Advance House Price Prediction USA

#Setting the Working Directory  
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))  
  
#clear environment & plot  
rm(list=ls())  
dev.off()

## null device   
## 1

options(scipen=999)  
  
#load relavent libraries  
library(pacman)

## Warning: package 'pacman' was built under R version 4.0.5

p\_load(rio, stargazer,Hmisc, PerformanceAnalytics, tidyverse)  
  
#import main data  
df <- import("housingData.xlsx")  
names(df) <- stringr::str\_to\_lower(names(df))  
str(df)

## 'data.frame': 510 obs. of 13 variables:  
## $ state : chr "Alabama" "Alaska" "Arizona" "Arkansas" ...  
## $ year : num 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...  
## $ avghomeprice : num 120294 253083 148090 105313 325668 ...  
## $ medianincome : num 49276 40933 57848 46896 38587 ...  
## $ minwage : num 7.25 7.75 7.25 7.25 8 7.24 8.25 7.25 8.25 7.25 ...  
## $ unemploymentrate : num 11.1 8.4 10.6 7.9 12.6 8.8 9.5 9.1 10.5 10.9 ...  
## $ mort30\_89 : num 5.1 2.3 3.9 4.3 3 2.7 3.5 4.1 3.1 3.9 ...  
## $ mort90plus : num 3.3 1.7 5.5 2.6 5.6 2.5 4.1 3.4 3.8 8.2 ...  
## $ subsidizedunitsavailable : num 94842 7757 39697 53128 485345 ...  
## $ pctsubsidizedunitsoccupied: num 88 98 93 92 96 96 93 91 85 92 ...  
## $ numhouseholdssubsidized : num 82471 7270 36808 46602 439799 ...  
## $ population : num 4785437 713910 6407172 2921964 37319502 ...  
## $ povertyrate : num 17.2 12.5 18.8 15.3 16.3 12.3 8.6 12.2 19.5 16 ...

#import gdp data  
gdp <- read.csv("current\_dollar\_gdp.csv", header = TRUE, sep = ",")  
colnames(gdp)[3:13] <- c("2009":"2019")  
gdp$`2009` <- NULL  
gdp %>% pivot\_longer(col = starts\_with("2"), names\_to = "Year", values\_to = "GDP" ) -> gdp  
str(gdp)

## tibble [510 x 4] (S3: tbl\_df/tbl/data.frame)  
## $ State : chr [1:510] "Alabama" "Alabama" "Alabama" "Alabama" ...  
## $ Region: chr [1:510] "Southeast" "Southeast" "Southeast" "Southeast" ...  
## $ Year : chr [1:510] "2010" "2011" "2012" "2013" ...  
## $ GDP : num [1:510] 175470 181930 186554 192167 195038 ...

names(gdp) <- stringr::str\_to\_lower(names(gdp))  
names <- c("state", "year", "region")   
gdp[,names] <- lapply(gdp[,names],factor)  
  
# merge temp & stores dataframes by store  
d <- merge(df, gdp, by=c("year", "state"))  
  
#categorize features,   
names <- c("state", "year")   
d[,names] <- lapply(d[,names],factor)  
colnames(d)[7:11] <- c("mort90plus\_delinquentrate","mort30to89\_delinquentrate", "subsd\_units\_avail",   
 "pctsubsd\_units\_occupied", "numhouse\_subsidized")  
  
#check for missing  
which(!complete.cases(d))

## integer(0)

colSums(is.na(d))

## year state avghomeprice   
## 0 0 0   
## medianincome minwage unemploymentrate   
## 0 0 0   
## mort90plus\_delinquentrate mort30to89\_delinquentrate subsd\_units\_avail   
## 0 0 0   
## pctsubsd\_units\_occupied numhouse\_subsidized population   
## 0 0 0   
## povertyrate region gdp   
## 0 0 0

#Data Visualizations  
hist(d$avghomeprice)  
hist(log(d$avghomeprice))  
  
library(lattice)  
histogram(~avghomeprice,data=d,  
 type="count",  
 xlab="Average Home Price",  
 main="Distribution of Avg Home Price")  
  
hist.data.frame(d)  
  
library(purrr)  
d %>% keep(is.numeric) %>% pairs(. , panel= panel.smooth)  
d %>% keep(is.numeric) %>% chart.Correlation(., histogram=TRUE, pch=19)  
  
