Longitudinal Data Analysis Exam

Joschka Hüllmann

2019-01-18

Introduction

Conventional wisdom suggests that the influence of the Supreme Court justices' ideology becomes more politically moderate over time. I test this hypothesis using a data set on the voting patterns of justices sitting on the Supreme Court during the Vinson, Warren, Burger, Rehnquist, and Roberts courts (1946-2017) with $N = 38, T = 72, \textit{unbalanced}$. In particular, I analyse the effect of the normed "Segal/Cover" (SC) score of the justice (time invariant) and the "Martin/Quinn" (MQ) score of the justice (time variant) on the percentage of liberal (left) votes casted by the justice. If the hypothesis holds, then SC should have a positive effect on the votes, while the interaction betweene SC and time should have a negative effect. Since MQ uses an inversed scale compared to SC, the results for MQ should be the opposite.

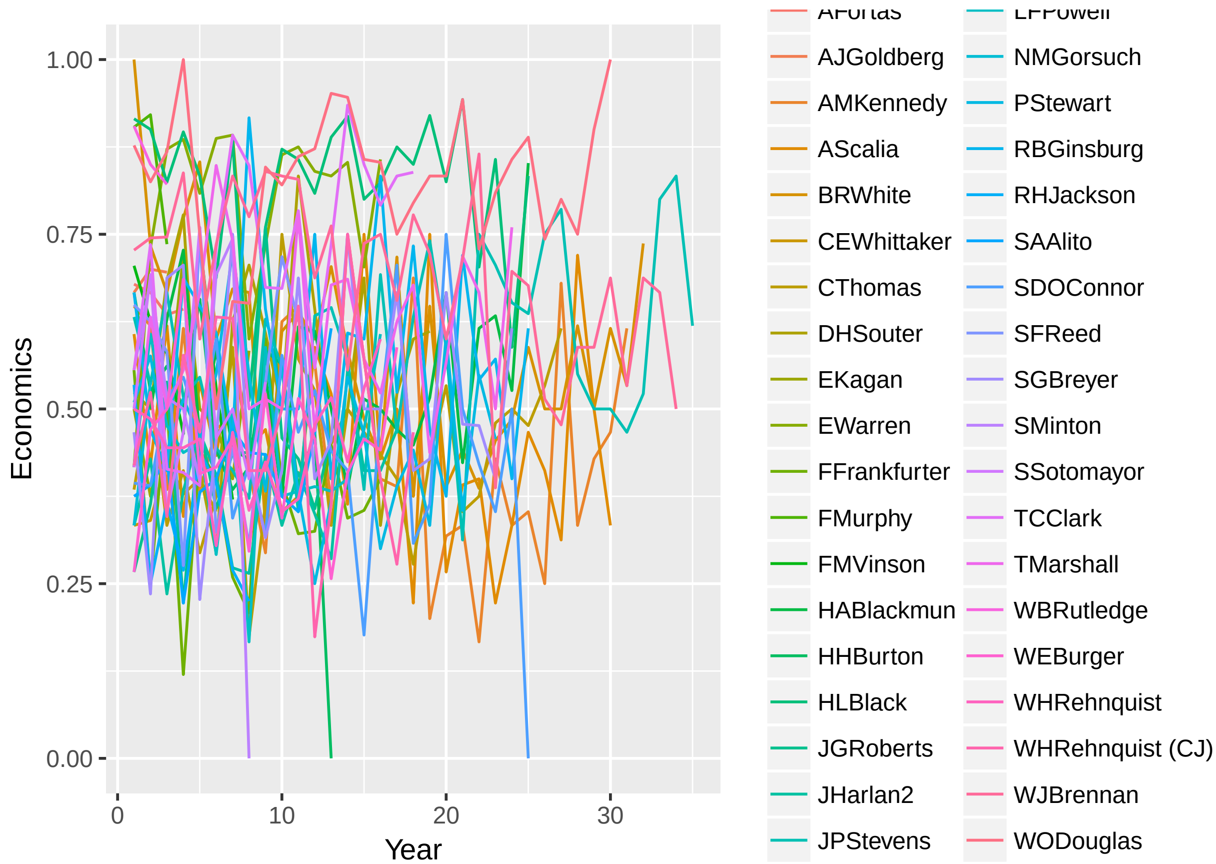
Analysis

Listing 1 shows the loading and pre-processing of the data.

