



Universidad del
Rosario

Un marco metodológico extendido de Diferencia en Diferencias

Autor

Carlos Fabian Ruiz Paredes

Director

Darwin Fauricio Cortés Cortés

Doctor en Economía

Facultad de Economía

Doctorado en Economía

Universidad del Rosario

Bogotá - Colombia

2026

Agradecimientos

A mi familia, mi mamá María Edilma y mis hermanos Jorge y Luis por su amor incondicional, paciencia y apoyo constante durante todo este proceso.

A mi director de tesis, Darwin Cortés por su guía académica, orientación y confianza en mi trabajo.

A mis compañeros de doctorado, José Tapias, Juan Urueña, Lina García, Renzo Clavijo, Roberto Perez, Julián Sánchez, Camilo Castillo, Nazly López, Heiner Rincón y demás estudiantes doctorales por el compañerismo, las conversaciones y los aprendizajes compartidos que enriquecieron esta experiencia.

Al personal docente y administrativo, por su dedicación y compromiso con la formación y la investigación.

Y de manera especial, a la Alianza EFI, en el marco del Ecosistema Científico de Colombia Científica, por el respaldo institucional y académico que hizo posible el desarrollo de este trabajo.

Resumen

Esta tesis propone un nuevo marco metodológico para estimar efectos causales en contextos donde las intervenciones no se aplican una sola vez, sino que se repiten, se intensifican o alteran la trayectoria del resultado. Mientras que la metodología tradicional de Diferencia en Diferencias (DiD) ha demostrado ser efectiva bajo supuestos de adopción única y tratamiento binario, su aplicación se vuelve problemática cuando las políticas se implementan por fases, con distinta intensidad, o de manera secuencial. Este trabajo responde a esa limitación al desarrollar una familia de diseños que permiten identificar efectos causales incluso en estos escenarios complejos.

El documento presenta cuatro diseños específicos que se adaptan a distintas formas de repetición e intensidad del tratamiento: desde escenarios donde la intervención produce saltos de nivel sin cambiar la tendencia, hasta casos con rupturas sostenidas en la trayectoria. Cada diseño incluye su propio esquema de identificación y estimación, lo cual permite diagnosticar qué tipo de dinámica está presente en los datos.

El marco propuesto busca ampliar el alcance práctico de los métodos DiD en estudios aplicados. En situaciones donde las políticas se despliegan en etapas, se repiten sobre las mismas unidades o varían en intensidad, esta tesis ofrece herramientas concretas para medir impactos acumulativos sin sacrificar rigor causal. Al integrar comparaciones unidad–unidad y tiempo–tiempo, incorporar covariables bajo supuestos débiles y descomponer efectos dinámicos, el enfoque permite adaptar la metodología DiD a realidades empíricas cada vez más frecuentes y desafiantes.

Palabras clave: Diferencia en Diferencias; tratamientos repetidos; efectos dinámicos; rupturas de tendencia; intensidad acumulativa; identificación causal.

Clasificación JEL: C13; C21; C23;

Abstract

This thesis proposes a new methodological framework to estimate causal effects in contexts where interventions are not applied only once, but are repeated, intensified, or alter the trajectory of the outcome. While the traditional Difference-in-Differences (DiD) methodology has proven effective under assumptions of single adoption and binary treatment, its application becomes problematic when policies are implemented in phases, with varying intensity, or sequentially. This work addresses that limitation by developing a family of designs that allow for the identification of causal effects even in these complex scenarios.

The document presents four specific designs that adapt to different forms of treatment repetition and intensity: from scenarios where the intervention produces level shifts without changing the trend, to cases with sustained trajectory breaks. Each design includes its own identification, estimation, and internal validation scheme, allowing the researcher to empirically diagnose which type of dynamic is present in the data. Throughout the thesis, a unified language is introduced that formalizes valid counterfactuals, systematizes identification conditions, and clarifies causal interpretation in each case.

