Bayesian Statistics in Management Research: Theory, Applications, and Opportunities

Tyson B. Mackey, Management, Clarkson University and Jeffrey P. Dotson, Marketing, Ohio State University

https://doi.org/10.1093/acrefore/9780190224851.013.287
Published online: 21 August 2024

Summary

Bayesian statistics deserve a more prominent place in strategic management methodology. Bayesian methodology shines in its ability to both model heterogeneity and simultaneously model endogenously chosen strategies and heterogeneity in the effect of a strategy on firm-level outcomes. Moreover, within the broader class of Bayesian methods, some methodologies present solutions to a pernicious problem—that of *essential heterogeneity*, in which typical methods of accounting for endogenously chosen strategies, e.g., instrumental variables, treatment effects models, or random-coefficient models, fail when the effects of strategies on outcomes are heterogeneous across firms. Research about Bayesian inference, including discussions of the theoretical underpinnings of Bayesian statistics and how to implement an empirical workflow using Bayesian statistics, can be useful for building bespoke models to fit theoretical questions pertaining to management research.

Keywords: Bayesian inference, causal inference, endogeneity, heterogeneity, resource-based theory

Subjects: Research Methods

Introduction

> The limits of my language mean the limits of my mind.
>
> —Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*

Language affects thought. In statistical analysis, the set of tools available to the researcher constitutes the language that drives the researchers' empirical hypotheses. This article provides a general overview of Bayesian statistics and its past and potential future impact on the field of management. Bayesian methods can provide management scholars with a new language that will allow them to engage in more rigorous, theory-based research, including hypothesis development and statistical testing. Bayesian methods enable the researcher to write down and estimate (potentially complex) models that reflect their beliefs about the data-generating process (i.e., how agents make decisions and how those decisions affect firm value). They allow researchers to move beyond the old adage of "If the only tool you have is a hammer, every problem looks like a nail" by equipping them with a toolbox of solutions that can be adapted to address any scenario they encounter.

As an illustration, consider the treatment of endogenous decision-making. Theory in strategic management has long understood that strategies are chosen endogenously based on the expected profitability of a strategy, yet empirical work for endogenous dichotomous strategies generally did not hypothesize along these lines until methodology caught up to theory (Hamilton & Nickerson, 2003; Shaver, 1998; Wolfolds & Siegel, 2019). While firm-level heterogeneity has always been a core construct in strategic management theory (Rumelt et al., 1995), empirical work modeling firm-level heterogeneity in strategy has been mostly limited to use in panel data models and interaction effects. As of the early 21st century, work has begun to model firm-specific heterogeneity through random-coefficient models (RCMs) (Alcácer et al., 2018). These models generate firm-specific point estimates, which can be aggregated into a distribution in order to present a major improvement in the ability of researchers to make empirical hypotheses that are consistent with strategic management theory. As Alcácer et al. (2018) explain,

> Note that identifying heterogeneous effects and understanding the sources of heterogeneous effects are two distinct and important objectives to a strategy researcher. Whereas identifying the presence of heterogeneous effects is the initial contribution of RCMs, researchers can then use distributional estimates as a jumping-off point to explore the reasons behind firm heterogeneity. (p. 538)

However, while multilevel RCMs are a major improvement over models that assume constant effects across primitive units of analysis (e.g., individuals, firms), multilevel Bayesian hierarchical models are a superior method of accounting for variation at the unit level (i.e., lower level) because they not only allow for a distribution across unit-level effects (as the frequentist RCMs do) but also estimate a distribution for each unit-level effect.

In addition to having superior multilevel modeling, Bayesian models are superior in their ability to estimate phenomena that are subject to endogeneity, which plagues most research in strategic management; it vastly expands the ability of scholars to conceive of and implement empirical studies in ways that are consistent with their conceptual theories.

The remainder of this article proceeds as follows: The "Literature Review" section provides an overview of Bayesian inference. The section "Primer on Bayesian Inference" summarizes the literature in management that has used Bayesian methods. The section "Summary and Next Steps" concludes the article by discussing opportunities for the application of Bayesian methods in management research.

Literature Review

While Bayesian methods have long since revolutionized modeling in marketing research (Rossi & Allenby, 2003), Bayesian methods are still nascent in management research, even though they are especially well suited for studying management phenomena. A core reason for the paucity of Bayesian research in management is that the evangelists of Bayes tend to emphasize the philosophical benefits of Bayesian statistics over frequentist statistics, to the neglect of preaching the practical benefits of Bayesian statistics. Outside of the management literature, scholarship in statistics has long since moved past the Bayesian-versus-frequentist philosophical debate. Writing in a *Journal of Management* special issue about Bayesian methods, statistician Andrew Gelman (2015) focuses fully on the problem of assuming constant treatment effects across units and considers it a main culprit of the replicability crisis. The solution he poses is to use Bayesian multilevel models, which will allow effects to vary across units of analysis (e.g., individuals).

Within the management literature, there are some notable works that stand out, either as a proof of concept for the usefulness of multilevel Bayesian models for management research or for advancing knowledge about a particular phenomenon. Work by Certo et al. (2024) and Kruschke et al. (2012) provides arguments in favor of a Bayesian approach to the study of organizations and management.

