Bayesian Multilevel Modeling and Its Application in Comparative Journalism Studies

CHUNG-HONG CHAN[1]
GESIS – Leibniz-Institut für Sozialwissenschaften, Germany

ADRIAN RAUCHFLEISCH
National Taiwan University, Taiwan

Comparative approaches are frequently used in communication research, especially journalism studies. The purpose of this article is to argue that Bayesian multilevel regression is the most justifiable option for analyzing comparative data. We argue that it is the only approach that can simultaneously account for the non-atomicity (nested nature) and non-stochasticity (nonrandom sampling) of comparative data. Using the openly available Worlds of Journalism Study and useNews data sets, we demonstrate how to apply the Bayesian approach for the analysis of comparative data. We address the common challenges when using the Bayesian approach and highlight the advantages of posterior predictive checks for modeling checking.

Keywords: Bayesian inference, multilevel model, comparative communication research, ecological effect

Comparative approaches are frequently used in communication research, especially in journalism studies, which use these approaches (Hanusch & Vos, 2019) to model journalistic culture (e.g., Esser & Umbricht, 2013), news value (e.g., Burggraaff & Trilling, 2017; Wilke, Heimprecht, & Cohen, 2012), news flow (e.g., Grasland, 2019; Wu, 2000), among others. Around 40% of recent comparative studies in journalism research are comparative content analyses (Hanusch & Vos, 2019), which include news articles from various outlets and usually also from various countries. Then the included news articles are coded either manually or automatically. Another 10% of recent comparative studies in journalism research is made up of surveys such as the Worlds of Journalism Study (WJS; Hanusch & Hanitzsch, 2017). Although this

Chung-Hong Chan: chung-hong.chan@gesis.org
Adrian Rauchfleisch: adrian.rauchfleisch@gmail.com
Date submitted: 2022-02-28

[1] Adrian Rauchfleisch's work was supported by the National Science and Technology Council, Taiwan (ROC), Grant No. 111-2628-H-002-003-. The authors would like to thank Professor Kasper Welbers (Vrije Universiteit Amsterdam) for his advice on the development of the Dutch dictionary; and Professor Junior Yuner Zhu (Hong Kong Baptist University) for sharing details about her study. Source code and data are available at https://github.com/chainsawriot/usenews_bayes. Preregistration: https://osf.io/2h4w8/

article uses examples from journalism studies, the comparative approach is not bound to journalism studies and has been widely deployed in various subfields of communication. Comparative communication research is not exclusive to comparison across countries. Based on De Vreese's (2017) classification, one can do comparisons across times, media, and other units (e.g., language regions within a multilingual country; see Vogler & Udris, 2021). In social media studies, for instance, it is common to compare content across sources (e.g., Zhao & Zhan, 2019). The methodological issues discussed in this article also apply to those cases.

Although some comparative studies are purely descriptive, most of these studies are aimed at studying how contextual conditions such as characteristics of media outlets (e.g., political orientation, online vs. offline) or their countries (e.g., postcommunist country) influence journalists' behaviors or the content of news articles. For example, some hypotheses are "the coverage of foreign countries by the news is primarily determined by geographical proximity and the status of the covered country" (Wilke et al., 2012, p. 306), and "the degree of opinion-orientation will be highest in newspapers from Polarized Mediterranean systems and lowest in those from Anglo-American systems" (Esser & Umbricht, 2013, p. 993). The effect on media content and journalistic behaviors in these hypotheses is assumed to be ecological (Susser, 1994). It assumes a macro contextual factor (e.g., newspapers from polarized Mediterranean systems), which is associated with an outcome at the micro-level (e.g., degree of opinion orientation), namely, the journalist level or article level.

This study of ecological effects on individual behaviors has a long tradition in communication research. There are hypothesis-generating and hypothesis-testing comparative studies. An example of hypothesis-generating comparative research is that of Chan and colleagues (2020), in which the sentiment profiles of terrorism coverage from Muslim- and Christian-majority countries were visualized. Hypothesis-generating comparative studies, however, are rare. As we see from the example hypotheses above, most of these studies propose hypotheses to test for ecological effects. Usually, traditional (frequentist) hypothesis-testing approaches were used in this context: Esser and Umbricht (2013) lumped multiple outlets from the same country together and used univariate analysis of variance to test for the differences in the proportion of opinion-orienting articles across countries. In Wu (2000), multiple stepwise regression was used. These approaches violate the underlying independence of observations assumption of linear regression (see non-atomicity below). It highlights the fact that comparative research introduces a feature that researchers usually overlook: Media contents are clustered in a multilevel structure. A news article is nested within its media outlet and its media outlet is, in turn, nested within its country. Such a data structure brings two problems: Non-atomicity and non-stochasticity. And specifically, we propose two solutions: Multilevel modeling and the Bayesian approach.

The goal of this article is twofold. We use examples from journalism studies as the most critical cases in communication science to highlight (1) the strength of multilevel models and (2) the advantages of Bayesian models. In our article, we first discuss these two issues and then suggest Bayesian multilevel regression as the most defensible option. Using the openly available WJS and useNews data sets, we then demonstrate how to apply a Bayesian approach to analyze comparative data.

Multilevel Models and Non-Atomicity

In comparative research, it is easy to assume that macro-level independent variables (e.g., Anglo-American systems) can be analyzed at the same level as the micro-level dependent variable (e.g., the degree of opinion orientation). The manifestation of this assumption is to enter macro-level variables as independent variables in multiple regression analysis and regress them against a micro-level dependent variable. Suppose one wants to study how the democratic performance of a country affects journalists' perceived professional autonomy using the WJS data (Wave 1: Reich & Hanitzsch, 2013; Wave 2: Hamada, 2021), the relationship can be expressed in the following regression equation: Let y_i be the dependent variable at the micro-level (journalist level: perceived professional autonomy) and x_i be the independent variable at the macro-level (country level: democratic performance). Suppose there are m journalists, where $i = 1,2,\dots,m$.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

This method is the so-called disaggregation approach. From an epistemic standpoint, searching for an ecological effect by studying the value of slope (β_1) leads to the so-called atomistic fallacy (Hox, Moerbeek, & Van de Schoot, 2017). From a statistical standpoint, this approach violates the underlying independence assumption. Almost all frequentist statistical tests assume that the observations are independent of each other: y_i given x_i are independently identically distributed (IID). Journalists from the same country are hardly IID: They are subjected to being governed by the same government and exposed to the same journalistic culture. Also, the way communication researchers survey journalists resembles (nonrandom) cluster sampling than pure random sampling: One usually starts from a specific country and then surveys journalists from it. The choice of countries is almost always not randomly selected. This sampling approach makes the independence assumption even more fragile. Previous simulation studies have shown that ignoring this dependence can lead to false-positive associations (e.g., Chen, 2012; Clarke, 2008).

There are two solutions to this.[2] The first is to aggregate the micro-level dependent variable across a macro-level independent variable. Suppose there are n countries in the above example. One can aggregate the average level of x as z_k for the k-th country, where $k = i,\dots,n$. Then, we can do a regression with a regression equation such as:

$$z_k = \beta_0 + \beta_1 x_k + \epsilon_k$$

Using this aggregation method, the unit of analysis is effectively switched from journalist to country. This method is useful when x is the only independent variable. It does not technically commit the atomistic fallacy when one uses the value of slope as evidence for an ecological effect. However, this method still has important drawbacks. First, it cannot be used for data with more than two levels. In those cases, the lower-level predictors are lumped together and, in effect, assumed to be homogeneous and discarded. Second, this method discards a massive amount of information. It is useful only for the analysis of a dependent micro-level variable with a reasonable aggregation function (e.g., counting a binary variable).

