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

Sebastián Soriano Pérez

3/8/2021

Abstract

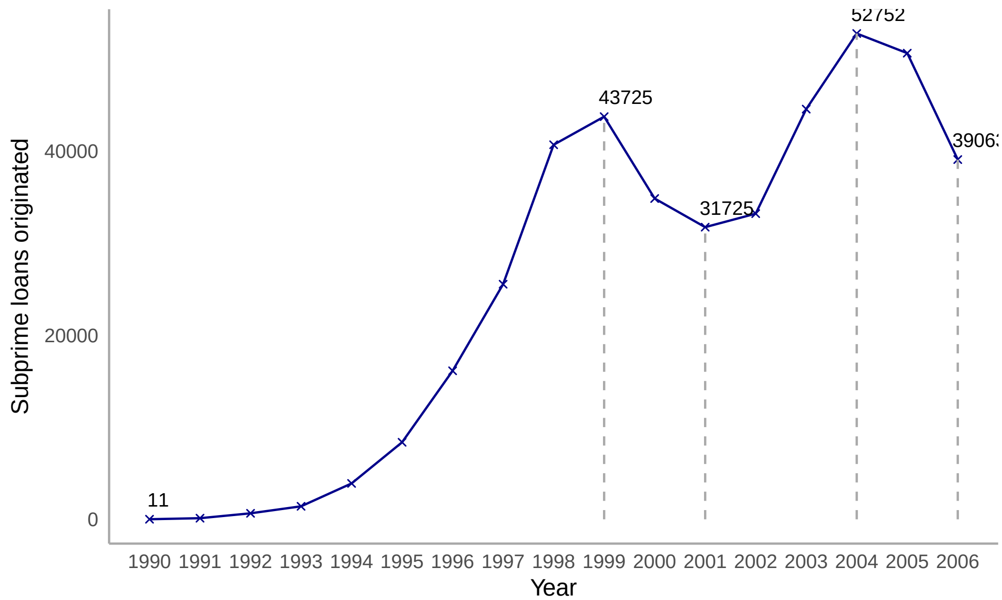
This report quantifies the impact of North Carolina's predatory lending law from 1999 in the state's subprime mortgage market. We used the Home Mortgage and Disclosure Act (HMDA) Loan Application Register (LAR) datasets in conjunction with the US Department of Housing and Urban Development (HUD)'s classification 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 mortgage market before the law's implementation (from 1993 to 1998), and after the intervention (from 1999 to 2006). We executed an Interrupted Time Series Analysis and a Differences in Differences model to compare North Carolina's subprime mortgage before and after the law was implemented, and we compared it to other states with similar socioeconomic characteristics. Our analysis shows the 1999 predatory lending law was effective as there was a significant decrease in the number of subprime loans originated in the state after the law was implemented, most likely due to a reduction in the number of predatory loans issued in the state.

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.

Figure 1

Subprime loans originated in North Carolina (1990−2006)



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 (See Figure 1). 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 in 2002 (Georgia Fair Lending Act), and South Carolina in 2000 (Predatory Mortgage Lending Prevention Act). Other states have passed minimal predatory lending laws, such as Alabama in 2003 (Payday Lending Law), Minnesota in 2004 (Prohibition of “equity stripping”), and Virginia in 2020 (Virginia Fairness in Lending Act).

Data

We used the HMDA LAR datasets from 1993 to 2006, to cover a 16 year period around 1999 when the predatory lending law went into effect in North Carolina. Each row in the HMDA LAR datasets contains the following information about a loan application: year, state, county, lender, loan type, loan purpose, loan amount, action taken (whether it was denied, originated, etc.), applicant race, applicant income, and more. As a first step, we classified each loan as subprime or non-subprime by merging the HMDA LAR data with the HUD lenders annual classification. 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…]. Additionally, we collected data from the Federal Reserve Economic Data (FRED) to append the state’s house price index (HPI), and median household for the same time period. Finally, using annual census estimates published by the National Cancer Institute, we included the population at a county level in our dataset.