Beyond solving a technical limitation, the proposed framework seeks to expand the practical scope of DiD methods in applied studies. In situations where policies unfold in stages, are repeated on the same units, or vary in intensity, this thesis offers concrete tools to measure cumulative impacts without sacrificing causal rigor. By integrating unit–unit and time–time comparisons, incorporating covariates under weak assumptions, and decomposing dynamic effects, the approach adapts the DiD methodology to increasingly frequent and challenging empirical realities.

Keywords: Difference-in-Differences; repeated treatments; dynamic effects; trend breaks; cumulative intensity; causal identification.

JEL Classification: C13; C21; C23.

Índice general

Introducción General	9
Convenciones y supuestos generales	14
1. Revisión de la literatura sobre Diferencia en Diferencias	16
2. Diseños DiD con tratamientos repetidos, asignación no simultánea y efectos de nivel sin ruptura de tendencia	33
2.1. Comparación unidad–unidad (promedio temporal)	36
2.2. Comparación tiempo–tiempo (promedio entre unidades)	59
3. Diseños DiD con tratamientos repetidos, asignación no simultánea y efectos de nivel con ruptura de tendencia	74
3.1. Comparación unidad–unidad (promedio temporal)	77
3.2. Comparación tiempo–tiempo con ruptura (promedio entre unidades)	92
4. Aplicación	110
4.1. Comparación unidad–unidad, sin ruptura	113
4.2. Comparación tiempo–tiempo, sin ruptura	117
4.3. Comparación unidad–unidad, con ruptura	120
4.4. Comparación tiempo–tiempo, con ruptura	122
4.5. Síntesis comparativa de resultados entre diseños	124
4.6. Análisis empírico comparativo con la literatura previa	126
5. Conclusiones	129
A. Demostraciones de resultados principales	139
A.1. Demostración de la Proposición 2.1	139
A.2. Demostración de la Proposición 2.2	140

A.3. Representaciones gráficas del estimador $ATT_{m,n}^{\circ}(t)$	141
A.4. Limitaciones de comparación entre unidades tratadas	142
A.5. Justificación del peso utilizado en la agregación de efectos	144
A.6. Justificación del peso utilizado en la agregación global	145
A.7. Demostraciones de los resultados asintóticos (diseño sin ruptura)	147
A.8. Ventanas admisibles y controles efectivos	148
A.9. Demostración de la Proposición 2.4	149
A.10. Demostración de la Proposición 2.3	150
A.11. Intuición del peso 2.2.10	151
A.12. Anexo: Invarianza del orden de agregación	151
A.13. Ventanas máxima y admisible	152
A.14. Demostraciones de los resultados asintóticos (diseño con ruptura)	155
A.15. Demostración de la Proposición 3.2	156
A.16. Demostración de la equivalencia entre las representaciones OR, IPW y DR	158
A.17. Demostración de los teoremas de inferencia	161
B. Anexo: Simulaciones	164
B.1. Generación de datos simulados (PGD)	164
B.2. Implementación del diseño sin ruptura de tendencia y de com- paración por cohortes de unidades	167
B.3. Implementación del diseño por cortes temporales (sin ruptura de tendencia)	172
B.4. Implementación del diseño con ruptura y comparación uni- dad–unidad	180
B.5. Implementación del diseño por cortes temporales (con ruptura de tendencia)	182

