Tourism Management 78 (2020) 104067

Contents lists available at ScienceDirect

Tourism Management

journal homepage: www.elsevier.com/locate/tourman

Explaining variability in tourist preferences: A Bayesian model well suited to small samples

Lendie Follett [a,1], Brian Vander Naald [b,1,*]

[a] Department of Information Management and Business Analytics, Drake University, Des Moines, IA, United States of America
[b] Department of Economics and Finance, Drake University, Des Moines, IA, United States of America

ARTICLE INFO

Keywords:
Bayesian estimation
Discrete choice experiment
LKJ distribution
Inverse Wishart distribution
Small samples

ABSTRACT

Discrete choice experiments are becoming more popular in the tourism and travel literature. While Bayesian methods to analyze discrete choice experiment data have been used in other disciplines, they have not been used in the tourism literature. In this article, we develop a Bayesian Mixed Logit Model in which we use a little known prior distribution developed by Lewandowski, Kurowicka, and Joe (LKJ) and half Cauchy distributions as an alternative to the more traditionally used inverse Wishart distribution as a prior scheme for the covariance matrix of random parameters in mixed logit estimation. Using multiple simulated data sets, we show that use of the LKJ prior scheme improves the estimation of coefficients, especially for small data sets. Finally, we test the model with an actual small discrete choice data set examining tourist preferences for reducing glacier recession, and discuss the implications of the model for research and policy.

1. Introduction

Over the past decade, discrete choice experiments (DCEs) have become increasingly popular in the tourism literature (Kim & Park, 2017; Kubo, Mieno, & Kuriyama, 2019; Landauer, Pröbstl-Haider, & Haider, 2012). When the market for a good does not exist, or when a good is a bundled composite of different attributes such as in many tourism applications (Chen, Masiero, & Hsu, 2019), the DCE is a sophisticated way to disentangle the roots of consumer preferences. DCEs are also widely used across many other disciplines, including economics (Johnston et al., 2017), health (de Bekker-Grob, Donkers, Jonker, & Stolk, 2015), marketing (Bryant & Hill, 2019), transportation (Hensher, 2010), and travel (Adhikari, 2015).

In recent years, the standard statistical approach used to estimate preference relationships from DCE data has been the mixed logit model (MLM). The MLM affords many benefits over other DCE estimation strategies, such as relaxing the IIA assumption and accounting for unobserved heterogeneity in the sample. Estimating the MLM using classical econometric methods means using a maximum simulated likelihood (MSL) algorithm. MSL relies on information provided by the likelihood along with asymptotic properties. As a result, complete separation (Albert & Anderson, 1984) can occur when analyzing small-sample DCE data sets using MSL, implying that the maximum likelihood estimates are unbounded and do not exist.

Yet despite the falling cost of collecting DCE data, small samples are not uncommon. In a survey of 69 DCEs, de Bekker-Grob et al. (2015) found 32% of them had samples of less than 100 respondents. Indeed, if research requires sampling a small population, the sample itself is likely to be small (Grijalva, Berrens, & Shaw, 2011; Huth & Morgan, 2011). A vast majority of the time, preference relationships from those small samples have been estimated using classical econometrics, which as mentioned above could be problematic. One potential solution to this problem is to use Bayesian methods to estimate preference relationships in small DCE samples. Bayesian methods of estimation are desirable in this context because (1) inference does not rely on asymptotic theory, (2) the introduction of prior information stabilizes estimates, especially in small samples, and (3) posterior distributions allow researchers to answer a richer set of questions.

The Bayesian approach has been used for other common methodologies in the tourism literature. Song, Qiu, and Park (2019) provide a comprehensive review of the methods that have been used in predicting tourism demand and note that Bayesian methods are promising. In particular, placing informative priors on regression coefficients has been shown to improve forecasting accuracy over classical methods that use unpenalized coefficient estimates. Wong, Song, and Chon (2006) employed variations of the Minnesota prior in a Bayesian vector autoregression (BVAR) model to improve forecasts of the demand for Hong Kong tourism. Gunter and Önder (2015) found similar success with

* Corresponding author.
E-mail address: brian.vandernaald@drake.edu (B. Vander Naald).
[1] Follett and Vander Naald contributed equal effort to this manuscript.

https://doi.org/10.1016/j.tourman.2019.104067
Received 25 September 2019; Received in revised form 13 December 2019; Accepted 14 December 2019
Available online 23 December 2019
0261-5177/

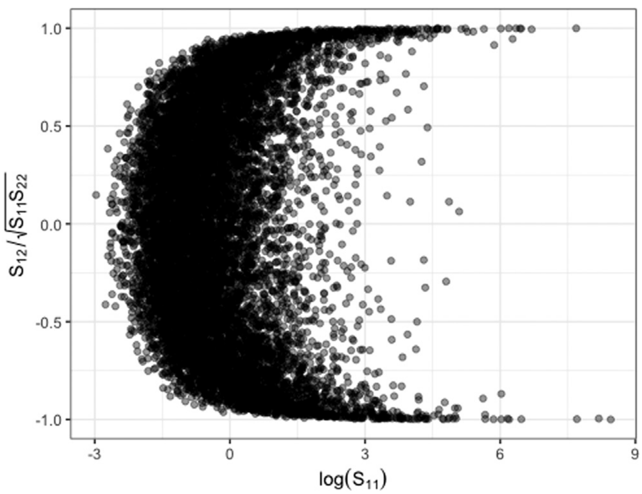

Fig. 1. Scatterplot of $log(S_{11})$, the log of the first variance component, by the implied correlation from 10,000 draws from an inverse Wishart ($k = 3, I_2$) distribution.

BVAR models in predicting tourism at select markets in Paris. More recently, Assaf, Li, Song, and Tsionas (2019) introduced the Bayesian Global VAR (BGVAR) as a way to simultaneously use priors and consider the interdependencies of multivariate tourism demands. Bayesian forecasting methods have also been used outside the vector autoregression framework. In predicting Hong Kong tourism demand, Wu, Law, and Xu (2012) used a sparse Gaussian process regression model along with Bayesian estimation to achieve probabilistic inference, as well as greater accuracy.

Bayesian methods have also been employed in structural equation models (SEMs). Assaf, Oh, and Tsionas (2016) propose a Bayesian finite mixture model as a general framework for SEMs. This allows the researcher to quantify unobserved heterogeneity. In a related work, a simulation study showed consistent superiority of the Bayesian SEMs over classical SEMs in terms of parameter bias across multiple sample sizes (Assaf, Tsionas, & Oh, 2018). Papastathopoulos, Ahmad, Al Sabri, and Kaminakis (2019) used Bayesian methods to quantify how demographics impact the relationship between tourism perception and support for tourism development. Importantly, this work notes that the Bayesian framework offers a natural way to use non-normal distributions for the data, resulting in potentially better representations of the data generating process.

To our knowledge, Bayesian methods have yet to be demonstrated for DCE analysis in the tourism literature. The use of Bayesian estimation to analyze DCE data began gaining in popularity around the same time Train (2003) generalized its use for that purpose. As computers have become more powerful, other researchers have employed Bayesian estimation techniques, especially in economics (Balcombe, Chalak, & Fraser, 2009; Rigby, Balcombe, & Burton, 2009; Rigby & Burton, 2006). Common across the aforementioned studies is the assignment of an inverse Wishart prior to the variance–covariance matrix in mixed logit estimation, which has historically been the default prior choice. That default prior assignment can be problematic when there is not a good reason for choosing it. Recently, Akinc and Vandebroek (2018) developed a Bayesian MLM, which they use to show that the choice of prior distribution in Bayesian estimation is important. Using a large simulated data set, they find that the default use of the inverse Wishart prior on the variance is inappropriate because it can exert too much influence on the results and cause biased estimates. Instead, they use a flexible prior for correlation matrices developed by Lewandowski, Kurowicka, and Joe (2009), which has desirable properties.

We extend the work of Akinc and Vandebroek (2018) and contribute to the nascent literature discussing the deliberate choice of prior distribution in Bayesian analysis. In particular, we develop a Bayesian MLM in which competing priors are assigned to the variance–covariance matrix of the random parameters. We then explore the model using both simulated and real data gathered during a 2015 DCE. We find that our model is able to reduce coefficient bias for both small and large data sets, with the bias reduction particularly pronounced for small data sets. We are also able to estimate higher order interaction effects to tease apart differences in willingness to pay (WTP) along sociodemographic lines, which would not be possible using MSL.

The rest of the paper is structured as follows: In Section 2, we present a general Bayesian model framework for analyzing DCE data. We then test the general model with a simulation study in Section 3, after which we extend the model for an application incorporating sociodemographic heterogeneity in Section 4. Section 5 discusses the results, and Section 6 concludes.

2. Model

In this section, we present a general model of consumer choice given a set of alternatives. Let $i = 1, \ldots, N$ represent individual, $j = 1, \ldots, J$ represent alternative, and $s = 1, \ldots S_i$ represent situation. Let $\boldsymbol{x}_{ijs}$ represent a K-dimensional vector of alternative j specific non-price attribute levels for situation s seen by individual i. If Y_{ijs} represents the random variable associated with the ith individual's choice regarding the jth alternative in the sth scenario, the data model can be specified as:

$$P(Y_{ijs} = 1 \mid \eta_{ijs}) = \frac{exp(\eta_{ijs})}{\sum_{j=1}^{J} exp(\eta_{ijs})} \tag{1}$$

where

$$\eta_{ijs} = \boldsymbol{x}_{ijs}^T \boldsymbol{\beta}_i. \tag{2}$$

We allow for preference heterogeneity across individuals by setting

$$\boldsymbol{\beta}_i \overset{ind}{\sim} N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \tag{3}$$

where $\boldsymbol{\mu}_\beta$ is a K-dimensional vector and $\boldsymbol{\Sigma}_\beta$ is a $K \times K$ covariance matrix. Thus, $\boldsymbol{\mu}_\beta$ is the overall mean of the coefficient vector and $\boldsymbol{\Sigma}_\beta$ describes the covariance. Often, $\boldsymbol{\Sigma}_\beta$ is constrained to have 0 elements on the off diagonals, implying independence across the parameters. Mariel and Meyerhoff (2018) detail several reasons why forcing correlations of zero may be inappropriate, including the risk of introducing bias into the coefficient vector. Thus, we aim to allow the elements of $\boldsymbol{\Sigma}_\beta$ to be estimated freely in a data-driven way.

Traditionally, $\boldsymbol{\Sigma}_\beta$ is assigned an inverse Wishart prior, e.g., $\boldsymbol{\Sigma}_\beta \sim iW(\nu, \boldsymbol{S})$ where ν is the degrees of freedom and $\boldsymbol{S}$ is a $k \times k$ scale matrix. This distribution is flexible in that it allows for non-zero correlations. The inverse Wishart distribution has been desirable as a prior on covariance matrices in multivariate Gaussian models at least partially because the resulting (conditional) posterior is conjugate and, thus, easily sampled from. However, conjugacy is no longer a restricting factor with sophisticated software available (Plummer, 2003; Stan Development Team, 2018). Akinc and Vandebroek (2018) describe several characteristics of the inverse Wishart distribution that may make it undesirable for use as a prior distribution on a covariance matrix. First, when we specify that a covariance matrix $\boldsymbol{\Sigma} \sim iW(\nu, \boldsymbol{S})$, we are placing a distinct correlation pattern between the variance and implied correlation. This is illustrated in Fig. 1. *A priori*, we are specifying a relationship where small correlations occur only with small variances. Thus, if the variance is large, the correlation will be pulled towards an extreme (-1 or 1) as well. Second, an inverse Wishart distribution implies, marginally, inverse Gamma priors on the variance components. The inverse gamma places very little weight on values close to zero and can be quite informative in some situations as is illustrated in Gelman et al. (2013). In particular, it is worrisome when the variances are truly small as they will be inflated in the posterior because of the prior influence rather than that of the data.

