Safe models for risky decisions

Steingröver, H.M.

Link to publication

Creative Commons License (see https://creativecommons.org/use-remix/cc-licenses):
Other

Citation for published version (APA):
Chapter 10

Using Bayesian Regression to Incorporate Covariates into Hierarchical Cognitive Models

This chapter has been submitted for publication as:
Using Bayesian regression to incorporate covariates into hierarchical cognitive models.

Abstract

An important tool in the advancement of cognitive science are quantitative models that represent different cognitive processes in terms of model parameters. To evaluate such models, their parameters are typically tested for relationships with behavioral and physiological variables that are thought to reflect specific cognitive processes. However, many models do not come equipped with the statistical framework needed to relate model parameters to covariates. Instead, researchers often revert to classifying participants into groups depending on their values on the covariates, and subsequently comparing the estimated model parameters between these groups. Here we develop a comprehensive solution to the covariate problem in the form of a Bayesian regression framework. Our framework can be easily added to existing cognitive models and allows researchers to quantify the evidential support for relationships between covariates and model parameters using Bayes factors. Moreover, we present a simulation study that demonstrates the superiority of the Bayesian regression framework to the conventional classification-based approach.

One of the major goals of cognitive science is to describe how cognition shapes human behavior. Cognitive models are an important tool in this endeavor as they offer a formal account of the relationship between cognitive processes, physiology, and behavior (Turner, Forstmann, Love, Palmeri, Thomas, & van Maanen, in press). More specifically, model parameters are often used to describe latent cognitive processes that are hypothesized to find their overt expression in physiological variables such as heart rate or EEG activity, and behavioral manifestations such as reaction times or confidence ratings (Forstmann, Wagenmakers, Eichele, Brown, & Serences, 2011).
10. Using Bayesian Regression to Incorporate Covariates into Hierarchical Cognitive Models

However, cognitive models usually do not come equipped with a principled way of evaluating these hypothesized relationships. Instead of directly modeling and assessing the relationship between latent cognitive processes and measurable covariates at the population level, researchers often resort to a multi-step procedure; they first fit a cognitive model to participants’ behavioral data, then group participants according to the values on a set of covariates, and subsequently test the groups of participants for differences in their estimated model parameters. For instance, Cooper, Worthy, and Maddox (2015) asked participants to fill out the Regulatory Focus Questionnaire (Higgins et al., 2001) which consists of two scales that measure participants’ tendency to either avoid new tasks for fear of failure (prevention focus) or approach new tasks with an anticipation of success (promotion focus). Cooper et al. categorized participants into two groups based on whether they scored higher on the prevention focus scale or on the promotion focus scale. Subsequently, participants performed 250 trials of the Mars Farming task in which they have to choose between two options with choice-dependent rewards. Cooper et al. analyzed their data by first fitting a reinforcement-learning model to participants’ choice data and in a second step used an ANOVA to compare the estimated model parameters between the two groups of participants. Although this multi-step procedure might seem a reasonable way of testing which covariates are related to which model parameters, it is cumbersome and potentially misleading.

There are two major statistical pitfalls associated with a multi-step analysis procedure. First, when researchers fit hierarchical cognitive models to data and subsequently extract parameter estimates for individual participants, they risk overstating the evidence for differences between groups of participants. Such practices have unfortunately become more frequent with the advent of hierarchical Bayesian models in recent years (e.g., Ahn et al., 2014; Chan et al., 2013; Chevalier, Chatham, & Munakata, 2014; Vassileva et al., 2013). We will give a brief intuition of the mechanism underlying this pitfall here (for an in-depth discussion of the problem see Boehm, Marsman, Matzke, & Wagenmakers, submitted). Hierarchical Bayesian methods use information about the group of participants to inform parameter estimates for individuals. Individual estimates that are less reliable are more strongly pulled towards the group mean, a phenomenon known as shrinkage (Efron & Morris, 1977; Gelman et al., 2013; Rouder & Lu, 2005). One side effect of this correction of individual estimates is that their variance is reduced because all estimates are pulled towards a common value. As a consequence, effect sizes and test statistics, which are usually computed as the ratio of differences between group means to the variance within groups, are inflated and consequently overstate the evidence against the null hypothesis. This problem can be easily avoided if differences between groups of participants are incorporated into the hierarchical structure of the cognitive model and statistical tests are carried out at the group-level, instead of applying the model separately to each group of participants and carrying out tests on the participant-level.

The second pitfall associated with the multi-step procedure is the dichotomization of continuous covariates which can lead to biased statistical tests. This problem has been discussed repeatedly in the context of frequentist statistics (Altman & Royston, 2006; Austin & Brunner, 2004; Cohen, 1983; MacCallum, Zhang, Preacher, & Rucker, 2002; Maxwell & Delaney, 1993; Royston, Altman, & Sauerbrei, 2006). Despite these repeated warnings, several authors have recently applied dichotomization of continuous covariates to test for relationships with model parameters (e.g., Cooper et al., 2015; Kwak, Pearson, & Huettel, 2014; Steingroever et al., submitted). The type of bias introduced by such dichotomization-based tests depends on the correlation between covariates; uncorrelated covariates lead to reduced power (i.e., tests missing true relationships between covariates and model parameters) whereas correlated covariates lead to an inflation of the Type I error rate (i.e., tests detecting spurious relationships between covariates and model parameters). Maxwell and Delaney (1993) provide an accessible explanation of the mechanisms underlying these biases which we briefly summarize here. In our explanation of the mechanisms
we will use linear regression as comparison standard which is the correct analysis approach for the scenario with continuous, normally distributed model parameters and covariates considered here.

