STATA CODING SAMPLE: PLOTTING AND ESTIMATION FOR SILAS KWOK HONOURS THESIS

```
***********************************INIT***********************************************
clear
cd "/Users/silas/ECON495/PROJ/STATA/Do_Files"

set scheme s1color
set textsize 300
graph set window fontface "Times New Roman"

***********************************INIT***********************************************
clear
cd "/Users/silas/ECON495/PROJ/STATA/Raw_data"

import delimited "/Users/silas/ECON495/CLOSER/zhvi_50_est.csv"

drop if missing(log_zhvi)
drop if log_zhvi == .

duplicates drop store_year_month zipcode, force

//Use 2010 ACS data for heterogeneity by county
egen mean_income_2010 = mean(cond(year == 2010, house_income, .))
egen mean_density_2010 = mean(cond(year == 2010, pop_density, .))
egen mean_white_2010 = mean(cond(year == 2010, white, .))

gen above_avg_income_2010 = cond(year == 2010, house_income > mean_income_2010, .)
gen above_avg_density_2010 = cond(year == 2010, pop_density > mean_density_2010, .)
gen above_avg_white_2010 = cond(year == 2010, white > mean_white_2010, .)

egen max_above_avg_income_2010 = max(above_avg_income_2010), by(zipcode)
egen max_above_avg_density_2010 = max(above_avg_density_2010), by(zipcode)
egen max_above_avg_white_2010 = max(above_avg_white_2010), by(zipcode)

replace above_avg_income_2010 = max_above_avg_income_2010
replace above_avg_density_2010 = max_above_avg_density_2010
replace above_avg_white_2010 = max_above_avg_white_2010

rename above_avg_income_2010 above_avg_income
rename above_avg_density_2010 above_avg_density
rename above_avg_white_2010 above_avg_white

drop max_above_avg_income_2010 max_above_avg_density_2010 max_above_avg_white_2010

save temp_acs_complete.dta, replace
sort zipcode quarters_from_opening
***********************************INIT***********************************************

//DATA VERSION (KEEP ZIPCODES W DATA FOR -10 TO 10 QUARTERS_FROM_OPENING)
keep if quarters_from_opening >= -10 & quarters_from_opening <= 10
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 63
drop obs_count

save temp_acs.dta, replace

***********************************INIT***********************************************
//PARALLEL TRENDS ASSUMPTION EVIDENCE : GRAPH OF SIMPLE MEANS

use temp_acs.dta, clear

collapse (mean) log_zhvi, by(distance_category quarters_from_opening)
```

```
twoway (line log_zhvi quarters_from_opening if distance_category == "0-5 miles", lcolor(red)
lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "5-10 miles",
lcolor(orange) lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "10-15 miles",
lcolor(green) lpattern(solid)) ///
          (line log_zhvi quarters_from_opening if distance_category == "15-20 miles",
lcolor(blue) lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "20-25 miles",
lcolor(black) lpattern(solid) lwidth(thick)), ///
       legend(order(1 "0-5 miles" 2 "5-10 miles" 3 "10-15 miles" 4 "15-20 miles" 5 "20-25
miles")) ///
       title("Mean Log ZHVIs by Distance Category" "over Quarters from Opening", size(large))
///
       ytitle("Log ZHVI") xtitle("Quarters From Store Opening") xscale(range(-10 10)) ///
       xline(-10 -4 0 10, lstyle(grid) lcolor(maroon))

use temp_acs.dta, clear

replace log_zhvi = log_zhvi - 11.03357

collapse (mean) log_zhvi, by(distance_category quarters_from_opening)

qui sum log_zhvi if distance_category == "20-25 miles" & quarters_from_opening == -10
local ref_value = r(mean)

foreach cat in "0-5 miles" "5-10 miles" "10-15 miles" "15-20 miles" {
    if "`cat'" == "20-25 miles" continue
    qui sum log_zhvi if distance_category == "`cat'" & quarters_from_opening == -10
    local cat_value = r(mean)
    local diff = `ref_value' - `cat_value'
    replace log_zhvi = log_zhvi + `diff' if distance_category == "`cat'"
}

* NORMALIZED PLOT
twoway (line log_zhvi quarters_from_opening if distance_category == "0-5 miles", lcolor(red)
lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "5-10 miles",
lcolor(orange) lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "10-15 miles",
lcolor(green) lpattern(solid)) ///
          (line log_zhvi quarters_from_opening if distance_category == "15-20 miles",
lcolor(blue) lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "20-25 miles",
lcolor(black) lpattern(solid) lwidth(thick)), ///
       legend(order(1 "0-5 miles" 2 "5-10 miles" 3 "10-15 miles" 4 "15-20 miles" 5 "20-25
miles") size(small) position(5) cols(1) ring(0) region(margin(tiny)) region(lwidth(none))) ///
       title("Figure 4: Evidence for Parallel Trends" "Normalized Mean Log ZHVI Values over
Quarters from Opening", size(medlarge)) ///
       ytitle("Log ZHVI") xtitle("Quarters From Store Opening") xscale(range(-10 10)) ///
          yscale(range(1 1.25)) ///
       xline(-10 0 10, lstyle(grid) lcolor(maroon)) ///
          xline(-4, lcolor(maroon) lpattern(dash) lwidth(medthick))

**********************************INIT**********************************************
//SUMMARY STATS
use temp_acs_full.dta, clear

sort distance_category post_treatment
bysort distance_category post_treatment: egen mean_log_zhvi = mean(log_zhvi)
bysort distance_category post_treatment: egen sd_log_zhvi = sd(log_zhvi)
```

```
bysort distance_category post_treatment: gen freq = _N
collapse (mean) mean_log_zhvi sd_log_zhvi (count) freq, by(distance_category post_treatment)
export excel using summary_stats_a1.xlsx, firstrow(variables) replace

use temp_acs_full.dta, clear
tab distance_category post_treatment, summarize(zhvi)
bysort distance_category post_treatment: egen mean_zhvi = mean(zhvi)
bysort distance_category post_treatment: egen sd_zhvi = sd(zhvi)
bysort distance_category post_treatment: gen freq = _N
collapse (mean) mean_zhvi sd_zhvi (count) freq, by(distance_category post_treatment)
export excel using summary_stats_a2.xlsx, firstrow(variables) replace

use temp_acs_full.dta, clear
tab distance_category post_treatment, summarize(house_income)
bysort distance_category post_treatment: egen mean_house_income = mean(house_income)
bysort distance_category post_treatment: egen sd_house_income = sd(house_income)
bysort distance_category post_treatment: gen freq = _N
collapse (mean) mean_house_income sd_house_income (count) freq, by(distance_category
post_treatment)
export excel using summary_stats_a3.xlsx, firstrow(variables) replace

use temp_acs_full.dta, clear
tab distance_category post_treatment, summarize(pop_density)
bysort distance_category post_treatment: egen mean_pop_density = mean(pop_density)
bysort distance_category post_treatment: egen sd_pop_density = sd(pop_density)
bysort distance_category post_treatment: gen freq = _N
collapse (mean) mean_pop_density sd_pop_density (count) freq, by(distance_category
post_treatment)
export excel using summary_stats_a4.xlsx, firstrow(variables) replace

// combine summary_stats_a to d for full summary stats.

