

An Analysis of Intending to Seek Treatment for a Mental Health Issue

STA2201 - Final Research Project

Maksim Helmann

12/04/23

The ability to anticipate an employee's inclination to seek treatment for mental health problems could significantly improve the offered mental health care by the tech companies by allowing for early interventions and tailored support for the respective employee. To achieve this, we utilized the dataset from the Mental Health in Tech Survey and fitted a Bayesian logistic regression model to estimate participants' intentions to seek treatment for mental health issues. Although the model we developed produces mostly accurate predictions, we note that additional latent variables, including interaction terms, may be necessary to fully address this objective.

Introduction and Related work

Mental health and physical health are the two major categories of health, whereby, the former fosters a balance between emotional and mental well-being and illness. Thus, it is as important as physical health because having a well mental health is essential to navigate smoothly through every stage of life, from childhood, through teenage years, and into adulthood [1]. Unfortunately, mental health is not accorded the same level of importance as physical health but identifying the need for treatment for individuals with mental health disorders can be a crucial step in fostering their overall well-being. This is particularly important in tech companies where employees are exposed to significant levels of stress, high work pressure, and other mentally burdensome factors. Since addressing mental health issues in tech companies can have several positive outcomes, including reducing the incidence of suicides, facilitating faster reintegration into the

workplace after taking time off for mental health reasons, and lowering the number of individuals who become unemployed due to mental health issues.

Therefore, our main research question aims to explore how demographic factors, such as age, gender, and location, as well as work-related factors, such as self-employment and employment at a tech company, influence an individual's inclination to seek treatment for a mental health condition. A further related work that solely focuses on the analysis of the data is covered in the paper "Mental Health Analysis in Tech Workplace" [2].

Data

To address the research question at hand, we will utilize the dataset [Mental Health in Tech Survey](#) from a 2014 survey that gauges attitudes towards mental health and the prevalence of mental health disorders in the tech workplace. This dataset has in total 27 variables that cover different categories, i.e., demographic, employment, mental health aspects, mental health benefits and support at work and attitudes and perceptions towards mental health at work. The variable of interest, among the listed variables, is "treatment", which reflects whether the individual has sought treatment for a mental health issue or not. Prior to running analysis of the data some data cleaning is performed. More precisely we only select the wanted variables (10 out of 27), where 9, namely gender, country, self-employed, family history, work interference, remote work, tech company, anonymity, age group, depict the independent and 1 the dependent variable. One of the reasons is that after performing the Fisher's exact test some of the variables turn out to be highly correlated. Next, the lack of computational power forced to choose wisely which variables to include. In addition, due to the desire to look at a more coarse level the variable, e.g., states is excluded. Looking at the selected variables, further processing steps are executed. For countries only a subset, Canada, Germany, US and UK, is picked because these countries had at least 40 observations and their population size, as well as their development status and living standard, are more or less similar. The neglected countries had only one observation recorded. Furthermore, the gender entries are coded as "male" and "female", as some showed different notations, e.g., "femail", "F", "women".

Next step is to look for missing values. The only two variables with missing values are self-employed (17 missing values) and work interference (209 missing values). The 209 missing values for work interference represent approximately 1/5 of all recorded observations. Moreover, all participants who did not answer this question are not seeking treatment, except for one participant. Hence, we cannot drop these observations, as the proportion for the group "not-treated" would be significantly influenced. One possible reasoning for the unanswered question is that the participants cannot make a statement

because they do not experience mental health issues. Therefore, we will decode the missing values as “Don’t know”. For the 17 missing entries of the self employed variable we first investigate whether the values are missing at random or not. A check of the data frame shows that those observations do not show any patterns and thus, removing them does not lead to concerning changes in the data distribution compared to the one observed previously.

The age variable is binned into groups to provide more data points for each age category, enabling us to draw more reliable conclusions about the outcome variable. If we were to analyze the data using individual age values, we would need a sufficiently large sample size to obtain statistically significant results. The age groups are 18-25, 26-32, 33-40, 40+. The split is done because for the first age group there can be individuals that just finished their studies so most likely a lot of them are just experiencing the work life. The second group address participants where people are already working for a longer period of time then the participant in the first age group. For the second last group, we assume that respondents are employed for a long period of time or that they have changed their jobs. Furthermore, the age groups 26-32 and 33-40 are the one where some employees get children, this can also affect the daily life stress which is connected to mental health issues. Last group includes participants that most of the time settled down and there were not many respondents for higher ages.

Observing the proportions of participants that work remotely, we notice that only 27% work remotely, whereby, $\sim 16\%$ sought treatment. The proportion of tech company employees sums up to 80%, 40% of them stated that they sought treatment. Further, 60% indicated that they experienced work interference due to a mental health issue.



Lastly, we want to investigate the proportion of participants seeking treatment by age group, country and gender (figure on the left) and at the same time pay attention to the corresponding sample size by age group, country, and gender (figure on the right). It should be noted that the number of male participants (78%) is approximately 4 times that of women (21%). First, observe that largest number of participants comes from the US followed by the UK. Moreover, for the male participants each cell has at least three respondents and the proportion of sought treatment is not zero as well. The two age groups which account for the biggest sample size are 26-32 and 33-40. Another observation is that despite the smaller sample size of female participants all the proportions of sought treatment for female are above 50%, except Germany where only one female participant is recorded that did not seek treatment for mental health issues. In general, there seem to be a slight relationship between the proportion of sought treatment and age group, especially for the countries US and UK. While this trend is positive for the mentioned countries, we observe that for Canada the trend is negative, in the sense that with increasing age group the proportion of sought treatment declines. These last observations were regarding the male participants because the sample size distribution for the female participants is a bit different compared to the male one. For female respondents, despite the mentioned fact that the proportions for seeking treatment are above 50%, we observe that the trend is mostly constant with some fluctuations. Although not shown, similar figures were produced for the remaining predictors of interest. Some observations are that for participants that are not self employed the proportion of sought treatment has its average around 50%, except for Canada where the proportion is 60%. For participants with a family history of mental illness, the proportion of sought treatment is 70% for the age group 26-32 and 33-40. The other two age groups, 18-25 and 40+, depict a value of at least 50%.