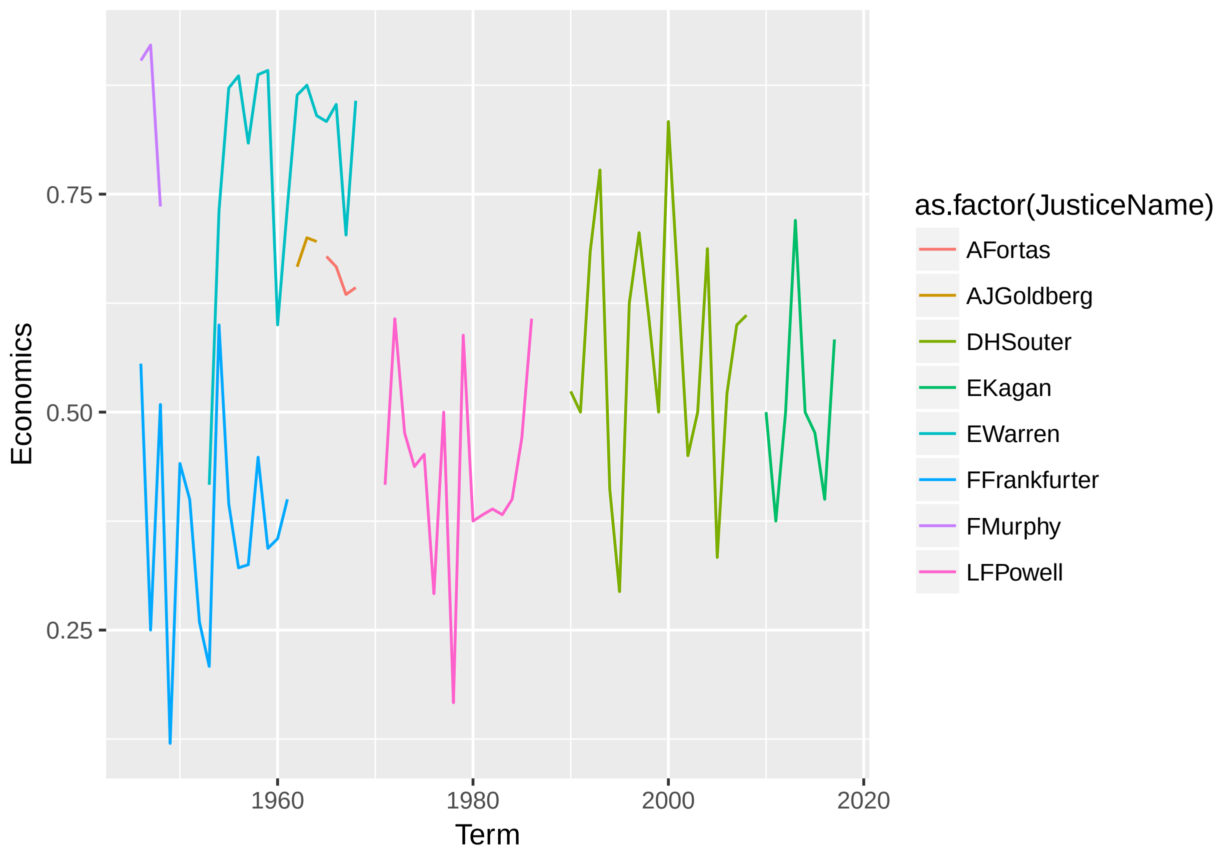
```
library(RCurl)
library(dplyr)
library(geepack)
library(glmmML)
library(ggplot2)
library(plm)
library(lme4)
library(lmtest)
library(zoo)
library(survival)
setwd("C:\\dev\\workspace\\GSERM-Oslo-2019-git\\Final Exam-Solution")
df<-read.csv("..\\Final Exam\\GSERM-Oslo-2019-Exam-Q1-Data.csv")
df<-df[,c(2,3,6,8,9)]
df<-df %>% arrange(JusticeName, Term) %>% group_by(JusticeName) %>% mutate(rank(Term, ties.method="firs
names(df)[6]<-"Year"
df$Economics<-df$Economics/100
df<-filter(df, is.na(Economics)==FALSE, is.na(SCIdeology)==FALSE,
      is.na(MQScore)==FALSE, is.na(Term)==FALSE,
      is.na(Year)==FALSE) %>% as.data.frame()
```

Listing 2 shows the descriptive statistics of the data. On a first glance, I can see no patterns that would support the hypothesis.

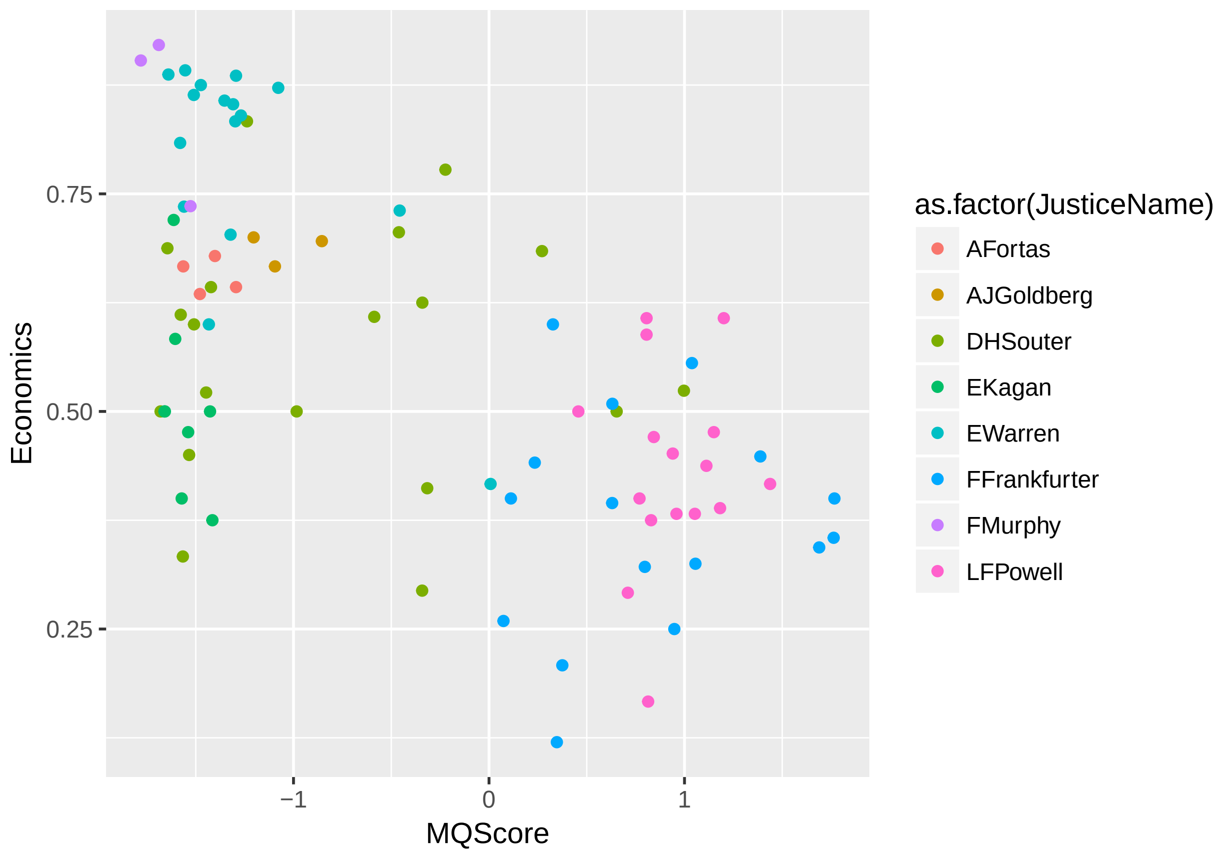
```
summary(df)
ggplot(df, aes(Year, Economics)) + geom_line(aes(colour=as.factor(JusticeName)))
```




```
ggplot(df[which(df$JusticeName %in% c("AFortas","AJGoldberg","LFPowell", "DHSouter", "EKagan", "FMurphy
```

