Preliminary Report

Sebastián Soriano Pérez

7/15/2020

Quantifying the Effects of the 1998 Predatory Lending Law in North Carolina

Abstract

This report quantifies the impact of North Carolina's Predatory Lending Law from 1999 in the state's subprime market. We used the Home Mortgage and Disclosure Act (HMDA) Loan Application Register (LAR) datasets in conjunction with the US Department of Housing and Urban Development (HUD)'s classification of subprime lenders and additional datasets from the Federal Reserve Economic Data (FRED) to measure any any changes that occured in North Carolina's subprime market before the law's implementation (from 1990 to 1998), and after the intervention (from 2001 to 2006). We executed an Interrupted Time Series Analysis and a Differences in Differences model to compare North Carolina's subprime market to those of other five states with similar socioeconomic characteristics: Alabama, Georgia, Minnesotta, South Carolina, and Virginia.

[Summary of the results here.]

Introduction

The mortgage market saw an increase during the 1990s, especially in the subprime market. The subprime market is typically defined as those loans that are lended to high-risk borrowers with the following characteristics: low credit scores (below 660), at least two 30 day or at least one 60-day delinquencies in the past two years, at least one foreclosure in the past two years or at least one bankrupcy in the past five years (Quercia 2004). Loans with high loan to value rates (LTV), high payment to income rates, for borrowers with limited documentation of income, or unstable income also represent additional risk to lenders. To make up for the higher risks associated with this kind of borrowers, subprime loans usually have characteristics that are more expensive to borrowers in the long term despite typically having small monthly payments at first: higher interest rates, high prepayment penalties, higher fees, balloon payments, among others. The predatory market is a subset of the subprime market that is abusive to borrowers. Some sectors of the population are especially vulnerable to this type of practices.

Lending to previously underserved population including immigrants and minorities grew fast during a short period of time. From 1990 to 1999 the number of loans originated by subprime lenders grew exponentially in North Carolina. In 1990 the HMDA data registers only 73 subprime loans originated by subprime lenders in the state, and this number kept growing until it reached 43,725 in 1999, an increase of 598% in one decade. The Home Ownership and Equity Protection Act of 1994 (HOEPA) was the first attempt at a national level to regulate the predatory lending market. In 1999 North Carolina became the first state to implement an anti-predatory lending law in order to protect borrowers from this practices. Other states have also passed strong anti-predatory lending legislation, including:

- Georgia: Georgia Fair Lending Act in October 1, 2002
- South Carolina: Predatory Mortgage Lending Prevention Act in July 1 and October 1, 2000

Other states that have passed minimal predatory lending laws:

- Alabama: Payday Lending Law in 2003 (minimal impact)
- Minnesota: prohibition of "equity stripping" in 2004
- Virginia: Virginia Fairness in Lending Act in 2020

Data

State Selection

For our analysis we selected five states that were socioeconomically similar to North Carolina. Using census data and yearly estimates from the National Cancer Institute, three different clustering techniques were applied to the 50 states' population data from 1990 to 2009, to find those that are more similar to North Carolina. The techniques used were K-means clustering, agglomerative clustering, and TSNE for dimensionality reduction.

The population dataset was simplified to contain the following variables: - Year (from 1990 to 2009) - State (all 50 states plus DC were considered) - Race (simplified into 1 - white or 2 - other) - Age (simplified into 1 - 0-19 years old, 2 - 20-34 years old, 3 - 35-49 years old, 4 - 50-64 years old, 5 - 65+ years old) - Sex (1 - male or 2 - female)

The dataset was transposed to create a vector for each of the 50 states plus DC, with the population for each year-race-age-sex group. After running the clustering algorithms on this data, the states that coincide across the three different models were analyzed and selected to our discretion. The final states are the following:

- Alabama
- Georgia
- Minnesota
- North Carolina
- South Carolina
- Virginia

[Explain income data here.]

HMDA and HUD data

We used the HMDA LAR datasets from 1990 to 2006, to cover a 16 year period around 1999 when the predatory lending law went into effect in North Carolina. The data was collected for the six selected states. In addition, each row in this dataset will be classified as subprime or not by merging the data with information from HUD. The HUD catalogue on subprime lenders classifies a lender as "subprime" if more than half of the loans they issue were subprime [research their definition of a subprime loan...]. Additional FRED data for house price index (HPI), and median household income was collected.

The final dataset uses this sources along with the NHI population data to identify the following variables for each of the six states per year:

- Year
- State
- Number of subprime loans issued
- Percentage of loans issued that are subprime
- Total loan amount for loans issued
- Percentage of total loan amount that went to subprime loans
- Percentage of borrowers that belong to racial minorities
- Average income for subprime borrowers
- State population
- Median household income
- House Price Index

```
county_data <- read.csv('../Data/HMDASubprime/county_data.csv')
county_data_2 <- read.csv('../Data/HMDASubprime/county_data_2.csv')

county_data$prime <- county_data$total - county_data$subprime
county_data$prime_originated <- county_data$loans_originated - county_data$subprime_originated
county_data$prime_amount <- county_data$loan_amount - county_data$subprime_amount
county_data$state <- 'AL'
county_data[county_data$state_code == 13,]$state <- 'GA'
```

```
county_data[county_data$state_code == 27,]$state <- 'MN'
county_data[county_data$state_code == 37,]$state <- 'NC'
county_data[county_data$state_code == 45,]$state <- 'SC'
county_data[county_data$state_code == 51,]$state <- 'VA'
county_data$state <- as.factor(county_data$state)
county_data_2$prime <- 1 - county_data_2$subprime

county_population_data <- read.csv('../Data/Population/county_pop2.csv')
county_population_data <- county_population_data[county_population_data$year >= 1990 & county_populatio
county_population_data <- county_population_data[county_population_data$state %in% c('AL', 'GA', 'MN',
county_population_data <- county_population_data %>% group_by(year, state, state_fips_x, county_fips) %

county_data <- inner_join(county_data, county_population_data, by=c('as_of_year' = 'year', 'state' = 's
                                                                    'state_code' = 'state_fips_x',
                                                                    'county_code' = 'county_fips'))
```

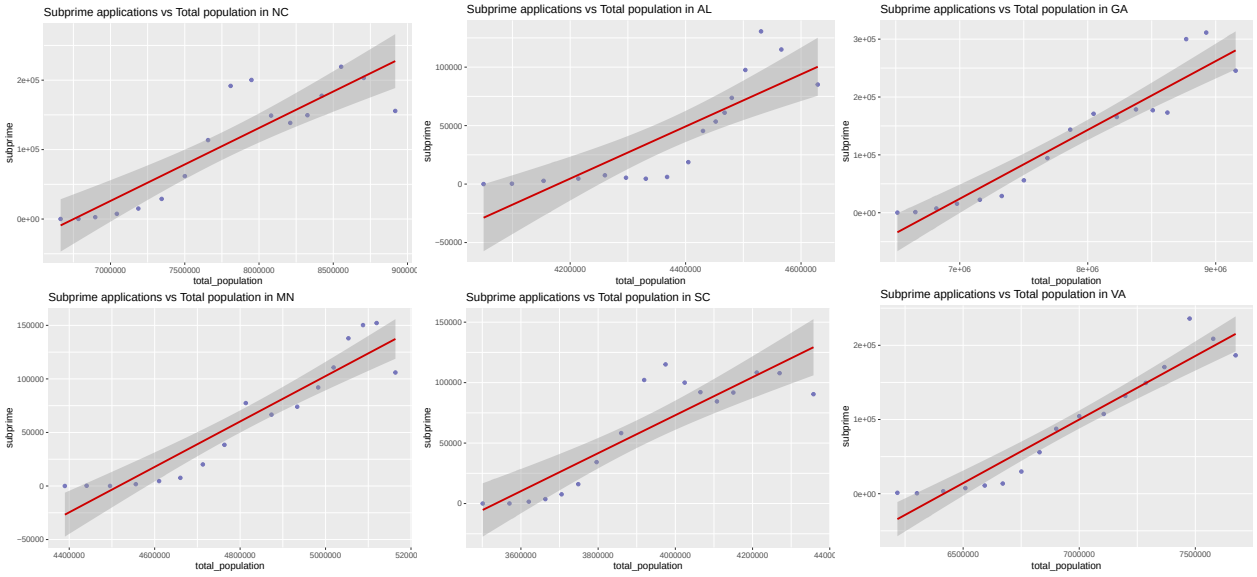
```
## Warning: Column `state` joining factors with different levels, coercing to
## character vector
```