//T-tests for difference in means (take the p-value)
use temp_acs_full.dta, clear
drop zhvi nearest_store quarters_from_opening months_from_opening treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 control treat5_post treat10_post treat15_post treat20_post
control_post year month pop house_income bachelors_deg_plus white land_area_sqm pop_density
timevar nearest_store_num store_year_month_num city_num county_num state_num store_time
store_time_num
keep if distance_category == "0-5 miles" | distance_category == "20-25 miles"
reshape wide log_zhvi, i(zipcode distance_category) j(post_treatment)
gen diff_log_zhvi = log_zhvi1 - log_zhvi0
drop log_zhvi1 log_zhvi0
ttest diff_log_zhvi, by(distance_category)

use temp_acs_full.dta, clear
drop zhvi nearest_store quarters_from_opening months_from_opening treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 control treat5_post treat10_post treat15_post treat20_post
control_post year month pop house_income bachelors_deg_plus white land_area_sqm pop_density
timevar nearest_store_num store_year_month_num city_num county_num state_num store_time
store_time_num
keep if distance_category == "5-10 miles" | distance_category == "20-25 miles"
reshape wide log_zhvi, i(zipcode distance_category) j(post_treatment)
gen diff_log_zhvi = log_zhvi1 - log_zhvi0
drop log_zhvi1 log_zhvi0
ttest diff_log_zhvi, by(distance_category)

use temp_acs_full.dta, clear
drop zhvi nearest_store quarters_from_opening months_from_opening treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 control treat5_post treat10_post treat15_post treat20_post
control_post year month pop house_income bachelors_deg_plus white land_area_sqm pop_density
```

```
timevar nearest_store_num store_year_month_num city_num county_num state_num store_time
store_time_num
keep if distance_category == "10-15 miles" | distance_category == "20-25 miles"
reshape wide log_zhvi, i(zipcode distance_category) j(post_treatment)
gen diff_log_zhvi = log_zhvi1 - log_zhvi0
drop log_zhvi1 log_zhvi0
ttest diff_log_zhvi, by(distance_category)

use temp_acs_full.dta, clear
drop zhvi nearest_store quarters_from_opening months_from_opening treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 control treat5_post treat10_post treat15_post treat20_post
control_post year month pop house_income bachelors_deg_plus white land_area_sqm pop_density
timevar nearest_store_num store_year_month_num city_num county_num state_num store_time
store_time_num
keep if distance_category == "15-20 miles" | distance_category == "20-25 miles"
reshape wide log_zhvi, i(zipcode distance_category) j(post_treatment)
gen diff_log_zhvi = log_zhvi1 - log_zhvi0
drop log_zhvi1 log_zhvi0
ttest diff_log_zhvi, by(distance_category)

**********************************INIT***********************************************
// DATA CLEANING
clear
cd "/Users/silas/ECON495/PROJ/STATA/Raw_data"
use temp_acs.dta, clear

keep if months_from_opening == -30 | months_from_opening == 30

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)   // Using Jan 31, 2000 as the reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both

sort zipcode

encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode name, generate(county_num)
encode state_full, generate(state_num)

gen int month_int = month
drop month
rename month_int month

gen store_time = nearest_store + "_" + string(date_stata,"%12.0g")
encode store_time, generate(store_time_num)

order year month, after(control_post)

drop geo_id fips_code name state_full city countyname latitude longitude geometry
min_distance_to_store min_distance_to_store_miles store_year_month date_stata

//Drop Alaskan Costco opening data points
drop if nearest_store == "Fairbanks"

save temp_acs_full.dta, replace
```

```
**********************************INIT**************************************************
//DIFF-IN-DIFF SPECIFICATIONS

clear
cd "/Users/silas/ECON495/PROJ/STATA/Raw_data"

use temp_acs_full.dta, clear

drop zhvi pop land_area_sqm bachelors_deg_plus white

sort zipcode timevar
xtset zipcode timevar, monthly

//Model 1 (Year FE + State FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.state_num i.month, vce(cluster nearest_store)

outreg2 using all_data_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) replace
ctitle(ln(ZHVI)) e(r2_o) addtext(Year FE, YES, Month FE, YES, State FE, YES, County FE, NO,
City FE, NO, Controls, YES)

//Model 2 (Year FE + County FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using all_data_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(ln(ZHVI)) e(r2_o) addtext(Year FE, YES, County FE, YES, Month FE, YES, Controls, YES)

//Model 3 (Year FE + City FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.city_num i.month, vce(cluster nearest_store)

outreg2 using all_data_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(ln(ZHVI)) e(r2_o) addtext(Year FE, YES, City FE, YES, Month FE, YES, Controls, YES)

**********************************INIT**************************************************
//No Controls Version
clear
cd "/Users/silas/ECON495/PROJ/STATA/Raw_data"

use temp_acs_full.dta, clear

drop zhvi pop land_area_sqm bachelors_deg_plus white house_income pop_density mean_income_2010
mean_density_2010 mean_white_2010 above_avg_income above_avg_density above_avg_white

sort zipcode timevar
xtset zipcode timevar, monthly

//Model 1 (Year FE + State FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post i.year i.state_num i.month, vce(cluster nearest_store)

outreg2 using nocontrol_data_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) replace
```

```
ctitle(ln(ZHVI)) e(r2_o) addtext(Year FE, YES, Month FE, YES, State FE, YES, County FE, NO,
City FE, NO, Controls, NO)

//Model 2 (Year FE + County FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using nocontrol_data_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(ln(ZHVI)) e(r2_o) addtext(Year FE, YES, County FE, YES, Month FE, YES, Controls, NO)

//Model 3 (Year FE + City FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post i.year i.city_num i.month, vce(cluster nearest_store)

outreg2 using nocontrol_data_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(ln(ZHVI)) e(r2_o) addtext(Year FE, YES, City FE, YES, Month FE, YES, Controls, NO)


**********************************INIT**********************************************
//Coefficient Plot for Model 2
use temp_acs_full.dta, clear

drop pop land_area_sqm bachelors_deg_plus white
xtset zipcode timevar, monthly

xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

local varlist treat5_post treat10_post treat15_post treat20_post
coefplot, keep(treat5_post treat10_post treat15_post treat20_post) vertical ///
    yline(0, lcolor(maroon)) ///
    coeflabels(treat5_post = "0-5 miles" treat10_post = "5-10 miles" treat15_post = "10-15
miles" ///
    treat20_post = "15-20 miles") ///
    xtitle("Change in Log ZHVI by Distance Post-Treatment") ///
    ytitle("Coefficient (Treatment Effect)") ///
       ysize(3) ///
       name(TreatmentEffects, replace) ///
       graphregion(margin(large))

//Coefficient Plot for Model 3
use temp_acs_full.dta, clear

drop pop land_area_sqm bachelors_deg_plus white
xtset zipcode timevar, monthly

xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.city_num i.month, vce(cluster nearest_store)

local varlist treat5_post treat10_post treat15_post treat20_post
coefplot, keep(treat5_post treat10_post treat15_post treat20_post) vertical ///
    yline(0, lcolor(maroon)) ///
    coeflabels(treat5_post = "0-5 miles" treat10_post = "5-10 miles" treat15_post = "10-15
miles" ///
    treat20_post = "15-20 miles") ///
    xtitle("Change in Log ZHVI by Distance Post-Treatment") ///
    ytitle("Coefficient (Treatment Effect)") ///
       ysize(3) ///
```

```
        name(TreatmentEffects, replace) ///
        graphregion(margin(large))


**********************************INIT***********************************************
// CLEANED DATA FOR PLOTTING HETEROGENEOUS DATA
clear
cd "/Users/silas/ECON495/PROJ/STATA/Raw_data"
use temp_acs.dta, clear

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)   // Using Jan 31, 2000 as the reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both

sort zipcode

encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode name, generate(county_num)
encode state_full, generate(state_num)

gen int month_int = month
drop month
rename month_int month

gen store_time = nearest_store + "_" + string(date_stata,"%12.0g")
encode store_time, generate(store_time_num)

order year month, after(control_post)

drop geo_id fips_code name state_full city countyname latitude longitude geometry
min_distance_to_store min_distance_to_store_miles store_year_month date_stata

//Drop Alaskan Costco opening data points
drop if nearest_store == "Fairbanks"

save temp_acs_plot.dta, replace
**********************************INIT***********************************************
//HETEROGENEITY ANALYSIS

//>Avg. Population Density
// use temp_acs_full.dta, clear
use temp_acs_plot.dta, clear

keep if above_avg_density == 1
drop above_avg_income above_avg_white
drop bachelors_deg_plus white

//CHECK PARALLEL TRENDS FOR SUBGROUP
//normalize values to log_zhvi = 1 @ -10 quarters_from_opening
replace log_zhvi = log_zhvi - 11.3334473

collapse (mean) log_zhvi, by(distance_category quarters_from_opening)

qui sum log_zhvi if distance_category == "20-25 miles" & quarters_from_opening == -10
```

```
local ref_value = r(mean)

foreach cat in "0-5 miles" "5-10 miles" "10-15 miles" "15-20 miles" {
    if "`cat'" == "20-25 miles" continue
    qui sum log_zhvi if distance_category == "`cat'" & quarters_from_opening == -10
    local cat_value = r(mean)
    local diff = `ref_value' - `cat_value'
    replace log_zhvi = log_zhvi + `diff' if distance_category == "`cat'"
}

* NORMALIZED PLOT
twoway (line log_zhvi quarters_from_opening if distance_category == "0-5 miles", lcolor(red)
lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "5-10 miles",
lcolor(orange) lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "10-15 miles",
lcolor(green) lpattern(solid)) ///
          (line log_zhvi quarters_from_opening if distance_category == "15-20 miles",
lcolor(blue) lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "20-25 miles",
lcolor(black) lpattern(solid) lwidth(thick)), ///
       legend(order(1 "0-5 miles" 2 "5-10 miles" 3 "10-15 miles" 4 "15-20 miles" 5 "20-25
miles") size(small) position(5) cols(1) ring(0) region(margin(tiny)) region(lwidth(none))) ///
       ytitle("Log ZHVI") xtitle("Quarters From Store Opening") xscale(range(-10 10)) ///
       xline(-10 0 10, lstyle(grid) lcolor(maroon)) ///
          xline(-4, lcolor(maroon) lpattern(dash) lwidth(medthick))

use temp_acs_plot.dta, clear

keep if above_avg_density == 1
drop above_avg_income above_avg_white
drop bachelors_deg_plus white

keep if months_from_opening == -30 | months_from_opening == 30

sort zipcode post_treatment

//sanity check
// sort zipcode
// by zipcode: egen has_zero = total(post_treatment == 0)
// by zipcode: egen has_one = total(post_treatment == 1)
// gen missing_either = (has_zero == 0) | (has_one == 0)
// list zipcode missing_either if missing_either == 1


xtset zipcode timevar, monthly

//Model 2 for >Avg. Population Density (Year FE + County FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using het_estimation_results.xls, keep(treatment_0_5 treatment_5_10 treatment_10_15
treatment_15_20 treat5_post treat10_post treat15_post treat20_post) replace ctitle(> Avg. Pop.