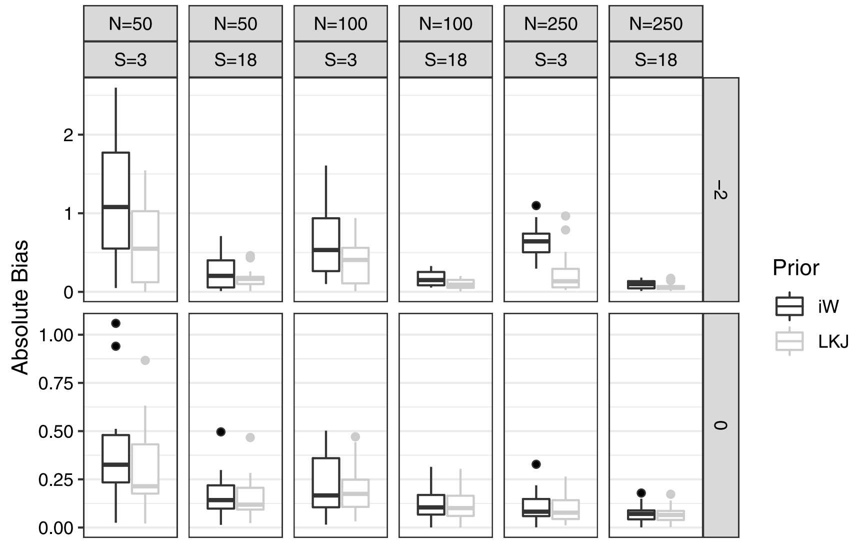

Fig. 2. Boxplots representing the distribution of absolute bias observed between the posterior means and the true values of $\boldsymbol{\mu}_\beta$ elements. Shown separately for parameters with true value -2 (top) and true value 0 (bottom).

We place a prior on the covariance matrix, $\boldsymbol{\Sigma}_\beta$, by placing independent priors on the lower triangular Cholesky factorization of the correlation matrix and the standard deviation terms. That is, we decompose the covariance matrix as

$$\boldsymbol{\Sigma}_\beta = diag(\sigma_{\beta 1}, \dots, \sigma_{\beta K})\boldsymbol{\Omega}\boldsymbol{\Omega}^T diag(\sigma_{\beta 1}, \dots, \sigma_{\beta K}). \tag{4}$$

This parameterization allows us to specify our prior beliefs about correlations between parameters ($\boldsymbol{\Omega}\boldsymbol{\Omega}^T$) independently from those about the variances of the parameters. The LKJ (Lewandowski et al., 2009) distribution is a convenient and flexible way to express prior beliefs about a correlation matrix. The LKJ(η) distribution is a distribution over all positive definite correlation matrices where the shape is determined by a single parameter, $\eta > 0$. If a correlation matrix, $\boldsymbol{\Omega} \sim LKJ(\eta)$, then $p(\boldsymbol{\Omega}) \propto |\boldsymbol{\Omega}|^{\eta-1}$. Setting $\eta = 1$ results in a uniform density over all positive definite correlation matrices. A $\eta < 1$ results in a density that is lowest at the identity matrix (independence) while a $\eta > 1$ results in the mode of the distribution positioned at the identity matrix. While one can place an LKJ prior directly on the correlation matrix, numerical problems can occur during the MCMC process. Placing the LKJ prior on the lower triangular Cholesky factor avoids these numerical instability problems. In our application, we set $\boldsymbol{\Omega} \sim LKJ(1)$. The joint prior of the unknowns is completed by specifying $\sigma_{\beta k} \sim$ Half Cauchy $(0, 2.5)$ independently for all k.

Gelman, Jakulin, Pittau, Su, et al. (2008) suggests placing independent student-t prior distributions on coefficients involved in generalized linear models and carefully choosing the degrees of freedom and scale parameters to reflect one's prior uncertainty or expectations. We impose weakly informative priors on the elements of $\boldsymbol{\mu}_\beta$ by taking the degrees of freedom to be 1 and the scale to be 1. That is, each of the coefficients listed above are assigned independent Cauchy(0,1) prior distributions.

We have discussed some of the known drawbacks of the inverse Wishart distribution and Akinc and Vandebroek (2018) has explored the behavior of the inverse Wishart as compared to the LKJ prior scheme for $N = 200$ and $S = 18$. To further inform our choice of prior for the covariance matrix in our data application, we ran a simulation study to illustrate the behaviors of these two prior schemes as they relate to the bias introduced into the coefficient vector.

3. Simulation study

We compare the performance of the inverse Wishart and LKJ prior schemes by varying the sample sizes (N) and number of choice scenarios (S) in the survey and then examining how this affects the estimated parameter bias. We do this for $N \in \{50, 100, 250\}$ and $S \in \{3, 18\}$. This includes much of the upper and lower bounds of the distribution of sample sizes and number of alternatives used in practice (Birol, Karousakis, & Koundouri, 2006; Campbell, Boeri, Doherty, & Hutchinson, 2015; Collins & Vossler, 2009; Czajkowski, Vossler, Budziński, Wiśniewska, & Zawojska, 2017; de Bekker-Grob et al., 2015; Interis & Petrolia, 2016). Akinc and Vandebroek (2018) performed a similar simulation study comparing multiple parameterizations of the two prior schemes but fixing $N = 200$ and $S = 18$.

For each of the sample size scenarios, we simulate choices according to (1)–(2) where there are three alternatives ($J = 3$) described by three factors, each with three levels. These choice sets are generated using the R package `choiceDes`. We simulate random parameters $\{\boldsymbol{\beta}_i : i = 1, \dots, N\}$ according to (3) where we fix $\boldsymbol{\mu}_\beta^T = (-2, 0, -2, 0, -2, 0)$ and

$$\boldsymbol{\Sigma}_\beta = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0 & 1 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 & 1 \end{bmatrix}.$$

For the inverse Wishart prior, we set $\boldsymbol{\Sigma}_\beta \sim$ inverse Wishart$(K + 1, (K + 1)I_K)$ where I_K is a $K \times K$ identity matrix. For the LKJ prior scheme, we set $\mu_{\beta k} \overset{iid}{\sim} Cauchy(0, 1)$, $\sigma_{\beta k} \overset{iid}{\sim} Half\ Cauchy(0, 1)$ for $k = 1, \dots, 6$, and $\boldsymbol{\Omega} \sim LKJ(1)$.

Fig. 2 summarizes the absolute bias, calculated as $|\hat{\mu}_{\beta k} - \mu_{\beta k}|$, for each of 5 replications of each sample size scenario where $\hat{\mu}_{\beta k}$ represents the posterior mean of $\mu_{\beta k}$. We show the results separately for true values of -2 and 0 as the performance seems to depend on whether or not there is a true signal. As expected, each of the prior schemes benefits from a larger N and S. In addition, we find no practical difference between the estimated bias when $N = 250$ and $S = 18$. However, in surveys with either small sample sizes or a small number of choice scenarios, we see that the inverse Wishart clearly suffers more than the LKJ prior scheme in estimating non-zero signals ($\mu_{\beta k} = -2$). In fact, even when the number of respondents is high ($N = 250$), a small number of choice scenarios ($S = 3$) results in somewhat unstable estimates.

$\boldsymbol{\Sigma}_\beta$ controls, to a large extent, the values that the individual level coefficients are able to take on and, thus, the estimates for $\boldsymbol{\mu}_\beta$. Fig. 3 summarizes absolute bias for the covariance and variance elements of $\boldsymbol{\Sigma}_\beta$. Again, we show the bias separately depending on the true value of the parameter. We suspect the bias in the estimates for the

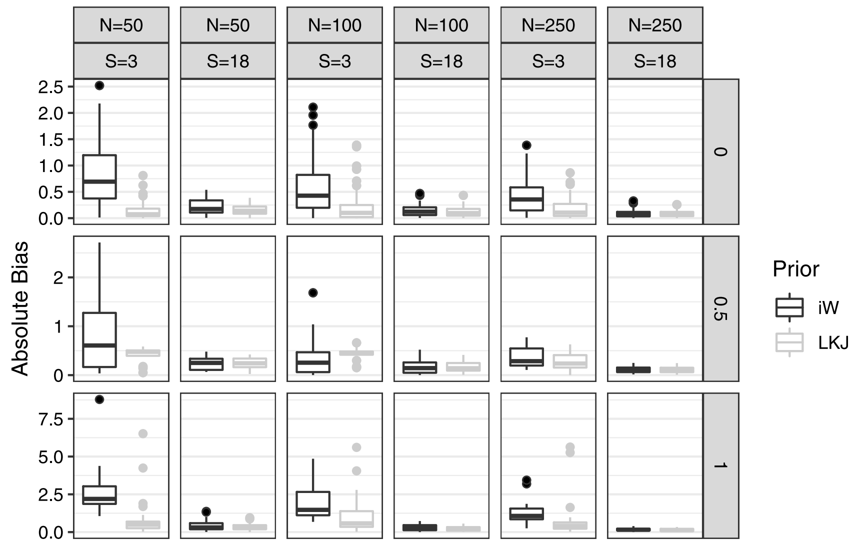

Fig. 3. Boxplots representing the distribution of absolute bias observed between the posterior means and the true values of Σ_β elements. Shown separately for covariance parameters with true value 0 (top), true value 0.5 (center), and 1 (bottom).

Table 1
Attribute levels used in survey.

Attribute	Levels used
Annual glacier loss (measured in volume)	0.15 km^3, 0.20 km^3, 0.25 km^3, 0.35 km^3
Insulating ice blanket	Yes, No
Climate agreement (CO_2 abatement)	None, Limited, Vigorous
Monthly program cost	\$0, \$5, \$10, \$20, \$25, \$40, \$50, \$60, \$100

mean vector μ_β is introduced as a result of the relatively restrictive nature of the inverse Wishart prior. In particular, it limits the ability of the variance components to reach small values, as opposed to the Half Cauchy distribution, which tends to shrink variance estimates towards smaller values. Fig. 4 illustrates the differences in posterior distributions resulting from the two prior schemes. While this is shown for $N = 250$ and $S = 3$, it is representative of the general pattern of outcomes one would see when comparing posteriors resulting from these two priors under small sample size situations. The posterior distribution of the variance components can be seen on the diagonal positions. In these we can see that the inverse Wishart consistently results in more posterior probability associated with larger variances. For several of the elements, we can see that most of the posterior mass is larger than the true value, which is indicated by red vertical lines. Compare this with Fig. 5, which shows the posterior distributions from a replication when $N = 250$ and $S = 18$. It takes a large amount of data to overwhelm a prior that has so little mass in potentially important areas of the joint parameter space.