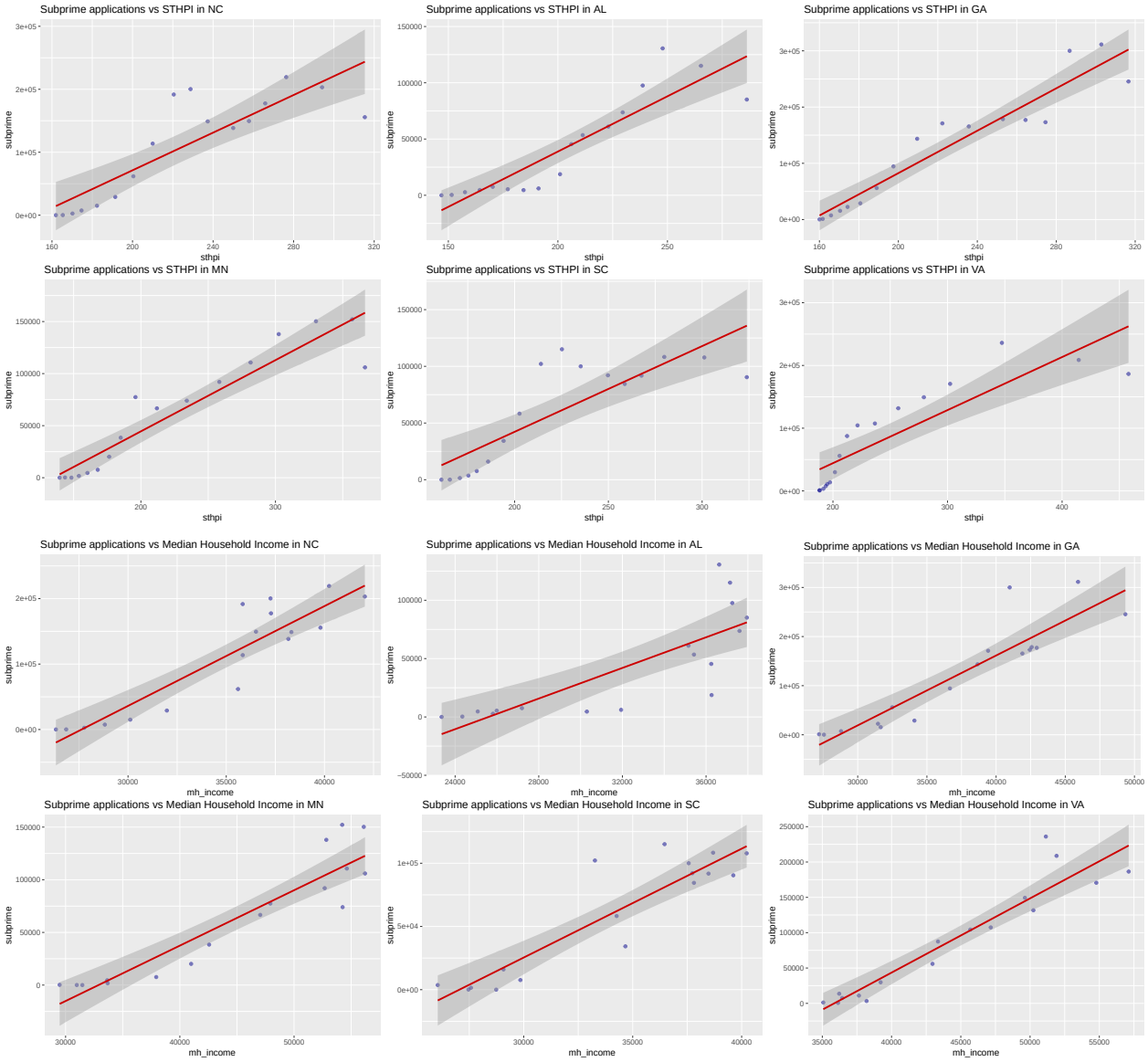
```
county_data <- inner_join(county_data, hpi, by=c('as_of_year' = 'year', 'state' = 'state'))

county_data <- inner_join(county_data, income, by=c('as_of_year' = 'year', 'state' = 'state'))
```

Exploratory Data Analysis

```
aux <- state_data[, c(2, 3, 22, 23, 24, 25, 26, 28, 29, 31, 32, 33, 34)]
aux$subprime_perc <- state_data_2$subprime
aux$subprime_originated_perc <- state_data_2$subprime_originated
```





Interrupted Time Series Analysis

```
#North Carolina
st_data <- aux[aux$state == 'NC',]

prelaw <- st_data[st_data$as_of_year <= 1998,]
transition <- st_data[(st_data$as_of_year > 1998) & (st_data$as_of_year <= 2000),]
postlaw <- st_data[st_data$as_of_year > 2000,]

prelaw_model <- lm(formula=(subprime ~ as_of_year), data=prelaw)
summary(prelaw_model)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year), data = prelaw)
##
## Residuals:
```

```
##    Min     1Q Median     3Q    Max
## -38647 -26590  -2649  15677  61729
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -41386980    9209682  -4.494  0.00282 **
## as_of_year      20779       4619   4.499  0.00280 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35780 on 7 degrees of freedom
## Multiple R-squared:  0.743,  Adjusted R-squared:  0.7063
## F-statistic: 20.24 on 1 and 7 DF,  p-value: 0.002802
```

```
postlaw_model <- lm(formula=(subprime ~ as_of_year), data=postlaw)
summary(postlaw_model)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year), data = postlaw)
##
## Residuals:
##     70     76     82     88     94    100
## -14852 -12066   7746  41223  16837 -38889
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -16353466   14995315  -1.091    0.337
## as_of_year       8249       7485   1.102    0.332
##
## Residual standard error: 31310 on 4 degrees of freedom
## Multiple R-squared:  0.2329, Adjusted R-squared:  0.04119
## F-statistic: 1.215 on 1 and 4 DF,  p-value: 0.3322
```

```
predicted <- data.frame(st_data$as_of_year, predict(prelaw_model, newdata=st_data))
names(predicted) <- c('as_of_year', 'subprime')

prelaw_model2 <- lm(formula=(subprime ~ as_of_year + total_population + sthpi + mh_income), data=prelaw
summary(prelaw_model2)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year + total_population + sthpi +
##     mh_income), data = prelaw)
##
## Residuals:
##        4       10       16       22       28       34       40       46
##  -3562.3   4325.6  -2955.3  15212.5  -2646.1 -19113.3   4180.1   -547.7
##       52
##   5106.4
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -3.458e+07  1.627e+08  -0.212   0.8421
## as_of_year        1.826e+04  8.415e+04   0.217   0.8388
```