For our dataset, we aggregated the individual loan applications data to a county level so its statistical analysis would be possible. The final dataset contains panel data with the following variables:

- `year` (numerical): Year when the loan application was created
- `elapsed_time` (numerical): Years elapsed since 1993
- `state`(categorical): State where the loan application was received
- `county_code`(numerical): County within the state where the loan application was received
- `subprime`(numerical): Number of subprime loan applications made in the county

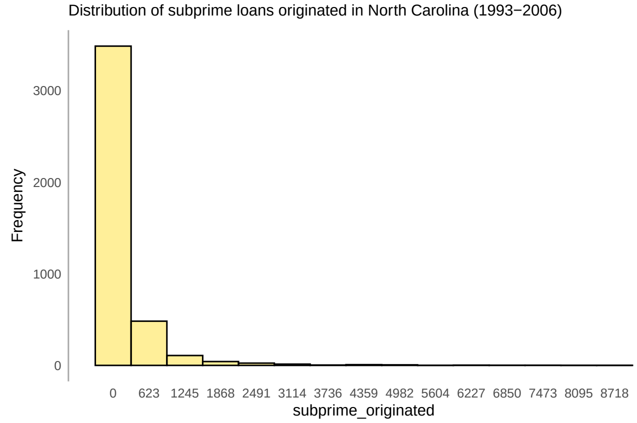
- `subprime_originated`(numerical): Number of subprime loans originated in the county
- `subprime_amount`(numerical): Total loan amount for subprime loans originated in the county
- `prime` (numerical): Number of non-subprime loan applications made in the county
- `prime_originated` (numerical): Number of non-subprime loans originated in the county
- `prime_amount` (numerical): Total loan amount for non-subprime loans originated in the county
- `loans_originated` (numerical): Number of loans (subprime and non-subprime) originated in the county
- `population` (numerical): County population
- `mh_income` (numerical): Median household income (state-level)
- `sthpi` (numerical): House Price Index (state-level)

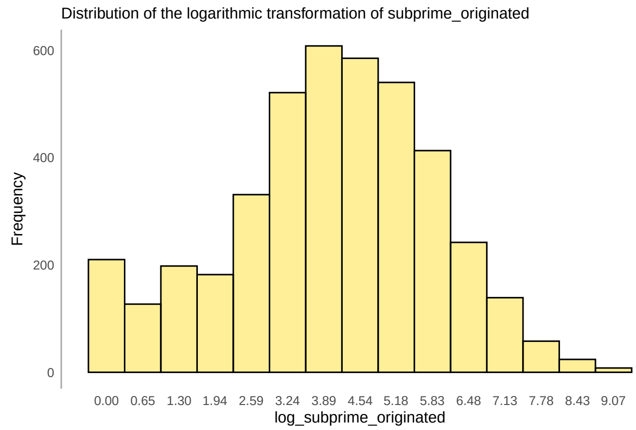
The potential response or target variables to measure any potential changes in the subprime mortgage market are `subprime`, `subprime_originated`, and `subprime_amount`. Additionally, one could create other response variables such as the proportion of all loans that are subprime (i.e.: $\text{subprime} \div (\text{subprime} + \text{prime})$). We decided to use the number of subprime loans originated each year at a county level `subprime_originated` as the response variable for our analysis as it is the most direct indicator of how the subprime mortgage market changes each year. We can then compare our results to a similar analysis on the non-subprime market `prime_originated`, to compare both sectors of the market.

The rows that contained zero values for the variable `loans_originated` were removed from our analysis because they contained no information on the size of the subprime mortgage market for that county and year, as no loans (subprime or not) were originated then. The final dataset contains $4,186$ rows ($1,400$ for North Carolina, 938 for Alabama, and 1848 for Virginia). In order to perform an interrupted time series analysis using a normal OSL regression, we observed the distribution of the response variable `subprime_originated` and performed a logarithmic transformation on it (see Figure 2) to create a new response variable that is normally distributed `log_subprime_originated` ($\text{log\_subprime\_originated} = \log(\text{subprime\_originated} + 1)$). A summary of the data variables for each state being analyzed can be found on Appendix 1.1.

