

# Replicación de Leonardi y Moretti (2023): aglomeración de restaurantes en Milán

Luis Alejandro Rubiano Guerrero

202013482

la.rubiano@uniandes.edu.co

Andrés Felipe Rosas Castillo

202013471

a.rosas@uniandes.edu.co

Carlos Andrés Castillo Cabrera

202116837

ca.castillolc1@uniandes.edu.co

## I. REPLICACIÓN FIGURA 1 LEONARDI Y MORETTI (2023)

### Introducción

En este ejercicio replicamos la Figura 1 de Leonardi y Moretti (2023), que describe la distribución espacial de restaurantes en Milán y su evolución entre 2004 y 2012. El objetivo es reconstruir los mapas de restaurantes per cápita por barrio y el crecimiento porcentual en el número de restaurantes per cápita.

### Datos y preparación

Trabajamos con 180 zonas censales (zona180) de la ciudad. Para cada zona contamos con: (i) geometría de los barrios (barrios), (ii) población nocturna y diurna (poblacion, variables nite\_pop y day\_pop) y (iii) registros de restaurantes en 2004 y 2012 (restaurants). A partir de esta información construimos, para cada año, el número de restaurantes por barrio.

### Metodología

Para cada barrio calculamos el número de restaurantes por mil habitantes diurnos, pc\_rest\_2004 y pc\_rest\_2012, usando day\_pop. Luego obtenemos la desviación respecto al promedio de la ciudad en cada año (diff\_city\_2004 y diff\_city\_2012) y el crecimiento porcentual del número de restaurantes per cápita entre 2004 y 2012 (growth\_pc). Estas variables se representan mediante mapas temáticos por barrio con intervalos fijos, siguiendo la escala del artículo original.

**Supuestos de medición:** Consideramos población diurna (day\_pop) como denominador de los indicadores per cápita, dado que los restaurantes atienden principalmente a la población presente durante el día. Un establecimiento se clasifica como restaurante activo en un año si tiene coordenadas reportadas o aparece identificado como ethnic, Michelin o sit-down. Los mapas coropléticos se construyen con cortes fijos en la variable de interés, siguiendo la escala de la Figura 1 del artículo original.

### Resultados

La parte superior de la Figura 1 muestra que ya en 2004 existían diferencias en la dotación de restaurantes per cápita por barrio, pero la dispersión alrededor del promedio de la ciudad era relativamente moderada: muchos barrios se encuentran en el rango de más o menos dos restaurantes por cada mil

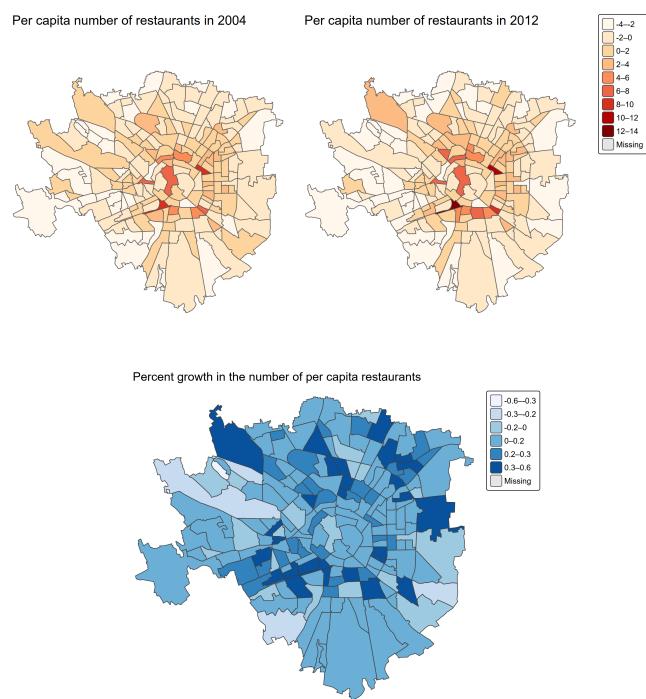


Figura 1: Número de restaurantes per cápita por barrio en 2004 y 2012 (panel superior) y crecimiento porcentual del número de restaurantes per cápita entre 2004 y 2012 (panel inferior), replicando la Figura 1 de Leonardi y Moretti (2023).

habitantes diurnos, y sólo unos pocos presentan concentraciones claramente por encima de la media. Esto es consistente con el rol de la antigua regulación de distancias mínimas, que mantenía una distribución de restaurantes relativamente uniforme entre zonas, a pesar de las diferencias subyacentes en densidad de empleo, renta y atractivo urbano.