In probably the most important early work, Hansen et al. (2004) use hierarchical Bayesian (HB) methods to estimate unit-level heterogeneity in the outcomes of a variety of firm-level choices and argue that HB is very well suited for the study of resource-based theory. Their article evinces the insight that while empirical models estimate the average effect between resources and firm performance, resource-based theory is fundamentally a theory of superior performance or competitive advantage, and thus empiricists need methods that can estimate firm-level effects. They note,

> A positive association between a resource and performance says nothing about superior economic performance or competitive advantage. Such a finding does suggest that firms without that resource may be at a disadvantage, but one cannot conclude that possessing that resource confers a competitive advantage. These characteristics of traditional statistical approaches lead to the conclusion that there is an important lack of congruency between resource-based view (RBV) theory and regression-type analysis.
>
> (Hansen et al., 2004, p. 1283)

Rather than extolling the philosophical benefits of Bayesian over frequentist statistics, Hansen et al. (2004) show the practical benefits of Bayesian analysis in estimating a model that is more consistent with strategic management theory—a model that cannot easily be estimated with frequentist methods. Instead of point estimates, Hansen et al. (2004) report graphical depictions of the distributions of parameters; instead of *p*-values, their article reports the percent of the distribution that is greater than zero, which in the context of the present article can be interpreted as the probability that a decision will have a positive outcome for a firm.

Hahn and Doh (2006) build upon Hansen et al. (2004) and add to the literature by noting that Bayesian methods, which can estimate unit-level effects, can be used in a variety of theoretical lenses beyond resource-based theory—in this case, dynamic capabilities and industry evolution (Hahn & Doh, 2006). They illustrate this by using a Bayesian model to estimate a model of entry order.

Hahn and Doh (2006) also argue for the utility of Bayesian methods when the sample size is limited (as Bayesian models need not appeal to the central limit theorem). They mention that Bayesian models are superior when the distribution of the data is asymmetric or non-normal in other ways, as asymmetry can lead to incorrect standard errors and, hence, confidence intervals. This work also argues that Bayesian methods for handling the bias caused by missing data are superior to the expectation-maximization (EM) algorithm for data augmentation, as the EM algorithm can only give a point estimate for the missing data while Bayes can account for the uncertainty of the estimated data.

Bayesian data augmentation underlies Denrell et al. (2013) in their approach to modeling firm capabilities and their effect on superior performance. Since firm capabilities are not observed directly, they estimate capabilities based on the degree to which industry performance is based on previous performance. Their article shows how a Bayesian model of these capabilities is far superior to maximum likelihood estimation.

While HB methods allow for the estimation of unit-level effects, using a Bayesian methodology does not absolve the researcher of the need to deal with the endogeneity problems that plague most empirical strategy research. Nandialath et al. (2014) acknowledge this and build upon Hansen et al. (2004) by accounting for endogeneity in their test of resource-based theory. They note that typical solutions for the endogeneity problem, instrumental variable (IV) or treatment effects models (for a review of these methods, see Hamilton & Nickerson, 2003), may not be well suited to solving endogeneity in strategy. This is for two reasons. First, the endogenous choices found in strategy data also have unit-level heterogeneity. Second, the manager's heterogeneous skill in resource picking is likely to be correlated with the heterogeneous unit-level effects. This second condition may be surprising, as it is assumed that the whole point of IV or treatment effects models is to account for managerial choice. However, Nandialath et al. (2014) show that a valid instrument in the IV or treatment effects model requires that the managerial skill must be uncorrelated with the strategic choice, which defeats the purpose of firms hiring managers to make strategic decisions.

This problem, dubbed *essential heterogeneity* by Heckman et al. (2006), is a data problem and not a problem specific to Bayesian methodology. However, Nandialath et al. (2014) demonstrate that HB methods are very well suited to handling the problem of essential heterogeneity. Rather than using IV or treatment effects models, their article uses a structural model (also denoted as a full information model) to jointly estimate the resource-picking decision and the value of the choice of resource.

Following the approach of Nandialath et al. (2014), Mackey et al. (2017) use hierarchical Bayes methods to take on a well-worn empirical debate in strategy and finance about the value of corporate diversification, often referred to as the *diversification discount* literature. In this line of research, early researchers found a negative coefficient for the diversification-performance relationship, which implied that diversified firms tend to have lower valuations than a portfolio of single-business firms does (Berger & Ofek, 1995; Lang & Stulz, 1994). Later work (Campa & Kedia, 2002; Miller, 2006; Villalonga, 2004) used IV models, treatment effects models, or matching models to find a positive coefficient for the diversification-performance relationship, which implied that diversified firms tend to have higher valuations than a portfolio of single-business firms does.

Mackey et al. (2017) account for both firm-level heterogeneity in the value of diversification and for the essential endogeneity in the decision about whether to diversify. Using an HB method to estimate the distribution of the value of diversification, they find that there is neither a broad diversification discount nor a diversification premium; rather, diversified firms tend to create value from diversification, but focused firms are not likely to create value by diversifying. In other words, firms tend to be rational in their diversification decisions.

The intriguing part of the work of Mackey et al. (2017) is not the unsurprising finding that firms tend to be rational in their diversification choices—this is exactly what economic theory would predict—but that previous empirical work did not reach similar conclusions. Previous work could not reach these conclusions because the methodology did not allow for it. One is left to wonder how many other core debates in management research would find that managers tend to be rational in their strategic decision-making.

