Appalachian
STATE UNIVERSITY

Department of Economics Working Paper

Number 21-03| May 2021

Measuring consequentiality: A "Knife-Edge" versus Continuous Approach

Tanga M. Mohr
Appalachian State University

Peter A. Groothuis
Appalachian State University

John C. Whitehead
Appalachian State University

Kristan Cockerill
Appalachian State University

William P. Anderson, Jr.
Appalachian State University

Chuanhui Gu
Appalachian State University

Department of Economics
Appalachian State University
Boone, NC 28608
Phone: (828) 262-2148
Fax: (828) 262-6105
www.business.appstate.edu/economics

Measuring consequentiality: A "knife-edge" versus continuous approach

Tanga M. Mohr (Corresponding Author)
Department of Economics
Appalachian State University
mohrtm1@appstate.edu

Peter A. Groothuis
Department of Economics
Appalachian State University

John C. Whitehead
Department of Economics
Appalachian State University

Kristan Cockerill
Department of Interdisciplinary Studies
Appalachian State University

William P. Anderson, Jr.
Department of Geological and Environmental Sciences
Appalachian State University

and

Chuanhui Gu
Department of Geological and Environmental Sciences
Appalachian State University

May 18, 2021

Acknowledgements: We would like to thank Appalachian State University for funding the Water Cluster: Integrating Geology, Economics and Policy Analysis through the CONducting ComplEx Research Together (CONCERT) Grants.

Abstract: A survey is consequential to a respondent if they believe their answer could influence the policy being addressed in the survey and if they will have to pay for the policy if implemented. We show that separating out respondents who find the survey inconsequential, even by very simple metrics such as a single question, goes a long way if the goal is to improve willingness to pay estimates. Using various follow up questions, we develop multiple thresholds to classify respondents into groups based on whether or not their responses satisfy the consequentiality criteria. Independent of the threshold, we find that respondents in the inconsequential group have a willingness to pay that is insignificantly different from zero. For those in the consequential group, marginal willingness to pay does not significantly depend on the threshold. These results lend additional support to the 'knife-edge' hypothesis. To provide additional insights we explore consequentiality using a hybrid choice model and find that the likelihood of payment consequentiality increases with income while respondents who find the survey policy consequential are more likely to be in favor of the policy.

Key Words: consequentiality, stormwater management, stated preferences, hybrid choice models, generalized structural equation method

"If a survey's results are seen by the agent as potentially influencing an agency's actions and the agent cares about the outcomes of those actions, the agent should treat the survey questions as an opportunity to influence those actions. In such a case, standard economic theory applies and the response to the question should be interpretable using mechanism design theory concerning incentive structures." Carson and Groves (2007)

1 Introduction

A contingent valuation survey is consequential when respondents believe their answers may influence a business or government action, respondents care about the subsequent outcomes, and respondents believe there are real financial consequences to those outcomes (Carson and Groves 2007). To separate respondents who find the survey consequential from those who do not, researchers use survey follow up questions. These questions include many types such as "Do you believe that the survey results will be shared with policy makers and influence policy?" Alternatively, they may include a Likert scale statement with intensity of agreement such as "If the referendum passes my taxes will rise." The

researcher then applies a rule to these questions separating the respondents into consequential and inconsequential belief groups. The consequential group will consider the survey incentive compatible and answer truthfully. The inconsequential group will potentially suffer from hypothetical bias because they have no incentive to answer truthfully.

A basis for separating respondents into these groups is Carson, Groves and List (2014)[1] who show support for the proposition that, as long as the probability of affecting policy is positive, the proportion of respondents who vote in favor of the policy is not dependent on that probability. This has become known as a knife-edge rule used to separate respondents who find the survey consequential and treat the survey as having real consequences and those who find the survey inconsequential and treat the survey as hypothetical.[2] This knife-edge approach has been followed in several studies including Herriges et al. (2010), Groothuis et al. (2017) and Morgan et al (2018).

More recent literature, instead of applying a knife-edge rule, has used latent variable hybrid choice models to add new insights about consequentiality including the potentially separate components related to affecting policy on the one hand and having to pay on the other. Zawojska et al. (2019) used a hybrid choice model that introduces risk attitudes as a component and conclude that policy and payment consequentiality should be separated. Czajkowski et al. (2017) used a hybrid choice model that incorporated consequential scripts in contingent valuation analysis to study the causes of consequentiality.

In this study we apply the knife-edge approach using various measures of consequentiality to understand the sensitivity of our estimates to the question or combination of questions used. We then use a hybrid choice model to learn more about the determinants of the underlying latent consequentiality

[1] Based on earlier research by Carson, Groves, List and Machina (2002).
[2] This proposition is in the context of a single binary choice so strategic preferences do not come into play.

measures. To our knowledge we are the first to compare these approaches. We conclude, based on our data, that controlling for consequentiality matters, but the fineness of the consequentiality description is less important. Using split samples, we observe that willingness to pay estimates for respondents in the inconsequential group are not significantly different from zero while those in the consequential group are positive and similar across various measurements of consequentiality. We find this result encouraging if the goal is to estimate willingness to pay using simple models.

In addition, we find that willingness to pay estimates from our comparable hybrid choice model are lower because this model produces a continuous measure of consequentiality and uses the full sample of respondents, including the respondents who find the survey inconsequential. Hybrid choice models, however, can give us a deeper understanding of respondents' behavior or motivation. This may improve survey design but, such level of detail may not be necessary to generate willingness to pay estimates that do not suffer from hypothetical bias.

2 Related Literature

Carson and Groves (2007, 2011) and Carson (2012) argue that stated preference surveys will generate accurate statements of willingness-to-pay if the valuation question is incentive compatible and consequential. An incentive compatible question is one where respondents have incentives to truthfully reveal their preferences (e.g., a referendum tax vote with majority or plurality rule). According to Carson and Groves (2007) a preference survey question is consequential when a respondent believes their answers will be "potentially influencing to the agency's [business or government] actions" and the agent cares about the subsequent outcomes. There is mounting empirical evidence from laboratory and field experiments that consequential questions are not prone to hypothetical bias (Landry and List 2007, Vossler and Evans 2009, Vossler and Poe 2011, Vossler, Doyon and Rondeau 2012 and Carson, Groves

and List 2014).[3] That is, these studies suggest that hypothetical behavior will be similar to real behavior if there is a positive chance that the hypothetical choices will have real consequences. This strand of enquiry has had important implications for contingent valuation (CV) research which by necessity is primarily concerned with the elicitation of preferences in hypothetical scenarios.