En 2012, siete años después de la liberalización de la entrada, el mapa de restaurantes per cápita exhibe un patrón mucho más polarizado. Aparece un conjunto de barrios con una alta concentración de restaurantes (valores muy por encima de la media de Milán) y, al mismo tiempo, un grupo amplio de zonas con niveles sistemáticamente inferiores al promedio. Desde la perspectiva de economía urbana, esto es lo que cabe esperar si existen externalidades de demanda y economías de aglomeración en el sector de restaurantes: cuando la

entrada deja de estar administrativamente restringida, los nuevos establecimientos tienden a ubicarse precisamente donde ya hay muchos restaurantes, porque allí encuentran más flujo de personas, mayor variedad que atrae consumidores, mejor información y complementariedades en la oferta.

El panel inferior, que muestra el crecimiento porcentual de restaurantes per cápita entre 2004 y 2012, confirma esta dinámica de divergencia. Los mayores crecimientos se concentran en los mismos barrios que ya tenían una situación relativamente ventajosa, mientras que muchas zonas parten de niveles bajos y crecen poco o incluso pierden restaurantes per cápita. Esta evolución es coherente con un proceso de “ganadores y perdedores” propio de modelos con externalidades auto-reforzadas, donde algunos barrios se consolidan como polos de amenidades de consumo, mientras otros quedan rezagados, generando una estructura urbana más desigual en el acceso local a restaurantes.

## II. DISTRIBUCIÓN DE PRECIOS

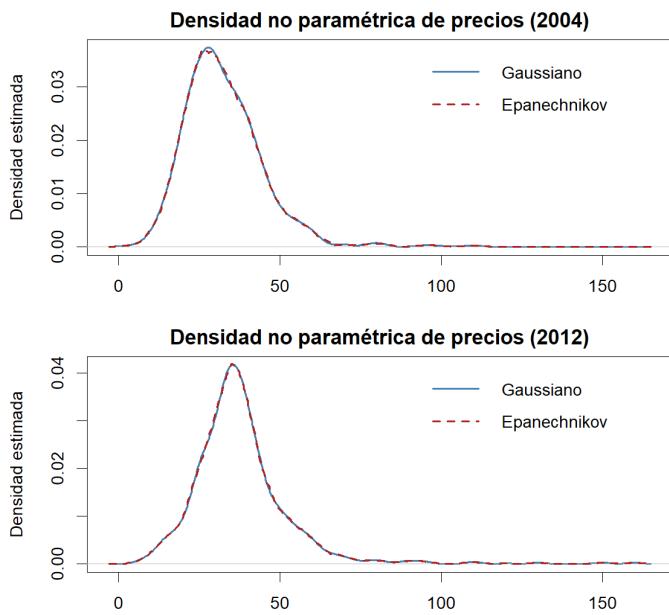


Figura 2: Distribución no paramétrica de precios de restaurantes en 2004 y 2012. Las líneas continuas corresponden al kernel gaussiano y las líneas punteadas al kernel Epanechnikov. En ambos casos el ancho de banda se elige mediante la regla de Silverman (rule-of-thumb).

### Distribución no paramétrica de precios (kernels gaussiano y Epanechnikov)

A partir de las observaciones de precios no faltantes (810 restaurantes en 2004 y 981 en 2012), estimamos densidades kernel usando dos funciones de ponderación: gaussiano y Epanechnikov. En ambos años el ancho de banda se eligió mediante la regla de Silverman (rule-of-thumb), lo que produce valores de  $h_{2004} \approx 3,1$  y  $h_{2012} \approx 2,6$ .

En el cálculo implementamos la versión robusta de la regla de Silverman,

$$h = 1,06 \min\{\hat{\sigma}, \text{IQR}/1,34\} n^{-1/5},$$

de modo que el ancho de banda se determina por el menor entre la desviación estándar muestral y una medida robusta de dispersión basada en el rango intercuartíl. Esta elección tiende a producir valores de  $h$  algo menores que la versión “clásica” basada solo en  $\hat{\sigma}$ , por lo que nuestras densidades muestran algo más de variación local que estimaciones alternativas con un  $h$  más grande.

Las estimaciones para los dos kernels son prácticamente indistinguibles: las curvas gaussiana y Epanechnikov se superponen casi por completo tanto en 2004 como en 2012. Esto sugiere que, dado el tamaño muestral y el ancho de banda empleado, la forma estimada de la distribución es muy poco sensible a la elección del kernel.

