, for the past 25 years, a non-profit organization named the *Centre d'Information sur l'Eau* (<https://www.cieau.com/observatoire-de-leau/>) (“Water Information Center”) has published the results of an annual opinion survey entitled *Les Français et l'Eau* (“French people and Water”). The 2012 survey (CIEAU, 2012) includes several questions for which the wording is identical to the questions featured in the data we obtained from CGDD (2014). CIEAU (2012) reports that 61% of the respondents ($N = 1,023$) do not know the price of tap water. The mean value given by the remaining 49% is €5.30 (standard deviation not reported). Additionally, the CIEAU (2012) reported that 62% of the respondents considered tap water to be “somewhat expensive,” while 29% considered tap water to be “somewhat affordable” and 9% reported that they do not know whether tap water is expensive or affordable. These figures are in line with the descriptive statistics reported in Table 4 and are found to be very stable over a 6-year range (2006–2012). This provides evidence that some of the data on tap water price perception used in this paper are reliable, despite the inherent flaws and limitations of the use of secondary data.

Other tests were performed to confirm the coherence of our findings. For example, respondents who reported knowing how much one cubic meter of tap water costs are also 17.01% less likely to answer “I don’t know” to question *Price2* and 20.54% more likely to choose “100 times more expensive.” These results are significant at the 1% level. They are also 2.23% less likely (at the 10% level) to select the midpoint option (3—*Tap water is 5 times more expensive*), which could capture a weak “I don’t know” effect for the respondents who reported not knowing the price of one cubic meter of tap water. These results come from a multinomial logit model (Other modeling strategies such as generalized ordered models were also suitable, but we considered the multinomial logit to be a more straightforward approach to ensure the clarity of our findings) estimated on the 3,506 observations also used for estimating the hybrid choice model presented in Table 6.

