Bayesian modeling for pro-environmental behavior data: sorting and selecting relevant variables

Paula Reveco-Quiroz1,2 · José Sandoval-Díaz3 · Danilo Alvares1,2

Accepted: 18 April 2022 / Published online: 18 May 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Pro-environmental behaviors towards climate change can be measured and evaluated in different fields. Typically, surveys are the standard tool for extracting personal information regarding this phenomenon. However, statistical modeling for these surveys is not straightforward, as the response variable is often not explicit. Hence, we propose a set of methodological procedures to deal with pro-environmental behavior data. First, validity evidence through a factorial analysis. Second, indexes are created from factor scores, where one of the latent factors summarizes a target variable. Third, a Beta regression is used to model the index of interest. Fourth, the inferential process is performed from a Bayesian perspective, in which posterior probabilities are used to sort and select the relevant variables. Finally, suitable models are obtained, and conclusions can be drawn from them. As a motivation, we used data from two Chilean surveys to illustrate our methodology as well as interpret and discuss the results.

Keywords Bayesian inference · Beta regression · Climate change · Factor analysis · Indexes · Likert scale

1 Introduction
Climate change is one of the main problems the world is facing (Gligo 2020), with irrefutable evidence that the main cause is human behavior (IPCC 2021). This implies that through various human activities, there is an emission of greenhouse gases (e.g., methane and CO2) that cause an increase in global surface temperature (Gallardo 2021; Rojas and Gallardo 2021) which it has been observed more conspicuously from 1900 to date (IPCC 2021).

This phenomenon has generated a state of climatic emergency, the scope of which can be summarized in the physical and social dimensions.

From the first dimension of the problem, changes have been observed in climatic patterns and physical phenomena, where the rise in sea level (Priestley et al. 2021), droughts (Pokhrel et al. 2021), reduced access to water (Padron et al. 2020) and natural disasters (Cappelli et al. 2021), are presented as the main challenges to face (Gallardo 2021; Rojas and Gallardo 2021).

From the second dimension of the problem, influences have been observed in processes of psychosocial conflict (Suh et al. 2021), the need for new forms of governance (Bremer et al. 2021), a new economic coexistence (Mele et al. 2021) and the need to introduce new consumer practices (Gallardo 2021).

In this context, there is scientific agreement that the cause and solution of this problem lies in human behavior, however, one aspect where there is divergence in this regard is in the political will to combine criteria for confronting the problem in conjunction with the interests of the different governments and their territories (Gallardo 2021).
However, where there are no doubts in this regard is the need for a rapid and sustained confrontation over time, the generation of adaptation plans (Gligo 2020), mitigation (Masson and Fritsche 2021), citizen participation (Mees et al. 2019) and the implementation of transitions to non-conventional energies or renewable (Gallardo 2021). Also, the incorporation of local approaches that allow responding to the idiosyncratic characteristics of the territories and thus be able to orient a sustainable development (Gligo 2020) that allows to re-think the way we live (Gallardo 2021).

In this scenario, research and scientific advances are essential to contribute to the existing knowledge on the subject, however, it is possible to detect some gaps in this regard. For example, most climate change studies are based on developed countries (van der Linden 2017) while other regions have recently been interested in this type of study (Hardoy and Romero-Lankao 2011), such is the case of Latin America. For its part, the methodological focus has initially been on qualitative analysis techniques (Sapiains and Ugarte 2017b; Libarkin et al. 2018; Bostrom et al. 1994; Andersson and Wallin 2000; Gautier et al. 2006). Thus, when describing the qualitative advances in Latin America and the Caribbean, Forero et al. (2014) indicate that between 1997 and 2012, 92.3% of the documents used qualitative methodological strategies, based mainly on field observation and semi-structured interviews.

Regarding the attempts to generate quantitative assessment instruments, these have focused on multiple-choice, true-false, or Likert-type questions (Boyes and Stanisstreet 1997; Keller 2006; Lambert et al. 2012).

A topic of study with investigative potential is the sociocultural contributions and perceptions of climate patterns (Adger 2001; van der Linden 2017) then, motivated by these gaps, in this study analyzes a pair of surveys that contain a perceptual instrument on climate change, applied in the Southern region of Chile.

Chile complying with seven of the nine vulnerability criteria for climate change (IPCC 2014; Chilean Government 2017): arid and semi-arid areas forest areas, low elevation coastal areas, areas prone to drought and desertification, mountain ecosystems such as the Coast and Andes Mountain ranges, urban areas with air pollution problems, and high susceptibility to extreme hazards. Therefore, according to the (ECLAC 2012), Chile’s challenge is to face vulnerabilities and opportunities for adaptation and mitigation to climate change, for which local studies are an imperative.

The applied climate change questionnaire was proposed by Spanish authors who compiled items previously tested by other researchers and research centers and also generated their own items, reaching a total of 19 items (Hidalgo and Pisano 2010). When using this questionnaire, some questions of interest arise, such as “what is the factorial structure of the Spanish scale when evaluated with Chilean observational data?”, “do some items in the questionnaire summarize pro-environmental behavior?”, and “can pro-environmental behavior be explained by other variables?”. Hence, we propose a sequence of statistical procedures that answer these questions. Part of our proposal includes statistical regression modeling, in which recent studies have addressed environmental issues from another methodological perspective, using the technique of artificial neural networks for predictions in protected areas (Saffariha et al. 2020), evaluation of the effect of human activities on vegetation (Jahani and Saffariha 2021), prediction of ecological varieties conditional on habitat (Saffariha et al. 2021), influence of factors on the perception of the landscape (Jahani et al. 2022), and uncertainty quantification for water resources applications (Ciriello et al. 2021). These methodologies could be incorporated into our procedural proposal without major difficulties.

The remainder of the paper is organized as follows. Section 2 introduces the origin of the data, the sampling design, the applied instruments and a descriptive analysis. Section 3 discusses theoretical aspects of axiomatic measures, validity evidence, and constructs. Section 4 presents the methodological framework of our proposal. Section 5 applies our methodology to Chilean data. Section 6 discusses the findings by applying our set of statistical procedures. Finally, Section 7 summarizes the methodology presented in this paper and points out some possible further research directions.