Índice de figuras

2.1.	Estimación puntual del efecto del tratamiento para la unidad m en $t = k$: $ATT_{m,n}^{\circ}(k)$	39
2.2.	Estimación puntual del efecto marginal del tratamiento: $ATT_{m,n}^{+}(k_3)$	43
2.3.	Evolución de y_t para las primeras ocho unidades.	56
2.4.	Ilustración esquemática de $ATT^{t^-,t^+}(k, h)$ para una ventana $[t^-, t^+]$ alrededor de k	62
3.1.	Ilustración de $ATT_{m,n}^{rup}(k_5)$ con control en $h-1$ en k_5 . La ventana de comparación es $[k_4, k_6] = [10, 18]$; el contrafactual extiende la tendencia en $h-1$ y el efecto puntual se marca en $t = 16$	79
3.2.	Trayectorias de y_t para las primeras ocho unidades (escenario con ruptura, sin dosis).	89
3.3.	Ilustración esquemática de $ATT_{rup}^{t^-,t^+}(k, h)$ para una ventana $[t^-, t^+]$ alrededor de k en el caso con ruptura.	94
A.1.	Estimación de $ATT_{m,n}^{\circ}(k)$, correspondiente al cambio de nivel de tratamiento de $h-2$ a $h-1$ en la unidad m . Se representa la trayectoria contrafactual esperada bajo el supuesto de tendencias paralelas.	141
A.2.	Estimación de $ATT_{m,n}^{\circ}(\bar{k})$, correspondiente al cambio de nivel de tratamiento de h a $h+1$ en la unidad m	142
A.3.	Estimación válida de $ATT_{n,m}^{+}(k_1)$ y limitaciones para otros efectos.	143
A.4.	Configuración ilustrativa de tiempos intercalados y niveles de intervención.	148
A.5.	Primer salto de m en k_1 : la ventana admisible (post) coincide con la ventana máxima (post), esto es, $[k_1, k_2]$	153

A.6. Tercer salto de m en k_5 : la ventana admisible (post) $[k_5, k_6]$ es un subconjunto de la ventana máxima (post) $[k_5, T]$. El tramo $[k_6, T]$ no es identificable con este control por falta de soporte en $h-1$ 154

Introducción General

En las últimas décadas, la economía aplicada ha experimentado lo que se conoce como una “revolución de la credibilidad” en la investigación empírica, resaltando la importancia de los diseños de evaluación de impacto rigurosos para identificar relaciones causales de manera confiable (Angrist y Pischke (2010)). La formulación de políticas públicas y la asignación eficiente de recursos dependen cada vez más de evidencia causal sólida proveniente de evaluaciones de impacto. Estas evaluaciones permiten capturar los efectos reales de intervenciones o programas, aislando su influencia de factores externos confusos. En este contexto, los métodos cuasi-experimentales han cobrado especial relevancia cuando los experimentos controlados aleatorios no son factibles por razones éticas, logísticas o de costo. En particular, métodos como la regresión discontinua, los experimentos naturales y el método de diferencia en diferencias (DiD) se han convertido en herramientas estándar de la evaluación de impacto en ciencias sociales (Angrist y Pischke (2010)). En conclusión, la capacidad de medir rigurosamente el impacto causal de intervenciones es esencial para validar teorías, mejorar la efectividad de programas sociales y legitimar el gasto público ante la sociedad, lo que explica la importancia central de los estudios de evaluación de impacto en la investigación contemporánea.

En este contexto, una de las metodologías más usadas para capturar el efecto de una intervención en escenarios específicos es el estimador conocido como *Two Way Fixed Effect* (TWFE, por sus siglas en inglés), numerosos trabajos han señalado que este estimador tradicional de efectos fijos de doble vía puede producir estimaciones sesgadas cuando existen efectos heterogéneos en el tiempo o entre grupos. En particular, de Chaisemartin y D’Haultfoeuille (2020) desarrollan una descomposición analítica del estimador TWFE, demostrando que éste equivale a un promedio ponderado de múltiples comparaciones 2x2, algunas de las cuales involucran grupos tratados en diferentes momentos. Dichas comparaciones contaminadas pueden asignar pesos negativos a ciertos