[2] In general, non-atomicity alone only influences the inference (e.g., *p* values, confidence intervals) by increasing the chance of Type I errors. The magnitude of the regression coefficient should still be consistent. Another approach, which we are not going to discuss in detail, is to use the clustered-robust standard error. For a survey of the technique, see Wooldridge (2003). The technique can only address the non-atomicity issue alone, not non-stochasticity.

For numerical variables, aggregation functions such as taking a mean or median cannot capture the spread of the micro-level variable (Bryk & Raudenbush, 1988). Also, the effective sample size is reduced from the number of journalists (m) to the number of countries (n). Third, if one specifies hypotheses at the journalist level but bases one's conclusion on the aggregated analysis, there is a risk of interpreting the macro-level association wrongly at the micro level, otherwise known as ecological fallacy. Nonetheless, this aggregation method, although inflexible, is still useful when the number of groups (e.g., n) is large. It is also useful to collapse micro levels (e.g., article level) that are not useful in answering one's research questions.

Another solution is to use the multilevel model (linear mixed model, or hierarchical model). In a multilevel model, the effect on the micro-level dependent variable (y) is modeled with equations at different levels. Using the above example, y_{ik} denotes the perceived professional autonomy of the i-th journalist in the k-th country; x_k denotes the democratic performance of the k-th country.

$$\begin{aligned} y_{ik} &= \beta_0 + \epsilon_{ik} \\ \beta_0 &= \gamma_{00} + \gamma_{01} x_k + \mu_{0k} \end{aligned}$$

In these equations, γ_{00} is the average slope, while μ_{0k} is the group-dependent deviation of the slope from the average. It is usually set as having a normal distribution with a variance τ_{00}, that is, $\mu_{0k} \sim \mathcal{N}(0, \tau_{00})$. Instead of a single value, the regression coefficient γ_{00} is assumed to be a distribution of values depending on a macro-level group.[3] It addresses the problem of clustering articles by macro-level variables. This model is called the varying-intercept model and is frequently used in social science research. We can then study the magnitude of γ_{01} to determine the ecological effect.

The advantage of using multilevel modeling lies in its flexibility in handling multilevel data. Suppose we also want to consider the clustering of journalists around o different media organizations in the above example and j-th media organization of the i-th journalist, where $j = 1, \ldots, o$. If such data were available,[4] the multilevel regression equations would be rewritten as

$$\begin{aligned} y_{ijk} &= \pi_0 + \epsilon_{ijk} \\ \pi_0 &= \beta_{00} + \mu_{0j} \\ \beta_{00} &= \gamma_{000} + \mu_{00k} + \gamma_{001} x_k \end{aligned}$$

This flexibility is demonstrated by Rinke (2016). He studied the likelihood of opinion justification in 1,559 utterances nested in 329 news items, which were in turn nested in 101 news broadcasts. Multilevel logistic regression was used to model the natural three-level hierarchy of his data.

[3] One can also think about this in terms of pooling; that is, how the data from different groups are pooled together: Complete pooling (the disaggregation approach, one slope), no pooling (each group is modeled individually, and each group has a different slope), and partial pooling (each group has a unique slope, but the distribution of all slopes follows a distribution) are three scenarios. Multilevel modeling uses partial pooling.

[4] The data on media organizations are not available in the second wave of the WJS data set.

Another way to look at the flexibility of multilevel modeling is by varying other parameters in the regression equation as well. Suppose in another situation, a researcher is interested in analyzing the relationship between experience (XP) and perceived professional autonomy. Instead of assuming the effect to be uniform across countries, as in the following equation:

$$y_i = \beta_0 + \beta_1 XP_i + \epsilon_i$$

one can assume the slope (β_1) to be a distribution that is governed by countries, that is,

$$\begin{aligned} y_{ik} &= \beta_0 + \beta_1 XP_{ik} + \epsilon_{ik} \\ \beta_1 &= \gamma_{10} + \mu_{1k} \end{aligned}$$

This model is called the varying slope model and is useful to establish a robust estimation of effect (in this case γ_{10}) across all countries. In this article, we will not go into detail about this kind of model. For an example of its application in comparative communication research, see Barnidge, Huber, Gil de Zúñiga, and Liu (2018).

In sum, multilevel modeling is more justifiable for analyzing non-atomic data from comparative research. For surveys using probabilistic sampling, the conventional frequentist approach (maximum likelihood estimation; MLE) might still be valid and is available in most statistical packages (see an SPSS tutorial for communication researchers by Hayes, 2006).

Bayesian Models and Non-Stochasticity

"If Czech history *could* be repeated, we should of course find it desirable to test the other possibility each time and compare the results. Without such an experiment, all considerations of this kind remain a game of hypotheses" (Kundera, 1984, p. 223; emphasis added).

Stegmueller (2013) demonstrates that MLE for multilevel modeling is associated with shrinkage (reduction of standard error, i.e., more false positives), and the shrinkage is more severe when the number of macro-level units (e.g., countries) is small. Stegmueller (2013) proposes to use Bayesian analysis as a robust alternative (see counterarguments from Elff, Heisig, Schaeffer, & Shikano, 2020). Although less-restrictive methods such as restricted maximum likelihood have been demonstrated to remediate the shrinkage issue of MLE (Elff et al., 2020), we still agree with Stegmueller's proposal for theoretical reasons (Western & Jackman, 1994).

Before diving into our theoretical reasoning, it is important to revisit what frequentist inference is. Under the frequentist framework, each experiment is assumed to be one of infinite independent, *repeatable* experiments on randomly drawn samples from a population. Random experiments are assumed to be often repeated arbitrarily. Based on this assumption, and with just one experiment from the current study, we make an estimation about the population. The discrepancy between the estimation from that one experiment and the actual value of the population is due to sampling error alone, that is, which subjects were randomly sampled from the population. Randomized surveys, for example, are assumed to be repeatable through repeated random sampling of the population. Suppose we replicated the same survey 100 times; we would obtain a slightly different sample every time. Then we would calculate the 90% confidence interval of the mean for each of these 100 surveys. We should have to anticipate that roughly 90 of these 100 confidence intervals would include the true mean of the population. We cannot say for sure exactly 90 of these 100 confidence intervals would include the true mean of the population because repeated random sampling is indeed random, and the process is **stochastic**. However, we can say 90 is more probable than zero or 100.

Most of the comparative content analytic studies, for example, often collect all available content data from a bunch of selected media outlets. In contrast to cluster sampling where media outlets are sampled randomly from a sampling frame of all media outlets, these studies collect the entire population of observations from some non-probabilistically selected media outlets. In these census-like situations, there is no way to get more data unless the scope of these studies is changed (Berk, Western, & Weiss, 1995). It is especially true for modern large-*N* studies using automated content analytic techniques. Burggraaff and Trilling (2017), for instance, "collected *all* available news items from *a selection of major Dutch news outlets*, both online and print" (p. 6, emphasis added), and that amounted to 762,095 articles from nine outlets in the period of 2014–2015. In that study, one could only get more data by including more media outlets or widening the time window.[5] Unlike a (theoretically) repeatable survey like the ones done by WJS, one can randomly select more journalists from the sampling frame of all journalists in the respective country.