Consequential CV surveys are expected to be more accurate but there are no predictions on the direction of bias for inconsequential surveys. For example, inconsequential stated preferences may be understated if respondents answer with 'protest noes' or overstated if respondents 'yea say'. The upwards bias or "yea saying" might be attributable to lack of payment consequentiality while the downwards bias might be attributable to lack of policy consequentiality.

For many CV applications researchers have found that respondents who believe the survey results are consequential are more likely to support the policy (Forbes et al. 2015, Hwang et al. 2014, Groothuis et al. 2017, and Li et al. 2018). Herriges et al. (2010) suggests individuals who hold strong beliefs that the survey is consequential and will influence policy are more likely to vote yes on a referendum. However, others have not found this positive relationship (e.g., Oehlman and Meyerhoff 2017, and Vossler et al. 2012). Kabaya (2020) find that scripts related to policy relevance increase support for the policy, while scripts emphasizing payment consequences reduce support.

Interis and Petrolia (2014) further explore the effects of consequentiality in binary and multiple discrete choice experiments.[4] Willingness-to-pay is greatest for respondents who believe it is very likely that policy makers will take survey results into consideration and lowest when respondents think that this is unlikely. In addition, they find that respondents who believed the survey was inconsequential were

[3] Consequentiality may also improve results when the hypothetical question is not incentive compatible. Bulte et al. (2005), using an implicit donation payment vehicle, find that a hypothetical question with a consequential script generates lower willingness-to-pay estimates than the hypothetical question without a consequential script.

[4] They do not find the knife-edge result with a binary discrete choice experiment but do with a multiple discrete choice experiment question.

less sensitive to scope effects. Vossler and Watson (2013) conduct sensitivity analysis of their results by incorporating consequentiality in the empirical willingness-to-pay model. Using a dummy variable for respondents who find the survey to be inconsequential, they find a negative effect on willingness-to-pay, and deleting respondents who find the survey to be inconsequential increases the theoretical validity of the willingness-to-pay model. Both Interis and Petrolia (2014) and Vossler and Watson (2013) find evidence to support an important implication of Carson and Groves (2007); respondents who perceive the survey to be inconsequential may care little about the outcome of the survey, so they have little reason to invest in well thought out responses.

Integrated Choice and Latent Variable (ICLV) models, or hybrid choice models, gained popularity in the transportation literature (e.g., Walker and Ben-Akiva, 2002; Kim et al., 2012) and have been more recently applied in the stated choice literature (e.g., Hess and Beharry-Borg, 2012, Czajkowski et al., 2017; Zawojska et al., 2019). The estimation procedure mitigates the problem of measurement error in survey responses to attitudinal questions. Because consequentiality has been measured in the literature using attitudinal questions, the application of hybrid choice models in this context makes sense, especially to analyze multiple aspects of consequentiality. One benefit of a hybrid choice model is that it could provide insights into the cause and direction of hypothetical bias.

3 Components of consequentiality

Herriges et al. (2010) break down Carson and Groves (2007) consequentiality requirements into two conditions they jointly label *strong* consequentiality. These conditions are policy consequentiality, which occurs when the survey respondent "believes that the survey will influence a policy that they care about," and payment consequentiality, which occurs when the respondent "perceives that there is a positive probability that they believe they will have to pay". If both of these conditions are met then a

survey respondent's dominant strategy is to answer truthfully. This separation has subsequently become common in the literature (e.g., Mitani and Flores, 2014; Johnston et al., 2017; Vossler and Holladay, 2018; Borger et al., 2020; Kabaya, 2020).

Most surveys measure consequentiality with follow up questions to the stated preference analysis. The structure of these questions differs across studies. We use three follow up questions to measure different aspects of policy consequentiality. We ask if respondents believe the survey results will be *shared* with policy makers; if respondents believe the survey will *affect* policy, and if respondents have *confidence* in the local government to implement the policy. Using three measures of policy consequentiality provides information on why respondents may or may not perceive the survey as policy relevant. For a respondent to believe that their answer may influence a business or government action they must first believe that the information obtained in the survey will be shared with policy makers, second that that information will be used by policy makers and affect their actions, and third have confidence in their government to carry out proposed changes. Most studies ask some form of the second component, i.e., whether or not the information in the survey will be used to affect policy. As we discuss later, single questions may well be sufficient to segment respondents into groups. Using multiple questions, however, can help disentangle why individuals fall into given groups.

We measure payment consequentiality by asking if respondents believe that their county sales, income or property tax will increase to pay for the proposed policy. All previous studies have essentially used the same payment consequentiality question. When respondents believe that their taxes will increase they are also more likely to consider the contingent valuation scenario more seriously.

One component of consequentiality that has been overlooked, or perhaps assumed, in the literature is whether the respondent cares about the agent's actions regarding the survey outcome. There are a number of ways respondents might care about the outcome; they may care that they will have to

pay higher taxes if a new policy is implemented, for example. In this study, we use a respondent's concern for the amenity being valued as a proxy for their caring about the outcome. While imperfect, this provides a measure of the saliency of the proposed policy to the respondent. As described later, we ask respondents for their level of concern about stormwater runoff and their level of concern for aquatic health.

Our strictest measure of consequentiality takes all three channels into consideration: is the survey policy relevant, are financial consequences credible, and is there concern about the amenity being valued. For our split sample models, we then use less strict measures, more in line with the current literature, to gauge the sensitivity of willingness to pay estimates. For comparison we use a hybrid choice model with three latent variables to learn more about factors influencing consequentiality.

