AQUA
Water Infrastructure, Ecosystems and Society

 AQUA — Water Infrastructure, Ecosystems and Society Vol 72 No 8, 1512 doi: 10.2166/aqua.2023.063

Unobserved heterogeneity and temporal instability in an analysis of household water consumption under block rate pricing

Ahmad Altarabsheh[a,*], Dulcy Abraham[b] and Ibrahim Altarabsheh[c]

[a] Department of Construction Engineering and Management, Hijjawi Faculty for Engineering Technology, Yarmouk University, Shafiq Irshidatst, Irbid 21163, Jordan
[b] Department of Construction Engineering and Management, Lyles School of Civil Engineering, Purdue University, 550 W Stadium Ave., West Lafayette, IN 47907, USA
[c] Department of Civil Engineering, Materials and Infrastructure Systems concentration, Faculty of Engineering, Louisiana Tech University, 1310 Railroad Avenue #7923, Ruston, LA 71270, USA
*Corresponding author. E-mail: ahmad.gt@yu.edu.jo

AA, 0000-0002-8998-3263

ABSTRACT

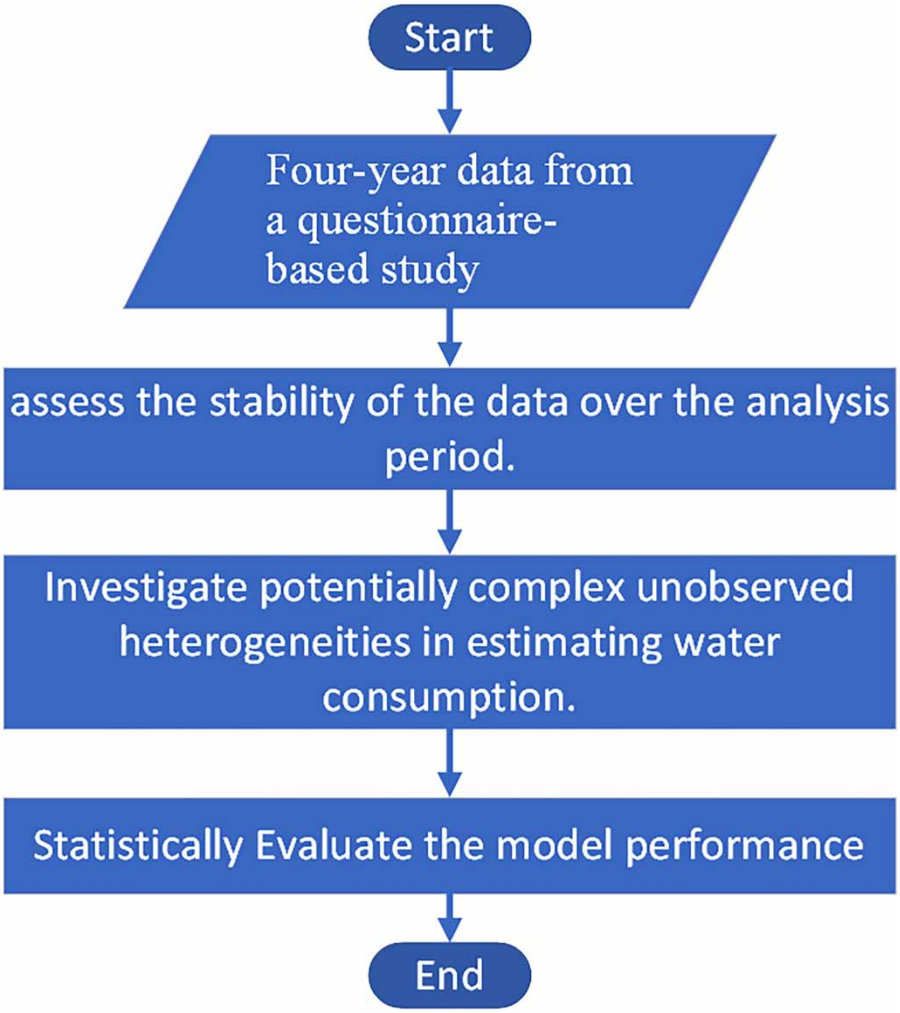
There is a lack of studies investigating household water consumption while considering the possible heterogeneity in the observed data and its temporal instability, often resulting in inconsistent and biased parameter estimations and, as a result, inaccurate forecasting of household water consumption. To address these constraints, the current study investigates temporal shifts in the household water consumption pattern and the effects of different socioeconomic factors on forecasting household water consumption. Using the results of a household survey performed seasonally in three major cities in Northern Jordan over a 4-year period, separate seasonal models of water consumption are estimated using three alternate modeling approaches to account for possible unobserved heterogeneity. Likelihood ratio tests were performed to investigate the temporal stability of the models' estimations over different seasons across the 4-year period. The findings of these tests indicated that the data are temporally stable over two datasets (the summer and winter seasons). Also, the findings revealed that household water consumption is influenced by a variety of factors, with the impact of many of these factors varying across observations. Finally, the findings highlight the need for additional research into how unobserved heterogeneity can be best modeled in temporal contexts for accurate water consumption forecasting.

Key words: ordered probit model, temporal instability, unobserved heterogeneity, water consumption

HIGHLIGHTS

- This study provides a framework for investigating the impact of socioeconomic factors on household water consumption while accounting for possible unobserved heterogeneity in temporal contexts.
- Separate seasonal models of water consumption are estimated using three alternate statistical modeling approaches to account for possible unobserved heterogeneity and data temporal instability.

GRAPHICAL ABSTRACT



1. INTRODUCTION

For any civilization to thrive and flourish, water is required; however, many nations across the world face acute water scarcity. Over 1.1 billion people worldwide lack access to safe drinking water, and 2.7 billion suffer from water shortages for at least 1 month out of the year (Motoshita *et al.* 2018).

In terms of water resources, Jordan is classified as the fourth poorest country in the world (World Bank 2013). This problem is further exaggerated by climate change and a sharp increase in the population because of unexpected waves of refugees from surrounding countries (UNHCR 2015). In response to this issue, Jordan has established a national water plan for 2008–2022, with the goal of institutionalizing the country's water demand-management process (MWI 2009). The plan's primary goal is to raise public awareness about the country's water scarcity, as well as to encourage individuals to use water-saving devices and to install such devices in government buildings and schools (MWI 2009).

Towards addressing the goals of this plan, this study provides a framework for investigating the impact of socioeconomic factors on household water consumption while accounting for possible unobserved heterogeneity in temporal contexts.

Attempts to forecast household water consumption in terms of socioeconomic characteristics are undeniably confounded by individual consumer behavior, which can vary greatly throughout the population and over time because of customers' diverse physical, economic, environmental, and attitudinal traits, as well as their varied responses to external stimuli (Fricke 2013; Donkor *et al.* 2014).

Given the complexities of forecasting household water consumption, to the best of the authors' knowledge, the current regression models in the literature do not account for unobserved factors in the data or temporal variability of the data over time, which may result in biased and inefficiently calculated parameters, hence resulting in erroneous predictions.

Accounting for unobserved factors is critical because the analyst is unlikely to have access to all factors influencing the prediction of household water consumption. Ignoring the impact of unobserved factors (which constitute unobserved heterogeneity) on predicting household water consumption can result in erroneous attribution of the impact of observed variables on water consumption.

Data instability may exist due to the change in consumers' water consumption over the analysis period. For example, predicting seasonal variation in household water consumption using only data on climate (such as temperature and humidity) can result in biased prediction, since household water consumption can vary within the same season and across the analysis period. Such variation can be attributed to the differences in consumers' characteristics and the change in their behavior over the analysis period.

For example, consumers may have different dwelling unit characteristics (such as the area of the dwelling unit or the presence of a garden) and household sociodemographic characteristics (such as number of the family members living in the household, age of household members, and the average monthly income of the household) among other characteristics, in addition, consumers may reduce their consumption over time by employing different techniques including the use of water-conservation fixtures, the collecting of rainwater, or the recirculation of gray water, among others. The different characteristics of consumers along with their changing behavior throughout the analysis period may result in unique water consumption values for each consumer at each season.

To address the challenges of unobserved heterogeneities and temporal variability of data, a two-step procedure is proposed in the present study. First, a series of likelihood ratio (LR) tests were performed to statistically test the stability of the data over the analysis period.

Following this, three distinct models were used in this study to address the data's diversity and variability, which could not be fully explained by the observed variables alone. The first model, known as the random parameter ordered probit model with heterogeneity in means, used random parameters that differed between individuals. This allowed for differences in the means of the variables to be accommodated, effectively accounting for individual-level variations that may impact water consumption.

The second model, the random thresholds random parameters hierarchical ordered probit model, expanded on the previous one by introducing random thresholds in addition to random parameters. It was recognized that different criteria for water consumption choices may exist among people by taking into account variations in decision-making thresholds. The third model, the latent class ordered probit model with class probability functions, used a latent class framework to divide individuals into distinct latent classes or groups based on their water consumption characteristics and preferences. The class probability functions estimated the probabilities of belonging to each class, providing valuable insights into the population's heterogeneity.

The application of these three methodological approaches allowed for a thorough evaluation of the temporal complexity of the socioeconomic factors influencing water consumption. Each model was able to capture various facets of heterogeneity, allowing for a thorough analysis of the data and the illumination of the various influences influencing water consumption patterns.

The proposed procedure was applied to 4-year data from a questionnaire-based study of household water users in three main cities in Northern Jordan: Irbid, Jerash, and Ajloun. The population size for the survey was calculated using stratified proportional sampling to prevent bias in the response rates (Chang & Vowles 2013). Data were gathered across the summer, spring, autumn, and winter seasons on five key issues: household sociodemographic factors, dwelling unit characteristics, water supply service, water-use behaviors, household sociodemographic characteristics, and attitudes about water conservation.

In terms of statistical fit and forecasting accuracy, four commonly used statistical metrics were utilized to evaluate the performance of the selected models: LR test, adjusted McFadden pseudo-R^2 (Adj-R^2), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC).

Following a review of prior research efforts in forecasting water consumption, the paper describes the methodological approach and the models developed to investigate the unobserved heterogeneity in factors influencing water consumption in a temporal context. Then, key features of the data collection and qualitative assessment of survey data are described, followed by tests for determining the temporal stability of the data over the analysis period. Next, the validation process for the three models was discussed, followed by a discussion of the results. Finally, the implications of the findings of the study are described, along with recommendations for future improvements.

2. LITERATURE REVIEW

A wide variety of statistical models have been developed in the literature to forecast household water consumption. These models can be broadly classified into two groups: parametric statistical (regression) models and machine-learning models. Lee & Derrible (2020) argued that although parametric models are more widely used, both groups have their own benefits and drawbacks. Although parametric models are more insightful and interpretable than machine-learning models, they tend to be less accurate than machine-learning models.

On the other hand, machine-learning models have the ability to capture nonlinear and complex interactions between the variables in a model; however, they are less interpretable than parametric models (Lee & Derrible 2020).

Examples of parametric regression models developed in the literature include multiple linear regression (Polebitski & Palmer 2010; Bakker *et al.* 2014; Piasecki *et al.* 2018), time series model (Michelsen *et al.* 1999; Alvisi *et al.* 2007; Dey 2022), two-step seemingly unrelated regression models (Acharya & Barbier 2002), and probit model (Cheesman *et al.* 2008).

On the other hand, examples of regression-based machine-learning models developed in the literature include artificial neural networks (Firat *et al.* 2009; Babel & Shinde 2011; Behboudian *et al.* 2014; Altunkaynak & Nigussie 2017), fuzzy- and neuro-fuzzy-based methods (Atsalakis & Minoudaki 2007; Tabesh & Dini 2009), random forests (Bolorinos *et al.* 2020), graph convolutional recurrent neural network (Zanfei *et al.* 2022), and support vector regression (Bai *et al.* 2014; Brentan *et al.* 2017; Vitter & Webber 2018; Erdogan *et al.* 2021).

Recently, Bich-Ngoc *et al.* (2022) implemented both fixed and mixed-effects linear regression with spatial random intercepts to investigate the spatial patterns of household water consumption in Wallonia, Belgium. The study investigated the impact of six different factors including household characteristics, alternative sources of water, dwelling properties, water appliances, consumers' attitudes, and urban form. The study found that household water consumption is directly proportional to the area of the household dwelling and inversely proportional to the utilization of rainwater as alternative source.