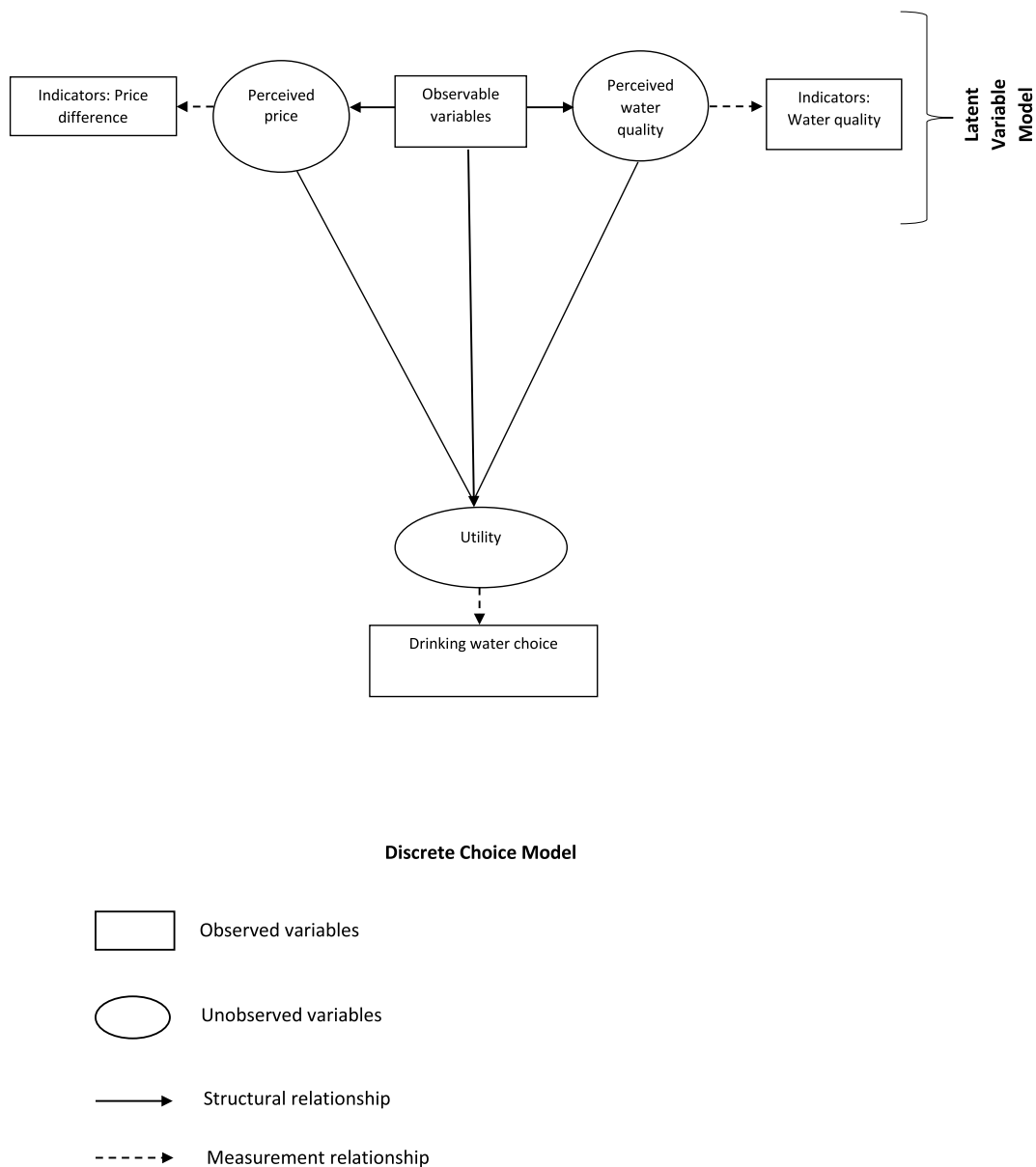


Figure A1. The hybrid choice model adapted from Ben-Akiva et al. (2002).

Table A1
Detailed Sample Enumeration Results

Nitrates level	Choice	Share	T-ratio	Difference with median level	
				Share	T-ratio
Max.	<i>Tap</i>	15.16%	7.96	-25.58	-13.71
	<i>Bottled</i>	56.24%	13.79	17.93	4.48
	<i>Filtered</i>	28.60%	7.46	7.64	1.99
q75	<i>Tap</i>	35.02%	37.04	-5.71	-9.82
	<i>Bottled</i>	42.19%	41.42	3.88	5.83
	<i>Filtered</i>	22.79%	27.13	1.83	3.16
Med.	<i>Tap</i>	40.74%	52.95		
	<i>Bottled</i>	38.31%	51.13		
	<i>Filtered</i>	20.96%	33.27		
q25	<i>Tap</i>	46.34%	47.58	5.6	9.42
	<i>Bottled</i>	34.54%	37.4	-3.77	-6.53
	<i>Filtered</i>	19.12%	24.15	-1.83	-3.85
Min.	<i>Tap</i>	51.15%	38.25	10.41	9.45
	<i>Bottled</i>	31.33%	25.51	-6.98	-6.87
	<i>Filtered</i>	17.52%	17	-3.43	-4.18

Education level	Choice	Share	T-ratio	Difference with master level and higher	
				Share	T-ratio
No degree	<i>Tap</i>	39.46%	8.78	-4.29	-0.91
	<i>Bottled</i>	41.00%	7.83	6.35	1.14
	<i>Filtered</i>	19.54%	4.82	-2.05	-0.47
Middle school	<i>Tap</i>	35.97%	20.4	-7.78	-3.09
	<i>Bottled</i>	41.06%	22.65	6.4	2.41
	<i>Filtered</i>	22.97%	14.02	1.38	0.58
High school	<i>Tap</i>	38.27%	17.54	-5.48	-2.05
	<i>Bottled</i>	40.18%	19.89	5.52	2.01
	<i>Filtered</i>	21.55%	11.35	-0.04	-0.02
Bachelor	<i>Tap</i>	41.16%	24.23	-2.59	-1.09
	<i>Bottled</i>	38.96%	23.45	4.3	1.9
	<i>Filtered</i>	19.88%	14.97	-1.71	-0.82
Master and higher	<i>Tap</i>	43.75%	21.42		
	<i>Bottled</i>	34.66%	17.08		
	<i>Filtered</i>	21.59%	11.89		