This simulation study shows the effect of sample size (N and S) on bias. However, we limited the results to one parameterization of the inverse Wishart (inverse Wishart($(K+1), (K+1)I_K$)). It is likely, especially in small sample size scenarios, that the values of the hyperparameters will be highly influential on these results. Therefore, as an additional sensitivity analysis, we present the simulation study results for six commonly used inverse Wishart hyperparameter schemes in Figs. 6 and 7. Our findings agree with the Akinc and Vandebroek (2018) result that the performance of the inverse Wishart is sensitive to the values of the hyperparameters. We also find that the posterior distributions can be influenced by the scale parameter of the half Cauchy distribution, especially in small sample size scenarios. In this simulation study, we set the scale parameter to 1. However, much larger parameter values may allow for variance components that are much too large, which can be especially troubling in situations when the parameter estimates tend to veer off in the extremes (e.g., in cases where we have quasi-complete separation). Thus, we recommend setting the scale parameter to more reasonable values to shrink the standard errors towards 0.

4. An application - glacier recession and climate change in the U.S.

Next, we present an application with real DCE data to illustrate how researchers might use our model in a practical situation. The data for this application come from a DCE conducted in summer 2015 at the Mendenhall Glacier Visitor Center (MGVC) in Juneau, Alaska. The objective of the survey was to estimate tourists' willingness to pay (WTP) to slow the rate at which glaciers are receding, as well as the policies to achieve those outcomes. As such, the choice experiment design included four attributes, which are described along with their levels in Table 1. The survey was broken into three sections. The first section asked preliminary screening and salience questions. The second section presented respondents with either two or four choice scenarios in which they were asked to make trade offs between annual rates of glacier loss at a monthly cost to their household. The final section collected sociodemographic information. A sample choice scenario is shown in Fig. 8. The DCE was administered by pen and paper using random intercept sampling of cruise ship passengers at the MGVC during the peak of the summer 2015 tourist season. 166 surveys were administered and 149 were completed. In this application, we use a subset of 98 observations for which full sociodemographic information was reported. Of those 98 observations, 30 saw four choice scenarios and 68 saw two choice scenarios, which resulted in 1280 observations. Descriptive statistics are presented in Table 2. Further details of the survey and its development can be found in Vander Naald (2019).

4.1. Baseline model

We estimate an MLM using the inverse Wishart prior scheme, the LKJ prior scheme, and classical MSL for comparison. The models share the same data distribution specification (Eqs. (1)–(2)) and mixing distribution (Eq. (3)), so the priors (or lack thereof) on μ_β and Σ_β are what differentiate the three models.

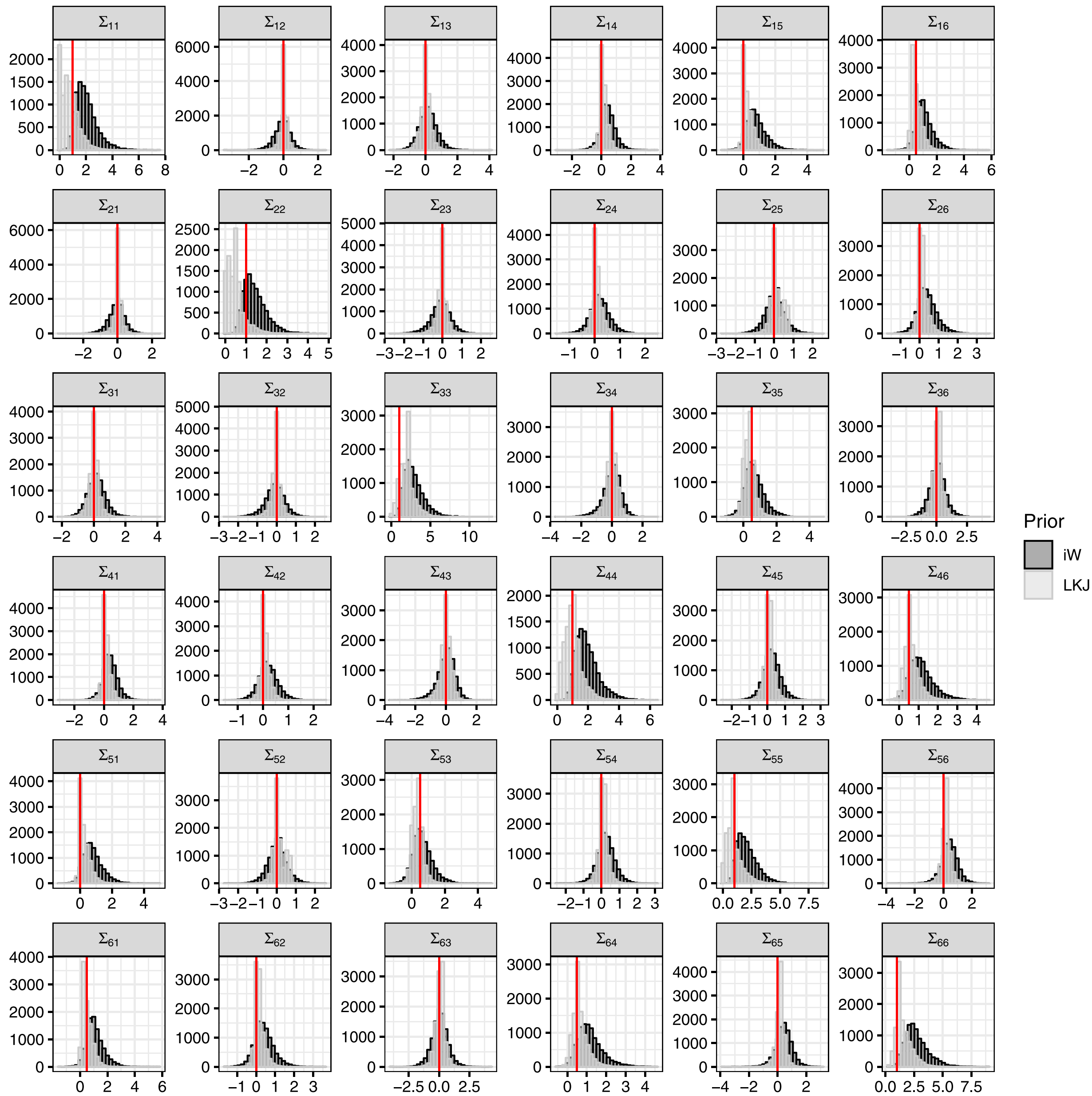

Fig. 4. Histogram of posterior draws from $\boldsymbol{\Sigma}_\beta$ from one representative repetition where $N = 250$ and $S = 3$.

Table 2
Summary statistics.

Variable	N	Mean	Std. Dev.	Min.	Max.
Age	98	60.76	12.60	24	82
Is female	98	0.49	0.50	0	1
Income	98	117.91	61.39	7.50	220
Is nonwhite	96	0.03	0.17	0	1
Politically liberal	98	0.33	0.47	0	1
Politically centrist	98	0.34	0.48	0	1
Is retired	98	0.47	0.50	0	1
Is unemployed	98	0.02	0.14	0	1
At least Bachelor's degree	98	0.70	0.46	0	1
Member of env. organization	98	0.46	0.50	0	1
Is aware cc caused by GHGs	98	0.99	0.10	0	1
Topic has high salience to respondent	98	0.67	0.47	0	1

Fig. 9 presents point and interval estimates for the three models. Importantly, the sign of the point estimates are consistent across the all models. This makes sense as we know that, for reasonably specified priors, Bayesian estimates will converge towards MSL estimates as the size of the data set increases.

In cases of magnitude discrepancies between the three models, the MSL estimates tend to be large in absolute value while the Bayesian posterior means are shrunk towards zero. In these cases, the classical 95% confidence intervals are typically quite wide, indicating a large standard error on the parameter. In unpenalized maximum likelihood methods for estimating parameters that lie near their space boundaries, these characteristics are symptomatic of complete separation (i.e., when the MSL estimates do not exist). Complete separation is particularly troublesome in small samples, like this one, that do not yield enough heterogeneity in the data to provide stable estimates.

The baseline model gives an idea of the preferences of the average person sampled. However, we aim to explain the variability between individuals by incorporating demographic characteristics. Since the classical analysis showed signs of complete separation without this added granularity, we now focus on Bayesian techniques which use priors to stabilize estimates.

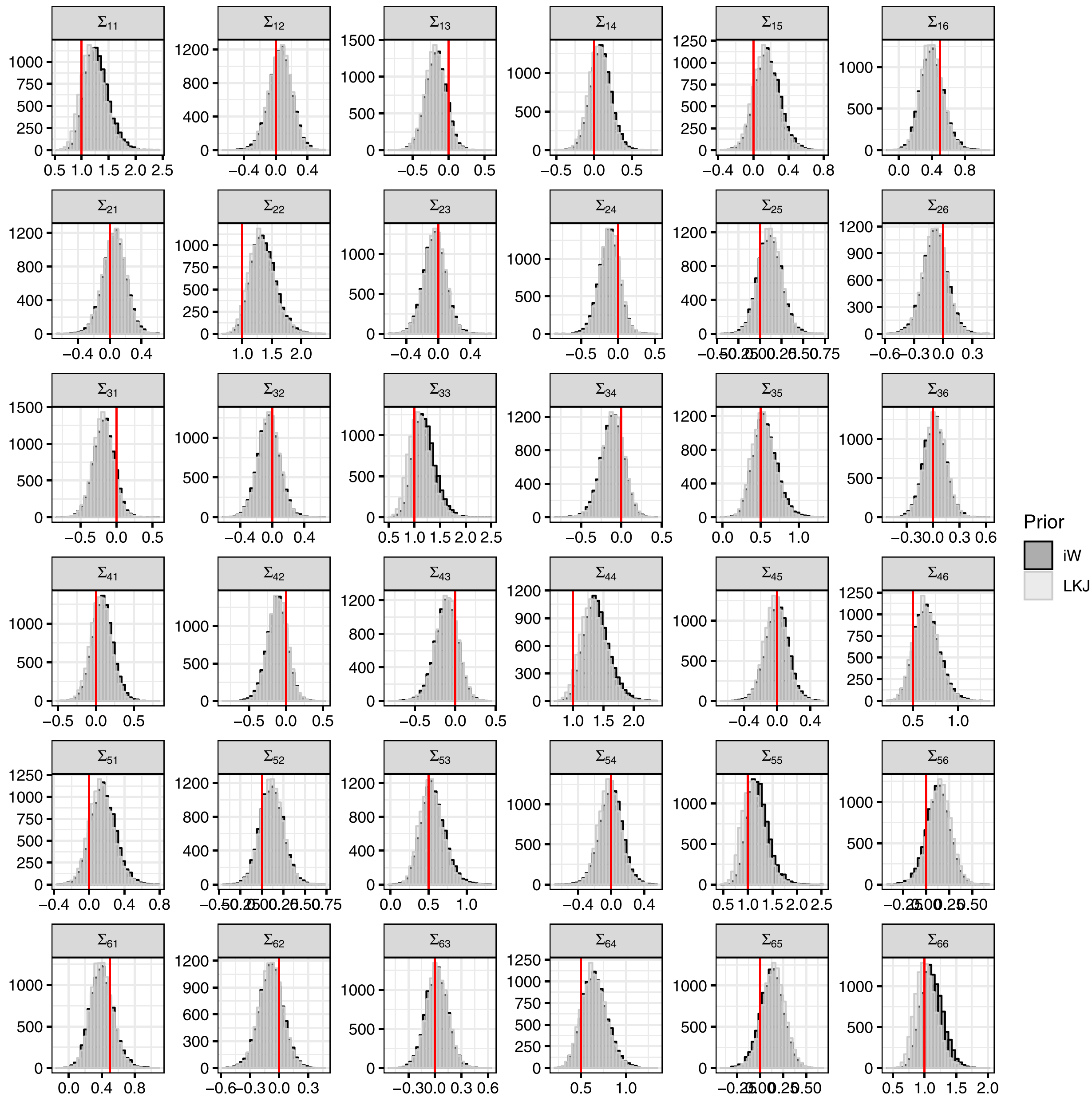

Fig. 5. Histogram of posterior draws from Σ from one representative repetition where $N = 250$ and $S = 18$.

