BAYESIAN MODEL AVERAGING AND THE OPTIMIZATION OF WORKFORCE PREDICTIONS

Gordon Goodwin, May, 2022

Director of Thesis: Dr. Alexander Schoemann

Major Department: Department of Psychology

ABSTRACT

This thesis explores the viability of Bayesian model averaging (BMA) as an alternative to traditional general linear models (GLMs) and ensemble-based machine-learning methods commonly used to predict workforce outcomes. In contrast to both the practices that focus on selecting a single "best" GLM and set of predictors, and the ensemble-based machine-learning methods that combine many simpler models, the BMA approach explores the space of all models to be considered and assigns probabilistic weights to each. These posterior model probabilities (PMPs) can then be used to generate optimal predictions regarding future data observations via a weighted-average of the model-specific predictions. By averaging over models, BMA routines are well-suited to addressing the model uncertainty that arises when a researcher has numerous potential predictor variables. Rather than condition inferences upon a single model and set of predictors, or upon a collection of poorer models and simpler subsets, the BMA routine can average predictions across all possible combinations of predictor variables. Focusing upon this form of model uncertainty, this thesis demonstrates how BMA might be employed to optimally forecast workforce outcomes in both classification and regression contexts by way of two illustrative case studies related to the prediction of employee turnover intentions.

BAYESIAN MODEL AVERAGING AND THE OPTIMIZATION OF WORKFORCE PREDICTIONS

A Thesis

Presented To the Faculty of the Department of Psychology

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Arts in Psychology

by

Gordon Goodwin

May, 2022

BAYESIAN MODEL AVERAGING AND THE OPTIMIZATION OF WORKFORCE PREDICTIONS

By

Gordon Goodwin

APPROVED BY:

Director of Thesis _____

Alexander M. Schoemann, PhD

Committee Member _____

Karl L. Wuensch, PhD

Committee Member _____

Mark C. Bowler, PhD

Committee Member _____

Ascher K. Munion, PhD

Chair, Department of Psychology _____

Alan J. Christensen, PhD

Dean of the Graduate School _____

Paul J. Gemperline, PhD

Acknowledgements

This work could not have been possible without the direction of Dr. Alex Schoemann, whose constant positivity, encouragement, and wealth of knowledge proved invaluable. I also owe a debt of gratitude to both Dr. Karl Wuensch, who first inspired my passion for all things quantitative and has probably forgotten more than I will ever know, and Dr. Mark Bowler, who constantly reminded me to connect the theoretical to the practical. Lastly, but certainly not least, I would like to thank Dr. Ascher Munion for making time to have multiple random hour-long conversations with me and for always reminding me to prioritize what's really important. All four faculty members have provided me with an indescribable amount of support and have gone above and beyond to help me personally, academically, and professionally. Thanks as well to my fellow I/O Pirates for your friendship and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF SYMBOLS

CHAPTER 1: INTRODUCTION

In recent years, fueled in part by an exponential increase in data availability and computational power (Oswald et al., 2020), the use of regression and machine learning-based predictive models as a means of workforce analytics has become increasingly critical to organizational success. Alternately referred to at times as "human capital analytics", "people analytics", or "HR analytics", *workforce analytics* can be understood as an evidence-based approach which seeks to understand, quantify, manage, and optimize the performance and impact of an organization's workforce (Cheng et al., 2021; Fitz-Enz & Mattox, 2014). Subsequently, when employed in furtherance of workforce analytics goals, predictive models offer value through their ability to forecast, predict, or classify outcomes related to organizations' people and teams (Putka, 2018). Predictive modeling has been applied to forecast workforce-related outcomes such as labor market dynamics (e.g., Liu et al., 2017), employee retention and turnover (e.g., Gao et al., 2019), future job performance (e.g., Sajjadiani et al., 2019), and employee safety and health (e.g., Ott-Holland et al., 2019). Through the optimization of such outcomes, predictive models are of critical importance to the ability of organizations to remain competitive amidst the current global market, in which highly-skilled workers are considered one of the most important corporate resources available (Chambers et al., 1998; Vaiman et al., 2012).

Despite the organizational utility thus afforded by such applications, there has been a growing perception that industrial/organizational psychology (IOP) and human resources management (HRM) IOP/HRM practitioners lack familiarity with the substantial advances made in predictive modeling in recent years and are inadequately prepared for the challenges and opportunities presented by a growing organizational demand for predictive routines (McCartney

et al., 2020; Oswald et al., 2020; Putka et al., 2018). This is especially concerning in the context of workforce-related applications, where IOP and HRM professionals are uniquely trained to navigate the specific legal, ethical, and sociopolitical intricacies and liabilities inherent to workforce decision-making, such as those surrounding employment law, selection practices, and the impacts of policy changes (Marler & Boudreau, 2017). Further, one particular aspect of predictive modeling that has been largely absent in the organizational research literature is that pertaining to model uncertainty.

Model Uncertainty

Statistical models represent a simplified version of a real-world generative process, constructed by the researcher in a manner that formalizes the relationships between the known components, such as the observed data (y, x), and the unknown components, such as the hypothesized parameters θ or future observations ($\tilde{x}, \tilde{y}$) (Levy & Mislevy, 2016). Critically, any such model is a function of decisions made by the researcher regarding which features relevant to the problem at hand are represented, and which are suppressed or ignored (Levy & Mislevy, 2016). As such, any inferences or predictions generated by way of the modeling routine are also implicitly conditioned upon the model itself, which creates an additional source of uncertainty commonly ignored by the researcher. In the context of predictive models, predictions generated under the traditional modeling framework commonly account for uncertainty regarding both the underlying parameter estimates and predictions of the unknown $\tilde{y}$, but ignore structural uncertainty regarding the generative model itself (Draper, 1995). Simply put, as noted by Steel (2020), model uncertainty is always a given; whether the goal is to identify a "true" data-generating model or to establish an optimal proxy for the "true" model, ultimately uncertainty

always exists regarding the generative process and the specifications and configurations assigned to the chosen model.

It is hard to overstate the importance of model uncertainty for predictive modeling of workforce-related organizational outcomes. While all domains incur uncertainty inherent to model specifications, the obstacle it presents is particularly challenging for IOP and HRM practitioners engaged in empirical work who must address both the natural uncertainty encountered when attempting to predict human thoughts and behaviors in general, as well as that associated with evolving business needs, changing market dynamics, and the ever-changing social, political, and economic climate (Fitz-Enz & Mattox II, 2014). For example, outcomes commonly modelled by IOP practitioners include labor availability and costs, skill development and measurement, and employee engagement and retention, all of which are shaped by both internal and external forces. Amidst this backdrop of uncertainty, and particularly in the context of the current heightened workforce-related tensions (Han & Hart, 2021), the decision to rely upon a single model, or even a combination of simpler submodels, can have significant negative repercussions for an organization's success. Further, while many forms of model uncertainty are relevant to organizational research, of particular importance is the uncertainty regarding which predictor variables to include in a given model

Variable Selection Uncertainty

For a given outcome measure of interest **y**, and a set of potential predictor variables $\{\boldsymbol{X}\} = \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$, uncertainty regarding which predictors to include in a predictive model is one of the most fundamental problems (see Clyde & George, 2004; George, 2000; Miller, 2002). In its simplest form, predictor inclusion uncertainty is present when the researcher is given a number of candidate predictor variables $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots \boldsymbol{x}_k\}$, and desires to find a subset that

optimally predicts the outcome measure of interest (McCullah & Nelder, 1989). In such a context, the set of all models to be considered is collectively referred to as the *model space* $\mathcal{M}$, and each constituent model corresponds to a distinct subset of $\{\boldsymbol{X}\}$, such that the full pool of *k* potential predictors yields 2^k possible models. Following the notation of Clyde and George (2004), the model space corresponding to the variable uncertainty problem can thus be indexed by $\boldsymbol{\gamma}$, a *k*-dimensional vector of indicators within which $\gamma_k = 1$ denotes that a given predictor $\boldsymbol{x}_k$ is included, and $\gamma_k = 0$ denotes $\boldsymbol{x}_k$ is excluded. Following this notation, M_γ can then be used to compactly represent the subset of predictor variables included in a given model.

Model uncertainty related to variable inclusion is a problem that has become increasingly important for organizational researchers in recent years as data availability has grown exponentially (Oswald et al., 2020). In contrast to the deliberate collection of organizational data pertaining to specific prespecified constructs and associated measures, regarding which the majority of IOP training and ethical standards are designed to address (e.g. Society for Industrial Organizational Psychology, 2018), organizational researchers are increasingly presented with unstructured data collected incidentally and/or in real time (Oswald, 2020). In such a context, organizational researchers tasked with the development of predictive models face a heightened degree of certainty regarding which variables to include. Critically, this issue is particularly prescient to IOP and HRM practitioners relative to other domains, as even when the available data are demonstrably relevant to the prediction of organizational outcomes, the choice regarding which predictor variables to include in a model can involve ethical and legal concerns in addition to the standard statistical considerations (Oswald et al., 2020). Consequently, this thesis will focus specifically upon model uncertainty in the context of predictor variable inclusion. At a

high level, two main strategies have been employed to address model uncertainty, referred to respectively as *model selection* and *model averaging.*

Model Selection

Model selection effectively represents an attempt by the researcher to mitigate, rather than account for, model uncertainty. Specifically, under a selection-based predictive modelling approach, the researcher chooses a single "best" model M^*, as measured by some performance metric, and then uses that model to draw inferences and generate predictions, essentially proceeding as if M^* were in fact the "true" generative model (Draper, 1995; Steel, 2020). As noted by Draper (1995) however, the very act of using a data-driven process to search for the optimal M^* implies an implicit acknowledgement by the researcher regarding the existence of some degree of structural model uncertainty. Importantly, predictions generated by model selection methods can only be relied upon insofar as the chosen model embodies a strong approximation of the real-world data-generating process. Given that some aspects are always suppressed by the researcher during the model specification process (Levy and Mislevy, 2016), predictions generated upon a model selection routine are overly precise and often fail to generalize well to out-of-sample contexts (Draper, 1995; Steel, 2020).

Within the traditional single-model framework, two of the methods most commonly used to predict workforce-related outcomes are multiple linear regression and logistic regression, used for regression and classification contexts respectively (Hunter & Schmidt, 1990; Putka et al., 2018; Raju et al., 1991; Stauffer & Ree, 1996).

Multiple Linear Regression

Among the most predominant single-model predictive methods, multiple regression models have long been used in organizational research to predict a variety of workforce-related

outcomes, such as personnel selection (Gatewood et al., 2011) job validation (Schmidt & Hunter, 1998), and employee engagement (Xu & Thomas, 2011). In their simplest form, linear regression analyses yield closed-form mathematical equations for the linear model that best estimates a continuous target variable from a set of optimally-weighted predictor variables (Darlington & Hayes, 2017). Specifically, in the classical linear model, the observed continuous outcome measure values $\boldsymbol{y} = \{y_1, \ldots, y_n\}^T$ are modelled as a weighted linear combination of the observed predictor variable values $\boldsymbol{x} = \{\boldsymbol{x_1}, \ldots \boldsymbol{x_p}\}$ for each of the *n* observations, arranged as an *n* x *p* predictor variable matrix **X** where each row corresponds to the i^{th} participant or unit, and each column corresponds to a different predictor variable. The set of weighting coefficients for the predictor variables is denoted as the vector $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_p\}^T$, and when combined with the predictor matrix **X** yields an equation for the vector of predicted response variable scores $\hat{\boldsymbol{y}} = \{\hat{y}_1, \ldots \hat{y}_n\}^{\boldsymbol{T}}$.

In matrix notation, the resultant equation for the predicted outcome scores can be written as $\hat{\boldsymbol{y}} = \boldsymbol{X\beta}$, and the error between the actual and predicted target variable scores can further be denoted as $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{X\beta}$. Putting these two components together, the equation for the vector $\boldsymbol{y}$ of the observed outcome scores can thus be written as $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}$ in matrix form. Further, the equations for the predicted and observed target scores for the i^{th} observation can be written respectively as

$$\hat{y}_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_j X_{ij}, \tag{1}$$

and

$$y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_j X_{ij} + e_i \tag{2}$$

Written more succinctly, the predicted outcome measure score for a given individual, conditional upon their scores on the included predictor variables, is of the form:

$$\hat{y}_i = E(Y_i|\boldsymbol{x}_i) = \sum_{1}^{p} x_{ij}\beta_j \tag{3}$$

Lastly, under the normal linear regression model, the corresponding error components are assumed to be independent and normally distributed around a mean of zero with a constant error variance. In combination, the multiple regression equation and its assumptions can be written as

$$y_i = \sum_1^p x_{ij}\beta_j + e_i \text{ , with } e_i \sim N(0, \sigma_e^2),$$

where the assumptions are equivalent to assuming that in the population the values of $\boldsymbol{Y}$ conditional on any $\boldsymbol{X}_i$ are distributed normally with constant variance (Darlington & Hayes, 2017). This last assumption is particularly important for probabilistic predictions, as the probabilistic forecast of any future observation $\tilde{y}_i$, conditional upon its corresponding predictor scores $\tilde{x}_i$, to be derived as

$$p(\tilde{y}_i|\tilde{x}_i) = p(\tilde{y}_i|\hat{y}_i) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(\tilde{y}_i - \hat{y}_i)^2}{2\sigma^2}\right)$$

Lastly, given this error structure, the normal linear regression model is only appropriate for the modelling of continuous outcome measures.

Under the normal linear regression model, the regression coefficients are typically estimated using the *least-squares criterion*, wherein the parameter estimates are chosen such that the resultant predicted outcome measure scores minimize the total squared error. Shown below as

$$\boldsymbol{e^T e} = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - \hat{y})^2 \,, \tag{5}$$

models that utilize this criterion are commonly referred to as *Ordinary Least Squares* (OLS) regression models and are computationally efficient due to their closed-form solution that can be derived through matrix algebra (Clyde & George, 2004).

Linear Regression Variable Inclusion Uncertainty. When uncertainty regarding which predictor variables to include arises in the context of normal linear regression models, letting $\boldsymbol{\gamma}$ index the subsets, each possible model to be considered is of the form

$$\boldsymbol{M_\gamma}: \boldsymbol{y} = \boldsymbol{X_\gamma \beta_\gamma} + \boldsymbol{\varepsilon}\,, \tag{6}$$

where $\boldsymbol{X_\gamma}$ is the predictor matrix whose columns correspond to the $\boldsymbol{\gamma}^{th}$ subset, $\boldsymbol{\beta_\gamma}$ is the vector of non-zero regression coefficients for that subset, and $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n)$.

Logistic Regression

Logistic regression models are an extension of the classic linear model to contexts where the outcome measure is categorical rather than continuous, wherein the weighted set of predictor variables are used to yield probabilistic predictions regarding the classification or outcome occurrence of an observation (McCullagh & Nelder, 1989). In organizational research, logistic regression models have been used for such purposes as assessing occupational fit/orientation (e.g. Leong et al., 2013), predicting employee turnover (e.g. Somers, 1996; Vandenberghe, 1999), and predicting future career advancement (e.g. Chan, 1996). Under the binomial logistic regression model, the expectation of the binary response measure, conditional upon the predictor variable scores, is the probabilistic forecast $\boldsymbol{p}_i$, shown as

$$\mathbb{E}[\boldsymbol{y}|\boldsymbol{X}] = \boldsymbol{p}_i = p(y = 1|\boldsymbol{x}_i) \tag{7}$$

Similar to their normal linear counterparts, the systematic, non-random component of the expected outcome measure, referred to as the linear predictor η_i, is modeled as an optimally weighted linear combination of the included predictor variables, as seen below.

$$\eta_i = \sum_{1}^{p} x_{ij}\beta_j \tag{8}$$

However, unless restrictions are imposed upon these coefficients, this systematic portion of the model yields fitted values over the entirety of the real line $-\infty < \eta < \infty$, which is incongruent with the [0,1] range of probabilistic forecasts (McCullagh & Nelder, 1989). Consequently, the logistic model proceeds by using a logit link function (g) to map the probability of a given observation's class membership or outcome occurrence (p_i) to the real line ($-\infty$,$+\infty$), such that the log odds of occurrence are modeled as the linear predictor instead, as shown below

$$g(p_i) = \eta_i = \ln\left(\frac{p_i}{1-p_i}\right) = \sum_{j=1}^{p} x_{ij}\beta_j \tag{9}$$

After fitting the model, exponentiation of the regression coefficients can be used to yield odds ratios, and the inverse logit function (g^{-1}) can be used to map the log-odds back to the unit interval and derive the desired probabilistic predictions regarding the binary outcome.

$$p(y_i = 1|\boldsymbol{x}_i) = logit^{-1}(\eta_i) = logit^{-1}\left(\sum_{j=1}^{p} x_{ij}\beta_j\right) = \frac{1}{1+e^{-\boldsymbol{\beta X_i}}} \tag{10}$$

Further, the probabilistic forecasts generated by logistic regression models can also be used to directly classify observations regarding class status or event occurrence. In such cases, the binary classification is derived by assigning a threshold value to the probabilistic output (Bishop, 2006).

Parameter estimates for logistic regression models and other non-normal generalized linear models (GLMs) are not available via a deterministic closed-form solution, as is the case under normal linear regression models (McCullagh & Nelder, 1989). Instead, optimal weights for logistic regression models are typically derived through an algorithmic process based on maximum likelihood estimation (MLE). While a detailed review is beyond the scope of this thesis, the parameter estimates under MLE estimation are obtained through a derivative-based

process in which candidate parameter values $\widehat{\boldsymbol{\beta}}^*$ are iteratively considered until the estimate is obtained for which the log likelihood function is maximized. For logistic regression models, the log likelihood is written as

$$LL(\theta) = \sum_{i=1}^{n} \left[y_i * log\left(\frac{1}{1+e^{-\boldsymbol{\beta X_i}}}\right) + (1-y_i) * log\left(\frac{1}{1+e^{-\boldsymbol{\beta X_i}}}\right)\right], \quad (11)$$

and is used by the MLE algorithms such as the *weighted least squares* algorithm (McCullagh & Nelder, 1989) to select the parameter estimates under which the probability of the observed data is maximized.

Logistic Regression Variable Inclusion Uncertainty. When uncertainty regarding predictor variable inclusion arises in the logistic regression context, again letting $\boldsymbol{\gamma}$ index the predictor subsets, each possible model to be considered is of the form

$$\boldsymbol{M_\gamma}: p(\boldsymbol{y} = \mathbf{1}|\boldsymbol{X}) = logit^{-1}\left(\boldsymbol{X_\gamma \beta_\gamma}\right) = \frac{1}{1+e^{-(\boldsymbol{X_\gamma \beta_\gamma})}}, \quad (12)$$

where $\boldsymbol{X_\gamma}$ is the predictor matrix whose columns correspond to the $\boldsymbol{\gamma}^{th}$ subset and $\boldsymbol{\beta_\gamma}$ is the vector of non-zero regression coefficients for that subset.

Model Selection for Linear and Logistic Regression Models

In response to the problem of predictor variable inclusion uncertainty in linear and logistic regression models, a variety of approaches have been proposed. While a comprehensive review is beyond the scope of this thesis, some of the most common approaches have included theory-driven approaches based on construct relevance (as discussed in Putka et al., 2018), stepwise procedures where covariates are sequentially included or removed based on the statistical significance of the coefficient estimates (Steel, 2020), regularization/shrinkage methods that favor parsimony through the inclusion of complexity penalties in the fitting process, and nested model evaluation procedures that examine the effects of predictor inclusion

upon performance metrics such as information criteria. Critically, regardless of how "good" the model ultimately selected is, and regardless of the method used to select said model, all of the single-model approaches described above fail to consider alternative models (Hoeting et al., 1999).