Methods

In order to analyze the intention to seek treatment for mental health issues, two Bayesian models will be fitted in Stan. The first model will be used as a baseline model, as this helps to evaluate the usefulness of the additional complexity introduced by the hierarchical model later. Thus, the baseline model will be a simple fixed-effects model and include all variables mentioned in section 2, namely age, gender, country, self employed, family history, work interference, remote work, tech company and anonymity, as predictors in the model. Given that the response variable “treatment” only has two possible outcomes (“Yes” or “No”), it is appropriate to use logistic regression to model the binary nature of the response variable. As the survey records a one-time response of individual participants, the Bernoulli distribution is a suitable probability distribution in this case. In the following, the fixed-effects model is defined.

$$\begin{aligned}
y_i | \pi_i &\sim \text{Bern}(\pi_i) \\
\pi_i &= \text{logit}^{-1}(\beta_0 + x_i^T \beta) \\
\beta_0, \beta &\sim \text{N}(0, 1)
\end{aligned}$$

where y_i depicts the outcome variable that returns the value 1 if the participant seeks treatment for a mental health issue, and 0 otherwise. π_i is the parameter of the Bernoulli distribution, typically called the “success probability”, here for participant i . The vector x_i contains the values of all predictor variables for participant i . β_0, β correspond to the intercept and the coefficients of the predictors, respectively, whereby, standard normal priors are placed on all of them.

The model of interest (second model) will be a Bayesian non-nested hierarchical model. This modeling approach will include fixed-effects for the following independent variables, namely gender, self-employment status, family history of mental illness, work interference, remote work status, tech company employment, and treatment anonymity. In addition, a hierarchical structure will be incorporated for age group and country to account for structured patterns related to these factors. By modeling these dependencies, the resulting model can estimate unique predictive effects for the participants. Additionally, due to the layout of this model, further priors must be placed on the variance parameters of age group and country. In the EDA part, we observed that for all age groups, or at least for the first three, the proportion of participants seeking treatment depicted a slight relationship, which is an increase in proportions over the age groups. To utilize this information, we will use first-order random walk for the prior of the age group intercept, starting from the second age group. For the first group, 18-25, a standard normal prior is set. The hierarchical model is mathematically formulated as follows.

$$\begin{aligned}
y_i | \pi_i &\sim \text{Bern}(\pi_i) \\
\pi_i &= \text{logit}^{-1}(\beta_0 + x_i^T \beta + \alpha_{j[i]}^{\text{age}} + \alpha_{k[i]}^{\text{country}}) \\
\alpha_j^{\text{age}} &\sim \text{N}(\alpha_{j-1}^{\text{age}}, \sigma_{\text{age}}^2), \text{ for } j = 2, \dots, 4 \\
\alpha_k^{\text{country}} &\sim \text{N}(0, \sigma_{\text{country}}^2), \text{ for } k = 1, \dots, 4 \\
\sigma_{\text{age}}^2, \sigma_{\text{country}}^2 &\sim \text{N}^+(0, 1) \\
\alpha_1^{\text{age}} &\sim \text{N}(0, 1) \\
\beta_0, \beta &\sim \text{N}(0, 1)
\end{aligned}$$

where y_i depicts the outcome variable that returns the value 1 if the participant seeks treatment for a mental health issue, and 0 otherwise. π_i is the parameter of the Bernoulli distribution, typically called the “success probability”, here for participant i . The vector

x_i contains the values of all predictor variables for participant i , except for age group and country.

$$\alpha^{\text{age}}$$

depicts the age-specific intercept and j is an index for the age group. α^{country} depicts the country-specific intercept and k is an index for the country. The variance parameters of age group and country are depicted as $\sigma_{\text{age}}^2, \sigma_{\text{country}}^2$ and are initialized with standard half-normal priors. α_1^{age} depicts the first age group and is assigned a standard normal prior. β_0, β correspond to the intercept and the coefficients of the predictors, except for age group and country effects, respectively, whereby, standard normal priors are placed on all of them.

To fit the model in Stan an appropriate data frame needs to be provided. Since we are using a vectorized model our vector x_i will be provided in form of a matrix, whereby, each row represents the data for a participant, namely the mentioned predictor variables. A further possible adjustment would be to add a unit column vector to the matrix which is reserved for the intercept parameter. In this case, the intercept value is added simply through addition, separately during the simulation. To elaborate more on what is meant by providing all categories of the predictor variable, it should be said that when all categories of one predictor are given to the model, the problem of multicollinearity arises because one of the categories can be simply deduced from the others. Thus, this category does not provide any new information. Therefore, one category from each explanatory variable will be excluded and serve as the reference category for that variable later when interpreting the key coefficient estimates. This results in the following baseline category: female participant, who does not know if it interferes with her work, is self employed, has no family history of mental health issues and does not know if taking mental health care remains anonymous at the tech company.

Furthermore, during the simulation, the code will generate replications of the dataset using the obtained posterior samples and it will also calculate the point-wise log-likelihood estimates. After, fitting the model in Stan using 4 chains and 1000 iteration for each of them, the next step is to look at the summary of the model to check whether the effective sample size is large enough and \hat{R} is smaller than 1.05. In addition, traceplots and pairs plots [3] need to be analyzed to make sure that the chains are mixing well and that they are converging to the stationary distribution. Lastly, the sampling behavior is adequate, meaning the sampler can access the entire parameter space for sampling. Once these quantities are looking good, posterior predictive checks are performed to assess the adequacy of the model's fit to the data. Specifically, we will inspect the `ppc_dens_overlay`, binned residual, `ppc_stat` and `ppc_stat_grouped` plots. Moreover, to justify that the hierarchical model performs better than the baseline model, the LOO elpd will be calculated for each model and compared. When all the checks and

validations return satisfactory results, we are off to a good start to try using this model for predictions.

Results

Before interpreting the final key coefficient estimates, their CIs and looking at the corresponding the diagnostics, we looked at three different models. The difference between them is depicted in how the priors for the effects of age were set. Recalling the EDA findings for the relationship between the proportion of participants seeking treatment across the age groups, there are three possible ways to set the priors. First, set the effects for age such that they are centered around zero. Second, an assumption is made that the age effect in group j is similar to that in the previous age group, plus some noise, which corresponds to modeling the age effect via a first order random walk (this model is mentioned in the Methods section). Lastly, a second-order random walk can be placed on the age effect, assuming that there is some structure over age with longer memory than the one captured by a first-order random walk. After the simulations were done, the convergence checks showed that, for the first option, the highest correlation within the age group effects, compared to the other options, is depicted. Additionally, the country effects showed high correlation too, which is not observed for the other fits. For the model with the second-order random walk, the model had too many divergent transitions after warmup even for a high iteration number and a low Tail Effective Samples Size (ESS). Furthermore, for both options, the t-statistics provided worse results than the model with a first-order random walk on the age group effects. In light of these findings, the final model includes the second option, which applies a first-order random walk to the age group effects. With respect to the final model, we found that the inclusion of the tech company covariate worsens the fit of the model to the data, as shown by all possible diagnostics and posterior predictive tests presented later for the final model with the tech company variable omitted. The most crucial observation was made for the binned plot where, for the two shown plots, clustering of points is depicted. This indicates that the model is most likely not adequately capturing some underlying structure in the data. Therefore, an analysis of the model is performed where we exclude variables to observe whether there is an improvement in the diagnostics. The analysis showed that the explanatory variable that caused this issue is work interference. Thus, after removing this covariate, we obtained a much better result, which is presented below.