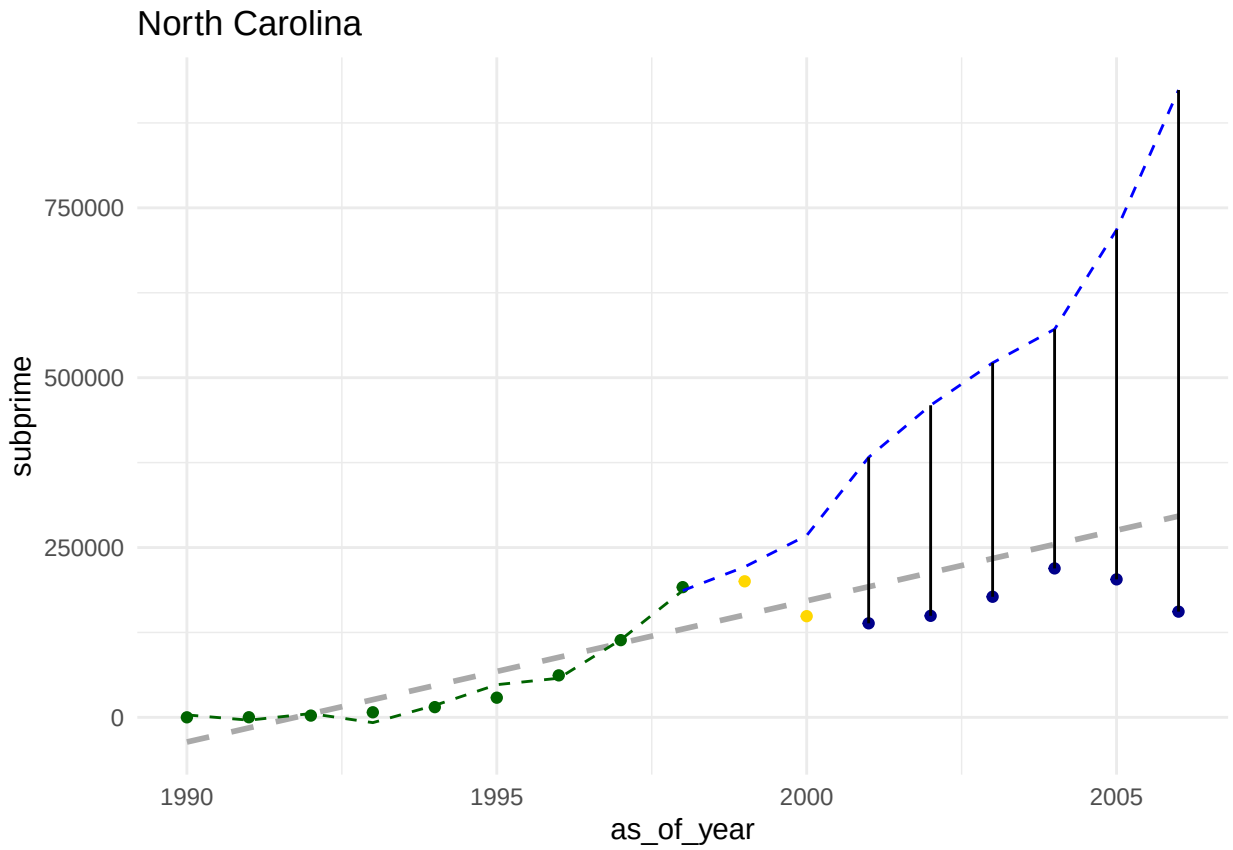
```
## total_population -5.331e-01  8.325e-01  -0.640   0.5568
## sthpi             1.293e+04  4.641e+03   2.786   0.0495 *
## mh_income        -1.150e+01  6.358e+00  -1.809   0.1447
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13110 on 4 degrees of freedom
## Multiple R-squared:  0.9803, Adjusted R-squared:  0.9606
## F-statistic:  49.7 on 4 and 4 DF,  p-value: 0.001152
```

```
predicted2 <- data.frame(st_data$as_of_year, predict(prelaw_model2, newdata=st_data))
names(predicted2) <- c('as_of_year', 'subprime')

predicted2_pre <- predicted2[predicted2$as_of_year <= 1998,]
predicted2_pos <- predicted2[predicted2$as_of_year >= 1998,]

differences <- postlaw[, c('as_of_year', 'subprime')]
differences$subprime_predicted <- predicted2[predicted2$as_of_year > 2000,]$subprime
differences$difference <- differences$subprime - differences$subprime_predicted

ggplot(NULL, aes(x=as_of_year, y=subprime)) + geom_point(data=prelaw, color='darkgreen') +
  geom_point(data=transition, color='gold') +
  geom_point(data=postlaw, color='darkblue') +
  geom_smooth(data=prelaw, method=lm, se=FALSE, color='darkgray', fullrange=TRUE, linetype='dashed') +
  geom_line(data=predicted2_pre, color='darkgreen', linetype='dashed') +
  geom_line(data=predicted2_pos, color='blue', linetype='dashed') +
  #geom_text(data=prelaw, x=1995, y=150000, label=lm_eqn(prelaw_model), parse=TRUE) +
  geom_segment(data=differences,
               aes(x=as_of_year, xend=as_of_year, y=subprime, yend=subprime_predicted)) +
  theme_minimal() + ggtitle('North Carolina')
```




```
#Alabama
st_data <- aux[aux$state == 'AL',]

prelaw <- st_data[st_data$as_of_year <= 2006,]
#transition <- st_data[(st_data$as_of_year > 1998) & (st_data$as_of_year <= 2000),]
#postlaw <- st_data[st_data$as_of_year > 2000,]

prelaw_model <- lm(formula=(subprime ~ as_of_year), data=prelaw)
summary(prelaw_model)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year), data = prelaw)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -27753 -12324  -2165  14776  40389
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.604e+07  1.876e+06  -8.549 3.77e-07 ***
## as_of_year   8.050e+03  9.392e+02   8.571 3.65e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18970 on 15 degrees of freedom
## Multiple R-squared:  0.8304, Adjusted R-squared:  0.8191
```

```
## F-statistic: 73.47 on 1 and 15 DF,  p-value: 3.646e-07
```

```
postlaw_model <- lm(formula=(subprime ~ as_of_year), data=postlaw)
summary(postlaw_model)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year), data = postlaw)
##
## Residuals:
##     70     76     82     88     94    100
## -14852 -12066   7746  41223  16837 -38889
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -16353466   14995315  -1.091    0.337
## as_of_year       8249       7485   1.102    0.332
##
## Residual standard error: 31310 on 4 degrees of freedom
## Multiple R-squared:  0.2329, Adjusted R-squared:  0.04119
## F-statistic: 1.215 on 1 and 4 DF,  p-value: 0.3322
```

```
predicted <- data.frame(st_data$as_of_year, predict(prelaw_model, newdata=st_data))
names(predicted) <- c('as_of_year', 'subprime')

prelaw_model2 <- lm(formula=(subprime ~ as_of_year + total_population + sthpi + mh_income), data=prelaw
summary(prelaw_model2)
```