Dang *et al.* (2022) implemented cluster tree and factor analyses to investigate the factors affecting spatiotemporal variations of household water consumption in Northwest China. The study found that increasing the use of water-saving appliances and water-use efficiency would decrease water consumption in Northwest China.

The aforementioned models show an increasing trend in investigating spatiotemporal variation in household water consumption and the socioeconomic factors that may cause such variations. Despite such advancements in the prior models for predicting household water consumption, they have few drawbacks. First, these models do not take into account the impact of unobserved factors in the data on predicting water consumption and instead base their predictions on the data that are readily available.

Predicting household water consumption is a complex process that includes a wide range of consumer responses to external stimuli, as well as complex interactions among consumers, water network features and conditions, supply-related factors, environmental conditions, and other factors. With such complexities, access to all of the data that could potentially predict the household water consumption is impractical. The absence of such critical data can pose serious specification issues for traditional statistical analyses, resulting in biased and inconsistent parameter estimates, incorrect inferences, and incorrect water consumption predictions.

Another drawback of the prior household water consumption forecasting models is that they assume temporally stable parameters, thus they do not account for temporal variation of household water consumption in their prediction. Temporal variations in household water consumption can be linked to a variety of factors, including changes in climate parameters and consumers behavior. While the former's impact on water consumption was extensively covered in the literature by predicting seasonal variation in water consumption, as shown in Table 1, the latter's impact, to the best of the authors' knowledge, has received little attention in the literature.

Examples of temporal variation in consumers' behavior would include increased awareness of water scarcity and increased adoption rate of water-conservation fixtures, among others. Considering temporal variation in consumer behavior allows analysts to more accurately predict not only seasonal variation in water consumption, but also variation from year to year for the same season, whereas ignoring such variation may result in incorrect conclusions and ineffective water-conservation policies.

Table 1 | Summary of factors affecting water consumption

Variable	Influence on water consumption	References	Location
Climate	Low variation between winter and summer household water consumption	Willis *et al.* (2011), Makki *et al.* (2015), Manouseli *et al.* (2019)	Gold Coast, Australia; Southeast Queensland, Australia; South Essex, England
	Water consumption is directly proportional to both rainfall rates and temperature.	Grafton *et al.* (2011), Gato-Trinidad *et al.* (2011)	Melbourne city, Australia
	Household water consumption is directly proportional to temperature, while precipitation and humidity have no effect.	Zapata (2015)	Ecuador
	Household water consumption is, on average, higher in the summer than in the winter, owing to the increased outdoor and indoor water-dependent activities such as irrigation, and it is indifferent to both temperature and rainfall rates.	Gato-Trinidad *et al.* (2011), Kenney *et al.* (2008), Rathnayaka *et al.* (2015, 2017)	East Doncaster, Australia; Melbourne city, Australia; and Aurora, Colorado, USA
	Per capita water consumption is inversely proportional to annual precipitation.		Phoenix, Unites States
Education and Household Income	Water consumption is inversely proportional to level of education and household income, which is attributed to the tendency of consumers with high income and high educational level to use conservation measures.	Arbués *et al.* (2010), Willis *et al.* (2011, 2013), Makki *et al.* (2015), Watson (2017), Vieira *et al.* (2018)	Zaragoza, Spain; Gold Coast, Australia; Southeast Queensland, Australia; Bexley and Greenwich, England; Portugal
	The observed correlation between household income and water consumption is not always accurate because low-income consumers typically practice water conservation.	Makki *et al.* (2015), Olmstead & Stavins (2009)	Southeast Queensland, Australia
Occupants' ages	Household water consumption is inversely proportional to occupant age, which is attributed to occupants' proclivity to adopt water-conservation measures and have end-use events (e.g., showers) with shorter duration and frequency as they age.	Aprile & Fiorillo (2017), Beal *et al.* (2011), Makki *et al.* (2015), Willis *et al.* (2013)	Italy; Southeast Queensland, Australia; Southeast Queensland, Australia; Gold Coast, Australia
Size of dwelling unit	Household water consumption is directly proportional to the area of the dwelling unit.	Renwick & Green (2000), Guhathakurta & Gober (2007), Crouch *et al.* (2021)	California and Phoenix City, Arizona, USA, and South Africa
	Household water consumption is indifferent to the increase in the area of the dwelling unit	Kontokosta & Jain (2015)	New York City, USA
Technology	Household water consumption is inversely proportional to the use of water-conservation devices, such as low-flow toilets tools, showerheads tools, water-efficient irrigation devices, and low-flow taps	Renwick & Archibald (1998), Fielding *et al.* (2012)	California, USA and Southeast Queensland, Australia
	The use of water-saving devices must be accompanied by water-conservation behaviors to avoid offsetting consumer behaviors; indeed, the researchers discovered that higher water conservation was achieved when consumers were unaware of the installed efficiency devices.	Fielding *et al.* (2012), Bartos & Chester (2014)	Southeast Queensland, Australia and Arizona, USA

The aforementioned flaws in the developed prediction models are demonstrated by comparing the factors found in the literature to influence household water consumption, as summarized in Table 1.

Table 1 shows that different effects on water consumption have been reported in the literature for many of the variables (such as the seasonal variation of water consumption, the impact of the increase in temperature and rainfall rates on water consumption, the impact of household income on water consumption, and the efficiency of using water on reducing water consumption).

As shown in the table, several researchers (Willis *et al.* 2011; Makki *et al.* 2015; Manouseli *et al.* 2019) found low variation between winter and summer household water consumption, while others found significant variation of water consumption in term of different climate parameters including both rainfall rates and temperature (Gato-Trinidad *et al.* 2011; Grafton *et al.* 2011) and outdoor and indoor water-dependent activities (Kenney *et al.* 2008; Rathnayaka *et al.* 2015, 2017).

The observed differences in the findings between different studies can be attributed to the simplicity of traditional regression models: these models do not account for the impact of unobserved factors on water consumption or the temporal stability of the data, which results in biased and inaccurate outcomes.

3. SURVEY DEVELOPMENT AND DEPLOYMENT

In this study, a household survey was created and deployed to understand the water consumption behavior of household consumers in three major cities in Northern Jordan: Irbid, Jerash, and Ajloun. Table 2 depicts the disparities between these cities in terms of size, population, yearly precipitation, and average water usage, as obtained from the Department of Statistics (DOS 2022), Government of Jordan (http://dosweb.dos.gov.jo/).

The three cities share a Mediterranean climate, with hot, dry summers and cold, and rainy winters. The Yarmouk Water Company is the water service provider for the three cities, supplying water from local springs and aquifers (AFD 2015; JICA 2015). However, these springs and aquifers are being exploited beyond their capacity for long-term renewal (AFD 2015).

Water service is intermittent in the three cities, owing to a scarcity of water resources in these cities. The Yarmouk Water Company provides water to the three cities for 6–24 h per week (JICA 2015). Because of the limited supply, consumers install water tanks and wells at their homes to store the water they receive from the water company and compensate for any shortage by collecting rainwater and storing it in household wells, as well as purchasing additional water from the water company or from external licensed sources.

The developed survey, which is presented in Appendix A, gathered data for different seasons on several socioeconomic factors that impact the water consumption behavior of household consumers. The surveys were conducted over a 4-year period (from November 2017 to January 2022). In particular, the designed survey gathered information on five main types of data: (1) dwelling unit characteristics, (2) water supply service, (3) water-use behaviors, (4) household sociodemographic characteristics, and (5) attitudes related to water conservation (Table 3).

The survey questionnaire was approved by the Institutional Review Board (IRB) of Yarmouk University, with an exempt designation and protocol number IRB/2021/41. The survey was disseminated using a proportionate stratified sample method to account for the variance in population size between the three cities, hence reducing any bias in the response rate. Based on this method, the sample sizes (based on a confidence level (CL) of 95%) for Irbid, Jerash, and Ajloun cities are 381, 40, and 30, respectively.

For this study, a total of 645 complete surveys were collected from the water consumers in the three cities, among which there were 471 surveys for Irbid city, 84 surveys for Jerash, and 67 surveys for Ajloun. Twenty-six surveys did not include information about the city of the respondents and were thus excluded from the analysis.

The survey was based on a deep literature search to identify potential socioeconomic factors influencing water consumption. The survey was designed and distributed to the sample population between November 2017 and January 2022, using Google Forms (a web-based survey software tool that was also used to collect responses).

Table 2 | Demographic and weather characteristics of the three adopted cities

City	Area (km^2)	Population	Annual precipitation (mm)	Average water consumption (liter/day per capita)
Irbid	1572	1,911,600	273	100
Jerash	410	237,000	204	87
Ajloun	420	181,000	230	83

Table 3 | Summary of the variable ranges used in the survey

Variable expression	Variables description	Mean	Maximum value
Dwelling unit characteristics data			
X1	Hometown (Irbid, Ajloun, Jerash)	0.34965	2
X2	Area of dwelling unit (m^2) (50–100, 100–150, 150–200, 200, or above)	1.977273	3
X3	Type of dwelling unit (studio, apartment, house)	1.657596	2
X4	Ownership type (own, rent)	0.16129	1
X5	Garden area (m^2) (no Garden, 50–100, 100–300, 300–600, above 600)	1.657767	4
X6	Garden watering frequency per week (0, 1, 2, 3 times, or above)	1.070732	3
X7	Number of bathrooms in the household (1, 2, 3, $\geq$4)	1.373853	3
X8	Total capacity of household water tanks and wells (m^3) ($<$2, 2–5, 5–8, $>$ 8)	1.74463	3
Water supply service data			
X9	Household main water source (multiple choices are allowed) (water company, rainwater storage wells at residence, water delivery by independent suppliers)	0.46127	6
X10	Frequency of receiving water from water company (assuming duration of 6 h for each delivery) (at most 1, 2, 3 times, or above per week)	0.143519	2
X11	Sufficient water is received weekly from the main household water source (i.e., water company, rainwater storage wells at residence, water delivery by independent suppliers) (yes, no)	0.473804	1
Water-use behavior data			
X12	Resources to compensate for weekly water shortage (multiple choices are allowed) (water delivery by independent company, rainwater storage wells at residence)	0.673913	2
X13	Average monthly payment to compensate for water shortage (JOD) (0–10, 10–20, 20–30, 30–40, 40, or above)	1.039409	4
X14	Quarterly water bill (JOD) (0–15, 15–30, 30, or above)	1.047191	2
X15	Average duration for showers each week (hours) ($<$1, 1–3, 3–5, 5, or above) for the whole household.	0.776978	3
Household sociodemographic data			
X16	Number of the family members living in the household (1–3, 3–6, 6–9, $>$ 9)	1.472973	3
X17	Average age of the family members (years) (5–15, 15–30, 30–50, $<$ 50)	1.554217	3
X18	Average monthly family income (JOD) (80–200, 200–400, 400–800, 800, or above)	1.981777	3
Water-conservation attitudes			
X19	Utilization of water-conservation fixtures (I do not utilize water-conservation fixtures, I utilize water-saving fixtures in at least one location)	0.686567	1
X20	Maximum willingness to pay for a single water-conservation fixture (JOD) (10–25, 25–50, 50–75, 75, or above)	0.411616	3
X21	Utilization of the recycled water (I do not utilize recycled water; I utilize recycled water for at least one purpose)	0.584507	1
X22	Utilization of the collected rainwater (I do not collect rainwater; I collect rainwater for at least one purpose)	0.538462	1
X23	Awareness of the availability of water-saving fixtures (aware, not aware)	0.238876	1
X24	Factors affecting consumers' decision to keep using water-conservation fixtures (20% saving in water bill, 10% saving or more in water bill, I will continue using it regardless of the expected savings)	1.039181	2
X25	Awareness on water scarcity in Jordan (aware, not aware)	0.206497	1

The survey design and measurement accuracy were measured using the method proposed by Mora (2011), which includes testing for both face and content validity. Face validity refers to a brief review of survey questionnaires by people with limited knowledge of the subject matter, whereas content validity refers to a subjective measure of how appropriate the questionnaires appear to a group of reviewers with knowledge of the subject matter.

First, the survey was pre-deployed to 12 consumers with limited knowledge of water consumption, price structure, distribution system, and engineering issues to assess face validity. This step is critical to ensuring that consumers with limited knowledge of these issues can understand and respond to the survey questions. Following that, the content validity of the survey was determined through a content review by six subject matter experts with backgrounds in water distribution systems, price-rate regulations, and public perception surveys.