As a starting point for evaluating our final hierarchical model, we want to look at the model summary. More precisely, we are interested in the key parameters of interest along with the corresponding 95% credible intervals (CI), depicted on the odds scale.

Table 1: Estimated Beta Coefficients from Fixed Effects Model (Odds Scale)

predictor	mean_beta	lower_beta	upper_beta
male	0.3885883	0.2751505	0.5364513
self_emplo_not	1.0512029	0.6926334	1.6103800
fam_yes	0.9774772	0.1424461	6.4663157
remote_not	1.0721175	0.7862228	1.4079175
anonym	2.0094150	1.5150247	2.6391447
anonym_not	2.1078620	1.0447708	4.1844806
intercept	1.1635950	0.7826828	1.7556704
age_26-32	1.3188941	0.8701298	2.0052148
age_33-40	1.7403755	1.0904489	2.8518876
age_40+	1.0279229	0.1361301	7.9937747
coun_UK	1.0187329	0.1320277	6.8255543
coun_CAN	1.0093302	0.1364643	7.4404072
coun_GER	1.3854887	0.7593404	2.4865542

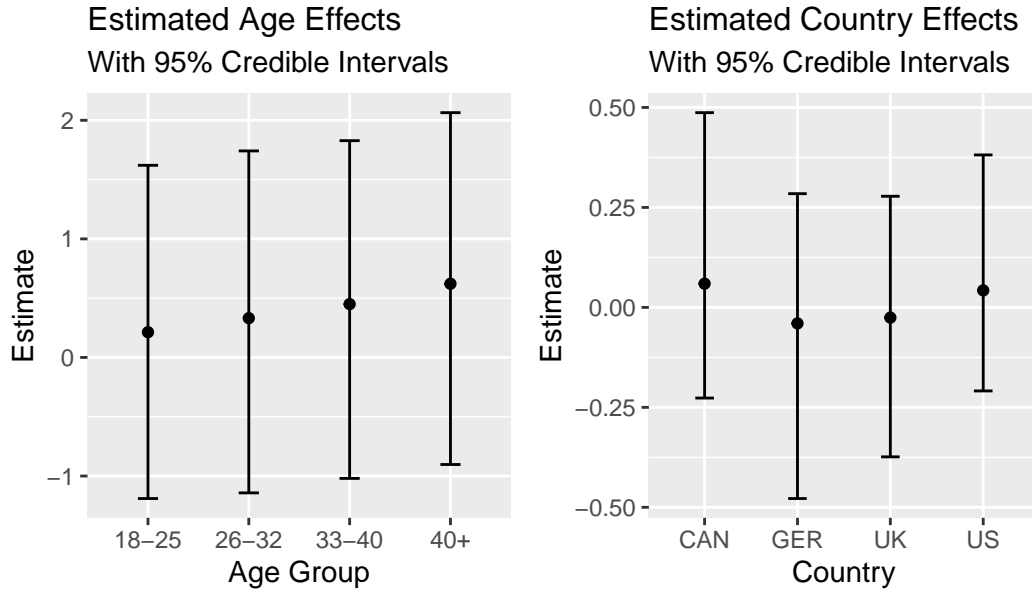
Especially of interest is to observe whether 1 lies within the CI bounds because this indicates that the effect of the predictor variable is not statistically significant. In other words, this suggests that the predictor variable may not be a significant predictor of the outcome variable and may not be important in explaining the variation in the response variable. The two tables show that the CI bounds for the male coefficient have similar range and magnitude values and that the CIs do not cover 1. Additionally, the mean value of the mentioned coefficient is below 1. Other variables, such as non-remote working and non-self-employed, show relatively narrow confidence intervals (CIs), which do cover 1. However, the difference for the coefficient of non-remote working is that its mean value is below 1 for the hierarchical model. Another similarity can be observed in the CI bounds of the family history coefficient, which cover 1 in both tables. An important observation is that the mean value of the family history coefficient is equal to approximately 1. The CIs for the coefficients of the anonymous and non-anonymous variables are above 1, as well as the mean values. Finally, the mean value of the intercept estimate is above 1 for both models but for the fixed-effects model, the CIs are much narrower than for the hierarchical model. Nevertheless, for both intercepts, the CI bounds cover 1.

For the fixed-effects model, the age group and country covariates, which are also included in the vector x mentioned in the Methods section, the table shows that the estimates for the age group coefficients cover 1 with their CI bounds, except for the 33-40 age group. For the countries all CIs bounds do cover 1. The hierarchical model estimates and 95%

Table 2: Estimated Beta Coefficients from Hierarchical Model (Odds Scale)

predictor	mean_beta	lower_beta	upper_beta
male	0.3956232	0.2883715	0.5404504
self_emplo_not	1.0317848	0.6603640	1.6219318
fam_yes	1.0061099	0.1506184	6.7449148
remote_not	0.9376423	0.7017657	1.2665236
anonym	2.0004071	1.5205453	2.6236253
anonym_not	2.1032111	1.1032241	4.2836345
intercept	1.2247533	0.2929706	5.2275571

CIs for the age group and country specific intercepts are depicted in the extra figure. For the estimated age group specific intercepts, the depicted mean values are above zero. Additionally, the CI bounds are wider than for the country specific effects but the CI bounds for both effects do include the value zero. The mean value of the estimated country specific intercepts is below zero for Germany and UK and above zero for the Canada and the US. Comparing this to the plots presented in the EDA, we observe that these estimated effects values are aligning well. There might be small over- and underestimation that can be partially attributed to the use of the first-order random walk effect chosen to model the priors of the age group effects but overall the odds of seeking treatment are higher than the odds of not seeking treatment for all age groups. Country effects also show a similar trend to the observations obtained from looking at the figure in the Data section: individuals in Canada and the US have higher odds of seeking treatment than not seeking it, while individuals in Germany and the UK have higher odds of not seeking treatment than seeking it.



