Online Appendix – Full study report

**Fundão Dam Incident**

**Economic Valuation of Environmental and Heritage Damages**

**Contingent Valuation Study**

**Final Version**

**August, 02nd 2020**

**Authors**

**Brazilian Team** **International Team**

Ronaldo Seroa da Motta Norman Meade

Ramon Arigoni Ortiz Mike Welsh

Luis Cesar Perisse Mikołaj Czajkowski

**Table of contents**

Executive Summary 4

1. Introduction 9

2. International Experience in Natural Resource Damage Assessment 11

2.1. Introduction to Economic Valuation Methods 11

2.2. How Economic Valuation Has Been Applied Elsewhere for NRDA 13

2.3. The BP Deepwater Horizon oil spill NRDA 14

3. The Adopted Methodology Approach 15

3.1. Methods for Estimating Environmental Values 15

Choice Experiments 15

Contingent Valuation versus Choice Experiments 16

3.2. Guidelines for Conducting a CV Study 17

3.3. Statistical analysis framework 19

Estimation of mean WTP based on binary choice data 20

Non-parametric estimator of mean WTP 23

Parametric estimator of mean WTP 25

Estimation 25

Parametric distributions to consider 26

Zero-inflation 27

Simulation of mean WTP 27

Models for investigating sensitivity of voting behavior to explanatory variables 28

4. General Survey Design Criteria 29

4.1. Rationale for City Selection 31

4.2. Survey Instrument 31

4.3. Main Challenges 33

Impacts to lifestyles, municipalities and infrastructure 34

Corporate Assets, Criminal/Negligent Behavior and Accountability 34

Recreation 35

Impacts to Commercial Fishing and Agriculture 35

Impacts to Local Water Supplies 35

Loss of Life 36

Dry Tailings Technology 37

Confidence That the Program Would be Carried Out 38

4.4. Damage Scenario 40

5. Pilot Studies Analysis 44

5.1. Tax Amounts Choice and Acceptance Rates Analysis 44

5.2. WTP analysis 46

5.3. Analysis of the indicators of validity of the study 47

5.4. Questionnaire Changes 48

6. Final Survey Administration and Results 50

6.1 Operational indicators 50

Sampling 50

Survey administration and verification 54

Intercept interviews 56

Interview Dispositions 57

6.2 Survey Results 58

Non-parametric estimate of population WTP 58

Parametric estimate of WTP 59

Aggregate Lost Value 61

Analysis of the indicators of validity of the study 62

Sensitivity to tax amounts and income levels 63

Attitudinal and debriefing questions 66

Quality-indicating questions 68

7. Conclusions 70

Appendix 1: Interview Protocol Report 72

Appendix 2: Regression used to impute missing income responses 76

Appendix 3: Analysis of reasons for voting ‘against’ and the identification of protest responses 79

Appendix 4: Sensitivity of voting behavior to explanatory variables specified as binary probit 85

References 91

# Executive Summary

The Fundao Dam Incident (FDI), which occurred on November 5, 2015 in Mariana, Minas Gerais, Brazil, caused the sudden release of over 40 million m3 of mine tailings and mud into the Doce River Basin.

Apart from the deaths to people, the loss in economic activities and adverse impacts to property and municipal infrastructure, the injuries to environmental and heritage/cultural resources within the river basin were also severe. The greatest damage in the Doce River Basin occurred within the first 110 km downstream of the (collapsed) Fundao Dam where approximately 50% of the tailings and mud severely impacted the river banks and riverside forest and caused a large number of animal deaths. Some 25% of the tailings and mud was stopped behind the Risoleta Neves hydro-electric dam and, consequently, reduced damage intensity downstream with no further adverse impacts to the riverside forest. The remaining tailings and mud (25%) went all the way down river reaching the sea, some 660 km away.

Following the FDI, the Federal Public Prosecutor’s Office commissioned a comprehensive damage assessment of all the environmental injuries and harm caused to heritage resources and their lost value.

The goal of this study was to estimate a monetary value for losses to the citizens of Brazil from the injuries to the environment and heritage resources caused by the FDI. The scope of the study did not include valuing loss of life, human health impacts, private losses to individuals and companies, or adverse impacts to municipal governments and public infrastructure.

The study adopted a stated preference methodology, known as the contingent valuation (CV) method, following standard implementation practices prescribed in the economics literature (e.g., Arrow et al, 1992, Freeman, Herriges and Kling (2014); Champ, Boyle and Brown (2017); Johnston et al. (2017)). It also followed closely the procedures used by Bishop et al. (2017) to estimate the monetary value of the economic loss suffered by American households as a result of injuries to natural resources caused by the BP Deep Water Horizon oil spill in the Gulf of Mexico in 2010.

The study team was composed of Brazilian and international experts with extensive academic and governmental experience carrying out environmental (i.e. non-market) valuation studies, including natural resource damage assessments. The study team was led by Ronaldo Seroa da Motta and Ramon Ortiz, who previously undertook several studies and authored articles on environmental valuation in Brazil. The international team was composed of Norman Meade, Mike Welsh and Mikołaj Czajkowski, who are environmental economists with expertise in conducting contingent valuation studies, including for natural resource damage assessments. Survey fieldwork was conducted by Quest Inteligencia de Mercado, led by Luis Cesar Perisse, which previously participated in a number of survey research projects in Brazil. Members of the team worked closely with physical and biological scientists at Lactec to develop the damage scenario of the environmental/heritage services injuries resulting from the FDI that was used in the CV study.

The CV study estimates the maximum willingness to pay (WTP) of Brazilian households to avoid a specified subset of the injuries to environmental/heritage resource services as occurred in the FDI. This estimate represents how much the Brazilian society values the moral damage of the psychological and emotional impacts in face of the damages caused to the environment and to the historical cultural heritage and also the losses of services and resources throughout and at the end of the recovery period. That is, how much the Brazilian society values these damages in terms of loss of well-being, even aware that most of the damages will be recovered by the responsible parties.

The survey instrument is a questionnaire presenting respondents with an object of choice, a context in which a choice is to be made, and a vote either in favor or against the object of choice. First the injury scenario was presented, with the FDI environmental and heritage resource injuries described in detail, then respondents were asked to vote whether or not they would be willing to pay to finance a program (the object of choice) to prevent one future tailings dam failure that would have the same environmental/heritage impacts as occurred in the FDI.

Each element of the questionnaire was composed of carefully tested wording and graphic information shown to be understandable to respondents and which did not make them feel pushed to make a choice either for or against the program. This ensured that the choice expressed by respondents reflected their personal assessment of the cost of the dam safety program relative to the benefits that would be provided if it was carried out. We note that the design of the study satisfied the necessary conditions for incentive compatibility, which means that a single dominant strategy for rational respondents was to reveal their preferences truthfully.

A conservative approach to survey design and analysis was followed by the study team throughout this project. That is, whenever it was faced with deciding on a particular course of action between two or more equally compelling (from a technical, scientific and/or practicable perspective) alternatives, where possible the study team tried to choose the one that would more likely result in either lowering, or having a neutral impact, on the final estimate of total damages (when compared to the other alternatives under consideration). In keeping with this conservative approach, for example, the wording “incident,” instead of “accident” or “disaster,” was used throughout to avoid any perception that would anticipate the injury levels before their detailed presentation.

The study commenced in October 2018 with development of the survey questionnaire and sample design. The final survey was initiated in November 2019 and concluded in early March 2020.

Each injury was described in the questionnaire using quantitative indicators for losses and expected recovery times, as provided by Lactec scientists. It was also made very clear to respondents that no species had become extinct or disappeared permanently due to the FDI. Estimated recovery times for each injury were typically specified in ranges spanning several years.

Survey respondents were told that one more incident like the Mariana incident would likely occur in the next 10 years, but that this future incident could be prevented with the adoption of a dry tailings’ safety program. They were told that in order to carry out the program to prevent a tailings dam failure incident from occurring within the next 10 years, the government would have to pay for the capital costs of the program and that mining companies would pay for the operational expenses, but that the government would not do so unless a majority of respondents to this survey voted in favor of it. Funding of this program would come from a temporary increase in the tax on household electricity bills, to be paid in ten, equal monthly instalments. To mitigate concerns about the tax collection and potential program mismanagement, a Dry Tailings Fund would be created with several types of administrative controls, obligations and guarantees applied.

Each respondent was asked to vote for or against the proposed tax increase, with a random tax amount selected from a previously determined possible range. The safety program and voting narratives were hypothetical, but respondents were not told that in order to create appropriate conditions for them to reveal their true preferences for or against the program.

To develop the final version of the questionnaire, initial drafts were subjected to extensive pretesting using a series of focus groups and personal/one-on-one interviews. At each stage of the pretesting, research team members evaluated the materials in terms of comprehension and plausibility for the participants and adjusted survey wording and materials as needed.

In addition, three pilot tests of the draft survey instrument were conducted. These pilot tests took place in the field, allowing for the evaluation of survey implementation procedures, an assessment of alternative tax payment amounts, and to get a preliminary understanding of how well the survey instrument was performing on test samples of respondents in various locations around Brazil.

These extensive testing procedures made it possible to minimize non-economic motivations or considerations when survey respondents voted for or against the program (object of choice).

Our sampling targeted the population of urban, income-earning, households located in 10 major metropolitan regions and municipalities across Brazil, with the goal of obtaining 5,200 completed survey interviews. The rural population of Brazil, which accounts for 14% of the country’s total, was not surveyed due to logistical reasons arising from its spatial dispersion.

Survey data collection was undertaken following standard, high-quality, survey research protocols. The fieldwork coordination team, QUEST Inteligencia de Mercado, based in São Paulo, was responsible for training local interviewers, supervising survey administration, verifying completed interviews and preparing the final dataset (see Appendix 1 for more details).

Completed interviews were carefully verified. The verification process began in the field with local teams, composed of at least one supervisor and two reviewers/verifiers, evaluating at least 20% of the completed interviews. Quality control measures for each completed interview employed consistency checks, such as the time respondents took to answer each question and the total interview time, the declared income compared to other household characteristics and the number of people contributing to household income.

The mean WTP can be estimated based on individual votes in a simulated referendum in which respondents are given the description of the proposed policy and its effects, as well as the cost to them, if the policy is implemented.

There are two approaches to estimating mean WTP from a CV study that is based on a referendum question, each with different strengths. The non-parametric approach estimates a lower bound of mean WTP and does not rely upon any parametric inferences (statistical assumptions). The parametric approach also produces an estimate of mean WTP, but it relies on additional statistical structure imposed by the use of a parametric model, particularly the necessary assumptions about vote probabilities between the different tax amounts used in the experimental design (i.e. the specified tax amounts).

Our non-parametric estimator of mean WTP was based on the work of [Lewbel (2000)](#_heading=h.1d96cc0) and [Watanabe (2010)](#_heading=h.3bj1y38). For the parametric estimate, we considered a large number of candidate parametric distributions including the zero-inflated model ([Gurmu and Trivedi, 1996](#_heading=h.zu0gcz)).

For both the parametric and non-parametric models, it was found that the percent of respondents voting “against” the program increased significantly as the tax amount increased. This is consistent with economic theory. To further demonstrate the validity of our study, we also tested the sensitivity of the votes to (i) income levels, (ii) socio-demographic characteristics, (iii) attitudinal characteristics; (iv) quality-related variables resulting from the survey; and (v) for the parametric estimates of WTP, sensitivity to econometric treatment of the data (model uncertainty). Overall, our analysis confirmed that data collected using our survey was well-behaved from an economic theory perspective.

Mean WTP values are estimated with parametric and non-parametric models. The non-parametric model approximates the distribution of WTP using a step function based on the observed proportions of respondents voting “against”. Non-parametric assumes conservatively rejection rates observed for the lower value of the interval and, therefore, it always generates a lower mean value regardless of the distribution function assumed for the parametric model. It is, however, unclear whether the mean WTP from the parametric approach results in an over or under-estimate of the true population mean WTP.

The non-parametric analysis implies a lower-bound estimate of mean WTP of 563.38 reais per urban household, with a 95% confidence interval (range) of 508.39 - 618.37 reais. The parametric analysis instead implies a higher mean WTP of 945.78 reais per urban household, with a 95% confidence interval of 798.11 – 1,093.45 reais.

An estimate of the average DAP makes it possible to calculate the value of damage to the environment and historical / cultural heritage resulting from the IBF for the Brazilian society that represents the moral damage and environmental and cultural heritage services and resources lost throughout and at the end of the recovery activities.

Calculating the aggregate value of damages for the society requires multiplying the mean WTP by the total number of urban households in Brazil (which represents 86% of the total number of households found in Brazil).

Considering the 55,949,131 households in Brazil, **our conservative estimate (lower limit) of the environmental damage and the historical and cultural heritage caused to the Brazilian society by the FDI is R $ 31.52 billion** (with a 95% confidence interval of 28.44 - 34.60 billion reais), even after all the technically feasible recovery actions.

# Introduction

The Fundao Dam Incident (FDI)[[1]](#footnote-1) occurred on November 5, 2015 in Mariana, Minas Gerais. The dam was operated by the Samarco mining company. It was part of its Germano Iron Ore Mining Complex. FDI caused the sudden release of over 40 million m3 of mine tailings and mud into the Doce River Basin.

The incident caused the deaths of 19 people leaving two thousand people homeless and destroying farming properties and the infrastructure of nearby villages. It also adversely affected agriculture, livestock, fisheries, and tourism.

Apart from the deaths to people, the loss in economic activities and adverse impacts to property and municipal infrastructure, the injuries to environmental and heritage/cultural resources within the river basin were also severe. The greatest damage in the Doce River Basin occurred within the first 110 km downstream of the (collapsed) Fundao Dam where approximately 50% of the tailings and mud severely impacted the riverbanks and riverside forest and caused a large number of animal deaths. Some 25% of the tailings and mud was stopped behind the Risoleta Neves hydro-electric dam and, consequently, reduced damage intensity downstream with no further adverse impacts to the riverside forest. The remaining tailings and mud (25%) went all the way down river reaching the sea, some 660 km away.

In areas where mud impacts were higher, the food chain and animal habitats were more impaired. Some animals died by burial or lack of food, but there was no permanent disappearance or extinction of any species. The mud flow also harmed portions of the historical and cultural heritage assets located near the collapsed dam.

In 2016, a Brazilian court ordered Samarco to pay 12 billion reais as a provisional indemnity for the environmental and heritage damages. In addition, at the request of the Federal Public Ministry, it was determined to estimate the amount of damages that could be paid to society as compensation for moral damage and losses of services and resources during and at the end of the recovery activities.

Consequently, the Federal Public Prosecutor’s Office commissioned a comprehensive damage assessment of all the environmental injuries and harm caused to heritage resources and their lost value. The damage assessment framework considered two separate valuation categories: (1) socio-economic and (2) environmental and heritage/cultural resource impacts.

Socio-economic impacts are comprised of market-based economic activities, including reduced income, and property and infrastructure losses which occurred in the affected area.

The scope of this study encompasses non-market, passive (or existence) and direct, value losses resulting from injuries to environmental and heritage/cultural resources. This report describes the contingent valuation (CV) study conducted to estimate how much the urban Brazilian population values the FDI environmental and heritage/cultural impacts. As will be later discussed in detail, the natural resource damage assessment covered losses to aquatic biota, terrestrial wildlife, birds and riparian forest, as well as to heritage resources and related activities.

The CV study team was composed of Brazilian and international experts with extensive academic and governmental experience carrying out environmental (i.e. non-market) valuation studies, including natural resource damage assessments. The study team was led by Ronaldo Seroa da Motta and Ramon Ortiz, who previously undertook several studies and authored articles on environmental valuation in Brazil. Norman Meade, Mike Welsh and Mikołaj Czajkowski are environmental economists with expertise in conducting contingent valuation studies, including for natural resource damage assessments. Survey fieldwork was led by Quest Inteligencia de Mercado[[2]](#footnote-2), which previously participated in a number of survey research projects in Brazil. Members of the team worked closely with physical and biological experts at Lactec to develop the environmental/heritage resource injury scenario used in the study.

The study commenced in October 2018 with development of the survey questionnaire and sample design. Initially, several rounds of focus groups and in-depth personal interviews were conducted. They were followed in 2019 by three large-scale pilot surveys. The final, or main survey, began in November 2019 and concluded in early March, 2020.

The next section of the report provides a brief overview of the history of natural resource damage assessment. This is followed by a discussion of standard, non-market damage valuation techniques, including a detailed description of the adopted methodological approach. Next, the general design of the study is described, along with a discussion of issues that arose during the process of developing the questionnaire, including results from the focus groups, in-depth interviews and pilot surveys. Results of field surveys are then analyzed with a presentation of estimates of lost value, or damages, and final conclusions. Three technical appendices are provided at the end to report covering missing value treatments, protest votes and interview protocols.[[3]](#footnote-3)

# International Experience in Natural Resource Damage Assessment

This section addresses the experience in natural resource damage assessment elsewhere with emphasis on the oil spill case of the BP Deepwater Horizon.

## Introduction to Economic Valuation Methods

The natural environment provides a wide range of goods and services. Some of the goods and services are private, with well-defined prices and quantities, which can be individually owned or appropriated, such as commercial fish. Others are public, such as rivers and parks, which are collectively owned and are not readily traded in markets. Commercial fish can be captured and sold in an organized market and fishermen can earn income from it. Profits to the fishermen net of production costs are known as producer surplus. Buyers of the fish can consume it to obtain nutrition and if the value of the nutrition is greater than the fish price, economists refer to it as consumer surplus. The total market value of the fish in this example is equal to the sum of the producer and consumer surplus.

The consumer surplus of many environmental resources is comprised of market and non-market goods and services. Non-market goods and services are typically unpriced and not traded in organized markets. Using the fish from the above example, the total consumer surplus of the fish is made up of market and non-market values. The market component can be thought of as the value received by the consumers of the fish who buy it to eat and the non-market component as the satisfaction/value obtained by recreational fishermen from angling and the general public who receive satisfaction/value from knowing the fish simply exist (i.e. existence value), or who like to watch them in the water (viewing value) and/or from its contribution to the proper functioning of the ecosystem. The use or consumption of some non-market public goods is revealed through individual behavior, such as visits to a public park, or a trip to a river to recreationally fish. Passive use of an environmental resource does not necessarily reveal itself in individual behavior. For example, someone can enjoy knowing the park exists without visiting it or knowing that there are fish in a given water body without going there to fish or observe them. When someone’s behavior demonstrates a use, and therefore some level of satisfaction/value for an environmental resource’s goods or services, economists label this as preference revealing. From revealed preferences economists are able to value non-market goods and services. When someone enjoys an environmental resource passively, they doo not necessarily reveal their preferences for the goods or services provided. In this latter case, it is necessary to ask users to state what their preferences are in order to value them.

The concept of economic value is well defined under the economic theory of consumer behavior. Individual choices in markets reveal individual preferences. The choices imply tradeoffs individuals are willing to make, as in a market, where something of value is foregone in order to obtain something else of value. This tradeoff allows analysts to estimate value from clearly defined objects of choice and the consequences of those choices. When an individual gives up specific sum of money in order to obtain an object of choice the monetary payment represents a lower bound on the value of that object.

In the case of the public, as opposed to private, environmental and heritage/cultural goods and services that were lost due to the FDI, there is no well-organized market comprised of traded quantities of these goods and services with well-defined prices that allows economists to directly estimate the value of what was lost. Instead, they have to construct a market to allow representative members of the public to make choices and express their preferences in a referendum-type valuation mechanism called a stated preference method. This valuation approach is described later.

Under U.S. law (and some European laws), private businesses and public entities, such as water utilities, can often claim damages and receive compensation for lost producer surplus or the cost of cleanup and disruption in public services from responsible parties who cause environmental harm. These are usually market-based losses of goods and services with (relatively) well defined prices. Public goods losses, however, such as impacts on parks and rivers, which are owned collectively and are not traded in organized markets, are unpriced. They are considered to be non-market goods and services. Generally, individuals cannot make private legal claims for compensation of losses to public goods and services. Rather a governmental entity, sometimes called trustee, usually has to make them. And because they are unpriced, it can be very challenging to value such legal claims. Natural resource damage assessment (NRDA) is a scientific and legal process to determine the amount of compensation owed to the public for injuries to natural resources caused by a release of hazardous substances so society is just as well off as it was before the release occurred.

The goal of this study is to estimate the lost value to the public from the FDI, which is a monetary measure of the public’s losses due to the injuries to the environmental and heritage/cultural resources from the time of the incident in 2015 considering moral damage and services and resources lost during and at the end the recovery activities. The NRDA focuses on estimating the value of the loss in consumer surplus to the public due to the adverse impacts on non-market environmental, and heritage/cultural goods and services. The estimate of lost total value is based on the willingness to pay (WTP) of survey respondents (a representative sample of 5,200 Brazilian households) for a program to prevent another dam failure like the FDI with very similar injuries to environmental and heritage/cultural resources.

As mentioned above, this NRDA does not include market-based business or private losses to individuals, groups, and companies, nor decreases in profits and wages. It also excludes the value of the loss of life and adverse health impacts and the cost of cleanup and restoration of the injured environment and the increased costs of providing public services and the loss of tax revenues.

## How Economic Valuation Has Been Applied Elsewhere for NRDA

The first known NRDA undertaken for an environmental incident took place following the Santa Barbara Oil Spill off the Coast of California in 1969. The study was an attempt by economists to quantify the value of non-market losses to recreationists from the loss of use of beaches that were coated in oil for many months from a leaking pipeline connected to an offshore oil drilling platform. In 1978, an international team of economists and physical scientists from France and the United States undertook a comprehensive NRDA of the environmental and economic consequences of Amoco Cadiz tanker spill on the coast of Brittany, France. Crude oil released from the grounded tanker coated hundreds of miles of coastline disrupting commercial and recreational uses of marine resources, killing countless quantities of sea birds and other marine life, while adversely impacting the recreational use of coastal resources for over two years.

After a number of highly-publicized and injurious oil spills and other hazardous waste releases occurred in the United States in the 1970s, the Comprehensive Environmental Response, Compensation and Liability Act of 1980 (CERCLA) – sometimes known as “The Superfund” – was enacted into law by the U.S. Government. Among other things, this law created the ability, or requirement, of State Governments and certain Federal Agencies to act as trustees on behalf of the public to sue responsible parties for damages for the loss of use of natural resources and the environment as a result of releases of certain hazardous substances. Recoveries for losses under CERCLA are limited to the total value of the public use of injured environmental and natural resources, and do not include private, commercial losses (producers’ surplus) and must be used for restoration of the injured natural resources. Recoveries can also include the cost to the public of cleanup. Once CERCLA was passed, there followed a number of NRDA cases filed against responsible parties in U.S. federal courts for injuries to the environment and the loss of use on the part of the public from hazardous waste releases from industrial operations, leaking mineral mining sites, and oil and hazardous substances spills from pipelines, tankers and other forms of transportation.

The methods for assessing injuries to environmental resources and the cost of the damages to the public began to evolve rapidly under CERCLA. That evolution in applied natural science and economics continued with the extensive NRDA conducted for the Exxon Valdez oil spill of 1989. With the settlement of public claims in that case reaching a half-billion dollars, industry began to vigorously challenge the methodologies of the trustees in assessing injuries and damages in subsequent cases. That trend continues today, wherever NRDA cases are implemented by public trustees, including in Europe and South America. These trends in the evolution of applied science and economics and the opposition by responsible parties to claims for natural resource damages are well-demonstrated in the case of the NRDA conducted for the BP Deepwater Horizon oil spill of 2010 in the U.S. Gulf of Mexico, as described below.

## The BP Deepwater Horizon oil spill NRDA

Following the 2010 BP Deepwater Horizon oil spill in the Gulf of Mexico (GOM), federal and state trustee agencies in the U.S. undertook a state-of-the-art NRDA of the effects of the oil spill on the environment and the resulting economic damages to the public. The economic damage study included estimates of losses to the public for both direct use, such as beach recreation impacts, and passive use, such as the death of seabirds and other biological organisms injured or killed by the oil. Claims by those engaged in business activities that were adversely affected, such as tourism and commercial fisheries, were not included in the public NRDA, but rather were handled in separate, private legal actions.

Trustees undertook over a hundred studies to quantify the injuries to the environment caused by the spill. The economic assessment used a survey-based, stated preference approach known as contingent valuation (CV). The study interviewed over 3,500 randomly selected adults from around the country to represent the population of English speaking, U.S. households. Respondents were told about the pre-spill baseline condition of the GOM, what the environmental injuries were based on the best scientific evidence available, how long they would take to recover, what the proposed governmental prevention program was to avoid a similar oil spill in the future, and how much they would have to pay in increased taxes if the program were implemented. The study was completed in 2016. The aggregate estimate of economic damage was: $17 billion (U.S.). The natural resource damage claim of the trustees was settled by BP prior to trial, despite the objections of some of the company’s technical consultants and advisors to various components of the trustees’ NRDA studies, in 2016 for $8.8 billion (U.S.). By law, the recovery must be used by the trustees to restore the injured natural resources of the GOM and for reimbursement of their NRDA costs.

# The Adopted Methodology Approach

This section overviews environmental valuation methods for estimating environmental values, presents the adopted methodology and estimation procedures.

## Methods for Estimating Environmental Values

The charge provided to the research team was to estimate the aggregate public (as opposed to the private) economic value (social cost/damages) lost by Brazilian households due to the impacts to environmental/cultural/historical resources resulting from the FDI. When a change in services provided by natural resources doesn’t result in observable change in human behavior (i.e. revealed preferences), economists rely on what are known as stated preference methods[[4]](#footnote-4) to estimate the economic value of the change. Because of the nature of the injuries to environmental/ cultural/historical resource resulting from the FDI, the team chose a stated preference method to estimate the value of the losses. Stated preference methods are the only known way for estimating the value of changes in such public goods when revealed preference data are not available. In addition, contrary to revealed preference methods, they allow for estimation of passive-use component of the economic value – the component which stems from consumers valuing and willing to pay for things they do not directly use (e.g., nature reserves).