Density) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)


drop house_income pop_density
//NO CONTROLS
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post i.year i.county_num i.month, vce(cluster nearest_store)
```

```
outreg2 using nocontrol_het_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) replace
ctitle(> Avg. Pop. Density) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES,
Controls, YES)

**********************************INIT**************************************************
**********************************INIT**************************************************
//>Avg. Household Income
use temp_acs_plot.dta, clear

keep if above_avg_income == 1
drop above_avg_density above_avg_white
drop bachelors_deg_plus white

//CHECK PARALLEL TRENDS FOR SUBGROUP
//normalize values to log_zhvi = 1 @ -10 quarters_from_opening
replace log_zhvi = log_zhvi - 11.292113

collapse (mean) log_zhvi, by(distance_category quarters_from_opening)

qui sum log_zhvi if distance_category == "20-25 miles" & quarters_from_opening == -10
local ref_value = r(mean)

foreach cat in "0-5 miles" "5-10 miles" "10-15 miles" "15-20 miles" {
    if "`cat'" == "20-25 miles" continue
    qui sum log_zhvi if distance_category == "`cat'" & quarters_from_opening == -10
    local cat_value = r(mean)
    local diff = `ref_value' - `cat_value'
    replace log_zhvi = log_zhvi + `diff' if distance_category == "`cat'"
}

* NORMALIZED PLOT
twoway (line log_zhvi quarters_from_opening if distance_category == "0-5 miles", lcolor(red)
lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "5-10 miles",
lcolor(orange) lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "10-15 miles",
lcolor(green) lpattern(solid)) ///
          (line log_zhvi quarters_from_opening if distance_category == "15-20 miles",
lcolor(blue) lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "20-25 miles",
lcolor(black) lpattern(solid) lwidth(thick)), ///
       legend(order(1 "0-5 miles" 2 "5-10 miles" 3 "10-15 miles" 4 "15-20 miles" 5 "20-25
miles") size(small) position(5) cols(1) ring(0) region(margin(tiny)) region(lwidth(none))) ///
       ytitle("Log ZHVI") xtitle("Quarters From Store Opening") xscale(range(-10 10)) ///
       xline(-10 0 10, lstyle(grid) lcolor(maroon)) ///
          xline(-4, lcolor(maroon) lpattern(dash) lwidth(medthick))

use temp_acs_plot.dta, clear

keep if above_avg_income == 1
drop above_avg_density above_avg_white
drop bachelors_deg_plus white

keep if months_from_opening == -30 | months_from_opening == 30

xtset zipcode timevar, monthly

//Model 2 for >Avg. Household Income (Year FE + County FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)
```

```
outreg2 using het_estimation_results.xls, keep(treatment_0_5 treatment_5_10 treatment_10_15
treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append ctitle(> Avg.
Income) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)

drop house_income pop_density
//NO CONTROLS
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using nocontrol_het_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(> Avg. Household Income) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES,
Controls, YES)

**********************************INIT**************************************************
**********************************INIT**************************************************
//>Avg. WHITE
use temp_acs_plot.dta, clear

keep if above_avg_white == 1
drop above_avg_density above_avg_income
drop bachelors_deg_plus white

//CHECK PARALLEL TRENDS FOR SUBGROUP
//normalize values to log_zhvi = 1 @ -10 quarters_from_opening
replace log_zhvi = log_zhvi - 11.2690753

collapse (mean) log_zhvi, by(distance_category quarters_from_opening)

qui sum log_zhvi if distance_category == "20-25 miles" & quarters_from_opening == -10
local ref_value = r(mean)

foreach cat in "0-5 miles" "5-10 miles" "10-15 miles" "15-20 miles" {
    if "`cat'" == "20-25 miles" continue
    qui sum log_zhvi if distance_category == "`cat'" & quarters_from_opening == -10
    local cat_value = r(mean)
    local diff = `ref_value' - `cat_value'
    replace log_zhvi = log_zhvi + `diff' if distance_category == "`cat'"
}

* NORMALIZED PLOT
twoway (line log_zhvi quarters_from_opening if distance_category == "0-5 miles", lcolor(red)
lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "5-10 miles",
lcolor(orange) lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "10-15 miles",
lcolor(green) lpattern(solid)) ///
          (line log_zhvi quarters_from_opening if distance_category == "15-20 miles",
lcolor(blue) lpattern(solid)) ///
       (line log_zhvi quarters_from_opening if distance_category == "20-25 miles",
lcolor(black) lpattern(solid) lwidth(thick)), ///
       legend(order(1 "0-5 miles" 2 "5-10 miles" 3 "10-15 miles" 4 "15-20 miles" 5 "20-25
miles") size(small) position(5) cols(1) ring(0) region(margin(tiny)) region(lwidth(none))) ///
       ytitle("Log ZHVI") xtitle("Quarters From Store Opening") xscale(range(-10 10)) ///
       xline(-10 0 10, lstyle(grid) lcolor(maroon)) ///
          xline(-4, lcolor(maroon) lpattern(dash) lwidth(medthick))

use temp_acs_plot.dta, clear

keep if above_avg_white == 1
drop above_avg_density above_avg_income
```

```
drop bachelors_deg_plus white

keep if months_from_opening == -30 | months_from_opening == 30

xtset zipcode timevar, monthly

//Model 2 for >Avg. White (Year FE + County FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using het_estimation_results.xls, keep(treatment_0_5 treatment_5_10 treatment_10_15
treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append ctitle(> Avg.
White) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)

drop house_income pop_density
//NO CONTROLS
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using nocontrol_het_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(> Avg. White) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls,
YES)

***********************************INIT***************************************************
***********************************INIT***************************************************
//OPPOSITE SUBSET CHECKS


//HETEROGENEITY ANALYSIS

//<Avg. Population Density
// use temp_acs_full.dta, clear
use temp_acs_plot.dta, clear

keep if above_avg_density == 0
drop above_avg_income above_avg_white
drop bachelors_deg_plus white

keep if months_from_opening == -30 | months_from_opening == 30

xtset zipcode timevar, monthly

//Model 2 for <Avg. Population Density (Year FE + County FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using opposite_check.xls, keep(treatment_0_5 treatment_5_10 treatment_10_15
treatment_15_20 treat5_post treat10_post treat15_post treat20_post) replace ctitle(< Avg. Pop.
Density) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)

***********************************INIT***************************************************
***********************************INIT***************************************************
//>Avg. Household Income
use temp_acs_plot.dta, clear

keep if above_avg_income == 0
drop above_avg_density above_avg_white
drop bachelors_deg_plus white
```

```
keep if months_from_opening == -30 | months_from_opening == 30

xtset zipcode timevar, monthly

//Model 2 for <Avg. Household Income (Year FE + County FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using opposite_check.xls, keep(treatment_0_5 treatment_5_10 treatment_10_15
treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append ctitle(< Avg.
Income) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)

**********************************INIT**************************************************

//<Avg. WHITE
use temp_acs_plot.dta, clear

keep if above_avg_white == 0
drop above_avg_density above_avg_income
drop bachelors_deg_plus white

keep if months_from_opening == -30 | months_from_opening == 30

xtset zipcode timevar, monthly

//Model 2 for <Avg. White (Year FE + County FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using opposite_check.xls, keep(treatment_0_5 treatment_5_10 treatment_10_15
treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append ctitle(< Avg.
White) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)



**********************************INIT**************************************************
**********************************INIT**************************************************

//FALSIFICATION/PLACEBO TESTS
//(1) 3 years earlier

clear
import delimited "/Users/silas/ECON495/CLOSER/plus3yrs_zhvi_50_complete_gdf.csv"

drop pop bachelors_deg_plus white land_area_sqm

destring log_zhvi, replace force
drop if missing(log_zhvi)

keep if quarters_from_opening >= -10 & quarters_from_opening <= 10
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 63
drop obs_count

keep if months_from_opening == -30 | months_from_opening == 30

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)  // Assuming you want to use Jan 31, 2000 as the
reference date
drop date
```

```
egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both
sort zipcode
encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode countyname, generate(county_num)

drop city countyname
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post

gen treatment_0_5 = (distance_category == "0-5 miles")
gen treatment_5_10 = (distance_category == "5-10 miles")
gen treatment_10_15 = (distance_category == "10-15 miles")
gen treatment_15_20 = (distance_category == "15-20 miles")
gen control = (distance_category == "20-25 miles")

replace treatment_0_5 = 0 if missing(treatment_0_5)
replace treatment_5_10 = 0 if missing(treatment_5_10)
replace treatment_10_15 = 0 if missing(treatment_10_15)
replace treatment_15_20 = 0 if missing(treatment_15_20)
replace control = 0 if missing(control)

gen treat5_post = float(treatment_0_5) * post_treatment
gen treat10_post = float(treatment_5_10) * post_treatment
gen treat15_post = float(treatment_10_15) * post_treatment
gen treat20_post = float(treatment_15_20) * post_treatment
gen control_post = float(control) * post_treatment

replace treat5_post = 0.0 if missing(treat5_post)
replace treat10_post = 0.0 if missing(treat10_post)
replace treat15_post = 0.0 if missing(treat15_post)
replace treat20_post = 0.0 if missing(treat20_post)
replace control_post = 0.0 if missing(control_post)

//Estimate fake openings
xtset zipcode timevar, monthly

//Model 2 (Year FE + County FE + Month FE + County by Year Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using fake_estimation_results.xls, keep(treatment_0_5 treatment_5_10 treatment_10_15
treatment_15_20 treat5_post treat10_post treat15_post treat20_post) replace ctitle(3 yrs)
e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)

**********************************INIT**********************************************

//FALSIFICATION/PLACEBO TESTS
//(2) 2.5 years earlier
clear
import delimited "/Users/silas/ECON495/CLOSER/plus2.5yrs_zhvi_50_complete_gdf.csv"

drop pop bachelors_deg_plus white land_area_sqm

//Clean + Prepare
```

```
destring log_zhvi, replace force
drop if missing(log_zhvi)

keep if quarters_from_opening >= -10 & quarters_from_opening <= 10
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 63
drop obs_count

keep if months_from_opening == -30 | months_from_opening == 30

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)  // Assuming you want to use Jan 31, 2000 as the
reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both
sort zipcode
encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode countyname, generate(county_num)

drop city countyname
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post

gen treatment_0_5 = (distance_category == "0-5 miles")
gen treatment_5_10 = (distance_category == "5-10 miles")
gen treatment_10_15 = (distance_category == "10-15 miles")
gen treatment_15_20 = (distance_category == "15-20 miles")
gen control = (distance_category == "20-25 miles")

replace treatment_0_5 = 0 if missing(treatment_0_5)
replace treatment_5_10 = 0 if missing(treatment_5_10)
replace treatment_10_15 = 0 if missing(treatment_10_15)
replace treatment_15_20 = 0 if missing(treatment_15_20)
replace control = 0 if missing(control)

gen treat5_post = float(treatment_0_5) * post_treatment
gen treat10_post = float(treatment_5_10) * post_treatment
gen treat15_post = float(treatment_10_15) * post_treatment
gen treat20_post = float(treatment_15_20) * post_treatment
gen control_post = float(control) * post_treatment

replace treat5_post = 0.0 if missing(treat5_post)
replace treat10_post = 0.0 if missing(treat10_post)
replace treat15_post = 0.0 if missing(treat15_post)
replace treat20_post = 0.0 if missing(treat20_post)
replace control_post = 0.0 if missing(control_post)

//Estimate fake openings
xtset zipcode timevar, monthly

//Model 2 (Year FE + County FE + Month FE + County by Year Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)
```

```
outreg2 using fake_estimation_results.xls, keep(treatment_0_5 treatment_5_10 treatment_10_15
treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append ctitle(2.5 yrs)
e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)

***********************************INIT*********************************************
//FALSIFICATION/PLACEBO TESTS
//(2) 2 years earlier

clear
import delimited "/Users/silas/ECON495/CLOSER/plus2yrs_zhvi_50_complete_gdf.csv"

drop pop bachelors_deg_plus white land_area_sqm

//Clean + Prepare
destring log_zhvi, replace force
drop if missing(log_zhvi)

keep if quarters_from_opening >= -10 & quarters_from_opening <= 10
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 63
drop obs_count

keep if months_from_opening == -30 | months_from_opening == 30

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)  // Assuming you want to use Jan 31, 2000 as the
reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both
sort zipcode
encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode countyname, generate(county_num)

drop city countyname
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post

gen treatment_0_5 = (distance_category == "0-5 miles")
gen treatment_5_10 = (distance_category == "5-10 miles")
gen treatment_10_15 = (distance_category == "10-15 miles")
gen treatment_15_20 = (distance_category == "15-20 miles")
gen control = (distance_category == "20-25 miles")

replace treatment_0_5 = 0 if missing(treatment_0_5)
replace treatment_5_10 = 0 if missing(treatment_5_10)
replace treatment_10_15 = 0 if missing(treatment_10_15)
replace treatment_15_20 = 0 if missing(treatment_15_20)
replace control = 0 if missing(control)

gen treat5_post = float(treatment_0_5) * post_treatment
gen treat10_post = float(treatment_5_10) * post_treatment
gen treat15_post = float(treatment_10_15) * post_treatment
gen treat20_post = float(treatment_15_20) * post_treatment
gen control_post = float(control) * post_treatment
```

```
replace treat5_post = 0.0 if missing(treat5_post)
replace treat10_post = 0.0 if missing(treat10_post)
replace treat15_post = 0.0 if missing(treat15_post)
replace treat20_post = 0.0 if missing(treat20_post)
replace control_post = 0.0 if missing(control_post)

//Estimate fake openings
xtset zipcode timevar, monthly

//Model 2 (Year FE + County FE + Month FE + County by Year Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using fake_estimation_results.xls, keep(treatment_0_5 treatment_5_10 treatment_10_15
treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append ctitle(2 yrs)
e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES,Controls, YES)


************************************INIT**************************************************
************************************INIT**************************************************
************************************INIT**************************************************
// EVENT STUDY CODE

//CLEANING
clear
cd "/Users/silas/ECON495/PROJ/STATA/Raw_data"

import delimited "/Users/silas/ECON495/CLOSER/zhvi_50_est.csv"

drop if missing(log_zhvi)
drop if log_zhvi == .

duplicates drop store_year_month zipcode, force

sort zipcode quarters_from_opening

keep if quarters_from_opening >= -10 & quarters_from_opening <= 10
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 63
drop obs_count

encode name, generate(county_num)
encode nearest_store, generate(nearest_store_num)
gen date_stata = date(date, "YMD")
drop date

drop geo_id fips_code name state_full city countyname latitude longitude zhvi geometry
nearest_store min_distance_to_store min_distance_to_store_miles bachelors_deg_plus white
land_area_sqm store_year_month pop

save eventstudy.dta, replace
************************************INIT**************************************************
//Cleaning so treated group is only 0-5 miles

use eventstudy.dta, clear

keep if distance_category == "0-5 miles" | distance_category == "20-25 miles"
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post quarters_from_opening
gen treated = (distance_category == "0-5 miles")
replace treated = 0 if distance_category == "20-25 miles"
```

```
gen post_treated = treated * post_treatment
replace post_treated = 0 if missing(post_treated)
drop post_treatment treated
replace months_from_opening = . if post_treated == 0 & distance_category == "20-25 miles"
bysort zipcode: egen event = max(cond(months_from_opening == 0, date_stata, .))
order zipcode date_stata event post_treated months_from_opening

// impute missing ACS 2008 date with 2007 values
sort zipcode year
bysort zipcode (year): egen min_year = min(year)
bysort zipcode (year): egen max_year = max(year)
fillin zipcode year
by zipcode (year): carryforward house_income pop_density, replace
by zipcode: gen temp_house_income2009 = cond(year==2009, house_income, .)
