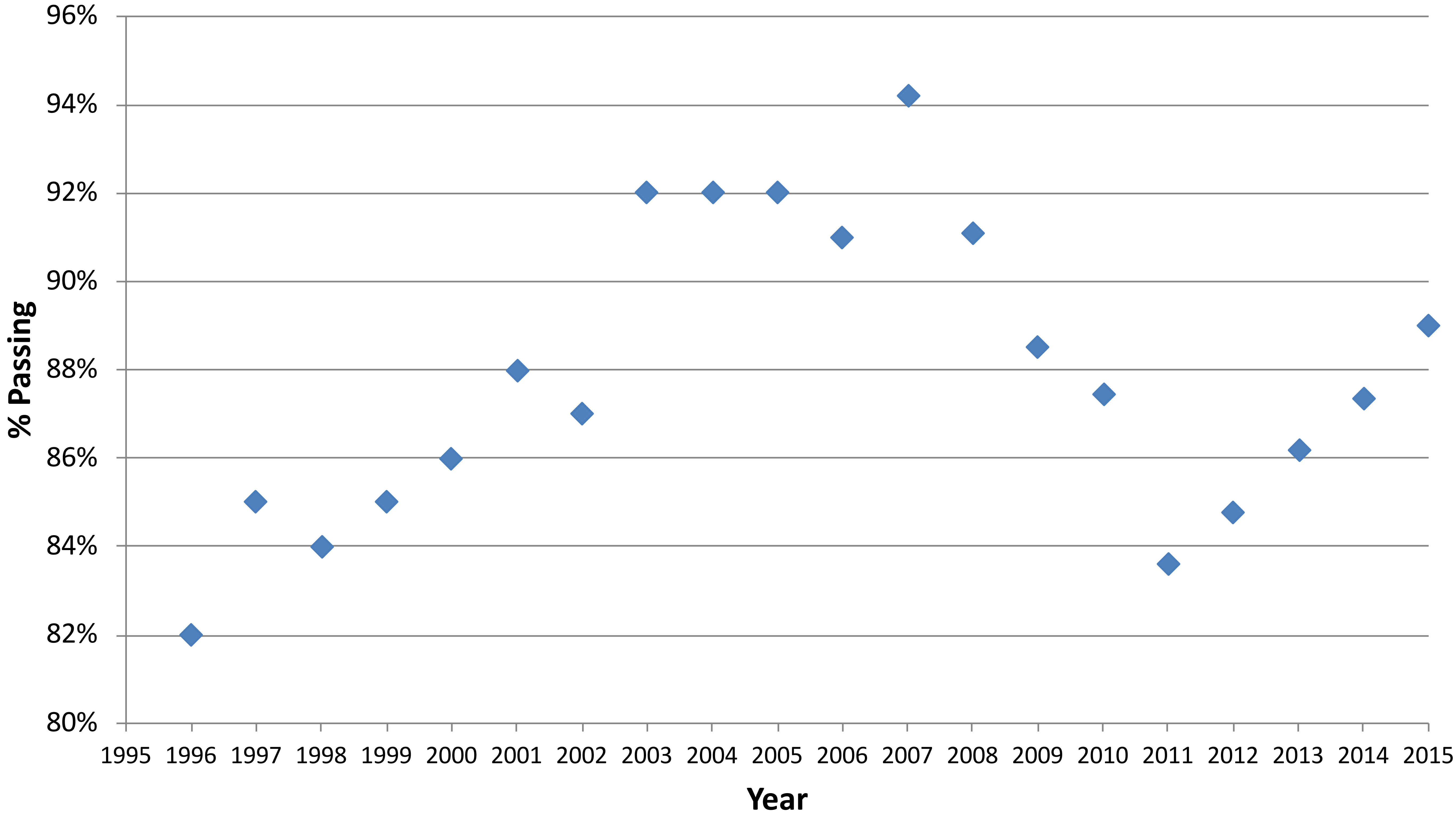
Research Question

Did reforms in 2003 and 2011 affect the pass rates of the certification exam for internal medicine medical residents?

Background

- Medical residents have completed an MD and are in training programs (residencies) before they can practice medicine full-time
- Residencies are difficult positions to have, mostly because of the demanding work schedules and long hour required
- Reforms were passed in 2003 and 2011 to curb the number of hours that a medical resident could work in a given week
  - Goal of reforms: improve patient care, decrease stress on residents
  - Potential side effects: alter the pass rates of the certification exam at the end of residency
    - Less demanding work schedules could mean more time to study and improved pass rates
    - Less exposure to clinical work (fewer total hours worked) could mean lower pass rates

The Data


96%
94%
92%
90%
88%
86%
84%
82%
80%
% Passing
1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
Year

The Data

2003-2010
First reforms active

2011-2015
Second reforms active

Year	N	% Pass
1996	6964	82%
1997	7173	85%
1998	7348	84%
1999	7311	85%
2000	7048	86%
2001	6802	88%
2002	7074	87%
2003	6751	92%
2004	7056	92%
2005	7051	92%
2006	7006	91%
2007	7090	94%
2008	7194	91%
2009	7226	88%
2010	7335	87%
2011	7337	84%
2012	7303	85%
2013	7482	86%
2014	7601	87%
2015	7839	89%

Analysis Questions

- What type of statistical analysis is appropriate to tell if there are differences in pass rates on the exam between these time periods: 1996-2002, 2003-2010, 2011-2015?
- How does the fact that the data points are proportions (percentages) affect the analysis?
- Should the data from different years within a time period be treated as a set to compare between time periods? Or should they be examined for trends within each time period and compared to other time periods for differences in trends?