In this study, these pricing blocks are used to represent household water consumption. This allows us to account for the ordinal nature of household water values by limiting the number of responses to three ordered responses (the pricing blocks) instead of discrete and indefinite number of responses (water consumption values). This is important because ignoring the ordinal nature of the data in the modeling process results in inefficient model parameters' estimation. Additionally, by expressing water consumption in terms of pricing blocks, data collection is made easier because participants are more likely to recall the cost of their water bill than their actual water consumption.

Subject matter experts provided insight on water block pricing based on the most common water bill ranges encountered by household customers of the Yarmouk Water Company. They indicated that 30 Jordanian Dinar (JOD) and higher was the top tier of quarterly water bill pricing for household water customers. Hence, in the survey process, three quarterly water bill price blocks were examined (Block 1 < 15 JOD, Block 2 – 15–30 JOD, Block 3 above 30 JOD), as advised by the subject matter experts. Customers who participated in the face validity stage also provided additional validation for the chosen ranges, stating that their water bill value always falls within one of these ranges.

Table 3 summarizes these ranges of responses to the survey variables, as well as descriptive statistics for each variable based on the entire dataset over the analysis period. The responses are grouped into classes, and each class is encoded as integer in the model (values provided by the respondent are not individually encoded as integers). The first variable (X1) in the table, for example, seeks information about the participants' hometown; the three options for this variable were encoded as integers into the model, with the choice of Irbid city being assigned zero, the choice of Ajloun city being assigned one, and the choice of Jerash city being assigned two. The same procedure for variable selection is followed for all variables.

First, for dwelling unit characteristics variables, the table shows that, 76% of the respondents were from Irbid city, 13% were from Jerash city, and 11% were from Ajloun city, 39% of the respondents live in a dwelling unit with area between 150 and 200 m^2, 30% live in a dwelling unit with area equal to 200 m^2 or larger, 25% of the respondents live in a dwelling unit with area between 100 and 150 m^2, and 6% of the respondents live in a dwelling unit with area between 50 and 100 m^2, 68% of the respondents live in a house, 30% live in an apartment, and 2% live in a studio, 80% of the respondents own their dwelling unit, while 20% rent their dwelling unit.

In addition, 73% of the respondents have a garden at their dwelling unit, with 74% of these respondents not having to water their garden once a week, which can be attributed to the use of rainfed agriculture (such as olive trees and other fruit trees). To address the intermittent water supply, 8% of the respondents installed household tanks and wells with total capacity of less than 2 m^3, while 48% of the respondents installed household tanks and wells with total capacity of 5 m^3 or more.

Second, for the water supply service variables, the survey results indicate that 65% of the respondents reported the water company as the main water source for their dwelling, 24% reported rainwater storage wells at residence as their main water source, and 11% reported purchasing water from independent suppliers as their main water source. Eighty-seven per cent of the respondents reported receiving water at most once a week from the water company, assuming a duration of 6 h for each delivery (JICA 2015), while 3% reported receiving water three times or more from the water company, and 53% of the participants stated that they receive enough water from the water company each week.

Third, for water-use behavior variables, the table shows that 90% of the respondents either purchase water from independent suppliers or install rainwater storage well at their dwelling (in an equal proportion) to compensate for their water shortage, while only 10% use both options (water purchase and rainwater storage). Sixty-nine per cent of the respondents pay less than 20 JOD per month to compensate for their water shortage. Forty-one per cent of the respondents have a quarterly water bill in the range of 15–30 JOD, 30% have a quarterly water bill of 30 JOD or more, and 29% of the respondents have quarterly water bills in the range of 0–15 JOD.

Fourth, for the household sociodemographic variables, the results show that the average number of family members for respondents is between 3 and 6 members, with fewer than 10% of the households having more than 9 members or fewer than 3 members. The average age of the respondents' family is between 15 and 30, and the average monthly income of the respondents' family is between 400 and 800 JOD, with 32% of the respondents having a family monthly income of more than 800 JOD, and 30% having family income of less than 400 JOD.

Finally, the table shows that 69% of the respondents stated that they utilize water-saving fixtures in at least one location in their dwelling, 58% stated that they recycle water and utilize it for at least one purpose in their dwelling, and 54% stated that they collect rainwater and utilize it for at least one purpose in their dwelling. Sixty-nine per cent of the respondents are willing to pay less than 10 JOD for a single water-saving fixture. Twenty-four per cent of the respondents were not aware of the availability of water-saving fixtures, while 20% were not aware of the water scarcity in Jordan.

4. METHODOLOGICAL APPROACH

In order to account for different potential unobserved heterogeneity in estimating water consumption and to validate the temporal stability of the observed data, this study investigates the use of three alternative modeling approaches, each of which accounts for a different type of unobserved heterogeneity in the observed data. These approaches are (1) a random parameter ordered probit model with heterogeneity in means, (2) a random thresholds random parameters hierarchical ordered probit model, and (3) a latent class ordered probit model with class probability functions. Table 4 shows the key feature and the value of using each approach.

4.1. Random parameters ordered probit model with heterogeneity in the means

The random parameters ordered probit approach with heterogeneity in means estimates unique parameters value for each participant by assuming that each estimated parameter varies across participants based on a certain statistical distribution (i.e., normal distribution) and that each estimated parameter is a function of the participant characteristics.

Table 4 | Key features of three models used in this study

Model	Features	Value added by the model
Random parameters ordered probit model with heterogeneity in the means.	Determines the factors that impact the consumption of each participant by: 1. Assigning different parameter value for each participant. 2. Assuming each estimated parameter is a function of the participants' characteristics	Accounts for unobserved heterogeneity in the data by assuming that the estimated parameters vary across observations based on the statistical distribution (i.e., normal distribution) and the distribution mean is a function of the participants' characteristics.
Random threshold random parameter hierarchical ordered probit model.	Determines the factors that impact the consumption of each participant by: 1. Assigning different parameter value for each participant. 2. Assigning different threshold value for each participant. 3. Assuming each estimated threshold is a function of the participants' characteristics.	Accounts for unobserved heterogeneity in the data by: 1. Assuming that the estimated parameters vary across observations based on statistical distribution (i.e., normal distribution) with constant mean. 2. Assuming that the estimated thresholds vary across observations based on statistical distribution (i.e., normal distribution) and the distribution mean is a function of the participants' characteristics.
Latent class ordered probit model with class probability function.	Determines the factors that impact the consumption of each participant by dividing participants into latent classes based on a certain probability, and assuming that the probability varies among participants based on their characteristics.	Accounts for unobserved heterogeneity in the data by assuming that estimated parameters vary across latent classes and such variation is unique for each participant based on the participant's characteristics.

To formulate this model, first a function that determines household water consumption outcomes is created (Greene 2012)

$$y_i' = \beta X_i + \varepsilon_i,\ y_i = k, \quad if\ \ \mu_{k-1} < y_i' < \mu_k,\ k = 1,\ \cdots, K \tag{1}$$

where y_i is the observed value of the water consumption level corresponding to observation (i), y_i' is a latent variable corresponding to observation (i), β is a vector of estimable parameters, X_i is a vector of explanatory variables corresponding to observation (i), μ is a vector of threshold parameters that define y and are estimated with β, k is a vector of integers that represents the ordered water consumption levels, and ε_i is a normally distributed random error term with zero mean and variance equal to one.

The possibility that the estimated parameters may vary significantly across observations can be achieved by adding a randomly distributed error term (ω), as shown in the following equation (Greene 2012):

$$\beta_i = \beta + \omega_i \tag{2}$$

where β_i is a vector of estimable parameters that differs across observations, β is a vector of estimated mean parameters, and ω_i is a vector of randomly distributed terms.

The random parameters are allowed to vary according to the explanatory variables. In particular, the current study aims to capture heterogeneity in the means of random parameters by rewriting Equation (2), as follows (Greene 2012):

$$\beta_i = \beta + \omega_i + \emptyset Z_i \tag{3}$$

where Z_i is a vector of covariates related to observation i and affect the mean of β_i, $\emptyset$ is a vector of estimable parameters, and all other terms are as previously defined.

4.2. Random threshold random parameter hierarchical ordered probit model

A random threshold random parameter hierarchical ordered probit model is applied to account for the change of the model's estimated thresholds using explanatory factors and the variation of the estimated parameters across observations.

In particular, this model estimates unique parameters and thresholds value for each participant by assuming that they both vary across participants based on a certain statistical distribution (i.e., normal distribution), in addition, the model assumes that the estimated thresholds for each participant are a function of the participant characteristics.

First, the random thresholds can be defined as follows:

$$\mu_{i,k} = \mu_{i,k-1} + e^{t_k + \gamma_k u_{ik} + d_k S_i} \tag{4}$$

where u_{ik} is a standard normally distributed term corresponding to observation (i) and threshold (k), t_k and γ_k are the mean and standard deviation for the intercept of a threshold (k), respectively, S_i is a vector of variables affecting the thresholds corresponding to observation (i), d_k is a vector of estimable parameters for S_i, and all other terms are as previously defined. The variations in the covariates across the observations can be defined by rewriting Equation (2) as follows:

$$\beta_i = \beta + \Gamma \omega_i \tag{5}$$

where Γ is the diagonal matrix of the standard deviations, and all other terms are as previously defined.

In this context, the ordered probability of each consumption level (k) for each observation (i) can be calculated as follows (Greene 2012):

$$P_i(k) = \varphi(\mu_k - \beta X_i) - \varphi(\mu_{k-1} - \beta X_i) \tag{6}$$

where $\varphi(.)$ is the cumulative normal distribution, which is given by Equation (7), and all other terms are as previously defined.

$$\varphi(\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-0.5\alpha^2} \partial\alpha \tag{7}$$

4.3. Latent class ordered probit model

The latent class ordered probit model is applied to account for data heterogeneity by grouping the data and allowing estimated parameters to vary only across the groups while remaining constant within each group. In this context, the probability $P_i(m)$ that an observation i belongs to latent class m is given by Greene (2012), as follows:

$$P_i(m) = \frac{e^{\alpha_m Z_i}}{\sum_{\forall M} e^{\alpha_m Z_i}} \tag{8}$$

where Z_i is a vector of covariates that indicates the latent class probabilities for each observation i, and α_m is a vector of related estimable parameters.

In addition, Equation (6) can be rewritten in terms of latent class (m), as follows (Greene 2012):

$$P_i(k|m) = \varphi(\mu_{km} - \beta_m X_i) - \varphi(\mu_{k-1,m} - \beta_m X_i) \tag{9}$$

where all terms are as previously defined. Furthermore, the unconditional probability $P_i(k)$ of an observation, i, which results in a consumption level, k, can be calculated as follows (Greene 2012):

$$P_i(k) = \sum_{\forall M} P_i(m) * P_i(k|m) \tag{10}$$

where all terms are as previously defined.

To improve estimation efficiency, a Halton sequence approach (Halton 1960; Bhat 2003) is used to simulate the maximum likelihood estimation process. In addition, marginal effects are used to capture the influence of the covariates on the probability of each water consumption level. This is achieved by taking the first derivative of the unconditional probability in Equations (6) or (10) with respect to the covariate under interest. Thus, the marginal effects can be used to evaluate the impact of a 'one-unit' change in each covariate on the consumption level (Greene & Hensher 2010).

4.4. Temporal stability tests

The current study employs three different tests to assess data temporal stability over the study period. First, a series of LR tests were performed to statistically examine whether water consumption models differ significantly across the seasons from 2018 to 2021. The Chi (X^2) distributed test statistic was used for this purpose, with degrees of freedom equal to the number of parameters estimated:

$$X^2 = -2*[LL_{s2t1,s1t1} - LL_{s1t1}] \tag{11}$$

where $LL_{s2t1,s1t1}$ is the log-likelihood at convergence for each model (for a given year $(t1)$), estimated using converged parameters from specific season $(s2)$ data on different season $(s1)$ data for that year and LL_{s1t1} is the log-likelihood at convergence of each model using season $(s1)$ data. This test was also reversed between season $(s1)$ and $(s2)$ to give two test results for season-to-season comparisons.

The results from applying Equation (11) are shown in Table 5 and Tables B.1 through B.11 in Appendix B for the three adopted heterogeneity models. As can be seen in the tables, season $(s1)$ data represent four different datasets each year: (1) data collected from November to January; (2) data collected from February to April; (3) data collected from May to July; and (4) data collected from August to October. On the other hand, season $(s2)$ data represent the same four datasets excluding season $(s1)$ dataset for the same year.