There are two possible scenarios for the correlation structure among covariates and model parameters that give rise to two different types of bias. In the first scenario, assume a researcher measures two uncorrelated continuous covariates, one of which is correlated with a model parameter whilst the other is not. For example, the researcher might administer a questionnaire with two uncorrelated subscales that measure participants’ preference for deliberate and intuitive decision making, respectively, and ask participants to complete 100 trials of a risky decision-making task. The researcher then fits a reinforcement-learning model with a loss aversion parameter to participants’ choice data, where, in fact, the loss aversion parameter is related to participants’ preference for intuitive decision making, but not to their preference for deliberate decision making. To test for relationships between the model parameter and the covariates, the researcher splits participants’ scores on each subscale into two halves based on, say, the median score of each subscale, and, for each subscale, uses a t-test to compare the parameter values of participants scoring above-median to the values of participants scoring below-median. Panel A of Figure 10.1 illustrates this scenario for the intuition scale which is positively correlated with the loss aversion parameter. The two horizontal lines show the mean parameter values of each group, the black diagonal line is the result of the correct regression analysis. As can be seen, within each group the deviation of most individual data points from the regression line, that is, the error variance, is much smaller than the deviation from the corresponding group mean. Consequently, a t-test for a difference in group means, which is just the ratio of the mean differences to the error variance, will be biased towards the null hypothesis. A t-test of the regression slope, on the other hand, uses the correct estimate for the error variance and will therefore not show such a bias.

For the second scenario, assume a researcher measures two correlated continuous covariates, one of which is correlated with a model parameter whilst the other is not. In our previous example, the deliberate decision-making subscale and the intuitive decision-making subscale might be correlated with each other, and the loss aversion model parameter might be correlated with the intuitive decision-making subscale, but not with the deliberate decision-making subscale. To test for relationships between the model parameter and the covariates, the researcher again splits each subscale into two halves and, for each subscale, uses a t-test to compare the parameter values of participants scoring above-median to the values of participants scoring below-median. In this case the covariate of interest is the deliberation subscale which is not correlated with the loss aversion parameter. Panel B of Figure 10.1 shows a scatterplot of participants’ scores on the two subscales with the intuition scale on the x-axis and the deliberation scale on the y-axis; the dark grey squares indicate the means of both subscales for each group created by splitting the deliberation scale into two halves. As can be seen, the mean value on the intuition scale is higher for one group than for the other. However, because the two subscales are correlated, the two groups also differ in their mean on the deliberation scale, which is correlated with the loss aversion parameter. Therefore, a t-test for a mean difference in the model parameter between the two groups might suggest a relationship between the deliberation scale and the model parameter due to the difference in means on the intuition scale. A regression analysis, on the other hand, avoids this problem because it partials out the correlation between the two covariates before relating the deliberation scale to the model parameter.

It should be clear from the above examples that dichotomization of continuous covariates is a problematic practice and the associated biases can be easily avoided by using an appropriate regression analysis. The goal of the present work is therefore to develop a hierarchical regression framework for cognitive models that allows researchers to directly relate model parameters to covariates. Specifically, we will use a hierarchical Bayesian approach as it allows for a principled
10. Using Bayesian Regression to Incorporate Covariates into Hierarchical Cognitive Models

Figure 10.1: Biases in analyzing dichotomized variables. (A) Error variance in t-test based on dichotomization compared to regression analysis. The scatterplot shows the relationship between a covariate and a model parameter (grey dots), the dashed line indicates the median of the covariate, horizontal black lines show the mean parameter values for each group obtained by dichotomization along the median, the grey arrow indicates the resulting error for one data point. The diagonal black line shows the least-squares regression line, the black arrow indicates the associated error. (B) Scatterplot showing the relationship between two correlated covariates (grey dots). The dashed line indicates the median of covariate 2, the dark grey squares show the mean value on both covariates of each group obtained by dichotomizing covariate 2 along the median.

quantification of the evidence for relationships between model parameters and covariates.

Within Bayesian statistics, the principled way of quantifying evidential support for scientific hypotheses is by computing Bayes factors. Bayes factors hold a number of advantages over conventional tests of statistical significance. First, significance tests can only ever reject but never accept the null hypothesis. Bayes factors, on the other hand, can express support for the null hypothesis as well as the alternative hypothesis (Rouder et al., 2009). Second, whilst significance tests force a binary choice upon researchers between rejecting the null hypothesis or remaining in a state of suspended disbelief, Bayes factors allow researchers a graded expression of the evidence for the competing hypotheses provided by their data. Third, conventional significance tests require researchers to commit to a sampling plan before data collection begins and to continue collecting data even if a hypothesis can be confidently rejected or accepted before the full sample has been acquired. Bayes factors, on the other hand, allow researchers to assess the support for competing hypotheses repeatedly during the sampling process and stop collecting data when a hypothesis is supported or rejected to a satisfying degree (Edwards et al., 1963; Kass & Raftery, 1995; Rouder, 2014).

In the next section we develop a Bayesian regression framework that can be easily attached to existing cognitive models and allows researchers to compute Bayes factors for the regression weights relating model parameters to covariates. We subsequently use a simulation study to illustrate the superiority of our Bayesian regression framework compared to a typical dichotomization-based analysis and point out the biases associated with the latter method.
10.1 Regression Framework for Relating Cognitive Model Parameters to Covariates

The regression framework we develop in this section can easily be applied to a wide range of cognitive models, such as multinomial processing trees (Batchelder & Riefer, 1999; Matzke, Dolan, et al., 2015; Riefer & Batchelder, 1988), reinforcement-learning models (Busemeyer & Stout, 2002; Sutton & Barto, 1998), or sequential sampling models (S. D. Brown & Heathcote, 2008; van Ravenzwaaij, Provost, & Brown, in press). As an illustrative example, we use the PVL-Delta model (Ahn et al., 2008; Fridberg et al., 2010; Steingroever, Wetzels, & Wagenmakers, 2013b; Steingroever et al., 2014)—a popular reinforcement-learning model for in the Iowa gambling task (IGT; Bechara et al., 1994). We will therefore first briefly outline the structure of the IGT and give a short summary of the PVL-Delta model and its hierarchical Bayesian implementation before explaining the regression extension of the model.

Iowa Gambling Task and Hierarchical PVL-Delta Model

The IGT is an economic decision-making task that aims to measure decision-making deficits in clinical populations. In the computerized version of the IGT, participants are given an initial credit of $2000 and are presented with four decks of cards, each of which is associated with a characteristic payoff structure. On each trial, participants pick a card and receive feedback about the wins and losses for that card, as well as the running tally. Participants are instructed to choose cards from the decks in a way that maximizes their long-term net outcomes (see Bechara et al., 1994, for more details on the task).