Table A2
Efficiency—Comparison Between the HCM and a Naïve MNL Counterpart

Model	Alternative	Obs. Freq.	Probability	Std. Err.	T-ratio
HCM	<i>Tap</i>	0.4044	0.4040	0.0079	51.23
	<i>Bottled</i>	0.3859	0.3855	0.0079	49.00
	<i>Filtered</i>	0.2096	0.2105	0.0064	32.72
Naive	<i>Tap</i>	0.4044	0.4036	0.0081	49.67
	<i>Bottled</i>	0.3859	0.3850	0.0083	46.32
	<i>Filtered</i>	0.2096	0.2114	0.0065	32.34

A3. Data Collection and Ethical Considerations

As an international polling and market research firm, IFOP is a member of ESOMAR (European Society for Opinion and Market Research) and abides by the ICC (International Chamber of Commerce)/ESOMAR international code. According to the ESOMAR website, “The Code ensures that researchers and analysts working with both traditional and new sources of data continue to meet their ethical, professional and legal responsibilities to the individuals whose data they use in research and to the clients and organizations they serve. It also is intended to protect the right of researchers to seek, receive, and impart information as stated in Article 19 of the United Nations International Covenant on Civil and Political Rights.” (Figure A1) (Table A1) (Table A2).

Efficiency has been measured by following the method outlined by Abou-Zeid and Ben-Akiva (2014) for comparing a hybrid choice model to its standard counterpart. More precisely:

- For each model, the distributions of the choice probabilities have been simulated by taking 200 draws from the multivariate distribution of the parameter estimators (using parameter estimates as means and the robust variance covariance matrix).
- This gives the standard errors of the choice probabilities.
- The model for which the standard errors are the smallest is the most efficient (in practice, T-ratios are used for comparing results, as predicted probabilities vary slightly across models).
- We find that the HCM presented in the article is marginally more efficient than a naïve approach.

Data Availability Statement

Data and supporting material are available at <https://doi.org/10.17605/OSF.IO/6PQ9E>.

Acknowledgments

We thank the three anonymous reviewers as well as Paul Veyronnet for their insightful comments. The opinions and any omissions in this paper remain our responsibility.

