



**Tesella: aplicación para la evaluación espacial de impacto de fenómenos urbanos**

**Wady Millett Pardo Russi**

**Universidad del Rosario  
Escuela de Ciencias e Ingeniería  
Bogotá D.C., Colombia  
2025**

**Tesella: aplicación para la evaluación espacial de impacto de fenómenos urbanos**

**Wady Millett Pardo Russi**

Trabajo presentado como requisito para obtener el título de:

**Magíster en Ciudades Inteligentes y Sostenibles**

Director

**Guibor Camargo Salamanca**

**Escuela de Ciencias e Ingeniería  
Maestría en Ciudades Inteligentes y Sostenibles  
Universidad del Rosario  
Bogotá D.C., Colombia  
2025**

## AGRADECIMIENTOS

Agradezco a Catalina Bejarano por facilitar oportunamente el acceso a la información y, sobre todo, por su generosidad intelectual al acompañarme en la reflexión sobre las necesidades de la ciudad y el diseño de herramientas como la aquí presentada. Su apoyo fue decisivo para la materialización de este trabajo.

A Carolina Carvajal, cuyo amor, presencia y cuidado me sostuvieron cuando me sentía sin norte; gracias por ese último empujón anímico que me permitió cerrar este trabajo. Su acompañamiento — más que confianza, una compañía cotidiana— hizo posible llegar hasta aquí.

## ABSTRACT

This project introduces a reproducible tool for exploratory impact evaluation of urban interventions, integrating spatial data processing and econometric estimation within an R+Shiny workflow. The tool operates on 5- and 10-hectare hexagonal tiles; treatment is defined by Project-tile intersection ( $t$ ), and spillovers are modeled through topological neighbor rings ( $n1...n5$ ). Identification relies on a two-way fixed-effects difference-in-differences (TWFE) design with pre-trend checks and placebo tests, producing fully auditable outputs.

While the demonstrative application focuses on economic outcomes, the tool is designed to evaluate a broad range of urban impacts (economic, social, spatial or environmental), depending on the datasets provided. The case study analyzes the depressed interchange at Calle 94–NQS (intervention year 2017) using a 2010–2023 panel built from Bogotá’s Censo Inmobiliario and Cámara de Comercio de Bogotá business registry. These sources were used solely to illustrate the tool’s functionality under real data conditions. In the 10-ha baseline, average effects on core economic outcomes are near zero and statistically insignificant for the  $t$  vs.  $n1$  comparison. At 5 ha, null results persist for  $t$  vs.  $n1$ , while a positive and statistically significant signal emerges for medium-sized firms when controls are moved to more distant rings ( $t$  vs.  $n2...n5$ ), consistent with first-ring contamination and fine-scale spatial diffusion.

The contribution is a standardized, transparent and replicable protocol for institutional monitoring of projects. While its current implementation emphasizes geographic and econometric modules, it operates as a flexible prototype that can be adapted to diverse datasets and outcome domains. Limitations include the partial automation of the current version and the need to scale up for institutional use. The framework is designed for extensibility, enabling its application to multiple types of urban interventions across different contexts worldwide.

**Keywords:** spatial econometrics; difference-in-differences; urban interventions; Bogotá; hexagonal tessellations; spillovers; R Shiny; MAUP.

## RESUMEN

Este proyecto presenta una aplicación replicable para la evaluación exploratoria del impacto de intervenciones urbanas, que integra el procesamiento de datos espaciales y la estimación econométrica en un flujo de trabajo desarrollado en R y Shiny. La herramienta opera sobre teselas hexagonales de 5 y 10 hectáreas; el tratamiento se define por la intersección entre el proyecto y la tesela ( $t$ ), y los efectos de derrame se modelan mediante anillos de vecindad topológicos ( $n1 \dots n5$ ).

Aunque el caso demostrativo se centra en impactos sobre el tejido económico, la herramienta está diseñada para evaluar una amplia gama de efectos urbanos (económicos, sociales, espaciales o ambientales), según las bases de datos que se integren. El caso de estudio analiza el intercambiador deprimido de la Calle 94 con la NQS (año de intervención 2017), utilizando un panel 2010–2023 construido a partir del Censo Inmobiliario y de los registros empresariales de la Cámara de Comercio de Bogotá. Estas fuentes se emplean exclusivamente para demostrar la funcionalidad del prototipo con datos reales, sin constituir su núcleo metodológico. En el escenario de 10 hectáreas, los efectos promedio son cercanos a cero y no significativos; en 5 hectáreas, los resultados se mantienen nulos para  $t$  vs.  $n1$ , pero se observa una señal positiva y significativa para las empresas medianas en anillos más lejanos ( $t$  vs.  $n2 \dots n5$ ), coherente con procesos de difusión espacial del impacto.

Entre las principales limitaciones se encuentran la automatización parcial del prototipo y la necesidad de ampliar su escalabilidad para uso institucional. En conjunto, Tesella ofrece un protocolo estandarizado, transparente y replicable para el monitoreo de proyectos urbanos y la toma de decisiones basadas en evidencia.

**Palabras clave: econometría espacial; diferencias en diferencias; intervenciones urbanas; Bogotá; teselaciones hexagonales; efectos de derrame; R Shiny; problema de la unidad espacial modificable (MAUP).**

## CONTENIDO

CONTENIDO.....	6
INTRODUCCIÓN .....	8
1. OBJETIVOS.....	10
1.1 Objetivo general .....	10
1.2 Objetivos específicos.....	10
2. PROBLEMA Y JUSTIFICACIÓN .....	11
3. MARCO TEÓRICO Y ESTADO DEL ARTE .....	13
3.1 Economías de aglomeración e infraestructura.....	13
3.2 Evaluación de impacto y toma de decisiones .....	15
3.3 Econometría espacial, spillovers y sesgos.....	16
3.4 Casos internacionales: transporte, logística y sostenibilidad .....	17
3.5 Gobernanza, instituciones y distribución de beneficios .....	18
4. METODOLOGÍA PARA LA CONSTRUCCIÓN DEL PROTOTIPO.....	21
4.1 Homogenización espacial con teselas de 5 y 10 hectáreas.....	22
4.2 Módulo geográfico .....	24
4.3 Módulo econométrico.....	25
4.4 Visualización .....	27
5. CONSTRUCCIÓN DEL PROTOTIPO .....	28
5.1 Preparación de datos.....	28
5.2 Módulo geográfico .....	34
5.3 Módulo econométrico.....	41
5.4 Caso de uso.....	46
CONCLUSIONES .....	59
REFERENCIAS .....	62
ANEXOS	64

## INDICE DE FIGURAS

Figura 1 Arquitectura básica del prototipo.....	21
Figura 2 Flujo del módulo geográfico.....	24
Figura 3 Flujo módulo econométrico.....	25
Figura 4 Total matrículas (2010-2023).....	32
Figura 5 Matrículas por tamaño (escala log, 2010–2023).....	32
Figura 6 Matrículas por tamaño (escala log, 2010–2023).....	33
Figura 7 Vista (t–n1 a n5) para 5 ha y otra para 10 ha (comparación de resolución).....	40
Figura 8 Interfaz de procesamiento de vecindades (carga de shapefile de tratamiento y selección de niveles).....	47
Figura 9 Interfaz de procesamiento del módulo econométrico.....	48
Figura 10 Mapa de vecindades (teselas 5 ha) con etiquetas t, n1–n5 — caso Calle 94–NQS.....	50
Figura 11 Mapa de vecindades (teselas 10 ha) con etiquetas t, n1–n5 — caso Calle 94–NQS.....	51
Figura 12 Vista de la tabla de resultados de DiD para 10Ha.....	52
Figura 13 Vista del gráfico de coeficientes para 10Ha.....	53
Figura 14 Vista de las gráficas de pretendencia para 10Ha.....	54
Figura 15 Vista del gráfico de coeficientes para el placebo para 10Ha.....	55

## INTRODUCCIÓN

La evaluación de impacto es una disciplina que permite medir, de manera cuantitativa y causal, el efecto de un fenómeno (como una política pública, una intervención o un evento) sobre la realidad. Sin embargo, a pesar de su relevancia teórica y práctica, esta disciplina enfrenta dos limitaciones sustantivas en el contexto de la gestión urbana. En primer lugar, es costosa: requiere personal altamente calificado, procesamiento intensivo de datos y un tiempo considerable por estudio, recursos con los que las instituciones públicas no siempre cuentan. En segundo lugar, en las ciudades el ejercicio resulta aún más complejo, pues las intervenciones urbanas tienen una naturaleza espacial intrincada que demanda combinar la econometría con análisis geográfico avanzado. La mayoría de las entidades públicas locales carecen de equipos con este conjunto especializado de habilidades, lo cual es comprensible, ya que (a) estos estudios no hacen parte del núcleo misional de dichas entidades y (b) sus altos costos restringen su aplicación en la mayoría de las decisiones. Un contrato promedio de evaluación de impacto en el SECOP ronda hoy los USD \$130 000 y puede requerir al menos seis meses de ejecución.

Si las ciudades pudieran analizar más decisiones bajo estos marcos metodológicos, podrían no solo medir la eficacia de sus políticas (es decir, si se cumplieron las metas físicas, como construir colegios, vías o emitir normas), sino también su efectividad, entendida como la capacidad de dichas intervenciones para modificar comportamientos y resultados sociales o económicos (por ejemplo, cómo cambiaron la criminalidad, los precios del suelo o la generación de ingresos).

Tesella es una aplicación que automatiza la evaluación econométrica del impacto de cualquier fenómeno que tenga expresión geográfica. Desarrollada en el marco de esta tesis, Tesella busca reducir sustancialmente el costo y el tiempo requeridos para que un gobierno urbano realice una evaluación de impacto detallada y espacialmente especializada de una intervención o política pública. La herramienta combina econometría y análisis espacial avanzado, permitiendo que, con solo cargar los polígonos de las áreas tratadas e ingresar un breve formulario, los usuarios generen un estudio completo y reproducible de evaluación de impacto.

Además de su propósito operativo, Tesella se sustenta en una idea central: los efectos de una intervención urbana no se limitan al área directamente tratada, sino que suelen propagarse a zonas vecinas con intensidades y temporalidades diferenciadas. Para capturar esta complejidad, la herramienta estructura la ciudad en teselas hexagonales homogéneas y define anillos de vecindad ( $n_1$ ,  $n_2$ ,  $n_3$ , etc.) que permiten estimar tanto los efectos directos como los derrames espaciales de cada intervención. Sobre esta base, el módulo econométrico implementa modelos de diferencias en diferencias con efectos fijos y pruebas de robustez (pretendencias y placebos), garantizando resultados comparables y auditables.

De este modo, Tesella no es únicamente un experimento técnico, sino un prototipo de política pública basada en evidencia. Al reducir los costos en capital humano y tiempo, la herramienta busca multiplicar exponencialmente la cantidad de decisiones urbanas que pueden tomarse con sustento empírico. Su arquitectura modular (compuesta por un módulo geográfico, uno econométrico y una interfaz interactiva en R Shiny) permite ampliar progresivamente su alcance, incorporar nuevas fuentes de información y adaptar sus funciones a distintos contextos urbanos. En consecuencia, Tesella constituye un paso hacia la institucionalización de mecanismos de monitoreo y evaluación sistemática de las intervenciones urbanas, aplicable en Bogotá y replicable en ciudades de todo el mundo.

Para desarrollar esta propuesta y mostrar el potencial de Tesella en la evaluación de impactos territoriales, el documento se estructura en cinco capítulos. El capítulo 1 presenta el objetivo general y los objetivos específicos del proyecto. El capítulo 2 desarrolla el problema de investigación y la justificación de la herramienta. El capítulo 3 expone el marco teórico y el estado del arte, articulando economías de aglomeración, evaluación de impacto, econometría espacial y experiencias internacionales. El capítulo 4 describe la metodología de construcción del prototipo, incluyendo la homogenización espacial mediante teselas y el diseño de los módulos geográfico y econométrico. El capítulo 5 documenta la implementación de Tesella y el caso de uso aplicado al intercambiador deprimido de la Calle 94 con la NQS. Finalmente, se presentan las conclusiones, seguidas de las referencias y los anexos técnicos.

## 1. OBJETIVOS

### 1.1 Objetivo general

Desarrollar un prototipo en R y Shiny que evalúe de manera exploratoria los impactos de intervenciones urbanas; el cual este basado en diferencias en diferencias y análisis de vecindades, integrando datos espaciales y temporales, y que se pruebe en un caso de uso sobre impactos en el tejido económico de Bogotá.

### 1.2 Objetivos específicos

- 1.2.1 Diseñar un módulo geográfico que combine información geoespacial y de series de tiempo en paneles de datos, permitiendo la identificación de áreas de tratamiento y sus vecindades.
- 1.2.2 Implementar un módulo econométrico inicial basado en diferencias-en-diferencias con efectos fijos, que permita estimar impactos de intervenciones en áreas de tratamiento frente a zonas de control, incorporando controles básicos y pruebas de robustez.
- 1.2.3 Construir una interfaz interactiva en R Shiny que facilite la carga de datos, la ejecución de estimaciones, la visualización de resultados y la descarga de salidas para su análisis posterior.

## 2. PROBLEMA Y JUSTIFICACIÓN

En contextos urbanos a nivel global, evaluar con precisión los impactos de las intervenciones urbanísticas enfrenta tres obstáculos principales: (1) dificultades para medir efectos de infraestructuras existentes y proyectadas; (2) carencias en integración de datos espacio-temporales; y (3) baja capacidad de monitoreo para la toma de decisiones basada en datos. La literatura aplicada en economía urbana muestra que, con el avance de técnicas econométricas y la mayor disponibilidad de datos, es posible identificar impactos con mayor precisión, lo cual es crucial en ciudades contemporáneas y entornos urbanos (rápida urbanización, informalidad, déficits de infraestructura, etc.) [1].

Desde el punto de vista metodológico, ignorar interacciones espaciales y efectos de vecindad induce sesgos en la evaluación de políticas urbanas. La econometría espacial ofrece herramientas para incorporar dependencia espacial en resultados y covariables, y para distinguir difusión de impactos de la mera correlación geográfica [5], [9], [24]. Asimismo, los diseños de diferencias en diferencias (DiD) con múltiples períodos, combinados con pruebas de pretendencias, placebos y correcciones por dependencia espacial, constituyen un modelo de riguroso para intervenciones escalonadas y con potencial de derramamientos (spillover) [21], [22]. En los contextos urbanos actuales, donde las infraestructuras de movilidad (entre otras infraestructuras urbanas) y el reordenamiento urbano generan encadenamientos más allá del área intervenida, estos enfoques son particularmente pertinentes.

La pertinencia de construir una herramienta en R y Shiny radica en cerrar la brecha entre la técnica y la toma de decisiones. El prototipo permitirá procesar información urbana diversa y ofrecer estimaciones basadas en diferencias en diferencias con análisis de vecindades, generando reportes que pueden utilizarse de manera directa en procesos de toma de decisiones y en la formulación de políticas públicas. El objetivo es construir un prototipo funcional, probado en condiciones cercanas a la aplicación real, con capacidad de validar su utilidad en la evaluación de intervenciones urbanas, que capture impactos directos y de derrame (spillover), y que pueda evolucionar hacia modelos

espaciales más exigentes[5], [9], [24], apoyándose en criterios metodológicos robustos [22], [21], [23].

La evidencia internacional reciente refuerza esta motivación: a escala agregada, la infraestructura y las intervenciones urbanas suelen asociarse con transformaciones sostenidas en variables urbanas como el valor del suelo, la distribución de actividades económicas, la intensidad del uso del suelo, la configuración de dinámicas territoriales, etc. Sin embargo, a nivel de proyectos específicos los efectos son heterogéneos y muy dependientes del diseño, el contexto urbano particular y la gobernanza [13]. Casos como Luas Finglas (Dublín) muestran cómo proyectos de transporte masivo, integrados con criterios de sostenibilidad y accesibilidad, pueden dinamizar la actividad comercial de proximidad [14]. A escala interregional, la Belt and Road favorece ingresos y variedad exportadora, incluso con efectos indirectos en territorios vecinos, subrayando la importancia de medir redes y tiempos [15]. En Rusia, estimaciones de panel ilustran impactos positivos del transporte sobre el PIB per cápita cuando se complementa con inversión extranjera, a la vez que alertan sobre ineficiencias modales [16]. En América Latina, evidencia cualitativa en Medellín documenta efectos en cadena por fallas en infraestructura, con impacto diferencial según vulnerabilidad territorial [17]. Y, desde la gobernanza, contrastes entre Indonesia (obra pública como patronazgo) e Iloilo (alineación multiescalar público–privada) revelan cómo las reglas del juego distribuyen beneficios y riesgos sobre el tejido urbano [18], [19].

