Using Bayesian Hierarchical Models to Estimate Heterogeneous Effects

Joshua Alley
Assistant Professor
University College Dublin*
joshua.alley@ucd.ie

September 11, 2023

Abstract

Heterogeneous effects are common in social science. In this note, I introduce a Bayesian hierarchical approach to estimating heterogeneous effects. The modeling strategy uses varying slopes and intercepts, along with predictors of the slopes, to capture heterogeneous effects by groups. Researchers can specify groups and sources of heterogeneity based on the quantities of interest, including heterogeneous treatments, treatment heterogeneity, and policy. This approach provides an intermediate tool between interactions or subgroup analyses and machine-learning approaches for discovering complex heterogeneity. It is more flexible than interactions and reduces the risk of underpowered subgroup comparisons. At the same time, it is more theoretically driven and interpretable than some machine-learning approaches, as well as easier to implement in small datasets. Researchers can thus use hierarchical models alongside other approaches to understand heterogeneous effects for scholarship and policy.

*Thanks to Carlisle Rainey for helpful comments.

1 Introduction

Whether in observational and experimental studies, every independent variable social scientists examine impacts some units more or less than others. Aggregate relationships often mask heterogeneous effects.[1] These averages are useful in some instances, but they often obscure interesting and important variation.

As a result, understanding varying responses to a given stimulus is essential for policy and scholarship. Estimating heterogeneity allows scholars to better elucidate the process linking their independent variable and outcome. Policymakers can target finite resources and focus interventions where they will have the most impact.

This note introduces a hierarchical Bayesian approach to estimating heterogeneous effects. The technique estimates heterogeneous effects using varying slopes and intercepts, along with covariates that predict slopes. Modeling heterogeneous effects in this way is produces easily interpretable results, which facilitates argument testing. It also allows researchers to compare different sources of heterogeneous effects, and can be extended in many ways. Implementation using the brms package for R is straightforward, and I provide example code here.[2]

This approach to heterogeneous effects fills a niche between existing tools for estimating heterogeneous effects. Parametric interactions and subgroup analyses are a common way to examine individual modifiers. While these techniques are easy to implement and interpret, they lose interpretability with more than three dimensions and can be misleadingly underpowered (Simmons, Nelson and Simonsohn, 2011).[3] To capture more complex variation, recent work employs random forests (Green and Kern, 2012; Wager and Athey, 2018), support vector machines (Imai and Ratkovic, 2013), and ensemble methods (Grimmer, Messing and

[1]For example, Abramson, Koçak and Magazinnik (2022) note that the average marginal component effect (AMCE) of conjoint experiments reflects the direction and intensity of respondent preferences, and gives more weight to intense preferences.

[2]Researchers can also calculate substantive effects with the marginaleffects package (Arel-Bundock, N.d.).

[3]Blackwell and Olson (2022) describe a lasso approach to interactions that is also an intermediate step between machine-learning and linear regressions.

Westwood, 2017; Künzel et al., 2019; Dorie et al., 2022). These machine learning algorithms capture complex patterns, but can be difficult to interpret and implement, especially in smaller datasets.

Using a hierarchical model where other variables predict heterogeneous effects is more flexible than parametric interactions but more straightforward than machine learning approaches. It preserves a simple and interpretable structure like interactions, while accommodating more factors and reducing the risks of subgroup analysis. This facilitates argument testing and is easier to interpret than some machine-learning techniques. The hierarchical approach lacks the flexibility to discover high-dimensional heterogeneity, however. As a result, this approach is best used in concert with other heterogeneous effects techniques, and could be a useful addition to ensemble models.

In the remainder of this note, I describe the model and demonstrate how it works by analyzing a study of how military alliances shape public support for war by Tomz and Weeks (2021). While substantiating the original findings, the reanalysis also reveals that alliances increase support for intervention most among men who support international engagement but are otherwise skeptical of using force. This suggests that alliances exert a large influence on mass support for war by impacting individuals who otherwise prefer peaceful collaboration in international affairs.

