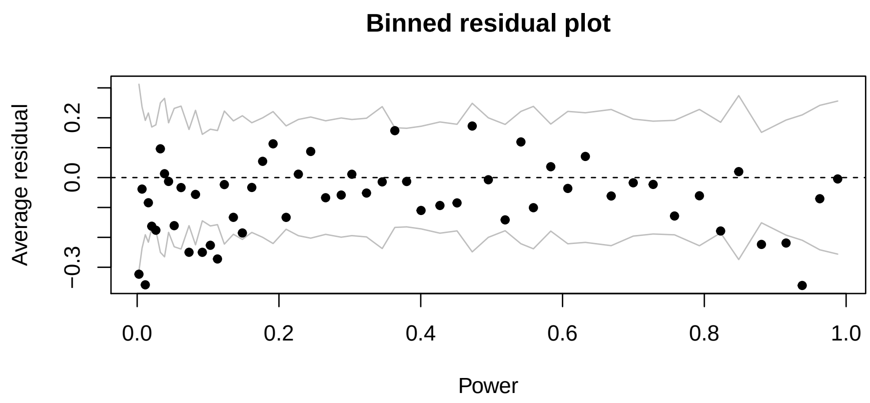
```r
rm(list = ls())
library(Zelig)
library(arm)
data(mid)
```

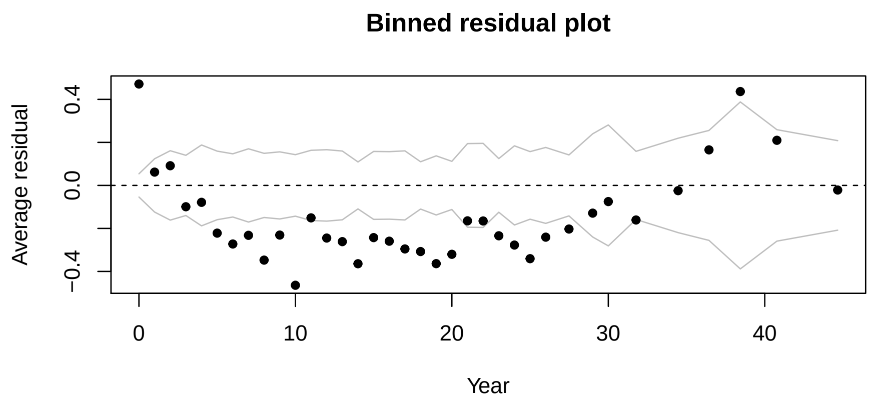
1 Probit model of conflict

```r
m_1 <- glm(conflict ~ major + contig + power + years, data = mid,
           family = binomial(link = "probit"))
```

1.1 binnedplot of residuals against powers and years

```r
par(mfrow = c(2, 1))
binnedplot(m_1$data$power, residuals(m_1), xlab = "Power")
binnedplot(m_1$data$year, residuals(m_1), xlab = "Year")
```