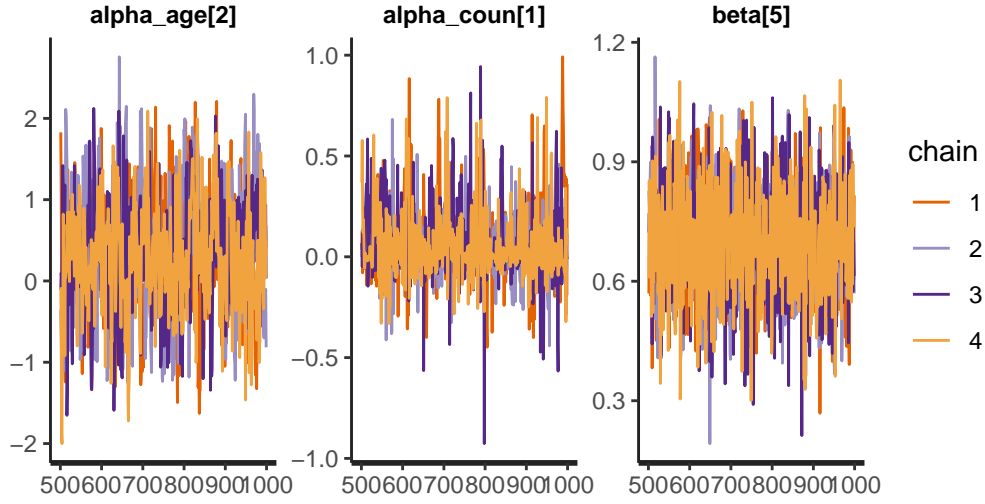
After taking a look at the key coefficient estimates and their CIs, we want to interpret them. Since logistic regression is carried out, the coefficients must be exponentiated in order to interpret them on the odds scale. We will only present the interpretation for each categorical variable depicted in the table for the hierarchical model, as this is the model of interest.

- The odds of a participant who is not a self-employed individual that and is female, works remotely, has no family history of mental health issues and does not know if taking mental health care remains anonymous at the tech company, to indicate intention to seek mental health treatment is ≈ 1.03 times that of a non-sometimes work interference participant with the other characteristics remaining the same.
- Next, for a non-remote working person who is female, is self-employed, has no family history of mental health issues and does not know if taking mental health care remains anonymous at the tech company, the odds of seeking treatment are ≈ 0.94 times that of a remote working person who possesses the other mentioned characteristics.
- At companies where seeking mental health care remains anonymous, a person who is female, works remotely, is self-employed and has no family history of mental health issues has approximately twice the odds of seeking treatment compared to a person who works at a company where it is not clear if seeking mental health care remains anonymous and possesses the other mentioned characteristics. Similar

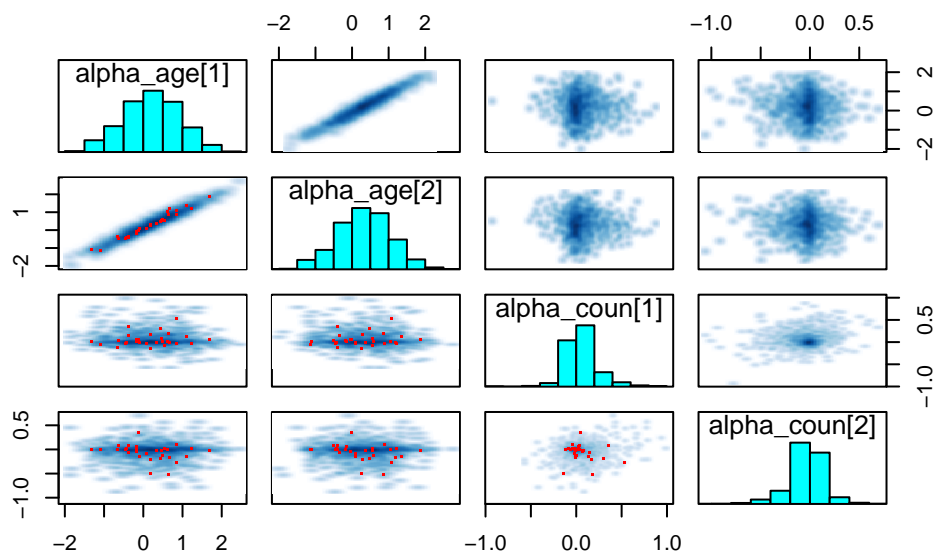
odds are observed for the same person where seeking mental health care is not anonymous.

- The male coefficient provides the odds of a male participant who is self employed, works remotely, has no family history of mental health issues and does not know if taking mental health care remains anonymous at the tech company. These odds, that the male participant will intend to seek treatment, are ≈ 0.4 times that of a female respondent with the other mentioned characteristics.
- Last but not least, the odds of a participant with family history of mental health issues who is female, is self employed, works remotely and does not know if taking mental health care remains anonymous at the tech company, to seek treatment are ≈ 1 times that of respondent who has no ancestors with mental health issues and posses the other mentioned characteristics. For the predictor family history, we can conclude that this covariate might not be significant. Thus, an additional model is fit to assess whether this variable can be dropped. The result showed that by excluding family history from the model, the correlation between the country-specific intercepts could be slightly reduced but overall, no significant improvement was observed. Therefore, the model remains the same and this observation can be included into further work.

The diagnostic checks for both models show that the \hat{R} values are all smaller than 1.05. In addition, S_{eff} is greater than 100. This further confirms that there are no significant convergence problems in either model. Besides this, we examine the sampling behavior via the trace and pairs plots. This is a necessary step to check if the chains are mixing adequately well and whether the parameters are correlated, depicted through very narrow, elliptical-shaped space, as this is a cause for concern because this means that the sampler does not have access to the entire parameter space to sample from. It is important to note that the analysis was conducted on all estimates from both models. The traceplots of all estimates for the fixed-effects and hierarchical model depict a good chain mixing and no signs of issues regarding convergence. However, to maintain clarity for the reader, only a subset of the trace plots for the estimates will be displayed. Below, three traceplots that belong to the hierarchical model illustrate that for each (randomly chosen) parameter i of β , α^{age} and α^{country} , the chains seem to mix quite well.



