Housing Vouchers in Washington State

Ben Warren

2023-05-29

Abstract:

An analysis of data collected in a study of over 800 families from Washington's Seattle and King counties demonstrated that offering families who qualified for Section 8 vouchers an additional suite of public housing services was correlated with positive outcomes in those families leasing in high-opportunity zones. In particular, those who took advantage of the services in King County were more likely to lease in these zones, while the predicted effect of offering additional services was an increase of 2.5 percentage points in household income.

Analysis: In this study, investigating the impact of housing vouchers for low-income families, allowing them to move into high-opportunity areas, *pha* acts as an assignment variable. It represents which housing authority the family falls in, either Seattle or King County. So, this is a discrete variable where each value is a different level of the treatment, making it an assignment variable. Since the two counties handled recruitment into the study differently, this assignment variable may have some impact on the findings for families in each county.

Though they handled recruitment differently, both counties randomly assigned families into treatment and control groups. To ascertain whether the assignments are effectively random, we should look at the values of different variables compared between treatment and control groups.

The variables I will look at are: hoh_age (head of household's age at baseline), child_count (number of children at baseline), child_age (average age of children at baseline), and hh_income (household income at baseline).

```
##          treated        control
## 1      34.042755      34.513784
## 2       2.453682       2.413534
## 3       6.532946       6.744230
## 4   19190.004751   19658.644110
```

In each of these categories, the treatment group is roughly identical to the control group.

If we instead look at a number of variables that are discrete, with either 0s or 1s to describe the families who participated in the experiment we can check if the distribution of these variables across the control and treatment groups is roughly equal.

If we now look at the difference between the means of the treatment and control groups, we see a slight difference between the two groups.

The raw differences in means (treated minus control) are as follows:

```
##   Head of Household Age Child Count Average Child Age Household Income
## 1                -0.471        0.04            -0.211           -468.6
```

Meanwhile, the standard errors are the following:

```
##    Head of Household Age Child Count Average Child Age Household Income
## 1                 0.549        0.089             0.258            843.6
```

Each of the variables shows a high level of randomness. None of the variables I looked at seem to skew based on the group families are assigned to.

An important aspect of any randomly-assigned study that involves the provision of services is the *compliance rate*, or what proportion of the subjects assigned to the treatment group actually received the treatment. In this case, we must consider what proportion of the families assigned to the treatment group in the study received services through CMTO.

When regressing the likelihood of receiving CMTO services on assignment to the treatment group, the model utility test returns a very low p-value, indicating that there is a linear relationship, while the R-squared value returns 0.62. Assignment to the treatment group (participation in the program) is responsible for ~62% of the variation in the target variable, a fairly strong relationship.

Still, this means a fair number of the families involved in the study assigned to the treatment group did not take advantage of the services (otherwise the R-squared value would be around 1). This may influence the predicted influence of the vouchers on families who receive it.

If we consider the intent-to-treat effect, instead of the overall effect, include all of the families in the treatment group, whether or not they used CMTO services. To do this, we can regress assignment to the treatment group on signing a lease in the high-opportunity area.

```
##
## Call:
## lm(formula = leased_up_opp ~ treatment_group + pha, data = cmto_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3293 -0.2843 -0.1489 -0.1039  0.8961
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.14895    0.02424   6.145 1.25e-09 ***
## treatment_group   0.18040    0.02829   6.377 3.02e-10 ***
## pha              -0.04509    0.02836  -1.590    0.112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4049 on 817 degrees of freedom
## Multiple R-squared:  0.0505, Adjusted R-squared:  0.04818
## F-statistic: 21.73 on 2 and 817 DF,  p-value: 6.401e-10
```

The results of this regression show that though there appears to be a linear relationship between assignment to the treatment group and leasing up into a high-opportunity zone (according to the model utility test, with a p-value of < 0.05), the relationship is not large, with an R-squared of just 0.05.