2 Materials

2.1 Regional context

The local reality treated in this paper corresponds to that of the Ñuble Region inhabitants, in Southern region of Chile (see Fig. 1). In administrative terms, Ñuble is located in the south-central zone of Chile, whose regional capital is Chillán. According to data from the last population census (INE 2017) the region has a total area of 13178 km² (with only 138 km² of urban area) and a population of 480609 inhabitants, corresponding to 2.7% of the Chilean population. In sociodemographic terms, the average age of the population is 37.6 years, with a slightly higher proportion of women (248022) than men (232587) and with an average formal schooling of 9.4 years, by the heads of households. Regarding residential location, 69.4% live in urban areas and 30.6% in rural areas, thus being the region with the largest rural population nationwide. In terms of people living in poverty and multidimensional poverty, the region presents the second rate at a national level (16.1%), doubling the national average (8.6%) (Chilean Government 2018).
On the other hand, according to the Atlas of Climate Risks for Chile (Chilean Government 2021), Chilla´n presents a strong increase not only in the risk index for urban water security at a domestic level (as a consequence of the meteorological drought) but also an increase in health risks due to the effects of urban heat waves, which are relevant characteristics for the case study.

In the same way, one aspect to emphasize is the imminent decentralization process that has been under discussion in Chile for some years and is currently declared by Chilean Government (2021) as part of the new political and economic structure of the country, in which the productive units are independent and competitive, taking full advantage of the benefits offered by a market system, which are extended to social, cultural and welfare development.

2.2 Sampling design

The relevance of a longitudinal trend study for the Chilean case, period in which the data were collected, is that between one measurement and the other, two fundamental events occurred in the social context of the country, such as the “social outbreak” of 2019, which implied a series of social mobilizations at a national level due to generalized discontent, as well as the Covid-19 global pandemic that has affected all spheres of society, for which there is the possibility of analyzing possible comparative changes in the perceptions of the participants between both moments of application.

It was carried out using a questionnaire survey applied in two moments, The first was conducted from July to August 2019, reaching a total of \( n = 266 \) final participants and the second was conducted from December 2020 and January 2021 reaching a total of \( n = 346 \) final participants.

The participants were different between one application and the other (moment 1 and moment 2) however, their heterogeneity will be evaluated and described later in the paper, in order to have evidence about whether both samples behave in a similar way to the instrument and thus resolve the doubt about whether they can join for the next steps of the methodology to be presented.

The conformation of the sample followed two inclusion criteria: (i) to be an adult of eighteen years and over, and (ii) residents of the Nuable Region of Chile.

Another aspect to be mentioned was that the first moment follow the form of a household survey with a pollster, corresponding to a university student formally trained for this occasion, considering the following aspects: clarification of objectives, voluntariness, confidentiality, application time, and clarification of emerging questions.

Meanwhile, the second moment follows the form of an online survey using Google forms, due to the adjustment of the precautionary measures taken in the Covid-19 pandemic period.

An aspect to consider was that, before filling out the instrument, the participants were informed about the research and its objectives through an informed consent, which they read and accepted before starting. Finally, the average time for both applications was 20 minutes per respondent.

2.3 Instruments

The real survey or test is seen as being composed of several instruments (Wilson 2005). In the case of this research, its surveys were composed by three types of different instruments: a behavior scale in the face of climate change (Likert type), a sociodemographic questionnaire and two dichotomous questions about climate change assumptions.

Behavior scale in the face of climate change (Hidalgo and Pisano 2010) made of the following nineteen items: (i) five items on the knowledge of the climate change causes (CC) (anthropogenic, natural or greenhouse gas causes); (ii) three items on the attitude towards CC (current concern about the current CC, sanctions for non-compliance with positioning towards environmental risks); (iii) two self-efficacy items; (iv) eight risk perception items (occurrence and consequences of CC) and (v) one item for behavioral intention (changes in lifestyle).

In the original version, for each item, the scores were given on a seven point Likert-type item with responses ranging from strongly disagree = 1 to strongly agree = 7, modifying the present version on a five point Likert-type item, according to psychometric guidelines by Crocker and Algina (2008) which reinforces the widely used format for inventories that was suggested by Likert (1932) where respondents read each statement and select a response from...
strongly disagree = 1 to strongly agree = 5. In the Chilean version, two items were also paraphrased with respect to the Spanish version. In the first item (V9), the reference to the Kyoto Protocol was omitted because it was considered a very specific treaty (not the only one) and the second item (V14) was modified because the question changes the sense of the response format. Furthermore, the order in which these items were administered also changed following a correlative order with which each item is numbered (V+number) in Table 6 of the Appendix A.

The sociodemographic questionnaire consists in variables associated with the measured sociodemographic characteristics and conditions such as: sex, age, territorial location, political tendency, income, and educational level.

In order to use a standard format for the mentioned measured variables, we used the questions (and response options) based on CASEN Post Earthquake Panel Survey (Chilean Government 2010).

The reason for including demographic variables and others related to climate change issues in the analysis of the Likert type scale, which is the central instrument of the analysis, is justified because exists theoretical background regarding variables that would be linked to environmental trends in people. For example, there is evidence that those individuals who have a higher education, that are older or are women exhibit higher pro-environmental behavior, between other characteristics of interests (Blankenberg and Alhusen 2019).

However, and according to Beiser-McGrath and Huber (2018), these sociodemographic characteristics would be insufficient to predict climatic/environmental attitudes, to which their contextual variability would be added, as dimensions permeable to particular culture and geography.

The dichotomous questions consists in two statements that are related to the climate change and where their two possible answers are yes or no. First was: Do you know what climate change is? and the content of this knowledge was verified through open responses that complemented it in front of the instruction “Briefly define what you mean by climate change”. Second was: Is it possible to reduce the impact of climate change? which also was complemented with an open answer when consulting them, how?

In relation with these questions, some cognitive factors are relevant as facilitators and/or barriers for pro-environmental behavior, such as explicit knowledge about climate change and self-perceived ability to reduce it (Grothmann and Patt 2005; Corral-Verdugo 2021).

