

Modelo de Resultados Potenciales

Alvaro J. Riascos Villegas¹

¹Universidad de los Andes y Quantil

18 de enero de 2026

Resultado potencial

- Rubin, D. 1974. Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies. Journal of Educational Psychology.
- Supongamos que una parte de una población indexada por i , es sometida a un tratamiento (i.e., intervencion binaria, todos al mismo tiempo, una única vez, etc).
- Sea $y_i(1)$ el **resultado potencial** del individuo i en caso de ser tratado y $y_i(0)$ el **resultado potencial** en caso de no ser tratado.
- Para cada i se observa solo uno de ellos. Es decir, $y_i(1)$ se observa solo para algunos i y $y_i(0)$ para los demás.
- Si $T_i \in \{0, 1\}$ denota el tratamiento ($T_i = 1$ denota que el individuo i fue tratado) entonces se observa la variable Y_i :

$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0) \quad (1)$$

Efectos tratamiento: ATE

- El efecto individual del tratamiento se define como:

$$Y_i(1) - Y_i(0) \quad (2)$$

Además de no ser observado (para ningún i), requiere hipótesis fuertes para estimarlo.

- Una alternativa menos ambiciosa es estimar el efecto promedio en la población:

$$ATE = E[Y(1) - Y(0)] \quad (3)$$

donde el valor esperado se calcula sobre la distribución de individuos. Tampoco es observado pero, bajo condiciones naturales, es estimable.

- Implícitamente estamos haciendo dos hipótesis: (1) Consistencia: cuando $T = t$, observamos $Y_i(t)$ y (2) No interferencia: el resultado potencial de un individuo solo depende de su tratamiento y no de los demás.

Estas hipótesis se conocen como: *Stable Unit Treatment Value Assumption* - SUTVA.

Efectos tratamiento: ATT y ATU

- Ahora, quizás si se pueden estimar los siguientes efectos, de ineteres por si mismos, sea posible estimar el ATE.
- Esto son:

$$ATT = E[Y(1) - Y(0) \mid T = 1] \quad (4)$$

$$ATU = E[Y(1) - Y(0) \mid T = 0] \quad (5)$$

Una vez más estos no son observados. Enfoquémonos en el efecto del tratamiento sobre los tratados (ATT). El caso de ATU es similar.

Efecto del tratamiento sobre los tratados: ATT

- Una primera aproximación para estimar el ATT sería usando la:

$$\text{Asociación} = E[Y \mid T = 1] - E[Y \mid T = 0] \quad (6)$$

- Sin embargo esto no es correcto excepto bajo condiciones especiales. Para ver esto, obsérvese que:

$$\begin{aligned} E[Y \mid T = 1] - E[Y \mid T = 0] &= \\ E[Y(1) \mid T = 1] - E[Y(0) \mid T = 0] &= \\ E[Y(1) \mid T = 1] - E[Y(0) \mid T = 1] + E[Y(0) \mid T = 1] - E[Y(0) \mid T = 0] &= \\ ATT + E[Y(0) \mid T = 1] - E[Y(0) \mid T = 0] &= \\ ATT + \text{Sesgo} \end{aligned}$$

- El **Sesgo** mide la diferencia ex ante (en el resultado potencial de no ser tratado) entre el grupo tratado y de control..
- Cuando del sesgo es cero: Asociación = ATT.**

Efecto del tratamiento sobre los no tratados: ATU

- De forma análoga al caso anterior, si no hay sesgo ex ante (en el resultado potencial de ser tratado) entre el grupo tratado y de control, entonces: Asociación=ATU
- Luego, si ex ante los grupos tratados y de control son comparables:

$$E[Y(0) \mid T = 1] = E[Y(0) \mid T = 0] \quad (7)$$

$$E[Y(1) \mid T = 1] - E[Y(1) \mid T = 0] \quad (8)$$

entonces:

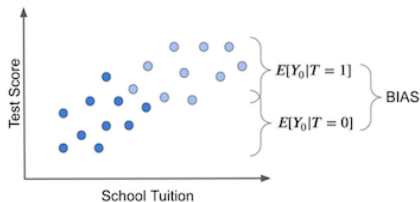
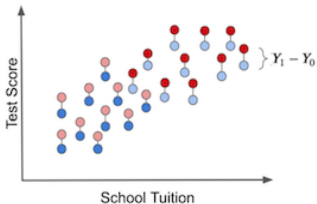
$$E[Y \mid T = 1] - E[Y \mid T = 0] = ATT = ATU$$

- La condición de comparabilidad se llama independencia en media y es una consecuencia de independencia.
- Vamos a hacer el supuesto más fuerte de independencia: $T \perp (Y(0), Y(1))$.

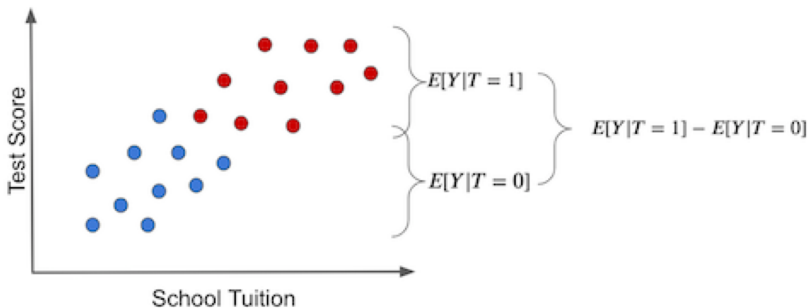
Ejemplo

- Un ejemplo obvio: supongamos que algunos colegios deciden darle una tablet a sus estudiantes y estamos interesados en conocer el efecto de esa intervención sobre el desempeño académico de los estudiantes (medido como puntaje en pruebas estandarizadas).
- Es evidente que los colegios que lo hicieron quizás son colegios con mayores recursos en muchos aspectos, incluyendo profesores mejores pagos, salones más pequeños, padres profesionales, etc.
- Luego, la intervención no es independiente ex ante del resultado potencial: el grupo de intervenidos con o sin tablet, seguramente les iría mejor que al grupo de control.
- Las siguientes gráficas ilustran el problema.

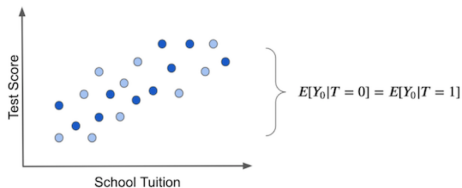
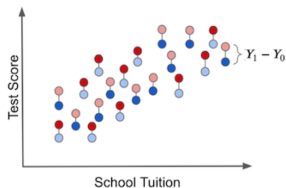
ATE gráficamente: con sesgo



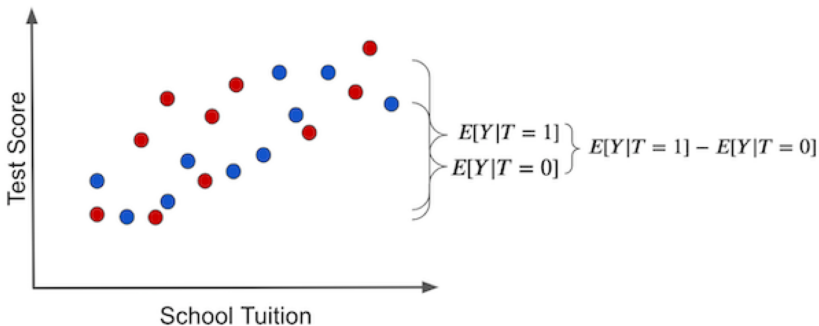
ATE gráficamente: con sesgo



ATE gráficamente: sin sesgo



ATE gráficamente: sin sesgo



Resumen identificación efecto del tratamiento

- Bajo el supuesto más fuerte de **independencia del tratamiento y los resultados potenciales**:

$T \perp (Y(0), Y(1))$. En particular:

$$E[Y(0) \mid T = 1] = E[Y(0) \mid T = 0] \quad (9)$$

$$E[Y(1) \mid T = 1] - E[Y(1) \mid T = 0] \quad (10)$$

el **efecto causal** del tratamiento se puede identificar por:

$$ATE = E[Y \mid T = 1] - E[Y \mid T = 0] = ATT = ATU$$

El estándar de oro que garantiza que esta hipótesis se cumple, es cuando el tratamiento es asignado de forma aleatoria en la población.