Our study addresses respondents' likelihood of voting for a stormwater management plan to reduce the heat and salinity associated with stormwater runoff, which threaten stream quality. We suggest that using multiple follow up questions to measure policy consequentiality and adding measures of concern for the environmental amenity provides additional insights into factors affecting respondents' preferences for environmental amenities that previous studies have not addressed. The use of latent variables in a hybrid choice model can help reveal reasons behind respondents' choices such as why they do not believe their responses will be taken into consideration.

In our case, we find that the respondent's income positively influences their perceptions about having to pay (i.e., payment consequentiality), and confidence in the local government has a positive association with perceptions about influencing policy (i.e., policy consequentiality). In addition, we find that policy consequentiality as well as concern for the amenity being valued are positively related to being in favor of the contingent valuation scenario. Payment consequentiality does not affect the vote for our respondents.

4 Survey design and implementation

This study uses data from a stated preference survey, employing a discrete choice experiment to assess public perceptions about stormwater and attitudes toward stormwater management (Groothuis et al. forthcoming). Many urban streams face rising temperatures and increasing salinity levels caused by stormwater runoff, yet public attitudes toward managing these issues remains understudied. We surveyed residents in the Appalachian region that experience snow in the winter and hence apply salt to various surfaces, which can then runoff into local streams. Our respondents represent multiple states from North Carolina to New York. To help improve survey salience, we asked respondents to note their home county and then explicitly referenced that county in subsequent survey questions.

Our survey questions focused on assessing public knowledge about stormwater runoff, perceptions about stormwater runoff and stream quality, and preferences for stormwater management approaches in the respondent's county. We conducted a large pilot study with particular attention paid to consequentiality and revised the survey based on those results (Groothuis et al. 2017). We then pre-tested the revised survey with 78 Survey Sampling International (SSI) panelists and asked a convenience sample of people with diverse backgrounds to take the draft survey and provide feedback. We made some minor changes based on these results. In 2019 we used the SSI online respondent panel and the SurveyMonkey platform to field the final survey.[5] We received 737 complete responses that are used in this analysis.

The survey included several questions about stormwater runoff and asked respondents about their concern regarding various stormwater issues, including temperature and salinity. Respondents then read a hypothetical scenario explaining how stormwater can be managed that included three photographs

[5] These "opt-in" panels are becoming popular in social science research due to their relatively low cost and ability to quickly collect a large amount of data. Opt-in samples are useful for exploratory research, such as here, but use of these estimates for policy analysis should be done with caution (Baker et al. 2010, Yeager et al. 2011, Lindhjem, Henrik, and Navrud 2011).

showing a rain garden, a rain barrel, and permeable pavement with captions noting how each can be used to reduce runoff. Respondents were then told about scientific evidence suggesting that if nothing is done to address stormwater runoff and subsequent long-term salt levels, rivers and streams will suffer compromised aquatic health within the next few years. This text was accompanied by a visual aid based on current research about salinity in freshwater streams that illustrates that salinity levels have increased over time, posing a risk to aquatic health.

To provide context for a test of the effect of the scope of the management plan on referendum votes, the survey included illustrations of how stormwater management practices can reduce salinity levels by ten percent, twenty-five percent, or fifty percent. Larger values of reduction should increase the probability of a vote in favor of the management program (Whitehead, 2016) and including three levels of salinity reduction allows us to assess this.

The survey included three randomly ordered referendum questions using all three levels of the scope of the management plan individually, coupled with a randomly assigned one-time tax payment vehicle. The choice task was framed as a referendum voting question. One referendum question was then presented for each level of the scope of the management plan that included a randomly chosen value for the one-time tax increase ($A) for each level of scope.

We assigned higher payment levels for greater amounts of stream quality improvement. We added or subtracted a small random number so that the tax amounts would be different in each scenario to enhance realism. The one-time tax amounts were estimated based on the potential range of stormwater management plan costs developed from engineering studies and pretested in Groothuis et al. (2017). A statement was included to enhance perceived consequentiality (Carson and Groves 2007). In our analysis, we coded all undecided voters as *no* votes as suggested by Groothuis and Whitehead (2002) and Caudill and Groothuis (2005). Respondents who said they would not vote were excluded (n = 59).

Table 1 shows the vote percentages by tax and scope levels.

Our strictest measure of consequentiality considers whether the respondent is concerned about the amenity, whether they believe the survey will influence policy, and whether or not they believe they will have to pay the amount specified in the survey. We measure respondents' concern about the amenity using two 4-point Likert scale questions. Respondents were asked:

I. How concerned are you about stormwater runoff in [respondent's county]?
II. How concerned are you about compromised aquatic health in rivers and streams in [respondent's county]?

The responses are shown in Table 2. Overall, about 70% of respondents report being at least somewhat concerned about stormwater runoff while 78% are concerned about aquatic health. This suggests that the topic is relevant for most respondents. Still, 20% to 30% of respondents were unconcerned about stream quality.

To provide insights on respondents' beliefs about policy consequentiality, we asked three 5-point Likert scale questions about the survey's influence on policy: Do you agree or disagree with the following statements?

III. I believe the results of this survey will be shared with policy makers in [respondent's county]
IV. I believe the results of this survey could affect decisions about stormwater management policy in [respondent's county]
V. I have confidence in the [respondent's county] government.

Responses are summarized in Table 3. A majority of respondents believe the results will be shared with policymakers and affect decisions, but only 40% report confidence in their local government.[6]

[6] Conditional on having confidence in the county government, 78% think the survey will have an effect.

Regarding respondents' beliefs about payment consequentiality, we asked a 5-point Likert scale question about the survey's influence on the likelihood of higher tax payments that respondents will have to pay if a policy change is made. We asked respondents:

VI. If there were a county sales, income or property tax increase for the stormwater management plan in [respondent's county], do you think that your own tax bill would increase?

These responses are shown in Table 4. Most people view the survey as payment consequential with eighty-three percent of respondents believing their tax bill will probably or definitely increase if the stormwater management plan were implemented.