2 A Hierarchical Model of Heterogeneous Effects

The heterogeneous effects model uses at least two equations, and is easy estimate with Bayesian methods.[4] The first equation links the treatment and outcome. The second equation estimates heterogeneous effects as a function of unit characteristics, other treatments, contextual factors, or whatever else the researcher is interested in. The estimates give heterogeneous

[4]Priors for most parameters depend on the problem and researcher knowledge.

effects for groups with unique combinations of variables that modify the treatment and correlates of differences in the treatment.[5]

This approach can apply to many problems, but the following example addresses a common scenario; making between-unit comparisons based on an experimental treatment. Start with N units indexed by i, some of which receive a binary treatment T. For simplicity, I assume that the outcome variable y is normally distributed with mean μ_i and standard deviation σ.[6]

The first equation predicts the outcome mean. The outcome for each unit is then a function of varying intercepts α_g, a matrix of control variables $\mathbf{X}$,[7] and a set of group treatment effects θ_g, which are normally distributed with mean η_g and standard deviation σ_θ. The researcher divides all units into g groups based on unique combinations predictors of heterogeneous effects $\mathbf{Z}$. Each θ parameter estimates the treatment effect in group g, and is often referred to as a varying slope.

$$
\begin{aligned}
y &\sim N(\mu_i, \sigma) && \text{(Likelihood)} \\
\mu_i &= \alpha + \alpha_g + \theta_g T + \mathbf{X}\beta && \text{(Mean Equation)} \\
\theta_g &\sim N(\eta_g, \sigma_\theta) && \\
\eta_g &= \lambda_0 + \mathbf{Z}\lambda && \text{(Heterogeneous Effects)}
\end{aligned} \tag{1}
$$

The second equation then predicts the treatment effects with a second set of variables in the matrix **Z**. **Z** can contain anything that modifies the impact of treatment, including unit characteristics, other treatments, or contextual factors. The researcher specifies these variables and uses them to define the groups. The second equation also includes an intercept λ_0 that estimates the impact of treatment when all the heterogeneous effect variables are zero.[8]

[5]Adding additional heterogeneous effect equations to estimate heterogeneous effects for multiple variables is straightforward.

[6]Researchers should use binary, categorical and other outcome likelihoods as needed.

[7]This can be omitted, depending on the application. Adding additional grouping structures for more complex data is also straightforward.

[8]In most applications, the random intercepts α_g and varying slopes θ_g should have a common multivariate normal prior to capture correlations between group slopes and intercepts.

The θ parameters are the key estimates in this model. These give the impact of a treatment within each group. All θs are a function of a systematic component where the group-level variables in $\mathbf{Z}$ modify the varying slope directly, and a random component of varying slopes. In most applications, the systematic component will be the dominant influence.

Specifying the groups in which treatment slopes vary is the most important task for researchers. As in most social science applications, researchers should identify what variation is most important and interesting. Theory, policy concerns, or normative factors are all possible motivations, and they yield three general approaches.

There are three general ways to set groups. First, if an intervention has multiple dimensions, researchers might set groups using combinations of other treatments. Here, the experimental design determines groups. The resulting hierarchical model estimates heterogeneous treatments. A second approach uses unit and contextual factors to create groups and estimate treatment effect heterogeneity. In this instance, researchers examine what within or around units shapes their reaction to an intervention.[9] Third, researchers may have specific policy aims, for example if they want to understand how an intervention affects individuals in a specific population in a given geography. Some of these policy goals may have normative motivations.

The number of grouping factors should depend first on a researcher's theoretical interest, but there may be some practical constraints. Dividing groups based on many factors will create many small groups and increase the risk of model fitting problems. Using only one factor will create an unidentified model, and researchers should use interactions if they only care about one modifier.

Estimating heterogeneous effects in this way has three advantages. First, this model allows researchers to account for multiple potential sources of heterogeneous effects in an easy to interpret framework. Researchers can thus examine theories of heterogeneous effects and com-

[9] For example, Alley (2021) uses alliance characteristics such as treaty design and membership to examine when alliance membership increases or decreases military spending.

pare sources of variation.[10] Partial pooling also facilitates reasonable estimates for small groups by sharing information across treated groups and leveraging the predictors in the heterogeneous effects equation. Finally, this approach will be faster than machine learning approaches for many datasets, and may scale better than models that attempt to estimate individual treatment effects.

Like all methods, this technique has downsides, which can be partially ameliorated by altering the basic framework above. Because groups are based on unique combinations of heterogeneous effect variables, using multiple continuous variables in the heterogeneous effects equation creates many small groups or individual treatment effects, which increases the risk of sampling problems, especially in small datasets. If using several continuous variables hinders model convergence, researchers can bin continuous variables.

Furthermore, unlike machine learning approaches, this model will not discover high-dimensional interactions. That said, researchers can inject substantial flexibility if they want using additional interactions or non-linear specifications in either level of the model. Third, this model can show general trends, but will not make powerful comparisons between every groups. Researchers who want to compare a few specific groups may not be able to, especially if the groups are small.

