Addressing empirical challenges related to the incentive compatibility of stated preference methods

Mikołaj Czajkowski[1], Christian A. Vossler[2,*], Wiktor Budziński[1],
Aleksandra Wiśniewska[1] and Ewa Zawojska[1]

The published version of this study is in the *Journal of Economic Behavior & Organization*.
The final publication is available at https://doi.org/10.1016/j.jebo.2017.07.023.

---

[1] Department of Economics, University of Warsaw, Długa 44/50, Warsaw, Poland

[*] Corresponding author, cvossler@utk.edu

[2] Department of Economics and Howard H. Baker Jr. Center for Public Policy, University of Tennessee, Knoxville, Tennessee, USA

Addressing empirical challenges related to the incentive compatibility of stated preferences methods

Abstract: An emerging theoretical literature focused on the incentive compatibility of stated preference surveys offers a new lens through which to view extant evidence on external validity, and provides guidance for practitioners. However, critical theoretical assumptions rest on *latent* respondent beliefs, such as the belief that respondents view surveys as potentially influencing policy (i.e., policy consequentiality), which gives rise to pressing empirical challenges. In this study, we develop a Hybrid Mixed Logit model capable of integrating multiple latent beliefs, and subjective measures of these beliefs, into a discrete choice model of stated preferences. Planned use of a resource, which can also be considered a latent variable, is frequently an important consideration when modelling stated preferences for a change in a good, and we demonstrate how our framework can be used to incorporate this information simultaneously. We further explore whether simple information treatments, which vary the degree to which the potential role of surveys in informing policy is emphasized, can influence respondent beliefs. Our results suggest that latent beliefs over consequentiality, current use and, to a much lesser extent, the information treatments significantly influence elicited willingness to pay.

Keywords: discrete choice experiment; stated preferences; consequentiality; field experiment; hybrid mixed logit model

JEL classification: C93, C35, H41, Q51

1. Introduction

Stated preference surveys continue to be the leading approach for estimating the economic value of products not available in markets, including proposed public policies with passive use value. Although the methodology has been in use for over fifty years, concerns over the ability of surveys to provide valid welfare measures remain, serving as an obstacle to widespread adoption in the legal and policy arenas. Recently, theoretical work has identified conditions for a stated preference survey to be incentive compatible in the sense that it provides incentives for truthful preference revelation. These conditions for incentive compatibility rely heavily on *latent* (i.e., unobserved) respondent beliefs. For instance, when a single binary choice (SBC) question is used, respondents must perceive that the stated cost can be coercively collected upon implementation of the project (i.e., payment consequentiality) and that a response in favor of the proposal weakly monotonically increases the chance of its implementation (i.e., policy consequentiality) (Carson and Groves, 2007; Vossler, Doyon and Rondeau, 2012). In addition to these beliefs, incentive compatibility for the increasingly popular repeated binary discrete choice experiment (binary DCE) requires respondents to believe that at most one of the proposed policies can be implemented, and that the perceived implementation rule induces independence between choice sets (i.e., that a vote on one policy in one choice set has no effect on the implementation probability of another policy from another choice set) (Vossler, Doyon and Rondeau, 2012). Our study proposes methods for addressing two of the significant challenges that arise when undertaking empirical work that endeavors to satisfy and/or test the theoretical assumptions tied to respondent beliefs.

One empirical challenge is how to incorporate stated measures of latent/unobservable beliefs, such as Likert-scale responses to a question about perceived policy consequentiality, into models of stated preferences. Direct inclusion of stated measures of beliefs may be problematic for two reasons. First, stated beliefs are measured imprecisely, giving rise to issues of measurement error. Second, stated beliefs may be correlated with other unobserved factors that influence respondents' choices. In prior work, Herriges *et al*. (2010) develop a Bayesian treatment effect model for SBC data that uses instrumental variables to identify the relationship between stated policy consequentiality and willingness to pay (WTP). Vossler, Doyon and Rondeau (2012) and Vossler and Watson (2013) briefly mention binary probit instrumental variable models, with the former study suggesting statistical evidence that measured beliefs can be considered exogenous and the latter citing a weak instruments problem. Here, we propose a

Hybrid Mixed Logit (HMXL) approach, which models an unobserved belief as a latent variable in a Random Utility Maximization (RUM) framework, and specifies a measurement equation where a stated measure of the belief is a function of the latent variable and an error term, thus recognizing the presence of measurement error.[3] Relative to prior approaches, the proposed HMXL model can: (1) be applied generally to both SBC and DCE data; (2) accommodate multiple latent factors in flexible ways; and (3) incorporate flexible specifications for measurement equations (e.g., ordered choice, multinomial choice, count data models, etc.). Further, as with standard mixed logit models, the HMXL allows the analyst to incorporate various forms of preference heterogeneity. Identification relies on there being available measures of the latent variables, rather than instrumental variables in the case where stated beliefs are directly included in the choice model.

A second challenge, assuming the theoretical incentive compatibility conditions tied to beliefs are not universally met, is how to modify survey design to induce desired beliefs (i.e., make respondents believe in real consequences following from a survey outcome). In most studies, aside from controlled experiments, researchers do not have the ability to manipulate the *actual* consequentiality of a survey. Further, whether and to what extent a survey is actually consequential is rarely known *ex ante*. Under these circumstances, beliefs over consequentiality remain important theoretically, and these beliefs are likely influenced by many aspects of survey design. In their critical review of the literature, Kling, Phaneuf and Zhao (2012) point out that "the effect of consequentiality scripts in stated preference surveys is in its infancy". Using as a case study a binary DCE survey focusing on the public financing of municipal theaters in Warsaw, Poland, we employ a split-sample approach to investigate four information scripts that vary in their emphasis of policy consequentiality.[4] The baseline treatment provides information at a level that is common in stated preference surveys, with additional emphasis placed in the other treatments. As an ancillary benefit, this exogenous variation allows us to identify whether there is a *causal* effect of policy consequentiality on elicited values. As acknowledged in prior work, follow-up consequentiality questions are themselves

[3] Many researchers state that hybrid choice models address general endogeneity issues (e.g., Daly *et al.*, 2011), such as may arise when unobservables underlying a stated belief measure are correlated with the choice model errors. Budziński and Czajkowski (2017) undertake a Monte Carlo analysis, demonstrating that this is not necessarily the case, and propose an extension to the HMXL that allows errors to be correlated across equations.

[4] Our application is of potential interest in its own right, as few non-market valuation studies have examined the value of the performing arts (Forrest, Grime and Woods, 2000; Hansen, 1997; Willis and Snowball, 2009; Grisolía and Willis, 2010; Grisolía and Willis, 2012; Willis *et al.*, 2012).

*in*consequential constructs. This opens up the possibility that identified correlations may be spurious and, similarly, that drivers of responses to consequentiality questions may have little to do with actual beliefs.

In the field survey context, and in a similar spirit to our research, a few prior studies exogenously vary information provided to respondents with the intent of altering perceptions over policy consequentiality.[5] Bulte *et al.* (2005) find that providing a statement that alerts respondents that the results of the study "will be made available to policymakers, and could serve as a guide for future decisions" decreases WTP. Oehlmann and Meyerhoff (2017) and Drichoutis *et al*. (2015), in contrast, observe no WTP changes resulting from the inclusion/exclusion of a consequentiality script. Herriges *et al*. (2010) make use of a published article that provided evidence that survey results directly affected related policy decisions in the past, and find that stated beliefs over policy consequentiality increase when respondents are provided this information. All of these studies used SBC elicitation, and as such our exploration provides primary evidence on this type of inducement for DCE surveys. Further, the scripts we explore can further be easily incorporated into general practice. Indeed, it is presumably rare to have relevant media coverage available that provides a clear third-party link between surveys and policy. An ancillary benefit of the HMXL framework in this context is that it allows one to not only measure whether information treatments alter stated beliefs but also whether such treatments influence stated WTP.

The empirical study provides several important insights. First, similar to Vossler and Watson (2013), for our application we are not able to identify (strong) instrumental variables from the extensive information collected through the survey.[6] This provides further impetus for the proposed HMXL estimator. Second, we find that latent beliefs over policy consequentiality have a discernible effect on elicited WTP for the proposed discounted theater ticket programs. Importantly, these latent beliefs are strongly correlated with stated beliefs, using as the measurement device a Likert-scale policy consequentiality question now prevalent in the

[5] A handful of controlled experiment studies exogenously vary the level of *actual* policy consequentiality, for example, by manipulating the probability a vote is binding (Landry and List, 2007; Mitani and Flores, 2012; Carson, Groves and List, 2014), or by introducing treatments the vary the proportion of respondent and "regulator" votes (Collins and Vossler, 2009; Vossler and Evans, 2009).

[6] In particular, we estimated a version of the mixed logit model that incorporates a control-function approach to deal with the potential endogeneity of stated perceptions of policy consequentiality. We refer the interested reader to Guevara and Ben-Akiva (2010) for a comparison of control-function and latent-variable approaches.

literature.[7] Third, WTP is significantly correlated with the information treatments, which emphasizes the empirical importance of the theoretical assumption regarding policy consequentiality; indeed, this can be taken as evidence in favor of construct validity.[8] Fourth, somewhat surprisingly, the information treatments have no significant effect on stated beliefs. Thus, although the econometric results provide empirical support that a responses to a follow-up question about policy consequentiality carries useful information, the findings emphasize the importance of developing follow-up questions that elicit beliefs more precisely. We further highlight the need to use multiple belief elicitation questions in order to help discriminate underlying motives. Our econometric framework can be applied to such an investigation.

2. Econometric approach – the Hybrid Mixed Logit model

Hybrid choice models are very flexible tools that allow analysts to incorporate perceptions and cognitive processes into a RUM model framework. They thus provide a link between behavioral sciences (e.g., psychology) and fields oriented on estimation such as engineering and economics. As stated in Ben-Akiva *et al.* (2002), hybrid choice models (also known as Integrated Choice and Latent Variable models) are a general class of models which may include additional random disturbances in the form of error components or random parameters, psychological factors in the form of latent variables and latent classes, with possible non-RUM decision processes in some of them. Of course, latent variables do not need to be limited only to psychological constructs, but can also be used for other features which are not straightforward to measure, such as social interactions (Kamargianni, Ben-Akiva and Polydoropoulou, 2014), social influences (Kim, Rasouli and Timmermans, 2014) or response quality (Hess and Stathopoulos, 2013). In this study we develop a HMXL model which combines the framework widely adopted for analyzing DCE data, the Mixed Logit (MXL) model (Greene, 2011), with the Multiple Indicators and Multiple Causes (MIMIC) model (Jöreskog and Goldberger, 1975).