For instance, if season $(s1)$ dataset represents data collected from November to January 2018, then season $(s2)$ data represent three datasets: (1) data collected from February to April; (2) data collected from May to July; and (3) data collected from August to October.

The test results show that there is no difference in water consumption models when using the data collected from November to January and from February to April. Similarly, there is no difference in the water consumption models when using data collected from May to July and August to October. A series of LR tests were then performed (with degrees of freedom equal to the number of estimated parameters of each model) to test the null hypothesis that separately estimated water consumption

Table 5 | Confidence levels in percentage and (degrees of freedom in parentheses) and [X^2 in brackets] results from applying random parameters ordered probit model with heterogeneity in means for 2018 data (refer to Equation (11))

Year	Seasons	Nov–Jan	Feb–Apr	May–Jul	Aug–Oct
2018	Nov–Jan	–	0.00(0)[0]	99.99(11)[79]	99.99(10)[70]
	Feb–Apr	0.00(0)[0]	–	99.99(9)[74]	99.99(8)[74]
	May–Jul	99.99(10)[71]	99.99(9)[78]	–	0.00(0)[0]
	Aug–Oct	99.99(8[76]	99.99(11)[77]	0.00(0)[0]	–
2019	Nov–Jan	–	0.00(0)[0]	99.99(8)[175]	99.99(9)[179]
	Feb–Apr	0.00(0)[0]	–	99.99(10)[177]	99.99(8)[179]
	May–Jul	99.99(11)[179]	99.99(9)[172]	–	0.00(0)[0]
	Aug–Oct	99.99(11)[179]	99.99(8)[174]	0.00(0)[0]	–
2020	Nov–Jan	–	0.00(0)[0]	99.99(11)[124]	99.99(9)[135]
	Feb–Apr	0.00(0)[0]	–	99.99(10)[137]	99.99(9)[141]
	May–Jul	99.99(8)[134]	99.99(10)[124]	–	0.00(0)[0]
	Aug–Oct	99.99(8)[125]	99.99(8)[125]	0.00(0)[0]	–
2021	Nov–Jan	–	0.00(0)[0]	99.99(8)[84]	99.99(11)[73]
	Feb–Apr	0.00(0)[0]	–	99.99(8)[80]	99.99(11)[71]
	May–Jul	99.99(9)[78]	99.99(10)[83]	–	0.00(0)[0]
	Aug–Oct	99.99(10)[76]	99.99(9)[81]	0.00(0)[0]	–

models for each season would be temporally stable over the 4-year study period, based on Equation (12):

$$X^2 = -2 * [LL_{s1t2,s1t1} - LL_{s1t1}] \tag{12}$$

where $LL_{s1t2,s1t1}$ is the log-likelihood at convergence for each model (for a given season ($s1$)) estimated using the converged parameters from specific year ($t2$) data on different year ($t1$) data for that season, and LL_{s1t1} is the log-likelihood at convergence of each model using year ($t1$) data. This test was also reversed between year ($t1$) and ($t2$) to give two test results for year-to-year comparisons.

The results from applying Equation (12) are shown in Table 6 and in Tables B.12 through B.22 in Appendix B for the three adopted heterogeneity models.

Table 6 | Confidence levels in percentage and (degrees of freedom in parentheses) and [X^2 in brackets] results from applying random parameters ordered probit model with heterogeneity in means for 2018 data (refer to Equation (12))

Year	Seasons	Nov–Jan	Feb–Apr	May–Jul	Aug–Oct
2018	Nov–Jan	–	–	–	–
	Feb–Apr	–	–	–	–
	May–Jul	–	–	–	–
	Aug–Oct	–	–	–	–
2019	Nov–Jan	0.00(0)[0]	–	–	–
	Feb–Apr		0.00(0)[0]	–	–
	May–Jul	–		0.00(0)[0]	
	Aug–Oct	–	–	–	0.00(0)[0]
2020	Nov–Jan	0.00(0)[0]	–	–	–
	Feb–Apr	–	0.00(0)[0]	–	–
	May–July	–	–	0.00(0)[0]	–
	Aug–Oct	–	–	–	0.00(0)[0]
2021	Nov–Jan	0.00(0)[0]	–	–	–
	Feb–Apr	–	0.00(0)[0]	–	–
	May–Jul	–	–	0.00(0)[0]	–
	Aug–Oct	–	–	–	0.00(0)[0]

The results show that across the analysis period, there is no difference in the water consumption models developed using data from the same season. Based on the results obtained from the two tests above, the data were divided into two sets. The first set includes data collected in the period from November to April (winter dataset), and the second set contains data collected in the period from May to October (summer dataset), for the entire analysis period. Next, a global test was performed, based on Equation (13), to statistically assess whether the water consumption models differ significantly for the two datasets:

$$X^2 = -2*[LL_T - LL_{\text{Nov-Apr}} - LL_{\text{May-Oct}}] \tag{13}$$

where LL_T is the log-likelihood of the converged model using all the data from the 4-year period, $LL_{\text{Nov-Apr}}$ is the log-likelihood of the converged model for the first dataset, and $LL_{\text{May-Oct}}$ is the log-likelihood of the converged model for the second dataset. The test results reject the null hypothesis that the parameters are the same for the two datasets, with 99.99% confidence for the three adopted models.

4.5. Model evaluation procedure

For the two datasets (summer and winter datasets), the statistical performance of the three adopted models was evaluated using four widely used statistical metrics in the literature: LR, Adj-R^2, the AIC, and the BIC.

The LR test was used to compare the developed models based on their overall statistical fit. The X^2 distributed test statistic was used for this purpose, with degrees of freedom equal to the number of parameters being estimated, as described by the following equation:

$$LR = -2*[LL(\beta_0) - LL(\beta_1)] \tag{14}$$

where $LL(\beta_0)$ is the log-likelihood at convergence of the random parameters ordered probit model with heterogeneity in means (i.e., restricted model) and $LL(\beta_1)$ is the log-likelihood at convergence for each of the remaining two models (i.e., unrestricted models).

The Adj-R^2 statistical metric was used to evaluate the overall performance of each model, based on the following equation (Long & Long 1997):

$$R^2_{\text{adj}} = 1 - \frac{LL(\beta_1) - K}{LL(\beta_0)} \tag{15}$$

where $LL(\beta_0)$ is the restricted log-likelihood at convergence, $LL(\beta_1)$ is the unrestricted log-likelihood at convergence, and K is the number of all estimated parameters in the model (including the intercept and threshold).

The AIC (Akaike 1973) and BIC (Schwarz 1978) statistical metrics were also used to evaluate the overall performance of the models. Both criteria assess each model's performance by minimizing the quantity $-2*l + A(n)*P$, such that as l is the likelihood function for each model, $A(n)$ is some function of the sample size n, and P is the number of parameters in the model.

Despite the fact that the two criteria minimize the same function, they use different values for $A(n)$. For the AIC, $A(n)$ is a constant equal to two, but for the BIC, it is a function of n equal to $\ln(n)$. According to Dziak *et al.* (2020), the BIC is the most commonly used technique in the literature because it is sample size dependent, increasing the likelihood of selecting the correct model as the sample size increases.

5. RESULTS AND DISCUSSION

The application of the methodological approach and the subsequent evaluation procedure described in Section 4 are visually summarized in Figure 1.

This section discusses the outcomes of using the methodology proposed in the current study to forecast seasonal water consumption in terms of pricing blocks (variable X14 in Table 3) over a 4-year period (2018–2021). The analysis was performed on two analysis periods, based on the results of the temporal stability tests discussed in section 4.4. For the entire 4-year period, the first analysis period contains data collected from November to April (winter period), while the second analysis period contains data collected from May to October (summer period).

Figure 1 | Methodological approach and evaluation procedure flowchart.

Tables 7–12 provide summary statistics for the results obtained from applying the three adopted models for the two analysis periods. For all the models, different distributions were investigated to represent the variations of the estimated parameters across observations, with the normal distribution providing the best fit. In addition, a trial-and-error procedure was applied to select statistically significant variables for each model.

5.1. Random parameters ordered probit model with heterogeneity in means

Table 7 provides summary statistics for the results obtained from applying the random parameters ordered probit model with heterogeneity in means over the two analysis periods, and Table 8 provides summary statistics for the resulting marginal effects. The entries in Tables 6 and 7 show that all parameters were estimated with a CL greater than 95%. Table 7 shows that seven factors with fixed parameters have a significant impact on water consumption for both periods, with three factors shared between the two.

These factors are classified into four main groups: characteristics of dwelling unit (i.e., garden area, number of bathrooms in the household, and total capacity of household's water tanks and wells (m^3)), water supply service (household main water source), water-use behavior (average monthly payment to compensate for water shortage (JOD)), and attitudes related to water conservation (awareness of water scarcity in Jordan and utilization of collected rainwater).

Table 7 | Results – applying random parameters ordered probit model with heterogeneity in the means

Variables description	Coefficient (*P*-value)	
Nonrandom parameters	**Winter analysis period (Nov–Apr)**	**Summer analysis period (May–Oct)**
Garden area (m^2) (50–100, 100–300, 300–600, above 600)	–	0.10524 (0.0064)
Number of bathrooms in the household (1, 2, 3, ≥4)	0.2407 (0.0012)	0.54027 (0.0007)
Total capacity of household water tanks and wells (m^3) (<2, 2–5, 5–8, > 8)	0.20961 (0.0149)	0.4518 (0.00198)
Household water source (multiple choices are allowed) (water company, water wells, water tanks)	–	−0.2105 (0.0019)
Utilization of the collected rainwater (I do not collect rainwater; I collect rainwater for at least one purpose)	−0.3511 (0.0341)	–
Average monthly payment to compensate for water shortage (JOD) (0–10, 10–20, 20–30, 30–40, 40, or above)	–	0.1472 (0.0041)
Awareness on water scarcity in Jordan (aware, not aware)	−0.4578 (0.0034)	−0.6314 (0.00087)
Means for random parameters	**Winter analysis period (Nov–Apr)**	**Summer analysis period (May–Oct)**
Constant	−0.5239 (0.0129)	−0.4329 (0.0019)
Sufficient water is received weekly (yes, no)	0.2781 (0.0095)	–
Number of the family members living in the household (1–3, 3–6, 6–9, > 9)	–	0.3618 (0.0023)
Standard deviation for random parameters	**Winter analysis period (Nov–Apr)**	**Summer analysis period (May–Oct)**
Constant	0.47013 (0.014)	0.5370 (0.0084)
Sufficient water is received weekly (yes, no)	0.5928 (0.0072)	–
Number of the family members living in the household (1–3, 3–6, 6–9, > 9)	–	0.4912 (0.00149)
Heterogeneity in the means of random parameters	**Winter analysis period (Nov–Apr)**	**Summer analysis period (May–Oct)**
Area of dwelling unit (m^2) (50–100, 100–150, 150–200, 200, or above): Sufficient water is received weekly (yes, no)	−0.2034 (0.01361)	–
Ownership type (own, rent): Number of the family members living in the household (1–3, 3–6, 6–9, > 9)	–	0.19539 (0.01469)
Threshold parameters for probabilities	**Winter analysis period (Nov–Apr)**	**Summer analysis period (May–Oct)**
Mu(01)	1.4386 (0)	1.3482 (0)

The results in both tables show that for the two periods, the water consumption is directly proportional to an increase in the value of the household structure variables. For example, increasing the capacity of the household's water well by 1 m^2 or adding one additional bathroom to the household would increase the probability of receiving a water bill of 30 JOD or higher (maximum price block) by 0.1283 and 0.1320, respectively, for the winter period and by 0.1419 and 0.1191, respectively, for the summer period.

However, increasing the garden area by 1 m^2 has no impact on water consumption in the winter period but increases the probability of receiving a water bill of 30 JOD or higher by 0.0237 in the summer period. A similar relationship was observed for the variable of average monthly payment to compensate for the water scarcity, where an increase by 1 JOD increases the probability of receiving a water bill of 30 JOD or more by 0.006 in the summer period while having no impact in the winter period.