The PVL-Delta model aims to explain the cognitive processes that drive participants’ choices on the IGT. We will focus here on a conceptual description of the model, its parameters, and their interpretation. The PVL-Delta model conceptualizes risky decision making as a three-step process that is governed by four model parameters. First, a loss aversion parameter $w \in [0, 5]$ describes the weighting of net losses relative to net gains, where a value of $w > 1$ means that negative net outcomes impact the subjective utility more strongly than positive net outcomes. Values close to 1 imply equal weighting of net losses and net wins, and as $w$ approaches 0, net losses are increasingly neglected. Second, the outcome sensitivity parameter $A \in [0, 1]$ determines the shape of the utility function. As $A$ approaches 1, the utility function becomes more linear, meaning that the subjective utility of the decks increases proportionally with increasing net outcomes, whereas as $A$ approaches 0, the utility function approximates a step function, meaning that the subjective utility is determined only by the sign of the net outcomes but not their actual value. Third, an updating parameter $a \in [0, 1]$ determines how past expectancies influence the evaluation of the current outcome. A value of $a$ close to 1 indicates quick forgetting and strong recency effects whilst a value of $a$ close to 0 indicates slow forgetting and weak recency effects. Finally, the response consistency parameter $c \in [0, 5]$ determines the relative amount of exploitation vs. exploration, with values close to 0 leading to random choice behavior and larger values leading to more deterministic behavior.

Steingroever et al. (submitted) have presented a Bayesian hierarchical implementation of the PVL-Delta model (solid arrows in Figure 10.2; see also Steingroever, Wetzels, & Wagenmakers, 2013b; and see Wetzels, Vandekerckhove, et al., 2010 for a hierarchical implementation of the related Expectancy Valence model). In their hierarchical implementation of the model, trials $t$ of the IGT (inner plate) are nested within participants $i$ (outer plate). For each trial $t$ of participant $i$ the choice of a deck of cards on the subsequent trial $Ch_{i,t+1}$, and the wins $W_{i,t}$ and losses $L_{i,t}$ on the current trial are observed nodes (grey rectangles); the utility $U_{i,t}$, the expected
utility \( E v_{k,i,t} \), the sensitivity parameter \( \theta_i \), and the probability of choosing deck \( k \) on the next trial \( P[S_k(t + 1)] \) are deterministic nodes (double-bordered circles) as they are fully determined by the model equations and parameters. Moreover, the individual-level model parameters \( z_i \in \{A_i, w_i, a_i, c_i\} \) are modelled based on their probit-transforms, which means that the individual-level model parameters \( z_i \) are treated as deterministic nodes. Their probit-transforms \( z_i' \), on the other hand, are stochastic nodes (single-bordered circles) sampled from a group-level normal distribution with mean \( \mu_{z'} \) and standard deviation \( \sigma_{z'} \). The priors for the group-level parameters are a standard normal distribution \( \mu_{z'} \sim N(0,1) \) and a uniform distribution \( \sigma_{z'} \sim U(0,1.5) \).

### Regression Extension

Our implementation of the regression model is based on the framework for Bayesian regression analysis suggested by Liang, Rui, German, Clyde, and Berger (2008), and Rouder and Morey (2012). The group-level normal priors for the probit-transformed model parameters in Steingroever et al.’s model (submitted) are replaced by a regression extension that relates the individual-level model parameters \( z_i' \) to the covariates \( j = 1 \ldots P \). Specifically, each individual-level probit-transformed model parameter is sampled from the following normal distribution:

\[
z_i' \sim N(\mu_{z'}, \beta_{z'} \beta_{z'}, \sigma_{z'}^2).
\]

Here \( \mu_{z'} \) is the intercept of the regression line, \( x_i' \) is a transposed \( P \times 1 \) vector of \( P \) centered covariate values for participant \( i \) (i.e., \( x_i' = [x_i1 - \bar{x}_1, \ldots, x_iP - \bar{x}_P] \) with \( \bar{x}_j \) the mean of covariate \( j \)), \( \beta_{z'} \) is the \( P \times 1 \) vector of conventional regression weights for model parameter \( z' \), and \( \sigma_{z'}^2 \) is the residual variance of the model parameter \( z' \). The standardized effect size for covariate \( j \) is a transformation of the conventional regression weight:

\[
\alpha_{z'j} = \beta_{z'j} \frac{s_j}{\sigma_{z'}},
\]

where \( s_j \) is the standard deviation of the covariate. The advantage of this reparameterization is that the standardized effect sizes \( \alpha \) can be assigned a (multivariate) Cauchy prior. Cauchy priors have a number of favorable theoretical properties, such as leading to consistent Bayes factors (see Rouder & Morey, 2012 for a discussion), but are computationally inefficient. To improve computational efficiency, the Cauchy distribution can be expressed as a continuous mixture of normal distributions, which is known as the mixture of g-prior (Zellner & Siow, 1980). Specifically, in the case of a single model parameter and a single covariate there is only one standardized effect size \( \alpha_{z'} \) that is assigned a Cauchy prior. To express this Cauchy prior as a mixture of normal distributions, the standardized effect size is assigned a normal prior that depends on a random variable \( g \):

\[
\alpha_{z'} \mid g \sim N(0, g),
\]

and \( g \) is assigned an inverse gamma prior:

\[
g \sim IG(1/2, s^2/2),
\]

where the shape parameter of the inverse gamma distribution is set to 1/2 and the scale parameter is set to \( s^2/2 \). By integrating over \( g \), one obtains a Cauchy prior with scale parameter \( s \), where \( s \) describes the interquartile range of plausible values for \( \alpha_{z'} \).

In the more general case of multiple covariates, a multivariate Cauchy prior can be constructed by assigning the vector of regression weights for each model parameter a multivariate normal prior that depends on a random variable \( g \) and the matrix of covariate values:

\[
\beta_{z'} \mid g \sim N(0, g \sigma_{z'}^2 (X^T X/N)^{-1}),
\]

180
\(N\) denotes the multivariate normal distribution, the mean of the distribution is set to 0 which is a \(P \times 1\) vector of 0s, and the variance-covariance matrix is \(g\sigma_z^2 (X'X/N)^{-1}\). Moreover, \(X\) is the centered \(N \times P\) design matrix that contains the \(P\) centered covariate values for each of the \(N\) participants. \(X^T\) is the transpose of the centered design matrix, and \((X^TX/N)^{-1}\) is the inverse of \((X^TX/N)\). Finally, \(g\) is assigned an inverse gamma prior:

\[
g \sim IG(1/2, s^2/2),
\]  

(10.6)

where the shape parameter of the inverse gamma is 1/2, and the scale parameter is \(s^2/2\). The multivariate Cauchy prior that results from integrating over \(g\) has scale parameter \(s\), which again describes the interquartile range of plausible values for \(\alpha\) (see Appendix H for details on the prior distribution).