[7] There are multiple possible interpretations of what latent variables in hybrid choice models are actually capturing. Following the literature, and for ease of exposition, we interpret a latent variable in a manner consistent with the (observed) variable that we assume provides a measurement of the latent variable; e.g., the latent variable we measure through stated responses to a policy consequentiality question is labelled *latent consequentiality*. In the Discussion section we elaborate on issues of interpretation.

[8] Given we do not have a criterion measure of actual demand for the proposed programs, we can make no claim that conditioning WTP estimates on the latent belief or on the information treatments produces more accurate or externally valid measures.

Connecting discrete choice models with a MIMIC model is an emerging approach for incorporating psychological factors in the RUM framework. Most of the applications to date appear in the transportation literature (e.g., Vredin Johansson, Heldt and Johansson, 2006; Hess, Hensher and Daly, 2012; Daziano and Bolduc, 2013). Applications in the economics literature include Hess and Beharry-Borg (2012), who incorporate latent attitudes towards coastal water quality protection in a model of stated preferences regarding watery quality improvements in Tobago, and Dekker *et al.* (2016), who treat preference uncertainty as a latent variable in a choice model of flood risk policies in the Netherlands. Further, Hoyos, Mariel and Hess (2015) employ latent constructs to explain preferences regarding land management in Spain, and Czajkowski, Hanley and Nyborg (2017) use the approach for investigating motives for household recycling. Vij and Walker (2016) analyze the possible advantages of employing a hybrid framework and provide general criteria for assessment of whether its use is justified. Corresponding with their second criterion, in our setting we use the HMXL in order to find a relationship between survey consequentiality and respondents' preferences, which can be used to inform policy and practice.

In the context of our application, we assume there are two latent factors in the discrete choice model. The first is a psychological factor assumed to explain responses to a Likert-scale question gauging beliefs over policy consequentiality. For ease of exposition, we interpret this factor as *latent consequentiality*. The second factor is assumed to explain the variation in past visits to local theaters, which is presumably an important underlying driver of demand for proposed programs, which we interpret as *latent theater use*. The first latent factor is of course motivated by theory work related to stated preferences, whereas the second captures heterogeneity from possible changes in theater ticket demand due to proposed discount ticket programs.

Hybrid choice models can consist of up to three parts: a discrete choice model, structural equations, and measurement equations. Below we describe each part in turn. In doing so, we describe the general framework as well as provide estimation details relevant for our particular application.

2.1. Discrete choice model

The theoretical foundation for the discrete choice model is RUM theory, which assumes that the utility a person derives depends on observed characteristics of choices and unobserved idiosyncrasies, represented by a stochastic component (McFadden, 1974). As a result, individual i's utility from alternative j in choice set t can be expressed as:

$$V_{ijt} = a_i c_{ijt} + \boldsymbol{b}_i' \boldsymbol{X}_{ijt} + e_{ijt}, \tag{1}$$

where utility is assumed to be additively separable in the cost of the alternative, c_{ijt}, and other attributes, $\boldsymbol{X}_{ijt}$; a_i and $\boldsymbol{b}_i$ denote estimable parameters; and e_{ijt} is a stochastic component allowing for factors not observed by the econometrician to affect utility and choices.

There are two aspects of the specification to emphasize. First of all, a_i and $\boldsymbol{b}_i$ are *individual*-specific, thus allowing for heterogeneous preferences amongst respondents and motivating the MXL model. Assuming instead that parameters are the same for all respondents implies homogenous preferences and results in the multinomial logit (MNL) model as a special case. Second, the stochastic component of the utility function (e_{ijt}) is of unknown, possibly heteroskedastic variance $\left(\operatorname{var}\left(e_{ijt}\right) = s_i^2\right)$. Identification of the model typically relies on normalizing this variance, such that the error term $\varepsilon_{ijt} = e_{ijt} \dfrac{\pi}{\sqrt{6} s_i}$ is i.i.d. type I extreme value with constant variance $\operatorname{var}\left(\varepsilon_{ijt}\right) = \pi^2 / 6$. This leads to the following specification:

$$U_{ijt} = \sigma_i a_i c_{ijt} + \sigma_i \boldsymbol{b}_i' \boldsymbol{X}_{ijt} + \varepsilon_{ijt}. \tag{2}$$

Note that due to the ordinal nature of utility, specification (2) represents the same preferences as (1). The estimates $\sigma_i a_i$ and $\sigma_i \boldsymbol{b}_i$ do not have a direct interpretation anyway, but if interpreted in relation to each other the scale coefficient ($\sigma_i = \pi / \left(\sqrt{6} s_i\right)$) cancels out.

Finally, given the interest in establishing estimates of WTP for the non-cost attributes, $\boldsymbol{X}_{ijt}$, it is convenient to introduce the following modification which is equivalent to using a money-metric utility function (a.k.a. estimating the parameters in WTP space; see Scarpa, Thiene and Train, 2008):

$$U_{ijt} = \sigma_i a_i \left(c_{ijt} + \frac{\boldsymbol{b}_i'}{a_i} \boldsymbol{X}_{ijt} \right) + \varepsilon_{ijt} = \lambda_i \left(c_{ijt} + \boldsymbol{\beta}_i' \boldsymbol{X}_{ijt} \right) + \varepsilon_{ijt}. \tag{3}$$

Under this specification the vector of parameters $\boldsymbol{\beta}_i$ is now scale-free and can be directly interpreted as a vector of implicit values for the attributes, $\boldsymbol{X}_{ijt}$. In MXL models, an additional advantage of this formulation is that the econometrician can specify a particular distribution of

WTP in the population, by defining the distribution of $\boldsymbol{\beta}_i$, rather than the distribution of the underlying utility parameters, $\boldsymbol{b}_i$.[9]

We assume there is a vector of individual-specific latent variables, denoted by $\mathbf{LV}_i$, that depend on the random parameters of the utility function. The functional form of this dependence may vary due to distributional assumptions. In the analysis we use two distributions, normal (for all non-cost attributes) and log-normal (for the cost attribute). For a normally distributed $\boldsymbol{\beta}_i$, this dependence is of the form:

$$\boldsymbol{\beta}_i = \boldsymbol{\Lambda}'\mathbf{LV}_i + \boldsymbol{\Xi}'\mathbf{SD}_i + \boldsymbol{\beta}_i^*, \tag{4}$$

where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Xi}$ are matrices of estimable coefficients and $\boldsymbol{\beta}_i^*$ has a multivariate normal distribution with a vector of means and a covariance matrix to be estimated.[10] This specification allows for inclusion of socio-demographic or other directly observable variables (such as different treatments in the survey) in the vector $\mathbf{SD}_i$. Similarly, we assume that the cost coefficient depends on latent variables in the following way:

$$\lambda_i = \exp\left(\boldsymbol{\tau}'\mathbf{LV}_i + \boldsymbol{\zeta}'\mathbf{SD}_i + \lambda_i^*\right), \tag{5}$$

where $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}$ are vectors of estimable coefficients and λ_i^* follows a normal distribution with the parameters describing its mean and standard deviation to be estimated.[11] As a result, the conditional probability of individual i's choices $\mathbf{y}_i$, for all T_i choice tasks, is given by:

$$P\left(\mathbf{y}_i \mid \mathbf{X}_i, \boldsymbol{\beta}_i^*, \lambda_i^*, \mathbf{LV}_i, \boldsymbol{\Lambda}, \boldsymbol{\Xi}, \boldsymbol{\tau}, \boldsymbol{\zeta}, \boldsymbol{\theta}\right) = \prod_{t=1}^{T_i} \frac{\exp\left(\lambda_i\left(c_{ijt} + \boldsymbol{\beta}_i'\mathbf{X}_{ijt}\right)\right)}{\sum_{k=1}^{C}\exp\left(\lambda_i\left(c_{ikt} + \boldsymbol{\beta}_i'\mathbf{X}_{ikt}\right)\right)}, \tag{6}$$

where $\boldsymbol{\theta}$ is a vector of parameters on which λ_i^* and $\boldsymbol{\beta}_i^*$ depend.

---

[9] As translating utility parameters into a money equivalent requires dividing them by a (possibly also random) cost coefficient, this can imply what are often implausible assumptions about the distribution of WTP (Carson and Czajkowski, 2013).

[10] The number of columns in Λ is equal to the number of latent variables, and the number of rows is equal to the number of non-cost attributes.

[11] λ_i^* can also be correlated with $\boldsymbol{\beta}_i^*$.

2.2. Structural equations

Latent variables can depend on exogenous factors, such as socio-demographic variables, which are stacked in the vector $\mathbf{X}_i^{str}$. Vector $\mathbf{X}_i^{str}$ may, in principle, overlap with vector $\mathbf{SD}_i$. This relationship is described by the following structural equations:

$$\mathbf{LV}_i = \mathbf{\Psi}'\mathbf{X}_i^{str} + \xi_i, \tag{7}$$

with a matrix of coefficients $\mathbf{\Psi}$ and error terms ξ_i which are assumed to come from a multivariate normal distribution.[12] Generally, linking socio-demographic variables with latent variables through structural equations is not necessary. In the absence of such structural equations, latent variables become similar to random parameters – they capture the correlation between individuals' preferences and measurement variables.

In order to make identification of hybrid choice models possible, the scale of every latent variable needs to be normalized (Hess, Hensher and Daly, 2012). This can be done by normalizing variances of the error terms in the structural equations or by normalizing some coefficients for each latent variable in the measurement equations. Raveau, Yáñez and Ortúzar (2012) conducted a simulation study which indicates that normalizing variances leads to better convergence of the models and results in the recovered estimates being closer to the underlying data generating process. Normalizing the variances of the error terms is thus the approach we adopt here. Contrary to most studies conducted to date, we do not normalize the variance of ξ_i to one. Instead, we use normalization to assure that the variance of every latent variable in $\mathbf{LV}_i$ is equal to one. Although such an approach introduces additional nonlinearities into the model, it is very useful, as now all latent variables have the same scale (even with socio-demographic variables in structural equations) and therefore their relative importance (e.g., in measurement equations) can be easily assessed. We do not observe any additional issues with convergence due to this normalization.