4.2. Model incorporating demographics

We extend the previous model described in Eqs. (1)–(3) by (1) including price and status quo into η_{ijs}; and (2) by allowing some preference heterogeneity to be explained by demographic factors, leaving the remaining unexplained variability to be absorbed into $\boldsymbol{\Sigma}_\beta$. Let $price_{ijs}$ represent the price, and sq_{ijs} represent an alternative specific constant for the status quo scenario. Now, our linear predictor is:

$$\eta_{ijs} = \gamma_1 price_{ijs} + \gamma_2 sq_{ijs} + \mathbf{x}_{ijs}^T \boldsymbol{\beta}_i.$$

Inclusion of observed heterogeneity is informed by the climate change literature. O'Connor, Bord, Yarnal, and Wiefek (2002) find political divisions in preferences to mitigate climate change through public policy, but no political divide with respect to voluntary and private actions. Leiserowitz (2006) finds that females, members of environmental groups, and politically liberal respondents perceived climate change as a greater risk than males, respondents who were not members of environmental groups, and politically conservative respondents. Hence, we interact political preference, membership in an environmental organization, and gender with outcome and policy variables to see how much variation in WTP estimates they can help explain. Let $\mathbf{z}_i^T$ represent the vector of demographic information for individual i, including the chosen interaction terms. Available demographic information is as described in Table 2. We set

$$\boldsymbol{\beta}_i \overset{ind}{\sim} N(\boldsymbol{\theta}_0 + \boldsymbol{\Theta}\mathbf{z}_i, \boldsymbol{\Sigma}_\beta) \tag{5}$$

where $\boldsymbol{\Theta}$ is a matrix coefficients $\{\theta_{kh} : k = 1, \dots, K;; \quad h = 1, \dots, H\}$, $\boldsymbol{\theta}_0$ is a K-dimensional intercept vector, and $\boldsymbol{\Sigma}_\beta$ is a $K \times K$ covariance matrix. We impose weakly informative priors on the coefficients $\{\theta_{kh} : k = 0, 1, \dots, K; \quad h = 1, \dots, H\}$, γ_1, and γ_2 by assigning student-t distributions where we set the degrees of freedom to be 1 and the scale to be 1. Finally, we collect our unknowns into $\boldsymbol{\Xi} = \{\gamma_1, \gamma_2, \boldsymbol{\beta}, \theta_0, \boldsymbol{\Theta}, \boldsymbol{\Omega}, \boldsymbol{\sigma}_\beta\}$. With the joint prior on $\boldsymbol{\Xi}$ as described above, and the data likelihood implied by (1), our goal in a Bayesian analysis is to examine $f(\boldsymbol{\Xi} \mid \mathbf{y})$, the joint posterior distribution of the unknowns.

$$f(\boldsymbol{\Xi} \mid \mathbf{y}) \propto \left(\prod_{ijs} f(y_{ijs} \mid \gamma_1, \gamma_2, \boldsymbol{\beta}_i) \right) p(\gamma_1) p(\gamma_2) \left(\prod_i p(\boldsymbol{\beta}_i \mid \theta_0, \boldsymbol{\Theta}, \boldsymbol{\Omega}, \boldsymbol{\sigma}_\beta) \right)$$

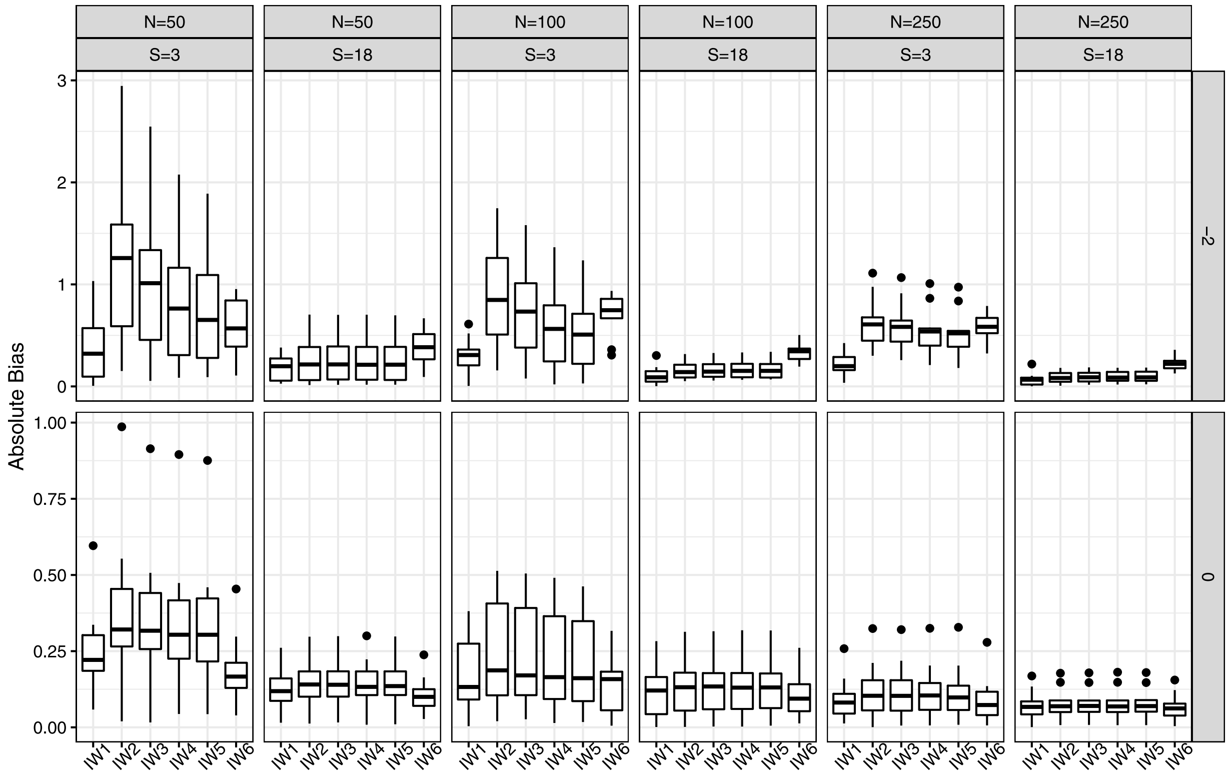

Fig. 6. Boxplots representing the distribution of absolute bias observed between the posterior means and the true values of μ_β elements for 6 specifications of the inverse Wishart prior: (1) $iW(K, \boldsymbol{I}_K)$, (2)$iW(K, K\boldsymbol{I}_K)$, (3)$iW(K+1,(K+1)\boldsymbol{I}_K)$, (4)$iW(K+3,(K+3)\boldsymbol{I}_K)$, (5)$iW(K+4,(K+4)\boldsymbol{I}_K)$, (6)$iW(0.5K(K+1), 0.05K(K+1)\mathbf{I}_K)$. Shown separately for parameters with true value -2 (top) and true value 0 (bottom).

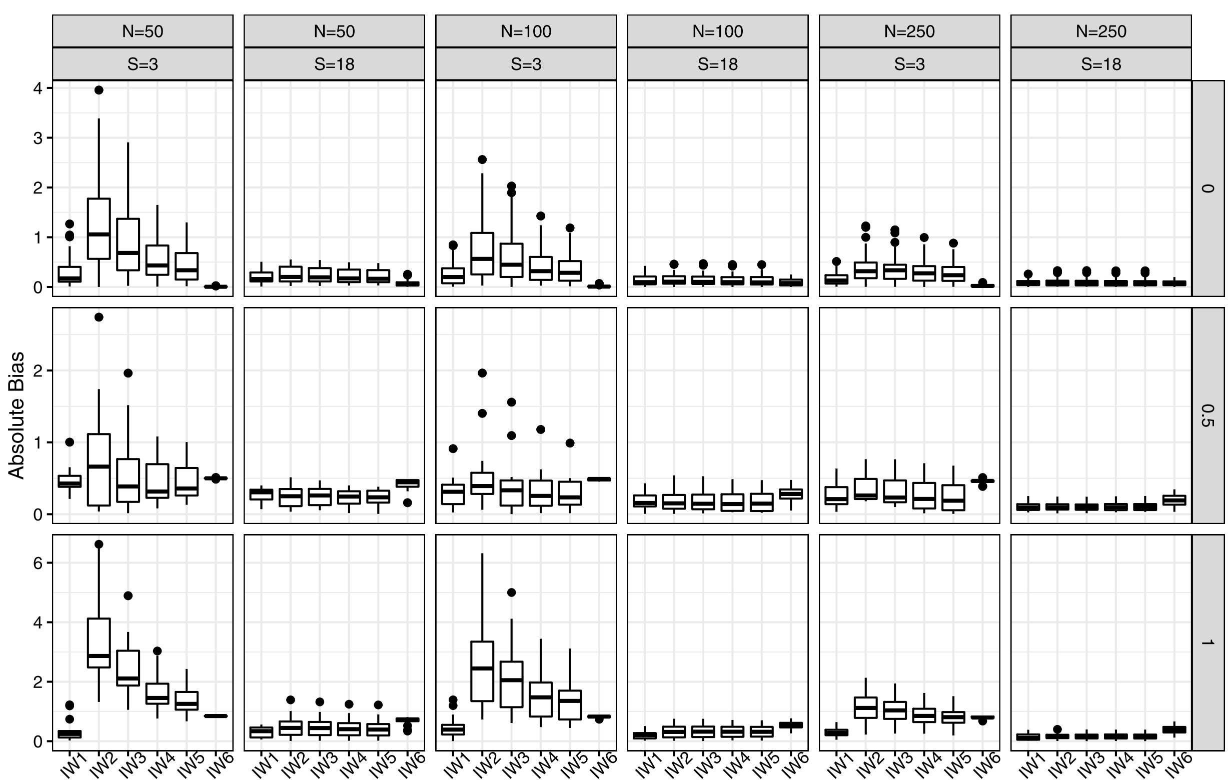

Fig. 7. Boxplots representing the distribution of absolute bias observed between the posterior means and the true values of $\boldsymbol{\Sigma}_\beta$ elements for 6 specifications of the inverse Wishart prior: (1) $iW(K, \boldsymbol{I}_K)$, (2)$iW(K, K\boldsymbol{I}_K)$, (3)$iW(K+1,(K+1)\boldsymbol{I}_K)$, (4)$iW(K+3,(K+3)\boldsymbol{I}_K)$, (5)$iW(K+4,(K+4)\boldsymbol{I}_K)$, (6)$iW(0.5K(K+1), 0.05K(K+1)\mathbf{I}_K)$. Shown separately for covariance parameters with true value 0 (top), true value 0.5 (center), and 1 (bottom).

Choice Scenario 1

	Program (A) No Action	Program (B) Ice blankets	Program (C) Limited GHG Reduction	Program (D) Limited GHG Reduction & Ice blankets	Program (E) Vigorous GHG Reduction
Policy	No reduction in GHG.	No reduction in GHG.	Reduces GHG.	Reduces GHG.	Reduces GHG more.
Ice blankets	No	Yes	No	Yes	No
Glacier Loss in the next 60 years	0.35km³/yr	0.20 km³/yr	0.20 km³/yr	0.15 km³/yr	0.15 km³/yr
View after 60 years	Picture 4	Picture 2	Picture 2	Picture 1	Picture 1
Cost to HH/mo	$0	$10	$50	$60	$100
Preferred program:	☐	☐	☐	☐	☐

Fig. 8. Example choice scenario.