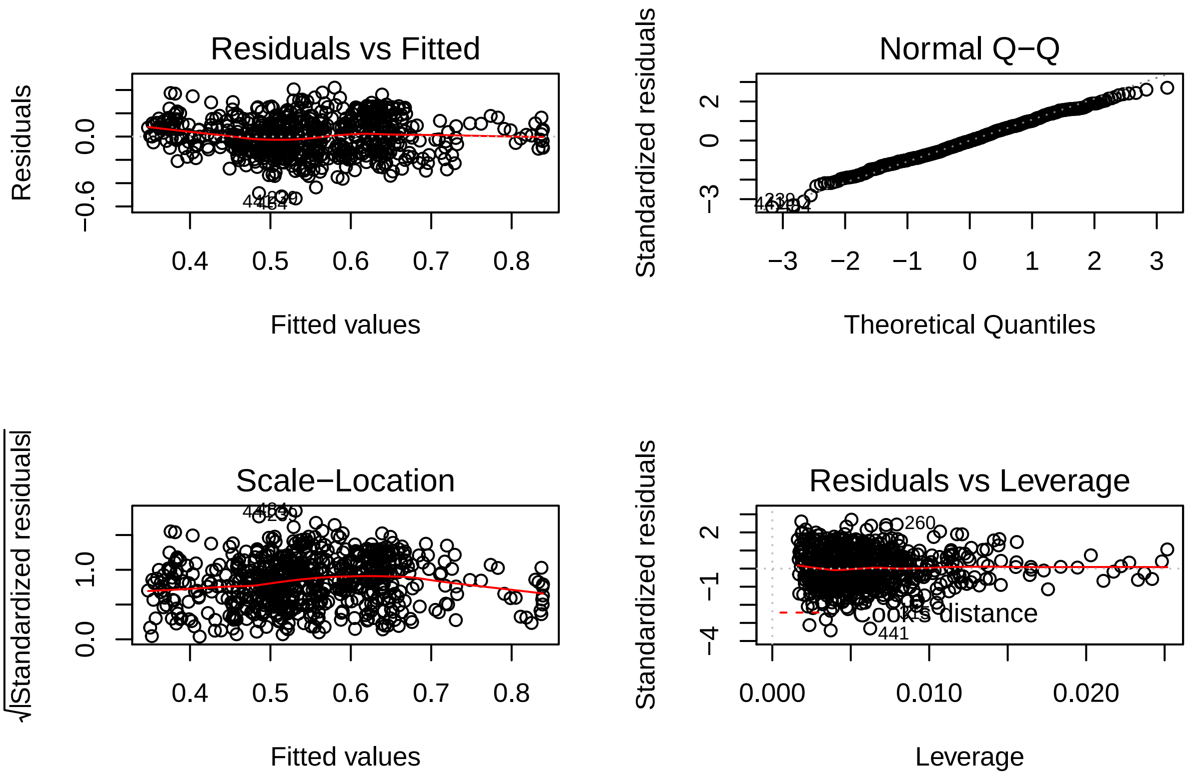



```
ggplot(df[which(df$JusticeName %in% c("AFortas","AJGoldberg","LFPowell", "DHSouter", "EKagan", "FMurphy
```

