

Homework Assignment 4

Greg Forkutza
Student ID: 400277514

18 December, 2023

1

In table 1 we see the results of the two different models. Comparing the difference in coefficients and standard error estimates, we see that for the two non interaction terms **glmer** overestimates compared to the **gam** fit and underestimates for the interaction terms coefficients and overestimates for the standard error. Looking at the magnitude of the differences, which were computed relative to the **gam** fit, we see that that the non interaction terms coefficient estimates are very similar and their standard errors are practically identical. The same is true for the interaction term except that the coefficient estimate is an order of magnitude closer more similar.

gam uses IRLS and **glmer** uses Laplace Approximation (LA). IRLS is robust to estimating complex, non-linear effects, which might explain the slightly lower estimates for the interaction term. LA is has conservative estimates when dealing with complex interactions or non-linearities, potentially leading to the observed slight overestimation in non-interaction terms.

Table 1: Comparison of fitting `gam()` and `glmer()` to the `mlmRev::contraception` dataset. The response use is converted to a binary numeric factor and age is scaled and centered. The fixed effects are age and urban and their interaction. Random effects include a random intercept for the district. The difference is computed relative to the `gam` fit.

	Estimate_glmer	StdError_glmer	Estimate_gam	StdError_gam	EstimateDiff	StdErrorDiff
age_sc	0.1013925	0.0574346	0.1001017	0.0570726	-0.0012908	-0.0003621
urbanY	0.6538503	0.1155389	0.6399185	0.1148419	-0.0139318	-0.0006970
age_sc:urbanY	-0.0711136	0.1078674	-0.0695667	0.1071936	0.0015469	-0.0006737