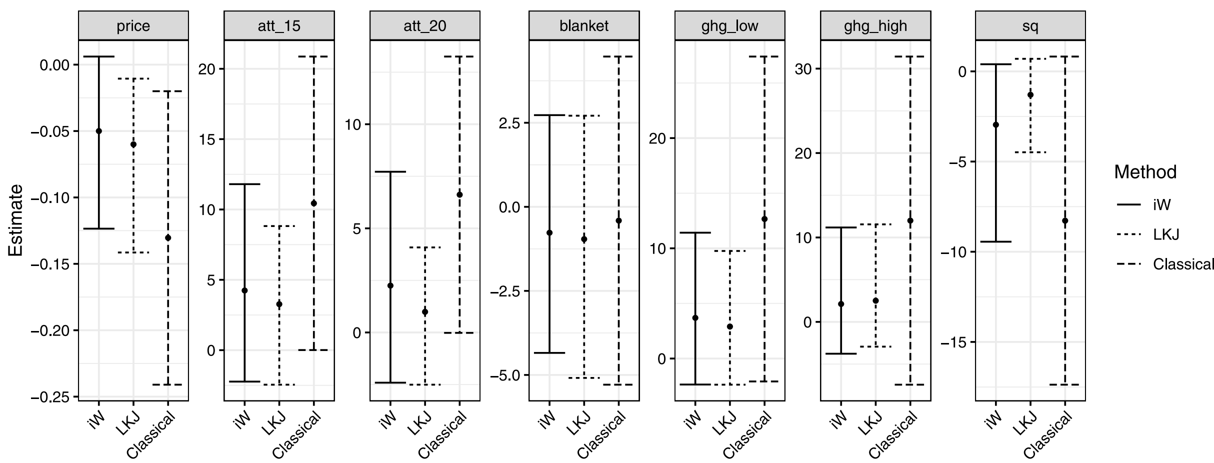

Fig. 9. Bayesian posterior means (using inverse Wishart and LKJ) and classical maximum likelihood point estimates shown with 95% Bayesian credible intervals and classical confidence intervals.

$$\times \left(\prod_{kh} p(\theta_{kh})\right)\left(\prod_{k} p(\theta_{0k})\right) p(\boldsymbol{\Sigma}_{\beta})$$

$$= \left(\prod_{ijs}\left(\frac{exp(\gamma_1 price_{ijs} + \gamma_2 sq_{ijs} + \mathbf{x}_{ijs}^T \boldsymbol{\beta}_i)}{\sum_{j=1}^{J} \ exp(\gamma_1 price_{ijs} + \gamma_2 sq_{ijs} + \mathbf{x}_{ijs}^T \boldsymbol{\beta}_i)}\right)^{y_{ijs}}\right)$$

$$\times \left(\pi(1+\gamma_1^2)\right)^{-1} \left(\pi(1+\gamma_2^2)\right)^{-1}$$

$$\times \left(\prod_{i} |2\pi \boldsymbol{\Sigma}_{\beta}|^{-\frac{1}{2}} exp\left(-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\theta}_0 - \boldsymbol{\Theta} \mathbf{z}_i)^T \boldsymbol{\Sigma}_{\beta}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\theta}_0 - \boldsymbol{\Theta} \mathbf{z}_i)\right)\right)$$

$$\times \left(\prod_{kh} \left(\pi(1+\theta_{kh}^2)\right)^{-1}\right)\left(\prod_{k} \left(2.5\pi(1+(\theta_{0h}/2.5)^2)\right)^{-1}\right) p(\boldsymbol{\Sigma}_{\beta})$$

Regardless of the form we choose for $p(\boldsymbol{\Sigma}_{\beta})$, the above distribution is not available in a closed analytical form, nor can we sample from it directly. However, we can use Markov chain Monte Carlo (MCMC) algorithms to obtain samples that are approximately from the joint posterior distribution. In particular, we use Stan and the No-U-Turn sampler (NUTS) as an efficient way to obtain these samples which can then be used to draw inference.

5. Results

Before discussing the results in detail, we make a choice about which prior regime is best for these data. After making the decision about the prior, we illustrate how a researcher might use this fine level of granularity to answer meaningful policy questions. We obtain samples from the posterior distribution by using an inverse Wishart prior on the variance covariance matrix, and also with the half Cauchy and LKJ prior on the standard deviation components and correlation matrices respectively.

Table 3
Posterior summaries of WTP for various attributes. Column 4 (Proportion positive) represents the proportion of the 98 individuals for which the posterior mean WTP > 0.

	Mean	Lower	Upper	Proportion positive
0.15 km³ recession	201	−243	604	0.78
0.20 km³ recession	51	−47	203	0.82
Blanket	−118	−361	161	0.23
Limited GHG agreement	11	−268	284	0.51
Vigorous GHG agreement	−46	−436	185	0.43

5.1. Model choice

We assess how well each model describes the data generating process first by creating a variety of posterior predictive p-values (see, for example, Meng et al. (1994) and Gelman et al. (2013)). The intuition behind posterior predictive p-values is that, if an estimated model is a good representation of a data set, any data set simulated from that model should resemble the original data set in all of the important aspects. We estimate

$$p(\tilde{\mathbf{y}} \mid \mathbf{y}) = \int p(\tilde{\mathbf{y}} \mid \Xi) p(\Xi \mid \mathbf{y}) d\Xi$$

by simulating draws from the data model (1) conditional on $\Xi^{(r)}$ for each posterior draw $r = 1, \ldots, R$. From each of these R new data

Table 4
WTP summaries for outcomes and policies for different representative agents in the sample.

Variable	Female	cons	env_org	Mean	Lower	Upper	$P(WTP > 0 \mid \boldsymbol{y})$
0.15 km^3 recession	Female	not cons	not env org	151	−406	1398	0.94
0.15 km^3 recession	Female	not cons	env org	169	−467	1427	0.95
0.15 km^3 recession	Female	cons	not env org	143	−496	1054	0.91
0.15 km^3 recession	Female	cons	env org	160	−463	1198	0.92
0.15 km^3 recession	Male	not cons	not env org	249	−689	1531	0.95
0.15 km^3 recession	Male	not cons	env org	290	−824	1828	0.96
0.15 km^3 recession	Male	cons	not env org	204	−664	1091	0.91
0.15 km^3 recession	Male	cons	env org	244	−793	1294	0.93
0.20 km^3 recession	Female	not cons	not env org	28	−146	271	0.74
0.20 km^3 recession	Female	not cons	env org	27	−160	282	0.73
0.20 km^3 recession	Female	cons	not env org	−6	−203	215	0.50
0.20 km^3 recession	Female	cons	env org	−7	−216	228	0.50
0.20 km^3 recession	Male	not cons	not env org	98	−263	675	0.95
0.20 km^3 recession	Male	not cons	env org	106	−331	685	0.95
0.20 km^3 recession	Male	cons	not env org	48	−164	340	0.80
0.20 km^3 recession	Male	cons	env org	55	−205	339	0.81
Limited GHG agreement	Female	not cons	not env org	−22	−468	304	0.53
Limited GHG agreement	Female	not cons	env org	9	−445	344	0.65
Limited GHG agreement	Female	cons	not env org	−40	−578	265	0.44
Limited GHG agreement	Female	cons	env org	−8	−495	350	0.57
Limited GHG agreement	Male	not cons	not env org	−1	−469	310	0.56
Limited GHG agreement	Male	not cons	env org	32	−389	479	0.71
Limited GHG agreement	Male	cons	not env org	−28	−482	336	0.44
Limited GHG agreement	Male	cons	env org	6	−450	403	0.60
Vigorous GHG agreement	Female	not cons	not env org	−35	−715	438	0.49
Vigorous GHG agreement	Female	not cons	env org	47	−531	595	0.72
Vigorous GHG agreement	Female	cons	not env org	−101	−1252	489	0.29
Vigorous GHG agreement	Female	cons	env org	−19	−984	554	0.51
Vigorous GHG agreement	Male	not cons	not env org	−79	−820	580	0.33
Vigorous GHG agreement	Male	not cons	env org	12	−525	503	0.62
Vigorous GHG agreement	Male	cons	not env org	−239	−1787	894	0.14
Vigorous GHG agreement	Male	cons	env org	−147	−1370	581	0.29

sets, we calculate a test quantity, $T(\tilde{\boldsymbol{y}})$, such as the number of times a conservative female who is part of an environmental group chooses the status quo option. We can compute the same quantity for the observed data set and then quantify how extreme the model-based simulated quantities are when compared to the observed. That is, we calculate the posterior predictive p-value, $p_B = P(T(\tilde{\boldsymbol{y}}) > T(\boldsymbol{y}))$. A p_B value close to 0 (indicating a high probability the model is underestimating the test quantity) or 1 (indicating a high probability the model is overestimating the test quantity) is evidence that the estimated model is not capturing that characteristic well.

We calculate posterior predictive p-values for a variety of policies and outcomes within each unique demographic group as an additional sensitivity analysis. For example, let the function $T()$ represent the number of times respondents chose a vigorous GHG agreement. Fig. 10 illustrates $T(\boldsymbol{y})$ and the distributions of $T(\tilde{\boldsymbol{y}})$ for both prior schemes along with the corresponding posterior predictive p-values. For both the LKJ and inverse Wishart prior schemes, we find little evidence that the models fail to capture the relative frequency with which individuals select this particular policy. Similar analyses examining the remaining policies and outcomes result in the same conclusion. Thus, from a prediction standpoint, neither model seems to have an advantage.[2]

However, models with very different coefficient estimates can yield the same or similar predictions. If interpretation of effects is of concern, as it is in this application, we should consider other sources of evidence toward using one model over another. The Widely Applicable Information Criterion (WAIC) is an information criterion that can be used to compare model fit in this scenario. WAIC estimates the log predictive density in a similar manner to DIC, but WAIC uses the samples from the posterior distribution while DIC plugs in a point estimate such as the posterior mean. Gelman, Hwang, and Vehtari (2014) thus recommends using the WAIC for model evaluation. We calculate $WAIC_{LKJ} - WAIC_{iW} = -18.5$ with an estimated standard error of 3.7. This indicates that the LKJ prior regime is preferred over the inverse Wishart. So, while the posterior predictive p-values indicate that both models are acceptable, the WAIC analysis along with the results of the simulation study for studies of a similar size motivate us to report the results from the LKJ prior scheme.

5.2. Posterior distributions

Summaries of WTP values across individuals are displayed in Table 3. For each person, the posterior mean WTP was calculated, then the mean of those means was taken to calculate an overall mean. This overall mean is shown in column 2 while the 2.5 and 97.5 quantiles are shown in columns 3 and 4. Average WTP for both non-status quo outcomes is greater than zero. As the outcome improves, average WTP increases; however, the proportion of individuals with a positive willingness to pay decreases. Next, average WTP for a limited GHG reduction agreement is \$11, which is greater than the average WTP for a vigorous GHG reduction agreement (\$-46). Further, there is a smaller proportion of people willing to pay a positive amount for a vigorous versus a limited GHG reduction agreement. This indicates that respondents prefer a less aggressive international agreement over a more modest one, on average. Further, mean WTP for an insulated ice blanket is -\$118, and the posterior probability that it is greater than zero is 0.23. Not shown is the WTP for the alternative specific constant status quo, which is calculated as $WTP_{sq} = -\frac{\gamma_2}{\gamma_1}$, with γ_2 entering the analysis as a fixed parameter. $WTP_{sq} = -\$52$, and the posterior probability that $WTP_{sq} > 0 = 0.13$, which indicates that people would prefer not to do nothing.

