Understanding Consequences of Higher Marginal Property Tax Rate

Misbah Arshad

Abstract An NGO raised concerns about the fairness of tax contributions, specifically whether wealthier households pay their fair share. To explore whether a higher marginal tax rate leads to higher rates of tax evasion, a regression discontinuity (RD) analysis was conducted. The analysis found that higher marginal tax rates are linked to higher rates of tax evasion.

Method To assess whether higher tax rates contribute to increased tax evasion, a Regression Discontinuity (RD) design was employed, a quasi-experimental method that exploits sharp discontinuities in the tax policy. The analysis was conducted across three counties, utilizing three separate datasets to enhance the robustness of the results through replication. Of these counties, two met the validity criteria required for a valid RD design. As a result, a fuzzy RD analysis was performed on one dataset and a sharp RD on the other. The third dataset, which did not pass the validity tests, was excluded from the RD analysis but retained for further exploration of the policy implications.

Data The analysis utilized a simulated dataset representing three counties—Alameda, Santa Clara, and YOLO—each characterized by different tax rate structures. In this dataset, homes purchased before 1978 are subject to a lower marginal tax rate of 15%, while homes purchased after 1978 face a higher marginal tax rate of 50%, in accordance with California's Proposition 13 tax law. The discontinuity in tax rates based on the purchase date of properties provides the basis for the regression discontinuity design.

Regression Discontinuity Validity Tests: Bunching and Smoothness

It is not appropriate to simply compare the difference in tax evasion rates between high and low marginal tax rates to determine if higher marginal tax rates cause increased tax evasion, as there may be an omitted variable that has not been accounted for in the difference.

A suitable research design for this analysis is regression discontinuity (RD), using 1978 as the cutoff. By zooming in on the observations near the cutoff, it is possible to measure the difference in tax evasion rates as individuals cross from a low to a high marginal tax rate. This approach helps determine whether the higher tax rate itself is responsible for the observed increase in tax evasion.

The identifying assumption for an RD is that all observed and unobserved determinants of tax evasion, aside from the tax rate change from 15% to 50%, are continuous around the cutoff (i.e., all covariates remain smooth between the pre-treatment and post-treatment periods).

Mathematically: $E[Yi(1)|Xi=x]$ and $E[Yi(0)|Xi=x]$ are continuous at $x = c$

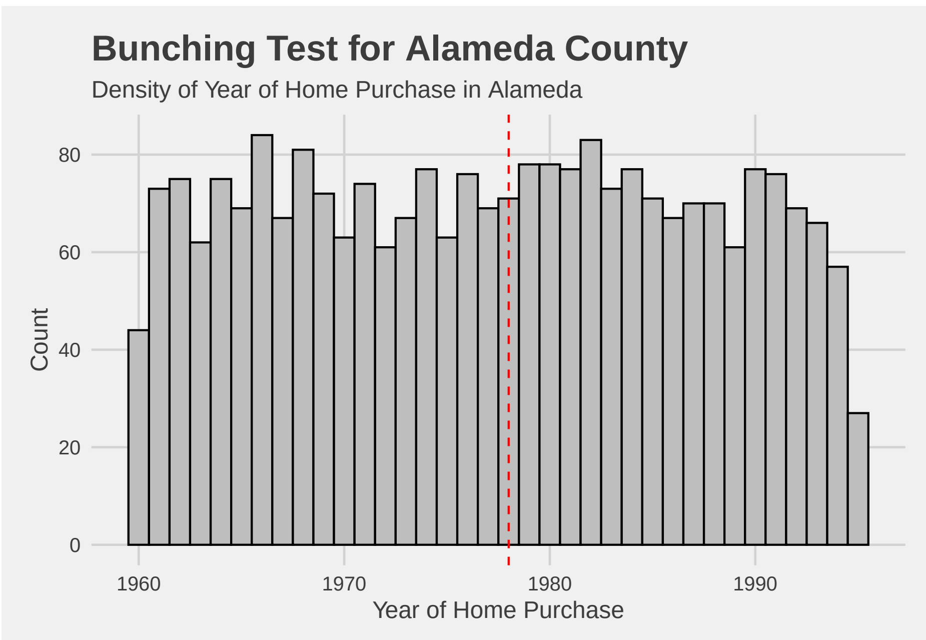
The dataset will be separated by county to evaluate each case individually.

```
county_A <- data %>% filter(county == "ALAMEDA")
county_B <- data %>% filter(county == "SANTA CLARA")
county_C <- data %>% filter(county == "YOLO")
```