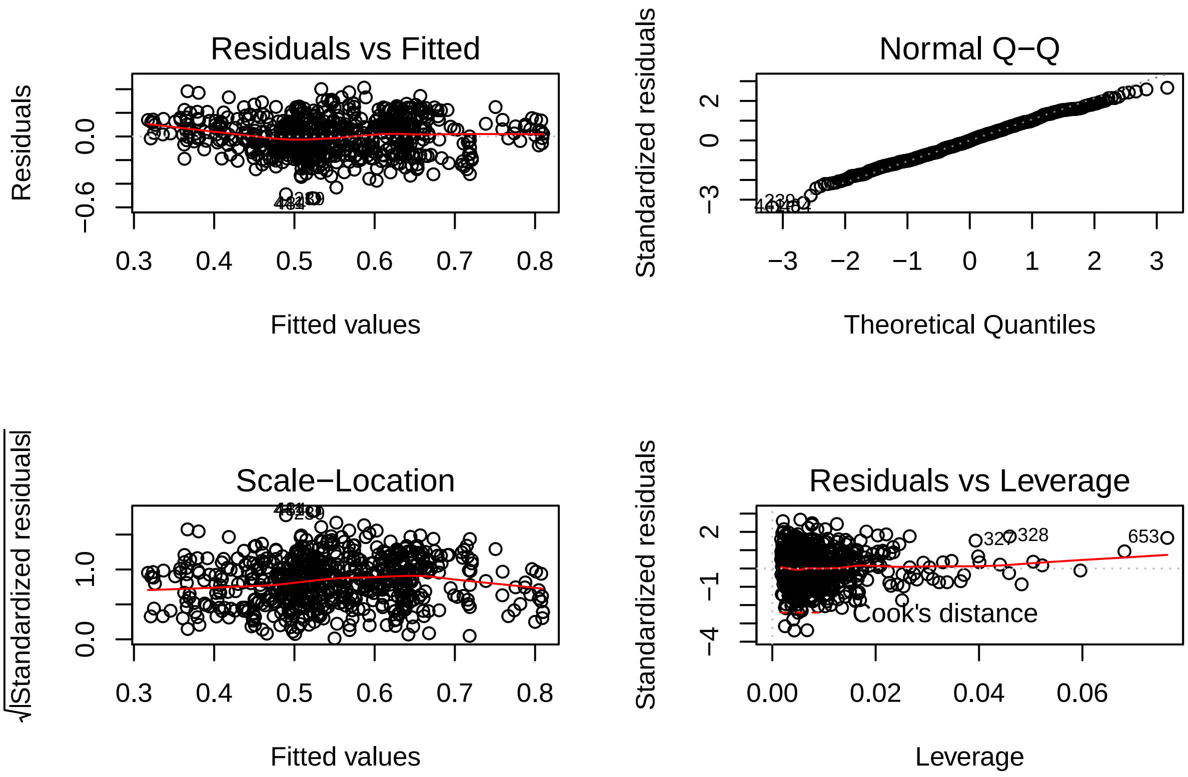


Listing 3 shows basic OLS models. Ideology is positively associated with the votes as expected, supporting the hypothesis.

```
fit.ols<-lm(Economics~SCIdeology+MQScore+Year, df)
fit.ols.ie<-lm(Economics~SCIdeology+MQScore+SCIdeology*Year, df)
fit.ols.ie2<-lm(Economics~SCIdeology+MQScore+SCIdeology*Year+MQScore*Year, df)
summary(fit.ols)
summary(fit.ols.ie)
summary(fit.ols.ie2)
par(mfrow=c(2,2))
plot(fit.ols)
```




```
plot(fit.ols.ie2)
```



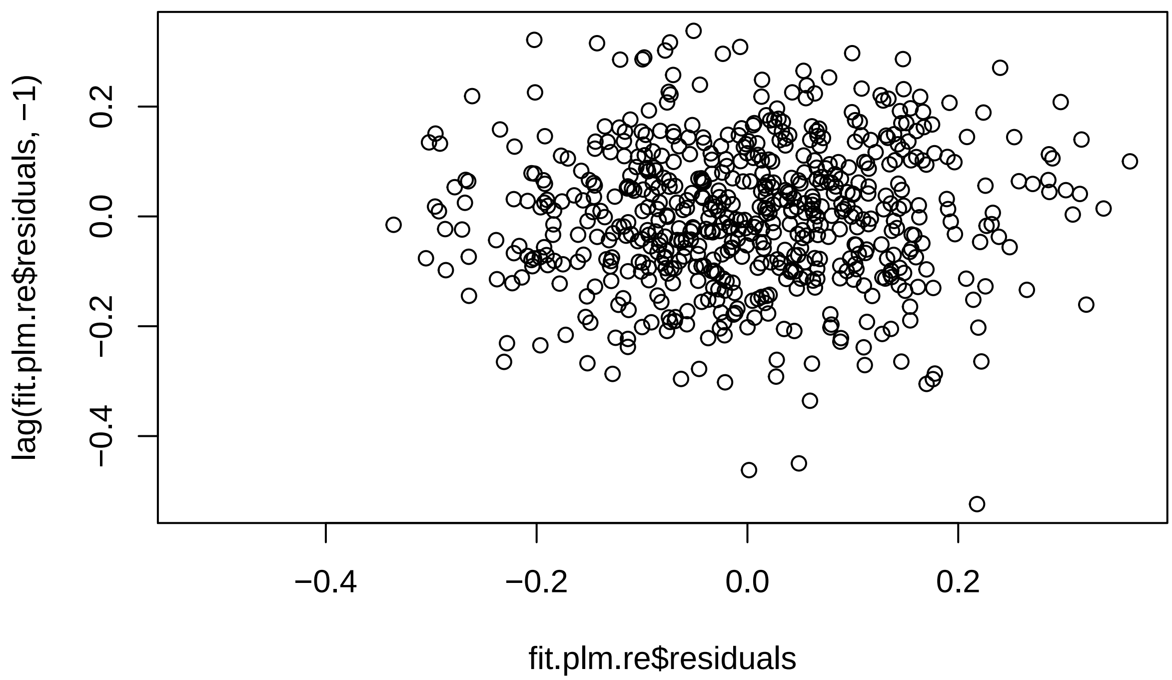
Listing 4 shows fixed and random effects models. Fixed effect does not make any sense, because our covariate of main interest, i.e. SCIdeology, is time invariant and therefore is excluded from model. The Hausman test suggests likewise. For the random effects model ($Y_{it} = f(X_{it}\beta + \alpha_i + u_{it})$, f is logit link for `glmm` and `glmer` and identify link for `plm` package) the random effect must be independent of the other independent variables ($cov(x_{it}, \alpha_i) = 0$, with x_{it} being the covariates and α_i being the random effects). Substantively, it is a little bit difficult to justify this, because the political ideology of a justice is influenced by many individual-specific effects, e.g. where they grew up, the ideology of their parents and social circle, past experience, professional history, etc. Another assumption to discuss is within-unit correlation. So far we have assumed no correlation within units. The plots show that for the linear random effects panel model, the residuals are uncorrelated and look fine. However, for the generalised models, there seems to be a systematic correlation. Substantively, this is related to the conventional wisdom that justices become more moderate over time. Analytically, the durbin watson test can be used to check this. In our case it confirms our assumptions about the models. To deal with this dynamic, we already included the time parameter in the model in a linear, non-unit specific way. For auto correlation, we can also fit the models with a correlation structure. Unfortunately, the packages for generalised linear models that we use here do not support this (though, it seems that the function glmmPQL from the MASS package can do this). Since we are not explicitly interested in unit level effect, but rather acknowledge their existence, we can choose a straight-forward way out and use a generalised estimating equation (GEE) model, i.e. a population average model.