by zipcode: gen temp_pop_density2009 = cond(year==2009, pop_density, .)
by zipcode (year): replace house_income = temp_house_income2009 if year==2008 &
missing(house_income)
by zipcode (year): replace pop_density = temp_pop_density2009 if year==2008 &
missing(pop_density)

drop temp_house_income2009 temp_pop_density2009
drop if missing(min_year) | missing(max_year)
drop min_year max_year _fillin

save eventstudy_1.dta, replace

************************************INIT**********************************************
//Cleaning so treated group is only 5-10 miles

use eventstudy.dta, clear

keep if distance_category == "5-10 miles" | distance_category == "20-25 miles"
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post quarters_from_opening
gen treated = (distance_category == "5-10 miles")
replace treated = 0 if distance_category == "20-25 miles"
gen post_treated = treated * post_treatment
replace post_treated = 0 if missing(post_treated)
drop post_treatment treated
replace months_from_opening = . if post_treated == 0 & distance_category == "20-25 miles"
bysort zipcode: egen event = max(cond(months_from_opening == 0, date_stata, .))
order zipcode date_stata event post_treated months_from_opening

// impute missing ACS 2008 date with 2007 values
sort zipcode year
bysort zipcode (year): egen min_year = min(year)
bysort zipcode (year): egen max_year = max(year)
fillin zipcode year
by zipcode (year): carryforward house_income pop_density, replace
by zipcode: gen temp_house_income2009 = cond(year==2009, house_income, .)
by zipcode: gen temp_pop_density2009 = cond(year==2009, pop_density, .)
by zipcode (year): replace house_income = temp_house_income2009 if year==2008 &
missing(house_income)
by zipcode (year): replace pop_density = temp_pop_density2009 if year==2008 &
missing(pop_density)

drop temp_house_income2009 temp_pop_density2009
drop if missing(min_year) | missing(max_year)
drop min_year max_year _fillin

save eventstudy_2.dta, replace
```

```
************************************INIT**********************************************
//Cleaning so treated group is only 10-15 miles

use eventstudy.dta, clear

keep if distance_category == "10-15 miles" | distance_category == "20-25 miles"
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post quarters_from_opening
gen treated = (distance_category == "10-15 miles")
replace treated = 0 if distance_category == "20-25 miles"
gen post_treated = treated * post_treatment
replace post_treated = 0 if missing(post_treated)
drop post_treatment treated
replace months_from_opening = . if post_treated == 0 & distance_category == "20-25 miles"
bysort zipcode: egen event = max(cond(months_from_opening == 0, date_stata, .))
order zipcode date_stata event post_treated months_from_opening

// impute missing ACS 2008 date with 2007 values
sort zipcode year
bysort zipcode (year): egen min_year = min(year)
bysort zipcode (year): egen max_year = max(year)
fillin zipcode year
by zipcode (year): carryforward house_income pop_density, replace
by zipcode: gen temp_house_income2009 = cond(year==2009, house_income, .)
by zipcode: gen temp_pop_density2009 = cond(year==2009, pop_density, .)
by zipcode (year): replace house_income = temp_house_income2009 if year==2008 &
missing(house_income)
by zipcode (year): replace pop_density = temp_pop_density2009 if year==2008 &
missing(pop_density)

drop temp_house_income2009 temp_pop_density2009
drop if missing(min_year) | missing(max_year)
drop min_year max_year _fillin

save eventstudy_3.dta, replace

************************************INIT**********************************************
//Cleaning so treated group is only 15-20 miles

use eventstudy.dta, clear

keep if distance_category == "15-20 miles" | distance_category == "20-25 miles"
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post quarters_from_opening
gen treated = (distance_category == "15-20 miles")
replace treated = 0 if distance_category == "20-25 miles"
gen post_treated = treated * post_treatment
replace post_treated = 0 if missing(post_treated)
drop post_treatment treated
replace months_from_opening = . if post_treated == 0 & distance_category == "20-25 miles"
bysort zipcode: egen event = max(cond(months_from_opening == 0, date_stata, .))
order zipcode date_stata event post_treated months_from_opening

// impute missing ACS 2008 date with 2007 values
sort zipcode year
bysort zipcode (year): egen min_year = min(year)
bysort zipcode (year): egen max_year = max(year)
fillin zipcode year
by zipcode (year): carryforward house_income pop_density, replace
by zipcode: gen temp_house_income2009 = cond(year==2009, house_income, .)
by zipcode: gen temp_pop_density2009 = cond(year==2009, pop_density, .)
```

```
by zipcode (year): replace house_income = temp_house_income2009 if year==2008 &
missing(house_income)
by zipcode (year): replace pop_density = temp_pop_density2009 if year==2008 &
missing(pop_density)

drop temp_house_income2009 temp_pop_density2009
drop if missing(min_year) | missing(max_year)
drop min_year max_year _fillin

save eventstudy_4.dta, replace

***********************************INIT*********************************************
// EVENT STUDY ESTIMATION
use eventstudy_1.dta, clear
xtset zipcode date_stata

eventdd log_zhvi house_income pop_density i.date_stata i.county_num,
timevar(months_from_opening) method(fe, cluster(nearest_store_num)) leads(30) lags(30)
keepbal(zipcode) graph_op(ytitle("Coefficient") xtitle("Months from Opening") title("Event
Study of 0-5 miles vs. Control group") xlabel(-30(5)30) scheme(s1color))

estat eventdd

test lag30 lag29 lag28 lag27 lag26 lag25 lag24 lag23 lag22 lag21 lag20 lag19 lag18 lag17 lag16
lag15 lag14 lag13 lag12 lag11 lag10 lag9 lag8 lag7 lag6 lag5 lag4 lag3 lag2


use eventstudy_2.dta, clear
xtset zipcode date_stata

eventdd log_zhvi house_income pop_density i.date_stata i.county_num,
timevar(months_from_opening) method(fe, cluster(nearest_store_num)) leads(30) lags(30)
keepbal(zipcode) graph_op(ytitle("Coefficient") xtitle("Months from Opening") title("Event
Study of 5-10 miles vs. Control group") xlabel(-30(5)30) scheme(s1color))

estat eventdd

test lag30 lag29 lag28 lag27 lag26 lag25 lag24 lag23 lag22 lag21 lag20 lag19 lag18 lag17 lag16
lag15 lag14 lag13 lag12 lag11 lag10 lag9 lag8 lag7 lag6 lag5 lag4 lag3 lag2


use eventstudy_3.dta, clear
xtset zipcode date_stata

eventdd log_zhvi house_income pop_density i.date_stata i.county_num,
timevar(months_from_opening) method(fe, cluster(nearest_store_num)) leads(30) lags(30)
keepbal(zipcode) graph_op(ytitle("Coefficient") xtitle("Months from Opening") title("Event
Study of 10-15 miles vs. Control group") xlabel(-30(5)30) scheme(s1color))

estat eventdd

test lag30 lag29 lag28 lag27 lag26 lag25 lag24 lag23 lag22 lag21 lag20 lag19 lag18 lag17 lag16
lag15 lag14 lag13 lag12 lag11 lag10 lag9 lag8 lag7 lag6 lag5 lag4 lag3 lag2


use eventstudy_4.dta, clear
xtset zipcode date_stata

eventdd log_zhvi house_income pop_density i.date_stata i.county_num,
timevar(months_from_opening) method(fe, cluster(nearest_store_num)) leads(30) lags(30)
keepbal(zipcode) graph_op(ytitle("Coefficient") xtitle("Months from Opening") title("Event
Study of 15-20 miles vs. Control group") xlabel(-30(5)30) scheme(s1color))
```

```
estat eventdd

test lag30 lag29 lag28 lag27 lag26 lag25 lag24 lag23 lag22 lag21 lag20 lag19 lag18 lag17 lag16
lag15 lag14 lag13 lag12 lag11 lag10 lag9 lag8 lag7 lag6 lag5 lag4 lag3 lag2

**********************************INIT**********************************************
**********************************INIT**********************************************
**********************************INIT**********************************************
//DIFF-IN-DIFF TEMPORAL ADJUSTMENT

use temp_acs_complete.dta, clear

//DATA VERSION: 3 YEARS (KEEP ZIPCODES W DATA FOR -12 TO 12 QUARTERS_FROM_OPENING)
keep if months_from_opening >= -36 & months_from_opening <= 36
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 73
drop obs_count

keep if months_from_opening == -36 | months_from_opening == 36

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)   // Using Jan 31, 2000 as the reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both

sort zipcode

encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode name, generate(county_num)
encode state_full, generate(state_num)

gen int month_int = month
drop month
rename month_int month

gen store_time = nearest_store + "_" + string(date_stata,"%12.0g")
encode store_time, generate(store_time_num)

order year month, after(control_post)