Stated preference methods use responses to survey questionnaires to estimate economic value. Two stated preference methods have been widely used to estimate the value of environmental injuries such as those caused by the FDI:

* Choice experiments
* Contingent valuation

### Choice Experiments

Natural resource damage assessment studies using the choice experiment approach first identify attributes of the environmental impacts. For example, attributes could include the number and population sizes of the affected species and the time required for their recovery, the number of miles of rivers impacted and some measure of the contamination (perhaps contaminant concentration), as well as the time required to return to pre-impact baseline, etc. Then respondents are presented with an array of choices, with each choice described in terms of a specified level of environmental quality or attribute and a cost to obtain that level of environmental quality. Figure 1 shows an example of such a question.[[5]](#footnote-5) By carefully specifying the types of attributes and their levels in each choice made by a survey respondent, the marginal value of changes in the level of each attribute can be statistically determined.

Please consider the two plans for restoring the environment after a tailings dam failure.

How would you vote if you had to choose between these two plans? (CIRCLE ONE NUMBER)

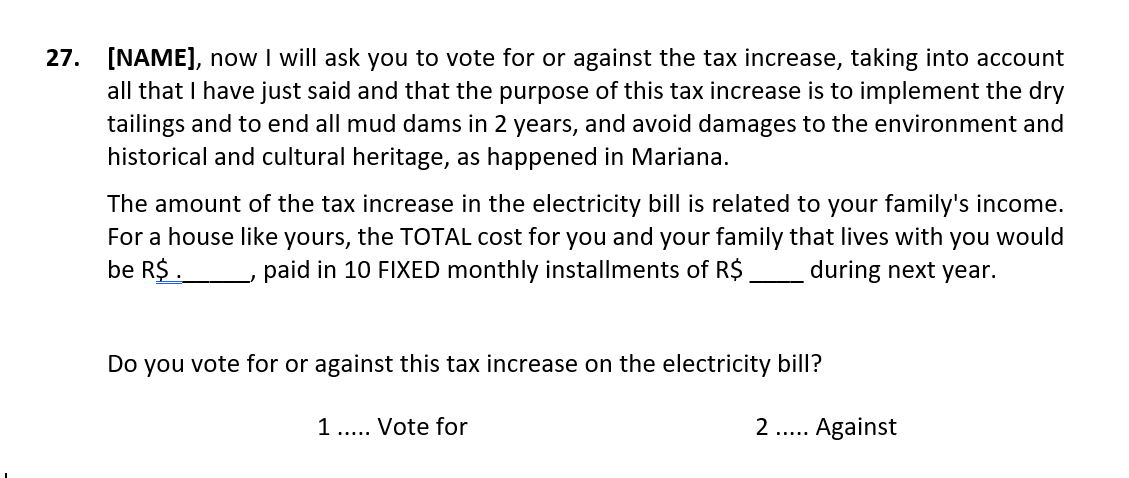
1 I would vote for Plan 1

2 I would vote for Plan 2

*Figure 1. Example of a discrete choice question*

### Contingent Valuation versus Choice Experiments

Studies using the contingent valuation (CV) method describe the environmental injury to be valued and then ask survey respondents if they would be willing to pay a specified amount to either “undo” environmental injury or to prevent a similar injury from happening in the future. Again, statistical analysis can be used to determine the average value of the environmental injury. Figure 2 shows an example of a CV question.



*Figure 2. Example of a contingent valuation question*

Ultimately the research team chose to use the CV method in this study. Several considerations played a role in this decision. First, it is much more common to use CV rather than CE for estimating welfare losses (including passive/non-use values) in NRDAs. Second, the CE method requires the provision of more detailed information to each respondent about individual attributes of the goods and services being valued. It also gathers more information by asking each respondent to make a number of choices, or tradeoffs, among the various attributes and prices. Likewise, the CE method offers the possibility of identifying values associated with marginal changes in attributes. But, incremental, marginal changes to environmental/cultural/historical services in this case did not appear highly relevant, or appropriate, since the goal of the study was to estimate the value of the environmental/cultural/historical injury. In addition, the research team was provided with a single description of the injuries to be valued.

At the same time, the CE method is also subject to several drawbacks. For example, it can be quite complicated to construct choice questions that are incentive-compatible, i.e. elicit the true preferences of each respondent. And the statistical modeling required as a result of observing repeated choices for each respondent imposes an additional level of complication – both in the actual analysis and in the exposition to policy makers as to how values are derived.

For these reasons, the research team decided that a study based on the CV method would allow the team to fulfill its basic charge without introducing the additional complexities arising with the use of CE.

## Guidelines for Conducting a CV Study

Since inferences of value in a CV study are based on stated preferences, as opposed to observed behavior, CV studies used to estimate values in high profile contexts can be controversial. The National Oceanic and Atmospheric Administration (NOAA) is the lead US agency responsible for conducting natural resource damage assessments for spills of oil and hazardous wastes into federal marine waters. In 1991 it commissioned a five-member panel of experts (“NOAA Panel”)[[6]](#footnote-6) to conduct an evaluation of the validity of the CV method for use in natural resource damage assessments.

The Panel concluded that if properly conducted, CV was capable of providing estimates of lost value suitable for use in legal proceedings. The panel also proposed a set of guidelines for conducting reliable CV studies.

These guidelines addressed the following elements associated with CV studies:

* Use of probability sampling
* Minimization of non-response
* Use of personal interviews
* Examination of interviewer effects
* Careful and complete reporting, including:
  + basic cross tabulations,
  + descriptions of procedures,
  + exact wording and other communication used during interviews
  + production of a documented data set that can be provided to interested parties
* Questionnaire pretesting
* Use of a conservative design
* Use of a willingness-to-pay format (WTP)
* Framing the contingent valuation question in terms of a referendum
* Providing an accurate description of the program or policy
* Pretesting of photographs (and graphics)
* Inclusion of reminders of undamaged substitute commodities
* Adequate time lapse from the incident
* Use of a No-Answer option
* Use of follow up to Yes/No responses
* Cross tabulations
* Checks on understanding and acceptance

Similar guidelines are also included in more up-to-date methodological recommendations for stated preference studies (e.g., Freeman, Herriges and Kling (2014); Champ, Boyle and Brown (2017); Johnston et al. (2017)).

The present study followed the recommendations listed above. It also followed closely the procedures used by Bishop et al. (2017) to estimate the monetary value of the total economic loss suffered by American households as a result of injuries to natural resources caused by the BP Deep Water Horizon oil spill in the Gulf of Mexico in 2010.

Two exceptions to the BP oil spill study were made in the procedures followed for the FDI study. First, we did not employ follow-up attempts to contact respondents who were not available at their randomly drawn households. Instead, another household with similar characteristics was chosen to be interviewed. This procedure was adopted because it was considered more feasible for surveying conditions in Brazil.[[7]](#footnote-7)

Second, the NOAA guidelines include a concern that the results of a CV study might not be deemed reliable if they demonstrated an “Inadequate responsiveness to the scope of the environmental insult.” In the context of conducting a CV study using a referendum format, carrying out a scope test requires that the researchers conduct, in essence, two separate CV studies to test the difference in the estimated mean values between two different size environmental injuries. Conducting such a scope test would require the devotion of considerable additional time and resources to the estimation of the value of a second, hypothetical injury that is either larger or smaller (in some attributes or dimensions) than the actual injury being valued in the damage assessment. Subsequent to the NOAA guidelines, numerous CV studies have been carried out and readily passed scope tests (Kling, et al., 2012). Given these results and limitations, the research team determined that it was not necessary or prudent to devote the time and resources required to conduct a second study needed for a formal scope test.

We also note that the design of our study satisfied the necessary conditions for incentive compatibility, which means that a single dominant strategy for rational respondents is to reveal their preferences truthfully ([Carson and Groves, 2007](#_heading=h.279ka65)). This was achieved by framing the choices as participating in an advisory referendum on a policy that would lead to preventing the described damages in the future at a given cost for a respondent. In particular, in this design:

1. respondents saw the survey as consequential – they viewed their responses as potentially influencing the supply of a good and the costs of this change to them ([Vossler, Doyon and Rondeau, 2012](#_heading=h.rjefff));
2. the payment was coercive – the payment vehicle was able to impose costs on all individuals if the government was to undertake the project ([Carson and Louviere, 2011](#_heading=h.184mhaj));
3. the referendum format was viewed by respondents as a take-it-or-leave-it offer, which means that they did not see their stated choices as influencing any other future choice situations ([Carson, Groves and List, 2014](#_heading=h.meukdy)).

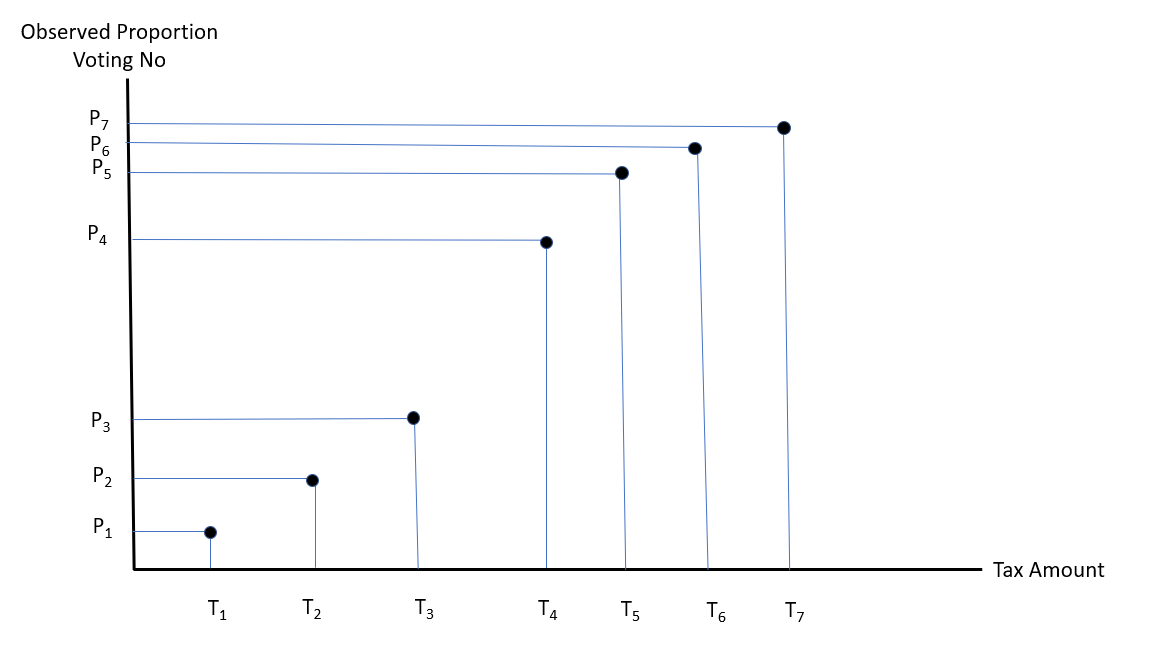
Overall, our study followed best practices” concerning SP study design and implementation and hence the estimates from the study are valid measures of respondents’ true preferences and provide useful information for adjudication of damages in the FDI.

## Statistical analysis framework

The ultimate goal of a stated-preference valuation study is to estimate the economic value of a non-market good (in our case – the value of the damages resulting from the FDI). In the case of a CV study, this is typically done by estimating mean willingness-to-pay (WTP) of individuals or households for the policy that would lead to delivering the non-market good (in our case – avoiding a future event that would result in exactly the same damages). The mean WTP can be estimated based on individual votes in a hypothetical referendum, in which respondents are given the description of the proposed policy and its effects as well as the cost for them in the case the policy is implemented. By presenting different respondents with different costs it is possible to observe average shares of respondents supporting the policy at each cost level, and apply statistical methods to estimate mean WTP.

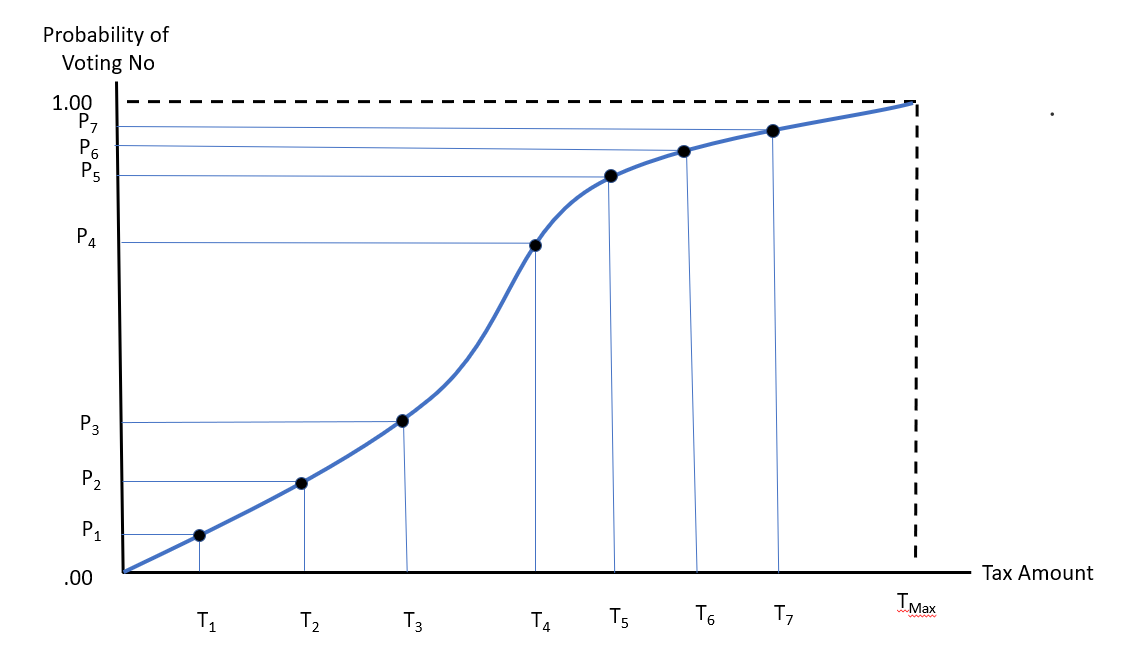
### Estimation of mean WTP based on binary choice data

To explain how mean WTP can be estimated from a referendum voting question (binary choice), suppose a set of 7 different tax amounts, T1 to T7, are arrayed in order from the lowest, T1, to the highest, T7, and that the observed proportion of respondents voting ‘no’ (or against) to each tax amount is designated P1 to P7 as shown in Figure 3.



*Figure 3.Tax amounts and proportions of No votes*

It is possible, using statistical techniques, to fit one of the well-known mathematical functions to the observed data – this is called a parametric approach to estimating mean WTP. As shown by the blue line in Figure 4, the parametric function predicts the probability of voting NO for tax amounts both above and below the range of tax amounts asked about in the voting question. In this example the parametric function predicts no one would vote for the referendum at an amount greater than TMax.

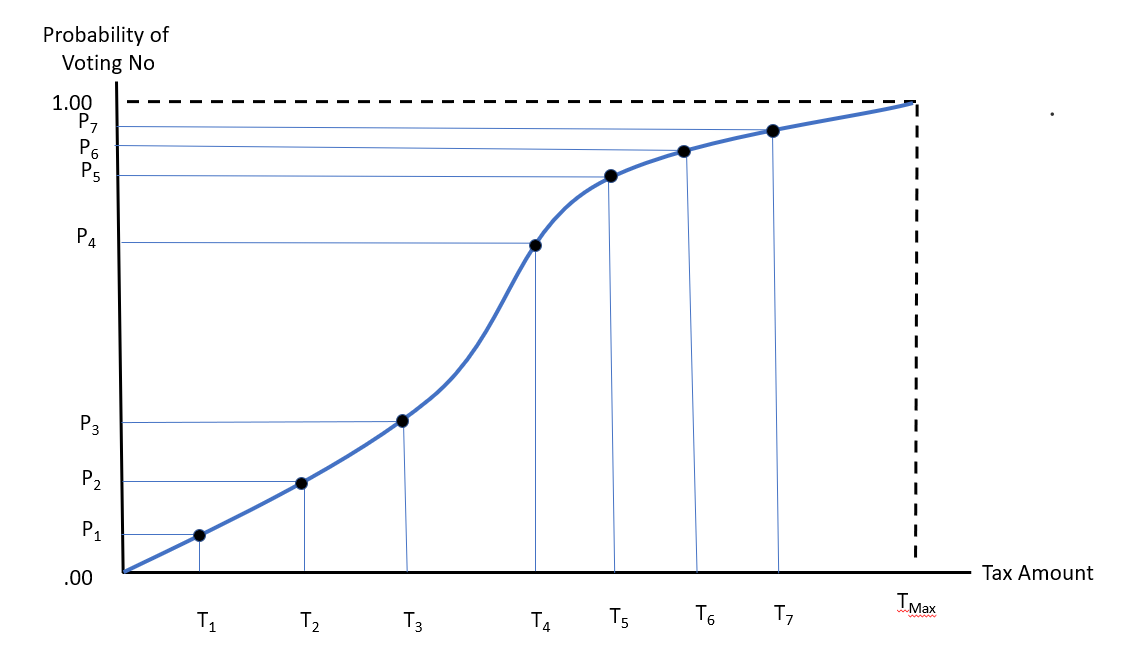
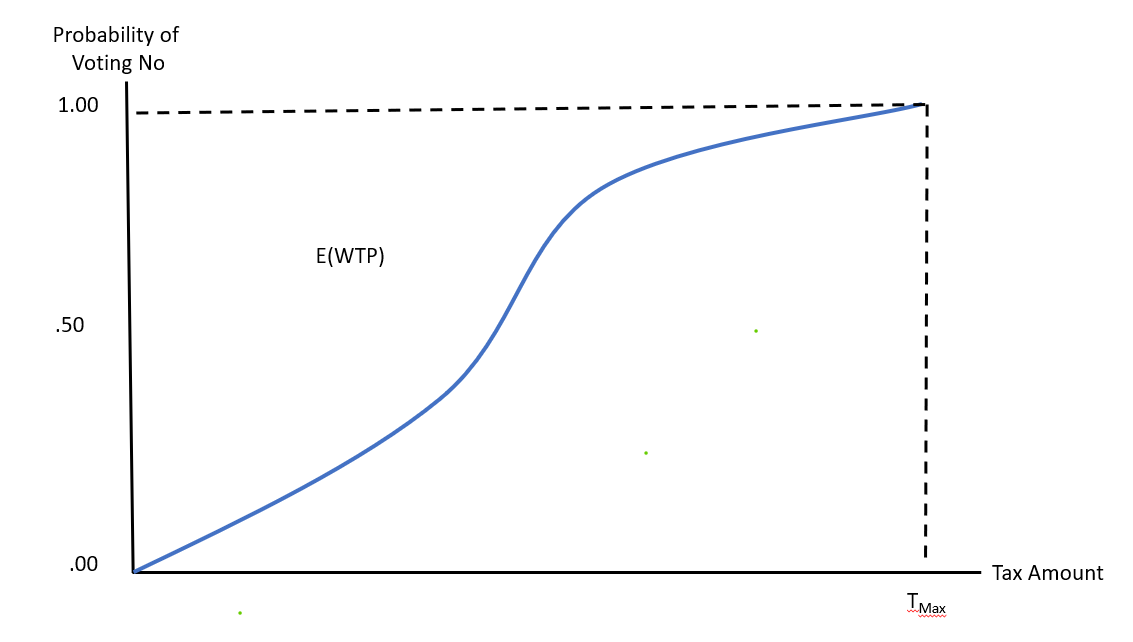


*Figure 4. Parametric function predicting the probability of No votes*

A well-known statistical result is that for a non-negative random variable X

 (1)

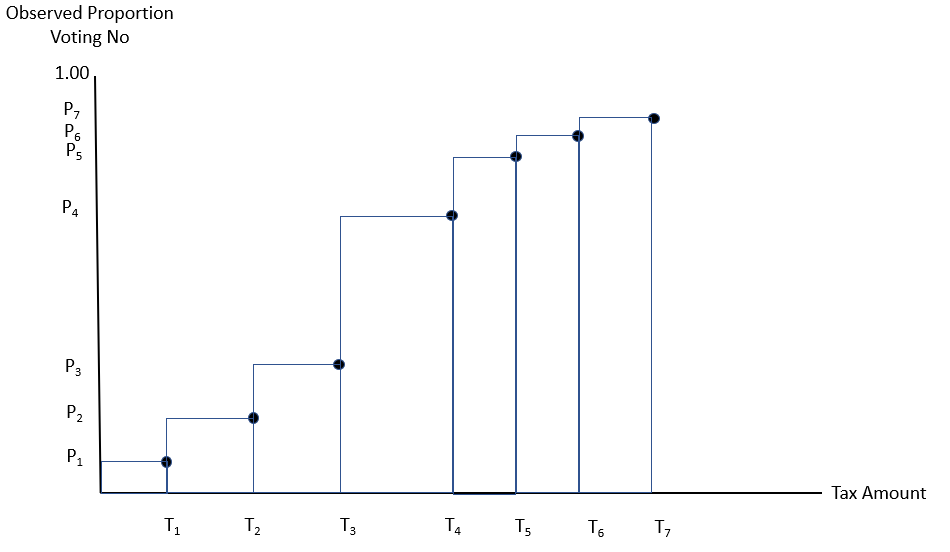
Consider a survey respondent with a WTP of *x*. In a CV study based on a referendum question, a survey respondent is asked if he/she would vote in favor of a referendum if passage of the referendum would cost him/her *t*. This survey respondent would vote ‘no’ when  and vote ‘yes’ otherwise. Once a parametric function predicting the probability of voting against the referendum is estimated, average WTP can be calculated as the area labeled as E(WTP) in the Figure 5.



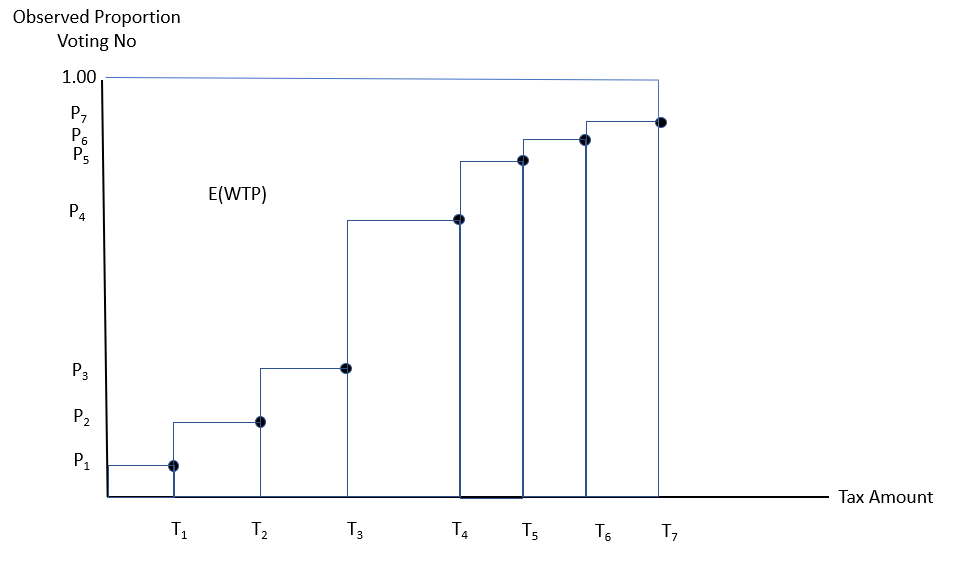
*Figure 5. Average WTP*

Note that the parametric approach uses mathematical functions that fit the data to interpolate expected probability of voting ‘no’ between observed tax amounts. A more conservative alternative is to simply assume that the probability of voting ‘no’ at levels even slightly higher than *t* immediately become equal to those observed for tax amount  (or equal to 1 for values higher that the highest observed tax amount). As this approach does not utilize any common probability functions, we will refer to this as a non-parametric approach.

In the non-parametric approach, an empirical step function is constructed (Figure 6). This step function conservatively approximates the parametric function and the mean willingness to pay can be estimated as the area above the step and below  from tax amount zero to T7 (Figure 7).

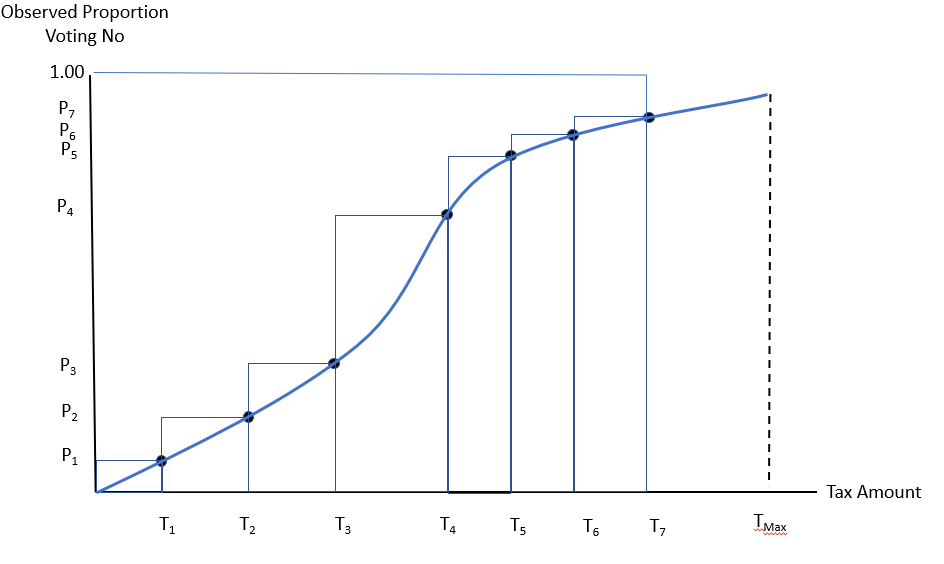


*Figure 6. Empirical step-function*



*Figure 7. Mean WTP approximation using step-function*

The nonparametric estimate of mean WTP will be lower than the parametric estimate of mean WTP as shown in Figure 8. This is because the parametric estimate of the mean WTP includes areas below the step function and above the fitted parametric curve and also includes the area below P=1 and above the fitted parametric curve from T7 to Tmax.