Conversely, the water consumption was found to be inversely proportional to the type of primary household water source, the utilization of the collected rainwater, and the awareness of water scarcity in Jordan. The marginal effect values for the

Table 8 | Marginal effect from applying random parameters ordered probit model with heterogeneity in the means

Variables description	Marginal effect (*P*-value)					
	Winter analysis period (Nov–Apr)			Summer analysis period (May–Oct)		
Nonrandom parameters	Prob (Price Block 1)	Prob (Price Block 2)	Prob (Price Block 3)	Prob (Price Block 1)	Prob (Price Block 2)	Prob (Price Block 3)
Number of bathrooms in the household (1, 2, 3, ≥4)	−0.136 (0)	0.0039 (0.001)	0.1320 (0.002)	−0.1341 (0.001)	0.015 (0.0017)	0.1191 (0.004)
Total capacity of household water tanks and wells (m^3) (<2, 2–5, 5–8, > 8)	−0.241 (0)	0.1126 (0.002)	0.1283 (0)	−0.1601 (0.006)	0.0182 (0.0017)	0.1419 (0.001)
Household main water source (multiple choices are allowed) (water company, rainwater storage wells at residence, water delivery by independent suppliers)				0.0154 (0.005)	−0.0015 (0.006)	−0.0139 (0.001)
Utilization of the collected rainwater (I don't collect rainwater; I collect rainwater for at least one purpose)	0.032 (0.003)	−0.0091 (0.004)	−0.0232 (0.004)	–	–	–
Garden area (m^2) (50–100, 100–300, 300–600, above 600)	–	–	–	−0.0127 (0.001)	−0.011 (0.005)	0.0237 (0.002)
Average monthly payment to compensate for water shortage (JOD) (0–10, 10–20, 20–30, 30–40, 40, or above)	–	–	–	−0.012 (0.003)	0.0068 (0.001)	0.006 (0.001)
Awareness on water scarcity in Jordan (aware, not aware)	0.129 (0)	−0.087 (0)	−0.042 (0)	0.152 (0)	−0.021 (0)	−0.1313 (0)
Means for random parameters	**Winter analysis period (Nov–Apr)**			**Summer analysis period (May–Oct)**		
Constant	–	–	–	–	–	–
Sufficient water is received weekly (yes, no)	−0.107 (0)	0.0503 (0.002)	0.057 (0.003)	–	–	–
Number of the family members living in the household (1–3, 3–6, 6–9, > 9)	–	–	–	−0.1532 (0.001)	0.134 (0.001)	0.018 (0)

three variables listed in Table 8 show that consumers who rely on a water company as their primary water source are 0.0139 less likely to receive a water bill of 30 JOD or more in the summer period than those who rely on household's water wells as their primary water source.

This can be attributed to the fact that relying on fluctuating water sources, such as a household rainwater storage well or purchasing water from an independent source, makes consumers feel less secure about their water storage and the availability of water in the event of a shortage, resulting in consumers storing and/or buying more water (from independent sources) than they actually need.

Furthermore, consumers who collect rainwater and utilize it for different purposes during the winter period are less likely to receive a water bill of 30 JOD or more by 0.0232 than those who do not collect rainwater. Finally, in the winter and summer periods, consumers who are unaware of Jordan's water scarcity are 0.042 and 0.1313, respectively, more likely to receive a water bill of 30 JOD or more.

In addition, the model shows two variables with random parameters; the variables are *sufficient water received weekly* and *number of family members living in the household*. The marginal effect values for the former variable show that customers whose water usage exceeds the amount of water provided by the water company during the winter period are more likely to receive a water bill of 30 JOD or more by 0.057, whereas the marginal effects for the latter variable show that increasing the family size in the household by one member increases the probability of receiving a water bill of 30 JOD or more during the summer period by 0.018.

It should be noted that for the two variables with random parameters, the model estimates a unique parameter value for each observation, as shown in Figure 2, which can be related to unobserved traits for each individual. For instance, the

Table 9 | Summary statistics for the random threshold random parameter hierarchical ordered probit model

Variables description	Coefficient (*P*-value)	
Nonrandom parameters	**Winter analysis period (Nov–Apr)**	**Summer analysis period (May–Oct)**
Number of bathrooms in the household (1, 2, 3, ≥4)	0.2711 (0.0024)	0.4942 (0.0004)
Total capacity of household water tanks and wells (m3) (<2, 2–5, 5–8, > 8)	0.20961 (0.0149)	0.4954 (0.0094)
Utilization of the collected rainwater (I do not collect rainwater; I collect rainwater for at least one purpose)	−0.4051 (0.0041)	–
Household main water source (multiple choices are allowed) (water company, rainwater storage wells at residence, water delivery by independent suppliers)	–	−0.1874 (0.0126)
Utilization of the recycled water (I do not utilize recycled water; I utilize recycled water for at least one purpose)	–	−0.3519 (0.00961)
Garden area (m^2) (50–100, 100–300, 300–600, above 600)	–	0.4278 (0.0011)
Awareness on water scarcity in Jordan (aware, not aware)	−0.0751 (0.00932)	−0.141 (0.00735)
Means for random parameters	**Winter analysis period (Nov–Apr)**	**Summer analysis period (May–Oct)**
Constant	−0.2154 (0.0133)	−0.2941 (0.0049)
Frequency of receiving water from water company (assuming duration of 6 h for each delivery) (at most 1, 2, 3 times, or above per week)	0.2781 (0.0015)	–
Area of dwelling unit (m^2) (50–100, 100–150, 150–200, 200, or above)	–	0.4416 (0.0044)
Standard deviation for random parameters	**Winter analysis period (Nov–Apr)**	**Summer analysis period (May–Oct)**
Constant	−0.3147 (0.003)	−0.4188 (0.001)
Frequency of receiving water from water company (assuming duration of 6 h for each delivery) (at most 1, 2, 3 times, or above per week)	0.3514 (0)	–
Area of dwelling unit (m^2) (50–100, 100–150, 150–200, 200, or above)	–	0.47851 (0)
Intercept for random threshold	**Winter analysis period (Nov–Apr)**	**Summer analysis period (May–Oct)**
Mu	1.1248 (0)	1.2157 (0)
Standard deviation for random threshold	**Winter analysis period (Nov–Apr)**	**Summer analysis period (May–Octo)**
Mu	0.4124 (.001)	0.5147 (.001)
Threshold covariates parameters	**Winter analysis period (Nov–Apr)**	**Summer analysis period (May–Oct)**
Sufficient water is received weekly (yes, no)	−0.00014 (0.0014)	–
Average monthly family income (JOD) (80–200, 200–400, 400–800, 800, or above)	–	−0.00041 (0.0189)

variable of *sufficient water received weekly* has a mean of 0.2781 and standard deviation of 0.5928, indicating, with a normal distribution, that this variable is positive for 68% of the observations and negative for 32% of the observations.

Similarly, the number of family members living in the household has a mean of −0.3618 and standard deviation of 0.4912, indicating, with a normal distribution, that this variable is negative for 77% of the observations and negative for 23% of the observations. These variations cannot be captured using the traditional regression methods developed in the literature and explain the contradiction in the impact of different variables on water consumption found in the literature and discussed in the background section (section 2).

Table 10 | Marginal effect from applying random threshold random parameter hierarchical ordered probit model

Variables description	Marginal effect (*P*-value)					
	Winter analysis period (Nov–Apr)			Summer analysis period (May–Oct)		
Nonrandom parameters	Prob (Price Block 1)	Prob (Price Block 2)	Prob (Price Block 3)	Prob (Price Block 1)	Prob (Price Block 2)	Prob (Price Block 3)
Number of bathrooms in the household (1, 2, 3, ≥4)	−0.1152 (0.0017)	0.0188 (0.0008)	0.0964 (0.0014)	−0.1009 (0.0042)	0.0085 (0.006)	0.0924 (0.0037)
Total capacity of household water tanks and wells (m^3) (<2, 2–5, 5–8, > 8)	−0.18 (0)	0.0837 (0.0087)	0.0963 (0.0019)	−0.1571 (0.0005)	0.047 (0.002)	0.1101 (0.0057)
Utilization of the collected rainwater (I do not collect rainwater; I collect rainwater for at least one purpose)	0.15 (0.0005)	−0.0501 (0.0017)	−0.0999 (0.005)	–	–	–
Household main water source (multiple choices are allowed) (water company, rainwater storage wells at residence, water delivery by independent suppliers)	–	–	–	0.1614 (0.0044)	−0.0608 (0.0063)	−0.1006 (0.0011)
Utilization of the recycled water (I do not utilize recycled water; I utilize recycled water for at least one purpose)	–	–	–	0.1579 (0.0006)	−0.0593 (0.0051)	−0.0986 (0.0074)
Garden area (m^2) (50–100, 100–300, 300–600, above 600)	–	–	–	−0.1403 (0.001)	0.1118 (0.0065)	0.0285 (0.0046)
Awareness on water scarcity in Jordan (aware, not aware)	0.1247 (0.0012)	−0.0128 (0.004)	−0.117 (0.0006)	0.4516 (0.004)	−0.0826 (0.007)	−0.717 (0.0001)
Means for random parameters	**Winter analysis period (Nov–Apr)**			**Summer analysis period (May–Oct)**		
Constant	–	–	–	–	–	–
Frequency of receiving water from water company (assuming duration of 6 h for each delivery) (at most 1, 2, 3 times, or above per week)	−0.1305 (0.002)	0.029 (0.0054)	0.1015 (0.0017)	–	–	–
Area of dwelling unit (m^2) (50–100, 100–150, 150–200, 200, or above)	–	–	–	−0.1801 (0.0039)	0.097 (0.0062)	0.0831 (0.0056)

To the best of the authors' knowledge, the majority of the prior studies in the literature have reported the average impact of different variables on water consumption without considering the variation of such impact across observations. It is important to understand the variations in actual water consumption patterns of consumers as a step towards preparing effective management plans to reduce water consumption.

Furthermore, the results in Table 7 indicate that the random parameter for *sufficient water received weekly* is a function of the variable for the *area of the dwelling unit* with its value decreasing as the value of the variable for dwelling unit area increases. In contrast, the random parameter for the variable regarding the number of family members living in the household is a function of the ownership type, with its value decreasing as the value of the variable for ownership type increases.

For instance, increasing the number of family members in owned dwelling units results in a higher increase in water consumption than increasing the number of family members in rented dwelling units.

5.2. Random threshold random parameter hierarchical ordered probit model

Table 9 provides summary statistics for the results obtained from applying the random threshold random parameter hierarchical ordered probit model over the two analysis periods. Table 10 provides summary statistics for the resulting marginal effects.

The entries in both tables show that all parameters have been estimated with a CL greater than 95%. Table 9 shows that seven variables with fixed parameters have a significant impact on water consumption for both periods. The majority of these variables were also selected in Table 7 by the random parameter ordered probit model with heterogeneity in means, except

Table 11 | Summary statistics for latent class ordered probit model with class probability function

Variables description	Coefficient (*P*-value) Winter analysis period (Nov–Apr) Latent class 1	Winter analysis period (Nov–Apr) Latent class 2	Summer analysis period (May–Oct) Latent class 1	Summer analysis period (May–Oct) Latent class 2
Constant	−0.27738 (0.0132)	3.78407 (0.041)	−0.7283 (0.0141)	4.0407 (0.011)
Garden area (m^2) (50–100, 100–300, 300–600, above 600)	–	–	0.3411 (0.003)	−0.5719 (0.0093)
Total capacity of household water tanks and wells (m^3) (<2, 2–5, 5–8, > 8)	0.4133 (0.0044)	−0.5071 (0.0031)	0.6844 (0.0057)	−0.8547 (0.0013)
Utilization of water-conservation fixtures (I don't utilize water-conservation fixtures, I utilize water-saving fixtures in at least one location)	–	–	−0.0718 (0.0191)	0.2611 (0.0126)
Utilization of the collected rainwater (I do not collect rainwater; I collect rainwater for at least one purpose)	−0.0941 (0.0012)	0.4381 (0.0016)	–	–
Awareness on water scarcity in Jordan (aware, not aware)	−0.2815 (0.0086)	1.4741 (0.00954)	−0.21181 (0.00716)	1.84437 (0.00754)
Latent class probability variables	**Winter analysis period (Nov-April)**		**Summer analysis period (May–Oct)**	
Constant	1.37706 (0.0106)	Fixed parameter	1.4067 (0.006)	Fixed Parameter
Utilization of the collected rainwater (I do not collect rainwater; I collect rainwater for at least one purpose)	−0.0133 (0.0028)	Fixed parameter	–	–
Garden area (m^2) (50–100, 100–300, 300–600, above 600)	–	–	−0.001 (0.0025)	Fixed Parameter
Threshold parameters for probabilities	**Winter analysis period (Nov–Apr)**		**Summer analysis period (May–Oct)**	
Mu(01)	1.0262 (0)	2.9427 (0.008)	1.162 (0)	2.471 (0)