drop geo_id fips_code name state_full city countyname latitude longitude geometry
min_distance_to_store min_distance_to_store_miles store_year_month date_stata

//Drop Alaskan Costco opening data points
drop if nearest_store == "Fairbanks"

drop zhvi pop land_area_sqm bachelors_deg_plus white

sort zipcode timevar
xtset zipcode timevar, monthly


//Model 2 (Year FE + County FE + Month FE + Controls)
```

```
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using temp1_data_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) replace
ctitle(ln(ZHVI)) e(r2_o) addtext(Year FE, YES, County FE, YES, Month FE, YES, Controls, YES)

***********************************INIT*********************************************

use temp_acs_complete.dta, clear

//DATA VERSION: 2 YEARS (KEEP ZIPCODES W DATA FOR -8 TO 8 QUARTERS_FROM_OPENING)
keep if months_from_opening >= -24 & months_from_opening <= 24
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 49
drop obs_count

keep if months_from_opening == -24 | months_from_opening == 24

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)   // Using Jan 31, 2000 as the reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both

sort zipcode

encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode name, generate(county_num)
encode state_full, generate(state_num)

gen int month_int = month
drop month
rename month_int month

gen store_time = nearest_store + "_" + string(date_stata,"%12.0g")
encode store_time, generate(store_time_num)

order year month, after(control_post)

drop geo_id fips_code name state_full city countyname latitude longitude geometry
min_distance_to_store min_distance_to_store_miles store_year_month date_stata

//Drop Alaskan Costco opening data points
drop if nearest_store == "Fairbanks"

drop zhvi pop land_area_sqm bachelors_deg_plus white

sort zipcode timevar
xtset zipcode timevar, monthly


//Model 2 (Year FE + County FE + Month FE + Controls)
```

```
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using temp1_data_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(ln(ZHVI)) e(r2_o) addtext(Year FE, YES, County FE, YES, Month FE, YES, Controls, YES)


***********************************INIT***********************************************
***********************************INIT***********************************************
***********************************INIT***********************************************

//ANNOUNCEMENT AS TREATMENT

use temp_acs_complete.dta, clear

//DATA VERSION: 2 YEARS (KEEP ZIPCODES W DATA FOR -8 TO 8 QUARTERS_FROM_OPENING)
keep if months_from_opening >= -42 & months_from_opening <= 18
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 61
drop obs_count

keep if months_from_opening == -42 | months_from_opening == 18

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)   // Using Jan 31, 2000 as the reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both

sort zipcode

encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode name, generate(county_num)
encode state_full, generate(state_num)

gen int month_int = month
drop month
rename month_int month

gen store_time = nearest_store + "_" + string(date_stata,"%12.0g")
encode store_time, generate(store_time_num)

order year month, after(control_post)

drop geo_id fips_code name state_full city countyname latitude longitude geometry
min_distance_to_store min_distance_to_store_miles store_year_month date_stata

//Drop Alaskan Costco opening data points
drop if nearest_store == "Fairbanks"

drop zhvi pop land_area_sqm bachelors_deg_plus white

sort zipcode timevar
xtset zipcode timevar, monthly
```

```
//Model 2 (Year FE + County FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using temp1_data_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(ln(ZHVI)) e(r2_o) addtext(Year FE, YES, County FE, YES, Month FE, YES, Controls, YES)


//Model 3 (Year FE + City FE + Month FE + Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.city_num i.month, vce(cluster nearest_store)

outreg2 using all_data_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(ln(ZHVI)) e(r2_o) addtext(Year FE, YES, City FE, YES, Month FE, YES, Controls, YES)


************************************INIT**********************************************
// Heterogeneous Placebo Test

//Heterogeneous FALSIFICATION/PLACEBO TESTS  (POP DENSITY)
//(1) 3 years earlier

clear
import delimited "/Users/silas/ECON495/CLOSER/plus3yrs_zhvi_50_complete_gdf.csv"

// drop pop bachelors_deg_plus white land_area_sqm

destring log_zhvi, replace force
drop if missing(log_zhvi)

//Use 2010 ACS data for heterogeneity by county
egen mean_income_2010 = mean(cond(year == 2010, house_income, .))
egen mean_density_2010 = mean(cond(year == 2010, pop_density, .))
egen mean_white_2010 = mean(cond(year == 2010, white, .))

gen above_avg_income_2010 = cond(year == 2010, house_income > mean_income_2010, .)
gen above_avg_density_2010 = cond(year == 2010, pop_density > mean_density_2010, .)
gen above_avg_white_2010 = cond(year == 2010, white > mean_white_2010, .)

egen max_above_avg_income_2010 = max(above_avg_income_2010), by(zipcode)
egen max_above_avg_density_2010 = max(above_avg_density_2010), by(zipcode)
egen max_above_avg_white_2010 = max(above_avg_white_2010), by(zipcode)

replace above_avg_income_2010 = max_above_avg_income_2010
replace above_avg_density_2010 = max_above_avg_density_2010
replace above_avg_white_2010 = max_above_avg_white_2010

rename above_avg_income_2010 above_avg_income
rename above_avg_density_2010 above_avg_density
rename above_avg_white_2010 above_avg_white

drop max_above_avg_income_2010 max_above_avg_density_2010 max_above_avg_white_2010

keep if above_avg_density == 1
drop above_avg_income above_avg_white
drop bachelors_deg_plus white
```

```
keep if quarters_from_opening >= -10 & quarters_from_opening <= 10
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 63
drop obs_count

keep if months_from_opening == -30 | months_from_opening == 30

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)  // Assuming you want to use Jan 31, 2000 as the
reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both
sort zipcode
encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode countyname, generate(county_num)

drop city countyname
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post

gen treatment_0_5 = (distance_category == "0-5 miles")
gen treatment_5_10 = (distance_category == "5-10 miles")
gen treatment_10_15 = (distance_category == "10-15 miles")
gen treatment_15_20 = (distance_category == "15-20 miles")
gen control = (distance_category == "20-25 miles")

replace treatment_0_5 = 0 if missing(treatment_0_5)
replace treatment_5_10 = 0 if missing(treatment_5_10)
replace treatment_10_15 = 0 if missing(treatment_10_15)
replace treatment_15_20 = 0 if missing(treatment_15_20)
replace control = 0 if missing(control)

gen treat5_post = float(treatment_0_5) * post_treatment
gen treat10_post = float(treatment_5_10) * post_treatment
gen treat15_post = float(treatment_10_15) * post_treatment
gen treat20_post = float(treatment_15_20) * post_treatment
gen control_post = float(control) * post_treatment

replace treat5_post = 0.0 if missing(treat5_post)
replace treat10_post = 0.0 if missing(treat10_post)
replace treat15_post = 0.0 if missing(treat15_post)
replace treat20_post = 0.0 if missing(treat20_post)
replace control_post = 0.0 if missing(control_post)

//Estimate fake openings
xtset zipcode timevar, monthly

//Model 2 (Year FE + County FE + Month FE + County by Year Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using het_fake_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) replace
ctitle(3 yrs) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)
```

```
************************************INIT**********************************************

//Heterogeneous FALSIFICATION/PLACEBO TESTS  (POP DENSITY)
//(2) 2.5 years earlier
clear
import delimited "/Users/silas/ECON495/CLOSER/plus2.5yrs_zhvi_50_complete_gdf.csv"

// drop pop bachelors_deg_plus white land_area_sqm

//Clean + Prepare
destring log_zhvi, replace force
drop if missing(log_zhvi)

//Use 2010 ACS data for heterogeneity by county
egen mean_income_2010 = mean(cond(year == 2010, house_income, .))
egen mean_density_2010 = mean(cond(year == 2010, pop_density, .))
egen mean_white_2010 = mean(cond(year == 2010, white, .))

gen above_avg_income_2010 = cond(year == 2010, house_income > mean_income_2010, .)
gen above_avg_density_2010 = cond(year == 2010, pop_density > mean_density_2010, .)
gen above_avg_white_2010 = cond(year == 2010, white > mean_white_2010, .)