*Figure 8. Mean WTP – parametric x step-function*

The two approaches to estimating mean WTP have different strengths. The non-parametric approach estimates a lower bound of mean willingness to pay. It does not rely upon parametric inferences about the cumulative density function of WTP between the observed tax amounts; rather it simply assumes it stays the same between the observed tax amounts. The parametric approach produces an estimate of mean WTP that is higher than the non-parametric, if one is willing to accept the additional structure imposed by the use of a parametric model, particularly the necessary assumptions about vote probabilities between the tax amounts used in the experimental design (i.e. the specified tax amounts).

Overall, the non-parametric approach has the advantage of being conservative – that is the true mean WTP is definitely not lower than the non-parametric estimate. On the other hand, the parametric estimate has the advantage of minimizing error of the estimate of the true (but unknown) mean WTP, however, the true mean WTP can be either somewhat lower or higher than the parametric estimate. Choosing between the use of a parametric or a non-parametric estimate ultimately depends on the purpose to which the estimate will be used.

### Non-parametric estimator of mean WTP

Our non-parametric estimator of mean WTP is based on [Lewbel (2000)](#_heading=h.1d96cc0) and [Watanabe (2010)](#_heading=h.3bj1y38), which in our case[[8]](#footnote-8) is equivalent to the ABERS ([Ayer et al., 1955](#_heading=h.4du1wux)) and Turnbull ([1976](#_heading=h.2ce457m)) estimator. For a random sample of  individuals who are offered different tax amounts  let individual ’s true WTP be . Respondent’s votes on the program provided at the cost of  reveal if their WTP is higher or lower than . The probability of observing a ‘no’ vote can therefore be mathematically represented using a cumulative distribution function (CDF):

. (2)

With random assignment of the tax amount, the non-parametric estimate of  becomes ([Haab and McConnell, 2003](#_heading=h.3jtnz0s)):

, (3)

where  represents the number of ‘no’ votes to a given tax amount  and  indicated the total number of individuals in the sample who were given the tax amount ,[[9]](#footnote-9) and the variance of the estimator becomes:

. (4)

In the case of continuous design for the tax amounts (and a maximum tax amount exceeding the maximum WTP in the population), the Lewbel-Watanabe estimator of mean WTP is unbiased and consistent. Unfortunately, these assumptions are rarely met in practice. For practical reasons, valuation studies usually use discrete design of tax amounts (a few discrete tax amounts randomly assigned to respondents). It is also infeasible to use the highest tax amount that is larger than the highest possible WTP in the population, as there is no a priori information on the maximum WTP, and it could be very large requiring that the continuous taxes span a very large range. As a result, the Lewbel-Watanabe estimate of mean WTP becomes a conservative (lower-bound) estimate of WTP, and in the case of no monotonicity violations it becomes equal to the traditional ABERS ([Ayer et al., 1955](#_heading=h.4du1wux)) and Turnbull ([1976](#_heading=h.2ce457m)) estimator.

For the case of using a discrete tax amount design, the CDF can only be evaluated at these tax amounts. Suppose there aretax amounts. We can obtain an unbiased and consistent estimator of the CDF of a ‘no’ vote at these  levels. For the other levels, our estimator conservatively extrapolates the next highest observed probability of voting ‘no’. For example, if we observe that the probability of a ‘no’ vote to be , ,  for tax amounts , , , respectively, we conservatively assume that the probability of voting ‘no’ for tax amounts between  and  is  (the same as for the highest observed tax amount in this range), and the probability of voting ‘no’ for tax amounts between  and  is . In reality, the probability of voting ‘no’ is likely to continuously increase between tax amounts, rather than experience a jump to the next higher observed level as soon as the tax amount raises above the observed level. For this reason, our non-parametric estimator of WTP can be considered a lower-bound for the actual (true) WTP. It is conservative in a sense of not relying on any assumptions for extrapolating the probability of ‘no’ votes between the levels observed for tax amounts – instead it assumes the highest possible probability of voting ‘no’, conditional on satisfying monotonicity of probability levels for increasing tax amounts.

Under these assumptions, the non-parametric estimate of mean WTP can be calculated as:

, (5)

where , , , and . The corresponding variance of the estimator is given by:

. (6)

### Parametric estimator of mean WTP

In our case, respondents were asked to indicate whether they would vote yes or no for a program of implementing the new policy at a given cost to them (tax amount). A response to a binary choice question reveals if one’s true maximum WTP is lower or higher than the tax amount. This information can be used to fit a parametric distribution describing people’s WTP.

Assuming the WTP distribution is of particular form (e.g., normal) with unknown parameters describing its mean and standard deviation, the probability of observing a particular choice is equal to the cumulative distribution function (CDF) of the assumed distribution evaluated at the upper bound (i.e., the probability that WTP is lower than the upper bound) less the CDF of this distribution evaluated at the lower bound (i.e., the probability that WTP is lower than the lower bound).[[10]](#footnote-10) The parameters of the selected parametric distribution can be found by maximizing the product of these probabilities for the observed choices of all respondents.

#### Estimation

Formally, the probability that individual ’s WTP lies between the accepted tax amount  (lower bound) and the next higher tax amount  (upper bound) can be expressed as

[[11]](#footnote-11) (7)

where  denotes a cumulative distribution function of the considered WTP distribution and  is a vector of the distribution parameters (for example, for a normal distribution,  consists of a mean and a standard deviation).

The parameters of the distribution can be estimated using the maximum likelihood method. The probability specified in (1) expresses individual ’s contribution to the likelihood function, while the log-likelihood function for a sample of  individuals can be formulated as:

, (8)

where  represent weights that account for the possible over- or underrepresentation of specific individuals in the sample relative to the target population.

#### Parametric distributions to consider

The above formulas are conditional on selecting a parametric distribution, whose CDFs are calculated. However, a researcher does not usually know what parametric distribution is best for approximating the distribution of WTP in the population. Instead, it is common practice that many parametric distributions are tried to select the one that fits data best. Because the distributions can vary with respect to the number of parameters, and distributions with more parameters can lead to better fit, one can use the Akaike information criterion (AIC) or the Bayesian information criterion (BIC), rather than simply the value of the log-likelihood function as a means for comparisons, to account for the cost of additional parameters (and penalize overfitting).

There is no theory guiding the choice of a parametric distribution – what parametric distribution fits the WTP distribution in the population is an empirical question. In what follows, we considered a large number of candidate parametric distributions. The specifications that performed best, according to the Bayesian Information Criterion, were: Exponential, Gamma, Negative Binomial, Nakagami, Birnbaum-Saunders, Uniform, Lognormal, Rayleigh, Loglogistic, Normal, Logistic, Extreme Value and Weibull (with or without zero-inflation).[[12]](#footnote-12)

#### Zero-inflation

It is usually found that there is a large share of respondents whose WTP is equal to zero coupled with observing relatively few very small WTP amounts.[[13]](#footnote-13) This can be represented by a jump discontinuity in a probability density function (PDF) of any parametric distribution and is typically called a spike ([Kriström, 1997](#_heading=h.2y3w247)) or a zero-inflated model ([Greene, 2011](#_heading=h.2koq656)). In the zero-inflated model, respondents’ WTP is modelled as a mixture of a Bernoulli distribution (a point mass at zero) and a given parametric distribution, allowing for over-proportional share of zero responses ([Gurmu and Trivedi, 1996](#_heading=h.zu0gcz)). As a result, the log-likelihood function becomes:

 (9)

where  represents the probability that individual  is a non-participant (meaning that his WTP for the program is zero).[[14]](#footnote-14)

#### Simulation of mean WTP

The estimated parameters of the best-fitting parametric distribution may not readily be interpretable as mean WTP. In this case, the required characteristics of the WTP distribution (mean) and its standard error can be simulated, following parametric bootstrapping method adapted from Krinsky and Robb ([1986](#_heading=h.1yyy98l); [1991](#_heading=h.4iylrwe)):

1. To account for the uncertainty with which the estimates are known, we use parameter estimates and the inverted Hessian at convergence[[15]](#footnote-15) to define a multivariate normal distribution[[16]](#footnote-16) and use it to draw a large number (e.g., 104) of new sets of parameters.
2. For each set of parameters simulated in step 1 draw 104 empirical WTP values. This follows the assumed parametric distribution of WTP (possibly with zero inflation component).
3. Observing variation in required characteristics (mean) of the WTP distributions simulated in step 2, driven by each set of parameters generated in step 1, allows estimating the uncertainty associated with the required WTP distribution characteristics.

### Models for investigating sensitivity of voting behavior to explanatory variables

In the analysis of the sensitivity of voting behavior to various explanatory variables (such as tax amounts, income levels of respondents, responses to debriefing questions, etc.) that are presented to demonstrate validity of our study we utilize simple linear regression models. We do so, even though the response variable is binary and hence models tailored for dealing with binary dependent variables could be used (e.g., binary probit). The reason for the selected approach is its simplicity and the ease of interpretability.

Following [Angrist and Pischke (2010)](#_heading=h.1tuee74) and [Wooldridge (2010)](#_heading=h.1qoc8b1), we note that binary response models (including linear probability models) do not have to be correctly specified in order to provide useful information and the linear probability model usually provides reliable estimates of directions and magnitudes of the effects of independent variables, on average, which is our goal in these analyses. The linear probability model estimated using ordinary least squares regression produces consistent and unbiased estimator, however, unless all coefficients are zero heteroskedascity is present. To account for the this issue we use standard heteroskedascity-robust standard errors ([MacKinnon and White, 1985](#_ENREF_2))

We note that for robustness check, we also used binary probit models. The results, presented in Appendix 4, show that the alternative modeling approach resulted in qualitatively equivalent findings.

# General Survey Design Criteria

This section discusses the criteria the study team followed in designing the survey instrument. But before doing so, it is worth noting here that a conservative approach to decision making was followed by the study team throughout this project. That is, whenever it was faced with deciding on a particular course of action between two or more equally compelling (from a technical, scientific and/or practicable perspective) alternatives, where possible the study team tried to choose the one that would more likely result in either lowering, or having a neutral impact, on the final estimate of total damages, when compared to the other alternatives under consideration. In line with this conservative approach the wording “incident” instead of “accident” or “disaster” was used throughout the survey instruments to avoid any perception that would anticipate the injury levels before their detailed presentation.

The research team undertook steps consistent with the NOAA guidelines in the design of the CV study. First, members of the team worked closely with physical and biological scientists at Lactec to gain a clear understanding of the environmental/cultural/historical services injuries resulting from the FDI. Members of the research team then wrote descriptions of these injuries in terms members of the general public could understand. The survey included textual and graphic information describing the injury to environmental/heritage resources. This information was tested in focus groups, in-depth personal interviews, and pilot surveys.

Graphic information was comprised of maps and diagrams presenting the spatial features of the incident. Infographics, with recovery actions and information from official sites, images and photos of the impacted species and heritage assets were also included. Images and photos illustrated the object or species without injuries to avoid advancing damage perception to respondents.

The research team also developed a description of a program that could prevent one future tailings dam failure that would have environmental injuries just like those resulting from the FDI. The program description was written in non-technical terminology that could be easily understood by the general public. The description was developed by research team members in conjunction with technical experts at Lactec.

Drafts of the description of the environmental injury and the program to prevent another such incident were then subjected to extensive pretesting using a series of focus groups and personal/one-on-one interviews.

As can be seen in Table 1**,** focus groups were held in a number of locations over a period of several months.

Table 1: List of Focus Groups (FG) conducted as part of qualitative pretesting of the study

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1st wave of  survey development** | | | **2nd wave  of survey development** | | | **3rd wave  of survey development** | | |
| **City** | **Date** | **Number of FG** | **City** | **Date** | **Number of FG** | **City** | **Date** | **Number of FG** |
| São Paulo | 22/01/19 | 2 | São Paulo | 04/04/19 | 2 | Recife | 23/05/19 | 2 |
| Belém | 24/01/19 | 2 | Goiânia | 09/04/19 | 2 | Rio de Janeiro | 28/05/19 | 2 |
| Goiânia | 29/01/19 | 2 | Belo Horizonte | 11/04/19 | 2 | Curitiba | 30/05/19 | 2 |
| Linhares | 31/01/19 | 2 | Belém | 16/04/19 | 2 | São Paulo | 06/06/19 | 1 |
| Rio de Janeiro | 12/02/19 | 2 | Curitiba | 23/04/19 | 2 |  |  |  |
| Recife | 14/02/19 | 2 |  | | | | | |
| Governador Valadares | 19/02/19 | 2 |
| Belo Horizonte | 21/02/19 | 2 |
| Curitiba | 26/02/19 | 2 |

Notes: All focus groups had up to 8 participants. Dates are in dd/mm/yy format.

Personal/one-on-one interviews (PI) were also held in many locations over several months, as can be seen in Table 2.

Table 2: List of Personal Interviews conducted as part of qualitative pretesting of the study

|  |  |  |  |
| --- | --- | --- | --- |
| **City** | **Date** | **Number of PI** | **Location** |
| São Paulo (SP) | 03/06/19 | 4 | Household |
| 06/06/19 | 3 | One-way mirror room |
| 10/06/19 | 4 | Household |
| 14/06/19 | 4 | Office |
| 17/06/19 | 4 | Household |
| 18/06/19 | 4 | One-way mirror room |
| 24/06/19 | 4 | Household |
| 27/06/19 | 4 | Office |
| Recife | 01/07/19 | 4 | household |
| Belo Horizonte | 04/07/19 | 3 | One-way mirror room |
| Total | --- | 38 | --- |

Note: Dates are in dd/mm/yy format.

At each stage of the pretesting, research team members evaluated the materials in terms of comprehension and plausibility by the participants and adjusted survey materials as needed.

In addition, three pilot tests were conducted. These pilot tests allowed evaluation of survey implementation procedures, assessment of the tax amounts to be used in the referendum question and to get a preliminary assessment of results for small samples in various locations around Brazil. Table 3 summarizes the locations, dates and sample sizes of the pilot studies.

Table 3: Pilot Studies

|  |  |  |  |
| --- | --- | --- | --- |
| **Pilot Survey** | **City** | **Interviews** | **Date** |
| 1 | São Paulo (SP) | 402 | 25/07/2019 to 16/08/2019 |
| 2 | São Paulo (SP) | 403 | 03/09/2019 to 20/09/2019 |
| 3 | Belo Horizonte (MG) | 402 | 15/10/2019 to 05/11/ 2019 |
| Recife (PE) | 405 | 17/10/2019 to 07/11/2019 |
| **Total** | - | 1,612 |  |

## Rationale for City Selection

Development of the questionnaire to meet the goals described above involved an iterative process carried out over a period of eleven months. This iterative process included preparing drafts of questionnaires, testing of draft materials in focus groups and personal interviews and testing of questionnaires in formal pilot tests.

Focus groups and personal interviews were scheduled in nine different locations across Brazil. A number of considerations were used to determine locations of these focus groups and interviews. The initial focus groups and interviews were concentrated in Sao Paulo and Rio de Janeiro for ease of logistics. Locations for later focus groups and interviews were selected in order to provide researchers with feedbacks from a variety of metropolitan areas in all regions including Recife (Northeast region), Belo Horizonte (Southeast region), Curitiba (South region), Goiania (Center-West region) and Belém (North region), as well as cities in the incident area (Governador Valadares and Linhares), as shown in Table 1, Table 2 and Table 3.

## Survey Instrument

In any stated preference study, a survey instrument presents respondents with an object of choice, describes a context in which a choice is to be made, and then asks the respondent to make a choice.

For this study, the object of choice is a program that will prevent one future tailings dam failure that would have the exact same impacts to the environment and cultural/historic sites that occurred as a result of the FDI[[17]](#footnote-17). However, by the end of the first wave of focus groups, it was clear that respondents did not think there was a high likelihood of future tailings dam failures in Brazil. Nonetheless, that thinking changed following the widely reported 2019 Brumadinho dam disaster and after being presented with information about the number of tailings dams in existence, or under construction. Thereafter, respondents readily believed another incident like the FDI could occur in Brazil.

A complete description of this object of choice requires that the survey instrument describes:

* the effects of the FDI
* why another dam failure like FDI will happen
* how a program to improve dam safety would, if carried out, prevent that future dam failure and the associated effects

Each of these items must be described in language that is understandable to the respondent. Furthermore, the language used should be believable to the respondent and minimize any possible ambiguities about the object of choice.

For this project, these objectives required that the respondents understand that the specific program about which they are asked to make a choice will, if carried out, prevent one future dam failure and will avoid the same exact effects that were observed as a result of the FDI.

In addition, respondents must not be left with the impression that the program, if carried out, will provide additional impacts (benefits) beyond the ones that were described to them. The research team was tasked with valuing the **environmental and heritage resource impacts**. This posed some special challenges. Dam failures like the FDI affected numerous other attributes besides the environment and cultural resources. For example, the dam incident affected local water supplies and commercial fishing, affected local agriculture, destroyed residences and killed people. The object of choice was designed to stress that these effects had been compensated for in the case of the FDI and would not be a feature of the future dam failure.[[18]](#footnote-18)

It was also important to design the questionnaire so that respondents did not feel pushed to make a choice, either for or against, the program. This ensures that the choice expressed by respondents during the survey reflects their personal assessment of the cost of the program relative to the benefits that would be provided if the program was carried out.

The final version of the questionnaire[[19]](#footnote-19) starts with a section collecting data from the household. The next section describes aspects of the FDI in terms of damages and injuries that have already been addressed through judicial actions and/or restoration actions. It then describes the specific environmental and heritage injuries that have not yet been addressed. These injuries are described using both text and graphics. The survey then describes current safety measures that are required of all mining companies. Next the questionnaire states that another incident similar to FDI will occur in the next ten years. Furthermore, the questionnaire states that this new incident will have impacts to environment and heritage resources equivalent to those experienced as a result of the FDI. The questionnaire then describes a technology, dry tailings, that would eliminate all mining dams within two years and, prevent an incident like FDI from occurring in the future.

The questionnaire states that rapid deployment of the dry tailings technology would require sharing costs between mining companies and the government and that the public financing would require an additional tax on household electricity bills. Then respondents were reminded of the environmental and heritage protection benefits that would result if the dry tailings program was carried out. Next, respondents were reminded of reasons they might vote against implementing the dry tailings project. The questionnaire then asked each respondent if he/she would vote for or against the dry tailings project if its implementation would cost them a specified amount in terms of an increased tax on their electric bills. Those voting against the project were asked to explain their vote by choosing from among a list of reasons, which included economic and protest motivations. The rest of the questionnaire included de-briefing questions to verify preferences, attitudinal characteristics and comprehension.

The damage scenario, as later described in more detail, was built with conservative parameters provided by Lactec experts. While the narrative describing the dry tailings program is hypothetical, it was designed in such a way as to create incentive compatibility and consequentiality in the minds of the respondents.

The next sections of this report present the testing process used to arrive at the final version of the questionnaire.

## Main Challenges

A number of challenges were addressed during the design of the survey. These challenges arose at various times during the design process and are discussed here by topic, rather than chronologically.

As discussed above, the research team was tasked with quantifying the economic value of impacts to the environment and cultural historical sites. A first set of challenges dealt with what respondents might have already known or felt about the FDI. The challenge in the design of the survey instrument was to make sure that this prior knowledge or feelings did **not affect** the object of choice in undesirable ways. This required designing an object of choice that did not embed values for any of the following items:

* Impacts to lifestyles, municipalities and public infrastructure
* Corporate assets, criminal/negligent behavior and accountability
* Damages to recreation
* Impacts to commercial fishing and agriculture
* Impacts to domestic water supplies
* Loss of life caused by the FDI

Compensation for damages related to the items in the list above have been, or will be, addressed through monetary compensation provided by Samarco, making them irrelevant to the charge provided to the research team. Thus, the survey instrument was designed in such a way that the object of choice did not include these items. We discuss each of these next.

### Impacts to lifestyles, municipalities and infrastructure

The FDI affected many aspects of the lifestyles of those residing in the area of impact, and included the need to relocate those people whose homes were destroyed by the resulting flood. In early focus groups and personal interviews, it was quite apparent that citizens of Brazil were aware of these impacts and felt quite strongly about them.

The final instrument included a discussion of the FDI, described in the survey instrument as the Mariana Incident.[[20]](#footnote-20) This discussion noted:

“This card shows the site of the rupture of the dam, which left about 2 thousand people homeless, destroyed houses, farming properties and the infrastructure of the villages around the dam.

It also affected trade, agriculture, livestock, fisheries, tourism, the environment and historical and cultural heritage. (See Item 1 in survey instrument)

Two lawsuits are in court, one CRIMINAL, against the companies responsible and their managers, and another to pay INDEMNIFICATIONS for all damages caused to people, municipalities and the environment.

According to a court order, the companies responsible were required to pay 12 BILLION reais to the Federal Public Prosecutor's Office to begin paying compensation and recovering the environment.

Until today, almost 1 billion and 400 million reais have already been spent to pay the damages of 25 thousand families. There are still many families and other losses that will be compensated. (See Item 2 in survey instrument)

…. From now on we will no longer talk about the economic damages, the deaths of people and indemnities that have already been paid or will be paid for these losses. These issues are being studied in other research projects. (See material presented to respondents just prior to Item 3 in survey instrument)”

The intent of these statements was to inform respondents that in situations like the Mariana incident, the law requires responsible parties to compensate individuals for impacts to their lifestyle and to municipalities for infrastructure losses and that this had indeed been done in the case of the Mariana Incident. The implication for respondents was that this compensation for lifestyle impacts and harm to municipalities was well established and would take place if future incidents were to occur.

### Corporate Assets, Criminal/Negligent Behavior and Accountability

Participants in the early focus groups were very aware that the Mariana incident occurred because of a dam failure at a facility run by a Samarco, a mining company owned in part by Vale S.A. By stressing that the law requires the responsible parties to pay compensation, the statements above were designed to avoid embedding of these motivations in the object of choice.

### Recreation

The Mariana incident affected recreation in two ways. Respondents were told that the Mariana incident affected tourism in the Doce River region. Economic losses were potentially incurred by businesses catering to tourism but respondents were told that Samarco would provide for these losses. As a result, these losses were not part of the scope for this research team.

In addition, there would be public losses incurred by recreators as a result of not taking trips to the affected areas. In theory these losses could be part of the charge to the research team. However, measurement of these losses would have required an extensive survey of recreational uses – an effort deemed to be beyond the time and resources available for this assessment[[21]](#footnote-21).

A significant effort was made to focus respondents’ attention on the Doce River and nearby riparian areas above the Risoleta hydro-electric dam. Information about spatial impacts did not mention coastal areas near the mouth of Doce River. Given that respondents had concerns (based on information gathered in focus groups and PIs) about beach recreation and possible impacts to marine life located there, minimizing the spatial extent of injury likely contributed to a conservative estimate of damages.

### Impacts to Commercial Fishing and Agriculture

Respondents were told the Mariana incident affected commercial fishing and agriculture and that Samarco would provide compensation for these losses. Respondents were also told that this type of compensation was required by law and would be provided in the event of a future incident.

### Impacts to Local Water Supplies

The Mariana incident affected the water supplies of cities along the Doce River. Respondents were told that initially these cities turned to special treatment and incurred additional costs. Respondents were also told that in the long run the affected water supplies had been found to be safe for human consumption and were again being used.

“After the incident, water companies took a special treatment of the water withdrawn from the river to remove heavy metals and other contaminants.

By analyzing this water provided by the supply companies, scientists have stated that treated water can be consumed WITHOUT affecting people's health.

Therefore, since 2016 the population normally consumes the treated water.”

By doing this, the object of choice did not embed respondent motivations for negative impacts to water supply.

### Loss of Life

A second dam failure occurred near Brumadihno, Minais Gerais on January 25, 2019. This dam failure, which resulted in the death of an estimated 270 people was a prominent news story for a period of time across Brazil. This second dam failure with its significant loss of life occurred shortly after initial drafts of the survey had been prepared. The survey instrument was modified to acknowledge the Brumadinho dam failure. In addition, respondents were told that the since this Brumadinho incident had happened so recently, scientists had not yet gained a complete understanding of its impacts, therefore, the survey would only address the Mariana incident. The questionnaire informs:

“Two major incidents involving mining dams occurred in Brazil, the first in Mariana and the most recent in Brumadinho. 19 people died in Mariana, while in Brumadinho about 270 people died.

Let's not talk about Brumadinho because we still do not know all the consequences of this incident.

Let's talk ONLY about the Mariana incident that we already know all the consequences.”

Respondents were, in turn, told that because of the Brumadinho dam failure, requirements had been implemented that would effectively eliminate loss of life in any future dam failures, as follows:

“In the incident of Brumadinho, the dam break alert failed, and the number of deaths was very high.

After the Brumadinho incident, all mining companies were forced to IMPROVE security measures to protect people near the dams.

The security measures used TODAY are:

1. Remove all employees from administrative areas, such as offices and restaurants, within the mining area;
2. Monitor 24 hours the degree of stability of the dams, with mechanical and electronic sensors, images of drones and satellites;
3. Remove all persons from the mud impact areas before reaching the maximum risk of breaking;
4. Automatically send alert messages by cellphone, radio, loudspeakers and sirens to the entire impact area;
5. Maintain transport system and escape routes properly signaled and unimpeded to remove the people; and
6. Train nearby populations of dams to properly follow emergency actions.

All 6 measures are already working TODAY and scientists say these measures have GREATLY increased safety, which will prevent many people from dying in the event of a new dam rupture.”

Most of these measures are already in place in some mining areas but they were presented as actual measures in such a way that respondents would not embed motivations related to loss of life in the object of choice.

In the next section we describe features of the object of choice that were designed to make sure respondents did embed certain other factors in their decision. These factors relate to:

* Confidence that the dry tailings technology could prevent a future dam failure
* Confidence that the program would be carried out
* Believability of mechanisms to pay for the program if carried out.