To understand the relationship between the indicator variables we conducted a principle component factor analysis. The results suggest that the variables are associated with two factors but *payment* (indicator question VI) is not highly correlated with either of those factors. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is above 0.55 for each of the variables and is 0.61 overall. This indicates that factor analysis is appropriate, but mediocre for this set of questions. The KMO values and rotated factor loadings are shown in Table 5. This result provides some support for the practice in the literature of differentiating between policy consequentiality and payment consequentiality (e.g., Herriges et al., 2010, Mitani and Flores, 2013) and shows support for consideration of respondents' level of concern. While we are assuming that concern for the amenity proxies for respondents caring about the outcome of the agency's decision, other variables may be more suitable for different studies.

5 Split sample results

In our first set of results we look at the sensitivity of our estimates to the complexity of the consequentiality classification. In particular, we look for differences in marginal willingness to pay which we measure with respect to the scope of the stormwater management plan; the estimates shown

are $/per unit of scope. For this analysis we apply the knife-edge definition to create binary variables equal to 0 when a respondent disagrees (or strongly disagrees with a statement) and 1 otherwise. We then use a split sample approach with respondents in either a consequential or inconsequential group. To account for the panel nature of our data, we cluster standard errors at the respondent level. In addition, we weighted our sample on gender and population to insure our sample is representative of the population of the Appalachian region.

The classifications (or thresholds) we use are shown in Table 6 and range from very strict (threshold 1) to less strict (threshold 4). Threshold 4 is most in line with early research on consequentiality (which considered primarily what is now called policy consequentiality) while threshold 3 separates policy and payment consequentiality. Threshold 2 uses additional indicator variables for policy consequentiality, while threshold 1 adds concern. Respondents are included if they are concerned about either runoff or aquatic health; this is less restrictive than requiring they be concerned about both. One motive for using the concern variables is to develop a proxy for whether respondents 'care about the outcome'. However, some readers may prefer to think of the concern variables as additional indicators of policy consequentiality. For this part of the analysis, the distinction is not important.

Using these classifications, we use split samples to produce the logit results in Tables 7a and 7b. Table 7a shows the results for respondents who fall into the inconsequential groups according to the thresholds in Table 6. Economic theory predicts that the taxes will be negatively related to voting for the proposal, the scope of the management plan positively related, and income positively related if storm water management is a normal good. For these respondents the tax is both negative and statistically significant as expected. Higher taxes result in a lower probability of voting in favor. However, the scope of the management plan is not statistically significant in any case. Income is likewise insignificant. We

also observe that the marginal willingness to pay estimates are not significantly different from zero in any of the four regressions. This lack of significance follows from the fact that scope is not significant for these respondents.

Table 7b shows the results for respondents who fall into the consequential groupings. For these respondents, the tax is negative and significant as theory would predict. Scope is positive and significant suggesting respondents are more likely to vote yes for more rigorous management plans. As with the inconsequential group, income is not significant for voting behavior. Turning attention to the willingness to pay estimates, we see what appears to be some differences in the mean levels. Since these estimates are calculated as $/per unit of scope, we conclude that willingness to pay is approximately $16.89 for the 10 percent management plan using consequentiality threshold 1, $16.88 using threshold 2, $15.07 using threshold 3 and $13.39 using threshold 4. We might expect the estimates to fall as the consequentiality threshold lowers, and we see evidence of that here, however, based on the confidence intervals including the point estimates from other distributions we cannot conclude they are statistically different. The similarity of the estimates is further illustrated in Figure 1 where we show histograms using simulation data.[7] The distributions are nearly identical for thresholds 1 and 2, and only the distribution for threshold 4 has a noticeably different mode.

What we conclude from this part of the analysis is that consequentiality matters when calculating willingness to pay estimates and determining whether respondents are attentive to scope. On the other hand, whether we consider single or multiple indicators of consequentiality is less important. The benefits of multiple indicator variables are non-trivial, however, and can give us a deeper understanding of respondent motivations, particularly if components of consequentiality have different effects on

[7] The simulations are based on 10,000 draws from a normal distribution using the mean and variances from the logit models.

voting behavior (as in Kabaya, 2020). For that level of analysis these variables can be incorporated using hybrid choice models. In the next section we discuss such a model using our data.

6 Hybrid Choice Model

We use a hybrid choice model in a random utility model framework to provide a comparison to the knife-edge models and provide insights on the influence of consequentiality on people's preferences for the stormwater management proposal. This technique allows researchers to incorporate perceptions and cognitive processes into a respondent's choice framework using latent variables. Our general approach follows Cajkowski et al. (2017) and Zawojska et al. (2019) who model the effect of consequentiality using Integrated Choice and Latent Variable (ICLV), or hybrid choice models.

Kim et al. (2014) provides an overview of and rationale for hybrid choice models, and our discussion of the model follows their description. For estimation we use Stata's generalized structural equation model (GSEM). The GSEM simultaneously estimates latent variable measurement models and a discrete choice model allowing us to incorporate the latent attitudinal variables as explanatory variables into the discrete choice utility model.[8]

The model we estimate is shown in Figure 2. The latent variable *pay* is estimated using a measurement equation with the indicator variable 'payment'. Payment is the respondent's answer to the Likert scale question VI above. The latent variable *policy* is estimated using a measurement equation with three indicator variables: shared, affect and confidence as defined in the previous section (questions III, IV and V). Unlike previous studies we add a third latent variable, concern. The latent variable *concern* is estimated in a measurement equation with two indicator variables: concern about stormwater

[8] The model assumes the variables have a conditional normal distribution.

runoff and concern for aquatic health (questions I and II above).[9] The indicator variables are modeled using ordered probit and use the full Likert scale of responses. They are coded from low to high in this analysis so that higher values represent more agreement.[10] The discrete choice model is estimated using binomial logit. The probability of voting for the stormwater management plan is estimated as a function of the tax, the scope of the plan, and income. The jointly estimated latent variables are also included as explanatory variables in the logit model.

We do not have theoretical predictions for the sign of the latent variables in the logit regression. The survey should be incentive compatible for those who believe their responses are consequential. On the other hand, that could make them more or less likely to vote in favor of the management plan. In particular, the results in the literature are mixed on the effect of payment and policy consequentiality. Vossler and Watson (2013) find that respondents who view the survey as consequential are more likely to vote in favor; Zawojska et al. (2019) find that payment consequentiality decreases willingness to pay while policy consequentiality increases willingness to pay.