Given the importance of firm-level heterogeneity, strategy researchers are beginning to use frequentist RCMs to estimate firm-specific parameters (for an excellent discussion, see Alcácer et al., 2018). However, frequentist implementations of the hierarchical model run into some problems that are handled more adroitly by the Bayesian implementations. For example, estimating an RCM with maximum likelihood estimation can very often result in flat regions in the likelihood, and this can prevent convergence in the model. This may force researchers to make concessions in their theory (e.g., reducing the number of parameters that are allowed to vary by unit) for the sake of convergence. Bayesian models, however, estimate the entire likelihood instead of looking for the maximum of the likelihood function.

As previously mentioned in this section regarding Nandialath et al. (2014), strategy models are fraught with essential heterogeneity. This creates problems for the use of RCMs in strategy research. Alcácer et al. (2018) note that a key assumption of RCMs is “that the firm-specific coefficient deviations are uncorrelated with the regressors, that is, $\text{cov}(u_i, X_{it}) = 0$. When [this assumption] is violated, RCMs do not provide consistent estimates of the model coefficients” (p. 542).

Another contribution of the HB model in Mackey et al. (2017) is that it uses a structural model much like Nandialath et al. (2014) use to deal with the endogeneity in the diversification choice and with heterogeneity in the value of diversification. Oldroyd et al. (2019) use a similar setup to estimate the effect of the endogenous choice of effort levels in codifying knowledge and its heterogeneous impact on project-level performance.

Wibbens (2019) uses an HB model to revisit previous research about the persistence of superior profits. The Bayesian aspect adds much to the time-series aspect of the work. Specifically, it allows an autoregressive term to be greater than 1 for a period of time, which enables the finding that competitive advantage persists for much longer periods of time than was previously thought in the literature.

A Primer on Bayesian Inference

Bayesian analysis, rooted in the work of the 18th-century statistician Thomas Bayes, has experienced a revival in recent decades, finding its way into diverse academic disciplines. This approach's allure lies in its ability to incorporate prior knowledge into statistical models, with the three core components being the prior distribution, the likelihood, and the posterior distribution. The prior distribution encapsulates preexisting knowledge about a parameter, the likelihood evaluates the model's fit to the observed data, and the posterior distribution, resulting from a synthesis of the two, reflects updated beliefs after data observation.

While the Bayesian philosophy is conceptually straightforward, its real-world application can be intricate, especially when dealing with complex models. Advanced computational techniques, such as Markov chain Monte Carlo (MCMC), have emerged as transformative tools, bridging the gap between theoretical models and tangible insights. These methods serve as the backbone of modern Bayesian studies by ensuring that models be methodologically robust, transparent, and reproducible.

Complementing the theoretical advancements are a myriad of software solutions tailored to Bayesian research. Specialized platforms like Stan, JAGS, BRMS, and PyMC equip the contemporary Bayesian researcher with an arsenal of tools to actualize their models. Delving into the nuances of Bayesian analysis can demystify its complexities and offer scholars a comprehensive toolkit for harnessing its potential by blending depth with application and laying out opportunities for applications in management.

The remainder of this section revolves around the following concepts:

1. Bayesian inference
2. Posterior distribution using modern Bayesian computational techniques
3. The interplay of priors, likelihoods, and posteriors by crafting a Bayesian workflow
4. Software tools for engaging Bayesian inference.

An Overview of Bayesian Inference

At the heart of any Bayesian study lies the posterior distribution. A commonly used rendition of Bayes's theorem posits that the posterior distribution is proportional to the product of the likelihood and the prior and is commonly expressed as

$$\text{posterior distribution} \propto \text{likelihood} \times \text{prior distribution} \tag{1}$$

Mathematically, this is often expressed more formally as

$$\pi(\vartheta|\text{data}) \propto \pi(\text{data}|\vartheta) \times \pi(\vartheta) \quad (2)$$

This relation encapsulates the building blocks of Bayesian inference, whereby

- Likelihood = π (data|θ): This represents the probability of observing data given the model parameters. It is sometimes useful to think of the likelihood as an un-normalized probability density.
- Prior distribution = π (θ): This depicts initial beliefs about the model parameters, before data observation.
- Posterior distribution = π (θ|data): This reflects beliefs that are updated after data observation.

As an illustration, consider the following example in which both the likelihood and prior are normally distributed as illustrated in Figure 1. As a result of a mathematical relationship called conjugacy, their product also results in a normal distribution, forming the posterior. If the initial strong belief (prior) is somewhat contradictory to what the data suggests (likelihood), the posterior will be an amalgamation of these two, balancing the prior beliefs with the new data insights. The posterior distribution is, quite literally, an information weighted amalgamation of both the prior distribution and likelihood.