Note: I am not sure why glmmML and glmer show such different results. They should both be fitted with the Gauss-Hermite Quadrature with equivalent model specifications. No hausman test is available for the packages. Some guy on StackOverflow ported it apparently (https://stackoverflow.com/a/23635004).

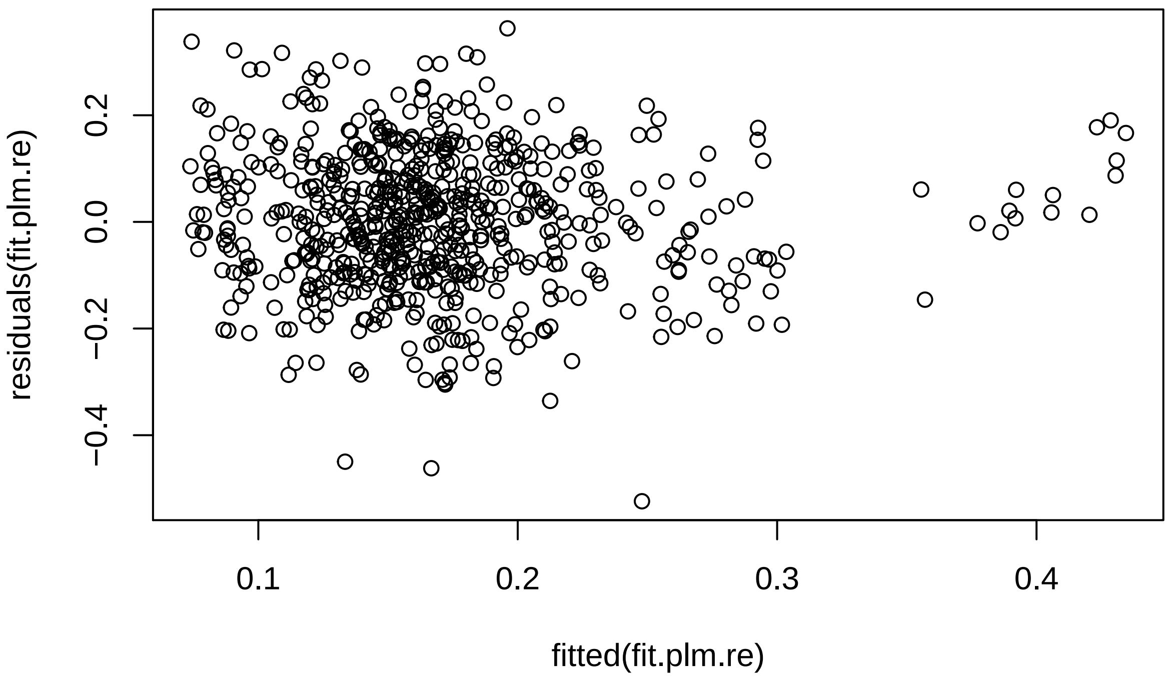
```
# fixed effects
fit.glmm.fe <- glmmboot(Economics~SCIdeology+MQScore+Year, data=df, family="binomial", cluster=JusticeNa
summary(fit.glmm.fe)
fit.glmm.fe.i <- glmmboot(Economics~SCIdeology+MQScore+Year+SCIdeology*Year+MQScore*Year, data=df, famil
```

```r
summary(fit.glmm.fe.i)


# random effects
fit.glmm.re <- glmmML(Economics~SCIdeology+MQScore+Year, data=df, family="binomial", cluster=JusticeName
summary(fit.glmm.re)
fit.glmm.re.i <- glmmML(Economics~SCIdeology+MQScore+Year+SCIdeology*Year+MQScore*Year, data=df, family=
summary(fit.glmm.re.i)


# random effects with glmer
fit.glmer.re2 <- glmer(Economics~SCIdeology+MQScore+Year + (1|JusticeName), data=df, family="binomial")
summary(fit.glmer.re2)
fit.glmer.re2.i <- glmer(Economics~SCIdeology+MQScore+Year+SCIdeology*Year+MQScore*Year + (1|JusticeName
summary(fit.glmer.re2.i)

# conditional fixed effects with coxph
fit.clogit <- clogit(Economics~SCIdeology+MQScore+Year+strata(JusticeName), data=df)
summary(fit.clogit)
fit.clogit.i <- clogit(Economics~SCIdeology+MQScore+Year+SCIdeology*Year+MQScore*Year+strata(JusticeName
summary(fit.clogit.i)

# fixed effect lsdv with glm
fit.lsdv <- glm(Economics~SCIdeology+MQScore+Year+as.factor(JusticeName), data=df,family=binomial)
fit.lsdv.i <- glm(Economics~SCIdeology+MQScore+SCIdeology*Year+MQScore*Year+as.factor(JusticeName), data

# fixed and random effect with plm (without logit as pglm + binomial family does not allow fractions)
fit.plm.fe <- plm(Economics~SCIdeology+MQScore+Year, data=df, index=c("JusticeName"), model="within")
summary(fit.plm.fe)
fit.plm.fe2.i <- plm(Economics~SCIdeology+MQScore+Year+SCIdeology*Year+MQScore*Year, data=df, index=c("
summary(fit.plm.fe2.i)
fit.plm.re <- plm(Economics~SCIdeology+MQScore+Year, data=df, index=c("JusticeName"), model="random")
summary(fit.plm.re)
fit.plm.re2.i <- plm(Economics~SCIdeology+MQScore+Year+SCIdeology*Year+MQScore*Year, data=df, index=c("
summary(fit.plm.re2.i)

# overview
overview <- cbind(
    c(0,coef(fit.glmm.fe)),
    coef(fit.lsdv)[1:4],
    c(0,coef(fit.plm.fe),0),
    coef(fit.glmm.re),
    coef(summary(fit.glmer.re2))[,1],
    coef(fit.plm.re)
  )
colnames(overview) <- c("glmmFE", "lsdv", "plmFE", "glmmRE","glmerRE","plmRE")
overview

overview.i <- cbind(
    coef(fit.glmm.re.i),
    coef(summary(fit.glmer.re2.i))[,1],
    coef(fit.plm.re2.i)
  )
```