Columns 3 and 4 in Table 3 indicate that there is a large amount of heterogeneity in WTP estimates. Using the climate change literature to

[2] We additionally compute an estimate of the leave-one-out cross-validation error (LOO), which estimates the predictive accuracy of the LKJ model to be better than that of the inverse Wishart model, although the difference is small relative to the standard error.

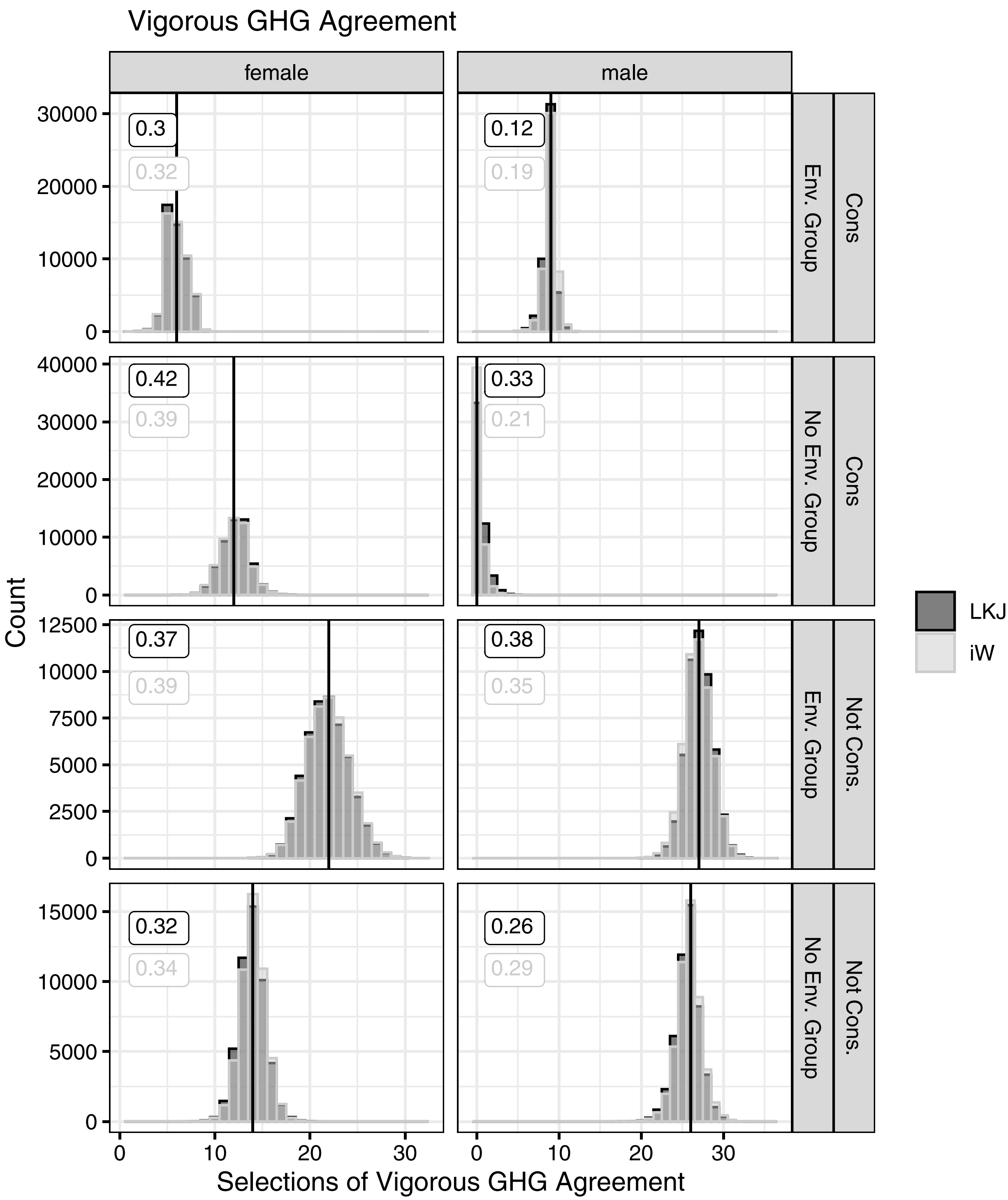

Fig. 10. Histogram of 50,000 samples of $T(\tilde{\mathbf{y}})$ where the test quantity function $T()$ represents the number of times respondents chose a vigorous GHG agreement. Shown for inverse Wishart and LKJ prior schemes. Vertical line shows $T(\mathbf{y})$. Posterior predictive p-values are shown in the upper left corner.

inform what we include, we explore some of the heterogeneity around WTP estimates with observed sociodemographic characteristics. We examine the posterior distribution of the quantity $\theta_0 + \boldsymbol{\Theta z}$ - $f(\theta_0 + \boldsymbol{\Theta z} \mid \mathbf{y})$ - for all possible vectors $\mathbf{z}$ describing a particular demographic group. In each scenario, income at $125,000, which is the median level observed in the data.

Table 4 shows posterior summaries of WTP for both policies and outcomes for different permutations of the interacted sociodemographics. Column 5 displays the posterior means, and Columns 6 and 7 display the upper and lower bounds of the 95% credible intervals. Column 8 displays the probability of a positive WTP for the specified representative individual and the specific attribute level. For example, a conservative female (male) who does not belong to an environmental organization is willing to pay an average of $5 ($44) per month to achieve a 0.20 km^3 rate of recession. Non-conservative males who are also not in an environmental organization are willing to pay $249, on average. Compare that to females, who are willing to pay an average of $151 for the best outcome of 0.15 km^3 recession per year.

Tables 5 thru 7 describe the posterior probability that WTP for a given attribute level is *different* for a particular sociodemographic characteristic. Probabilities that are farther away from 0.5 indicate a higher degree of separation between the demographic groups. For example, line 2 in Table 5 can be interpreted in the following way: For non-conservative individuals who are not members of an environmental organization, the posterior probability that females (males) are willing to pay more than males (females) for 0.20 km^3 recession is 10% (90%). On the other hand, lines 3 and 13 in Table 5 can be interpreted jointly in the following way: Controlling for political preferences and limiting our scope to individuals who are not members of an environmental organization, the posterior probability that females are willing to pay more than males for a program containing an insulated ice blanket is 94%. Tables 6 and 7 can be interpreted in an identical way, except they compare, respectively, membership to an environmental organization and political preferences. Fig. 13 succinctly describes the information given in Tables 5 thru 7.

5.3. Discussion

The role that sociodemographic characteristics play in explaining variation in the posterior distribution of WTP values depends on whether the issue at hand is a policy or an outcome, as well as how extreme that policy or outcome is. We examine two attractive features of the posterior distribution: the ability of explicit interaction terms to explain variation in baseline WTP distributions, and separation between categories within sociodemographic classes. We organize the discussion around three observed sociodemographic characteristics, which we examine in turn.

Table 5 illustrates that, in general, females are willing to pay more for policies and less for outcomes, relative to males. For example, lines 8 and 18 tell us that, controlling for political preferences and limiting our scope to individuals who belong (do not belong) to an environmental organization, the posterior probability that females are willing to pay more than males for a program containing an insulated ice blanket is 86% (94%). This insight is confirmed in Fig. 11. Contrast this with WTP for the 0.20 km^3 recession outcome. Regardless of political preferences or membership in an environmental organization, the posterior probability that females are willing to pay more than males for 0.20 km^3 recession is small, indicating that it is unlikely females are willing to pay more than males for this outcome. However, males and females tend to agree on their WTP for a limited GHG reduction agreement. The posterior probability that WTP for a limited GHG agreement is positive is greater than 0.50 for all representative agents except conservative individuals who are also not members of an environmental organization. Finally, gender disagreements can explain a large amount of variation in the posterior distribution of WTP values for 0.20 km^3 recession, an insulated ice blanket, and a vigorous GHG reduction agreement (as seen in Fig. 12).

Table 6 illustrates that environmental organization membership explains more of the variation in policy than in outcomes. Controlling for politics, the posterior probability that females (males) belonging to an environmental organization are willing to pay more for a vigorous GHG reduction agreement than females (males) not belonging to an environmental organization is 0.83 (0.87). Moreover, controlling for political preference, the posterior probability that females (males) who are members of an environmental organization are willing to pay more for a limited GHG reduction agreement than females (males) who are not members of an environmental organization is 71% (73%). In contrast, controlling for political preferences, the posterior probability that females (males) belonging to an environmental organization are willing to pay more for 0.20 km^3 recession than females (males) not belonging to an environmental organization is 0.51 (0.56).

Table 7 shows that self-identified non-conservative respondents tend to be willing to pay more than conservative respondents and that this tends to hold for both policies and outcomes. Moreover, there is more variability in the posterior distribution of WTP values attributed to politics when the policy is extreme. Controlling for environmental organization membership, the posterior probability that non-conservative females (males) are willing to pay more for a vigorous GHG reduction agreement than conservative females (males) is 83% (90%). In contrast, the posterior probability of non-conservatives being willing to pay more than conservatives for a limited GHG reduction agreement is smaller. Politics also explains variation in WTP for outcomes. Controlling for membership in an environmental organization, the posterior probability that non-conservative females (males) are willing to pay more than conservative females (males) for 0.20 km^3 recession is 78% (88%).

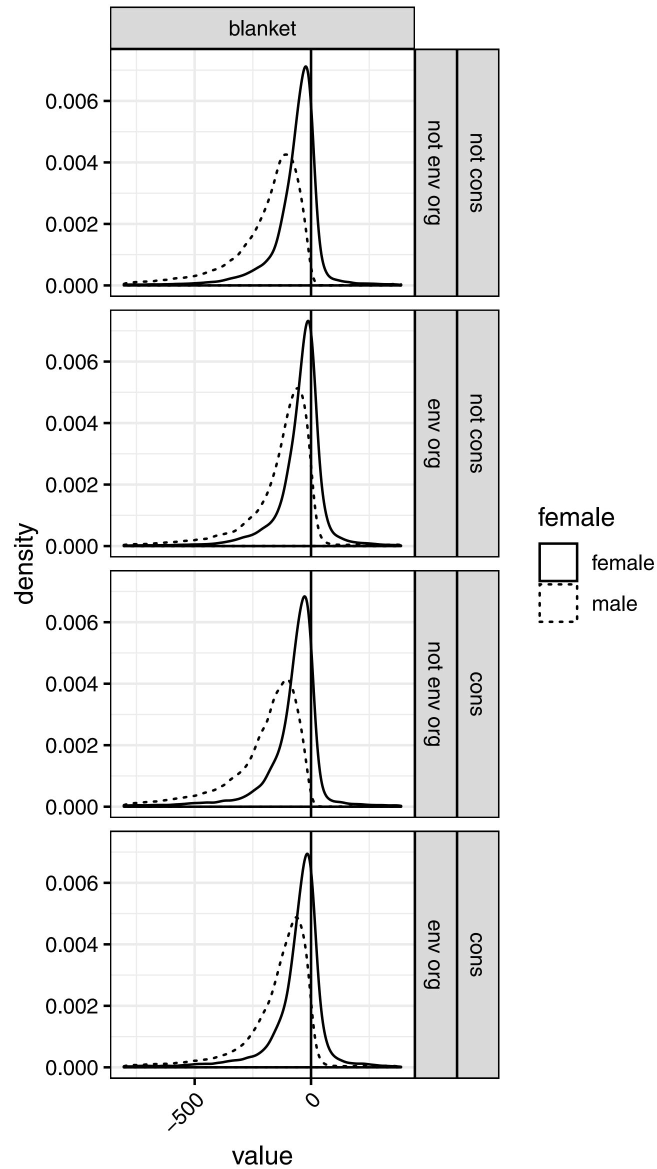

Fig. 11. Posterior distributions of the WTP for a blanket. This is shown for males and females as indicated by line type.