### 2.4 Descriptive analysis

From the surveys applied in two moments of time (2019 and 2020), two different databases were obtained, which were subjected to an initial process of data cleaning and variable debugging. As a result of the data cleaning, one observation from the ‘Survey 1’ database and 24 observations from the ‘Survey 2’ database were eliminated, and as a result of the variable cleaning, a total of 27 variables that could be included in the analysis were obtained, unifying their categories at a binary, categorical and continuous level, as appropriate.

Of the 27 variables of interest for the analysis, 19 correspond to variables of the perceptual scale on climate change. Then, six demographic variables were also included, which were: sex of the interviewee (sex), educational level (education), income of the participants (income), territorial location (location), political tendency (poltrend) and age of the participants (age). Finally, two variables were included, one on knowledge about climate change (know) and another on the possibility of reducing the impact of climate change (impreduc). The last eight variables are described in Table 1.

Considering a significance level of 5%, the variables sex, impreduc, location, and age are significantly different
comparing the two surveys. The sensitivity of these differences will be analyzed in the validation step (see Section 4).

3 Theoretical framework

3.1 Axiomatic measurement theory

Through the Axiomatic or Abstract Measurement Theory (AMT), Torres Irribarra (2021) explains a form to mathematically measure a representation based on Scott and Suppes (1958); Luce and Tukey (1964); Krantz et al. (2007) whose support is the existence of a relational structure based on axioms.

Torres Irribarra (2021) describes four necessary conditions for the measurement based in Narens and Luce (1986): (1) an ordered relational structure, (2) a set of axioms that describe the empirical structure (empirical laws), (3) a number-based relational structure, and (4) a proof of the existence of a structure that preserves the mapping, which can be an empirical or qualitative structure, with a representative structure and a mapping that preserves the structure as collection of all homomorphisms in the same representative structure scale.

Definitely, AMT provides a measurement framework that allows us to identify sets of ordered variables and their relationships to the abstract measurement in Psychology (Cliff 1992). In addition, Cliff (1992) highlights the little use given to this theory based on the hypothesis of the necessary mathematical background that it implies and empirically, on the other hand, “the problem of error” linked as an example to the analysis structure of covariance and latent variable models in general, which do not see randomness as a problem, but rather depend on it (Torres Irribarra 2021).

3.2 Validity evidence to interpret the scale scores

The most influential position on validity today holds that: “validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (AERA 2014).

In other words, validity issues can be understood as the inferences made from the measurement scores towards the use of these scores to reflect the definition of the construct (Wu et al. 2016).

To internalize what a “construct” is, it is also worth referring to the notion of “latent trait”, being both terms used to refer to the psycho-social attributes (Wu et al. 2016). These attributes can be interpreted in multiple ways (Skrondal and Rabe-Hesketh 2004) and such as highlights Torres Irribarra (2021) from classic authors, the constructs refers to the attribute that the researcher is trying to measure (Messick 1989; Wilson 2005).

However, van der Linden (1994) goes one step further by holding that from modern measurement theory is possible to verify laws or models that explain observable data using only unmeasured variables and if these are quantitative and empirically verified, the latent variables have quantitative scales on which the positions of the objects are known. Furthermore, points out that a model that contains only latent variables and are measured jointly related to one another, implies that the measurement of them is not derived from other measured variables. At this point, the term “common factor model” gains relevance, which, when it was introduced by Spearman (1904) allowed to lay the foundations for structural analysis (Lazarsfeld and Henry 1968) which can apply to model attributes with different structures, including continuous variables (latent feature models), categorical (latent classes), and ordinal (ordered latent classes) (Torres Irribarra 2021).

The foundations of the multivariate statistical procedures with factor analysis were developed “to test hypotheses about the correspondence between the scores of the observed variables (manifest) or indicators and the hypothetical constructs (latent variables), or factors, presumed to affect those scores” (Kline 2013).

The assumptions of the factor analysis are (a) the common variance is due to the effects of underlying factors and (b) the number of factors of interest is less than the number of indicators, its final objective being to be able to identify and interpret a smaller number of factors that explain most of the common variance (Kline 2013).

3.3 Outcome space to reach constructs

The term outcome space was introduced by Marton (1981) to describe a set of outcome categories developed from a detailed (phenomenographic) analysis, relative to the set of qualitatively described categories to record and/or judge how respondents have responded to the items. One of the ways to respond to an item is through Likert-style survey questions, a format that implies a qualitative understanding of what constitutes different levels of response, being inherent to the idea of categorization (Wilson 2005).

Two aspects of relevance about an outcome space highlighted by Dahlgren (1984) and reviewed in Wilson (2005) are a “kind of analytic map” which includes an intensive examination of empirical data and its specific content linked with the set of descriptive categories, which has not been determined a priori, but depends on the specific content of the item, being essential for the categories of an outcome space to be well defined, that these
are finite and exhaustive, ordered, context-specific, and research-based (Masters and Wilson 1997).

To meet this objective, categories must include (a) a general definition of what is being measured by that item, (b) background material; (c) examples of items, item responses, and their categorization; as well as (d) a training procedure, with the objective to reach a construct map whose base are qualitative levels between the extremes, that lets us assume that the respondents can be at any point in between, and which finally implies that the construct is continuous (Wilson 2005).

4 Methodology

4.1 Validation

4.1.1 Factor analysis

The factor analysis (FA) corresponds to the first of two stages in the generation of validity evidence on the data of the climate change scale. This factor analysis procedure consisted of three well-defined steps. In the first step, a confirmatory factor analysis (CFA) is applied to evaluate the item grouping structure proposed by the Spanish scale (Hidalgo and Pisano 2010). In the second step, an exploratory factor analysis (EFA) is applied to know the structure of this scale with Chilean data. Finally, in the third step, a CFA is applied again, this time to evaluate the structure found in the EFA with Chilean data.

For the first CFA, both sample 1 and 2 are analyzed, separately. For the EFA only sample 1 is analyzed and for the second CFA only sample 2 is analyzed Brown (2015).