Figure 10.2 shows the graphical implementation of the PVL-Delta model with our regression extension. The model components we added to the Steingroever et al.’s (submitted) hierarchical implementation of the PVL-Delta model are indicated by dashed lines. As in the hierarchical PVL-Delta model, the probit-transformed model parameters are stochastic nodes that are nested within participants. However, in addition to the group-level stochastic entities \(\mu_z\) and \(\sigma_z\), the model parameters also depend on the vector \(\beta_z\) and the observed vector of covariate values \(x\), that is nested within participants; the relationship between these entities is given by Equation 10.1. Moreover, the vector \(\beta_z\) depends on the vector of covariate values \(x\) via Equation 10.5. In line with Steingroever et al.’s (submitted) implementation of the hierarchical PVL-Delta model, we assigned the intercept \(\mu_z\) a standard normal prior \(\mu_z \sim \mathcal{N}(0, 1)\). We assigned the residual variance \(\sigma_z^2\), an inverse-gamma prior \(\sigma_z^2 \sim IG(2, 1/2)\) with shape parameter 2 and scale parameter 1/2, instead of the uniform prior used in the hierarchical PVL-Delta model. Our choice of a relatively informative prior was mainly made to speed up model convergence (see below). Nevertheless, we also tried a uniform prior which yielded qualitatively identical results. Finally, we assigned the vector of regression weights \(\beta_z\) the mixture of \(g\) prior described in Equations 10.5 – 10.6 and set the scale parameter \(s = 1\). The Stan code for the model can be downloaded from our Open Science Framework folder at osf.io/6tfz3.

### Computing Bayes Factors

Within the regression framework developed above, researchers can test for a relationship between a normally distributed model parameter, in our case the probit-transformed parameter \(z\), and a covariate \(x\) by computing the Bayes factor for the standardized effect size \(\alpha_{z,x}\). Bayes factors are the standard way of testing hypotheses within the Bayesian framework. They express the relative likelihood of the observed data \(y\) under two competing hypotheses, \(H_0\) and \(H_1\) (Rouder et al., 2009):

\[
BF_{01} = \frac{p(y \mid H_1)}{p(y \mid H_0)}.
\]  

(10.7)

A sensible null hypothesis might be that the standardized effect size for model parameter \(z\) on the covariate \(x\) is 0, \(H_0 : \alpha_{z,x} = 0\), and the alternative hypothesis might state that the standardized effect size is not 0, \(H_1 : \alpha_{z,x} \neq 0\), that is, a point-null hypothesis that is nested under the alternative hypothesis. For such hypotheses, the Bayes factor for the parameter in question can conveniently be obtained using the Savage-Dickey density ratio (Dickey & Lientz, 1970; Wagenmakers et al., 2010). According to this test, the Bayes factor is the ratio of the alternative hypothesis’s prior density over its posterior density at the point-null \(BF_{01} = p(\alpha_{z,x} = 0 \mid H_1)/p(\alpha_{z,x} = 0 \mid y, H_1)\).
10. Using Bayesian Regression to Incorporate Covariates into Hierarchical Cognitive Models

\[ \mu_{z'} \sim N(0, 1) \]
\[ \sigma_{z'}^2 \sim IG(2, 1/2) \]
\[ \beta_{x'} | g \sim N(0, \Sigma_{\beta_{x'}}) \]
\[ g \sim IG(1/2, s^2/2) \]
\[ \Sigma_{\beta_{x'}} = (g\sigma_{z'}^2 X^T X / N)^{-1} \]

\[ z_i' \sim N(\mu_{z'} + x_i' \beta_{x'}, \sigma_{z'}^2) \]
\[ A_i = \Phi(A_i') \]
\[ w_i = 5\Phi(w_i') \]
\[ a_i = \Phi(a_i') \]
\[ c_i = 5\Phi(c_i') \]

Figure 10.2: Hierarchical PVL-Delta model with regression extension. Solid lines indicate components of Steingroever et al.’s (submitted) hierarchical implementation of the PVL-Delta model; newly added regression components for relating model parameters to covariates are indicated by dashed lines.

10.2 Simulation Study

The goal of our simulation study is to demonstrate how dichotomizing continuous covariates biases Bayes factors and how these biases can be avoided using the regression framework developed above. To generate realistic data for our simulations, we first fitted the PVL-Delta model with the regression extension to a published data set (Steingroever et al., submitted; see also Steingroever, Davis, et al., 2015). We subsequently used the resulting parameter estimates to generate synthetic data for two scenarios, one in which covariates are not correlated with each other, and one in which covariates are correlated. To emulate a typical dichotomization-based analysis strategy, we applied the PVL-Delta model in combination with a median-split of the covariates to the simulated data. Finally, we compared the resulting Bayes factors from the dichotomization-based analysis to the Bayes factors obtained from the PVL-Delta model in combination with our regression extension.

Generating synthetic data

Data set

We based the setup for our simulated data on the data published in Steingroever et al.’s (submitted) because of the simple experimental design and the clear structure of the covariates measured. In
Steingroever et al.’s study 70 participants performed 100 trials of the IGT [Bechara et al. (1994)]. In addition, they completed Betsch and Ianello’s (2010) decision style questionnaire, which consists of 70 items that assess participants’ tendency to use an intuitive or deliberate decision style on a seven-point Likert scale. Steingroever et al. submitted participants’ responses to a principal component analysis and divided participants into two groups depending on their scores on the two factors, deliberation and intuition. In addition, they fitted the PVL-Delta model to the IGT choices from the deliberate and the intuitive decision makers, respectively, and related each participant’s factor scores to the estimated PVL-Delta parameters. A full description of the sample, IGT, and questionnaire data can be found in Steingroever et al. (submitted).