En resumen, muchas ciudades disponen de abundante información georreferenciada, pero carecen de herramientas para estimar y monitorear los efectos de las intervenciones urbanas. Un prototipo como el planteado, que integre paneles espacio–temporales, estimaciones de diferencias en diferencias y análisis de vecindades ofrece una forma concreta de aprovechar estos datos. En este sentido, más que en la proliferación de sensores o dispositivos IoT, la noción de Smart City se materializa cuando los datos existentes se convierten en insumos efectivos para la toma de decisiones y para una política pública basada en evidencia [4].

### 3. MARCO TEÓRICO Y ESTADO DEL ARTE

El desarrollo de una herramienta para monitorear y evaluar los impactos de las intervenciones urbanas requiere una base teórica que combine aportes de la economía urbana, la econometría espacial, la evaluación de impacto y los estudios de caso internacionales y regionales sobre infraestructura y desarrollo. La literatura muestra de manera consistente que la infraestructura urbana no solo constituye un componente físico del crecimiento, sino también un determinante de las dinámicas urbanas, al influir en la productividad, el valor del suelo, la localización de actividades, la movilidad, la generación de empleo y la distribución espacial de oportunidades.

#### 3.1 Economías de aglomeración e infraestructura

En el campo de la economía urbana, las economías de aglomeración constituyen un marco fundamental para entender cómo la proximidad espacial de empresas y la disponibilidad de infraestructura compartida generan externalidades positivas. Eberts y McMillen demostraron que la infraestructura urbana—redes de transporte, servicios públicos y equipamientos colectivos—actúa como un factor complementario que potencia la eficiencia y la productividad de las firmas localizadas en áreas urbanas densas [2]. Las ciudades ofrecen a las empresas acceso a mercados laborales amplios, mayor diversidad de insumos intermedios y un entorno que facilita la difusión de conocimiento. Este conjunto de externalidades explica por qué la productividad de las empresas manufactureras tiende a ser mayor en grandes ciudades respecto a localizaciones periféricas o dispersas.

En la literatura clásica, las economías de aglomeración se agrupan en tres categorías. En primer lugar, las economías de escala internas a la firma, asociadas a la posibilidad de distribuir costos fijos y especializar procesos a medida que aumenta el volumen de producción. En segundo lugar, las economías de localización, que emergen cuando empresas de un mismo sector se concentran en un territorio, lo que facilita encadenamientos productivos, mercados de trabajo especializados y aprendizajes sectoriales. Finalmente, las economías de urbanización, vinculadas a la diversidad de actividades y funciones presentes en la ciudad, que ofrecen a las firmas acceso a servicios

avanzados, infraestructura compartida y mercados laborales amplios y heterogéneos. Como señala Camagni, estas tres dimensiones permiten entender cómo la estructura urbana, y en particular la configuración de su infraestructura de transporte, condiciona la productividad y la localización de las actividades económicas, ofreciendo un marco para interpretar cómo se generan y distribuyen territorialmente las externalidades de aglomeración.

Desde esta perspectiva, la infraestructura de transporte puede leerse como un dispositivo que reconfigura, en el espacio, la intensidad relativa de las economías de localización y de urbanización, al modificar tiempos de viaje, accesibilidad efectiva a mercados y conectividad entre zonas urbanas y metropolitanas. Melo, Graham y Noland analizaron la relación entre la densidad de infraestructura de transporte y la formación de nuevas empresas en municipios portugueses, encontrando que la disponibilidad de autopistas y ferrocarriles está positivamente asociada con la apertura de plantas productivas [3]. Su evidencia empírica sugiere que las mejoras en la conectividad no solo reducen costos de transporte, sino que amplían mercados efectivos y generan incentivos para la localización de nuevas actividades. Dicho de otra forma, la infraestructura de transporte puede convertirse en un detonante de encadenamientos productivos que trascienden límites administrativos y regionales. Este hallazgo es particularmente relevante en contextos urbanos donde los efectos de nuevas infraestructuras, como sistemas de metro o ampliaciones de corredores de transporte masivo, no se restringen a las áreas directamente beneficiadas, sino que alcanzan barrios y municipios periféricos.

En términos espaciales, este tipo de intervenciones puede interpretarse como una modificación de las áreas de mercado de hogares y empresas. La mejora en la accesibilidad reduce el “costo de distancia” y expande el territorio desde el cual las firmas pueden acceder a clientes, proveedores y fuerza de trabajo, es decir, a sus mercados de producto, de insumos y de trabajo. Al mismo tiempo, amplía el conjunto de oportunidades de empleo y servicios a los que acceden los hogares dentro de un mismo umbral de tiempo de viaje. Esta lectura es consistente con la noción de acceso al mercado desarrollada en la literatura de economía espacial cuantitativa, donde los cambios en infraestructura y costos de transporte redefinen el tamaño efectivo de los mercados de bienes y factores [10], [12].

### 3.2 Evaluación de impacto y toma de decisiones

Uno de los retos en los contextos urbanos actuales es la ausencia de evaluaciones de impacto sobre los proyectos de intervención. Frecuentemente, las decisiones se sustentan en juicios subjetivos o de “expertos”, presiones políticas o diagnósticos incompletos. Bernal y Peña advierten que este déficit analítico no solo reduce la eficacia del gasto público, sino que también debilita la legitimidad de la política urbana, al carecer de evidencia que muestre con claridad los beneficios y costos de cada intervención [4]. Las evaluaciones de impacto permiten comparar resultados entre áreas tratadas y de control, identificar ganadores y perdedores y proponer ajustes que maximicen beneficios y minimicen efectos negativos.

En este sentido, la literatura metodológica aporta herramientas clave: Anselin subraya que ignorar la dependencia espacial en los análisis urbanos puede llevar a conclusiones erróneas [5]. La razón es que las ciudades no son agrupaciones de unidades aisladas, sino sistemas dinámicos e interdependientes: un cambio en la infraestructura de un barrio afecta el desempeño de barrios adyacentes, y estos efectos de vecindad se transmiten en múltiples escalas. Incorporar esta lógica es fundamental para los entornos urbanos actuales, donde los impactos de las intervenciones en movilidad, espacio público o renovación urbana se difunden por redes sociales y económicas que exceden los perímetros de intervención.

Desde la perspectiva de la interacción social, estos patrones de dependencia espacial pueden interpretarse como el resultado agregado de procesos de imitación, aprendizaje, coordinación y presión de pares entre agentes localizados en proximidad física. La literatura de economía urbana ha mostrado que fenómenos como la segregación residencial, la concentración de informalidad o la difusión de violencia se organizan espacialmente porque las decisiones de los hogares y las firmas responden al comportamiento y a las condiciones de su entorno inmediato [1]. De forma similar, estudios de fallas en infraestructura y vulnerabilidad urbana documentan cómo los impactos se propagan a través de redes de servicios y relaciones sociales, amplificando o atenuando choques iniciales según la estructura del tejido barrial [17]. En este marco, los modelos de econometría espacial no solo corrigen un problema estadístico, sino que ofrecen una forma

explícita de incorporar estas interacciones sociales mediadas por el espacio en las evaluaciones de impacto [5].

### 3.3 Econometría espacial, spillovers y sesgos

La econometría espacial se ha consolidado como un campo dedicado a estudiar cómo lo que ocurre en un lugar puede influir en lugares cercanos. En este ámbito se emplean distintos modelos: el SAR (Spatial Autoregressive Model), que analiza la dependencia de los resultados entre áreas; el SEM (Spatial Error Model), que identifica la autocorrelación en los errores o factores no observados; y el SDM (Spatial Durbin Model), que combina ambos enfoques al considerar rezagos tanto en los resultados como en las variables explicativas. Estos modelos ilustran la diversidad de herramientas disponibles para capturar interdependencias espaciales [9]. Sin embargo, para el desarrollo del prototipo que aquí se presenta resulta más adecuado emplear un modelo de diferencias en diferencias con efectos fijos, que permite trabajar con datos panel, identificar impactos a lo largo del tiempo y generar reportes sistemáticos para apoyar la toma de decisiones.

Butts argumenta que las metodologías tradicionales de diferencias en diferencias (DiD) suelen asumir independencia entre las unidades de análisis, lo que puede llevar a subestimar o sobrestimar impactos cuando en realidad existen derrames espaciales, conocidos como spillovers [7]. Este problema toma relevancia en proyectos urbanos como la construcción de líneas de transporte o la renovación de sectores consolidados, cuyos efectos se extienden más allá del área directamente intervenida y modifican la movilidad y la localización de actividades en zonas vecinas. El uso de diferencias en diferencias con efectos fijos resulta especialmente útil en este contexto, ya que controla características que no cambian en el tiempo y mejora la validez de las comparaciones entre áreas tratadas y no tratadas.

La experiencia de Medellín refuerza la necesidad de este enfoque. Un estudio cualitativo sobre fallas en infraestructura crítica mostró que los impactos no se limitaron al área afectada, sino que se propagaron a través de redes de servicios y de comercio, con efectos especialmente severos en asentamientos informales [17]. Este ejemplo muestra la importancia de que los modelos de evaluación capten no solo los efectos directos, sino también los impactos indirectos y distributivos. El modelo de diferencias en diferencias con efectos fijos, que es más sencillo que otros enfoques

espaciales, ofrece una alternativa práctica y adecuada para un prototipo, al permitir generar resultados comparables y útiles para la toma de decisiones públicas.

### **3.4 Casos internacionales: transporte, logística y sostenibilidad**

Un meta-análisis de impactos de infraestructura de transporte demuestra que, a nivel agregado, existe una asociación robusta entre inversión en infraestructura y crecimiento económico, mientras que a nivel de proyecto los resultados son heterogéneos y dependen del contexto y la calidad de la gobernanza [13]. Este contraste entre escalas sugiere que las herramientas de evaluación deben capturar tanto los efectos inmediatos de proyectos específicos como las transformaciones estructurales a largo plazo.

En Dublín, el proyecto Luas Finglas integró sostenibilidad, economía circular y movilidad activa en el diseño de la extensión de un tranvía. Más allá de mejorar la conectividad, el proyecto buscó dinamizar la economía local mediante el desarrollo orientado al transporte (TOD), facilitando comercio de proximidad, densificación y cohesión barrial [14]. Este caso muestra cómo criterios de diseño y participación condicionan los impactos económicos de las intervenciones urbanas.

A escala interregional, la iniciativa Belt and Road (BRI) evidenció incrementos en ingresos por exportaciones y diversidad de productos en países conectados directamente con nuevas redes ferroviarias, pero también beneficios indirectos en países vecinos [15]. Esto resalta la importancia de modelar efectos de red y de accesibilidad efectiva (tiempos y confiabilidad), en lugar de limitarse a medidas de distancia.

En el contexto ruso, un análisis de panel mostró que la infraestructura de transporte y la inversión extranjera directa tienen efectos positivos sobre el PIB per cápita, mientras que otros indicadores, como el número de buses per cápita o la inflación, pueden tener efectos negativos [16]. Este hallazgo sugiere que no basta con aumentar la inversión, sino que es necesario monitorear la composición modal, la eficiencia del gasto y su alineación con objetivos de desarrollo sostenible.

No todos los impactos económicos provienen de megaproyectos de transporte o logística. Las estrategias de regeneración urbana basadas en turismo cultural y emprendimiento también

muestran capacidad para reactivar el tejido económico. El caso del Albergo Diffuso en Sombor, Serbia, demostró que la recuperación patrimonial y el fomento al turismo cultural pueden dinamizar microempresas de servicios y comercio, aunque con riesgos de gentrificación y desplazamiento [20]. Este tipo de experiencias ofrece lecciones sobre cómo diseñar intervenciones que equilibren revitalización económica con inclusión social en contextos urbanos diversos.

### **3.5 Gobernanza, instituciones y distribución de beneficios**

La gobernanza es un factor decisivo en la traducción de infraestructura en beneficios económicos. En Indonesia, la obra pública suele utilizarse como mecanismo de patronazgo: contratos asignados a aliados políticos concentran beneficios en grupos reducidos y limitan los encadenamientos locales [18]. En contraste, el caso de Iloilo (Filipinas) muestra que cuando existe estabilidad política y alineación entre escalas de gobierno, los proyectos de infraestructura se convierten en catalizadores de transformación estructural. Bajo el liderazgo del gobierno central, con apoyo de gobiernos locales y empresarios, la ciudad consolidó un ciclo de grandes obras (como el aeropuerto, diques y la rehabilitación del río) que atrajeron inversión en el sector servicios y tecnológico [19].

Estos ejemplos confirman que la infraestructura no es un factor neutral, sino que sus beneficios dependen de la calidad de las instituciones y de la capacidad de generar consensos. En Bogotá, donde la fragmentación institucional y la rotación de liderazgos son altas, diseñar un sistema de monitoreo transparente puede servir como mecanismo para reducir discrecionalidad y construir confianza.

En resumen, la literatura metodológica contemporánea converge en que el análisis de impactos de intervenciones urbanas debe combinar:

- Paneles espacio–temporales que integren información socioeconómica y territorial en capas geoespaciales.
- Modelos de diferencias en diferencias con múltiples períodos y cohortes escalonadas, que permiten verificar tendencias, realizar placebos y explorar heterogeneidades [21].

- Modelos espaciales como SAR, SEM y SDM, útiles para capturar dependencia entre unidades y distinguir mecanismos de difusión [9], [24].
- Errores robustos (Conley, Driscoll–Kraay) que corrigen la dependencia espacial y temporal de largo alcance [22].
- Controles sintéticos como alternativa cuando no existen grupos de comparación adecuados [23].

En el caso del prototipo que aquí se presenta, la estrategia seleccionada es un modelo de diferencias en diferencias con efectos fijos. Este enfoque resulta más adecuado para una herramienta en etapa inicial, ya que permite trabajar con datos panel, controlar diferencias invariables en el tiempo y generar resultados verificables y comparables de manera sistemática.

En la práctica, los diseños de diferencias en diferencias se han consolidado como el estándar para la evaluación de impacto con datos administrativos y paneles espacio–temporales, porque explotan de manera directa la variación temporal entre áreas tratadas y de control y permiten controlar heterogeneidad no observable que permanece constante en el tiempo [21]. Frente a enfoques basados en propensity score matching, que dependen fuertemente de la correcta especificación de covariables observables y no capturan shocks no observados, el DiD con efectos fijos ofrece una estructura más estable para automatizar análisis repetibles en distintos proyectos. Además, se integra de forma natural con extensiones como estudios de eventos, correcciones de errores robustos ante dependencia espacial y temporal [22] y, en etapas posteriores, con alternativas como el control sintético en contextos con pocos tratados [23]. Por estas razones, el prototipo adopta el modelo de diferencias en diferencias como columna vertebral de la herramienta, dejando métodos basados en matching o controles sintéticos para desarrollos futuros y para aplicaciones que requieran diseños más intensivos en supervisión analítica.

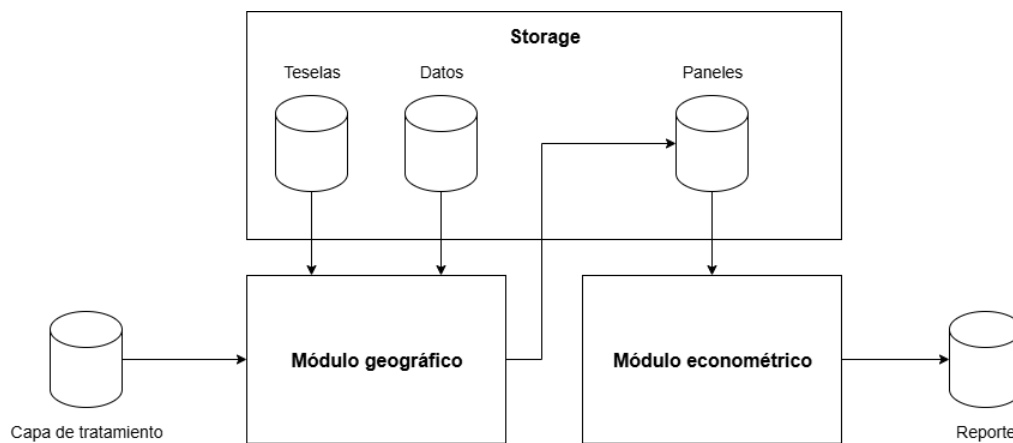
De tal manera que, la construcción del prototipo en R y Shiny debe incorporar los componentes de manera progresiva: comenzar con un módulo de diferencias en diferencias y análisis de vecindades, y avanzar hacia modelos espaciales más sofisticados. Esta evolución no solo permitirá estimar impactos directos, sino también capturar efectos de derrame, heterogeneidades sectoriales y diferencias territoriales, lo que contribuirá a una planificación urbana más efectiva y equitativa.