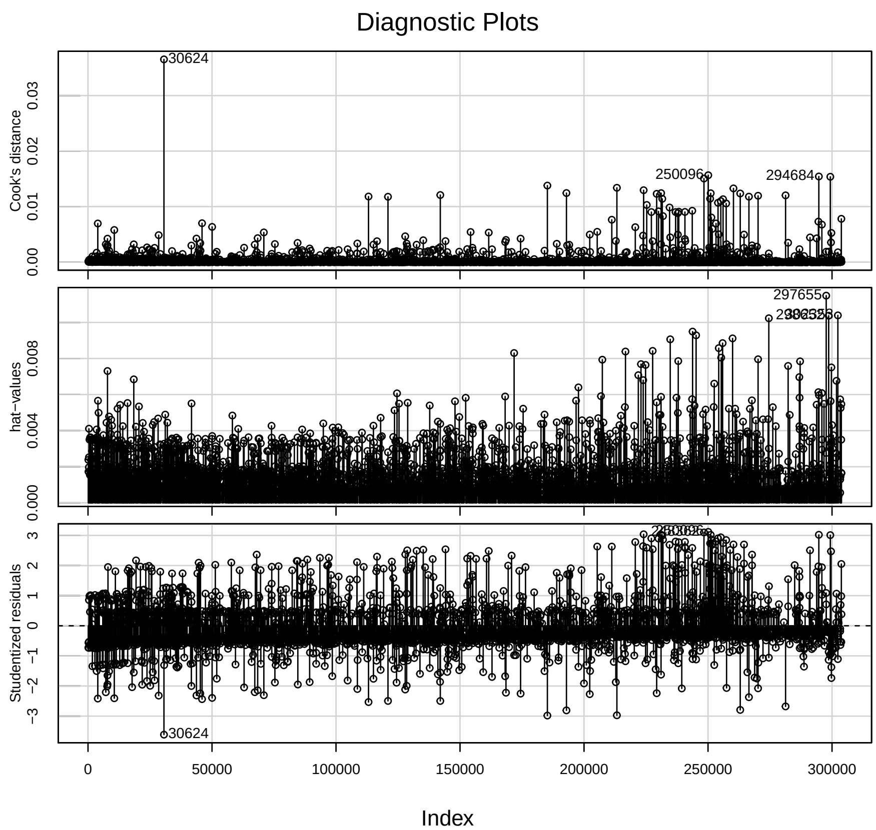

```
par(mfrow = c(1, 1))
```

The fit looks pretty good especially for power (all residuals hover around 0). For year, the residuals seem to have a quadratic relationship and also big residuals near the two extremes of the distribution.

1.2 Influence statistics to find problematic data point

```
library(car)

##
## Attaching package:  'car'
## The following object is masked from 'package:arm':
##
##     logit
```

```r
influenceIndexPlot(m_1, vars=c("Cook","hat","Studentized"), id.n=3)
```



Point 30624 seems problematic according to Cook's D and Studentized residuals. (hat-values plot indicate other problematic points, may be worth investigating).

```r
compareCoefs(m_1, update(m_1, subset=-c(which(row.names(mid) == 30624))))

##
## Call:
## 1: glm(formula = conflict ~ major + contig + power + years, family =
##   binomial(link = "probit"), data = mid)
## 2: glm(formula = conflict ~ major + contig + power + years, family =
##   binomial(link = "probit"), data = mid, subset =
##   -c(which(row.names(mid) == 30624)))
##                Est. 1     SE 1   Est. 2     SE 2
## (Intercept) -1.12893  0.07229 -1.14451  0.07268
```

```
## major        1.37940  0.08361  1.40589  0.08402
## contig       2.25593  0.07804  2.27203  0.07851
## power        0.53912  0.11060  0.57045  0.11106
## years       -0.03184  0.00293 -0.03204  0.00294
```

Deleting this point doesn't change the estimate too much

2 Robit model, same variables, using t-distribution with 3 df

```
library(bbmle)

## Loading required package:  stats4

LL_robit_3 <- function(params,y,X){
  B <- params
  p <- pt(X %*% B, 3) #t link w/ 3 df
  minusll  = -sum(y*log(p) + (1-y)*log(1-p))
  return(minusll)
}

parnames(LL_robit_3) <- c("Intercept", "Major", "Contig", "Power", "Years")

m_2 <- mle2(LL_robit_3, start = c(Intercept=0, Major=0, Contig=0, Power=0, Years=0),
             data=list(y=mid$conflict,
                       X=cbind(1, as.matrix(mid[ , c("major", "contig", "power", "years")]))),
            vecpar = TRUE)
summary(m_2)

## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = LL_robit_3, start = c(Intercept = 0, Major = 0,
##     Contig = 0, Power = 0, Years = 0), data = list(y = mid$conflict,
##     X = cbind(1, as.matrix(mid[, c("major", "contig", "power",
##         "years")]))), vecpar = TRUE)
##
## Coefficients:
##             Estimate Std. Error  z value     Pr(z)
## Intercept -1.4402460  0.1258706 -11.4423 < 2.2e-16 ***
## Major      2.0055530  0.1309166  15.3193 < 2.2e-16 ***
## Contig     3.0663944  0.1342645  22.8385 < 2.2e-16 ***
## Power      0.8610372  0.1862135   4.6239 3.765e-06 ***
## Years     -0.0604773  0.0053128 -11.3834 < 2.2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 1879.132
```

3 Model with same variables, using complementary log log link

```r
m_3 <- glm(conflict ~ major + contig + power + years, data = mid,
           family = binomial(link = "cloglog"))
summary(m_3)

##
## Call:
## glm(formula = conflict ~ major + contig + power + years, family = binomial(link = "clogl
##     data = mid)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.3435  -0.5130  -0.3396   0.3162   2.8420
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.588622   0.100125 -15.866  < 2e-16 ***
## major        1.521090   0.109602  13.878  < 2e-16 ***
## contig       2.464128   0.089778  27.447  < 2e-16 ***
## power        0.401678   0.139130   2.887  0.00389 **
## years       -0.057889   0.004195 -13.799  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3979.5  on 3125  degrees of freedom
## Residual deviance: 1951.5  on 3121  degrees of freedom
## AIC: 1961.5
##
## Number of Fisher Scoring iterations: 7
```