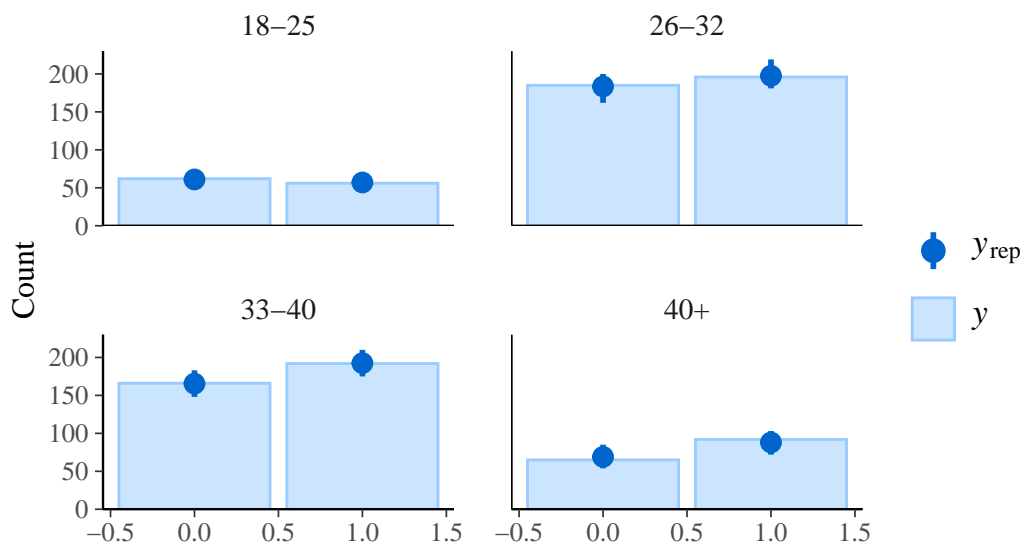
Next, the pair plots will be analyzed. In the fixed-effects model, the histograms exhibit a unimodal distribution of observations centered around a certain point. The bivariate scatterplots do not depict a narrow distribution of points. Therefore, the sampler is able to effectively access the complete sampling space. For the hierarchical model, the pairs plots look well for all estimates β but not that good for the estimates of α^{age} and α^{country} . To illustrate the issue, the pairs plot for two α^{age} and two α^{country} parameters is shown. We observe that the histograms for α^{age} do not depict any problematic behavior but the bivariate scatterplots depict a narrow, elliptical-shaped space between the two α^{age} parameters. This is concerning because it suggests that the sampler lacks access to the entire parameter space to sample from. For α^{country} parameter, the histograms appear to be narrower, indicating a lesser spread of observations around the central point. The corresponding scatterplots do not show an ideal point cloud but are still reasonably well spread out than the scatterplots observed for the α^{age} parameters. No strong relationship is observed between the α^{age} and α^{country} parameters.



The first posterior predictive check illustrates that our estimations do a fairly good job of fitting the true binary nature of the response variable. We observe that for the age group 18-25 the y_{rep} medians are either slightly below or above the corresponding y value. A similar observation holds for the oldest age group. For the age groups 26-32 and 33-40, the largest uncertainty intervals are returned.

Distribution of Observed vs Predicted Sought Treatments

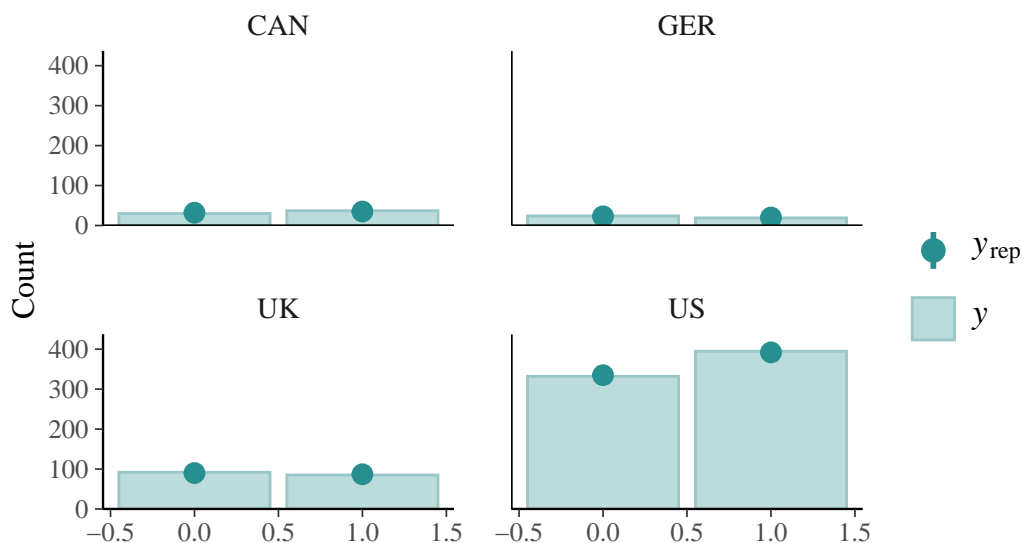
By age group for non-nested hierarchical model



The median values of y_{rep} do correspond well with the y value for the countries. For the US, we observe the largest uncertainty intervals.

Distribution of Observed vs Predicted Sought Treatments

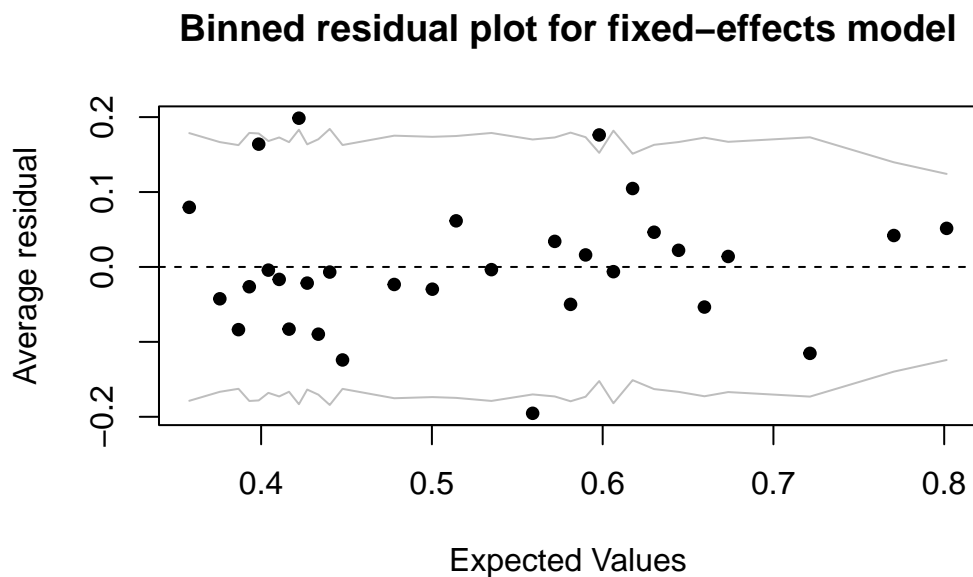
By country for non-nested hierarchical model



In the next step, the LOO elpd (expected log pointwise predictive density) for each model is calculated and compared, whereby the point-wise log likelihood estimates for