References

- Abou-Zeid, M., & Ben-Akiva, M. (2014). Hybrid choice models. In *Handbook of choice modelling* (pp. 383–412).
- Abrahams, N., Hubell, B., & Jordan, J. (2000). Joint production and averting expenditure measures of willingness to pay: Do water expenditures really measure avoidance costs? *American Journal of Agricultural Economics*, 82(2), 427–437. <https://doi.org/10.1111/0002-9092.00036>
- Allcott, H. (2011). Consumers’ perceptions and misperceptions of energy costs. *American Economic Review, Papers and Proceedings*, 101(3), 98–104. <https://doi.org/10.1257/aer.101.3.98>
- Allcott, H. (2013). The welfare effects of misperceived product costs: Data and calibrations from the automobile market. *American Economic Journal: Economic Policy*, 5(3), 30–66. <https://doi.org/10.1257/pol.5.3.30>
- Allcott, H., & Sunstein, C. R. (2015). Regulating externalities. *Journal of Policy Analysis and Management*, 34(3), 698–705. <https://doi.org/10.1002/pam.21843>
- Amaris, G., Hess, S., Gironás, J., & de Dios Ortúzar, J. (2021). Using hybrid choice models to capture the impact of attitudes on residential greywater reuse preferences. *Resources, Conservation and Recycling*, 164, 105171. <https://doi.org/10.1016/j.resconrec.2020.105171>
- Artell, J., Ahtiainen, H., & Pouta, E. (2013). Subjective vs. objective measures in the valuation of water quality. *Journal of Environmental Management*, 130, 288–296. <https://doi.org/10.1016/j.jenvman.2013.09.007>
- Beaumais, O., & Veyronnet, P. (2017). Hétérogénéité des préférences individuelles, perception de la qualité et choix d’alimentation en eau potable en France. Un Logit multinomial à facteurs d’échelle. *La Revue Économique*, 68(6), 1077–1102. <https://doi.org/10.3917/reco.pr2.0102>
- Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., et al. (2002). Hybrid choice models: Progress and challenges. *Marketing Letters*, 13(3), 163–175. <https://doi.org/10.1023/a:1020254301302>
- BIPE. (2019). *Public water and waste water services in France—Economic, social and environmental data*. Report (7th ed., p. 112). BIPE.
- Bontemps, C., & Nauges, C. (2009). Carafe ou bouteille? Le rôle de la qualité de l’environnement dans la décision du consommateur. *Économie & Prévision*, 188(2), 61–79. <https://doi.org/10.3406/ecop.2009.7900>
- Bontemps, C., & Nauges, C. (2016). The impact of perceptions and measurement bias in averting-decision models: An application of the special regressor method to drinking water choices. *American Journal of Agricultural Economics*, 98(1), 297–313. <https://doi.org/10.1093/ajae/aav046>
- Brent, D. A., & Ward, M. B. (2019). Price perceptions in water demand. *Journal of Environmental Economics and Management*, 98, 21. <https://doi.org/10.1016/j.jeem.2019.102266>
- Brent, D. A., & Wichman, C. J. (2022). Do behavioral nudges interact with prevailing economic incentives? Pairing experimental and quasi-experimental evidence from water consumption. <https://doi.org/10.2139/ssrn.4162438>
- Buckell, J., Hensher, D. A., & Hess, S. (2021). Kicking the habit is hard: A hybrid choice model investigation into the role of addiction in smoking behavior. *Health Economics*, 30(1), 3–19. <https://doi.org/10.1002/hec.4173>
- Budziński, W., & Czajkowski, M. (2022). Endogeneity and measurement bias of the indicator variables in hybrid choice models: A Monte Carlo investigation. *Environmental and Resource Economics*, 83(3), 605–629. <https://doi.org/10.1007/s10640-022-00702-0>
- CGDD. (2014). Ressources en eau: perception et consommation des français. Résultats d’enquête. *Études et Documents*, 106, 1–80.
- CIEAU. (2012). *Les français et l’eau. Baromètre 2012. 17^{ème} édition*. Centre d’information sur l’eau/tns sofiés.
- Doria, M. F. (2006). Bottled water versus tap water: Understanding consumers’ preferences. *Journal of Water and Health*, 4(2), 271–276. <https://doi.org/10.2166/wh.2006.0023>
- Dupont, D., Adamowicz, W. L., Krupnick, A., & Krupnick, A. (2010). Differences in water consumption choices in Canada: The role of sociodemographics, experiences, and perceptions of health risks. *Journal of Water and Health*, 8(4), 671–686. <https://doi.org/10.2166/wh.2010.143>
- Groothuis, P. A., Mohr, T. M., Whitehead, J. C., Cockerill, K., William, P., Anderson, W. P., & Gu, C. (2021). Measuring the direct and indirect effect of scientific information on valuing storm water management programs with a hybrid choice model. *Water Resources Research*, 57(5), 1–17. <https://doi.org/10.1029/2020wr027552>