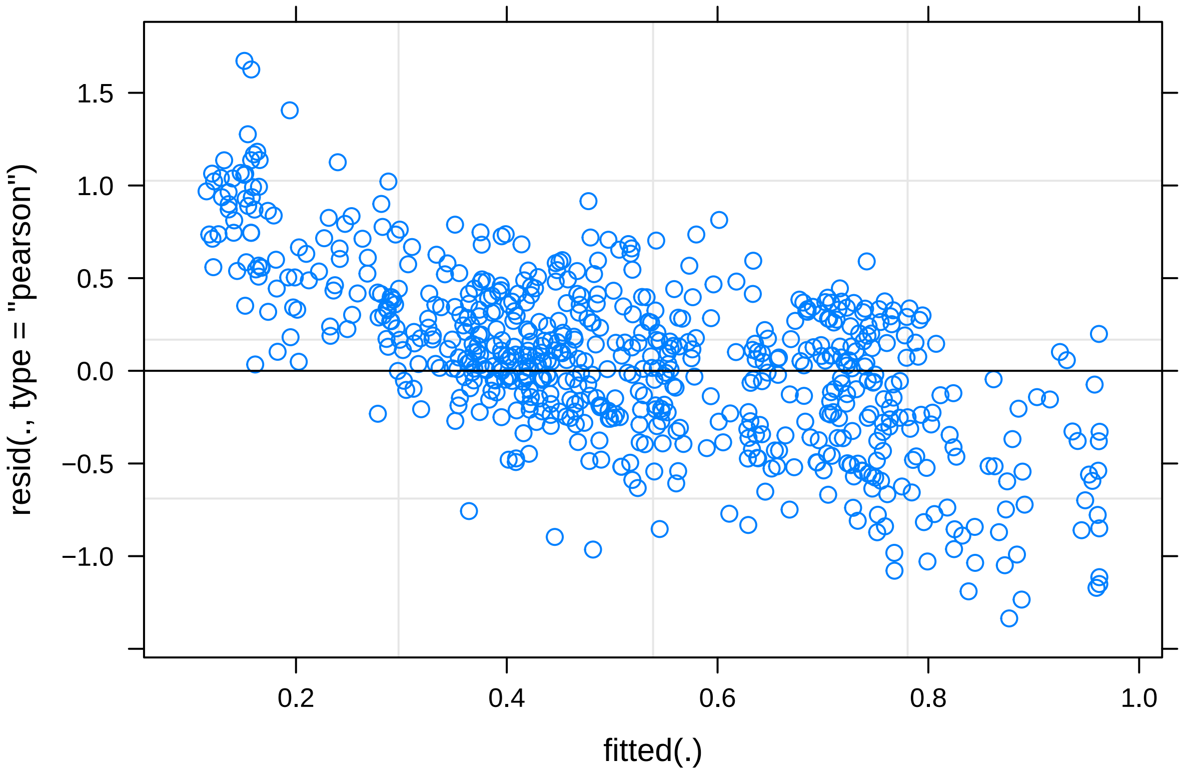
```r
colnames(overview.i) <- c("glmmRE","glmerRE","plmRE")
overview.i

# hausman test and breusch-godfrey test
phtest(fit.plm.fe,fit.plm.re)
phtest(fit.plm.fe2.i,fit.plm.re2.i)
pbgtest(fit.plm.re)
pbgtest(fit.plm.re2.i)

# no pbgtest for glmer and glmm...

# durbin watson
pdwtest(fit.plm.re)
pdwtest(fit.plm.re)
pdwtest(fit.plm.re2.i)
dwtest(fit.glmm.re)

plot(fit.plm.re$residuals, lag(fit.plm.re$residuals, -1))
```