We fitted the PVL-Delta model with the regression extension to Steingroever et al.’s IGT data and used participants’ scores on the deliberation and intuition scales as covariates $x_1$ and $x_2$, respectively. In contrast to Steingroever et al., whose analysis only included the data of participants who scored high on one scale and low on the other, we included the data of all participants in our analysis. As Steingroever et al. reported relatively small effects of the covariates on the model parameters, we expected to also find relatively small standardized effect sizes $\alpha_{z'j}$ and therefore set the scale parameter of the Cauchy prior to $s = 1/3$ (Equation 10.6). To fit the PVL-Delta model to the data, we implemented the model with the regression extension in Stan (Carpenter et al., in press; Stan Development Team, 2016a, 2016b) and obtained samples from the posterior distributions of the model parameters. For each model parameter we ran three MCMC chains and collected 50000 posterior samples per chain. We discarded the first 5000 samples of each chain as burn-in samples and furthermore thinned each chain, discarding 4 out of every 5 samples. Starting values for the population means $\mu_{z'}$ were randomly drawn from standard normal distributions, starting values for the population standard deviations $\sigma_{z'}$ were randomly drawn from exponential distributions with scale parameter 1, and starting values for the standardized effect sizes $\alpha_{z'}$ were randomly drawn from normal distributions with mean 0 and standard deviation 2. All chains were run until convergence (Gelman-Rubin diagnostic $\hat{R} \leq 1.004$, Gelman & Rubin, 1992).

Model fit and generating parameter values

The left half of Table 10.1 shows the estimated posterior means for our fit of Steingroever et al.’s data. As can be seen, the regression weights $\beta_{z'j}$ for the regression of participants’ model parameters on their covariate values are relatively small; the strongest relationships are between the outcome sensitivity parameter $A$ and the deliberation scale, and between the loss aversion parameter $w$ and the intuition scale (i.e., $\beta_{A'1} = 0.61$ and $\beta_{w'2} = -0.51$, respectively). Nevertheless, compared to the ratio of the standard deviations of the model parameters and the covariates, the regression weights $\beta_{z'j}$ are very small. However, to be able to demonstrate the adverse effects of dichotomizing covariates, we needed to generate data with clearly identifiable relationships between model parameters and covariates (recall that, in the case of uncorrelated covariates, dichotomizing covariates should result in statistical tests missing existent effects). We therefore set $\beta_{A'1} = 1$ and $\beta_{A'2} = 0$, which means that outcome sensitivity should be associated with deliberation, but not intuition. In addition, we set $\beta_{w'1} = 0$ and $\beta_{w'2} = -0.9$, which means that loss aversion should be negatively associated with intuition, but not deliberation. Because the regression weights $\beta_{A'}$ and $\beta_{w'}$ were now larger than the values estimated in our model fit, we needed to reduce the residual variance for the corresponding model parameters to maintain reasonable variance in the covariate scores between participants (compare Equation 10.2). We therefore set the residual variances $\sigma_A^2$ and $\sigma_w^2$ to $3/8$ the values estimated in our model fit. The resulting parameter values used to generate data in our simulations are shown in the right half of Table 10.1.
10. Using Bayesian Regression to Incorporate Covariates into Hierarchical Cognitive Models

Table 10.1: Posterior estimates of parameter values for Steingroever et al.'s (submitted) data and adjusted parameter values used to generate synthetic data. Subscript 1 indicates effect sizes for the deliberation scale, subscript 2 indicates effect sizes for the intuition scale.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated</th>
<th>Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z'$</td>
<td>$\beta_{z'1}$</td>
<td>$\mu_{z'}$</td>
</tr>
<tr>
<td>$A'$</td>
<td>0.61</td>
<td>0.31</td>
</tr>
<tr>
<td>$w'$</td>
<td>-0.04</td>
<td>-0.51</td>
</tr>
<tr>
<td>$a'$</td>
<td>-0.08</td>
<td>0.24</td>
</tr>
<tr>
<td>$c'$</td>
<td>-0.02</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

Adjusted $\beta_{z'1}$ | $\beta_{z'2}$ | $\mu_{z'}$ | $\sigma_{z'}$ |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1.06</td>
</tr>
</tbody>
</table>

Note. Empty cells indicate that the parameter values for generating synthetic data equaled the values estimated for Steingroever et al.’s data.

Data generation

For our simulations we generated 50 synthetic data sets. Each simulated data set consisted of 150 synthetic participants, which should allow our regression analysis to reliably detect relationships between model parameters and covariates. For each participant we generated two covariate values, $x_{i1}$ and $x_{i2}$, as well as one value for each of the four PVL-Delta parameters. To obtain covariate values that were related to a specific model parameter, but not to others, we started by generating a $2 \times 1$ vector of covariate values for each participant from a multivariate normal distribution, $x_i \sim N(\mu, \Sigma)$, with $2 \times 1$ mean vector $\mu = 0$, and $2 \times 2$ covariance matrix $\Sigma$. In the case of covariates that did not correlate with each other the covariance matrix was the identity matrix. In the case of covariates that were correlated with each other the covariance matrix had diagonal entries 1 and off-diagonal entries 0.7. In a second step, we generated probit-transformed PVL-Delta parameters for each participant using Equation 10.1, $z_i' \sim N(\mu_{z'} + x_i'\beta_{z'}, \sigma_{z'})$. We set the data-generating group-level parameter values for the regression weights $\beta_{z'}$, mean group-level parameters $\mu_{z'}$, and residual variances $\sigma_{z'}^2$ to the values given in Table 10.1.

Finally, based on the four PVL-Delta parameters, we simulated 200 trials of the IGT for each participant. We doubled the number of trials per participant compared to the data in Steingroever et al.’s (submitted) to reduce the impact of imprecise estimates of the PVL-Delta parameters on the estimation of the regression weights. To generate simulated IGT trials for each participant, we first spawned a set of payoffs for each deck of cards based on the payoff scheme used in Steingroever et al.’s (submitted) study. We then applied the cumulative distribution function of the standard normal distribution to the probit-scaled model parameters $z_i'$ generated previously to obtain the corresponding PVL-Delta parameter $z$. We furthermore initialized the expected utilities for all four decks of cards to be 0, meaning that the choice of the first card was entirely random for all simulated participants. After generating a random choice on the first trial for each participant, we evaluated the outcome, updated the expected utilities, and generated the participant’s choice on the next trial and the parameter values for that simulated participant. We continued this process iteratively until we had accumulated 200 simulated choices. Further details and the R code used to generate the simulated data can be found at osf.io/6tfz3.