Follow-up Questions

- What happens with residents who fail, do they retake the exam in the following year? If so, might variation in the pass rate have a lag effect on itself?
- The data I provided are only for first-time test takers, but for the complete data set this point is quite valid.

Question 2

- How long is a residency in internal medicine? If longer than one year, would you expect the effects of a change in policy to phase in over more than one year?
- Residency for internal medicine is 3 years, so in this case it is possible that changes in policy would phase in over the course of a few years.

Question 3

- Many standardized tests are continually adjusted to achieve a target pass rate. Is that so here? If so, would this suggest you should be looking for a reversion to a standard following each new change in policy?
- The tests are standardized to what a passing score is based on content experts' expectations of performance on each question. There is no standardization of passing rate (percentage that obtain a passing score or above).

Question 4

- Are there earlier data that can be used to get a better sense of stable (i.e. not influenced by changes to the system) year-to-year variation in pass rates?
- These are the only data currently available to us.

Question 5

- What is the nature of the two reforms and how did they differ between 2003 and 2011?
- Reforms of 2003 placed an 80-hour cap on the number of hours worked in a given week. Reforms of 2011 placed smaller guidelines to bolster the 80-hour cap with more specific limitations for types of activities that are or are not permitted.

Question 6

- Did these reforms affect other medical disciplines and, if so, could you increase power to detect a policy effect by co-analyzing several related time series?
- All medical disciplines were affected, and we are attempting to obtain multi-year data sets from a few others. Each discipline is a little protective of the long-term data.

Question 7

- Are there any relevant macroeconomic trends that may influence supply/demand for residents, their numbers & quality that may, at least in part, explain some of the temporal variability evident in the data?
- We are not aware of other macroeconomic factors affecting the pass rates at this time. We acknowledge that these data will not be sufficient to draw meaningful trends on their own. We'd like to include these data as a piece of evidence among others to investigate how true current narratives are around the effects of the reforms.

STA470/851 — Statistics Consulting Workshop

Overdispersed Binary Outcome GLMs/GLMMs in R

November 21, 2019

- Medical Residents Data
- Two Structures, One Model
- Overdispersion Problem
- Model for Cluster–Sampled Data
- Quasi–Binomial Models: Two Structures, Two Fits!
- Comparable GLMM Model
- Comparable LM Model