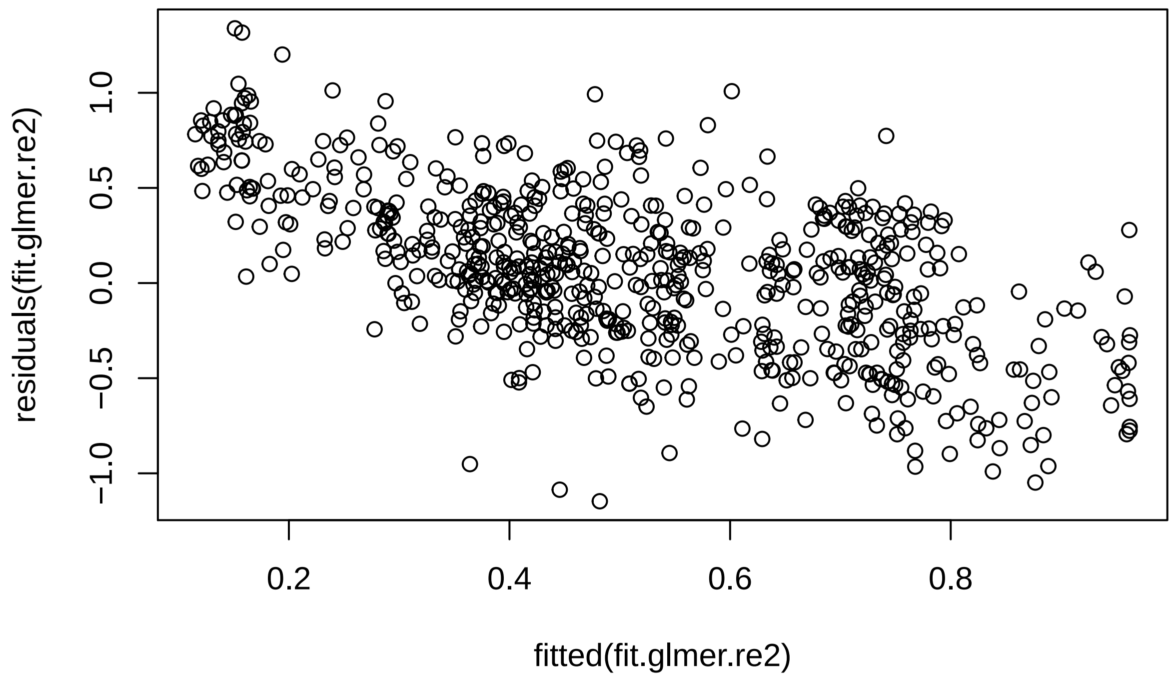

```r
plot(fitted(fit.plm.re), residuals(fit.plm.re))
```




```
plot(fit.glmer.re2)
```




```r
plot(fitted(fit.glmer.re2), residuals(fit.glmer.re2))
```

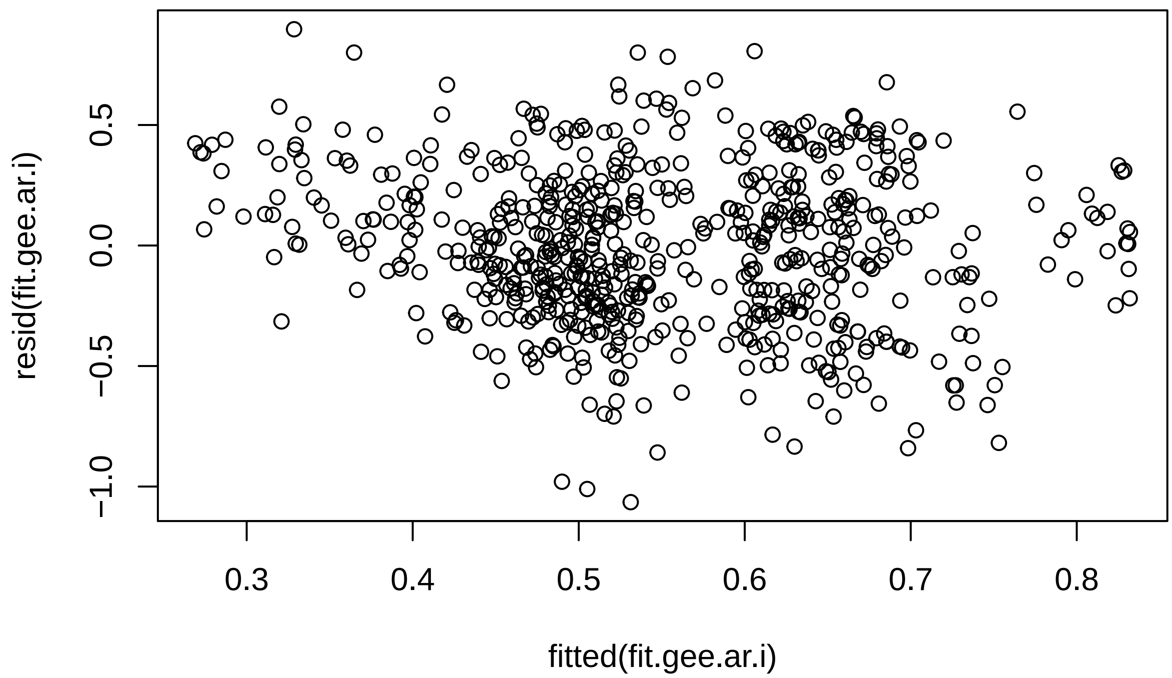



```
#plot(fitted(fit.glmm.re), residuals(fit.glmer.re)) # cannot extract residuals this way...
par(mfrow=c(1,1))
```

Listing 5 shows the GEE model. The results are similar to the previous model and also comparable across the three GEE models. Notably, the coefficient for SCIdeology changed the sign in the AR(1) model. However, the standard error is so high compared to the estimated coefficient that there probably is no underlying systematic association between SC and how the justices vote. Conversly, the MQ metric is significant and shows a high value for that Wald test $10 \leq T_W \leq 21$. On population average, a higher MQ value is negatively associated in margin to the percentage of left votes of justices.