Therefore, these modern comparative content analytic studies are often not repeatable, and thus they generate non-stochastic data. It could be argued that data from these comparative content analytic studies with a census-like approach are fundamentally irrelevant for frequentist inference (Western & Jackman, 1994). The confidence intervals generated do not have the same meaning as those from repeatable studies. According to Western and Jackman (1994), these values from non-stochastic data "lack meaning even as abstract propositions" (p. 413).[6]

A common counterargument to this is the classic one from Deming and Stephan (1941), who suggested that "as a basis for scientific generalizations and decisions for action, a census is only a sample" (p. 45). This notion assumes a census of a population can be used to make inferences on a theoretical device called *superpopulation*, which "theoretically could exist, may have existed, or may exist in the future" (Gibbs, Shafer, & Miles, 2015, p. 3). In other words, a finite census-as-a-sample census is assumed to be a "representative sample" of an infinite superpopulation. This counterargument could be useful, but unlike a regular random sample (cluster sample included), whose representativeness can be assessed, the

---

[5] Comparative content analytic studies can also be performed without this census-like approach. The traditional approach of making a probabilistic sample from all units can also be used (Krippendorff, 2018). As there is an explicit sampling frame of all units, the population of content from which the sample is drawing inferences is known in this case. Any further inference made beyond the population is speculative.

[6] In this article, we will not go into detail of the null hypothesis statistical testing and *p* values, except in the online appendix (https://doi.org/10.17605/OSF.IO/2H4W8). Researchers are usually not only interested in whether there is or is not an effect. Instead, they are interested in how large the magnitude and in what direction the effect is. Therefore, we will focus only on the point and interval estimations from various regression models.

representativeness of the census-like data with respect to the theoretical superpopulation can never be assessed. Echoing the quote at the beginning of this paragraph, there is no way to tell if the current Czech history is representative of all possible Czech histories in the multiple parallel universes.

Instead of invoking the theoretical device of superpopulation, we follow the arguments from Western and Jackman (1994) and Stegmueller (2013) for comparative studies: Bayesian inference should be used for analyzing data from comparative research in our field. Choosing a Bayesian approach also solves the problem of misinterpretation of confidence intervals (Rinke & Schneider, 2018). Unlike the frequentist confidence intervals, Bayesian "credible intervals support an interpretation of probability in terms of plausibility" (Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2015, p. 120).

Before we move on to the next section, it is important to point out that *non-stochastic data is not an essential condition for applying the Bayesian approach.* The same approach is equally applicable to both non-stochastic and stochastic comparative data sets.

Bayesian Analysis as Alternative

What is the probability of this article being accepted by *IJoC*? Under the frequentist framework (and if *IJoC* were accepting articles stochastically), we could only find this out by repeatedly submitting this article to *IJoC*, say 100 times, and then count the frequency of acceptance in these repeated submissions. It is indeed impractical as well as inhumane to the editorial team of *IJoC*. Instead, we assert before submission that this article has a 24% probability of being accepted. That is the published acceptance rate of a similar journal. After this article is submitted and is not desk rejected, the probability might be around 24%–30%. After months of waiting and having our confidence shaken a little, the probability might decrease to 10%–20%. After the article is mixed reviewed by three reviewers and a resubmission is invited, the probability might increase to 40%–60%. If you see this article on *IJoC*'s website, then the probability is beyond doubt 100%. If we (or other researchers) "repeat the experiment" and resubmit the same article again to *IJoC*, the probability of the resubmitted article being accepted is 0%.

Without repeated experiments, these probabilities quantify our certainty on how plausible it is for the article to be accepted, given the currently available data. We revise our old beliefs (or *prior*, $p(\theta)$) with the new data (X) and form our revised belief (or *posterior*, $p(\theta|X)$). This can be summarized in the following equation (Gelman et al., 2020):

$$P(\theta|X) \propto P(\theta)P(X|\theta)$$

The $P(X|\theta)$ part is called the likelihood function. In the *IJoC* example, the likelihood function is based on rough rules from our experience and thus is not systematic. In actual analysis, we need to derive such a likelihood function based on the available data using methods such as Markov Chain Monte Carlo (MCMC).[7] Nonetheless, the above equation indicates that there are only three ingredients in any Bayesian

[7] How concretely MCMC can facilitate the estimation of the likelihood function, given the data and prior, is too technical to be included in this article. Readers can get an idea of how it works in chapter 8 of McElreath

analysis: (1) data (X), which require no elaboration, (2) a method to derive the likelihood function $P(X|\theta)$ from the data, and (3) prior, $P(\theta)$.

R interfaces to Stan (the probabilistic programming language for conducting MCMC), such as *brms* (Bürkner, 2018), can enable the derivation of the likelihood function (Part 2). But still, it is important to be mindful that Bayesian analysis is much more computationally intensive than methods such as MLE. Our benchmark suggests that Bayesian analysis needs at least 100 times more running time than MLE.

Part 3 (prior) is arguably the most controversial part of Bayesian analysis. In the *IJoC* example, we can select a reasonable prior (or *informative* prior) of 24% from published information. Finding previous studies for an informative prior is the logical first thing to do. Keating and Totzkay (2019) show that one in every seven communication research articles published in major communication journals was a form of replication attempt. For this one-seventh of communication research, there should be previous studies available to base one's informative prior in a noncontroversial manner. For the remaining six-seventh, one might not have any information to set an informative prior. One option is to consult experts or make an educated guess. Expert elicitation is a way to probe how the experts in the field think about the hypotheses. A standardized protocol for expert elicitation is available (Hanea et al., 2017), and there are many software tools available to facilitate the process.[8] But one person's expert opinion could be another person's wishful thinking. And this perceived subjectivity of specifying priors by experts' judgment attracts widespread criticism from both statisticians (e.g., Efron, 1986) and social scientists (e.g., Elff et al., 2020).

Undoubtedly, setting prior is consequential to the analysis. But the influence from priors is greatly weakened when the data get bigger. Other than expert elicitation, another noncontroversial way to specify priors—at least in our opinion—is to use a weakly informative prior (Lemoine, 2019). In this way, one specifies only the possible range of the posterior. Some so-called default weakly informative priors, for example, $t(1,0,0.25)$, have been suggested for typical regression models (Gelman, Jakulin, Pittau, & Su, 2008).

Bayesian Communication Research

Bayesian analysis is still a minority statistical method in social sciences. As far as we know, the only available comparative communication studies that used Bayesian multilevel regression to study ecological effect are by de Leeuw, Azrout, Rekker, and Van Spanje (2020) and Heidenreich, Eberl, Lind, and Boomgaarden (2022). As argued in the two previous sections, comparative communication research fits the use case of Bayesian multilevel regression analysis. This article demonstrates how to do the analysis using the R package *brms* (Bürkner, 2018) (see the source code in the online appendix https://www.doi.org/10.17605/OSF.IO/2H4W8). As the interface of *brms* is almost the same as that of

---

(2020) and understand how the method works in regression models by referring to chapters 9, 10, and 12 of McElreath (2020). For a comparative perspective between Bayesian and frequentist approaches, please refer to Samaniego (2010).

