# Análisis de Variable Instrumental en Homeless

Mateo, Sebastián, Genaro

#### **Problema**

En numerosos condados a lo largo de Estados Unidos, el incremento en los costos de vivienda ha sido una preocupación creciente, no solo por su impacto directo en la asequibilidad de la vivienda, sino también por sus posibles efectos en el aumento de la población sin hogar. Identificar las dinámicas precisas entre estos factores es crucial para desarrollar políticas efectivas de vivienda y programas de asistencia social. Sin embargo, la relación entre el costo de la vivienda y el número de personas sin hogar es compleja y puede estar confundida por variables omitidas, como el desempleo o las políticas locales, lo que dificulta obtener estimaciones causales claras.

Nuestra hipótesis plantea que el costo promedio de renta en un condado tiene un impacto significativo en la cantidad de personas sin hogar en ese mismo condado. Para explorar esta relación de manera más precisa y controlar la posible endogeneidad de la variable de costo de vivienda, utilizamos el costo promedio de renta de los condados colindantes en el periodo de medida anterior como variable instrumental. Esperamos demostrar que, al aislar la influencia de otros factores, un aumento en el costo de la renta contribuye directamente al incremento en el número de personas sin hogar, proporcionando así evidencia que podría ser utilizada para guiar intervenciones y políticas públicas efectivas.

#### Variables Instrumentales

En análisis estadístico, a menudo nos enfrentamos al problema de la endogeneidad, que ocurre cuando hay una correlación entre una variable explicativa X y el término de error U. Esta correlación puede surgir de una variable omitida que influye simultáneamente en X, y en la variable dependiente Y, generando estimaciones sesgadas y conclusiones erróneas en modelos de regresión lineal. Para abordar esta complicación, uno de los enfoques más efectivos es el uso de variables instrumentales. Este método nos permite aislar el efecto verdadero de X sobre Y, proporcionando una estimación más precisa de la relación causal.

# Definición y supuestos

Una variable Z se considera instrumental si cumple con los siguientes supuestos:

• Relevancia: Z debe estar correlacionada con la variable explicativa X, pero no necesariamente con la variable dependiente Y (Que Z tenga efecto sobre Y). Esto se verifica a través de la condición:

$$cov(X_i, Z_i) \neq 0$$

• Exogeneidad: Z no debe estar correlacionada con el término de error U, garantizando que no está afectada por las variables omitidas que influyen en Y. Esto asegura que los efectos estimados no están sesgados por variables no observadas:

$$cov(U_i, Z_i) = 0$$

• Restricción de Exclusión: Z solo debe influir en Y a través de X y no debe tener ningún efecto directo o a través de otros canales no observados. Esta condición es crucial para asegurar que la relación causal que medimos es exclusivamente a través de X:

$$cov(Y_i,Z_i|X_i)=0$$

# **Datos**

```
data <- read.csv("./metodos_lineales_data.csv")
#data[is.na(data$Previous_CoC_Neighbor_Median_Rent), ]
# los primeros periodos están en null por lo que para hacer la regresión filtramos
data <- na.omit(data, cols = "Previous_CoC_Neighbor_Median_Rent")
summary(lm(Overall_Homeless ~ Previous_CoC_Neighbor_Median_Rent, data = data))</pre>
```