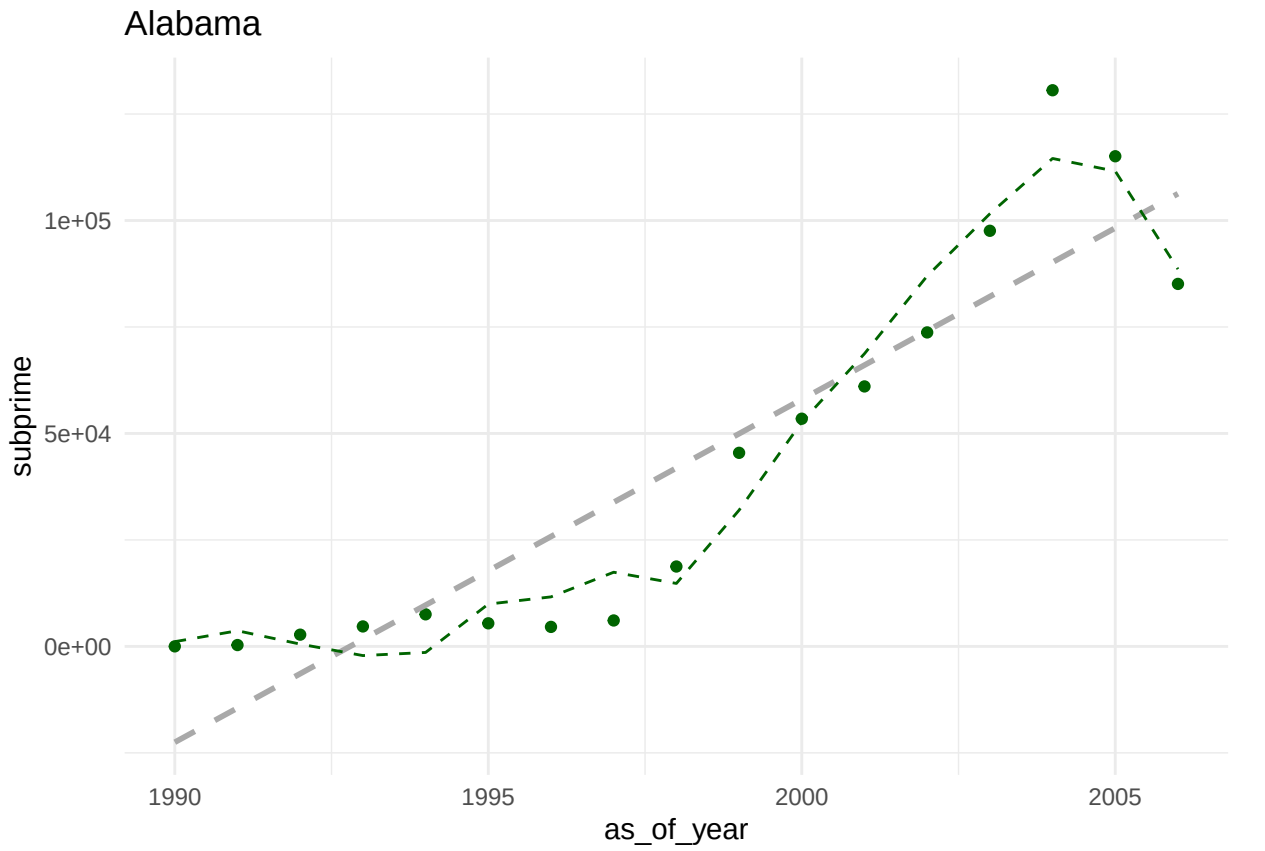
```
##
## Call:
## lm(formula = (subprime ~ as_of_year + total_population + sthpi +
##     mh_income), data = prelaw)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -13249  -4516  -1110   3972  16018
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -7.160e+07  1.060e+07  -6.754 2.03e-05 ***
## as_of_year        3.723e+04  5.471e+03   6.804 1.89e-05 ***
## total_population -5.618e-01  8.821e-02  -6.368 3.57e-05 ***
## sthpi            -1.112e+03  4.314e+02  -2.578   0.0242 *
## mh_income        -1.917e+00  1.479e+00  -1.296   0.2194
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9450 on 12 degrees of freedom
## Multiple R-squared:  0.9663, Adjusted R-squared:  0.9551
## F-statistic: 86.12 on 4 and 12 DF,  p-value: 9.889e-09
```

```
predicted2 <- data.frame(st_data$as_of_year, predict(prelaw_model2, newdata=st_data))
names(predicted2) <- c('as_of_year', 'subprime')

predicted2_pre <- predicted2[predicted2$as_of_year <= 2006,]
predicted2_pos <- predicted2[predicted2$as_of_year >= 2006,]
```

```
differences <- postlaw[, c('as_of_year', 'subprime')]
differences$subprime_predicted <- predicted2[predicted2$as_of_year >= 2006,]$subprime
differences$difference <- differences$subprime - differences$subprime_predicted

ggplot(NULL, aes(x=as_of_year, y=subprime)) + geom_point(data=prelaw, color='darkgreen') +
  #geom_point(data=transition, color='gold') +
  #geom_point(data=postlaw, color='darkblue') +
  geom_smooth(data=prelaw, method=lm, se=FALSE, color='darkgray', fullrange=TRUE, linetype='dashed') +
  geom_line(data=predicted2_pre, color='darkgreen', linetype='dashed') +
  #geom_line(data=predicted2_pos, color='blue', linetype='dashed') +
  #geom_text(data=prelaw, x=1995, y=150000, label=lm_eqn(prelaw_model), parse=TRUE) +
  #geom_segment(data=differences,
               #aes(x=as_of_year, xend=as_of_year, y=subprime, yend=subprime_predicted)) +
  theme_minimal() + ggtitle('Alabama')
```




```
#Georgia
st_data <- aux[aux$state == 'GA',]

prelaw <- st_data[st_data$as_of_year <= 2001,]
transition <- st_data[(st_data$as_of_year > 2001) & (st_data$as_of_year <= 2002),]
postlaw <- st_data[st_data$as_of_year > 2002,]

prelaw_model <- lm(formula=(subprime ~ as_of_year), data=prelaw)
summary(prelaw_model)
```

```
##
## Call:
```

```
## lm(formula = (subprime ~ as_of_year), data = prelaw)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -35275 -13431    107  15592  31698
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -38095515    3839803  -9.921 1.71e-06 ***
## as_of_year      19128       1924   9.940 1.68e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23010 on 10 degrees of freedom
## Multiple R-squared:  0.9081, Adjusted R-squared:  0.8989
## F-statistic: 98.81 on 1 and 10 DF,  p-value: 1.679e-06
```

```
postlaw_model <- lm(formula=(subprime ~ as_of_year), data=postlaw)
summary(postlaw_model)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year), data = postlaw)
##
## Residuals:
##     80     86     92     98
## -50119  53968  42421 -46270
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -45559369   61342011  -0.743    0.535
## as_of_year      22857      30602   0.747    0.533
##
## Residual standard error: 68430 on 2 degrees of freedom
## Multiple R-squared:  0.2181, Adjusted R-squared:  -0.1728
## F-statistic: 0.5579 on 1 and 2 DF,  p-value: 0.533
```

```
predicted <- data.frame(st_data$as_of_year, predict(prelaw_model, newdata=st_data))
names(predicted) <- c('as_of_year', 'subprime')

prelaw_model2 <- lm(formula=(subprime ~ as_of_year + total_population + sthpi + mh_income), data=prelaw
summary(prelaw_model2)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year + total_population + sthpi +
##     mh_income), data = prelaw)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -18830  -7471  -2682   7508  21179
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.907e+08  1.310e+08   2.218   0.0620 .
```