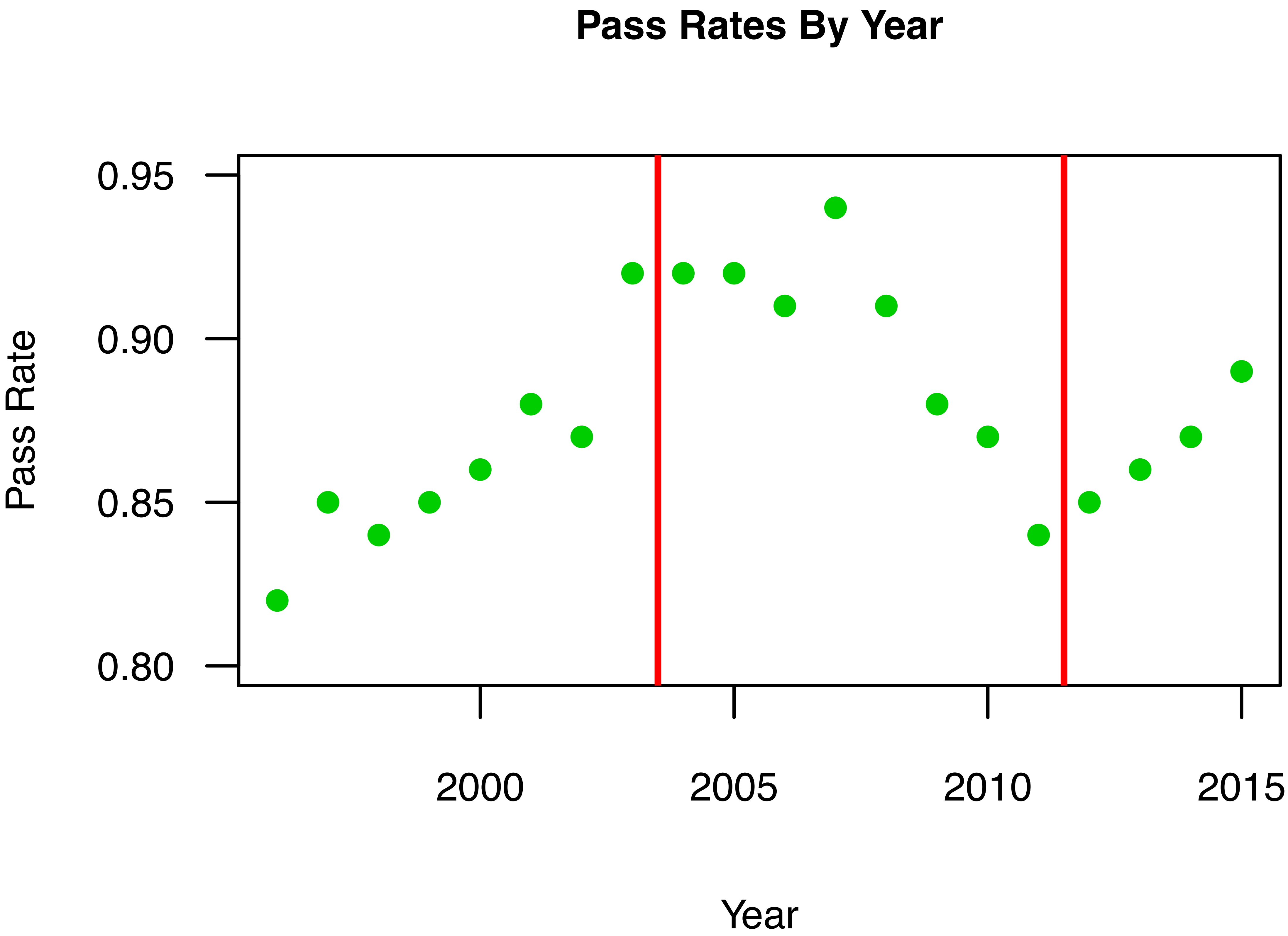
Med Resident Exam Data

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

## [1] 20  3

x$Pass<-round(x$N*x$Pct)
x$Fail<-(x$N - x$Pass)
x$timeperiod<-rep(1,nrow(x))
x$timeperiod[x$Year>2003]<-2
x$timeperiod[x$Year>2011]<-3
x$timeperiod<-factor(x$timeperiod,levels=c(1,2,3),
                     labels=c("tp1","tp2","tp3"))
```

Med Resident Exam Data, Ctd.

```
x$Response<-cbind(x$Pass,x$Fail)
head(x,7)

##   Year    N  Pct Pass Fail timeperiod Response.1 Response.2
## 1 1996 6964 0.82 5710 1254        