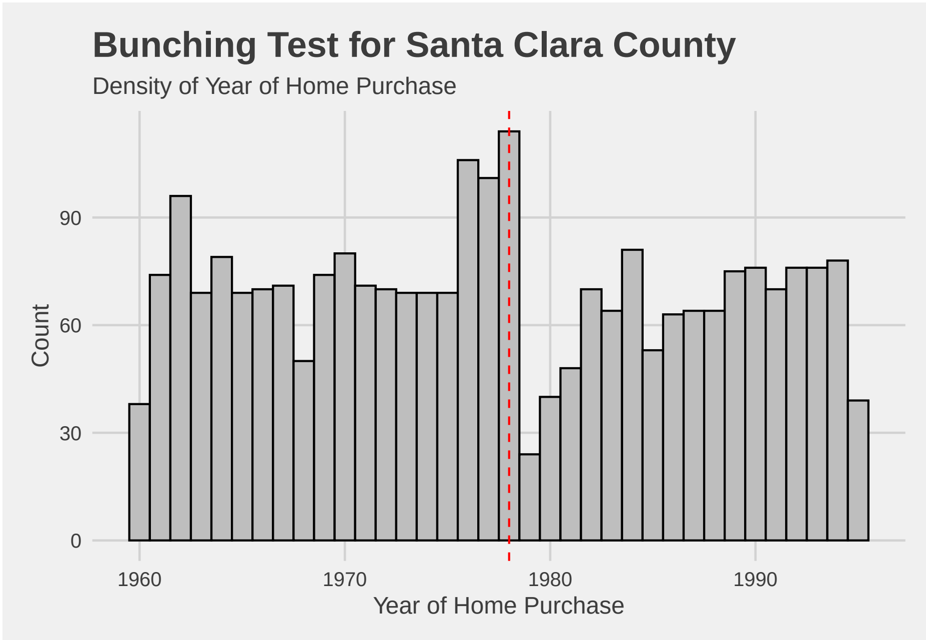
Bunching Test: Alameda and YOLO County pass the bunching test for the regression discontinuity (RD) to be valid, as there is no bunching observed around the cutoff. However, Santa Clara County does not pass the test due to a sharp decline in the number of homes purchased following the cutoff.

As a result, the Santa Clara dataset (county B) is invalid for RD analysis. Nevertheless, this finding provides valuable insights into the response of individuals to changes in the tax structure. The significant reduction in home purchases following the tax rate increase suggests a behavioral response to the policy change. This contrast with the other counties raises the possibility that residents in Santa Clara had prior knowledge of the tax change, enabling them to adjust their behavior, whereas residents in Alameda and YOLO Counties were not informed in advance, which may explain why their data sets remain valid for RD analysis.