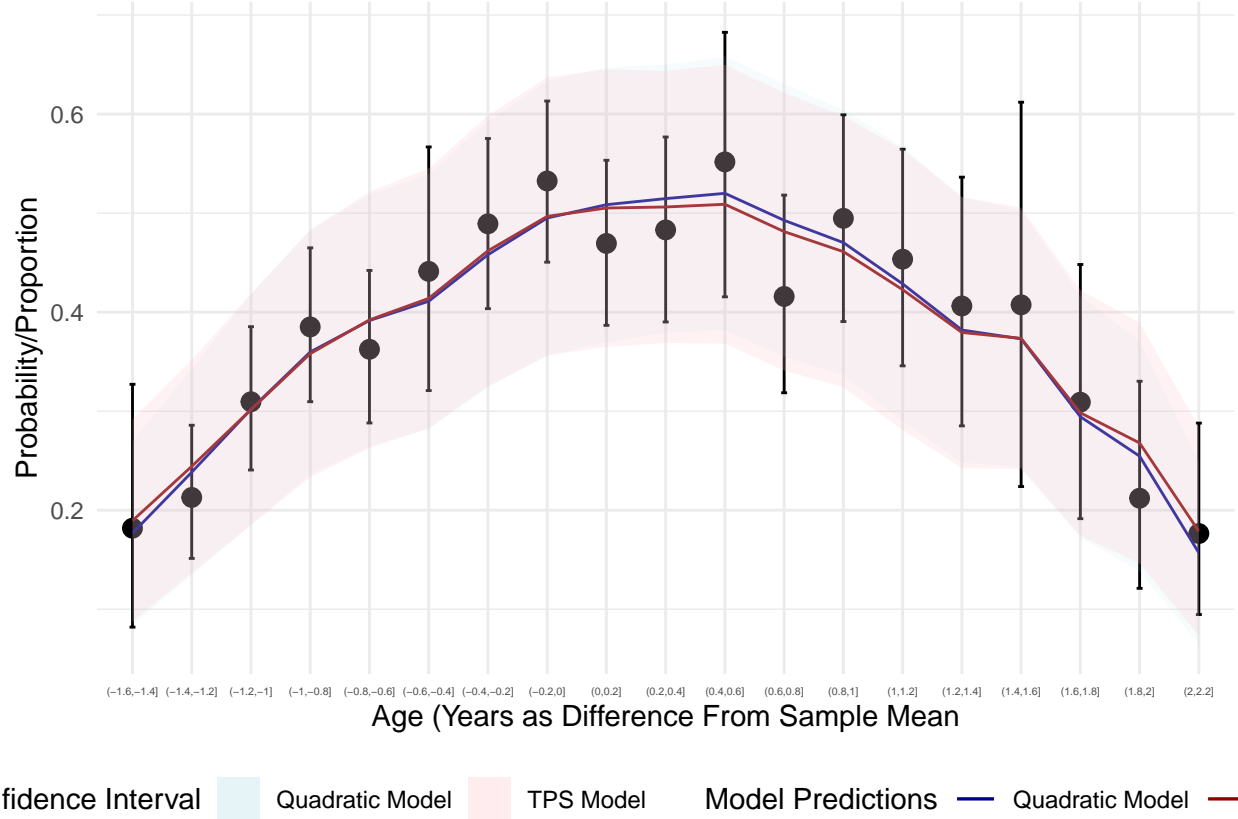


Figure 1: The figure displays observed and predicted probabilities of contraceptive use based on the Contraception dataset. The black bars represent the binned proportions of contraceptive use, with confidence intervals (CIs) calculated using beta quantiles. These observed proportions are juxtaposed against two predictive models. Both models include 'urban' as a fixed effect and 'district' as a random intercept, but they differ in how the age variable is treated. The 'Quad Model' employs a quadratic polynomial to model the fixed effect of scaled and centered age, while the 'TPS Model' uses a thin-plate spline for the same purpose. The curves on the plot indicate the mean predicted probabilities for each age bin, as estimated by `mgcv::gam`.

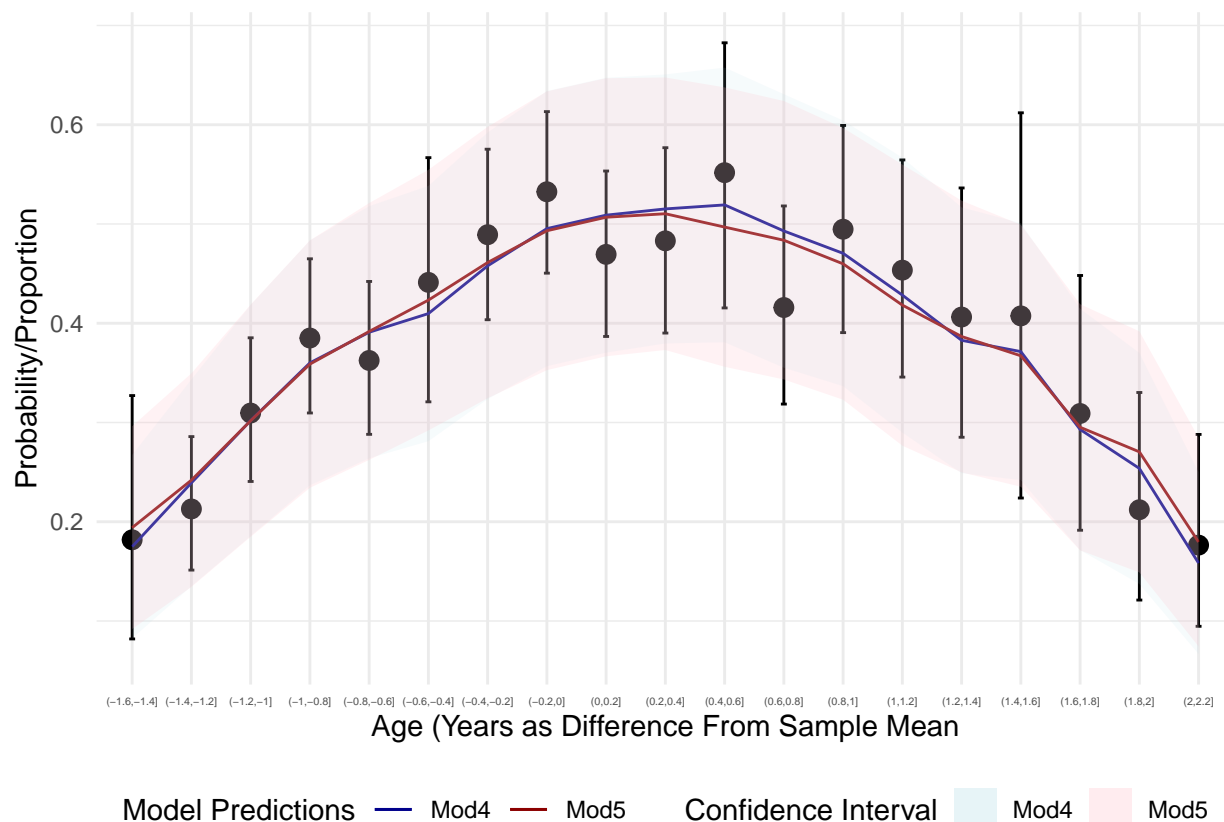


Figure 2: The figure displays observed and predicted probabilities of contraceptive use based on the Contraception dataset. The black bars represent the binned proportions of contraceptive use, with confidence intervals (CIs) calculated using beta quantiles. These observed proportions are juxtaposed against two predictive models. Both models include 'urban' as a fixed effect and 'district' as a random intercept. The 'Mod4' includes an additional interaction between a quadratic polynomial of age and urban status. 'Mod5' uses a split thin-plate spline term for urban status. The curves on the plot indicate the mean predicted probabilities for each age bin, as estimated by `mgcv::gam`.