Ejercicio: Mostrar que cuando no hay sesgo en el tratamiento de los resultados potenciales, esto no implica que no hay sesgo en el resultado observado. Esto ocurre si y solo si $E[Y(0)] = E[Y(1)]$.

Independencia Condicional

- Una extensión natural del problema anterior es cuando podemos suponer que, condicional a **variables observadas** X , el tratamiento es independiente de los resultados potenciales: $Y(0), Y(1) \perp T \mid X$.
- En este caso un argumento similar muestra que:

$$E[Y \mid X, T = 1] - E[Y \mid X, T = 0] = ATE(X) = ATT(X) = ATU(X) \quad (11)$$

donde $ATE(X)$, $ATT(X)$, $ATU(X)$ son el efecto del tratamiento en la población, sobre los tratados y sobre los no tratados, respectivamente, condicional a X .

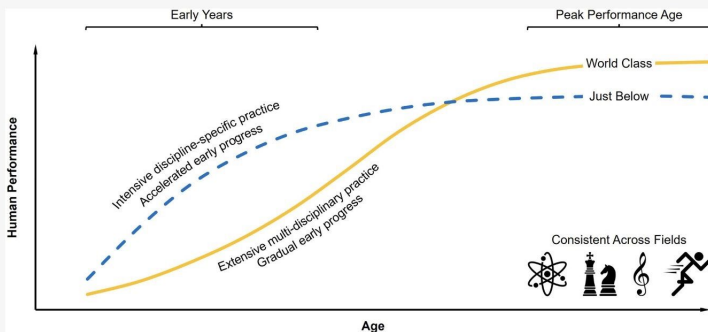
Sesgo

- El sesgo puede tener diferentes orígenes. Aquí estudiaremos dos que son muy relevantes:
 - 1 *Confounding*: ocurre cuando un conjunto de variables X causa tanto el resultado potencial como el tratamiento. En este caso claramente se viola el supuesto de independencia y ocurre cuando no condicionamos por todos los confounders (en principio solo es posible hacerlo por los observados). Ejemplos: (1) Crimen, consumo de helados. (2) Ocupación hotelera, precios habitaciones.
 - 2 Sesgo de selección: ocurre cuando estudiamos el efecto en subgrupos de la población y cada subgrupo revela parcialmente información sobre el tratamiento y el resultado potencial en el subgrupo. Ejemplos: (1) Paradoja de Berkson. (2) Paradoja de Simpson. (3) Monty Hall.

Sesgo de selección



Sesgo de selección



The development of the highest levels of human achievement.

Across domains, world-class performers, compared with peers performing just below this level, engaged in more multidisciplinary practice and showed more gradual performance progress through their early years.

Introducción

- Hasta este punto nos hemos enfocado principalmente en la **identificación** del efecto causal.
- En esta parte nos concentramos en la **estimación** de este efecto.
- Comenzamos con un repaso del modelo de regresión lineal.

Modelo de regresión lineal

- Considere la siguiente especificación lineal (poblacional):

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K + \mu \quad (12)$$

- Típicamente no conocemos la verdadera forma funcional (i.e., no conocemos el verdadero proceso generador de los datos) y solo observamos una muestra de datos.
- Suponiendo una muestra aleatoria de observaciones (i.e., i.i.d), una hipótesis clave para poder estimar β de forma consistente¹ es:

$$E(\mu) = 0 \quad (13)$$

$$\text{Cov}(x_j, \mu) = 0 \text{ hipótesis de exogeneidad.} \quad (14)$$

¹La consistencia de un estimador es lo mínimo que podemos pedir para que este sea de interés.