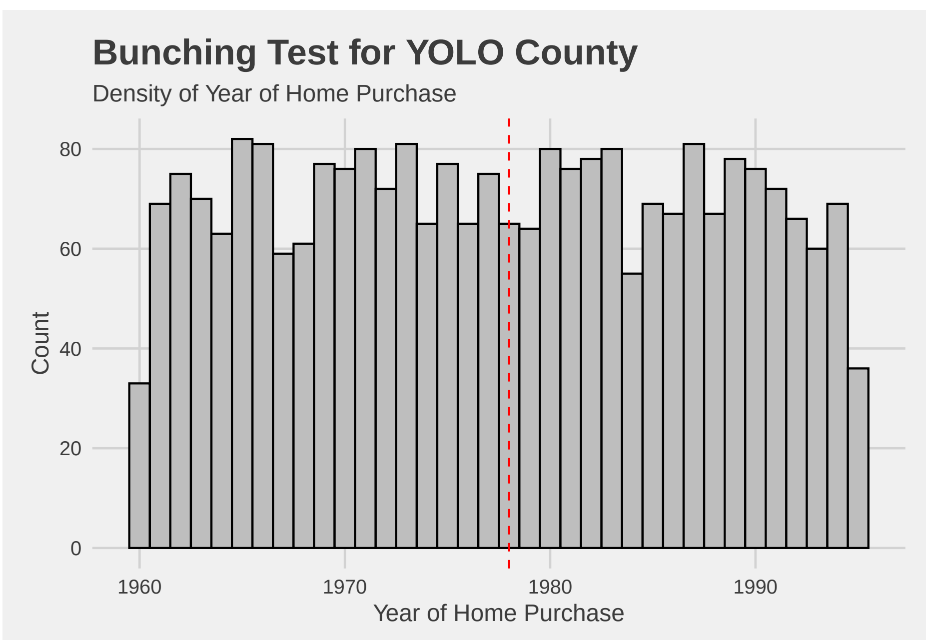
```r
cutoff_year = 1978

ggplot(county_A, aes(x = year_of_home_purchase)) +
  geom_histogram(binwidth = 1, color = "black", fill = "gray") +
  geom_vline(xintercept = cutoff_year, linetype = "dashed", color = "red") +
  labs(title = "Bunching Test for Alameda County",
      subtitle = "Density of Year of Home Purchase in Alameda",
       x = "Year of Home Purchase", y = "Count") +
  theme_fivethirtyeight() +
  theme(axis.title = element_text())
```




```r
ggplot(county_B, aes(x = year_of_home_purchase)) +
  geom_histogram(binwidth = 1, color = "black", fill = "gray") +
  geom_vline(xintercept = cutoff_year, linetype = "dashed", color = "red") +
  labs(title = "Bunching Test for Santa Clara County",
      subtitle = "Density of Year of Home Purchase",
       x = "Year of Home Purchase", y = "Count") +
  theme_fivethirtyeight() +
  theme(axis.title = element_text())
```




```r
ggplot(county_C, aes(x = year_of_home_purchase)) +
  geom_histogram(binwidth = 1, color = "black", fill = "gray") +
  geom_vline(xintercept = cutoff_year, linetype = "dashed", color = "red") +
  labs(title = "Bunching Test for YOLO County",
      subtitle = "Density of Year of Home Purchase",
       x = "Year of Home Purchase", y = "Count") +
  theme_fivethirtyeight() +
  theme(axis.title = element_text())
```



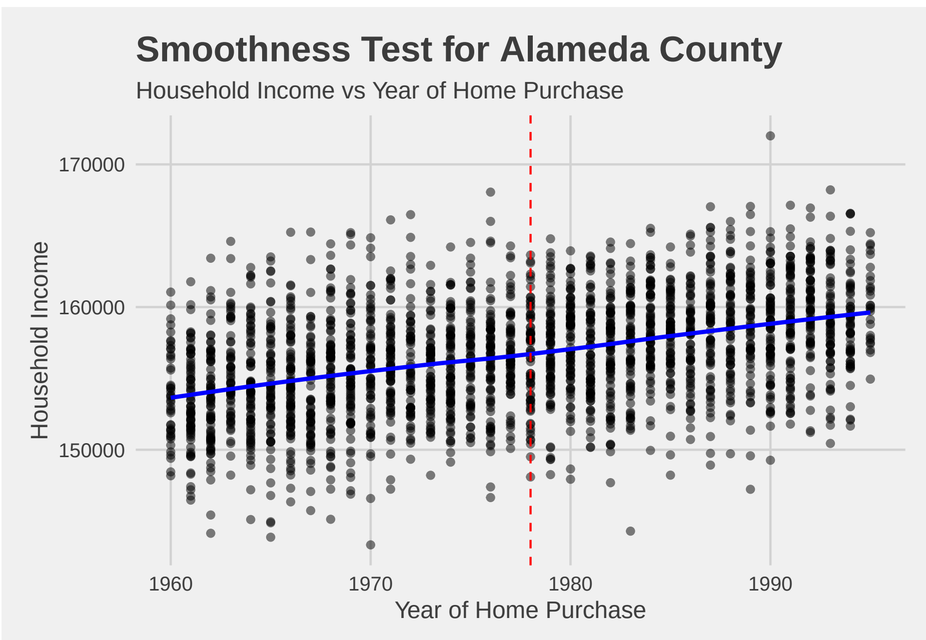
Smoothness Test: The dataset also includes household income as a variable, necessitating a smoothness test to ensure its continuity around the cutoff for the regression discontinuity (RD) to be valid.