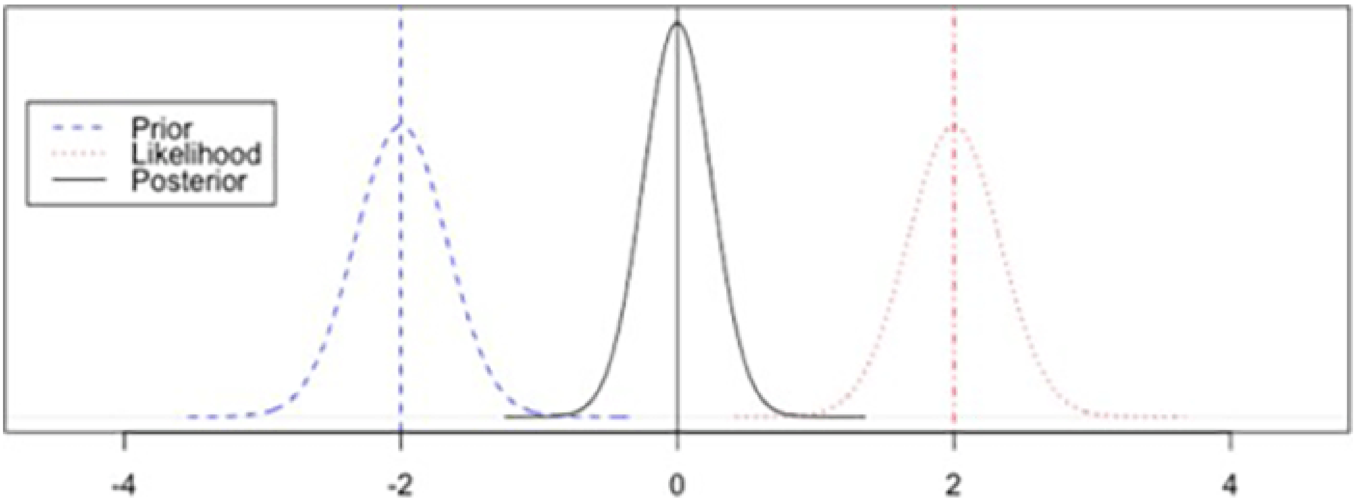

Figure 1. Illustration of combining a normal prior and normal likelihood to generate a normal posterior.

A central result from Bayesian statistics is that a posterior distribution is guaranteed to exist, provided that one begins with a proper prior distribution (i.e., one that integrates to one over its range) and a well-defined likelihood. The methodologies used to learn about the posterior, however, can vary. This is a topic discussed further in the "Modern Bayesian Computational Techniques" section.

In essence, Bayesian inference aims to understand the posterior distribution. Once the form of this distribution (either analytically or numerically) is known, statistical inference becomes straightforward. The posterior distribution can be used to calculate point estimates, variance estimates, intervals, and more with ease. The subsequent sections will further expand on these concepts and the tools facilitating Bayesian inference.

Modern Bayesian Computational Techniques

Thus far in this article, core concepts from the world of Bayesian inference—including prior distributions, likelihoods, and, most importantly, posterior distributions—have been introduced and discussed. A consistent theme has been the posterior distribution, which stands as the bedrock of inference for any Bayesian study. It exists, but how is its form deciphered? One traditional method for learning about the posterior distribution is through the use of conjugate distributions. Thoughtful model specification and prior selection can enable the computation of the posterior distribution in closed form. Thus, a mathematical formula will describe the posterior.

A classic example entails combining a normal prior with a normal likelihood. The product of these two distributions is also normal, which results in their being marked as conjugate. For example, suppose the prior distribution can be expressed in the following way:

$$\text{prior distribution}: \pi(\vartheta) \sim N\left(\vartheta_0, \sigma_0^2\right) \tag{3}$$

And suppose that the likelihood for the model can be written as

$$\text{likelihood}: \pi(\text{data}|\vartheta) \sim N\left(\vartheta_x, \sigma_x^2\right) \tag{4}$$

As the product (by virtue of conjugacy) of two normal distributions is also a normal distribution, the posterior distribution can be expressed in closed form as

$$\text{posterior}: \pi(\vartheta|\text{data}) \sim N\left(\frac{\sigma_0^2}{\sigma_x^2+\sigma_0^2}\vartheta_x + \frac{\sigma_x^2}{\sigma_x^2+\sigma_0^2}\vartheta_0, \frac{1}{\sigma_x^2}+\frac{1}{\sigma_0^2}\right)^{-1} \tag{5}$$

As with the graphical illustration in Figure 1, the posterior mean is a precision weighted average of the means from the prior and likelihood. Likewise, the posterior variance is an average of the variances of its constituent components.

Many other distributions also exhibit this property. For instance, when a beta distribution meets a binomial distribution, the product is another beta distribution. Until the mid-1990s, this conjugate approach was the standard for practical Bayesian inference for empirical studies. It was done, however, for computational feasibility, with likelihood (i.e., model) and prior choices based more on computational ease than on genuine beliefs about data generation or prior distribution properties.

The mid-1990s brought a seismic shift to Bayesian computation with the advent of MCMC techniques. Specifically, the Metropolis-Hastings (M-H) algorithm, introduced in 1994, made posterior computation feasible, even in situations in which the likelihood and prior distributions were not conjugate. The M-H and other simulation-based algorithms rely on a fundamental principle: If a sufficiently large random sample can be drawn from the posterior distribution, then it can indicate as much about that distribution as it could if it were expressed in closed form (i.e., if a formula described the distribution exactly). This principle has profound implications: Posterior distributions, even those of unknown form, can be understood, quantified, and used for statistical inference.