  
#https://liuyanguu.github.io/post/2019/04/17/ggplot-heatmap-us-50-states-map-and-china-province-map/  
par(mfrow = c(1,2))  
  
start = levels(d$year)[1]  
end = levels(d$year)[10]  
library(usmap)

## Warning: package 'usmap' was built under R version 4.0.5

for (i in c(start,end)){  
 df = subset(d, year == i)  
 map <- plot\_usmap(data = df, values = "minwage", labels = T) +  
 labs(fill = 'Min Wage') +  
 scale\_fill\_gradientn(colours=hcl.colors(10, rev = TRUE),na.value="grey",  
 guide = guide\_colourbar(barwidth = 25, barheight = 0.4)) +  
 theme(legend.position = "bottom",  
 legend.title=element\_text(size=10),  
 legend.text=element\_text(size=10))  
   
 print(map)  
  
}  
  
d %>% subset(d$year == 2019) %>% group\_by(region) %>% summarise(mean = mean(avghomeprice)) %>% arrange(mean)

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 8 x 2  
## region mean  
## <fct> <dbl>  
## 1 Great Lakes 176781.  
## 2 Southeast 179646.  
## 3 Plains 194056.  
## 4 Southwest 203856   
## 5 New England 297905.  
## 6 Rocky Mountain 319248.  
## 7 Mideast 346760.  
## 8 Far West 431471.

d %>% subset(d$year == 2019) %>% group\_by(state) %>% summarise(homeprice = avghomeprice) %>% arrange(homeprice) %>% top\_n(5)

## `summarise()` ungrouping output (override with `.groups` argument)

## Selecting by homeprice

## # A tibble: 5 x 2  
## state homeprice  
## <fct> <dbl>  
## 1 Washington 413619  
## 2 Massachusetts 427860  
## 3 California 562678  
## 4 District of Columbia 647177  
## 5 Hawaii 652969

d %>% subset(d$year == 2010) %>% group\_by(state) %>% summarise(homeprice = avghomeprice) %>% arrange(homeprice) %>% top\_n(5)

## `summarise()` ungrouping output (override with `.groups` argument)  
## Selecting by homeprice

## # A tibble: 5 x 2  
## state homeprice  
## <fct> <dbl>  
## 1 New Jersey 315053  
## 2 Massachusetts 319441  
## 3 California 325668  
## 4 District of Columbia 432179  
## 5 Hawaii 483688

d %>% subset(d$year == 2010) %>% group\_by(state) %>% summarise(homeprice = minwage) %>% arrange(homeprice) %>% top\_n(5)

## `summarise()` ungrouping output (override with `.groups` argument)  
## Selecting by homeprice

## # A tibble: 6 x 2  
## state homeprice  
## <fct> <dbl>  
## 1 Connecticut 8.25  
## 2 District of Columbia 8.25  
## 3 Illinois 8.25  
## 4 Nevada 8.25  
## 5 Oregon 8.4   
## 6 Washington 8.55

temp <- d %>% subset(d$year == 2019) %>% group\_by(state) %>% summarise(homeprice = minwage) %>% arrange(homeprice) %>% top\_n(-5)

## `summarise()` ungrouping output (override with `.groups` argument)  
## Selecting by homeprice

library(corrplot)

## corrplot 0.84 loaded

d %>% keep(is.numeric) %>% cor() %>% corrplot(., method = "number", type = "upper", order = "hclust", tl.col = "black", tl.srt = 45)

## Warning in corrplot(., method = "number", type = "upper", order = "hclust", :  
## Not been able to calculate text margin, please try again with a clean new empty  
## window using {plot.new(); dev.off()} or reduce tl.cex