Table 12 | Marginal effect from applying latent class ordered probit model with class probability function

Variables description / Nonrandom parameters	Marginal effect (*P*-value) Winter analysis period (Nov–Apr) Prob (Price Block 1)	Winter Prob (Price Block 2)	Winter Prob (Price Block 3)	Summer analysis period (May–Oct) Prob (Price Block 1)	Summer Prob (Price Block 2)	Summer Prob (Price Block 3)
Garden area (m^2) (50–100, 100–300, 300–600, above 600)	–	–	–	−0.1043 (0.0039)	0.0074 (0.0037)	0.0969 (0.0027)
Capacity of total capacity of household water tanks and wells (m^3) (<2, 2–5, 5–8, > 8)	−0.1639 (0.0006)	0.0613 (0.0001)	0.1026 (0.0002)	−0.1844 (0.0009)	0.1836 (0.0023)	0.0008 (0.0003)
Utilization of water-conservation fixtures (I don't utilize water-conservation fixtures, I utilize water-saving fixtures in at least one location)	–	–	–	0.1664 (0.0009)	−0.0708 (0.0005)	−0.0956 (0.0002)
Utilization of the collected rainwater (I do not collect rainwater; I collect rainwater for at least one purpose)	0.1438 (0.0026)	−0.0135 (0.0006)	−0.1303 (0.0009)	–	–	–
Awareness on water scarcity in Jordan (aware, not aware)	0.1888 (0.0074)	−0.1679 (0.0007)	−0.0209 (0.0007)	0.1295 (0.0018)	−0.1146 (0.0038)	−0.0149 (0.0018)

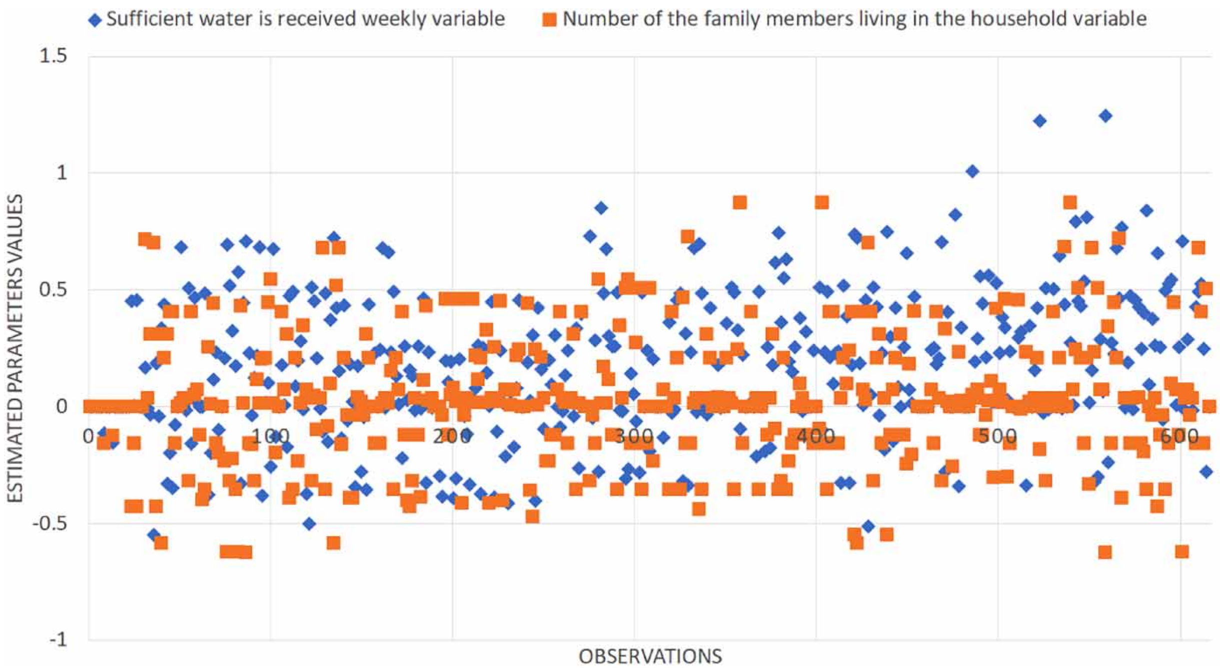


Figure 2 | Estimated parameters value for each observation from applying random parameters ordered probit model with heterogeneity in the means.

for the utilization of the recycled water. The marginal effects values for this variable, as listed in Table 10, indicate that consumers who recycle water and utilize it for different purposes during the summer period are less likely to receive a water bill of 30 JOD or more by 0.0986 than those who do not recycle water during the same period.

In addition, the model shows two variables with random parameters: *frequency of receiving water from the water company* and the *area of the dwelling unit*. The marginal effect values for the former variable show that consumers who received water from the water company three or more times per week (during the winter period) had a 0.1015 higher probability of receiving a water bill of 30 JOD or more, whereas the marginal effects for the latter variable show that increasing the area of the dwelling unit by 1 m^2 increases the probability of receiving a water bill of 30 JOD or more during the summer period by 0.0831.

For the two variables with random parameters, the model estimated a unique parameter value for each observation, as shown in Figure 3, which can be related to the unobserved traits for each individual. For instance, the variable for frequency of receiving water from the water company has a mean of 0.2781 and a standard deviation of 0.3514, indicating, with a normal distribution, that this variable is positive for 79% of the observations and negative for 21% of the observations. Similarly, the variable for the area of the dwelling unit has a mean of 0.4416 and a standard deviation of 0.47851, indicating, with a normal distribution, that this variable is positive for 82% of the observations and negative for 18% of the observations.

Finally, the model shows that the estimated threshold varies across observations as a function of two variables: *sufficient water is received weekly* and *average monthly family income*. Table 9 shows that the threshold determined for the winter period is inversely proportional to an increase in the value of the first variable (sufficient water is received weekly), with an intercept value (t_k) of 1.048 and standard deviation (γ_k) of 0.9124. This finding indicates that, with a normal distribution, the threshold variable is inversely proportional to an increase in the variable value for 87% of the observations and positively proportional to an increase in the variable value for 13% of the observations.

Similarly, the threshold determined for the summer period is negatively proportional to an increase in the value of the latter variable (*average monthly family income*), with an intercept value (t_k) of 1.0157 and standard deviation (γ_k) of 0.947; indicating that, with a normal distribution, the threshold variable is inversely proportional to an increase in the variable value for 86% of the observations and positively proportional to an increase in the variable value for 14% of the observations.

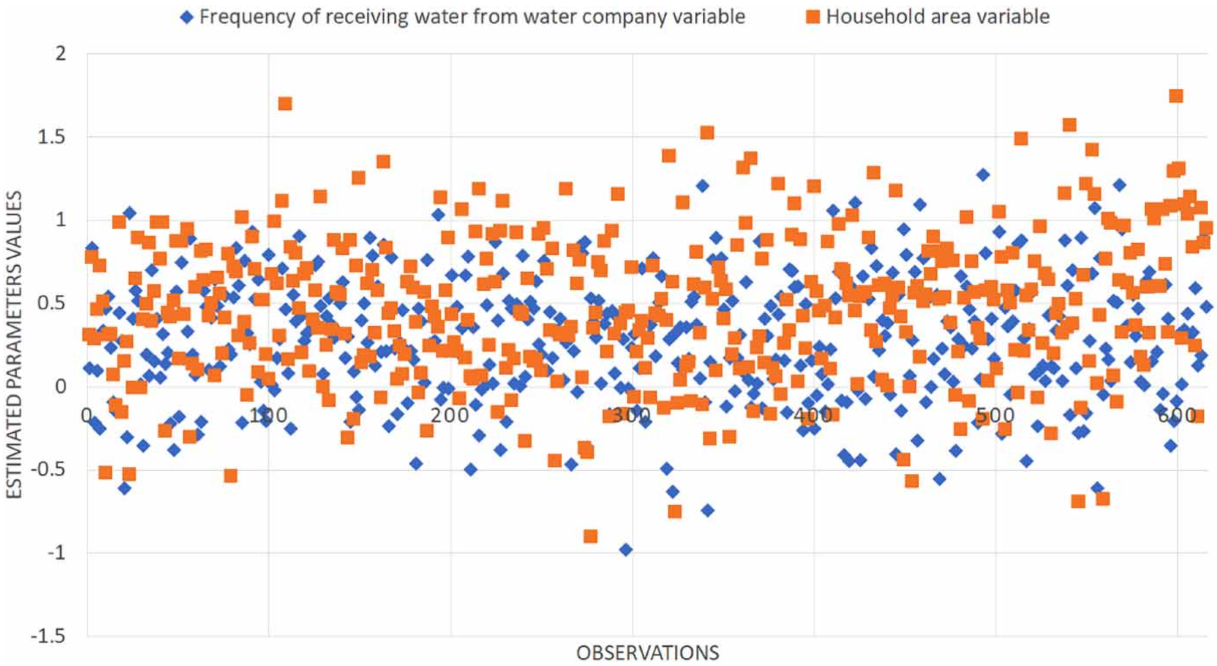


Figure 3 | Estimated parameters value for each observation from applying random threshold random parameter hierarchical ordered probit model.

5.3. Latent class ordered probit model

The latent class ordered probit model was initially applied with a class probability function. The results are compared with those obtained with fixed class probabilities; the former produces a superior match, and its findings are presented in Table 11, with the marginal effect values presented in Table 12.

The results in Tables 10 and 11 demonstrate that two latent classes are statistically significant for the two periods; thus, they were employed for model evaluation. For the two classes, five variables impact the water consumption level. These variables can be divided into two main groups: household structure data (*garden area* and capacity of the household's water well) and data related to water-conservation attitudes (*utilization of water-conservation fixtures, utilization of the collected rainwater,* and *awareness of water scarcity in Jordan*).

Among these variables, the variable for the influence of the *utilization of water-conservation fixtures* on water consumption was not captured by the two other models. The marginal effect values for this variable indicate that utilizing water-conservation fixtures in at least one location in the household reduces the probability of receiving a water bill of 30 JOD or more during the summer period by 0.0956. In addition, the first latent class probabilities for both analysis periods are inversely proportional to two covariates: the utilization of the collected rainwater for the winter period and the garden area for the summer period.

5.4. Model evaluation

Finally, the performance of the three models was compared using the approach outlined in the model evaluation procedure section (section 6), which includes the use of four statistical indicators, as shown in Table 13. First, the LR test was applied to statistically investigate the difference in the performance of the three models using both winter and summer datasets. The outcomes in Table 13 demonstrate that the difference in the performance of the three models for both analysis periods is statistically significant, with a 99.99% CL. Next, AIC, BIC, and Adj-R^2 metrics were applied to compare the performance of the three models using both winter and summer datasets.

The results in Table 13 for the three statistical indicators, over the two analysis periods, show that the random parameter ordered probit model with heterogeneity in means has the highest Adj-R^2 value and the lowest AIC and BIC values, indicating that this model outperforms the other two models statistically over the two periods (winter and summer periods). However,

Table 13 | Comparing the models' performance

Model	Winter analysis period (Nov–Apr)				Summer analysis period (May–Oct)			
	CL% (DOF) [X^2]	Adj-R^2	AIC	BIC	CL% (DOF) [X^2]	Adj-R^2	AIC	BIC
Random parameter ordered probit model with heterogeneity in the means	–	0.948	1.23	1.31	–	0.951	1.24	1.33
Random threshold random parameters Hierarchical ordered probit model	99.99 (1) [65.40]	0.921	1.26	1.34	99.99 (1) [68.33]	0.929	1.26	1.36
Latent class ordered probit model	99.99 (3) [44.82]	0.934	1.25	1.32	99.99 (4) [46.78]	0.940	1.25	1.33

the high Adj-R^2 value for the three models over the two analysis periods (i.e., Adj-R^2 values greater than 0.9) indicates that the three models' performance is statistically acceptable.

Finally, the use of these three alternative modeling approaches in this study can be justified by the fact that each model addresses a different type of unobserved heterogeneity in the observed data (as discussed in the methodological approach section and as illustrated by Table 4), allowing us to account for different potential unobserved heterogeneity in estimating water consumption and to validate the temporal stability of the observed data.

6. IMPLICATION FOR POLICY MAKERS

This study investigated unobserved heterogeneity in factors affecting household water consumption under data temporal instability and demonstrated the impact of these factors by applying three distinctive ordered probit model approaches within the context of household water consumption three main cities in northern Jordan. Understanding the factors affecting household water consumption patterns could help water utilities to implement appropriate water management policies and regulations.