Modelo de regresión lineal

- Si $\mathbf{x}$ tiene una constante, las condiciones anteriores son equivalentes a $E[\mathbf{x}'u] = 0$, condición de ortogonalidad.
- Ambas condiciones son una consecuencia de $E[\mu | \mathbf{x}] = 0$, hipótesis de esperanza condicional cero.

Formas de endogeneidad

- Tres formas:
 - 1 Variable omitida: el salario de una persona lo determina el nivel de educación observable y la habilidad (no observable). Otro caso común es la autoselección: las personas que reciben un tratamiento son elegibles de forma aleatoria pero el tratamiento es voluntario y puede depender de variables no observadas.
 - 2 Error de medición: medición de sobredosis de consumo de drogas con base en entrada a urgencias en IPS.
 - 3 Simultaneidad: demanda de un bien se determina de forma simultánea con oferta.

Estimación: OLS

- Si se cumple la hipótesis de ortogonalidad y del rango completo entonces β queda **identificado**:

$$\beta = (E[x'x])^{-1}E[x'y] \quad (15)$$

- Dada una muestra aleatoria el estimador muestral de β es **consistente** (estimador OLS).
- Si se cumple la hipótesis ligeramente más fuerte de esperanza condicional cero, entonces el estimador OLS está identificado, es consistente y es **no sesgado**.

Estimación efecto causal

- Considere la siguiente especificación lineal (poblacional):

$$y = \beta_0 + \alpha T + \beta_1 x_1 + \dots + \beta_K x_K + \mu \quad (16)$$

y supongamos que $E[\mu \mid T, \mathbf{x}] = 0$, donde $\mathbf{x} = (1, x_1, \dots, x_K)$.

- Entonces:

$$E[y \mid \mathbf{x}, T = 1] - E[y \mid \mathbf{x}, T = 0] = \quad (17)$$

$$\alpha + E[\mu \mid \mathbf{x}] = \quad (18)$$

$$\alpha \quad (19)$$

y el efecto causal condicional a X queda indentificado por el parámetro α que puede ser estimado por OLS y se puede utilizar toda la teoría de inferencia para el estimador de OLS.

Variables instrumentales

- Suponga que x_K esta correlacionado con μ .
- El método de VI consiste en econtrar una variable z_1 tal que:
 - 1 $Cov(z_1, \mu) = 0$.
 - 2 Si proyectamos x_K en $[1, x_1, \dots, x_{K-1}, z_1]$:

$$x_K = \delta_0 + \delta_1 x_1 + \dots + \theta_1 z_1 + r_K \quad (20)$$

$\theta_1 \neq 0$ (obsérvese que por construcción se cumple las condición de ortogonalidad de r_K y $[1, x_1, \dots, x_{K-1}, z_1]$).

- Esta hipótesis se puede enunciar como, una vez se controla por las demás variables exógenas, x_K está correlacionado con z_1 .
- En este caso decimos que z_1 es un candidato a VI de x_K .

VI

- La segunda ecuación se llama la ecuación en **forma reducida**.
- Con las hipótesis mencionadas se pueden identificar, los coeficientes β como el estimador en dos etapas (TSLS) de VI.
- La primera ecuación, ortogonalidad del instrumento, no es testeable. La forma reducida sí.

Ejemplos: Evaluación de Programas

- En la evaluación de un programa los elegibles se determinan de forma aleatoria.
- Los tratados se autoseleccionan.
- La variable de elegibilidad puede usarse como un instrumento de la variable participación en el programa.

Ejemplos: Salario y educación

- Regresar salarios en experiencia y educación.
- Educación puede no ser exógena porque el error puede incluir variables omitidas no observables: habilidad intelectual, características familiares o variables potencialmente observables: educación de la madre.
- Si usamos la educación de la madre como instrumento de educación probablemente se cumple la condición de la ecuación en forma reducida.
- Sin embargo, la primera queda en duda: características familiares (i.e., educación del padre) pueden correlacionarse con la educación de la madre (i.e., no se cumpliría la hipótesis de exogeneidad del instrumento).
- Alternativamente, el último dígito de la cédula de la persona cumpliría exogeneidad pero no estaría correlacionado con la educación de la persona.