efectos, induciendo sesgos cuando los efectos del tratamiento varían entre cohortes o en el tiempo. Goodman-Bacon (2021) llega a conclusiones similares mediante una descomposición alternativa, y varios artículos recientes proponen soluciones para corregir o evitar este sesgo de heterogeneidad. Por ejemplo, Sun y Abraham (2021) introducen un estimador de evento dinámico que elimina los pesos negativos, esencialmente omitiendo comparaciones inapropiadas entre tratados tempranos y tardíos. Sant’Anna y Zhao (2020) desarrollan estimadores *doblemente robustos* para Diferencia en Diferencias (DiD por sus siglas en inglés), combinando modelos de resultado y de asignación a tratamiento de modo que la estimación del efecto de tratamiento promedio sobre los tratados, comúnmente conocido como ATT (Average Treatment Effect on the Treated) sea consistente incluso si uno de dichos modelos está mal especificado. Por su parte, Callaway y Sant’Anna (2021) proponen una estrategia basada en calcular efectos promedio por cohorte y período (los llamados *group-time ATTs*) y luego agregarlos de forma consistente; este enfoque permite efectos heterogéneos entre grupos y momentos de tratamiento, al tiempo que relaja la necesidad de un supuesto de no anticipación estricto. Por su parte. Otra contribución importante es la de Borusyak et al. (2024), quienes sugieren un método de imputación de contrafactuales (a veces denominado estimador de “regresión imputada”), el cual reconstruye la trayectoria contrafactual de los tratados usando unidades no tratadas y restaura la interpretabilidad de los coeficientes incluso con efectos dinámicos heterogéneos. Esta prolífica línea de investigación reciente ha quedado recogida en varias revisiones de literatura (Roth et al. (2023)), subrayando nuevas y mejores prácticas para diseños con tratamiento escalonado o efectos que varían en el tiempo. A pesar de estos avances, persisten importantes vacíos en la metodología DiD aplicada a contextos complejos. En primer lugar, la mayoría de trabajos asume que cada unidad recibe el tratamiento a lo sumo una vez (por ejemplo, en diseños escalonados cada unidad pasa de no tratada a tratada en una única ocasión). Sin embargo, en muchos programas de políticas públicas o intervenciones económicas, una misma unidad puede ser sometida repetidamente al *mismo* tratamiento en diferentes momentos. Por ejemplo, un país puede implementar reiteradas campañas similares de vacunación año tras año, o una empresa puede aplicar sucesivas rondas de un programa de formación a sus empleados. Los métodos existentes no abordan directamente cómo identificar y estimar efectos cuando el tratamiento se repite varias veces para las mismas unidades a lo largo del horizonte temporal. En segundo lugar, los diseños usuales de DiD suponen que el impacto del

tratamiento se manifiesta como un cambio de nivel (un *salto* discreto en el resultado) mantenido en el tiempo bajo la condición de tendencias paralelas post-tratamiento. No consideran la posibilidad de que una intervención altere la *tendencia* o pendiente de la trayectoria temporal del resultado, es decir, que provoque un cambio gradual en la tasa de crecimiento o declive de Y (variable de interés). En algunas aplicaciones, una política puede no generar un impacto inmediato considerable, pero sí modificar la evolución futura (por ejemplo, una reforma educativa podría no afectar instantáneamente los puntajes estudiantiles, pero sí cambiar su ritmo de mejora año con año). La literatura reciente en DiD típicamente modela efectos dinámicos mediante coeficientes de rezago (*event-study*) que eventualmente se estabilizan, mas no como cambios permanentes de pendiente en cada intervención; por tanto, se carece de métodos específicos para identificar rupturas de tendencia sucesivas dentro de un marco DiD. En tercer lugar, prácticamente todos los desarrollos mencionados tratan el tratamiento como una variable dicotómica (tratado/no tratado). En escenarios reales, las intervenciones pueden tener distintas *dosis o intensidades*. Por ejemplo, un subsidio gubernamental puede variar en monto, un programa de entrenamiento puede impartirse con distinta duración, o una política ambiental puede tener distintos grados de implementación según la región. Ignorar la intensidad del tratamiento puede llevar a conclusiones incompletas, por tanto, interesa estimar no solo si hubo un efecto, sino cuánto varía dicho efecto con la dosis recibida y cómo se acumula el impacto de dosis repetidas. Algunos trabajos han empezado a considerar tratamientos continuos dentro del enfoque DiD de Chaisemartin et al. (2025) extienden el estimador DID multiplicativo a casos de tratamiento no binario), pero no abordan el caso de tratamientos dosificados aplicados secuencialmente en el tiempo. En resumen, el problema de investigación central que aborda esta tesis es cómo extender los métodos de diferencias en diferencias para cubrir tres situaciones hasta ahora poco exploradas: (1) unidades con tratamientos repetidos en el tiempo; (2) intervenciones cuyo efecto puede manifestarse como un cambio en la pendiente de la trayectoria del resultado (además o en lugar de un cambio de nivel); y (3) tratamientos de intensidad variable o dosificados, cuyos efectos podrían acumularse de forma no lineal. Estos vacíos representan obstáculos para la evaluación rigurosa de políticas en contextos complejos, y llenarlos contribuirá a acercar los diseños DiD a situaciones empíricas más realistas.