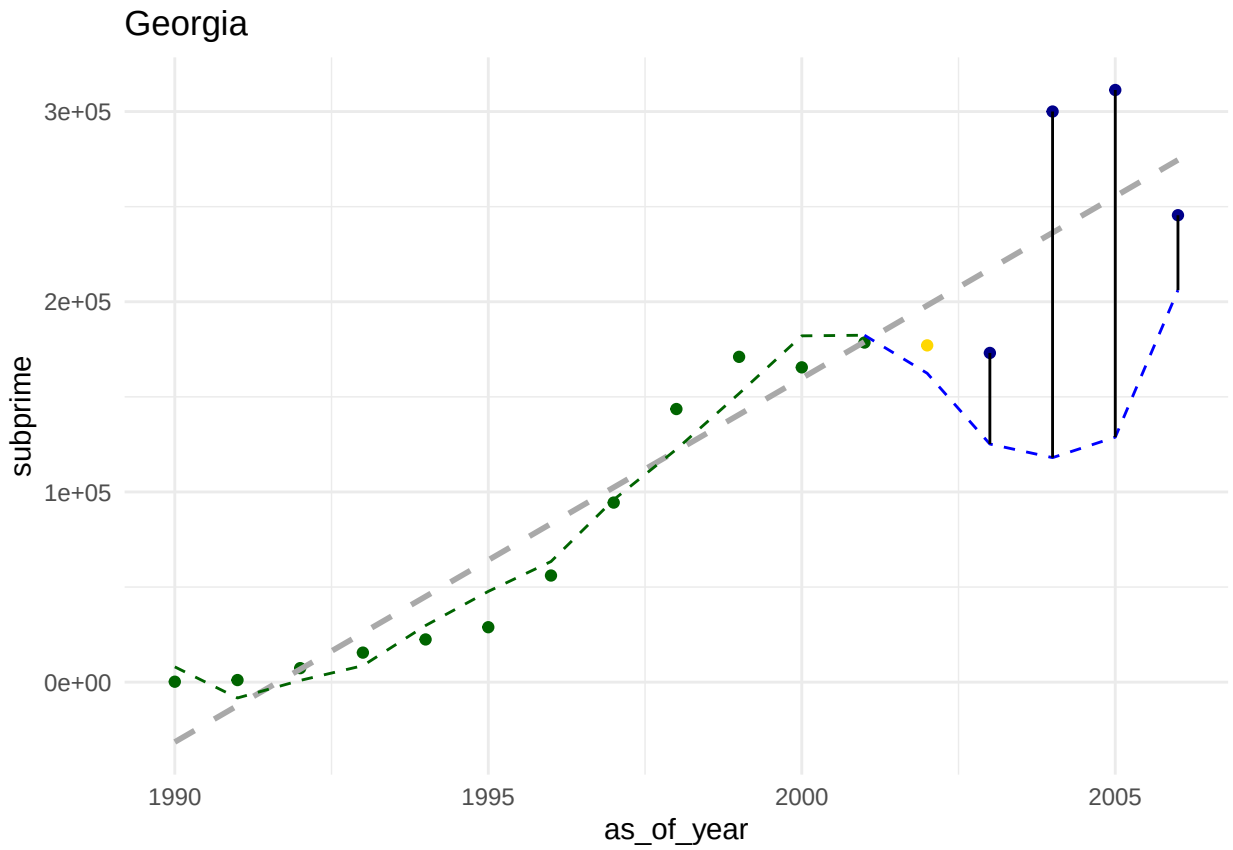
```
## as_of_year       -1.492e+05  6.716e+04  -2.221   0.0618 .
## total_population  9.437e-01  4.218e-01   2.237   0.0603 .
## sthpi             4.354e+02  7.999e+02   0.544   0.6031
## mh_income         1.036e+00  5.384e+00   0.193   0.8528
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16100 on 7 degrees of freedom
## Multiple R-squared:  0.9685, Adjusted R-squared:  0.9505
## F-statistic: 53.84 on 4 and 7 DF,  p-value: 2.429e-05
```

```
predicted2 <- data.frame(st_data$as_of_year, predict(prelaw_model2, newdata=st_data))
names(predicted2) <- c('as_of_year', 'subprime')

predicted2_pre <- predicted2[predicted2$as_of_year <= 2001,]
predicted2_pos <- predicted2[predicted2$as_of_year >= 2001,]

differences <- postlaw[, c('as_of_year', 'subprime')]
differences$subprime_predicted <- predicted2[predicted2$as_of_year > 2002,]$subprime
differences$difference <- differences$subprime - differences$subprime_predicted

ggplot(NULL, aes(x=as_of_year, y=subprime)) + geom_point(data=prelaw, color='darkgreen') +
  geom_point(data=transition, color='gold') +
  geom_point(data=postlaw, color='darkblue') +
  geom_smooth(data=prelaw, method=lm, se=FALSE, color='darkgray', fullrange=TRUE, linetype='dashed') +
  geom_line(data=predicted2_pre, color='darkgreen', linetype='dashed') +
  geom_line(data=predicted2_pos, color='blue', linetype='dashed') +
  #geom_text(data=prelaw, x=1995, y=150000, label=lm_eqn(prelaw_model), parse=TRUE) +
  geom_segment(data=differences,
               aes(x=as_of_year, xend=as_of_year, y=subprime, yend=subprime_predicted)) +
  theme_minimal() + ggtitle('Georgia')
```




```
#Minnesota
st_data <- aux[aux$state == 'MN',]

prelaw <- st_data[st_data$as_of_year <= 2003,]
transition <- st_data[(st_data$as_of_year > 2003) & (st_data$as_of_year <= 2004),]
postlaw <- st_data[st_data$as_of_year > 2004,]

prelaw_model <- lm(formula=(subprime ~ as_of_year), data=prelaw)
summary(prelaw_model)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year), data = prelaw)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -21400 -11178  -3304  12222  24621
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -21365648    2082199  -10.26 2.71e-07 ***
## as_of_year      10724       1043   10.28 2.65e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15730 on 12 degrees of freedom
## Multiple R-squared:  0.8981, Adjusted R-squared:  0.8896
```

```
## F-statistic: 105.7 on 1 and 12 DF,  p-value: 2.647e-07
```

```
postlaw_model <- lm(formula=(subprime ~ as_of_year), data=postlaw)
summary(postlaw_model)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year), data = postlaw)
##
## Residuals:
## ALL 2 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 92735034         NA      NA       NA
## as_of_year    -46176         NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 1 and 0 DF,  p-value: NA
```

```
predicted <- data.frame(st_data$as_of_year, predict(prelaw_model, newdata=st_data))
names(predicted) <- c('as_of_year', 'subprime')

prelaw_model2 <- lm(formula=(subprime ~ as_of_year + total_population + sthpi + mh_income), data=prelaw
summary(prelaw_model2)
```