Ejemplos: Salario y educación (Angrist y Krueger (1991))

- Solución de Angrist y Krueger (1991). Considere como instrumento si la persona nació en el primer trimestre del año.
- Exogeneidad es creíble, correlación en forma reducida: la ley obliga por lo menos a algunas personas a matricularse al colegio, los que nacen en el primer trimestre estudian más.
- Problema: que el efecto que en realidad se está midiendo sea solo el de aquellos que fueron obligados a ir a la escuela y por lo tanto tener más tiempo de educación.

- En ocasiones tenemos una variación exógena (aleatoria) en el contexto económico que se correlaciona con la variable endógena: anuncios en TV de desmovilización de las FARC y el clima en el país.

Ejemplos: Experimentos naturales (Salario y participación Vietnam)

- Angrist - Krueger (1991).
- Estudian el efecto de participar en la guerra de Vietnam en los ingresos de los hombres.
- Participar puede no ser exógena a características no observadas.
- Usan como instrumento el número de la balota de prestación del servicio: entre menor el valor más probable ser llamado a prestar servicio.
- Pueden persistir problemas de endogeneidad: si no se controla por educación en la ecuación de ingresos, hombres con balota con número bajo saben que su probabilidad aumenta y eligen tener mayor educación buscando una oportunidad de no prestar servicio.

Ejemplos: No Experimentales (Evans y Schab (1995))

- Estudian el efecto de asistir a colegios católicos en diferentes resultados de los estudiantes.
- Instrumento de atender a un colegio católico: ser católico o no.
- En ocasiones se desea estimar el efecto de una variable endógena y se usa como instrumento la variación en precios de las regiones o impuestos como instrumentos.
- Ejemplo: Si se desea estimar el efecto del consumo de alcohol en el rendimiento universitario se puede usar el precio del alcohol como instrumento del consumo.

Ejemplos: Sesgo de usar OLS

- Usar IV no siempre disminuye el efecto de la variable de interés.
- Card (1995) usa como instrumento si una persona nació cerca a una universidad como instrumento de años de escolaridad.
- Al estimar el efecto de los retornos a la educación los resultados pasan de 7.5 % (OLS) a 13.2 % (IV).

- Otra alternativa para estimar:

$$E[Y | X, T = 1] - E[Y | X, T = 0] \quad (21)$$

bajo el supuesto de independencia condicional se puede motivar con base en el estimador de la media condicional a X, T . Esto requiere de un número adecuado (no vacío) de ejemplos en cada *celdas* de tratamiento y control.

- Una forma de mitigar el problema es usar una función de propensity (al tratamiento) $P(T = 1 | X)$ que, para cada individuo con características X , estime la probabilidad de ser tratados.
- Bajo la hipótesis de independencia:

$$E[Y | X, T] = E[Y | P(X), T] \quad (22)$$

luego la única hipótesis adicional que necesitamos será para mitigar el problema del número adecuado de datos por celda.

- Hipótesis de soporte común. Para todo X

$$0 < P(T = 1 | X) < 1 \quad (23)$$

Propensity score matching

- Entonces la estrategia de estimación es:
 - 1 Estimar una función de propension (i.e., logit, probit).
 - 2 Definir un algoritmo de emparejamiento (i.e., vecinos más cercanos).
 - 3 Estimar el efecto causal y errores estándar.

Motivation

- Between january 2012 and september de 2015, **all homicides and 25 % of all crimes reported in Bogotá occured in 2 % of street segments.**
- During the same period these segments received only 10 % **of the attention of police resources** (Blattman et.al 2017).
- We would like to predict.²

²Á. J. R. Villegas, J. S. M. Pabón, M. Dulce Rubio, S. Quintero, J. G. Vargas and H. García, "Spatio Temporal Sparsity in Homicide Prediction Models," in IEEE Access, vol. 10, pp. 14359-14367, 2022.