En términos de forma, en ambos años la densidad presenta una única moda en el rango de 30–40 euros y una cola derecha alargada, que refleja la presencia de algunos restaurantes con precios considerablemente más altos. Entre 2004 y 2012 la distribución se desplaza hacia la derecha (mayores precios típicos) y la cola derecha se hace algo más pesada, lo que indica un aumento tanto en el nivel medio como en la dispersión de precios del sector de restaurantes en Milán.

### Distribución no paramétrica con kernel Epanechnikov

Para cada año estimamos la distribución no paramétrica de precios usando un kernel Epanechnikov. Como referencia tomamos el ancho de banda “rule-of-thumb” calculado en el apartado anterior,  $h_{2004}$  y  $h_{2012}$ , y construimos tres estimaciones por año: (i) el ancho de banda “rule-of-thumb”  $h$ , (ii) un ancho de banda reducido  $h/2$  (sub-suavizado) y (iii) un ancho de banda ampliado  $2h$  (sobre-suavizado). Las seis curvas resultantes se muestran en la Figura 3.

En ambos años la distribución estimada de precios es unimodal y se concentra en el rango aproximado de 20–40 euros. La estimación con  $h/2$  presenta más variación local y varios “picos” pequeños, reflejando una mayor varianza de la estimación. En contraste, la curva asociada a  $2h$  es mucho más suave y tiende a aplandar la cola derecha, lo que indica un sesgo mayor por sobre-suavización. El ancho de banda “rule-of-thumb”  $h$  ofrece un compromiso razonable entre ambos extremos, capturando bien la forma general de la distribución sin introducir demasiada irregularidad.

Comparando 2004 y 2012, las formas de las distribuciones con kernel Epanechnikov son muy similares a las obtenidas con el kernel gaussiano: la distribución de 2012 presenta una moda ligeramente más alta y una concentración algo mayor alrededor del pico, lo que sugiere precios algo más homogéneos en 2012. En todo caso, las conclusiones cualitativas sobre la distribución de precios son robustas tanto a la elección del kernel como a cambios razonables en el ancho de banda.

### Concentración de precios y explicación teórica

Las densidades no paramétricas muestran una masa de probabilidad elevada en un rango relativamente estrecho de precios (en torno a 20–40 euros), con una cola derecha más bien delgada. Esto indica una clara concentración alrededor de un “precio de referencia” del mercado: muchos restaurantes fijan

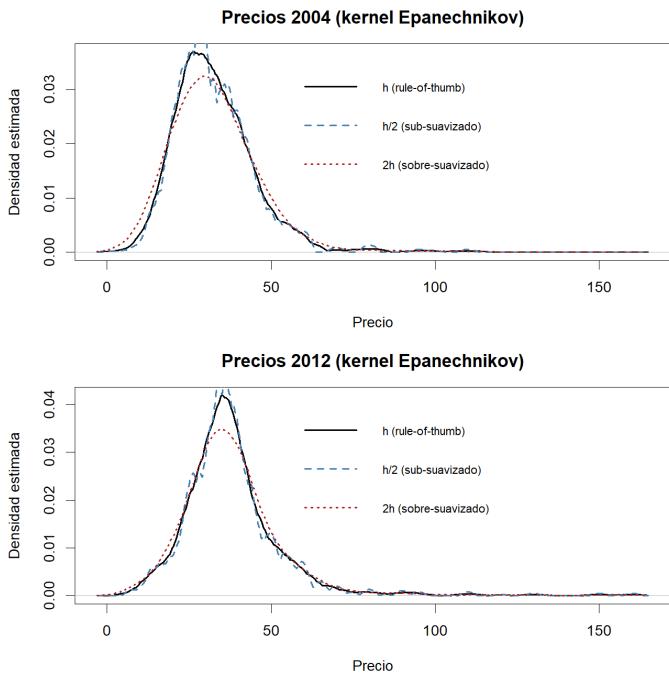


Figura 3: Estimación no paramétrica de la distribución de precios con kernel Epanechnikov para 2004 y 2012, utilizando tres anchos de banda:  $h$  (rule-of-thumb),  $h/2$  y  $2h$ .

precios muy similares para menús comparables y la dispersión observada es acotada.