3 Example Application

In the following, I demonstrate how the model works by reanalyzing a study by Tomz and Weeks (2021) (TW hereafter). TW examine how military alliances shape public support for war. In a factorial experiment with vignettes, they find that alliances increase support for war by 33% on average.

I use hierarchical models to estimate heterogeneous treatment and treatment heterogene-

[10]Rescaling variables in the heterogeneous effects equation, for example by rescaling continuous variables by two standard deviations (Gelman, 2008), can aid model fitting and direct coefficient comparisons.

ity. First, I examine how the impact of alliances varies with other factors in the experiment, especially costs, stakes, region and partner democracy. The heterogeneous treatment model corroborates TW's conclusion that alliances exert the greatest impact in instances when public opinion is otherwise skeptical of intervention, such as supporting an autocracy with high costs and low stakes. A second model examines treatment heterogeneity- how respondent demographics change the impact of alliances. This suggests that alliances exert the most impact on individuals who otherwise prefer peaceful cooperation in foreign policy.

Figure 1 supports TW's findings that alliances exert the most influence in situations where the public is otherwise skeptical of intervention. This figure shows the impact of alliances in unique combinations of all other experimental treatments. For a hypothetical democracy in Eastern Europe where intervention has high stakes, alliances exert minimal impact on public attitudes. In low-stakes and high cost interventions to support African, Asian or Latin American dictators, alliances increase support for intervention by 50%. Military alliances generally increase support for intervention, but the magnitude of the effect varies widely with context.

The impact of alliances also varies with individual respondent characteristics, as I show in Figure 2. This uses race, gender, hawkishness and internationalism to create groups and predict the impact of alliances on support for using force. I selected these variables because foreign policy dispositions like militant assertiveness shape general willingness to intervene (Kertzer et al., 2014) as does gender (Barnhart et al., 2020) and race.

The treatment heterogeneity estimates indicate that alliances exert the most influence on support for foreign interventions among white men, especially those with low hawkishness and high internationalism, who can be labeled as "cooperative internationalists." Among white men with minimal hawkishness and maximum internationalism, alliances increase support for using force by 50%, which is roughly double the typical effect. By contrast, alliances have little impact on support for war among non-white females who are skeptical of international engagement. Militant assertiveness reduces the impact of alliances, perhaps because high militant