#population: subsd\_units\_avaial, %sub\_units\_occupied  
# unemploymentrate: mort\_30to89  
  
d$numhouse\_subsidized <- NULL  
d$gdp <- NULL  
d$subsd\_units\_avail <- NULL  
  
with(d, interaction.plot(year,minwage,avghomeprice )) #interaction check  
  
  
# #standardize the variables using min max if needed  
# normalized <- function(x){  
# minimum = min(x)  
# maximum = max(x)  
#   
# (x-minimum)/ (maximum-minimum)  
# }  
#   
# denormalized <- function(normalized,x){  
# (normalized)\*(max(x)-min(x))+min(x)  
# }  
#   
  
  
#Model  
library(lme4)

## Warning: package 'lme4' was built under R version 4.0.4

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

##   
## Attaching package: 'lme4'

## The following object is masked from 'package:rio':  
##   
## factorize

m1 <- lmer(log(avghomeprice) ~ log(medianincome) + unemploymentrate+ mort90plus\_delinquentrate+   
 mort30to89\_delinquentrate + pctsubsd\_units\_occupied + minwage\*year + povertyrate +   
 log(population) + (1 | state), data = d, REML=FALSE)  
m2 <- lmer(log(avghomeprice) ~ log(medianincome) + minwage\*year +unemploymentrate+ mort90plus\_delinquentrate+   
 mort30to89\_delinquentrate + pctsubsd\_units\_occupied+ povertyrate +   
 log(population) + (1 | state), data = d, REML=FALSE)  
m3 <- lmer(log(avghomeprice) ~ log(medianincome) + minwage + unemploymentrate+ mort90plus\_delinquentrate+   
 mort30to89\_delinquentrate + pctsubsd\_units\_occupied+ povertyrate +   
 log(population) + (1 | year) + (1 | state), data = d, REML=FALSE)  
summary(m1)

## Linear mixed model fit by maximum likelihood ['lmerMod']  
## Formula:   
## log(avghomeprice) ~ log(medianincome) + unemploymentrate + mort90plus\_delinquentrate +   
## mort30to89\_delinquentrate + pctsubsd\_units\_occupied + minwage \*   
## year + povertyrate + log(population) + (1 | state)  
## Data: d  
##   
## AIC BIC logLik deviance df.resid   
## -1243.9 -1121.1 650.9 -1301.9 481   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.6866 -0.4408 0.0122 0.5020 3.2599   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## state (Intercept) 4.596801 2.14402   
## Residual 0.001637 0.04046   
## Number of obs: 510, groups: state, 51  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) -18.265282 1.773041 -10.302  
## log(medianincome) 0.044678 0.041392 1.079  
## unemploymentrate -0.017747 0.003064 -5.791  
## mort90plus\_delinquentrate -0.003358 0.010735 -0.313  
## mort30to89\_delinquentrate -0.017113 0.004318 -3.963  
## pctsubsd\_units\_occupied -0.003328 0.001238 -2.688  
## minwage -0.001220 0.015923 -0.077  
## year2011 0.062871 0.141757 0.444  
## year2012 0.002671 0.135716 0.020  
## year2013 -0.152807 0.133172 -1.147  
## year2014 -0.162988 0.128320 -1.270  
## year2015 -0.143842 0.123776 -1.162  
## year2016 -0.106125 0.119763 -0.886  
## year2017 -0.077661 0.118130 -0.657  
## year2018 -0.046224 0.118129 -0.391  
## year2019 0.007045 0.118098 0.060  
## povertyrate -0.001027 0.001600 -0.642  
## log(population) 2.012141 0.109961 18.299  
## minwage:year2011 -0.017647 0.019052 -0.926  
## minwage:year2012 -0.011311 0.018175 -0.622  
## minwage:year2013 0.012372 0.017813 0.695  
## minwage:year2014 0.012229 0.017051 0.717  
## minwage:year2015 0.010100 0.016362 0.617  
## minwage:year2016 0.008057 0.015906 0.507  
## minwage:year2017 0.008180 0.015672 0.522  
## minwage:year2018 0.007040 0.015642 0.450  
## minwage:year2019 0.002704 0.015609 0.173

##   
## Correlation matrix not shown by default, as p = 27 > 12.  
## Use print(x, correlation=TRUE) or  
## vcov(x) if you need it

summary(m2)