Model Averaging

In contrast to single-model selection-based approaches, model averaging methods aggregate the model-specific predictions generated by *all* models contained within a specified model space. In doing so, model-averaging methods capitalize upon the recognition that the beliefs encoded in each single model are a subset of a richer, fuller understanding of the real-world generative process (Levy & Mislevy, 2016). While model averaging can be applied to virtually any form of model uncertainty, such as that related to theory uncertainty, functional form specification, and fixed/random effects inclusion, uncertainty regarding variable inclusion in the context of predictive modelling is an application particularly well-suited to model-averaging routines. As noted by Steel (2020), model-averaging presents a natural solution to problems where the researcher is ultimately interested in quantities that are themselves not model-specific, such as predictions regarding unknown data. Specifically, in the context of predictive modelling, where the researcher's goal is to generate optimal out-of-sample predictive performance, conditioning upon only a single model is both unnecessary and undesirable. By averaging predictions across all available models in the model space, the researcher avoids the overly confident or potentially biased results that can result from single-model routines.

Machine Learning Ensemble Methods

Despite the wide range of the model-averaging methods documented in the general literature pertaining to predictive modelling, a class of modelling routines referred to as *ensemble*

methods (e.g. Rokach, 2019) have been predominantly featured in organizational research (e.g. Oswald et al, 2020; Putka et al., 2018; Tonidandel et al., 2018; Yuan et al., 2021). At a high level, ensemble models comprise a subset of machine learning methods, which themselves encompass inductive modelling routines wherein computerized algorithms are used to identify and represent patterns in data without being given explicit instructions (Bishop, 2006). Underneath the general umbrella of machine learning, ensemble methods are technically considered a form of model averaging in that they attempt to combine a set of simpler "weak learner" models and their respective individual predictions in order to obtain a final prediction that is more robust and generalizes well to out-of-sample contexts (Rokach, 2019). Amongst the ensemble methods, *random forest* models, which aggregate the individual predictions of many simpler decision trees, are commonly proffered as a solution to model uncertainty in organizational contexts (Putka et al., 2018; Oswald et al., 2020; Tonidandel et al., 2018; Yuan et al., 2021).

Decision Trees. Considered a foundational predecessor of more advanced ensemble-based methods, decision trees sequentially stratify the predictor variable space into a number of regions, each of which corresponds to a single simple decision rule (James et al., 2013). The decision rules used to split the predictor space are selected algorithmically during the model-fit process based on their ability to segment the observations in each layer in a manner that maximizes homogeneity with respect to the outcome measure (Breiman et al., 1984). In contexts where the outcome measure is continuous the decision tree is referred to as a *regression tree*, and homogeneity maximization is evaluated with respect to a reduction in variance. In contexts where the outcome measure is binary, the decision tree is referred to as a *classification tree*, and

homogeneity maximization is typically evaluated using so-called "purity" measures closely related to entropy, such as the Gini index (see Myles et al., 2004).

This partitioning of the predictor space continues until a final segmented region is derived, wherein the predictions regarding the outcome measure are generated (Rokach, 2019). For classification trees, predictions are generated via majority vote, wherein a given observation is assigned the same class status or outcome occurrence as that of the majority of the training observations in the same terminal region. For regression trees, the predicted outcome value is generated by averaging the continuous outcome measure values of the training observations in the same terminal region (Breiman et al., 1994). Importantly, predictions generated by single decision trees tend to generalize poorly to out-sample-performance, which arises from their tendency to identify solutions that are locally optimal (Brieman et al., 1984).

Random Forest Models. Seeking to mitigate such problems, random forest models represent a limited form of model averaging, wherein many simpler decision trees and their respective predictions are combined to generate a single more-robust prediction (Breiman, 1996; 2001). Specifically, as proposed by Breiman (1996, 2001), random forest models involve: (a) repeatedly drawing bootstrapped training data subsets, (b) fitting a single classification or regression tree to each bootstrapped training subset, and then (c) aggregating the predictions generated by each separate tree through either a majority vote (classification) or averaging (regression) procedure (Putka et al., 2018). In addition, in order to mitigate the redundancy of predictions that can arise from correlated constituent trees (Breiman, 2001), random forest models constrain each stratification split and its respective decision rule such that only a random subset of the available predictor variables is considered (Putka et al., 2018). Further, the specific number of predictors that are randomly sampled stays constant and is treated as a hyper-

parameter that can either be specified by the researcher or selected algorithmically via the optimization of a performance metric (James, 2013; Putka et al., 2018). Through this process, the random forest model attempts to increase both the robustness and generalizability of future predictions to out-of-sample contexts (Putka et al., 2018).

Random Forest Models and Predictor Uncertainty. Ostensibly, random forest models partially mitigate uncertainty regarding predictor inclusion by aggregating predictions generated by random subsets of predictor variables. Importantly, however, while ensemble methods such as random forests are nominally considered to be a form of model averaging (e.g., Oswald et al., 2020; Putka et al., 2018; Todinandel et al., 2018), there are several critical distinctions that separate them from "full" model-averaging approaches (Steel, 2020). First, in only considering decision trees composed of a constant number of predictor variables, random forest models ignore all other possible model sizes. Second, the underlying approach of combining multiple base "weak learner" models is inherently different from comprehensively exploring the entirety of a given model space and all possible combinations of predictor variables (Steel, 2020). Third, random forest models grant equal weight to each of the individual predictions generated by the constituent trees and fail to consider the relative probabilities or fits of their respective generative models, in contrast to the weighted-averaging approaches discussed below. In sum, while random forest models do average predictions across multiple models and their respective predictor variable subsets, their focus upon a fixed number of predictor variables and their inclusion of only simpler base models precludes them from being considered an adequate solution to the problem of model uncertainty. In contrast, Bayesian model averaging (BMA) is presented as an alternative that fully addresses the problem of model uncertainty in the context of predictor variable inclusion.

Bayesian Model Averaging

The Bayesian model averaging (BMA) approach consists of a weighted model averaging procedure in which model-specific quantities of interest (i.e., predictions) are averaged across the entirety of a specified model space, in a manner such that the weighting coefficients correspond to posterior probabilities assigned to the models themselves (George & Clyde, 2004; Steel, 2020). As noted by Steel (2020), BMA can thus be thought of as a natural extension to the single-model Bayesian inferential approach, in which parameter uncertainty is addressed by averaging predictions across all possible values of the model-specific posterior distribution. When extended to model-averaging, BMA simply takes this process one step further by also averaging predictions over the range of possible models in a given model space, such that the final predictions thus represent a combination of averaging procedures performed across both the continuous parameter space and the discrete model space (Steel, 2020).

In addition to the conceptual appeal of incorporating the inferential information offered by multiple models rather than relying solely upon a single model, BMA has been shown to have the desirable property of producing optimal predictions in GLM-based contexts under a variety of scoring rules (e.g., Hoeting et al., 1999; Kaplan & Lee, 2018). While applications of BMA have been documented across a number of domains such as economics (e.g., Steel, 2020), political science (e.g., Montgomery & Nyhan, 2010), and educational psychology (e.g., Kaplan & Lee, 2018; Kaplan, 2021), it has been noticeably absent from the organizational literature. Consequently, this thesis seeks to illustrate how IOP and HRM researchers might employ BMA to optimally predict workforce-related outcomes.

Bayesian Modeling

As a precursor to discussing the underlying methodology of Bayesian model averaging, a brief overview of the Bayesian paradigm is discussed below.

Frequentist Inference

Broadly speaking, traditional "frequentist" methods primarily conceive of probability as the expectation of long-run frequencies under repeated sampling (Neyman, 1977). In frequentist models, parameters are considered to be fixed but unknown, and direct probabilistic statements about the parameter values themselves are intractable (Levy & Mislevy, 2016). Instead, the frequentist approach focuses solely upon the conditional probability of the observed data given a particular parameter estimate, $p(y|\theta)$, referred to as the *likelihood*. As noted by Levy and Mislevy (2016), likelihood-based estimation essentially seeks to answer the question, "What are the values of the parameters that yield the highest conditional probability of observing the values of the data that were in fact observed?" (p.25).

Under frequentist estimation routines such as the least-squared-error criterion and maximum likelihood estimation, the true parameter values are considered fixed and the estimators themselves are considered random variables with their own sampling distributions, such that each specific estimate (i.e. the MLE) is a realization derived from a specific sample of data (Levy & Mislevy, 2016). While standard error estimates can be used to construct confidence intervals to capture the uncertainty in a given estimator and its subsequent predictions, frequentist notions of uncertainty only concern the variability of the parameter estimators themselves under repeated sampling of the data (Levy & Mislevy, 2016).

Single-Model Bayesian Inference

In contrast to the frequentist approach, Bayesian inference allows for explicit probabilistic statements about both the unknown parameters θ of a given model and any unobserved data $\tilde{y}$, conditioned upon the observed data *x*, *y*. Referred to as *posterior distributions*, such statements are facilitated by derivations of Bayes Theorem that center around the following functional form:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)} \tag{13}$$

As seen in EQ13, the first step in a single-model Bayesian analysis begins with the specification of the joint distribution for both the observable data and the unknown parameters, and should be elicited in a manner consistent with the researcher's knowledge (or lack thereof) about the underlying data generation and collection process. In doing so, the joint distribution can further be split into both a *likelihood function* $p(y|\theta)$ that models the hypothesized data-generating process, and a *prior distribution* $p(\theta)$ that incorporates the researcher's prior knowledge or beliefs regarding the parameters that govern that generative process. Put simply, this initial step entails two questions: a) How does the researcher want to model the data generating process?; b) How does the researcher want to model their prior knowledge and beliefs about the parameters that govern said process?.

Conceptually, the bulk of the Bayesian data analysis in the single-model context is thus equivalent to the researcher first probabilistically encoding their *a priori* beliefs about the unknown parameter values and then allowing for an updating of those *a priori* beliefs by the observed data via the specification of a sampling distribution. The final term in EQ 13 represents the joint probability of the observed data and the parameter values integrated over the entire

range of *all* parameter values possible under the given model and is referred to as the *marginal likelihood* $p(y)$.

$$p(y) = \int p(\theta, y)d\theta = \int p(y|\theta)p(\theta)d\theta \quad (14)$$

In single-model Bayesian inference, the marginal likelihood is treated as a normalization constant that simply ensures that the model-specific posterior distribution is a proper probability distribution. Consequently, given that all dependency upon θ has been integrated out, in model-specific Bayesian analyses the marginal likelihood is typically replaced by the simpler, un-normalized version of the posterior, written as

$$p(\theta|y)\ \alpha\ p(y, \theta) = \ p(y|\theta)p(\theta) \quad (15)$$

This alternate version of the posterior has proven essential to fostering many of the advances in single-model Bayesian inference, as the integral required to compute $p(y)$ is typically multi-dimensional and highly intractable (Hoeting et al., 1999). The importance of the unnormalized posterior in facilitating Bayesian computation can be seen when evaluating the relative posterior probabilities of any two model-specific parameter values under consideration, wherein the posterior odds ratio of the two candidate values can be expressed as:

$$\frac{p(\theta^*|y)}{p(\theta|y)} = \frac{\frac{p(y|\theta^*)p(\theta^*)}{p(y)}}{\frac{p(y|\theta)p(\theta)}{p(y)}} = \frac{p(y|\theta^*)p(\theta^*)}{p(y|\theta)p(\theta)} = \frac{p(\theta^*)}{p(\theta)}\ x\ \frac{p(y|\theta^*)}{p(y|\theta)} \quad (16)$$

As seen in EQ16, the posterior evaluation of any two proposed parameter values under a given model thus reduces to a comparison of their relative prior probabilities and likelihoods. This has proved critical to computational advances made in Bayesian inference, as discussed further below.

Model-Specific Bayesian Prediction. After deriving the model-specific posterior distribution for the parameter values $p(\boldsymbol{\theta}|y)$, the researcher can then predict an unknown observable $\tilde{y}$ (i.e. a future data value), conditioned upon the future observed explanatory values $\tilde{x}$ and the observed data *y*, in a manner that incorporates all of the information and uncertainty about the generative parameter values by integrating the prediction over the entire model-specific posterior distribution (Steel, 2020). Conceptually, this is equivalent to a weighted average of the predictions associated with each possible specific parameter value across the range of the entire posterior distribution, with weights corresponding to the posterior probabilities of the generative parameter values. This process, referred to as the *posterior predictive distribution* for the unknown observables $\tilde{y}$ under a given model, can be written in the single-model context as follows:

$$p(\tilde{y}|\tilde{x}, y) = \int p(\tilde{y}|\tilde{x}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|y) d\boldsymbol{\theta} \tag{17}$$

Simply put, the model-specific posterior predictive distribution thus describes beliefs about the future observation given the observed data and the specified model.

Multi-Model Bayesian Inference

Crucially, while the single-model Bayesian derivations shown thus far have all required an implicit conditioning upon a given model, the same principles used to account for model-specific parameter uncertainty can intuitively be extended to the multi-model context of variable selection uncertainty. Specifically, letting $\boldsymbol{\gamma}$ index the predictor variable subsets, the model-specific Bayesian posterior distribution shown in EQ 13 can be rewritten as

$$p(\boldsymbol{\theta}_{\boldsymbol{\gamma}}|y, M_{\boldsymbol{\gamma}}) = \frac{p(\boldsymbol{\theta}_{\boldsymbol{\gamma}}, y|M_{\boldsymbol{\gamma}})}{p(y|M_{\boldsymbol{\gamma}})} = \frac{p(y|\boldsymbol{\theta}_{\boldsymbol{\gamma}}, M_{\boldsymbol{\gamma}}) p(\boldsymbol{\theta}_{\boldsymbol{\gamma}}|M_{\boldsymbol{\gamma}})}{p(y|M_{\boldsymbol{\gamma}})} \tag{18}$$

Importantly, through this extension to the multi-model context, Bayesian inference can be used to address model uncertainty by way of a procedure known as *Bayesian Model Averaging*.

Bayesian Model Averaging

Following Steel's (2020) definition, Bayesian model averaging (BMA) consists of a weighted model averaging procedure in which the model-specific parameter values are integrated out, where the weights used for averaging the quantity of interest correspond to the posterior probabilities of the models themselves. As further noted by Steel (2020), Bayesian model averaging can thus be thought of as a natural extension of the Bayesian tendency to address uncertainty through averaging, wherein the resultant posterior distributions are averaged over both the continuous parameter space and the discrete model space. Subsequently, the final BMA posterior distribution thus equates to a weighted sum of the model-specific posterior distributions under each of the $\{M_1, \dots, M_k\}$ models considered. This can be seen below, where, for a given quantity of interest (i.e. parameter estimates or future predictions) Δ, the BMA posterior distribution can be written as

$$p_{BMA}(\Delta|y) = \sum_{k=1}^{K} p(\Delta|y, M_{\boldsymbol{\gamma}}) p(M_{\boldsymbol{\gamma}}|y) \tag{18}$$

The above equation highlights that the final model-averaged posterior distribution in a BMA routine incorporates all of the model-specific probabilistic forecasts $p(\Delta|y, M_{\boldsymbol{\gamma}})$ by weighting each in accordance with the posterior probability of the model $p(M_{\boldsymbol{\gamma}}|y)$ that generated it.

Bayesian Model Averaging Predictive Distribution

Building upon the general BMA approach to calculating quantities of interest described in EQ18, the model-specific posterior predictive distribution seen earlier in EQ17 can thus be expanded to produce the *Bayesian model averaging predictive distribution*. Denoting the scores

on the future observed predictor variables and the unknown outcome measure as $(\tilde{x}, \tilde{y})$, respectively, the BMA predictive distribution is expressed as:

$$p_{BMA}(\tilde{y}|\tilde{x}, y) = \sum_{k=1}^{K} \left[\int p(\tilde{y}|\tilde{x}, y, \theta_{\gamma}, M_{\gamma}) p(\theta_{\gamma}|y, M_{\gamma}) d\theta_{\gamma} \right] p(M_{\gamma}|y), \quad (19)$$

which reduces to

$$p_{BMA}(\tilde{y}|\tilde{x}, y) = \sum_{k=1}^{K} \left[\int p(\tilde{y}, \theta_{\gamma}|\tilde{x}, y, M_{\gamma}) d\theta_{\gamma} \right] p(M_{\gamma}|y), \quad (20)$$

and finally yields:

$$p_{BMA}(\tilde{y}|\tilde{x}, y) = \sum_{k=1}^{K} p(\tilde{y}|\tilde{x}, y, M_{\gamma}) p(M_{\gamma}|y) \quad (21)$$

Thus, as seen above, under a BMA predictive routine, the predictions are averaged over both the continuous range of model-specific parameter values and the discrete model space. Finally, and of particular relevance to applied settings, the BMA routine can easily be adapted to yield point predictions that are optimal under the log-score criterion (Hoeting et al., 1999). As noted by Clyde and George (2004), the optimal such estimate is derived simply via a weighted average of the model-specific posterior means:

$$\mathbb{E}_{BMA}[\tilde{y}|\tilde{x}, y] = \sum_{k=1}^{K} \left[\mathbb{E}(\tilde{y}|\tilde{x}, y, M_k) \times p(M_{\gamma}|y) \right] \quad (22)$$

In using a predictive point estimate derived via the equation shown above, the researcher can thus obtain a prediction that is both optimal under the log score rule and simple to compute, but still adequately incorporates both parameter and model uncertainty.

Posterior Model Probabilities

The critical component of a BMA routine is the set of posterior model probabilities (PMP) used to weight the model-specific inferences and predictions. The PMPs are derived by integrating out the model-specific parameter values and can be written as

$$p(M_{\boldsymbol{\gamma}}|y) = \frac{p(y|M_{\boldsymbol{\gamma}})p(M_{\boldsymbol{\gamma}})}{\sum_{k=1}^{K} p(y|M_{\boldsymbol{\gamma}})p(M_{\boldsymbol{\gamma}})}\,, \tag{23}$$

where the numerator is comprised of the marginal likelihood $p(y|M_{\boldsymbol{\gamma}})$ and the prior probability $p(M_{\boldsymbol{\gamma}})$ of the model $M_{\boldsymbol{\gamma}}$ under consideration, and the denominator represents a normalization constant analogous to that in the single-model context. As in the single-model context, the normalized posterior model distribution is typically discarded in favor of the unnormalized version, written as

$$p(M_{\boldsymbol{\gamma}}|y) \propto p(y|M_{\boldsymbol{\gamma}})p(M_{\boldsymbol{\gamma}}) \tag{24}$$

Implementation

Conceptually, implementation of a BMA routine involves two main components, namely the specification and exploration of the model space, and the calculation or approximation of the posterior model probabilities that serve as weighting coefficients. When the pool of candidate predictor variables is large, as is often the case in modern organizational research applications (Oswald et al., 2020; Putka et al., 2018), the exponential growth of the model space formed by the potential combinations of predictor variables quickly precludes an exhaustive enumeration of the model space and deterministic calculations of the posterior model probabilities (Hoeting et al., 1999). Consequently, aside from contexts where the pool of potential predictor variables is relatively small (e.g. $k < 15$), modern BMA routines typically utilize methods that seek to explore the model space and compute posterior model probabilities simultaneously.

Exploration of the Model Space

Construction/specification of the model space to be considered should be closely aligned with the exact type of model uncertainty the researcher is attempting to address (Steel, 2020). When implementing a BMA routine to address uncertainty regarding the inclusion of predictor variables $\{\boldsymbol{X} = \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$, the model space is typically specified as the set of all 2^k possible predictor variable subsets that can be formed. Given the exponentially large model spaces that can arise when the pool of available predictors is large, as is often the case in larger organizational settings (Oswald, 2020), an exhaustive enumeration of the entire model space often quickly becomes infeasible (Clyde & George, 2004; Hoeting et al., 1999). As such, exploration of the model space has been a primary barrier that has directly shaped the historical development of BMA implementation.

Reduced-Subsets Approach to Model Space Exploration

The first approach to the problem of model space exploration has been to average over a reduced subset of models best supported by the data, rather than over the entirety of the model space (Hoeting et al., 1999). One method for this is *Occam's Window*, originally proposed by Madigan and Raftery (1994) for the identification of a data-supported subset of graphical models, and later extended to regression contexts by Raftery and colleagues (1997).