tp1       5710       1254
## 2 1997 7173 0.85 6097 1076        tp1       6097       1076
## 3 1998 7348 0.84 6172 1176        tp1       6172       1176
## 4 1999 7311 0.85 6214 1097        tp1       6214       1097
## 5 2000 7048 0.86 6061  987        tp1       6061        987
## 6 2001 6802 0.88 5986  816        tp1       5986        816
## 7 2002 7074 0.87 6154  920        tp1       6154        920
```

Time Plot


Pass Rates By Year
0.95
0.90
0.85
0.80
Pass Rate
2000
2005
2010
2015
Year

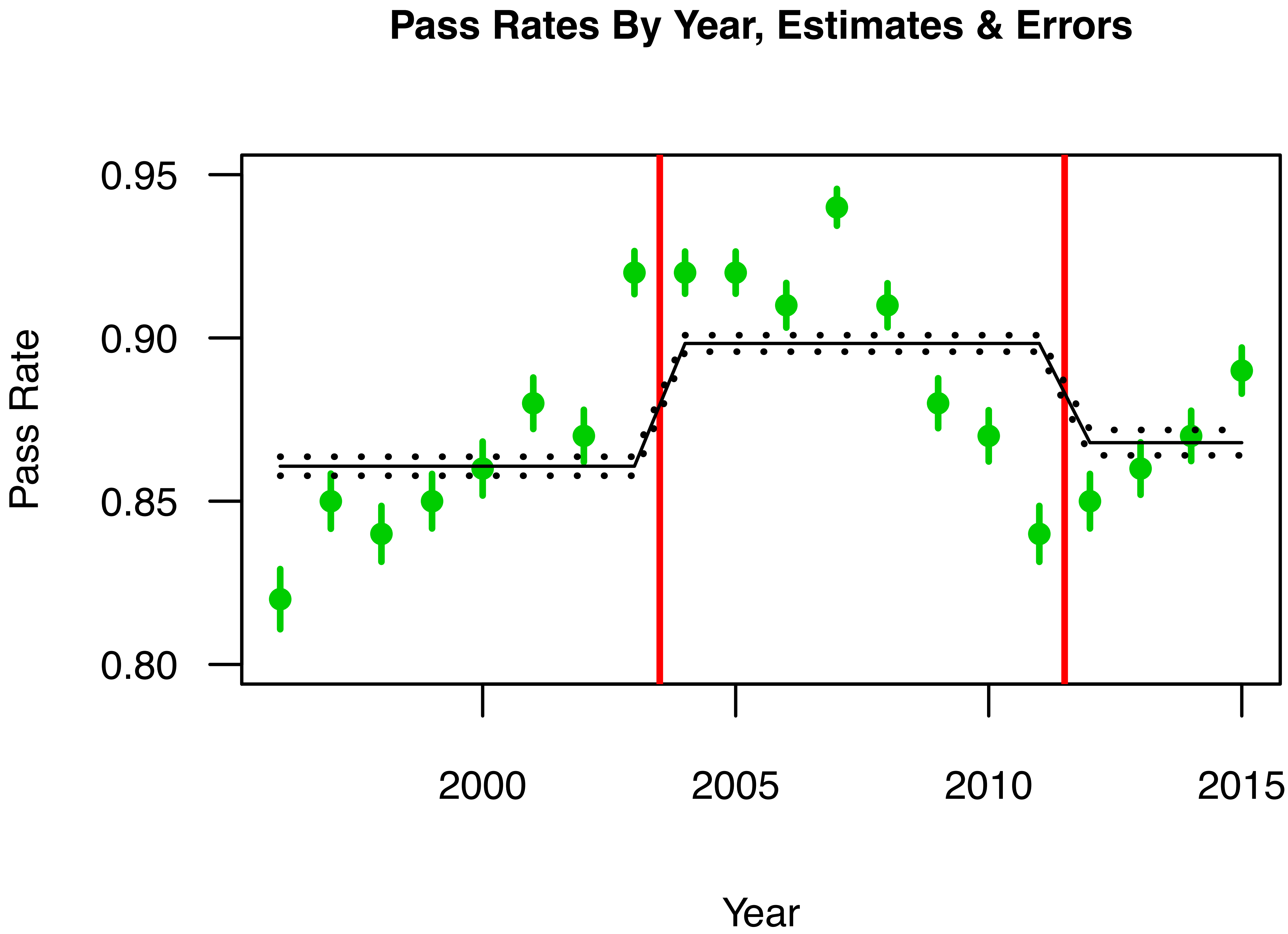
Data, 2nd Representation