### Dry Tailings Technology

Survey respondents were told that one more incident like the Mariana incident would occur in the next 10 years. Participants in focus groups and personal interviews readily accepted this fact. A number of technologies were considered for a program that would prevent this future incident. Ultimately, the research team determined that a dry tailings program was both understandable and credible to most survey respondents. This program was described in the following terms:

“This new technology is known as dry tailing, as you can see on that card.

In a very simple explanation of how to draw the water from the tailings, the mud is centrifuged as in a washing machine and then pressed, leaving only the dry tailings.

These photos show the equipment to dry the tailings and how the tailings get dry.

The withdrawn water is REUSED in the mining process.

All dry tailings are USED to produce cement.

To prevent future dam incidents, miners have to dry up and do away with all mud dams that already exist.

Without mud dams, there will be no risk of damage to the environment and historical and cultural heritage, as happened in the Mariana incident. This technology, although recent, is already in use in almost 10% of the world's mines.

This worldwide experience shows that dry tailings can be implanted in 2 years in Brazil, while all mud dams go through the drying process.

This eliminates the risk of breaking another dam.

The cost is very high because the drying and deactivation of the dams and the implantation of the dry tailings facilities must be done AT THE SAME TIME in all existing dams.

Because it is very expensive, all countries that use dry tail, rich or less developed, have been only able to install this technology QUICKLY with the financial help of their governments.”

### Confidence That the Program Would be Carried Out

Participants in focus groups frequently expressed a belief that the Brazilian government and mining companies might engage in corrupt activities that could prevent a dry tailings plan from being carried out. To counteract these beliefs, the process by which the program would be implemented, while taking steps to minimize the potential for corrupt practices, was described in careful detail.

Some participants expressed a desire that mining companies should pay for the program. To address this concern, respondents were told that mining companies could carry out the program but that it would take them more than 10 years:

“…according to the experts, if left to the 90 miners, the implantation of dry tailings in the 280 dams will take more than 10 years.

Because of these obstacles, without a quick resolution within the law, another dam incident, such as Mariana's, will happen in the next ten years.”

Respondents were told that to carry out the program in time to prevent the dam incident that would occur in the next 10 years, the government would pay for the capital costs of the program and that mining companies would pay operational expenses:

“As this card shows, to implement dry tailings in 2 years, the proposal is to divide spending between mining and government:

Mining companies PAY all costs to close and dry mud of all 280 dams and PAY all expenses to run dry tailing technology; and

Government PAYS investment in design, machinery, equipment and installation of dry tailing technology.

With this division of spending, the government will make LEGALLY mandatory that ALL mining companies to end the 280 mud dams forever within 2 YEARS, otherwise mines that do not deploy dry tailings will be closed.

This solution makes it possible for miners to deploy dry tailing quickly in 2 years and thereby eliminate the possibility of another incident like Mariana's”

Framing the dry tailings project in this way accomplished two goals. First, testing of this description revealed that it addressed the desire of survey respondents that mining companies pay for the program. Second, it established the rationale that the government would pay for part of the program, thus motivating the need for taxes to be raised for this project.

Early focus groups and interviews revealed that many people were concerned that the project would be subject to mismanagement and project funds mis-used. Respondents were told there would be a high degree of oversight and transparency if the dry tailings project was carried out. Specifically, they were told:

“To make sure that money of the Dry Tailings Fund will be efficiently and honestly used, there will be several ways of controls, obligations and guarantees, as follows:

1. BY LAW, the government WILL NOT be able to use this money for any other purpose;

2. The use of money will follow a plan with goals to perform all actions efficiently;

3. The release of money to the mining companies will follow deadlines and targets that, if not met by the mining companies, will force the mine to close immediately and the mining executives be prosecuted;

4. The oversight will be carried out by a Task Force led by the Federal Public Prosecution Service, with the participation of the Federal Court of Auditors, the Federal Police and university specialists;

5. Expenses will be posted every month on the Internet and may be overseen by the population, including mining specialists, journalists and organizations that monitor government spending; and

6. Task Force members shall be held criminally liable for failure to monitor or for any other irregularity”

Finally, the research team had to design a payment mechanism that was believable to respondents and that was binding upon them. Participants in focus groups and interviews easily understood that the national budget was fully allocated and that the national government was perceived to be constrained in its ability to quickly enact new laws that might be used to pay for a dry tailings program. To address these issues, respondents were told that it was possible, without enacting new laws, to temporarily increase an existing tax. In this case the respondents were told that the dry tailings program would be funded by a temporary increase in the tax on their households’ electricity bill. Specifically, they were told:

“But the government does not have the money now to pay for dry tailing technology and make drying of the 280 dams mandatory IN 2 YEARS.

In the current economic crisis, all government money is already set aside for use in health, education, security, and various other public sectors. [Pause 2 seconds]

To immediately implement dry tailings, without the need of the Parliamentary approval, the only solution will be to increase a tax that already exists for a period of ONLY 10 MONTHS, starting next year.

…

The government is considering whether to adopt this proposal that guarantees the IMMEDIATE and MANDATORY implementation of dry tailings in 2 years, with the tax increase on the electricity bill for a period of ten months.

But the government will only adopt this proposal if most people vote in favor of the tax increase.”

Most households in Brazil understand the tax they pay on their electricity bill and accepted the fact that the government could quickly make adjustments to this tax.

Based on extensive testing in focus groups, personal interviews and pilot tests, the research team believes we have designed an object of choice that:

* minimizes, to the extent possible, the consideration of unwanted motivations when survey respondents vote on the object of choice and
* maximizes the consideration of the effect of the impacts to relevant environmental cultural resources when survey respondents vote on the object of choice.

## Damage Scenario

The injury scenario presented to respondents specifies the environmental and heritage/cultural resource injuries that they will value by indicating whether or not they are willing to pay to avoid the same resource injuries from occurring again in the future. As mentioned previously, the injury scenario must be specific and only reflect the same environmental and heritage/cultural injuries that occurred in the FDI. All of the scientific information contained in the injury scenario was provided to the study team by scientists from Lactec.

There were a number of challenges to overcome in developing the injury scenario, as already discussed in the previous section. First, the Brumadinho incident led us to make changes in the questionnaire in order to avoid the possibility of respondents combining damages which occurred in the FDI with those which occurred in the Brumadinho incident. Second, an understanding of the ecosystem and habitat functions needed to be stressed before the injury scenario was presented to respondents so that they could properly understand the context from a scientific perspective. Therefore, the injury scenario began by presenting the baseline ecological functions and interactions among forest, watershed and fauna and how the mud flow from the incident affected them. This helped make the transition to detailed descriptions of the environmental and heritage/cultural injuries work smoothly with minimal confusion or misunderstanding. Third, it was necessary to ensure that the environmental and heritage/cultural injury scenario did not get combined by respondents with the socio-economic impacts related to economic productivity foregone, as well as any altruism associated with loss of life.

Due to the scientific uncertainty concerning recovery from the injuries and synergy among recovery functions, it was decided to adopt the most conservative (i.e. shortest) recovery times that were scientifically verifiable.

In order to minimize respondent uncertainty concerning the severity of the injuries and estimated recovery times, emphasis was placed on the role of the scientific community in assessing the injuries and in the design of the restoration/recovery plan. The questionnaire notes that restoration actions have been continuously monitored and that based on the resulting data, scientists have been able to calculate the time required for the injured components of the environment to return to baseline (i.e. pre-incident). Additionally, official data on current expenditures for restoration of the injured environment were presented to respondents, along with information on the total number of scientists, technicians and workers conducting the environmental restoration of the Doce River Basin.

Each injury was described in the questionnaire using available quantitative indicators for losses and expected recovery times. It was also made very clear that no species had become extinct or disappeared due to the FDI. For example, respondents were told that birds could fly to other nearby areas with unaffected habitats and would return later after the injured riparian forest along the Doce River recovered. Similarly, it was explained that fish species did not go extinct. Estimated recovery times for each injury were usually specified in ranges spanning several years, as provided by Lactec scientists.

For aquatic life, apart from showing the number collected dead fish, the narrative pointed out that the increase of heavy metals in the water from the mine tailings had caused tumors, difficulty breathing and reproduction problems in some of the fish. As a consequence, the recovery pathway for the fish will be adversely affected.

The extent of injuries to the riparian forest was provided in the narrative with an emphasis on its connection to animals living along the river and the possibility of reforestation/recovery from the injuries caused by the FDI.

In the case of wildlife, most individual animals died from suffocation under the mud and it was not possible to find and count them. Therefore, emphasis was placed on a description of the habitat recovery times for species living in the Doce River Basin. Longer recovery times were given for animals that are highly dependent on water quality for food and habitat.

Injuries to birds was described by presenting the number of species which fled to nearby regions after the dam rupture and the time that would be required for them to return and recover to pre-FDI baseline.

The injury descriptions included images of the affected resources presented on cards to the respondents. The images were fully tested for comprehension in focus groups and in-depth interviews before inclusion in the final questionnaire.

For heritage/cultural losses, it was possible to present information on some of the quantities of assets totally or partially destroyed, such as archaeological sites, cultural sites, such as mansions, and for religious property such as chapels, churches, altars and religious images. For adversely affected cultural events, such as holiday celebrations, processions, festivities and traditional cultural events that took place in various communities located in the Doce River Basin, it was possible to present estimates of some of the quantities affected.

Table 4 and Table 5 summarize the pre-FDI baseline conditions and injuries to environmental and heritage/cultural resources, respectively, that are described in the questionnaire.

Table 4: Classification of environmental damages of the FDI

|  |  |  |  |
| --- | --- | --- | --- |
|  | Before the incident | The injury | Recovery time |
| Aquatic Life | Habitat of 100 species | The habitats of 90 species suffered injuries | 60 to 70 years |
| Riparian Forest | Present at 660 km on the banks of the river | The forest was impaired, on average, 30 m on either side of the river along the first 110 km | 25 to 30 years |
| Wildlife | Habitat of 440 species | Alligator, otter, turtles | 60 to 70 years |
| Other wildlife species | 25 to 30 years |
| Birds | 125 species were observed | 25 species were no longer seen in the region | 25 to 30 years |

Table 5: Classification of heritage damages of the FDI

|  |  |  |  |
| --- | --- | --- | --- |
| Category | Total existing in the region | Totally destroyed | Partially destroyed |
| Archaeological sites  Indigenous cemeteries, ruins of gold mines and very old paintings | 450 | 15 | 55 |
| Cultural Goods  Old mansions, chapels, churches and religious images | 4,300 | 1,500 | 1,100 |
| Cultural manifestations  Celebrations, processions, festivities and traditional events | 1,200 | 5 | 55 |

Finally, respondents were informed that the greatest damage in the Doce River Basin occurred within the first 110 km downstream of the (collapsed) Fundao Dam where around 50% of the tailings and mud severely impacted the river banks and riverside forest and caused a large number of animal deaths. Some 25% of the tailings and mud was stopped behind the Risoleta Neves hydro-electric dam and, consequently, reducing damage intensity downstream with no further adverse impacts to the riverside forest. The remaining of the tailings and mud went all the way down river reaching the sea, some 660 km away. To help convey this information, a map was provided showing that the volume of mud decreased along the river, along with the injuries, the further it proceeded downstream of the Fundao Dam.

# Pilot Studies Analysis

This section reports the results of the pilot surveys that contribute to the definition of tax amounts and the validation of the questionnaire design changes. Section 4 shows where and when all pilot studies occurred.

## Tax Amounts Choice and Acceptance Rates Analysis

One of the main goals for the pilot studies was to gather preliminary information about the WTP distribution. This was achieved by an iterative design of tax amounts and the analysis of resulting information about respondents’ WTP. The first pilot used a relatively large number of arbitrarily selected tax amounts of 5 to 200 reais. The corresponding acceptance (votes ‘for’ the program) rates of the hypothetical program described in the survey to be delivered at these cost levels varied from 45 to 14%.

The information collected in Pilot 1 allowed us to calculate initial estimates of mean WTP. Moreover, they also provided the first insight into the overall distribution of respondents’ WTP and allowed for suggestions on revising the tax amounts used in the survey for a more efficient (precise) estimation of mean WTP. In particular, the main changes in the tax amounts used for Pilot 2 were:

* reducing the lowest tax amount to 3 reais to observe preferences at very low costs, in particular by individuals with low incomes,
* increasing the highest tax amount to 315 reais, to make it closer to the (‘choke point’) value at which almost all respondents would vote ‘against’ the program and hence be able to provide a better inference about the higher-end of the WTP distribution,
* using non-round tax amounts in an effort to make them seem more precisely calculated, and hence the entire scenario more realistic, and the survey more consequential,
* reducing the number of tax amounts to 7, to reduce the risk of observing non-monotonicity in the acceptance rates for increasing tax amounts, and
* increasing the density (reducing the spread) of tax amounts in the lower end of the scale, where the slope of the expected cumulative distribution function of WTP was steeper, and reducing the density (increasing the spread) of tax amounts in the higher end of the scale, where it was flatter.

In Pilot 2 the observed acceptance rates varied from 55 to 4% and the results allowed for deriving more precise non-parametric and parametric estimates of mean WTP. Perhaps more importantly, however, the combined data collected in Pilot 1 and Pilot 2 allowed for more precise insights into the shape and location of the WTP distribution. This information was further used in a simulation study designed to determine the most a priori efficient tax amounts to be used in Pilot 3 and the main study.

The simulation study proceeded as follows:

1. the data regarding observed acceptance rates for different tax amounts was used to fit various parametric probability distributions and choose the ones that fit the WTP-distribution best,
2. the two best-fitting parametric distributions (gamma and exponential with zero inflation) were used to simulate a large number (10.000) of respondents with WTP randomly drawn from these distributions,
3. for each of the simulated WTP distributions, a large number (107) of randomly drawn tax vectors composed of 3 to 10 integer tax amounts between 3 and 460 was evaluated for its efficiency, understood as the difference between the lower-bound non-parametric mean WTP estimated using the selected tax amounts and the mean WTP implied by the data generating process (the assumed WTP distribution),
4. in each case, the tax vector that minimized the bias (the difference between the assumed and the observed mean WTP) was recorded,
5. additionally, the simulation was run while imposing a constraint that all tax amounts needed to be placed sufficiently away from each other for the probability of observing monotonicity violations was to be lower than 5%.[[22]](#footnote-22)

The simulation study allowed to select optimal (bias-minimizing) tax amounts in each case, quantify the expected extent of the bias resulting from choosing a different number of tax amounts, and the bias resulting from imposing constraints on the spacing of tax amounts that help alleviate the risk of monotonicity violations.

Additionally, the bias-minimizing tax amounts were compared with tax amounts that would maximize the differences between expected acceptance rates, and hence minimize the probability of observing monotonicity violations.[[23]](#footnote-23) Since observing monotonicity violations would likely lead to additional inefficiencies in the estimation, the team decided to adopt measures to limit this risk. For Pilot 3 this was achieved by imposing constraints on the minimum spacing between the tax amounts, as described in point (5) above. Similar concerns led to a further reduction of the number of tax amounts, from 7 as used in Pilot 2, to 6.

The results of pilot 3 led to further refinement of the tax amounts selected for the main study. Using both (1) all data and (2) pilot 3 data only, we repeated the simulation study and quantified the expected extent of the bias resulting from using the lower-bound non-parametric estimator with a vector of 5 or 6 tax amounts.[[24]](#footnote-24) The three best fitting parametric distributions of WTP assumed for this exercise were the negative binomial, gamma and gamma with zero inflation component. The new simulations estimated the expected bias if the highest tax amount was set at the level corresponding to 10% and 5% acceptance rates, while the lowest tax amount was retained at the level of 3 reais. In addition to selecting tax amounts that minimized expected bias (while imposing spacing that limited the expected probability of monotonicity violations to 5%), we also simulated the case of tax amounts minimizing the monotonicity violations altogether, corresponding to placing tax amounts in such a way that the expected distances between acceptance rates were maximized (evenly spaced between the expected acceptance rate for the minimum and the maximum tax amount).

After considering the expected bias resulting from the considered options, we decided to choose the most conservative one of them – 5 tax amounts placed in a way that minimized the risk of monotonicity violations. The highest tax amount was established at the level of 375, which corresponds to an expected acceptance rate of approximately 5%. This reduced the expected bias (relative to the tax amount corresponding to the 10% level acceptance rate) and further reduced the risk of observing monotonicity violations.

Overall, it must be noted that the final choices we made were conservative, in terms of resulting in the low risk of monotonicity violations but a higher bias (conditional on the data generating process assumed in our simulations). This leads to the expected lower-bound non-parametric estimate of WTP to be lower but associated with lower uncertainty and more robust to methodological assumptions. The expected bias for the lower-bound non-parametric mean WTP estimate (based on the data collected in the pilot studies) was 45%. This means that, in our simulations the lower-bound estimate of mean WTP was expected to be 45% lower than the mean WTP in the data generating process.

## WTP analysis

Using the data from each pilot[[25]](#footnote-25) we analyzed the mean WTP estimate following from non-parametric (lower-bound) and parametric approach, including the analysis of the shape of the best-fitting parametric distributions of WTP. In addition to allowing for the efficient design of the tax amounts for the final study, as described in the previous section, our analysis aimed at making sure the WTP estimates are stable – that they are not unintentionally influenced by other changes in the survey. All our analyses showed that (conditional on the introduced changes) while the estimates were becoming more efficient (less uncertain), there were no significant or unexpected changes in the estimates of mean WTP resulting from different pilot studies’ data batches.

The pilot-based analyses were also used to confirm that our adopted estimation approach was practical and methodologically correct. We found that the adopted approach worked with no problems, confirming that, for example, the mean WTP implied by the best fitting parametric distributions were very similar. This corroborated the theoretical expectation that the large range of parametric distributions considered granted enough flexibility, and that the selection criteria for best fit were robust, for the implied mean WTP not to be significantly affected by the arbitrary choice of a parametric distribution selected to model the underlying WTP distribution.

The analysis of WTP changes in the pilot studies also included simulating the quantiles of the implied (expected) distributions of WTP (e.g., the implied WTP of the 10% of respondents who were willing to pay the most). This allowed us to confirm that they remained relatively stable between pilot studies, and between the best fitting parametric distributions considered. In addition, the observed results allowed us to verify that our mean WTP results are not influenced by the experimental design of tax amounts.

Finally, the analysis of each of the pilot study data sets included all of the elements mentioned above. However, we used sub-samples that included/excluded respondents who could be classified as not disclosing their true preferences in the survey but instead used the survey to express their opinions about other issues, e.g., the government, corruption, etc. (protest responses). This analysis was focused on better understanding the reasons why some respondents might choose to use the opportunity of participating in the survey to express their opinions about other issues, instead of accepting the scenario described and truthfully answering the survey questions. It allowed us to identify potential problems in our descriptions of the program and to revise the instrument to minimize the number of respondents who rejected the scenario and may have provided responses unrelated to their true (but unknown) preferences.

## Analysis of the indicators of validity of the study

The three pilot studies allowed us to refine the instrument and make sure the descriptions and questions were clear, relevant, engaging, and well understood. We also used the pilot studies to observe the usual indicators of the validity and reliability of a stated preference study. In particular, these analyses included:

* Sensitivity of votes to tax amounts – the theory of economics predicts that higher taxes should be associated with lower acceptance rates.
* Sensitivity of votes to respondents’ income levels – there are no theoretically expected relationships for respondents income; the theory of economics predicts that respondent’s WTP for any given public good can both increase and decrease with income ([Flores and Carson, 1997](#_heading=h.45jfvxd)).
* Sensitivity of votes to other socio-demographic characteristics – there are no theoretically expected relationships for respondents socio-demographics and probability of voting ‘yes’, however, we conducted this analysis to see if there are any unexpected results that could indicate problems in the survey instrument.
* Sensitivity of votes to attitudinal variables collected in the survey – for several attitudinal statements it can be expected that higher agreement rates are associated with larger WTP; we verified whether this was the case.
* Sensitivity of votes to other quality-related variables resulting from the survey (e.g., interviewer effects, perceived pressure, completion time, interviewer’s evaluation of respondents’ engagement, objective factors jeopardizing survey scenario, such as not paying for electricity, etc.) – this analysis aimed at identifying specific problems with survey administration.
* Sensitivity of parametric estimates of WTP to econometric treatment of data (model uncertainty)

Overall, our analysis confirmed that data collected using our survey was well-behaved and that no changes introduced to the survey in the process of its development led to violating its validity.

In addition, the pilot studies’ results were used to investigate the sensitivity of protest responses to tax amounts, respondents’ income levels and other socio-demographic and attitudinal variables collected in the survey, as well as other quality-related variables resulting from the survey. These analyses confirmed that the definition of protest ‘no’ (‘against’) votes (in contrast to legitimate ‘no’ votes) was valid and identified respondents who used the survey to express their opinions about other issues, rather than the good being valued. We also used this analysis to identify the elements of the survey that had the highest potential of reducing the rejection of the hypothetical scenario and making respondents answer the survey questions truthfully. This led us to, for example, refine the descriptions of the safety program and payment consequentiality. This allowed us to make sure that the final instrument was as objective, valid and robust as possible.

## Questionnaire Changes

The pilot studies were also used to refine the wording of the survey to remove redundant or irrelevant questions and response items, as well as make other changes to improve the credibility and perceived consequentiality of the instrument.

One of the specific issues that we tested in the pilot studies was the placement of the income question. While many studies ask about respondents’ income (and other sensitive information) at the end of the survey ([Dillman, Smyth and Christian, 2014](#_heading=h.1ljsd9k)), more recent stated preference valuation studies ([e.g., Bishop et al., 2017](#_heading=h.2szc72q)) ask about income before the vote question. The reason for this is that respondents who are asked about their income first may treat the tax amount provided in the survey as more tailored to their income level (since most survey scenarios say the individual tax amounts will be income-dependent). As a result, the respondents may find the survey questions more credible, and consider their responses more consequential. A potential drawback of this procedure is that if respondents are not sincere in providing their income level, they may treat the tax amount they see next as based on the incorrect (exaggerated or diminished) income and expect that the actual cost, should it be based on their actual income, will be different (lower or higher, respectively).

In the pilot studies we used a split-sample treatment to test if the placement of income question before or after the vote question matters for vote responses, observed income levels, income non-response, protest behavior or survey completion. We did not find any significant differences and hence in the final version of the survey we proceeded with asking the income question before the vote question.

Another issue tested in the pilot studies was reversing the order of responses in survey questions, for example, from ‘definitely agree’ to ‘definitely disagree’ or in reversed order. We did not observe any significant differences associated with the order of survey responses.

Similar issues studied in the pilot questions included: the selection items for the close-ended list of reasons to vote ‘against’, the selection of ‘pause’ questions (meant to keep respondent engaged and testing their comprehension of the descriptions), and the order of survey questions. In each case, we did not observe any quantitatively observable differences (such as differences in voting patterns, protest responses, comprehension and engagement levels) resulting from the changes that were introduced.

# Final Survey Administration and Results

This section presents the main results of the final survey, including an estimate of the damages, as well as a number of validity checks of the data and models. We start by summarizing the sampling process and the main principles guiding the final survey administration. Then, results are presented, starting with the most conservative, non-parametric estimates of the damages, followed by parametric estimates. In order to validate our results, we performed (validity) tests on our sample aiming to check if respondents voted following patterns that would be expected based on the economic valuation literature. It is important to note that all validity tests used non-weighted sample data. However, as indicated in the statistical literature, sample weights were applied for estimation of the population mean WTP estimates, and the aggregated damage values presented in section 6.2.

Weights have been used for adjusting the final distribution of interviews to the targeted population and are based on three parameters: 1) distribution of family income within income groups within each metropolitan region / municipality; 2) weight of each metropolitan region / municipality in relation to the region of Brazil to which it belongs and within each income bracket, and 3) weight of each region of Brazil in relation to the total households in urban areas of the country, also within each income bracket.

We used Brazilian Census and Annual Household Survey data estimates of population, income and region of residence to estimate individual weights in our sample. Since 67 respondents in our sample did not disclose their income levels, these 67 observations would have been lost in our estimation of population weighted mean WTP. In order to avoid this loss, however, we used regression analysis to predict (‘impute’) the missing information on income. The regression analysis was based on respondents’ social-demographic characteristics (i.e. city of residence, working status, number of household residents, number of children in the household, respondent’s gender, and education level), and resulted in a model used to impute the missing income levels for the 67 respondents.[[26]](#footnote-26)

## Operational indicators

This section describes the sampling design and implementation procedures.

### Sampling[[27]](#footnote-27)

Our sampling plan specified and defined the following aspects of the final survey: the population of interest to the research; the units or sample points where the interviews were conducted; the procedure for determining the sample size, its probabilistic confidence level and sample errors, the stratification criteria, and the procedures for selecting or drawing sample units. They are summarized below.

The sample was stratified by income. It was also stratified by metropolitan regions. Rural households, which account for 14% of all Brazilian households, were not surveyed due to logistical reasons due to its spatial dimension that would have created barriers to randomization and imposed restrictions on the timescale of the study.

The universe or target population of the research was defined by the total number of permanent private households, with monthly income, families or conglomerate of people residing throughout the country. In the absence of the 2019 census data, the methodology adopted to carry out household estimates and projections for 2019, was based on data from the 2010 Demographic Census and the 2018 National Household Sample Survey (PNAD) allowing estimating the annual geometric mean growth rate of households in the period to project the number of households for the year 2019. This procedure was adopted for Brazil as a whole and for all regions and municipalities selected for the field research.

In addition, once the number of households in each location of interest for the survey was estimated, the structure of relative composition of households in the 2010 Demographic Census was used to estimate the number of households by monthly household income range for these locations for 2019.

Ideally, in surveys like ours, the sample size is calculated as a function of the maximum error desired in the estimates, using the sample data, of the population parameters of the random variables or characteristics of interest that are the object of the investigation. However, the sample size determination process is usually inverted because of limited time and budgets. Therefore, the number of interviews to be carried out is defined first and later the intrinsic error of the sample is calculated. It was no different in this study. We started with the prospect of obtaining a maximum of 5,200 interviews, given time and budget constraints.