Both Alameda and YOLO Counties demonstrate smoothness around the cutoff, with no significant jump or drop in household income at the 1978 threshold that would interfere with the validity of the RD analysis. In contrast, Santa Clara County does not pass the smoothness test. A noticeable increase in household income after the tax change cutoff could introduce confounding factors, potentially compromising the ability to attribute changes in tax evasion rates solely to the marginal tax rate increase.

As Santa Clara fails both the smoothness and bunching tests, it is excluded from the RD analysis to estimate the effect of a higher tax rate on tax evasion rates. However, the observed changes in home purchases and household income around the cutoff in Santa Clara may provide valuable insights into how behavior adjusts in response to changes in tax policy for futher research.

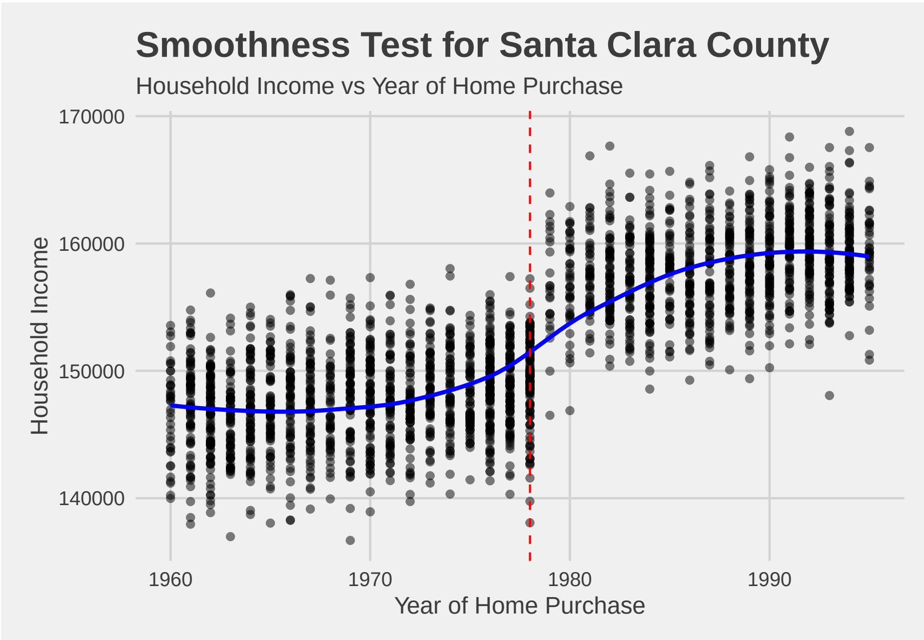
```
ggplot(county_A, aes(x = year_of_home_purchase, y = household_income)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", se = FALSE, color = "blue") +
  labs(title = "Smoothness Test for Alameda County",
  subtitle = "Household Income vs Year of Home Purchase",
       x = "Year of Home Purchase", y = "Household Income") +
  geom_vline(xintercept = 1978, linetype = "dashed", color = "red") +
  theme_fivethirtyeight() +
  theme(axis.title = element_text())
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



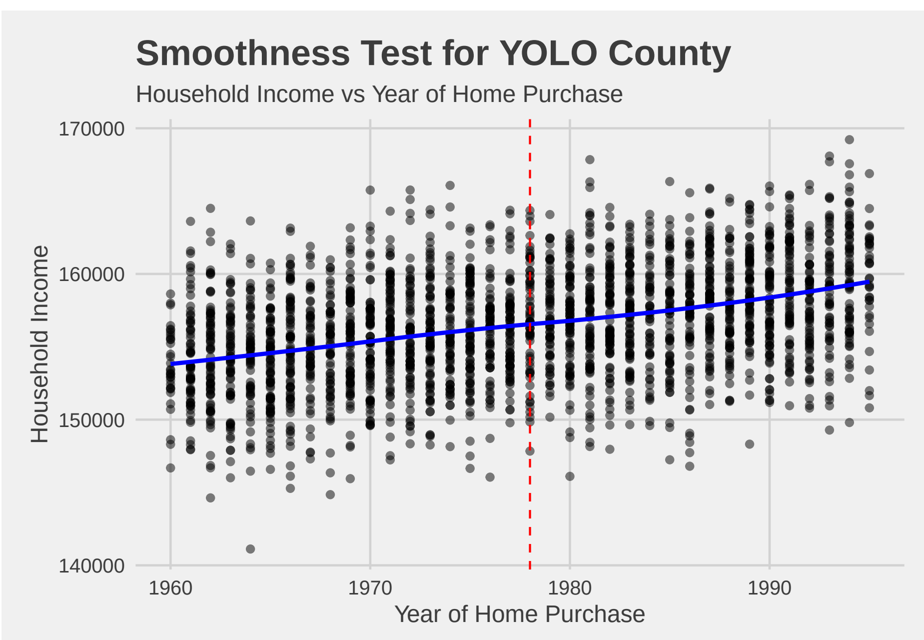