El objetivo general de esta tesis es desarrollar métodos econométricos de Diferencia en Diferencias que permitan identificar y estimar efectos causales

en escenarios con múltiples tratamientos secuenciales y posibles cambios de tendencia en la variable de interés. Este objetivo general se desglosa en los siguientes objetivos específicos:

- **Objetivo específico 1:** Proponer un modelo DiD para escenarios de tratamientos *repetidos* que genere efectos de nivel en el resultado, y derivar un estimador consistente del efecto medio de cada intervención repetida sobre las unidades tratadas.
- **Objetivo específico 2:** Extender el modelo anterior para permitir rupturas de pendiente en cada intervención, incorporando metodologías para estimar tanto el efecto inmediato como el cambio en la tendencia causada por tratamientos sucesivos.

Estos objetivos específicos se abordan en los capítulos metodológicos de la tesis, procurando en cada caso contrastar las propuestas con los métodos existentes y evaluar su desempeño teórico y empírico. A continuación, se presenta un resumen de cada capítulo y sus contribuciones principales.

El Capítulo 1 ofrece un recorrido histórico por la evolución de los métodos DiD. Se discuten los orígenes del diseño 2x2 clásico, el surgimiento del modelo de efectos fijos de dos vías y su uso generalizado, y finalmente la “revolución” de métodos recientes (incluyendo las aportaciones de de Chaisemartin y D’Haultfoeuille (2020), Sun y Abraham (2021), Callaway y Sant’Anna (2021), Sant’Anna y Zhao (2020), Borusyak et al. (2024), entre otros) que corrigen las limitaciones del estimador tradicional. Este marco contextualiza las brechas identificadas y establece el punto de partida para las contribuciones de los capítulos subsiguientes.

El Capítulo 2 desarrolla un marco metodológico para estimar efectos causales en contextos con *tratamientos repetidos, asignación no simultánea y efectos de nivel sin ruptura de tendencia*. A diferencia de los diseños DiD tradicionales con una única intervención, aquí se asume que las unidades pueden recibir múltiples exposiciones al mismo tratamiento a lo largo del tiempo, acumulando un historial creciente de intervenciones. Cada nueva intervención genera un *salto permanente en el nivel del resultado*, sin alterar su tendencia estructural.

El capítulo propone dos estrategias complementarias de identificación y estimación. La primera, presentada en la Sección 2.1, compara unidades tratadas y no tratadas en un mismo instante, promediando los cambios antes y después del evento dentro de ventanas temporales. Este diseño permite

construir un efecto promedio entre unidades con diferente historial de tratamiento, y su principal aporte adicional consiste en introducir una comparación paralela con unidades nunca tratadas, lo cual da lugar a dos estimadores diferenciados —uno *entre tratadas* y otro *contrafactual puro*— que sirven como mecanismo de validación interna del supuesto de tendencias paralelas. La segunda estrategia, desarrollada en la Sección 2.2, centra el análisis en comparaciones *entre tiempos*, evaluando la variación promedio del resultado antes y después de cada tratamiento dentro de ventanas específicas. Como extensión metodológica, esta sección incorpora *placebos pre y post* para verificar la estabilidad de las trayectorias y detectar posibles desviaciones del diseño esperado.