Formally, we define $\mathbf{LV}_i^* = \mathbf{\Psi}^{*}{}'\mathbf{X}_i^{str} + \xi_i^*$, with $\mathbf{\Psi}^*$ being a matrix of parameters to be estimated and ξ_i^* being a vector of independent normally distributed variables with mean zero and unit standard deviation. With $\mathbf{LV}_{\bullet k}^*$ representing a vector of values of the

[12] This is a common assumption, although Bhat, Dubey and Nagel (2015) introduce a specification that allows for non-normal error terms.

k-th non-normalized latent variable for all individuals, and $\delta_k = std\left(\mathbf{LV}^*_{\bullet k}\right)$ its standard deviations, we have $\mathbf{LV}_{\bullet k} = \mathbf{LV}^*_{\bullet k} / \delta_k$, $\mathbf{\Psi}_k = \mathbf{\Psi}^*_k / \delta_k$ and $\boldsymbol{\xi}_{\bullet k} = \boldsymbol{\xi}^*_{\bullet k} / \delta_k$.[13]

2.3. Measurement equations

The main purpose of including latent variables in the models is that they describe psychological or other factors that cannot be measured in a direct way, unlike other individual characteristics such as age and gender. Instead, a researcher must use various indicator questions in a survey, responses to which are hypothesized to be determined by the latent variables.

The model choice for the indicator equations depends on the particular application. The measurement equations can be linear, ordered, binary, multinomial or count regressions – whatever fits an interpretation of each indicator best. In this study we include one indicator for the latent belief over the policy consequentiality of a survey, measured on a five-point Likert scale. The measurement equation is modelled using ordered probit. We include a second indicator, which is a count variable of prior theater visits. We model this as a Poisson regression.

First, consider the case of the ordinal indicator. We can specify the following index function:

$$I_i^* = \boldsymbol{\rho}_1' \mathbf{LV}_i + \boldsymbol{\upsilon}_1' \mathbf{X}_i^{Mea} + \eta_i , \tag{8}$$

where $\boldsymbol{\rho}_1$ and $\boldsymbol{\upsilon}_1$ are vectors of coefficients and η_i denotes an error term assumed to have a normal distribution with zero mean and unit standard deviation.[14] $\mathbf{X}_i^{Mea}$ is a vector of socio-demographic variables, which directly explain the indicator variable. Such a formulation allows for greater flexibility of the model and may provide additional insight into how socio-demographic variables influence respondents' choices modeled through the HMXL.

For our application, in the first measurement equation, we define I_i as the indicator for policy consequentiality for individual *i*. Policy consequentiality is measured through a Likert-

---

[13] $\mathbf{\Psi}_k$ denotes k-th row of $\mathbf{\Psi}$ matrix, and $\boldsymbol{\xi}_{\bullet k}$ denotes stacked values of the random term in the k-th structural equation for all individuals.

[14] It is important to note that the number of measurement equations does not need to equal the number of latent variables. For instance, cases may arise where more than one indicator for a latent variable is available (e.g., there may be two survey questions measuring beliefs over policy consequentiality). This framework can accommodate such a setting by specifying multiple measurement equations for a single latent variable.

scale question to gauge the degree of perceived policy consequentiality, giving rise to an ordinal measurement variable that takes on five values. Accordingly, we adopt a model of ordered choices.[15] Under this specification, we define the relationship between I_i and I_i^* as follows:

$$\begin{aligned} I_i = 1 \quad & \text{if} \quad I_i^* < \alpha_1 \\ I_i = 2 \quad & \text{if} \quad \alpha_1 \le I_i^* < \alpha_2 \\ I_i = 3 \quad & \text{if} \quad \alpha_2 \le I_i^* < \alpha_3, \\ I_i = 4 \quad & \text{if} \quad \alpha_3 \le I_i^* < \alpha_4 \\ I_i = 5 \quad & \text{if} \quad \alpha_4 \le I_i^* \end{aligned} \tag{9}$$

where the α 's are the threshold parameters to be estimated. Assuming a normal distribution for η_i, this leads to the well-known ordered probit likelihood form for I_i:

$$P\left(I_i \mid \mathbf{X}_i^{str}, \boldsymbol{\xi}_i^*, \mathbf{X}_i^{Mea}, \boldsymbol{\rho}_1, \boldsymbol{\upsilon}_1, \boldsymbol{\alpha}\right) = \Phi\left(\alpha_l - \boldsymbol{\rho}_1 '\mathbf{LV}_i - \boldsymbol{\upsilon}_1 '\mathbf{X}_i^{Mea}\right) - \Phi\left(\alpha_{l-1} - \boldsymbol{\rho}_1 '\mathbf{LV}_i - \boldsymbol{\upsilon}_1 '\mathbf{X}_i^{Mea}\right), \tag{10}$$

where $\Phi(\cdot)$ denotes the normal cumulative distribution function and $I_i = l$.[16]

For the second measurement equation, the dependent variable is a count of past theater attendances, denoted as J_i. Employing a Poisson model, we define $\mu_i = \exp\left(\boldsymbol{\rho}_2 '\mathbf{LV}_i + \mathbf{v}_2 '\mathbf{X}_i^{Mea}\right)$, where $\boldsymbol{\rho}_2, \boldsymbol{\upsilon}_2$ are sets of vectors to be estimated, and the corresponding probability is:

$$P\left(J_i \mid \mathbf{X}_i^{str}, \boldsymbol{\xi}_i^*, \mathbf{X}_i^{Mea}, \boldsymbol{\rho}_2, \boldsymbol{\upsilon}_2\right) = \frac{\exp\left(-\mu_i\right) \mu_i^{J_i}}{J_i!}. \tag{11}$$

2.4. Hybrid Mixed Logit Model estimation

Finally, we combine the discrete choice model specified in (6), the structural equations described in (7) and the measurement equations defined in (10) and (11) to obtain the full-information likelihood function for our HMXL model (for ease of exposition, we stack the parameter vectors $\boldsymbol{\Lambda}, \boldsymbol{\Xi}, \boldsymbol{\tau}, \boldsymbol{\zeta}, \boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\rho}_1, \boldsymbol{\upsilon}_1, \boldsymbol{\rho}_2, \boldsymbol{\upsilon}_2, \boldsymbol{\alpha}$ into the single vector $\boldsymbol{\Omega}$):

---

[15] Many early hybrid choice model applications used a simple, linear regression even in cases where the dependent variable was clearly ordered (e.g., Daly *et al.*, 2012). Although we specify this measurement equation as an ordered probit, nothing precludes one from assuming a different error distribution, giving rise to alternative ordered choice models (e.g., ordered logit).

[16] We assume $\alpha_0 = -\infty$ and $\alpha_5 = \infty$.

$$L_i = \int P\left(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{X}_i^{str}, \boldsymbol{\beta}_i^*, \lambda_i^*, \boldsymbol{\xi}_i^*, \boldsymbol{\Omega}\right) P\left(I_i \mid \mathbf{X}_i^{str}, \boldsymbol{\xi}_i^*, \mathbf{X}_i^{Mea}, \boldsymbol{\Omega}\right) \\ P\left(J_i \mid \mathbf{X}_i^{str}, \boldsymbol{\xi}_i^*, \mathbf{X}_i^{Mea}, \boldsymbol{\Omega}\right) f\left(\boldsymbol{\beta}_i^*, \lambda_i^*, \boldsymbol{\xi}_i^* \mid \boldsymbol{\theta}\right) d\left(\boldsymbol{\beta}_i^*, \lambda_i^*, \boldsymbol{\xi}_i^*\right). \tag{12}$$

As random disturbances of $\boldsymbol{\beta}_i^*, \lambda_i^*$ and (non-normalized) error terms in structural equations $\boldsymbol{\xi}_i^*$ are not directly observed, they must be integrated out of the conditional likelihood. This multidimensional integral can be approximated using a simulated maximum likelihood approach.[17] As can be seen, we use one-step estimation. This approach has two main advantages over a two-step (or multi-step) method. First, it is more efficient and, second, it allows for identification of more flexible specifications because it has more degrees of freedom.

Unfortunately, the exact conditions for HMXL identification are not known – this depends on the number of latent variables and measurement equations. We follow Bollen and Davis (2009) to ensure the necessary condition for identification of structural equation models holds; in particular, our specification satisfies the “2+ emitted paths rule” (we assume that each latent variable has exactly one unique indicator in the measurement equation and is linked with six preference parameters in the discrete choice component). In addition, we tested our model using simulations – we generated artificial datasets and validated our model by recovering the underlying parameters. Our model encountered no problems in identification and produces stable results.[18, 19]

[17] Our model assumes no correlations between error terms in the measurement, structural and choice components. This assumption has been relaxed in some related albeit simpler models (e.g., Bhat, Varin and Ferdous, 2010), and it is possible that allowing for some of these correlations can improve model performance or better address specification issues. The tradeoff in introducing correlations, however, is that this imposes estimation challenges to what is already a complex estimation problem.

[18] Econometric models estimated using maximum simulated likelihood are known to be relatively sensitive to starting values, optimization techniques and selection of convergence criteria. Our model is no exception in this respect and to make sure we reached the global maximum in optimization, we used different optimization methods, derived gradients analytically and used multiple starting points. In addition, since using longer low-discrepancy sequences (as opposed to shorter sequences or using pseudo-random draws) is found to facilitate reaching the global optimum or revealing identification problems (Chiou and Walker, 2007; Czajkowski and Budziński, 2015) in simulation of the log-likelihood function, we used 10,000 Sobol draws with a random linear scramble and a random digital shift.

[19] The models were estimated using a DCE package, which among other things can be used to estimate HMXL models. The package has been developed in Matlab and is available at https://github.com/czaj/DCE. The code and data for estimating the specific models presented in this study are available from http://czaj.org/research/supplementary-materials.

3. Stated preference survey and information treatments

3.1. Survey design

Our application focuses on the public financing of municipal theaters in Warsaw, Poland. The survey scenarios describe proposed programs for discounted theater tickets for Warsaw citizens. In particular the considered programs would result in a uniform ticket price of 5 Polish złoty (PLN) for a single admission,[20] for up to four different categories of theaters defined by the type of productions they offer. Entertainment theaters stage light comedies and music shows. Drama repertory theaters aim at combining reading classics with ambitious comedies. Children's theaters are focused on the youngest audience, while experimental theaters reject typical forms to introduce an alternative perspective into the performing arts. The proposed payment vehicle is an additional annual tax levied on citizens of Warsaw. Our use of a coercive payment vehicle is important from an incentive standpoint. Moreover, as all of about 50 municipal theaters in Warsaw are already subsidized by the city on the order of 40 to 60%, the scenarios can be seen as increasing the provision of a public good. The cost amounts considered are 10, 20, 50 and 100 PLN per year.

The development of the survey was informed by in-depth interviews and a pilot study, which confirmed respondents found it easy to distinguish between theater types, and that the tax payment vehicle was credible. Moreover, the majority of respondents had visited one or more types of theaters, suggesting that the possible policies described in the survey were straightforward to understand. A detailed description of the development of the study, the policy context and policy relevance of the results are provided in Wiśniewska and Czajkowski (2017).