```
ggplot(county_B, aes(x = year_of_home_purchase, y = household_income)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", se = FALSE, color = "blue") +
  labs(title = "Smoothness Test for Santa Clara County",
  subtitle = "Household Income vs Year of Home Purchase",
       x = "Year of Home Purchase", y = "Household Income") +
  geom_vline(xintercept = 1978, linetype = "dashed", color = "red") +
  theme_fivethirtyeight() +
  theme(axis.title = element_text())
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
ggplot(county_C, aes(x = year_of_home_purchase, y = household_income)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", se = FALSE, color = "blue") +
  labs(title = "Smoothness Test for YOLO County",
       subtitle = "Household Income vs Year of Home Purchase",
       x = "Year of Home Purchase", y = "Household Income") +
  geom_vline(xintercept = 1978, linetype = "dashed", color = "red") +
  theme_fivethirtyeight() +
  theme(axis.title = element_text())
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Results: Regression Discontinuity

The first step is to create a dummy variable for the high marginal tax rate to convert all households having a 50% or higher marginal tax rate to 1 and all households below 50% to 0.

```
county_A <- county_A %>%
  mutate(high_tax_dummy = ifelse(marginal_property_tax_rate >= 50, 1, 0))

county_C <- county_C %>%
  mutate(high_tax_dummy = ifelse(marginal_property_tax_rate >= 50, 1, 0))
```

Aggregate:

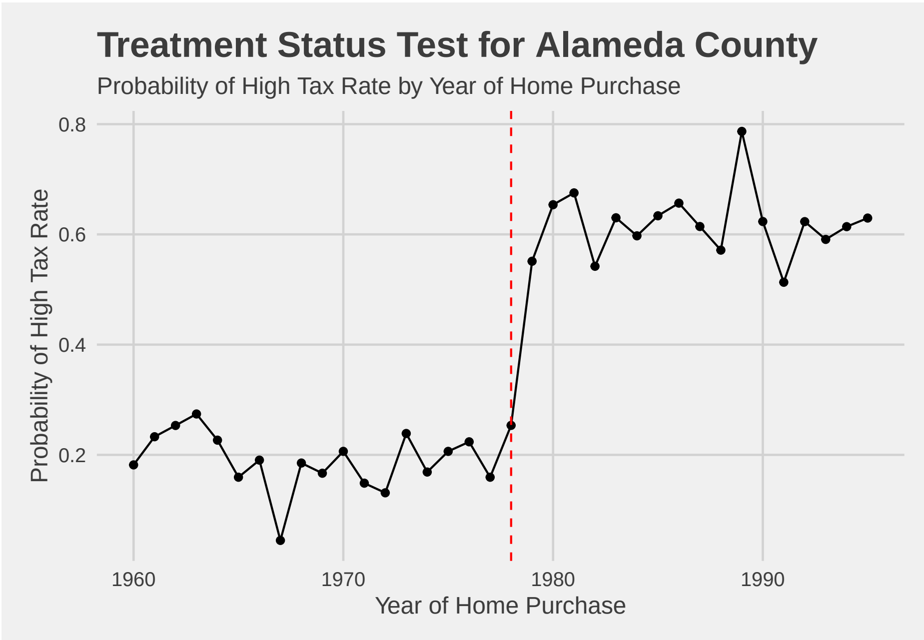
```
tax_rate_by_year_A <- county_A %>%
  group_by(year_of_home_purchase) %>%
  summarise(prob_high_tax = mean(high_tax_dummy, na.rm = TRUE))

tax_rate_by_year_C <- county_C %>%
  group_by(year_of_home_purchase) %>%
  summarise(prob_high_tax = mean(high_tax_dummy, na.rm = TRUE))
```