[8] Some examples are the Web-based MATCH Uncertainty Elicitation Tool (http://optics.eee.nottingham.ac.uk/match/uncertainty.php) and the R package SHELF.

lme4 (Bates, Mächler, Bolker, & Walker, 2015; another R package for fitting multilevel models using MLE), *lme4* users might find *brms* extremely familiar.

The following two examples show how the Bayesian multilevel regression analysis can be deployed to the common archetypes of comparative journalism studies. The first example is an analysis of a comparative, stochastic survey data set (WJS). This example was chosen to highlight the multilevel (non-atomicity) aspect and represent the situation of replication studies. The second example is an analysis of a comparative, non-stochastic content analytic data set. This example was chosen to highlight the necessity of using the Bayesian approach to analyze non-stochastic comparative data and represent the situation of studies where noninformative priors were needed.

Example 1: The WJS

In this example, the Bayesian approach is applied to a stochastic data set from a survey in which some randomization is involved. The starting point of this example is the recent study by Hamada (2021). Using the second-wave data from WJS (2012–2016), the study seeks to study this hypothesis (the H2a in the article): *The greater the level of democracy in a country, the more perceived professional autonomy journalists enjoy*. To rephrase this, the level of democracy in a country has an ecological effect on the journalists' perceived professional autonomy.

It is important to point out that the study is a replication. The same hypothesis has been studied in an earlier study by Reich and Hanitzsch (2013), which uses the first-wave data from WJS and has a remarkably similar title to that of Hamada (2021). The earlier study studied this hypothesis (H4): *Journalists' perceived professional autonomy is positively associated with democratic performance and press freedom, and it is negatively related to political parallelism and state intervention.* The operationalizations of both professional autonomy (based on two questions in WJS) and level of democracy (based on the Economist Intelligence Unit's Index of Democracy) are the same in the two studies. It is important to recite the operationalization of professional autonomy here. The two questions in the survey are (1) "Thinking of your work overall, how much freedom do you personally have in selecting news stories you work on?" (2) "How much freedom do you personally have in deciding which aspects of a story should be emphasized?" The possible answers ranged from 5 = "complete freedom" to 1 = "no freedom at all." The two answers were averaged.

Interestingly, the analytical approaches are also similar in the two studies. The study by Reich and Hanitzsch (2013) contains multiple regression models on how journalist-level characters predict perceived professional autonomy. However, for country-level predictors such as the Index of Democracy, Reich and Hanitzsch (2013) apply an aggregated approach due to "methodological considerations" of "includ[ing] substantive country-level predictors in an OLS regression" (p. 146). Effectively, the analysis boils down to aggregating the perceived professional autonomy of all journalists into mean values according to baskets of the Index of Democracy and then comparing those mean values by analysis of variance (ANOVA). The subsequent study by Hamada (2021) applies the same aggregated approach by studying the bivariate correlation between the mean perceived professional autonomy of all journalists in a country and the Index of Democracy of a country.

Bayesian Multilevel Analysis of WJS

Let us assume we were in the shoes of Hamada (2021) and wanted to replicate the study by Reich and Hanitzsch (2013) with the second-wave WJS data. Bayesian multilevel regression provides several advantages over the aggregated approach.

First, it can take advantage of the hierarchical structure of the data and estimate the contextual effect of democratic performance on the journalist-level perceived professional autonomy. It allows for adjustment of other journalist-level predictors that are known to influence perceived professional autonomy, for example, rank ($RANK$), experience (XP), gender (GEN), and having a university degree ($UNIV$). It also makes it possible to adjust for the possible confounding effects of other country-level variables. For example, GDP per capita (according to the World Bank) is moderately correlated with the Index of Democracy (r = 0.53). The correlation between the Index of Democracy ($DEMO$) and perceived professional autonomy (PPA) found by Reich and Hanitzsch (2013) could be spurious, and the GDP per capita was the actual determinant of perceived professional autonomy. We can adjust for the effect of GDP per capita by entering it also as an independent variable. The data can also provide such variance because there are countries with high GDP per capita but with low democracy (e.g., the United Arab Emirates and Qatar) and vice versa (e.g., Botswana and India).

Priors

The idea of choosing priors is to select a probable probability distribution for each of the unknown parameters in a model. Usually, the first step in choosing priors is to review previous studies and look for possible values to be used as our informative priors. Incorporating prior information from Reich and Hanitzsch (2013) in this Bayesian analysis makes the intention to replicate clearer.

Reich and Hanitzsch (2013) suggest that the relationship between Index of Democracy and perceived professional autonomy is in "J shape," with the average perceived professional autonomy of journalists in authoritarian regimes (the lowest end of the democratic performance) being higher than those in hybrid regimes (3.92 vs. 3.65). We can incorporate this prior information into our model by modeling a cubic relationship between Index of Democracy and perceived professional autonomy. Like the general procedure of conducting polynomial regression, we also create a parsimonious model assuming only a linear relationship to study whether the cubic regression improves the model fit. In other words, whether the relationship is really in "J shape" is studied. It is important to spell out all the equations so that we can have an idea of all the estimands. For the parsimonious model, the regression equations are the following:

$$\begin{aligned} PPA_{ik} &= \beta_0 + \beta_1 XP_i + \beta_2 RANK_i + \beta_3 GEN_i + \beta_4 UNIV_i + \epsilon_{ik} \\ \beta_0 &= \gamma_{00} + \gamma_{01} DEMO_k + \gamma_{02} \log GDP_k + \mu_{0k} \end{aligned}$$

For the cubic model, the regression equations are the following:

$$\begin{aligned} PPA_{ik} &= \beta_0 + \beta_1 XP_i + \beta_2 RANK_i + \beta_3 GEN_i + \beta_4 UNIV_i + \epsilon_{ik} \\ \beta_0 &= \gamma_{00} + \gamma_{010} DEMO_k + \gamma_{011} DEMO_k^2 + \gamma_{012} DEMO_k^3 + \gamma_{02} \log GDP_k + \mu_{0k} \end{aligned}$$

From model 4 in Reich and Hanitzsch (2013), we know that the standardized beta coefficients for the variables rank (β_1) and professional experience (β_2) are .15 and .07, respectively. Unfortunately, the associated standard errors (or standard deviations) were not reported but only asterisks representing statistical significance (p < .001 for rank, p < .01 for professional experience). We estimated the maximum standard deviation based on the significance level. For example, the standard deviation of the standardized regression coefficient ($\sigma_{\beta'_1}$) of rank can be estimated by solving the following:

$$\begin{aligned} Pr(\frac{\beta'_1}{\sigma_{\beta'_1}}) &= \frac{0.001}{2} \sim \mathcal{N}(0,1) \\ \frac{\beta'_1}{\sigma_{\beta'_1}} &= 3.29 \\ \frac{0.15}{\sigma_{\beta'_1}} &= 3.29 \\ \sigma_{\beta'_1} &= 0.045 \end{aligned}$$

The above information can be used as an informative prior of the current analysis. Unfortunately, the ANOVA result in relation to γ_{01} from Reich and Hanitzsch (2013) cannot be used as the prior, and we need to use weakly informative priors suggested by Lemoine (2019). For regression coefficients, for example, γ_{01}, we used a normal distribution of $\mathcal{N}(0,1)$. For variance terms, for example, μ_{0k}, a student *t*-distribution of $t(3,0,2.5)$ was used. All, except the prior for regression coefficients, are default priors suggested by *brms*. In practice, we suggest preregistering the priors before the data collection. As this is a secondary data analysis, we cannot do that.