The latent variable (LV) model can be composed of both measurement and structural components. Each measurement model has the form:

$$I_n = \theta LV_n + \varepsilon_n$$

where I_n is the response of individual n on the relevant indicator question, LV_n is the latent variable underlying survey responses, θ is a parameter to be estimated and ε_n is a random disturbance term. In Figure 2, the measurement component is illustrated by an arrow from the latent variable to the indicator variable. The indicator variables are observed; they are answers to survey questions representing

[9] In structural equation models ovals are used to indicate latent variables and rectangles to indicate exogenous variable.

[10] For the payment indicator variable in Table 4, we have coded don't know as a middle response. The order for this analysis is: definitely no, probably no, don't know, probably yes, definitely yes.

respondents' concern or beliefs. Theoretically, the answers to these questions are based on the unobserved variables we are interested in. For that reason, the indicator variables are functions of the latent variables.

The structural component of an LV model contains exogenous variables related to the information the respondent received in the survey. For example, the latent variables could be a function of demographic variables. For each latent variable the structural equation has the form:

$$LV_n = \Gamma X_n + \xi_n$$

where X_n is a set of exogenous variables with parameters Γ, and ξ is a random disturbance term. Here, we use income as an exogenous variable in the structural model.

The discrete choice logit model has the form:

$$y_n = \beta_z X_n^z + \beta_{LV} X_n^{LV} + \varepsilon_n$$

where subscript and superscript z refer to exogenous variables and LV refers to latent variables. The error term ε_n has a conditionally normal distribution. We estimate the likelihood a respondent votes for the stormwater management plan as a function of observed variables: tax increase, the scope of the plan and income, and unobserved latent variables: concern, policy and pay. Estimation of the GSEM model is by maximum likelihood. Since the scale of indicator variables is arbitrary, the LV models require normalization to ensure identification. We constrain the variance of the error term in the measurement models to be one. As in the previous section, we use clustered standard errors and population weights.

7 Hybrid Choice Model Results

We present the results of the model in three parts, the measurement models, the structural model and the discrete choice model. The results from the measurement models are shown in Table 8c. Recall that the measurement models assume the indicator variables (i.e., survey question responses) are a

function of some underlying latent variable. The names we supply for these latent variables are arbitrary but are chosen here to be in line with the consequentiality literature and the corresponding theory. The estimates in Table 8c suggest the indicator variables are all significantly affected by the underlying latent variables. The LV variable we label *concern*, which proxies for respondents caring about the outcome, is significantly affecting concern about stormwater runoff, and concern for aquatic health. The LV *policy* significantly influences respondents' beliefs about the survey being shared, affecting policy and their confidence in local government. Our results suggest that various beliefs in how the information is used as well as confidence in government are all potentially components of policy consequentiality. The LV *pay* significantly affects respondents' belief about having to pay if a new management program is implemented.

In Table 8b we show the results of the structural model where the latent variables are modelled as functions of income. We find that *pay* consequentiality is significantly affected by income. Respondents with higher income are more likely to find it credible that their taxes will increase and they will have to pay for the management plan. According to the structural model, however, income does not affect whether respondents find the survey policy relevant nor their concern for the amenity.

We show the results for the discrete choice model in Table 8a. The tax variable is negative and statistically significant as theory would predict. The likelihood of voting for the management plan declines as the tax amount increases. The scope variable is positive and significant suggesting that respondents are more likely to vote in favor of plans with more benefits. Income is also positive but not significant. The jointly estimated latent variables are included as explanatory variables in the discrete choice model. Concern about the problems associated with stormwater runoff, and beliefs about influencing policy both positively and significantly affect the likelihood of voting for the stormwater management plan. However, beliefs about having to pay for policy changes have an insignificant affect.

An increase in concern and policy relevance should increase the saliency and importance of the survey, and for our respondents, leads to an increase in support for the proposed management plan. Given the variation in responses to the policy indicator variables, respondents who believe the survey will have an effect and have confidence in their local government to carry out the management plan are more likely to vote in favor of the policy change. Since 83% of respondents believe they would have to pay if the management plan were implemented, this could explain why payment consequentiality does not significantly affect respondents' vote.

The hybrid choice model provides information we missed with the split sample logits in the previous section. For instance, the results suggest that the chosen indicator variables are functions of underlying latent variables that we are associating with consequentiality. As Czajkowskia et al. (2017, p. 61) conclude: "Although empirical evidence has demonstrated that conditioning on stated beliefs over policy consequentiality can enhance external validity, more evidence is needed and it remains an open question as to what exactly is being captured by belief questions in this context. We conclude our paper with a proposed approach for investigating this. Specifically, one can include multiple survey questions to measure a particular belief." Our analysis demonstrates that incorporating multiple indicator questions can help us understand more about what is being captured by these questions. For our data, for example, we find that a lack of confidence in the local government affects perceptions about the policy consequences of the survey. Future research could look more closely at respondents' attitudes towards government and their likelihood of voting for policy changes.

Another interesting finding that the logit models did not capture was the effect of income. While income does not affect the likelihood of voting for the management plan for our data, we do see that income affects respondents' beliefs about payment consequentiality. This might be important if a researcher is trying to determine the extent of hypothetical bias resulting from behavior such as 'yea-

saying'. Income could play a factor in those cases where respondents vote in favor even though they do not believe they will have to pay. Future research could further explore this relationship.

While Czajkowski et al. (2017) and Zawojska et al. (2019) use their models to show how policy consequentiality influences willing to pay estimates, it was not our goal in this paper to show how our three latent variables affect willingness to pay. Instead, we look at the average willingness to pay from the model as a comparison to the results from the split samples in the previous section. The latent variables in the hybrid choice model enter the discrete choice model as continuous variables. This is in contrast to our split sample models which put respondents into one of two groups. As such, the estimates include respondents who fall on the lower end of the consequentiality scale, and we might expect the willingness to pay estimates to be lower. That is what the estimates in Table 8c suggest (although, based on the confidence intervals there is still quite a bit of overlap with the results in Table 7b).