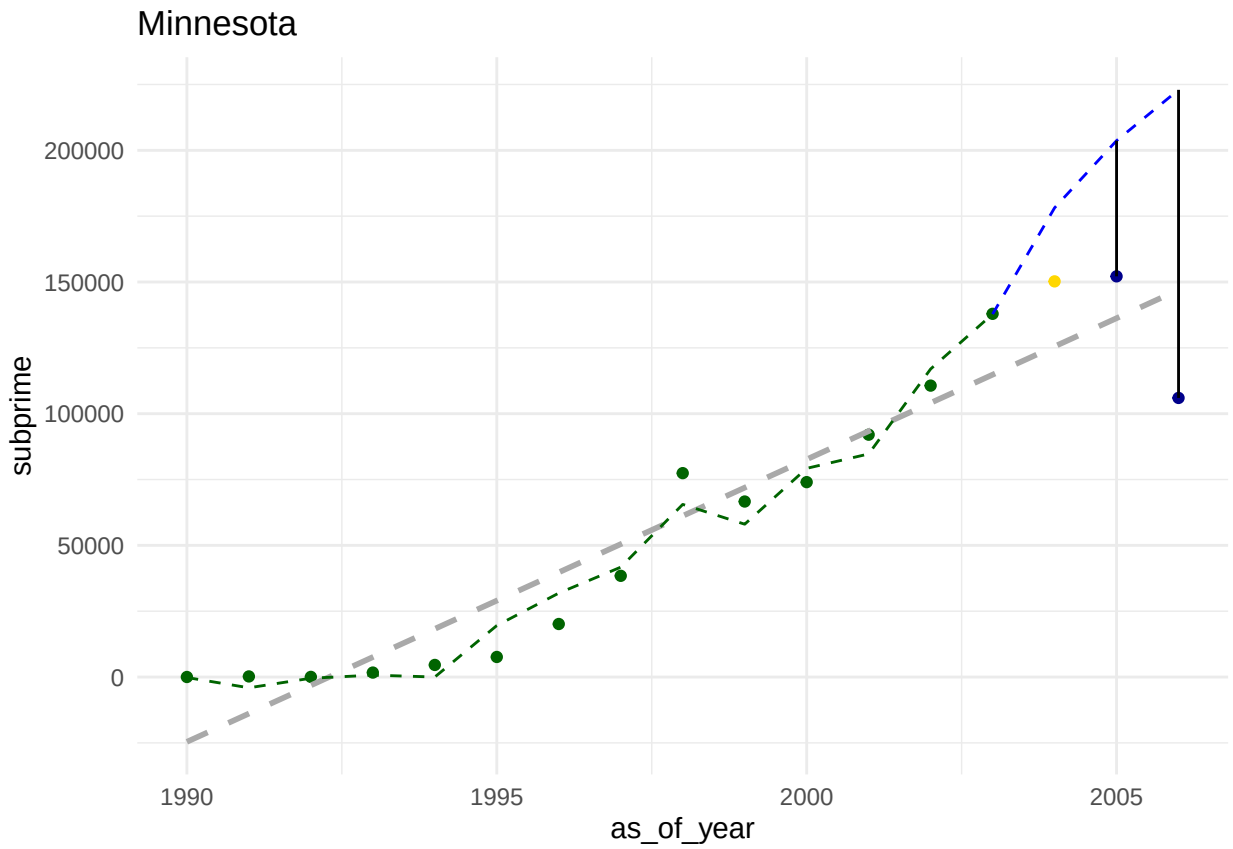
```
##
## Call:
## lm(formula = (subprime ~ as_of_year + total_population + sthpi +
##     mh_income), data = prelaw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11927.1  -4739.7    352.4   4506.8  11876.8
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.179e+08  3.954e+07  -2.982   0.0154 *
## as_of_year        6.178e+04  2.071e+04   2.983   0.0154 *
## total_population -1.180e+00  3.825e-01  -3.084   0.0131 *
## sthpi             2.842e+02  2.071e+02   1.373   0.2031
## mh_income         3.347e+00  1.211e+00   2.764   0.0220 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8610 on 9 degrees of freedom
## Multiple R-squared:  0.9771, Adjusted R-squared:  0.9669
## F-statistic: 95.99 on 4 and 9 DF,  p-value: 2.247e-07
```

```
predicted2 <- data.frame(st_data$as_of_year, predict(prelaw_model2, newdata=st_data))
names(predicted2) <- c('as_of_year', 'subprime')

predicted2_pre <- predicted2[predicted2$as_of_year <= 2003,]
predicted2_pos <- predicted2[predicted2$as_of_year >= 2003,]
```

```
differences <- postlaw[, c('as_of_year', 'subprime')]
differences$subprime_predicted <- predicted2[predicted2$as_of_year > 2004,]$subprime
differences$difference <- differences$subprime - differences$subprime_predicted

ggplot(NULL, aes(x=as_of_year, y=subprime)) + geom_point(data=prelaw, color='darkgreen') +
  geom_point(data=transition, color='gold') +
  geom_point(data=postlaw, color='darkblue') +
  geom_smooth(data=prelaw, method=lm, se=FALSE, color='darkgray', fullrange=TRUE, linetype='dashed') +
  geom_line(data=predicted2_pre, color='darkgreen', linetype='dashed') +
  geom_line(data=predicted2_pos, color='blue', linetype='dashed') +
  #geom_text(data=prelaw, x=1995, y=150000, label=lm_eqn(prelaw_model), parse=TRUE) +
  geom_segment(data=differences,
               aes(x=as_of_year, xend=as_of_year, y=subprime, yend=subprime_predicted)) +
  theme_minimal() + ggtitle('Minnesota')
```




```
#South Carolina
st_data <- aux[aux$state == 'SC',]

prelaw <- st_data[st_data$as_of_year <= 1999,]
transition <- st_data[(st_data$as_of_year > 1999) & (st_data$as_of_year <= 2000),]
postlaw <- st_data[st_data$as_of_year > 2000,]

prelaw_model <- lm(formula=(subprime ~ as_of_year), data=prelaw)
summary(prelaw_model)
```

```
##
## Call:
```

```
## lm(formula = (subprime ~ as_of_year), data = prelaw)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -24319 -16945  -3987  20122  24363
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -25762153    4464682  -5.770 0.000419 ***
## as_of_year      12934       2238   5.778 0.000416 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20330 on 8 degrees of freedom
## Multiple R-squared:  0.8067, Adjusted R-squared:  0.7825
## F-statistic: 33.38 on 1 and 8 DF,  p-value: 0.0004156
```

```
postlaw_model <- lm(formula=(subprime ~ as_of_year), data=postlaw)
summary(postlaw_model)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year), data = postlaw)
##
## Residuals:
##     71     77     83     89     95    101
##   1872  -8003  -2912  11370   8651 -10978
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4392190    4793372  -0.916    0.411
## as_of_year      2240       2392   0.936    0.402
##
## Residual standard error: 10010 on 4 degrees of freedom
## Multiple R-squared:  0.1798, Adjusted R-squared:  -0.02529
## F-statistic: 0.8766 on 1 and 4 DF,  p-value: 0.4021
```

```
predicted <- data.frame(st_data$as_of_year, predict(prelaw_model, newdata=st_data))
names(predicted) <- c('as_of_year', 'subprime')

prelaw_model2 <- lm(formula=(subprime ~ as_of_year + total_population + sthpi + mh_income), data=prelaw
summary(prelaw_model2)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year + total_population + sthpi +
##     mh_income), data = prelaw)
##
## Residuals:
##        5       11       17       23       29       35       41       47
##  1742.89  2118.67 -2105.49 -3521.15    94.53 -2226.30 -1552.97  2554.54
##       53       59
## 11212.37 -8317.09
##
## Coefficients:
```