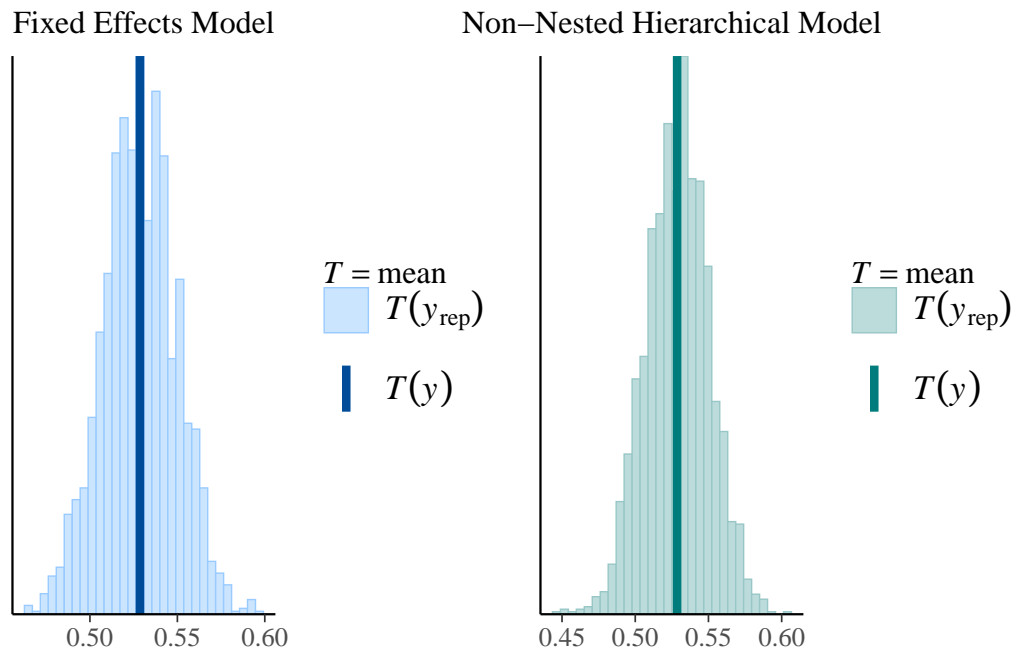
each model are calculated in Stan during the simulation. The result showed that the expected predictive accuracy of the models were quite close in value. To be precise, the ELPD of the hierarchical model exceeded that of the fixed effects model by 0.3 units. However, the standard error linked with this difference is 0.9, which is greater than the difference in ELPD's. Therefore, it can be inferred that there is no strong preference for the hierarchical model compared to the fixed effects model. Looking at the PSIS diagnostic plot, we can observe that the Pareto shape k values do not depict any outliers (k values above 0.7). One interesting observation is that for the fixed-effects model, there is an accumulation of the Pareto shape k -values, e.g., around -0.09. Please refer to the Appendix for the plots.

Another possibility to assess the overall model fit is by looking at the binned residual plot. The two shown plots depict a mostly good distribution of the binned residuals. This indicates that the model is most likely able to adequately capture some underlying structure in the data. Further, for the hierarchical model, we observe slightly fewer points outside the ± 2 standard error bounds than for the fixed-effects model. Thus, judging by the binned residual plots, the hierarchical model appears to be doing a slightly better job of fitting the data.





Final check is to compare the mean outcome of the replicated data to the mean outcome of the original data for each level of education and age group, via the `ppc_stat_grouped()` function. The resulting histograms for fixed-effects model show that the distributions is nearly centered around the true mean outcome for each age group. The biggest deviation is observed for the age group 18-25. For the countries, the model fails to capture the true mean for Canada and Germany. Histograms for the age group of the hierarchical model show a distribution nearly centered around the true mean outcome, largest deviation can be observed for the oldest age group. For the countries, the histograms are narrower than those of the fixed-effects model but for Germany, the true value is still not well captured. Please see the Appendix for the plots of both models. Overall, this means that the true observed mean is captured quite well by the obtained distribution, suggesting a good fit to the true data. Below is a comprehensive overview of the posterior predictive check, provided through ungrouped histograms. The blue histogram on the left represents the fixed-effects model, while the green histogram on the right corresponds to the hierarchical model. For both models, the predicted mean proportion of participants seeking treatment aligns well with the one from observed data.



Discussion

Although the presented non-nested hierarchical model shows some slightly better results regarding the data fit when looking at the age group and country effects, the overall results are not significantly better than those of the fixed-effects model as observed by the posterior predictive checks. The model fit to the data improved when the work interference variable was excluded, as evidenced by the absence of grouping in the binned residual plot. Furthermore, we note that all posterior predictive checks showed no grave signs of concern. Additionally, the convergence checks revealed no issues with, for example, sampling from the stationary distribution. Nevertheless, we should not forget the saying of the late British statistician George Box: “all models are wrong, but some are useful” [4]. In this case, the hierarchical model allows us to estimate more accurate odds for the intention of participants to seek treatment for their mental health issue because it enables us to directly estimate the intention for seeking treatment by age group and country. Further, the estimates are not far off from the observed data. However, the complexity of the model is most likely too simple to completely rely on the estimated odds, since mental health issue analysis poses many more latent variables than those we can capture with the data at hand!

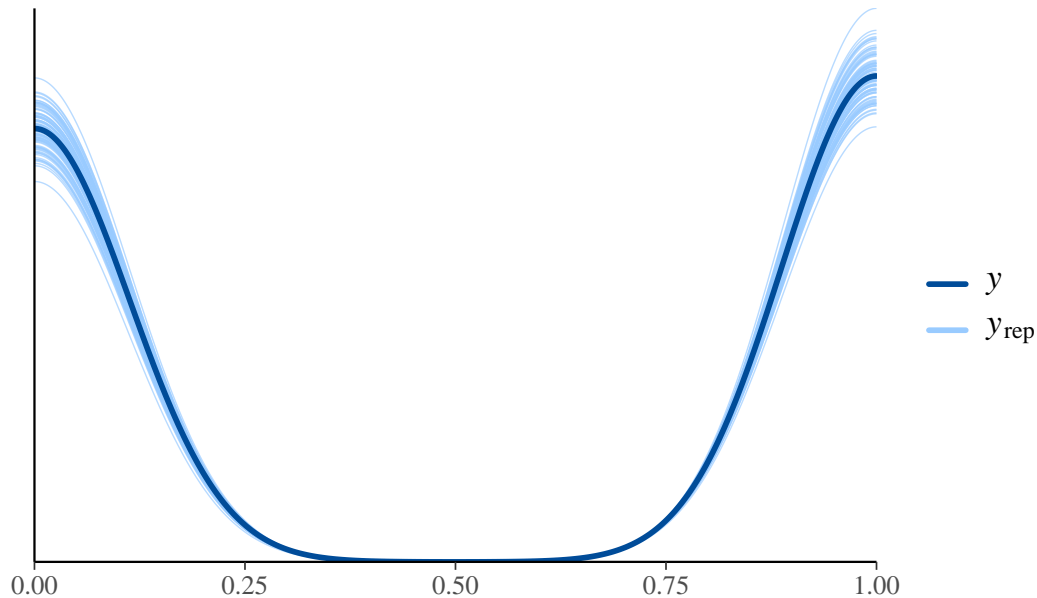
A surprising result is that the covariate tech company was not significant and even produced a worse model fit to the data. Our expectation was that this predictor variable

would be important as it highly relates to the research question but it turns out that most likely the information can be deduced from the remaining covariates. The result for the work interference variable is somewhat expected, as this variable has a strong impact on the outcome. People who are impaired at work due to mental health problems, hopefully, seek treatment.