One of the methodological decisions made before implementation of the EFA was to use the Robust Diagonally Weighted Least Squares (RDWLS) estimation method with PROMIN rotation, which was consistent with the nature of the correlation matrix (polychoric) and the robustness of the selected rotation method (Lloret-Segura et al. 2014), all using the FACTOR program (Lorenzo-Seva and Ferrando 2006). In addition, before implementation of the CFA, we used the parallel analysis method (Lorenzo-Seva et al. 2011) for the number of factors to retain and then applied the CFA with the DWLS estimation method, after reviewing the coefficients of adaptation through the JASP program version 0.14.1 (JASP Team 2020).

4.1.2 Construct map to climate change scale

Once the factors at the base of the scale have been obtained, the second validation process consists of an inferential procedure of categorical analysis of the statements of the items (Dahlgren 1984) in the style of the outcome space of Marton (1981) whose final objective is the emergence of a construct map, which allows explaining from a qualitative perspective, the way in which the constructs represent attributes with a theoretical sense, capable of being measured by the scale under study (Messick 1989; Wilson 2005).

This is carried out through the definition of categories and the categorical convergence between items of the same factor (Kline 2013). In addition, it will make it possible to relate the responses of the subjects to the intermediate levels of the constructs (Wilson 2005), which in this case have well-defined limits. This is carried out, considering some indicators of the quality of the process (Masters and Wilson 1997) such as the finiteness and exhaustivity of the items, that the items are ordered, that there are contextual specifications and that they are based on research.

4.2 Creating indexes

After the validation step, we will apply a confirmatory factor analysis (Thompson 2004) using the data from both surveys together. Each latent variable is defined according to the survey validation result (see Sect. 3.2). So, the values of latent variables (sometimes referred to as factor scores) are estimated for each individual. Computationally, this can be done using the lavaan R-package (Rosseel 2012), where the cfa function performs the confirmatory factor analysis and the lavPredict function computes estimated scores.

These scores (one for each latent variable) can take any value, making interpretation difficult. Therefore, we propose a scale transformation to the unit interval, given by the following expression:

$$\text{Index}_i = \frac{\text{Score}_i - \min(\text{Score})}{\max(\text{Score}) - \min(\text{Score})},$$

(1)

where $\text{Index}_i$ represents a standardized factor score in the unit interval (we will call it index from now on), $\text{Score}_i$ is the estimated score for individual $i$ from the confirmatory factor analysis, and $\min(\text{Score})$ and $\max(\text{Score})$ are the minimum and maximum scores, respectively, considering the estimated values for all individuals.

4.3 Bayesian Beta regression

One of the indexes obtained will be our response variable. By construction, such index is defined in the unit interval (see Equation (1)) and so a standard linear regression to associate it with other variables is inappropriate, since this type of regression must be applied when the response variable is not bounded (Smithson and Shou 2020). To get around this problem, we will use a Beta regression model.
Methods have been developed in this field (Gelman 2013). Mathematically, the response variable $y_i \in (0, 1)$, for $i = 1, \ldots, n$, can be modeled using a Bayesian Beta regression given by:

$$y_i \sim \text{Beta}(\mu_i, \phi),$$

$$\logit(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} = x_i^T \beta,$$

where $\mu_i$ and $\phi$ are the Beta mean and precision parameters, respectively. The Beta mean is modeled, through a logit link, by a set of variables $x_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})^T$ weighted by its coefficients $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)^T$. We assume independent and diffuse prior distributions, where the $\beta$'s and $\log(\phi)$ follow a Normal distribution with zero mean and large variance. The Bayesian Beta regression model can be easily implemented using the `stan_betareg` function from the R-package `rstanarm` (Muth et al. 2018).

### 4.4 Bayesian model comparison

Suppose we have $m$ Bayesian models, say $M_1, \ldots, M_m$, to be compared. So, the relative plausibility of a particular model $M_v$ given its prior probability and the evidence from the data is quantified by the so-called posterior model probability (PMP) (Berger and Molina 2005), defined as follows:

$$\text{PMP}_v = P(M_v \mid \text{Data}) = \frac{P(\text{Data} \mid M_v)P(M_v)}{\sum_{j=1}^{m} P(\text{Data} \mid M_j)P(M_j)},$$

where we typically assume that the models are equally probable a priori, i.e., $P(M_v) = 1/m$ for $v = 1, \ldots, m$.

### 4.5 Sorting and selecting relevant variables

Reducing processing time for Bayesian inference is one of the main challenges of this methodological perspective. Therefore, over the last decades, many computational methods have been developed in this field (Robert 2014). Still, the inferential process remains very time-consuming, making the implementation of models with all combinations of variables impractical. To get around this problem, we used a forward approach (Heinze et al. 2018) in order to build different Bayesian models based on the relevance of each variable to explain the phenomenon studied. Our Bayesian forward-selection rule starts with no explanatory variables and then adds variables, one by one, based on which model is the a posteriori most likely, until there are no remaining variables. This strategy produces an ordering of relevance of the variables and avoids exploring the overall model space.

To illustrate our proposal, let’s assume that we have a response variable that can potentially be explained by three explanatory variables, $x_1$, $x_2$ and $x_3$, through a Bayesian model (e.g., the Beta regression introduced in Sect. 4.3). The first step would be to fit a model for each of these variables and then calculate the posterior model probability (PMP, see Equation (3)) for these three models. Hypothetically, the model with the variable $x_2$ (let’s name it as $M_1$) has the highest PMP, then this variable will be ranked as the most relevant to explain the response variable. The second step is analogous to the first, but now the variable $x_2$ is fixed and so there are only two competing models ($x_2 + x_1$ vs $x_2 + x_3$). Let’s assume the model with variables $x_2$ and $x_3$ (let’s name it as $M_2$) has a higher PMP than the model with variables $x_2$ and $x_1$, then the variable $x_3$ is the second most important and therefore $x_1$ is the least important among the three variables. Finally, we define as $M_3$ the model composed of $x_2 + x_3 + x_1$.

Note that we sorted the relevance of the variables and at the same time built three models, $M_1$, $M_2$ and $M_3$. The final step is to rank these three models using PMP as well. In practice, our methodology provides the researcher with the possibility of interpreting the best ranked models according to an objective measure of model selection. This approach enriches the discussions between the differences and similarities of such selected models.