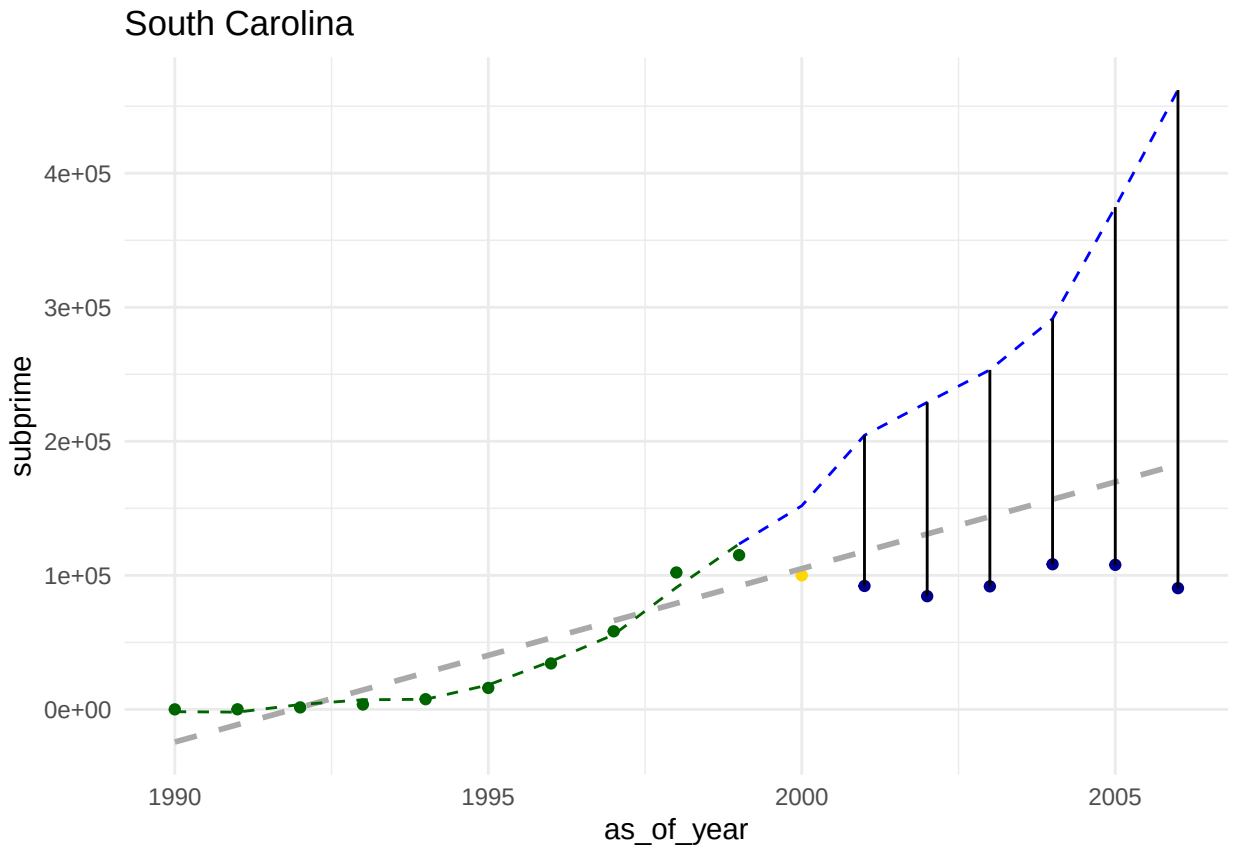
```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.801e+07  2.943e+07   0.952  0.38482
## as_of_year       -1.424e+04  1.540e+04  -0.925  0.39757
## total_population -1.305e-01  3.839e-01  -0.340  0.74775
## sthpi             4.970e+03  1.028e+03   4.834  0.00474 **
## mh_income        -5.462e-01  1.718e+00  -0.318  0.76342
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6829 on 5 degrees of freedom
## Multiple R-squared:  0.9864, Adjusted R-squared:  0.9755
## F-statistic: 90.47 on 4 and 5 DF,  p-value: 7.515e-05
```

```
predicted2 <- data.frame(st_data$as_of_year, predict(prelaw_model2, newdata=st_data))
names(predicted2) <- c('as_of_year', 'subprime')

predicted2_pre <- predicted2[predicted2$as_of_year <= 1999,]
predicted2_pos <- predicted2[predicted2$as_of_year >= 1999,]

differences <- postlaw[, c('as_of_year', 'subprime')]
differences$subprime_predicted <- predicted2[predicted2$as_of_year > 2000,]$subprime
differences$difference <- differences$subprime - differences$subprime_predicted

ggplot(NULL, aes(x=as_of_year, y=subprime)) + geom_point(data=prelaw, color='darkgreen') +
  geom_point(data=transition, color='gold') +
  geom_point(data=postlaw, color='darkblue') +
  geom_smooth(data=prelaw, method=lm, se=FALSE, color='darkgray', fullrange=TRUE, linetype='dashed') +
  geom_line(data=predicted2_pre, color='darkgreen', linetype='dashed') +
  geom_line(data=predicted2_pos, color='blue', linetype='dashed') +
  #geom_text(data=prelaw, x=1995, y=150000, label=lm_eqn(prelaw_model), parse=TRUE) +
  geom_segment(data=differences,
               aes(x=as_of_year, xend=as_of_year, y=subprime, yend=subprime_predicted)) +
  theme_minimal() + ggtitle('South Carolina')
```




```
#Virginia
st_data <- aux[aux$state == 'VA',]

prelaw <- st_data[st_data$as_of_year <= 2006,]
#transition <- st_data[(st_data$as_of_year > 1999) & (st_data$as_of_year <= 2000),]
#postlaw <- st_data[st_data$as_of_year > 2000,]

prelaw_model <- lm(formula=(subprime ~ as_of_year), data=prelaw)
summary(prelaw_model)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year), data = prelaw)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -28843 -16186  -1086   6864  55370
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -30572931    2270208  -13.47 8.80e-10 ***
## as_of_year      15346       1136   13.51 8.45e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22950 on 15 degrees of freedom
## Multiple R-squared:  0.924,  Adjusted R-squared:  0.919
```

```
## F-statistic: 182.4 on 1 and 15 DF,  p-value: 8.455e-10
```

```
postlaw_model <- lm(formula=(subprime ~ as_of_year), data=postlaw)
summary(postlaw_model)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year), data = postlaw)
##
## Residuals:
##     71     77     83     89     95    101
##   1872  -8003  -2912  11370   8651 -10978
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4392190    4793372  -0.916    0.411
## as_of_year      2240       2392   0.936    0.402
##
## Residual standard error: 10010 on 4 degrees of freedom
## Multiple R-squared:  0.1798, Adjusted R-squared:  -0.02529
## F-statistic: 0.8766 on 1 and 4 DF,  p-value: 0.4021
```

```
predicted <- data.frame(st_data$as_of_year, predict(prelaw_model, newdata=st_data))
names(predicted) <- c('as_of_year', 'subprime')

prelaw_model2 <- lm(formula=(subprime ~ as_of_year + total_population + sthpi + mh_income), data=prelaw
summary(prelaw_model2)
```

```
##
## Call:
## lm(formula = (subprime ~ as_of_year + total_population + sthpi +
##     mh_income), data = prelaw)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -39427  -6049    790   5842  57871
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.100e+08  8.923e+07  -1.233    0.241
## as_of_year        5.689e+04  4.649e+04   1.224    0.245
## total_population -5.317e-01  5.268e-01  -1.009    0.333
## sthpi             2.751e+02  2.103e+02   1.308    0.215
## mh_income         1.497e+00  3.196e+00   0.468    0.648
##
## Residual standard error: 23230 on 12 degrees of freedom
## Multiple R-squared:  0.9377, Adjusted R-squared:  0.917
## F-statistic: 45.18 on 4 and 12 DF,  p-value: 3.864e-07
```