Occam's Window. In regression contexts, the Occam's Window method consists of two principle tenets. First, if a model predicts the data far less well than the model with the strongest predictive performance, the poorer-performing model should be discarded. Under this first tenet, the researcher specifies *a priori* the width of a constraint, *C*, that sets the threshold ratio of predictive performance. Thus, models not belonging to the subset *A'*, in accordance with the criteria shown below, are excluded from consideration.

$$A' = \left\{ M_k : \frac{max_l\{p(M_l|y)\}}{p(M_k|y)} \right\} \leq C \tag{25}$$

Under the second tenet of Occam's Window, if a simpler model exists that produces better predictions than a more complex model, the more complex model is discarded. Formally, letting *B* denote the subset of models to be excluded accordingly, this second tenet is shown as follows:

$$B = \left\{ M_k : \exists M_l \in A', M_l \subset M_k, \frac{p(M_l|y)}{p(M_k|y)} > 1 \right\} \tag{26}$$

Put simply, the above equation states that this secondary subset of models to be discarded, *B*, is to contain any model M_k for which there exists a simpler model M_l that generates better predictive performance. In combination, the remaining subset of models to be considered constitutes the final reduced model space *A*, over which any quantity of interest is to be averaged during the BMA routine as follows:

$$p(\Delta|y, A) = \sum_{M_\gamma \in A} p\left(\Delta\middle|y, M_\gamma\right) p\left(M_\gamma\middle|y\right) \tag{27}$$

Importantly, reduced-subsets methods require not only a method for identifying models to be discarded, but also a method for searching the remaining subspace and identifying the constituent models. The most common search strategies employ derivations of an algorithmic approach originally introduced by Raftery et al. (1997), which utilized the "leaps and bounds" algorithm developed by Furnival and Wilson (1974) to rapidly explore the reduced model space. While increases in computational power have largely eliminated the need for reduced-subsets approaches, various algorithmic approaches have also been proposed in recent years as well, such as the *Bayesian Adaptive Sampling* method developed by Clyde and colleagues (2011), and the *Evolutionary Stochastic Search* by Bottolo and Richardson (2010).

MCMC Sampling. Exploration of large model spaces in modern BMA routines is now almost exclusively conducted via the use of *Markov chain Monte Carlo* (MCMC) methods, which approximate the posterior model distribution by sampling models from the model space with a frequency proportional to their posterior model probabilities. Crucially, MCMC samplers have thus greatly reduced the computational complexity that previously served as a barrier to large-scale BMA implementation, as model space exploration and PMP approximation are addressed simultaneously during the sampling routine. While a variety of samplers have been proposed, most are derivations of the MCMC Model Composition (MC^3) algorithm, a random-walk Metropolis sampler developed by Madigan and York (1995).

Letting $\boldsymbol{M}$ denote the model space under consideration, the samplers proceed by constructing a Markov chain $\{M(t)\}, t = 1,2, \dots N$ with state space $\boldsymbol{M}$ and equilibrium distribution equal to the posterior model probability $p(M_i|y)$ in order to simulate from the Markov chain and obtain observations $M(1), \dots M(N)$. Then, for any function $g(M_i)$ defined on $\boldsymbol{M}$, the average,

$$\hat{G} = \frac{1}{N}\sum_{t=1}^{N} g\big(M(t)\big), \tag{28}$$

is an estimate of $E\big(g(M)\big)$. The application of standard MCMC results yields

$$\hat{G} \to \boldsymbol{E}\big(g(M)\big)\ as\ N \to \infty, \tag{29}$$

which simply states that the expectation of a given quantity can be estimated through sampling by averaging over the individual simulations, and that those estimations will converge to the true expectation as sample size increases asymptotically. Applied to predictive contexts, the above equation can be used to derive model-averaged predictions under a sampling-based BMA routine

by setting $g(M) = p(\tilde{y}|\tilde{x}, y, M)$, such that the expectation represents the BMA predictive estimate over all models.

When constructing the Markov chain over the model space for the Metropolis-based samplers typically used in BMA routines, a neighborhood and a transition matrix are both specified in order to guide the manner in which a candidate model is proposed and either accepted or rejected. Specifically, at any given iteration the sampler stands at the current model M_i, and in order to select a model M_{i+1} for the next step in the chain, a candidate model $M^*_{i+1}$is proposed from the current model's neighborhood. In regression contexts, the neighborhood for each predictive model is typically comprised of the set of models that include one greater or one fewer predictor variables than the current model. After a candidate model is randomly drawn from the proposal neighborhood, the sampler will either accept the candidate model with probability

$$min\left\{1, \frac{p(M^*|y)}{p(M|y)}\right\}$$

or reject the candidate model in favor of the current model. As seen in the equation above, the sampler will automatically accept any candidate model with a higher posterior probability than the current model, but will also still accept less probable models with a frequency proportional to the ratio of their relative posterior probabilities. As the sampling frequencies converge asymptotically to the posterior model distribution, empirical approximations of the posterior model probabilities can easily be derived through the sampling routine.

Marginal Likelihood. The first component of the posterior model probability is the marginal likelihood of a given model M_k, and represents the conditional probability of the observed data *y*, evaluated at and integrated across all possible values of the parameters specified by the model. Conceptually, as can be seen in the equation below, the marginal likelihood

equates to a weighted average of the probability of the observed data under each possible set of parameter values $p(y|\boldsymbol{\theta}_{\boldsymbol{\gamma}}, M_{\boldsymbol{\gamma}})$, with the weights corresponding to the prior probability distribution over the model parameters $p(\boldsymbol{\theta}_{\boldsymbol{\gamma}}|M_{\boldsymbol{\gamma}})$ specified by the researcher.

$$p(y|M_{\boldsymbol{\gamma}}) = \int p(y|\boldsymbol{\theta}_{\boldsymbol{\gamma}}, M_{\boldsymbol{\gamma}}) p(\boldsymbol{\theta}_{\boldsymbol{\gamma}}|M_{\boldsymbol{\gamma}}) d\boldsymbol{\theta}_{\boldsymbol{\gamma}} \tag{32}$$

Critically, computation of the marginal likelihood often entails the calculation of a complex multi-dimensional integral that is very sensitive to both the researcher's choice of model parameter prior (Berger et al., 2001; Kass & Raftery, 1995; Steel, 2020) and the functional form of the model-specific likelihood function. Further, while the marginal likelihood serves only as a normalization constant within single-model Bayesian posterior calculations, this is not the case in BMA routines, where sensitivity of a model's marginal likelihood has a direct impact upon its subsequent posterior probability (Steel, 2020). Consequently, BMA implementations have largely been constrained to a narrow range of specific prior structures that yield analytical closed-form solutions or approximations to the marginal likelihood.

Applying a Bayesian Model Averaging Routine

In practice, the application of a BMA routine is largely similar across all predictive contexts, consisting of a sampling-based exploration of the model space and empirical approximation of posterior model probabilities. Implementing such a routine requires that the researcher specify the model space, the functional form of the constituent models, and a set of priors to be placed upon both the parameter space and the model space (Steel, 2020). In the context of model uncertainty pertaining to predictor variable inclusion, the model space is typically considered to be comprised of all subsets of predictors possible under the full model, such that all of the constituent models share the same functional form. When these constraints

are applicable, the researcher's main concerns are those related to specification of the prior structures.

Prior Over Models

The prior model probability $p(M_{\boldsymbol{\gamma}})$, is the prior probability assigned by the researcher that a given model is the "true" actual data-generating model (Steel, 2020). While authors have debated the theoretical appropriateness of probabilistic statements regarding "true" generative models (e.g., Clyde & Iverson, 2013), such concerns are typically assuaged by instead treating model space priors as being reflective of the prior probability that a given model is a useful proxy for the "true" model (e.g. Steel, 2020). Theoretical considerations aside, in model uncertainty contexts related to variable inclusion, model space priors are typically derived simply by specifying an inclusion probability for the potential predictors. (Steel, 2020). Rather than elicit custom inclusion probabilities separately for each available predictor, most implementations in the literature have relied upon variations of the Bernoulli prior (e.g., George & McCulloch, 1993, 1997; Raftery et al., 1997), wherein each potential regressor is independently included in the model with probability ω (Ley & Steel, 2012). Written generally as

$$p\left(M_{\gamma}\middle|\omega\right) = \omega^{k_{\gamma}}(1-\omega)^{k-k_{\gamma}}\,, \tag{33}$$

the Bernoulli-based model priors require only the specification of a single hyperparameter $\omega \in (0,1)$ (Clyde & George, 2004), and are structurally equivalent to placing a prior distribution over the range of model sizes (Clyde & George, 2004; Steel, 2020). Consequently, the above equation implies that prior selections of $\omega > 0.5$ will favor larger models, $\omega < 0.5$ will favor smaller models, and $\omega = 0.5$ assigns equal probability $\frac{1}{k_{\gamma}}$ to all models (Steel, 2020).

Given that the choice of ω can thus have a substantial effect upon the results of a BMA routine, the current literature standard is to place a Beta-distributed hyperprior upon ω, which precludes researcher misspecification by allowing the posterior distribution of ω to vary in accordance with the data (e.g., Brown et al., 1998; Clyde & George, 2004; Ley and Steel, 2012; Liang et al., 2008; Scott & Berger, 2010). The resulting hierarchical prior over the model space is less restrictive than fixed-term model priors (Ley & Steel, 2009) and only requires the researcher to specify a prior expected model size m, the selection of which has been shown to have a limited impact in practice due to the adaptiveness of the prior (Ley & Steel, 2009; Steel, 2020). Formalized through the relationship

$$\omega \sim Beta\left(1, \frac{k_\gamma - m}{m}\right), \tag{34}$$

the random-ω Beta-Binomial model-space prior is thus desirable due to both its simplicity and its robustness to researcher misspecification (Ley & Steel, 2009; Steel, 2020).

Priors Over Parameters

In regression contexts where the BMA routine is being employed to address model uncertainty pertaining to predictor variable inclusion, subjective prior distributions typically become infeasible due to the exponentially large model spaces that can arise (Berger et al., 2001; Clyde & George, 2004; Steel, 2020). Consequently, research thus far has predominantly focused upon variations of a common class of prior structures developed by Zellner (1983, 1986) referred to as *g*-priors, which yield computationally simple solutions or approximations to the marginal likelihoods and induce a regularization penalty in alignment with the researcher's prior beliefs (Ley & Steel, 2012).

Importantly, the exact parameter prior structure recommended depends greatly upon the specific functional form of the generative model specified in the likelihood function. As such, for

ease of discussion, the review of current best practices regarding prior structures that follows has been divided into two contexts: those where the model space under consideration is comprised of normal linear regression models, and those that rely upon the broader generalized linear model (GLM), such as logistic regression models

Normal Linear Regression BMA

Many of the fundamental developments in the use of BMA routines to address model uncertainty related to predictor variable inclusion have occurred in the context of the normal linear regression model (e.g. George & McCulloch, 1993; Raftery et al., 1997; Zellner, 1983; Zellner & Siow, 1980), a trend due in large part to the closed-form integrals for the marginal likelihood that arise when certain prior structures are placed upon the parameter space of normal linear models (Hoeting et al., 1999; Raftery et al., 1997). Specifically, for normal linear regression contexts, a set of *natural conjugate priors* exists, referred to as such because they share same functional form as the likelihood (i.e., the normal distribution) and yield a posterior distribution that is *also* of the same form (Gelman et al., 2013). Further, natural conjugate priors also yield exact marginal likelihood integrals, a property that greatly facilitates fast and efficient sampling over the model space in BMA routines even the model space is very large (George & Clyde, 2004).

Consequently, a popular approach has been to follow the common model-specific prior structure proposed by Fernandez and colleagues (2001), under which improper non-informative prior structures are placed upon the intercept and error variance, and variations of the weakly-informative natural-conjugate *g*-prior (Zellner, 1983; 1986) are placed upon the regression coefficients. Developed first for normal linear regression models, the *g*-prior framework places a multivariate-normal distribution upon the regression coefficients, with a prior mean of zero and a

covariance matrix equivalent to a scaled version of the covariance matrix of the maximum likelihood estimator (Clyde & George, 2004; Zellner, 1983; 1986). Again letting γ index the predictor variable subsets included in each model M_γ, placement of a g-prior upon the regression coefficients $\boldsymbol{\beta}_\gamma$ can be written as

$$p(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|g, M_\gamma) \sim N\left(\mathbf{0}, \sigma^2 g\left(\boldsymbol{X}_{\boldsymbol{\gamma}}'\boldsymbol{X}_{\boldsymbol{\gamma}}\right)^{-1}\right), \tag{35}$$

which simplifies elicitation of the prior structure to specification of the hyper-parameter g. Thus, in selecting g, the researcher is effectively encoding how conservative his or her prior beliefs are regarding $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$, with a small g signifying the researcher is quite certain the coefficients are equal to zero. Importantly, under the g-prior shown above, the marginal posterior distribution for the regression coefficients follows a k_γ-variate Student's-t distribution (Fernandez et al., 2001; Steel, 2020) with an expected value of

$$E(\boldsymbol{\beta}_\gamma|y, \boldsymbol{X}_\gamma, g, M_\gamma) = \tfrac{g}{1+g}\widehat{\boldsymbol{\beta}}_\gamma^*, \tag{36}$$

where $\widehat{\boldsymbol{\beta}}_\gamma^*$ is equivalent to the OLS estimate and $\frac{g}{1+g}$ serves as a regularization factor (Liang et al., 2008). Subsequently, the posterior expectation under the g-prior represents a compromise between the OLS estimate and the prior mean of zero, with the relative influence of the prior determined by the selection of g (Fernandez et al., 2001). Similarly, the out-of-sample predictive distribution for each regression model, given the predictor values for the future observations, also follows a Student-t distribution (Fernandez et al, 2001; Steel, 2020). Lastly, the integral for marginal likelihood for any normal linear regression model M_γ simplifies under the g-prior to the tractable form

$$p(y|\boldsymbol{X}_\gamma, g, M_\gamma) \propto (1+g)^{-\frac{k_\gamma}{2}}\left(1 - \frac{g}{1+g}R_\gamma^2\right)^{-\frac{N-1}{2}}, \tag{37}$$

which includes a model-complexity penalty and fosters a computational efficiency that greatly facilitates posterior sampling in a BMA routine (Steel, 2020).

Selection of *g*. Much research has been done regarding the relative efficacy of various selections for *g* (e.g. Fernandez et al, 2001; George and Foster, 2000; Hansen and Yu, 2003). Under uniform prior model probabilities, the choice of *g* drastically impacts the posterior model distribution, with large *g* selections tending to result in parsimonious models with a few large coefficients, and small *g* selections tending to result in a larger number of saturated models with small coefficients (George & Foster, 2000; Liang et al., 2008). While earlier research focused upon fixed-*g* specifications (e.g., Fernandez et al., 2001; Foster & George, 1994; Kass & Wasserman, 1995), the sensitivity of the BMA routine to *g* misspecifications has led researchers to recommend various mixtures of *g*-priors (e.g., Clyde & George, 2000; George & Foster, 2000; Liang et al., 2008), wherein the marginal likelihoods and posterior distributions are integrated over both the parameter values and *g* values possible under each model. Written conceptually as

$$p(y|M_\gamma) = \int p(y|\boldsymbol{\theta}_\gamma, g, M_y)p(\boldsymbol{\theta}_\gamma|g, M_y)p(g|M_\gamma)dg\,, \tag{26}$$

mixtures of *g*-priors for linear regression models include variations such as the *Cauchy prior* (Zellner & Snow, 1980) and the *hyper-g* and *hyper-g*/*n* priors (Cui & George, 2008; Liang et al., 2008).

Of the above methods, the *hyper-g* prior of Liang and colleagues (2008) has been particularly popular, under which a Beta-distributed hyperprior is placed upon the shrinkage factor $\frac{g}{1+g}$ (e.g. Liang et al., 2008; Zellner & Siow, 1980). Written as

$$\frac{g}{1+g} \sim Beta\left(1, \frac{a}{2} - 1\right)\,, \tag{38}$$

the hyper-g prior only requires specification of values for the hyperparameter $a \in \{2,4\}$, wherein values closer to 2 favor parsimonious models with large coefficients, and values closer to 4 favor a larger number of saturated models with small coefficients (Ley & Steel, 2012; Liang et al., 2008). In the absence of strong prior beliefs held by the researcher, both Liang and colleagues (2008) and Ley and Steel (2011) recommend selecting a value of $a = 3$, which predominantly favors shrinkage factors close to 1 and thus yields posterior estimates that converge to the OLS estimates (Liang et al., 2008). Importantly, in requiring only the selection of a single scalar value *a*, and by allowing the selection of *g* to vary via a data-dependent process, concerns regarding the impact of potential researcher misspecification of prior beliefs are thus largely mitigated (Ley & Steel, 2012; Liang et al., 2008).

Finally, under the random-g priors, the optimal BMA predictions under the squared-error loss criterion for a vector of future values $\widehat{\tilde{\boldsymbol{y}}}$, given the associated predictor scores $\tilde{\boldsymbol{X}}$, can be written as

$$\mathbb{E}[\tilde{\boldsymbol{y}}|\boldsymbol{y}] = \sum_{k=1}^{K} \mathbb{E}\left[\frac{g}{1+g}|M_\gamma, \boldsymbol{y}\right] \tilde{\boldsymbol{X}}_\gamma \widehat{\boldsymbol{\beta}}_\gamma \times p(M_\gamma|\boldsymbol{y}), \tag{39}$$

where $\mathbb{E}\left[\left(\frac{g}{1+g}\right)|M_\gamma, \boldsymbol{y}\right]$ represents the posterior expected value of the shrinkage factor, and $\widehat{\boldsymbol{\beta}}_\gamma$ represents the posterior mean for the regression coefficients under model M_γ. In combination, use of the random-g priors under a BMA routine thus results in predictive point estimates that represent a hierarchical weighted average of the model-specific predictions, where the first set of weights corresponds to a data-dependent regularization term expected under model M_γ, and the second set of weights corresponds to the posterior model probability of M_γ.

Logistic Regression/GLM Priors Over Parameters

Eliciting a prior distribution over the parameter space for logistic regression models and other GLM derivations can quickly become complex, as the closed-form expressions for the marginal likelihood integrals possible under normal linear regression models are typically unavailable (Li & Clyde, 2018). Consequently, numerous researchers have proposed variations or mixtures of the original *g*-priors discussed earlier as a means of eliciting objective prior structures that yield computationally-efficient integral approximations for non-normal linear models, including Hansen and Yu (2003), Kass and Wasserman (1995), and Li and Clyde (2018), among others. Of particular promise is the research by Li and Clyde (2018), who have shown that by employing large-sample Laplace approximations to the marginal likelihood integrals (Tierney et al., 1989; Tierney & Kadane, 1986), the usage of specific *g*-priors can lead to analytic marginal likelihoods and conditionally conjugate posterior distributions for non-normal GLMs.

Although a detailed review of Laplace approximations is beyond the scope of this thesis, at a high level the Laplace method provides a means of integral approximation based upon the volume of a Gaussian distribution centered over the posterior mode, which asymptotically converges to the maximum likelihood estimate (MLE). Given that the joint posterior distribution converges to normality asymptotically (Gelman et al., 2013), Laplace approximations to marginal likelihood integration can be appropriately applied to the BMA routines in large-sample contexts (Li & Clyde, 2018; Tierney & Kadane, 1986). Consequently, under a modelling routine that employs such approximations, Li and Clyde (2018) advocate for the use of a flexible generalized Beta distribution referred to as the truncated compound confluent hypergeometric (tCCH) distribution (Gordy 1998), of which mixtures of *g*-priors such as the *hyper-g* prior of

Liang and colleagues (2018) are included as special cases. As with the hyper-g prior, the tCCH prior is an adaptive prior under which the values of g are allowed to vary and which requires only the specification of a single scalar hyperparameter by the researcher (Li & Clyde, 2018).