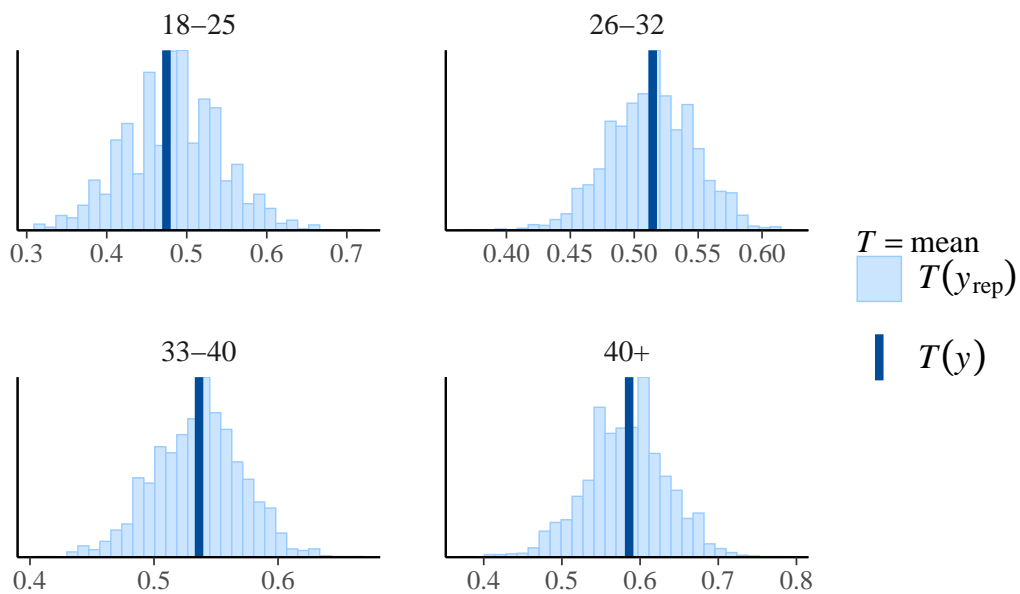
Potential future work, if we had more data or more time, could be to investigate interaction terms, as a possible improved estimation for the intention to seek treatment can be obtained through a latent relationship between the two predictor variables. Nevertheless, the added explanatory variable should still relate to the fact that we want to estimate the intention of seeking treatment for a mental health issue. Another area of interest would be to consider states within the countries if data for this would be available. Finally, comparing the results to newer data would be really interesting because nowadays the tech company situation has changed again compared to 2014.

Appendix

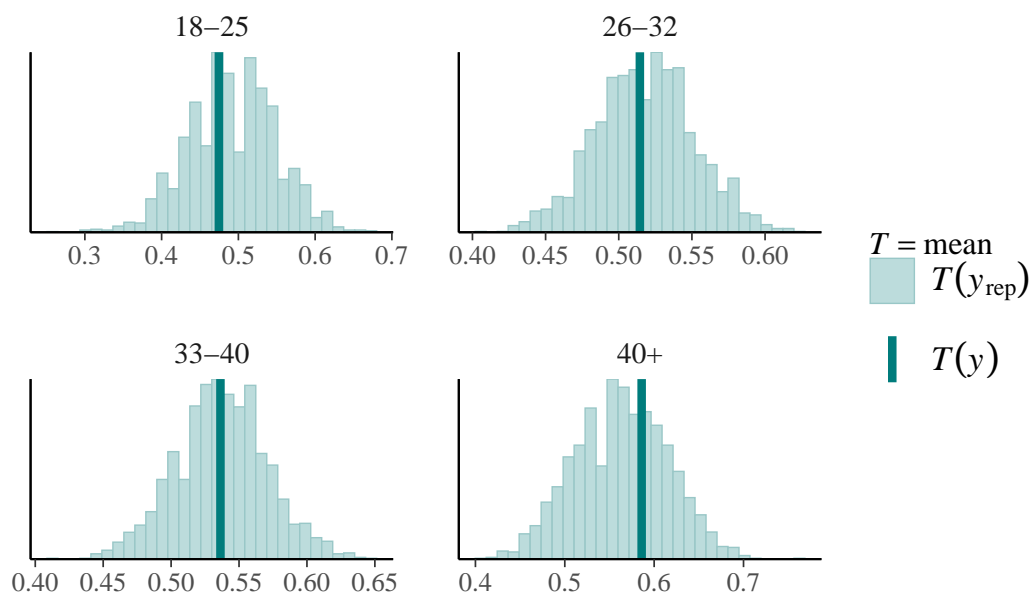
Distribution of Observed vs Predicted Sought Treatments



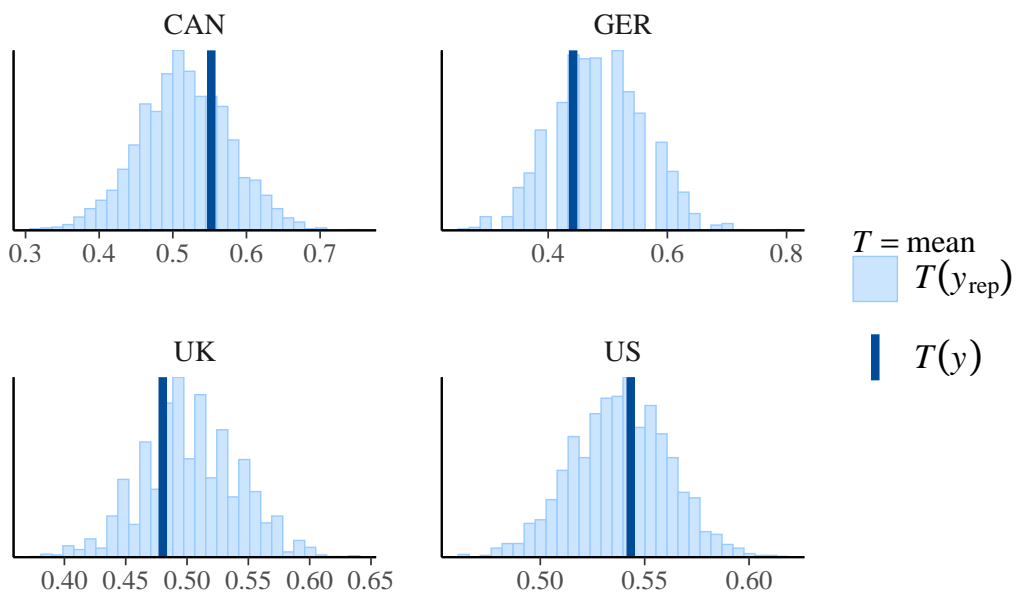
Fixed Effects Model grouped by age group