As an illustration of this idea, consider Bayesian quantile regression. Traditional linear regression fits a line through the conditional mean of data points. In contrast, quantile regression fits a line through the conditional quantiles of distribution. For example, fitting a line through the 50th percentile of a distribution gives rise to median regression. In general, quantile regression clarifies linear relationships between a set of independent variables X and any conditional quantile of the response variable y.

While the standard approach to quantile regression relies on frequentist methods, a Bayesian counterpart also exists. One approach to Bayesian quantile regression uses a likelihood derived from an asymmetric Laplacian (AL) distribution and a normal prior. The challenge with this approach is that these distributions are not conjugate to each other (i.e., their product does not result in a known parametric form). However, in this example, the product of a proper prior distribution and well-defined likelihood will result in the existence of a posterior distribution, even when there is no name for that distribution. Using modern computational techniques, like the M-H algorithm, it is possible to draw a random sample from this resulting posterior distribution.

The likelihood for the quantile regression model can be written as

$$\pi(y|\beta_p, \sigma) \sim \mathrm{AL}(X'\beta_p, p, \sigma) \tag{6}$$

in which AL denotes the density function for the AL distribution and p is the quantile of interest. Prior distribution for β_p and σ can be written as

$$p(\beta_p) \sim \mathrm{N}(0, 100\mathrm{I})$$

(7)

$$p(\sigma) \sim \text{Uniform}(0, 100)$$

(8)

These inputs can be fed into a routine that uses the M-H algorithm to sample from the resulting posterior distribution. This is illustrated in the plot presented in Figure 2. This plot illustrates 1,000 sequential draws from the posterior distribution of one of the model parameters, β_p. The red line in the plot represents the true value of the parameter, and the black line represents each subsequent sample from the posterior. Although the routine is initialized at 0, by draw 200 it has converged (in distribution) to the true value of the parameter. The empirical sample generated after achieving convergence can then be used to compute the mean, interval estimates, or any other quantity of interest about β_p.

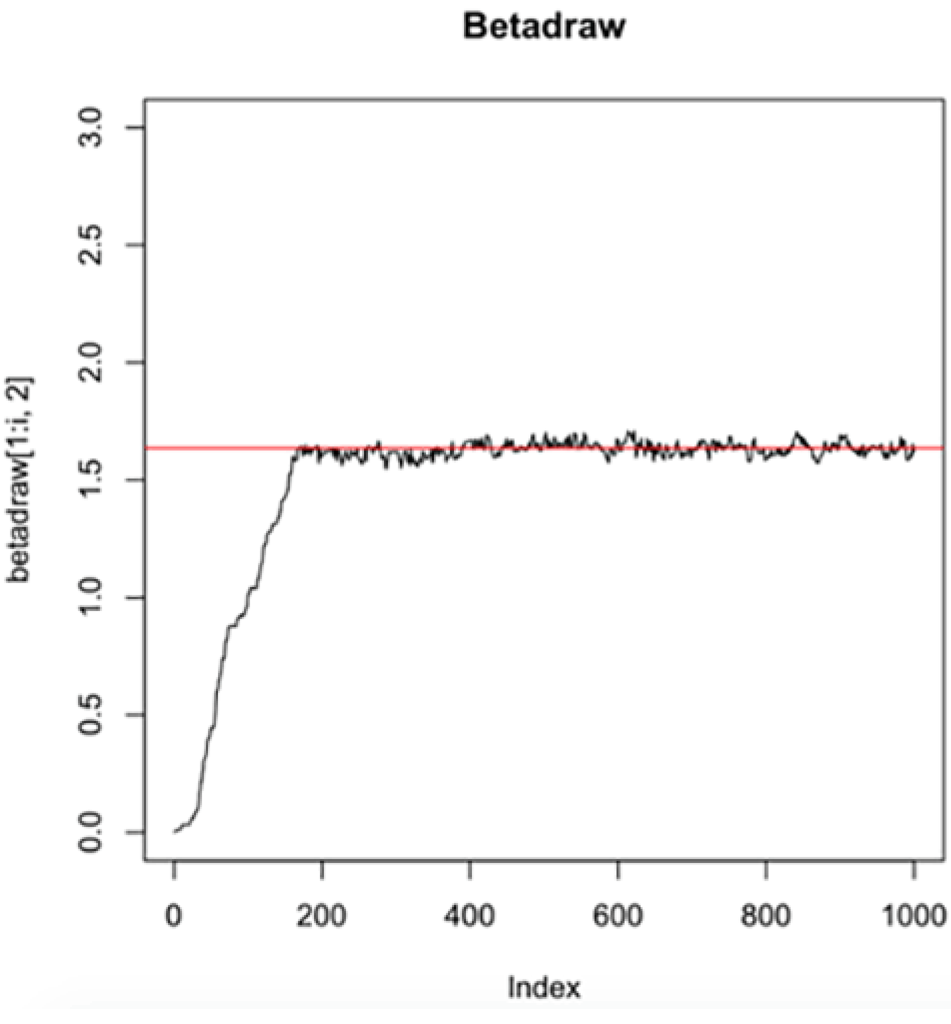

Figure 2. Illustration of posterior draws (black line) for a quantile regression parameter, β_p, constructed using the M-H algorithm. The red line represents the true value of the parameter.

In summary, the foundational principle of modern Bayesian computation is this: Knowing a distribution in closed form is great, but drawing samples from it is just as informative. This principle is at the center of the development of a myriad of MCMC techniques, with the M-H algorithm being just the tip of the iceberg.