The main population parameter of interest is the mean WTP. However, in order to obtain a consistent mean WTP estimate, it is necessary and sufficient to have first a consistent estimate of the proportion of people willing to pay, since for both parametric and non-parametric approaches the estimators of mean WTP are functions of the proportions of people willing to pay.

We then determined that the proportion of people willing to pay and the mean WTP values were the population parameters of greatest interest.[[28]](#footnote-28) Then, to determine the sample size for each metropolitan region/municipality:

Let  be a discrete random variable of type  that represents an individual’s willingness to pay. That is,  if the person is willing to pay and , otherwise, where:

. (10)

Let  be a sample of size  of this population. Let  be an unbiased and consistent estimator of . So, the problem is getting the sample size  such that:

, (11)

where:

 proportion to be estimated;

 estimator of proportion ;

 sampling error;

 level of significance or the probability that the estimate error be greater than the sampling error ;

 level of confidence or probabilistic reliability of the sample (usually 95%).

Therefore, the problem is to obtain the sample size  that guarantees an estimation error less than or equal to  with high probabilistic reliability. After some probabilistic calculations, which can be found in textbooks on sampling theory ([Cochran, 1977](#_heading=h.36ei31r)), the solution to the problem is given by the following expression:

 (12)

and:

 sample size for finite population;

 size of the target population;

 abscissa value of the  distribution that determines ;

 variance of the variable of interest  (since it is unknown, usually the maximum variance is assumed, ).

Evaluating the final sample size, it can be seen in Table 6 that sample error for each tax amount presented is at the desired level of statistical significance, calculated with a 95% confidence level and maximum variance, sufficient to estimate the population parameters of interest. For each metropolitan region/municipality related to average tax amount, the sample errors shown in Table 7 are equally low.

Sample stratification did not require that interviews be conducted in the cities of Ipatinga (MG), Mariana (MG) and Linhares (ES), all located in the FDI-impacted region. However, by including small numbers of residents of these cities in the final study, it was possible to explore whether the survey results would be significantly different in those localities compared to other locations in Brazil that were not directly affected by the FDI. Of particular interest was whether residents of these cities tended to select possible protest motivations for their vote on the dry tailings program. As will be shown in Appendix 3, responses from these localities did not differ in this regard from other locations.

Table 6: Sample error by tax amount

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tax amount [REAIS]** | **Number of ‘against’ votes** | **Number of ‘for’ votes** | **Number of votes (total)** | **Share of ‘against’ votes (p)** | **Share of ‘for’ votes (1-p)** | **Sampling Error, with probabilistic confidence level of 95%** |
| 3 | 602 | 429 | 1031 | 0,5839 | 0,4161 | 0,03009 |
| 12 | 713 | 336 | 1049 | 0,6797 | 0,3203 | 0,02824 |
| 38 | 754 | 278 | 1032 | 0,7306 | 0,2694 | 0,02707 |
| 109 | 860 | 183 | 1043 | 0,8245 | 0,1755 | 0,02308 |
| 375 | 915 | 125 | 1040 | 0,8798 | 0,1202 | 0,01976 |
| **Total** | **3844** | **1351** | **5195** | **0,7399** | **0,2601** | **0,01193** |

Table 7: Sample size and error by region/municipality

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Brazil, Metropolitan Region and Municipality** | **Population: Urban Particular Households, with income, in year 2019** | **Sample size surveyed (excluding incompletes)** | **Sampling Error, with probabilistic confidence level of 95%** | **Survey**  **Dates** |
| RM - São Paulo | 6,607,581 | 1,050 | 0,02719 | 26/11/2019 to 27/01/2020 |
| RM - Rio de Janeiro | 4,276,364 | 660 | 0,03368 | 13/12/2019 to 29/01/2020 |
| RM - Belo Horizonte | 1,894,453 | 404 | 0,04120 | 21/01/2020 to 20/02/2020 |
| Mun. - Ipatinga (MG) | 78,119 | 203 | 0,05254 | 07/02/2020 to 15/02/2020 |
| Mun. - Linhares (ES) | 43,548 | 203 | 0,06042 | 05/02/2020 to 14/02/2020 |
| Mun. - Mariana (MG) | 15,425 | 204 | 0,06190 | 05/02/2020 to 13/02/2020 |
| RM - Recife | 1,279,086 | 953 | 0,02821 | 26/11/2019 to 29/01/2020 |
| RM - Curitiba | 1,082,243 | 618 | 0,03627 | 22/01/2020 to 09/03/2020 |
| RM - Goiânia | 827,643 | 398 | 0,04465 | 29/01/2020 to 13/03/2020 |
| RM - Belém | 640,142 | 502 | 0,03169 | 28/11/2019 to 13/01/2020 |
| **Total (Brasil)** | **55,949,131** | **5,195** | 0,01193 | **26/11/2019 to 13/03/2020** |

Source: IBGE - Instituto Brasileiro de Geografia e Estatística

Note: Estimated number of permanent private households with household income, for year 2019, based on the 2010 Demographic Census expanded with data from the 2018 Annual Household Survey (PNAD).

The research team was concerned that a higher awareness and possibly deeply felt emotions about the FDI in these localities could increase the risk of dissemination of the hypothetical scenario in the questionnaire through local press and social media. Large samples in these areas would increase that risk and perhaps compromise the survey. For that reason, smaller sample sizes were used for these three cities. And because of that, a greater sampling error occurred in there. It did not adversely affect the overall aggregate estimate of mean WTP, however, because the total sample from these cities represented a small fraction of the country total (see Table 7).

Finally, it is important to note that the randomness of our stratified sample is maintained by the random selection of regions, neighborhoods and blocks within each metropolitan region (stratum). Once a block was selected, a pre-established procedure was in place for selecting the households in the block (see item 5 of Appendix 1). Such procedure does not affect the sample randomness since there is no a-priori information about the residents.

### Survey administration and verification

High standards of quality control were maintained throughout the pilot tests and the final survey. The fieldwork team, Quest Inteligencia de Mercado, based in São Paulo, was responsible for training local interviewers, supervising survey administration, verification of completed interviews and codification of the data. A coordination team was always present in the metropolitan regions where interviews took place, with at least 3 coordinators present during the initial five days of effort in each metropolitan region/municipality to evaluate all aspects of the fieldwork. From the sixth day on one coordinator stayed with each local interview team.

In each metropolitan region, a local team consisting of a supervisor and one or two verifiers, depending on the sample size, were trained in the proper use of the survey materials and administration protocols in order to guarantee effective use of the computer assisted personal interviewing (CAPI) software (“Survey to Go”). Local interviewers received training consisting of reading and understanding of the questionnaire and of the protocols to be applied in the interviews (see Appendix 1: Interview Protocol Report)

All interviewers were tested regarding their reading proficiency and use of the protocols. The test consisted of a simulation of an interview with a supervisor. For each interviewer who passed the test, their first days of work were monitored by a supervisor.

Three fieldwork teams (a supervisor, a verifier and interviewers in each team) who conducted the interviews in the cities of Mariana, Ipatinga and Linhares were sent from the cities of São Paulo and Rio de Janeiro after the completion of the fieldwork survey in these metropolitan regions.

Completed interviews were carefully verified.[[29]](#footnote-29) At least 20% of the completed questionnaires for each interviewer were evaluated and verified by the local supervisor and verifier(s) team. Final approval, or rejection, of each completed interview, was the responsibility of the Sao Paulo coordination team, which analyzed each completed interview before final approval. Completion of the review and verification process occurred within two weeks of the completion of interview.

Quality control measures for each completed interview employed consistency checks, including a review of the time respondents took to answer each question and the total time of interview, and a review of the declared income compared to other household characteristics and number of people contributing to the household income.

Interviews with completion times much lower than two standard deviations from the average can indicate a survey interview that was accomplished in too short period of time to be valid[[30]](#footnote-30). In those cases, questionnaires were separated to be evaluated and verified. Six questionnaires were refused due to total interview time lower than 2 standard deviations from the mean value. Also, questions answered in a very short time can indicate that the question was accidentally jumped by the interviewee and invalidate the questionnaire. Additionally, interviews with completion times longer than two standard deviations in relation to the average may indicate an unusual pause in the interview and were separated to be analyzed. The verification protocol specified that if there was a question time longer than 30 minutes, that questionnaire would be refused, but no such cases were reported.

In order to review the declared income, a score created by the Brazilian Association of Research Companies (ABEP) based on the household declared possessions was used to reflect the respondent socioeconomic classification, and compared to the declared household income. Interviews with potential inconsistencies were separated and analyzed for later verification. A total of 233 questionnaires were verified for possible income inconsistencies, from which 26 were rejected and 207 approved (approximately 4% of the total sample).

Quality control of field work took into account capabilities that the online CAPI management system offered, such as GPS, which allowed us to check if the interviews were conducted in the specified neighborhoods at the pre-determined path

Quest Inteligencia de Mercado sent all tablets to fieldwork previously configured with only the CAPI software (Survey To Go). The tablet included software (Pulsus) that locked the configuration of the survey. This prevented interviewers from changing any configuration pre-determined by the Sao Paulo coordination team.

Daily production of each interviewer, as measured in the number of completed interviews, was also carefully monitored. The average daily production per interviewer was between 3 and 4 completed interviews, rarely over these figures. Owing to the combination of software georeferenced (GPS) data collection, with date, time, interview length, IMEI (universal unique identification of the tablet) and a photo of the interviewer, all obtained automatically by the tablet computer, the veracity of the work done by each interviewer could be assured.

As an additional check on the validity of the completed interviews, some key points were verified by either a telephone contact with the respondents, or a second visit of a supervisor. The purpose of these contacts was to confirm details such as age, socioeconomic classification, income, and who contributed to household expenses. Table 8 summarizes the number of verifications conducted by metropolitan region.

Table 8: Review and verification of completed interviews

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Region** | **% verification** | **Verification On-site** | **Verification By phone** | **No verification (only review)** | **Total** |
| RM São Paulo | 30.4% | 22 | 297 | 731 | 1050 |
| RM Rio de Janeiro | 27.1% | 49 | 130 | 481 | 660 |
| RM Belo Horizonte | 30.0% | 28 | 93 | 283 | 404 |
| Mariana | 30.9% | 13 | 50 | 141 | 204 |
| Ipatinga | 30.5% | 14 | 48 | 141 | 203 |
| Linhares | 37.4% | 15 | 61 | 127 | 203 |
| RM Curitiba | 44.8% | 45 | 232 | 341 | 618 |
| RM Goiânia | 35.2% | 28 | 112 | 258 | 398 |
| RM Recife | 28.8% | 56 | 218 | 679 | 953 |
| RM Belém | 27.7% | 17 | 122 | 363 | 502 |
| **Total** | **31.8%** | **287** | **1363** | **3545** | **5195** |

#### Intercept interviews

According to our sampling strategy, the sample unit is composed of individuals with income who reside in private urban households. However, due to security reasons and personal safety concerns, interviewers were not permitted in certain buildings or neighborhoods. For example, in multi-residence apartment buildings or condominiums, doormen often do not allow any strangers to enter unless authorized by a resident. For this reason, some of the interviews were obtained via intercepts on the street, in general in multi-unit buildings areas. As can be seen in Table 9, less than 2% of our final sample was obtained via intercept interviews.

Table 9: Household and Intercept Interviews

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Region** | **Household surveys** | **Intercept surveys** | **Share of intercept surveys** | **Total** |
| RM São Paulo | 1047 | 3 | 0.29% | 1050 |
| RM Rio de Janeiro | 626 | 34 | 5.15% | 660 |
| RM Belo Horizonte | 393 | 11 | 2.72% | 404 |
| Mariana | 204 | 0 | 0.00% | 204 |
| Ipatinga | 203 | 0 | 0.00% | 203 |
| Linhares | 203 | 0 | 0.00% | 203 |
| RM Curitiba | 577 | 41 | 6.63% | 618 |
| RM Goiânia | 395 | 3 | 0.75% | 398 |
| RM Recife | 946 | 7 | 0.73% | 953 |
| RM Belém | 502 | 0 | 0.00% | 502 |
| **Total** | **5096** | **99** | **1.91%** | **5195** |

#### Interview Dispositions

Table 10 summarizes interview dispositions for each city/metropolitan region. It can be seen that only 0.3% of interviews that were actually started, were broken off or “abandoned” before completion. Nearly one-third (32%) of all interview attempts were “refusals” (potential respondents who refused requests for interviews).

Table 10: Interviews Outcomes

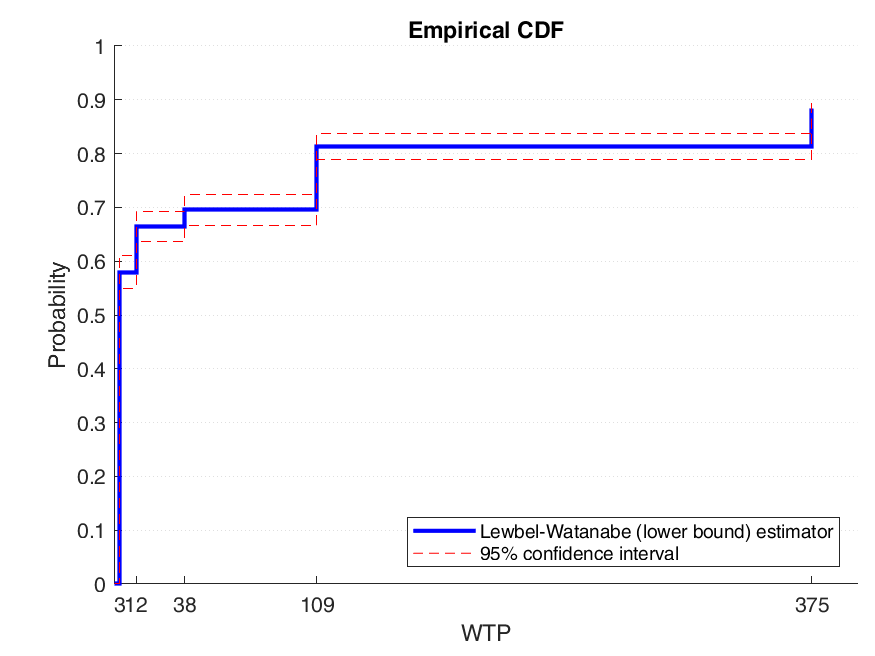
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Occurrence | São Paulo | Rio de Janeiro | Belo Hzte | Mariana | Ipatinga | Linhares | Curitiba | Goiânia | Recife | Belém | Total | **Total (%)** |
| Missing respondent (appointment), empty household or nobody opens door | 7,384 | 2,980 | 1,431 | 453 | 1,674 | 741 | 2,253 | 423 | 2,527 | 2,995 | 22,861 | **39.0%** |
| No one qualified in the household (screening) | 2,606 | 348 | 551 | 326 | 507 | 114 | 379 | 141 | 741 | 970 | 6,683 | **11.4%** |
| Risky area, not allowed into the building or difficult access | 168 | 53 | 30 | 6 | 42 | 22 | 44 | 10 | 173 | 224 | 772 | **1.3%** |
| Refused interview request | 4,764 | 3,312 | 1,501 | 248 | 895 | 827 | 1,164 | 142 | 3,175 | 2,719 | 18,747 | **32.0%** |
| Abandoned interview | 47 | 32 | 31 | 8 | 9 | 8 | 1 | 32 | 7 | 2 | 177 | **0.3%** |
| Out of quota | 574 | 305 | 252 | 48 | 127 | 73 | 152 | 171 | 312 | 1,461 | 3,475 | **5.9%** |
| Completed interview invalidated due to quality control or interviewer falsification | 9 | 15 | 27 | 0 | 2 | 0 | 26 | 9 | 35 | 6 | 129 | **0.2%** |
| Other reasons (block with no households, equipment failure) | 204 | 52 | 25 | 4 | 59 | 18 | 79 | 2 | 15 | 93 | 551 | **0.9%** |
| Completed Interviews | 1,050 | 660 | 404 | 204 | 203 | 203 | 618 | 398 | 953 | 502 | 5,195 | **8.9%** |
| **Total attempts** | **16,806** | **7,757** | **4,252** | **1,297** | **3,518** | **2,006** | **4,716** | **1,328** | **7,938** | **8,972** | **58,590** | **100.0%** |
| **Success ratio** | **6.2%** | **8.5%** | **9.5%** | **15.7%** | **5.8%** | **10.1%** | **13.1%** | **30.0%** | **12.0%** | **5.6%** | **8.9%** |  |

## Survey Results

This section describes the estimation procedures and results.

### Non-parametric estimate of population WTP

The respondents’ weighted votes for each tax amount make it possible to estimate the CDF, as explained in section 3.3. This is equivalent to Lewbel-Watanabe non-parametric estimator of the population WTP. The empirical CDF is presented in Figure 9.



*Figure 9. The Lewbel-Watanabe weighted non-parametric estimator of the population WTP distribution*

These results allow for calculating the non-parametric estimate of mean WTP   
as the area above the CDF plotted in Figure 9. The lower-bound (non-parametric-based) estimate of mean WTP, along with its standard error and 95% confidence interval is presented in Table 11.

Table 11: The lower-bound (non-parametric-based) estimate of WTP [REAIS]

|  |  |  |
| --- | --- | --- |
| **WTP  (mean)** | **WTP  (s.e.)** | **WTP 95% c.i.** |
| 56.34 | 2.81 | 50.84 - 61.84 |

Considering that the program was associated with 10 fixed monthly installments, the non-parametric analysis implies that lower bound estimate of total WTP is 563.38 reais with a 95% confidence interval of 508.39 - 618.37 reais.

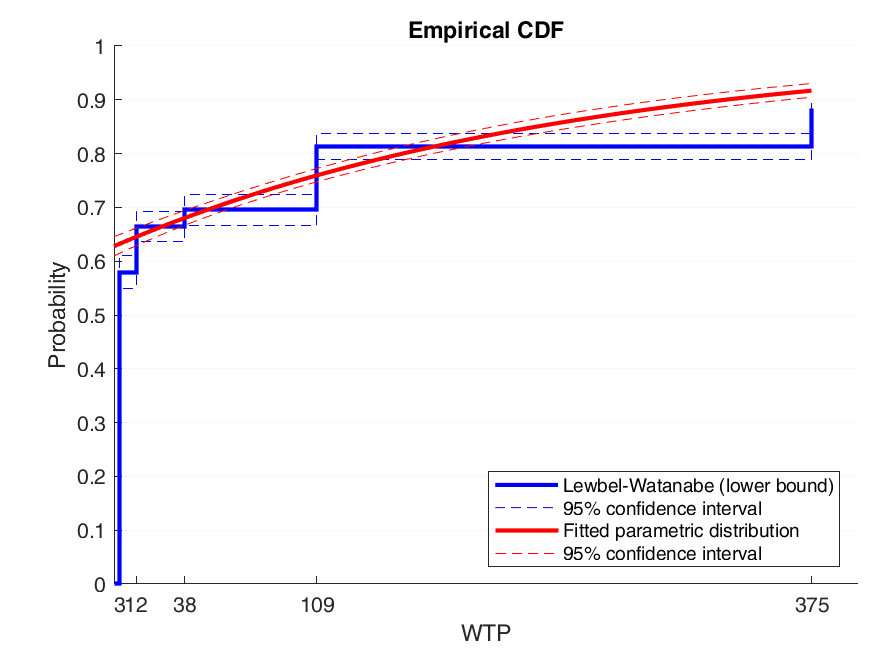
### Parametric estimate of WTP

Following the procedure described in section 3.3, we fitted many typical parametric distributions to the weighted voting data. The results are presented in Table 12. Each distribution was tried with and without a zero-inflation component.[[31]](#footnote-31) The fit of each distribution is indicated by the value of the log-likelihood function (the higher the better). However, it is expected that distributions using more parameters will generally fit the data better. In order to avoid over-parameterization of the data, we based our comparisons on the Bayesian Information Criterion (BIC), which penalizes specifications for additional parameters they require. A lower BIC implies a better fit. The specifications in Table 12 are sorted in terms of fit from best to worst. Finally, for each specification, the last two columns of Table 12 present the implied (simulated) mean WTP along with its standard error.

The results presented in Table 12 indicate that in the case of our data, the best-fitting parametric distribution was exponential with a zero-inflation component. However, it is worth noting that in terms of fit indicators, the results (and irrespective of whether BIC or AIC is used as the fit indicator) are very close for the top 19 specifications and result in similar WTP measures in the range of approximately 87-118 reais. The CDF of the best fitting parametric distribution is presented in Figure 10, along with the non-parametric weighted (Lewbel-Watanabe) estimator.

Table 12: The results of the parametric estimation of respondents’ WTP

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Distribution** | **Log-likelihood** | **Zero inflation** | **AIC** | **BIC** | **Param.** | **WTP (mean)** | **WTP  (s.e.)** |
| Exponential | -2909.53 | 1 | 1.1209 | 1.1234 | 2 | 94.58 | 7.53 |
| Gamma | -2905.35 | 1 | 1.1197 | 1.1235 | 3 | 94.30 | 10.03 |
| Negative\_Binomial | -2905.52 | 1 | 1.1197 | 1.1235 | 3 | 94.54 | 9.85 |
| Weibull | -2908.17 | 1 | 1.1208 | 1.1245 | 3 | 86.63 | 7.99 |
| Nakagami | -2912.64 | 1 | 1.1225 | 1.1263 | 3 | 98.42 | 6.61 |
| BirnbaumSaunders | -2914.25 | 1 | 1.1231 | 1.1269 | 3 | 94.81 | 9.39 |
| Uniform | -2916.98 | 1 | 1.1241 | 1.1279 | 3 | 87.89 | 4.61 |
| Inverse\_Gaussian | -2926.36 | 1 | 1.1278 | 1.1315 | 3 | 95.09 | 8.70 |
| Lognormal | -2928.09 | 1 | 1.1284 | 1.1322 | 3 | 96.37 | 9.11 |
| Rayleigh | -2934.62 | 1 | 1.1306 | 1.1331 | 2 | 101.36 | 5.18 |
| Negative\_Binomial | -2900.98 | 0 | 1.1322 | 1.1348 | 2 | 94.41 | 10.22 |
| Gamma | -2902.98 | 0 | 1.1330 | 1.1355 | 2 | 94.31 | 10.51 |
| Loglogistic | -2942.52 | 1 | 1.1340 | 1.1378 | 3 | 113.30 | 17.87 |
| Poisson | -2950.07 | 1 | 1.1365 | 1.1390 | 2 | 118.18 | 0.64 |
| Generalized\_Extreme\_Value | -2943.88 | 1 | 1.1349 | 1.1399 | 4 | 111.34 | 7.70 |
| Normal | -2948.53 | 1 | 1.1363 | 1.1401 | 3 | 101.79 | 3.49 |
| Logistic | -2949.92 | 1 | 1.1368 | 1.1406 | 3 | 107.42 | 3.19 |
| Extreme\_Value | -2949.98 | 1 | 1.1369 | 1.1406 | 3 | 117.01 | 3.85 |
| Weibull | -2974.93 | 0 | 1.1610 | 1.1636 | 2 | 115.77 | 23.10 |
| Inverse\_Gaussian | -3147.04 | 0 | 1.2282 | 1.2307 | 2 | 86.90 | 21.16 |
| Exponential | -4980.09 | 0 | 1.9427 | 1.9440 | 1 | 46.13 | 1.69 |
| Nakagami | -6425.17 | 0 | 2.5067 | 2.5092 | 2 | 77.54 | 1.89 |
| Logistic | -7904.43 | 0 | 3.0836 | 3.0862 | 2 | 46.16 | 1.62 |
| Uniform | -8510.86 | 0 | 3.3202 | 3.3227 | 2 | 193.89 | 3.72 |
| Normal | -8684.97 | 0 | 3.3881 | 3.3906 | 2 | 75.05 | 2.45 |
| Extreme\_Value | -10185.68 | 0 | 3.9734 | 3.9759 | 2 | 105.01 | 3.40 |
| Rayleigh | -10806.60 | 0 | 4.2151 | 4.2164 | 1 | 82.89 | 1.61 |
| Poisson | -100179.41 | 0 | 39.0719 | 39.0732 | 1 | 44.09 | 0.21 |



*Figure 10. Weighted non-parametric and the parametric estimators of the population WTP distribution*

Figure 10 shows the empirical CDF of the best-fitting parametric distribution for our data, along with the non-parametric weighted Lewbel-Watanabe estimator. For the tax amounts used in our study (3, 12, 109, 375) both estimators are very close, resulting in similar expected probability of voting ‘against’.

Our parametric estimate of the population mean WTP, along with its standard error and 95% confidence interval is presented in Table 13.

Table 13: The parametric-based estimate of WTP [REAIS]

|  |  |  |
| --- | --- | --- |
| **WTP  (mean)** | **WTP  (s.e.)** | **WTP 95% c.i.** |
| 94.58 | 7.53 | 79.81 - 109.35 |

Considering that the program was associated with 10 fixed monthly installments, the parametric analysis implies that mean WTP is 945.78 reais with a 95% confidence interval of 798.11 – 1,093.45 reais.

### Aggregate Lost Value

The average DAP estimates allow to determine the value of environment and historical and cultural heritage damages calculated with an aggregated DAP measure for Brazilian society for the moral damage and for the loss of services and environmental resources and of the historical cultural heritage over and at the end of recovery activities.

The WTP inferred from our contingent valuation approach represents mean WTP of an average urban Brazilian household. Calculating the aggregate value of damages therefore requires multiplying the mean WTP by the number of total urban households in Brazil (Table 14).

Table 14: Number of Permanent Private Households with Household Income in Brazil, Total, Urban and Rural, according to Major Regions

|  |  |  |  |
| --- | --- | --- | --- |
| Brazil, Regions | **2019** | | |
| **Total** | **Urban** | **Rural** |
| **Brasil** | 64,836,233 | **55,949,131** | 8,887,102 |
| South-east | 28,569,397 | 26,702,062 | 1,867,334 |
| North-east | 16,631,919 | 12,609,267 | 4,022,653 |
| South | 10,243,165 | 8,786,688 | 1,456,477 |
| Center-west | 4,955,421 | 4,428,491 | 526,929 |
| North | 4,436,331 | 3,422,622 | 1,013,709 |

Source: IBGE - Instituto Brasileiro de Geografia e Estatística

Note: Estimated number of permanent private households with household income, for year 2019, based on the 2010 Demographic Census expanded with data from the 2018 Annual Household Survey (PNAD).