8 Conclusions

The stated preference literature has benefited from a rigorous analysis of consequentiality and its components over the last decade. We add to the literature by showing that separating out respondents who find the survey inconsequential, even by very simple metrics such as a single question, goes a long way toward improving willingness to pay estimates. Using multiple thresholds to classify respondents into groups based on whether or not their responses satisfy the consequentiality criteria, we find that respondents in the inconsequential group have a willingness to pay that is insignificantly different from zero. For those in the consequential group, willingness to pay decreases as the threshold is relaxed; however, based on confidence intervals these differences are not significant. These results lend additional support to the 'knife-edge' hypothesis of Carson, Groves and List (2014) that a perceived

positive probability of a binding result should be sufficient for respondents to behave as if the survey is real and consequential.

To understand what we gain from a more complex definition of consequentiality, we estimate a hybrid choice model with three latent variables we label as policy, pay and concern. Policy and payment consequentiality have been explored in the literature. The latent variable we call concern proxies for whether respondents care about the agency's decision. While there are unrelated reasons a respondent may care, we suggest that those who are concerned about the amenity being valued are also likely to care about the agency's decision.

Most studies seem to agree that policy consequentiality positively affects respondents' support for a proposal. On the other hand, the role of payment consequentiality is more complicated. As shown by (e.g., Groothuis et al., 2017) potential endogeneity between the magnitude of the payment and beliefs about having to pay can interplay in multiple ways that affect a respondent's vote/choice. For this study, the payment vehicle was perceived as credible enough that more than 80% believed they would have to pay for policy changes, thus in our study we do not find that payment consequentiality affects the likelihood of voting in favor of the proposed stormwater management plan.

Our research joins other studies showing that controlling for consequentiality through follow up questions is an important feature of survey design. Understanding which aspects of a proposed policy render it non-credible to respondents is necessary for deciphering stated preferences. Incorporating multiple Likert scale indicator questions to measure latent attitudes may be beneficial in that respect. Moreover, gauging respondents' level of concern for the policy issue can be revealing in a number of ways. Lack of concern may indicate that the survey issue is not perceived to be a problem or important. Further research could seek to tease apart the importance of respondents' lack of concern and to distinguish between low levels of concern and ambivalence towards the policy outcome.

References.

Bulte, Erwin, Shelby Gerking, John A. List and Aart de Zeeuw. 2005. “The effect of varying the causes of environmental problems on stated WTP values: evidence from a field study” *Journal of Environmental Economics and Management* 49:330–342.

Carson, Richard T. 2012. "Contingent Valuation: A Practical Alternative When Prices Aren't Available." *Journal of Economic Perspectives*, 26 (4): 27-42.

Carson, Richard T., Theodore Groves, John A. List, and Mark J. Machina. 2002. “Probabilistic influence and supplemental benefits: A field test of the two key assumptions underlying stated preferences,” Unpublished manuscript, Department of Economics, University of California, San Diego.

Carson, Richard T and Theodore Groves. 2007. “Incentive and informational properties of preference.” *Environmental and Resource Economics*, 37:181–210.

Carson, Richard T and Theodore Groves. 2011. “Incentive and information properties of preference questions: commentary and extensions” The International Handbook on Non-Market Environmental Valuation, Edward Elgar Publishing.

Carson, Richard T., Theodore Groves, John A. List. 2014. “Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice,” *Journal of the Association of Environmental and Resource Economists* 1:171-207.

Czajkowski, Mikołaj, Christian A. Vossler, Wiktor Budzińskia, Aleksandra Wiśniewska, Ewa Zawojska. 2017. Addressing empirical challenges related to the incentive compatibility of stated preferences methods, *Journal of Economic Behavior & Organization* 142: 47–63.

Groothuis, P.A., Tanga M. Mohr, John C. Whitehead, and Kristan Cockerill. 2017. “Endogenous Consequentiality in Stated Preference Referendum Data: The Influence of the Randomly Assigned Tax Amount” *Land Economics*, 93(2): 258-268.

Groothuis, P.A., Tanga M. Mohr, John C. Whitehead, Kristan Cockerill, William P. Anderson and Chuanhui Gu (forthcoming) "Measuring the Direct and Indirect Effect of Scientific Information on Valuing Stormwater Management Programs with a Hybrid Choice Model," *Water Resources Research*.

Herriges, Joseph, Catherine Kling, Chih-Chen Liu and Justin Tobias.2010. "What are the consequences of consequentiality?" *Journal of Environmental Economics and Management* 59(1): 67-81.

Hess, Stephane and Nesha Beharry-Borg. 2012. "Accounting for Latent Attitudes in Willingness-to-Pay Studies: The Case of Coastal Water Quality Improvements in Tobago," *Environmental and Resource Economics* 52:109–131

Interis, Matthew, and Daniel Petrolia. 2014. "The effects of consequentiality in binary-and multinomial-choice surveys." *Journal of Agricultural and Resource Economic*s 39(2): 201-216.

Kabaya, Kei. 2020. "Opposite impacts of policy and payment consequentiality treatments on willingness-to-pay in a contingent valuation study," *Journal of Environmental Economics and Policy*, DOI: 10.1080/21606544.2020.1816218.

Kim, Jinhee, Soora Rasouli and Harry Timmermans. 2014. "Hybrid choice models: Principles and recent progress incorporating social influence and nonlinear utility functions," *Procedia Environmental Sciences* 22: 20 – 34.

Landry, Craig, and John List. 2007. "Using ex ante approaches to obtain credible signals for value in contingent markets: Evidence from the field." *American Journal of Agricultural Economics* 89: 420-429.

Matani, Yohei and Nicholas E. Flores. 2014. "Hypothetical Bias Reconsidered: Payment and Provision Uncertainties in a Threshold Provision Mechanism," *Environmental and Resource Economics* 59:433–454.

Morgan, O Ashton, William L Huth, Paul Hindsley. 2018. "Examining the perceptions and effects of survey consequentiality across population subgroups." *Journal of Benefit-Cost Analysis* 9:2, 305-322.