Este patrón es coherente con modelos de competencia en precios con costos de búsqueda y con economías de aglomeración en el sector de restaurantes. Cuando muchos locales se concentran en una misma zona, la aglomeración genera externalidades de demanda: la presencia de varios restaurantes cercanos atrae más clientes al área y aumenta el mercado potencial de cada firma. Al mismo tiempo, la proximidad geográfica facilita la comparación de precios, de modo que un restaurante que se desvía demasiado por encima del nivel modal corre el riesgo de perder una fracción importante de su clientela.

En equilibrio, la combinación de presión competitiva (por comparación de precios) y cierta diferenciación de producto (calidad, tipo de cocina, servicio) induce una distribución de precios con un pico marcado alrededor de un precio de referencia y colas relativamente delgadas, tal como se observa en las estimaciones empíricas. A diferencia del sector transable, donde las economías de aglomeración suelen modelarse como aumentos de productividad (vía labor pooling o knowledge spillovers como lo exponía Moretti y Greenstone), en el sector de restaurantes el canal dominante son estas externalidades de demanda, en línea con los modelos teóricos de aglomeración para comercio minorista y servicios locales.<sup>1</sup>

<sup>1</sup>Ver, por ejemplo, Varian (1980), Stahl (1982), Wolinsky (1993), Dudey (1990), Fischer y Harrington (1996), Bester (1998) y Konishi (2005).

### III. TEST DE DURANTON Y OVERMAN (2005)

En este apartado implementamos el test de localización de Duranton y Overman (2005) utilizando las variables construidas en el punto 1, en particular el número de restaurantes por mil habitantes diurnos en 2004 y 2012 (`pc_rest_2004`, `pc_rest_2012`) y el crecimiento porcentual per cápita (`growth_pc`). A partir de `growth_pc` identificamos los cinco barrios con mayor crecimiento en el número de restaurantes per cápita entre 2004 y 2012 y nos concentraremos en la distribución espacial de los restaurantes dentro de esa submuestra de barrios.

#### Implementación del test.

Primero seleccionamos los cinco barrios con mayor `growth_pc` y extraemos sus polígonos a partir del objeto espacial `milan`, transformando las geometrías al sistema de coordenadas proyectado UTM 32N para trabajar en metros. Luego, para cada año  $t \in \{2004, 2012\}$ , construimos el conjunto de restaurantes ubicados en esos cinco barrios que cuentan con coordenadas geográficas (`lat_t`, `long_t`) y calculamos todas las distancias bilaterales entre restaurantes. Las distancias se expresan en kilómetros y se recortan al intervalo  $(0, 1]$ , tal como indica el enunciado.

Sea  $d$  la distancia entre dos restaurantes. Para cada año estimamos la densidad de la distribución de distancias observadas  $f_{obs,t}(d)$  sobre  $[0, 1]$  mediante un estimador de densidad kernel univariado,

$$\hat{f}_{obs,t}(d) = \frac{1}{n_t h_t} \sum_{i=1}^{n_t} K\left(\frac{d - d_i}{h_t}\right),$$

donde  $K(\cdot)$  es un kernel simétrico (en la implementación utilizamos un kernel de Epanechnikov, aunque los resultados son robustos a usar un kernel Gaussiano) y  $h_t$  es el ancho de banda elegido mediante la regla de Silverman (`bw = "nrd"` en R). La densidad se evalúa en una malla regular de 60 puntos entre 0 y 1 km.

Bajo la hipótesis nula de Duranton y Overman, las ubicaciones de los restaurantes en los barrios analizados constituyen una muestra aleatoria simple del conjunto de ubicaciones potenciales dentro de la región de análisis, condicional en  $N_t$ , todo subconjunto de  $N_t$  sitios potenciales es igualmente probable. Para aproximar esta distribución nula, tomamos la unión de los cinco barrios de mayor crecimiento como espacio de ubicaciones potenciales y realizamos  $B = 500$  simulaciones. En cada simulación  $b$ , muestramos aleatoriamente  $N_t$  puntos dentro de los polígonos (donde  $N_t$  es el número de restaurantes observados en esos barrios en el año  $t$ ), calculamos las distancias bilaterales simuladas en el intervalo  $(0, 1]$  km y estimamos la densidad kernel  $\hat{f}_{b,t}(d)$  usando el mismo kernel y ancho de banda que en el caso observado. Para cada punto de la malla  $d_j$  construimos una banda de simulación a partir de los percentiles empíricos 5 y 95 de  $\{\hat{f}_{b,t}(d_j)\}_{b=1}^B$ , obteniendo así funciones  $\hat{f}_t^{5\%}(d)$  y  $\hat{f}_t^{95\%}(d)$ .