A Bayesian Workflow

In the realm of Bayesian statistics, understanding the dynamics of priors, likelihoods, and posteriors is paramount. This section presents a streamlined workflow that can serve as a compass for conducting Bayesian studies. As the section navigates this landscape, it will elucidate the steps that are indispensable to any Bayesian exploration.

Every Bayesian study follows a fundamental sequence: model specification, prior determination, sampling from the posterior distribution, and summarizing these posterior draws. This four-step sequence forms the crux of the methodology.

Model Specification

Begin by defining a model for the data. Although there is a tendency to write down familiar models (e.g., linear relationships), model specification should result from thinking deeply about the specific process that generated the observed data. Writing down the generative process often involves thinking deeply about theory and may lead to model specifications that transcend the typical linear relationships often found in management papers. Justifying the model choice is a critical step in this process.

Prior Determination

Every parameter in the model requires a prior. For a model with parameters like β and σ^2, one might consider assigning a normal prior with mean 0 and variance 100 for β. This represents a weakly informative prior, which would essentially allow the data to be the primary influencer. For the variance term, an inverse chi-square distribution can be chosen, given its conjugacy in such contexts. The formal inclusion of a prior distribution is one of the big differences between Bayesian and frequentist inference, and historically it has been a source of skepticism regarding the use of Bayesian inference. As such, it is important to justify the choice of prior. Does it add significant value? Or is it merely an embodiment of a lack of prior knowledge? A potential approach to assuage concerns over prior specification is to experiment with alternative priors, thereby testing the robustness of the resulting findings. A more formal discussion of this topic can be found in McCann and Schwab (2023).

Sampling the Posterior Distribution

For standard models and established priors, often one can resort to canned software packages. Jebb and Woo (2015) provide an introduction to the BuggXLA program in the context of organizational research. Platforms like R, with packages such as bayesm and brms, also provide a plethora of routines tailored for Bayesian inferences, including various regression scenarios. Figure 3 presents a snippet of code used to sample from the posterior distribution for a univariate regression. Note that while there is substantial complexity in the function called in this code, access requires relatively little effort. In this case, it requires fewer than 26 lines of code. More complex and bespoke models may require custom coding or the use of more flexible computational environments like that of Stan. Independent of the source, the hallmark of successful sampling is an assurance of model convergence coupled with an effective representation of the posterior distribution.

```
library(bayesm)

n=100 #sample size
X=cbind(rep(1,n),runif(n)) #simulate X matrix
beta=c(1.5,3) #set regression coefficients
sigsq=.25 #set standard deviation

## Simulate data (y) given X, beta, sigma
y=X%*%beta+rnorm(n,sd=sqrt(sigsq))

## Create input objects required by bayesm
Data1=list(y=y,X=X)
Mcmc1=list(R=1000,keep=1)

## Run regression
out=runiregGibbs(Data=Data1,Mcmc=Mcmc1)

## Plot output
par(mfrow = c(1,2)) #plot beta and sigma side by side
matplot(out$betadraw,t="l")
abline(h=beta,col=c(1,2))
plot.ts(out$sigmasqdraw)
abline(h=sigsq)
```

Figure 3. Illustration of code to estimated a hierarchical linear model using the bayesm package in R.

Summarizing Posterior Draws

Once an empirical distribution has been drawn from the posterior, a graphical representation can be immensely insightful. For example, consider the histograms of posterior parameters presented in Figure 4. They illustrate the general range, mode, and shape of the posterior. Extracting quantitative statistical measures from these distributions, such as mean, standard deviation, and credible intervals (akin to confidence intervals), also becomes feasible.

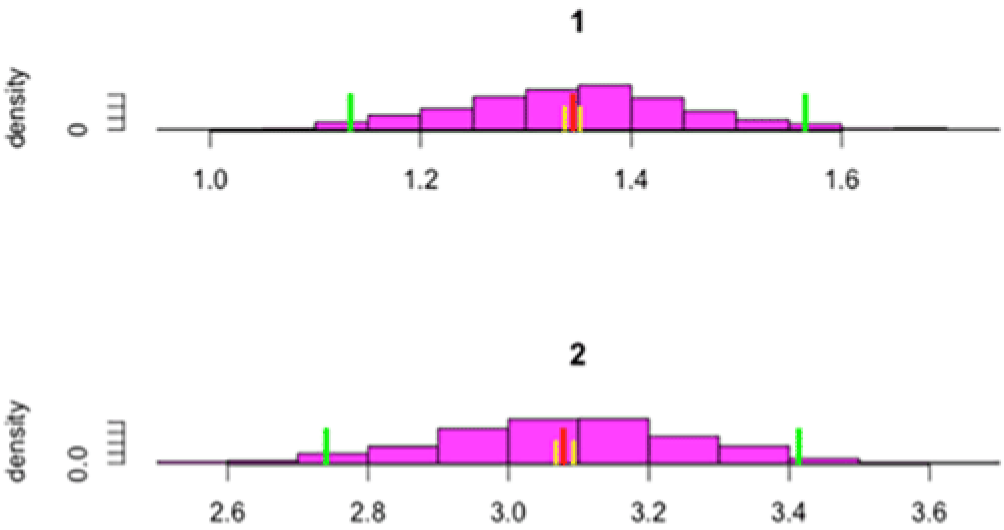

Figure 4. Graphical summaries of the posterior distribution.