#### Call:

```
lm(formula = Overall_Homeless ~ Previous_CoC_Neighbor_Median_Rent,
    data = data)
```

# Residuals:

```
Min 1Q Median 3Q Max -5909 -1078 -489 149 75360
```

#### Coefficients:

```
Estimate Std. Error t value Pr(>|t|)

(Intercept) -1412.600 268.700 -5.26 1.6e-07 ***

Previous_CoC_Neighbor_Median_Rent 4.012 0.354 11.33 < 2e-16 ***
---

Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 4780 on 3378 degrees of freedom Multiple R-squared: 0.0366, Adjusted R-squared: 0.0363 F-statistic: 128 on 1 and 3378 DF, p-value: <2e-16

- Year: Año del registro. (2013 2022)
- CoC\_Number: Identificador numérico del Continuum of Care (CoC).
- CoC\_Name: Nombre descriptivo del CoC.
- Overall\_Homeless: Cantidad total de personas sin hogar.
- Chronic\_Homeless: Cantidad de personas sin hogar crónicamente.
- Non-Chronic\_Homeless: Cantidad de personas sin hogar no crónicamente.
- Actual\_CoC\_Median\_Rent: Renta media en el área del CoC.
- Previous\_CoC\_Neighbor\_Median\_Rent: Renta media en los CoCs vecinos.
- State: Estado en EE.UU. donde se encuentra el Continuum of Care (CoC).
- CoC\_AVG\_Median\_Household\_Income: Ingreso medio del hogar promedio en el área del CoC.
- CoC\_Population\_Estimate: Estimación de la población total en el área del CoC.
- Poverty\_Percentage: Porcentaje de la población bajo el umbral de pobreza en el área del CoC.
- CoC\_Civilian\_Labor\_Force: Tamaño de la fuerza laboral civil en el área del CoC.
- Unemployment\_Rate: Tasa de desempleo en el área del CoC.
- Party\_Affiliation: Afiliación política predominante en el área del CoC.
- Vote\_Percentage: Porcentaje de votos obtenidos por el partido predominante en las últimas elecciones.
- CoC\_Poverty\_Estimate\_Age\_0\_17: Estimación del número de menores de 18 años en situación de pobreza en el área del CoC.
- Birth\_Rate: Tasa de natalidad en el área del CoC.
- Death Rate: Tasa de mortalidad en el área del CoC.

- International\_Migration\_Rate: Tasa de migración internacional en el área del CoC.
- Domestic\_Migration\_Rate: Tasa de migración doméstica dentro del país en el área del CoC.

Agregamos una varibale dummy llamada Party\_Democrat

```
data$Party_Democrat <- as.integer(data$Party_Affiliation == 'Democrat')</pre>
```

#### str(data)

```
'data.frame':
              3380 obs. of 22 variables:
$ Year
                                  $ CoC_Number
                                        "AL-500" "AL-501" "AL-502" "AL-503" ...
                                  : chr
$ CoC_Name
                                       "Birmingham/Jefferson, St. Clair, Shelby Counties
                                  : chr
                                        "AL" "AL" "AL" "AL" ...
$ State
                                  : chr
$ Overall_Homeless
                                        943 585 232 549 278 40 935 859 343 814 ...
                                  : int
$ Chronic_Homeless
                                  : int
                                        267 76 11 118 57 0 22 290 49 111 ...
$ Non.Chronic_Homeless
                                  : int 676 509 221 431 221 40 913 569 294 703 ...
$ Actual_CoC_Median_Rent
                                  : num 834 783 554 682 674 ...
$ Previous_CoC_Neighbor_Median_Rent: num 728 636 551 534 539 ...
$ CoC_AVG_Median_Household_Income
                                  : num 61996 63010 52939 75282 44976 ...
                                        665084 657929 271888 638704 246227 236690 1383394
$ CoC_Population_Estimate
                                  : int
$ Poverty_Percentage
                                       15.9 16.4 15.7 11.1 18.5 ...
                                  : num
$ CoC_Civilian_Labor_Force
                                  : int 323290 295161 120142 308386 115243 103432 591702
$ Unemployment_Rate
                                  : num 0.0258 0.0293 0.0262 0.0212 0.0313 ...
                                        "Democrat" "Republican" "Republican" "Republican"
$ Party_Affiliation
                                  : chr
$ Vote_Percentage
                                        0.516 0.627 0.746 0.614 0.628 ...
                                  : num
$ CoC_Poverty_Estimate_Age_0_17
                                  : int 31449 33537 12649 19794 16108 9540 69752 22117 18
$ Birth_Rate
                                  : num 11.9 11.5 10.8 11 13.1 ...
$ Death_Rate
                                  : num 13.2 13.1 15.2 10.9 12.2 ...
$ International_Migration_Rate
                                  : num 0.952 0.813 0.427 1.359 1.474 ...
$ Domestic_Migration_Rate
                                  : num -6.72 8.37 10 15.78 -5.23 ...
$ Party_Democrat
                                  : int 1000100100...
```

# summary(data)

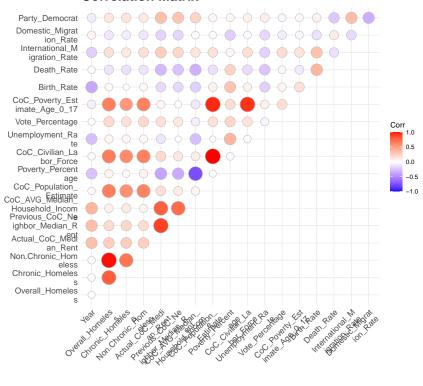
Year	CoC_Number	CoC_Name	State	Overall_Homeless Ch
Min. :2013	Length:3380	Length:3380	Length:3380	Min. : 23 Mi
1st Qu.:2015	Class :character	Class :character	Class :character	1st Qu.: 310 1s
Median :2018	Mode :character	Mode :character	Mode :character	Median: 610 Me
Mean :2018				Mean : 1485 Me

3rd Qu.:2020			3rd Qu.:	1351 3rd
Max. :2022			Max. :	78676 Ma:
Previous_CoC_Neigh	bor_Median_Rent	CoC_AVG_Median_Househ	old_Income CoC_Population	_Estimate Po
Min. : 387		Min. : 30141	Min. : 57	91 M:
1st Qu.: 556		1st Qu.: 49215	1st Qu.: 2618	47 1:
Median : 658		Median : 57575	Median: 4957	40 Me
Mean : 722		Mean : 61497	Mean : 8118	26 Me
3rd Qu.: 828		3rd Qu.: 69529	3rd Qu.: 8431	10 3:
Max. :2158		Max. :150502	Max. :102642	68 Ma
Party_Affiliation	Vote_Percentag	e CoC_Poverty_Estimate	_Age_0_17 Birth_Rate	Death_Ra
Length:3380	Min. :0.420	Min. : 491	Min. : 1.36	Min. : 1
Class :character	1st Qu.:0.541	1st Qu.: 8631	1st Qu.: 9.69	1st Qu.: 7
Mode :character	Median :0.589	Median : 17174	Median :11.09	Median : 9
	Mean :0.600	Mean : 33415	Mean :10.60	Mean : 8
	3rd Qu.:0.646	3rd Qu.: 33459	3rd Qu.:12.45	3rd Qu.:10
	Max. :0.853	Max. :624784	Max. :21.77	Max. :19
Party Democrat				

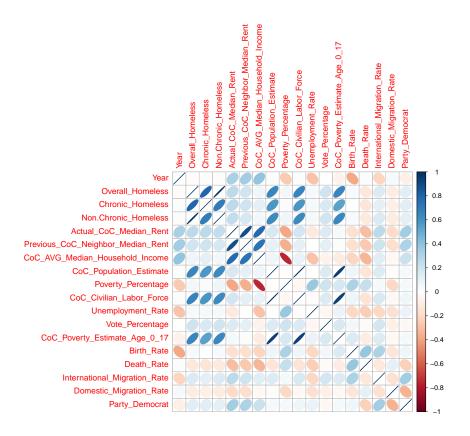
Party\_Democrat

Min. :0.000 1st Qu.:0.000 Median :1.000 Mean :0.521 3rd Qu.:1.000 Max. :1.000

### **Correlation Matrix**