Analysis Using Dichotomized Covariates

Dichotomization-based analysis strategies take several forms. One that is frequently seen in practice is the median-split. In a median-split analysis, participants are divided into two groups based on whether their value on the covariate lies above or below the median. To emulate this analysis
strategy in the context of our simulation study, we developed a version of the PVL-Delta model that estimates separate group-level means $\mu_{z'}$ for participants scoring above the median on one covariate and below the median on another covariate. Note that including these separate group-level means in the model constitutes a relatively sophisticated version of a dichotomization-based analysis; in practice, researchers are more likely engage in a two-step analysis approach, first fitting the cognitive model separately to the groups of participants scoring above and below the median, and subsequently testing the estimated model parameters for differences between groups. However, such a two-step procedure introduces additional biases beyond those introduced by dichotomization which is beyond the scope of the present work.

Our median-split model assumes the same hierarchical structure as the PVL-Delta model, with trials being nested within participants whose probit-transformed parameter values are sampled from a group-level normal distribution. The mean of the group-level distribution from which a participant’s probit-transformed parameter values are drawn depends on the person’s values on the covariates. Therefore, each participant $i$ is assigned a $P \times 1$ vector $d_i$ of dummy variables, where the $j$th entry of the vector is 1 if the person’s score on covariate $j$ is greater than the median, and 0 otherwise:

$$z' \sim \mathcal{N}(\mu_{z'}, \delta_{z'}d_i\sigma_{z'}, \sigma_{z'}^2).$$

(10.8)

Here $\mu_{z'}$ is the mean for a person who scores below-median on all covariates (i.e., the person’s dummy variables are all 0). Furthermore, $\delta_{z'}$ is the transposed $P \times 1$ vector of standardized effect sizes (i.e., $\delta_{z'} = [\delta_{z'}1, \ldots, \delta_{z'}P]'$) and $\delta_{z'}1$ is the standardized effect size indicating the difference, in standard deviations, between the group-level mean for a person with below-median values on all covariates, and the group-level mean for a person with an above-median value on covariate $j$ and below-median values on all other covariates. Finally, $\sigma_{z'}^2$ is the variance of the model parameter $z'$.

As in the PVL-Delta model with the regression extension, we assigned the group-level means $\mu_{z'}$ a standard normal prior $\mu_{z'} \sim \mathcal{N}(0, 1)$, and the group-level variance $\sigma_{z'}^2$ an inverse-gamma prior $\sigma_{z'}^2 \sim IG(2, 1/2)$ with shape parameter 2 and scale parameter 1/2. Finally, we assigned the standardized effect sizes $\delta_{z'}$ independent Cauchy priors $\delta_{z'} \sim C(1)$ with location parameter 0 and scale parameter 1.

**Data Analysis**

We analyzed the simulated data using the PVL-Delta model with regression extension and the median-split version of the PVL-Delta model. For both models we computed Bayes factors contrasting the null hypothesis that there is no relationship between model parameters and covariates with the alternative hypothesis that there is such a relationship. More specifically, in the case of the regression model, the null hypothesis stated that the standardized effect size for a specific model parameter $z'$ on a specific covariate $x_j$ is 0, $\mathcal{H}_0: \alpha_{z' j} = 0$, and the alternative hypothesis stated that the standardized effect size is not 0, $\mathcal{H}_1: \alpha_{z' j} \neq 0$. In the case of the median-split model, the null hypothesis stated that the standardized difference in group means is 0, $\mathcal{H}_0: \delta_{z' j} = 0$, and the alternative hypothesis stated that the difference in group means is not 0, $\mathcal{H}_1: \delta_{z' j} \neq 0$.

We based our computation of the Bayes factors for both models on the Savage-Dickey density ratio (Dickey & Lientz, 1970; Wagenmakers et al., 2010). To obtain estimates of the posterior density for each model’s effect size parameters, we first implemented both models in Stan (Carpenter et al., in press; Stan Development Team, 2016a, 2016b). As we expected sizable effects in the simulated data, we set the scale parameter for the regression model’s Cauchy prior to $s = 1$. We subsequently ran MCMC chains until convergence (Gelman-Rubin diagnostic $R \leq 1.001$, Gelman & Rubin, 1992). For each model parameter we ran two MCMC chains and collected 45000
posterior samples per chain. We discarded the first 5000 samples of each chain as burn-in samples and furthermore thinned each chain, discarding four out of every five samples, which left us with a total of 8000 samples per chain. As the prior distribution for the regression weights $\beta_{zj}$ in the model with regression extension cannot easily be expressed in closed form, we approximated the density of the prior distribution for each regression weight using MCMC sampling with the same setup as for the model fit. We estimated the density of the posteriors and the priors for the $\beta_{zj}$ using log-spline functions, and computed the exact value of the Cauchy priors for the $\delta_{zj}$. Finally, we computed the Bayes factors by taking the ratio of posterior densities to the prior densities at 0.

Results

Figure 10.3 shows the log Bayes factors for the alternative hypothesis obtained in our simulations. We chose to plot the log of the Bayes factors here, rather than the Bayes factors, because the Bayes factors spanned up to five orders of magnitude, which means that, on the linear scale, large Bayes factors would obscure differences in Bayes factors at the low end of the scale. Moreover, because we generated our data in such a way that only the PVL-Delta parameters $A$ and $w$ had sizable relationships with the covariates, we will only present the results for these parameters here. The full results for all model parameters as well as details on the estimated effect sizes can be found in Appendix II. The left panel of Figure 10.3 shows the log Bayes factors for our simulations with uncorrelated covariates. As can be seen, the Bayes factors obtained from the regression analysis showed strong evidence for an effect of the first covariate on the $A$ parameter (dark grey dots, left column in the top row), whereas the median-split analysis provided much weaker evidence for such an effect (light grey dots, left column in the top row). To quantify the difference between Bayes factors on the linear scale, we took the negative reciprocal of Bayes factors favoring the null hypothesis (i.e., $BF < 1 \rightarrow -BF^{-1}$) and computed the relative difference defined as: $(BF_{MS} - BF_{RG})/|BF_{RG}|$, where $BF_{MS}$ are the Bayes factors from the median-split analysis and $BF_{RG}$ are the Bayes factors from the regression analysis. For the effect of the first covariate on the $A$ parameter, the median relative difference was -0.85, indicating a strong underestimation of the evidence in the median-split analysis. Similarly, the regression analysis strongly supported an effect of the second covariate on the $w$ parameter (dark grey dots, right column in the bottom row), whereas the median-split analysis provided much weaker evidence for such an effect (light grey dots, right column in the bottom row). The median relative difference in Bayes factors was -0.95, indicating a tremendous underestimation of the evidence in the median-split analysis. For the null-effects of the first covariate on the $w$ parameter (right column, top row) and of the second covariate on the $A$ parameter (left column, bottom panel), both analyses performed similarly, with median relative differences in Bayes factors of 0.48 and 0.40, respectively, indicating that the median-split analysis favored the alternative hypothesis slightly more strongly than the regression analysis.