To make the result from this replication comparable with that of Reich and Hanitzsch (2013), we also calculate the standardized regression coefficients, that is, all variables are transformed to z-scores.

Other Parameters

Other parameters such as *adapt_delta* control how the MCMC should be performed. It is in general safe to use the default values unless MCMC shows evidence of non-convergence. We provide a short guide on how to diagnose convergence (online appendix).

Modeling

For this analysis, we tried three different approaches: (1) disaggregation approach (ignoring the multilevel structure); (2) multilevel regression using MLE; and (3) Bayesian multilevel regression. In the disaggregation case, the multilevel structure is ignored, and the regression equations become

$$\begin{aligned} PPA_{ik} &= \beta_0 + \beta_1 XP_i + \beta_2 RANK_i + \beta_3 GEN_i \\ &+\beta_4 UNIV_i + \beta_5 DEMO_k \\ &+\beta_6 GDP_k + \epsilon_i \\ PPA_{ik} &= \beta_0 + \beta_1 XP_i + \beta_2 RANK_i + \beta_3 GEN_i \\ &+\beta_4 UNIV_i + \beta_5 DEMO_k \\ &+\beta_6 GDP_k + \beta_7 DEMO_k^2 \\ &+\beta_8 DEMO_k^3 + \epsilon_i \end{aligned}$$

It is important to remind our readers that the results from the three are likely to be similar. However, it should be noted that this similarity does not provide evidence that the disaggregation approach and the MLE approach are replacements for the Bayesian approach (Morey et al., 2015). The interval estimates have different meanings and cannot be directly compared. Comparing the three approaches, the disaggregation approach is the least useful because it violates the underlying independence assumption.

Results

The Bayesian multilevel regression gave the *posterior distribution*, $P(\theta|X)$, of the regression coefficients. It is a distribution, and therefore we need ways to display both the central tendency and spread of the distribution. By default, *brms* displays the mean and the 95% highest density interval (HDI). These two values represent the point and interval estimates of the regression coefficient.

Table 1. Point and Interval Estimates from WJS Analysis Using Three Different Analytic Strategies: Bayesian Multilevel Modeling, Multilevel Modeling Using MLE, and Disaggregation Approach.

Term	Bayesian	MLE	Disaggregation
(Intercept)	0.02 (−0.06, 0.11)	0.03 (−0.06, 0.11)	0.01 (0, 0.02)
Experience	0.08 (0.07, 0.09)	0.08 (0.07, 0.09)	0.09 (0.08, 0.1)
Rank	0.15 (0.14, 0.16)	0.16 (0.15, 0.17)	0.16 (0.15, 0.17)
Gender (female)	−0.06 (−0.07, −0.05)	−0.06 (−0.07, −0.05)	−0.04 (−0.05, −0.03)
University degree	−0.01 (−0.03, 0)	−0.01 (−0.02, 0)	−0.04 (−0.05, −0.02)
Index of Democracy	0.24 (0.13, 0.34)	0.24 (0.14, 0.34)	0.24 (0.22, 0.25)
GDP per capita	−0.05 (−0.15, 0.05)	−0.04 (−0.14, 0.05)	−0.06 (−0.07, −0.04)

The regression coefficient and the 95% HDI of the parsimonious model (Table 1) show that the standardized regression coefficient for Index of Democracy is 0.24 (95% HDI: 0.13–0.34). The Index of Democracy is the strongest predictor among all other predictors (rank: 0.15, experience: 0.08, gender: −0.06, university degree: −0.01). The standardized regression coefficient for GDP per capita is not large (−0.05). It supports Hamada's (2021) H2a, even after adjusting for the possible confounding effect of GDP per capita.

A comparison of the point and interval estimates using different approaches shows that the point estimates are remarkably similar. Bayesian multilevel regression gave a wider 95% HDI than confidence intervals from frequentist methods. This is similar to findings of previous studies (Elff et al., 2020; Stegmueller, 2013). While the frequentist confidence intervals and the Bayesian credible interval seem to be similar, they are totally different from a philosophical point of view and lead to different forms of inference. The Bayesian approach allows us to say that *given the observed data, the true value of the standardized regression coefficient has a 95% probability of falling between 0.13 and 0.34*. In contrast, the frequentist confidence interval only allows us to say that *when computing a confidence interval from the same type of data in repeated studies (if possible), 95% of the confidence intervals will include the true value of the standardized regression coefficient*. One of the misconceptions of frequentist confidence intervals is that they can be interpreted in a Bayesian way as described above (Morey et al., 2015). This example illustrates that a Bayesian approach allows us to interpret the results in a more intuitive way.

J Shape

The conditional effects plot of the cubic model (Figure 1) suggests that the relationship is not in J-shape.[9]

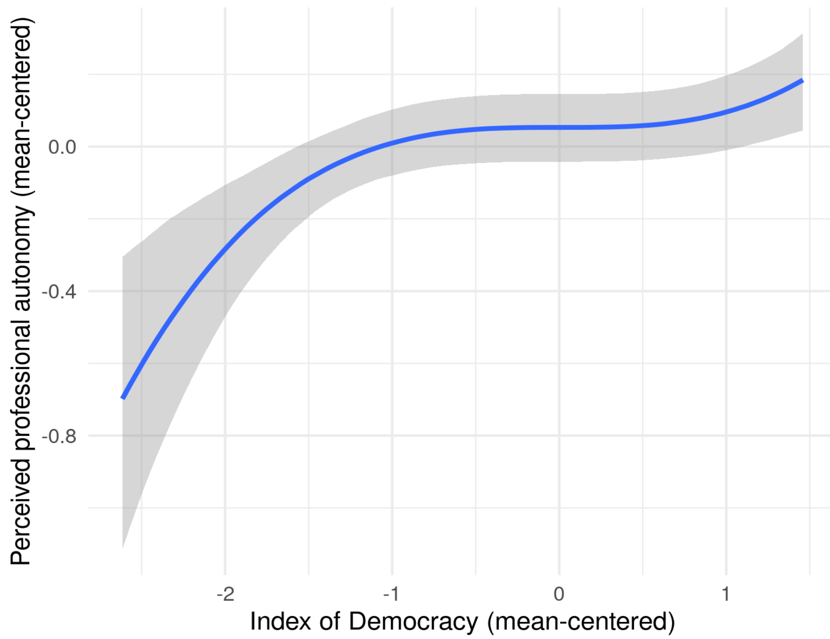

Figure 1. Conditional effects plot showing no J-shaped relationship between perceived professional autonomy and the Index of Democracy.

Posterior Predictive Checks

One unique feature of the Bayesian approach is posterior predictive checks (PPCs) for studying how our model works. Bayesian analysts have a strong culture of model checking (e.g., Gelman, 2007; Mimno, Blei, & Engelhardt, 2015), while many analyses—not only in communication research—end with regression tables and inferences.

A useful model for theory testing should pin down the data-generating process in the real world, not just whether the model fits the data. Therefore, a model should be able to generate simulated data that are like the observed data (Gelman, 2007). The gist of a PPC is to use the fitted model to generate some simulated data (y_{rep}) and compare them with the original observed data (y). If the model is useful, the model should display a similar probability distribution of y_{rep} and y. (Figure 2).