If we wish to isolate the treatment group families that actually used CMTO services, we could analyze the treatment-on-the-treated effect.This can be calculated by the same regression, but instead of using assignment to the treatment group as the independent variable, we use actual use of the services (represented by the variable received_cmto_services).

```
##
## Call:
## lm(formula = leased_up_opp ~ received_cmto_services + pha, data = cmto_data)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3817 -0.3231 -0.1595 -0.1008  0.8992
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.15948    0.02186   7.295 7.05e-13 ***
## received_cmto_services 0.22224    0.02860   7.770 2.35e-14 ***
## pha                   -0.05864    0.02808  -2.089   0.0371 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4003 on 817 degrees of freedom
## Multiple R-squared:  0.07183,    Adjusted R-squared:  0.06956
## F-statistic: 31.61 on 2 and 817 DF,  p-value: 5.961e-14
```

There is a higher p-value on the model utility test when using the tot variable, as well as a larger r-squared value, demonstrating that actual use of CMTO services had a larger impact on the likelihood that a participant leased up into a high opportunity zone than simply being assigned to the treatment group.

This makes sense when we consider that the second regression accounts for non-compliance. Just because a family was assigned into the treatment group doesn't mean that they actually used the service. We would assume then that for these families, assignment to the treatment group would not have an effect.

An effective measure of predicted change for each family environment should be the change (positive or negative) in the long-run earnings of the origin and destination census tract.