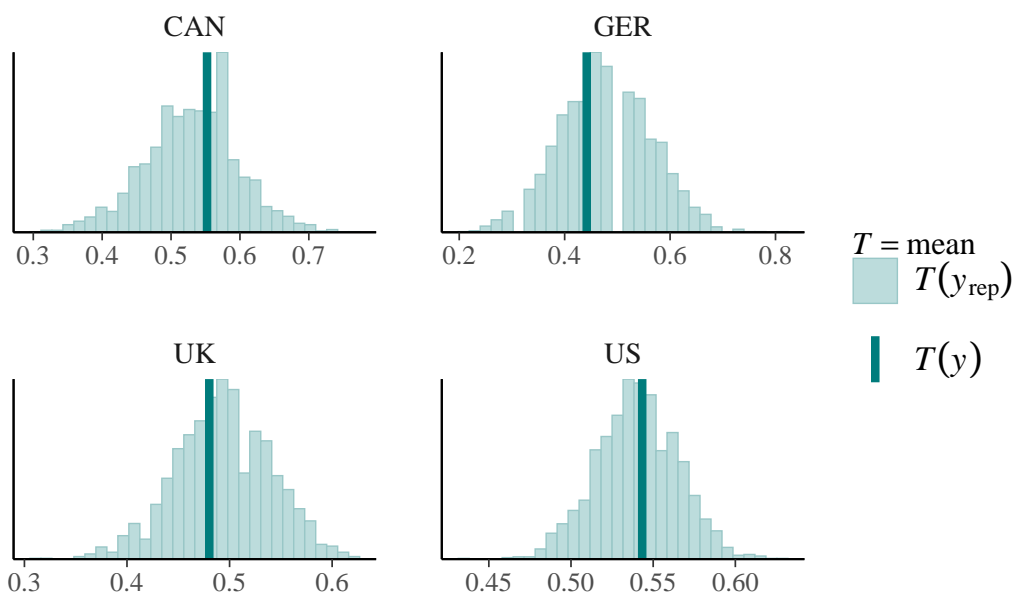
Non-Nested Hierarchical Model grouped by age group



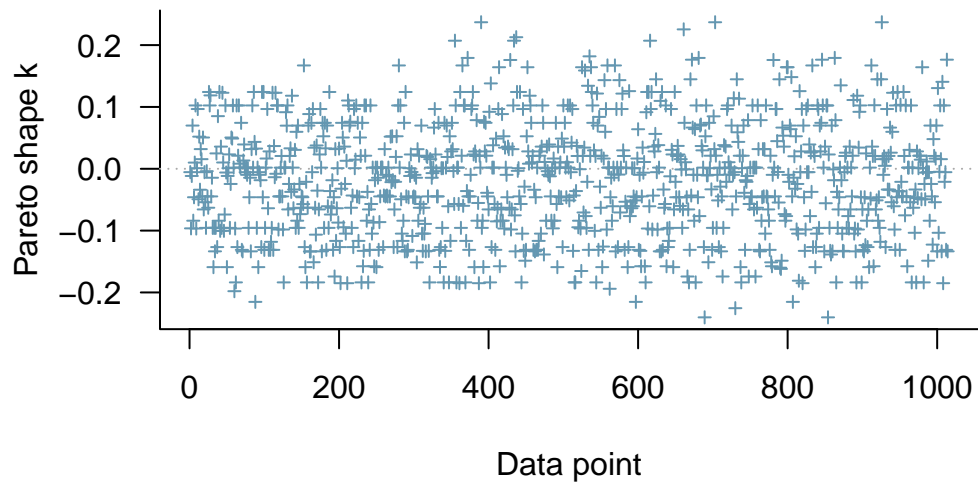
Fixed Effects Model grouped by country



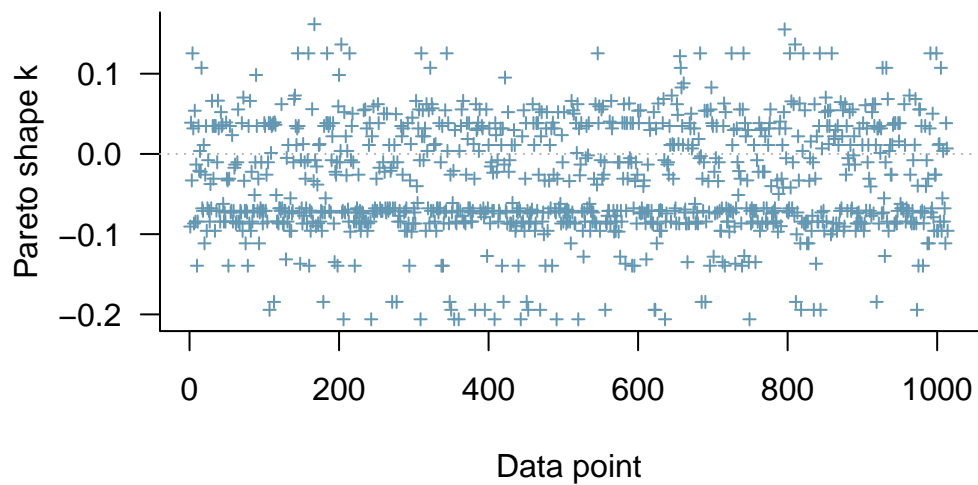
Non-Nested Hierarchical Model grouped by country



PSIS diagnostic plot for non-nested hierarchical model



PSIS diagnostic plot for fixed-effects model



References

- [1] BH, S., K, N., C, D., P, B., DURAIPANDIAN, K., RAJAN, S. and SATHIYANARAYANAN, M. (2022). [Mental health analysis of employees using machine learning techniques](#). In *2022 14th international conference on COMmunication systems & NETworkS (COMSNETS)* pp 1–6.
- [2] UDDIN, M. M., FARJANA, A., MAMUN, M. and AL MAMUN, M. (2022). Mental health analysis in tech workplace.
- [3] GABRY, M., J. & Modrák. (2023). Visual MCMC diagnostics using the bayesplot package. Retrieved april 03, 2023, from <https://cran.r-project.org/web/packages/bayesplot/vignettes/visual-mcmc-diagnostics.html>.
- [4] BOX, G. E. P. and DRAPER, N. R. (1987). *Empirical model-building and response surfaces*. Wiley, New York.