Finalmente, el estado del arte demuestra que la infraestructura urbana es un determinante clave de las transformaciones en la ciudad, pero que sus efectos son complejos, heterogéneos y dependen de la calidad de las instituciones. En muchos contextos urbanos, la falta de herramientas de evaluación ha limitado la capacidad de planificar con base en evidencia. El desarrollo de un prototipo de monitoreo y evaluación, sustentado en econometría espacial y metodologías de impacto, permitirá superar estas limitaciones. Al integrar experiencias internacionales (Dublín, BRI, Rusia, Indonesia, Iloilo, Medellín, Serbia) con enfoques metodológicos de vanguardia, las ciudades podrán avanzar hacia sistemas de planificación más transparentes, inclusivos y sostenibles.

#### 4. METODOLOGÍA PARA LA CONSTRUCCIÓN DEL PROTOTIPO

La metodología se organiza en torno a dos módulos principales que articulan el funcionamiento del prototipo. El módulo geográfico procesa la información alfanumérica y espacial, la estandariza en unidades homogéneas de análisis (teselas hexagonales), construye paneles espacio-temporales y define áreas de tratamiento y vecindad. Sobre estos insumos, el módulo econométrico aplica un modelo de diferencias en diferencias con efectos fijos, extendido a un estudio de eventos, que permite identificar impactos directos e indirectos de las intervenciones. Esta secuencia metodológica sienta las bases para una arquitectura modular que integra la preparación de datos, la estimación econométrica y la visualización de resultados en una misma herramienta.

Figura 1 Arquitectura básica del prototipo



Fuente: elaboración propia para este proyecto.

El prototipo se organiza en una arquitectura modular que articula el procesamiento geográfico y econométrico a partir de un repositorio común de datos. El flujo general se compone de los siguientes pasos:

- Ingesta de datos y teselas: se cargan las fuentes alfanuméricas y espaciales, junto con la grilla de teselas hexagonales de 5 y 10 hectáreas que homogeniza el espacio de análisis.
- Almacenamiento inicial (Storage): tanto las teselas como los datos alfanuméricos forman parte nativa del prototipo y se integran en un repositorio único. Actualmente se cargan de forma manual o semiautomatizada.

- **Tratamiento:** la capa correspondiente a la obra o proyecto que se quiere evaluar se incorpora de manera independiente y constituye la base para identificar las teselas de tratamiento y sus vecindades.
- **Módulo geográfico:** a partir de la intersección entre teselas, datos y capa de tratamiento, se construyen los paneles espacio-temporales. En esta etapa se asignan las teselas tratadas y se generan los anillos de vecindad que permiten capturar efectos de derrame. La salida de este módulo son los paneles listos para la estimación econométrica.
- **Módulo econométrico:** los paneles resultantes se utilizan para estimar impactos mediante un modelo de diferencias en diferencias con efectos fijos. La información de vecindades (anillos  $t$ ,  $n1$ ,  $n2$ , ...) no se trata como “tratamientos” adicionales, sino como niveles de exposición espacial: permite comparar teselas tratadas con distintos anillos de control y estimar gradientes de impacto según el orden de vecindad. En la práctica, los primeros anillos capturan los efectos directos y de derrame más intensos, mientras que anillos más lejanos operan como zonas de comparación donde se espera que el efecto tienda a desaparecer; el prototipo permite fijar un número máximo de anillos y, en el caso de uso, trabaja con varios órdenes de vecindad alrededor del tratamiento. La salida de este módulo son los resultados cuantitativos de impacto para cada comparación tratamiento-vecindad.
- **Reportes y visualizaciones:** los resultados se consolidan en tablas, gráficos y mapas que muestran la asignación de tratamiento y los impactos estimados, y pueden exportarse junto con los parámetros y metadatos de cada ejecución.

#### **4.1 Homogenización espacial con teselas de 5 y 10 hectáreas**

Un aspecto central de la metodología es la elección de las teselas hexagonales de 5 y 10 hectáreas como unidades de análisis. La decisión se orienta a mitigar algunas manifestaciones del problema de la unidad espacial modificable (MAUP), que afecta a los análisis cuando se usan unidades administrativas de tamaño desigual o con límites que cambian en el tiempo. En la medida en que se desconoce la verdadera extensión espacial de los procesos sociales y económicos, cualquier partición del territorio conserva un componente de arbitrariedad; la teselación no resuelve este problema de fondo, pero sustituye límites político-administrativos variables por unidades de igual área y geometría regular. Esto permite homogenizar el espacio urbano, facilitar comparaciones

longitudinales y reducir sesgos asociados a bordes o a divisiones administrativas contingentes. El hexágono, frente al cuadrado, reduce la proporción de perímetro sobre área y minimiza problemas de orientación.

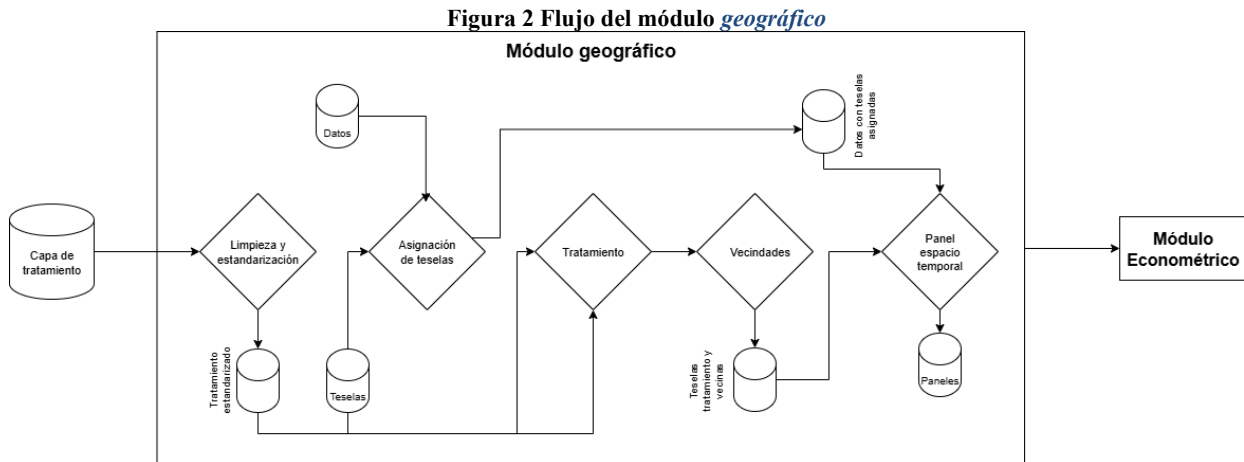
La escala de 5 hectáreas permite capturar micro dinámicas barriales y fenómenos muy localizados, como la apertura de locales comerciales en corredores o la consolidación de micro clústeres de servicios. En contraste, la escala de 10 hectáreas suaviza la varianza y permite observar patrones macro, como la difusión de impactos hacia áreas colindantes o la sustitución espacial de actividades. Usar ambas escalas se plantea como una prueba de robustez frente al MAUP: permite evaluar hasta qué punto los resultados cambian al modificar el tamaño de la unidad de análisis y según la naturaleza de los datos (eventos puntuales agregados a teselas o variables ya observadas por área). Los hallazgos deben interpretarse, por tanto, como válidos para el rango de escalas considerado y para el tipo de dato disponible, reconociendo que el problema de la unidad espacial modificable no desaparece, sino que se hace explícito y se somete a contraste empírico.

En el prototipo actual, las escalas de 5 y 10 hectáreas fueron definidas por el autor como un compromiso entre la necesidad de representar fenómenos con huella muy localizada y la de captar patrones de difusión a escala urbana y metropolitana. Esta elección debe entenderse como una decisión metodológica informada, coherente con la resolución de los datos disponibles y con la escala espacial del caso de estudio, y no como una definición universal de la unidad “correcta” para analizar impactos territoriales. En otros contextos, procesos más micro (por ejemplo, dinámicas intra-manzana) o más agregados (corredores metropolitanos completos) podrían requerir unidades de observación de menor o mayor tamaño.

De cara al desarrollo futuro de Tesella, la teselación se concibe como un marco flexible en el que la escala pueda parametrizarse según la extensión espacial del fenómeno a evaluar y la calidad de la información de entrada. El prototipo documentado en este trabajo fija las teselas en 5 y 10 hectáreas, pero la arquitectura metodológica está pensada para admitir otras combinaciones de tamaño de celda y análisis multiescala. En esa perspectiva, la comparación sistemática de resultados entre distintas escalas se vuelve parte del propio ejercicio de evaluación de impacto, en lugar de un supuesto implícito sobre el que no se reflexiona.

## 4.2 Módulo geográfico

El módulo geográfico se encarga de preparar la información espacial y alfanumérica, homogenizarla en unidades comparables y generar los paneles que servirán como insumo para la estimación econométrica.



Fuente: elaboración propia para este proyecto.

Su funcionamiento parte de un repositorio de insumos nativos, compuesto por la grilla de teselas y los datos alfanuméricos. En el caso del prototipo, se utilizan teselas hexagonales de 5 y 10 hectáreas, conforme a la discusión metodológica de la sección 4.1, pero el tamaño de la tesela es un parámetro que el evaluador puede ajustar de acuerdo con la escala espacial del fenómeno que se desee analizar (por ejemplo, dinámicas microbarriales frente a patrones metropolitanos). A partir de ellos, el módulo organiza una secuencia de etapas que culminan en paneles espacio–temporales con tratamiento y vecindades definidos. Los pasos que sigue el módulo son los siguientes:

- Limpieza y estandarización de capa de tratamiento: se corrigen geometrías, se unifican campos y se asegura que todos los datos estén en el mismo sistema de referencia. El resultado son capas homogéneas y listas para cruzarse con las teselas.
- Asignación de teselas: los datos alfanuméricos se intersectan con la grilla de 5 y 10 hectáreas, de modo que cada observación queda asociada a una tesela y un año.

- Tratamiento: la capa del proyecto o intervención que se desea analizar se cruza con las teselas. De este modo se identifican las unidades tratadas y el año de inicio de la exposición. Esta definición es fundamental para diferenciar áreas tratadas de no tratadas.
- Vecindades: con base en las teselas tratadas se generan anillos de proximidad (n1, n2, n3, etc.) que representan áreas de vecindad. Estos anillos permiten capturar los efectos indirectos o de derrame que suelen acompañar a las intervenciones urbanas.
- Panel espaciotemporal: el módulo concluye con la integración de todas las piezas en un panel que combina variables, estado de tratamiento y vecindades, todo organizado por tesela y año. Este panel constituye el producto final del módulo geográfico y el insumo directo para el módulo econométrico.

### 4.3 Módulo econométrico

La estrategia empírica se basa en un modelo de diferencias en diferencias de dos vías con efectos fijos (TWFE), que permite trabajar con datos panel y controlar factores invariables en el tiempo. La especificación incorpora efectos fijos de unidad y de año, y *dummies* de vecindad para capturar impactos tanto en las teselas tratadas como en los anillos colindantes. De esta manera es posible estimar efectos directos de las intervenciones y los derrames que se producen hacia áreas vecinas.

$$Y_{ti} = \alpha_i + \delta_t + \sum_k \beta_k [1\{vecindad_i = k\} * 1\{t \geq t^*\}] + X_{it}\gamma + \varepsilon_{it}$$

donde:

$i$  indexa las teselas y  $t$  los períodos de tiempo;

$Y_{ti}$  es la variable de resultado en la tesela  $i$  y el período  $t$  (por ejemplo, actividad económica, densidad de locales, indicadores socioeconómicos);

$X_{it}$  es el vector de variables de control observables;

$\alpha_i$  son efectos fijos por tesela, que capturan heterogeneidad no observable invariante en el tiempo;

$\delta_t$  son efectos fijos de tiempo, que recogen choques comunes a todas las teselas en cada período;

$\mathbf{1}\{\text{vecindad}_i = k\}$  es un indicador que toma valor 1 si la tesela  $i$  pertenece al anillo de vecindad  $k$  (por ejemplo,  $k = 0$  teselas tratadas,  $k = 1$  primer anillo,  $k = 2$  segundo anillo, ...,  $k = K$ );

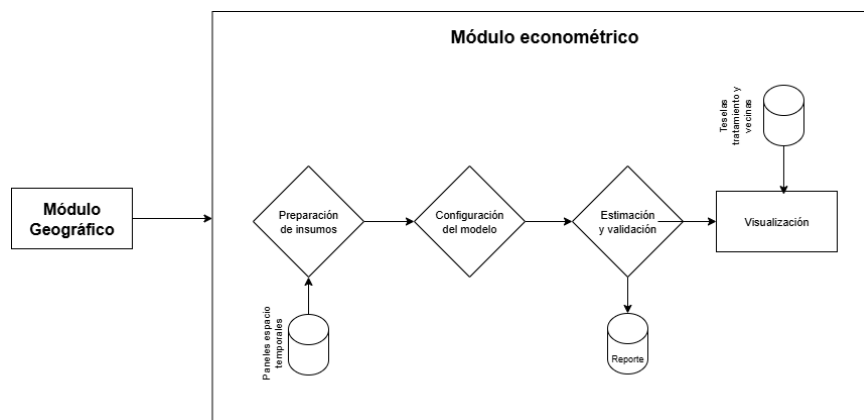
$\mathbf{1}\{t \geq t^*\}$  es el indicador de período posterior a la intervención (post-tratamiento), donde  $t^*$  es el momento de entrada en operación del proyecto;

los coeficientes  $\beta_k$  representan el efecto medio del tratamiento en el anillo  $k$  en comparación con la categoría de referencia, resultado de la interacción entre pertenencia al anillo  $k$  y el período posterior a la intervención;

$\varepsilon_{it}$  es el término de error idiosincrático.

En este diseño, la estructura espacial del modelo se introduce a través de la variable vecindad  $i$ , que clasifica cada tesela según su orden de vecindad respecto a la infraestructura intervenida (tratada, primer anillo, segundo anillo, etc.). Esto permite estimar no solo el impacto directo sobre las teselas tratadas, sino también un gradiente de efectos de derrame por anillos de vecindad, manteniendo la lógica de diferencias en diferencias con efectos fijos por tesela y por período.

Figura 3 Flujo módulo econométrico



Fuente: elaboración propia para este proyecto.

El módulo econométrico toma los paneles espacio-temporales generados por el módulo geográfico y aplica esta estrategia para producir resultados comparables y útiles en la toma de decisiones.

Además de las estimaciones principales, el prototipo incorpora verificaciones metodológicas como pretendencias, placebos y winsorización opcional, lo que refuerza la validez de los resultados.

El flujo metodológico se desarrolla en los siguientes pasos:

- Preparación de insumos: se cargan los paneles generados por el módulo geográfico, que contienen variables de resultado, de control, estado de tratamiento y vecindades.
- Especificación del modelo: se define un esquema de diferencias en diferencias con efectos fijos de unidad y de año, adecuado para capturar impactos directos e indirectos.
- Estimación y validación: se ajusta el modelo a los datos e incorpora diagnósticos de pretendencias, placebos y procedimientos de robustez como la winsorización de valores extremos.
- Resultados econométricos: el módulo produce coeficientes, errores estándar, intervalos de confianza y estadísticas que permiten evaluar la magnitud y significancia de los impactos.
- Paneles de salida: las estimaciones se organizan en tablas y gráficos, incluyendo comparaciones entre áreas tratadas y no tratadas.
- Visualización y exportación: los resultados se presentan en la interfaz Shiny y pueden exportarse junto con metadatos que documentan parámetros y especificaciones.

#### 4.4 Visualización

La plataforma desarrollada en Shiny integra todos los módulos en una interfaz gráfica que permite cargar datos, definir parámetros (año de tratamiento, *outcomes*, covariables), correr estimaciones y visualizar resultados. Los usuarios pueden inspeccionar tablas dinámicas, gráficos de eventos con intervalos de confianza y mapas de impactos directos e indirectos. Los resultados pueden exportarse en formato Excel con metadatos que documentan la versión de datos, parámetros y especificaciones utilizadas. Este diseño asegura trazabilidad y reproducibilidad, elementos clave para la apropiación institucional y la transparencia en la toma de decisiones.