## Linear mixed model fit by maximum likelihood ['lmerMod']  
## Formula:   
## log(avghomeprice) ~ log(medianincome) + minwage \* year + unemploymentrate +   
## mort90plus\_delinquentrate + mort30to89\_delinquentrate + pctsubsd\_units\_occupied +   
## povertyrate + log(population) + (1 | state)  
## Data: d  
##   
## AIC BIC logLik deviance df.resid   
## -1243.9 -1121.1 650.9 -1301.9 481   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.6866 -0.4408 0.0122 0.5020 3.2599   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## state (Intercept) 4.596801 2.14402   
## Residual 0.001637 0.04046   
## Number of obs: 510, groups: state, 51  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) -18.265282 1.773041 -10.302  
## log(medianincome) 0.044678 0.041392 1.079  
## minwage -0.001220 0.015923 -0.077  
## year2011 0.062871 0.141757 0.444  
## year2012 0.002671 0.135716 0.020  
## year2013 -0.152807 0.133172 -1.147  
## year2014 -0.162988 0.128320 -1.270  
## year2015 -0.143842 0.123776 -1.162  
## year2016 -0.106125 0.119763 -0.886  
## year2017 -0.077661 0.118130 -0.657  
## year2018 -0.046224 0.118129 -0.391  
## year2019 0.007045 0.118098 0.060  
## unemploymentrate -0.017747 0.003064 -5.791  
## mort90plus\_delinquentrate -0.003358 0.010735 -0.313  
## mort30to89\_delinquentrate -0.017113 0.004318 -3.963  
## pctsubsd\_units\_occupied -0.003328 0.001238 -2.688  
## povertyrate -0.001027 0.001600 -0.642  
## log(population) 2.012141 0.109961 18.299  
## minwage:year2011 -0.017647 0.019052 -0.926  
## minwage:year2012 -0.011311 0.018175 -0.622  
## minwage:year2013 0.012372 0.017813 0.695  
## minwage:year2014 0.012229 0.017051 0.717  
## minwage:year2015 0.010100 0.016362 0.617  
## minwage:year2016 0.008057 0.015906 0.507  
## minwage:year2017 0.008180 0.015672 0.522  
## minwage:year2018 0.007040 0.015642 0.450  
## minwage:year2019 0.002704 0.015609 0.173

##   
## Correlation matrix not shown by default, as p = 27 > 12.  
## Use print(x, correlation=TRUE) or  
## vcov(x) if you need it

stargazer(m1, m2, m3, type='text', single.row = TRUE)

##   
## ==================================================================================  
## Dependent variable:   
## --------------------------------------------------------  
## log(avghomeprice)   
## (1) (2) (3)   
## ----------------------------------------------------------------------------------  
## log(medianincome) 0.045 (0.041) 0.045 (0.041) 0.063 (0.039)   
## unemploymentrate -0.018\*\*\* (0.003) -0.018\*\*\* (0.003) -0.019\*\*\* (0.003)   
## mort90plus\_delinquentrate -0.003 (0.011) -0.003 (0.011) -0.001 (0.010)   
## mort30to89\_delinquentrate -0.017\*\*\* (0.004) -0.017\*\*\* (0.004) -0.019\*\*\* (0.004)   
## pctsubsd\_units\_occupied -0.003\*\*\* (0.001) -0.003\*\*\* (0.001) -0.004\*\*\* (0.001)   
## minwage -0.001 (0.016) -0.001 (0.016) 0.005 (0.003)   
## year2011 0.063 (0.142) 0.063 (0.142)   
## year2012 0.003 (0.136) 0.003 (0.136)   
## year2013 -0.153 (0.133) -0.153 (0.133)   
## year2014 -0.163 (0.128) -0.163 (0.128)   
## year2015 -0.144 (0.124) -0.144 (0.124)   
## year2016 -0.106 (0.120) -0.106 (0.120)   
## year2017 -0.078 (0.118) -0.078 (0.118)   
## year2018 -0.046 (0.118) -0.046 (0.118)   
## year2019 0.007 (0.118) 0.007 (0.118)   
## povertyrate -0.001 (0.002) -0.001 (0.002) -0.001 (0.002)   
## log(population) 2.012\*\*\* (0.110) 2.012\*\*\* (0.110) 2.004\*\*\* (0.111)   
## minwage:year2011 -0.018 (0.019) -0.018 (0.019)   
## minwage:year2012 -0.011 (0.018) -0.011 (0.018)   
## minwage:year2013 0.012 (0.018) 0.012 (0.018)   
## minwage:year2014 0.012 (0.017) 0.012 (0.017)   
## minwage:year2015 0.010 (0.016) 0.010 (0.016)   
## minwage:year2016 0.008 (0.016) 0.008 (0.016)   
## minwage:year2017 0.008 (0.016) 0.008 (0.016)   
## minwage:year2018 0.007 (0.016) 0.007 (0.016)   
## minwage:year2019 0.003 (0.016) 0.003 (0.016)   
## Constant -18.265\*\*\* (1.773) -18.265\*\*\* (1.773) -18.378\*\*\* (1.777)  
## ----------------------------------------------------------------------------------  
## Observations 510 510 510   
## Log Likelihood 650.930 650.930 626.239   
## Akaike Inf. Crit. -1,243.861 -1,243.861 -1,228.478   
## Bayesian Inf. Crit. -1,121.063 -1,121.063 -1,177.665   
## ==================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#State-wise and yearly effects  
dotplot(ranef(m3, condVar=TRUE))

