Using Bayesian Hierarchical Models to Estimate Heterogeneous Effects

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

August 25, 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 use the results to infer how different factors modify an effect and how groups with unique combinations of those factors respond to an independent variable. This approach provides a flexible intermediate tool between interactions or subgroup analyses and machine-learning approaches for discovering complex heterogeneity. It provides more nuance than interactions and reduces the risk of underpowered subgroup comparisons. At the same time, it is more theoretically driven and interpretable, as well as easier to implement and use with small datasets, than some machine-learning approaches. Researchers can thus use hierarchical models alongside other approaches to provide a fuller accounting of heterogeneous effects for scholarship and policy.

1 Introduction

Whether in observational and experimental studies, every independent variable social scientists study impacts some units more or less than others. Aggregate relationships often mask heterogeneous effects.[1] As a result, understanding variation in how different units respond to the same stimulus is essential for policy and scholarship. Scholars can better elucidate the process linking their independent variable and outcome, while policymakers can target finite resources.

This note introduces a hierarchical Bayesian approach to estimating heterogeneous effects. The technique uses varying slopes and intercepts, along with covariates, to estimate heterogeneous effects. 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.

This approach to heterogeneous effects fills a niche between existing tools for estimating heterogeneous effects. Parametric interactions and subgroup analyses are a common way that scholars examine individual modifiers. While these techniques are easy to implement and interpret, they lose their ease of interpretation with more than three dimensions and are often misleadingly underpowered (Simmons, Nelson and Simonsohn, 2011).[2] To capture more complex variation, later 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 Westwood, 2017; Künzel et al., 2019; Dorie et al., 2022). These machine learning algorithms capture complex patterns, but can be harder to interpret and

---

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

implement, especially in smaller datasets.

Using a hierarchical model where other variables predict heterogeneous effects is more flexible than simple parametric interactions but easier to employ 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 foreign side-taking in US elections impacts support for economic engagement by Bush and Prather (2020). While substantiating the original findings, the reanalysis also reveals that gender and political engagement shape how foreign side-taking in elections impacts support for economic engagement.

2   A Hierarchical Model of Heterogeneous Effects

The heterogeneous effects model uses at least two equations, and is easiest to estimate with Bayesian methods.[3] 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 effects for groups with unique combinations of variables that modify the treatment and correlates of differences in the treatment.[4]

This approach can apply to many problems, but the following example addresses a common

---

[3]Priors for most parameters depend on the problem and researcher knowledge. I estimate the example model below with STAN (Carpenter et al., 2016).

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

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.[5]

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 $\mathbf{X}$,[6] 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.

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

The second equation then predicts the treatment effects with a second set of variables in the matrix $\mathbf{Z}$. $\mathbf{Z}$ 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_g that estimates the impact of treatment when all the heterogeneous effect variables are zero. As in other regressions, this intercept is not always substantively meaningful.[7]

The θ parameters are the key estimates in this model. These give the impact of a treatment within each group as a function of a systematic component where the group-level variables in $\mathbf{Z}$ modify the varying slope directly, and a random component. The impact of a treatment

---

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

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

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

within groups is also correlated with the group mean α_g.

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 compare sources of variation.[8] 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. First, relying on varying intercepts makes it most straightforward for binary independent variables. Extensions to continuous variables should use varying slopes within groups, not varying intercepts like the model above.[9] Furthermore, 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.

Second, unlike machine learning approaches, this model will not discover high-dimensional interactions. That said, researchers can inject substantial flexibility if they want using 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 pair of groups. Researchers who want to compare a few specific groups may not be able to, especially if the groups are small.

---

[8]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.

[9]See Alley (2021) for an example.

3 Example Application

In the following, I demonstrate how the model works by reanalyzing a study by Bush and Prather (2020) (BP hereafter). This study examines how foreign meddling in elections impacts support for economic engagement with the meddler. One of their experiments examines how Russian or German engagement in the 2016 US election impacts mass support for trade and investment with those countries. I use a heterogeneous effects model to check their results and further explore their findings.

BP employ a 2x2x2 factorial experiment. This design randomizes whether a foreign country is interfering in the 2016 election, the country and their attitude towards Trump and whether the potential economic ties entail greater trade or investment in the United States. In one treatment, Russia expresses support for Trump and in the other, Germany expresses opposition to Trump.

BP hypothesize that individuals will prefer economic engagement with states that support their candidate. They thus examine how the impact of side-taking varies with the direction of the endorsement, individual political preferences, and the economic issue. To do this, they use four tables and figures, and rely on eight t-tests of differences in means between groups of roughly 25 respondents.