4 Rare event logit, same variables

```r
m_4 <- zelig(conflict ~ major + contig + power + years, data = mid,
             model = "relogit", cite = FALSE)
summary(m_4)
```

```
## Model:
##
## Call:
## z5$zelig(formula = conflict ~ major + contig + power + years,
##     data = mid)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2808  -0.4563  -0.2873   0.3630   3.0275
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.925968   0.140435 -13.714  < 2e-16
## major        2.527912   0.153441  16.475  < 2e-16
## contig       3.943343   0.150047  26.281  < 2e-16
## power        1.024513   0.214138   4.784 1.72e-06
## years       -0.065851   0.005641 -11.674  < 2e-16
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3979.5  on 3125  degrees of freedom
## Residual deviance: 1905.8  on 3121  degrees of freedom
## AIC: 1915.8
##
## Number of Fisher Scoring iterations: 5
##
## Next step: Use 'setx' method
```

5 Logit model, same variables, weakly informative priors on all coefs

```r
m_5 <- bayesglm(conflict ~ major + contig + power + years, data = mid,
         family = binomial(link = "logit"))
summary(m_5)
```

```
##
## Call:
## bayesglm(formula = conflict ~ major + contig + power + years,
##     family = binomial(link = "logit"), data = mid)
```

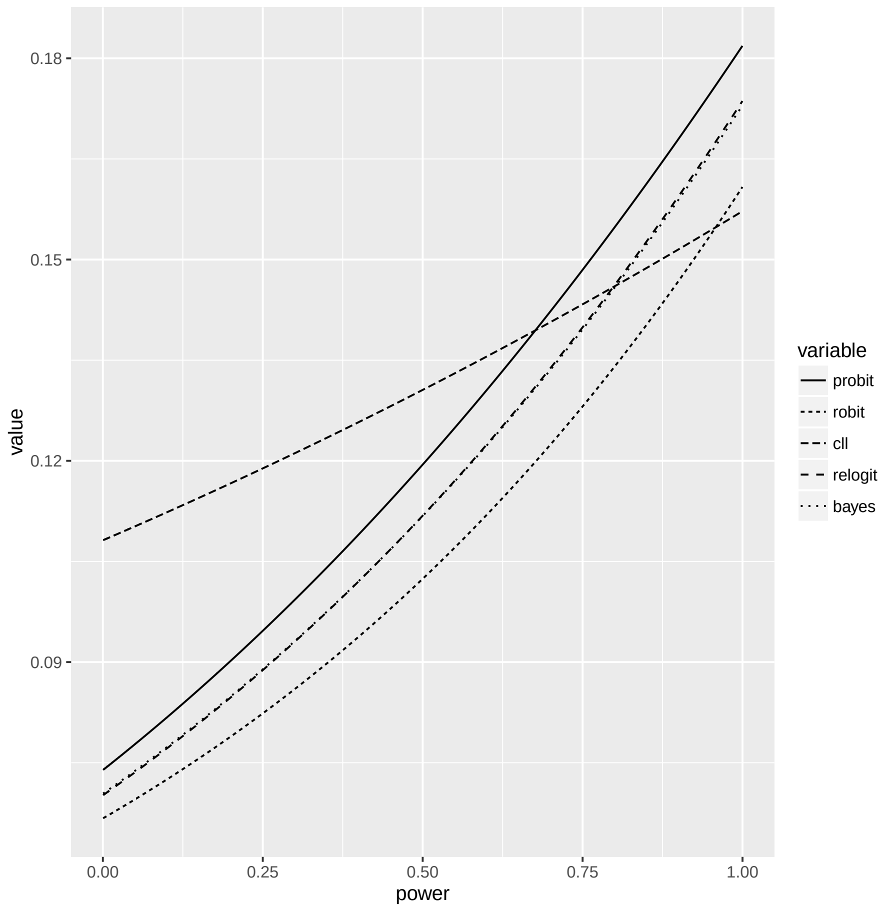
```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2788  -0.4563  -0.2870   0.3626   3.0292
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.919626   0.139821 -13.729  < 2e-16 ***
## major        2.521050   0.152794  16.500  < 2e-16 ***
## contig       3.945553   0.149553  26.382  < 2e-16 ***
## power        1.016201   0.213135   4.768 1.86e-06 ***
## years       -0.066121   0.005635 -11.734  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3979.5  on 3125  degrees of freedom
## Residual deviance: 1905.8  on 3121  degrees of freedom
## AIC: 1915.8
##
## Number of Fisher Scoring iterations: 6
```