## 5. CONSTRUCCIÓN DEL PROTOTIPO

Este apartado pretende mostrar no solo el funcionamiento técnico de los módulos de la herramienta, sino también la calidad y relevancia de los productos obtenidos. Cada módulo entregó salidas verificables —paneles de datos, indicadores derivados, rutinas de control de calidad— que constituyen la base empírica del análisis econométrico. El objetivo de esta sección no es repetir la descripción de código ya detallada en la metodología, sino mostrar qué se obtuvo concretamente, cómo esos resultados respaldan la solidez del modelo y qué implicaciones tienen para el análisis del tejido económico de Bogotá.

### 5.1 Preparación de datos

Un primer componente corresponde al censo inmobiliario, que aporta información sobre la estructura física y funcional del territorio; el segundo a la Cámara de Comercio de Bogotá, que refleja la dinámica empresarial. La discusión se centra en cómo estos dos paneles, depurados y armonizados, proveen datos consistentes para estimar impactos de intervenciones urbanas sobre el tejido económico y sus heterogeneidades.

#### 5.1.1 Censo inmobiliario

El procesamiento de la información del censo inmobiliario culminó en la construcción de un panel espacial-temporal de 646.310 registros, cada uno asociado a unidades prediales agregadas por manzanas catastrales (*Codigo.manzana*) y referenciadas por coordenadas (*coord\_x*, *coord\_y*). El panel cubre un horizonte temporal de múltiples años (columna *Year*, 2010 a 2023), permitiendo seguir la evolución de la capacidad física del territorio y su uso a lo largo del tiempo. En esta subsección se describen únicamente las variables originales del censo; su agregación posterior a teselas u otras unidades espaciales y la eventual construcción de indicadores relativos al área se detalla en la sección del módulo geográfico.

Las variables se almacenan principalmente en niveles (metros cuadrados construidos, número de predios) y, en algunos casos, como razones simples entre usos; en esta etapa no se ha aplicado una normalización espacial adicional respecto al área de la unidad de análisis.

Entre las variables principales destacan:

- AreaTotal: superficie total construida, indicador de capacidad edificada en cada manzana.
- PredioTotal: número total de predios, que permite aproximar la densidad parcelaria.
- AreaResi / AreaNoResi: área destinada a uso residencial y no residencial, respectivamente, fundamentales para calcular la mezcla funcional.
- PredioResi / PredioNoResi: número de predios por uso, que complementan la medición anterior.
- ManArea: superficie total de la manzana, empleada como denominador común para homogenizar densidades e índices.

El panel resultante cumple dos funciones críticas para la herramienta: primero, ofrece variables de control para los modelos econométricos (área construida, intensidad edificatoria, relación residencial/no residencial), que ayudan a separar el efecto de una intervención del efecto de condiciones físicas preexistentes. Segundo, permite calcular indicadores derivados como índices de construcción y razones de uso, que enriquecen la lectura de la estructura territorial.

La discusión más importante es que este panel no se construyó con un flujo ETL totalmente automatizado, sino a través de un proceso manual y supervisado desde Mapas Bogotá, complementado con códigos auxiliares para homogenizar formatos. Este esfuerzo manual permitió resolver inconsistencias de origen (variación de nombres de campos, codificación de categorías, cambios en los esquemas de exportación) y garantizar que los datos mantuvieran coherencia interanual. El resultado fue un panel consistente, libre de duplicados y con control espacial uniforme, indispensable para estudios longitudinales.

Los beneficios son claros: (i) se reducen riesgos de sesgo por errores de manzana o duplicación de registros; (ii) se asegura la comparabilidad longitudinal, esencial para análisis dinámicos; y (iii) se garantiza la compatibilidad espacial con la malla de teselas, que constituye la unidad de análisis homogénea en los modelos.

En síntesis, el resultado más relevante del módulo de censo inmobiliario es un panel robusto de variables físicas y de uso del suelo, con 646 mil observaciones distribuidas en múltiples años, que

no solo aporta controles confiables a la estimación, sino que también constituye una fuente independiente para interpretar los cambios observados en el tejido económico.

### 5.1.2 Cámara de Comercio de Bogotá

El módulo de la Cámara de Comercio de Bogotá (CCB) transforma los archivos brutos de matrículas activas y canceladas en un panel anual georreferenciado por tesela. A diferencia de interpretar “*activa/cancelada*” como equivalentes de *nacimiento/muerte*, aquí se reconstruye el *stock* de matrículas vigentes por año, complementado con variables de establecimientos, empleo y estructura del tejido (tamaño y naturaleza jurídica), además de métricas derivadas de existencia y renovación. Esta lógica reduce el ruido propio de los registros administrativos (rezagos de renovación, cancelaciones no económicas) y entrega series comparables para el análisis de impacto.

En esta subsección se describe exclusivamente la fuente CCB y el procedimiento para construir la serie histórica de matrículas. La base conserva el código CIU reportado para cada registro (a cuatro dígitos), de modo que admite futuras agregaciones o desagregaciones sectoriales; sin embargo, en la versión actual del prototipo estos códigos no se utilizan para construir el indicador de diversidad, que se deriva de los usos reportados en el Censo Inmobiliario.

#### 5.1.2.1 Funciones clave en *utils*: qué hacen y por qué importan.

- **Limpieza y estandarización de datos (*clean\_and\_prepare\_data*):** selecciona columnas obligatorias (identificadores, atributos empresariales, coordenadas) y excluye registros con georreferenciación inválida (lat/long nulas o cero). Estandariza tipos y normaliza categorías (p. ej., tamaños de empresa), evitando que errores de origen contaminen el panel espacial.
- **Asignación temporal de registros (*calculate\_year*):** asigna un año explícito a cada lote, condición necesaria para el apilado anual y para diseños de panel. La presencia de *Year* también facilita trazabilidad durante auditorías y reproducibilidad.
- **Construcción de indicadores derivados (*calculate\_derived\_fields*):** construye indicadores de madurez y formalidad: (i) tiempo de existencia (desde *ANO\_CREACION*) y (ii) tiempo desde última renovación (desde *ANO\_RENOVA*). Prepara, según corresponda,

transformaciones como Existencia (log). Estas métricas permiten diferenciar caídas por rezagos administrativos de cambios reales en el tejido.

- **Aplicación de reglas de validación (*apply\_rules*):** implementa reglas de inclusión/exclusión que aproximan la realidad vigente: ventana de última renovación (típicamente 0–5 años), eliminación de casos con cancelaciones en el año actual/futuro, y verificación de campos críticos (i.e. *TIPO\_PERSONA*). El resultado es un panel depurado con mejor validez interna.

#### **5.1.2.2 Cómo funciona el main.**

El archivo *main* coordina de manera integral el flujo del *pipeline*. El proceso inicia con la lectura de los archivos de matrículas activas y canceladas por año y continúa con la validación de las columnas mínimas requeridas. Luego se ejecuta la limpieza y estandarización de datos, la asignación temporal de los registros y la concatenación de los distintos años para construir el horizonte histórico. A partir de esta base consolidada, se generan los indicadores derivados, se aplican las reglas de validación e inclusión y finalmente se realiza el agregado por tesela-año, con la exportación del archivo *csv* resultante. En el *front-end*, la interfaz de *Shiny* permite cargar archivos, controlar la ejecución, visualizar resultados preliminares y gestionar alertas o mensajes de error, dejando trazabilidad operativa y reduciendo significativamente los errores del usuario.

#### **5.1.2.3 Horizonte temporal y alineación con censo.**

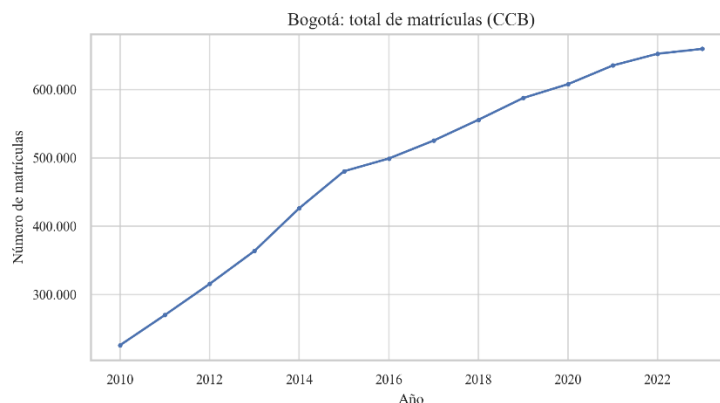
Para asegurar comparabilidad con el censo inmobiliario, restringimos el horizonte a 2010–2023 (intersección de años completos en ambas fuentes). Así evitamos sesgos de cortes incompletos (i.e. 2024 parcialmente observado) y garantizamos que las tendencias reflejen fenómenos económicos y no artefactos de cobertura.

#### **5.1.2.4 Resultados descriptivos.**

La Figura 4 presenta la serie anual del total de matrículas en Bogotá (2010–2023). En el contexto del modelo, el número de matrículas vigentes se define como la variable dependiente principal (outcome), porque permite observar la trayectoria secular del tejido empresarial sin sobrereaccionar a los baches administrativos de renovación. En una lectura preliminar, los

períodos de aceleración/contracción del total suelen alinearse con choques macro o con hitos sectoriales; esta curva es la referencia para las posteriores estimaciones de impacto (DiD) y verificaciones de pretensiones.

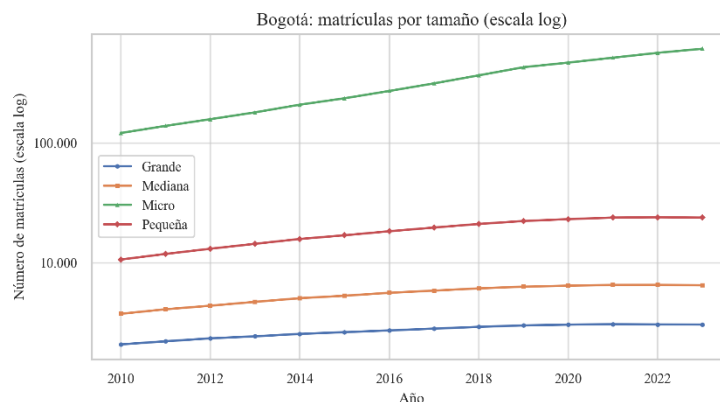
**Figura 4 Total matrículas (2010-2023)**



Fuente: elaboración propia para este proyecto, Secretaría Distrital de Desarrollo Económico.

La Figura 5 desagrega por tamaño de empresa (Micro, Pequeña, Mediana, Grande) en escala logarítmica, resolviendo el problema de sobrerrepresentación de microempresas. En niveles, las micro dominan; en log, se observa con claridad la co-evolución de Pequeñas y Medianas, y la estabilidad relativa de Grandes. Esta desagregación es crucial para discutir heterogeneidad de impactos: por ejemplo, intervenciones de accesibilidad tienden a incidir primero en micro y pequeñas del comercio/servicios, mientras que los cambios sostenidos en Medianas podrían estar indicando reconfiguraciones de cadenas de valor o efectos de derrame más estructurales.

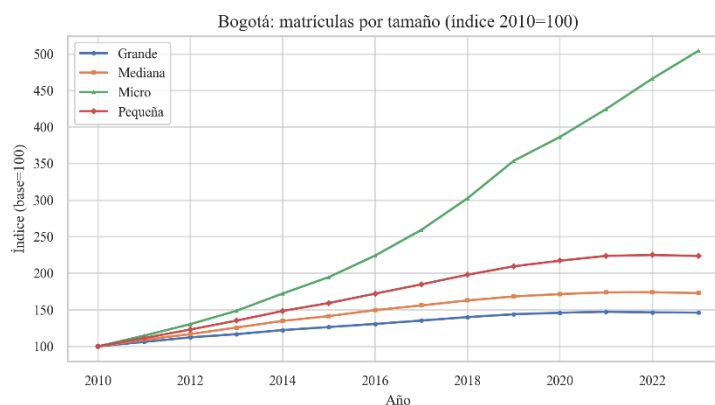
**Figura 5 Matrículas por tamaño (escala log, 2010–2023)**



Fuente: elaboración propia para este proyecto, Secretaría Distrital de Desarrollo Económico.

La Figura 6 muestra la misma desagregación por tamaño, pero indexada (base=100 en el primer año). Este formato facilita comparar trayectorias relativas aun cuando los niveles sean muy distintos. En particular, permite detectar si Pequeñas/Medianas crecen proporcionalmente más que Micro en ciertos tramos, hipótesis que después podemos contrastar con *dummies* de tratamiento y anillos de vecindad. La lectura conjunta de Figuras 4 y 5 aporta dos lentes: magnitudes (*log*) y dinámica porcentual (índice), ambas necesarias para una discusión rigurosa.

**Figura 6 Matrículas por tamaño (escala log, 2010–2023)**



Fuente: elaboración propia para este proyecto, Secretaría Distrital de Desarrollo Económico.

### 5.1.2.5 Implicaciones para el análisis econométrico.

Los resultados confirman tres condiciones necesarias para la estimación: (1) consistencia interna (sin duplicaciones y con coordenadas válidas), (2) comparabilidad temporal (2010–2023, con reglas homogéneas), y (3) compatibilidad espacial con la malla de teselas (condición para modelar *spillovers*).

En este marco, el tejido económico se aproxima a través del *stock* de matrículas vigentes, entendido como la presencia y permanencia de unidades productivas formales en el territorio. Esta medida captura principalmente el margen extensivo del tejido (entrada y salida de firmas, creación y desaparición de establecimientos) y permite seguir la trayectoria del número de unidades económicas sin sobrerreaccionar a los baches administrativos de renovación. Al usar el *stock* de

matrículas como *outcome* evitamos sesgos por rezagos de renovación; al incluir tamaños y naturaleza jurídica podemos explorar heterogeneidades de tratamiento, y con los indicadores derivados (existencia/renovación) mitigamos interpretaciones espurias por formalización administrativa.

Conviene precisar que el “tamaño” empresarial (micro, pequeña, mediana, grande) no es una etiqueta nominal, sino una clasificación definida normativamente a partir de rangos de ingresos por actividad ordinaria. Es decir, la base ya incorpora una dimensión monetaria de escala: cuando desagregamos el stock de matrículas por tamaño estamos distinguiendo, de manera indirecta, entre territorios donde predominan microempresas y territorios donde tienen mayor peso empresas con mayores ingresos. En versiones futuras, una posible extensión sería explotar directamente los valores continuos de ingresos o ventas para construir *outcomes* alternativos o ponderar el stock por tamaño, pero ello exige series monetarias homogéneas y consistentes para todo el período de análisis.

En resumen, el módulo CCB entrega una base robusta y auditable para estimar efectos directos e indirectos de intervenciones urbanas sobre el tejido económico (en términos de presencia y permanencia de unidades formales) y para evaluar su distribución por estratos empresariales.

## **5.2 Módulo geográfico**

El módulo geoespacial asegura que el análisis causal descansa sobre una estructura espacial homogénea y auditable. Sus tres objetivos son: (1) estandarizar geometrías y referencias espaciales (tratamiento y teselas); (2) construir vecindades por contigüidad topológica alrededor del tratamiento (*anillos*  $t, n1, n2, \dots$ ); y (3) cruzar esa estructura con los paneles (Censo Inmobiliario y CCB) para obtener, por tesela-año-vecindad, los indicadores requeridos por diferencias-en-diferencias (*DiD*) con exposición espacial. El algoritmo de vecindades es el núcleo: define “dosis de cercanía” al tratamiento con base en contigüidad real entre teselas y, por tanto, determina la métrica espacial con la que se interpretan los derrames (*spillovers*).

### 5.2.1 Estandarización geométrica: entrada y coherencia de proyección

El flujo inicia con *prepare\_shapefile*, que valida y lee los tres componentes del *shapefile* del polígono de tratamiento (.shp, .shx, .dbf) como objeto *sf*. Luego, *adjust\_crs* fuerza consistencia de sistema de referencia entre tratamiento, malla de teselas y capas panel, utilizando *st\_transform*. Esta coherencia es crítica: operaciones topológicas (i.e. *st\_touches*, *st\_intersects*) y agregaciones espaciales dependen de trabajar en un CRS métrico común (recomendado: MAGNA-SIRGAS u otro proyectado para Bogotá, teniendo en cuenta que es el caso de uso). Antes de cualquier topología, se verifica la validez geométrica (*st\_is\_valid*) y, si es necesario, reparar con *st\_make\_valid* para eliminar *self-intersections*, polígonos tipo “*bow-tie*”, “*slivers*” de área cero o anillos invertidos que rompen la adyacencia.