The heterogeneous effects model encapsulates the argument and all tests by estimating the impact of side-taking on economic engagement as a function of which country is involved, a dummy indicator of intention to vote for Clinton, and their interaction. The interaction captures the hypothesis that Clinton voters will support engagement with Germany because Germany opposed Trump, and should be positive. I also add an indicator of whether the experiment deals with trade or investment. Last the heterogeneous effects equation also includes indicators of female gender, active employment, college education, political knowledge, and political interest.

The outcome variable is a scale from one to four that measures support for greater economic engagement with the potential partner. To approximate BP's analyses, I use a normal outcome likelihood. Because this experimental randomization produced a roughly balanced sample, I do not include any control variables in the outcome equation.
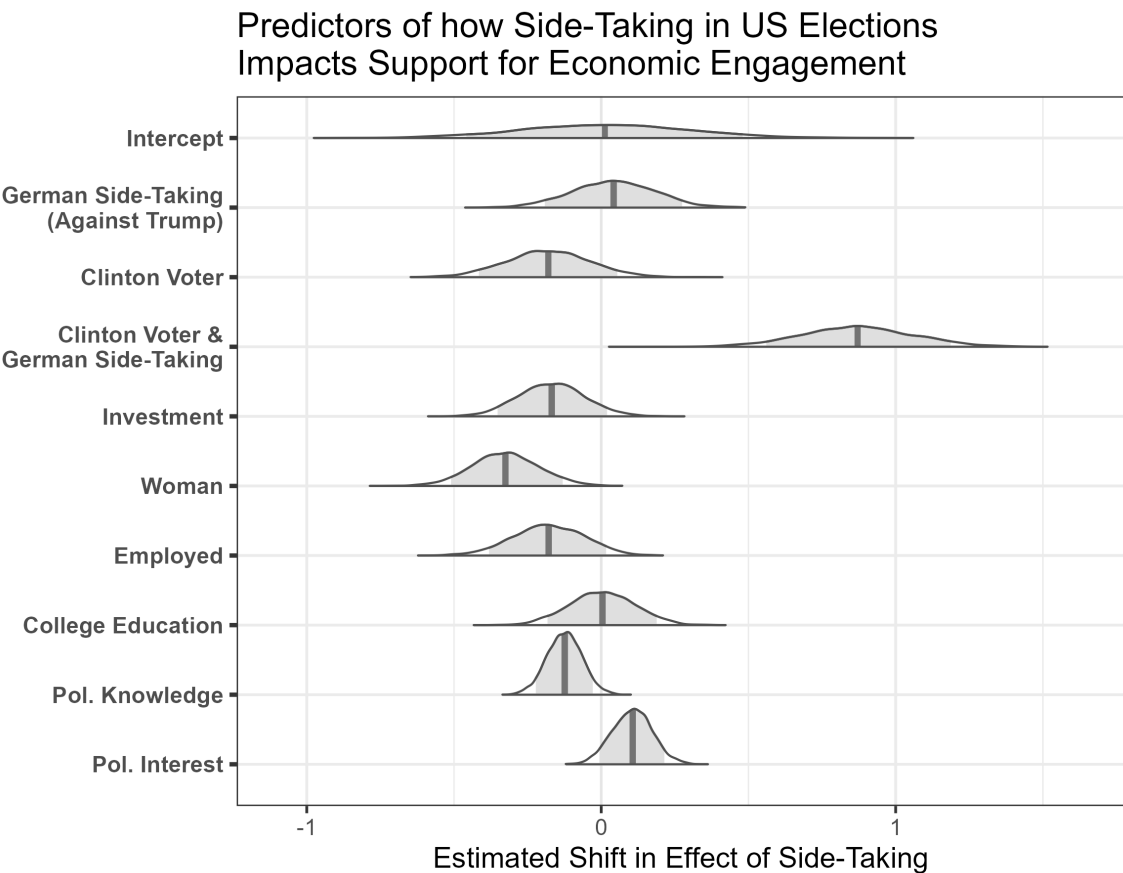
Predictors of how Side-Taking in US Elections Impacts Support for Economic Engagement

Figure 1

Figure 1 summarizes the estimates from the heterogeneous effects equation. As BP predicted and find, side-taking increases support for economic engagement with voters whose candidate the intervention supports. Relative to the impact of Russian side-taking on Trump voters, German side-taking increases the impact of side taking by almost 1, which is a large difference on an outcome that ranges from 1 to 4. German side-taking does not impact Trump voter support for economic engagement, and Clinton voters are marginally but not clearly less