6

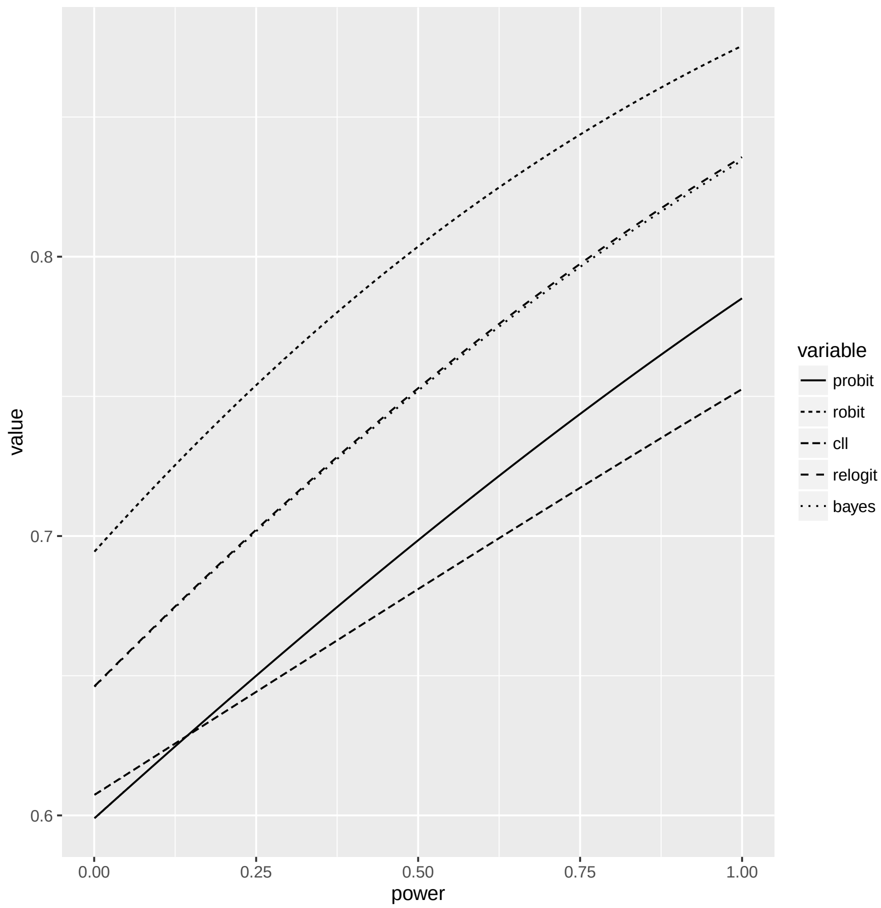
Plot the relationship between the predicted probability that two states are in conflict in a given year and the balance of power between the states for all 5 models on the same plot. Hold contig and major at 0 and years at 10. Use a different line type for each model. Describe what you find. Then create a second plot where you change major to 1 and years to 0. Again, describe what you find. You do NOT need to plot confidence intervals for any of the estimates.

```
library(ggplot2)
library(reshape2)

f_predict <- function(model, contig, major, years) {
  newdata <- model$data
  newdata$contig <- contig
  newdata$major <- major
  newdata$years <- years
  predict(model, newdata = newdata, type = "response")
}

f_predict_robit <- function(model, contig, major, years) {
```

```
  X <- model@data$X
  X[ , "contig"] <- contig
  X[ , "major"] <- major
  X[ , "years"] <- years

  B <- coef(model)
  p <- pt(X %*% B, 3) #t link w/ 3 df
  return(p)
}

f_predict_zelig <- function(model, contig, major, years) {
  X <- cbind(1, major = major, contig = contig, power = mid$power, years = years)
  B <- coef(m_4)[[1]]
  return(plogis(X %*% B))
}

pdata <- data.frame(
  power = mid$power,
  probit = f_predict(m_1, contig = 0, major = 0, years = 10),
  robit = f_predict_robit(m_2, contig = 0, major = 0, years = 10),
  cll = f_predict(m_3, contig = 0, major = 0, years = 10),
  relogit = f_predict_zelig(m_4, contig = 0, major = 0, years = 10),
  bayes = f_predict(m_5, contig = 0, major = 0, years = 10)
)

ggplot(data = melt(pdata, id.vars = "power"), aes(power, value)) +
  geom_line(aes(linetype = variable))
```




```r
pdata <- data.frame(
  power = mid$power,
  probit = f_predict(m_1, contig = 0, major = 1, years = 0),
  robit = f_predict_robit(m_2, contig = 0, major = 1, years = 0),
  cll = f_predict(m_3, contig = 0, major = 1, years = 0),
  relogit = f_predict_zelig(m_4, contig = 0, major = 1, years = 0),
  bayes = f_predict(m_5, contig = 0, major = 1, years = 0)
)

ggplot(data = melt(pdata, id.vars = "power"), aes(power, value)) +
  geom_line(aes(linetype = variable))
```


0.8
0.7
0.6
value
0.00
0.25
0.50
0.75
1.00
power
variable
probit
robit
cll
relogit
bayes