The distribution of the variable is as follows. The distribution overall appears fairly normal, centered close to 0, with a slight skew towards a positive change.
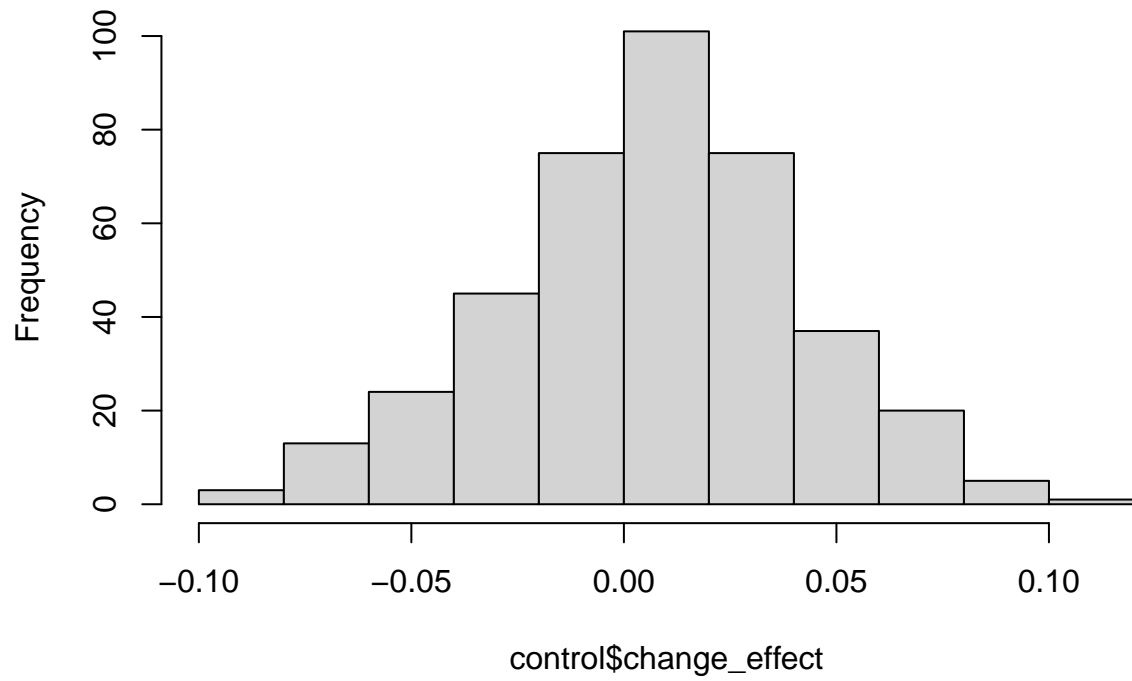
Dist. of predicted change of environment
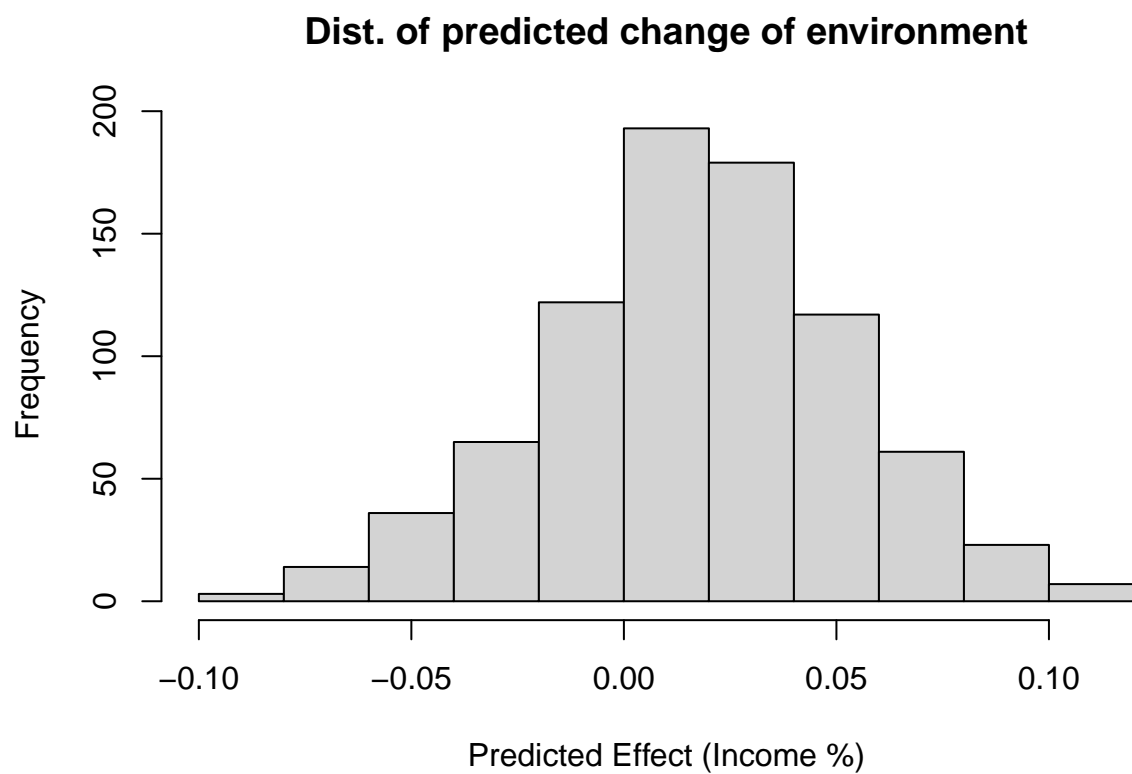