Treatment Status Test To conduct a sharp RD, the plot should show a distinct jump from 0% to 100% without variation. This is measured by a treatment status test by plotting the relationship between the probability of a high tax rate and year of home purchase to ensure there is a no variation before and after.

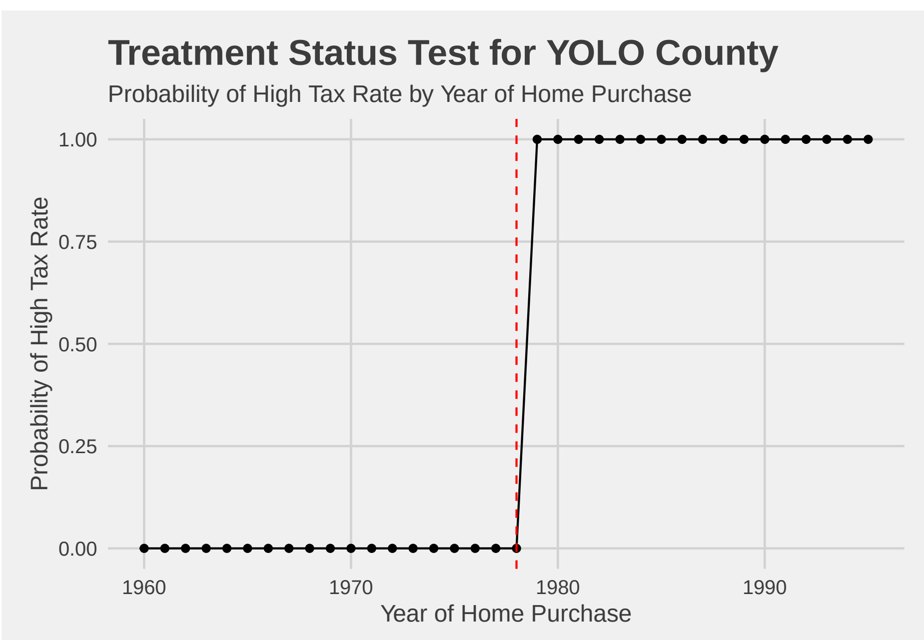
In Alameda County, the cutoff affects the probability of receiving treatment but does not guarantee it. Since there is some variation in the tax rates before and after the cutoff, Alameda County requires a fuzzy RD instead of a sharp RD.

In YOLO County, the cutoff completely determines whether treatment is received (i.e. the treatment status jumps from 0% to 100%) and therefore can be used for a sharp RD.

```
ggplot(tax_rate_by_year_A, aes(x = year_of_home_purchase, y = prob_high_tax)) +
  geom_point() +
  geom_line() +
  labs(title = "Treatment Status Test for Alameda County",
       subtitle = "Probability of High Tax Rate by Year of Home Purchase",
        x = "Year of Home Purchase", y = "Probability of High Tax Rate") +
  geom_vline(xintercept = 1978, linetype = "dashed", color = "red") +
  theme_fivethirtyeight() +
  theme(axis.title = element_text())
```




```r
ggplot(tax_rate_by_year_C, aes(x = year_of_home_purchase, y = prob_high_tax)) +
  geom_point() +
  geom_line() +
  labs(title = "Treatment Status Test for YOLO County",
      subtitle = "Probability of High Tax Rate by Year of Home Purchase",
       x = "Year of Home Purchase", y = "Probability of High Tax Rate") +
  geom_vline(xintercept = 1978, linetype = "dashed", color = "red") +
  theme_fivethirtyeight() +
  theme(axis.title = element_text())
```