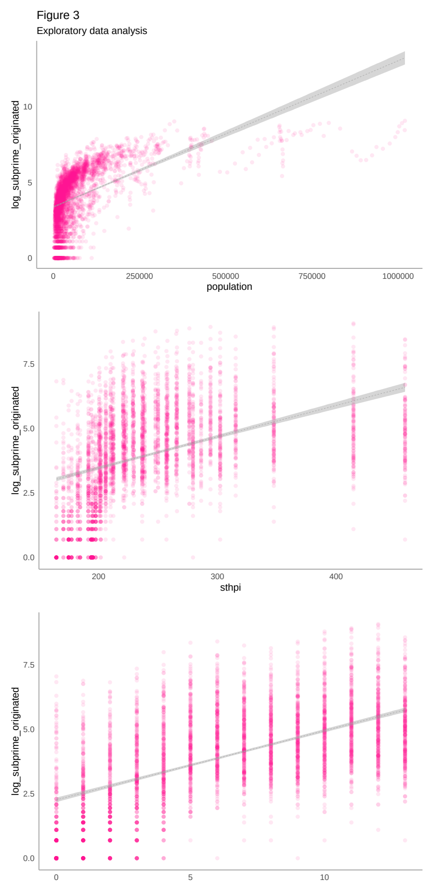
Figure 2

Distribution of subprime loans originated in North Carolina (1993−2006)





As part of an exploratory data analysis, we plotted all potential predictor variables versus the response variable `log_subprime_originated` to determine if they appear to have an effect on the response variable and therefore could be included in our analysis (See Figure 3). It appears that all numerical variables being considered have a positive impact on the response variable `log_subprime_originated`. We included these variables in our analysis.



Interrupted Time Series Analysis

To quantify the impact of the 1999 North Carolina predatory lending law on the amount of subprime loans being originated in the state, we used a time series model with fixed effects for our panel data, using only the data from North Carolina. For this purpose, we created a dummy variable `intervention` with a value of 0 for all observations from 1993 to 1998, the period before the predatory lending law was passed, and a value of

1 for the observations from 1999 to 2006, indicating when the law was in effect. For simplicity purposes, we assume the effects of the law can be observed starting on the year of its implementation, even though its policies were implemented in two stages in 1999 and 2000. To execute the interrupted time series analysis, we created an interaction variable `intervention_time` that indicate the number of years the intervention has been in effect (intervention_time = intervention $\times$ (elapsed_time $-$ 5), where 5 is the number of years elapsed since 1993 whem the law was implemented).

We used a fixed effects model with fixed effects for counties to measure how the 1999 law affected the number of subprime loans originated in each county, assuming each of them has very different time invariant county-level effects or baseline levels. For our panel data, the addition of county-level fixed effects allows us to control for any constant effects within the same county and focus on changes for each county over time. This is reasonable to assume since counties vary a lot in terms of population size, race, median household income, etc. Larger counties should have larger fixed-effects, as their mortgage markets should be larger too. The final model has the following formula:

$$\begin{aligned}\widehat{\text{log\_subprime\_originated}}_{i,t} = \alpha_i &+ \beta_1 \text{elapsed\_time}_{i,t} + \beta_2 \text{intervention}_{i,t} + \beta_3 \text{intervention\_time}_{i,t} \\ &+ \beta_4 \text{population}_{i,t} + \beta_5 \text{sthpi}_{i,t} + \epsilon_{i,t}\end{aligned}$$

Where $i \in \{1, 3, 5, ..., 199\}$ corresponds to each one of the 100 counties in North Carolina, $t \in \{1993, 1994, ..., 2006\}$ corresponds to each one of the yearly observations in the dataset, $\widehat{\text{log\_subprime\_originated}}_{i,t}$ is the response variable for county i at time t, α_i is time invariant county-level effect, $\beta_1, ..., \beta_5$ are the coefficients of the predictors or explanatory variables, $\text{elapsed\_time}_{i,t}, ..., \text{sthpi}_{i,t}$ are the predictor variables observed for each county i at time t, and $\epsilon_{i,t}$ is the error term for county i at time t.