```
yP<-data.frame(Year=rep(x$Year,x$Pass),Pass=rep(1,sum(x$Pass)))
yF<-data.frame(Year=rep(x$Year,x$Fail),Pass=rep(0,sum(x$Fail)))
y<-rbind(yP,yF); rm(yP,yF)
y$timeperiod<-rep(1,nrow(y))
y$timeperiod[y$Year>2003]<-2
y$timeperiod[y$Year>2011]<-3
y$timeperiod<-factor(y$timeperiod,levels=c(1,2,3),
                     labels=c("tp1","tp2","tp3"))
head(y,2)

##   Year Pass timeperiod
## 1 1996    1        tp1
## 2 1996    1        tp1
```

Equivalent GLM Fits

```
glm.out0<-glm(Pass ~ timeperiod - 1,
              family=binomial(link=logit),data=y)
summary(glm.out0)$coefficients

##                Estimate Std. Error  z value Pr(>|z|)
## timeperiodtp1 1.821177 0.01215333 149.8501        0
## timeperiodtp2 2.178649 0.01382198 157.6221        0
## timeperiodtp3 1.882726 0.01698883 110.8214        0

glm.out1<-glm(Response ~ timeperiod - 1,
              family=binomial(link=logit),data=x)
summary(glm.out1)$coefficients

##                Estimate Std. Error  z value Pr(>|z|)
## timeperiodtp1 1.821177 0.01215333 149.8500        0
## timeperiodtp2 2.178649 0.01382293 157.6112        0
## timeperiodtp3 1.882726 0.01698885 110.8213        0
```

The Overdispersion Problem



A Model for Clustered Samples

Binomial sampling model conditional on year, y:

$$S_y \sim \text{binomial}(n_y, \pi_y).$$

For convenience, let the number of trials $n_y = m$ for all y, so

$$\begin{aligned} \mathrm{E}(S_y) &= m\pi_y \\ \mathrm{V}(S_y) &= m\pi_y(1-\pi_y). \end{aligned}$$

Now suppose

$$\begin{aligned} \mathrm{E}(\pi_y) &= \pi \\ \mathrm{V}(\pi_y) &= \tau^2\pi(1-\pi). \end{aligned}$$

Model for Clustered Samples, Ctd.

Integrating over the cluster–specific success probabilities,

$$
\begin{aligned}
\mathrm{E}(S_y \,|\, \pi,\, \tau^2) &= \mathrm{E}\big(\mathrm{E}(S_y \,|\, \pi_y) \,\big|\, \pi,\, \tau^2\big) \\
&= \mathrm{E}\big(m\pi_y \,\big|\, \pi,\, \tau^2\big) \\
&= m\pi, \\
\mathrm{V}(S_y \,|\, \pi,\, \tau^2) &= \mathrm{V}\big(\mathrm{E}(S_y \,|\, \pi_y) \,\big|\, \pi,\, \tau^2\big) \;+\; \mathrm{E}\big(\mathrm{V}(S_y \,|\, \pi_y) \,\big|\, \pi,\, \tau^2\big) \\
&= \mathrm{V}\big(m\pi_y \,\big|\, \pi,\, \tau^2\big) \;+\; \mathrm{E}\big(m\pi_y(1-\pi_y)) \,\big|\, \pi,\, \tau^2\big) \\
&= \big((m-1)\tau^2 \;+\; 1\big)\, m\pi(1-\pi) \\
&= \sigma^2\, m\pi(1-\pi),
\end{aligned}
$$

$$
\mathrm{E}(S \,|\, \pi,\, \tau^2) = \mathrm{E}\Big(\sum_{y=1}^{n} S_y \,|\, \pi,\, \tau^2\Big) = n\, m\pi = N\pi,
$$

$$
\mathrm{V}(S \,|\, \pi,\, \tau^2) = \mathrm{V}\Big(\sum_{y=1}^{n} S_y \,|\, \pi,\, \tau^2\Big) = \sigma^2\, n\, m\pi(1-\pi) = \sigma^2\, N\pi(1-\pi).
$$

Notes/Implications

- When $\tau^2 = 0$ there is no overdispersion.
- When $m = 1$ it isn't detectable.
- Underdispersion is not possible.
- When model for S_y given y is correctly specified, i.e. $S_y \sim \text{binomial}(n_y, \pi_y)$, then there is, by definition, no overdispersion.
- The point of the marginalization exercise above is to demonstrate the effect of a particular form of model misspecification.

Notes/Implications, Continued

- The Beta–Binomial model is an alternative.
- Its dispersion parameter is $\sigma^2 = ((N-1)\tau^2 + 1)$,
- where τ^2 a function of the beta distribution's parameters.
- Compare with cluster sampling model's $\sigma^2 = ((m-1)\tau^2 + 1)$.
- There are other options, too.
- Reference: McCullagh & Nelder (1989). Generalized Linear Models, Second Edition. New York: Chapman and Hall. Chapter 4.5.

Back to R: Binary Outcome Formulation

$$\int \left(\frac{Y-\mu}{\sigma^2 V(\mu)} \right)$$