[9] Another way to evaluate the model fit is to use the leave-one-out cross validation. See the online appendix for the analysis.

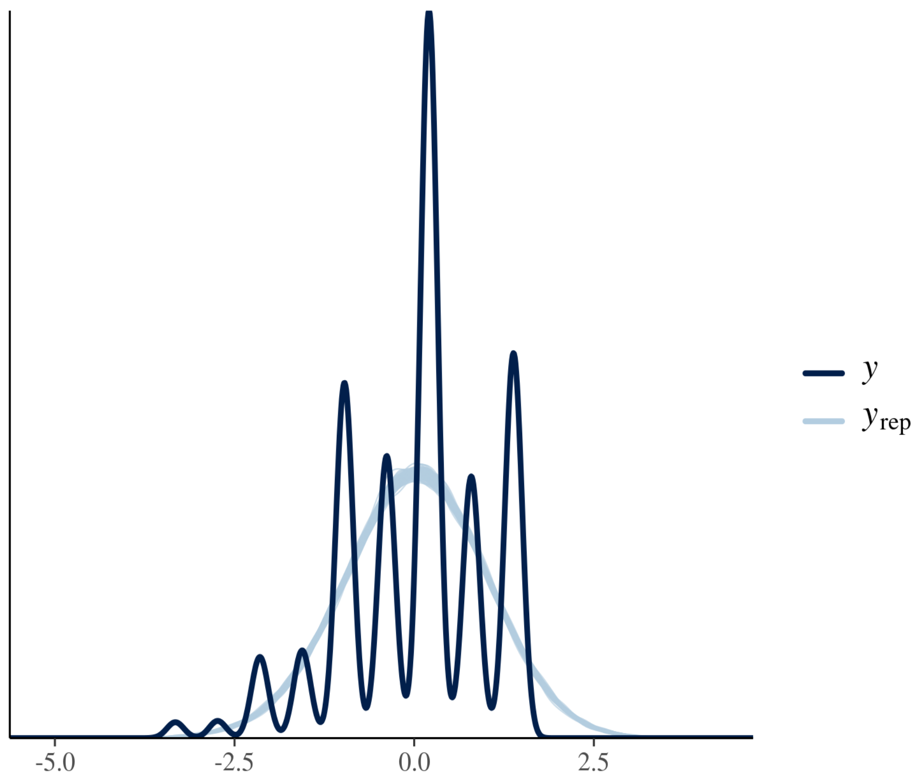

Figure 2. PPCs with 100 sets of simulated data based on the parsimonious model.

As one can see in Figure 2, the fitted model can get the range and the peak of the original data about right. But the model can only give simulated data y_{rep} with a Gaussian distribution, while the shape of the original data y is not Gaussian at all. In the online appendix, we model the finger-like shape of the data with another technique.

Summary of Example 1

In this example, the Bayesian multilevel regression approach allows the study of ecological effect hypotheses of a macro-level variable (e.g., democracy) on a micro-level variable with a comparative, semi-randomized data set. The inference of the model is at the level specified in the hypotheses (micro level, i.e., journalists) and allows adjustment for other confounding variables (e.g., GDP per capita). Visualizations such as conditional effects plots and PPCs assist model checking.

Example 2: useNews

We used the useNews data set (Puschmann & Haim, 2020) to demonstrate how to perform a Bayesian multilevel regression analysis on a non-stochastic comparative data set. The analysis was guided by a classic question in news value research: Does the distance between China and the host country of a media outlet increase the frequency of China coverage? In other words, we hypothesized that there is an ecological effect between a closer distance to China and an increased frequency of China coverage. Several distance measures were used to formulate our preregistered hypotheses. Due to limited space, only the analysis of trade volume between two countries is displayed here. For other analyses, please refer to the online appendix.

For demonstrative purposes, our analysis was separated into two levels: Outlet-level analysis (here) and article-level analysis (in the online appendix). The outlet-level analysis is an aggregated version of article-level analysis. It is used to demonstrate the flexibility of Bayesian analysis to handle two-level data with very small data sizes. The article-level analysis is used to demonstrate the analysis of a massive three-level data set.

Data

The useNews is an openly available data set (Puschmann & Haim, 2020) that includes 2019–2020 media content data from an array of worldwide media outlets. The media content data were collected from MediaCloud and made available as document-term matrices.

We used only the data from 2019 as the baseline and excluded media outlets that contributed fewer than 1,000 articles that year. This threshold allowed us to retain media outlets that the frequency of China coverage can be reliably estimated. In total, 61 media outlets contributed 1,525,871 articles.

In the outlet-level analysis, the data have 61 rows, and each row represents a media outlet. The data contain a count of articles covering China (z), total number of articles (n), country of the outlet (k), and distance measures (x).

Coverage of China

A dictionary-based approach was used. The seed English and German dictionaries from the R package newsmap (Watanabe, 2017) were used as the basis. The seed dictionaries contain words about Chinese, China, Beijing, and Shanghai. We developed further the Spanish, Romanian, Korean, Portuguese, Norwegian, and Dutch dictionaries.[10] The dictionaries were applied to the 61 document-term matrices (format of the provided data set) of all included media outlets. An article is classified as China coverage when at least one dictionary match is detected.

Distance Measure: Trade Volume

The volumes of import to and export from China with the host country of a media outlet were extracted from the 2019 edition of the *China Statistical Yearbook* (National Bureau of Statistics of China,

[10] The validation of the dictionary is available in the online appendix.

2019). This measure was log-transformed because previous studies showed the relationship between news coverage and these distance measures as not linear (Grasland, 2019; Wu, 2000).

Outlet-Level Analysis

We used negative binomial regression for this analysis.[11] Applying the same notation under the introductory section of this article, we arrive at the following multilevel regression equations

$$\begin{aligned} \log z_{jk} &= \gamma_{00} + \mu_{0k} + \gamma_{01}\log x_k + \log n_j + \epsilon_{jk} \\ \log z_{jk} - \log n_j &= \gamma_{00} + \mu_{0k} + \gamma_{01}\log x_k + \epsilon_{jk} \\ \log \frac{z_{jk}}{n_j} &= \gamma_{00} + \mu_{0k} + \gamma_{01}\log x_k + \epsilon_{jk} \end{aligned}$$

Media outlets are nested in countries. Thus, a country-based variance is added (μ_{0k}). We also added $\log n_j$ as an offset value. An offset value is a term that does not have the associated regression coefficient at the right-hand side of the regression equation. Effectively, we modeled the *rate* of China coverage.

The estimand of interest, γ_{01}, is interpreted as the average unit change in the log rate of China coverage, $\log \frac{z_{jk}}{n_j}$, for each unit change in the log distance measure of the outlet k with China.

Priors

There are four unknown parameters to be estimated: γ_{00}, μ_{0k}, γ_{01}, and the negative binomial shape parameter ϕ. Although Wu (2000) is a possible reference point, we select not to use this as priors because China was not studied. In this analysis, we used weakly informative priors suggested by Lemoine (2019). The priors were the same as the ones used in example 1. An additional gamma distribution of $\mathcal{Gamma}(0.01,0.01)$ was used for the shape parameter ϕ.

Regression Results

Table 2 shows a summary of all models from the outlet-level analysis using three different ways of modeling. As in example 1, the three methods give a similar point estimate but a wider interval estimate from the Bayesian model.