## $state

##   
## $year

ranef <- ranef(m3)  
ranef\_df <- as.data.frame(ranef)  
ranef\_df$term <- NULL  
ranef\_df$condsd <- NULL  
df\_out <- split( ranef\_df , f = ranef\_df$grpvar )  
df\_state <- df\_out$state  
df\_year <- df\_out$year  
df\_state <- df\_state[order(df\_state$condval, decreasing = TRUE), ]  
df\_year <- df\_year[order(df\_year$condval, decreasing = TRUE), ]  
df\_state$grpvar <- NULL  
df\_year$grpvar <- NULL  
colnames(df\_state) <- c("StateName", "Coefficient")  
colnames(df\_year) <- c("Year", "Coefficient")  
row.names(df\_state) <- NULL  
row.names(df\_year) <- NULL  
  
fixef(m3)

## (Intercept) log(medianincome) minwage   
## -18.378444228 0.062616687 0.005033967   
## unemploymentrate mort90plus\_delinquentrate mort30to89\_delinquentrate   
## -0.018710292 -0.001416813 -0.019148093   
## pctsubsd\_units\_occupied povertyrate log(population)   
## -0.003753771 -0.001193539 2.004108940

fixef <- fixef(m3)  
fixef\_df <- as.data.frame(fixef)  
temp <- fixef\_df$fixef[1]  
df\_state$betacoeff <- df\_state$Coefficient + temp  
df\_state$Coefficient <- NULL  
  
df\_year$betacoeff <- df\_year$Coefficient + temp  
df\_year$Coefficient <- NULL  
  
head(df\_state,5)

## StateName betacoeff  
## 1 District of Columbia -13.86590  
## 2 Wyoming -14.54219  
## 3 Vermont -14.65186  
## 4 Alaska -14.73207  
## 5 North Dakota -15.14814

tail(df\_state,5)

## StateName betacoeff  
## 47 Ohio -21.06684  
## 48 New York -21.27171  
## 49 Florida -21.73045  
## 50 California -22.19947  
## 51 Texas -22.50693

df\_year

## Year betacoeff  
## 1 2019 -18.32779  
## 2 2010 -18.33508  
## 3 2018 -18.34043  
## 4 2017 -18.35988  
## 5 2016 -18.38750  
## 6 2013 -18.39740  
## 7 2011 -18.40202  
## 8 2015 -18.40770  
## 9 2014 -18.40952  
## 10 2012 -18.41713

#model assumptions  
library(lmtest)

## Warning: package 'lmtest' was built under R version 4.0.4

library(car)

## Loading required package: carData

## Registered S3 methods overwritten by 'car':  
## method from  
## influence.merMod lme4  
## cooks.distance.influence.merMod lme4  
## dfbeta.influence.merMod lme4  
## dfbetas.influence.merMod lme4