```
qb.out0<- glm(Pass ~ timeperiod - 1,
              family=quasibinomial(link=logit),data=y)
summary(qb.out0)$coefficients

##                Estimate Std. Error  t value Pr(>|t|)
## timeperiodtp1 1.821177 0.01215379 149.8443        0
## timeperiodtp2 2.178649 0.01382251 157.6160        0
## timeperiodtp3 1.882726 0.01698948 110.8172        0

summary(qb.out0)$dispersion

## [1] 1.000077
```

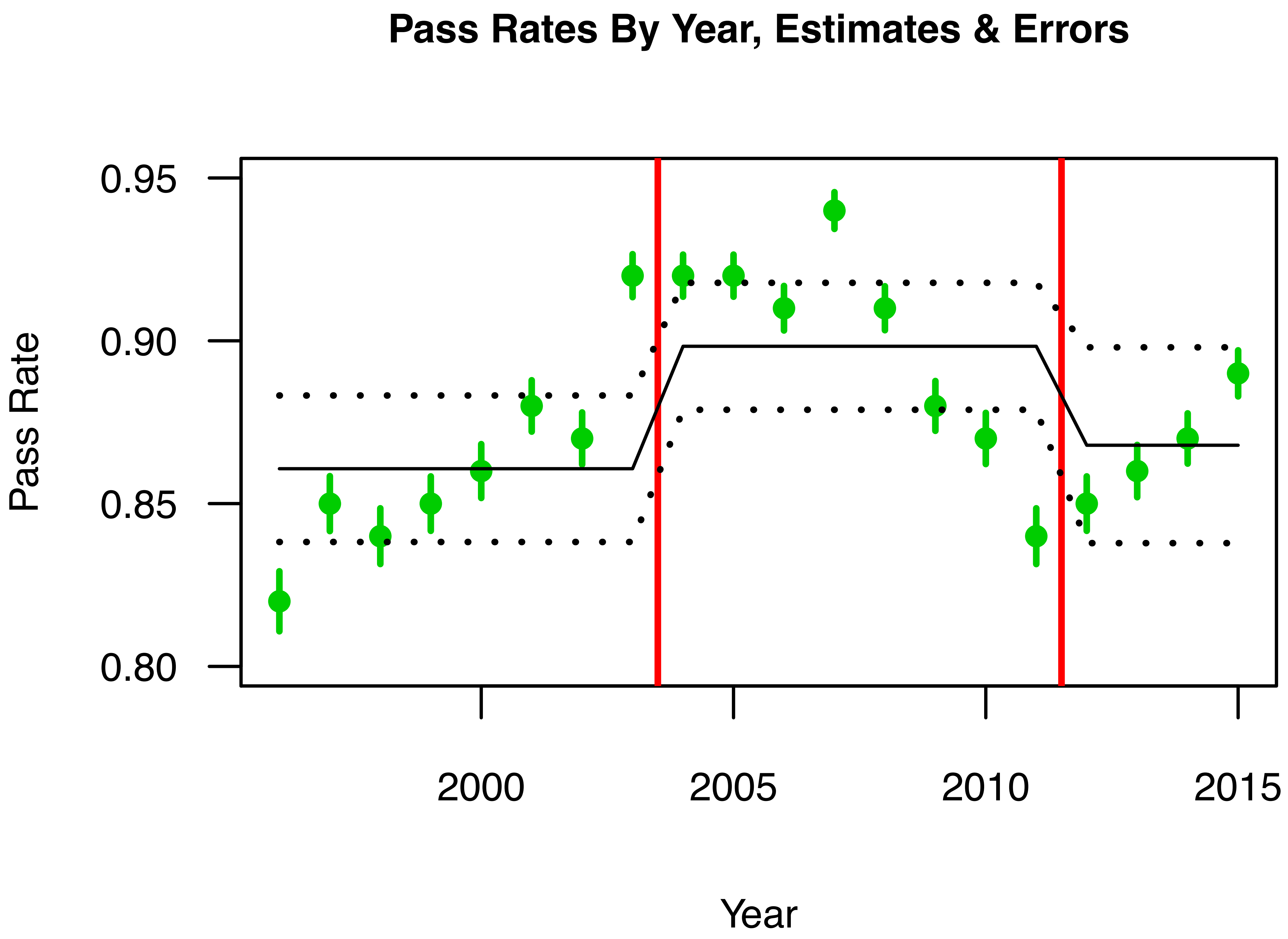
Do you believe this?!

Binomial/Grouped Formulation

```
qb.out1<-glm(Response ~ timeperiod - 1,
             family=quasibinomial(link=logit),data=x)
summary(qb.out1)$coefficients

##                Estimate Std. Error  t value     Pr(>|t|)
## timeperiodtp1 1.821177 0.09381746 19.41192 4.871359e-13
## timeperiodtp2 2.178649 0.10670594 20.41732 2.134866e-13
## timeperiodtp3 1.882726 0.13114517 14.35604 6.198861e-11

summary(qb.out1)$dispersion

## [1] 59.59047
```

How about this?

Plot With Overdispersion Model Estimates



Sanity/Model Consistency Check

```
qb.null<-glm(Response ~ 1,
             family=quasibinomial(link=logit),data=x)
summary(qb.null)$dispersion

## [1] 72.16216

phat<-mean(x$Pct)
vhat<-(phat*(1-phat))
tau2<-var(x$Pct)/vhat    ## Tau^2
## Cluster--Sampled DP
1 + (mean(x$N) - 1)*tau2

## [1] 74.22329
```

GLMM Implementation?

Hierarchical Model:

$$\begin{aligned} S_y &\sim \text{binomial}(n_y, \pi_y) \\ \text{logit}(\pi_y) &= \alpha + \beta_y \\ \beta_y &\sim \text{N}(0, \tau^2) \end{aligned}$$

Fit to medical residents data:

```
library(lme4)
x$fYear<-as.factor(x$Year)
glmm.null<-glmer(Response ~ (1|fYear),
                 family=binomial(link=logit),data=x)
```

GLMM Fit & Comparison

```
summary(glmm.null)$coefficients

##             Estimate Std. Error z value      Pr(>|z|)
## (Intercept) 2.006964  0.0728941 27.5326 7.151449e-167

summary(qb.null)$coefficients; qb.null$df.residual

##             Estimate Std. Error  t value     Pr(>|t|)
## (Intercept) 1.966055 0.06820535 28.82553 3.805403e-17
## [1] 19

glm.null<-glm(Response ~ 1,
              family=binomial(link=logit),data=x)
summary(glm.null)$coefficients

##             Estimate  Std. Error z value Pr(>|z|)
## (Intercept) 1.966055 0.008029041 244.868        0
```

GLMM Fit & Comparison, Continued

```
eta<-rnorm(10e5,mean=2.00696,sd=0.3238)
pi.hat<-exp(eta)/(1 + exp(eta))
var(pi.hat)

## [1] 0.00122288

var(x$Pct)

## [1] 0.001093421

pi.mu<-mean(pi.hat)
var(pi.hat)/(pi.mu*(1-pi.mu)) ## comp to tau^2

## [1] 0.01136711

tau2  ## Quasi Binomial Dispersion Parameter tau^2

## [1] 0.01017195
```

GLMM Z–Score

Recall, the issue:

```
summary(qb.null)$coefficients

##               Estimate Std. Error  t value     Pr(>|t|)
## (Intercept) 1.966055 0.06820535 28.82553 3.805403e-17

qb.null$df.residual

## [1] 19

summary(glmm.null)$coefficients

##               Estimate Std. Error z value       Pr(>|z|)
## (Intercept) 2.006964  0.0728941 27.5326 7.151449e-167

library(pbkrtest)
glmm.nullNull<-glmer(Response ~ -1 + (1|fYear),
                     family=binomial(link=logit),data=x)
```

GLMM Z–Score, Continued

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Response ~ -1 + (1 | fYear)
##    Data: x
##
##      AIC      BIC   logLik deviance df.resid
##    351.6    352.6   -174.8    349.6       19
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## 0.01142 0.01383 0.01608 0.02310 0.03323
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  fYear  (Intercept) 4.138    2.034
## Number of obs: 20, groups:  fYear, 20
```

