

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:

$$\text{cov}(X_i, Z_i) \neq 0$$

- 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:

$$\text{cov}(U_i, Z_i) = 0$$

- 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 :

$$\text{cov}(Y_i, Z_i | X_i) = 0$$

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))
```

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)	-1412.6000	268.6999	-5.257	1.55e-07 ***
Previous_CoC_Neighbor_Median_Rent	4.0116	0.3542	11.325	< 2e-16 ***

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

Residual standard error: 4780 on 3378 degrees of freedom

Multiple R-squared: 0.03658, Adjusted R-squared: 0.0363

F-statistic: 128.3 on 1 and 3378 DF, p-value: < 2.2e-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 variable dummy llamada `Party_Democrat`

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

```
str(data)
```

```
'data.frame':  3380 obs. of  22 variables:
 $ Year                : int  2022 2022 2022 2022 2022 2022 2022 2022 2022 2022 2022
 $ CoC_Number          : chr   "AL-500" "AL-501" "AL-502" "AL-503" ...
 $ CoC_Name            : chr   "Birmingham/Jefferson, St. Clair, Shelby Counties
 $ State               : chr   "AL" "AL" "AL" "AL" ...
 $ Overall_Homeless    : int   943 585 232 549 278 40 935 859 343 814 ...
 $ 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 ...
 $ CoC_Population_Estimate : int  665084 657929 271888 638704 246227 236690 1383394
 $ Poverty_Percentage   : num   15.9 16.4 15.7 11.1 18.5 ...
 $ 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 ...
 $ Party_Affiliation    : chr   "Democrat" "Republican" "Republican" "Republican"
 $ Vote_Percentage      : num   0.516 0.627 0.746 0.614 0.628 ...
 $ 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    1 0 0 0 1 0 0 1 0 0 ...
```

```
summary(data)
```

	Year	CoC_Number	CoC_Name	State
Min.	:2013	Length:3380	Length:3380	Length:3380
1st Qu.	:2015	Class :character	Class :character	Class :character
Median	:2018	Mode :character	Mode :character	Mode :character
Mean	:2018			

3rd Qu.:2020

Max. :2022

Overall_Homeless Chronic_Homeless Non.Chronic_Homeless

Min. : 23.0 Min. : 0.0 Min. : 6.0

1st Qu.: 309.8 1st Qu.: 34.0 1st Qu.: 259.8

Median : 609.5 Median : 89.0 Median : 501.0

Mean : 1484.5 Mean : 267.5 Mean : 1217.0

3rd Qu.: 1351.2 3rd Qu.: 228.2 3rd Qu.: 1097.5

Max. :78676.0 Max. :26985.0 Max. :72378.0

Actual_CoC_Median_Rent Previous_CoC_Neighbor_Median_Rent

Min. : 400.0 Min. : 387.4

1st Qu.: 570.0 1st Qu.: 556.2

Median : 677.6 Median : 657.8

Mean : 763.0 Mean : 722.2

3rd Qu.: 872.5 3rd Qu.: 827.6

Max. :2447.0 Max. :2157.5

CoC_AVG_Median_Household_Income CoC_Population_Estimate Poverty_Percentage

Min. : 30141 Min. : 5791 Min. : 4.337

1st Qu.: 49215 1st Qu.: 261847 1st Qu.:10.301

Median : 57575 Median : 495740 Median :13.098

Mean : 61497 Mean : 811826 Mean :13.343

3rd Qu.: 69529 3rd Qu.: 843110 3rd Qu.:15.985

Max. :150502 Max. :10264268 Max. :35.164

CoC_Civilian_Labor_Force Unemployment_Rate Party_Affiliation Vote_Percentage

Min. : 1954 Min. :0.01589 Length:3380 Min. :0.4199

1st Qu.: 122947 1st Qu.:0.03810 Class :character 1st Qu.:0.5413

Median : 244582 Median :0.04864 Mode :character Median :0.5891

Mean : 399014 Mean :0.05324 Mean :0.5997

3rd Qu.: 434365 3rd Qu.:0.06333 3rd Qu.:0.6464

Max. :5151546 Max. :0.25298 Max. :0.8529

CoC_Poverty_Estimate_Age_0_17 Birth_Rate Death_Rate

Min. : 491 Min. : 1.356 Min. : 1.058

1st Qu.: 8631 1st Qu.: 9.690 1st Qu.: 7.102

Median : 17175 Median :11.089 Median : 9.095

Mean : 33415 Mean :10.597 Mean : 8.847

3rd Qu.: 33459 3rd Qu.:12.447 3rd Qu.:10.878

Max. :624784 Max. :21.767 Max. :19.086

International_Migration_Rate Domestic_Migration_Rate Party_Democrat

Min. : -1.5099 Min. : -70.4207 Min. : 0.000

1st Qu.: 0.4777 1st Qu.: -4.4478 1st Qu.:0.000

Median : 1.2118 Median : -0.4604 Median :1.000

Mean : 1.8847 Mean : 0.4435 Mean :0.521

3rd Qu.: 2.5789 3rd Qu.: 4.7348 3rd Qu.:1.000

Max. :19.8348 Max. : 61.8406 Max. :1.000

Análisis

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

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)	-1902.6512	229.7788	-8.28	<2e-16 ***
Actual_CoC_Median_Rent	4.4394	0.2819	15.75	<2e-16 ***

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

Residual standard error: 4701 on 3378 degrees of freedom

Multiple R-squared: 0.0684, Adjusted R-squared: 0.06812

F-statistic: 248 on 1 and 3378 DF, p-value: < 2.2e-16

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

Variable Instrumental

```
iv_model <- ivreg(Overall_Homeless ~ Actual_CoC_Median_Rent |
                  Previous_CoC_Neighbor_Median_Rent, data = data)
summary(iv_model)
```

Call:

```
ivreg(formula = Overall_Homeless ~ Actual_CoC_Median_Rent | Previous_CoC_Neighbor_Median_Rent,
      data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-7424.3	-1023.8	-443.4	183.8	73948.0

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1308.5986	255.9324	-5.113	3.35e-07 ***
Actual_CoC_Median_Rent	3.6608	0.3182	11.504	< 2e-16 ***

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

Residual standard error: 4706 on 3378 degrees of freedom

Multiple R-Squared: 0.06629, Adjusted R-squared: 0.06602

Wald test: 132.3 on 1 and 3378 DF, p-value: < 2.2e-16

de aquí pa abajo creo que no lo pelen por ahora

Primera etapa

Realizamos la regresión sobre la variable endógena `Actual_CoC_Median_Rent` sobre la variable instrumental (IV) `Previous_CoC_Neighbor_Median_Rent`

```
# Primera etapa: Regresión de la renta actual sobre la renta de los vecinos
model_stage1 <- lm(Actual_CoC_Median_Rent ~ Previous_CoC_Neighbor_Median_Rent, data = data)

data$Fitted_CoC_Median_Rent <- fitted(model_stage1)

head(data$Fitted_CoC_Median_Rent)
```

```
[1] 769.6639 668.3781 574.9581 556.3106 562.6573 793.9245
```