### 5.2.2 Formulación formal del problema de vecindades

Sea una malla de teselas  $V = \{1, \dots, N\}$ , y una relación de contigüidad topológica  $E \subset V \times V$  inducida por *sf::st\_touches* (contigüidad *queen*: comparte borde o vértice). Esto define un grafo no dirigido  $G = (V, E)$ . Sea  $T \subset V$  el conjunto de teselas tratadas (las que intersectan el polígono de intervención; en la práctica se marca  $t$  por *st\_intersects*). Para  $k \geq 0$ , definimos el anillo  $k$  como el conjunto de vértices a distancia geodésica en el grafo igual a  $k$  desde  $T$ :

- Anillo 0:  $A_0 = T$ .
- Anillo 1:  $A_1 = \{v \in V \setminus A_0 : \exists u \in A_0 \text{ con } (u,v) \in E\}$ .
- Anillo 2:  $A_2 = \{v \in V \setminus (A_0 \cup A_1) : \exists u \in A_1 \text{ con } (u,v) \in E\}$ .
- ...
- Resto sin vecindad (si se corta a  $K$ ):  $S_n = V \setminus (A_0 \cup \dots \cup A_K)$ .

La distancia de grafo captura la mínima cantidad de saltos topológicos para alcanzar el tratamiento y es invariante a escalas métricas locales; esto es crucial para no confundir proximidad euclidiana con conexión espacial efectiva (compartir frontera o vértice).

### 5.2.3 Implementación: crecimiento por capas (BFS) y determinismo

La función *mark\_neighbors* implementa este esquema mediante un BFS (*breadth-first search*) sobre E. El algoritmo:

- Inicialización de etiquetas. Se crea un vector ordenado de etiquetas *labels* =  $c("t", "n1", "n2", "n3", \dots)$ . Las teselas que intersecan el tratamiento (*st\_intersects*) reciben t. Se marca un arreglo *visited* para evitar revisitas.
- Expansión por capas. Para cada nivel *k* desde 0 hasta  $K-1$ :
  - Se toman las teselas con etiqueta  $t\{k\}$ .
  - Se obtiene su frontera por contigüidad usando *st\_touches* (contigüidad *queen*).
  - A toda tesela vecina no visitada se le asigna  $t\{k+1\}$  y se marca como visitada.
- Las teselas no alcanzadas tras *K* expansiones se etiquetan como "sn" (sin vecindad), pudiendo luego excluirse con *remove\_sn* para acotar universo analítico a tratamiento + anillos.

Para la reproducibilidad, el *BFS* fija: (1) un orden canónico de teselas (por ID), (2) el vector *labels* y (3) el uso sistemático de *queen*. Ello asegura que, a igualdad de insumos, las etiquetas son invariantes. En términos de coste computacional, con mallas limpias el *BFS* es  $O(|V|+|E|)$ . Para acelerar, puede precomputarse la lista de adyacencia a partir de *st\_touches* una sola vez por malla (5 ha y 10 ha), almacenándola y reusándola para múltiples tratamientos.

### 5.2.4 Por qué contigüidad topológica (y no buffers de distancia)

Emplear *st\_touches* en lugar de buffers evita falsos positivos de “vecindad” por cercanía euclidiana que no conlleva contacto territorial. En contextos urbanos, corredores viales, ríos o infraestructuras pueden generar proximidades métricas sin interacción espacial efectiva. Los anillos topológicos se interpretan naturalmente como órdenes de vecindad (primer, segundo, tercer anillo), lo que se corresponde con la noción de *lag espacial* discreto en econometría (e.g., variables de exposición  $n1, n2$ ). Adicionalmente, el esquema *queen* (borde o vértice) captura continencias en esquina muy comunes en mallas regulares; si se prefiere una noción más estricta, la variante *rook* (solo borde) puede activarse como parámetro, con implicaciones en la densidad del grafo (menos *edges*) y la extensión de anillos (más “estrechos”).

De tal manera que, en vez de medir con “una regla” qué tan cerca está cada tesela del proyecto, medimos si se tocan de verdad. Primero marcamos las que intersectan el proyecto ( $t$ ); luego las que tocan a  $t$  por borde o vértice ( $n1$ ), después las que tocan a  $n1$  ( $n2$ ), y así sucesivamente: cada “paso” es un anillo, esto evita llamar vecinas a teselas que están cerca pero separadas. Si contamos vértice y borde usamos el esquema *queen* (más conexiones); si solo borde, el *rook* (menos conexiones y anillos más estrechos). En econometría eso se traduce en niveles discretos de exposición ( $n0$ ,  $n1$ ,  $n2...$ ) fáciles de interpretar y de comparar.

En tejidos urbanos con barreras físico-territoriales (ríos, cerros, autopistas sin paso peatonal), cabe restringir E imponiendo máscaras de conectividad (i.e. recortar con una capa de impedancias y borrar *edges* que crucen la barrera). Esto produce anillos geoméricamente plausibles en términos de accesibilidad real. El módulo admite esta extensión: basta con filtrar la matriz de contigüidad antes de ejecutar el *BFS*. La alternativa, más costosa, es constreñir contigüidad por red vial (vecindad de primer orden por grafo de calles), pero ese es otro régimen de vecindades (*network-based*) fuera del alcance del diseño base.

En este trabajo no se aplican máscaras de barreras, Bogotá se particionó en teselas regulares y la vecindad se definió únicamente por contigüidad topológica en la malla (esquema *queen*), sin restricciones por ríos, cerros o infraestructuras. Esto entrega una línea base neutra y reproducible para el análisis. Es importante subrayar que el método es extensible para cuando se requiera mayor realismo de accesibilidad, basta incorporar una máscara de conectividad (eliminar aristas que crucen barreras) o, en una versión más especializada, pasar a un régimen basado en red vial. Es decir, aunque aquí no contemplamos barreras físico-espaciales, la arquitectura permite integrarlas sin reescribir el motor de vecindades.

#### **5.2.4.1 Casuística geométrica y control de calidad**

- **Multipolígonos y huecos.** Si el tratamiento es multipoligonal,  $t$  cubre todas las teselas que interceptan cualquiera de los componentes. Si el tratamiento tiene huecos (agujeros), el

anillo  $0$  no incluye las teselas del hueco (salvo que también intercepten el borde externo); el *BFS* desde  $t$  no “salta” al hueco si no hay contigüidad topológica.

- **Componentes desconectados.** El *BFS* opera naturalmente sobre múltiples componentes del tratamiento: la expansión se hace en paralelo; la primera etiqueta asignada a una tesela prevalece, evitando duplicidades.
- ***Slivers* y *micro-gaps*.** Geometrías con microfragmentos pueden crear “contigüidades espurias”. De allí el uso de *st\_make\_valid* y, si persisten, una disolución previa de bordes (buffer de tolerancia mínima seguido de *st\_buffer(-tol)* para limpiar el borde).
- **CRS y tolerancia.** Trabajar en CRS proyectado métrico evita decisiones “a ojo” sobre tolerancias en grados. Si se necesitan tolerancias, estas deben expresarse en metros y quedar documentadas.

#### 5.2.4.2 Homogeneización espacial: teselas de 5 ha y 10 ha

La malla de teselas regulares (5 ha y 10 ha) no elimina el problema de la unidad espacial modificable (MAUP), pero sí permite controlar algunas de sus manifestaciones prácticas al homogeneizar el área y la forma de las unidades. Esto facilita expresar indicadores en términos de densidades (por ejemplo, m<sup>2</sup> construidos por hectárea, matrículas por hectárea, empleo por hectárea) y hace más comparables los resultados entre unidades y a lo largo del tiempo, siempre dentro del rango de escalas analizado.

- 5 ha maximiza resolución cerca del tratamiento (mejor detección de gradientes y “efectos de borde”), pero puede incrementar la varianza de conteos.
- 10 ha suaviza ruido local y resulta útil para efectos más amplios o evaluaciones de robustez.
- La regla práctica es estimar con 5 ha y replicar con 10 ha, reportando estabilidad de signos y magnitudes.

#### 5.2.5 Cruce espacial con paneles y preparación del panel para econometría

Con la malla etiquetada (*tipo\_vecindad*), *cross\_panels\_with\_teselas* ejecuta el cruce espacial de paneles con teselas:

- Para puntos (i.e. matrículas CCB geocodificadas), *st\_intersects* asigna unívocamente el *ID* de tesela y su anillo.

- Para polígonos (i.e. manzanas del Censo), si existe intersección múltiple con varias teselas, el flujo base suma por tesela. En casos con polígonos grandes, una mejora sería usar intersección área-ponderada para variables de área y centroide-en-polígono para conteos, evitando doble contabilización. Esta extensión es compatible con la arquitectura actual.

Luego, *calculate\_indicators* agrega por tesela-año-vecindad y construye el panel final con *outcomes* y controles:

#### Censo Inmobiliario (estructura física y usos)

- Sumas: *Area\_T\_Cons*, *AreaResi*, *AreaNoResi*, *Predio\_Total*, *PredioResi*, *PredioNoResi*.
- Índices:  $I\_Cons\_Total$  ( $= Area\_T\_Cons/ManArea$ ) e índices por uso; razones de mezcla ( $R\_AreaRes\_NoRes$ ).

#### CCB (tejido económico)

- Vigentes (stock en t con renovación al día, *Establecimientos* (con la regla: si micro reporta 0, se imputa 1), *Empleo*).
- Desagregaciones: tamaño (*Microempresa*, *Pequena*, *Mediana*, *Grande*) y naturaleza jurídica (*P\_Natural*, *P\_Juridica*) para vigentes.
- Madurez/formalidad: *Existencia* (años desde creación) y *Renovación* (años desde última renovación).
- Diversidad sectorial (a partir de *CIJU\_1*).

Antes de exportar, *replace\_inf\_na* sana divisiones por cero y faltantes. El flujo genera al menos *panel\_total\_5ha.csv* y *panel\_total\_10ha.csv*, que soportan los dos niveles de resolución.

#### **5.2.5.1 Visualización de teselas y lectura exploratoria**

Se desarrolló una pestaña de visualización de teselas del módulo econométrico que ofrece un visor cartográfico interactivo que carga la malla regular (5 ha y 10 ha) ya etiquetada por vecindad ( $t$ ,  $n1$ ,



Para evitar endogeneidades, el mapa de anillos se mantiene fijo en el tiempo (definido sobre la geometría base), y las fechas de tratamiento entran como shocks temporales; si se acumulan nuevas intervenciones, puede calcularse un “mínimo anillo” por tesela (distancia al tratamiento más cercano), evitando superposiciones conflictivas de múltiples mapas.

### 5.3 Módulo econométrico

Este módulo toma los paneles espaciales ya preparados (tesela-año, con etiquetas de *vecindad t, n1, n2, ...*) y ejecuta diferencias-en-diferencias (*DiD*) con efectos fijos bidireccionales para estimar el impacto de la intervención sobre *outcomes* del tejido económico (i.e. Vigentes, Establecimientos, Empleo, o los que indiques). El foco está en comparaciones “*tratamiento vs. control*” definidas por anillos de vecindad, con controles de calidad previos (esquema/variables, atípicos, transformaciones) y con pruebas de pretendencias y placebos como chequeos de validez.

#### 5.3.1 Ingesta, validación de esquema y sanitización

- **Lectura y dependencias (*read\_csv\_auto*):** carga el panel (el *main* lo hace vía interfaz *Shiny*). Antes de correr modelos, *ensure\_packages* verifica que estén los paquetes necesarios (i.e. *fixest*, *dplyr*, *ggplot2*, *openxlsx*).
- **Validación de columnas (*validate\_schema*):** comprueba que existan las columnas mínimas: identificador espacial (por defecto *ID\_T5*), año (*Year*), vecindad (*tipo\_vecindad*) y las variables de resultado seleccionadas. Si falta algo, lanza error de forma explícita.
- **Tipos y limpieza (*sanitize\_df*):** normaliza tipos: *ID* como *factor/character*, *Year* como numérico/entero, vecindad como *character*, y corrige detalles frecuentes (espacios, etiquetas). Esta etapa evita errores silenciosos en la construcción de *dummies* y en los efectos fijos.
- **Winsorización opcional:** para atenuar la influencia de *outliers* en *outcomes* continuos, *winsorize\_df* aplica winsorización simétrica al 1%–99% en las columnas que indiques. Esto es opcional y transparente (queda registrado en la exportación).
- **Transformaciones (*transform\_outcomes*):** permite trabajar en logaritmos de  $(1 + y)$  cuando la distribución está fuertemente asimétrica (típico con conteos de microempresas). En el *main*, esto se expone como selector; la transformación aplicada queda documentada.

### 5.3.2 Construcción del diseño *DiD* y estimación

- **Codificación de tratamiento y tiempo (*run\_did\_comparison*):** es el caballo de batalla: a partir del panel, filtra dos grupos definidos por vecindad (i.e. *t vs n1*, *o t vs n2*, etc.) y genera tres variables canónicas:
  - *treated* = 1 si la tesela pertenece al anillo tratado (etiqueta seleccionada), 0 si pertenece al control;
  - *time* = 1 a partir del año de tratamiento que indiques, 0 en los años previos<sup>1</sup>;
  - *did* = *treated* × *time*, la interacción que captura el efecto diferencial tras la intervención.El filtrado a dos grupos (tratamiento y un control a la vez) hace explícita cada comparación y evita solapamientos.
- **Fórmula y efectos fijos (*build\_did\_formula*):** compone la fórmula *outcome* ~ *treated* + *time* + *did* (+ *covariables*) | *ID* + *Year*, en el formato de *fixest::feols*. Así se controlan efectos fijos de tesela y año (*TWFE*).
- **Estimación y clúster (*feols\_did*):** estimula *fixest::feols* con clúster de errores bidireccional por *ID* y *Year* (si *cluster\_tw=TRUE*, que es la opción por defecto en el *main*). La estimación principal de interés es el coeficiente asociado a *DiD*, interpretable—bajo la condición de tendencias paralelas pre-tratamiento—como ATT del anillo tratado frente al anillo de control seleccionado.
- **Multiplicidad de comparaciones (*run\_all\_comparisons*):** itera sobre varios controles (i.e. estimar *t vs n1*, *t vs n2*, *t vs n3*) y varios *outcomes* en un solo pase, y concatena resultados en una tabla *tidy*. Esto permite ver gradientes espaciales del efecto (si el impacto decae de *n1* a *n3*) y robustez por variable.
- **Exportación y traza (*export\_results\_to\_excel*):** guarda todo en un *.xlsx* con una hoja “*results*” (coeficientes, errores estándar, p-values, *fit stats*) y otra “*meta*” (paquetes, versión de R, timestamp).

---

<sup>1</sup> Por ejemplo, si el panel va de 2010 a 2023 y la intervención entra en operación en 2020, se fija  $t^* = 2020$ . En ese caso, para todas las teselas (tratadas y de control)  $time = 0$  entre 2010 y 2019 y  $time = 1$  entre 2020 y 2023;  $treated = 1$  solo para las teselas del anillo 0 (tratamiento) y  $treated = 0$  para las teselas del anillo de control. La variable  $did = treated \times time$  toma valor 1 únicamente para las teselas tratadas en los años posteriores a 2020, que son las que identifican el efecto del tratamiento.

### 5.3.3 Comprobaciones de validez: pretendencias y placebo

- **Pretendencias (*pretrend\_summary*):** construye, antes del año de tratamiento, promedios anuales del *outcome* por grupo (tratado/control). *plot\_pretrends()* grafica estas medias con líneas sólidas y punteadas.  
Lectura: si las trayectorias divergen sistemáticamente antes del tratamiento, el supuesto de tendencias paralelas puede estar comprometido; si se mueven parejo y solo se separan después, hay más confianza en la identificación; aquí es un chequeo descriptivo.
- **Placebo temporal (*run\_placebo*):** replica la *DiD*, pero forzando un año falso de tratamiento anterior al real. Si el algoritmo devuelve un *DiD\_placebo* estadísticamente nulo, es evidencia a favor de que no hay rupturas espurias previas; si sale significativo, sugiere alerta (posible confusión por shocks preexistentes o mala elección del control).