GLMM Z–Score, Parametric Bootstrap P–Value

```
##system.time(
##   pb.out<-PBmodcomp(largeModel=glmm.null,
##                     smallModel=glmm.nullNull,
##                     nsim=500000)
##)
pb.out

## Parametric bootstrap test; time: 48448.81 sec; samples: 500000 extre
## Requested samples: 500000 Used samples: 498978 Extremes: 0
## large : Response ~ (1 | fYear)
## small : Response ~ -1 + (1 | fYear)
##          stat df   p.value
## LRT    73.187  1 < 2.2e-16 ***
## PBtest 73.187    2.004e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

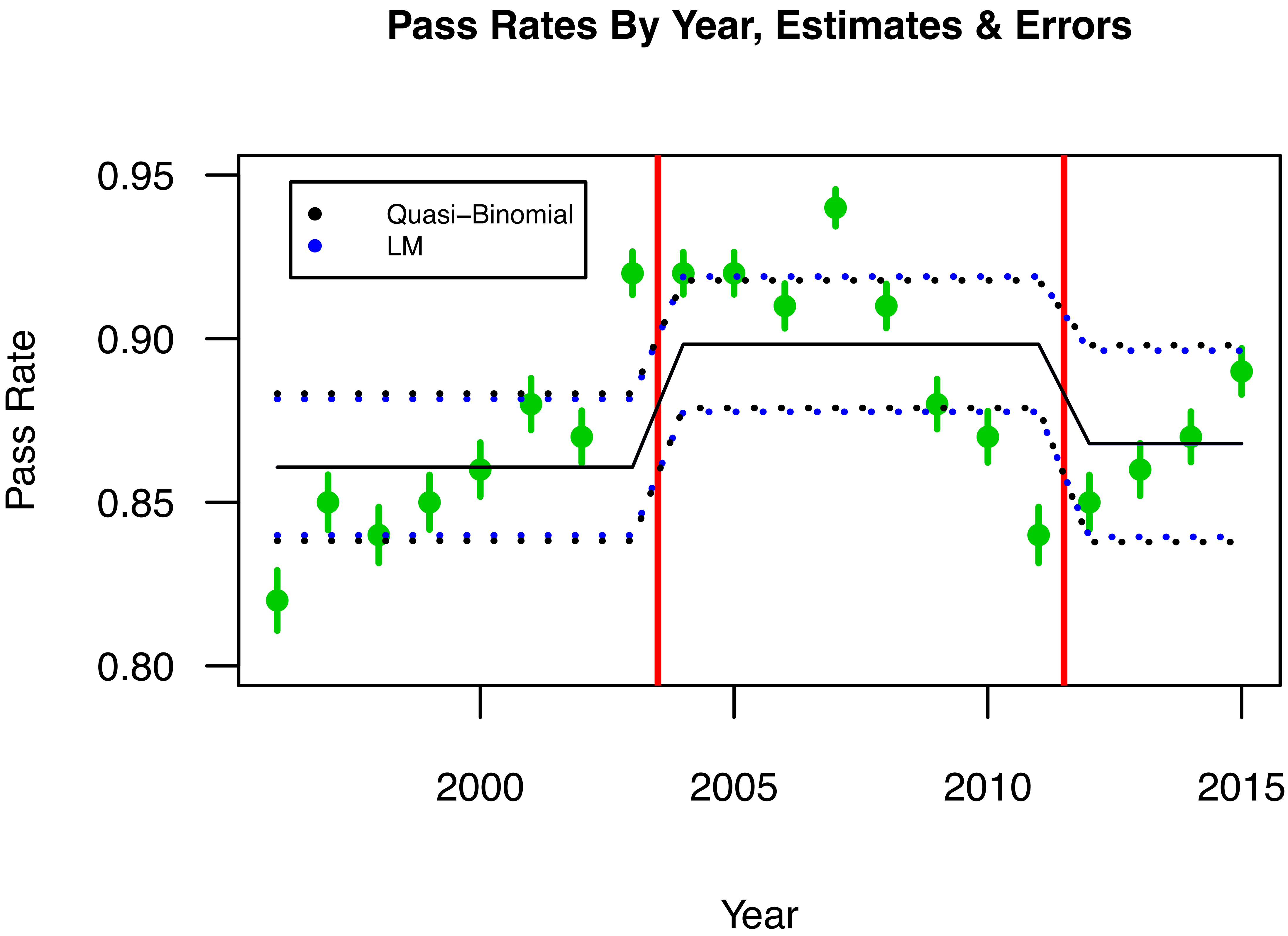
GLMM Z–Score, Continued

```
## Parametric bootstrap test; time: 48448.81 sec; samples: 500000 extre
## Requested samples: 500000 Used samples: 498978 Extremes: 0
## large : Response ~ (1 | fYear)
## small : Response ~ -1 + (1 | fYear)
##              stat     df    ddf   p.value
## PBtest   73.187                 2.004e-06 ***
## Gamma    73.187                 2.220e-16 ***
## Bartlett 67.544  1.000          2.220e-16 ***
## F        73.187  1.000 25.939 5.012e-09 ***
## LRT      73.187  1.000          < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