### 5 Results

#### 5.1 Factor analyzes

The results of the factor analysis in each of the three steps described in Validation section are the following:

The first CFA, evaluated on the Spanish structure, showed two complexities for its correct implementation. Firstly, when using the JASP program (JASP Team 2020), it does not allow analysis with less than two items per factor, which made it impossible to introduce item 6 (V6) contained as a single component in the last category of the Spanish grouping. Secondly, it was not possible to obtain results from the other four categories proposed in the Spanish version, since when introducing the items corresponding to the program, it reported according to the lavaan R-package (Rosseel 2012) that the model was not admissible, as long as the matrix of covariance of latent variables was not positively defined. Therefore, Spanish structure is not admissible to the Chilean data.

Based on what was found in the attempt to carry out the CFA, the decision was made to apply an EFA in order to investigate what were the factors that made up the structure of the Chilean observational data.
Parallel Analysis method was used to determine the number of factors to extract. Two or three factors were identified as potential candidates, but the three factors specification presented a better interpretability and that is why it was chosen. In addition, it was also decided to rotate an item in the database, corresponding to item 4 (V4) because in this way the logic of their response options against the statement was consistent with the magnitude of the assignment score. Table 2 show the grouping of items by factor, where all the items have significant loads.

Regarding the (re)configuration found from the grouping of the items from the exploratory factor analysis, a confirmatory factor analysis of the structure of the scale items was carried out, this time for the second sample under study. The results of which will be described below.

The CFA applied to sample two shows that initially tested three-factor model presented an acceptable fit of the model to the data $\chi^2 = 213.423$ (p-value < 0.001); Root mean square error of approximation (RMSEA) = 0.035 (90% CI = 0.024 – 0.046); Tucker-Lewis Index (TLI) = 0.951; Comparative Fit Index (CFI) = 0.957; Bentler-Bonnet Non-normed Fit Index (NNFI) = 0.951.

Based on these indicators, it was possible to confirm that the responses of both scales (applied in two moments of time) are homogeneous between both and therefore, it is possible to join them to continue with the following steps.

### 5.2 Construct map

For the elaboration of the outcome space construct map of the climate change scale, some relevant elements will be described to ensure its quality.

1. **Order** in questions of a Likert-type survey, the order is implicit in the nature of the response options (Wilson 2005). In this case, those options are: ‘Strongly disagree’, ‘Disagree’, ‘Neither agree nor disagree’, ‘Agree’ and ‘Strongly agree’ with the statement; that is to say, 5 categories to which then values from 0 to 4 are respectively associated. But also, in the process of preparing the construct map, there is a second ordering conferred by the factor analysis, which resulted in 3 differentiated factors at the base of the climate change scale with Chilean data. The fact that 3 factors have been defined implies the grouping of items into 3 latent variables, whose theoretical meaning will be clarified in subsequent steps.

2. **Finitude** for the elaboration of this construct map, a sample (of two moments in time) has been taken as a reference among all the possible samples of the population of inhabitants of the region of Chile, from which the data comes. In addition, the questions on the scale are closed statements whose response options are also limited to only 5 possible options. Therefore, the measurement process in this case is associated with a limited context that does not intend to generalize its results.

3. **Categorical exhaustivity** all items are included at least in one of the defined categories, these categories being of two types: contextual to the Chilean case and based on previous research (see Table 6 of the Appendix A).

As a result of the ordering into factors and the categorization of the items within each factor, it is possible to define three constructs on the basis of the measured instrument: pro-environmental behavior (proenvbeh), naturalistic perception (naturalistic) and anthropogenic perception (anthropogenic) of climate change. **Pro-environmental behavior** (proenvbeh = V4, V5, V6, V9, V10, and V14). The categories that make up this variable from the perspective of the authors of this study are: belief (e.g., cognitive, emotional), awareness, norm and perceived behavioral control, and the categories that make up the variable from the perspective of studies prior, include self-efficacy, risk perception, behavioral intention, and attitude.
The presumption that the construct framing all these categories is pro-environmental behavior and is supported by theoretical reviews that will be presented below.

Pro-environmental behavior is a complex and multi-caused concept (Stern 2000) which can be defined as individual behavior aimed at reducing negative environmental impact (Stern 2000; Kollmuss and Agyeman 2002).

Among the determinants of this behavior, Ajzen (1991) highlights the intention to perform this behavior. In turn, this intention is determined by attitudes towards behavior, by subjective norms connected to behavior and perceived behavioral control. Besides, awareness is “knowing the impact of human behavior on the environment” (Kollmuss and Agyeman 2002).

A greater precision of these subjective representations is provided by Klöckner (2013) stating that the attitudes would be a general measure of the preference that an individual has as a behavioral alternative. The subjective norms would be the perceived expectations of other relevant people about what alternative behavior should be carried out and mediated by the will to fulfill that expectation. In the case of perceived behavioral control, it would be a measure of the opportunity and capacity that a person has to carry out a certain behavioral alternative.

Attitudes are the sum of all beliefs (cognitive and affective) about an activated behavior in a given situation. Understanding belief as the expectation that a behavior could lead to a certain result, accompanied by the probability that it will occur, as well as the evaluation of the degree to which the result will be favorable (Klöckner 2013).

According to Klöckner (2013) previous analyzes of environmental behavior (Harland et al. 1999; Heath and Gifford 2002; Tonglet et al. 2004; Han et al. 2010) and of planned behavior theory (Ajzen 1991) support the idea that people perform a behavior with positive environmental outcomes if they hold a positive attitude towards them, if other people expect them to act in that way and support them in doing so, and if they perceive themselves as being able to implement their intentions and, in relation to perceived behavioral control, can under certain conditions have an additional direct impact on behavior, for example when conditions change, before the behavior is performed.

Finally, self-efficacy is a significant variable of the predictive model on risk perception (Brody et al. 2008) that is linked both to the perception of climate change and to personal efforts to combat it, being risk perception the willingness to carry out individual actions to mitigate the effects on the environment (Hidalgo and Pisano 2010).

Anthropogenic perception (anthropogenic = V8, V11, V12, V15, V16, V17, V18, and V19). In the case of this third variable, the authors detected the following basic categories, which are anthropic cause, cause of greenhouse gases greenhouse, industrial cause, cognition, consciousness, knowledge and negative consequences. In addition, from the perspective of previous studies, there are the categories of knowledge and perception of risk.