- Hess, S., & Palma, D. (2019). Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, 32, 100170. <https://doi.org/10.1016/j.jocm.2019.100170>
- Hess, S., Spitz, G., Bradley, M., & Coogan, M. (2018). Analysis of mode choice for intercity travel: Application of a hybrid choice model to two distinct US corridors. *Transportation Research Part A: Policy and Practice*, 116, 547–567. <https://doi.org/10.1016/j.tra.2018.05.019>
- Hess, S., & Stathopoulos, A. (2013). Linking response quality to survey engagement: A combined random scale and latent variable approach. *Journal of Choice Modelling*, 7, 1–12. <https://doi.org/10.1016/j.jocm.2013.03.005>
- Javidi, A., & Pierce, G. (2018). U.S. households' perception of drinking water as unsafe and its consequences: Examining alternative choices to the tap. *Water Resources Research*, 54(9), 6100–6113. <https://doi.org/10.1029/2017wr022186>
- Johnstone, N., & Serret, Y. (2012). Determinants of bottled and purified water consumption: Results based on an OECD survey. *Water Policy*, 14(4), 668–679. <https://doi.org/10.2166/wp.2011.048>
- Kim, J., Rasouli, S., & Timmermans, H. (2014). Hybrid choice models: Principles and recent progress incorporating social influence and nonlinear utility functions. *Procedia Environmental Sciences*, 22, 20–34. <https://doi.org/10.1016/j.proenv.2014.11.003>
- Krinsky, I., & Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, 68(4), 715–719. <https://doi.org/10.2307/1924536>
- Krinsky, I., & Robb, A. L. (1990). On approximating the statistical properties of elasticities: A correction. *The Review of Economics and Statistics*, 72(1), 189–190. <https://doi.org/10.2307/2109761>
- Levêque, J. G., & Burns, R. C. (2017). Predicting water filter and bottled water use in Appalachia: A community-scale case study. *Journal of Water and Health*, 15(3), 451–461. <https://doi.org/10.2166/wh.2017.219>
- Lloyd-Smith, P., Schram, C., Adamowicz, W., & Dupont, D. (2018). Endogeneity of risk perceptions in water expenditure models. *Environmental and Resource Economics*, 69(2), 217–246. <https://doi.org/10.1007/s10640-016-0075-6>
- Orgill, J., Jeuland, A., Brown, J., & Shaheed, A. (2013). Water quality perceptions and willingness to pay for clean water in peri-urban Cambodian communities. *Journal of Water and Health*, 11(3), 489–506. <https://doi.org/10.2166/wh.2013.212>
- Orgill-Meyer, J., Jeuland, M., Albert, J., & Cutler, N. (2018). Comparing contingent valuation and averting expenditure estimates of the costs of irregular water supply. *Ecological Economics*, 146, 250–264. <https://doi.org/10.1016/j.ecolecon.2017.10.016>
- Pattanayak, S. K., Yang, J.-C., Whittington, D., & Bal Kumar, K. C. (2005). Coping with unreliable public water supplies: Averting expenditures by households in Kathmandu, Nepal. *Water Resources Research*, 41(2), 1–11. <https://doi.org/10.1029/2003wr002443>
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, 68(1), 109–130. <https://doi.org/10.1093/poq/nfh008>
- Ruel, E., Wagner, W. E., III., & Gillespie, B. J. (2016). *The practice of survey research: Theory and applications*. Sage.
- Terza, J., Basu, A., & Rathouz, P. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modelling. *Journal of Health Economics*, 27(3), 531–543. <https://doi.org/10.1016/j.jhealeco.2007.09.009>
- Train, K. E. (2015). Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples. *Journal of Choice Modelling*, 16, 15–22. <https://doi.org/10.1016/j.jocm.2015.09.003>
- Vásquez, W. F., Mozumder, P., & Franceschi, D. (2015). Water quality, household perceptions and averting behavior: Evidence from Nicaragua. *Water Economics and Policy*, 4(21), 1550011. <https://doi.org/10.1142/s2382624x15500113>
- Vij, A., & Walker, J. L. (2016). How, when and why integrated choice and latent variable models are latently useful. *Transportation Research Part B: Methodological*, 90, 192–217. <https://doi.org/10.1016/j.trb.2016.04.021>
- Viscusi, K. W., Huber, J., & Bell, J. (2015). The private rationality of bottled water drinking. *Contemporary Economic Policy*, 33(3), 450–467. <https://doi.org/10.1111/coep.12088>
- Whitehead, J. C. (2006). Improving willingness to pay estimates for quality improvements through joint estimation with quality perceptions. *Southern Economic Journal*, 73(1), 100–111. <https://doi.org/10.2307/20111876>