With the objective of estimating the demand for programs involving varying opportunities for discounted theater tickets, we adopt a binary DCE approach. Specifically, each choice set includes the status quo (no program of theatre ticket reduction implemented) and one policy alternative.[21] Figure 1 presents an example choice set.[22] With four theater categories and

[20] During the time frame the survey was administered, 1 PLN ≈ 0.25 EUR ≈ 0.33 USD.

[21] Vossler, Doyon and Rondeau (2012) identify conditions for incentive compatibility of this elicitation approach, and through their complementary field experiment find favorable evidence of external validity conditional on policy consequentiality holding empirically. However, given the scant evidence available, there is not yet consensus regarding external validity of this preference elicitation format.

[22] The mean number of theater visits of the respondents in the 12 months before the survey was almost three, with approximately 20% of the sample not having visited a theater. This leads us to believe that respondents were familiar with the status quo, such as current ticket prices.

cost, there are five choice set attributes. With two levels for each theater attribute (i.e., the reduced ticket pricing is either offered or not), and four possible cost amounts, the associated full factorial design gives rise to $2^4 \times 4 = 64$ choice sets.

The policy alternatives were generated using Ngene software following a Bayesian (median) D-efficient design optimized for the MNL model (Scarpa and Rose, 2008), with priors for the choice parameters obtained from a pilot study administered to a sample of 119 respondents. In particular, we generated three "blocks" of 12 choice sets, and respondents were randomly assigned one of the blocks. To control for order effects, we randomized across respondents the order the choice sets were presented as well as the order in which the theater attributes appeared in the choice set.

The survey begins with "warm-up" questions gauging respondents' interest in culture and participation in cultural events, including past visits to Warsaw theaters. This is followed by a discussion of possible discounted theater programs, including descriptions of the four theater types, the proposed funding mechanism and a typical reminder about budget constraints. A sequence of 12 choice tasks is then presented, where respondents are instructed to indicate their preference between the proposed program and the status quo for each.

At the end of the survey, participants are asked to indicate their beliefs concerning policy consequentiality (hereafter, *stated consequentiality*), in addition to typical follow-up questions (e.g., gauging motivations for choices, attitudes towards the financing of Warsaw theaters, etc.). The wording of the policy consequentiality question is (translated from Polish): "To what extent do you agree with the statement that the results of the survey will influence future decisions regarding financing municipal theaters in Warsaw?". The level of respondents' agreement is measured on a five-point Likert scale (from 1 – "definitely disagree" to 5 – "definitely agree").

3.2. Experimental information treatments

In a field setting such as ours, it is not possible to vary the *actual* consequentiality of the survey as it pertains to its actual role in formulating policy. Instead, to the extent that survey design can influence beliefs over consequentiality, we introduce four treatments to examine the effect of consequentiality scripts, which vary the degree of emphasis placed on the potential role that the survey plays in informing policy regarding discounted theater programs. Table 1 presents the information provided in the various treatments. Each respondent participated only in one, randomly assigned treatment.

Consistent with survey design principles (Dillman, Smyth and Christian, 2014), it is common for surveys to emphasize the importance of the individuals' response, as well as the social usefulness of the survey, which often includes statements that results will be provided to government officials. In other words, common survey practice involves using language that may promote beliefs of policy consequences. In accordance with standard practice, regardless of a treatment, the preamble script indicates that the survey is being conducted by researchers of the University of Warsaw and that the purpose of the survey is to collect information on opinions regarding the financing of Warsaw theaters. Treatment 1 (T1) does not provide any additional information. The survey instrument in T2 states at the very beginning of the survey that respondents' choices may influence future policies. As such, either T1 or T2 is likely to characterize the typical written (or verbally articulated) information related to policy consequentiality provided in most stated preference surveys.

The remaining treatments, T3 and T4, further emphasize potential policy influence. Similar to common advertising techniques, these treatments involve repetition and alternative phrasing in attempt to increase awareness and importance. In addition to the information provided in T2, and in the context of providing background information relevant for the proposals, T3 reminds respondents of possible ties to actual policymaking. T4 includes all the information from T3 and adds an inducement immediately before the value elicitation questions and after the budget constraint reminder. We note that none of the information provided is deceptive. Relevant institutes (e.g., Office of Culture of Warsaw) were made aware of the study and our results were disseminated to them.

3.3. Survey implementation

The survey instrument is the outcome of extensive pretesting including individual interviews with potential survey respondents (verbal protocols) and a pilot study performed on a group of 119 Warsaw citizens. The survey was administered by a professional polling agency using Computer Assisted Web Interviews. Screening questions were used to restrict respondents to include only adult citizens of Warsaw who live and pay taxes in Warsaw. Quotas were implemented in order to obtain a close match of the treatment subsamples with the adult population of Warsaw with respect to gender, age, education and household size.

Data was collected in February and March 2014. In total, there are 1,700 respondents.[23] Detailed socio-demographic characteristics for each treatment sample and the adult population of Warsaw are presented in Table 2. As evident from the table, the targeted sampling strategy worked well to match the survey sample and population in terms of key demographics. However, since we cannot rule out important differences in unobservable factors between the sample and the population, we make no claim regarding the sample representativeness. We nevertheless have clean identification of information treatment effects given random treatment assignment.

4. Results

Using data from the stated preference survey, we apply the HMXL model detailed in Section 2. The structure of the model is illustrated in Figure 2 and the estimation results for the primary specification are presented in Table 3.[24] Overall, the model explicitly incorporates the link between the information treatments, *stated consequentiality*, two latent variables and respondents' observed preferences, and possibly also their socio-demographic characteristics into one, jointly estimated, framework.

The first latent variable considered, *latent consequentiality*, is the belief over policy consequentiality of the survey. The second latent variable, *latent theater use*, represents respondents' unobserved profile of theater use as it pertains to proposed scenarios. The ordered probit measurement equation links *latent consequentiality* with respondents' *stated consequentiality*, while the Poisson regression links *latent theater use* with the reported number of theater visits in the past 12 months.[25]

[23] As reported by the polling agency, the response rate to our survey was 29%.

[24] It is somewhat common in the literature to modify analyses based on protest respondents, such as those who indicate an unwillingness to pay any amount in response to some unanticipated characteristic of the valuation exercise, such as incomplete information. Adopting common convention in experimental economics, we do not eliminate participants – all were treated as legitimate respondents and left in the sample. We note that 16.18% of respondents consistently selected the no-cost status quo alternative; however, only 3.47% could be classified as protest zero responses when taking their stated motives into account. As a robustness check, Model 1.5 in the Online Appendix (available at http://czaj.org/research/supplementary-materials) presents the results from our primary specification when these possible protest responses are removed.

[25] Model 1.1 in the Online Appendix presents a parallel model where instead of controlling for the influence of the theater use profile via a latent variable, we include two indicators of theater use directly (whether a respondent made any theater trip at all and the log of the number of theater visits if a respondent made at least 1 trip). We find this specification to outperform many other non-linear direct use specifications. None of the alternative specifications qualitatively change our results.

The discrete choice model component of the HMXL explains stated choices from the binary DCE as a function of the attributes, *latent consequentiality*, *latent theater use* and the information treatments. Given that each of the DCE attributes has two levels, we use binary indicators for discounted theater tickets in a given theatre type. An indicator, *status quo*, is included to allow for differences in unobserved utility associated with choosing the status quo option relative to a proposed program. The latent variables and the information treatments appear as linear functions of the random coefficients tied to the policy attributes and the status quo indicator, as well as in an exponential function defining the random coefficient λ_i. The variation in information treatments is collapsed into a single linear variable and normalized (i.e., the original variable taking on the values 1 through 4, indicating the treatments from T1 to T4, respectively, is normalized for zero mean and unit standard deviation), and henceforth referred to as *consequentiality script*.[26]

The first panel of Table 3 presents the results for the ordered probit measurement equation for *stated consequentiality*. In addition to the estimates of the thresholds of the ordered probit model, we find that *latent consequentiality* significantly increases respondents' reported level of agreement with the statement that the survey results will indeed be used in designing future policies. This result thus suggests there is a causal link between latent beliefs over policy consequentiality and stated assessments of it. At the same time, we find an insignificant relationship between *consequentiality script* and *stated consequentiality*. Congruent with this result, as illustrated in Table 7, the raw data suggests that the information treatments did not systematically alter stated beliefs. In particular, using a Pearson chi-squared test, we fail to reject the null hypothesis that the frequency distributions for the four treatments are equal ($\chi^2_{(12)} = 6.759$, $p = 0.873$).

In the second panel of Table 3, the results for the Poisson measurement equation for reported theater trips are reported. We observe that *latent theater use* is positively correlated with the reported number of theater visits, thus lending itself to being interpreted as unobserved theater use intensity.[27]

---

[26] Model 1.2 in the Online Appendix presents a parallel model where instead of treating *consequentiality script* as a continuous variable, we include the information treatments into the choice model as separate treatment-specific indicators. The results justify the use of a single (linear, normalized) variable, which allows for easier interpretation of our results. This specification does not qualitatively change our results.

[27] We consider the possibility that the latent variable we label as *latent consequentiality* in fact captures some other latent construct, which would also manifest itself by giving higher scores to the question about survey consequentiality. One thing we can test for with our data is whether being a theater enthusiast (an active user) has

The third panel of Table 3 reports the results of the discrete choice (MXL) component of the model. The model is specified in WTP space and, accordingly, preference parameters can be directly interpreted in terms of WTP differences, in 100 PLN.[28] Noting that the expected value of *latent consequentiality* and *consequentiality script* are normalized to equal zero, the main effect for a particular theater type is the estimate of the mean change in WTP associated with inclusion of the theater type in the proposed program. Specifically, the model suggests that respondents, on average, are willing to pay the most – 33 PLN more per year – for a program that includes discounted tickets to entertainment theaters, followed by 21 PLN for drama repertory theaters and about 10 PLN for children's and experimental theaters each. The main effect corresponding with the status quo option is less significant, although the large and significant estimated standard deviation of this random parameter suggests substantial heterogeneity in individual preferences in this respect. In fact, all estimated standard deviations are both relatively large and statistically different from zero, providing evidence of significant preference heterogeneity and, accordingly, justification for the mixed logit relative to the restrictive multinomial logit.

From the perspective of evaluating the influence of consequentiality on respondents' observed preferences, the most interesting results are tied to the effects that *latent consequentiality* and *consequentiality script* have on the WTP for program attributes. For all theater types, the interactions with *latent consequentiality* are statistically significant and positive, indicating that a stronger belief in policy consequentiality is correlated with higher WTP for these attributes. Stronger latent beliefs over policy consequentiality has the additional effect of decreasing the scale of the unobserved component of utility (λ_i).