```
correlation_matrix <- cor(data %>% select_if(is.numeric))
R <- round(correlation_matrix, 3)
corrplot(correlation_matrix, method = "ellipse")</pre>
```



# **Análisis**

```
model <- lm(Overall_Homeless ~ Actual_CoC_Median_Rent, data = data)
summary(model)</pre>
```

# Call:

lm(formula = Overall\_Homeless ~ Actual\_CoC\_Median\_Rent, data = data)

#### Residuals:

Min 1Q Median 3Q Max -8735 -1048 -337 305 73258

# Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1902.651 229.779 -8.28 <2e-16 \*\*\*
Actual\_CoC\_Median\_Rent 4.439 0.282 15.75 <2e-16 \*\*\*

---

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 4700 on 3378 degrees of freedom Multiple R-squared: 0.0684, Adjusted R-squared: 0.0681

F-statistic: 248 on 1 and 3378 DF, p-value: <2e-16

R cuadrada bajo por lo que teníamos razón diciendo que hay más cosas afectando a los homeless aparte de la renta. (?)

#### Call.

```
lm(formula = Overall_Homeless ~ Actual_CoC_Median_Rent + CoC_AVG_Median_Household_Income +
    CoC_Population_Estimate + Poverty_Percentage + CoC_Civilian_Labor_Force +
    Unemployment_Rate + Vote_Percentage + CoC_Poverty_Estimate_Age_0_17 +
    Birth_Rate + Death_Rate + International_Migration_Rate +
    Domestic_Migration_Rate, data = data)
```

# Residuals:

Min 1Q Median 3Q Max -21122 -711 223 851 58551

#### Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-2.04e+03	8.87e+02	-2.30	0.02153 *
Actual_CoC_Median_Rent	5.45e+00	4.00e-01	13.64	< 2e-16 ***
CoC_AVG_Median_Household_Income	-5.97e-02	9.05e-03	-6.60	4.6e-11 ***
CoC_Population_Estimate	-1.65e-02	9.47e-04	-17.42	< 2e-16 ***
Poverty_Percentage	-3.92e+01	2.75e+01	-1.43	0.15286
CoC_Civilian_Labor_Force	3.23e-02	1.76e-03	18.40	< 2e-16 ***
Unemployment_Rate	6.12e+03	3.35e+03	1.83	0.06761 .
Vote_Percentage	3.01e+03	8.14e+02	3.70	0.00022 ***