The right panel for Figure 10.3 shows the log Bayes factors for our simulations with correlated covariates. The Bayes factors obtained from the regression analysis again showed stronger evidence for an effect of the first covariate on the $A$ parameter (dark grey dots, left column in the top row) than the median-split analysis (light grey dots, left column in the top row). However, the median relative difference in Bayes factors of -0.62 was much smaller than in the case of uncorrelated covariates. Similarly, the regression analysis provided stronger support for an effect of the second covariate on the $w$ parameter (dark grey dots, right column in the bottom row), than the median-split analysis (light grey dots, right column in the bottom row). The median relative difference in Bayes factors of -0.89 was slightly smaller than in the case of uncorrelated covariates. Unlike in the case of uncorrelated covariates, in the case of correlated covariates the median-split
\[ \rho(X_1, X_2) = 0 \]

\[ \rho(X_1, X_2) \neq 0 \]

Figure 10.3: Bayes factors from 50 simulated data sets for the regression and median-split analysis. Data points show the log Bayes factors for the alternative hypothesis (log(BF)) obtained in the regression (RG, dark grey dots) and median-split (MS, light grey dots) analysis for the PVL-Delta model’s \( A \) and \( w \) parameters (columns) and two covariates (rows). The left subplot shows the results for the case of uncorrelated covariates, the right subplot shows the results for the case of correlated covariates. Lines indicate the mean log BF. Data points are jittered along the x-axis for improved visibility.
10. Using Bayesian Regression to Incorporate Covariates into Hierarchical Cognitive Models

\[ \rho(X_1, X_2) = 0 \]

Figure 10.4: Posteriors of effect sizes for the case of uncorrelated covariates. Shown are the posterior distributions quantile-averaged across 50 simulated data sets. The left subplot shows the results for the \( A \) parameter, the right subplot shows the results for the \( w \) parameter. Thick black lines are the posteriors of the standardized effect sizes \( \alpha \) (left column in each subplot), thick grey lines are the posteriors of the standardized mean differences \( \delta \) (right column in each subplot), thin grey lines show the priors. The top row shows the results for the first covariate \( (X_1) \), the bottom row shows the results for the second covariate \( (X_2) \).

The biases inherent in the median-split analysis are also clearly visible in the posterior distributions for the effect sizes. Figure 10.4 shows the quantile-averaged posterior distributions of the standardized differences in group means, \( \delta \), and the standardized effect size, \( \alpha \), for the case of uncorrelated covariates. The left column of the top left subplot shows the prior (thin grey line) and the posterior (thick black line) for the regression of the \( A \) parameter on the first covariate. Compared to the prior, which has considerable mass at the point null hypothesis \( \alpha_{A,1} = 0 \), the posterior has nearly no mass at the point null, resulting in Bayes factors that strongly favor the alternative hypothesis. The right column in the same subplot shows the prior (thin grey line) and posterior (thick grey line) for the standardized difference in the \( A \) parameter between participants.
10.2. Simulation Study

\[ \rho(X_1, X_2) \neq 0 \]

Figure 10.5: Posteriors of effect sizes for the case of correlated covariates. Shown are the posterior distributions quantile-averaged across 50 simulated data sets. The left subplot shows the results for the \( A \) parameter, the right subplot shows the results for the \( w \) parameter. Thick black lines are the posteriors of the standardized effect sizes \( \alpha \) (left column in each subplot), thick grey lines are the posteriors of the standardized mean differences \( \delta \) (right column in each subplot), thin grey lines show the priors. The top row shows the results for the first covariate \((X_1)\), the bottom row shows the results for the second covariate \((X_2)\).

who score above-median on the first covariate and participants who score below-median. As can be seen, the posterior has little mass at the point null hypothesis \( \delta_{A,1} = 0 \), resulting in Bayes factorsfavoring the alternative hypothesis. However, compared to the posterior under the regression model, the posterior under the median-split model is considerably wider and has more mass at the point null, which results in the underestimation of the evidence against the null observed above. A comparable pattern can be seen in the bottom right subplot; the posterior under the median-split model has more mass at the point null than the posterior under the regression model, resulting in a strong underestimation of the evidence against the null. Finally, the top right and bottom left subplots show the comparison for the true null-effects of the first covariate on the \( A \) parameter and of the second covariate on the \( w \) parameter, respectively. Although the posterior under the median-split model again has less mass at the point null than the posterior under the regression model, the differences are less pronounced and both models favor the null hypothesis.

Figure 10.5 shows the quantile-averaged posterior distributions of the standardized differences
in group means, \( \delta \), and the standardized effect size, \( \alpha \), for the case of correlated covariates. The top left and bottom right subplots show comparable patterns to the case of uncorrelated covariates; the posterior under the regression and the median-split model both have much less mass at the point null than the respective prior, resulting in Bayes factors favoring the null hypothesis. However, compared to the prior, the posterior under the regression model is much narrower than the posterior under the median-split model, which leads to smaller Bayes factors under the median-split model. Finally, the top right and bottom left subplots show the comparison for the true null-effects of the first covariate on the \( A \) parameter and of the second covariate on the \( w \) parameter, respectively. As can be seen, the posterior for the regression weights is centered at 0 and has considerably more mass at the point null than the prior. Therefore, Bayes factors under the regression model correctly favor the null hypothesis. However, the posterior under the median-split model lies to the left of the point null for the \( A \) parameter and to the right of the point null for the \( w \) parameter, and thus has considerably less mass at the point null than the posterior under the regression model. Consequently, Bayes factors under the median-split model understate the evidence for the null and in many instances even support the alternative hypothesis, suggesting spurious associations between the first covariate and the \( w \) parameter and between the second covariate and the \( A \) parameter.