The interactions with *consequentiality script* show that emphasizing the potential role surveys have in shaping new policies significantly increases WTP for entertainment theaters and drama theaters, decreases the scale of unobserved utility and has a weak effect for reducing

such an effect. To explore this, we use the responses about individuals' past use of theaters as an explanatory variable in the measurement equation for *stated consequentiality* – either through the inclusion of *latent theater use* (model 1.3 in the Online Appendix) or directly, using two indicators of theater use (model 1.4). The latter specification performs better. We find that respondents who are more active theater users tend to respond that they perceive the survey as more consequential. Even with this effect controlled, however, our main results remain qualitatively the same.

[28] The cost parameter instead represents the scale of the unobserved component of utility $(\lambda_i = \sigma_i a_i)$ and hence does not have a money-metric interpretation. The cost parameter is assumed be log-normally distributed, but the reported results are the mean and the standard deviation of the underlying normal distribution, which are expressed in negative 100 PLN.

the preferences for the status quo alternative. These qualitative findings thus largely mirror those based on *latent consequentiality*. However, the magnitudes of the effects of information signals are much smaller. One interpretation of this result is that beliefs over policy consequences may largely be "homegrown" and entrenched, providing little room for the researcher to significantly influence them. Further, given our earlier finding of no significant correlation between *stated consequentiality* and *consequentiality script*, the implication here is that the Likert-scale question did not capture adequately the effects of the information treatments. One possibility is that the five-point Likert scale we used may not be precise enough to identify important distinctions.

To place the effects of the latent beliefs and the information treatments into better perspective, Tables 4, 5 and 6, and accompanying Figure 3 provide the simulated WTP changes associated with *consequentiality script*, *stated consequentiality* and *latent consequentiality*, respectively. The results tied to *latent consequentiality* are quite dramatic. Even when comparing the WTP for program attributes across respondents in the 25th and 75th percentiles, WTP increases by a factor of two for children's theaters, by a factor of three for drama theaters, by a factor of five for entertainment theaters and by a factor of four for experimental theaters.

Although *stated consequentiality* does not directly appear in the discrete choice component of the model, through simulation we can determine how WTP varies based on stated beliefs over policy consequentiality. These results are provided in Table 5. Overall, the results suggest that considerable information on preferences is conveyed through this simple indicator for policy consequentiality. Indeed, WTP for program attributes is negative or statistically zero at the two lowest levels of *stated consequentiality* and become positive and larger for higher levels. This positive relationship between stated consequentiality and elicited WTP mirrors a common but not universal finding in the literature (e.g., Herriges *et al.*, 2010). Similar to Vossler, Doyon and Rondeau (2012), we find a continuous relationship, rather than the "knife-edge" distinction between those with *in*consequential beliefs and those with consequential beliefs (regardless of intensity). Such a "knife-edge" result is predicted by the theory and found in the field survey studies of Herriges *et al.* (2010) and Vossler and Watson (2013). One possible explanation for this divide is the elicitation format; i.e., the latter two studies involved SBC rather than binary DCE elicitation.

In Figure 4, we present the effect of *latent theater use* on WTP. Generally, the effect is around two to three times weaker than the effect of *latent consequentiality*, although still significant for all attributes, except for children's theaters. Importantly, even for individuals whose *latent theater use* is in the bottom 2.5%, i.e., for the ones who with high probability will not visit any theater after program implementation, we still observe a positive WTP for most attributes. Experimental theaters are the exception, and this type of entertainment may only be valuable to theater users.

Finally, we note that while moving from a MXL to a HMXL model allows for richer specifications, potentially avoiding estimation biases, the analyst must make additional choices as it pertains to the specification of the measurement and structural equations. In order to help assess the sensitivity of our results to these specification choices, as well as to gain further insight, we estimate several additional models and include them in the Online Appendix.[29] Given that there are no meaningful differences in the qualitative results relative to those in our primary model, this provides evidence that the previously illustrated results regarding *latent consequentiality* are not simply an artefact driven by a particular model specification. In fact, the results of all our models consistently point to the importance of *stated consequentiality*, *latent consequentiality* and *consequentiality script*, while demonstrating that these three variables are not necessarily fully correlated nor easily measured.

5. Discussion

Using a novel Hybrid Mixed Logit (HMXL) model, this study provides primary empirical evidence that, consistent with mechanism design theory, latent beliefs regarding the policy consequentiality of a binary DCE survey are an important driver of stated preferences. In fact, we find the effects to be very pronounced, with beliefs supporting policy consequentiality leading to substantially higher estimates of willingness to pay. Importantly, we observe that the emerging approach for identifying beliefs over policy consequentiality – i.e., a simple Likert-scale survey question – provides a measure that is, in fact, strongly correlated with latent beliefs. Recognizing that the direct inclusion of stated measures of the beliefs in choice models can give rise to econometric issues, for example measurement error,

[29] In particular, models 2.0 and 3.0 in the Online Appendix present the results of specifications which include socio-demographic controls as explanatory variables in the measurement equations, and as additional latent variables capturing the combined effects of socio-demographic characteristics on preferences (Pakalniete *et al.*, 2017). We have considered other specifications of the count data model, namely zero inflated Poisson, negative binomial and zero inflated negative binomial. The results of these specifications were not, however, reliable enough for presenting here, indicating convergence problems we were not able to trace.

the HMXL model provides an alternative, flexible framework from which to tie stated measures and preferences.

We further find that the use of consequentiality scripts, which emphasize the potential role of surveys in formulating policy, lead to meaningful increases in elicited WTP. However, the magnitude of the effect on WTP is small in comparison to the effect of latent consequentiality beliefs. That our information treatments have only a small effect suggests some possibilities. First, it may be the case that in this and other field applications beliefs over policy consequentiality naturally emerge. Stated preference surveys are often of high quality, relate to socially important topics and describe plausible policy interventions. It thus may be instinctive for respondents to believe that, even if left unstated, policymakers are funding the study or otherwise care about its outcomes. Second, even in settings where beliefs over policy consequentiality are weak, the potential role that consequentiality scripts may play is limited. That is, barring some exceptional circumstances, survey researchers cannot assure respondents that results will absolutely inform policy nor precisely define the mechanism by which this would occur. These arguments, in turn, suggest that it may be interesting to explore consequentiality in opportunistic settings. This may include settings where it is possible to establish and convey a link between survey results and policy decisions (see Johnston, 2006). It is further possible that, at least for some people and some settings, beliefs about potential citizen involvement in policy making may be very entrenched, and researchers are thus unlikely able to alter them.

It bears noting that stated beliefs about policy consequences were somewhat weak in this study, with just 5.82% of respondents indicating they definitely agree, 21.21% stating they rather agree and almost half (49.53%) neither in agreement or disagreement with the statement that the results of the survey will influence future decisions. Klamer (2016) argues that economists have had little influence on cultural policy, and thus our respondents appear to have reasonable beliefs. From a theory standpoint, respondents need only have a non-zero probability of influencing agency action. On the other hand, an elicitation mechanism may only work as expected if this probability is high enough. This remains an important empirical issue.

Our exploration of policy consequentiality targets just one of the beliefs identified by theory as important for truthful preference revelation. Another important, latent belief is that payment can be coercively collected from respondents upon program implementation. This assumption has received less attention, although efforts have begun to explore the use of follow-up questions to elicit information tied to this belief. The HMXL model developed here provides an estimation framework that allows the analyst to incorporate many latent beliefs simultaneously. It accommodates the joint modelling of choices and responses to a multitude of survey questions charged with measuring beliefs, regardless of whether these questions are targeted towards the same belief or different sets of beliefs. In fact, related to the development of belief elicitation questions, the HMXL model can be used to determine the magnitude of the relationship between elicited and latent beliefs, thus providing a platform from which to discriminate and evaluate alternative approaches. The econometric framework further provides a way to determine the relative importance of distinct latent factors – tied to different beliefs relevant for incentive compatibility or otherwise – on elicited values.

In implementing hybrid models it is common for researchers, as we have done here, to hypothesize that certain latent factors exist, proceed to measure them based on available data, and interpret results as if they are in fact driven by the hypothesized latent variables. We note, however, that there are multiple possible interpretations of the latent variables in these models and our particular interpretations are thus subjective. The measurement variables may inadequately explain the latent variables or otherwise the latent variables identified in a model may be attributable to factors not considered by the analyst. From a technical perspective, in our model what we label and interpret as *latent consequentiality* is merely *some* unobserved construct that is positively correlated with responses to the Likert-scale survey question on policy consequences and positively correlated with the WTP for attributes of the proposed discounted theater ticket programs. For instance, the latent factor could instead be tied to forms of social desirability bias (e.g., yea-saying or nay-saying), or internally motivated expressions of attitudes that are disjoint from unobserved demand for proposed projects.

Other econometric approaches that incorporate stated beliefs, such as instrumental variables methods or simply treating belief indicators as exogenous variables, are of course not immune to interpretation issues. 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. The literature on consequentiality has relied on a single question, whereas it is more common in the broad literature on hybrid models to use multiple indicators to measure a latent factor. The use of multiple questions gives more opportunities for researchers to support or refute their proposed interpretation. Related to this, to help address concerns over social desirability bias or internal motivations, one can construct questions assumed to be correlated with for example yea-saying and incorporate these other possible factors as separate latent variables (with associated measurement variables) within our proposed econometric framework.

References

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D., Daly, A., De Palma, A., Gopinath, D., Karlstrom, A., and Munizaga, M., 2002. Hybrid Choice Models: Progress and Challenges. *Marketing Letters*, 13(3):163-175.

Bhat, C. R., Dubey, S. K., and Nagel, K., 2015. Introducing non-normality of latent psychological constructs in choice modeling with an application to bicyclist route choice. *Transportation Research Part B: Methodological*, 78:341-363.

Bhat, C. R., Varin, C., and Ferdous, N., 2010. A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered-response model. In: *Advances in Econometrics*, W. Greene and R. C. Hill, eds., Emerald Group Publishing Limited, 65-106.

Bollen, K. A., and Davis, W. R., 2009. Two Rules of Identification for Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3):523-536.

Budziński, W., and Czajkowski, M., 2017. Addressing Endogeneity in Hybrid Choice Models. Paper presented at the Conference of the European Association of Evironmental and Resource Economists, Athens, Greece.

Bulte, E., Gerking, S., List, J. A., and de Zeeuw, A., 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(2):330-342.

Carson, R., and Czajkowski, M., 2013. A New Baseline Model for Estimating Willingness to Pay from Discrete Choice Models. Paper presented at the International Choice Modelling Conference, Sydney, Australia.

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

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

Chiou, L., and Walker, J. L., 2007. Masking identification of discrete choice models under simulation methods. *Journal of Econometrics*, 141(2):683-703.