Alameda County (Fuzzy Regression Discontinuity): A fuzzy RD in this scenario requires the extra steps testing the instrument before the running the RD.

1) First Stage: Regress the treatment on the instrument (year of home purchase)

   This first stage regression demonstrates that year of home purchase is a strong instrument for the treatment variable (high tax rate dummy), as it significantly predicts whether a household in Alameda County faces a high tax rate. This can be seen by the high f-statistic that is significant at the 1% level and the p-value that is also significant at 1%.

```r
first_stage <- lm(high_tax_dummy ~ year_of_home_purchase, data = county_A)
stargazer(first_stage, type = "text", title = "First Stage Regression Results",
          dep.var.labels = "High Tax Rate (Dummy)",
          covariate.labels = c("Instrument: Year of Home Purchase"),
          out = "first_stage_output.txt")
```

```
##
## First Stage Regression Results
## =============================================================
##                                       Dependent variable:
##                                   ---------------------------
##                                      High Tax Rate (Dummy)
## -------------------------------------------------------------
## Instrument: Year of Home Purchase          0.018***
##                                             (0.001)
##
## Constant                                  -35.191***
##                                             (1.791)
##
## -------------------------------------------------------------
## Observations                                 2,500
## R2                                           0.136
## Adjusted R2                                  0.136
## Residual Std. Error                    0.454 (df = 2498)
## F Statistic                        394.507*** (df = 1; 2498)
## =============================================================
## Note:                             *p<0.1; **p<0.05; ***p<0.01
```

2) Reduced Form: Regress the outcome on the instrument (year of home purchase)

   This reduced form regression demonstrates that the year of home purchase significantly influences the likelihood of tax evasion in Alameda County. The positive and statistically significant coefficient for year of home purchase indicates that as the year of home purchase increases, the probability of tax evasion also increases. This suggests that the timing of when a household purchased their home is a significant predictor of tax evasion.

```r
reduced_form <- lm(evades_taxes_yn ~ year_of_home_purchase, data = county_A)
stargazer(reduced_form,
          type = "text",
          title = "Reduced Form Regression Results",
          dep.var.labels = "Tax Evasion (Yes/No)",
          covariate.labels = c("Year of Home Purchase"),
          out = "reduced_form_output.txt")
```

```
##
## Reduced Form Regression Results
## =================================================
##                            Dependent variable:
##                        ---------------------------
```

```
##                         Tax Evasion (Yes/No)
## -----------------------------------------------
## Year of Home Purchase          0.017***
##                                 (0.001)
##
## Constant                      -32.797***
##                                 (1.522)
##
## -----------------------------------------------
## Observations                     2,500
## R2                               0.159
## Adjusted R2                      0.158
## Residual Std. Error        0.385 (df = 2498)
## F Statistic            471.025*** (df = 1; 2498)
## ===============================================
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

3) Second Stage: Isolate the causal effect of the treatment (high marginal tax rate) on the outcome (tax evasion rate), using the instrument to correct for the fuzziness.

   The coefficient for high tax rate (dummy) is 0.928 and is statistically significant at the 1% level, which suggests households that have o a high tax rate are 92% more likely to engage in tax evasion compared to households with the low tax rate.

   The model explains about 30% of the variation in tax evasion (R-squared = 0.301). This suggests a moderately strong fit of the model. The results are robust with highly significant coefficients, and the standard errors are relatively small, further emphasizing the reliability of the estimates.