Both the order and the categories that make up this variable make it possible to verify that it isolates the effect that anthropogenic influence would have as a cause of climate change, individually as a generator of greenhouse gases or at an industrial level as promoters of climate change, linked to consequences negative as well as cognitive notions such as awareness, knowledge and perception of risk.

This point of view connects the attribution of human activity to global warming observed since the mid-twentieth century (Stocker et al. 2013) and is closely related to the theory of “detection and attribution related to anthropic climate change” (Hegerl et al. 2010).

In relation to the way of classifying the items of the anthropogenic factor, the anthropogenic perception will be defined as: awareness, knowledge and perception of the risk of the negative consequences that human triggers have on climate change.

Specifically, perception is the process by which the world is represented and the product of which constitutes the conscious experience available for verbal or other reporting (Milner and Goodale 1995); the cognitive component associated with thought and consciousness will be that knowledge regarding the impact of human behavior on the environment (Kollmuss and Agyeman 2002); knowledge will be the amount of information that individuals have on environmental issues and their ability to understand and evaluate their impact on the environment (Blankenberg and Alhusen 2019); risk perception will be the perceived probability of negative consequences for oneself and for society of a specific environmental phenomenon, for example, global warming (O’Connor et al. 1999), the cause will be what triggers the problem and the consequences, its effects and scope.

Naturalistic perception (naturalistic = V1, V2, V3, V7, and V13). This variable is composed of the categories positive consequence, natural cause and belief, which were assigned by the authors of this study and by the categories perception of risk, knowledge and attitude, when considering the perspective of previous studies on these same items.

The identified variable isolates the effect that nature would have as a cause of climate change, linking it to positive consequences and introducing notions such as belief, risk perception, attitude, and knowledge as the basis for the naturalistic perception of climate change.

This perspective would be linked to the exact and natural sciences approaches to climate change that address the transformations of the atmosphere and its interaction at various scales with the sea and the continent (Soares et al. 2018). An example is the scientific strategy of Climate and...
Land Use Change of the US Geological Survey, which uses as a framework to understand and respond to global change the investigation of biological, hydrological, and geological processes, fundamental chemical and physical underlyi ng global stressors and environmental response (Burkett et al. 2013).

Due to these antecedents, naturalistic perception will be defined as: beliefs, attitudes, knowledge and perception of risk, which tend to attribute positive consequences to the natural triggers of climate change.

Perception will be interpreted as a process by which the world is represented and whose product constitutes the conscious experience available for verbal or other reports (Milner and Goodale 1995). Belief will be understood as an expectation about the result of a certain behavior (Blankenberg and Alhusen 2019), while attitude will be defined as a personal predisposition to respond to the object in a consistently favorable or unfavorable way (Ajzen and Fishbein 1973). Knowledge will represent the amount of information that subjects have about environmental issues, as well as their ability to understand and evaluate the impact of these problems on the environment (Blankenberg and Alhusen 2019). The perception of risk will be the probability of experiencing negative consequences for oneself and society from an environmental phenomenon such as global warming (O’Connor et al. 1999). Finally, cause will be what the problem is attributed to and the consequences such as its effects and scope.

5.3 Indexes

From the items grouped by factor (see Table 2), we estimated the latent variables – proenvbeh, naturalistic and anthropogenic indexes – for each individual (see Sect. 4.2 for more details). Then, the transformation to the unit interval is performed using Equation (1). Figure 2 shows the distribution of each index.

Visibly, proenvbeh and anthropogenic indexes have a similar pattern, where values greater than 0.5 are much more present. On the other hand, the index naturalistic is more concentrated on values close to zero and has more variability.

5.4 Model selection

Our main objective is to explain the behavior of the proenvbeh index through sociodemographic variables, naturalistic and anthropogenic indexes, and contrasting whether there are differences between surveys 1 and 2. See the description of these variables in Table 7 of the Appendix A.

To model the proenvbeh index, we used the Bayesian Beta regression (2) introduced in Sect. 4.3. In addition, we employed the strategy presented in Sect. 4.5 to order the most relevant variables and select the best models. Table 3 shows the final result.

As we can see in Table 3, M2 and M3 are the two models that best explain the index proenvbeh. In particular, M2 has a high posterior model probability (PMP) of being the best model among the analyzed models. However, the PMP for M3 is not negligible. For this reason, we will analyze both models. Table 4 shows a posterior summary for model M2.

Interpretively, we can conclude that anthropogenic and naturalistic have a positive and negative effect, respectively, on proenvbeh. More specifically, we can say that
increasing the index anthropogenic by 10% (and remaining the index naturalistic unchanged), we expect that, on average, the relative proportion increases by a factor of 1.77 ($\exp(0.1 \times 5.69)$). On the other hand, by increasing the index naturalistic by 10% (and remaining the index anthropogenic unchanged), we expect that, on average, the relative proportion to decrease by a factor of 0.96 ($\exp(-0.1 \times 0.41)$).

Table 5 shows a posterior summary for model M₃.

The results of model M₃ maintain the same findings as model M₂ regarding indexes anthropogenic and naturalistic. However, education is now present in the model. In this case, the interpretation is that the odds of proenvbeh for participants who have reached higher education is 1.1 ($\exp(0.1)$) times that those who studied at most until secondary school, while all other variables remain constant.

### 6 Discussion

When evaluating the climate change questionnaire presented by Hidalgo and Pisano (2010) with our Chilean data, we identified three factors that group variables related to each other. It should be noted that this result comes from the analysis of empirical data, which differs from the theoretical proposal of five categories that would group the same items in the questionnaire proposed by the Spanish authors, in their pilot. Then, these results are not comparable with each other, as they present levels of approximation to the object of study that are different. However, generating new information or evidence about the same instrument would allow, from a scientific point of view, to give continuity to the generation of psychometric knowledge about the questionnaire, which could present new configurations when evaluated in samples from other territories. In the Chilean context, which has a multiplicity of different climates, this same questionnaire could be used to evaluate the factorial structures in different regions of the country. In addition, the translation and application of this questionnaire could be done to other countries and the results compared with our study.