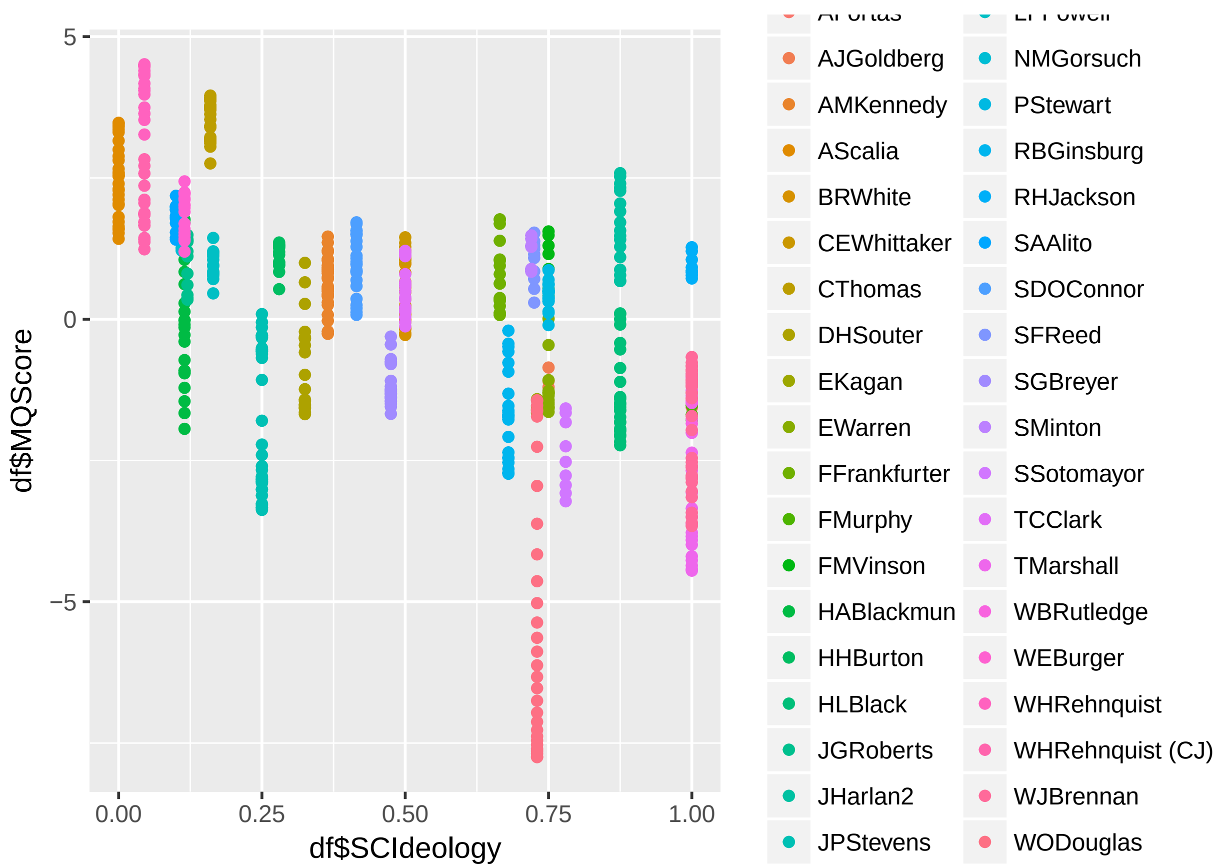
```
fit.gee.in <- geeglm(Economics~SCIdeology+MQScore+Year,
                     data=df,id=JusticeName,family=binomial, corstr="independence")
fit.gee.ex <- geeglm(Economics~SCIdeology+MQScore+Year,
                     data=df,id=JusticeName,family=binomial, corstr="exchangeable")
fit.gee.ar <- geeglm(Economics~SCIdeology+MQScore+Year,
                     data=df,id=JusticeName,family=binomial, corstr="ar1")
summary(fit.gee.in)
summary(fit.gee.ex)
summary(fit.gee.ar)
fit.gee.in.i <- geeglm(Economics~SCIdeology+MQScore+Year+SCIdeology*Year+MQScore*Year,
                     data=df,id=JusticeName,family=binomial, corstr="independence")
fit.gee.ex.i <- geeglm(Economics~SCIdeology+MQScore+Year+SCIdeology*Year+MQScore*Year,
                     data=df,id=JusticeName,family=binomial, corstr="exchangeable")
fit.gee.ar.i <- geeglm(Economics~SCIdeology+MQScore+Year+SCIdeology*Year+MQScore*Year,
                     data=df,id=JusticeName,family=binomial, corstr="ar1")
summary(fit.gee.in.i)
summary(fit.gee.ex.i)
```

```
summary(fit.gee.ar.i)
plot(fitted(fit.gee.ar.i), resid(fit.gee.ar.i))
```



The GEE model with AR correlation structure is a good fit (and for a paper, you would only report this one in particular), because it deals with the assumed autocorrelation and we because are not interested in the unit-level effects. Rather, we are interested in population average effect across all justices, whether they are getting more mild with time. Our covariate of main interest SC does not show such an effect, but the MQ covariate does, including the interaction effect at $P(>|W|) = 0.074$). Since both are supposed to capture the political ideology, a further inquiry should look at how both metrics are derived. Listing 6 shows that there is a mild correlation between the two.

```
cor(df$SCIdeology,df$MQScore)
ggplot(df, aes(df$SCIdeology, df$MQScore)) + geom_point(aes(colour=as.factor(JusticeName)))
```



Ultimately, I am not convinced that the hypothesis holds. For the model specification, there are probably other variables that are omitted skewing the results. Also, the basic OLS model is misleading in this case. It would be interested to hear to what conclusions the other participants came. Cheers.