Collins, J. P., and Vossler, C. A., 2009. Incentive compatibility tests of choice experiment value elicitation questions. *Journal of Environmental Economics and Management*, 58(2):226-235.

Czajkowski, M., and Budziński, W. (2015). "An insight into the numerical simulation bias – a comparison of efficiency and performance of different types of quasi Monte Carlo simulation methods under a wide range of experimental conditions." In: *Environmental Choice Modelling Conference*, Copenhagen.

Czajkowski, M., Hanley, N., and Nyborg, K. 2017. Social Norms, Morals and Self-interest as Determinants of Pro-environment Behaviours: The Case of Household Recycling. *Environmental and Resource Economics*, 66(4):647-670.

Daly, A., Hess, S., Patruni, B., Potoglou, D., and Rohr, C., 2011. Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour. *Transportation*, 39(2):267-297.

Daziano, R. A., and Bolduc, D., 2013. Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. *Transportmetrica A: Transport Science*, 9(1):74-106.

Dekker, T., Hess, S., Brouwer, R., and Hofkes, M., 2016. Decision uncertainty in multi-attribute stated preference studies. *Resource and Energy Economics*, 43:57-73.

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

Drichoutis, A. C., Vassilopoulos, A., Lusk, J. L., and Nayga, R. M., Jr. (2015). "Reference Dependence, Consequentiality and Social Desirability in Value Elicitation: A Study of Fair Labor Labeling." Paper presented at the 143rd Joint EAAE-AAEA Seminar, Naples, Italy.

Forrest, D., Grime, K., and Woods, R., 2000. Is it worth subsidising regional repertory theatre? *Oxford Economic Papers*, 52(2):381-397.

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

Grisolía, J., and Willis, K., 2012. A latent class model of theatre demand. *Journal of Cultural Economics*, 36(2):113-139.

Grisolía, J. M., and Willis, K. G., 2010. An evening at the theatre: using choice experiments to model preferences for theatres and theatrical productions. *Applied Economics*, 43(27):3987-3998.

Guevara, C. A., and Ben-Akiva, M., 2010. Addressing Endogeneity in Discrete Choice Models: Assessing Control-Function and Latent-Variable Methods. In: *Choice Modelling: The State-of-the-Art and the State-of-Practice. Proceedings from the Inaugural International Choice Modelling Conference*, S. Hess and A. Daly, eds., Emerald Group Publishing Limited, 353-370.

Hansen, T., 1997. The Willingness-to-Pay for the Royal Theatre in Copenhagen as a Public Good. *Journal of Cultural Economics*, 21(1):1-28.

Herriges, J., Kling, C., Liu, C.-C., and Tobias, J., 2010. What are the consequences of consequentiality? *Journal of Environmental Economics and Management*, 59(1):67-81.

Hess, S., and Beharry-Borg, N., 2012. Accounting for Latent Attitudes in Willingness-to-Pay Studies: The Case of Coastal Water Quality Improvements in Tobago. *Environmental and Resource Economics*, 52(1):109-131.

Hess, S., Hensher, D. A., and Daly, A., 2012. Not bored yet – Revisiting respondent fatigue in stated choice experiments. *Transportation Research Part A: Policy and Practice*, 46(3):626-644.

Hess, S., and Stathopoulos, A., 2013. Linking response quality to survey engagement: A combined random scale and latent variable approach. *Journal of Choice Modelling*, 7:1-12.

Hoyos, D., Mariel, P., and Hess, S., 2015. Incorporating environmental attitudes in discrete choice models: An exploration of the utility of the awareness of consequences scale. *Science of The Total Environment*, 505:1100-1111.

Johnston, R. J., 2006. Is hypothetical bias universal? Validating contingent valuation responses using a binding public referendum. *Journal of Environmental Economics and Management*, 52(1):469-481.

Jöreskog, K. G., and Goldberger, A. S., 1975. Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable. *Journal of the American Statistical Association*, 70(351a):631-639.

Kamargianni, M., Ben-Akiva, M., and Polydoropoulou, A., 2014. Incorporating social interaction into hybrid choice models. *Transportation*, 41(6):1263-1285.

Kim, J., Rasouli, S., and Timmermans, H., 2014. Expanding scope of hybrid choice models allowing for mixture of social influences and latent attitudes: Application to intended purchase of electric cars. *Transportation Research Part A: Policy and Practice*, 69:71-85.

Klamer, A., 2016. The value-based approach to cultural economics. *Journal of Cultural Economics*, 40(4):365-373.

Kling, C., Phaneuf, D. J., and Zhao, J., 2012. From Exxon to BP: Has Some Number Become Better than No Number? *Journal of Economic Perspectives*, 26(4):3-26.

Landry, C. E., and List, J. A., 2007. Using Ex Ante Approaches to Obtain Credible Signals for Value in Contingent Markets: Evidence from the Field. *American Journal of Agricultural Economics*, 89(2):420-429.

McFadden, D., 1974. Conditional Logit Analysis of Qualititative Choice Behaviour. In: *Frontiers in Econometrics*, P. Zarembka, ed., Academic Press, New York, NY, 105-142.

Mitani, Y., and Flores, N. E. (2012). "Robustness Tests of Incentive Compatible Referenda: Consequential Probability, Group Size, and Value-cost Difference." Paper presented at the Conference of the European Association of Environmental and Resource Economists, Prague, Czech Rebuplic.

Oehlmann, M., and Meyerhoff, J., 2017. Stated Preferences Towards Renewable Energy Alternatives in Germany – Do the Consequentiality of the Survey and Trust in Institutions Matter? *Journal of Environmental Economics and Policy*, 6(1):1-16.

Pakalniete, K., Aigars, J., Czajkowski, M., Strake, S., Zawojska, E., and Hanley, N., 2017. Understanding the distribution of economic benefits from improving coastal and marine ecosystems. *Science of The Total Environment*, 15:29-40.

Raveau, S., Yáñez, M. F., and Ortúzar, J. d. D., 2012. Practical and empirical identifiability of hybrid discrete choice models. *Transportation Research Part B: Methodological*, 46(10):1374-1383.

Scarpa, R., and Rose, J. M., 2008. Design Efficiency for Non-Market Valuation with Choice Modelling: How to Measure it, What to Report and Why. *Australian Journal of Agricultural and Resource Economics*, 52(3):253-282.

Scarpa, R., Thiene, M., and Train, K., 2008. Utility in Willingness to Pay Space: A Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps. *American Journal of Agricultural Economics*, 90(4):994-1010.

Vij, A., and Walker, J. L., 2016. How, when and why integrated choice and latent variable models are latently useful. *Transportation Research Part B: Methodological*, 90:192-217.

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

Vossler, C. A., and Evans, M. F., 2009. Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality. *Journal of Environmental Economics and Management*, 58(3):338-345.

Vossler, C. A., and Watson, S. B., 2013. Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. *Journal of Economic Behavior & Organization*, 86(0):137-147.

Vredin Johansson, M., Heldt, T., and Johansson, P., 2006. The effects of attitudes and personality traits on mode choice. *Transportation Research Part A: Policy and Practice*, 40(6):507-525.

Willis, K. G., and Snowball, J. D., 2009. Investigating how the attributes of live theatre productions influence consumption choices using conjoint analysis: the example of the National Arts Festival, South Africa. *Journal of Cultural Economics*, 33(3):167-183.

Willis, K. G., Snowball, J. D., Wymer, C., and Grisolía, J., 2012. A count data travel cost model of theatre demand using aggregate theatre booking data. *Journal of Cultural Economics*, 36(2):91-112.

Wiśniewska, A., and Czajkowski, M., 2017. Designing a socially efficient cultural policy: the case of municipal theaters in Warsaw. *International Journal of Cultural Policy*:1-13.

Table 1. Information treatments tied to policy consequentiality

Information statement	T1	T2	T3	T4
At the very beginning of the survey: "Your answers may be used in planning future policies of financing theaters in Warsaw."		included	included	included
By the general information on financing the city theaters: "Warsaw citizens might have an impact on the work of city theaters [as they are subsidized from the city budget]. This survey aims at finding out Your opinion on the city theaters, and its results will allow to determine how the city of Warsaw should finance them."			included	included
Near the description of the proposals: "Your answers in this survey will allow us to assess whether introduction of such a program is a good idea."			included	included
Directly before the value elicitation exercise, presented in a frame: "ATTENTION – Please, remember that Your answers might be used for planning the future financing of Warsaw theaters. In the case of an introduction of any policy, which can be influenced by the survey results, the changes will involve all citizens (including You)."				included

Note: all passages translated from Polish.

Table 2. Socio-demographic characteristics of survey information treatment samples and target population

Variable	T1	T2	T3	T4	Adult population of Warsaw
Female (proportion)	0.55 [0.50]	0.52 [0.50]	0.53 [0.50]	0.55 [0.50]	0.54
Age (in years)	44.90 [16.79]	43.32 [16.00]	43.73 [16.46]	43.71 [15.83]	43.92
Education attainment (share)					
Elementary school	0.07 [0.25]	0.06 [0.23]	0.06 [0.23]	0.07 [0.25]	0.16
Vocational pre-high school	0.09 [0.29]	0.12 [0.32]	0.09 [0.28]	0.09 [0.29]	
High school	0.30 [0.46]	0.31 [0.46]	0.32 [0.47]	0.35 [0.47]	0.43
Vocational school	0.11 [0.32]	0.12 [0.32]	0.12 [0.32]	0.10 [0.30]	
University	0.43 [0.49]	0.40 [0.49]	0.41 [0.49]	0.40 [0.49]	0.41
Household size (individuals)	2.97 [1.20]	2.98 [1.22]	2.86 [1.21]	3.03 [1.18]	2.87
Number of respondents	446	416	419	419	

Note: standard deviations appear in square brackets.