Using our methodology, it was found that in the regional context of Nuble in Chile, there are three relevant indicators to explain pro-environmental behavior (proenvbeh). First, a direct association of anthropogenic perception with pro-environmental behavior, which implies that higher scores in the index of cognitive-perceptual attributions about the negative consequences that human behavior has towards climate change, would increase the proenvbeh index. Second, an indirect association of naturalistic perception with pro-environmental behavior, which implies that higher scores in the index of subjective dispositions that assign a positive assessment to the consequences of climate change due to the effect of nature, would decrease the proenvbeh index. Third, a direct association between educational level and pro-environmental behavior, which...
implies that the higher educational level lead to an increase in the proenvbeh index.

Interestingly, our results provide local evidence where there seems to be a similar trend to the studies from developed countries, mainly in United States, Australia and United Kingdom, where greater questioning is expressed about anthropogenic or naturalistic attributions regarding from responsibilities to the base of the problem of climate change (Sapiains and Ugarte 2017a). This is important considering that to generate climate governance strategies (Bulkeley 2013), top-down and bottom-up models are required to integrate the local with scientific and global knowledge in order to contribute to climate change policies (Blhve et al. 2014; Butler et al. 2015). In relation to this last point, the promotion of educational interventions that emphasize raising awareness about the role that human behavior is having on negative consequences associated with climate change could promote favorable behaviors towards the environment, according to our results, which could collaborate actively with mitigation and adaptation measures against this phenomenon (Grafakos et al. 2018).

Among the limitations, our results can only be extended to populations similar to the ones we studied, because the samples were not probabilistic and therefore, the results are not generalizable to any population. However, it should be noted that the impacts of climate change are contextual (Lange and Dewitte 2019), by configuring the degree of vulnerability, capacities and danger. The data collection process was also a limitation in our study, since it is not standardized in both samples due to the global pandemic situation. However, exploratory and confirmatory factor analysis did not point out problems handling both samples together. Still, for future work, we recommend the use of probabilistic sampling (e.g., stratified).

7 Conclusion

Our proposal is based on a sequence of quantitative tools, such as factor analysis, creation of index, Bayesian Beta regression, and model selection. Constructing indexes from the factor analysis (or principal components analysis) is a common and recommended practice (OECD 2008). The interpretation of an index on the unit interval, instead of a variable with no specific limits, is usually straightforward, so its use is very useful in many applications. In our case study, the index also represents the response variable for the statistical model (see Sect. 4.3). By construction, this response variable is defined between zero and one and that is why the Beta regression was chosen. However, when the response variable is not an index, other models must be applied, such as linear and logistic regressions. It is noteworthy that the main role of statistical modeling is to find associations between the response variable and explanatory variables. In our application, the proenvbeh index was the response variable and the variables shown in Table 7 were the explanatory ones.

The use of Bayesian statistics in both modeling and model selection is one of the main contributions of this paper. The Bayesian paradigm has been introduced and defended in different fields of application (Kruschke and Liddell 2018; van de Schoot et al. 2021), but, in practical terms, many researchers still want to make decisions based on p-values. In this sense, we presented Bayesian tools for selecting models and interpreting the relevance of each variable (see Sects. 4.4, 4.5, and 5.4 for more details). Specifically, we would like to highlight the use of posterior model probabilities rather than Bayes factors for selecting models/variables, as the latter has some practical limitations, for example, decisions in favor of the null hypothesis are often based on rules-of-thumb and so may suffer the same criticism of “P-value < 0.05” (van der Linden and Chryst 2017). Furthermore, following Box’s famous aphorism “all models are wrong, but some are useful” (Box 1976), we advise (and illustrated in Sect. 5.4) the interpretation of more than one “better” model, since the application context can also support decision making.

As a future proposal, we intend to apply our methodology to other contexts and populations. From a methodological point of view, we would like to explore other regression models (including modeling by artificial neural networks using the references cited in Section 1) and compare our variable selection criteria with others proposed in the literature.

Appendix A

See Tables 6 and 7.
| Factor | Item | Statement | Chilean context proposed category | Category based on previous research |
|--------|------|-----------|-----------------------------------|-------------------------------------|
| Proenvbeh | V4   | It is very difficult for a person like me to do something for the environment. | Perceived behavioral control | Self-efficacy<sup>*</sup> |
|         | V5   | I think climate change is happening right now. | Cognitive belief; awareness | Risk perception<sup>†</sup> |
|         | V6   | I would be willing to make changes in my way of life to combat climate change. | Perceived behavioral control | Behavioral intention<sup>*</sup> |
|         | V9   | I believe that countries that violate international treaties to reduce climate change should be sanctioned in some way. | Belief; norm | Attitude<sup>*</sup> |
|         | V10  | There are simple things I can do that would have a significant effect in alleviating the negative consequences of climate change. | Perceived behavioral control | Self-efficacy<sup>*</sup> |
|         | V14  | I am concerned about the current climate change situation. | Emotional belief; awareness | Attitude<sup>*</sup> |
| Anthropogenic | V8   | The main cause of climate change is human activity. | Anthropic cause | Knowledge<sup>†</sup> |
|         | V11  | I have already noticed some signs of climate change. | Awareness | Risk perception<sup>†</sup> |
|         | V12  | The emission of greenhouse gases (CO2, Methane and others) is the main cause of climate change. | Greenhouse gases cause; knowledge | Knowledge<sup>†</sup> |
|         | V15  | It seems to me that the temperature is warmer now than in previous years. | Awareness | Risk perception<sup>†</sup> |
|         | V16  | Every time we use coal, oil or gas, we enhance the greenhouse effect. | Awareness; negative consequence | Knowledge<sup>*</sup> |
|         | V17  | I think rainfall has decreased in recent years. | Cognition | Risk perception<sup>*</sup> |
|         | V18  | The main sector causing climate change is industry (including energy production), ahead of transport, livestock and heating. | Industrial cause | Knowledge<sup>*</sup> |
|         | V19  | Overall climate variability and sudden and unpredictable changes have increased. | Awareness | Risk perception<sup>*</sup> |
| Naturalistic | V1   | There will be some positive consequences of climate change for the environment. | Positive consequence | Risk perception<sup>†</sup> |
|         | V2   | Climate change is due to natural causes. | Natural cause | Knowledge<sup>†</sup> |
|         | V3   | Many of the claims about environmental threats are exaggerated. | Belief | Attitude<sup>*</sup> |
|         | V7   | There will be some positive consequences of climate change on the sea/glaciers. | Positive consequence | Risk perception<sup>†</sup> |
|         | V13  | There will be some positive consequences of climate change on human health. | Positive consequence | Risk perception<sup>†</sup> |