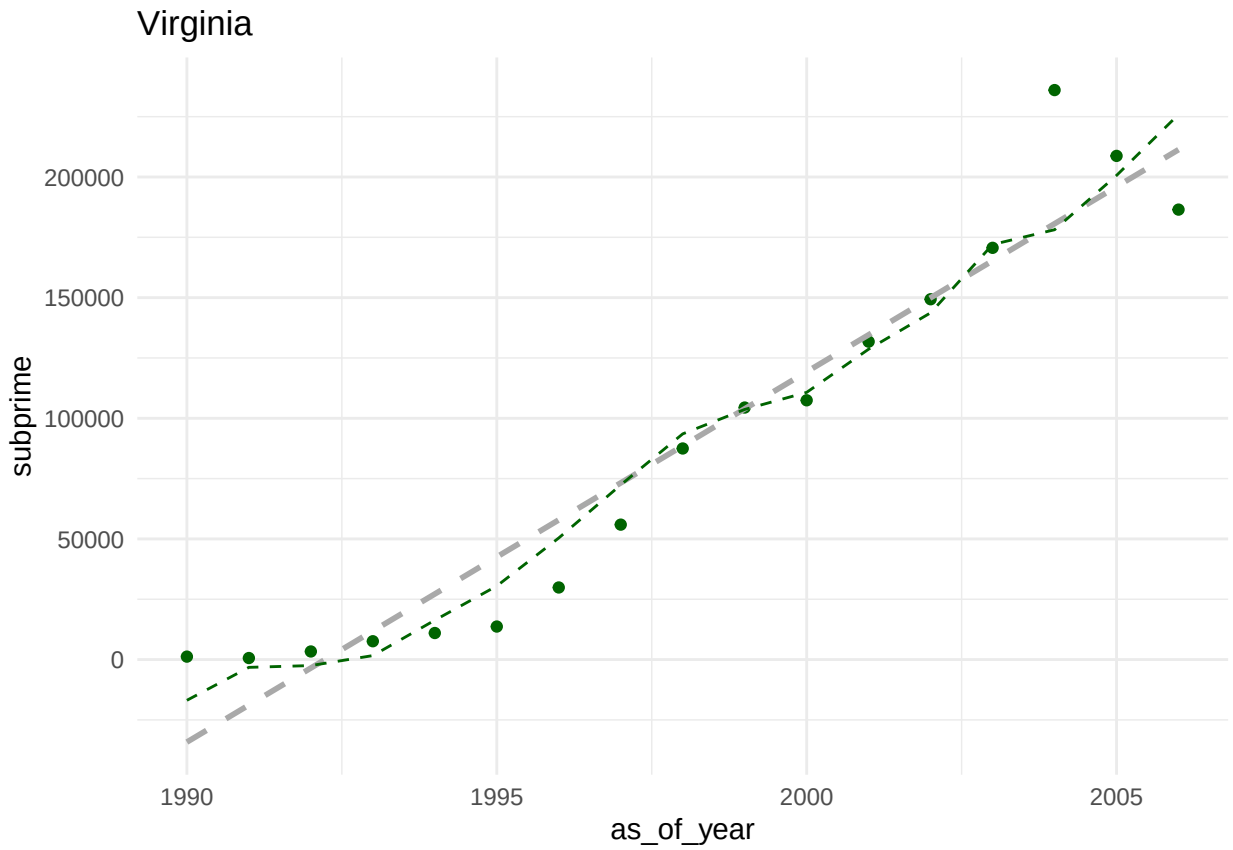
```
predicted2 <- data.frame(st_data$as_of_year, predict(prelaw_model2, newdata=st_data))
names(predicted2) <- c('as_of_year', 'subprime')

predicted2_pre <- predicted2[predicted2$as_of_year <= 2006,]
predicted2_pos <- predicted2[predicted2$as_of_year >= 2006,]

differences <- postlaw[, c('as_of_year', 'subprime')]
```

```
differences$subprime_predicted <- predicted2[predicted2$as_of_year >= 2006,]$subprime
differences$difference <- differences$subprime - differences$subprime_predicted

ggplot(NULL, aes(x=as_of_year, y=subprime)) + geom_point(data=prelaw, color='darkgreen') +
  #geom_point(data=transition, color='gold') +
  #geom_point(data=postlaw, color='darkblue') +
  geom_smooth(data=prelaw, method=lm, se=FALSE, color='darkgray', fullrange=TRUE, linetype='dashed') +
  geom_line(data=predicted2_pre, color='darkgreen', linetype='dashed') +
  #geom_line(data=predicted2_pos, color='blue', linetype='dashed') +
  #geom_text(data=prelaw, x=1995, y=150000, label=lm_eqn(prelaw_model), parse=TRUE) +
  #geom_segment(data=differences,
               #aes(x=as_of_year, xend=as_of_year, y=subprime, yend=subprime_predicted)) +
  theme_minimal() + ggtitle('Virginia')
```



Difference in Differences Model

```
state_model <- lm(formula=(subprime_amount_originated ~ as.factor(as_of_year) + as.factor(state_code) +
summary(state_model)
```

```
##
## Call:
## lm(formula = (subprime_amount_originated ~ as.factor(as_of_year) +
##     as.factor(state_code) + did + total_population + sthpi +
##     mh_income), data = state_data)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -1580641  -327079      587   330231  1741770
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -1.701e+07  1.372e+06 -12.397  < 2e-16 ***
## as.factor(as_of_year)1991 -3.759e+05  3.994e+05  -0.941 0.349550
## as.factor(as_of_year)1992 -6.631e+05  4.034e+05  -1.644 0.104328
## as.factor(as_of_year)1993 -9.550e+05  4.101e+05  -2.329 0.022538 *
## as.factor(as_of_year)1994 -1.418e+06  4.233e+05  -3.351 0.001258 **
## as.factor(as_of_year)1995 -1.929e+06  4.385e+05  -4.399 3.49e-05 ***
## as.factor(as_of_year)1996 -2.212e+06  4.807e+05  -4.601 1.65e-05 ***
## as.factor(as_of_year)1997 -2.125e+06  5.190e+05  -4.094 0.000105 ***
## as.factor(as_of_year)1998 -2.435e+06  5.623e+05  -4.331 4.48e-05 ***
## as.factor(as_of_year)1999 -2.541e+06  6.024e+05  -4.218 6.75e-05 ***
## as.factor(as_of_year)2000 -3.505e+06  6.580e+05  -5.326 9.86e-07 ***
## as.factor(as_of_year)2001 -4.197e+06  6.833e+05  -6.143 3.44e-08 ***
## as.factor(as_of_year)2002 -4.283e+06  7.122e+05  -6.013 5.91e-08 ***
## as.factor(as_of_year)2003 -3.316e+06  7.473e+05  -4.438 3.03e-05 ***
## as.factor(as_of_year)2004 -2.921e+06  7.905e+05  -3.696 0.000411 ***
## as.factor(as_of_year)2005 -3.973e+06  8.714e+05  -4.559 1.93e-05 ***
## as.factor(as_of_year)2006 -6.870e+06  9.500e+05  -7.232 3.20e-10 ***
## as.factor(state_code)13   -8.015e+06  1.102e+06  -7.270 2.71e-10 ***
## as.factor(state_code)27   -5.691e+05  5.070e+05  -1.122 0.265185
## as.factor(state_code)37   -8.613e+06  1.022e+06  -8.431 1.65e-12 ***
## as.factor(state_code)45    7.647e+05  2.888e+05   2.647 0.009855 **
## as.factor(state_code)51   -6.765e+06  9.014e+05  -7.505 9.69e-11 ***
## did                       -9.661e+05  4.232e+05  -2.283 0.025225 *
## total_population           2.832e+00  3.093e-01   9.158 6.65e-14 ***
## sthpi                      4.644e+04  4.593e+03  10.110 1.03e-15 ***
## mh_income                 -4.612e+01  3.809e+01  -1.211 0.229752
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 689500 on 76 degrees of freedom
## Multiple R-squared:  0.9447, Adjusted R-squared:  0.9265
## F-statistic: 51.91 on 25 and 76 DF,  p-value: < 2.2e-16
```

##Conclusions