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

library(DHARMa) # for autocorrelation test

## Warning: package 'DHARMa' was built under R version 4.0.5

## This is DHARMa 0.4.1. For overview type '?DHARMa'. For recent changes, type news(package = 'DHARMa') Note: Syntax of plotResiduals has changed in 0.3.0, see ?plotResiduals for details

lmer\_assumptions <- function(model){  
 #' Tests for assumptions  
 plot(resid(model) ~ fitted(model)) # Residual plot for Linearity  
 hist(resid(model))  
   
 qqnorm(resid(model)) # Q-Q plot for normality  
 qqline(resid(model), col="red")  
 print(shapiro.test(resid(m1)))  
 norm <- rnorm(200)  
 print(ks.test(norm, resid(model))) # Kolmogorov-Smirnov test for normality  
   
   
 print(bartlett.test(list(resid(model), fitted(model))))# Bartlett test of homoskedasticity  
   
 print(vif(model)) # Variance inflation factor  
   
 print(testTemporalAutocorrelation(model, time = NULL)) # Autocorrelation   
   
}  
  
  
lmer\_assumptions(m1)

##   
## Shapiro-Wilk normality test  
##   
## data: resid(m1)  
## W = 0.97091, p-value = 0.00000001592  
##   
##   
## Two-sample Kolmogorov-Smirnov test  
##   
## data: norm and resid(model)  
## D = 0.45147, p-value < 0.00000000000000022  
## alternative hypothesis: two-sided  
##   
##   
## Bartlett test of homogeneity of variances  
##   
## data: list(resid(model), fitted(model))  
## Bartlett's K-squared = 1717.5, df = 1, p-value < 0.00000000000000022  
##   
## GVIF Df GVIF^(1/(2\*Df))  
## log(medianincome) 6.747053 1 2.597509  
## unemploymentrate 11.321055 1 3.364678  
## mort90plus\_delinquentrate 11.251007 1 3.354252  
## mort30to89\_delinquentrate 6.712630 1 2.590874  
## pctsubsd\_units\_occupied 1.674454 1 1.294007  
## minwage 47.661064 1 6.903699  
## year 7329266518327103488.000000 9 11.170148  
## povertyrate 2.614234 1 1.616859  
## log(population) 2.337855 1 1.529005  
## minwage:year 10098445456480309248.000000 9 11.370823

## DHARMa::testTemporalAutocorrelation - no time argument provided, using random times for each data point

##   
## Durbin-Watson test  
##   
## data: simulationOutput$scaledResiduals ~ 1  
## DW = 1.8423, p-value = 0.07445  
## alternative hypothesis: true autocorrelation is not 0

lmer\_assumptions(m2)

##   
## Shapiro-Wilk normality test  
##   
## data: resid(m1)  
## W = 0.97091, p-value = 0.00000001592  
##   
##   
## Two-sample Kolmogorov-Smirnov test  
##   
## data: norm and resid(model)  
## D = 0.45127, p-value < 0.00000000000000022  
## alternative hypothesis: two-sided  
##   
##   
## Bartlett test of homogeneity of variances  
##   
## data: list(resid(model), fitted(model))  
## Bartlett's K-squared = 1717.5, df = 1, p-value < 0.00000000000000022  
##   
## GVIF Df GVIF^(1/(2\*Df))  
## log(medianincome) 6.747053 1 2.597509  
## minwage 47.661064 1 6.903699  
## year 7329266518182236160.000000 9 11.170148  
## unemploymentrate 11.321055 1 3.364678  
## mort90plus\_delinquentrate 11.251007 1 3.354252  
## mort30to89\_delinquentrate 6.712630 1 2.590874  
## pctsubsd\_units\_occupied 1.674454 1 1.294007  
## povertyrate 2.614234 1 1.616859  
## log(population) 2.337855 1 1.529005  
## minwage:year 10098445456215609344.000000 9 11.370823

## DHARMa::testTemporalAutocorrelation - no time argument provided, using random times for each data point

##   
## Durbin-Watson test  
##   
## data: simulationOutput$scaledResiduals ~ 1  
## DW = 1.8877, p-value = 0.204  
## alternative hypothesis: true autocorrelation is not 0

lmer\_assumptions(m3)