### 5.3.4 Controles, covariables y heterogeneidad

- **Covariables:** admite *covars* adicionales (i.e., indicadores físicos del censo inmobiliario como *I\_Cons\_Total*, razones de mezcla o densidades). El criterio es incluir controles exógenos o preexistentes, sin “sobrecontrolar” mecanismos del propio tratamiento. El *main* permite pasar estas variables de forma selectiva.
- **Elección de controles (anillos):** se compara *t* con el anillo inmediatamente siguiente (*n1*) para un control “cercano pero no tratado”; luego, como robustez, *t vs n2* y *t vs n3*. El módulo facilita ver los tres (o las vecindades que se definan) y cotejar si el efecto se atenúa con la distancia topológica (lo esperable si hay derrames decrecientes).
- **Heterogeneidad por tamaño:** aunque el código no fragmenta automáticamente por categorías empresariales, puedes pasar *outcomes* estratificados (p. ej., *Vigentes\_Micro*, *Vigentes\_Pequena*, etc., si se construyen en el panel) y correr *run\_all\_comparisons* para obtener perfiles de efecto por tamaño. Esto es útil cuando la teoría sugiere impactos distintos para micro vs. medianas.

### 5.3.5 Salidas y visualización

- **Tabla de resultados:** cada fila corresponde a un término del modelo. El interés se centra en  $term == "did"$ , con su estimación, error estándar,  $IC95\%$  y metadatos (comparación, *outcome*, *cluster*,  $R^2$ ,  $AIC/BIC$ ). Esta tabla es la que se exporta a Excel.
- **Gráfico de coeficientes (*generate\_coefficient\_plot*):** gráfica un *forest-plot* de las estimaciones *DiD* con  $IC95\%$ , ordenado por comparación. Es un resumen visual para Resultados y Discusión: si las barras de  $t$  vs  $n1$  y  $t$  vs  $n2$  están por encima de cero (*con IC que no cruza 0*), hay evidencia de efecto positivo decreciente con la vecindad.
- **Pretendencias:** se grafican para cada *outcome* y comparación (líneas tratadas vs control en la ventana *pre-tratamiento*).

### 5.3.6 Qué sí identifica y qué no

El modelo con efectos fijos por tesela y por año permite estimar el cambio diferencial que ocurre en las áreas tratadas después de una intervención, en comparación con las áreas de control. En otras palabras, el estimador mide cuánto cambia el resultado en las teselas intervenidas respecto a lo que habría pasado si no se hubiera hecho la intervención, usando como referencia las teselas que no fueron tratadas. Esta configuración controla dos tipos de factores: los que no cambian a lo largo del tiempo dentro de cada tesela (por ejemplo, su ubicación o características físicas estables) y los shocks comunes que afectan a todas las unidades en un mismo año (como una crisis económica o una política general).

El modelo también utiliza clústeres bidireccionales para estimar los errores estándar, lo que significa que se ajusta a la posibilidad de que los datos estén correlacionados tanto en el tiempo (dentro de una misma tesela) como entre teselas en un mismo año. Este procedimiento hace más estricta la inferencia estadística y reduce el riesgo de interpretar como significativos efectos que en realidad son producto de dependencias ocultas en los datos.

El supuesto más importante del modelo es el de tendencias paralelas, que plantea que antes del tratamiento las trayectorias de las áreas tratadas y de control deberían moverse de manera similar. Si esto se cumple, podemos atribuir los cambios observados después del

tratamiento al efecto de la intervención. Por eso el módulo incorpora pruebas como *pretrend\_summary* y *run\_placebo*, que permiten visualizar las tendencias previas y ensayar tratamientos falsos. Si se observan diferencias notorias antes de la intervención, es necesario proceder con cautela: puede ser conveniente ajustar el grupo de control, modificar la ventana temporal o incluir variables adicionales que capten diferencias estructurales entre las áreas.

En cuanto a los efectos espaciales (*spillovers*), el tratamiento y los controles se definen con base en anillos de vecindad generados por el módulo geoespacial. Esto quiere decir que las teselas se agrupan según su cercanía topológica al área tratada: el anillo  $t$  corresponde a las teselas directamente intervenidas,  $n1$  al primer anillo vecino,  $n2$  al siguiente, y así sucesivamente. El módulo econométrico no recalcula estas vecindades, sino que las utiliza como referencia fija para estimar el efecto en cada nivel de proximidad. Por eso, es recomendable reportar los resultados comparando el anillo tratado ( $t$ ) con sus vecindades ( $n1$ ,  $n2$ ,  $n3$ ), verificando que la magnitud del efecto disminuya conforme aumenta la distancia. Ese patrón decreciente suele ser una señal coherente de difusión espacial del impacto.

### **5.3.7 Flujo operativo en la aplicación**

La aplicación integra todos los componentes del módulo econométrico en una interfaz interactiva que facilita su uso por parte del analista o la institución. Desde esta interfaz, la aplicación identifica automáticamente las columnas clave del panel (el identificador espacial, el año y el tipo de vecindad) y permite configurar los principales parámetros del modelo: seleccionar el año de tratamiento, definir el grupo tratado (por ejemplo,  $t$ ) y los anillos de control (como  $n1$ ,  $n2$ ,  $n3$ ), elegir las variables de resultado, y activar opciones adicionales como la winsorización o la transformación logarítmica de las variables. Una vez definidos estos parámetros, el usuario puede ejecutar tres tipos de análisis: (1) la verificación de pretensiones entre grupos antes de la intervención, (2) la estimación principal de diferencias en diferencias, y (3) la prueba placebo, que permite comprobar la robustez de los resultados. Al finalizar, la aplicación genera un archivo Excel que incluye los resultados y los metadatos técnicos (como las versiones de R, los paquetes utilizados

y la fecha de ejecución), asegurando transparencia, trazabilidad y reproducibilidad en cada corrida.

En conjunto, la aplicación implementa un modelo de diferencias en diferencias con efectos fijos de dos vías (TWFE–DiD) coherente con la estructura espacial definida por los anillos de vecindad. El modelo compara los cambios en las teselas tratadas con los de los anillos de control, controlando las diferencias que permanecen constantes en el tiempo (efectos fijos por tesela) y los factores comunes a todos los años (efectos fijos por periodo). Además, los errores estándar se estiman mediante un clúster bidireccional, lo que fortalece la validez estadística de los resultados. Las pruebas de tendencias y placebos complementan el análisis y permiten verificar que los efectos observados después de la intervención son consistentes y atribuibles al tratamiento. En las salidas de la aplicación —tablas, gráficos y reportes— se muestra, por diseño, el cambio diferencial entre el anillo tratado y los anillos de control dentro de la ventana temporal posterior a la intervención.

#### **5.4 Caso de uso**

Como caso de uso se analiza el intercambiador deprimido de la Calle 94 con la NQS en Bogotá, una obra vial estratégica que reorganizó flujos y giros en un nodo de alta demanda del norte de la ciudad y cuya entrada en operación fue en 2017. Este hito ofrece una cronología clara para la evaluación de impactos: permite fijar 2017 como año de tratamiento, construir una ventana pre (2010–2016) y post (2017–2023), y contrastar con un placebo cercano (2015). Metodológicamente, el caso es idóneo para nuestra estrategia de diferencias-en-diferencias con efectos fijos y exposición espacial por anillos de vecindad (teselas de 10 y 5 ha): concentra efectos potencialmente locales sobre el tejido económico inmediato (actividad formal y su composición por tamaño empresarial) y, a la vez, es lo bastante conectado a corredores comerciales y residenciales como para revelar derrames espaciales graduales hacia  $n1-n3$ . En consecuencia, el estudio del 94–NQS opera como banco de pruebas del pipeline (geoespacial→ econométrico), combinando lectura cartográfica previa (mapa de teselas y etiquetas  $t/n1\dots$ ) con verificación de *pre-tendencias* y estimaciones *DiD* que priorizan Vigentes y su desagregación por tamaño, con  $I\_Cons\_Total$  como contraste físico de baja respuesta inmediata.

### 5.4.1 Configuración de los módulos

- Módulo geoespacial

El módulo geoespacial se parametriza de forma mínima y reproducible con dos entradas: (i) la geometría de tratamiento cargada como shapefile (los cuatro componentes .shp/.shx/.dbf/.prj), y (ii) el número de niveles de vecindad a generar (p. ej., 5 para obtener  $t, n1, n2, n3, n4, n5$ ).

Figura 8 Interfaz de procesamiento de vecindades (carga de shapefile de tratamiento y selección de niveles)



Fuente: elaboración propia para este proyecto.

Con esos parámetros, al ejecutar Procesar datos el motor valida el CRS del tratamiento y lo proyecta al de trabajo, *intersecta* el polígono con la malla de teselas preconfigurada (5 ha o 10 ha, según el panel del proyecto), construye el grafo de contigüidad topológica (*queen* por defecto) mediante *st\_touches()* y aplica un BFS para etiquetar anillos:  $t$  (teselas que intersectan el tratamiento),  $n1$  (teselas que tocan a  $t$ ),  $n2$  (que tocan a  $n1$ ), y así sucesivamente hasta el nivel solicitado. El resultado es un conjunto de archivos en storage/—shapes y/o CSV—con campos estandarizados ( $ID\_T5/ID\_T10, tipo\_vecindad, Year$ ), listos para el cruce con paneles y su consumo directo por el módulo econométrico.

- Modulo econométrico

Para el caso del deprimido de la Calle 94–NQS se ejecutó el módulo econométrico con una configuración replicable. La raíz del proyecto se fijó en la carpeta que

contiene *Storage/*, desde donde la aplicación detecta los paneles en *Storage/tasty\_data* y las teselas en *Storage/cook\_geo*. Se trabajó con el panel de 10 hectáreas (*panel\_total\_10ha.csv*) como especificación principal, mapeando las columnas *ID* de unidad a *ID\_T10*, año a *Year* y vecindad a *tipo\_vecindad*. La etiqueta de tratamiento en la variable de vecindad se definió como *t* y el año de tratamiento se estableció en 2017, coherente con la entrada en operación de la obra. Con esta base, el estimador TWFE compara el anillo tratado con los anillos de control definidos por contigüidad topológica.

**Figura 9** Interfaz de procesamiento del módulo econométrico.

The figure displays two screenshots of a web application interface for spatial econometrics. The left screenshot shows the main configuration page, titled "Econometría espacial: DiD". It includes a "Raíz del proyecto (contiene /Storage)" field with the path "C:/Users/wraed/OneDrive/Portátil/Proyecto/Código". Below this is a search bar for panels and tiles. The "Archivo del panel" section has a dropdown menu set to "panel\_total\_10ha.csv" and a "Buscar" button. There is also a "Cargar CSV (opcional)" section with a "Browse..." button. The "Columna ID unidad" is set to "ID\_T10", "Columna año" to "Year", "Columna vecindad" to "tipo\_vecindad", and "Etiqueta tratamiento (en vecindad)" to "t". The right screenshot shows the "Vecindades de control" section with a list of neighborhoods: "n1 n2 n3 n4 n5". The "Variables de resultado (outcomes)" section has a dropdown menu with options: "Vigentes", "Microempresa", "Pequeña", "Mediana", "Grande", "Empleo", and "I\_Cons\_Total". The "Covariables (separadas por coma)" field contains "R\_AreaRes\_NoRes,I\_Cons\_NoResi". There are three checked checkboxes: "Transformar outcomes con log1p()", "Winsorizar outcomes (1%-99%)", and "Cluster 2-vias (ID + Año)". The "Ejecutar placebo" checkbox is also checked. The "Año placebo" field is set to "2015". At the bottom, there is a blue "Ejecutar análisis" button and a "Descargar resultados (Excel)" button.

Fuente: elaboración propia para este proyecto.

El conjunto de controles espaciales incluyó los anillos *n1*, *n2*, *n3*, *n4* y *n5*. En la lectura de resultados, el contraste central es *t vs n1* (control más cercano no tratado), mientras que *t vs n2–n5* funciona como robustez y permite inspeccionar si la magnitud del efecto se atenúa con la distancia topológica, patrón esperado bajo derrames

decrecientes. Esta estructura de comparaciones se alinea con la definición de exposición por anillos de vecindad que proviene del módulo geoespacial.

En variables de resultado (*outcomes*) se seleccionaron siete indicadores: *Vigentes* (stock de matrículas activas), las desagregaciones por tamaño —*Microempresa, Pequeña, Mediana, Grande*—, *Empleo* y el indicador físico *I\_Cons\_Total* del Censo Inmobiliario. Dado que los conteos empresariales suelen mostrar asimetría y una representación dominante de microempresas, se activó la transformación  $\log Ip()$  para todos los *outcomes* y la *winsorización* al 1 % y 99 %, reduciendo la influencia de extremos. La estimación utilizó efectos fijos por tesela y por año y errores estándar con clúster bidireccional (ID y año), lo que resguarda la inferencia ante correlación serial por unidad y shocks comunes por periodo. En la discusión, *I\_Cons\_Total* sirve como placebo negativo físico: por su inercia, es menos probable que responda bruscamente al hito de 2017; si lo hiciera, sugeriría confusión o cambios no atribuibles a la obra.

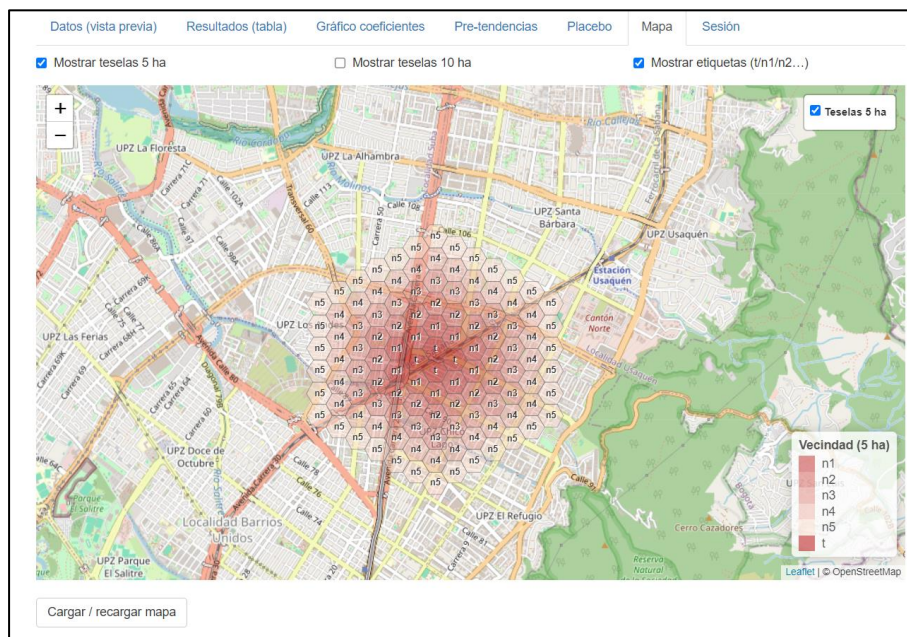
Respecto de covariables, la corrida mostró *R\_AreaRes\_NoRes* e *I\_Cons\_NoResi*. Su papel es estrictamente ajustar por diferencias temporales no explicadas entre teselas, siempre que se traten de predeterminadas (línea base o rezagos) y no de mediadores del tratamiento.

Finalmente, se activó el placebo temporal en 2015. Esta fecha es cercana al año tratado sin solaparse con la obra, aporta una ventana pre suficiente (2010–2014) para verificar paralelismo y funciona como prueba de falsación: un *did\_placebo* no significativo respalda que la ruptura observada en 2017 no replica un quiebre previo. Tras ejecutar el análisis, los resultados se descargan en Excel con traza de metadatos (fórmulas, paquetes, fecha), y se recomienda replicar la corrida con el panel de 5 hectáreas (*panel\_total\_5ha.csv, ID\_T5*) para confirmar que los signos y órdenes de magnitud no dependen de la resolución espacial.

## 5.4.2 Visualización de teselas y lectura exploratoria

En la pestaña Mapa del módulo econométrico se ofrece un visor interactivo que permite superponer la malla de teselas con la etiqueta de vecindad generada en el módulo geoespacial—tratadas ( $t$ ) y anillos  $n1$ – $n5$  por contigüidad *queen*—y alternar entre resoluciones de 5 ha y 10 ha, además de activar rótulos y la leyenda por anillo. Las Figuras muestran el caso del intercambiador 94–NQS: el polígono tratado se materializa como un núcleo rojo ( $t$ ) alrededor del cual se ordenan anillos concéntricos. Esta visualización cumple un rol metodológico doble: (i) es un control de calidad previo a la econometría (permite verificar que la exposición esté bien definida, sin “agujeros” ni teselas mal etiquetadas), y (ii) es un instrumento de lectura exploratoria para anticipar patrones plausibles (gradientes de intensidad que decaen con el orden del anillo, posibles asimetrías hacia corredores viales o zonas de actividad). La combinación de color categórico por vecindad y etiquetas  $t/n1 \dots$  facilita comunicar, de un vistazo, la lógica de identificación espacial que luego usa el estimador *DiD*.