```r
Alameda_Fuzzy_RD <- ivreg(evades_taxes_yn ~ high_tax_dummy | year_of_home_purchase, data = county_A)
stargazer(Alameda_Fuzzy_RD, type = "text", title = "Fuzzy RD Results",
          dep.var.labels = "Tax Evasion (Yes/No)",
          covariate.labels = c("High Tax Rate (Dummy)", "Instrument: Year of Home Purchase"),
          out = "fuzzy_rd_output.txt")
```

```
##
## Fuzzy RD Results
## ===============================================================
##                                       Dependent variable:
##                                   ---------------------------
##                                      Tax Evasion (Yes/No)
## ---------------------------------------------------------------
## High Tax Rate (Dummy)                       0.928***
##                                             (0.039)
##
## Instrument: Year of Home Purchase          -0.135***
##                                             (0.017)
##
## ---------------------------------------------------------------
## Observations                                 2,500
## R2                                           0.301
## Adjusted R2                                  0.301
## Residual Std. Error                    0.351 (df = 2498)
## ===============================================================
## Note:                               *p<0.1; **p<0.05; ***p<0.01
```

```r
summary(Alameda_Fuzzy_RD)
```

```
##
## Call:
## ivreg(formula = evades_taxes_yn ~ high_tax_dummy | year_of_home_purchase,
##     data = county_A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7935  0.1347  0.1347  0.1347  0.2065
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.13467    0.01680  -8.015 1.67e-15 ***
## high_tax_dummy  0.92816    0.03897  23.815  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3513 on 2498 degrees of freedom
## Multiple R-Squared: 0.3013,  Adjusted R-squared: 0.301
## Wald test: 567.2 on 1 and 2498 DF,  p-value: < 2.2e-16
```

YOLO County (Sharp Regression Discontinuity): Yi = a + T * Di + f(Xi) + Ei Yi = the probability of tax evasion (outcome variable)

a (alpha) = intercept

T (tau) = treatment effect at the cutoff (the difference in tax evasion rates between those who are just below and just above 1978 (cutoff year)

Di = treatment indicator (1 if year of home purchase is 1978 or greater and 0 otherwise)

f(Xi) = function of the year of home purchase (running variable)

Ei = error term

The estimated coefficient for the high tax rate is 0.0128, and it is statistically significant ($p < 2e-16$). This means that households with a high tax rate are 44.9 percentage points more likely to evade taxes than those with a low tax rate, holding other factors constant. The model explains 21% of the variance in tax evasion behavior.

```r
YOLO_Sharp_RD <- lm(evades_taxes_yn ~ high_tax_dummy, data = county_C)
stargazer(YOLO_Sharp_RD, type = "text", title = "Sharp RD Results for Yolo County",
          dep.var.labels = "Tax Evasion (Yes/No)",
          covariate.labels = c("High Tax Rate (Dummy)"),
          out = "sharp_rd_yolo_output.txt")
```

```
##
## Sharp RD Results for Yolo County
## ================================================
##                           Dependent variable:
##                       ---------------------------
##                          Tax Evasion (Yes/No)
## ------------------------------------------------
## High Tax Rate (Dummy)          0.449***
##                                 (0.017)
##
```

```
## Constant                         0.188***
##                                  (0.012)
##
## ------------------------------------------------
## Observations                      2,500
## R2                                0.210
## Adjusted R2                       0.209
## Residual Std. Error        0.435 (df = 2498)
## F Statistic           663.093*** (df = 1; 2498)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

Discussion

The analysis indicates that both counties with valid regression discontinuity (RD) models exhibit a statistically significant positive coefficient (p-value $<$ 2e-16), demonstrating that higher marginal tax rates are associated with an increased likelihood of tax evasion.

The RD analysis highlights that raising the tax rate from 15% to 50% corresponds to a measurable rise in tax evasion behaviors. This finding underscores the potential for the current tax structure to inadvertently incentivize non-compliance, particularly among wealthy households.

Although this analysis is based on a limited dataset from three counties, the results suggest the need for policy adjustments to address tax evasion while preserving progressive taxation goals. The NGO concerned with ensuring equitable tax contributions among wealthy households should advocate for strategies that reduce tax evasion incentives, such as enhancing enforcement mechanisms, increasing transparency, and simplifying tax compliance procedures.