If we look at the treatment and control groups separately, the distribution of the predicted effect for the treatment group is centered around 0.025, clearly to the right of the center of the distribution for the control group.
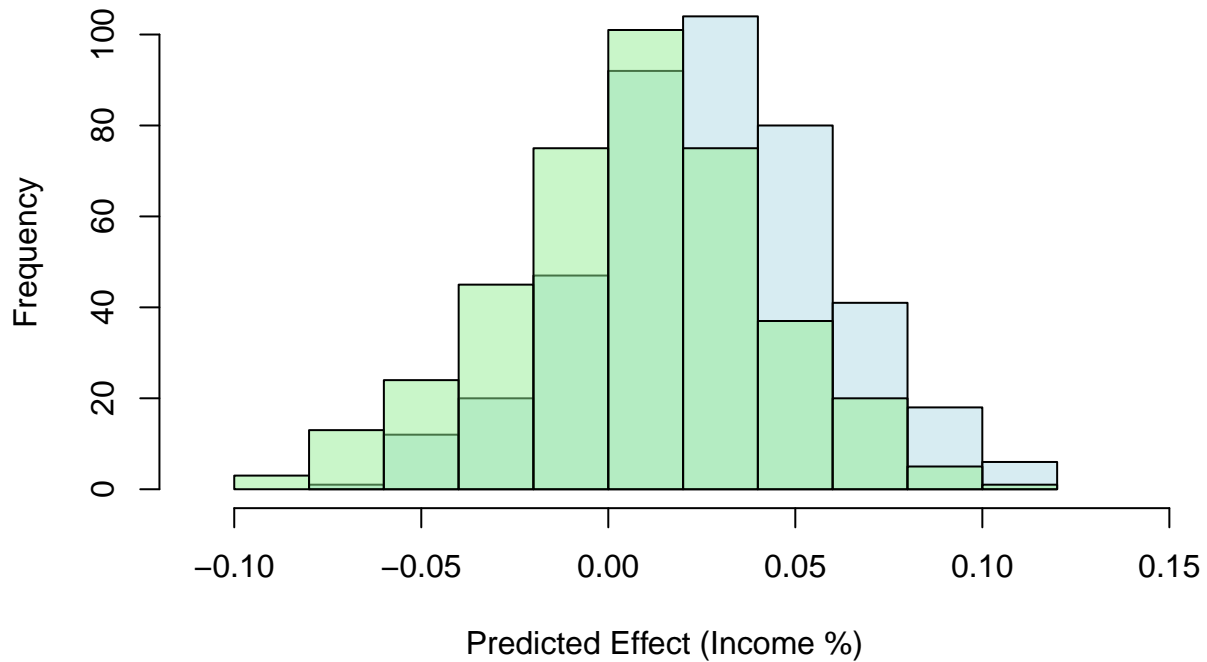
Histogram of treated$change_effect
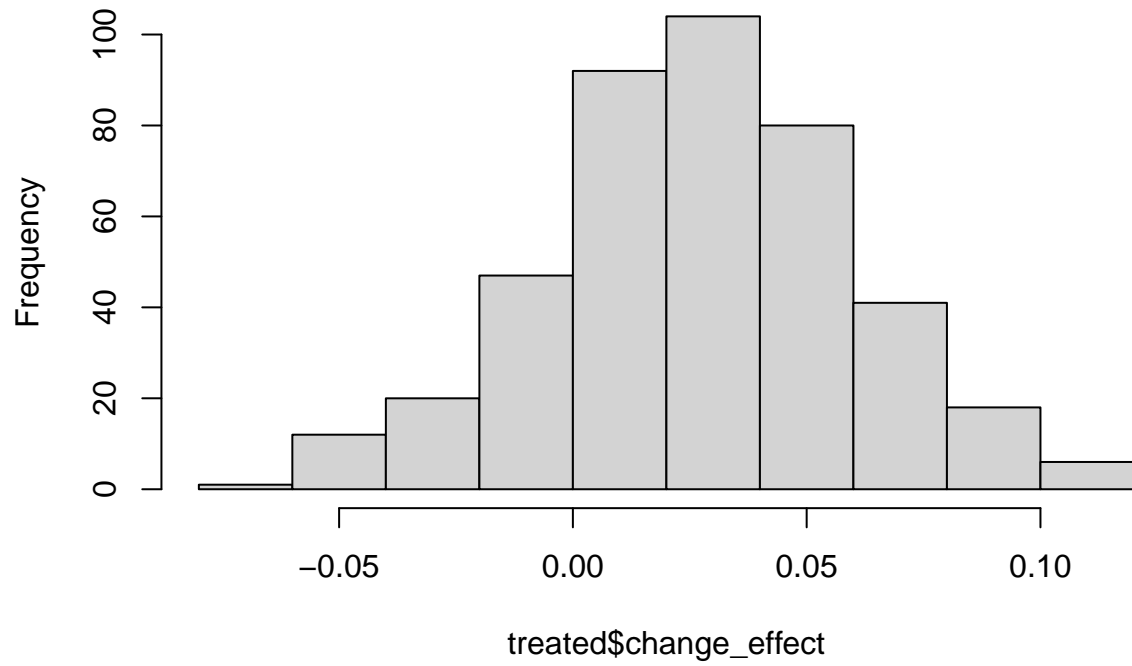
Histogram of control$change_effect