Considering the 55,949,131 urban households in Brazil, our **conservative (lower-bound) estimate of environmental, heritage damages caused by the FDI to the Brazilian society, in terms of moral damage and loss of services and resources throughout and at the end of the recovery, is 31.52 billion reais** (with a 95% confidence interval of 28.44 - 34.60 billion reais)[[32]](#footnote-32). The parametric estimate, based on choosing the best-fitting probability distribution to observed votes of survey participants, implies the aggregate value of damages equal to 52.92 billion reais (with a 95% confidence interval of 44.65 - 61.18 billion reais).[[33]](#footnote-33)

### Analysis of the indicators of validity of the study

This section presents the results of the statistical (regression) analyses that are aimed at investigating the validity of the study. In our analyses, we now use unweighted data, because we are no longer interested in calculating the mean WTP and extrapolating the results to the larger population, but instead we want to demonstrate the validity of our study for the available sample of respondents.

As mentioned before, the linear probability model estimated for these results using ordinary least squares regression produces consistent and unbiased estimator, however, unless all coefficients are zero heteroskedascity is present. To account for this issue we use standard heteroskedascity-robust standard errors (MacKinnon and White, 1985). We note that for robustness check, we also used binary probit models as presented in Appendix 4,

#### Sensitivity to tax amounts and income levels

The theory of economics predicts that higher tax amounts should be associated with lower acceptance rates (i.e. more ‘against’ votes). Similarly, respondents with higher incomes are expected to be more likely to accept a given tax amount (more ‘for’ votes), as found by most empirical studies. Table 15 presents the number of votes for each tax amounts used in the survey. Tax amounts were randomly selected for each survey respondent, so they are each used approximately the same number of times (total number of votes for each tax amount was approximately 1026). An important feature of the responses to the vote question is that as the tax amounts increase, the shares of ‘against’ votes increase and the shares of ‘for’ votes decrease. This is expected – the higher the cost, the lower the number of citizens who would likely be willing to accept it, and hence the lower the probability of observing the ‘for’ vote in the sample.

Table 15: Frequency of votes for tax amounts used in the survey

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tax amount [REAIS]** | **Number of ‘against’ votes** | **Number of ‘for’ votes** | **Number of votes (total)** | **Share of  ‘against’ votes** | **Share of  ‘for’ votes** |
| 3 | 602 | 429 | 1031 | 58.39% | 41.61% |
| 12 | 713 | 336 | 1049 | 67.97% | 32.03% |
| 38 | 754 | 278 | 1032 | 73.06% | 26.94% |
| 109 | 860 | 183 | 1043 | 82.45% | 17.55% |
| 375 | 915 | 125 | 1040 | 87.98% | 12.02% |

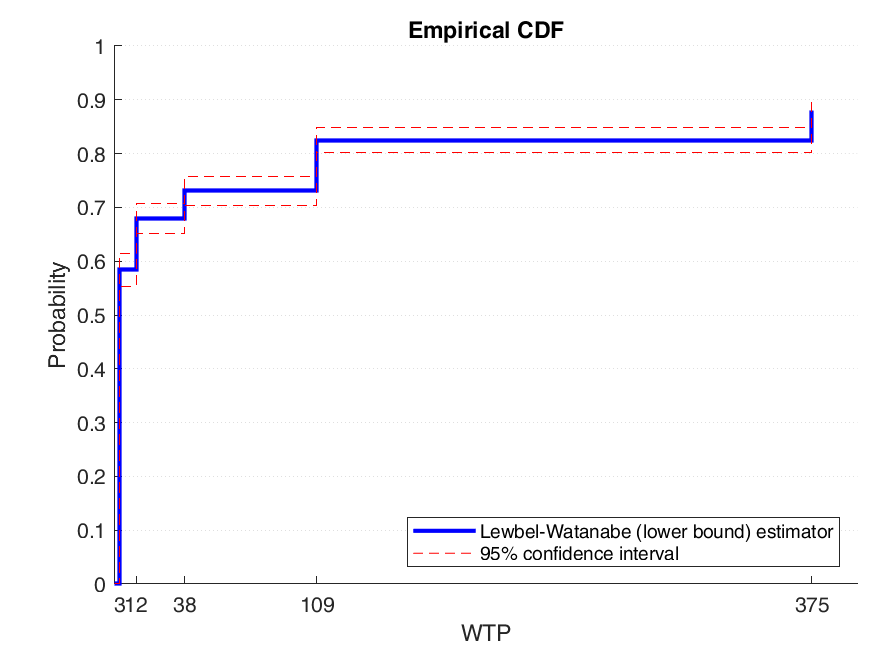
It is possible to test if the increase in the share of ‘against’ votes associated with increasing tax amounts is statistically significant. We first do this using the shares of ‘against’ votes corresponding to each paid of adjacent tax amounts. The null hypothesis is that the percentage of ‘against’ votes is the same, and the one-sided test alternative is that the vote percentage is larger for the larger of the two tax amounts. The test is performed using z-statistics and the results are presented in Table 16. We find that the observed increase in the share of ‘against’ votes is statistically significant for all cases at the 0.01 level (for most pairs of adjacent tax amounts the p-value is lower than 0.001). This implies that our results are consistent with the economic theory, in observing no monotonicity violations with respect to the shares of ‘against’ votes for increasing tax amounts.

Table 16: Pair-wise tests of proportions voting ‘against’ the program

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tax amount 1 [REAIS]** | **Tax amount 2 [REAIS]** | **Share of  ‘no’ votes 1** | **Share of  ‘no’ votes 2** | **Difference** | **P-value** |
| 3 | 12 | 58.39% | 67.97% | 9.58%\*\*\* | 0.0000 |
| 12 | 38 | 67.97% | 73.06% | 5.09%\*\*\* | 0.0054 |
| 38 | 109 | 73.06% | 82.45% | 9.39%\*\*\* | 0.0000 |
| 109 | 375 | 82.45% | 87.98% | 5.53%\*\*\* | 0.0002 |

Notes: \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively (one-sided test).

The observed shares of ‘against’ and ‘for’ votes for each tax amount make it possible to estimate the CDF, as explained in section 3.3. The empirical CDF for unweighted data is presented in Figure 11. In line with the numerical results presented in Table 16, the probability of observing an ‘against’ vote increases with increasing tax amounts. The red dashed lines indicate 95% confidence intervals, which demonstrate that the differences in the probability of an ‘against’ vote for adjacent tax amounts are statistically significant. This graphically illustrates that the data is well behaved in terms of observing no monotonicity violations.[[34]](#footnote-34)



*Figure 11. The Lewbel-Watanabe non-parametric estimator of the sample WTP distribution*

The relationships between votes and tax amounts and respondents’ income are further established in the seven models presented in Table 17. The models use simple linear regression, in line with the approach outlined in section 3.3. The models differ with respect to other controls (in addition to tax amount and income) that are included. They include socio-demographics, regional fixed-effects, city fixed-effects, responses to debriefing questions (analyzed separately below) and quality controls (such as the number of incorrect responses to ‘pause’ questions, perceived pressure to vote ‘for’ or ‘against’ the program, respondents’ involvement and understanding level, as perceived by interviewer, an indicator of whether the survey was taken while sitting or standing, and interviewer fixed effects).

The results of the models presented in Table 17 demonstrate the expected relationships for tax amounts and income levels. The positive and statistically significant coefficient of the logarithmic transformation[[35]](#footnote-35) of the tax amount indicates an increasing probability of observing a ‘against’ vote for increasing tax amounts. The negative and significant coefficients of the logarithmic transformation of respondent’s income in all but one case[[36]](#footnote-36) indicate that the probability of observing an ‘against’ vote decreases with increasing respondent’s income. Overall, this is an indication of the validity of our estimates.

Table 17: The evaluation of the effects of tax amounts and income on the probability of observing a ‘no’ vote

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Model 1** | **Model 2** | **Model 3** | **Model 4** | **Model 5** | **Model 6** | **Model 7** |
| Constant | 0.5184\*\*\* (0.0152) | 1.0363\*\*\* (0.0721) | 1.0011\*\*\* (0.1717) | 0.8731\*\*\* (0.0781) | 0.8676\*\*\* (0.0789) | 0.7917\*\*\* (0.0756) | 1.1137\*\*\* (0.0948) |
| log(Tax amount) | 0.0620\*\*\* (0.0035) | 0.0626\*\*\* (0.0035) | 0.0615\*\*\* (0.0035) | 0.0625\*\*\* (0.0034) | 0.0625\*\*\* (0.0034) | 0.0269\*\*\* (0.0034) | 0.0633\*\*\* (0.0034) |
| log(Income) |  | -0.0672\*\*\* (0.0092) | -0.0973\*\*\* (0.0130) | 0.0625\*\*\* (0.0034) | -0.0871\*\*\* (0.0126) | -0.0086 (0.0086) | -0.0578\*\*\* (0.0101) |
| Socio-demographic controls | no | no | yes | no | no | no | no |
| Region fixed-effects | no | no | no | yes | no | no | no |
| City fixed-effects | no | no | no | no | yes | no | no |
| Debriefing questions | no | no | no | no | no | yes | no |
| Quality controls | no | no | no | no | no | no | yes |
| R2 (adj) | 0.0561 | 0.0680 | 0.0808 | 0.0850 | 0.0866 | 0.3071 | 0.1100 |
| *k* (parameters) | 2 | 4 | 23 | 17 | 36 | 25 | 111 |
| *n* (observations) | 5195 | 5195 | 5195 | 5195 | 5195 | 5195 | 5195 |

Notes: heteroskedascity-robust standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

A more detailed analysis of the reasons for voting ‘against’, including protest responses identification and validation, is presented in Appendix 3.

#### Attitudinal and debriefing questions

The validity of the study can also be demonstrated with the sensitivity of votes to responses to attitudinal and debriefing questions in the survey. Even though there may not be theoretically clear links between them, we explore the impacts of these variables on the votes.

Table 18 presents the regression results that include attitudinal and debriefing responses as explanatory variables of respondents’ votes. For simplicity, the ordinal (Likert scale) responses are treated as if they were continuous, while missing responses are controlled with additional dummy variables.

The results show that respondents who had visited Mariana in the past, who say they consider themselves concerned about the environment, those who agreed that they trust scientists’ assessment of the damage caused by the FDI to the environment and to historical and cultural heritage, and those who believe that the implementation of the dry tailings is important to avoid forever another dam break in Brazil were not more likely to vote ‘against’ or ‘for’.

Respondents who believe in the terms given in years for environmental recovery that would be needed to return the Rio Doce basin back to what it was before the FDI, those who said that the implementation of the dry tailing in Brazil will be possible in up to two years with the financial help of the government were significantly less likely to vote ‘against’. Respondent’s belief in whether the government will implement dry tailings if the majority of the Brazilian population vote ‘for’ was also negatively correlated with the probability of voting ‘against’. Similar negative and statistically significant effects were found for those who believe that the government will actually raise the tax amount if dry tailings technology is deployed, and believe that with the financial help of the government mining companies will implement dry tailing technology. Respondents who believed that the tax amount they would actually have to pay would be higher (lower) than specified were more (less) likely to vote against the program, while we did not observe a statistically significant relationship for respondents who were more concerned that it would be difficult for their household to pay the specified tax amount.

Table 18: The evaluation of the relationship between debriefing questions responses and the probability of observing an ‘against’ vote

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Range of values** | **Mean value** | **Coefficient (st. err.)** |
| Constant | - | - | 0.7917\*\*\* (0.0756) |
| log(Tax amount) | [1.0986, 5.9269] | 3.5708 | 0.0269\*\*\* (0.0034) |
| log(Income) | [6.9078, 9.6158] | 7.7707 | -0.0086 (0.0086) |
| log(Income) – missing response | {0,1} | 0.0129 | 0.0782\*\* (0.0320) |
| Visited Mariana | {0,1} | 0.0957 | 0.0000 (0.0174) |
| Concerned about environment | {1,2,…,5} | 4.3875 | -0.0069 (0.0057) |
| Trust the scientists’ assessment of the damage | {1,2,…,5} | 3.2353 | -0.0036 (0.0051) |
| Believe recovery time | {1,2,…,5} | 3.0598 | -0.0109\*\* (0.0051) |
| Believe the implementation of dry tailing is important for stopping future dam breaks | {1,2,…,5} | 3.6622 | -0.0020 (0.0046) |
| Implementation of the dry tailing in Brazil with the financial help of the government possible in up to two years | {1,2,…,5} | 2.4534 | -0.0442\*\*\* (0.0054) |
| The government will implement dry tailings if the majority of the Brazilian population vote ‘for’ | {1,2,…,5} | 2.6385 | -0.0500\*\*\* (0.0055) |
| The government will actually raise the tax if dry tailings technology is deployed | {1,2,…,5} | 3.2498 | -0.0187\*\*\* (0.0040) |
| With the financial help of the government mining companies will implement dry tailing technology | {1,2,…,5} | 2.6178 | -0.0291\*\*\* (0.0055) |
| Specified tax amount would be lower, the same, or higher | {-1,0,1} | 0.3689 | 0.0277\*\*\* (0.0073) |
| Payment of the specified tax would be difficult for household | {1,2,3,4} | 3.1742 | 0.1153\*\*\* (0.0060) |
| Concerned about environment – missing response | {0,1} | 0.0010 | 0.0835 (0.2128) |
| Trust the scientists’ assessment of the damage – missing response | {0,1} | 0.0035 | 0.0945\* (0.0496) |
| Believe recovery time – missing response | {0,1} | 0.0021 | 0.0395 (0.0655) |
| Believe the implementation of dry tailing is important for stopping future dam breaks – missing response | {0,1} | 0.0033 | -0.0942 (0.0739) |
| Implementation of the dry tailing in Brazil with the financial help of the government possible in up to two years – missing response | {0,1} | 0.0023 | 0.1039\* (0.0578) |
| The government will implement dry tailings if the majority of the Brazilian population vote ‘for’ – missing response | {0,1} | 0.0027 | 0.0630 (0.0632) |
| The government will actually raise the tax if dry tailings technology is deployed – missing response | {0,1} | 0.0042 | 0.1351\*\*\* (0.0451) |
| With the financial help of the government mining companies will implement dry tailing technology – missing response | {0,1} | 0.0027 | 0.0407 (0.0845) |
| Specified tax amount would be lower, the same, or higher – missing response | {0,1} | 0.0373 | 0.0235 (0.0225) |
| Payment of the specified tax would be difficult for household – missing response | {0,1} | 0.0031 | 0.0698 (0.0667) |
| R2 (adj) |  |  | 0.3071 |
| *k* (parameters) |  |  | 25 |
| *n* (observations) |  |  | 5195 |

Notes: heteroskedascity-robust standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

Overall, we find that the responses to attitudinal and debriefing questions are linked with the probability to support the program in an expected way (or are not statistically significant). In particular, we note that the questions referring to perceived policy and payment consequentiality of the survey are linked with the probability of voting ‘against’ in a way that was expected from other empirical studies available in the literature (e.g., Zawojska, Bartczak and Czajkowski, 2019). These issues are of special significance, given the importance of perceived consequentiality for incentive compatibility of the survey (Carson and Groves, 2007; Carson, Groves and List, 2014).

#### Quality-indicating questions

Finally, Table 19 presents the regression results for the model including additional explanatory variables related to survey quality. Specifically, we find that respondents who answered incorrectly to more ‘pause’ questions, which may indicate inattention or rejection of the descriptions included in the survey, were more likely to vote ‘against’. Respondents who said they felt pressured to vote ‘for’ (6.1% of respondents) and those who said they felt pressured to vote ‘against’ (3.6% of respondents) were more likely to vote ‘against’, than respondents who said they felt free to vote.[[37]](#footnote-37)

Respondents who said they could not decide if they felt free of pressure to vote ‘for’ or ‘against’ the program or refused to answer were also more likely to vote ‘against’ (this result is only weakly significant, at 0.1 level). Other variables, such as respondents’ involvement or understanding (as perceived by the interviewer), whether the survey was taken sitting or standing (which can be an indication of whether interviewer was invited to respondent’s home or the interview was conducted standing in the door) and whether the survey was an intercept survey (conducted at random in public area – see Section 6.1) were not significant explanatory variables of the probability to vote ‘against’ at the 0.05 level.

Table 19: The evaluation of the relationship between survey quality indicators and the probability of observing a ‘no’ (‘against’) vote

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Range of values** | **Mean value** | **Coefficient (st. err.)** |
| Constant | - | - | 1.1148\*\*\* (0.0815) |
| log(Tax amount) | [1.0986, 5.9269] | 3.5708 | 0.0623\*\*\* (0.0034) |
| log(Income) | [6.9078, 9.6158] | 7.7707 | -0.0618\*\*\* (0.0095) |
| log(Income) – missing response | [0,1] | 0.0129 | 0.1746\*\*\* (0.0337) |
| Number of incorrect ‘pause’ questions | {1,2,…,6} | 0.0362 | 0.0540\*\*\* (0.0122) |
| Respondent felt pressured to vote ‘for’ the program | {0,1} | 0.0610 | 0.1454\*\*\* (0.0208) |
| Respondent felt pressured to vote ‘against’ the program | {0,1} | 0.0364 | 0.1828\*\*\* (0.0189) |
| Respondent could not respond if felt free of pressure to vote ‘for’ or against the program | {0,1} | 0.0050 | 0.1513\*\*\* (0.0547) |
| Respondent’s involvement as perceived by interviewer | {1,2,…,5} | 4.6135 | -0.0289\*\* (0.0139) |
| Respondent’s understanding as perceived by interviewer | {1,2,…,5} | 4.5879 | -0.0016 (0.0143) |
| Survey taken sitting | {1,2,…,5} | 3.5392 | 0.0010 (0.0032) |
| Intercept survey | {0,1} | 0.0191 | -0.0045 (0.0442) |
| R2 (adj) |  |  | 0.0828 |
| *k* (parameters) |  |  | 12 |
| *n* (observations) |  |  | 5195 |

Notes: heteroskedascity-robust standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

Overall, the results presented in Table 19 show that the data are ‘well-behaved’ – that is, they can be interpreted as an indication that the survey was conducted consistently and according to all quality-assuring protocols such that the conditions under which the survey was conducted do not appear to have influenced the results.

# Conclusions

The goal of this study was to estimate the monetary value of damages to the population of Brazil from injuries to environmental and heritage resources resulting from the FDI, using the contingent valuation methodology.

The study team was comprised of Brazilian and international experts with extensive academic and governmental experience carrying out environmental valuation studies, including natural resource damage assessments.

A conservative approach to the questionnaire design and analysis of results was followed by the study team throughout this project, when there were reasonable alternatives available on how best to proceed, in order to choose the one that would more likely result in either lowering, or having a neutral impact (when compared to the other alternatives under consideration), on the final estimate of total damages, including the choice of the wording in the questionnaire narratives.

This conservative approach also resulted in not including damages at the river mouth and sea coast, since those impacts were not known at the time of the survey instrument design and not excluding potential protest votes against the dry tailings program.

As in any contingent valuation study, the survey instrument presented respondents with an object of choice, described a context in which a choice was to be made, and then asked respondent to make a choice. For this study, the injuries from the FDI environmental and heritage resource impacts were presented and then respondents were asked to vote ‘for’ or ‘against’ a safety program to prevent one future tailings dam failure that would have the exact same consequences for environmental and heritage resources as the FDI, at a specific price (tax amount).

The injury scenario was described in the questionnaire using available quantitative indicators for the impacts and expected recovery times, as provided by Lactec scientists.

The survey questionnaire underwent extensive testing during the development phase of the study. The testing procedures minimized any potential for bias that could prevent the revelation of respondents’ true preferences when they were asked to vote for or against the safety program (i.e. the object of choice) in the final survey.

The survey sample targeted the population of urban Brazilian households, with income, residing in cities and metropolitan areas located in representative areas of the country. The fieldwork was undertaken to maintain a high standard of quality control.

The study applied parametric and non-parametric modelling to estimate a monetary value for the injuries caused by the FDI. If we followed a conservative approach, the aggregate environmental and heritage damages caused by the FDI would be the lower value of the non-parametric estimate. That is, the urban Brazilian population represented in this study revealed that the environmental and heritage damages from the FDI were worth at least 31.52 billion reais covering moral damage and losses of services and throughout and at the end of the recovery period. Considering other conservative decisions made in the course of the study, this aggregate damage estimate can be considered a (very) conservative, lower-bound value.

# Appendix 1: Interview Protocol Report

All interviewers, in each city where the survey was conducted, were trained for approximately eight hours and tested for competency before being hired. In this appendix, we present the main instructions that interviewers received in order to standardize their performance in conducting interviews while mitigating potential interviewer bias. At the end we present the training dates and the number of qualified interviewers. The instructions were as follows:

1. Contact with the respondent

* You should not express opinions or demonstrate feelings as they can influence the respondent's responses.
* Show the potential respondent the importance of the study for universities, but do so without mentioning the Mariana or Brumadinho incidents, or that the study discusses environmental issues, until this is mentioned during the questionnaire application.
* If the respondent wants to know more about the study, you can explain the three sections of the questionnaire:
  + The first section is about issues that the country currently faces.
  + The second regards a program studied by universities that affect the population.
  + The last section asks your opinion.
* It is important not to mention anything about voting for or against a tax and possible values.
* Avoid influencing the interviewee's responses during the interview. Conversations and praise can influence a respondent in ways that may bias their answers to the survey.
* Do not comment on respondents’ preferences regarding the research topic.
* Only talk about subjects totally distant from the research topic, such as, heat, rain, traffic, etc.
* Look directly at the respondent when conducting the interview. If you have difficulty looking from time to time, follow the “LOOK” instruction that appear in the middle of the text. So, when you return to reading, you will not get lost.
* Value your respondents. Make sure that they know how much their opinions are valued. Remind them that all answers will be read by the researchers who will analyze the data.
* Do not influence the interviewee. Avoid influencing the respondent to meet what he or she perceives as your expectations regarding specific answers to interview questions.
* Do not answer what was not asked.
* Be brief and kind.

1. Responses to potential refusals

* Importance of the interview: an opportunity to discover how Brazilians think about some problems and their solutions.
* Use of information: it will be presented in an aggregated form, by region and income class, to express an average of the opinions and never the opinion of only one individual.
* Privacy: information and opinions will be treated confidentially, as will the responses of all people interviewed.
* Who funds the research: resources from the scientific community that investigates issues that affect the population.
* Where the research will be conducted: Interviews are conducted in other cities, but I am only doing interviews in this city.

1. Interview instructions

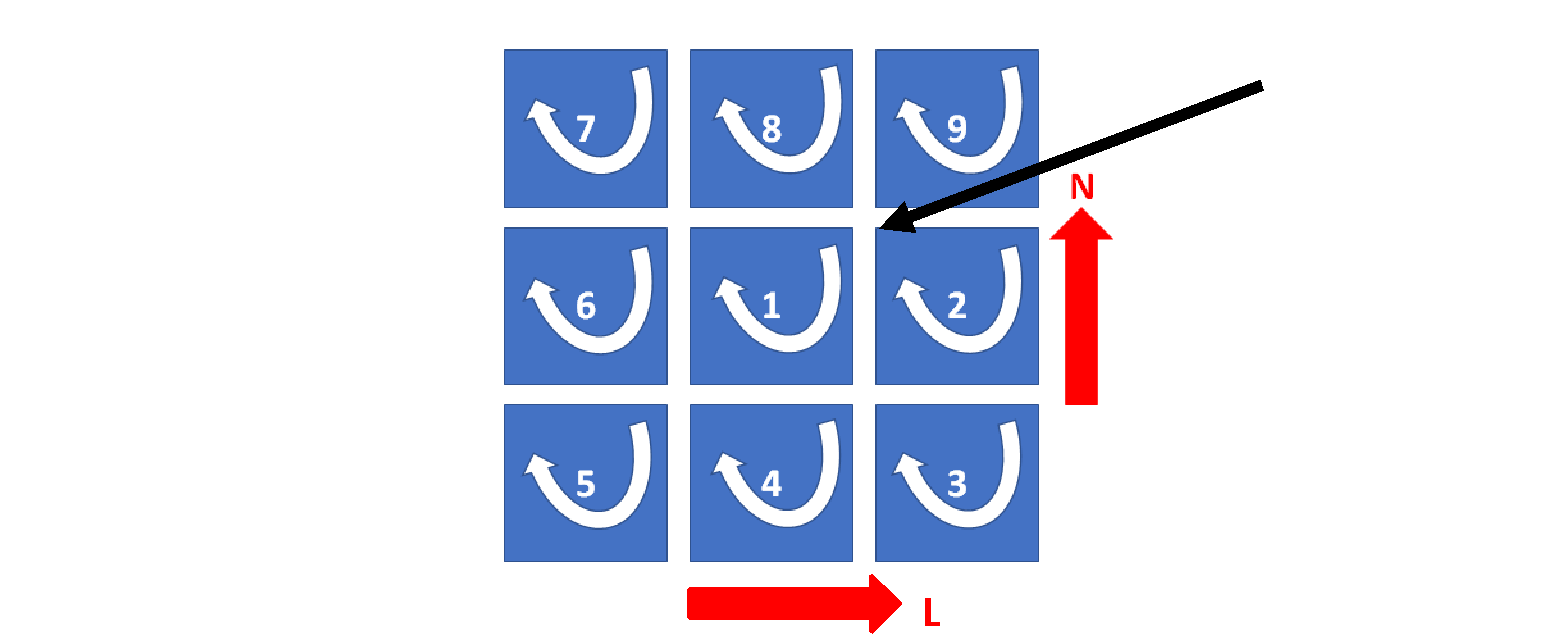
* Never re-phrase any instructions, questions or answers. Always read the questions as they are written. If you make a mistake while reading the text, apologize, take a two-second break, and read the sentence again.
* Do not emphasize any words or phrases, unless specified (such as underlining or bold text).
* Do not define any words, except as specified.
* Don't assume for yourself the answer to any question. Do not check an answer category unless the respondent chooses it explicitly.
* Do not paraphrase or change the answers to open-ended questions. Always record responses in full and in the first person.
* Do not react to any responses.
* Don't skip questions.
* Do not paraphrase or alter the writing to clarify the respondent's response.
* Read at the pace and time you were trained. Read more slowly if prompted. But, never faster than it was trained, even if asked.
* Whenever objections are raised, it is vitally important that the respondent knows that you have heard. The most effective phrase you can use is "I understand". Use as many times as you need.
* If the respondent points to the answer on one of the presented cards, without mentioning the answer, ask him or her to read the answer.
* At the beginning of the interview, write down the name of the respondent as he or she wants to be called. His or her name will be spoken only where it is in the text to be read.