LM

```
lm.out<-lm(Pct ~ timeperiod - 1,weights=N,data=x)
summary(lm.out)$coefficients

##                Estimate Std. Error  t value     Pr(>|t|)
## timeperiodtp1 0.8607345 0.01041069 82.67796 1.364067e-23
## timeperiodtp2 0.8983061 0.01033556 86.91416 5.844508e-24
## timeperiodtp3 0.8678792 0.01423014 60.98882 2.365630e-21
```

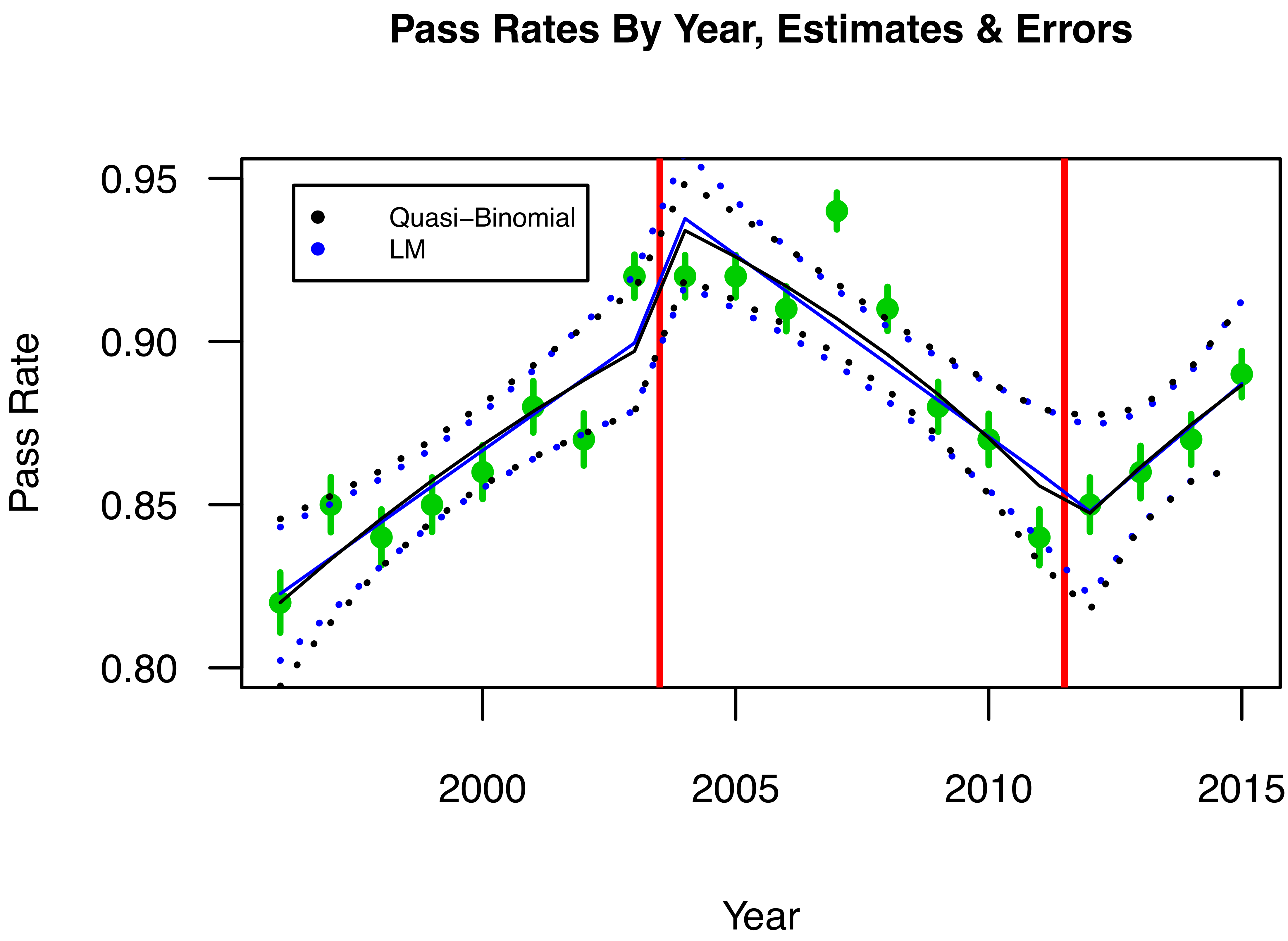
LM, Continued



QB and LM Model Formulation 2

```
qb.out2<-glm(Response ~ -1 + timeperiod*scale(Year),
             family=quasibinomial(link=logit),data=x)
lm.out2<-lm(Pct ~ -1 + timeperiod*scale(Year),weights=N,data=x)
```

Formulation 2, Continued



Take–Aways

1. Quasi–Binomial Models: Two Structures, Two Fits!
2. QB vs GLMM vs LM: Three Models, Similar Fits.
3. There are often several reasonable ways to formulate the same problem. See points 3 & 4 above.
4. Sanity–check results, e.g. against empirical summaries.
5. Results not believable? Triple–check them!