Vossler, Christian, and Mary Evans. 2009. "Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality." *Journal of Environmental Economics and Management* 58: 338-345.

Vossler, Christian, and Gregory Poe. 2011. "Consequentiality and contingent values: An emerging paradigm." Chapter 7 in The International Handbook on Non-Market Environmental Valuation, ed. Jeff Bennett. Northampton, MA: Edward Elgar, 122-141.

Vossler, Christian A., Maurice Doyon and Daniel Rondeau. 2012. "Truth in consequentiality: Theory and field evidence on discrete choice experiments." *American Economics Journal: Microeconomics* 4(44): 145-171.

Vossler, Christian A. and Sharon B. Watson. 2013. "Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field," *Journal of Economic Behavior & Organization* 86: 137–147.

Zawojska, Ewa; Anna Bartczak, and Mikołaj Czajkowski. 2019. "Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences," *Journal of Environmental Economics and Management* 93: 63-84.

Figure 1: WTP estimates based on simulations for respondents in consequential groups, by threshold.

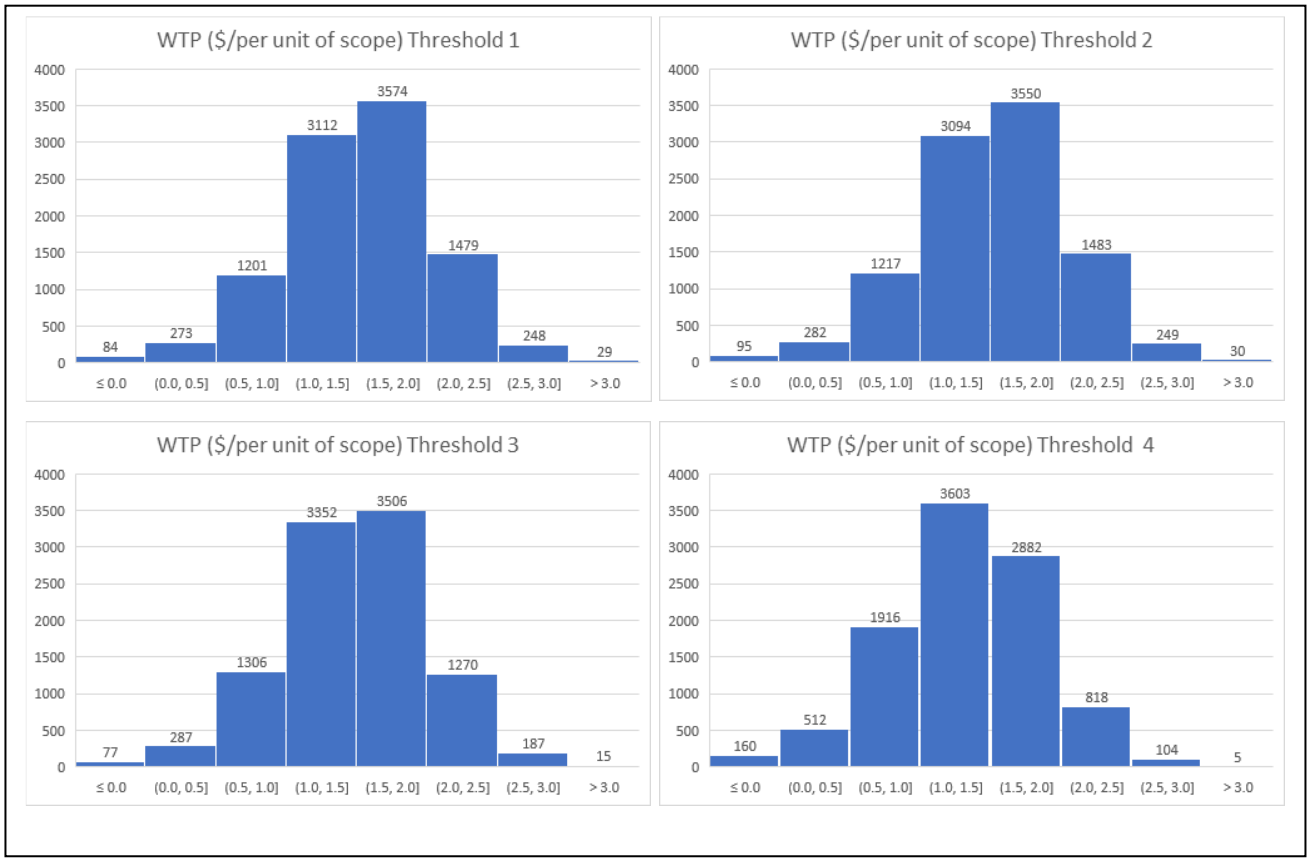

Figure 2: Hybrid choice model with three latent variables

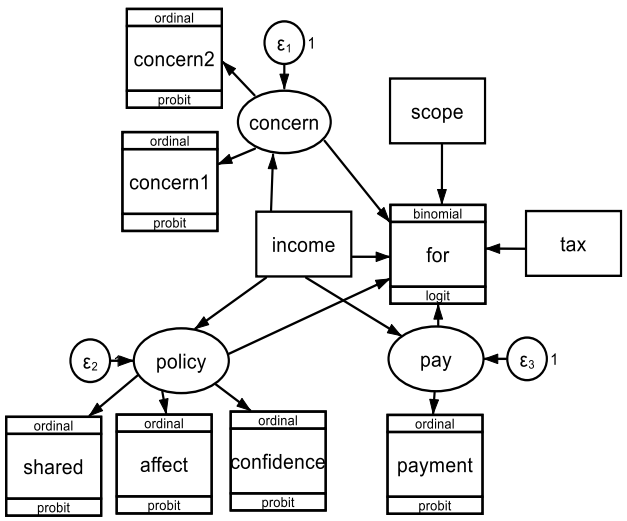

Table 1. Referendum Vote Responses by Tax and Scope

Scope = 10			Scope = 25			Scope = 50		
Tax	%For	Sample	Tax	%For	Sample	Tax	%For	Sample
28	60.00	155						
83	45.39	141	79	54.35	138			
129	38.85	157	120	47.79	136	122	45.51	156
162	34.85	132	171	33.99	153	179	40.69	145
226	22.37	152	224	30.30	165	231	33.80	142
			286	26.21	145	280	27.56	127
						329	25.75	167