1. Questions during the interview

* Reaction to the respondent's questions is one of the most critical points in an interview.
* Engagement: Do not interact with the respondent by exchanging opinions or clarifying technical questions or any other issues.
* Technical questions:
  + Questions from respondents should be answered by saying “let's read the question again”.
  + Those that are not answered in the re-reading, just say “let's continue and see what the questionnaire says later”. Shorten engagement by moving on to the next question.
  + If the respondent insists, say “I don't know; it's your understanding that matters”. Never respond with your opinion or knowledge.
* Objections to the yes or no answers: Whether the answer was yes followed by an objection or protest, you do not need to repeat it and accept yes.
* ‘no’ answers only: ask “did you understand the question?”.
  + If the answer is “I understand, but I don't agree”, check ‘yes’ because the respondent understood although he(she) disagrees.
  + If the answer is only “I don't agree”, ask if he(she) wants you to read the question again and re-read only the question.
  + If “I disagree” persists, ask “did you understand the question?”. If the answer remains “I don't agree”, reinforce “but did you understand the question or do you want me to read it again?”
* Doubts in choosing the categorized answer: For questions with category answers the respondent has to make a choice. If she responds vaguely, say "I will repeat the answer options and please choose one". If the answer is still vague, you should insist again saying, "It would be helpful if you could choose one of these answers" and then read the third time again. You cannot choose an answer that fits what he says.
* Response of a category before the end of the reading: continue reading all the categories until the end and then ask a question and listen to the answer again.
* I didn't understand: if respondent didn't understand, always read the whole question with the question at the end.
* Don't know the answer: Never read the answers “I don't know” or “don't want to answer” and only check them after reading the answers a third time.
* Photos: If asked which fish, animal or bird in the photo, say “these are examples of the species mentioned in the text”. If she insists, say “I don't know”. In the photos of the historical and cultural heritage if asked what is in the photo, say "that is what is written next to the photo". If she insists, say “I don't know”.
* Cards: Maintain vigilance so that the card in front of the respondent is always aligned with the question being read. And that, when going to another question without a card, check that the respondent is not yet reading the card from the previous question that had a card.
* Conversation at the end of the interview: At the end of the interview, just say that the results of the interviews will be analyzed by experts and sent to the government to decide the best option to deal with the public policy problem. Never express personal opinions or information that is not in the questionnaire. If asked about who sponsored the survey, just say “this survey is being conducted by some Universities”.

1. Household selection

* Starting point on the corner of the north and east faces of block 1.
* Search for the respondent household in a clockwise direction from the starting point of block 1 in the figure below.



* At the end of the route of block 1, restart the search for the corner of the north and east faces of block 2, and so on for the next blocks, according to the increasing number.
* After each interview, the interviewer should skip the next 3 households and try the next interview in the 4th household.
* Conduct the interview according to the quotas established for the neighborhood.
* When the interview screener is started and the first person in the household is under 17 years old, or does not share expenses, or does not speak Portuguese, try to speak to someone else in the house who does not have these disqualifying characteristics.
* If the interview is interrupted for 30 minute, it will be considered as non-response and search for another household will be according to the criterion above described.
* If you do an interview in an apartment building, do 1 interview per floor.
* If you find a village with a gate and several houses at the same address, do only 1 interview at that address.

Table 20 below shows training dates and qualified Interviewers per city.

Table 20: Training dates and qualified interviews

|  |  |  |  |
| --- | --- | --- | --- |
| **Metropolitan Region (RM) and Municipalities (Mun)** | **Training Dates**  **(day/month/year)** | **Fieldwork Period**  **(day/month/year)** | **Qualified Interviewers** |
| RM - São Paulo | 25 a 27/11/2019 | 26/11/2019 a 27/01/2020 | 17 |
| RM - Rio de Janeiro | 11 e 12/12/2019 | 13/12/2019 a 29/01/2020 | 14 |
| RM - Belo Horizonte | 20 e 21/01/2020 | 21/01/2020 a 20/02/2020 | 13 |
| Mun - Ipatinga (MG) | 06/02/2020 | 07/02/2020 a 15/02/2020 | 6 |
| Mun - Linhares (ES) | 04/02/2020 | 05/02/2020 a 14/02/2020 | 7 |
| Mun - Mariana (MG) | 04/02/2020 | 05/02/2020 a 13/02/2020 | 6 |
| RM – Recife | 24 a 26/11/2019 | 26/11/2019 a 29/01/2020 | 18 |
| RM – Curitiba | 20 e 21/01/2020 | 22/01/2020 a 09/03/2020 | 11 |
| RM – Goiânia | 27 e 28/01/2020 | 29/01/2020 a 13/03/2020 | 16 |
| RM – Belém | 26 e 27/11/2019 | 28/11/2019 a 13/01/2020 | 11 |

# Appendix 2: Regression used to impute missing income responses

Table 21 presents the parameters of the regression used to impute missing income responses.

Table 21: The results of the regression used to impute missing income responses

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Range of values** | **Mean value** | **Coefficient (st. err.)** |
| Constant | - | - | 12023.6597\*\*\* (868.3279) |
| City – Diadema | {0,1} | 0.0054 | -476.5715\*\* (232.8205) |
| City – Guarulhos | {0,1} | 0.0139 | 136.7245 (219.0716) |
| City – Osasco | {0,1} | 0.0083 | 19.5192 (306.0090) |
| City – Santo André | {0,1} | 0.0083 | 138.5088 (347.9177) |
| City – São Bernardo do Campo | {0,1} | 0.0098 | -289.1132 (238.0474) |
| City – São Caetano do Sul | {0,1} | 0.0021 | -543.1249 (392.5354) |
| City – Belo Horizonte | {0,1} | 0.0539 | 257.8396\*\* (131.0142) |
| City – Betim | {0,1} | 0.0065 | 9.4576 (181.9215) |
| City – Contagem | {0,1} | 0.0119 | -272.7340\* (149.2292) |
| City – Ribeirão da Neves | {0,1} | 0.0054 | -112.9911 (153.4425) |
| City – Recife | {0,1} | 0.1111 | -59.3757 (114.2787) |
| City – Jaboatão dos Guararapes | {0,1} | 0.0462 | -362.4831\*\*\* (114.8017) |
| City – Olinda | {0,1} | 0.0262 | -434.0293\*\*\* (132.7677) |
| City – Rio de Janeiro | {0,1} | 0.0882 | -43.0910 (115.6393) |
| City – Nova Iguaçu | {0,1} | 0.0117 | -341.1093\* (184.1972) |
| City – Duque de Caxias | {0,1} | 0.0133 | -409.3187\*\* (161.0433) |
| City – Niterói | {0,1} | 0.0079 | 571.1453 (370.2775) |
| City – São Gonçalo | {0,1} | 0.0060 | -63.7501 (350.4443) |
| City – Belém | {0,1} | 0.0722 | -272.4893\*\* (120.0006) |
| City – Ananindeua | {0,1} | 0.0244 | -166.8558 (179.2801) |
| City – Curitiba | {0,1} | 0.1055 | 6.0253 (108.8262) |
| City – São José dos Pinhais | {0,1} | 0.0135 | -178.2149 (216.7026) |
| City – Goiânia | {0,1} | 0.0614 | -210.6289 (148.1890) |
| City – Aparecida de Goiânia | {0,1} | 0.0152 | -268.3740 (278.3961) |
| City – Ipatinga | {0,1} | 0.0391 | 7.9369 (142.8908) |
| City – Linhares | {0,1} | 0.0383 | -6.4673 (113.4287) |
| City – Mariana | {0,1} | 0.0393 | -54.4824 (151.4936) |
| City – other | {0,1} | 0.0008 | -369.0572 (325.0813) |
| Work status – home maker | {0,1} | 0.0470 | -530.5404\*\*\* (107.8542) |
| Work status – retired | {0,1} | 0.1748 | -68.0131 (109.5672) |
| Work status – unemployed without benefits | {0,1} | 0.0152 | -701.1852\*\*\* (171.2108) |
| Work status – unemployed with benefits | {0,1} | 0.0200 | -532.4767\*\*\* (133.0671) |
| Work status – autonomous/freelance | {0,1} | 0.2745 | -177.3301\*\* (79.8062) |
| Work status – part–time freelance | {0,1} | 0.1309 | -586.5627\*\*\* (75.8133) |
| Work status – pensioner | {0,1} | 0.0510 | -245.0940\*\* (114.1948) |
| Work status – savings | {0,1} | 0.0167 | -325.2540 (238.2577) |
| Work status – family allowance | {0,1} | 0.0393 | -484.4472\*\*\* (85.9130) |
| Household size – 2 | {0,1} | 0.2225 | -440.3495 (485.5968) |
| Household size – 3 | {0,1} | 0.2551 | -379.3942 (486.7065) |
| Household size – 4 | {0,1} | 0.2135 | -156.2112 (487.6392) |
| Household size – 5 | {0,1} | 0.1209 | -6.5544 (493.1946) |
| Household size – 6 | {0,1} | 0.0560 | 191.8707 (501.9652) |
| Household size – 7 | {0,1} | 0.0189 | 355.4110 (518.5721) |
| Household size – 8 | {0,1} | 0.0144 | 245.5473 (522.0609) |
| Household size – 9 | {0,1} | 0.0038 | 972.6614 (636.0445) |
| Household size – 10 | {0,1} | 0.0025 | 1513.3205\* (889.8692) |
| Household size – 11 | {0,1} | 0.0008 | 1006.6244 (788.2873) |
| Household size – 12 | {0,1} | 0.0006 | 721.2775 (1092.6082) |
| Household size – 13 | {0,1} | 0.0002 | 3394.8324\*\*\* (499.7372) |
| Household size – 17 | {0,1} | 0.0002 | -298.0642 (555.6310) |
| Household size – 32 | {0,1} | 0.0002 | -968.2094\* (495.7372) |
| Household children – 1 | {0,1} | 0.2397 | -280.4401\*\*\* (75.2575) |
| Household children – 2 | {0,1} | 0.1436 | -313.0531\*\*\* (100.8992) |
| Household children – 3 | {0,1} | 0.0510 | -551.1734\*\*\* (129.5481) |
| Household children – 4 | {0,1} | 0.0187 | -469.8849\*\*\* (168.4793) |
| Household children – 5 | {0,1} | 0.0058 | -1082.6147\*\*\* (226.4933) |
| Household children – 6 | {0,1} | 0.0015 | -1607.7519\*\*\* (435.5753) |
| Household children – 7 | {0,1} | 0.0002 | -2172.0706\*\*\* (768.2010) |
| Female | {0,1} | 0.5034 | -425.8580\*\*\* (61.9839) |
| Age | [18,97] | 47.3936 | 6.6077\*\*\* (2.3808) |
| Household wealth | [5,74] | 25.3411 | 364.0233\*\*\* (14.3928) |
| log(Household wealth) | [1.6094, 4.3041] | 3.1530 | -5564.4078\*\*\* (305.5057) |
| Household head | {0,1} | 0.6117 | 284.9591\*\*\* (62.3268) |
| Education – primary (part) | {0,1} | 0.0589 | -488.2541\* (288.3953) |
| Education – primary | {0,1} | 0.2062 | -393.3274 (284.1399) |
| Education – secondary | {0,1} | 0.1680 | -166.4652 (286.9952) |
| Education – high school | {0,1} | 0.4071 | -222.2402 (286.9724) |
| Education – undergraduate | {0,1} | 0.1305 | 335.7893 (311.7788) |
| Education – post graduate | {0,1} | 0.0277 | 1471.7235\*\*\* (416.9449) |
| Intercept survey | {0,1} | 0.0191 | 1350.0912\*\*\* (336.5532) |
| R2 (adj) |  |  | 0.5748 |
| *k* (parameters) |  |  | 71 |
| *n* (observations) |  |  | 4659 |

Notes: heteroskedascity-robust standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

# Appendix 3: Analysis of reasons for voting ‘against’ and the identification of protest responses

A protest vote occurs when a respondent either refuses to provide an answer to the valuation/WTP (referendum) question, or votes ‘against’ for reasons unrelated to his or her ability to pay or true preference for the item being valued. Instead, protest votes are associated with discontent or disapproval of some aspect of the valuation process. Therefore, these votes may be considered an inaccurate representation of the respondent’s actual preferences for the item being valued, and indicate some other concern.

This appendix presents an analysis of the potential protest votes that occurred in the answers to the final survey’s valuation question. We start by investigating the main reasons for voting ‘against’ the program presented in the referendum question at a given tax amount. Then we present the results of statistical models used to evaluate socio-demographic characteristics of potential protest respondents, as well as their responses to debriefing questions and survey quality indicators.

**Reasons for voting ‘against’ and protest response identification**

Respondents who voted ‘against’ were asked to state the main reason for their vote. Table 22 lists the possible reasons for voting ‘against,’ along with the number of respondents choosing this option. Reasons 1, 2, 3, 8, 9, 10, 11, 14 and 15 may be considered legitimate reasons to vote ‘against’, reflecting respondents’ true preferences for the benefits of the program. On the other hand, reasons 4, 5, 6, 7, 12 and 13 (in bold), may indicate that respondents voted ‘against’ for reasons unrelated to their preferences for the dam safety program and avoiding environmental and heritage damages associated with future incident. These responses can be an indication of a potential protest vote, which does not reflect these respondents’ economic preferences. According to this classification, 39.52% of the respondents can be considered protest responses (53.41% of those who voted ‘against’). The majority of protests indicated a disbelief in government as a reliable provider of public services, given that corruption, mismanagement of public resources and doubts that the tax would be used for DT deployment led to rejection of the program.

Table 22: Reasons for voting ‘against’

|  |  |  |
| --- | --- | --- |
| **Reasons** | **Number  of cases** | **Share** |
| 1 – I have no income to pay this amount | 689 | 17.9% |
| 2 – I would rather use this money for another personal purpose | 187 | 4.9% |
| 3 – It is not important to stop the mud dams | 4 | 0.1% |
| **4 – I do not want to give my money to the miners** | **521** | **13.5%** |
| **5 – The money from this tax will be mismanaged** | **481** | **12.5%** |
| **6 – The money from this tax will not be used to deploy dry tailings** | **177** | **4.6%** |
| **7 – The money goes to corruption** | **858** | **22.3%** |
| 8 – I already paid a lot of taxes | 73 | 1.9% |
| 9 – Oher reason | 25 | 0.6% |
| 10 – Refuse to answer | 7 | 0.2% |
| 11 – The value is too high for me | 722 | 18.8% |
| **12 – The government should pay, not the population** | **51** | **1.3%** |
| **13 – The government and miners should pay, not the population** | **33** | **0.9%** |
| 14 – Money should be used for another purpose | 8 | 0.2% |
| 15 – Concerned about other payers | 8 | 0.2% |
| Total | 3844 | 100% |

**Analysis of socio-demographic characteristics, debriefing responses and survey quality indicators for protest respondents**

Table 23 presents the results of a linear regression model used to analyze socio-demographic characteristics of protest respondents[[38]](#footnote-38). We find that the tax amount presented to respondents did not increase the probability of a protest (‘against’) vote (while higher tax amounts did increase the probability of observing an ‘against’ vote – see Table 17 for details). Respondents’ income was only weakly significant as an explanatory variable for explaining a potential protest vote. However, respondents with higher household wealth were significantly more likely to protest. Regarding respondent’s other socio-demographic characteristics, we found that respondents who were retired, unemployed, worked part-time freelance jobs, were pensioners, or lived off of a family allowance were less likely to protest than those who were employed full time (reference group). Female respondents were less likely to protest. Age increased the probability of protest votes, however, at a decreasing marginal rate. Respondents’ education level and household size did not appear significantly correlated with the probability of protesting.

Table 23: Explanatory variables of the protest ‘against’ votes – socio-demographics

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Range of values** | **Mean value** | **Coefficient (st. err.)** |
| Constant | - | - | -0.3550\* (0.1958) |
| log(Tax amount) | [1.0986, 5.9269] | 3.5708 | 0.0029 (0.0040) |
| log(Income) | [6.9078, 9.6158] | 7.7707 | 0.0278\* (0.0149) |
| log(Income) – missing response | {0,1} | 0.0129 | 0.1242\*\* (0.0580) |
| Work status – home maker | {0,1} | 0.0470 | -0.0436 (0.0351) |
| Work status – retired | {0,1} | 0.1748 | -0.0547\*\* (0.0264) |
| Work status – unemployed without benefits | {0,1} | 0.0152 | -0.1151\*\* (0.0537) |
| Work status – unemployed with benefits | {0,1} | 0.0200 | -0.1052\*\* (0.0493) |
| Work status – autonomous/freelance | {0,1} | 0.2745 | 0.0087 (0.0194) |
| Work status – part-time freelance | {0,1} | 0.1309 | -0.0475\*\* (0.0240) |
| Work status – pensioner | {0,1} | 0.0510 | -0.0988\*\*\* (0.0340) |
| Work status – savings | {0,1} | 0.0167 | -0.0886 (0.0546) |
| Work status – family allowance | {0,1} | 0.0393 | -0.1212\*\*\* (0.0343) |
| Household size | {1,2,…,32} | 3.4296 | -0.0030 (0.0041) |
| Female | {0,1} | 0.5034 | -0.0552\*\*\* (0.0144) |
| Age | [18,97] | 47.3936 | 0.0109\*\*\* (0.0024) |
| Age2 | [3.2400, 94.0900] | 25.0756 | -0.0106\*\*\* (0.0025) |
| log(Household wealth) | [1.6094, 4.3041] | 3.1530 | 0.0761\*\*\* (0.0250) |
| Education – primary (part) | {0,1} | 0.0589 | 0.0411 (0.1639) |
| Education – primary | {0,1} | 0.2062 | 0.0174 (0.1626) |
| Education – secondary | {0,1} | 0.1680 | 0.1008 (0.1629) |
| Education – high school | {0,1} | 0.4071 | 0.1568 (0.1626) |
| Education – undergraduate | {0,1} | 0.1305 | 0.1565 (0.1638) |
| Education – post graduate | {0,1} | 0.0277 | 0.1751 (0.1681) |
| R2 (adj) |  |  | 0.0563 |
| *k* (parameters) |  |  | 24 |
| *n* (observations) |  |  | 5195 |

Notes: heteroskedascity-robust standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

Since protest vote is mostly driven by attitudes that make respondents distrustful towards the scenarios and proposals presented in the questionnaire, it is expected that debriefing questions can reveal such attitudes. Table 24 presents the linear probability model of observing a protest (‘against’) vote with respect to respondents’ answers to debriefing questions. We found that respondents who stated that they believe: (1) the estimated recovery times; (2) that the implementation of the dry tailing safety program in Brazil with financial help of the government was possible in up to two years; (3) that the government would implement dry tailings if the majority of the Brazilian population voted ‘for;’ and (4) that mining companies would successfully implement dry tailing technology with the financial help of the government – were all less likely to vote ‘against’ for protest reasons. These results confirm that lack of trust in the damage scenario and in the program to avoid another incident were correlated to protest voting.

Respondents who said: (1) that they believe the implementation of the dry tailing safety program is important for stopping future dam breaks; (2) that the actual tax amount would be higher; and (3) that payment of the specified tax amount would be difficult for their household, were more likely to vote against the program. However, the responses to the question on motivations for voting no may be also capturing voters who have economic reasons to vote against, but chose another explanation to justify a vote against the program. Such a possibility makes adjustments in the estimation of mean WTP problematic.

Table 24: Explanatory variables of the protest ‘against’ votes – debriefing questions

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Range of values** | **Mean value** | **Coefficient (st. err.)** |
| Constant | - | - | -0.4388\*\*\* (0.0912) |
| log(Tax level) | [1.0986, 5.9269] | 3.5708 | -0.0087\*\* (0.0042) |
| log(Income) | [6.9078, 9.6158] | 7.7707 | 0.1252\*\*\* (0.0102) |
| log(Income) – missing response | {0,1} | 0.0129 | 0.0708 (0.0602) |
| Visited Mariana | {0,1} | 0.0957 | 0.0213 (0.0219) |
| Concerned about environment | {1,2,…,5} | 4.3875 | 0.0091 (0.0076) |
| Trust the scientists’ assessment of the damage | {1,2,…,5} | 3.2353 | 0.0089 (0.0064) |
| Believe recovery time | {1,2,…,5} | 3.0598 | -0.0140\*\* (0.0064) |
| Believe the implementation of dry tailing is important for stopping future dam breaks | {1,2,…,5} | 3.6622 | 0.0234\*\*\* (0.0062) |
| Implementation of the dry tailing in Brazil with the financial help of the government possible in up to two years | {1,2,…,5} | 2.4534 | -0.0482\*\*\* (0.0066) |
| The government will implement dry tailings if the majority of the Brazilian population vote ‘for’ | {1,2,…,5} | 2.6385 | -0.0644\*\*\* (0.0067) |
| The government will actually raise the tax if dry tailings technology is deployed | {1,2,…,5} | 3.2498 | -0.0026 (0.0055) |
| With the financial help of the government mining companies will implement dry tailing technology | {1,2,…,5} | 2.6178 | -0.0268\*\*\* (0.0066) |
| Specified tax amount would be lower, the same, or higher | {-1,0,1} | 0.3689 | 0.0633\*\*\* (0.0092) |
| Payment of the specified tax would be difficult for household | {1,2,3,4} | 3.1742 | 0.0425\*\*\* (0.0065) |
| Concerned about environment – missing response | {0,1} | 0.0010 | -0.2737\*\*\* (0.0894) |
| Trust the scientists’ assessment of the damage – missing response | {0,1} | 0.0035 | -0.0206 (0.1083) |
| Believe recovery time – missing response | {0,1} | 0.0021 | 0.1743 (0.1089) |
| Believe the implementation of dry tailing is important for stopping future dam breaks – missing response | {0,1} | 0.0033 | -0.0677 (0.1129) |
| Implementation of the dry tailing in Brazil with the financial help of the government possible in up to two years – missing response | {0,1} | 0.0023 | -0.0022 (0.1431) |
| The government will implement dry tailings if the majority of the Brazilian population vote ‘for’ – missing response | {0,1} | 0.0027 | 0.0343 (0.1548) |
| The government will actually raise the tax if dry tailings technology is deployed – missing response | {0,1} | 0.0042 | 0.1445 (0.1036) |
| With the financial help of the government mining companies will implement dry tailing technology – missing response | {0,1} | 0.0027 | -0.1289 (0.1144) |
| Specified tax amount would be lower, the same, or higher – missing response | {0,1} | 0.0373 | -0.0494 (0.0368) |
| Payment of the specified tax would be difficult for household – missing response | {0,1} | 0.0031 | 0.1591 (0.1306) |
| R2 (adj) |  |  | 0.1479 |
| *k* (parameters) |  |  | 25 |
| *n* (observations) |  |  | 5195 |

Notes: heteroskedascity-robust standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

Finally, Table 25 presents the results of the model in which protest (‘against’) votes are explained using survey quality indicators. We found that respondents who incorrectly answered to more ‘pause’ questions, and those who said they felt pressured to vote ‘for’ or ‘against’ the program were more likely to provide protest (‘against’) responses. At the same time, respondents who participated in intercept surveys were found to be less likely to protest, although the statistical relationship in this case is above the usual 0.05 threshold.

Table 25: Explanatory variables of the protest ‘no’ votes – survey quality indicators

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Range of values** | **Mean value** | **Coefficient (st. err.)** |
| Constant | - | - | -0.5967\*\*\* (0.0911) |
| log(Tax level) | [1.0986, 5.9269] | 3.5708 | 0.0042 (0.0040) |
| log(Income) | [6.9078, 9.6158] | 7.7707 | 0.1137\*\*\* (0.0106) |
| log(Income) – missing response | [0,1] | 0.0129 | 0.1237\*\* (0.0629) |
| Number of incorrect ‘pause’ questions | {1,2,…,6} | 0.0362 | 0.0557\*\*\* (0.0212) |
| Respondent felt pressured to vote ‘for’ the program | {0,1} | 0.0610 | 0.1659\*\*\* (0.0278) |
| Respondent felt pressured to vote ‘against’ the program | {0,1} | 0.0364 | 0.1137\*\*\* (0.0361) |
| Respondent could not respond if felt free of pressure to vote ‘for’ or against the program | {0,1} | 0.0050 | -0.0159 (0.0951) |
| Respondent’s involvement as perceived by interviewer | {1,2,…,5} | 4.6135 | -0.0046 (0.0172) |
| Respondent’s understanding as perceived by interviewer | {1,2,…,5} | 4.5879 | 0.0253 (0.0171) |
| Survey taken sitting | {1,2,…,5} | 3.5392 | -0.0012 (0.0037) |
| Intercept survey | {0,1} | 0.0191 | -0.0843 (0.0513) |
| R2 (adj) |  |  | 0.0334 |
| *k* (parameters) |  |  | 12 |
| *n* (observations) |  |  | 5195 |

Notes: heteroskedascity-robust standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

# Appendix 4: Sensitivity of voting behavior to explanatory variables specified as binary probit

This appendix presents the alternative modeling approach for investigating sensitivity of voting behavior to explanatory variables using binary probity for robustness check presented of the ones using linear modeling discussed in Section 6.2, in Table 17 to Table 19 and Table 23 to Table 25, respectively. Results with binary model in the table set below show qualitatively equivalent findings of the linear one.