Dist. of predicted change of environment, treatment (G) vs. control (E

A similar disparity appears in the mean effects of the two groups. The average change effect in the treatment group is 4 times the average change effect in the control group, a significant difference.

```
## [1] "Avg. Change Effect, Treatment Group"
```

```
## [1] 0.02715177
```

```
## [1] "Avg. Change Effect, Control Group"
```

```
## [1] 0.006818682
```

```
## [1] "Ratio of change effects:"
```

```
## [1] 3.981967
```

If we regress the change effect on treatment group, the regression shows a strong model utility test, as well as a relatively strong R-squared, with treatment group accounting for over 20% of the variation in the change effect. There seems to be a correlation between whether families were in the treatment group and the environmental change effect.

```
##
## Call:
## lm(formula = change_effect ~ treatment_group + pha, data = cmto_data)
##
```

```
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.090542 -0.021274  0.000511  0.021069  0.096378
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.019026   0.001870  10.175   <2e-16 ***
## treatment_group 0.020005   0.002182   9.166   <2e-16 ***
## pha            -0.026046   0.002188 -11.904   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03123 on 817 degrees of freedom
## Multiple R-squared:  0.2186, Adjusted R-squared:  0.2167
## F-statistic: 114.3 on 2 and 817 DF,  p-value: < 2.2e-16
```

A heterogeneous treatment effect would occur if, for example, families in the study who made more money experienced a significantly greater effect from being in the treatment group than those who made less. This is an important consideration in analyzing the data from the study.

In order to consider this, I will re-run the earlier regressions on 4 different groups: those with family income greater than the median in the sample, less than the median in the sample, and those in each of the public housing authorities individually.

Family income greater than the median in the sample and less than (or equal to) the median in the sample:

```
##
## Call:
## lm(formula = leased_up_opp ~ received_cmto_services + pha, data = above)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4084 -0.3039 -0.1964 -0.1277  0.8723
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.19641    0.03357   5.852 1.00e-08 ***
## received_cmto_services 0.21200    0.04223   5.020 7.74e-07 ***
## pha                   -0.06874    0.04138  -1.661   0.0974 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4186 on 407 degrees of freedom
## Multiple R-squared:  0.06301,    Adjusted R-squared:  0.0584
## F-statistic: 13.68 on 2 and 407 DF,  p-value: 1.771e-06


##
## Call:
## lm(formula = leased_up_opp ~ received_cmto_services + pha, data = below)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36019 -0.30139 -0.12784 -0.06904  0.93096
##
```

```
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             0.12784    0.02841   4.500 8.88e-06 ***
## received_cmto_services  0.23236    0.03852   6.032 3.64e-09 ***
## pha                    -0.05880    0.03819  -1.539    0.124
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3805 on 407 degrees of freedom
## Multiple R-squared:  0.08478,    Adjusted R-squared:  0.08029
## F-statistic: 18.85 on 2 and 407 DF,  p-value: 1.48e-08
```

There is no significant difference between the two regressions. There does not appear to be a heteregeneous effect based on household income.

Each Public Housing Authority separately (KCHA and SHA):

```
##
## Call:
## lm(formula = leased_up_opp ~ received_cmto_services, data = kcha)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4424 -0.1232 -0.1232 -0.1232  0.8768
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             0.12319    0.02412   5.106 4.90e-07 ***
## received_cmto_services  0.31924    0.03944   8.094 5.71e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4008 on 439 degrees of freedom
## Multiple R-squared:  0.1299, Adjusted R-squared:  0.1279
## F-statistic: 65.52 on 1 and 439 DF,  p-value: 5.712e-15


##
## Call:
## lm(formula = leased_up_opp ~ received_cmto_services, data = sha)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2609 -0.2609 -0.1468 -0.1468  0.8532
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             0.14679    0.02664   5.509 6.69e-08 ***
## received_cmto_services  0.11408    0.04088   2.791  0.00553 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3934 on 377 degrees of freedom
## Multiple R-squared:  0.02024,    Adjusted R-squared:  0.01764
## F-statistic: 7.787 on 1 and 377 DF,  p-value: 0.005528
```

There *is* a significant difference in the r-squared value produced by the regression when the two different county housing authorities are considered. Families in King County experienced a 6 times greater effect of being assigned to the treatment group than those in Seattle.

The two housing authorities handled recruitment and randomization for the study separately, so there may be some structural difference between the way King County recruited and randomized families in the study and the way Seattle did so. This warrants further investigation.