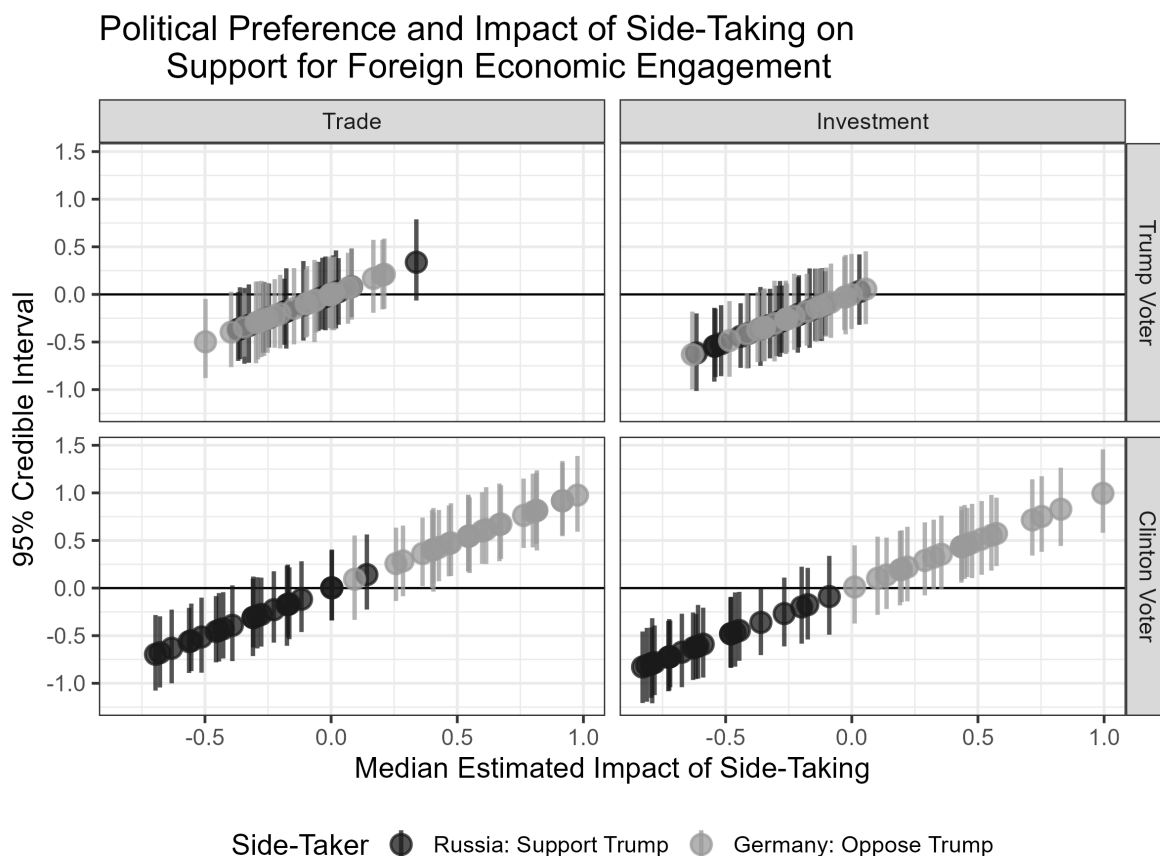supportive of engagement with Russia after Russian side-taking. These estimates are consistent with BP's results.

Figure 2. *Heterogeneous effects of foreign side-taking in US elections on support for economic engagement with the interfering country. Each estimate reflects a treated group with a unique combination of other treatments and demographic characteristics.*

The model also provides some new information about how side-taking impacts support for economic engagement. Electoral side-taking is more likely to reduce support for foreign investment in the United States, all else equal. Moreover, women largely respond more negatively to side-taking. Finally, increasing political knowledge decreases the the impact of side-taking, while interest in politics increases it.

Figure 2 substantiates these inferences and presents the group-specific treatment effects. In general, Clinton voters were more responsive to side-taking. Most Clinton voters were

less likely to support trade and investment with Russia after exposure to Russian support for Trump, and more likely to back economic cooperation with Germany in response to German opposition to Trump. Additional variation in this relationship comes from gender and political engagement. Some Trump voters reduce their support for investment regardless of who the intervention backs and are less likely to change their trade attitudes in response to side-taking as well.

The credible intervals on these estimates suggest that researchers should not rely on this technique to generate powerful tests between groups, unless predictors of heterogeneous effects create substantial differences. For instance, most Trump voters hold very similar views of economic engagement, and there are only a few that are clearly different. Among Clinton voters, however, the response to German side-taking and corresponding distaste for Russian side-taking creates huge differences between different treated groups.

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 new arguments and identify who responds best to a given intervention.

Again, this approach is a complement for existing tools, not a substitute. Hierarchical modeling provides an intermediate approach between simple interactions or subgroup analyses and complex machine-learning algorithms. Researchers can thus use this tool to check other results or inform those 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.

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

Bush, Sarah Sunn and Lauren Prather. 2020. "Foreign Meddling and Mass Attitudes Toward International Economic Engagement." *International Organization* 74(2):584–609.

Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li and Allen Riddell. 2016. "Stan: A probabilistic programming language." *Journal of Statistical Software* 20:1–37.

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.

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.

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.