In addition to specification simplicity, the tCCH prior is also desirable due to the computational efficiency it yields with respect to both the model-specific posterior distribution and the integrated marginal likelihood. Specifically, under the tCCH prior, both the posterior distribution of the regression coefficients and integrated marginal likelihood reduce to functions of the MLE estimates. Again letting γ index the predictor variable subsets included in each model M_γ, the posterior distribution for the regression parameters and the marginal likelihood can be written respectively as

$$p(\boldsymbol{\beta}_\gamma | y, M_\gamma) \overset{N}{\rightarrow} N\left(\frac{g}{1+g}\widehat{\boldsymbol{\beta}}^*_\gamma, \frac{g}{1+g}\mathcal{J}_n\left(\widehat{\boldsymbol{\beta}}^*_\gamma\right)^{-1}\right), \tag{40}$$

and

$$p(y|M_\gamma) \approx p(y|\widehat{\boldsymbol{\theta}}^*_\gamma)p(\widehat{\boldsymbol{\theta}}^*_\gamma)(2\pi)^{\frac{k_\gamma}{2}}|\Sigma|^{\frac{1}{2}}N^{-\frac{k_\gamma}{2}}, \tag{41}$$

where $\widehat{\boldsymbol{\beta}}^*_\gamma$ and $\widehat{\boldsymbol{\theta}}^*_\gamma$ denote the MLE estimates for the regression parameters, $\mathcal{J}_n\left(\widehat{\boldsymbol{\beta}}^*_\gamma\right)^{-1}$ represents the inverse of the observed Fisher information matrix evaluated at the MLE estimate, and Σ represents the Hessian matrix of second derivatives. Thus, as shown above, under the tCCH prior, closed form solutions dependent solely upon the MLE estimate can be derived for both the posterior distribution and the marginal likelihood, which greatly facilitates fast and efficient sampling from the posterior model distribution (Li & Clyde, 2018).

Evaluating Predictive Performance

In organizational research, as in all domains, the evaluation of a given model's predictive performance is often of critical importance. Intuitively, when a predictive model is used to

generate forecasts for unknown or future observations, a discrepancy between the predictions generated by a model and the actual observations suggest that that model is of limited utility (Vehtari & Ojanen, 2012). Consequently, decisions regarding the selection of a performance metric that can be used to quantify such a discrepancy are often of great importance to the researcher. As noted by Gelman and colleagues (2013), ideally such a measure should be tailored specifically for the application at hand and be adept at measuring both the costs and benefits associated with the predictions generated by a given model.

In organizational contexts, these costs and benefits are sometimes tangible and calculable, such as in the context of personnel selection models that can be associated with a specific predictive validity, adverse impact, and economic impact Schmidt & Hunter, 1998). However, when the costs and benefits associated with employing a predictive model are implicit or intractable, generic scoring functions and rules are needed (Gelman et al., 2013). In this regard, a distinction should first be made between *point predictions*, wherein a single value representing the unknown future observation $\tilde{y}$ is reported, and *probabilistic prediction*, where the inferences about $\tilde{y}$ are reported in a manner such that the full uncertainty over $\tilde{y}$ is acknowledged (Vehtari & Ojanen, 2012).

As discussed by Bernardo and Smith (1994), a modeling strategy is generally preferred over another when it consistently assigns higher probabilities to the events that actually occur. To this end, scoring rules provide measures of predictive accuracy for probabilistic prediction (Gelman, 2013), and should assess both its calibration (whether the assigned probabilities are compatible with the proportion of times the outcome occurred) and its sharpness (the concentration of the predictive distributions). Further, a scoring rule should be proper, such that the score is maximized when the forecasted probability of an event's occurrence is the same as

the true probability of the event's occurrence (Gelman, 2013). While there are numerous proper scoring rules cited in the literature, variations of the logarithm predictive score and the squared-error function are among the most popular.

Log Predictive Score

Perhaps the most common proper scoring rule is the logarithm predictive score (LPS) of Good (1952), also referred to as the "log-likelihood", or the "log predictive density", which seeks to measure the predictive ability of a given model by summing the logarithms of the predictive densities of each observation. Introduced by Good (1952), and implemented in the BMA context by Fernandez et al. (2001a, b), Kaplan and Lee (2018), Ley and Steel (2009b), and Madigan et al. (1995), the LPS for a given model M_γ can be written as

$$LPS_\gamma = -\sum_{i=1}^{n} log\left[p\left(\tilde{y}_i|\tilde{x}_i, y, M_\gamma\right)\right], \tag{42}$$

which can be written in the Bayesian context as

$$LPS_\gamma = -\sum_{i=1}^{N} log\left(\int p\left(\tilde{y}_i|\boldsymbol{\theta}_\gamma, \tilde{x}_i, M_\gamma\right)p(\boldsymbol{\theta}_\gamma|y, M_\gamma)d\boldsymbol{\theta}_\gamma\right), \tag{43}$$

where $\tilde{x}_i$ and $\tilde{y}_i$ denote the predictor and response scores, respectively, for the i^{th} future data point and $\boldsymbol{\theta}_\gamma$ represents the vector of parameter estimates under model M_γ. Further, when applied to a BMA predictive modelling routine, the LPS can be written as

$$LPS_{BMA} = -\sum_{i=1}^{N} log\left(\sum_{k=1}^{K} p\left(\tilde{y}_i|\tilde{x}_i, y, M_\gamma\right)p\left(M_\gamma|y\right)\right), \tag{44}$$

and thus represents the summation of the logarithms of the Bayesian model-averaged pointwise predictive densities of the unknown observations (Gelman et al., 2013).

Squared-Error Loss

The squared-error loss function is one of the simplest and most common predictive performance metrics (Gelman et al., 2013; Gneiting, 2011; Vehtari & Ojanen, 2012), and will yield equivalent results to the LPS when assumptions of normality and constant error variance are met (Gelman, 2013). Amongst the many variations of squared-error loss functions, two of the most common derivations are the root mean squared error (RMSE) function and the Brier score, intended for continuous and binary outcome measures respectively.

Root Mean Squared Error. The accuracy of the predictions generated for previously unseen data by a given model can be evaluated by way of the RMSE function. Written as follows,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\tilde{y}_i - \widehat{\tilde{y}}_i\right)^2}, \tag{45}$$

the RMSE statistic involves calculating the average pointwise discrepancy between the actual outcome values for previously unseen data $\tilde{y}_i$ and the predicted values $\widehat{\tilde{y}}_i$. The RMSE function has the desirable properties of simplicity and ease of computation, but it may not be appropriate in situations where the distribution of the outcome measures is significantly non-normal (Gelman et al., 2013).

Brier Score. For binary outcome measures, a popular squared-error derivation is the Brier (1950) score, a probabilistic analogue to the squared-error cost function written as

$$Brier = \frac{1}{N}\sum_{i=1}^{N}(p_i - \tilde{y}_i)^2, \tag{46}$$

where for the i^{th} observation, p_i is the probabilistic forecast and $\tilde{y}_i$ denotes the actual outcome or class status (coded as 1 when the outcome or class membership occurred and 0 otherwise).

Out-of-Sample Prediction

The generalizability of predictive models and decision-making to out-of-sample contexts is of critical importance when such models are employed in organizational settings (Putka et al., 2018), and should thus be a critical consideration in any model evaluation procedure. In order to estimate the out-of-sample predictive performance of a given modelling routine, two popular approaches are the *corrective adjustment-based* approach, and the *cross-validation* approach. While a detailed review is beyond the scope of this thesis, the former approach is comprised of measures referred to as *information criteria*, under which the within-sample predictive accuracy is calculated and a bias-corrected adjustment of within-sample error is applied (Gelman et al., 2013). There are numerous such metrics, but some of the more popular ones include the *Akaike Information Criterion* (Akaike, 1973), the *Watanabe-Akaike Information Criterion* (Watanabe, 2013), and the *Deviance Information Criterion* (Spiegelhalter et al., 2002, van der Linde, 2005).

In contrast, under a *cross-validation* approach, the researcher instead partitions the dataset into a training set, upon which the model is fit, and a testing set, upon which the predictive accuracy is evaluated (Bishop, 2006). Further, in a *k-fold* cross-validation approach, the researcher can elect to recursively partition the dataset into multiple (k) training and hold-out sets (*folds*), ranging from the simplest 2-fold case, to the extreme leave-one-out (LOO) case, consisting of n folds (Gelman et al., 2013). While computationally intensive (Gelman et al., 2013), cross-validation is now a staple component of modern predictive methods (Putka et al., 2018), and is considered to be the industry standard for model evaluation in organizational applications (Society of Industrial Organizational Psychology, 2018).

Predictive Efficacy of Bayesian Model Averaging

There is an abundance of literature regarding the efficacy of BMA relative to other predictive methods. Specifically, Madigan and Raftery (1994) found that BMA predicts at least as well as any single model with respect to the log score rule, and Min and Zellner (1993) have shown that the expected squared-error loss of point predictions is always minimized by BMA in situations where the model space includes the generative model. Further, on the basis of empirical results, Raftery and colleagues (1997) reported that predictive coverage under a BMA routine was improved relative to that observed under a single model. This finding was subsequently confirmed by Hoeting and colleagues (1999) and demonstrated again by Kaplan and Lee (2018). Similarly, Fernandez et al. (2001) and Ley and Steel (2009) both found that under the log score rule, BMA provided substantially better predictive performance than single models when applied to economic growth data. Lastly, while the majority of the literature has evaluated the predictive performance of BMA routines relative to model selection methods, research by Davidson and Fan (2006) found that BMA significantly outperformed various ensemble-based machine learning methods in certain situations.

Predictor Variable Importance

While the bulk of the current study focuses upon the benefits of the BMA routine that pertain specifically to optimizing predictive performance, the BMA approach does provide metrics that allow the researcher to gauge the relative "importance" of individual predictor variables. Consequently, a brief review of several of the predictor variable importance metrics available under the BMA, OLS/logistic regression, and Random Forest modelling routines is provided.

Posterior Inclusion Probabilities

The posterior inclusion probabilities (PIPs) are generated by summing the posterior model probabilities of all the models that contain a given predictor and can be interpreted as the marginal model-averaged probability that the regression coefficient for a given predictor is not equal to zero (Zeugner, 2011). When the PIP for a predictor variable is high (e.g. close to 1.00), this denotes that virtually all of the posterior model mass is distributed over models that contain that predictor. In contrast, when the PIP for a predictor variable is low (e.g. close to 0.00), this denotes that the majority of the most probable models do not include said predictor variable. Consequently, in contrast to frequentist *p*-values, predictors with high PIPs will have model-averaged coefficient estimates that are far from zero, and predictors with low PIPs will have coefficient estimates that are close to zero (Zeugner, 2011).

Rescaled Relative Weights

Rescaled Relative Weights (RRWs) are a specific type of relative importance index for regression models that can be employed to partition explained variance (R^2) into the portion that is attributable to each individual predictor (Tonidandel et al., 2009; Tonidandel & LeBreton, 2011). Specifically, relative weights analyses combat intercorrelations amongst the predictor variables by utilizing a variable transformation approach to create a new set of orthogonal predictors that the outcome measure is then regressed upon (Johnson, 2000; Tonidandel & LeBreton, 2011). Further, the RRWs represent a rescaled version of the "raw" relative weights in that they represent a percentage of the explained/predicted variance that is attributable to each predictor.

Increase in Node Purity

Variable importance metrics provided by Random Forest models are typically derived by calculating the total decrease in node impurities that results from splitting on a given predictor variable, averaged over all of the constituent bootstrapped trees (Liaw & Wiener, 2002). Specifically, node impurity is typically measured via the decrease in Residual Sums of Squares (RSS) in regression contexts, and via the decrease in the Gini homogeneity index in classification contexts (Rokach, 2019). Consequently, variable importance in random forest modelling routines is measured via the total decrease in either the RSS (regression contexts) or the Gini index (classification contexts), averaged over all the individual trees.

Current Study

In order to highlight how BMA can be employed to optimize both regression-based and classification-based predictions of organizational outcomes by acknowledging model uncertainty pertaining to covariate inclusion, two case studies based on an illustrative dataset used by Yuan and colleagues (2021) are presented. In Case Study #1, the predictive performance of BMA with respect to a continuous measure of employee turnover intentions is compared to that of both the full multiple linear regression model and a random forest regression routine. In Case Study #2, the predictive performance of BMA with respect to a binary classification of employee turnover risk is compared to that of both the full logistic regression model and a random forest classifier. Further details regarding the six predictive routines and model configurations are included below.

*Hypothesis 1***:** Bayesian Model Averaging should provide an increase in predictive performance relative to traditional single-model methods and machine-learning ensemble-based

methods when used to predict organizational outcomes in both classification and regression contexts.

CHAPTER II: METHODS

Data

Both case studies make use of a dataset used by Yuan and colleagues (2021) that was collected from a convenience sample of employees (N = 1,454) from 199 different small and medium-sized businesses in The Netherlands between 2017 and 2019. The dataset was obtained for the purposes of this thesis from the Open Science Framework page pertaining to the article published by Yuan and colleagues (2021), which can be viewed directly at https://osf.io/yx2us/.

Data Collection

The data were derived from two questionnaires, one of which was distributed to the individual employees, and one of which was given to the managing director of each organization. The questionnaire given to the individual employees contained validated measures of turnover intentions, demographic characteristics such as education, age, contract type, and job/work culture attitudes such as perceived fairness, pay satisfaction, and job proactivity. The questionnaire that was given to each organization's managing director covered a variety of organizational-level contextual factors, such as the company's size, governance structure, and past turnover rates. A complete list of the variables included in the current study can be seen below in Table 1, and a detailed description provided by Yuan et al. (2021) of the measures, their origins, exemplar items and scale/descriptive statistics can be seen in Appendices A and B. In addition, while data were initially collected from 1,510 respondents, the respondent pool was later screened by Yuan et al. (2021) to only include the 1,454 participants who responded to all of the constituent items for the variable of interest, turnover intentions.

Table 1.

Variables Selected for the Current Study

Level	Variable	Number of items	Scale
Person			
	Turnover intentions	3	Yes/No
	Education level	1	Years
	Management position	1	Yes/No
	Hours per week - contract	1	Hours
	Contract type	1	Temp/Perm
	Leader-member exchange	12	5pt Likert
	Information sharing	1	5pt Likert
	Employee voice	5	5pt Likert
	Pay satisfaction	3	5pt Likert
	Perceived fairness	16	5pt Likert
	Job proactivity	5	5pt Likert
	Career opportunities	3	5pt Likert
Organization			
	Total number of employees	1	Numeric
	Number of FTE current year	1	Numeric
	Number of FTE last year	1	Numeric
	Presence of employee council	1	Yes/No
	Presence of HRM on-site	1	Yes/No
	Number of hierarchical levels	1	Numeric
	Number of departments	1	Numeric
	Number of managers	1	Numeric
	Family business	1	Yes/No
	HRM 1: Past turnover rate	1	5pt Likert
	HRM 2: Attract qualified pers.	1	5pt Likert
	HRM 3: Retain key employees	1	5pt Likert
	HRM 4: Absenteeism	1	5pt Likert
	HRM 5: Labor disputes	1	5pt Likert
	HRM 6: Employee involvement	1	5pt Likert
	HRM 7: Innovation	1	5pt Likert
	HRM 8: Quality of ideas	1	5pt Likert
	HRM 9: Employee flexibility	1	5pt Likert
	Strategic planning	1	5pt Likert
	Corporate entrepreneurship	7	5pt Likert

Note. HRM = Human Resources Manager.

Demographics and Background

The following information regarding employee demographics and organizational context are derived directly from the original study conducted by Yuan et al. (2021). The largest group of businesses in the sample operated in the service industry (54.8%), followed by manufacturing and agriculture (20.1%), healthcare, education, and recreation (14.6%), and logistics (10.6%). The average company size, as measured by the average number of full-time equivalent employees (FTEs), was 22.92 FTEs (*SD* = 31.85). However, given that a proportion of the various organizations' employees worked part-time, the average number of "real" employees per organization was likely substantially higher, although no additional details were provided. Lastly, a slight majority (~ 61%) of the organizations sampled were self-characterized by the managing directors as being family-owned-and-operated. Regarding employee demographics, of the 1,454 employee respondents included in the final sample, slightly more than half (~57%) self-identified as male, and roughly 38% self-reported as having acquired at least a bachelor's degree from a four-year university. Finally, the average employee age was 38.44 years (*SD* = 12.87).

Outcomes of Interest

The target variable used by Yuan et al. (2021) and in the first case study of this thesis is a continuous measure of turnover intentions measured as a composite of three Likert-style items designed by Valentine and colleagues (2006) to relate thoughts about the likelihood of leaving with expected job search behaviors within a 3-year timeframe. In addition to this first outcome measure, a secondary target variable was created via dichotomization of the original variable such that participants with scores greater than 3.00 (out of a 5-pt Likert scale) were classified as being at greater risk of quitting (High-Risk), and respondents with scores of 3.00 or below were

classified as being at a lower risk of quitting (Low Risk). Construction of this alternate outcome measure via dichotomization was only made such that the current study could illustrate the implementation of a BMA routine in both a regression and a classification context. As noted by Hunter and Schmidt (1990), dichotomization of a continuous variable should rarely be conducted in practice.

Data Structure

Lastly, it should also be noted that the data is of a nested structure, as it consists of employees nested within organizations. While the relative efficacy of fixed, random, and mixed-effects approaches was a primary focus of Yuan and colleagues (2021), such an approach is beyond the scope of this thesis, which will utilize only a fixed-effects approach. Specifically, in both case studies, both the employee and organizational-level variables are included for consideration in the model space, with observations from employees belonging to the same organization all sharing the same values on the organizational factors. Prospective practitioners and researchers should ensure that real-life data nestedness is adequately addressed, as will be addressed further in Section IV.

Data Preparation and Pre-Processing

Upon receipt of the initial dataset, all of included predictor variables were already mean centered by Yuan and colleagues (2021), and no further scaling or standardization methods were utilized. Data missingness was then screened for and listwise deletion was employed, resulting in the deletion of 65 cases and a final sample consisting of 1,389 observations. Last, the original dataset included variables regarding the participants' age and gender. Due to concerns regarding protected-status groups, age and gender were removed from the final dataset and were not included as predictor variables in any of the modelling routines.

Data Partitioning into Training and Testing Subsets

In order to better approximate the out-of-sample predictive performance for each predictive modelling routine the full dataset was partitioned using a 70/30 split into a training subset (70%) and a testing subset (30%) using the *caret* package (Kuhn, 2021) in *R* (R Core Team, 2021). Importantly, the partitioning procedure was conducted using a stratified split with respect to the target variable, such that distribution of scores on the outcome measure, turnover intentions, was preserved in both the training and testing subsets. Stratified splits are particularly important in the context of classification problems, where any disjoint proportionality between the target class groups can greatly impact predictive performance. Consequently, preserving the relative distribution of the outcome measure scores across both the training and testing subsets is important in order to mitigate predictive performance variability (Sechidis et al., 2011). The end result of the data partitioning step was thus a single subset of training data (n = 973) and a single subset of testing data (n = 416) that was next used to fit and evaluate, respectively, each of the modelling routines. Importantly, this single test/train split was used for both the regression and classification contexts, with the only distinction being the selected outcome measure (continuous turnover intentions vs binary turnover risk).

Case Study 1: BMA for Linear Regression

Overview

For the first case study, the original continuous measure of turnover intentions was regressed upon both the employee-level and organizational-level factors. A fixed-effects approach was utilized, in which observations from employees belonging to the same organization all shared the same values on the various organizational factors. All employee-level and organizational-level variables shown in Table 1 were included as potential covariates, with

no hierarchical distinction made. As such, the full model space contained 34 predictor variables and 2^{34} separate submodels, making it an ideal candidate for prediction under a BMA routine.