The final model coefficients $\beta_1, ..., \beta_4$ have the following values:

Table 1: Model's Coefficients

Coefficient	Estimate	Cluster s.e.	t value	$\Pr(T > \lvert t \rvert)$
β_1	0.8909103	0.0265604	33.542844	0.0000000
β_2	-0.3949739	0.0351692	-11.230672	0.0000000
β_3	-0.7115399	0.0212101	-33.547178	0.0000000
β_4	-0.0000014	0.0000006	-2.202548	0.0299469
β_5	-0.0140619	0.0016855	-8.342998	0.0000000

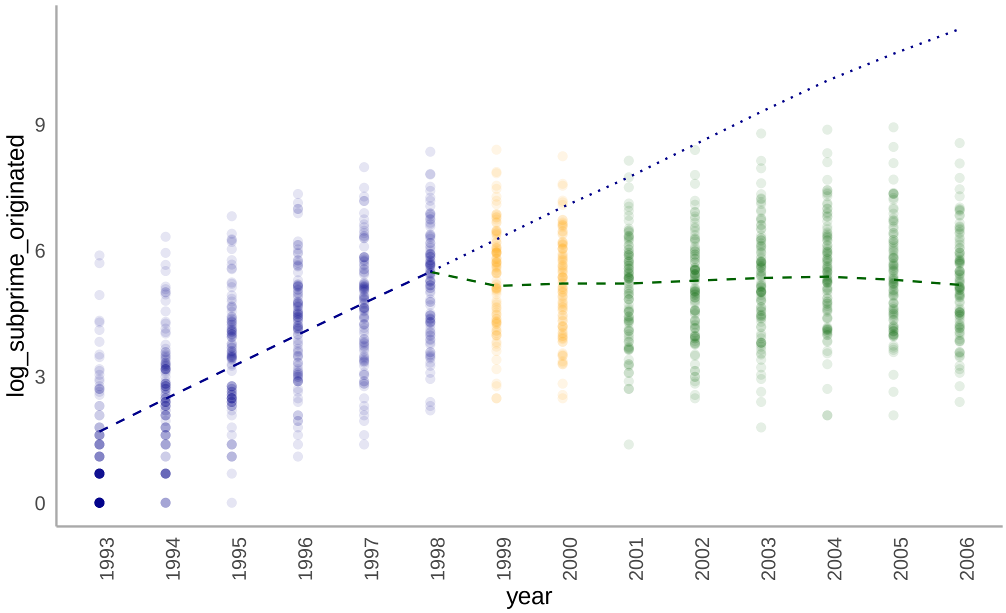
[I should interpret all this in terms of number of loans i.e.: undo the log transformation]

All of the coefficients in the model are statistically significant. These results indicate that before the intervention was implemented, the response variable was increasing at a rate of 0.8909 per year, while it dropped to $0.8909 - 0.7115 = 0.1794$ per year after 1999. Because of the intervention, there was a level drop of 0.3950 in the response variable for the intercepts after 1999. Additionally, each unit increase in population (i.e. for each additional person living in a county) there is a small decrease in the response variable of 1.4×10^{-6}, and for every unit increase in the state's house price index, there's a decrease of 0.0141 in the response variable. This implies that for larger counties and counties where the houses have higher values, the response variable decreases and thus the number of subprime loans originated is smaller.