### 10.3 Discussion

The goal of the present work was to develop a methodological framework that allows researchers to test hypotheses about associations between the cognitive processes and behavioral and physiological covariates in a principled way. To this end we showed how Bayesian linear regression can be used to relate the parameters of a cognitive model to covariates, and use Bayes factors to quantify the evidential support for specific associations between model parameters and covariates. As an example application, we chose the PVL-Delta model which aims to explain risky decision making as the result of a reinforcement-learning process. Adding a regression extension to the PVL-Delta model allowed us to simultaneously account for participants’ model parameters and measurements of participants’ preferred decision styles.

One major advantage of incorporating a Bayesian regression framework into cognitive models is that it obviates the need for dichotomization-based analysis strategies, circumventing the associated statistical biases. Despite repeated warnings against the use of dichotomization-based analyses (Austin & Brunner, 2004; Cohen, 1983; Maxwell & Delaney, 1993; MacCallum et al., 2002; Royston et al., 2006), a number of recent studies have relied on median splits (e.g., Beitz, Salthouse, & Hasker, 2014; Cooper et al., 2015; Kwak et al., 2014; Steingroever et al., submitted) to test for associations between the parameters of different cognitive models and covariates. We conducted a simulation study to illustrate the degree of bias introduced by such flawed analysis strategies. To this end, we generated simulated data under two scenarios. In one scenario covariates were not correlated with each other, and some of the covariates were correlated with some of the model parameters but not others. In the other scenario covariates were correlated with each other, and some of the covariates were correlated with some of the model parameters but not others. Our simulations showed that, in the first scenario, a median-split analysis leads to Bayes factors that understated the evidence for true effects compared to the Bayes factors obtained from a regression model. In the second scenario, Bayes factors from a median-split analysis again understated the evidence for true effects but, in addition, a median-split analysis also suggested spurious effects of covariates on model parameters that were, in fact, unrelated.

Interestingly, the Bayes factors for spurious effects suggested by the median-split analysis were
relatively small compared to the Bayes factors for true effects. This result is most likely due to the fact that the median-split analysis generally leads to wider posteriors than the regression analysis, resulting in overall smaller Bayes factors. Surprisingly, for the median-split as well as the regression analysis, Bayes factors favoring the null hypothesis were very modest across simulations. There are two probably causes for this result. First, due to the relatively large uncertainty in the estimates of the model parameters, posteriors were relatively wide compared to the priors, putting relatively little posterior mass at the point null. Second, the default scale parameter for the priors on the effect sizes might have been too small compared to the effects in the simulated data, putting too much prior mass at the point null, thus resulting in small Bayes factors for the null hypothesis.

With the goal of a general-purpose application in mind, one question that inevitably arises is how easily the Bayesian regression framework used here can be adapted to other cognitive models. The requirements with respect to the model are relatively modest. First, the cognitive model needs to be implemented in a hierarchical way to allow researchers to relate individual participants’ model parameters to measured covariates. For many popular models such hierarchical implementations are readily available (Matzke, Dolan, et al., 2015; Steingroever et al., 2014; Wiecki, Sofer, & Frank, 2013; Ahn, Haines, and Zhang's, 2016, R-package contains hierarchical implementations of several popular models of decision making) or can be easily developed using MCMC software packages such as JAGS (Plummer, 2003) or Stan (Carpenter et al., in press; Stan Development Team, 2016b). Second, the model parameters of interest need to be normally distributed. Although this assumption is often reasonable and can be readily adopted, in other cases, specific bounds on the parameter values are required due to the cognitive interpretation of the parameters or mathematical constraints. However, such constraints can often be overcome by using transformed versions of the model parameters, rather than the model parameters themselves, in the regression analysis. In the case of the PVL-Delta model, for instance, all model parameters are restricted to closed intervals, yet probit transforming the parameters allowed us to add the Bayesian regression extension to the PVL-Delta model. These two conditions are all that is required for our regression extension to be added to a cognitive model and are easily met by most existing models.

Although reinforcement-learning models, and the PVL-Delta model in particular, served merely as an example for our Bayesian regression framework, we believe that our regression extension can greatly facilitate research involving risky decision making. One potential application beyond identifying relationships between model parameters and physiological measurements is the statistical control of nuisance variables. A number of authors have suggested that performance on the IGT might be subject to practice effects (Ernst et al., 2003; Lejuez et al., 2003; Verdejo-García & Pérez-García, 2007), although no study to date has comprehensively addressed this problem (Buelow & Suhr, 2009). Including time-on-task as a covariate in model-based analyses might allow researchers not only to control for practice effects but also to pinpoint which cognitive processes are affected by practice and which processes remain stable over time. Similar model-based analyses in perceptual decision making, for example, have suggested that whilst participants’ processing of stimuli remains unaffected by practice, their response mode can change over time although considerable practice might be needed for participants to reach optimal performance (Hawkins, Forstmann, Wagenmakers, Ratcliff, & Brown, 2015; Simen et al., 2009).

To conclude, in the present work we presented a hierarchical Bayesian regression extension for cognitive models that relates participants’ estimated model parameters to behavioral and physiological covariates. This regression framework allows researchers to test hypotheses about relationships between model parameters and covariates in a principled manner using Bayes factors. Moreover, our regression framework overcomes many of the biases associated with alternative analysis strategies often seen in practice. In particular, we illustrated in our simulation study the superiority of our regression framework to the often-practiced median-split analysis that can
10. Using Bayesian Regression to Incorporate Covariates into Hierarchical Cognitive Models

lead researchers to either miss existing relationships between model parameters and covariates, or suggest spurious associations between model parameters and covariates, depending on whether the covariates are correlated with each other or not. Due to its relatively modest requirements for the cognitive model under consideration, our regression extension can be easily applied to a vast number of existing models and promises to obviate the need for inappropriate, cumbersome, and biased analysis strategies such as the median-split.

Acknowledgements

We thank Thorsten Pachur for providing the data used in this article. This research was supported by a Netherlands Organisation for Scientific Research (NWO) grant to UB (406-12-125) and to HS (404-10-086), and a European Research Council (ERC) grant to EJW (283876).