Equilibrium

- Prediction models ignore **strategic reaction**.
- We use a unique experimental data to identify the causal impact of police patrolling on crime.
- Use of a **structural model of crime location choice**.
- Estimate **own-and cross-elasticities of crime to patrolling time**.
- **Evaluate alternative patrolling strategies**.

Spatial Discrete Choice Model

- N potential criminal offenders with symmetric preferences, each of them deciding between $J + 1$ locations in the city to commit a crime.
- The associated utility u_{ij} , of agent i , of selecting location j , is given by

$$u_{ij} = \alpha P_j + X_j \beta + \xi_j + \varepsilon_{ij} \quad (24)$$

where:

- P_j police presence in location j .
- X_j : K observed characteristics of the location.
- ξ_j : unobserved characteristics of location j .
- ε_{ij} : idiosyncratic error term.

Spatial Discrete Choice Model

- Assume $\varepsilon_{ij}, \varepsilon_{ij'}$ are i.i.d. extreme value type I distributed, location choice probabilities are:

$$s_{ij}(P_j, X_j, \xi_j; \alpha, \beta) = \frac{\exp(\alpha P_j + X_j \beta + \xi_j)}{1 + \sum_{k=1}^J \exp(\alpha P_k + X_k \beta + \xi_k)} \quad (25)$$

where option $j = 0$ is assumed to be the **outside option**.

- By symmetry of preferences:

$$S_j(P_j, X_j, \xi_j; \alpha, \beta) = s_{ij}(P_j, X_j, \xi_j; \alpha, \beta)$$

Spatial Discrete Choice Model

- Own- and cross-elasticities of crime:

$$\frac{\partial S_j}{\partial P_\ell} = \begin{cases} \alpha S_j(1 - S_j) & \text{if } j = \ell \\ -\alpha S_j S_\ell & \text{if } j \neq \ell \end{cases} \quad (26)$$

and

$$E_{S_j, P_\ell} \equiv \frac{\partial S_j}{\partial P_\ell} \frac{P_\ell}{S_j} = \begin{cases} \alpha(1 - S_j) P_j & \text{if } j = \ell \\ -\alpha S_\ell P_\ell & \text{if } j \neq \ell \end{cases}. \quad (27)$$

Estimation

- To estimate the structural parameters $\theta = (\alpha, \beta)$ from equation (24) we note that:

$$\delta_j = \log(S_j) - \log(S_0) = \alpha P_j + X_j \beta + \xi_j, \quad (28)$$

Estimation: Endogeneity

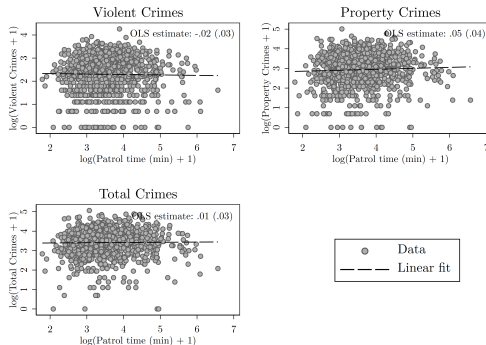


Figura 1: OLS estimation: Biased estimates

Estimation: TSLS

- Starting in January 2016 and during 8 months, 756 out of 1,919 street segments labeled as crime hot spots (out of the 136,984 street segments) received a doubled patrolling time
- We used this randomized treatment to **instrument** the police presence P_j and identify α .

Results: Estimation Direct Elasticities

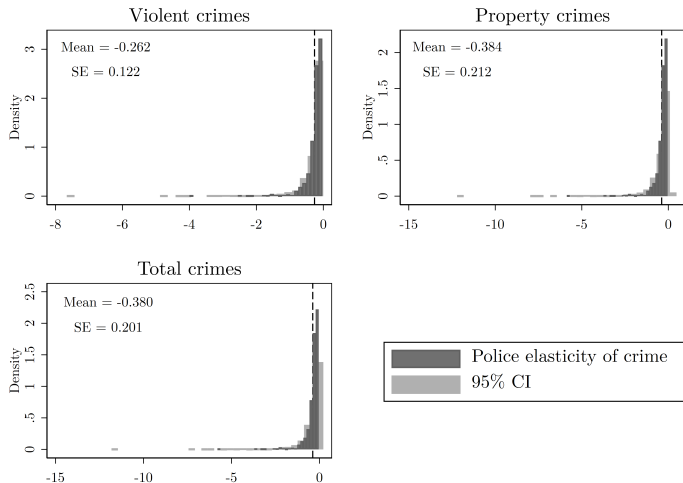


Figure 2: Impact of police presence on crime in the same location.