These summaries subsequently pave the way for hypothesis testing. For instance, if the credible interval for a parameter does not encompass zero, one can confidently infer that the parameter is not zero.

All Bayesian endeavors, however complex, can be distilled into these four pivotal steps. The essence of Bayesian modeling lies not just in meticulously following these steps but in understanding and, crucially, justifying every choice made along the way—be it about the model's architecture or the selection of priors.

Software Tools for Bayesian Inference

Modern Bayesian statistics relies heavily on access to the right software tools for effective analysis. As researchers aim to sample from the posterior distribution, the landscape of available tools presents a wide spectrum, ranging from fixed solutions to highly customizable platforms.

Software tools catering to Bayesian analyses can be perceived along a continuum, as illustrated in Figure 5. At one end are the "fixed" tools, optimized for standard models such as linear regressions, logistic regressions, or hierarchical linear models. These software solutions, though powerful, often come with preset likelihoods and limited options for customizing prior distributions. They offer ready-to-use models, which ensure reliability and efficiency.

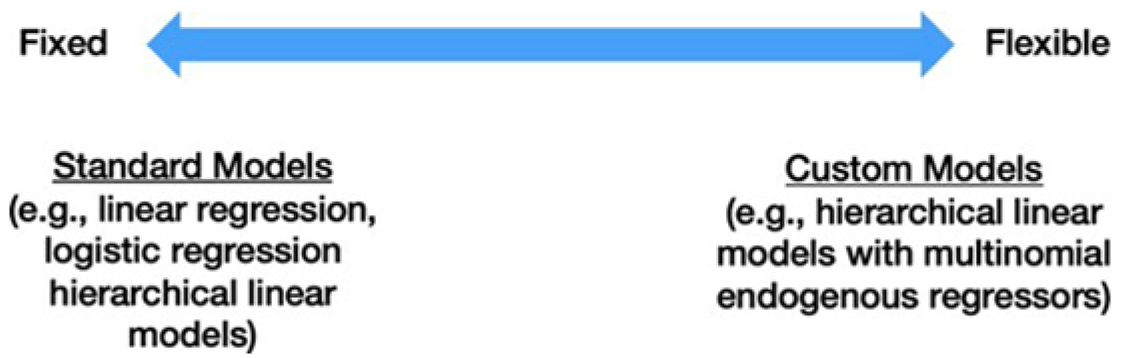

Figure 5. A continuum of Bayesian software solutions.

On the other end of this continuum lie the "flexible" tools, ideal for bespoke modeling. These tools accommodate nonstandard likelihoods and allow for flexible prior distributions, granting researchers a broader canvas to work on. Examples might include models with endogenous multinomial regressors or hierarchical linear models with unconventional temporal dependencies. The following are examples of some of the more popular solutions available, classified by where they lie on the fixed-to-flexible continuum:

1. **Fixed software solutions**: Packages like SAS and Stata have introduced Bayesian routines with an expansive range of standard models. Stata, for instance, has seamlessly integrated Bayesian methods into its environment. Simply adding a *Bayes* prefix to known functions can transform many existing routines into their Bayesian counterparts. These packages currently support over 50 likelihood models.
2. **Limited flexibility**: WinBUGS and similar software were among the pioneers in Bayesian analysis because they leveraged the Gibbs sampler and the M-H algorithm. Such software strikes a balance by allowing some degree of customization while retaining core functionalities.
3. **Full customization**: Stan has emerged as a game changer in the probabilistic programming domain. With Stan, researchers specify their model structure and priors, letting Stan's computational engine handle the heavy lifting by deploying the Hamiltonian Monte Carlo method for sampling. Such platforms offer an unprecedented degree of freedom, making them a top choice for advanced researchers.

For novices, it is advisable to commence with familiar platforms in order to fit simple models and to gain a foundational grasp. As proficiency develops and research questions become more nuanced, transitioning to platforms like Stan can unlock new horizons. The rise of machine learning and artificial intelligence has intensified interest in Bayesian methods as researchers seek to quantify the underlying uncertainty in predictions generated from these otherwise deterministic models. New computational and algorithmic approaches are being developed to approximate the Bayesian posterior distribution for large and complex data sets and models.

In summary, the world of Bayesian statistics is vast and evolving, but with the right tools and guidance, it becomes a navigable terrain. For beginners and seasoned researchers alike, the key is to match needs with the right software in order to ensure that analyses are both robust and insightful.

Summary and Next Steps

This article has provided an overview of Bayesian inference in the context of management research, including a review of the current management literature, an illustration of the contexts in which Bayesian methods have been applied, and a discussion of potential applications of the same. The article has also provided a broad overview of Bayesian methods, including definitions of the core building blocks (priors, likelihoods, and posteriors), a summary of modern computational algorithms, a general workflow for conducting Bayesian studies, and a review of software solutions.

For researchers looking to add a valuable and rare capability to their skill set, Bayesian methods can give them the ability to answer questions in ways that few in the field are able to estimate. In the parlance of strategy, this can be viewed as a dynamic capability (Teece et al., 1997), whereby the flexibility of the Bayesian approach allows for the bespoke construction of empirical research in response to the dynamic evolution of management theory. It is our opinion that this has the potential to increase the rigor and relevance of management as a field of academic inquiry.