Table 2. Point and Interval Estimates from Outlet-Level Analysis Using Three Different Analytic Strategies: Bayesian Multilevel Modeling, Multilevel Modeling Using MLE, and Disaggregation Approach.

Term	Bayesian	MLE	Disaggregation
(Intercept)	−6.69 (−8.01, −5.24)	−6.77 (−7.89, −5.66)	−6.77 (−7.88, -5.66)
Log (trade volume)	0.25 (0.15, 0.34)	0.25 (0.18, 0.33)	0.25 (0.18, 0.33)

[11] We propose four extensions: Hypothesis testing, investigating cross-level interaction, establishing informative prior, and testing for temporal changes (online appendix).

Posterior Predictive Checks

Similar to the previous example, we conducted the PPC (Figure 3). Our model can capture the original data's range, peak, and shape. But there are many variations in the simulated data, probably because the Bayesian model is based on just 61 data points and one single predictor.

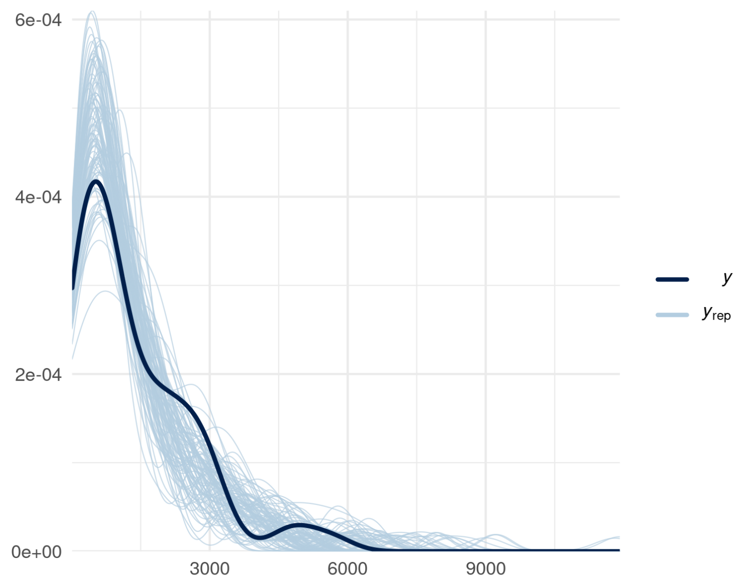

Figure 3. PPCs with 100 sets of simulated data.

Summary of Example 2

In this example, the Bayesian multilevel regression approach allows the study of ecological effect hypotheses in a nonrandom, comparative data set. Various extensions we demonstrated in the online appendix show the flexibility of the approach.

Conclusion

In this article, we argue the case for using Bayesian multilevel regression analysis to analyze data from comparative communication research. Using the openly available WJS and useNews data sets, we demonstrate that Bayesian analysis provides a valid inference of ecological effects and can be done easily with the R package *brms*. We mainly used the WJS data set to illustrate the strength of multilevel models, whereas the useNews data set was selected as an example where Bayesian models are conceptually the most defensible option. However, both multilevel models and Bayesian analysis are connected approaches.

The hierarchical data structure in both examples warrants the multilevel modeling approach. The question then is why we should use a Bayesian approach as these models can also be estimated with a frequentist approach. As the useNews data set represents a census-like situation, we argue a Bayesian approach is more appropriate. Of course, as the argument about the theoretical superpopulation shows, one could always justify a frequentist approach. However, there are also several practical reasons why a Bayesian approach is more suitable than a frequentist approach.

First, the Bayesian interpretation of confidence intervals is far more intuitive than the frequentist interpretation (Morey et al., 2015). Second, while all the models reported here can also be estimated within a frequentist framework, the use of a Bayesian framework and, more specifically, *brms*, allows estimating them all with the same package. This also makes it easier to compare different competing models, as we have illustrated in our analysis. Third, comparing models and checking the data-generating quality of the models is an essential part of a Bayesian framework. Of course, frequentist models can also be compared based on information criteria such as the Bayesian information criterion. However, PPCs that give by far the most detailed information about models and indicate in which area they potentially fail are only possible by using a Bayesian approach. Finally, while we discussed the question about priors as a potential challenge of Bayesian models, they are instead a strength of Bayesian models. Researchers must define priors to run a Bayesian regression analysis. As we were able to show, this limitation does not pose a controversial challenge as weakly informative priors can be chosen, and in most cases, enough data are available. Thus, the priors have almost no influence on the posterior distribution of the parameters. Moreover, if we have prior knowledge, it allows us to create even more useful models, as the first example has shown.

Using a Bayesian approach still has several limitations. First, it takes more time to get the results of a Bayesian regression analysis. Estimating a Bayesian regression model is a computationally demanding task. It took four days on a regular computer to get the results of the Bayesian regression model with 1,525,871 articles (example 2, article-level analysis). Still, eventually, it worked, and typical studies in comparative journalism research, for example, 27,567 journalists in example 1 could be estimated within a reasonable time frame (< 10 mins). Second, if sampling procedures are used for a typical experimental design that could be replicated, using a frequentist approach is probably the less complicated approach. In any other scenario, we believe the benefits of the Bayesian approach outweigh the potential limitations.

Coda: An Education Reform Proposal

Unlike what it was several years ago, software and educational materials for doing Bayesian analysis have been tremendously improved. We see the availability of the R package *brms* (Bürkner, 2018) and approachable textbooks such as *Regression and Other Stories* (Gelman, Hill, & Vehtari, 2020) and *Statistical Rethinking* (McElreath, 2020) as watershed moments in the (re)mainstreamization of the Bayesian approach. This development coincides with Rinke and Schneider's (2018) recommendation for a different teaching of statistics for communication research. In their opinion, the teaching of frequentist hypothesis testing would be less central: "Instead, more substantive concerns, replication, effect sizes, and better ways of drawing statistical inferences, including Bayesian methods, would take center stage" (Rinke & Schneider, 2018, p. 18). We concur with their proposal. The frequentist approach should be taught only in the context of analyzing data with explicit randomization, such as randomized experiments and simple random surveys. We hope that this article can convince the educators of our field to teach Bayesian methods, especially for multilevel regression analysis. We should also teach the issues concerning the non-atomicity and non-stochasticity of real-world data, as well as the necessity of model checking with procedures such as PPCs.

References

Barnidge, M., Huber, B., Gil de Zúñiga, H., & Liu, J. H. (2018). Social media as a sphere for "risky" political expression: A twenty-country multilevel comparative analysis. *The International Journal of Press/Politics, 23*(2), 161–182. doi:10.1177/1940161218773838

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01

Berk, R. A., Western, B., & Weiss, R. E. (1995). Statistical inference for apparent populations. *Sociological Methodology, 25*(1), 421–458. doi:10.2307/271073

Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin, 104*(3), 396–404. doi:10.1037/0033-2909.104.3.396

Burggraaff, C., & Trilling, D. (2017). Through a different gate: An automated content analysis of how online news and print news differ. *Journalism, 21*(1), 112–129. doi:10.1177/1464884917716699

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal, 10*(1), 395–411. doi:10.32614/RJ-2018-017

Chan, C.-H., Wessler, H., Rinke, E. M., Welbers, K., Van Atteveldt, W., & Althaus, S. (2020). How combining terrorism, Muslim, and refugee topics drives emotional tone in online news: A six-country cross-cultural sentiment analysis. *International Journal of Communication, 14*, 3569–3594. Retrieved from https://ijoc.org/index.php/ijoc/article/view/13247

Chen, Q. (2012). The impact of ignoring a level of nesting structure in multilevel mixture model. *SAGE Open, 2*(1). doi:10.1177/2158244012442518

Clarke, P. (2008). When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiology & Community Health, 62*(8), 752–758. doi:10.1136/jech.2007.060798

de Leeuw, S. E., Azrout, R., Rekker, R. S. B., & Van Spanje, J. H. P. (2020). After all this time? The impact of media and authoritarian history on political news coverage in twelve Western countries. *Journal of Communication, 70*(5), 744–767. doi:10.1093/joc/jqaa029

Deming, W. E., & Stephan, F. F. (1941). On the interpretation of censuses as samples. *Journal of the American Statistical Association, 36*(213), 45–49. doi:10.1080/01621459.1941.10502070

De Vreese, C. H. (2017). Comparative political communication research. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford handbook of political communication* (1st ed., pp. 287–300). Oxford, UK: Oxford University Press.