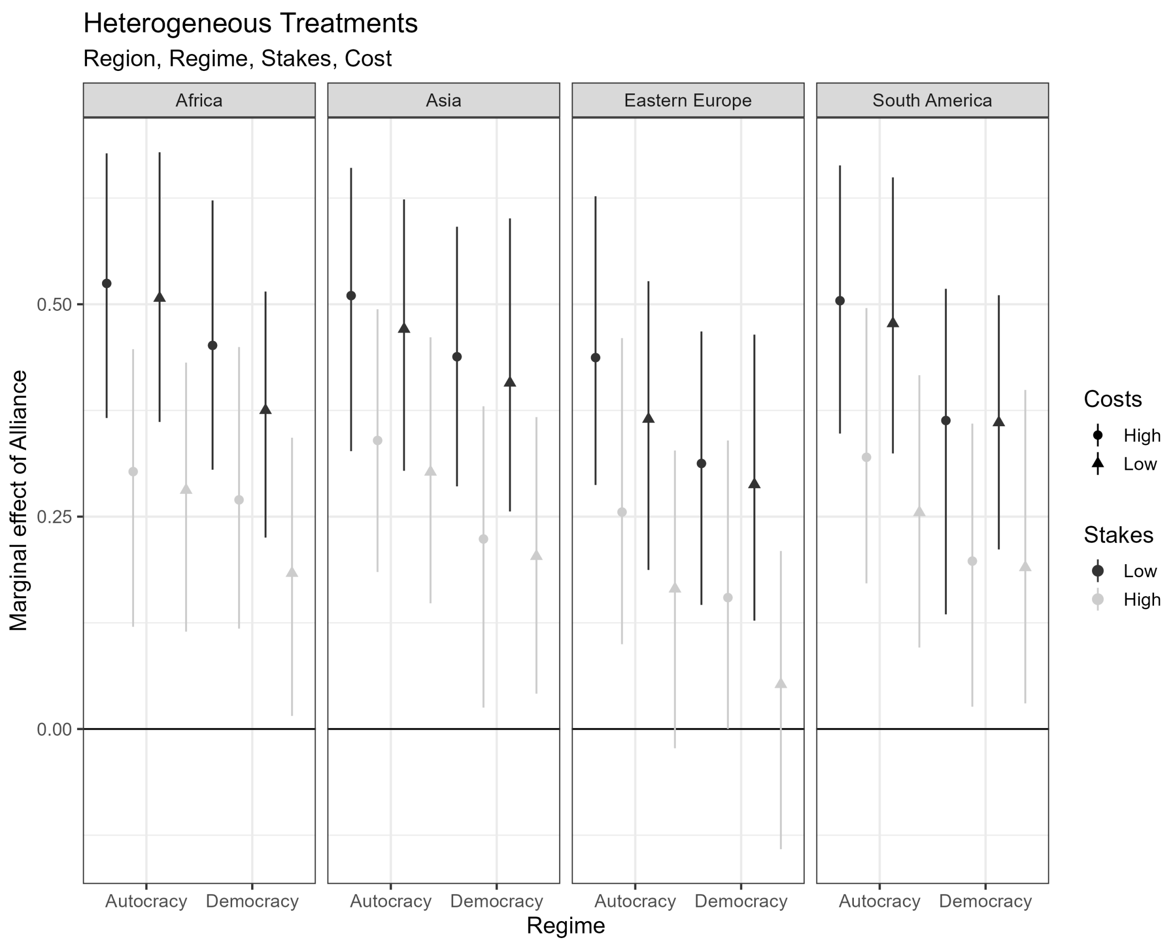

Figure 1. *Estimated impact of military alliances on public support for war across hypothetical region, costs, stakes and partner regime. Colors distinguish stakes, point shapes mark different costs, and estimates are grouped by regime type. Point estimates give the posterior median and error bars summarize the 95% credible interval.*

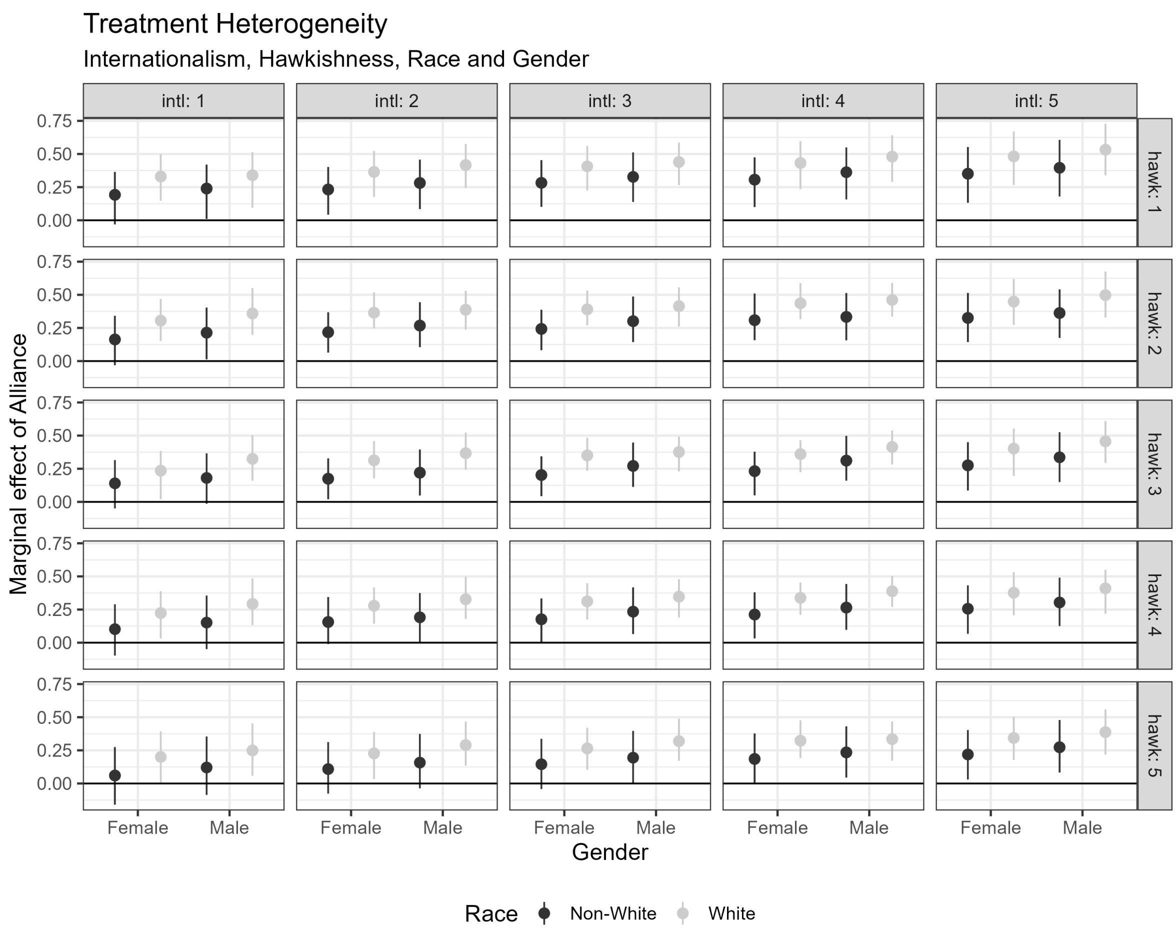

Figure 2. *Estimates of how the impact of military alliances on support for using force varies across different demographic groups. Points mark the posterior median and*