Finalmente, para cada año graficamos la densidad observada  $f_{obs,t}(d)$  junto con la región sombreada delimitada por las

curvas  $\hat{f}_t^{5\%}(d)$  y  $\hat{f}_t^{95\%}(d)$ . Valores de la densidad observada persistentemente por encima (por debajo) de la envolvente simulado indican localización (dispersión) estadísticamente significativa a las distancias correspondientes.

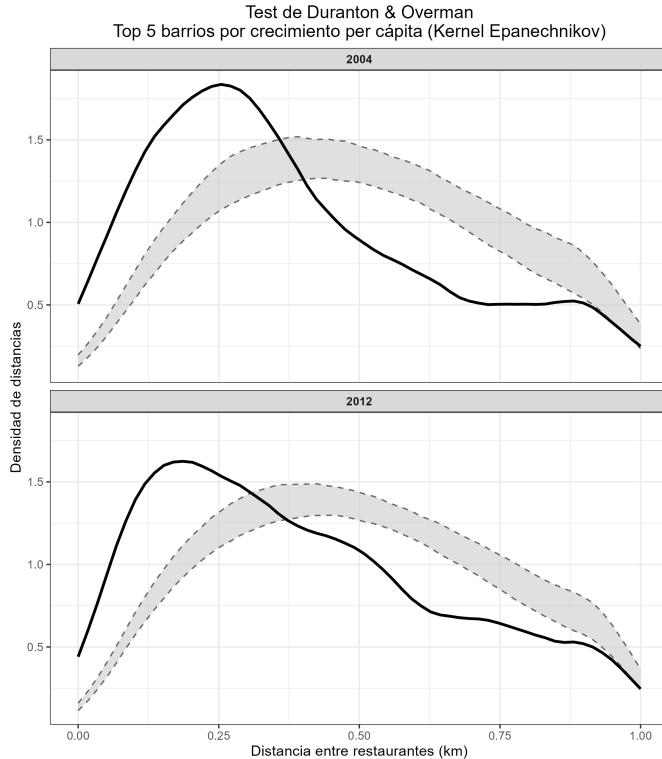


Figura 4: Test de Duranton y Overman para los cinco barrios con mayor crecimiento en restaurantes per cápita, 2004 y 2012. La línea continua muestra la densidad observada de distancias entre restaurantes y la banda sombreada corresponde al intervalo entre los percentiles 5 y 95 de las simulaciones bajo la hipótesis nula.

## Resultados

En ambos años la densidad de distancias observada se sitúa sistemáticamente por encima de la banda de simulaciones para distancias muy cortas, lo que indica localización (aglomeración) de restaurantes a pequeña escala dentro de los cinco barrios de mayor crecimiento. En 2004 la curva observada supera el límite superior de la envolvente para distancias aproximadamente entre 0 y 0.36 km, mientras que en 2012 esto ocurre para distancias entre 0 y 0.31 km. Es decir, hay más pares de restaurantes muy cercanos entre sí de lo que cabría esperar si las ubicaciones fueran puramente aleatorias dentro de los barrios considerados.

Para distancias ligeramente mayores no encontramos evidencia estadística de localización ni de dispersión: en 2004 la densidad observada se mantiene dentro de la banda de simulación aproximadamente entre 0.36 y 0.40 km, y en 2012 entre 0.31 y 0.375 km. A partir de estos puntos, la curva observada cae por debajo del límite inferior de la envolvente (en torno a 0.40 km en 2004 y 0.375 km en 2012) y permanece por debajo de la banda en buena parte del intervalo hasta 1 km,

lo que es consistente con dispersión relativa a distancias medias y largas: hay menos pares de restaurantes a estas distancias de los que se obtendrían bajo la hipótesis nula de ubicaciones aleatorias.

En conjunto, el test de Duranton y Overman proporciona evidencia robusta de localización de restaurantes a muy corta distancia y de dispersión a distancias medias en los cinco barrios que más expandieron su oferta per cápita entre 2004 y 2012, tanto al inicio como al final del período. Este patrón es coherente con la interpretación de las actividades de restauración como un sector fuertemente sujeto a economías de aglomeración vía externalidades de demanda: los restaurantes se benefician de ubicarse muy cerca de otros restaurantes porque atraen un mayor flujo de potenciales consumidores, aun a costa de una mayor presión competitiva en precios.