Ambas estrategias producen estimadores tipo DiD del efecto marginal de una nueva intervención, con pesos construidos siguiendo la lógica de descomposición de TWFE propuesta por Goodman-Bacon (2021) y las adaptaciones recientes de de Chaisemartin y D’Haultfoeuille (2020). En conjunto, el capítulo extiende la literatura sobre diferencias en diferencias acumulativas, integrando de forma explícita el número de tratamientos previos H_{it} como variable estructural y proporcionando estimadores confiables, robustos y alineados con los desafíos reales de evaluación de políticas con múltiples fases.

El Capítulo 3 extiende el análisis de diferencias en diferencias a escenarios en los que el tratamiento no solo tiene efectos sobre el nivel del resultado, sino también sobre su tendencia. Bajo este diseño, cada nueva intervención puede modificar la pendiente de la trayectoria de la variable de interés, rompiendo así la dinámica previa al tratamiento. Se parte nuevamente de una variable acumulativa H_{it} , y el objetivo es estimar efectos marginales entre niveles consecutivos de tratamiento $h-1 \rightarrow h$, permitiendo que el tratamiento altere la evolución futura del resultado.

El capítulo se estructura en dos estrategias metodológicas complementarias. La primera, desarrollada en la Sección 3.1, mantiene la comparación *entre unidades* en un mismo instante $t=k$, pero admite que las trayectorias posteriores difieran en su tendencia. Su aporte adicional consiste en la descomposición del efecto estimado en dos componentes: un *efecto instantáneo*, que mide el salto en el momento del tratamiento, y un *efecto total*, que incorpora la evolución posterior de las trayectorias. Esta distinción permite evaluar si las intervenciones producen únicamente un cambio de nivel o también un cambio persistente en la tendencia de los resultados.

La segunda estrategia, presentada en la Sección 3.2, se basa en comparaciones *entre tiempos*, fijando ventanas $[t^-, t^+]$ que enmarcan la ocurrencia del

tratamiento. En este diseño, se estiman diferencias promedio entre periodos previos y posteriores al evento para unidades tratadas y de control, capturando las rupturas estructurales inducidas por la intervención. Como extensión metodológica, esta sección incorpora un bloque adicional dedicado a la inclusión de covariables X , que permite relajar el supuesto de tendencias paralelas no condicionales. Allí se definen tres representaciones equivalentes del efecto ajustado —*Outcome Regression (OR)*, *Inverse Probability Weighting (IPW)* y *Doubly Robust (DR)*— y se plantea un nuevo esquema de inferencia centrado en el estimador DR, por ser el más eficiente y estable frente a posibles errores de especificación.

En conjunto, el capítulo aporta un marco integral para analizar tratamientos repetidos con ruptura de tendencia, combinando la identificación de saltos y cambios estructurales con mecanismos de inferencia robustos ante heterogeneidad y desequilibrio temporal.

El Capítulo 4 presenta la aplicación empírica de los desarrollos teóricos expuestos en los Capítulos 2 y 3. En particular, se analiza el impacto de la liberalización —o desregulación— bancaria a nivel estatal sobre la participación laboral en Estados Unidos durante las décadas de 1970, 1980 y 1990. Los resultados obtenidos muestran una consistencia notable con los hallazgos de Leblebicioğlu y Weinberger (2020) y Dube et al. (2023). Finalmente, se concluye la investigación con las conclusiones recogidas en la tesis y los retos subsiguientes a las metodologías aquí presentadas.