The `gam()` model failed to converge with the default `optimizer = "bobyqa"` or `optimizer = "NelderMead"`. It converged for `optimizer = "bobyqa"` and when I increased the number of iterations to `1e5`.

I struggled with getting the standard errors for the predictions for glmer model using `predict`. I don't think it's possible to extract the standard error from using `predict` on a `glmer` model? I tried using `lme4::bootMer` to bootstrap some samples and compute confidence intervals but it took almost 20 minutes and then it was all NA. So I said no to trying to get that to work. So that's why the plot for this question does not have CI's

for the predictions.

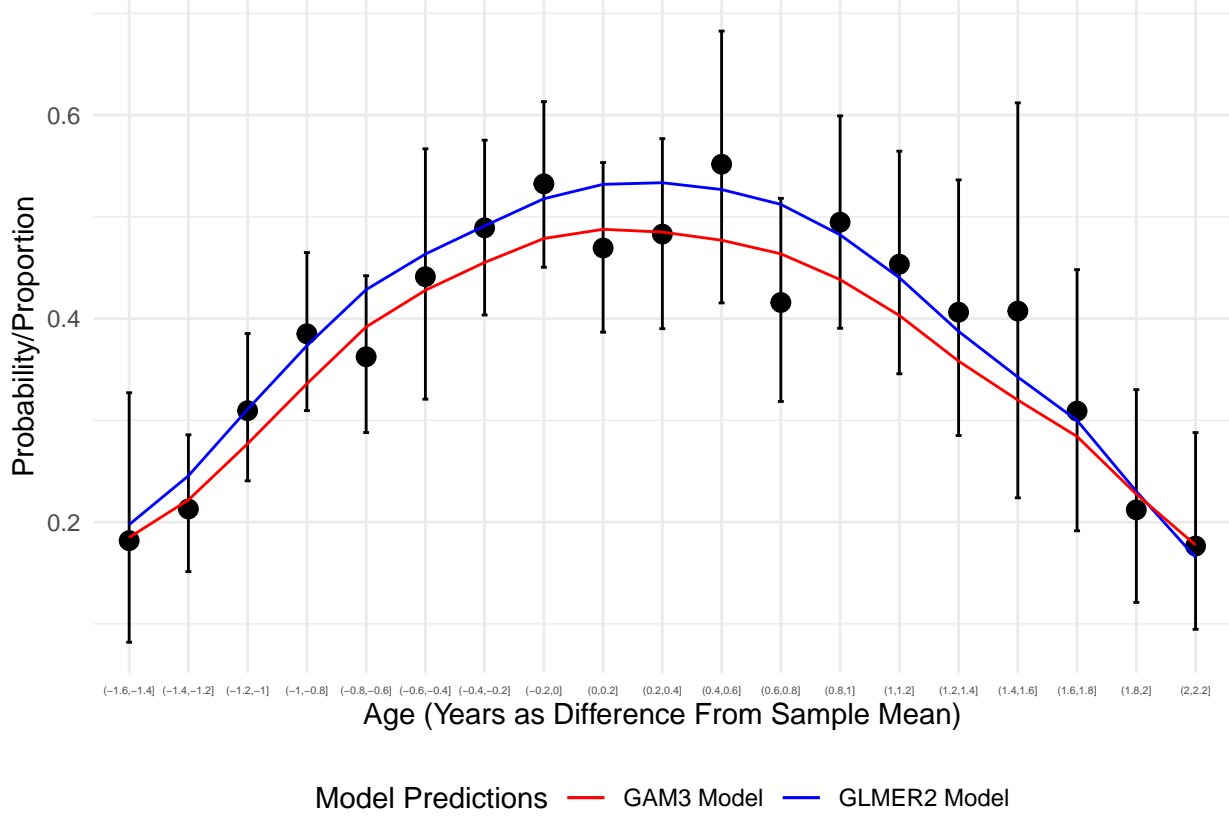


Figure 3: The figure displays observed and predicted probabilities of contraceptive use based on the Contraception dataset. The black bars represent the binned proportions of contraceptive use, with confidence intervals (CIs) calculated using beta quantiles. These observed proportions are juxtaposed against two predictive models. This figure is different than the previous figures in that it used the scaled and centered age predictor to plot the predictions. The 'GLMER2' model includes a fixed quadratic-age by urban-rural interaction and a random effect that allows the quadratic effect of age to vary across districts. The 'GAM3' model includes different population-level smooths for urban vs rural plus different age-smooths for each district. The curves on the plot indicate the mean predicted probabilities for each age bin, as estimated by `gam()` and `glm()`.