egen max_above_avg_income_2010 = max(above_avg_income_2010), by(zipcode)
egen max_above_avg_density_2010 = max(above_avg_density_2010), by(zipcode)
egen max_above_avg_white_2010 = max(above_avg_white_2010), by(zipcode)

replace above_avg_income_2010 = max_above_avg_income_2010
replace above_avg_density_2010 = max_above_avg_density_2010
replace above_avg_white_2010 = max_above_avg_white_2010

rename above_avg_income_2010 above_avg_income
rename above_avg_density_2010 above_avg_density
rename above_avg_white_2010 above_avg_white

drop max_above_avg_income_2010 max_above_avg_density_2010 max_above_avg_white_2010

keep if above_avg_density == 1
drop above_avg_income above_avg_white
drop bachelors_deg_plus white

keep if quarters_from_opening >= -10 & quarters_from_opening <= 10
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 63
drop obs_count

keep if months_from_opening == -30 | months_from_opening == 30

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)  // Assuming you want to use Jan 31, 2000 as the
reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both
sort zipcode
encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
```

```
encode countyname, generate(county_num)

drop city countyname
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post

gen treatment_0_5 = (distance_category == "0-5 miles")
gen treatment_5_10 = (distance_category == "5-10 miles")
gen treatment_10_15 = (distance_category == "10-15 miles")
gen treatment_15_20 = (distance_category == "15-20 miles")
gen control = (distance_category == "20-25 miles")

replace treatment_0_5 = 0 if missing(treatment_0_5)
replace treatment_5_10 = 0 if missing(treatment_5_10)
replace treatment_10_15 = 0 if missing(treatment_10_15)
replace treatment_15_20 = 0 if missing(treatment_15_20)
replace control = 0 if missing(control)

gen treat5_post = float(treatment_0_5) * post_treatment
gen treat10_post = float(treatment_5_10) * post_treatment
gen treat15_post = float(treatment_10_15) * post_treatment
gen treat20_post = float(treatment_15_20) * post_treatment
gen control_post = float(control) * post_treatment

replace treat5_post = 0.0 if missing(treat5_post)
replace treat10_post = 0.0 if missing(treat10_post)
replace treat15_post = 0.0 if missing(treat15_post)
replace treat20_post = 0.0 if missing(treat20_post)
replace control_post = 0.0 if missing(control_post)

//Estimate fake openings
xtset zipcode timevar, monthly

//Model 2 (Year FE + County FE + Month FE + County by Year Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using het_fake_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(2.5 yrs) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)

************************************INIT*********************************************
//Heterogeneous FALSIFICATION/PLACEBO TESTS  (POP DENSITY)
//(2) 2 years earlier

clear
import delimited "/Users/silas/ECON495/CLOSER/plus2yrs_zhvi_50_complete_gdf.csv"

// drop pop bachelors_deg_plus white land_area_sqm

//Clean + Prepare
destring log_zhvi, replace force
drop if missing(log_zhvi)

//Use 2010 ACS data for heterogeneity by county
egen mean_income_2010 = mean(cond(year == 2010, house_income, .))
egen mean_density_2010 = mean(cond(year == 2010, pop_density, .))
egen mean_white_2010 = mean(cond(year == 2010, white, .))

gen above_avg_income_2010 = cond(year == 2010, house_income > mean_income_2010, .)
gen above_avg_density_2010 = cond(year == 2010, pop_density > mean_density_2010, .)
```

```
gen above_avg_white_2010 = cond(year == 2010, white > mean_white_2010, .)

egen max_above_avg_income_2010 = max(above_avg_income_2010), by(zipcode)
egen max_above_avg_density_2010 = max(above_avg_density_2010), by(zipcode)
egen max_above_avg_white_2010 = max(above_avg_white_2010), by(zipcode)

replace above_avg_income_2010 = max_above_avg_income_2010
replace above_avg_density_2010 = max_above_avg_density_2010
replace above_avg_white_2010 = max_above_avg_white_2010

rename above_avg_income_2010 above_avg_income
rename above_avg_density_2010 above_avg_density
rename above_avg_white_2010 above_avg_white

drop max_above_avg_income_2010 max_above_avg_density_2010 max_above_avg_white_2010

keep if above_avg_density == 1
drop above_avg_income above_avg_white
drop bachelors_deg_plus white

keep if quarters_from_opening >= -10 & quarters_from_opening <= 10
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 63
drop obs_count

keep if months_from_opening == -30 | months_from_opening == 30

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)  // Assuming you want to use Jan 31, 2000 as the
reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both
sort zipcode
encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode countyname, generate(county_num)

drop city countyname
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post

gen treatment_0_5 = (distance_category == "0-5 miles")
gen treatment_5_10 = (distance_category == "5-10 miles")
gen treatment_10_15 = (distance_category == "10-15 miles")
gen treatment_15_20 = (distance_category == "15-20 miles")
gen control = (distance_category == "20-25 miles")

replace treatment_0_5 = 0 if missing(treatment_0_5)
replace treatment_5_10 = 0 if missing(treatment_5_10)
replace treatment_10_15 = 0 if missing(treatment_10_15)
replace treatment_15_20 = 0 if missing(treatment_15_20)
replace control = 0 if missing(control)

gen treat5_post = float(treatment_0_5) * post_treatment
gen treat10_post = float(treatment_5_10) * post_treatment
gen treat15_post = float(treatment_10_15) * post_treatment
```

```
gen treat20_post = float(treatment_15_20) * post_treatment
gen control_post = float(control) * post_treatment

replace treat5_post = 0.0 if missing(treat5_post)
replace treat10_post = 0.0 if missing(treat10_post)
replace treat15_post = 0.0 if missing(treat15_post)
replace treat20_post = 0.0 if missing(treat20_post)
replace control_post = 0.0 if missing(control_post)

//Estimate fake openings
xtset zipcode timevar, monthly

//Model 2 (Year FE + County FE + Month FE + County by Year Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using het_fake_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(2 yrs) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES,Controls, YES)

************************************INIT**********************************************
// Heterogeneous Placebo Test

//Heterogeneous FALSIFICATION/PLACEBO TESTS (INCOME)
//(1) 3 years earlier

clear
import delimited "/Users/silas/ECON495/CLOSER/plus3yrs_zhvi_50_complete_gdf.csv"

// drop pop bachelors_deg_plus white land_area_sqm

destring log_zhvi, replace force
drop if missing(log_zhvi)

//Use 2010 ACS data for heterogeneity by county
egen mean_income_2010 = mean(cond(year == 2010, house_income, .))
egen mean_density_2010 = mean(cond(year == 2010, pop_density, .))
egen mean_white_2010 = mean(cond(year == 2010, white, .))

gen above_avg_income_2010 = cond(year == 2010, house_income > mean_income_2010, .)
gen above_avg_density_2010 = cond(year == 2010, pop_density > mean_density_2010, .)
gen above_avg_white_2010 = cond(year == 2010, white > mean_white_2010, .)

egen max_above_avg_income_2010 = max(above_avg_income_2010), by(zipcode)
egen max_above_avg_density_2010 = max(above_avg_density_2010), by(zipcode)
egen max_above_avg_white_2010 = max(above_avg_white_2010), by(zipcode)

replace above_avg_income_2010 = max_above_avg_income_2010
replace above_avg_density_2010 = max_above_avg_density_2010
replace above_avg_white_2010 = max_above_avg_white_2010

rename above_avg_income_2010 above_avg_income
rename above_avg_density_2010 above_avg_density
rename above_avg_white_2010 above_avg_white

drop max_above_avg_income_2010 max_above_avg_density_2010 max_above_avg_white_2010

keep if above_avg_income == 1
drop above_avg_density above_avg_white
drop bachelors_deg_plus white
```

```
keep if quarters_from_opening >= -10 & quarters_from_opening <= 10
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 63
drop obs_count

keep if months_from_opening == -30 | months_from_opening == 30

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)  // Assuming you want to use Jan 31, 2000 as the
reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both
sort zipcode
encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode countyname, generate(county_num)

drop city countyname
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post

gen treatment_0_5 = (distance_category == "0-5 miles")
gen treatment_5_10 = (distance_category == "5-10 miles")
gen treatment_10_15 = (distance_category == "10-15 miles")
gen treatment_15_20 = (distance_category == "15-20 miles")
gen control = (distance_category == "20-25 miles")

replace treatment_0_5 = 0 if missing(treatment_0_5)
replace treatment_5_10 = 0 if missing(treatment_5_10)
replace treatment_10_15 = 0 if missing(treatment_10_15)
replace treatment_15_20 = 0 if missing(treatment_15_20)
replace control = 0 if missing(control)

gen treat5_post = float(treatment_0_5) * post_treatment
gen treat10_post = float(treatment_5_10) * post_treatment
gen treat15_post = float(treatment_10_15) * post_treatment
gen treat20_post = float(treatment_15_20) * post_treatment
gen control_post = float(control) * post_treatment

replace treat5_post = 0.0 if missing(treat5_post)
replace treat10_post = 0.0 if missing(treat10_post)
replace treat15_post = 0.0 if missing(treat15_post)
replace treat20_post = 0.0 if missing(treat20_post)
replace control_post = 0.0 if missing(control_post)

//Estimate fake openings
xtset zipcode timevar, monthly

//Model 2 (Year FE + County FE + Month FE + County by Year Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using het2_fake_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) replace
ctitle(3 yrs) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)
```

```
**********************************INIT**********************************************

//Heterogeneous FALSIFICATION/PLACEBO TESTS  (INCOME)
//(2) 2.5 years earlier
clear
import delimited "/Users/silas/ECON495/CLOSER/plus2.5yrs_zhvi_50_complete_gdf.csv"

// drop pop bachelors_deg_plus white land_area_sqm

//Clean + Prepare
destring log_zhvi, replace force
drop if missing(log_zhvi)

//Use 2010 ACS data for heterogeneity by county
egen mean_income_2010 = mean(cond(year == 2010, house_income, .))
egen mean_density_2010 = mean(cond(year == 2010, pop_density, .))
egen mean_white_2010 = mean(cond(year == 2010, white, .))