Table 3. HMXL reference model

	Measurement equation 1 (ordered probit) dependent variable: *stated consequentiality*	
	Coefficient	Std. err.
Latent consequentiality	0.1489***	0.0382
Consequentiality script	-0.0300	0.0260
Threshold parameters:		
α_1	-1.6132***	0.0506
α_2	-0.7342***	0.0178
α_3	0.6181***	0.0214
α_4	1.5885***	0.0670
	Measurement equation 2 (Poisson regression) dependent variable: *number of theater visits in the last 12 months*	
	Coefficient	Std. err.
Constant	0.7938***	0.0271
Latent theater use	0.7874***	0.0251

Willingness to pay model (mixed logit)

	Means – main effects	**Means – interactions with *latent consequentiality***	**Means – interactions with *latent theater use***	**Means – interactions with *consequentiality script***	**Standard deviations**
	Coefficient (s.e.)	Coefficient (s.e.)	Coefficient (s.e.)	Coefficient (s.e.)	Coefficient (s.e.)
Status quo	0.0285** (0.0140)	0.0010 (0.0161)	-0.1041*** (0.0143)	0.0232* (0.0124)	0.4294*** (0.0147)
Entertainment theaters	0.3278*** (0.0121)	0.3122*** (0.0119)	0.0997*** (0.0135)	0.0378*** (0.0098)	0.0372*** (0.0111)
Drama theaters	0.2096*** (0.0100)	0.1640*** (0.0129)	0.0616*** (0.0135)	0.0314*** (0.0088)	0.1151*** (0.0163)
Children's theaters	0.1085*** (0.0094)	0.0394*** (0.0121)	0.0089 (0.0120)	0.0044 (0.0087)	0.1592*** (0.0101)
Experimental theaters	0.0982*** (0.0093)	0.0864*** (0.0120)	0.0532*** (0.0109)	-0.0020 (0.0084)	0.1683*** (0.0079)
Cost (*scale*)	2.1947*** (0.0645)	-0.5084*** (0.0854)	-0.2791*** (0.0610)	-0.1692*** (0.0452)	1.0534*** (0.0718)

Model diagnostics	
Log-likelihood (constant only)	-20,529.90
Log-likelihood	-15,035.96
McFadden's pseudo R^2	0.2676
AIC/*n*	1.4778
n (observations)	20,400
k (parameters)	38

Note: *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.

Table 4. Simulated marginal willingness to pay (PLN per year) associated with information treatments [95% confidence interval]

	Status quo	*Entertainment theaters*	*Drama theaters*	*Children's theaters*	*Experimental theaters*
Consequentiality script = 1	-0.30 [-4.23 , 3.64]	27.67 [24.42 , 30.91]	16.71 [13.81 , 19.62]	10.25 [7.47 , 13.03]	10.10 [7.45 , 12.73]
Consequentiality script = 2	1.77 [-1.01 , 4.57]	31.02 [28.62 , 33.43]	19.50 [17.45 , 21.55]	10.64 [8.73 , 12.55]	9.92 [8.05 , 11.78]
Consequentiality script = 3	3.84 [0.76 , 6.93]	34.38 [31.76 , 37.00]	22.29 [20.13 , 24.45]	11.04 [8.99 , 13.09]	9.74 [7.69 , 11.79]
Consequentiality script = 4	5.91 [1.36 , 10.45]	37.74 [34.02 , 41.46]	25.08 [21.92 , 28.21]	11.43 [8.37 , 14.53]	9.56 [6.53 , 12.59]

Table 5. Simulated marginal willingness to pay (PLN per year) associated with *stated consequentiality* levels [95% confidence interval]

	Status quo	*Entertainment theaters*	*Drama theaters*	*Children's theaters*	*Experimental theaters*
Stated consequentiality = 1	1.75 [-31.71 , 35.41]	-305.50 [-330.28 , -281.00]	-156.71 [-183.51 , -129.79]	-31.82 [-56.90 , -6.68]	-83.85 [-108.96 , -58.68]
Stated consequentiality = 2	2.05 [-22.17 , 26.38]	-213.34 [-231.23 , -195.63]	-108.30 [-127.67 , -88.89]	-20.20 [-38.31 , -2.03]	-58.33 [-76.51 , -40.13]
Stated consequentiality = 3	2.81 [0.18 , 5.47]	20.61 [18.37 , 22.83]	14.56 [12.65 , 16.49]	9.31 [7.54 , 11.08]	6.45 [4.64 , 8.27]
Stated consequentiality = 4	3.61 [-20.91 , 27.97]	264.14 [245.95 , 282.52]	142.47 [123.03 , 161.77]	40.03 [21.81 , 58.21]	73.89 [55.83 , 91.87]
Stated consequentiality = 5	3.94 [-30.87 , 38.49]	365.89 [340.14 , 391.93]	195.90 [168.21 , 223.40]	52.87 [27.01 , 78.72]	102.06 [76.29 , 127.66]

Table 6. Simulated marginal willingness to pay (PLN per year) associated with *latent consequentiality* [95% confidence interval]

	Status quo	*Entertainment theaters*	*Drama theaters*	*Children's theaters*	*Experimental theaters*
2.5th percentile	2.65 [-3.31 , 8.65]	-28.42 [-32.87 , -24.01]	-11.18 [-15.93 , -6.44]	3.13 [-1.26 , 7.56]	-7.12 [-11.61 , -2.63]
25th percentile	2.79 [-0.14 , 5.72]	11.72 [9.35 , 14.09]	9.90 [7.71 , 12.07]	8.19 [6.18 , 10.21]	3.99 [1.91 , 6.06]
50th percentile	2.85 [0.11 , 5.60]	32.78 [30.41 , 35.14]	20.96 [19.00 , 22.93]	10.85 [9.02 , 12.68]	9.82 [7.99 , 11.65]
75th percentile	2.92 [-1.06 , 6.89]	53.84 [50.58 , 57.10]	32.02 [29.07 , 34.98]	13.50 [10.72 , 16.30]	15.66 [12.93 , 18.38]
97.5th percentile	3.05 [-4.45 , 10.58]	93.97 [88.16 , 99.81]	53.10 [47.26 , 58.95]	18.57 [13.04 , 24.07]	26.77 [21.37 , 32.17]

Table 7. Cross-tabulation of information treatments versus *stated consequentiality* – number of cases

		Stated consequentiality					
		1	2	3	4	5	Total
Consequentiality script	1	21	70	204	103	21	419
	2	24	75	207	87	26	419
	3	24	75	200	92	25	416
	4	25	83	231	80	27	446
	Total	94	303	842	362	99	1700

	Alternative A	Alternative B (continuation of the current policy)
Entertainment theaters	No change	No change
Drama repertory theaters	Tickets for 5 PLN	No change
Children's theaters	No change	No change
Experimental theaters	Tickets for 5 PLN	No change
Annual cost for You	100 PLN	0 PLN
Your choice	□	□

Figure 1. An example choice set (translated from Polish)

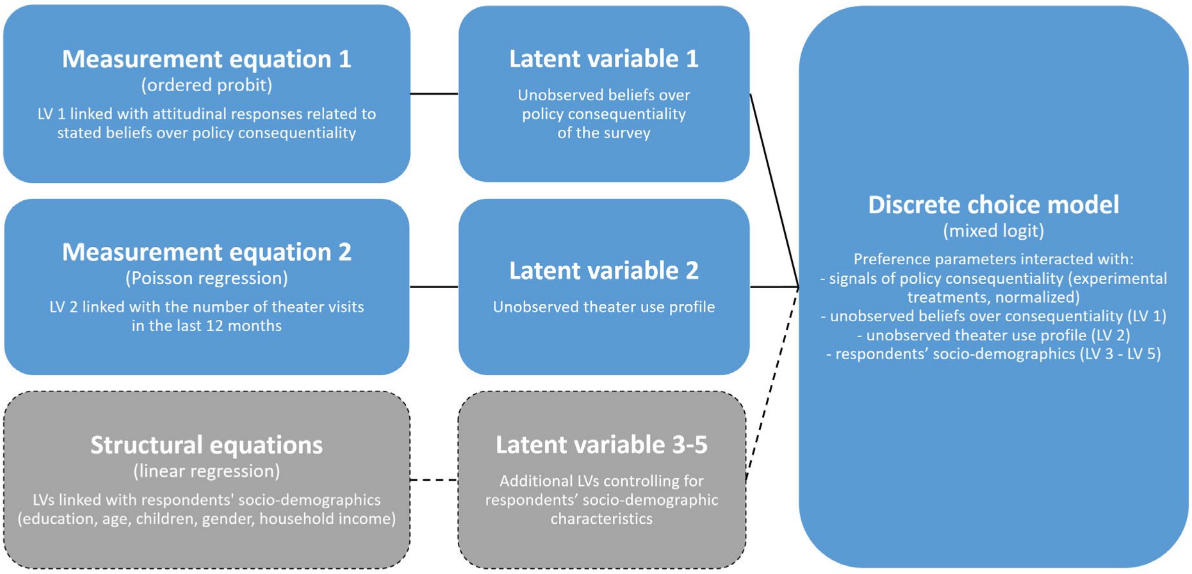
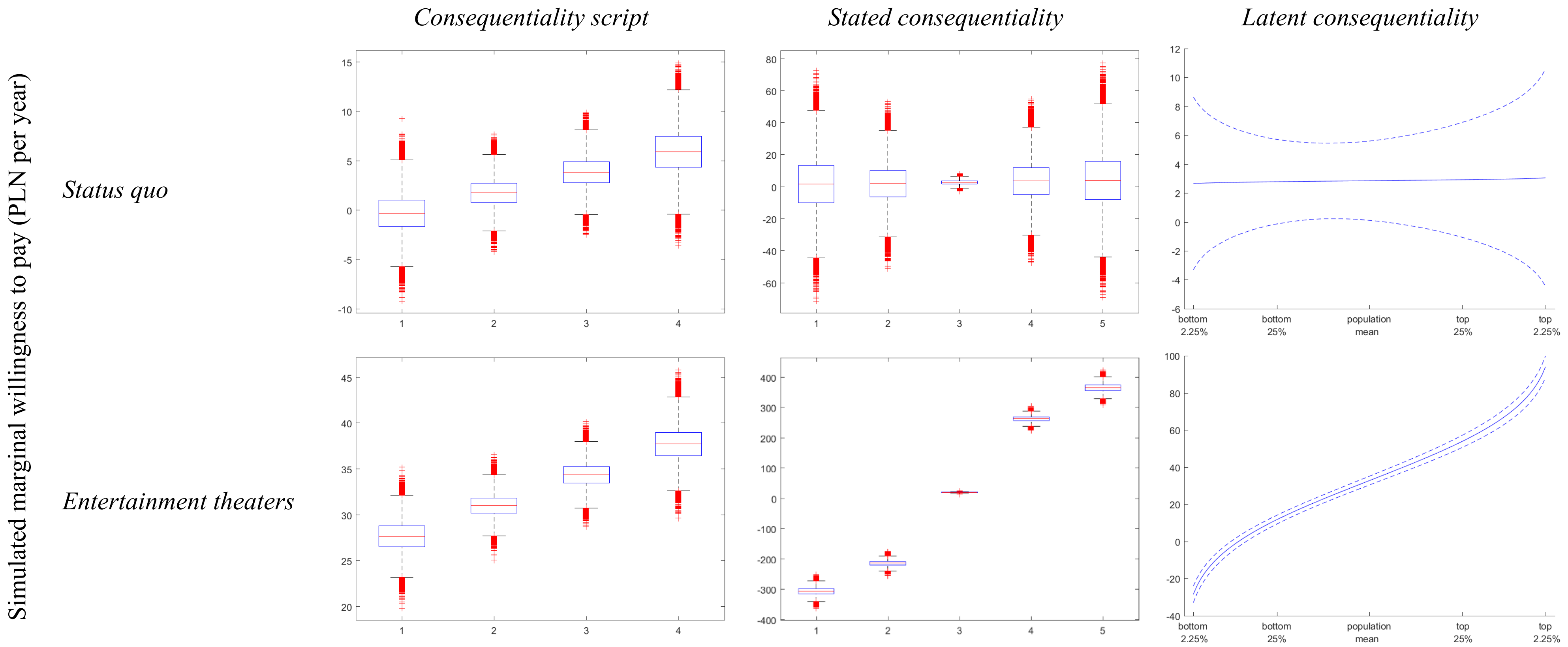
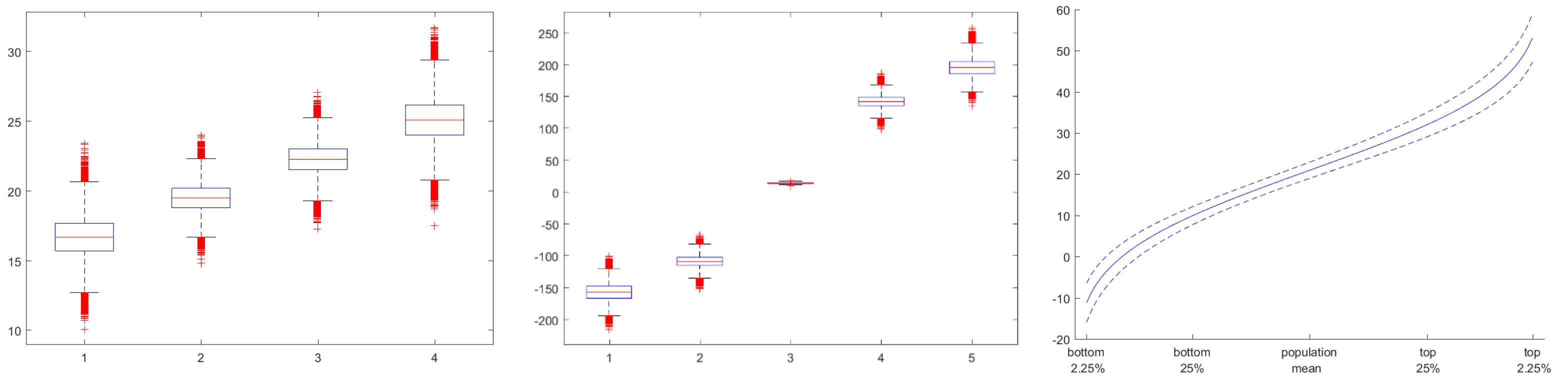