```
CoC_Poverty_Estimate_Age_0_17
                               7.25e-02
                                         4.28e-03
                                                    16.92 < 2e-16 ***
                              -1.00e+02 2.31e+01 -4.35 1.4e-05 ***
Birth_Rate
Death_Rate
                               8.58e+01 2.38e+01 3.60 0.00032 ***
International_Migration_Rate
                              -1.12e+02
                                         3.21e+01 -3.47 0.00053 ***
                               6.65e+00
                                         6.64e+00 1.00 0.31655
Domestic_Migration_Rate
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 3310 on 3367 degrees of freedom
Multiple R-squared: 0.539, Adjusted R-squared: 0.538
F-statistic: 328 on 12 and 3367 DF, p-value: <2e-16
```

Regresión solo con variables significativas

# Call:

```
lm(formula = Overall_Homeless ~ Actual_CoC_Median_Rent + CoC_AVG_Median_Household_Income +
    CoC_Population_Estimate + CoC_Civilian_Labor_Force + Vote_Percentage +
    CoC_Poverty_Estimate_Age_0_17 + Birth_Rate + Death_Rate +
    International_Migration_Rate, data = data)
```

# Residuals:

Min 1Q Median 3Q Max -20819 -676 216 857 58723

#### Coefficients:

	Estimate	Std. Error t	value	Pr(> t )
(Intercept)	-2.02e+03	6.23e+02	-3.24	0.00119 **
Actual_CoC_Median_Rent	5.25e+00	3.51e-01	14.99	< 2e-16 ***
${\tt CoC\_AVG\_Median\_Household\_Income}$	-5.33e-02	6.21e-03	-8.59	< 2e-16 ***
CoC_Population_Estimate	-1.62e-02	9.27e-04	-17.48	< 2e-16 ***
CoC_Civilian_Labor_Force	3.19e-02	1.71e-03	18.59	< 2e-16 ***

```
Vote_Percentage
                               2.69e+03
                                         7.99e+02 3.37 0.00077 ***
CoC_Poverty_Estimate_Age_0_17
                              7.13e-02 4.13e-03 17.27 < 2e-16 ***
Birth_Rate
                              -1.15e+02 2.22e+01 -5.19 2.2e-07 ***
Death_Rate
                               7.52e+01 2.24e+01
                                                   3.36 0.00079 ***
International_Migration_Rate
                              -1.22e+02 3.15e+01 -3.88 0.00011 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 3310 on 3370 degrees of freedom
Multiple R-squared: 0.539, Adjusted R-squared: 0.537
F-statistic: 437 on 9 and 3370 DF, p-value: <2e-16
```

#### Variable Instrumental

#### Call:

```
lm(formula = Overall_Homeless ~ fitted_Actual_CoC_Median_Rent,
    data = data)
```

#### Residuals:

```
Min 1Q Median 3Q Max -5909 -1078 -489 149 75360
```

# Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) -1308.599 259.973 -5.03 5.1e-07 ***
fitted_Actual_CoC_Median_Rent 3.661 0.323 11.33 < 2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 4780 on 3378 degrees of freedom Multiple R-squared: 0.0366, Adjusted R-squared: 0.0363

F-statistic: 128 on 1 and 3378 DF, p-value: <2e-16

Los errores estandar están mal pero las betas son correctas, la U del modelo es diferente que la U de la étapa en la segunda etapa estoy utilizando la renta actual predicha entonces tiene otro error sumado

me fijo más en como son las betas para ver el efecto sobre los homeless bartik instruments. problemas en probar exogeneidad restricción de exclución. problema por que la periferia influye

si instrumento es exogeno, controles no deberían afectar controles deben pasar una prueba de balance, pruebas placebo regresar controles y no debería haber efecto

#### Call:

```
lm(formula = Overall_Homeless ~ fitted_Actual_CoC_Median_Rent2 +
    CoC_AVG_Median_Household_Income + CoC_Population_Estimate +
    Poverty_Percentage + CoC_Civilian_Labor_Force + Unemployment_Rate +
    Vote_Percentage + CoC_Poverty_Estimate_Age_0_17 + Birth_Rate +
    Death_Rate + International_Migration_Rate + Domestic_Migration_Rate,
    data = data)
```

#### Residuals:

Min 1Q Median 3Q Max

#### Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
                           -1.97e+03 1.01e+03 -1.95 0.05170 .
(Intercept)
fitted_Actual_CoC_Median_Rent2
                            5.53e+00
                                      6.26e-01
                                               8.82 < 2e-16 ***
CoC_Population_Estimate
                           -1.65e-02
                                      9.62e-04 -17.14 < 2e-16 ***
Poverty_Percentage
                           -4.12e+01
                                      3.06e+01 -1.35 0.17836
CoC_Civilian_Labor_Force
                            3.23e-02
                                      1.78e-03 18.12 < 2e-16 ***
                                      3.42e+03 1.77 0.07616 .
Unemployment_Rate
                            6.06e+03
Vote_Percentage
                            2.98e+03
                                      8.54e+02
                                                3.49 0.00050 ***
CoC_Poverty_Estimate_Age_0_17
                                      4.36e-03 16.64 < 2e-16 ***
                            7.25e-02
Birth_Rate
                           -9.96e+01
                                      2.40e+01 -4.15 3.4e-05 ***
Death_Rate
                            8.55e+01
                                      2.43e+01
                                                3.51 0.00045 ***
International_Migration_Rate
                                      3.36e+01
                                               -3.36 0.00079 ***
                           -1.13e+02
Domestic_Migration_Rate
                            6.77e+00
                                      6.79e+00 1.00 0.31875
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 3360 on 3367 degrees of freedom Multiple R-squared: 0.525, Adjusted R-squared: 0.523 F-statistic: 310 on 12 and 3367 DF, p-value: <2e-16

# Variables instrumental solo con significativos

```
# First Stage: Regress Actual_CoC_Median_Rent on Previous_CoC_Neighbor_Median_Rent and other
first_stage_model3 <- lm(Actual_CoC_Median_Rent ~ Previous_CoC_Neighbor_Median_Rent +
                        CoC_AVG_Median_Household_Income + CoC_Population_Estimate +
                        CoC_Civilian_Labor_Force +
                        Vote_Percentage +
                        CoC_Poverty_Estimate_Age_0_17 + Birth_Rate + Death_Rate +
                        International_Migration_Rate,
                      data = data)
data$fitted_Actual_CoC_Median_Rent3 <- fitted(first_stage_model3)</pre>
# Second Stage: Regress Overall_Homeless on the fitted values from the first stage and other
second_stage_model3 <- lm(Overall_Homeless ~ fitted_Actual_CoC_Median_Rent3 +</pre>
                         CoC_AVG_Median_Household_Income + CoC_Population_Estimate +
                         CoC_Civilian_Labor_Force +
                         Vote_Percentage +
                         CoC_Poverty_Estimate_Age_0_17 + Birth_Rate + Death_Rate +
```

```
International_Migration_Rate, data = data)
summary(second_stage_model3)
Call:
lm(formula = Overall_Homeless ~ fitted_Actual_CoC_Median_Rent3 +
   CoC_AVG_Median_Household_Income + CoC_Population_Estimate +
   CoC_Civilian_Labor_Force + Vote_Percentage + CoC_Poverty_Estimate_Age_0_17 +
   Birth_Rate + Death_Rate + International_Migration_Rate, data = data)
Residuals:
  Min
          10 Median
                        3Q
                              Max
-21247
        -666
                179
                       807 61292
Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)
                               -1.96e+03
                                          6.49e+02 -3.01 0.00260 **
                                5.41e+00 4.97e-01 10.87 < 2e-16 ***
fitted_Actual_CoC_Median_Rent3
CoC_AVG_Median_Household_Income -5.53e-02 7.79e-03 -7.10 1.5e-12 ***
CoC_Population_Estimate
                               -1.62e-02 9.41e-04 -17.22 < 2e-16 ***
CoC_Civilian_Labor_Force
                                          1.74e-03
                                                    18.31 < 2e-16 ***
                                3.19e-02
Vote_Percentage
                                2.59e+03
                                          8.42e+02
                                                     3.08 0.00212 **
CoC_Poverty_Estimate_Age_0_17
                                7.12e-02
                                          4.21e-03 16.90 < 2e-16 ***
```

-1.14e+02

-1.25e+02

7.50e+01

2.26e+01

2.27e+01

3.26e+01

-5.03 5.1e-07 \*\*\*

3.30 0.00098 \*\*\*

-3.83 0.00013 \*\*\*

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3360 on 3370 degrees of freedom Multiple R-squared: 0.524, Adjusted R-squared: 0.523 F-statistic: 413 on 9 and 3370 DF, p-value: <2e-16

#ivreg()

Birth\_Rate

Death\_Rate

International\_Migration\_Rate

# Regresión Poisson