Efron, B. (1986). Why isn't everyone a Bayesian? *The American Statistician, 40*(1), 1–5.

Elff, M., Heisig, J. P., Schaeffer, M., & Shikano, S. (2020). Multilevel analysis with few clusters: Improving likelihood-based methods to provide unbiased estimates and accurate inference. *British Journal of Political Science, 51*(1), 412–426. doi:10.1017/s0007123419000097

Esser, F., & Umbricht, A. (2013). Competing models of journalism? Political affairs coverage in US, British, German, Swiss, French and Italian newspapers. *Journalism: Theory, Practice & Criticism, 14*(8), 989–1007. doi:10.1177/1464884913482551

Gelman, A. (2007). Comment: Bayesian checking of the second levels of hierarchical models. *Statistical Science, 22*(3), 349–352.

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge, UK: Cambridge University Press.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics, 2*(4), 1360–1383. doi:10.1214/08-aoas191

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., . . . Modrák, M. (2020). *Bayesian workflow*. arXiv. doi:10.48550/arXiv.2011.01808

Gibbs, B. G., Shafer, K., & Miles, A. (2015). Inferential statistics and the use of administrative data in US educational research. *International Journal of Research & Method in Education*, *40*(2), 214–220. doi:10.1080/1743727x.2015.1113249

Grasland, C. (2019). International news flow theory revisited through a space–time interaction model: Application to a sample of 320,000 international news stories published through RSS flows by 31 daily newspapers in 2015. *International Communication Gazette, 82*(3), 231–259. doi:10.1177/1748048518825091

Hamada, B. I. (2021). Determinants of journalists' autonomy and safety: Evidence from the Worlds of Journalism Study. *Journalism Practice*. Advance online publication. doi:10.1080/17512786.2021.1871861

Hanea, A. M., McBride, M. F., Burgman, M. A., Wintle, B. C., Fidler, F., Flander, L., . . . Mascaro, S. (2017). Investigate discuss estimate aggregate for structured expert judgement. *International Journal of Forecasting, 33*(1), 267–279. doi:10.1016/j.ijforecast.2016.02.008

Hanusch, F., & Hanitzsch, T. (2017). Comparing journalistic cultures across nations. *Journalism Studies, 18*(5), 525–535. doi:10.1080/1461670x.2017.1280229

Hanusch, F., & Vos, T. P. (2019). Charting the development of a field: A systematic review of comparative studies of journalism. *International Communication Gazette, 82*(4), 319–341. doi:10.1177/1748048518822606

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research, 32*(4), 385–410. doi:10.1111/j.1468-2958.2006.00281.x

Heidenreich, T., Eberl, J.-M., Lind, F., & Boomgaarden, H. G. (2022). Discontentment trumps euphoria: Interacting with European politicians' migration-related messages on social media. *New Media & Society*. Advance online publication. doi:10.1177/14614448221074648

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Keating, D. M., & Totzkay, D. (2019). We do publish (conceptual) replications (sometimes): Publication trends in communication science, 2007–2016. *Annals of the International Communication Association, 43*(3), 225–239. doi:10.1080/23808985.2019.1632218

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Newbury Park, CA: SAGE.

Kundera, M. (1984). *The unbearable lightness of being*. New York, NY: Harper & Row.

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos, 128*(7), 912–928. doi:10.1111/oik.05985

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton, FL: CRC Press.

Mimno, D., Blei, D. M., & Engelhardt, B. E. (2015). Posterior predictive checks to quantify lack-of-fit in admixture models of latent population structure. *Proceedings of the National Academy of Sciences, 112*(26), E3441–E3450. doi:10.1073/pnas.1412301112

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2015). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23*(1), 103–123. doi:10.3758/s13423-015-0947-8

National Bureau of Statistics of China. (2019). *China statistical yearbook 2019*. Retrieved from https://web.archive.org/web/20201022104956/http://www.stats.gov.cn/tjsj/ndsj/2019/indexeh.htm

Puschmann, C., & Haim, M. (2020). *UseNews* [Data set]. doi:10.17605/OSF.IO/UZCA3

Reich, Z., & Hanitzsch, T. (2013). Determinants of journalists' professional autonomy: Individual and national level factors matter more than organizational ones. *Mass Communication and Society, 16*(1), 133–156. doi:10.1080/15205436.2012.669002

Rinke, E. M. (2016). The impact of sound-bite journalism on public argument. *Journal of Communication, 66*(4), 625–645. doi:10.1111/jcom.12246

Rinke, E. M., & Schneider, F. M. (2018). *Probabilistic misconceptions are pervasive among communication researchers*. SocArXiv. doi:10.31235/osf.io/h8zbe

Samaniego, F. J. (2010). *A comparison of the Bayesian and frequentist approaches to estimation* (Vol. 24). New York, NY: Springer.

Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science, 57*(3), 748–761. doi:10.1111/ajps.12001

Susser, M. (1994). The logic in ecological: I. The logic of analysis. *American Journal of Public Health, 84*(5), 825–829. doi:10.2105/ajph.84.5.825

Vogler, D., & Udris, L. (2021). Transregional news media coverage in multilingual countries: The impact of market size, source, and media type in Switzerland. *Journalism Studies, 22*(13), 1793–1813. doi:10.1080/1461670x.2021.1965909

Watanabe, K. (2017). Newsmap. *Digital Journalism, 6*(3), 294–309. doi:10.1080/21670811.2017.1293487

Western, B., & Jackman, S. (1994). Bayesian inference for comparative research. *American Political Science Review, 88*(2), 412–423. doi:10.2307/2944713

Wilke, J., Heimprecht, C., & Cohen, A. (2012). The geography of foreign news on television. *International Communication Gazette, 74*(4), 301–322. doi:10.1177/1748048512439812

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review, 93*(2), 133–138. doi:10.1257/000282803321946930

Wu, H. D. (2000). Systemic determinants of international news coverage: A comparison of 38 countries. *Journal of Communication, 50*(2), 110–130. doi:10.1111/j.1460-2466.2000.tb02844.x

Zhao, X., & Zhan, M. M. (2019). Appealing to the heart: How social media communication characteristics affect users' liking behavior during the Manchester terrorist attack. *International Journal of Communication, 13*, 3826–3847. Retrieved from https://ijoc.org/index.php/ijoc/article/view/11816/2753