Table 26: The evaluation of the effects of tax amounts and income on the probability of observing a ‘no’ vote – binary probit

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Model 1** | **Model 2** | **Model 3** | **Model 4** | **Model 5** | **Model 6** | **Model 7** |
| Constant | -0.0298 (0.0430) | 1.5964\*\*\* (0.2280) | 1.4935\*\*\* (0.5705) | 1.0854\*\*\* (0.2523) | 1.1064\*\* (0.4875) | 1.3394\*\*\* (0.3100) | 1.9024\*\*\* (0.3277) |
| log(Tax amount) | 0.1984\*\*\* (0.0117) | 0.2024\*\*\* (0.0117) | 0.2022\*\*\* (0.0119) | 0.2065\*\*\* (0.0119) | 0.2076\*\*\* (0.0119) | 0.1168\*\*\* (0.0150) | 0.2169\*\*\* (0.0123) |
| log(Income) |  | -0.2114\*\*\* (0.0289) | -0.3178\*\*\* (0.0431) | -0.2861\*\*\* (0.0416) | -0.2886\*\*\* (0.0418) | -0.0465 (0.0348) | -0.1883\*\*\* (0.0337) |
| Socio-demographic controls | no | no | yes | no | no | no | no |
| Region fixed-effects | no | no | no | yes | no | no | no |
| City fixed-effects | no | no | no | no | yes | no | no |
| Debriefing questions | no | no | no | no | no | yes | no |
| Quality controls | no | no | no | no | no | no | yes |
| Log-likelihood | -2827.84 | -2793.92 | -2748.52 | -2737.03 | -2723.59 | -2085.37 | -2609.41 |
| AIC/*n* | 1.0894 | 1.0772 | 1.0670 | 1.0603 | 1.0628 | 0.8125 | 1.0473 |
| *k* (parameters) | 2 | 4 | 23 | 17 | 37 | 25 | 111 |
| *n* (observations) | 5195 | 5195 | 5195 | 5195 | 5195 | 5195 | 5195 |

Notes: standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

Table 27: The evaluation of the relationship between debriefing questions responses and the probability of observing an ‘against’ vote – binary probit

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Range of values** | **Mean value** | **Coefficient (st. err.)** |
| Constant | - | - | 1.3394\*\*\* (0.3100) |
| log(Tax amount) | [1.0986, 5.9269] | 3.5708 | 0.1168\*\*\* (0.0150) |
| log(Income) | [6.9078, 9.6158] | 7.7707 | -0.0465 (0.0348) |
| log(Income) – missing response | {0,1} | 0.0129 | 0.6400\*\* (0.2759) |
| Visited Mariana | {0,1} | 0.0957 | 0.0143 (0.0766) |
| Concerned about environment | {1,2,…,5} | 4.3875 | -0.0191 (0.0279) |
| Trust the scientists’ assessment of the damage | {1,2,…,5} | 3.2353 | -0.0235 (0.0223) |
| Believe recovery time | {1,2,…,5} | 3.0598 | -0.0413\* (0.0222) |
| Believe the implementation of dry tailing is important for stopping future dam breaks | {1,2,…,5} | 3.6622 | -0.0356 (0.0223) |
| Implementation of the dry tailing in Brazil with the financial help of the government possible in up to two years | {1,2,…,5} | 2.4534 | -0.1535\*\*\* (0.0216) |
| The government will implement dry tailings if the majority of the Brazilian population vote ‘for’ | {1,2,…,5} | 2.6385 | -0.1917\*\*\* (0.0219) |
| The government will actually raise the tax if dry tailings technology is deployed | {1,2,…,5} | 3.2498 | -0.0858\*\*\* (0.0192) |
| With the financial help of the government mining companies will implement dry tailing technology | {1,2,…,5} | 2.6178 | -0.1122\*\*\* (0.0219) |
| Specified tax amount would be lower, the same, or higher | {-1,0,1} | 0.3689 | 0.1205\*\*\* (0.0319) |
| Payment of the specified tax would be difficult for household | {1,2,3,4} | 3.1742 | 0.4051\*\*\* (0.0216) |
| Concerned about environment – missing response | {0,1} | 0.0010 | -0.8323 (0.9736) |
| Trust the scientists’ assessment of the damage – missing response | {0,1} | 0.0035 | 13.7317 (>1000) |
| Believe recovery time – missing response | {0,1} | 0.0021 | 0.0608 (0.8329) |
| Believe the implementation of dry tailing is important for stopping future dam breaks – missing response | {0,1} | 0.0033 | -0.2850 (0.5798) |
| Implementation of the dry tailing in Brazil with the financial help of the government possible in up to two years – missing response | {0,1} | 0.0023 | 11.1349 (>1000) |
| The government will implement dry tailings if the majority of the Brazilian population vote ‘for’ – missing response | {0,1} | 0.0027 | 11.2648 (>1000) |
| The government will actually raise the tax if dry tailings technology is deployed – missing response | {0,1} | 0.0042 | 12.7308 (>1000) |
| With the financial help of the government mining companies will implement dry tailing technology – missing response | {0,1} | 0.0027 | 1.1105 (0.8237) |
| Specified tax amount would be lower, the same, or higher – missing response | {0,1} | 0.0373 | 0.2327 (0.1516) |
| Payment of the specified tax would be difficult for household – missing response | {0,1} | 0.0031 | 0.4357 (0.6266) |
| Log-likelihood |  |  | -2085.37 |
| AIC/*n* |  |  | 0.8125 |
| *k* (parameters) |  |  | 25 |
| *n* (observations) |  |  | 5195 |

Notes: standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

Table 28: The evaluation of the relationship between survey quality indicators and the probability of observing a ‘no’ (‘against’) vote – binary probit

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Range of values** | **Mean value** | **Coefficient (st. err.)** |
| Constant | - | - | 1.8512\*\*\* (0.2671) |
| log(Tax amount) | [1.0986, 5.9269] | 3.5708 | 0.2044\*\*\* (0.0119) |
| log(Income) | [6.9078, 9.6158] | 7.7707 | -0.1953\*\*\* (0.0304) |
| log(Income) – missing response | [0,1] | 0.0129 | 0.7820\*\*\* (0.2310) |
| Number of incorrect ‘pause’ questions | {1,2,…,6} | 0.0362 | 0.2734\*\*\* (0.0977) |
| Respondent felt pressured to vote ‘for’ the program | {0,1} | 0.0610 | 0.5346\*\*\* (0.0923) |
| Respondent felt pressured to vote ‘against’ the program | {0,1} | 0.0364 | 0.9240\*\*\* (0.1562) |
| Respondent could not respond if felt free of pressure to vote ‘for’ or against the program | {0,1} | 0.0050 | 0.7189\* (0.3806) |
| Respondent’s involvement as perceived by interviewer | {1,2,…,5} | 4.6135 | -0.1014\*\* (0.0507) |
| Respondent’s understanding as perceived by interviewer | {1,2,…,5} | 4.5879 | 0.0029 (0.0513) |
| Survey taken sitting | {1,2,…,5} | 3.5392 | 0.0029 (0.0106) |
| Intercept survey | {0,1} | 0.0191 | 0.0024 (0.1399) |
| Log-likelihood |  |  | -2739.85 |
| AIC/*n* |  |  | 1.0594 |
| *k* (parameters) |  |  | 12 |
| *n* (observations) |  |  | 5195 |

Notes: standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

Table 29: Explanatory variables of the protest ‘against’ votes – socio-demographics – binary probit

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Range of values** | **Mean value** | **Coefficient (st. err.)** |
| Constant | - | - | -2.2873\*\*\* (0.5665) |
| log(Tax amount) | [1.0986, 5.9269] | 3.5708 | 0.0069 (0.0107) |
| log(Income) | [6.9078, 9.6158] | 7.7707 | 0.0718\* (0.0395) |
| log(Income) – missing response | {0,1} | 0.0129 | 0.3315\*\* (0.1590) |
| Work status – home maker | {0,1} | 0.0470 | -0.1089 (0.0950) |
| Work status – retired | {0,1} | 0.1748 | -0.1364\* (0.0714) |
| Work status – unemployed without benefits | {0,1} | 0.0152 | -0.3095\*\* (0.1538) |
| Work status – unemployed with benefits | {0,1} | 0.0200 | -0.2720\*\* (0.1329) |
| Work status – autonomous/freelance | {0,1} | 0.2745 | 0.0239 (0.0501) |
| Work status – part-time freelance | {0,1} | 0.1309 | -0.1215\* (0.0646) |
| Work status – pensioner | {0,1} | 0.0510 | -0.2738\*\*\* (0.0975) |
| Work status – savings | {0,1} | 0.0167 | -0.2274 (0.1454) |
| Work status – family allowance | {0,1} | 0.0393 | -0.3705\*\*\* (0.1126) |
| Household size | {1,2,…,32} | 3.4296 | -0.0079 (0.0113) |
| Female | {0,1} | 0.5034 | -0.1491\*\*\* (0.0386) |
| Age | [18,97] | 47.3936 | 0.0302\*\*\* (0.0066) |
| Age2 | [3.2400, 94.0900] | 25.0756 | -0.0297\*\*\* (0.0070) |
| log(Household wealth) | [1.6094, 4.3041] | 3.1530 | 0.2096\*\*\* (0.0697) |
| Education – primary (part) | {0,1} | 0.0589 | 0.0962 (0.4886) |
| Education – primary | {0,1} | 0.2062 | 0.0272 (0.4844) |
| Education – secondary | {0,1} | 0.1680 | 0.2654 (0.4849) |
| Education – high school | {0,1} | 0.4071 | 0.4091 (0.4841) |
| Education – undergraduate | {0,1} | 0.1305 | 0.4005 (0.4870) |
| Education – post graduate | {0,1} | 0.0277 | 0.4465 (0.4969) |
| Log-likelihood |  |  | -3350.20 |
| AIC/*n* |  |  | 1.2990 |
| *k* (parameters) |  |  | 24 |
| *n* (observations) |  |  | 5195 |

Notes: standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

Table 30: Explanatory variables of the protest ‘against’ votes – debriefing questions – binary probit

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Range of values** | **Mean value** | **Coefficient (st. err.)** |
| Constant | - | - | -2.7586\*\*\* (0.2682) |
| log(Tax level) | [1.0986, 5.9269] | 3.5708 | -0.0263\*\* (0.0124) |
| log(Income) | [6.9078, 9.6158] | 7.7707 | 0.3731\*\*\* (0.0305) |
| log(Income) – missing response | {0,1} | 0.0129 | 0.1746 (0.1637) |
| Visited Mariana | {0,1} | 0.0957 | 0.0660 (0.0631) |
| Concerned about environment | {1,2,…,5} | 4.3875 | 0.0241 (0.0220) |
| Trust the scientists’ assessment of the damage | {1,2,…,5} | 3.2353 | 0.0234 (0.0187) |
| Believe recovery time | {1,2,…,5} | 3.0598 | -0.0427\*\* (0.0185) |
| Believe the implementation of dry tailing is important for stopping future dam breaks | {1,2,…,5} | 3.6622 | 0.0708\*\*\* (0.0175) |
| Implementation of the dry tailing in Brazil with the financial help of the government possible in up to two years | {1,2,…,5} | 2.4534 | -0.1476\*\*\* (0.0192) |
| The government will implement dry tailings if the majority of the Brazilian population vote ‘for’ | {1,2,…,5} | 2.6385 | -0.1928\*\*\* (0.0195) |
| The government will actually raise the tax if dry tailings technology is deployed | {1,2,…,5} | 3.2498 | -0.0117 (0.0153) |
| With the financial help of the government mining companies will implement dry tailing technology | {1,2,…,5} | 2.6178 | -0.0806\*\*\* (0.0193) |
| Specified tax amount would be lower, the same, or higher | {-1,0,1} | 0.3689 | 0.1874\*\*\* (0.0269) |
| Payment of the specified tax would be difficult for household | {1,2,3,4} | 3.1742 | 0.1294\*\*\* (0.0199) |
| Concerned about environment – missing response | {0,1} | 0.0010 | -16.2666 (>1000) |
| Trust the scientists’ assessment of the damage – missing response | {0,1} | 0.0035 | -0.1217 (0.3446) |
| Believe recovery time – missing response | {0,1} | 0.0021 | 0.5031 (0.4348) |
| Believe the implementation of dry tailing is important for stopping future dam breaks – missing response | {0,1} | 0.0033 | -0.1311 (0.3718) |
| Implementation of the dry tailing in Brazil with the financial help of the government possible in up to two years – missing response | {0,1} | 0.0023 | 0.0921 (0.4568) |
| The government will implement dry tailings if the majority of the Brazilian population vote ‘for’ – missing response | {0,1} | 0.0027 | 0.1249 (0.4424) |
| The government will actually raise the tax if dry tailings technology is deployed – missing response | {0,1} | 0.0042 | 0.4303 (0.3174) |
| With the financial help of the government mining companies will implement dry tailing technology – missing response | {0,1} | 0.0027 | -0.4694 (0.4947) |
| Specified tax amount would be lower, the same, or higher – missing response | {0,1} | 0.0373 | -0.1390 (0.1027) |
| Payment of the specified tax would be difficult for household – missing response | {0,1} | 0.0031 | 0.4823 (0.3753) |
| Log-likelihood |  |  | -3077.13 |
| AIC/*n* |  |  | 1.1943 |
| *k* (parameters) |  |  | 25 |
| *n* (observations) |  |  | 5195 |

Notes: standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

Table 31: Explanatory variables of the protest ‘no’ votes – survey quality indicators – binary probit

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Range of values** | **Mean value** | **Coefficient (st. err.)** |
| Constant | - | - | -2.8696\*\*\* (0.2459) |
| log(Tax level) | [1.0986, 5.9269] | 3.5708 | 0.0107 (0.0106) |
| log(Income) | [6.9078, 9.6158] | 7.7707 | 0.2971\*\*\* (0.0282) |
| log(Income) – missing response | [0,1] | 0.0129 | 0.3130\*\* (0.1567) |
| Number of incorrect ‘pause’ questions | {1,2,…,6} | 0.0362 | 0.1495\*\* (0.0632) |
| Respondent felt pressured to vote ‘for’ the program | {0,1} | 0.0610 | 0.4308\*\*\* (0.0741) |
| Respondent felt pressured to vote ‘against’ the program | {0,1} | 0.0364 | 0.2963\*\*\* (0.0944) |
| Respondent could not respond if felt free of pressure to vote ‘for’ or against the program | {0,1} | 0.0050 | -0.0410 (0.2581) |
| Respondent’s involvement as perceived by interviewer | {1,2,…,5} | 4.6135 | -0.0117 (0.0453) |
| Respondent’s understanding as perceived by interviewer | {1,2,…,5} | 4.5879 | 0.0675 (0.0457) |
| Survey taken sitting | {1,2,…,5} | 3.5392 | -0.0034 (0.0097) |
| Intercept survey | {0,1} | 0.0191 | -0.2207\* (0.1323) |
| Log-likelihood |  |  | -3420.50 |
| AIC/*n* |  |  | 1.3215 |
| *k* (parameters) |  |  | 12 |
| *n* (observations) |  |  | 5195 |

Notes: standard errors given in parentheses; \*, \*\*, \*\*\* represent statistical significance at 0.1, 0.05, 0.01 level, respectively.

# References

Angrist, J. D., and Pischke, J.-S., 2010. The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24(2):3-30.

Arrow, K. et al, Report of the NOAA panel on Contingent Valuation, 1992, Federal register 58, National Oceanic and Atmospheric Administration.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E., 1955. An Empirical Distribution Function for Sampling with Incomplete Information. *Ann. Math. Statist.*, 26(4):641-647.

Bishop, R. C., Boyle, K. J., Carson, R. T., Chapman, D., Hanemann, W. M., Kanninen, B., Kopp, R. J., Krosnick, J. A., List, J., Meade, N., Paterson, R., Presser, S., Smith, V. K., Tourangeau, R., Welsh, M., Wooldridge, J. M., DeBell, M., Donovan, C., Konopka, M., and Scherer, N., 2017. Putting a value on injuries to natural assets: The BP oil spill. *Science*, 356(6335):253-254.

Carson, R., and Louviere, J., 2011. A Common Nomenclature for Stated Preference Elicitation Approaches. *Environmental and Resource Economics*, 49(4):539-559.

Carson, R. T., and Czajkowski, M., 2014. The Discrete Choice Experiment Approach to Environmental Contingent Valuation. In: *Handbook of choice modelling*, S. Hess and A. Daly, eds., Edward Elgar, Northampton, MA.

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

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

Cochran, W. G., 1977. Sampling Techniques. 3 Ed., John Wiley & Sons.

Dillman, D. A., Smyth, J. D., and Christian, L. M., 2014. Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method. 4 Ed., Wiley.

Flores, N. E., and Carson, R. T., 1997. The Relationship between the Income Elasticities of Demand and Willingness to Pay. *Journal of Environmental Economics and Management*, 33(3):287-295.

Greene, W. H., 2011. Econometric Analysis. 7 Ed., Prentice Hall, Upper Saddle River, NJ.

Gurmu, S., and Trivedi, P. K., 1996. Excess Zeros in Count Models for Recreational Trips. *Journal of Business & Economic Statistics*, 14(4):469-477.

Haab, T., and McConnell, K., 2003. The Econometrics Of Non-Market Valuation. Edward Elgar, Northampton, MA.

Krinsky, I., and Robb, A. L., 1986. On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, 68(4):715-719.

Krinsky, I., and Robb, A. L., 1991. Three methods for calculating the statistical properties of elasticities: A comparison. *Empirical Economics*, 16(2):199-209.

Kriström, B., 1997. Spike Models in Contingent Valuation. *American Journal of Agricultural Economics*, 79(3):1013-1023.

Lewbel, A., 2000. Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, 97(1):145-177.

MacKinnon, J. G., and White, H., 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305-325.

Parsons, G., and Myers, K., 2017. Fat tails and truncated bids in contingent valuation: an application to an endangered shorebird species. In: *Contingent valuation of environmental goods. A Comprehensive Critique*, D. McFadden and K. Train, eds., Edward Elgar, Northampton, USA.

Renova*, In the Repair Path*, Fundação Renova, Base year: 2018, Belo Horizonte, 2019

Turnbull, B. W., 1976. The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(3):290-295.

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

Watanabe, M., 2010. Nonparametric Estimation of Mean Willingness to Pay from Discrete Response Valuation Data. *American Journal of Agricultural Economics*, 92(4):1114-1135.

Wooldridge, J. M., 2010. Econometric Analysis of Cross Section and Panel Data. 2 Ed., The MIT Press.

Zawojska, E., Bartczak, A., and Czajkowski, M., 2019. Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences. *Journal of Environmental Economics and Management*, 93:63-84.

1. The wording “incident” instead of “accident” or “disaster” was used throughout the survey, in line with the conservative approach followed by the study team to prevent any perception that might anticipate the injury levels before their detailed presentation. [↑](#footnote-ref-1)
2. Coordinated by Luís César Périssé with the assistance of Francisco Teizen. [↑](#footnote-ref-2)
3. The final version of the questionnaire, including presentation cards, is provided in two annexed files. [↑](#footnote-ref-3)
4. Revealed preference methods are based on observations of human behavior to make inferences about value. For example, the value of lost recreation at beaches affected by an oil spill can be estimated by observing how beach visitation patterns and travel costs changed as a result of beach oiling. [↑](#footnote-ref-4)
5. Figure one is just an **example** of a generic CE question that might be used. [↑](#footnote-ref-5)
6. See Arrow et al. (1992). [↑](#footnote-ref-6)
7. See Section 6.1 for sampling and Appendix 1 on interview protocols for documentation on how this issue was handled. [↑](#footnote-ref-7)
8. We use a finite set of tax amounts and observe no monotonicity violations for the increased probability of voting ‘no’ with increasing tax amounts. [↑](#footnote-ref-8)
9. In the case of weighting the sample observations, instead of the number of votes/individuals the sum of individual weights can be used. [↑](#footnote-ref-9)
10. In the case of a binary choice, if an individual voted ‘yes’ for a program at a particular tax amount – the lower bound of his or her true WTP is the tax amount, and the upper bound is unknown, limited by individual income. A ‘no’ vote reveals that this respondent’s WTP is between zero (lower bound) and the presented tax amount (upper bound). The remainder of this subtraction is the probability that respondents’ WTP lies between the lower and the upper bound. [↑](#footnote-ref-10)
11. We operationalize the numerical estimation by using the value of the PDF evaluated at the lower bound instead of the difference of CDFs, whenever the calculated difference in the values of the CDF is numerically equal to zero. [↑](#footnote-ref-11)
12. For the distributions with support in negative numbers, censoring negative values to zero has been used when simulating mean WTP in the sample. [↑](#footnote-ref-12)
13. This is particularly expected for payment card and open-ended elicitation formats ([Carson and Czajkowski, 2014](#_heading=h.3s49zyc)). [↑](#footnote-ref-13)
14.  is usually assumed to be a CDF of normal or logistic distribution, leading to using the probit or logit model to account for zero-inflation. [↑](#footnote-ref-14)
15. To approximate asymptotic variance covariance matrix. [↑](#footnote-ref-15)
16. Maximum likelihood estimates are asymptotically normal. [↑](#footnote-ref-16)
17. The prevention of a future spill/incident/contamination event is often used as a proxy for the spill/incident/contamination event being evaluated. This approach was followed in the evaluation of public losses arising from the Deepwater Horizon oil spill in the Gulf of Mexico. [↑](#footnote-ref-17)
18. As will be seen in the next section, respondents were told that because of recent dam failures, steps had been taken that would minimize loss of life, damage to residences. [↑](#footnote-ref-18)
19. The final version of the applied questionnaire and respective cards are available in two annexed files. [↑](#footnote-ref-19)
20. It was described in this way to avoid inflaming negative feelings about the company operating the mine by not using the words Dam failure. Data on expenditures and compensations came from Renova (2019) [↑](#footnote-ref-20)
21. Measurement of recreational losses is often conducted through a careful analysis of recreational trips to the affected area. This analysis requires detailed counts of where recreational trips occur and information about how far each individual traveled. Data of this type was not readily available and would have been quite expensive to obtain. [↑](#footnote-ref-21)
22. The probability was calculated based on the average standard errors of the estimated probabilities of non-parametric approach for the available data that were extrapolated for the case the number of observations expected for the final study. [↑](#footnote-ref-22)
23. It is theoretically possible that the limited number of observations and the randomness of the sampling process may (with some non-zero probability) lead to observing that the acceptance rate of respondents who were faced with a higher tax amount is higher than the acceptance rate of respondents who were faced with a lower tax amount. While this is statistically possible, this has caused some concern in applied studies – particularly in ones that did not take the uncertainty associated with the observed results into account. This literature suggests that in the case where monotonicity violations are observed a more conservative non-parametric ([Turnbull, 1976](#_heading=h.2ce457m)) estimator should be used, which requires that adjacent tax amounts and the observed acceptance rates are merged until no monotonicity violations are observed. This would lead to a large loss of efficiency (in terms of both – bias and standard errors). [↑](#footnote-ref-23)
24. No substantial differences were found between the results based on all data and pilot 3 data only. [↑](#footnote-ref-24)
25. Additionally, we analyzed the data half-way through the collection in each pilot, to check if there were any immediate changes that would have to be made to the survey. [↑](#footnote-ref-25)
26. The results of the regression used for imputing missing income responses are presented in Appendix 2. [↑](#footnote-ref-26)
27. Prof. Waldir Lobão assisted with sample design and creation of sample weights. [↑](#footnote-ref-27)
28. The sample size determination was undertaken in a generic form for one single proportion (and tax value), which approximates the sample size determination for each of the five proportions (and tax values) used in the study. [↑](#footnote-ref-28)
29. Verification procedures were designed and tested during Pilot tests 1, 2 and 3. Most of the Sao Paulo team gained extensive experience with the verification procedures through participation in pilot tests 1, 2 and 3 [↑](#footnote-ref-29)
30. The average interview time was 41 minutes and 15 seconds (41’15”), with small differences among cities: São Paulo = 42’30”; Rio de Janeiro = 38’54”; Belo Horizonte = 41’10”; Mariana = 42’6”; Ipatinga = 45’36”; Linhares = 39’36”; Curitiba = 39’30”; Goiania = 39’18” ; Recife = 42’36” and Belem = 43’24”. [↑](#footnote-ref-30)
31. Zero inflation component was an add-on to the assumed standard distribution, accounting for a higher number of zero responses; see section 3.3 for details. [↑](#footnote-ref-31)
32. The assumptions behind our conservative (Lewbel-Watanabe-approach-based) estimate make model-driven overestimation impossible. On the other hand, this approach results in an estimate that is likely lower than the true extent of the damages and therefore can be considered a lower-bound estimate. [↑](#footnote-ref-32)
33. While the parametric approach has the advantage of minimizing the error of the estimate (relative to the true, unknown mean WTP), it is possible it underestimates or overestimates the true value. [↑](#footnote-ref-33)
34. As an aside, we note that the non-parametric analysis of unweighted data implies that the lower bound estimate of monthly WTP is 55.56 BRL with a 95% confidence interval of 50.01 - 61.12 BRL – a value that is very close (not statistically significantly different) to the estimate based on weighted data. [↑](#footnote-ref-34)
35. The logarithmic transformation resulted in a better fit than linear and several other functional forms tested. [↑](#footnote-ref-35)
36. In model 6, which includes responses to debriefing questions as explanatory variables of ‘against’ votes, the income was not statistically significant. This is likely a result of a high correlation between explanatory variables, and relatively high explanatory power of debriefing questions alone, as indicated by the adjusted r2 coefficient for this model, which was approximately three times higher than for any other model. [↑](#footnote-ref-36)
37. The share of respondents who said they felt free to vote was 0.8976 (4663 of 5195 respondents). [↑](#footnote-ref-37)
38. In line with the analysis presented in Section 6, all regression models control for tax amounts presented to respondents and their income levels. [↑](#footnote-ref-38)