Analytic Procedure

The predictive performance of the BMA routine, as applied to the regression context, was evaluated relative to that of the full OLS model and the random forest model using an analytic procedure that consisted of three general steps. In the first stage, the various applicable hyperparameters, configurations, and settings were tuned and/or specified for each of the three modelling routines, as described below.

The BMA linear regression modelling routine was conducted using the *BAS* package (Clyde, 2020) in *R* (R Core Team, 2021). Following the current literature standard, a "random" model space prior equivalent to the placement of a beta-binomial distribution over the prior variable inclusion probability was selected (Brown et al., 1998; Clyde & George, 2004; Ley & Steel, 2009). This has the desired effect of placing a uniform prior distribution over the expected model size (Ley & Steel, 2009). For the prior over the regression coefficients, the random hyper-g prior recommended by Liang et al. (2008) was used, with a hyperparameter value of $a = 3$ selected. As discussed in Section I, this corresponds to a placement of a Beta-distributed hyperprior over the shrinkage factor $\frac{g}{1+g}$, with the hyperparameter selection of $a = 3$ yielding model-specific coefficient estimates similar to those derived under a regularized OLS routine. Finally, 524,288 unique models were sampled from the model space using the *BAS* (Clyde, 2020) default *MC3*-based sampler with 50,000,000 iterations.

The OLS linear regression modelling routine used for comparison consisted of the full regression model with all available predictor variables included. There were no applicable

hyperparameters, specifications, or configurations to be considered, and the model was conducted using base *R* (R Core Team, 2021).

The random forest regression model was fit using the *caret* package (Kuhn, 2021) in *R* (R Core Team, 2021), which by default uses a cross-validation tuning process to select the optimal number of variables to be randomly sampled as candidates at each decision tree split. Specifically, the default selection for the *caret* package random forest model uses a *k*-fold cross-validation routine with a *RMSE* performance metric to evaluate a range of possible numbers of predictor variables to be randomly sampled as candidates at each split. In addition, the default number of 500 trees with a default bootstrapped resampling routine (as described in Section I) was utilized as well.

After specifying the applicable model configurations, the second step of the regression-based analytic procedure consisted of fitting the full OLS, random forest regression, and BMA linear regression models to the training data set using the software packages referenced above. This resulted in a set of applicable parameter estimates and decision rules for each of the three models, which were then formalized as the estimated regression equations generated by each trained modelling routine.

For the third and final step, each of the fitted models was used to generate predicted levels of turnover intentions for each of the testing set observations, such that three separate sets of predicted turnover level intentions were generated. The predictions generated by each of the three modelling routines were then compared to the actual turnover intention levels of the test set observations in order to evaluate and compare the predictive performance of the BMA approach with that of the full OLS model and the random forest regression model. In order to do so, the

root mean squared error (*RMSE*) and the percentage of variance explained (R^2) performance metrics were utilized, in alignment with Yuan and colleagues (2021).

Case Study 2: BMA for Logistic Regression/Classification

Overview

For the second case study, the binary measure of turnover risk (High Risk = 1, Low Risk = 0) was regressed upon the same employee-level and organizational-level factors as in the first case study. All of the variables shown in Table 1 were thus included again as potential predictors, utilizing the same fixed-effects approach with no hierarchical distinction. Consequently, the full model space under consideration for the second case study also contained 34 predictor variables and 2^{34} submodels, with the only distinction relative to the first case study being that of the functional form of the constituent generative models (logistic rather than normal linear).

Analytic Procedure

Similarly to the first case study, the predictive performance of the BMA routine, as applied to the classification context, was evaluated relative to that of the full logistic model and the random forest classifier using an analytic procedure that consisted of three general steps. In the first stage, the various applicable hyperparameters, configurations, and settings were tuned and/or specified for each of the three modelling routines, as described below.

The BMA logistic regression modelling routine was conducted using the *BAS* package (Clyde, 2020) in *R* (R Core Team, 2021). As in the regression context, a "random" model space prior equivalent to the placement of a beta-binomial distribution over the prior variable inclusion probability was selected (Brown et al., 1998; Clyde & George, 2004; Ley & Steel, 2009). This has the desired effect of placing a uniform prior distribution over the expected model size (Ley &

Steel, 2009). For the prior over the regression coefficients, the tCCH prior recommended by Li & Clyde (2018) for use in non-normal GLM contexts was used, with hyperparameter values selected that yield equivalent results to those produced under a normal linear regression context by the hyper-g ($a = 3$) prior of Liang et al. (2008). Similarly, to the normal linear context, this corresponds to the placement of a Beta-distributed hyperprior over the shrinkage factor $\frac{g}{1+g}$ that yields model-specific coefficient estimates similar to those derived under a regularized MLE estimation routine (Li & Clyde, 2018). In addition, given the classification context, the default binomial family and logistic link function were utilized for the functional form specification. Finally, as in the regression context, 524,288 unique models were sampled from the model space using the *BAS* (Clyde, 2020) default *MC3*-based sampler with 50,000,000 iterations.

The logistic regression modelling routine used for comparison consisted of the full logistic regression model with all available predictor variables included. There were no applicable hyperparameters, specifications, or configurations to be considered, and the model was conducted using base *R* (R Core Team, 2021).

Similarly to the random forest regression model, the random forest classifier model was fit using the *caret* package (Kuhn, 2021), which by default uses a cross validation tuning process to select the optimal number of variables to be randomly sampled as candidates at each decision tree split. Specifically, the default selection uses a *k*-fold cross-validation routine with an accuracy-based performance metric to evaluate a range of possible numbers of predictor variables to be randomly sampled as candidates at each split. In addition, the default number of 500 trees with a default bootstrapped training data resampling routine was utilized as well.

After specifying the applicable model configurations, the second step of the classification-based analytic procedure consisted of fitting the full logistic regression, random

forest classifier, and BMA logistic regression models to the training data set using the software packages referenced above. This resulted in a set of applicable parameter estimates and decision rules for each of the three models, which were then formalized as the estimated regression equations generated by each trained modelling routine.

For the third and final step, each of the fitted models was used to generate probabilistic predictions regarding turnover risk status for each of the testing set observations, such that three separate sets of probabilistic predictions regarding turnover risk status were generated. The probabilistic forecasts generated by each of the three modelling routines were then compared to the actual turnover risk status of the corresponding test set observations in order to evaluate predictive performance for each model. Specifically, following the example of Kaplan and Lee (2018), the Brier (1950) score was used to evaluate the efficacy of the BMA approach in generating a probabilistic classification forecast relative to that of the full logistic regression model and the random forest classifier. As discussed in Section I, the Brier score is analogous to a squared-error cost function and measures the discrepancy between the observed class status and the probabilistic forecast assigned by the modelling routine.

CHAPTER III: RESULTS

Case Study 1

Predictive Performance

In order to evaluate the predictive performance of the BMA regression routine relative to that of the full OLS and Random Forest regression models, the Root Mean Squared Error (*RMSE*) and the proportion of variance explained (R^2) and are provided below in Table 2. As hypothesized, the BMA regression routine produced higher R^2 and lower *RMSE* values than both the Random Forest and the full OLS regression models. It should be noted, however, that the difference in R^2 between the BMA and Random Forest routines was relatively small (~ 0.05%). Further, the Random Forest regression model produced lower *RMSE* and higher R^2 scores than the full OLS model.

Table 2.

Predictive Performance: Bayesian Model Averaging, OLS Regression, and Random Forest Comparison

Modeling Routine	Root Mean Squared Error	R^2
BMA	0.814	0.225
Full OLS	0.828	0.206
Random Forest	0.817	0.220

Note. BMA = Bayesian Model Averaging; OLS = Ordinary Least Squares regression

Parameter Estimates

Regression coefficient estimates and relevant statistics produced from the BMA regression routine and the full OLS regression model can be seen below in Table 3. Specifically, for the model-averaged regression coefficient estimates generated by the BMA routine, the marginal posterior mean $\mathbb{E}[\beta|y]$, the marginal 95% posterior credible interval, and the posterior

inclusion probability (PIP) are provided. For the coefficient estimates generated by the full OLS regression model, the unstandardized slopes *b*, the 95% confidence intervals for the unstandardized slopes, and the *p*-values for the tests of significance are provided. It should also be noted that random forest regression modelling routines do not employ or produce regression coefficients or their associated estimates, and as such are not included in Table 3.

As can be seen below, the coefficient estimates produced by the BMA routine tend to be of a smaller magnitude than those associated with the full OLS model. This is largely to be expected, as under the hyper-g prior placed upon the regression coefficients, the model-specific regression coefficients have conditional posterior expectations $\mathbb{E}\left[\beta_\gamma | y, g, M_\gamma\right]$ equivalent to a convex combination of the OLS estimates and a shrinkage term $\frac{g}{1+g}$, as discussed earlier in Section I (see Liang et al., 2008; Zeugner, 2011). Consequently, the results shown below suggest that in addition to the increase in predictive performance, the BMA regression routine also produces coefficient estimates that better incorporate uncertainty and mitigate potential overfitting.

Table 3.

Regression Coefficients: Bayesian Model Averaging and OLS Regression Comparison

	Bayesian Model Averaging		Full OLS Linear Regression		
Predictor	Mean (β | *y*)	95% CI (β | *y*)	*b*	95% CI	*p*
(Intercept)	2.128	[2.08, 2.18]	2.109	[1.97, 2.25]	< .001
Education level	0.107	[0.07, 0.15]	0.101	[0.06, 0.14]	< .001
Management position	0.003	[-0.03, 0.04]	0.032	[-0.10, 0.16]	.620
Hours per week - contract	-0.003	[-0.01, 0.00]	-0.006	[-0.01, 0.00]	.028
Contract type	0.001	[-0.02, 0.04]	0.022	[-0.11, 0.16]	.750
Leader-member exchange	-0.282	[-0.39, -0.16]	-0.286	[-0.40, -0.17]	< .001
Information sharing	-0.026	[-0.08, 0.00]	-0.059	[-0.11, -0.01]	.015
Employee voice	0.005	[0.00, 0.07]	0.047	[-0.02, 0.12]	.192
Pay satisfaction	-0.120	[-0.21, 0.00]	-0.114	[-0.20, -0.03]	.009
Perceived fairness	-0.161	[-0.31, 0.00]	-0.208	[-0.34, -0.07]	.003
Job proactivity	0.001	[-0.01, 0.03]	0.009	[-0.08, 0.10]	.847
Career opportunities	-0.207	[-0.28, -0.13]	-0.211	[-0.29, -0.14]	< .001
Total employees (*N*)	0.002	[0.00, 0.01]	0.001	[0.00, 0.00]	.610
Current FTE (*N*)	0.002	[0.00, 0.01]	0.007	[0.00, 0.01]	.027
Last year FTE (*N*)	-0.004	[-0.01, 0.00]	-0.007	[-0.01, 0.00]	.018
Employee council present	0.006	[0.00, 0.12]	-0.051	[-0.22, 0.12]	.559
HRM rep on-site	0.002	[-0.02, 0.03]	0.034	[-0.09, 0.15]	.579
Hierarchical levels (*N*)	0.013	[0.00, 0.03]	0.023	[0.01, 0.04]	.009
Departments (*N*)	-0.003	[-0.03, 0.00]	-0.023	[-0.05, 0.01]	.132
Managers (*N*)	0.001	[0.00, 0.01]	0.007	[-0.02, 0.03]	.545
Family business	-0.001	[-0.03, 0.01]	-0.003	[-0.12, 0.11]	.957
HR1: Past turnover rate	0.004	[0.00, 0.06]	0.04	[-0.03, 0.11]	.235
HR2: Attract qualified talent	-0.094	[-0.16, 0.00]	-0.1	[-0.18, -0.02]	.016
HR3: Retain key talent	0.002	[-0.01, 0.03]	0.042	[-0.05, 0.14]	.369
HR4: Absenteeism	0.000	[0.00, 0.02]	-0.014	[-0.07, 0.05]	.648
HR5: Labor disputes	0.002	[0.00, 0.04]	0.016	[-0.08, 0.11]	.728
HR6: Employee involvement	0.003	[0.00, 0.06]	0.004	[-0.10, 0.11]	.936
HR7: Innovation	0.004	[0.00, 0.06]	0.021	[-0.06, 0.11]	.628
HR8: Quality of ideas	-0.012	[-0.12, 0.01]	-0.115	[-0.23, -0.00]	.045
HR9: Employee flexibility	0.017	[0.00, 0.12]	0.098	[0.00, 0.20]	.048
Supervisory board	0.175	[0.00, 0.47]	0.331	[0.07, 0.59]	.012
Advisory board	-0.259	[-0.48, 0.00]	-0.379	[-0.60, -0.17]	< .001
Corporate entrepreneurship	-0.001	[-0.03, 0.01]	-0.007	[-0.09, 0.08]	.878
Strategic planning	0.073	[0.00, 0.25]	0.173	[0.02, 0.32]	.023
Entrepreneurial orientation	-0.021	[-0.10, 0.00]	-0.075	[-0.15, 0.00]	.039

Note. OLS = Ordinary Least Squares regression; Mean (β|*y*) = marginal posterior mean; 95% CI (β | y) = Bayesian credible interval; 95% CI = OLS confidence interval; FTE = Full-time employees; *N* = number/count. HR = Human Resources

Predictor Importance

The various predictor importance statistics associated with each of the BMA regression, full OLS regression, and Random Forest regression modelling routines can be seen below in Table 4. Specifically, for the BMA routine, the PIP $p(\beta \neq 0|y)$ is provided for each predictor and the predictor variables have been sorted in rank order from largest to smallest PIP accordingly. For the full OLS regression model, the Rescaled Relative Weight (RRW) is provided by way of the *rwa* (Chan, 2020) package in *R* (R Core Team, 2021). The rank of each predictor according to the RRW metric is also provided, although the predictors are still presented in rank order according to the PIP metric. Lastly, for the Random Forest regression model, the decrease in Residual Sums of Squares (RSS) is provided, along with the associated ranks generated for each of the predictor variables.

While the importance rankings observed under the full OLS and BMA regression routines seemed to be in relative agreement regarding the higher-ranked predictors, they became increasingly divergent as relative importance decreased. In contrast, the variable importance rankings observed under the random forest regression routine differed greatly from both the full OLS and BMA regression routines, with no clear divergence pattern being readily apparent. Also of note is the observation that out of the 32 available predictor variables, only 13 had PIPs greater than 0.50, which marks the median probabilistic inclusion point (Barbieri & Berger, 2004). In other words, only 13 out of the 32 available predictor variables were included in models that comprised the majority of the posterior mass. In contrast, under the full OLS model, 18 of the 32 predictors were statistically significant when evaluated at an alpha level of .05. Similarly, the random forest cross-validation tuning process identified a model size of 18 as being optimal under the RMSE metric.

Table 4.

Predictor Importance: BMA, Full OLS Regression, and Random Forest Comparison

	BMA		Full OLS Regression		Random Forest	
Predictor	Rank	PIP	Rank	RRW	Rank	RSS Decrease
Education level	1	1.00	5	6.04	10	26.62
Leader-member exchange	2	1.00	2	16.34	3	79.50
Career opportunities	3	1.00	1	19.98	2	104.43
Pay satisfaction	4	0.93	4	11.34	4	44.66
HR2: Attract qualified talent	5	0.92	7	3.09	20	10.31
Perceived fairness	6	0.85	3	16.28	1	124.95
Advisory board	7	0.85	8	2.91	34	3.01
Last year FTE (*N*)	8	0.69	16	1.01	12	24.86
Total employees (*N*)	9	0.68	12	1.41	11	25.59
Hierarchical levels (*N*)	10	0.66	11	1.41	16	19.14
Supervisory board	11	0.63	13	1.22	29	4.88
Hours per week - contract	12	0.55	10	2.18	8	28.01
Information sharing	13	0.55	6	4.34	14	24.13
Strategic planning	14	0.48	19	0.63	7	30.22
Current FTE (*N*)	15	0.39	14	1.12	15	23.94
Entrepreneurial orientation	16	0.36	20	0.54	13	24.25
HR9: Employee flexibility	17	0.24	25	0.41	23	8.89
HR8: Quality of ideas	18	0.17	17	1.01	25	8.32
Departments (*N*)	19	0.16	31	0.23	18	14.72
Employee voice	20	0.13	9	2.68	6	33.97
HR1: Past turnover rate	21	0.13	15	1.17	19	13.59
HR7: Innovation	22	0.13	30	0.24	21	9.89
Employee council	23	0.11	27	0.33	33	3.28
HR6: Employee involvement	24	0.10	28	0.25	26	7.22
Management position	25	0.09	34	0.11	28	6.32
Managers (*N*)	26	0.09	21	0.54	17	17.00
HR3: Retain key talent	27	0.09	22	0.52	24	8.47
HRM rep onsite	28	0.08	29	0.24	32	3.79
HR5: Labor dispute	29	0.08	26	0.37	27	7.08
Corporate entrepreneurship	30	0.08	23	0.47	9	26.63
Contract type	31	0.07	32	0.16	31	4.73
Job proactivity	32	0.07	18	0.79	5	37.17
Family business	33	0.07	24	0.43	30	4.78
HR4: Absenteeism	34	0.07	33	0.12	22	9.47

Note. BMA = Bayesian Model Averaging; RRW = Rescaled Raw Relative Weights; OLS = Ordinary Least Squares regression; RSS = Residual Sums of Squares; PIP = posterior inclusion probability, or the marginal probability that the coefficient is not equal to zero; *N* = number/count; HR = Human Resources.

Case Study 2

Predictive Performance

In order to evaluate the predictive performance of the BMA logistic regression routine relative to that of the full logistic regression and Random Forest classifier models, the overall accuracy, kappa, and Brier Score are provided below in Table 5. As hypothesized, lower Brier Score values were observed under the BMA routine than both the Random Forest classifier and the full logistic regression models. However, it should also be noted that the difference in Brier Score values between the BMA and full logistic regression routines was relatively small (~ 0.04). Further, the BMA routine had a lower accuracy rate than the full logistic model, and a lower kappa score than both the logistic regression model and the Random Forest classifier. A review of the associated confusion matrices suggested that these lower accuracy-based scores may be due to the BMA routine's regularization term, which "shrinks" the probabilistic forecasts and thus lowers the frequency of positive (target) classifications relative to the other models. Further, given the regularization of the probabilistic estimates, selection of a different classification threshold than the 50% level used for the current study may yield better accuracy-based results for the BMA routine.

Table 5.

Predictive Performance: BMA, Logistic Regression, and Random Forest Comparison

Modelling Routine	Accuracy	Kappa	Brier Score
BMA	0.868	0.21	0.800
Full Logistic	0.873	0.31	0.804
Random Forest	0.865	0.220	0.812

Note. BMA = Bayesian Model Averaging.

Parameter Estimates

Regression coefficient estimates and relevant statistics produced from the BMA logistic regression routine and the full logistic regression model can be seen below in Table 6. Specifically, for the model-averaged regression coefficient estimates generated by the BMA routine, the marginal posterior means $\mathbb{E}[\beta|y]$, the marginal 95% posterior credible intervals, the posterior inclusion probabilities (PIPs), and the odds ratios are provided. For the coefficient estimates generated by the full logistic regression model, the unstandardized slopes *b*, the 95% confidence intervals for the unstandardized slopes, the *p*-values for the tests of significance, and the odds ratios are provided. It should also be noted that as in the regression context, random forest classifier modelling routines do not employ or produce regression coefficients or their associated estimates, and as such are not included in Table 6.