The results of the Bayesian analysis presented in this section largely support the results of the classical analysis found in Vander Naald (2019). When it comes to policy, both analyses find that non-conservatives are willing to pay more than conservatives and that environmental membership has a positive impact on WTP for both limited and vigorous GHG agreements. The environmental organization membership and conservative political preference interactions are significant at the 5% level in the classical analysis, and imply large deviations in particular for the vigorous GHG agreement. Similarly, the

Table 5
Posterior probabilities that females are willing to pay more than males within fixed political and environmental groups.

	cons	env_org	Variable	$P(WTP_{female} > WTP_{male} \mid \boldsymbol{y})$
1	not cons	not env org	0.15 km^3 recession	0.30
2	not cons	not env org	0.20 km^3 recession	0.10
3	not cons	not env org	Blanket	0.94
4	not cons	not env org	Limited GHG agreement	0.45
5	not cons	not env org	Vigorous GHG agreement	0.68
6	not cons	env org	0.15 km^3 recession	0.24
7	not cons	env org	0.20 km^3 recession	0.10
8	not cons	env org	Blanket	0.86
9	not cons	env org	Limited GHG agreement	0.40
10	not cons	env org	Vigorous GHG agreement	0.61
11	cons	not env org	0.15 km^3 recession	0.38
12	cons	not env org	0.20 km^3 recession	0.23
13	cons	not env org	Blanket	0.94
14	cons	not env org	Limited GHG agreement	0.49
15	cons	not env org	Vigorous GHG agreement	0.79
16	cons	env org	0.15 km^3 recession	0.32
17	cons	env org	0.20 km^3 recession	0.21
18	cons	env org	Blanket	0.86
19	cons	env org	Limited GHG agreement	0.44
20	cons	env org	Vigorous GHG agreement	0.74

Table 6
Posterior probabilities that individuals who are part of an environmental organization are willing to pay more than those who are not within fixed political and gender groups.

	cons	Female	variable	$P(WTP_{env} > WTP_{not\ env} \mid \boldsymbol{y})$
1	not cons	Female	0.15 km^3 recession	0.56
2	not cons	Female	0.20 km^3 recession	0.51
3	not cons	Female	Blanket	0.72
4	not cons	Female	Limited GHG agreement	0.71
5	not cons	Female	Vigorous GHG agreement	0.83
6	not cons	Male	0.15 km^3 recession	0.66
7	not cons	Male	0.20 km^3 recession	0.56
8	not cons	Male	Blanket	0.87
9	not cons	Male	Limited GHG agreement	0.73
10	not cons	Male	Vigorous GHG agreement	0.87
11	cons	Female	0.15 km^3 recession	0.56
12	cons	Female	0.20 km^3 recession	0.51
13	cons	Female	Blanket	0.72
14	cons	Female	Limited GHG agreement	0.71
15	cons	Female	Vigorous GHG agreement	0.83
16	cons	Male	0.15 km^3 recession	0.66
17	cons	Male	0.20 km^3 recession	0.56
18	cons	Male	Blanket	0.87
19	cons	Male	Limited GHG agreement	0.73
20	cons	Male	Vigorous GHG agreement	0.87

Table 7
Posterior probabilities that individuals who do not identify as conservative are willing to pay more than those who do identify as conservative within fixed gender and environmental groups.

	env_org	Female	Variable	$P(WTP_{not\ cons} > WTP_{cons} \mid \boldsymbol{y})$
1	not env org	Female	0.15 km^3 recession	0.66
2	not env org	Female	0.20 km^3 recession	0.78
3	not env org	Female	Blanket	0.61
4	not env org	Female	Limited GHG agreement	0.66
5	not env org	Female	Vigorous GHG agreement	0.83
6	not env org	Male	0.15 km^3 recession	0.73
7	not env org	Male	0.20 km^3 recession	0.88
8	not env org	Male	Blanket	0.63
9	not env org	Male	Limited GHG agreement	0.68
10	not env org	Male	Vigorous GHG agreement	0.90
11	env org	Female	0.15 km^3 recession	0.66
12	env org	Female	0.20 km^3 recession	0.78
13	env org	Female	Blanket	0.61
14	env org	Female	Limited GHG agreement	0.66
15	env org	Female	Vigorous GHG agreement	0.83
16	env org	Male	0.15 km^3 recession	0.73
17	env org	Male	0.20 km^3 recession	0.88
18	env org	Male	Blanket	0.63
19	env org	Male	Limited GHG agreement	0.68
20	env org	Male	Vigorous GHG agreement	0.90

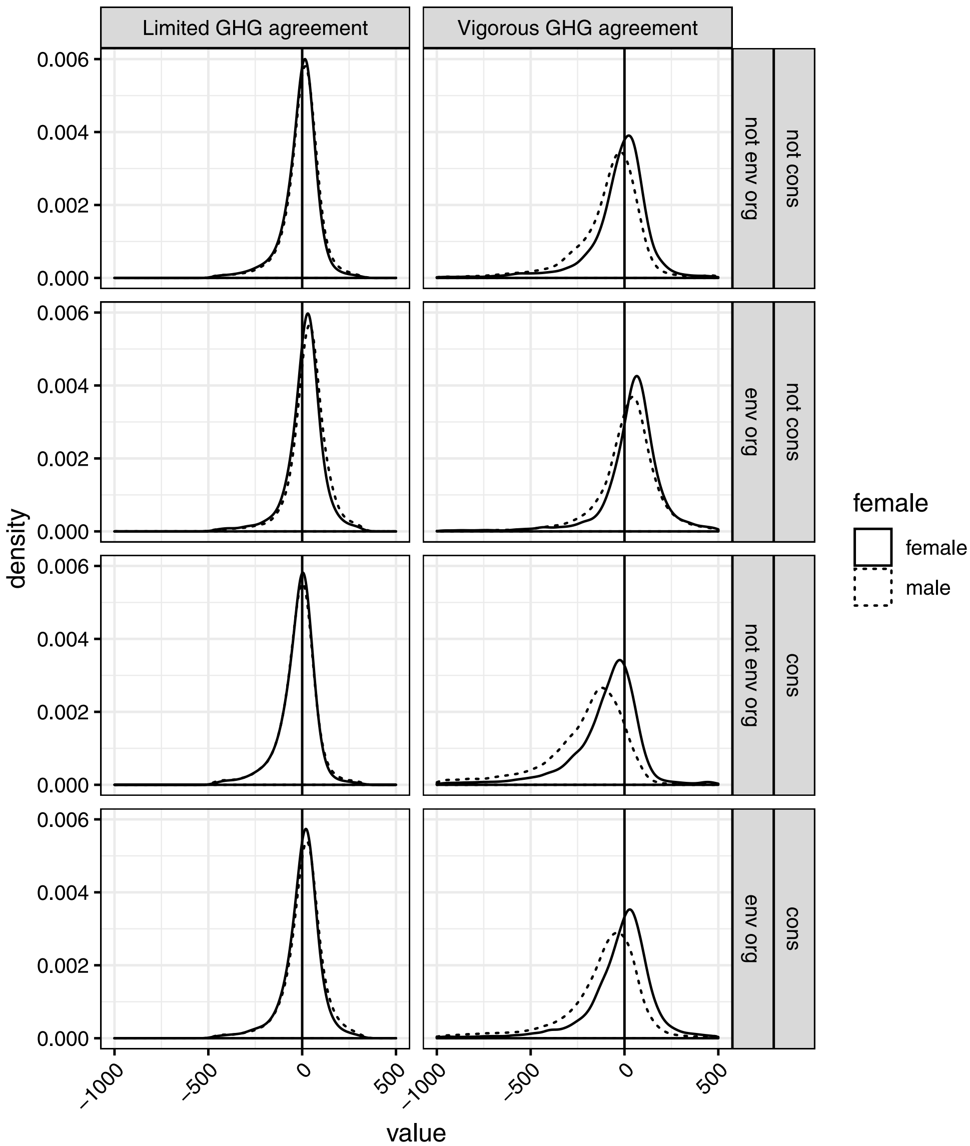

Fig. 12. Posterior distributions of the WTP for a vigorous climate agreement in the top facet and for a moderate climate agreement in the bottom facet. This is shown for males and females as indicated by line type.

Bayesian analysis finds moderately high probabilities of differing WTP for the limited agreement, and high probabilities of differing WTP for the vigorous agreement. Both analyses find that while most individuals are willing to pay a positive amount for non-status quo outcomes, individuals who do not identify as conservative are willing to pay more than conservatives. However, in the classical analysis, conservative status does not significantly impact WTP for these outcomes. In the Bayesian analysis, we find moderate probabilities (between 66% and 88%, depending on other demographic factors) that non conservatives are willing to pay more than conservatives. One area of disagreement between the two analyses is of the effect of environmental organization membership on WTP for non-status quo outcomes. The maximum likelihood estimates imply a different directional effect of environmental group membership than do the Bayesian posterior means. However, the classical analysis finds no significant effect of environmental group membership and the Bayesian analysis estimates the probability that environmental group members are willing to pay more than non members at around 50%. Thus, we can conclude no practically significant effect.

6. Conclusions

The purpose of this article was to introduce a Bayesian MLM with a non-standard prior regime to analyze small-sample DCE data. Most studies using Bayesian methods with DCE data assign an inverse Wishart prior to the variance–covariance matrix. This prior assignment can be problematic because it enforces a strict relationship between the correlation and variance components, and can bias posterior estimates of variances upwards. We developed a Bayesian MLM in which we decomposed the variance–covariance matrix and assigned an LKJ prior to the variance. We then compared the performance of this model to a Bayesian MLM in which we applied an inverse Wishart prior to the entire variance–covariance matrix using simulated data over a range of data set sizes from small to large. Based on the WAIC criterion and a measure of absolute bias, the LKJ prior regime consistently outperformed the inverse Wishart prior regime.