Despite its potential, there are two key impediments to expanding the adoption of Bayesian methods within management: training and journal reception. For training, the primary challenge lies in integrating Bayesian statistical methods into the academic curriculum. Institutions should aim to embed Bayesian statistics deeply within graduate programs by developing courses that not only teach the fundamentals but also emphasize their practical applications in management. Additionally, academic associations should develop and offer workshops and seminars that are taught by seasoned researchers with experience publishing Bayesian studies. This would offer hands-on experiences that allow researchers to bridge theoretical knowledge with practical application. Online resources, including webinars and tutorial series, should be designed to support self-paced learning and to help overcome geographical and logistical barriers.

A related impediment is the reception by academic journals, which plays a critical role in the acceptance and mainstreaming of Bayesian methods in scholarly research. Journals within the field must begin to actively advocate for and publish research that uses Bayesian approaches. This could be facilitated by editorial policies that encourage submissions based on Bayesian analysis and by dedicating special issues to showcasing innovative Bayesian research. Such initiatives would not only increase the visibility of Bayesian methods but also help in establishing them as a valid and valuable toolset for tackling complex research questions in management and strategy. Ultimately, addressing these impediments will require a concerted effort from academic institutions, journals, and the broader scholarly community to foster an environment in which Bayesian methods are both taught effectively and valued highly in scholarly discourse.

References

Alcácer, J., Chung, W., Hawk, A., & Pacheco-de Almeida, G. (2018). Applying random co-efficient models to strategy research: Identifying and exploring firm heterogeneous effects. *Strategy Science*, *3*(3), 533–553.

Berger, P. G., & Ofek, E. (1995). Diversification's effect on firm value. *Journal of Financial Economics*, *37*(1), 39–65.

Campa, J. M., & Kedia, S. (2002). Explaining the diversification discount. *The Journal of Finance*, *57*(4), 1731–1762.

Certo, S. T., Albader, L. A., Raney, K. E., & Busenbark, J. R. (2024). A Bayesian approach to nested data analysis: A primer for strategic management research. *Strategic Organization*, *22*(2), 241–268.

Denrell, J., Fang, C., & Zhao, Z. (2013). Inferring superior capabilities from sustained superior performance: A Bayesian analysis. *Strategic Management Journal*, *34*(2), 182–196.

Gelman, A. (2015). The connection between varying treatment effects and the crisis of unreplicable research a Bayesian perspective. *Journal of Management*, *41*(2), 632–643.

Hahn, E. D., & Doh, J. (2006). Using Bayesian methods in strategy research: An extension of Hansen et al. *Strategic Management Journal*, *27*(8), 783–798.

Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. *Strategic Organization*, *1*(1), 51–78.

Hansen, M. H., Perry, L. T., & Reese, C. S. (2004). A Bayesian operationalization of the resource-based view. *Strategic Management Journal*, *25*(13), 1279–1295.

Heckman, J. J., Urzua, S., & Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics*, *88*(3), 389–432.

Jebb, A., & Woo, S. (2015). A Bayesian primer for the organizational sciences: The "two sources" and an introduction to BugsXLA. *Organizational Research Methods*, *18*, 92–132.

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, *15*, 722–752.

Lang, L. H., & Stulz, R. M. (1994). Tobin's *q*, corporate diversification, and firm performance. *Journal of Political Economy*, *102*(6), 1248–1280.

Mackey, T. B., Barney, J. B., & Dotson, J. P. (2017). Corporate diversification and the value of individual firms: A Bayesian approach. *Strategic Management Journal*, *38*(2), 322–341.

McCann, B. T., & Schwab, A. (2023). Bayesian analysis in strategic management research: Time to update your priors. *Strategic Management Review*, *4*(1), 75–106.

Miller, D. J. (2006). Technological diversity, related diversification, and firm performance. *Strategic Management Journal*, *27*(7), 601–619.

Nandialath, A. M., Dotson, J. P., & Durand, R. (2014). A structural approach to handling endogeneity in strategic management: The case of RBV. *European Management Review*, *11*(1), 47–62.

Oldroyd, J. B., Morris, S. S., & Dotson, J. P. (2019). Principles or templates? The antecedents and performance effects of cross-border knowledge transfer. *Strategic Management Journal*, *40*(13), 2191–2213.

Rossi, P. E., & Allenby, G. M. (2003). Bayesian statistics and marketing. *Marketing Science*, *22*(3), 25.

Rumelt, R. P., Schendel, D. E., & Teece, D. J. (1995). *Fundamental issues in strategy: A research agenda*. Rutgers University Press.

Shaver, J. M. (1998). Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science*, *44*(4), 571–585.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509–533.

Villalonga, B. (2004). Does diversification cause the "diversification discount"? *Financial Management*, *33*(2), 5–27.

Wibbens, P. D. (2019). Performance persistence in the presence of higher-order resources. *Strategic Management Journal*, *40*(2), 118–202.

Wolfolds, S. E., & Siegel, J. (2019). Misaccounting for endogeneity: The peril of relying on the Heckman two-step method without a valid instrument. *Strategic Management Journal*, *40*(3), 432–462.

Related Articles

Inferential Statistics