As in the regression context, the coefficient estimates produced by the BMA routine in the classification context are analogous to those produced under the full logistic regression model, but of a smaller magnitude. This to be expected, as the tCCH prior over the regression coefficients utilized in the present study is the GLM-counterpart to the hyper-g prior used in the regression context (Li & Clyde, 2018). Consequently, the model-specific regression coefficients have conditional posterior expectations $\mathbb{E}\left[\beta_\gamma|y, g, M_\gamma\right]$ equivalent to a convex combination of the MLEs and a shrinkage term $\frac{g}{1+g}$, as discussed earlier in Section I (see Li & Clyde, 2018). As noted by Copas (1993) and Li & Clyde (2018), shrinking regression coefficients toward the prior mean of zero can mitigate the potential for overfitting and thereby help optimize out-of-sample predictive performance. Further, in classification and/or logistic regression contexts, this regularization has the additional consequence of shrinking the odds ratios towards 1.00, as can be seen when comparing the BMA and logistic regression results shown in Table 6.

Table 6.

Regression Coefficients: Bayesian Model Averaging and Logistic Regression Comparison

	Bayesian Model Averaging			Full Logistic Regression			
Predictor	Mean ($\beta \mid y$)	95% CI ($\beta \mid y$)	O.R.	*b*	95% CI	*p*	O.R.
(Intercept)	-2.243	[-2.74, -1.74]	0.11	-2.43	[-3.08, -1.82]	< .001	0.09
Education level	0.110	[0.00, 0.29]	1.12	0.18	[0.00, 0.36]	.045	1.20
Management position	0.001	[-0.30, 0.35]	1.00	0.00	[-0.54, 0.52]	.994	1.00
Hours per week – contract	-0.002	[-0.02, 0.01]	1.00	-0.01	[-0.03, 0.01]	.479	0.99
Contract type	0.079	[-0.18, 0.58]	1.08	0.35	[-0.21, 0.94]	.238	1.42
Leader-member exchange	-0.326	[-0.78, 0.00]	0.72	-0.48	[-0.93, -0.40]	.034	0.62
Information sharing	-0.004	[-0.12, 0.10]	1.00	-0.01	[-0.19, 0.17]	.898	0.99
Employee voice	-0.062	[-0.35, 0.04]	0.94	-0.12	[-0.41, 0.16]	.402	0.89
Pay satisfaction	-0.16	[-0.55, 0.02]	0.85	-0.24	[-0.57, 0.10]	.170	0.79
Perceived fairness	-0.452	[-1.00, 0.00]	0.64	-0.48	[-1.02, 0.06]	.084	0.62
Job proactivity	-0.003	[-0.26, 0.20]	1.00	-0.02	[-0.39, 0.35]	.914	0.98
Career opportunities	-0.625	[-0.90, -0.36]	0.54	-0.69	[-0.99, -0.4]	< .001	0.50
Total employees (*N*)	0.001	[-0.00, 0.01]	1.00	0.00	[-0.01, 0.01]	.855	1.00
Current FTE (*N*)	0.002	[-0.01, 0.02]	1.00	0.02	[-0.01, 0.04]	.164	1.02
Last year FTE (*N*)	-0.002	[-0.02, 0.01]	1.00	-0.01	[-0.03, 0.01]	.206	0.99
Employee council present	0.29	[-0.01, 0.90]	1.33	0.31	[-0.34, 0.95]	.350	1.36
HRM rep on-site	0.003	[-0.27, 0.35]	1.00	-0.03	[-0.53, 0.47]	.917	0.97
Hierarchical levels (*N*)	0.01	[-0.02, 0.07]	1.01	0.04	[-0.03, 0.11	.244	1.04
Departments (*N*)	-0.018	[-0.13, 0.04]	0.98	-0.08	[-0.21, 0.05]	.210	0.92
Managers (*N*)	0.008	[-0.03, 0.08]	1.01	0.02	[-0.08, 0.11]	.727	1.02
Family business	-0.113	[-0.60, 0.09]	0.89	-0.18	[-0.67, 0.30]	.459	0.84
HR1: Past turnover rate	0.185	[0.00, 0.43]	1.20	0.31	[0.02, 0.59]	.034	1.36
HR2: Attract qualified talent	-0.002	[-0.22, 0.20]	1.00	0.09	[-0.26, 0.43]	.613	1.09
HR3: Retain key talent	-0.012	[-0.27, 0.16]	0.99	-0.03	[-0.41, 0.35]	.861	0.97
HR4: Absenteeism	0.013	[-0.09, 0.19]	1.01	0.06	[-0.19, 0.31]	.618	1.06
HR5: Labor disputes	-0.01	[-0.26, 0.18]	0.99	-0.12	[-0.53, 0.28]	.569	0.89
HR6: Employee involvement	-0.038	[-0.34, 0.15]	0.96	-0.04	[-0.49, 0.41]	.851	0.96
HR7: Innovation	-0.083	[-0.41, 0.05]	0.92	-0.23	[-0.60, 0.14]	.234	0.79
HR8: Quality of ideas	-0.024	[-0.32, 0.16]	0.98	-0.08	[-0.55, 0.38]	.728	0.92
HR9: Employee flexibility	-0.005	[-0.24, 0.19]	1.00	-0.03	[-0.44, 0.39]	.891	0.97
Supervisory board	0.057	[-0.49, 0.79]	1.06	0.36	[-0.67, 1.33]	.481	1.43
Advisory board	-0.414	[-1.56, 0.16]	0.66	-1.10	[-2.43, -0.01]	.069	0.33
Corporate entrepreneurship	0.051	[-0.10, 0.37]	1.05	0.09	[-0.28, 0.46]	.640	1.09
Strategic planning	0.213	[-0.02, 0.86]	1.24	0.53	[-0.08, 1.15]	.086	1.70
Entrepreneurial orientation	-0.013	[-0.22, 0.14]	0.99	-0.06	[-0.37, 0.25]	.695	0.94

Note. Mean ($\beta \mid y$) = marginal posterior mean; 95% CI ($\beta \mid$ y) = Bayesian credible interval ; 95% CI = logistic regression confidence interval; O.R. = odds ratio; *N* = number/count; FTE = Full-time employees; HR = Human Resources; HRM = Human Resources Management

Predictor Importance

The various predictor importance statistics associated with each of the BMA logistic regression, full logistic regression, and Random Forest classifier modelling routines can be seen below in Table 7. Specifically, for the BMA routine, the PIP $p(\beta \neq 0|y)$ is provided for each predictor and the predictor variables have been sorted in rank order from largest to smallest PIP. For the full logistic regression model, the RRW metric and the associated predictor ranks are again provided, although as in the regression context, the predictors are still presented in rank order according to the PIP metric. Lastly, for the Random Forest classifier, the mean Gini decrease is provided, along with the associated ranks generated for each of the predictor variables.

As seen below, results indicate agreement between the three modelling routines regarding the top few predictor importance rankings, as career opportunities, fairness, leader-member exchange, and pay satisfaction are all highly ranked. However, aside from the agreement regarding the top few predictor importance rankings, no clear trend regarding rest of the predictor variables' importance rankings seems readily apparent. In addition, predictor importance agreement between the BMA and full logistic regression models does not appear to be significantly higher than the two routines' level of agreement with the random forest model, which is a noticeable divergence from the regression context. Lastly, only nine out of the 32 available predictors had PIPs greater than 0.50, which indicates that only nine predictors were included in models that comprised most of the posterior mass. In contrast, under the full logistic regression model, only four of the 32 predictors were statistically significant at the .05 alpha level. Similarly, the random forest cross-validation tuning process identified a model size of 2 as being optimal under the accuracy-based metrics.

Table 7.

Predictor Importance: Bayesian Model Averaging, Logistic Regression, and Random Forest Comparison

	BMA		Full Logistic Regression		Random Forest	
Predictor	Rank	PIP	Rank	RRW	Rank	Mean Gini Decrease
Career opportunities	1	1.00	1	27.80	2	18.98
Perceived fairness	2	0.82	2	12.85	1	19.44
HR1: Past turnover rate	3	0.79	18	0.78	20	3.87
Leader-member exchange	4	0.78	3	12.01	3	15.56
Education level	5	0.73	7	4.41	12	7.24
Employee council	6	0.66	23	0.60	32	1.92
Pay satisfaction	7	0.61	4	9.57	5	12.02
Advisory board	8	0.58	13	1.14	34	0.69
Strategic planning	9	0.55	8	2.62	14	6.90
HR7: Innovation	10	0.48	17	0.80	23	3.55
Family business	11	0.45	27	0.51	29	2.27
Employee voice	12	0.43	6	5.23	4	12.49
Hierarchical levels (*N*)	13	0.38	11	1.70	18	5.11
Departments (*N*)	14	0.38	30	0.19	17	5.59
Corp. entrepreneurship	15	0.38	12	1.31	13	7.13
Total employees (*N*)	16	0.37	16	0.84	8	8.28
Current FTE (*N*)	17	0.37	15	1.01	10	7.64
Contract type	18	0.36	31	0.16	30	2.17
Last year FTE (*N*)	19	0.36	14	1.08	11	7.42
HR6: Employee involvement	20	0.35	24	0.65	24	3.34
Hours per week – contract	21	0.32	10	2.09	9	7.82
Managers (*N*)	22	0.32	26	0.51	16	5.97
HR8: Quality of ideas	23	0.3	10	0.73	25	3.17
Supervisory board	24	0.3	24	0.58	33	1.00
Entrepreneurial orientation	25	0.3	29	0.24	15	6.85
Information sharing	26	0.29	5	5.67	7	8.69
HR2: Attract key talent	27	0.29	9	2.30	19	4.09
HR4: Absenteeism	28	0.29	33	0.12	21	3.80
Job proactivity	29	0.28	25	0.55	6	9.65
HRM rep on-site	30	0.28	34	0.12	31	2.15
HR3: Retain key talent	31	0.28	28	0.31	26	3.10
HR5: Labor disputes	32	0.28	19	0.77	27	2.67
HR9: Employee flex.	33	0.28	21	0.68	22	3.60
Management position	34	0.27	32	0.14	28	2.64

Note. RRW = Rescaled Raw Relative Weights; Mean Gini Decrease = average decrease in Gini index purity metric; PIP = posterior inclusion probability, or the marginal probability that the coefficient is not equal to zero; *N* = number/count.

CHAPTER IV: DISCUSSION

This study was conducted to investigate the predictive performance of BMA routines relative to that of random forest models and the full OLS and logistic regression models when used to forecast turnover intentions and turnover risk. Prior to conducting analyses, it was hypothesized that the BMA routine would outperform the random forest models and the full OLS and logistic regression models in both the regression and classification contexts (Hypothesis 1). This hypothesis appears to have been supported in both the regression and classification contexts, as the BMA routine resulted in higher R^2/lower *RMSE* values (regression context) and lower Brier Score (classification context) metrics than did the OLS/logistic regression and the Random Forest modelling routines. It should, however, be noted that the differences in performance between the BMA routine and the Random Forest model (regression context), as well as the BMA routine and the logistic regression model (classification context), were not very large. Lastly, regarding the classification context, while the BMA routine resulted in poorer scores on the accuracy-based metrics, this should not be interpreted as undermining support for Hypothesis I, as the BMA routine still provided optimal probabilistic forecasts regarding the binary classifications. Further, as will be discussed further below, the classification inaccuracy may be partially attributable to the target class imbalance observed in the present study.

In addition, while the present study's primary research goal was centered upon predictive performance, the findings regarding the predictor variable importance rankings were unexpected and warrant further discussion. Specifically, while the magnitude and direction of the slope estimates were as expected (i.e., roughly equivalent but slightly smaller slope estimates were observed under the BMA routines due to the associated regularization terms), there was a large degree of divergence across the three modelling routines with respect to the importance rankings. In the regression context, results indicated relative agreement between the full OLS and the

BMA routines regarding only the higher-ranked predictors, and both routines were noticeably divergent from the random forest model. In the classification context, all three routines seemed to agree regarding the top few predictors, but results differed greatly regarding all of the other predictors.

While little research has been conducted to date that specifically addresses discrepancies between the PIP rankings available under the BMA routine and the predictor importance metrics used by other modelling routines, one possible explanation for the especially high divergence observed in the classification context may be the larger shrinkage penalty that was applied by the BMA routine. As seen in Figures 1 and 2 in Appendix C, the posterior mass for the shrinkage factor $\frac{g}{1+g}$ in the classification context was centered over a much lower numeric range (and thus a higher shrinkage penalty) than that observed under the regression context. This greater degree of regularization may be due to improper dichotomization of the target variable and the resultant class imbalance (discussed further in the Limitations section), which can result in a greater dispersion of posterior mass over larger models with smaller individual PIPs (see Zeugner, 2011). In addition, variance inflation factors were also checked for all available predictors, as multicollinearity can result in inaccurate underestimation of the marginal PIPs (Clyde et al., 2011). However, upon doing so, all included predictors had relatively low variance inflation factors (VIF < 3.0).

Organizational Implications

The utilization of predictive modelling routines to forecast workforce-related outcomes is an essential component of organizational success (Cheng et al., 2020; Fitz-Enz & Mattox, 2014). While an increasing number of organizations have implemented or are implementing workforce analytics programs in recognition of this insight, most such implementations are either single-

model selection-based approaches, or ensemble-based machine learning approaches. Despite comprising a predominance of the current predictive modelling routines, predictions generated by model selection and ensemble-based approaches fail to adequately incorporate model uncertainty regarding the underlying data-generating process (Steel, 2020). In contrast, predictions generated by BMA routines fully incorporate model uncertainty by averaging over the entire set of possible models, and have been shown to provide optimal predictive performance under a variety of scoring rules, contexts, and subject matter domains (e.g. Davidson & Fan, 2006; Hoeting et al., 1999; Kaplan & Lee, 2018).

Predictive BMA routines offer particular value in workforce-related contexts, as the advent of "big data" and modern computing power have led to an exponential increase in the amount of data and potential predictor variables available to organizational researchers and practitioners (Oswald et al., 2020). Rather than employ a theory-based or data-driven model selection approach, or a constrained quasi-averaging ensemble-based approach, predictive BMA routines offer a means of fully incorporating information from *all* available predictors and mitigate the risk of researcher misspecification regarding the "best" subset of predictors. The present study has provided an illustrative guide for how organizational researchers might employ BMA routines to forecast workforce-related outcomes and offered insights into the resulting predictive efficacy relative to several competing modelling routines commonly employed in organizational settings.

Based on the results, organizational researchers and practitioners may benefit from implementing BMA modelling routines when the goal is the optimization of out-of-sample predictive performance and generalizability. Further, the potential for increased predictive performance associated with BMA routines is readily accessible via several open-source

software packages in *R* (R Core Team, 2021), most of which offer "out-of-the-box" default configurations that require little to no custom specifications on the part of the researcher or practitioner. Consequently, BMA routines can offer organizations a cost-effective means of generating accurate and efficient probabilistic forecasts of workforce-related outcomes.

Lastly, the present study has illustrated how BMA routines might also offer researchers insight into the relative importance of individual predictor variables as well and can thus serve as a supplement to more explanatory-based traditional modelling routines without sacrificing predictive accuracy. While future research is needed in order to fully investigate the nuances surrounding BMA-generated PIP importance rankings, results of the current study suggest that BMA routines have the potential to afford organizational researchers the ability to simultaneously generate optimal predictions, mitigate overfitting and spurious associations, and ascertain the relative importance of possible predictors, all in a single intuitive modelling routine. It should also be noted, however, that caution should be undertaken to ensure that any explanatory inference should be coupled with applicable theoretical, legal, and ethical considerations.

Limitations and Future Directions

Several limitations should be noted regarding the present study. First, as noted by Kaplan and Lee (2018), the current study operates under the theoretical assumption that the "true" generative model, denoted M^T, is one of the models $\{M_1, M_2, \ldots, M_k\}$ that comprise the model space $\boldsymbol{\mathcal{M}}$. Referred to as the "*M*-closed framework" (Bernardo & Smith, 1994; Clyde & Iverson, 2013), this assumption underlies most BMA applications and is a conceptual prerequisite to the placement of prior probabilities over the model space. Importantly, the current study follows Steel (2020) in regarding this as a theoretical limitation rather than a practical one, and instead

treating model space priors as being reflective of the prior probability that a given model is a useful proxy for the "true" model.

A second limitation should be noted with respect to several analytic decisions made regarding the current dataset. Specifically, the decisions to dichotomize the original continuous outcome measure of turnover intentions and to employ a fixed-effects approach despite a nested data structure were both made solely for illustrative purposes and should not be replicated in practice. Organizational researchers and practitioners dealing with nested data structures should explore other options to resolve said clustering, as fixed-effects approaches are prone to overfitting and inhibit generalizability (Snijders & Bosker, 2012). Similarly, researchers and practitioners should rarely dichotomize continuous measures, as doing so can result in a loss of data information and statistical power (Hunter & Schmidt, 1990). Further, the cut-point utilized in this study for the dichotomization of the original turnover intentions measure resulted in a class imbalance, with the resultant target class comprising only 15% of the data observations. This imbalance could partially explain the poor performance of the BMA routine on the accuracy-based metrics (Japkowicz & Stephen, 2002).

A third limitation should be noted regarding the generalizability of the current study's results beyond the contexts and configurations specific to this study. Primarily, the findings of the current study are at least partially dependent upon the characteristics of a dataset drawn from a single-sample context. In other words, results and conclusions drawn from a single-sample study should be verified and validated through comprehensive replication- or simulation-based reviews. A Monte Carlo simulation study should be conducted to supplement the single-sample case study-based approach currently used. In addition, the current results were also dependent upon the configurations and analytical choices used in the specification of each of the modelling

routines. Specifically, for the random forest routine, results may differ based on the number of bootstrapped constituent trees chosen. For the BMA routines, results may differ depending on the size of the shrinkage penalty imposed, as well as the form and precision of both the model-space and model-parameter priors chosen.

Lastly, BMA applications in general, and non-normal GLM-based BMA applications in particular, suffer from several technical limitations. First, the computational complexity inherent in the calculation of the marginal likelihoods largely constrains the choice of prior distributions over model parameters to conjugate, or quasi-conjugate prior structures (Li & Clyde, 2018; Liang et al., 2008). While research has been conducted regarding possible MCMC-based methods for marginal likelihood estimation (e.g. Carlin & Chib, 1995; Chib, 1995), such techniques have yet to be implemented consistently and effectively. As such, the "custom" priors commonly associated with subjectivist Bayesian methods are currently out of reach for practitioners wishing to implement a BMA routine. In addition, the approximation-based integration methods used for marginal likelihood estimation can be unstable and lead to MCMC convergence issues under certain conditions, particularly when $R_{\boldsymbol{\gamma}}^2$ is very low and the expectation of the resultant data-dependent shrinkage term $\frac{g}{1+g}$ is "pulled" towards the lower bound (Clyde et al., 2011; Liang et al., 2008). This is especially problematic in the context of non-normal likelihood functions such as the classification context discussed in the present study, as exact closed-form integration is unavailable (Liang et al., 2008).

Conclusions

Despite the limitations noted above, the current study offers provides a novel contribution to the extant organizational research literature by illustrating how researchers and practitioners can employ Bayesian Model Averaging routines to optimize predictive performance when

forecasting workforce outcomes. Results showed that the BMA routine resulted in a greater proportion of variance explained (higher R^2) and lower *RMSE* that both the full OLS and Random Forest regression models when used to predict the continuous measure of employee turnover intentions. In the classification context, the BMA routine produced lower probabilistic-based Brier Score values than both the full logistic regression and the Random Forest models but appeared to underperform on the simpler accuracy-based metrics. Consequently, practitioners and researchers may benefit from using BMA routines to optimize both predictions of continuous organizational outcome measures and probabilistic forecasts of binary organizational outcomes.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *2nd International symposium on information theory* (pp. 267-281). Akademia Kiado.

Barbieri, M. M., & Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, *32*(3), 870-897.

Berger, J. O., Pericchi, L. R., Ghosh, J. K., Samanta, T., De Santis, F., Berger, J. O., & Pericchi, L. R. (2001). Objective Bayesian methods for model selection: Introduction and comparison. *Lecture Notes-Monograph Series*, *38*, 135-207.

Bernardo, J. M., & Smith, A. F. (1994). *Bayesian theory*. John Wiley & Sons.

Bishop, C. (2006). *Pattern recognition and machine learning*. Springer.