Figura 10 Mapa de vecindades (teselas 5 ha) con etiquetas  $t, n1$ – $n5$  — caso Calle 94–NQS.

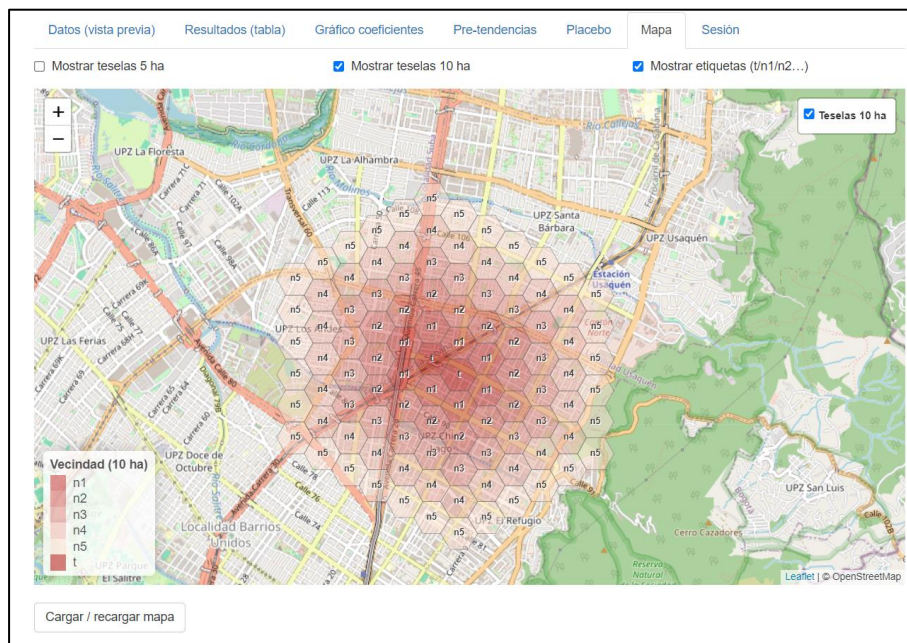


Fuente: elaboración propia para este proyecto.

Comparar 5 ha (Figura 9) con 10 ha (Figura 10) aporta evidencia visual sobre el efecto de la resolución espacial. La malla de 5 ha capta mejor micro heterogeneidad y

contornos irregulares del tratamiento, útil para detectar bordes finos o transiciones bruscas en barrios fragmentados; la malla de 10 ha ofrece una lectura más suavizada y estable, conveniente para atenuar ruido y evitar celdas con conteos muy bajos. Esta contrastación ayuda a mitigar riesgos de MAUP y a justificar la robustez de los resultados: si los patrones (posición del núcleo tratado, continuidad de anillos, orientación del gradiente) se repiten en ambas resoluciones, aumenta la confianza en que la señal no es un artefacto de discretización. Nótese que en este ejercicio las vecindades se definen por contacto topológico (no por red vial ni por barreras físico-territoriales); sin embargo, el flujo admite, de ser necesario, máscaras de conectividad en geoespacial para excluir cruces por ríos, cerros o autopistas y la pestaña del módulo econométrico las visualizará de igual forma.

**Figura 11 Mapa de vecindades (teselas 10 ha) con etiquetas t, n1–n5 — caso Calle 94–NQS.**



Fuente: elaboración propia para este proyecto.

### 5.4.3 Resultados econométricos

La vista Resultados (tabla) lista, para cada comparación espacial ( $t$  vs  $n1 \dots n5$ ) y para cada *outcome*, los términos estimados del modelo *TWFE*: el efecto de interés (*did*), las covariables (cuando se activan), su error estándar, el estadístico de prueba y el p-valor, junto con la marca del grupo tratado. En la corrida mostrada, los coeficientes *did* aparecen cercanos a cero y con intervalos amplios, lo que sugiere no significación estadística al 95 % en varias comparaciones; en contraste, algunas covariables (p. ej., la razón residencial/no residencial o la intensidad no residencial) sí exhiben señal estadística, coherente con su papel de ajuste. Esta lectura implica que, bajo la especificación y el periodo seleccionados, no se detecta un quiebre diferencial robusto entre  $t$  y los anillos de control, aunque el balance y la potencia deben contrastarse con las otras vistas y con la exportación completa.

Figura 12 Vista de la tabla de resultados de DiD para 10Ha

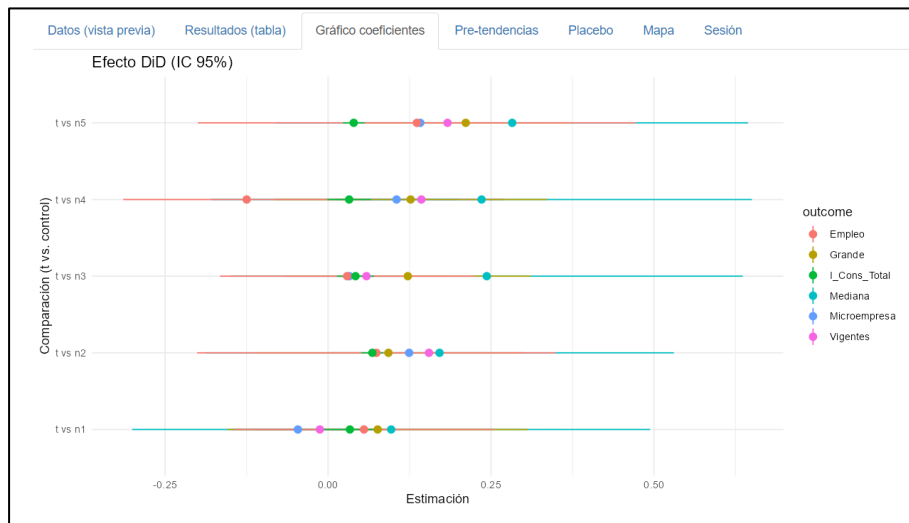
Datos (vista previa)		Resultados (tabla)		Gráfico coeficientes		Pre-tendencias		Placebo		Mapa		Sesión	
Show		25		entries		Search:							
term	estimate	std.error	statistic	p.value	treatment								
1 did	-0.01251539761303999	0.05168701635572157	-0.2421381324645678	0.8141005478459552	t								
2 R_AreaRes_NoRes	0.02783481601960086	0.01121485534843879	2.481959432804965	0.03487912843201397	t								
3 I_Cons_NoResi	-0.2028870903912387	0.04605892834921681	-4.404945960812581	0.001708071011875349	t								
4 did	-0.04625387460398078	0.05167487208847244	-0.8950941286277326	0.3940453683136487	t								
5 R_AreaRes_NoRes	0.003629393705931655	0.02800039470024822	0.1296193766118404	0.8997182457598997	t								
6 I_Cons_NoResi	-0.2149353539690616	0.08225569073866285	-2.613015002839615	0.02813132118671334	t								
7 did	0.09685335366836068	0.2027009566079094	0.4778139940193132	0.6441739763377149	t								
8 R_AreaRes_NoRes	-0.1192947651865557	0.07395947022102063	-1.612974847305625	0.1412083198718534	t								
9 I_Cons_NoResi	-0.004710202644724104	0.2793094945689756	-0.01686373981662452	0.9869132497120467	t								
10 did	0.07614014661577717	0.1178067074107033	0.647413300585675	0.5335282629522011	t								
11 R_AreaRes_NoRes	0.02072188147224306	0.03119358574527033	0.6642994377581288	0.5231518145555605	t								

Fuente: elaboración propia para este proyecto.

El *forest plot* resume el *did* por comparación ( $t$  vs  $n1 \dots n5$ ) y *outcome*, con IC95 %. Visualmente, la mayoría de los puntos se ubican próximos al cero y con bandas que lo cruzan, reforzando la idea de ausencia de efectos claramente distintos de cero en la especificación de referencia. La organización por anillos permite verificar si existe gradiente espacial (magnitud decreciente al pasar de  $n1$  a  $n3/n5$ ); en la evidencia

mostrada, no se observa un patrón monótono fuerte, lo que sugiere que, de existir, los efectos serían pequeños o difusos en el horizonte temporal post-tratamiento considerado. Esta vista es útil para comunicar de un vistazo qué *outcomes* merecen exploraciones adicionales (p. ej., cambiar resolución 10→5 ha, acotar ventana, o refinar controles).

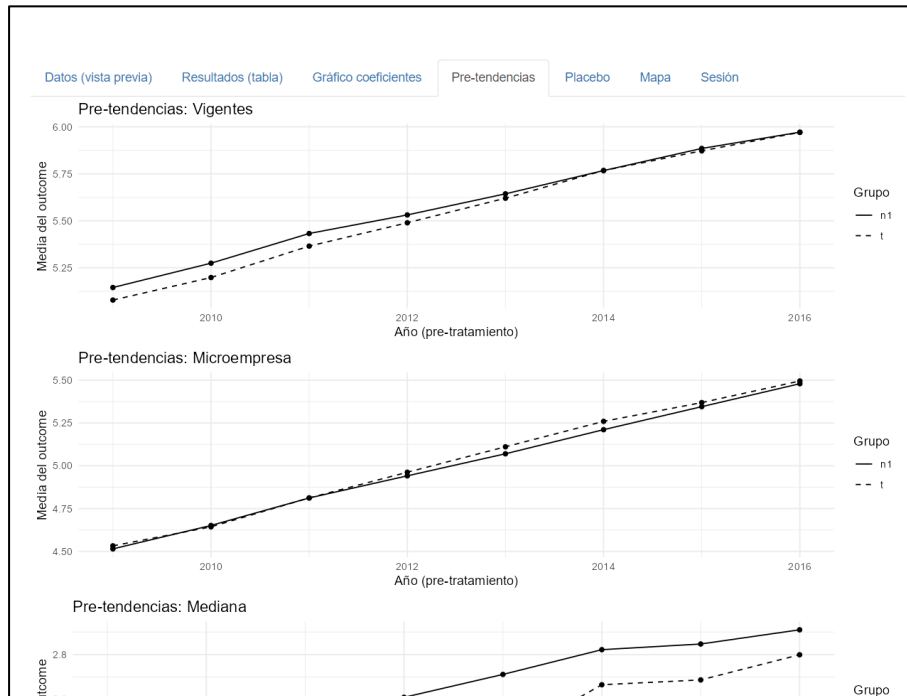
Figura 13 Vista del gráfico de coeficientes para 10Ha.



Fuente: Elaboración propia para este proyecto.

La pestaña *Pre-tendencias* compara la trayectoria media pre-tratamiento de  $t$  y del control principal ( $n1$ ). Para Vigentes y Microempresa se aprecia una evolución paralela y ascendente hasta el año inmediatamente anterior al hito, sin divergencias sistemáticas, lo que apoya el supuesto de paralelismo. En Mediana la separación entre curvas es ligeramente mayor en algunos años pre, aunque mantiene pendiente semejante; ello aconseja reportar la ventana exacta y, si se desea máxima cautela, probar un control alternativo ( $n2$ ) o una ventana pre más corta en una corrida de robustez. En conjunto, esta vista sugiere que el contraste  $t-n1$  es razonable para la identificación, si bien la potencia puede ser limitada para tamaños empresariales menos frecuentes.

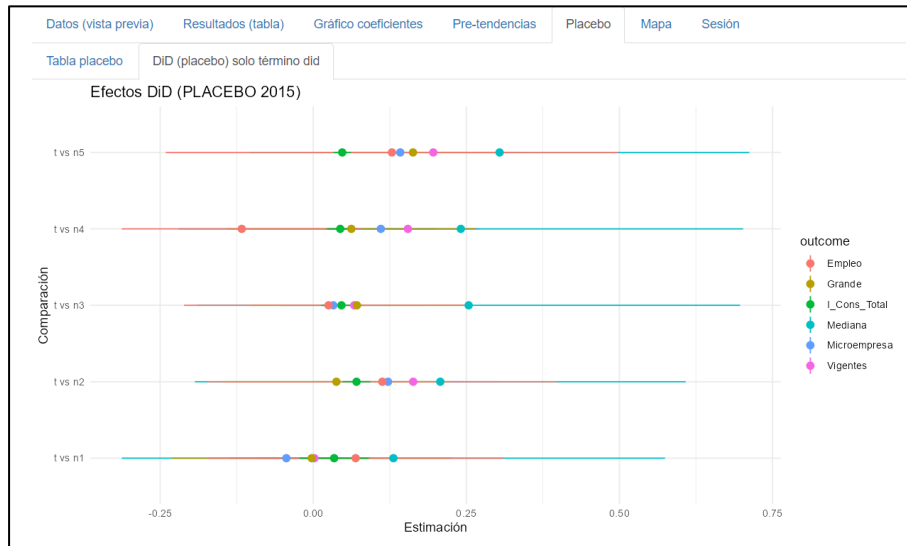
**Figura 14 Vista de las gráficas de pretendencia para 10Ha.**



Fuente: elaboración propia para este proyecto.

El panel Placebo muestra el *did* con año falso (2015). La mayor parte de puntos se concentra alrededor de cero y los IC95 % lo cruzan, lo que respalda la ausencia de quiebres espurios antes del hito (2017) y, por tanto, refuerza la validez del diseño. Alguna señal aislada con banda amplia no cambia la conclusión general, pero justifica informar explícitamente la especificación (*outcomes*, anillos, opciones activadas). Como se indicó, el uso de un indicador físico inercial (p. ej., *I\_Cons\_Total*) funciona además como placebo negativo: si apareciera significativo, sugeriría revisar emparejamientos o *shocks* concurrentes.

**Figura 15 Vista del gráfico de coeficientes para el placebo para 10Ha.**



Fuente: elaboración propia para este proyecto.

Para garantizar trazabilidad y reproducibilidad, el análisis se acompaña del libro **DiD\_Results\_YYYY-MM-DD.xlsx**, que consolida, para cada combinación *outcome* × comparación espacial (p. ej., *t vs n1*, *t vs n2*), el coeficiente *DiD* (*did*), su error estándar, *valor-p*, tamaño muestral (N) y la especificación completa utilizada (fórmula, efectos fijos por tesela y año, esquema de clúster bidireccional, y transformaciones/limpieza activadas: *logIp()* y winsorización 1–99%; además, lista los anillos incluidos). Este libro constituye el insumo de referencia del paper: de él se derivan la tabla principal, que reporta *t vs n1* para los *outcomes* priorizados del tejido económico (Vigentes y Microempresa) junto con los IC95%, y la tabla de robustez, que contrapone resultados entre 10 ha y 5 ha e incluye el placebo temporal. El resto de las combinaciones (p. ej., *t vs n3–n5*, desagregaciones por tamaño y contrastes físicos) se consignan en el apéndice. El archivo conserva metadatos de corrida (fecha/versión), lo que permite auditar el flujo y replicar exactamente la estimación presentada en el manuscrito.

#### 5.4.4 Discusión de resultados.

Por otro lado, con base en los resultados consolidados, la lectura global es consistente entre hojas y vistas: en la comparación principal ( $t$  vs  $n1$ ) los coeficientes  $DiD$  para los *outcomes* del tejido económico (*Vigentes*, *Microempresa*, *Mediana*, *Grande*, *Empleo*) aparecen cercanos a cero y estadísticamente no significativos bajo la especificación de referencia (TWFE con  $\log Ip$ , *winsorización* 1–99 % y clúster bidireccional por tesela y año). Esto sugiere que, al menos con la resolución de 10 ha y la ventana temporal utilizada, no se identifica un quiebre diferencial robusto entre el área tratada y su primer anillo de control inmediatamente posterior al hito de 2017.

