Residency Exams

```r
library(lme4)
```

```
Loading required package: Matrix
```

```
Warning in check_dep_version(): ABI version mismatch:
lme4 was built with Matrix ABI version 2
Current Matrix ABI version is 1
Please re-install lme4 from source or restore original 'Matrix' package
```

```r
library(knitr)
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
```

```
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x tidyr::expand() masks Matrix::expand()
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
x tidyr::pack()   masks Matrix::pack()
x tidyr::unpack() masks Matrix::unpack()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```r
library(glmmTMB)
```

```
Warning in checkMatrixPackageVersion(getOption("TMB.check.Matrix", TRUE)): Package version in
TMB was built with Matrix ABI version 2
Current Matrix ABI version is 1
Please re-install 'TMB' from source using install.packages('TMB', type = 'source') or ask CRA
```

```
library(broom)
library(emmeans)
```

```
Welcome to emmeans.
Caution: You lose important information if you filter this package's results.
See '? untidy'
```

```
data <- read.table("data.txt",header = TRUE, as.is = TRUE)
```

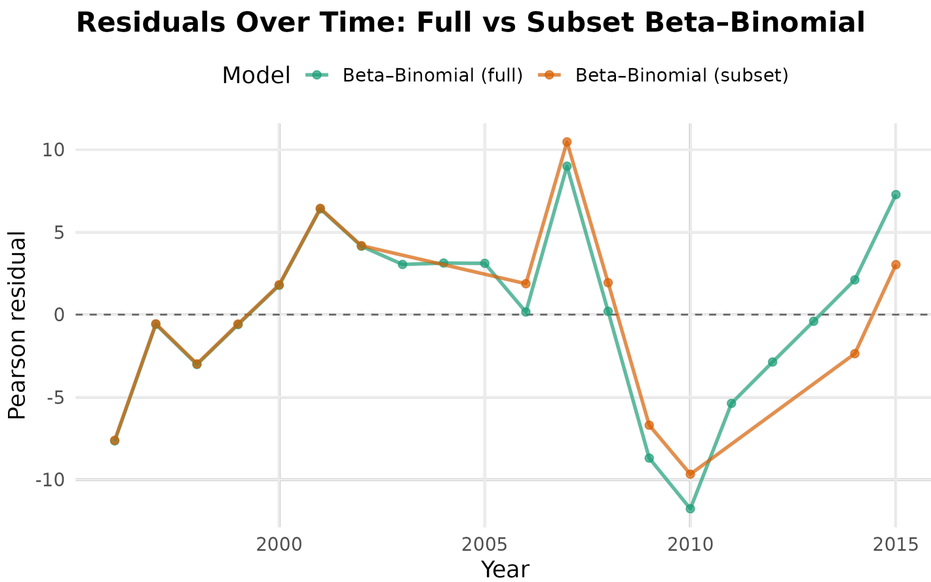
```
data$Pass <- round(data$N * data$Pct)
data$Fail <- (data$N - data$Pass)
data$timeperiod <- rep(1, nrow(data))
data$timeperiod[data$Year > 2002] <- 2
data$timeperiod[data$Year > 2010] <- 3
data$timeperiod <- factor(data$timeperiod, levels = c(1, 2, 3), labels = c("tp1", "tp2", "tp3
```

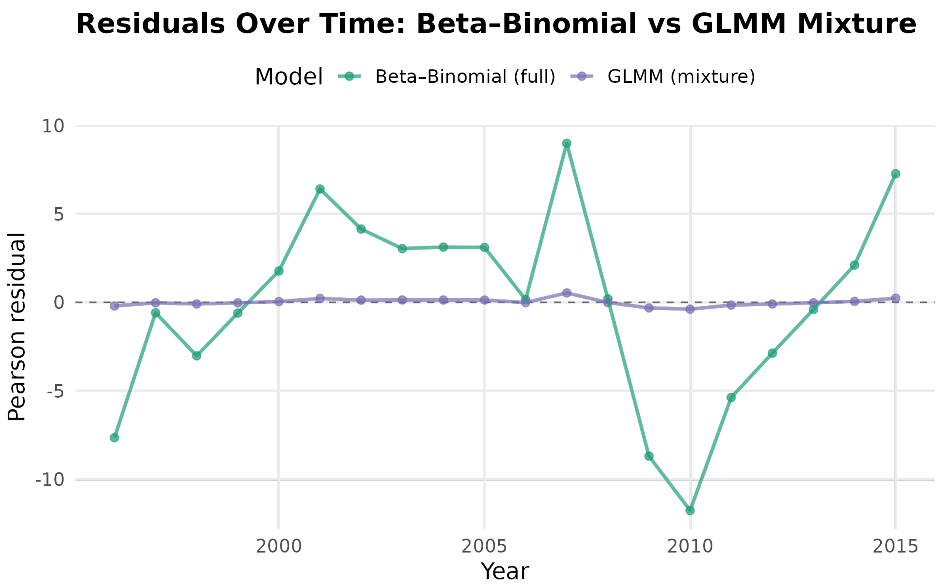
Background & Motivation

Model Implementation Details

Model Evaluation

To check goodness of fit, we plotted Pearson residuals over time. We appear to minimize residuals over time in the mixture model, as seen below. Putting this together, we concluded that the full beta-binomial mixture model is the best fit.





Shortcomings

An issue came with finding good evaluation metrics that compare across models. We initially attempted AIC across all models, but soon learned this was not comparable for any of the models because the mixture modeling package and beta binomial packages use different formulations of AIC. In addition, AIC is dependent on the likelihood and because of this comparing AIC across a full or subset model since they utilize variable number of observations.

Looking ahead, a next step would be to use Bayesian models, which would allow us to incorporate auxiliary information into the analysis.

Conclusion

We concluded that the binomial mixture model is the best fit. With that model, we found that the first reform increased the odds of passing, while the second reform was associated with a decline in performance. Both of these were confirmed by significant coefficients.