Bottolo, L., & Richardson, S. (2010). Evolutionary stochastic search for Bayesian model exploration. *Bayesian Analysis*, *5*(3), 583-618.

Box, G. E. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*(356), 791-799.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123-140.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Breiman, L, Friedman, J., Olshen. R., & Stone, C. (1984). *Classification and Regression Trees*. Wadsworth.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*(1), 1-3.

Brown, P. J., Vannucci, M., & Fearn, T. (1998). Multivariate Bayesian variable selection and prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 60*(3), 627-641.

Carlin, B. P., & Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(3), 473-484.

Chambers, E. G., Foulon, M., Handfield-Jones, H., Hankin, S. M., & Michaels III, E. G. (1998). The war for talent. *The McKinsey Quarterly, 3*, 44-57.

Chan, D. (1996). Criterion and construct validation of an assessment center. *Journal of Occupational and Organizational Psychology, 69*(2), 167-181.

Chan, M. (2020). *rwa: Perform a Relative Weights Analysis* [Computer software]. https://CRAN.R-project.org/package=rwa

Cheng, M. M., & Hackett, R. D. (2021). A critical review of algorithms in HRM: Definition, theory, and practice. *Human Resource Management Review, 31*(1), 1-14.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association, 90*(432), 1313-1321.

Clyde, M. (2020). *Bayesian variable selection and model averaging using Bayesian Adaptive Sampling* [Computer software]. https://www2.stat.duke.edu/~clyde/BAS/

Clyde, M., & George, E. I. (2000). Flexible empirical Bayes estimation for wavelets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62*(4), 681-698.

Clyde, M., & George, E. I. (2004). Model uncertainty. *Statistical Science, 19*(1), 81-94.

Clyde, M. A., Ghosh, J., & Littman, M. L. (2011). Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics, 20*(1), 80-101.

Clyde, M. and E. Iversen (2013). Bayesian model averaging in the M-open framework. In P. Damien, P. Dellaportas, N. Polson, & D. Stephens (Eds.), *Bayesian theory and applications* (pp. 493-98). Oxford University Press.

Copas, J. B. (1993). The shrinkage of point scoring methods. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 42*(2), 315-331.

Cui, W., & George, E. I. (2008). Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference, 138*(4), 888-900.

Darlington, R., & Hayes, A. F. (2017). *Regression analysis and linear models*. Guilford Press.

Davidson, I., & Fan, W. (2006). When efficient model averaging outperforms boosting and bagging. In J. Furnkranz, T. Scheffer, & M. Spiliopoulou (Eds.), *Machine learning and knowledge discovery in databases: Lecture notes in computer science*. Springer.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 45-70.

Fernandez, C., Ley, E., & Steel, M. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics, 100*, 381-427.

Fitz-Enz, J., & Mattox II, J. R. (2014). *Predictive analytics for human resources*. John Wiley & Sons.

Foster, D. P., & George, E. I. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics, 22*(4), 1947-1975.

Furnival, G. M., & Wilson, R. W. (1974). Regression by leaps and bounds. *Technometrics, 16*, 499-511.

Gao, X., Wen, J., & Zhang, C. (2019). An improved random forest algorithm for predicting employee turnover. *Mathematical Problems in Engineering, 2019*, 1-12.

Gatewood, R. D., Field, H. S., & Barrick, M. *Human resource selection* (7th ed.). Cengage Learning.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman & Hall/CRC.

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

George, E. I. (2000). The variable selection problem. *Journal of the American Statistical Association, 95*(452), 1304-1308.

George, E. I., & Foster, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika, 87*(4), 731-747.

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association, 88*(423), 881-889.

George, E. I., & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica, 7*(2), 339-373.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association, 106*(494), 746-762.

Good, I. (1952). Rational decisions. *Journal of the Royal Statistical Society*, *14*, 107–114.

Gordy, M. B. (1998). *A generalization of generalized beta distributions*. Board of Governors of the Federal Reserve System.

Guest, D. E., & Peccei, R. (2001). Partnership at work: Mutuality and the balance of advantage. *British Journal of Industrial Relations, 39*(2), 207–236.

Han, W. J., & Hart, J. (2021). Job precarity and economic prospects during the COVID-19 public health crisis. *Social Science Quarterly, 102*(5), 2394-2411.

Hansen, M. H., & Yu, B. (2003). Minimum description length model selection criteria for generalized linear models. *Lecture Notes-Monograph Series, 40*, 145-163.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science, 14*(4), 382-417.

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology, 75*(1), 28-42.

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology, 75*(3), 334-339.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (2nd ed.). Springer.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis, 6*(5), 429-449.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research, 35*(1), 1-19.

Kaplan, D. (2021). On the quantification of model uncertainty: A Bayesian perspective. *Psychometrika, 86*(1), 215-238.

Kaplan, D., & Lee, C. (2018). Optimizing prediction using Bayesian model averaging: Examples using large-scale educational assessments. *Evaluation Review, 42*(4), 423-457.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association, 90*(431), 928-934.

Kellermanns, F. W., & Eddleston, K. A. (2006). Corporate entrepreneurship in family firms: A family perspective. *Entrepreneurship, Theory and Practice, 30*(6), 809–830.

Kroon, B., van de Voorde, K., & Timmers, J. (2012). High performance work practices in small firms: A resource-poverty and strategic decision-making perspective. *Small Business Economics, 41*(1), 71–91.

Kuhn, M. (2021). *caret: Classification and regression training* [Computer software]. https://CRAN.R-project/package=caret.

Leong, F. T., Pearce, M., & Huang, J. L. (2013). Assessing scientist and practitioner orientations in industrial/organizational psychology. *Journal of Career Assessment, 21*(3), 452-468.

Levy, R., & Mislevy, R. J. (2016). *Bayesian psychometric modeling*. Chapman and Hall/CRC.

Ley, E., & Steel, M. F. (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics, 24*(4), 651-674.

Ley, E., & Steel, M. F. (2012). Mixtures of g-priors for Bayesian model averaging with economic applications. *Journal of Econometrics, 171*(2), 251-266.

Li, Y., & Clyde, M. A. (2018). Mixtures of g-priors in generalized linear models. *Journal of the American Statistical Association, 113*(524), 1828-1845.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association, 103*(481), 410-423.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18-22.

Liden, R. C., & Maslyn, J. M. (1998). Multidimensionality of leader-member exchange : An empirical assessment through scale development. *Journal of Management, 24*(1), 43–72.

Liu, J. X., Goryakin, Y., Maeda, A., Bruckner, T., & Scheffler, R. (2017). Global health workforce labor market projections for 2030. *Human Resources for Health, 15*(1), 1-12.

Madigan, D., & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association, 89*(428), 1535-1546.

Madigan, D., York, J., & Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 215-232.

Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR Analytics. The *International Journal of Human Resource Management, 28*(1), 3-26.

McCartney, S., Murphy, C., & Mccarthy, J. (2020). 21st century HR: a competency model for the emerging role of HR Analysts. *Personnel Review, 50*(6), 1495-1513.

McCullagh. P., & Nelder, J.A. (1989). *Generalized Linear Models*. Chapman and Hall

Miller, A. (2002). *Subset selection in regression*. Chapman and Hall/CRC.

Min, C. K., & Zellner, A. (1993). Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics, 56*(1-2), 89-118.

Montgomery, J. M., & Nyhan, B. (2010). Bayesian model averaging: Theoretical developments and practical applications. *Political Analysis, 18*(2), 245-270.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society, 18*(6), 275-285.

Neyman, J. (1977). Frequentist probability and frequentist statistics. *Synthese, 36*(1), 97-131.

Oswald, F. L., Behrend, T. S., Putka, D. J., & Sinar, E. (2020). Big data in industrial-organizational psychology and human resource management: Forward progress for organizational research and practice. *Annual Review of Organizational Psychology and Organizational Behavior, 7*, 505-533.

Ott-Holland, C. J., Shepherd, W. J., & Ryan, A. M. (2019). Examining wellness programs over time: Predicting participation and workplace outcomes. *Journal of Occupational Health Psychology, 24*(1), 163.

Pearce, J. A., Robbins, D. K., & Robinson, R. B. (1987). The impact of grand strategy and planning formality on financial performance. *Strategic Management Journal, 8*(2), 125–134.

Putka, D. J., Beatty, A. S., & Reeder, M. C. (2018). Modern prediction methods: New perspectives on a common problem. *Organizational Research Methods, 21*(3), 689-732.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association, 92*(437), 179-191.

Raju, N. S., Steinhaus, S. D., Edwards, J. E., & DeLessio, J. (1991). A logistic regression model for personnel selection. *Applied Psychological Measurement, 15*(2), 139-152.

Rokach, L. (2019). *Ensemble learning: Pattern classification using ensemble methods*. World Scientific.

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/.

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. *Journal of Applied Psychology, 104*(10), 1207-1225.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274.

Scott, J. G., & Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics, 38*(5), 2587-2619.

Sechidis, K., Tsoumakas, G., Vlahavas, I. (2011). On the stratification of multi-label data. In D. Gunopulos, T. Hofmann, D. Malerba, and M. Vazirgiannis (Eds.), *Machine learning and knowledge discovery in databases: Lecture notes in computer science*. Springer.

Snijders, T. A., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.

Society for Industrial Organizational Psychology. (2018). *Principles for the validation and use of personnel selection procedures* (5th ed.). American Psychological Association.

Somers, M. J. (1996). Modelling employee withdrawal behavior over time: A study of turnover using survival analysis. *Journal of Occupational and Organizational Psychology, 69*(4), 315-326.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583-639.

Stauffer, J., & Ree, M. J. (1996). Predicting with logistic or linear regression: Will it make a difference in who is selected for pilot training?. *The International Journal of Aviation Psychology, 6*(3), 233-240.

Steel, M. F. (2020). Model averaging and its use in economics. *Journal of Economic Literature, 58*(3), 644-719.

Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology, 26*(1), 1-9.

Tonidandel, S., LeBreton, J. M., & Johnson, J. W. (2009). Determining the statistical significance of relative weights. *Psychological Methods, 14*(4), 387.

Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *The Journal of the American Statistical Association, 81*, 82-86.

Tierney, L., Kass, R. E., & Kadane, J. B. (1989). Approximate marginal densities of non-linear functions. *Biometrika, 76*, 425-433.

Tsui, A. S., Pearce, J. L., Porter, L. W., & Tripoli, A. M. (1997). Alternative approaches to the employee-organization relationship: Does investment in employees pay off?. *Academy of Management Journal, 40*(5), 1089–1121.

Vaiman, V., Scullion, H., & Collings, D. (2012). Talent management decision making. *Management Decision, 50*(5), 925-941.

Valentine, S., Greller, M.M. and Richtermeyer, S.B. (2006). Employee job response as a function of ethical context and perceived organization support. *Journal of Business Research, 59*(5), 582-588

Vandenberghe, C. (1999). Organizational culture, person–culture fit, and turnover: A replication in the health care industry. *The International Journal of Industrial, Occupational, and Organizational Psychology and Behavior, 20*(2), 175-184.

Van Der Linde, A. (2005). DIC in variable selection. *Statistica Neerlandica, 59*(1), 45-56.

Van Veldhoven, M., & Dorenbosch, L. (2008). Age, proactivity and career development. *Career Development International, 13*(2), 112–131.

Vehtari, A., & Ojanen, J. (2012). A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys, 6*, 142-228.

Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research, 14*, 867-897.

Xu, J., & Thomas, H. C. (2011). How can leaders achieve high employee engagement?. *Leadership & Organization Development Journal, 32*(4), 399-416.

Yuan, S., Kroon, B., & Kramer, A. (2021). Building prediction models with grouped data: A case study on the prediction of turnover intention. *Human Resource Management Journal*, 1-19.

Zellner, A. (1983). Applications of Bayesian analysis in econometrics. *Journal of the Royal Statistical Society: Series D (The Statistician), 32*, 23-34.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In P. K. Goel & A. Zellner (Eds.), *Bayesian inference and decision techniques: Essays in honor of Bruno de Finetti* (pp. 233-240). Amsterdam.

Zellner, A. and A. Siow (1980). Posterior odds ratios for selected regression hypotheses (with discussion). In J. Bernardo, M. DeGroot, D. Lindley, & A. Smith (Eds.), *Bayesian statistics* (pp. 585-603). University Press.

Zeugner, S. (2011). *Bayesian model averaging with BMS*. http://bms.zeugner.eu/tutorials/

Appendix A: Complete Employee-Level Variables, Measures, and Descriptive Statistics

Measure	N items	α[1]	Scale	Reference	*M*	*SD*
Intention to leave	3	.888	A[2]	Valentine, Greller & Richtmeyer, 2006	2.19	.932
Gender (male)	1				57%	
Age	1		years		38.44	12.867
Educational level > Ba.	1				38%	
Management position	1		No/Yes		24%	
Hours/week - contract	1		hours		31.43	12.39
Contract type (1-2)	1		B[3]		83%	
Leader-Member Exchange	12	.889	A	Liden & Maslyn (1998)	3.74	.426
Information sharing	2	.686	C[4]	Kroon, et al. (2012)	3.52	1.291
Voice	5	.774	C	Kroon, et al. (2012)	3.05	.922
Pay satisfaction	3	.812	A	QEEW (2014)	3.37	.776
Perceived fairness	16	.941	A	Tsui, et al. (1997)	3.37	.616
Job proactivity	5	.889	A	Van Veldhoven & Dorenbosch (2008)	3.94	.603
Career opportunities	3	.891	A	QEEW (2014)	3.29	.918

1 α = Cronbach's alpha.
2 Scale A is a 5pt scale (completely disagree, disagree, neither, agree, strongly agree).
3 Scale B is a 4pt scale (1 = permanent employee, 2 = temporary with outlook on permanent, 3 = temporary with no outlook for permanent status, 4 = on call).
4 Scale C is a 5pt scale (1= not applicable to any employee, 2 = not applicable to most employees, 3 = applicable to some, 4 = applicable to most, 5 = applicable to all employees)

Appendix B: Complete Organization-Level Variables, Measures, and Descriptive Statistics

Measure	N items	α[1]	Scale	Reference	*M*	*SD*
Number of employees	1				31.27	40.02
Number of FTE current year	1		FTE		22.92	31.80
Number of FTE last year	1		FTE		20.67	28.06
Availability works council	1		No/Yes		12.69%	
Availability HR professional	1		No/Yes		39.90%	
Number of hierarchical levels	1				2.26	2.57
Number of departments	1				3.34	2.11
Number of managers	1				2.90	2.54
Family business	1		No/Yes		60.47%	
HRM: turnover	1		A[2]		2.11	0.96
HRM: attract well-qualified personnel	1		A		3.15	0.78
HRM: retain key employees	1		A		3.78	0.71
HRM: Absenteeism level	1		A		2.08	0.98
HRM: Amount of labor disputes.	1		A		1.52	0.69
HRM: quality of ideas and suggestions	1		A	Guest, & Peccei(2001)	3.2	0.74
HRM: Contribution to innovation	1		A		3.14	0.85
HRM: Involvement of employees.	1		A		4.03	0.66
HRM: Flexibility of employees	1		A		3.85	0.74
Availability of a supervisory board	1		No/Yes		5.05%	
Availability of an advisory board	1		No/Yes		8.67%	
Corporate entrepreneurship	7	.824	B[3]	Kellermanns & Eddleston (2006),	3.10	0.77
Strategic planning	13	.696	B	Pearce et al., (1987), Powell (1992)	3.54	0.41
Entrepreneurial orientation	9	.827	C[4]	Kroon et al., (2012)	4.60	0.94

[1] α = Cronbach's alpha.
[2] Scale A is a 5pt scale (very low, low, average, high, very high).
[3] Scale B is a 5pt scale (completely disagree, disagree, neither, agree, strongly agree)
[4] Scale C is a 7pt scale ranging from 1 (very low) to 7 (very high

Appendix C: MCMC Diagnostics

Figure 1.

BMA Regression Context: Posterior Distribution of the Shrinkage Factor

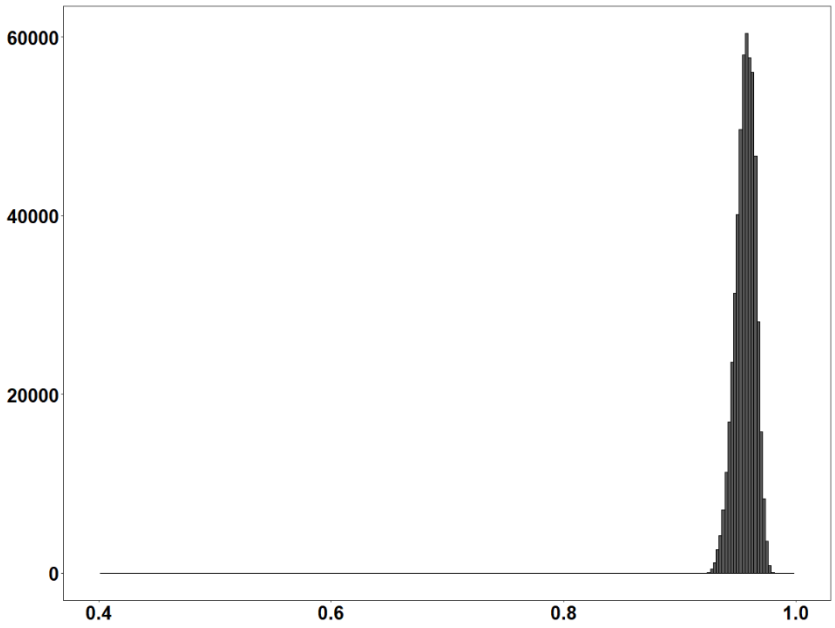
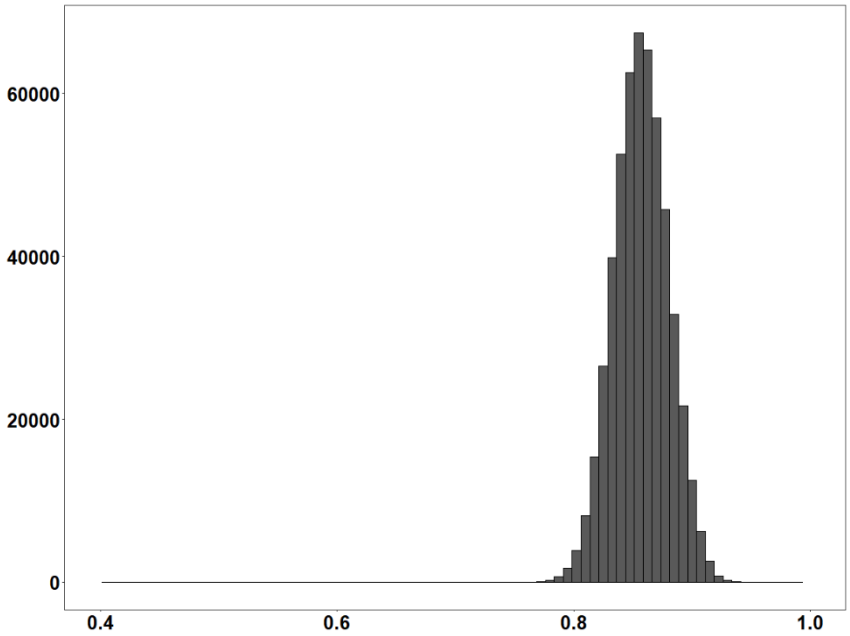


Figure 2.