assertiveness makes This implies that alliances help convince individuals who back international engagement but are less inclined to use force. As a result, internationalism is more important than hawkishness for understanding who is willing to fight for U.S. allies.

These results show some of the strengths and weaknesses of the hierarchical approach to estimating heterogeneous effects. A relatively simple model based on demographic groups provides new insights about who responds to alliances. At the same time, because some demographic groups are relatively small, the within-group effect estimates can have substantial uncertainty. That uncertainty makes comparing groups and inferring precise effects more challenging. Smaller groups would have less uncertainty but perhaps average out interesting variation in the impact of alliances.

4 Conclusion

This note introduced a simple and interpretable hierarchical technique for estimating heterogeneous effects. The approach above can apply to a wide range of outcomes, data structures, and theories. Explicitly modeling how different groups respond to an independent variable can help test arguments and identify who responds best to a given intervention.

Hierarchical modeling provides an intermediate approach between simple interactions or subgroup analyses and complex machine-learning algorithms. As a result, this approach is a complement for existing tools, not a substitute. Researchers can should use this tool to check or inform other techniques. With this and other tools, scholars and policymakers can better understand heterogeneous effects.

References

Abramson, Scott F, Korhan Koçak and Asya Magazinnik. 2022. "What Do We Learn about Voter Preferences from Conjoint Experiments?" *American Journal of Political Science*

66(4):1008–1020.

Alley, Joshua. 2021. "Alliance Participation, Treaty Depth and Military Spending." *International Studies Quarterly* 65(4):929–943.

Arel-Bundock, Vincent. N.d. *marginaleffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests*. R package version 0.14.0.9000.
URL: *https://vincentarelbundock.github.io/marginaleffects/*

Barnhart, Joslyn N, Robert F Trager, Elizabeth N Saunders and Allan Dafoe. 2020. "The Suffragist Peace." *International Organization* 74(4):633–670.

Blackwell, Matthew and Michael P Olson. 2022. "Reducing Model Misspecification and Bias in the Estimation of Interactions." *Political Analysis* 30(4):495–514.

Dorie, Vincent, George Perrett, Jennifer L Hill and Benjamin Goodrich. 2022. "Stan and BART for Causal Inference: Estimating Heterogeneous Treatment Effects Using the Power of Stan and the Flexibility of Machine Learning." *Entropy* 24(12):1782.

Gelman, Andrew. 2008. "Scaling regression inputs by dividing by two standard deviations." *Statistics in medicine* 27(15):2865–2873.

Green, Donald P and Holger L Kern. 2012. "Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees." *Public Opinion Quarterly* 76(3):491–511.

Grimmer, Justin, Solomon Messing and Sean J Westwood. 2017. "Estimating Heterogeneous Treatment Effects and the Effects of Heterogeneous Treatments with Ensemble Methods." *Political Analysis* 25(4):413–434.

Imai, Kosuke and Marc Ratkovic. 2013. "Estimating treatment effect heterogeneity in randomized program evaluation." *The Annals of Applied Statistics* 7(1):443–470.

Kertzer, Joshua D., Kathleen E. Powers, Brian C. Rathbun and Ravi Iyer. 2014. "Moral Support: How Moral Values Shape Foreign Policy Attitudes." *The Journal of Politics* 76(3):825–840.

Künzel, Sören R, Jasjeet S Sekhon, Peter J Bickel and Bin Yu. 2019. "Metalearners for estimating heterogeneous treatment effects using machine learning." *Proceedings of the national academy of sciences* 116(10):4156–4165.

Simmons, Joseph P, Leif D Nelson and Uri Simonsohn. 2011. "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant." *Psychological Science* 22(11):1359–1366.

Tomz, Michael and Jessica L.P. Weeks. 2021. "Military Alliances and Public Support for War." *International Studies Quarterly* 65(3):811–824.

Wager, Stefan and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association* 113(523):1228–1242.