We tested the model using actual DCE data. We compared baseline point estimates across the two prior regimes and MSL. Under both prior regimes, the Bayesian MLM models performed better than the classically estimated MLM using MSL. An additional benefit of our model is

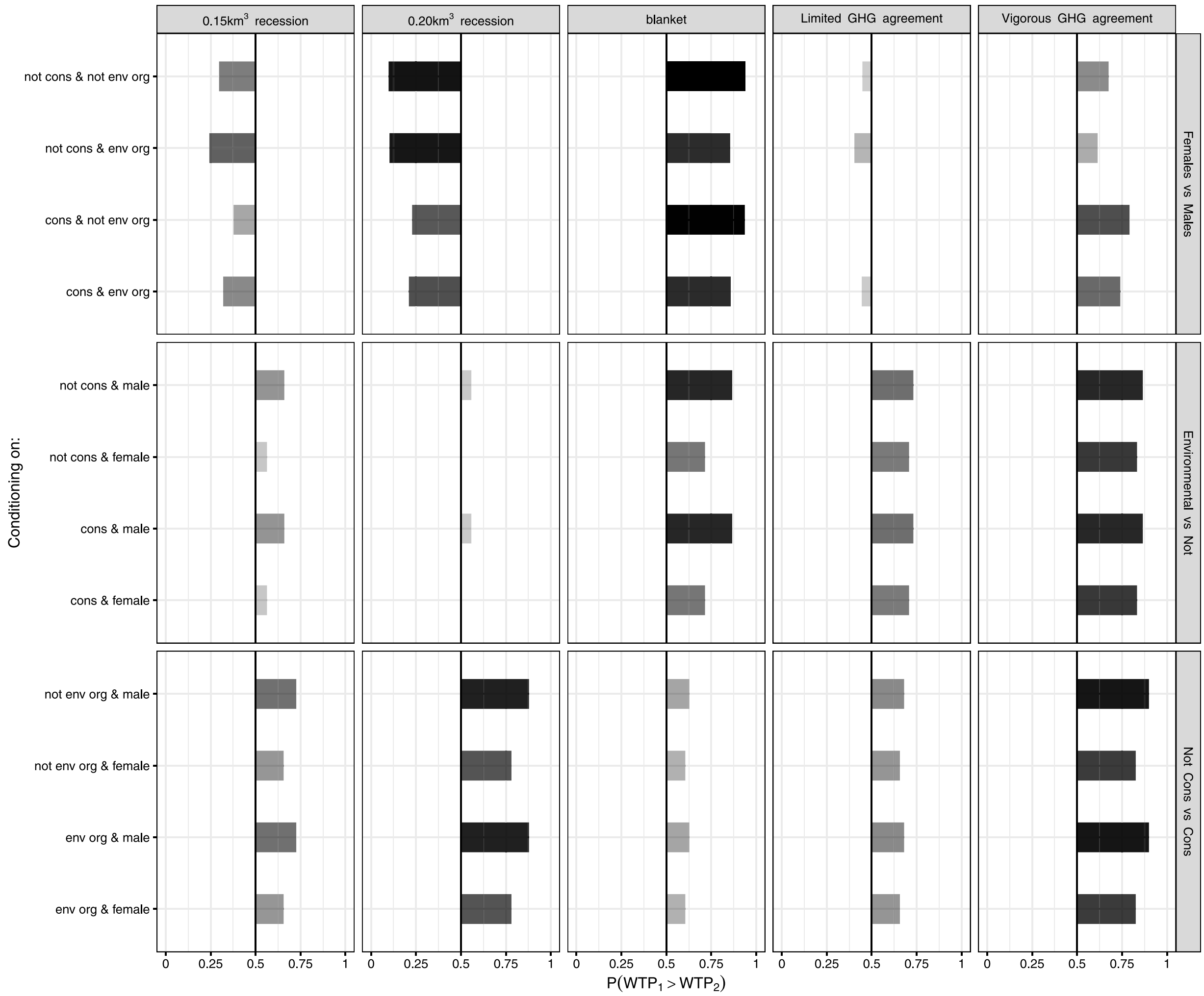

Fig. 13. Posterior probabilities that demographic group 1 is willing to pay more than demographic group 2. Demographic groups being compared is indicated by row facet. Outcomes (columns 1 and 2) and policies (columns 3–5) are indicated by column facet.

that it can be extended to allow for the analysis of second and third order interaction effects among observed respondent characteristics for small DCE data sets — something that would not be possible using classical econometric analysis. Our application illustrated how one would conduct such an analysis in practice. This is likely to be useful for situations in which data collection is expensive and the number of observations is small. Moreover, in many areas of the world, whether a public policy is enacted depends on whether a majority of voters approves. This approach provides a straightforward way to determine exactly what proportion of the population with specific characteristics might be receptive to possible public policies around environmental issues, with relatively small data requirements. Further, this approach has the potential to benefit the market segmentation literature as it allows for finer separation on observable heterogeneity than is possible using classical methods.

In our simulation study, we found that the LKJ prior scheme dominated the inverse Wishart prior scheme. One limitation of the current study is that this result may not generalize to other situations. Depending on the specific application, prior beliefs might be better represented by alternative distributional schemes. Future research should include additional empirical applications, with particular attention paid to choice of prior scheme. In addition, we make the assumption that the $\boldsymbol{\beta}$ vectors are independently distributed multivariate normal. Different distributional assumptions should be tested across different applications. DCEs are a popular tool in other disciplines, including economics, health, marketing, transportation, and travel. While Bayesian MLM has been used for DCE analysis in some of these disciplines (Balcombe et al., 2009), it does not seem to have gotten much attention in others. Future nonmarket valuation research should consider using Bayesian analysis for inference. And when using Bayesian analysis, choice of prior distribution should be carefully considered.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Data collection for the empirical application was funding in part by Alaska EPSCoR NSF award #OIA-1208927. We thank Andrew Bryant (Drake University) and Jerrod Penn (LSU) for their helpful comments and insight. Finally, we are grateful to participants of the Santa Fe edition of the W-4133 Meeting.

References

Adhikari, A. (2015). Differentiating subjective and objective attributes of experience products to estimate willingness to pay price premium. *Journal of Travel Research, 54*(5), 634–644.

Akinc, D., & Vandebroek, M. (2018). Bayesian estimation of mixed logit models: Selecting an appropriate prior for the covariance matrix. *Journal of Choice Modelling*, *29*, 133–151.

Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, *71*(1), 1–10.

Assaf, A. G., Li, G., Song, H., & Tsionas, M. G. (2019). Modeling and forecasting regional tourism demand using the Bayesian global vector autoregressive (BGVAR) model. *Journal of Travel Research*, *58*(3), 383–397.

Assaf, A. G., Oh, H., & Tsionas, M. G. (2016). Unobserved heterogeneity in hospitality and tourism research. *Journal of Travel Research*, *55*(6), 774–788.

Assaf, A. G., Tsionas, M., & Oh, H. (2018). The time has come: Toward Bayesian SEM estimation in tourism research. *Tourism Management*, *64*, 98–109.

Balcombe, K., Chalak, A., & Fraser, I. (2009). Model selection for the mixed logit with Bayesian estimation. *Journal of Environmental Economics and Management*, *57*(2), 226–237.

Birol, E., Karousakis, K., & Koundouri, P. (2006). Using a choice experiment to account for preference heterogeneity in wetland attributes: The case of Cheimaditida wetland in Greece. *Ecological Economics*, *60*(1), 145–156.

Bryant, A., & Hill, R. P. (2019). Poverty, consumption, and counterintuitive behavior. *Marketing Letters*, 1–11.

Campbell, D., Boeri, M., Doherty, E., & Hutchinson, W. G. (2015). Learning, fatigue and preference formation in discrete choice experiments. *Journal of Economic Behaviour and Organization*, *119*, 345–363.

Chen, N., Masiero, L., & Hsu, C. H. (2019). Chinese outbound tourist preferences for all-inclusive group package tours: A latent class choice model. *Journal of Travel Research*, *58*(6), 916–931.

Collins, J. P., & 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., Vossler, C. A., Budziński, W., Wiśniewska, A., & Zawojska, E. (2017). Addressing empirical challenges related to the incentive compatibility of stated preferences methods. *Journal of Economic Behaviour and Organization*, *142*, 47–63.

de Bekker-Grob, E. W., Donkers, B., Jonker, M. F., & Stolk, E. A. (2015). Sample size requirements for discrete-choice experiments in healthcare: A practical guide. *The Patient-Patient-Centered Outcomes Research*, *8*(5), 373–384.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016.

Gelman, A., Jakulin, A., Pittau, M. G., Su, Y.-S., et al. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Grijalva, T., Berrens, R. P., & Shaw, W. (2011). Species preservation versus development: An experimental investigation under uncertainty. *Ecological Economics*, *70*(5), 995–1005.

Gunter, U., & Önder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni-and multivariate models employing monthly data. *Tourism Management*, *46*, 123–135.

Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. *Transportation Research, Part B (Methodological)*, *44*(6), 735–752.

Huth, W. L., & Morgan, O. (2011). Measuring the willingness to pay for cave diving. *Marine Resource Economics*, *26*(2), 151–166.

Interis, M. G., & Petrolia, D. R. (2016). Location, location, habitat: How the value of ecosystem services varies across location and by habitat. *Land Economics*, *92*(2), 292–307.

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., et al. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, *4*(2), 319–405.

Kim, D., & Park, B.-J. R. (2017). The moderating role of context in the effects of choice attributes on hotel choice: A discrete choice experiment. *Tourism Management*, *63*, 439–451.

Kubo, T., Mieno, T., & Kuriyama, K. (2019). Wildlife viewing: The impact of money-back guarantees. *Tourism Management*, *70*, 49–55.

Landauer, M., Pröbstl-Haider, U., & Haider, W. (2012). Managing cross-country skiing destinations under the conditions of climate change–Scenarios for destinations in Austria and Finland. *Tourism Management*, *33*(4), 741–751.

Leiserowitz, A. (2006). Climate change risk perception and policy preferences: The role of affect, imagery, and values. *Climatic Change*, *77*(1–2), 45–72.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001.

Mariel, P., & Meyerhoff, J. (2018). A more flexible model or simply more effort? On the use of correlated random parameters in applied choice studies. *Ecological Economics*, *154*, 419–429.

Meng, X.-L., et al. (1994). Posterior predictive p-values. *The Annals of Statistics*, *22*(3), 1142–1160.

O'Connor, R. E., Bord, R. J., Yarnal, B., & Wiefek, N. (2002). Who wants to reduce greenhouse gas emissions? *Social Science Quarterly*, *83*(1), 1–17.

Papastathopoulos, A., Ahmad, S. Z., Al Sabri, N., & Kaminakis, K. (2019). Demographic analysis of residents' support for tourism development in the UAE: A Bayesian structural equation modeling multigroup approach. *Journal of Travel Research*, 0047287519874131.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

Rigby, D., Balcombe, K., & Burton, M. (2009). Mixed logit model performance and distributional assumptions: Preferences and GM foods. *Environmental and Resource Economics*, *42*(3), 279–295.

Rigby, D., & Burton, M. (2006). Modeling disinterest and dislike: A bounded Bayesian mixed logit model of the UK market for GM food. *Environmental and Resource Economics*, *33*(4), 485–509.

Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting. *Annals of Tourism Research*, *75*, 338–362.

Stan Development Team (2018). *RStan: the R interface to Stan. R package version 2.18.2*. URL: http://mc-stan.org/.

Train, K. E. (2003). *Discrete choice methods with simulation*. Cambridge University Press.

Vander Naald, B. (2019). Examining tourist preferences to slow glacier loss: Evidence from Alaska. *Tourism Recreation Research* 1–11, http://dx.doi.org/10.1080/02508281.2019.1606978.

Wong, K. K., Song, H., & Chon, K. S. (2006). Bayesian Models for tourism demand forecasting. *Tourism Management*, *27*(5), 773–780.

Wu, Q., Law, R., & Xu, X. (2012). A sparse Gaussian process regression model for tourism demand forecasting in Hong Kong. *Expert Systems with Applications*, *39*(5), 4769–4774.

Lendie Follett is an Assistant Professor of Business Analytics at Drake University. Her research is primarily focused on Bayesian methodology and computation.

Brian Vander Naaldis an Assistant Professor in the Department of Economics and Finance at Drake University. His research focuses on nonmarket valuation of ecosystem goods and services.