Results: Estimation Cross Elasticities

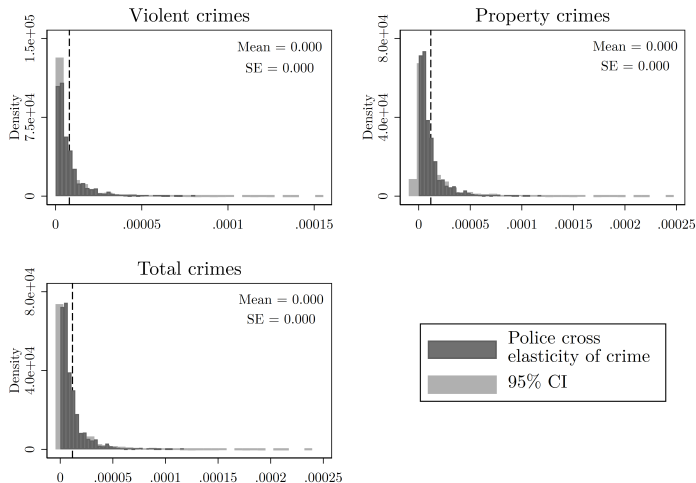
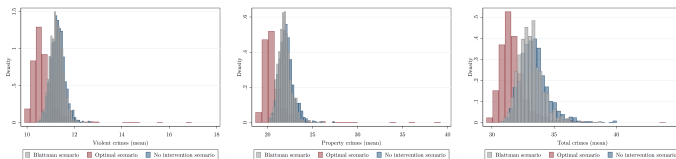


Figure 3: Impact of police presence on crime in different locations.

Optimality

Panel A. Bootstrap distributions of the predicted average amount of crimes per quadrant

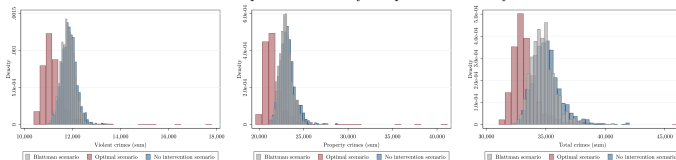


(a) Violent crimes

(b) Property crimes

(c) Total crimes

Panel B. Bootstrap distribution of the predicted sum of crimes



(d) Violent crimes

(e) Property crimes

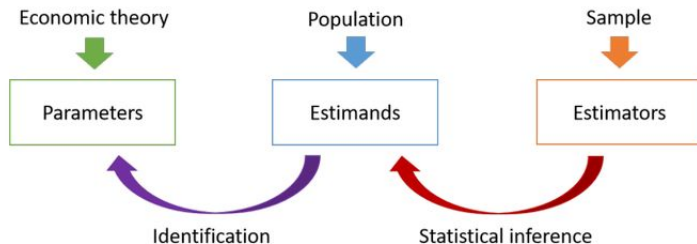
(f) Total crimes

Figure 2: Each figure displays the bootstrap distributions of the predicted average amount of crimes per quadrant (Panel A) and predicted sum of crimes in the city (Panel B) under the original police patrol scenario (gray bars), under the optimal police patrol scenario (red bars), and under the no-intervention scenario (blue bars).

Conclusions

- Switching to the optimal time allocation policy, average reduction of
 - -7 % violent crimes per quadrant.
 - -8.5 % in property crimes.
 - -5.2 % in total crimes.
- That is a reduction of 862 violent crimes, 1,919 property crimes and 1,763 in one year.

Econometrics: The Big Picture



Make life easier by separating the *statistical task* (inferring estimands from data) from the *modeling task* (picking estimands that identify parameters)