gen above_avg_income_2010 = cond(year == 2010, house_income > mean_income_2010, .)
gen above_avg_density_2010 = cond(year == 2010, pop_density > mean_density_2010, .)
gen above_avg_white_2010 = cond(year == 2010, white > mean_white_2010, .)

egen max_above_avg_income_2010 = max(above_avg_income_2010), by(zipcode)
egen max_above_avg_density_2010 = max(above_avg_density_2010), by(zipcode)
egen max_above_avg_white_2010 = max(above_avg_white_2010), by(zipcode)

replace above_avg_income_2010 = max_above_avg_income_2010
replace above_avg_density_2010 = max_above_avg_density_2010
replace above_avg_white_2010 = max_above_avg_white_2010

rename above_avg_income_2010 above_avg_income
rename above_avg_density_2010 above_avg_density
rename above_avg_white_2010 above_avg_white

drop max_above_avg_income_2010 max_above_avg_density_2010 max_above_avg_white_2010

keep if above_avg_income == 1
drop above_avg_density above_avg_white
drop bachelors_deg_plus white

keep if quarters_from_opening >= -10 & quarters_from_opening <= 10
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 63
drop obs_count

keep if months_from_opening == -30 | months_from_opening == 30

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)  // Assuming you want to use Jan 31, 2000 as the
reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both
sort zipcode
encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
```

```
encode countyname, generate(county_num)

drop city countyname
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post

gen treatment_0_5 = (distance_category == "0-5 miles")
gen treatment_5_10 = (distance_category == "5-10 miles")
gen treatment_10_15 = (distance_category == "10-15 miles")
gen treatment_15_20 = (distance_category == "15-20 miles")
gen control = (distance_category == "20-25 miles")

replace treatment_0_5 = 0 if missing(treatment_0_5)
replace treatment_5_10 = 0 if missing(treatment_5_10)
replace treatment_10_15 = 0 if missing(treatment_10_15)
replace treatment_15_20 = 0 if missing(treatment_15_20)
replace control = 0 if missing(control)

gen treat5_post = float(treatment_0_5) * post_treatment
gen treat10_post = float(treatment_5_10) * post_treatment
gen treat15_post = float(treatment_10_15) * post_treatment
gen treat20_post = float(treatment_15_20) * post_treatment
gen control_post = float(control) * post_treatment

replace treat5_post = 0.0 if missing(treat5_post)
replace treat10_post = 0.0 if missing(treat10_post)
replace treat15_post = 0.0 if missing(treat15_post)
replace treat20_post = 0.0 if missing(treat20_post)
replace control_post = 0.0 if missing(control_post)

//Estimate fake openings
xtset zipcode timevar, monthly

//Model 2 (Year FE + County FE + Month FE + County by Year Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using het2_fake_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(2.5 yrs) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES, Controls, YES)

************************************INIT*********************************************
//Heterogeneous FALSIFICATION/PLACEBO TESTS  (INCOME)
//(2) 2 years earlier

clear
import delimited "/Users/silas/ECON495/CLOSER/plus2yrs_zhvi_50_complete_gdf.csv"

// drop pop bachelors_deg_plus white land_area_sqm

//Clean + Prepare
destring log_zhvi, replace force
drop if missing(log_zhvi)

//Use 2010 ACS data for heterogeneity by county
egen mean_income_2010 = mean(cond(year == 2010, house_income, .))
egen mean_density_2010 = mean(cond(year == 2010, pop_density, .))
egen mean_white_2010 = mean(cond(year == 2010, white, .))

gen above_avg_income_2010 = cond(year == 2010, house_income > mean_income_2010, .)
gen above_avg_density_2010 = cond(year == 2010, pop_density > mean_density_2010, .)
```

```
gen above_avg_white_2010 = cond(year == 2010, white > mean_white_2010, .)

egen max_above_avg_income_2010 = max(above_avg_income_2010), by(zipcode)
egen max_above_avg_density_2010 = max(above_avg_density_2010), by(zipcode)
egen max_above_avg_white_2010 = max(above_avg_white_2010), by(zipcode)

replace above_avg_income_2010 = max_above_avg_income_2010
replace above_avg_density_2010 = max_above_avg_density_2010
replace above_avg_white_2010 = max_above_avg_white_2010

rename above_avg_income_2010 above_avg_income
rename above_avg_density_2010 above_avg_density
rename above_avg_white_2010 above_avg_white

drop max_above_avg_income_2010 max_above_avg_density_2010 max_above_avg_white_2010

keep if above_avg_income == 1
drop above_avg_density above_avg_white
drop bachelors_deg_plus white

keep if quarters_from_opening >= -10 & quarters_from_opening <= 10
bysort zipcode distance_category: egen obs_count = count(quarters_from_opening)
keep if obs_count >= 63
drop obs_count

keep if months_from_opening == -30 | months_from_opening == 30

gen date_stata = date(date, "YMD")
gen timevar = date_stata - mdy(1, 31, 2000)  // Assuming you want to use Jan 31, 2000 as the
reference date
drop date

egen tag1 = total(post_treatment == 1), by(zipcode)
egen tag0 = total(post_treatment == 0), by(zipcode)
gen both = tag1 & tag0
keep if both == 1
drop tag1 tag0 both
sort zipcode
encode nearest_store, generate(nearest_store_num)
encode store_year_month, generate(store_year_month_num)
encode city, generate(city_num)
encode countyname, generate(county_num)

drop city countyname
drop treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 control treat5_post
treat10_post treat15_post treat20_post control_post

gen treatment_0_5 = (distance_category == "0-5 miles")
gen treatment_5_10 = (distance_category == "5-10 miles")
gen treatment_10_15 = (distance_category == "10-15 miles")
gen treatment_15_20 = (distance_category == "15-20 miles")
gen control = (distance_category == "20-25 miles")

replace treatment_0_5 = 0 if missing(treatment_0_5)
replace treatment_5_10 = 0 if missing(treatment_5_10)
replace treatment_10_15 = 0 if missing(treatment_10_15)
replace treatment_15_20 = 0 if missing(treatment_15_20)
replace control = 0 if missing(control)

gen treat5_post = float(treatment_0_5) * post_treatment
gen treat10_post = float(treatment_5_10) * post_treatment
gen treat15_post = float(treatment_10_15) * post_treatment
```

```
gen treat20_post = float(treatment_15_20) * post_treatment
gen control_post = float(control) * post_treatment

replace treat5_post = 0.0 if missing(treat5_post)
replace treat10_post = 0.0 if missing(treat10_post)
replace treat15_post = 0.0 if missing(treat15_post)
replace treat20_post = 0.0 if missing(treat20_post)
replace control_post = 0.0 if missing(control_post)

//Estimate fake openings
xtset zipcode timevar, monthly

//Model 2 (Year FE + County FE + Month FE + County by Year Controls)
xtreg log_zhvi treatment_0_5 treatment_5_10 treatment_10_15 treatment_15_20 treat5_post
treat10_post treat15_post treat20_post house_income pop_density year##c.house_income
year##c.pop_density i.year i.county_num i.month, vce(cluster nearest_store)

outreg2 using het2_fake_estimation_results.xls, keep(treatment_0_5 treatment_5_10
treatment_10_15 treatment_15_20 treat5_post treat10_post treat15_post treat20_post) append
ctitle(2 yrs) e(r2_o) addtext(Year FE, YES, Month FE, YES, County FE, YES,Controls, YES)
```