##   
## Shapiro-Wilk normality test  
##   
## data: resid(m1)  
## W = 0.97091, p-value = 0.00000001592  
##   
##   
## Two-sample Kolmogorov-Smirnov test  
##   
## data: norm and resid(model)  
## D = 0.48716, p-value < 0.00000000000000022  
## alternative hypothesis: two-sided  
##   
##   
## Bartlett test of homogeneity of variances  
##   
## data: list(resid(model), fitted(model))  
## Bartlett's K-squared = 1708, df = 1, p-value < 0.00000000000000022  
##   
## log(medianincome) minwage unemploymentrate   
## 1.042529 1.092325 1.846450   
## mort90plus\_delinquentrate mort30to89\_delinquentrate pctsubsd\_units\_occupied   
## 1.550547 1.813582 1.108704   
## povertyrate log(population)   
## 1.173789 1.181631

## DHARMa::testTemporalAutocorrelation - no time argument provided, using random times for each data point

##   
## Durbin-Watson test  
##   
## data: simulationOutput$scaledResiduals ~ 1  
## DW = 1.9869, p-value = 0.8826  
## alternative hypothesis: true autocorrelation is not 0

summary(m3)

## Linear mixed model fit by maximum likelihood ['lmerMod']  
## Formula: log(avghomeprice) ~ log(medianincome) + minwage + unemploymentrate +   
## mort90plus\_delinquentrate + mort30to89\_delinquentrate + pctsubsd\_units\_occupied +   
## povertyrate + log(population) + (1 | year) + (1 | state)  
## Data: d  
##   
## AIC BIC logLik deviance df.resid   
## -1228.5 -1177.7 626.2 -1252.5 498   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.6003 -0.4800 0.0333 0.5261 3.1237   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## state (Intercept) 4.557553 2.13484   
## year (Intercept) 0.001205 0.03471   
## Residual 0.001700 0.04123   
## Number of obs: 510, groups: state, 51; year, 10  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) -18.378444 1.776737 -10.344  
## log(medianincome) 0.062617 0.039382 1.590  
## minwage 0.005034 0.003183 1.581  
## unemploymentrate -0.018710 0.003047 -6.140  
## mort90plus\_delinquentrate -0.001417 0.010418 -0.136  
## mort30to89\_delinquentrate -0.019148 0.004283 -4.471  
## pctsubsd\_units\_occupied -0.003754 0.001233 -3.043  
## povertyrate -0.001194 0.001612 -0.740  
## log(population) 2.004109 0.111087 18.041  
##   
## Correlation of Fixed Effects:  
## (Intr) lg(md) minwag unmply mrt90\_ m3089\_ pcts\_\_ pvrtyr  
## log(mdnncm) -0.199   
## minwage -0.035 0.037   
## unmplymntrt 0.048 0.057 0.008   
## mrt90pls\_dl -0.119 0.098 -0.159 -0.313   
## mrt30t89\_dl -0.153 -0.021 0.206 -0.447 -0.251   
## pctsbsd\_nt\_ -0.286 0.092 0.121 -0.033 0.023 0.106   
## povertyrate -0.165 -0.035 0.073 -0.177 -0.047 -0.004 -0.116   
## log(popltn) -0.951 -0.054 0.005 -0.064 0.090 0.161 0.210 0.180

confint(m3)

## Computing profile confidence intervals ...

## 2.5 % 97.5 %  
## .sig01 1.699763519 2.710615089  
## .sig02 0.021972016 0.062051745  
## .sigma 0.038619636 0.044160547  
## (Intercept) -22.471674303 -14.108970331  
## log(medianincome) -0.018205750 0.142778384  
## minwage -0.001254331 0.011327134  
## unemploymentrate -0.024705603 -0.012700369  
## mort90plus\_delinquentrate -0.021978586 0.019084123  
## mort30to89\_delinquentrate -0.027635146 -0.010709298  
## pctsubsd\_units\_occupied -0.006213441 -0.001313394  
## povertyrate -0.004423256 0.002007510  
## log(population) 1.731749913 2.262707525