Table 2: Concern about the amenity

	Concerned about runoff (%)	Concern about aquatic health (%)
1 Not concerned at all	12	7
2 Somewhat not concerned	18	15
3 Somewhat concerned	49	44
4 Very concerned	21	34

Table 3: Beliefs about the survey's influence on policy

	Shared (%)	Affect (%)	Confidence (%)
1 Strongly disagree	3	3	10
2 Disagree	6	8	15
3 Neither agree nor disagree	30	31	35
4 Agree	39	41	31
5 Strongly agree	22	17	9

Table 4: Beliefs about having to pay

	Do you think your tax bill will increase (%)
1 Definitely no	3
2 Probably no	5
3 I don't know	9
4 Probably yes	39
5 Definitely yes	44

Table 5: Rotated factor loadings and Kaiser-Meyer-Olkin statistics

	Factor 1:Policy Consequentiality	Factor 2: Concern	KMO
Concern- runoff	0.110	**0.843**	0.559
Concern- aquatic health	0.132	**0.862**	0.577
Shared	**0.809**	0.183	0.615
Affect	**0.825**	0.208	0.639
Confidence	**0.723**	-0.121	0.741
Payment	-0.036	0.415	0.616
Overall			0.611

Table 6: Consequentiality thresholds and dummy variable definitions

Consequentiality Threshold	Indicator Questions	Indicator Variables	The split sample Dummy variables = 1 if
1	I - VI	• Concern about runoff OR Concern about aquatic health • Shared, Affect, Confidence • Payment	From Tables 2 – 4 Concern > 1 Shared > 2 Affect > 2 Confidence > 2 Payment > 2
2	III - VI	• Shared, Affect, Confidence • Payment	
3	IV and VI	• Affect • Payment	
4	IV	• Affect	

Table 7a: Logit model for probability of voting in favor for respondents in the inconsequential groups.

	Inconsequential 1	Inconsequential 2	Inconsequential 3	Inconsequential 4
Tax	-0.005*** (0.001)	-0.005*** (0.001)	-0.007*** (0.002)	-0.011*** (0.003)
Scope	0.003 (0.005)	0.003 (0.005)	0.001 (0.007)	0.010 (0.010)
Income	0.003 (0.003)	0.003 (0.003)	0.004 (0.005)	0.003 (0.006)
Constant	-0.171 (0.277)	-0.058 (0.285)	0.457 (0.520)	-0.017 (0.611)
n (clusters)	304	286	141	85
R^2	0.034	0.034	0.049	0.093
WTP (Mean)	0.642	0.632	0.186	0.880
WTP CI	-1.165 to 2.449	-1.197 to 2.461	-1.875 to 2.247	-0.866 to 2.626
LogLiklihood	-602.629	-584.117	-250.4971	-139.386
AIC	1213.259	1176.233	508.9942	286.772
BIC	1232.521	1176.233	525.1837	300.937

Standard errors in parentheses. Significance Level: 1% (***), 5%(**), 10%(*)

Table 7b: Logit model for probability of voting in favor for respondents in the consequential groups.

	Consequential 1	Consequential 2	Consequential 3	Consequential 4
Tax	-0.006*** (0.001)	-0.006*** (0.001)	-0.006*** (0.001)	-0.006*** (0.001)
Scope	0.010** (0.004)	0.010** (0.004)	0.009** (0.004)	0.007** (0.003)
Income	0.002 (0.002)	0.003 (0.002)	0.003 (0.002)	0.002 (0.002)
Constant	0.427* (0.247)	0.287 (0.240)	0.180 (0.201)	0.198 (0.193)
n (clusters)	433	451	596	652
R^2	0.041	0.039	0.036	0.033
WTP (Mean)	1.536	1.535	1.493	1.339
WTP CI	0.535 to 2.535	0.654 to 2.722	0.528 to 2.458	0.336 to 2.342
LogLiklihood	-852.824	-821.137	-1146.665	-1251.63
AIC	1598.284	1650.274	2301.331	2511.265
BIC	1618.962	1671.114	2323.287	2533.579

Standard errors in parentheses. Significance Level: 1% (***), 5%(**), 10%(*)

Table 8a. Logit model from the hybrid choice model with three latent variables

Variable	Vote
Tax	-0.007*** (0.001)
Scope	0.009*** (0.004)
Income	0.002 (0.002)
Concern	0.551*** (0.126)
Policy	0.604*** (0.124)
Pay	0.004 (0.167)
Constant	0.168 (0.214)
n (clusters)	737
LogLiklihood	-17298.788
AIC	34673.58
BIC	34890.22
WTP (Mean) WTP (confidence Interval)	1.287 0.441 to 2.134

Table 8b: Structural model

	Concern	Policy	Pay
Income	0.001 (0.001)	0.001 (0.001)	0.008*** (0.002)

Table 8a – 8c: Standard errors in parentheses. Significance Level: 1% (***), 5%(**), 10%(*)

Table 8c: Measurement models

Latent/Indicator (dependent variable)	Runoff	Aquatic Health	Shared	Affect	Confidence	Payment	Error variance
Concern	1.316*** (0.329)	1.940*** (0.974)					1
Policy			1.206*** (0.165)	2.276*** (0.753)	0.468*** (0.085)		1
Pay						0.704*** (0.079)	1
Cut 1	-1.798 (0.301)	-2.76 (1.129)	-2.906 (0.279)	-4.318 (1.258)	-1.300 (0.107)	-2.006 (0.169)	
Cut 2	-0.643 (0.164)	-1.378 (0.574)	-2.024 (0.201)	-2.684 (0.782)	-0.569 (0.086)	-1.462 (0.114)	
Cut 3	1.424 (0.291)	1.169 (0.502)	-0.378 (0.134)	-0.388 (0.251)	0.337 (0.078)	-0.937 (0.106)	
Cut 4			1.340 (0.186)	2.612 (0.748)	1.573 (0.114)	0.448 (0.142)	