The estimated results of the three applied models revealed that variables related to dwelling unit characteristics (i.e., garden area and total capacity of household water tanks and wells) and those related to water-conservation attitude (i.e., collecting and utilizing rainwater and awareness of water scarcity in Jordan) were the most commonly chosen. First, the results of the three models show that, during summer seasons, water consumption increases as the household garden area increases, which is attributed to the irrigation during summer in household gardens.

This finding is consistent with previous research (Renwick & Green 2000; Guhathakurta & Gober 2007). In addition, the results of the three models revealed the fact that water consumption is directly proportional to the total capacity of household water tanks and wells during both summer and winter seasons. The practical policy implication of these findings is that combining structured water pricing policy based on dwelling unit characteristics with public education campaigns that can better inform individuals on changing their water-use habits is likely to produce better results for both the water utility and households.

The study findings also show that the water-conservation beliefs of consumers may affect their water consumption patterns. In particular, the results from the three models show that water consumption is inversely proportional to consumers' beliefs about the importance of collecting and utilizing rainwater and their awareness of water scarcity in Jordan.

The practical policy implication from these findings is that establishing long-term programs to educate individuals about the water scarcity situation in Jordan and about best practices to reuse alternate water sources (for instance, collecting and utilizing rainwater) and reduce their water consumption (for instance, using water-conservation fixtures) has the potential to change daily water practices, and thus reduce water consumption in the long term.

The Jordanian government has implemented several actions to inform people about the country's water scarcity and to educate them about sustainable practices for reducing their consumption. Aside from the national water plan established by the government for 2008–2022 (explained in the introduction section of this paper), the Ministry of Water and Irrigation (MWI) and Miyahuna Water Company (MWI) – the primary water company in the capital city of Amman, Jordan, ran a water-conservation campaign in 2019 called *Water Efficiency and Public Information for Action*.

The campaign aimed to educate people about Jordan's water crisis. Following the campaign, MWI conducted a survey to assess the campaign's impact on the targeted population, which revealed that people's awareness of Jordan's water crisis had increased, with 24% of the targeted population requesting a leakage check at their residences.

In addition, the study's findings revealed that the impact of other variables related to dwelling unit characteristics (i.e., the number of family members living in the household and the area of the dwelling unit) and those related to water supply service (i.e., whether sufficient water is received weekly and the frequency of receiving water from the water company) on water consumption varies across individuals, which can be attributed to unobserved traits for each individual.

These findings have a practical policy implication in that identifying specific groups of households with similar characteristics and specifying different consumption patterns for each of these groups is likely to result in better water consumption estimates, allowing for the development of more effective long-term demand-management strategies.

Finally, the results revealed that water-conservation behavior does not occur on the spur of the moment and does not change from one day to the next. It is rather based on a latent trait known as water-conservation attitude. This underlying disposition is manifested through various behavioral means, which may differ from person to person. As a result, people with the same level of water-conservation attitude may not take the same individual actions depending on their other traits (as listed in Table 3) (Zietlow 2016; Liu *et al.* 2022).

Therefore, it is crucial for decision makers to understand how different consumers' traits influence their water-conservation attitude as a first step toward developing effective management plans to reduce water consumption (Rosenberg *et al.* 2008; Potter *et al.* 2010; Potter & Darmame 2010; Zietlow 2016).

7. DISCUSSION

Understanding the influence of socioeconomic factors on household water consumption behavior is crucial for developing efficient water resource management strategies. The current study aimed to investigate unobserved heterogeneity in water consumption data by considering data stability over the analysis periods. For this purpose, a two-step procedure was proposed. First, a series of LR tests were performed to examine data stability across different seasons throughout the analysis periods. Next, three distinct ordered probit models were applied to forecast household water consumption while considering potentially complex unobserved heterogeneities in the data, as well as data temporal stability.

The proposed approach was applied to the results of a household survey conducted in three main cities in Northern Jordan: Irbid, Jerash, and Ajloun. The survey collected data on five main aspects: household sociodemographic factors, data related to the dwelling unit, water supply service data, water-use behavior data, and data related to water-conservation attitudes. To the best of the authors' knowledge, no previous study has investigated the unobserved heterogeneity in factors influencing water consumption in a temporal context.

However, exploring these factors is required to improve prediction accuracy and eliminate potential biases in the estimation results, allowing water consumers to identify the factors that have the greatest impact on their water usage and to prepare mitigation plans to reduce their consumption.

The current study has shown that the data were temporally stable over the two analysis periods. The winter period includes data collected between the months of November and April from 2018 to 2021, while the summer period includes data collected between the months of May and October from 2018 to 2021. This finding aligns with those reported by several studies discussed in the literature review section (section 2), including Gato-Trinidad *et al.* (2011), Kenney *et al.* (2008), and Rathnayaka *et al.* (2015, 2017). For each analysis period, the three heterogeneity-ordered probit models were applied to capture different potential unobserved heterogeneity in the data and to allow for an evaluation of the data's temporal complexities in predicting household water consumption.

The study revealed several influential factors on water consumption, with a particular focus on the characteristics of the dwelling unit and data related to water-conservation attitudes, as evident in Table 14. Furthermore, Table 14 includes references to various studies in the literature that have reported similar findings. Despite the shared outcomes between this study and prior research, the proposed approach in this study is different from previous approaches and provides added value. Specifically, this study employs a two-step analysis of water consumption data: (1) an investigation into the temporal stability of the data, and (2) an examination of potential unobserved heterogeneity within each stable temporal segment. The first step involves conducting a series of LR ratio tests, while the second step assumes that water consumption differs among

Table 14 | The commonly selected variables to influence water consumption (across the three models)

Variable expression	Variable description	Studies reporting similar findings
Dwelling unit characteristics variables		
X5	Garden area (m^2) (50–100, 100–300, 300–600, above 600)	Rosenberg *et al.* (2008), Potter & Darmame (2010), Potter *et al.* (2010), Zietlow (2016), Bich-Ngoc *et al.* (2022)
X8	Total capacity of household water tanks and wells (m^3) (<2, 2–5, 5–8, > 8)	Rosenberg *et al.* (2008), Potter & Darmame (2010), Potter *et al.* (2010)
Variables related to water-conservation attitudes		
X22	Utilization of collected rainwater (I do not collect rainwater; I collect rainwater for at least one purpose)	Rosenberg *et al.* (2008), Potter & Darmame (2010), Zietlow (2016), Bich-Ngoc *et al.* (2022), Tzanakakis *et al.* (2020)
X25	Awareness of water scarcity in Jordan (aware, not aware)	Potter & Darmame (2010), Zietlow (2016), Tzanakakis *et al.* (2020), Dang et. al. (2022)

consumers based on their individual characteristics. These two steps are illustrated in Figure 1. Using this two-step approach enables a deeper investigation of the unobserved heterogeneity in data and the interplay between different factors in the context of household water consumption. These findings underscore the importance of investing time and effort in modifying consumers' consumption behavior.

Conducting education programs that inform consumers about the factors affecting their water consumption and educating them on practices that would reduce their water consumption could impact their water-use habits. Similarly, informing consumers about the impact of household garden areas or the total capacity of household water tanks and wells on their water consumption and educating them about possible remedies for reducing such impacts (i.e., relying on rainfed agriculture or optimizing the capacity of household water tanks and wells based on their needs) could promote more efficient water consumption.

In addition, conducting campaigns to raise consumer awareness of water scarcity and the availability of various sustainable practices (for example, using water-saving fixtures or collecting and utilizing rainwater for various purposes in the household) would improve consumers' water consumption patterns.

Moreover, the application of both the random parameters ordered probit model with heterogeneity in means and random thresholds random parameters hierarchical ordered probit model shows that for each period, the estimated parameters for several factors, as well as the estimated thresholds, have varying impacts across observations. In addition, the application of the latent class ordered probit model for each analysis period shows that the data can be divided into two latent classes, with the probability of the variables in the first latent class varying across observations using a class probability function.

8. CONCLUDING REMARKS

The value added from this research lies in its specific focus on unobserved heterogeneity and the implications it holds for policymakers. While it is well understood that factors listed in Table 14 can influence water consumption per household, this study goes beyond that by investigating the unobserved heterogeneity in the dataset. By considering unobserved traits and characteristics of individuals, the study provides a deeper understanding of how these factors interact and impact water consumption patterns.

The significance of this research for policymakers is that it provides empirical evidence on the extent to which unobserved heterogeneity matters in the context of household water consumption. By applying three distinctive ordered probit model approaches, the study demonstrates the impact of these unobserved factors on water consumption. This knowledge is crucial for policymakers as it helps them better comprehend the complexity of water consumption behaviors and make informed decisions in designing effective policies and regulations.

Additionally, the research identifies specific variables related to dwelling unit characteristics and water-conservation attitudes that are particularly influential in determining water consumption patterns. This insight enables policymakers to target these factors when formulating policies and interventions. By focusing on variables such as garden area, the total capacity of household water tanks and wells, collecting and utilizing rainwater, and awareness of water scarcity, policymakers can develop targeted strategies that address these key drivers of water consumption.

Furthermore, the study emphasizes the importance of combining structured water pricing policies with public education campaigns. By aligning water pricing with dwelling unit characteristics and promoting awareness about water scarcity and conservation practices, policymakers can incentivize behavior change and achieve more sustainable water use. This integrated approach, supported by the research findings, offers a practical and holistic solution for policymakers to address water consumption challenges.

By providing empirical evidence, identifying influential variables, and emphasizing the need for targeted policies and integrated approaches, the study equips policymakers with invaluable insights. These insights can guide the development of evidence-based and effective policies, ultimately leading to improved water management, conservation, and sustainability outcomes.

Despite these encouraging results and illustration of a methodological approach that may be extensively applied to consumer and water system studies, the current study has a few limitations. First, household sociodemographic factors such as family income, average age and education were found to be statistically insignificant in this study. Applying data mining techniques, such as clustering, reclassification, and affinity grouping, may provide deeper insight about the statistical significance of these factors.

Also, considering households in Jordan's other regions that are equally or even more severely affected by drought and water limitations (such as the southern area) will give new insights into the factors that impact water consumption for such investigations, other factors, such as differences in the weather between the investigated regions, may be incorporated to enhance the ability of the applied models to explain the data.

In future research, a random parameter ordered probit model with heterogeneity in means and variance, as well as a latent class random parameter ordered probit model, could be explored to account for unobserved heterogeneity in random parameters' variance and unobserved heterogeneity within and across latent classes, respectively.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Acharya, G. & Barbier, E. 2002 Using domestic water analysis to value groundwater recharge in the Hadejia'Jama'are Floodplain, Northern Nigeria. *American Journal of Agricultural Economics* **84** (2), 415–426.

Agence Française de Développement (AFD) 2015 *Improving Water Supply in the Governorates of Irbid, Ajloun and Jerash*. Agence Française de Développement, Amman. Available from: https://www.afd.fr/en/carte-des-projets/improving-water-supply-governorates-irbid-ajloun-and-jerash (accessed 27 May 2022).

Akaike, H. 1973 Information theory and an extension of the maximum likelihood principle. In: *Proc., 2nd Int. Symp. on Information*, Springer, New York, pp. 199–213.

Altunkaynak, A. & Nigussie, T. A. 2017 Monthly water consumption prediction using season algorithm and wavelet transform-based models. *Journal of Water Resources Planning and Management* **143** (6), 04017011.

Alvisi, S., Franchini, M. & Marinelli, A. 2007 A short-term, pattern-based model for water-demand forecasting. *Journal of Hydroinformatics* **9** (1), 39–50.

Aprile, M. C. & Fiorillo, D. 2017 Water conservation behavior and environmental concerns: evidence from a representative sample of Italian individuals. *Journal of Cleaner Production* **159**, 119–129.

Arbués, F., Villanúa, I. & Barberán, R. 2010 Household size and residential water demand: an empirical approach. *Australian Journal of Agricultural and Resource Economics* **54** (1), 61–80.

Atsalakis, G. & Minoudaki, C. 2007 Daily irrigation water demand prediction using adaptive neuro-fuzzy inferences systems (ANFIS). In: *Proceedings of the 3rd IASME/WSEAS International Conference on Energy, Environment, Ecosystems and Sustainable Development (EEESD'07), Agios Nikolaos, Crete Island, Greece, 24–26 July, 2007*. World Scientific and Engineering Academy and Society Press (WSEAS Press), pp. 368–373.