Cuando la comparación se desplaza hacia controles más lejanos ( $n2$ – $n5$ ) emerge un patrón más informativo. En primer lugar, el indicador físico  $I\_Cons\_Total$  del Censo Inmobiliario muestra diferencias sistemáticas a favor del tratado en anillos intermedios ( $n2$ – $n3$ ), coherentes con la mayor inercia y la lógica de inversión física alrededor de la obra: la señal aparece antes en variables de intensidad constructiva que en variables estrictamente empresariales. En segundo lugar, para el tejido económico la señal tiende a fortalecerse en anillos más lejanos ( $n4$ – $n5$ ), con evidencia de efecto en *Vigentes* y *Microempresa* y, de manera más tenue, en *Grande*. Esta configuración es interpretable: el primer anillo ( $n1$ ) comparte exposición a shocks locales con el tratado (obras, cambios de accesibilidad y circulación, externalidades de corto plazo), de modo que la diferencia estimada se diluye; el contraste con anillos más distantes, menos expuestos, revela mejor la brecha atribuible a la intervención.

Dos resultados adicionales merecen mención. Por un lado, *Empleo* y el tamaño *Mediana* no exhiben evidencia robusta en ninguna comparación, lo que sugiere menor potencia estadística (por varianza alta y/o tamaños muestrales relativamente más bajos) o rezagos en la traslación de actividad a empleo y a segmentos empresariales de tamaño medio.

En términos de gradiente espacial, el patrón de promedios por anillo sugiere señales crecientes al pasar de  $n1$  hacia  $n4$ – $n5$ , aunque no estrictamente monótonas para todos

los *outcomes*. Este comportamiento es compatible con derrames decrecientes y, a la vez, con el hecho de que  $n1$  constituye un control “demasiado parecido” al tratado (comparten condiciones y *shocks*), lo que reduce la diferencia estimada. La consistencia interna del archivo —coeficientes *did* próximos a cero en  $t$  vs  $n1$ , presencia de señal en  $I\_Cons\_Total$  en  $n2-n3$  y en *outcomes* económicos en  $n4-n5$ — refuerza esta lectura.

En resumen, los resultados permiten construir una narrativa clara: (i) tabla principal con  $t$  vs  $n1$  para Vigentes y Microempresa (coeficiente *DiD*, EE e IC95 %), explicitando que no se detectan efectos robustos en la comparación más estricta; (ii) tabla de robustez con los contrastes  $t$  vs  $n2-n5$ , donde se evidencia la aparición del efecto (físico primero, económico después) y se documenta el gradiente espacial.

En seguida, se corrió el panel de 5 hectáreas para contrastar con los resultados anteriores (10 ha) y así evaluar la sensibilidad de los resultados al MAUP y a la micro heterogeneidad espacial. Con celdas más pequeñas se atenúa la mezcla interna de cada unidad (menos “promedio dentro de tesela”), lo que puede revelar variaciones locales que la malla de 10 ha suaviza; a la vez, se asume el costo estadístico de mayor varianza en *outcomes* poco frecuentes. En términos de identificación, el diseño y los anillos de vecindad son los mismos; únicamente cambia el nivel de agregación espacial sobre el que se estiman las mismas especificaciones *DiD TWFE* (*log1p*, *winsor 1-99 %*, *clúster bidireccional*).

Los resultados de 5 ha son coherentes con la lectura de 10 ha: en  $t$  vs  $n1$  los coeficientes *DiD* de los *outcomes* económicos siguen pequeños y no significativos. La diferencia aparece al mover el control a anillos más lejanos: surge una señal positiva y estadísticamente significativa para el tamaño “Mediana” en  $t$  vs  $n2$  y se mantiene en  $n3-n5$ ; en cambio, *Vigentes*, *Microempresa*, *Grande* y *Empleo* permanecen sin evidencia robusta a este nivel de granularidad. La interpretación es consistente: el primer anillo comparte buena parte de la exposición del tratado (la diferencia se diluye), mientras que controles más distantes permiten discernir mejor el contraste; además, las

empresas medianas parecen reaccionar primero en la escala fina, mientras que la agregación de 10 favorece la detección de cambios físicos en anillos intermedios. En conjunto, 5 funciona como robustez informativa: no cambia la conclusión principal para  $nl$ , pero aporta evidencia puntual sobre reconfiguración en empresas medianas al ampliar la distancia del control.

## CONCLUSIONES

Este trabajo entrega un prototipo operativo, reproducible y auditable para evaluar, de manera exploratoria, impactos de intervenciones urbanas sobre el tejido económico de Bogotá. La contribución es doble. En lo técnico, consolida un flujo que integra preparación de datos administrativos y físicos, homogenización espacial con teselas regulares, construcción de vecindades topológicas por anillos y estimaciones *DiD* con diagnósticos de validez. En lo sustantivo, ofrece una lectura empírica del caso 94–NQS (2017) que, aun cuando no identifica efectos promedio robustos en la especificación de referencia, evidencia la relevancia de las decisiones espaciales (resolución, definición de controles) para la inferencia y para la política pública.

Primero, sobre los hallazgos empíricos. En la resolución de 10 hectáreas —especificación principal por estabilidad—, los coeficientes *DiD* para los *outcomes* económicos nucleares se ubican próximos a cero y con intervalos de confianza amplios; los “*forest plots*” refuerzan la ausencia de señales claramente distintas de cero y no exhiben un gradiente espacial monótono por anillo. Esta lectura sugiere que, con la ventana 2010–2016/2017–2023 y la comparación *t vs. n1*, no se detecta un quiebre diferencial estadísticamente significativo atribuible a la puesta en operación del intercambiador. Al mismo tiempo, las covariables físicas (p. ej., razón residencial/no residencial, intensidad no residencial) sí muestran señal, coherente con su papel de ajuste.

Segundo, sobre los diagnósticos de validez. Los gráficos de *pre-tendencias* apoyan el supuesto de paralelismo para “*Vigentes*” y “*Microempresa*” en el contraste *t–n1*, y recomiendan cautela —sin comprometer la identificación— en “*Mediana*”, donde se observa separación algo mayor en algunos años pre. La inclusión del placebo temporal (2015) opera como prueba de falsación adicional. En conjunto, estos chequeos fortalecen la interpretación de los nullos como resultado sustantivo (y no de ruptura previa mal atribuida), aunque invitan a atender potencia y selección de controles.

Tercero, sobre sensibilidad espacial (5 ha). Al aumentar la resolución, se preserva el diseño econométrico y la lógica de identificación, pero se reduce la mezcla intra-tesela; el costo es mayor varianza en *outcomes* infrecuentes. En esta corrida, los nullos persisten en *t vs. n1*, mientras aparece

una señal positiva y significativa para “*Mediana*” cuando el control se desplaza a anillos más lejanos ( $n2\dots n5$ ). Dos lecturas son plausibles y complementarias: (i) contaminación del primer anillo, que comparte parte de la exposición del tratado y diluye el contraste; y (ii) difusión espacial fina, consistente con reconfiguraciones que emergen ligeramente fuera del núcleo tratado y que se detectan mejor cuando el control se aleja un paso. Esta sensibilidad confirma que la resolución espacial y la elección del control no son detalles cosméticos: forman parte de la identificación.

Cuarto, implicaciones metodológicas. El ejercicio valida una arquitectura que: (a) define exposición espacial por contigüidad topológica y anillos (BFS), (b) utiliza unidades homogéneas (5 y 10 ha) para mitigar MAUP y (c) ancla la inferencia en TWFE con diagnósticos básicos (pretendencias, placebo) y exportaciones trazables. Ello estandariza el “cómo” y reduce discrecionalidad en el “cuándo/dónde” de los impactos. No obstante, los resultados sugieren evolucionar hacia: (1) modelos espaciales explícitos (SAR/SEM/SDM) para distinguir autocorrelación en resultados, covariables o errores; (2) contigüidad con barreras (ríos, cerros, autopistas sin paso) o, en su caso, contigüidad por red vial, a fin de alinear vecindades con accesibilidad efectiva; (3) adopción escalonada/event-study con cohortes y ventanas móviles, para capturar dinámicas lentas; y (4) errores robustos de largo alcance (Conley/Driscoll-Kraay) paramétricamente documentados en el reporte. La propia herramienta contempla varios de estos desarrollos como línea de extensión.

Quinto, limitaciones. (i) Potencia y granularidad: el uso del stock de “*Vigentes*” es deliberado (evita confundir rezagos de renovación con “*muertes*”), pero su inercia puede amortiguar respuestas inmediatas; resultados nulos a corto plazo no descartan ajustes a mediano plazo en composición o diversidad. (ii) Definición de control:  $n1$  es intuitivo pero expuesto a derrames; la evidencia a 5 ha sugiere complementar con  $n2\dots n3$  como especificaciones de robustez. (iii) Simultaneidad de shocks: aun con efectos fijos, otras obras o tendencias sectoriales pueden interferir; por ello, el MVP incorpora *pre-trends* y placebo y permite documentar parámetros y versiones de datos para auditoría. (iv) MAUP residual: el uso de teselas regulares reduce, pero no elimina, sesgos de agregación; la confrontación 5/10 ha es necesaria. (v) Medición física: el uso de *I\_Cons\_Total* como “placebo negativo” físico es pertinente por su inercia; si esta respondiera bruscamente, sugeriría confusiones de identificación.

Sexto, implicaciones para política pública. La principal utilidad del prototipo no radica en “probar” efectos grandes en un caso, sino en institucionalizar un método replicable para monitoreo sistemático: (a) comparar tratamiento vs. anillos, (b) verificar pretendencias y placebo, (c) documentar metadatos, (d) exportar resultados estandarizados y (e) repetir ejercicios en múltiples proyectos y corredores. Con ello, la administración puede pasar de afirmaciones “*impresionistas*” a evidencia verificable sobre magnitud, dirección y alcance espacial de impactos, incorporando de manera progresiva elementos más sofisticados (modelos espaciales, máscaras de conectividad, cohortes). En términos estratégicos, el patrón observado —nulos locales robustos con señales puntuales a distancia para empresas medianas en alta resolución— es coherente con intervenciones cuya asimetría espacial se manifiesta fuera del núcleo (efectos de borde, reasignaciones), lo cual respalda decisiones de política que miren más allá de la línea de obra al diseñar mitigaciones y aprovechamientos.

Finalmente, agenda de trabajo. A corto plazo: (1) consolidar un banco de casos con parametrización homogénea; (2) sistematizar reportes con tablas principales (*t vs. n1*) y apéndices (*t vs. n2–n5, 5 vs. 10 ha, placebo*); (3) incorporar el índice de diversidad y cortes por tamaño/naturaleza jurídica como *outcomes* estándar; (4) explorar ventanas/event-study con rezagos conducidos por teoría (p. ej., tres años post-tratamiento para respuestas lentas). A mediano plazo: (5) integrar matrices de pesos con barreras/Red vial; (6) evaluar SAR/SEM/SDM; (7) habilitar adopción escalonada y comparación entre obras; y (8) conectar el módulo con tableros institucionales para seguimiento continuo. Con estos pasos, la ciudad contará con un sistema de evaluación espacialmente consciente que reduzca discrecionalidad, eleve calidad de la evidencia y permita decisiones más transparentes, inclusivas y eficaces.

## REFERENCIAS

- [1] P. Cheshire y E. S. Mills, "Introduction: Applied Urban Economics," in *Handbook of Regional and Urban Economics*, vol. 3, 1999.
- [2] R. W. Eberts y D. P. McMillen, "Agglomeration Economies and Urban Public Infrastructure," in *Handbook of Regional and Urban Economics*, vol. 4, 1999.
- [3] P. C. Melo, D. J. Graham y R. B. Noland, "The Impact of Transport Infrastructure on Firm Formation: Evidence from Portuguese Municipalities," *Transportation Research Record*, vol. 2163, pp. 133–143, 2010.
- [4] R. Bernal y X. Peña, *Guía práctica para la evaluación de impacto*. Bogotá: Universidad de los Andes, 2011.
- [5] L. Anselin, *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer, 1988. doi: 10.1007/978-94-015-7799-1.
- [6] G. Camargo, *Slum management effects on informal and formal urban expansion: A comparative impact evaluation between neighborhood legalization and comprehensive neighborhood improvement–MIB programs in Bogotá, Colombia*, Master's thesis, Universidad de los Andes, Bogotá, 2022.
- [7] K. Butts, "Difference-in-Differences with Spatial Spillovers," Working Paper, University of Colorado Boulder, Department of Economics, Jun. 2023.
- [8] S. Gibbons, H. G. Overman y E. Resende, "Spatial Methods in Econometrics: Theory and Applications," *Journal of Economic Geography*, 2015.
- [9] J. P. Elhorst, *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Berlin: Springer, 2014. doi: 10.1007/978-3-642-40340-8.
- [10] S. J. Redding y E. Rossi-Hansberg, "Quantitative Spatial Economics," *Annual Review of Economics*, vol. 9, 2017. doi: 10.1146/annurev-economics-063016-103713.
- [11] G. Duranton y M. A. Turner, "The Fundamental Law of Road Congestion," *American Economic Review*, vol. 101, no. 6, 2011. doi: 10.1257/aer.101.6.2616.
- [12] D. Donaldson, "Railroads of the Raj," *Quarterly Journal of Economics*, 2016. doi: 10.1093/qje/qjw002.
- [13] "The Economic Impact of Transport Infrastructure: A Review of Project-Level vs Aggregate-Level Evidence," *Transport Reviews*, 2025. doi: 10.1080/01441647.2025.2476012.

- [14] “Luas Finglas – Delivering a Sustainable Infrastructure: Sustainability, Circular Economy, Active Mobility,” in Springer (cap./vol.), 2025. doi: 10.1007/978-3-031-95284-5.
- [15] “One belt, one road, one way? Where European exporters benefit from the New Silk Road,” *Review of World Economics*, 2023. doi: 10.1007/s10290-022-00467-0.
- [16] S. Gutman y M. Malashenko, “The Impact of Transport Infrastructure on Sustainable Economic Development of Russian Regions,” *Sustainability*, 2025. doi: 10.3390/su17093776.
- [17] D. Purwar, J. Flacke, E. A. Guzmán y R. Sliuzas, “A qualitative analysis of cascading effects of critical infrastructure service failure post torrential floods in formal & informal settlement: the study-case of Medellin city, Colombia,” *Sustainable and Resilient Infrastructure*, vol. 9, no. 5, pp. 496-512, 2024, doi: 10.1080/23789689.2024.2340924.
- [18] M. Haryanto y M. Mahsun, “Business and Politics in Urban Indonesia: Patrimonialism, Oligarchy and the State in Two Towns,” *Journal of Current Southeast Asian Affairs*, vol. 43, no. 2, pp. 177-197, 2024, doi: 10.1177/18681034241264574.
- [19] R. A. Hall, “National–Local Synergies for Development: How a Local Political Machine Delivered Infrastructure Results in Iloilo City,” *Journal of Current Southeast Asian Affairs*, vol. 43, no. 2, 2024, doi: 10.1177/18681034241265724.
- [20] B. Antičić y A. Stupar, “Urban Regeneration through Cultural–Tourism Entrepreneurship... Sombor, Serbia,” *Land*, 2024. doi: 10.3390/land13091379.
- [21] B. Callaway y P. H. C. Sant’Anna, “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 2021. doi: 10.1016/j.jeconom.2020.12.001.
- [22] T. Conley, “GMM Estimation with Cross-Sectional Dependence,” *Journal of Econometrics*, 1999. doi: 10.1016/S0304-4076(98)00084-0.
- [23] A. Abadie, A. Diamond y J. Hainmueller, “Synthetic Control Methods for Comparative Case Studies,” *JASA*, 2010. doi: 10.1198/jasa.2009.ap08746.
- [24] J. P. LeSage y R. K. Pace, *Introduction to Spatial Econometrics*. CRC Press, 2009.
- [25] [Camagni] R. Camagni, *Economía urbana*, trad. V. Galletto. Barcelona, España: Antoni Bosch Editor, 2005.

## **ANEXOS**

### **ANEXO 1. Códigos y datos**

En este anexo se entrega de manera digital: (i) los scripts en R (utils\_geo.R, maingeo.R, utilseco.R, maineco.R); (ii) la carpeta storage/ con los datos que produce el flujo (artefactos intermedios y salidas finales); y (iii) los datos nativos necesarios para la réplica: mallas de teselas regulares (5 ha y 10 ha), insumos de la Cámara de Comercio de Bogotá preprocesados (panel georreferenciado de matrículas y establecimientos) y el Censo Inmobiliario consolidado/estandarizado. La entrega respeta la estructura de carpetas y rutas relativas utilizada por los scripts para permitir su ejecución inmediata.

### **ANEXO 2. Tabla completa de coeficientes y especificaciones del caso demostrativo**

Se anexan resultados en Excel:

- DiD\_Results\_2025-08-31\_5Ha.xlsx
- DiD\_Results\_2025-08-31\_10Ha.xlsx