<sup>*</sup> Hidalgo and Pisano (2010); <sup>†</sup> Heath and Gifford (2006); <sup>‡</sup> Sundblad et al. (2007)
Acknowledgements Paula Reveco-Quiroz would like to thank the National Agency for Investigation and Development (ANID) for funding her postgraduate studies through the Beca Magíster Nacional (number 22181231).

Author contributions P.R.Q. and J.S.D. reviewed the state-of-the-art on climate change research and quantitative methodologies. J.S.D. and D.A. contributed to conceptualization, formal analysis, supervision, review, and editing of the manuscript. D.A. and P.R.Q. performed the data curation and software implementation. J.S.D. applied the validation factor analysis and got funding. All authors contributed to the writing of methodologies, results and discussions.

Funding National Fund for Scientific and Technological Development (FONDECYT) grant 11200683.

Availability of data and materials Not applicable.

Declarations

Conflicts of interest The authors declare no competing interests.

Ethics approval The authors declare that there is no violation of ethical protocol.

Consent to participate The authors declare that they have taken necessary approval.

Consent for publication The authors consent the publication.

Table 7 Description of variables

| Variable | Description | Type | Codification |
|----------|-------------|------|--------------|
| Proenvbeh | Individual behavior aimed at reducing negative environmental impact (Stern 2000; Kollmuss and Agyeman 2002). | Index | Range (0, 1) |
| Anthropogenic | Subjective disposition of a cognitive type that includes awareness, knowledge and perception of risk, towards human triggers in the negative consequences of climate change. | Index | Range (0, 1) |
| Naturalistic | Subjective dispositions such as beliefs, attitudes, knowledge and perception of risk, which attribute positive consequences to the natural triggers of climate change. | Index | Range (0, 1) |
| Survey | Indicates whether the case in study belongs to survey one or two. | Binary | 0 = Survey 1 & 1 = Survey 2 |
| Sex | Indicates the sex of each participants. | Demographic | 0 = Woman & 1 = Man |
| Education | Indicates whether the participants have reached the higher education. | Demographic | 0 = Secondary or less & 1 = College or higher |
| Location | Indicates whether the participants belongs to a rural or urban territory. | Demographic | 0 = Rural & 1 = Urban |
| Age | Indicates the age of the participants | Demographic | Continuous values |
| Income | Indicates whether the participant receives income or not. | Demographic | 0 = Without paid work & 1 = With paid work |
| Poltrend | Indicates what is the political tendency of the participants. | Demographic | 0 = Left, 1 = Center, 2 = Right, & 3 = Apolitical |
| Know | Indicates whether or not the participant knows what climate change is. | Binary | 0 = No & 1 = Yes |
| Impreduc | Indicates whether or not the participant agrees that it is possible to reduce the impact of climate change. | Binary | 0 = No & 1 = Yes |

References

Adger WN (2001) Scales of governance and environmental justice for adaptation and mitigation of climate change. J Int Dev 13(7):921–931

AERA (2014) The standards for educational and psychological testing American Educational Research Association

Ajzen I (1991) The theory of planned behavior. Organ Behav Hum Decis Process 50(2):179–211

Ajzen I, Fishbein M (1973) Attitudinal and normative variables as predictors of specific behaviors. J Pers Soc Psychol 27(1):41–57

Andersson B, Wallin A (2000) Students’ understanding of the greenhouse effect, societal consequences of reducing CO2 emissions and why ozone layer depletion is a problem. J Res Sci Teach 37:1096–1111

Beiser-McGrath LF, Huber RA (2018) Assessing the relative importance of psychological and demographic factors for predicting climate and environmental attitudes. Clim Change 149(3):335–347

Berger J, Molina G (2005) Posterior model probabilities via path-based pairwise priors. Stat Neerl 59(1):3–15

Bhave AG, Mishra A, Raghwanshi NS (2014) A combined bottom-up and top-down approach for assessment of climate change adaptation options. J Hydrol 518:150–161

Blankenberg AK, Allhusen H (2019) On the determinants of pro-environmental behavior: a literature review and guide for the empirical economist. CEGE Discussion Papers 1(350):1–28

Boström A, Morgan M, Fischhoff B, Read D (1994) What do people know about global climate change? Risk Anal 14:959–970

Box GEP (1976) Science and statistics. J Am Stat Assoc 71(356):791–799

Boyes E, Stanisstreet M (1997) Children’s models of understanding of two major global environmental issues (ozone layer and greenhouse effect). Res Sci Technol Education 15:19–28
Torres Irribarra D (2021) A pragmatic perspective of measurement. Springer, Cham, Switzerland
van de Schoot R, Depaoli S, King R, Kramer B, Mårtens K, Tadesse MG, Vannucci M, Gelman A, Veen D, Willemsen J, Yau C (2021) Bayesian statistics and modelling. Nat Rev Methods Primers 1(1):1–26
van der Linden S (2017) Determinants and measurement of climate change risk perception, worry, and concern In: The Oxford encyclopedia of climate change communication. Oxford University Press, Oxford UK, pp 1–53
van der Linden S, Chryst B (2017) No need for Bayes factors: a fully Bayesian evidence synthesis. Front Appl Math Statistics 3:1–3
van der Linden WJ (1994) Fundamental measurement and the fundamentals of Rasch measurement In: Objective measurement: theory into practice, Ablex, Norwood USA, pp 3–24
Wilson M (2005) Constructing measures: an item response modeling approach. Routledge, New York USA
Wu M, Tam HP, Jen T-H (2016) Educational measurement for applied researchers Springer, Singapore

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.