Como comprobación de robustez, re-estimamos el test de Duranton–Overman utilizando un kernel gaussiano. La forma de  $\hat{f}_{obs,t}(d)$  y los tramos en los que la densidad observada se sitúa por encima (aglomeración a corta distancia) o por debajo (dispersión a distancias medias) de la envolvente de simulación son muy similares al caso Epanechnikov, por lo que las conclusiones sobre la localización de restaurantes resultan prácticamente invariantes a la elección del kernel (en este caso en particular, Gaussiano-Epanechnikov).

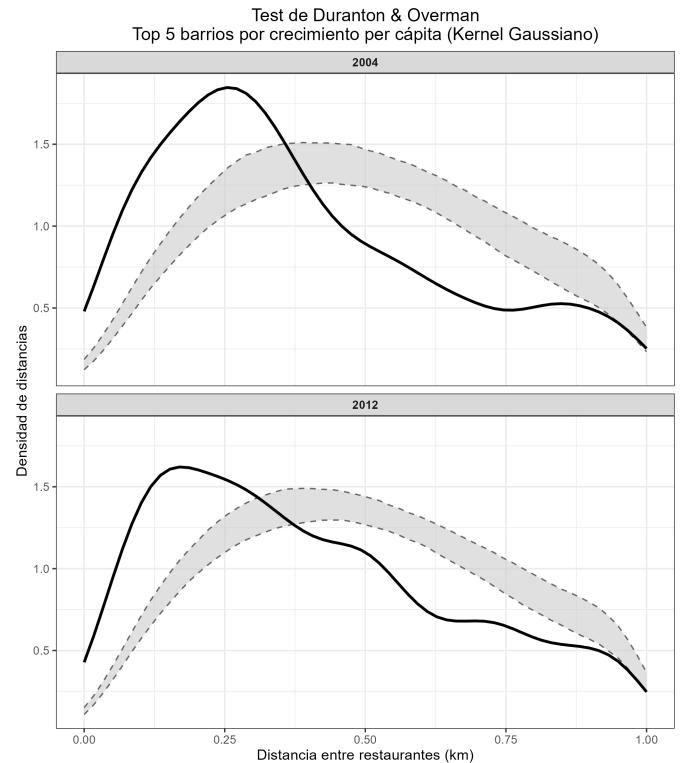


Figura 5: Test de Duranton y Overman para los cinco barrios con mayor crecimiento en restaurantes per cápita, 2004 y 2012. La línea continua muestra la densidad observada de distancias entre restaurantes y la banda sombreada corresponde al intervalo entre los percentiles 5 y 95 de las simulaciones bajo la hipótesis nula.

#### IV. CONCLUSIONES FINALES

En conjunto, los ejercicios muestran un cuadro coherente de cómo la liberalización de la entrada en el sector de restaurantes en Milán se traduce en una reconfiguración espacial y en cambios en la estructura de precios. La replicación de la Figura 1 del paper de Leonardi Moretti (2023) revela que, tras la reforma, algunos barrios se consolidan como polos de alta densidad de restaurantes per cápita mientras otros quedan rezagados, en línea con un proceso de divergencia impulsado por economías de aglomeración. El análisis no paramétrico de la distribución de precios muestra una fuerte concentración alrededor de un rango relativamente estrecho y colas derechas moderadas, consistente con modelos de competencia en precios con búsqueda y diferenciación de producto, en los que la proximidad de establecimientos limita la dispersión de precios. El test de Duranton y Overman aplicado a los cinco barrios de mayor crecimiento confirma la presencia de localización significativa a distancias muy cortas y de dispersión relativa a distancias medias, resultado robusto a la elección del kernel en la estimación de densidades. En conjunto, la evidencia empírica respalda la interpretación de los restaurantes como un sector fuertemente afectado por externalidades de demanda: los establecimientos tienden a concentrarse en unos pocos barrios donde se benefician de mayor tráfico de clientes y complementariedades en la oferta, a la vez que se genera una estructura urbana más desigual en el acceso local a amenidades de consumo.

#### REFERENCIAS

- Duranton, G., y H. G. Overman (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4), 1077–1106.
- Leonardi, M., y E. Moretti (2023). The Agglomeration of Urban Amenities: Evidence from Milan Restaurants. *American Economic Review: Insights*, 5(2), 141–157.
- Notas de clase y material del curso (códigos) de Economía Urbana, Universidad de los Andes, 2025.