Babel, M. S. & Shinde, V. R. 2011 Identifying prominent explanatory variables for water demand prediction using artificial neural networks: a case study of Bangkok. *Water Resources Management* **25** (6), 1653–1676.

Bai, Y., Wang, P., Li, C., Xie, J. & Wang, Y. 2014 A multi-scale relevance vector regression approach for daily urban water demand forecasting. *Journal of Hydrology* **517**, 236–245.

Bakker, M., Van Duist, H., Van Schagen, K., Vreeburg, J. & Rietveld, L. 2014 Improving the performance of water demand forecasting models by using weather input. *Procedia Engineering* **70**, 93–102.

Bartos, M. D. & Chester, M. V. 2014 The conservation nexus: valuing interdependent water and energy savings in Arizona. *Environmental Science & Technology* **48** (4), 2139–2149.
Beal, C., Stewart, R. A., Spinks, A. & Fielding, K. 2011 Using smart meters to identify social and technological impacts on residential water consumption. *Water Science and Technology: Water Supply* **11** (5), 527–533.
Behboudian, S., Tabesh, M., Falahnezhad, M. & Ghavanini, F. A. 2014 A long-term prediction of domestic water demand using preprocessing in artificial neural network. *Journal of Water Supply: Research and Technology – AQUA* **63** (1), 31–42.
Bhat, C. R. 2003 Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B: Methodological* **37** (9), 837–855.
Bich-Ngoc, N., Prevedello, C., Cools, M. & Teller, J. 2022 Factors influencing residential water consumption in Wallonia, Belgium. *Utilities Policy* **74**, 101281.
Bolorinos, J., Ajami, N. K. & Rajagopal, R. 2020 Consumption change detection for urban planning: monitoring and segmenting water customers during drought. *Water Resources Research* **56** (3), e2019WR025812.
Brentan, B. M., Luvizotto Jr, E., Herrera, M., Izquierdo, J. & Pérez-García, R. 2017 Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics* **309**, 532–541.
Chang, T. Z. D. & Vowles, N. 2013 Strategies for improving data reliability for online surveys: a case study. *International Journal of Electronic Commerce Studies* **4** (1), 121–130.
Cheesman, J., Bennett, J. & Son, T. V. H. 2008 Estimating household water demand using revealed and contingent behaviors: evidence from Vietnam. *Water Resources Research* **44** (11), 1–11.
Crouch, M. L., Jacobs, H. E. & Speight, V. L. 2021 Defining domestic water consumption based on personal water use activities. *Journal of Water Supply: Research and Technology-Aqua* **70** (7), 1002–1011.
Dang, X., Li, L. & Fan, L. 2022 Spatiotemporal variation of household water consumption and coping strategies in water-stressed city of arid and semiarid Northwest China. *Frontiers in Environmental Science* **1144**, 1–11.
Dey, S. 2022 Impacts of major crop production on the agricultural sector in Bangladesh: a time series analysis. *International Journal of Sustainable Agricultural Research* **9** (1), 10–17.
Donkor, E. A., Mazzuchi, T. A., Soyer, R. & Alan Roberson, J. 2014 Urban water demand forecasting: review of methods and models. *Journal of Water Resources Planning and Management* **140** (2), 146–159.
DOS 2022 Department of Statistics, Jordan. Retrieved June 1. 2022, from www.dos.gov.jo: Analytical Reports – Department of Statistics (dos.gov.jo).
Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R. & Jermiin, L. S. 2020 Sensitivity and specificity of information criteria. *Briefings in Bioinformatics* **21** (2), 553–565.
Erdogan, B. E., Özöğür-Akyüz, S. & Ataş, P. K. 2021 A novel approach for panel data: an ensemble of weighted functional margin SVM models. *Information Sciences* **557**, 373–381.
Fielding, K. S., Russell, S., Spinks, A. & Mankad, A. 2012 Determinants of household water conservation: the role of demographic, infrastructure, behavior, and psychosocial variables. *Water Resources Research* **48** (10), 1–12.
Firat, M., Yurdusev, M. A. & Turan, M. E. 2009 Evaluation of artificial neural network techniques for municipal water consumption modeling. *Water Resources Management* **23** (4), 617–632.
Fricke, K. 2013 *Analysis and Modelling of Water Supply and Demand Under Climate Change, Land use Transformation and Socio-Economic Development: The Water Resource Challenge and Adaptation Measures for Urumqi Region, Northwest China*. Springer Science & Business Media, Heidelberg.
Gato-Trinidad, S., Jayasuriya, N. & Roberts, P. 2011 Understanding urban residential end uses of water. *Water Science and Technology* **64** (1), 36–42.
Grafton, R. Q., Ward, M. B., To, H. & Kompas, T. 2011 Determinants of residential water consumption: evidence and analysis from a 10-country household survey. *Water Resources Research* **47** (8), 1–14.
Greene, W. H. 2012 NLOGIT version 5 reference guide. *Plainview, NY*, *11803*.
Greene, W. H. & Hensher, D. A. 2010 *Modeling Ordered Choices: A Primer*. Cambridge University Press, New York.
Guhathakurta, S. & Gober, P. 2007 The impact of the Phoenix urban heat island on residential water use. *Journal of the American Planning Association* **73** (3), 317–329.
Halton, J. H. 1960 On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik* **2** (1), 84–90.
Japan International Cooperation Agency (JICA) 2015 The Project for the Study on Water Sector for the Host Communities of Syrian Refugees in Northern Governorates in the Hashemite Kingdom of Jordan. Ministry of Water and Irrigation, Amman. Accessed May 27, 2021. Microsoft Word – MPWS20150130_ver2.docx (jica.go.jp).
Kenney, D. S., Goemans, C., Klein, R., Lowrey, J. & Reidy, K. 2008 Residential water demand management: lessons from Aurora, Colorado 1. *JAWRA Journal of the American Water Resources Association* **44** (1), 192–207.
Kontokosta, C. E. & Jain, R. K. 2015 Modeling the determinants of large-scale building water use: implications for data-driven urban sustainability policy. *Sustainable Cities and Society* **18**, 44–55.
Lee, D. & Derrible, S. 2020 Predicting residential water demand with machine-based statistical learning. *Journal of Water Resources Planning and Management* **146** (1), 04019067.
Liu, C., Liu, A. & Xu, D. 2022 Influence of China's overall planning for urban and rural medical insurance on citizen health. *American Journal of Health Behavior* **46** (3), 248–258.

Long, J. S. & Long, J. S. 1997 *Regression Models for Categorical and Limited Dependent Variables*, Vol. 7. Sage, Bloomington.
Makki, A. A., Stewart, R. A., Beal, C. D. & Panuwatwanich, K. 2015 Novel bottom-up urban water demand forecasting model: revealing the determinants, drivers and predictors of residential indoor end-use consumption. *Resources, Conservation and Recycling* **95**, 15–37.
Manouseli, D., Kayaga, S. M. & Kalawsky, R. 2019 Evaluating the effectiveness of residential water efficiency initiatives in England: influencing factors and policy implications. *Water Resources Management* **33** (7), 2219–2238.
Michelsen, A. M., McGuckin, J. T. & Stumpf, D. 1999 Nonprice water conservation programs as a demand management tool 1. *JAWRA Journal of the American Water Resources Association* **35** (3), 593–602.
Ministry of Water and Irrigation (MWI). 2009 *Water for Life: Jordan's Water Strategy 2008–2022*. Ministry of Water and Irrigation, Amman. Available from: http://web.idrc.ca/uploads/user-S/12431464431JO_Water-Strategy09.pdf (accessed 22 August 2021).
Mora, M. (2011). *Validity and Reliability in Surveys*. Available from: http://www.relevantinsights.com/validity-and-reliability (accessed 22 February 2018).
Motoshita, M., Ono, Y., Pfister, S., Boulay, A. M., Berger, M., Nansai, K. & Inaba, A. 2018 Consistent characterisation factors at midpoint and endpoint relevant to agricultural water scarcity arising from freshwater consumption. *The International Journal of Life Cycle Assessment* **23** (12), 2276–2287.
Olmstead, S. M. & Stavins, R. N. 2009 Comparing price and nonprice approaches to urban water conservation. *Water Resources Research* **45** (4), 1–10.
Piasecki, A., Jurasz, J. & Kaźmierczak, B. 2018 Forecasting daily water consumption: a case study in Torun, Poland. *Periodica Polytechnica Civil Engineering* **62** (3), 818–824.
Polebitski, A. S. & Palmer, R. N. 2010 Seasonal residential water demand forecasting for census tracts. *Journal of Water Resources Planning and Management* **136** (1), 27–36.
Potter, R. B. & Darmame, K. 2010 Contemporary social variations in household water use, management strategies and awareness under conditions of 'water stress': the case of Greater Amman, Jordan. *Habitat International* **34** (1), 115–124.
Potter, R. B., Darmame, K. & Nortcliff, S. 2010 Issues of water supply and contemporary urban society: the case of Greater Amman, Jordan. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **368** (1931), 5299–5313.
Rathnayaka, K., Malano, H., Maheepala, S., George, B., Nawarathna, B., Arora, M. & Roberts, P. 2015 Seasonal demand dynamics of residential water end-uses. *Water* **7** (1), 202–216.
Rathnayaka, K., Malano, H., Arora, M., George, B., Maheepala, S. & Nawarathna, B. 2017 Prediction of urban residential end-use water demands by integrating known and unknown water demand drivers at multiple scales II: model application and validation. *Resources, Conservation and Recycling* **118**, 1–12.
Renwick, M. E. & Archibald, S. O. 1998 Demand side management policies for residential water use: who bears the conservation burden? *Land economics* **74**, 343–359.
Renwick, M. E. & Green, R. D. 2000 Do residential water demand side management policies measure up? an analysis of eight California water agencies. *Journal of Environmental Economics and Management* **40** (1), 37–55.
Rosenberg, D. E., Talozi, S. & Lund, J. R. 2008 Intermittent water supplies: challenges and opportunities for residential water users in Jordan. *Water International* **33** (4), 488–504.
Schwarz, G. 1978 Estimating the dimension of a model. *Annals of Statistics* **6** (2), 461–464. https://doi.org/10.1214/aos/1176344136.
Tabesh, M. & Dini, M. 2009 Fuzzy and neuro-fuzzy models for short-term water demand forecasting in Tehran. *Iranian Journal of Science and Technology* **33**, 61–77.
Tzanakakis, V. A., Paranychianakis, N. V. & Angelakis, A. N. 2020 Water supply and water scarcity. *Water* **12** (9), 2347.
UNHCR 2015 *UNHCR Report Shows World's Poorest Countries Host Most Refugees*. UNHCR Central Europe Accessed August 22, 2021.
Vieira, P., Jorge, C. & Covas, D. 2018 Efficiency assessment of household water use. *Urban Water Journal* **15** (5), 407–417.
Vitter, J. S. & Webber, M. E. 2018 A non-intrusive approach for classifying residential water events using coincident electricity data. *Environmental Modelling & Software* **100**, 302–313.
Watson, S. 2017 Consuming water smartly: the significance of sociocultural differences to water-saving initiatives. *Local Environment* **22** (10), 1237–1251.
Willis, R. M., Stewart, R. A., Panuwatwanich, K., Williams, P. R. & Hollingsworth, A. L. 2011 Quantifying the influence of environmental and water conservation attitudes on household end use water consumption. *Journal of Environmental Management* **92** (8), 1996–2009.
Willis, R. M., Stewart, R. A., Giurco, D. P., Talebpour, M. R. & Mousavinejad, A. 2013 End use water consumption in households: impact of socio-demographic factors and efficient devices. *Journal of Cleaner Production* **60**, 107–115.
World Bank 2013 *Jordan Overview*. Online document. Development News, Research, Data | World Bank. Viewed February 14, 2021.
Zanfei, A., Brentan, B. M., Menapace, A., Righetti, M. & Herrera, M. 2022 Graph convolutional recurrent neural networks for water demand forecasting. *Water Resources Research* **58** (7), e2022WR032299.
Zapata, O. 2015 More water please, it's getting hot! The effect of climate on residential water demand. *Water Economics and Policy* **1** (03), 1550007.
Zietlow, K. J. 2016 *Socio-demographic and Psychological Determinants of Water Conservation Behavior: Evidence From Germany and Jordan*. Doctoral dissertation, Humboldt Universität zu Berlin, Berlin.

First received 15 March 2023; accepted in revised form 9 July 2023. Available online 22 July 2023