For the full list of the model's fixed effects see Appendix 1.4. The average county intercept $\bar{\alpha} = \left(\sum_{t \in \{1993,1994,...,2006\}} \sum_{i \in \{1,3,...,199\}} \alpha_i\right) \div (T \times N) = 4.2465$, where $T = 14$ and $N = 100$. This means that for the average county in North Carolina (where each county's weight is proportional only to the number of observations it has in the dataset), the response variable was 4.2465 in 1993, with a population and house price index of 0. See Figure 4 for the model's predicted trends for the average county in North Carolina. The blue dashed line represents the trend pre-intervention and the green dashed line represents the trend post intervention. The blue dotted line is obtained using the model's predicitons for the counterfacutals (using the same data except for the `intervention` and `intervention_time` variables which are set to zero), which represent what would have happened had the 1999 law never been implemented. It is evident that the linear trends decreased significantly after the law was implemented.

Figure 4

Response variable estimates for average county & observed values



Appendix

##1.1 State Selection For our difference in differences analysis we selected five states that were socioeconomically similar to North Carolina. Using the census 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 Alabama, Georgia, Minnesota, South Carolina, and Virginia. We selected those states that passed no anti-predatory lending laws from 1990 to 2006, so they could be used as a control group for North Carolina: Virginia and Alabama.

##1.2 Dataset Summary *North Carolina*

```
## [1] "NC_data"
```

```
##       year         state            county_code     elapsed_time
##  Min.   :1993   Length:1400        Min.   :  1.0   Min.   : 0.0
##  1st Qu.:1996   Class :character   1st Qu.: 50.5   1st Qu.: 3.0
##  Median :2000   Mode  :character   Median :100.0   Median : 6.5
##  Mean   :2000                      Mean   :100.0   Mean   : 6.5
##  3rd Qu.:2003                      3rd Qu.:149.5   3rd Qu.:10.0
##  Max.   :2006                      Max.   :199.0   Max.   :13.0
##   intervention    intervention_time   population         sthpi
##  Min.   :0.0000   Min.   :0.000     Min.   :  3775   Min.   :174.7
##  1st Qu.:0.0000   1st Qu.:0.000     1st Qu.: 23594   1st Qu.:200.4
##  Median :1.0000   Median :1.500     Median : 46376   Median :233.1
##  Mean   :0.5714   Mean   :2.571     Mean   : 79791   Mean   :236.1
##  3rd Qu.:1.0000   3rd Qu.:5.000     3rd Qu.: 92720   3rd Qu.:265.9
##  Max.   :1.0000   Max.   :8.000     Max.   :831507   Max.   :315.4
##  subprime_originated_lagged log_subprime_originated
##  Min.   :   0.0             Min.   :0.000
##  1st Qu.:   1.0             1st Qu.:3.555
##  Median :  24.0             Median :4.758
##  Mean   : 163.7             Mean   :4.552
##  3rd Qu.: 101.0             3rd Qu.:5.762
##  Max.   :6738.0             Max.   :8.932
```