Figure 2. Overview of the HMXL model framework used for analyzing stated preferences for proposed discounted theater programs

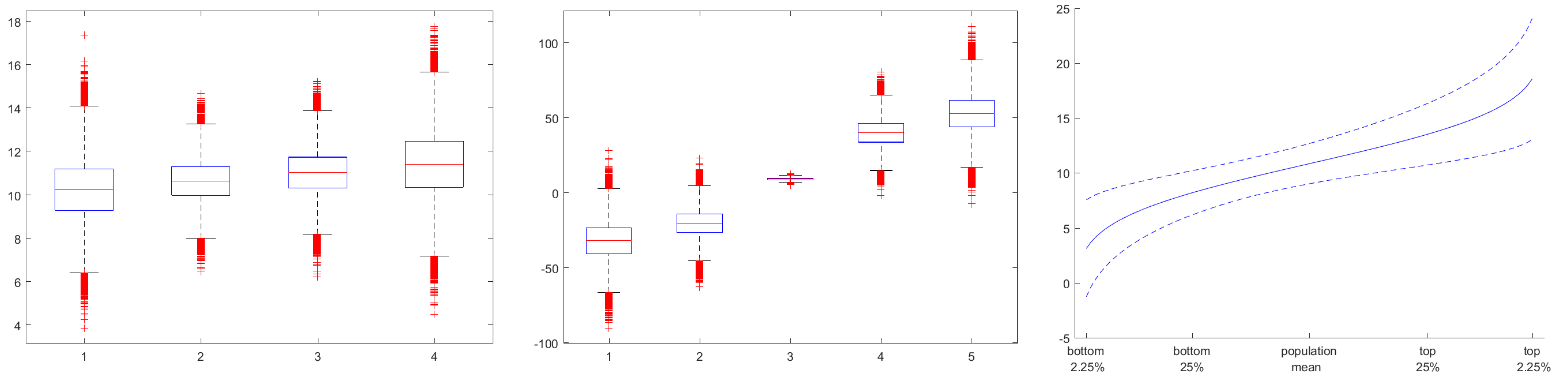
Figure 3. Simulated marginal willingness to pay for different levels of *consequentiality script*, *stated consequentiality* and *latent consequentiality*



Drama theaters



Children's theaters



Experimental theaters

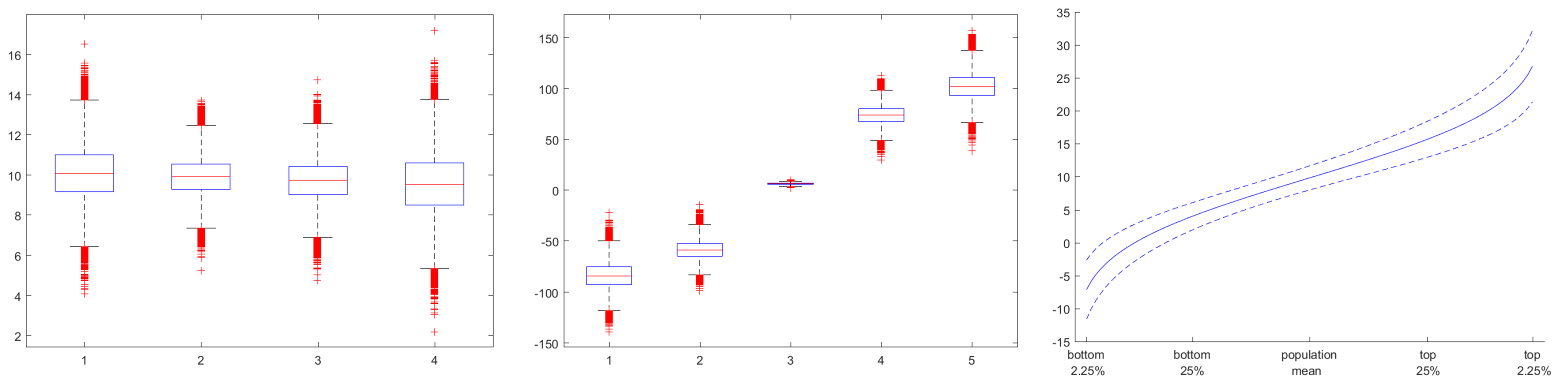
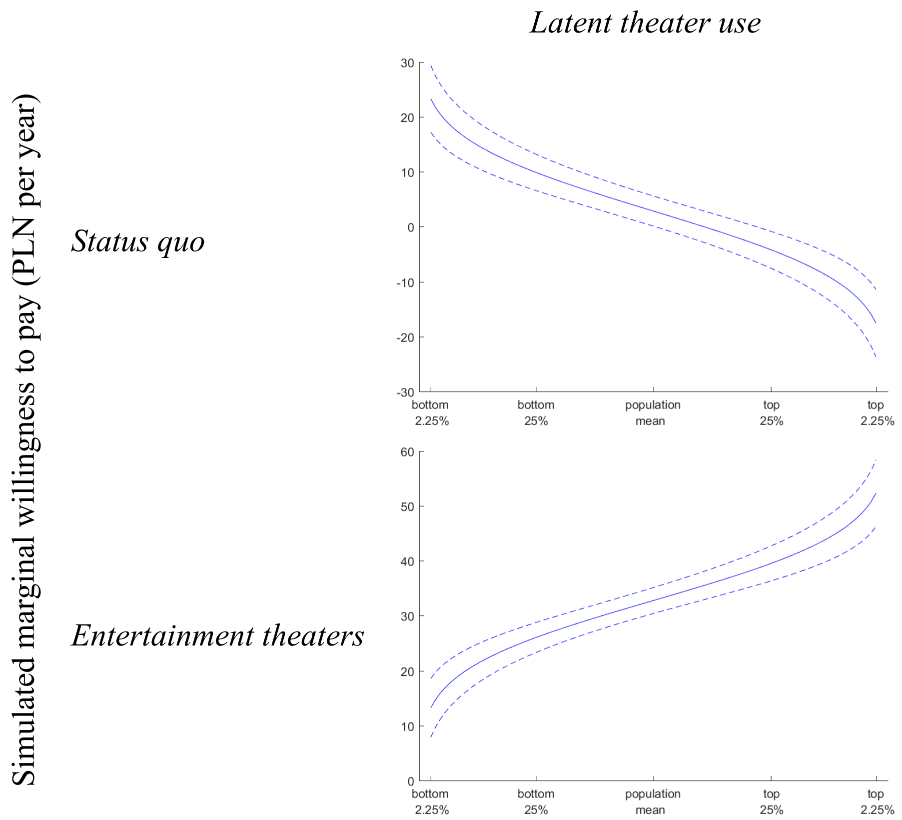
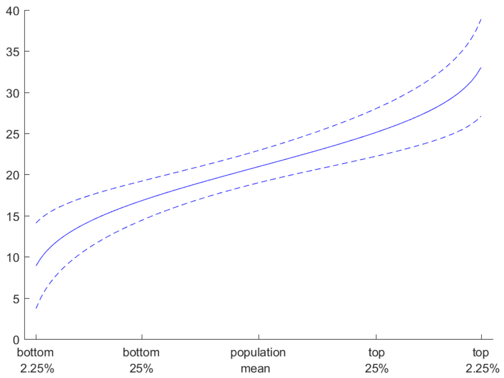


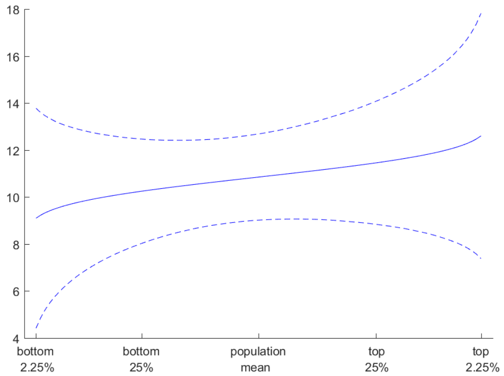
Figure 4. Simulated marginal willingness to pay for different levels of *latent theater use*



Drama theaters



Children's theaters



Experimental theaters