BMA Classification Context: Posterior Distribution of the Shrinkage Factor



Appendix D: Syntax

```
# Syntax for Using BMA for Classification and/or Regression

################################################################################
# Housekeeping & Data Prep
################################################################################


# Packages

library(psych)

library(BAS)

library(tidyverse)

library(caret)

library(randomForest)

library(DescTools)

library(rwa)

################################################################################

# Data Prep (All variables except T/O intent mean-centered already)

dat1 <- read.csv("turnover_data.csv") # Load original dataset

# Remove company code and protected status variables (Age & Gender)

dat1 <- dat1[, -c(2, 3, 4)]

# Convert categorical/indicators to factors

dat1$manager <- factor(dat1$manager) # Manager Status (N/Y)
levels(dat1$manager) <- c("No", "Yes")

dat1$contype <- factor(dat1$contype) # Contract Type (Perm/Temp)
levels(dat1$contype) <- c("Temp", "Perm")

dat1$employee_council <- factor(dat1$employee_council) # Employee council
(N/Y)
levels(dat1$employee_council) <- c("No", "Yes")

dat1$hr_rep <- factor(dat1$hr_rep) # HRM on-site (N/Y)
levels(dat1$hr_rep) <- c("No", "Yes")

dat1$familybiz <- factor(dat1$familybiz) # Family Business (N/Y)
levels(dat1$familybiz) <- c("No", "Yes")

dat1$superboard <- factor(dat1$superboard) # Supervisory Board (N/Y)
levels(dat1$superboard) <- c("No", "Yes")
```

```
dat1$adv_board <- factor(dat1$adv_board) # Advisory Board (N/Y)
levels(dat1$adv_board) <- c("No", "Yes")

################################################################################

# Because the train/test partitions are stratified with respect to the
# outcome measures, we need to make 2 different full datasets prior to
# train/test partitioning.

# The first full dataset will have a continuous outcome measure, and the
# second full dataset will have a dichotomous outcome measure.

# This way we can ensure that the train/test partitions will be stratified
# correctly according to their respective outcome measure


dat2 <- dat1 # Make 2nd full dataset

dat2$TI_risk <- ifelse(dat2$TI > 3, 1, 0) # Dichotomous outcome measure

dat2$TI_risk <- factor(dat2$TI_risk) # Make factor

dat2 <- dat2 %>% relocate(TI_risk, .before = TI) # Relocate

dat2 <- dat2[-2] # Delete continuous outcome from this 2nd  dataset

# Double-checking datasets 1 & 2 are correct

View(dat1) # View dataset 1

View(dat2) # View dataset 2

str(dat1) # Check Variable descriptions/structure for datasets 1 & 2

str(dat2)

################################################################################

# Listwise Deletion for both datasets

dat1<-na.omit(dat1)

dat2<-na.omit(dat2)

nrow(dat1) # Check 1

nrow(dat2) # Check 2

################################################################################

# Partitions full datasets 1 & 2 into training and testing sets.

# We will end up with a training and test set for dataset1 (continuous
# outcome),
# and a training and test set for dataset2 (classification outcome)
```

```
set.seed(300)

# Partitions via 70/30 train/test split

# First train/test set

indxTrain1 <- createDataPartition(y = dat1$TI, p = 0.7, list = FALSE) # index

training1 <- dat1[indxTrain1, ] # training set 1

testing1 <- dat1[-indxTrain1, ] # testing set 1

# Second train/test set

indxTrain2 <- createDataPartition(y = dat2$TI_risk, p = 0.7, list = FALSE)

training2 <- dat2[indxTrain2, ] # training set 2

testing2 <- dat2[-indxTrain2, ] # testing set 2

#################################################################

# Check outcome distributions are correctly stratified

# Continuous outcome distributions

summary(training1$TI) # training1

summary(testing1$TI) # testing1

summary(dat1$TI) # overall

# Classification outcome distributions

prop.table(table(training2$TI_risk)) * 100 # Outcome balance for training2

prop.table(table(testing2$TI_risk)) * 100 # Outcome balance for testing2

prop.table(table(dat2$TI_risk)) * 100 # full/non-partioned dataset2

#################################################################################
#################################################################################

# BMA Case Study 1 (Continuous Outcome Measure via Normal Linear Regression)

# 'prior' sets the regression coefficient prior to a hyper-g coefficient
# prior
# with Laplace approximation for the integration over 'g'.

# 'method' is set to "MCMC" to specify a MCMC sampling routine should be used

# "MCMC.iterations" specifies the number of iterations.

# "Force.heredity" requires the factor levels to be kept together
```

```
#  "na.action=omit" is the default for missing data

#  "model prior" defaults to Beta-Binomial (1, 1) on predictor inclusion # #
  prob.

# "Renormalization" is specified in order to calculate posterior model
# probabilities
# exactly after model space has been sampled, rather than used frequency-
# based approximations
# (done when data is noisy and convergence of MCMC routine is poor).



# Model Specification

BMA1 <- bas.lm(TI ~ edu + manager + conhour + contype + lmx + infshare +
voice +
                 paysatis + fair + proact + caropp + N + FTEN + FTELY +
employee_council +
                 hr_rep + n_levels + n_deps + n_magers + familybiz + HR1 +
HR2 +
                 HR3 + HR4 + HR5 + HR6 + HR7 + HR8 + HR9 + superboard +
adv_board +
                 corp_entr + stratplan + entr_orient,
  data = training1, prior = "hyper-g-laplace", alpha = 3, method = "MCMC",
  MCMC.iterations = 50000000, force.heredity = TRUE, renormalize = TRUE
)


################################################################################

# MCMC Convergence Diagnostics (for use if "renormalized=FALSE" was used)



# MCMC approximated predictor inclusion probs vs calculated predictor
# inclusion probs

diagnostics(BMA1, type = "pip", pch = 16)



# MCMC approximated posterior model probabilities vs calculated posterior
# model probs

diagnostics(BMA1, type = "model", pch = 16)



# Number of Unique Models Sampled/Explored

BMA1$n.Unique


################################################################################

# Results & Diagnostics
```

```
# initial results summary

summary(BMA1)



# Plots -> (1) Resids vs Fitted, (2) Cumulative Model Probs,
# (3) Marginal Likelihood vs complexity, Marginal Predictor inclusion probs

plot(BMA1, ask = F) # Plots



# Visualization of Posterior Model Probabilities and Variable Inclusion.
# Models with indistinguishable log posterior odds have the same color.

image(BMA1, rotate = F)


# Regression Coefficients

BMA1coefs <- coef(BMA1, digits = 2)

BMA1coefs


# 95% Credible Interval (Highest Posterior Density) for Coefficients

confint(BMA1coefs)


# Plot of 95% Credible Intervals (HPD) from above

# "parm" argument can specify specific predictors (helpful with large #)


plot(confint(BMA1coefs, parm = 1:8))
plot(confint(BMA1coefs, parm = 9:15))
plot(confint(BMA1coefs, parm = 16:22))
plot(confint(BMA1coefs, parm = 23:29))
plot(confint(BMA1coefs, parm = 30:35))


# Marginal Posterior Distributions for coefficients.

# The vertical bar is the posterior probability that the coefficient is 0

# bell shaped curve represents the density of plausible values from all the
# models where the coefficient is non-zero.

# This is scaled so that the density height for non-zero values is the
# probability
# that the coefficient is non-zero
```

```
plot(BMA1coefs, ask = F) # subset argument can be used to specify specific
predictors


# Histogram of shrinkage/regularization term sample frequencies (g/1+g)
hist(BMA1$shrinkage)

##############################################################################

# Fitted Values for current data and Prediction of new data


# fitted values for current data

fitted_BMA1 <- fitted(BMA1, estimator = "BMA")


# prediction for new data (testing set 1)

predicted_BMA1 <- predict(BMA1, newdata = testing1, estimator = "BMA")


# View Available attributes

names(predicted_BMA1)


# Predicted Outcome Measure for newdata

predY_BMA1 <- predicted_BMA1$fit

head(predY_BMA1)


# Predictive intervals for predictions (can't get this to work due to vector
# size)

BMA1.pred <- predict(BMA1, estimator = "BMA", predict = FALSE, se.fit = TRUE)

confint(BMA1.pred)


##############################################################################


# Evaluation of BMA1 Predictive Performance (RMSE, R^2, MAE)


postResample(predY_BMA1, obs = testing1$TI)


##############################################################################
##############################################################################


# Full OLS model Case Study 1 (Continuous Turnover Intent Outcome Measure)
```

```
# Fit OLS model using training dataset 1

olsm1 <- lm(TI ~ edu + manager + conhour + contype + lmx + infshare + voice +
              paysatis + fair + proact + caropp + N + FTEN + FTELY +
employee_council +
              hr_rep + n_levels + n_deps + n_magers + familybiz + HR1 + HR2 +
              HR3 + HR4 + HR5 + HR6 + HR7 + HR8 + HR9 + superboard +
adv_board +
              corp_entr + stratplan + entr_orient, data = training1)


# Model Summary

summary(olsm1)

# 95% CI for Regression Coefficients (Standardized)

confint(olsm1)    # Unstandardized
effectsize(olsm1) # Standardized

##############################################################################

# OLS Model Raw Relative Weights Analysis (Predictor "Importance")

training1b <- training1

levels(training1b$manager) <- c(0, 1)
training1b$manager <- as.numeric(training1b$manager)

levels(training1b$contype) <- c(0, 1)
training1b$contype <- as.numeric(training1b$contype)

levels(training1b$employee_council) <- c(0, 1)
training1b$employee_council <- as.numeric(training1b$employee_council)

levels(training1b$hr_rep) <- c(0, 1)
training1b$hr_rep <- as.numeric(training1b$hr_rep)

levels(training1b$familybiz) <- c(0, 1)
training1b$familybiz <- as.numeric(training1b$familybiz)

levels(training1b$superboard) <- c(0, 1)
training1b$superboard <- as.numeric(training1b$superboard)

levels(training1b$adv_board) <- c(0, 1)
training1b$adv_board <- as.numeric(training1b$adv_board)

rwa(df = training1b,
    outcome = "TI",
    predictors = c("edu","manager","conhour","contype","lmx", "infshare",
"voice","paysatis","fair","proact", "caropp", "N",
"FTEN", "FTELY", "employee_council", "hr_rep", "n_levels",
"n_deps", "n_magers", "familybiz", "HR1", "HR2", "HR3",
"HR4", "HR5", "HR6", "HR7", "HR8", "HR9", "superboard",
"adv_board", "corp_entr", "stratplan", "entr_orient"))
```

```
#################################################################

# OLS Model Predictions for new data (testing set 1)

predY_OLS1 <- predict(olsm1, newdata = testing1)

head(predY_OLS1)


# Evaluation of OLS Predictive Performance (BMA wins!)

postResample(predY_OLS1, obs = testing1$TI)

#################################################################
#################################################################

# Random Forest Regression Case Study 1 (Continuous Outcome Measure of T/O
# Intent)


set.seed(300)

# Set training control to use 10-fold Cross Validation

ctrl1 <- trainControl(method = "repeatedcv", repeats = 10)


# Train/Fit Random Forest Regression model to training dataset 1

rf1 <- train(TI ~ edu + manager + conhour + contype + lmx + infshare + voice
+ paysatis + fair + proact + caropp + N + FTEN + FTELY + employee_council +
hr_rep + n_levels + n_deps + n_magers + familybiz + HR1 + HR2 + HR3 + HR4 +
HR5 + HR6 + HR7 + HR8 + HR9 + superboard + adv_board + corp_entr + stratplan
+ entr_orient, data = training1, trControl = ctrl1, method = "rf", na.action
= na.omit)



# Predictions for new data (testing set 1)

predY_RF1 <- predict(rf1, newdata = testing1)

head(predY_RF1)


# Evaluation of RF Regression Predictive Performance (RMSE, R^2, MAE)  #BMA
does better!

postResample(predY_RF1, obs = testing1$TI)


# Variable Importance for RF

varImp(rf1)

#################################################################
```

```
#################################################################

# BMA Case Study 2 (Binary Turnover Risk Outcome via Logistic Regression)

# "family" sets binomial logistic regression likelihood family

# 'betaprior' sets the regression coefficient prior to a CCH(1, 2, 0)
# coefficient prior
#  equivalent to the hyper-g (alpha = 3) prior used for regression contexts.

# "modelprior" sets beta.binomial(1, 1) prior over predictor inclusion prob.

# 'method' is set to "MCMC" to specify a MCMC sampling routine should be used

# "MCMC.iterations" specifies the number of iterations.

# "Force.heredity" requires the factor levels to be kept together

#  "na.action=omit" is the default for missing data

# "laplace" = TRUE/FALSE specifies Laplace integration/Cephes estimation for
# marginal likelihood

# "Renormalization" =TRUE/FALSE specifies whether posterior model probs are
# calculated

# using renormalization or approximated based on sampling frequencies.



# Model Specification

BMA2 <- bas.glm(TI_risk ~ edu + manager + conhour + contype + lmx + infshare
+ voice + paysatis + fair + proact + caropp + N + FTEN + FTELY +
employee_council + hr_rep + n_levels + n_deps + n_magers + familybiz + HR1 +
HR2 + HR3 + HR4 + HR5 + HR6 + HR7 + HR8 + HR9 + superboard + adv_board +
corp_entr + stratplan + entr_orient, data = training2, family = binomial(link
= "logit"), betaprior = CCH(1, 2, 0), modelprior = beta.binomial(1,1), method
= "MCMC", MCMC.iterations = 50000000, force.heredity = TRUE, renormalize =
FALSE)

#################################################################################

# MCMC Convergence Diagnostics (for use if "renormalized=FALSE" was used)


# MCMC approximated predictor inclusion probs vs calculated predictor
# inclusion probs

diagnostics(BMA2, type = "pip", pch = 16)


# MCMC approximated posterior model probabilities vs calculated posterior
model probs

diagnostics(BMA2, type = "model", pch = 16)
```

```
# Number of Unique Models Sampled/Explored

BMA2$n.Unique


################################################################################
# Results & Diagnostics


# initial results summary

summary(BMA2)



# Plots -> (1) Resids vs Fitted, (2) Cumulative Model Probs,
# (3) Marginal Likelihood vs complexity, Marginal Predictor inclusion probs

plot(BMA2, ask = F) # Plots



# Visualization of Posterior Model Probabilities and Variable Inclusion.
# Models with indistinguishable log posterior odds have the same color.

image(BMA2, rotate = F)


# Regression Coefficients

BMA2coefs <- coef(BMA2)

BMA2coefs


# 95% Credible Interval (Highest Posterior Density) for Coefficients

confint(BMA2coefs)


# Plot of 95% Credible Intervals (HPD) from above

# "parm" argument can specify specific predictors (helpful with large #)


plot(confint(BMA2coefs, parm = 1:8))
plot(confint(BMA2coefs, parm = 9:15))
plot(confint(BMA2coefs, parm = 16:22))
plot(confint(BMA2coefs, parm = 23:29))
plot(confint(BMA2coefs, parm = 30:35))


# Marginal Posterior Distributions for coefficients.

# The vertical bar is the posterior probability that the coefficient is 0
```

```
# bell shaped curve represents the density of plausible values from all the
# models
# where the coefficient is non-zero.

# This is scaled so that the density height for non-zero values is the
# probability
# that the coefficient is non-zero

plot(BMA2coefs, ask = F) # subset argument can be used to specify specific
predictors


# Histogram of shrinkage/regularization term sample frequencies (g/1+g)
hist(BMA2$shrinkage)


################################################################################

# Fitted Values for current data and Prediction of new data


# fitted values for current data

fitted_BMA2 <- fitted(BMA2, estimator = "BMA")


# prediction for new data (testing set 2)

predicted_BMA2 <- predict(BMA2, newdata = testing2, estimator = "BMA",
type="response")


# View Available attributes

names(predicted_BMA2)

# Ybma = linear predictor scale predictions (log odds)

# fit = response scale predictions (probability of class membership)

# Predicted probability of class membership  for new data

predY_BMA2 <- predicted_BMA2$fit


# Predictive intervals for predictions (can't get this to work due to vector
size)

BMA2.pred <- predict(BMA2, estimator = "BMA", predict = FALSE, se.fit = TRUE)

confint(BMA2.pred)


# Convert predicted probabilities to classifications

pred_class_BMA2<- ifelse(predY_BMA2 >= 0.5, 1, 0)
```

```
pred_class_BMA2 <- factor(pred_class_BMA2)


# Evaluation of BMA2 Predictive Performance


postResample(pred_class_BMA2, obs = testing2$TI_risk) # Accuracy/Kappa

BrierScore(as.numeric(testing2$TI_risk), predY_BMA2)  # Brier Score

confusionMatrix(pred_class_BMA2, testing2$TI_risk)    # Confusion Matrix

##############################################################################
##############################################################################


# Full Logistic Regression model Case Study 2 (Binary Turnover Risk Outcome)


# Fit Logistic Regression model using training dataset 2

logregm2 <- glm(TI_risk ~ edu + manager + conhour + contype + lmx + infshare
+ voice + paysatis + fair + proact + caropp + N + FTEN + FTELY +
employee_council + hr_rep + n_levels + n_deps + n_magers + familybiz + HR1 +
HR2 + HR3 + HR4 + HR5 + HR6 + HR7 + HR8 + HR9 + superboard + adv_board +
corp_entr + stratplan + entr_orient, data = training2, family=binomial())


# Model Summary

summary(logregm2)

# 95% CI for Regression Coefficients

confint(logregm2)    # Unstandardized
effectsize(logregm2) # Standardized

##############################################################################

# Logistic Regression Model Raw Relative Weights Analysis (Predictor
"Importance")

rwa(df = training1b,
    outcome = "TI_risk",
 predictors = c("edu","manager","conhour","contype","lmx", "infshare",
"voice","paysatis","fair","proact", "caropp", "N",
  "FTEN", "FTELY", "employee_council", "hr_rep", "n_levels",
 "n_deps", "n_magers", "familybiz", "HR1", "HR2", "HR3",
"HR4", "HR5", "HR6", "HR7", "HR8", "HR9", "superboard",
"adv_board", "corp_entr", "stratplan", "entr_orient"))

##############################################################################

# Predicted Probabilities for new data (testing set 2)

predY_logreg2 <- predict(logregm2, newdata = testing2, type = "response")
```

```
head(predY_logreg2)

# Convert Predicted Probabilities to Classifications

pred_class_logreg2<- ifelse(predY_logreg2 >= 0.5, 1, 0)

pred_class_logreg2 <- factor(pred_class_logreg2)

# Evaluation of LogReg Predictive Performance

postResample(pred_class_logreg2, obs = testing2$TI_risk) # Accuracy/Kappa

BrierScore(as.numeric(testing2$TI_risk), predY_logreg2)  # Brier Score

confusionMatrix(pred_class_logreg2, testing2$TI_risk)    # Confusion Matrix

##############################################################################
##############################################################################

# Random Forest Classifier Case Study 2 (Binary Turnover Risk Outcome)

set.seed(300)

# Set training control to use 10-fold Cross Validation

ctrl1 <- trainControl(method = "repeatedcv", repeats = 10)

# Train/Fit Random Forest Regression model to training dataset 1

rf2 <- train(TI_risk ~ edu + manager + conhour + contype + lmx + infshare +
voice + paysatis + fair + proact + caropp + N + FTEN + FTELY +
employee_council + hr_rep + n_levels + n_deps + n_magers + familybiz + HR1 +
HR2 + HR3 + HR4 + HR5 + HR6 + HR7 + HR8 + HR9 + superboard + adv_board +
corp_entr + stratplan + entr_orient, data = training2, trControl = ctrl1,
method = "rf", na.action = na.omit)

# Predictions for new data (testing set 2)

predY_RF2 <- predict(rf2, newdata = testing2, type = "prob")

head(predY_RF2)

# Convert Predicted Probabilities to Classifications

pred_class_RF2<- ifelse(predY_RF2[,2] >= 0.5, 1, 0)

pred_class_RF2<-factor(pred_class_RF2)
```

```
# Predictor Variable Importance

varImp(rf2)

# Evaluation of RF Regression Predictive Performance

postResample(pred_class_RF2, obs = testing2$TI_risk) # Accuracy/Kappa

BrierScore(as.numeric(testing2$TI_risk), predY_RF2[,2])  # Brier Score

confusionMatrix(pred_class_RF2, testing2$TI_risk)    # Confusion Matrix

#################################################################
#################################################################
```