Alabama

```
## [1] "AL_data"
```

Virginia

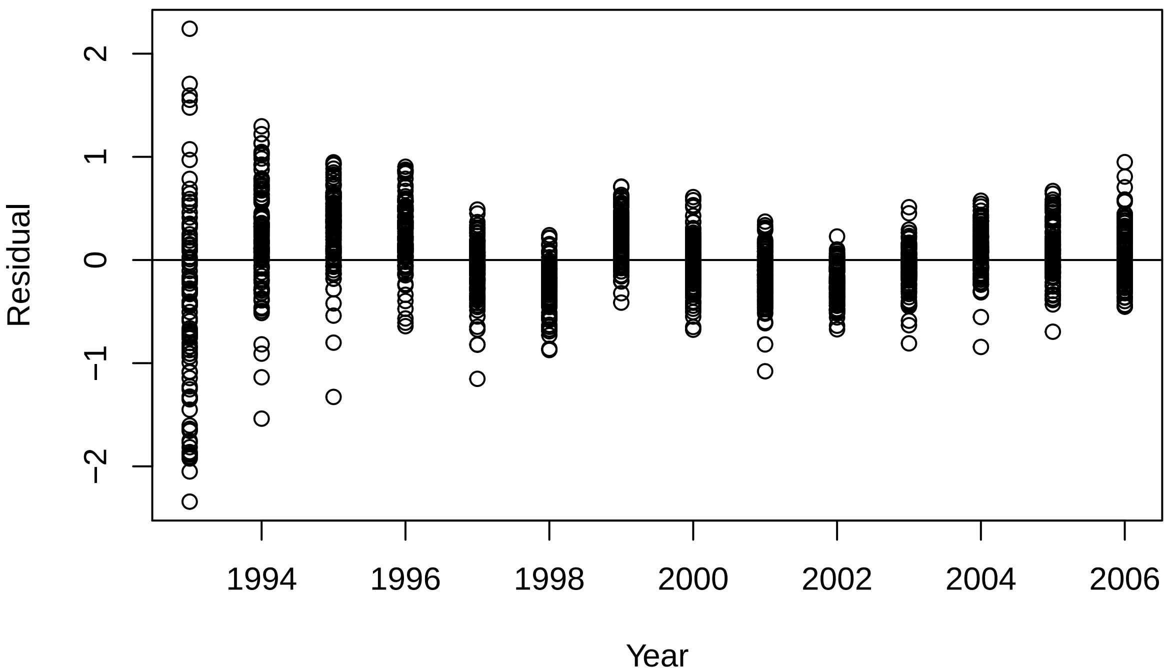
```
## [1] "VA_data"
```

##1.3. Exploratory Data Analysis

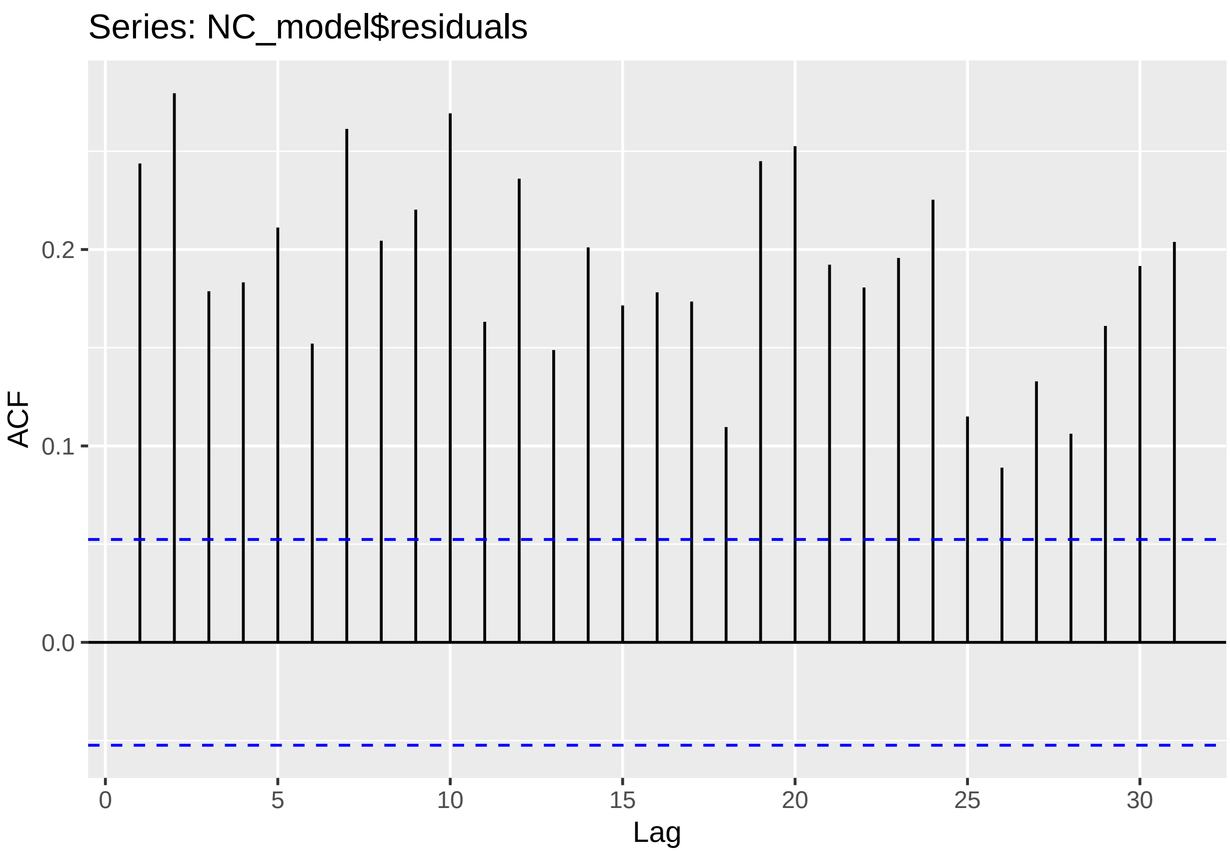
##1.4 Model's fixed effects by county

##1.5 Model assessment

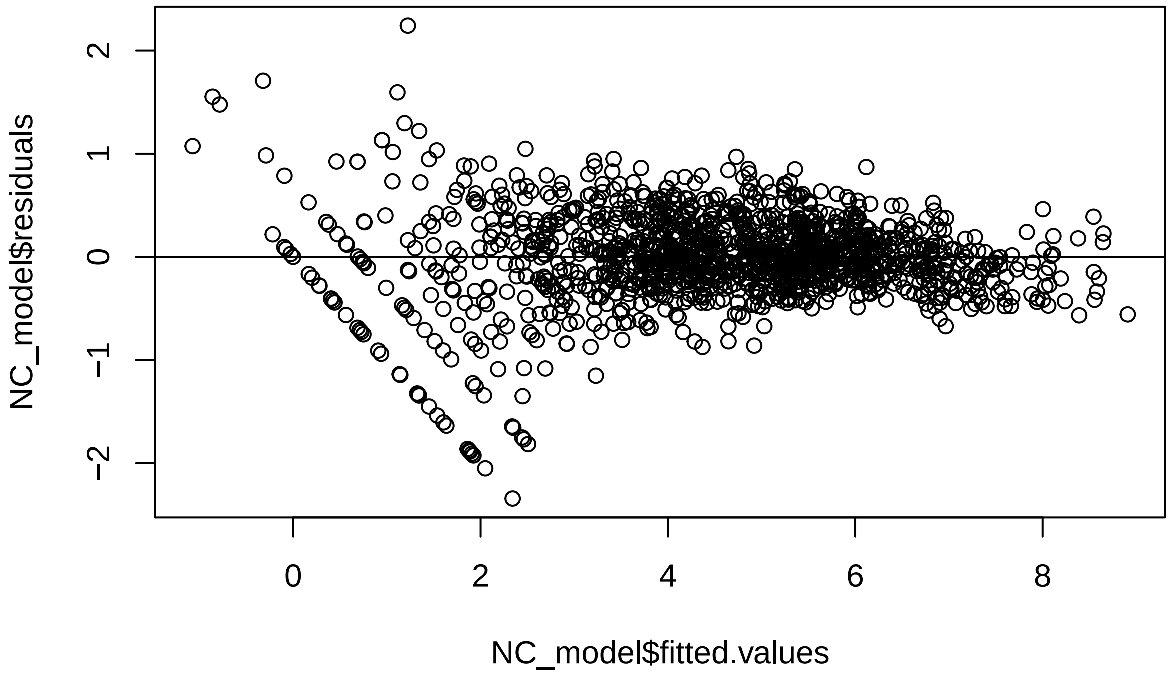
```r
#r.squaredGLMM(NC_model)
plot(y=NC_model$residuals, x=NC_data$year, xlab = "Year", ylab = "Residual")
abline(0,0)
```




```r
ggAcf(NC_model$residuals)
```



```r
plot(y=NC_model$residuals, x=NC_model$fitted.values)
abline(0,0)
```




```r
#plot(NC_model, which=1)
#plot(NC_model, which=2)
#durbinWatsonTest(NC_model)
```