Convenciones y supuestos generales

Para no repetir definiciones a lo largo del documento, se fija aquí la notación básica que se usará en todos los capítulos. Se establecen unidades $i \in \mathcal{I} = \{1, \dots, I\}$ y periodos $t \in \mathcal{T} = \{1, \dots, T\}$. En casos donde de establezca el instante k , simplemente se considera $t=k$.

Se usa H_{it} para el *nivel de intervención*: el número acumulado de intervenciones que ha recibido la unidad i hasta t . El resultado observado se denota y_{it} . En el caso de resultados potenciales, se define $Y_{it}(h)$ para referir el valor que tomaría y si en t la unidad i tuviera nivel h ; por consistencia, $y_{it} = Y_{it}(H_{it})$.

Se denota por ATT el *parámetro causal* definido en términos de resultados potenciales y por \widehat{ATT} su *forma identificada en observables* bajo los supuestos explicitados en cada sección. Según el contexto, se añaden subíndices ($ATT_{i,j}$), superíndices (ATT^{var}) y/o argumentos entre paréntesis ($ATT(k, h)$) para

indicar índices, variantes y configuraciones; su significado se precisará en cada sección, y la versión identificada seguirá la misma convención: $\widehat{ATT}_{i,j}^{\text{var}}(k, h)$.

En adelante se definen tres supuestos comunes a toda la tesis:

1 Supuesto: Irreversibilidad del tratamiento

Para toda unidad $i \in \mathcal{I}$ y todo par $s \leq t$: $H_{i,s} \leq H_{i,t}$.

2 Supuesto: Incremento unitario por periodo

Para toda unidad $i \in \mathcal{I}$ y todo $t \in \mathcal{T}$: $H_{i,t} - H_{i,t-1} \leq 1$.

3 Supuesto: Muestreo por unidad (i.i.d.) y no interferencia

Las secuencias $\{(Y_{i,t}, H_{i,t})_{t=1}^T\}_{i=1}^I$ son independientes e idénticamente distribuidas en i . Se permite dependencia temporal dentro de cada unidad, pero independencia entre unidades, y no hay interferencia entre unidades.

En esencia, el acumulado H_{it} no retrocede, como mucho crece de a una unidad por periodo, y asumimos independencia a nivel de unidad junto con ausencia de interferencia entre unidades.

Capítulo 1

Revisión de la literatura sobre Diferencia en Diferencias

Orígenes y modelo canónico 2x2

El enfoque de diferencia en diferencias (DiD) constituye uno de los métodos cuasiexperimentales más antiguos y ampliamente utilizados para la evaluación de impacto. Su formulación más básica compara la evolución del resultado de interés entre un grupo tratado y un grupo de control, antes y después de la implementación de una intervención. Este diseño elemental, conocido como modelo 2x2, se remonta a estudios clásicos como el análisis de la epidemia de cólera en Londres realizado por Snow (1849), y ha sido formalizado en numerosos textos de econometría aplicada (Angrist y Krueger (1999), Meyer (1995) y J. M. Wooldridge (2010)).

En este esquema, la estimación del efecto promedio del tratamiento sobre los tratados (ATT, por sus siglas en inglés) se define a partir del marco de resultados potenciales. Sea $D_i \in \{0, 1\}$ un indicador binario que toma el valor uno si la unidad i recibe tratamiento, y cero en caso contrario. Definimos $Y_{it}(1)$ como el resultado potencial que habría observado la unidad i en el periodo t bajo tratamiento, y $Y_{it}(0)$ como el que habría tenido sin tratamiento. Bajo esta notación, el ATT se expresa como:

$$\text{ATT} = \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) \mid D_i = 1]$$

Dado que no es posible observar simultáneamente ambos resultados potenciales para la misma unidad, se requiere un supuesto que permita identificar el contrafactual no observado. El supuesto fundamental en el diseño 2x2 es el

de *tendencias paralelas* o *tendencias comunes*, el cual postula que, en ausencia de tratamiento, el cambio promedio en el resultado para el grupo tratado habría sido igual al del grupo control:

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0]$$

Bajo este supuesto, el ATT puede estimarse a partir de datos observados mediante la siguiente expresión de diferencias en diferencias:

$$\widehat{\text{ATT}} = (\mathbb{E}[y_{i1} \mid D_i = 1] - \mathbb{E}[y_{i0} \mid D_i = 1]) - (\mathbb{E}[y_{i1} \mid D_i = 0] - \mathbb{E}[y_{i0} \mid D_i = 0])$$

Una forma equivalente de estimar este parámetro es mediante una regresión lineal con interacción, de la forma:

$$y_{it} = \alpha + \beta D_i + \gamma \text{Post}_t + \delta(D_i \times \text{Post}_t) + \varepsilon_{it},$$

donde y_{it} representa el resultado observado para la unidad i en el periodo t , D_i indica pertenencia al grupo tratado, Post_t toma el valor uno para el periodo posterior al tratamiento, y el término de interacción captura el efecto diferencial. Bajo los supuestos del modelo canónico, en particular el de tendencias paralelas, el coeficiente δ estimado mediante mínimos cuadrados ordinarios (MCO) coincide con la estimación del ATT. Esta equivalencia ha hecho del modelo DiD un instrumento ampliamente utilizado en la práctica empírica, debido a su sencillez, interpretabilidad y facilidad de implementación.

Generalización con efectos fijos: el modelo TWFE y su contraparte en primeras diferencias

En la práctica moderna muchas aplicaciones del enfoque de diferencia en diferencias (DiD) superan el esquema 2x2 y utilizan paneles con múltiples periodos, en los que distintas unidades reciben el tratamiento en diferentes momentos. Este patrón, conocido como *adopción escalonada* (*staggered adoption*), refleja una situación común en la implementación progresiva de políticas públicas a lo largo del tiempo.

Para este tipo de datos, el modelo por defecto en la literatura empírica ha sido la regresión con efectos fijos por unidad y por tiempo, conocida como TWFE (por sus siglas en inglés, *two-way fixed effects*). La forma funcional del modelo TWFE es:

$$y_{it} = \alpha_i + \lambda_t + \beta_{fe}D_{it} + \varepsilon_{it}, \quad (1.0.1)$$

donde y_{it} representa el resultado observado para la unidad i en el periodo t , α_i captura heterogeneidad no observable constante por unidad, λ_t controla por shocks comunes a todos los grupos en el tiempo, y D_{it} es un indicador binario que toma el valor uno si la unidad i ha recibido tratamiento en el periodo t , y cero en caso contrario.

Este modelo asume un tratamiento binario, aplicado una única vez a cada unidad, con efectos constantes y persistentes. Bajo los supuestos de tendencias paralelas y ausencia de efectos heterogéneos, el coeficiente β_{fe} estimado mediante Mínimos Cuadrados Ordinarios (MCO) identifica el efecto promedio del tratamiento sobre las unidades tratadas. La especificación TWFE se ha vuelto dominante en estudios empíricos debido a su facilidad de implementación y a su capacidad para controlar confusores no observados fijos en el tiempo. Ejemplos representativos incluyen el análisis del EITC y la salud infantil en Estados Unidos (Hoynes et al. (2015)), o la expansión del seguro de salud en la infancia (Currie y Gruber (1996)).

de Chaisemartin y D’Haultfoeuille (2020) documentan que aproximadamente el 19 % de los artículos empíricos publicados por la *American Economic Review* entre 2010 y 2012 utilizaron regresiones TWFE para estimar efectos causales. No obstante, también identifican una estrategia alternativa comúnmente utilizada: la regresión en primeras diferencias, cuya forma funcional es la siguiente:

$$\Delta y_{it} = \lambda_t + \beta_{fd}\Delta D_{it} + \varepsilon_{it}, \quad (1.0.2)$$

donde $\Delta y_{it} = y_{it} - y_{i,t-1}$ representa el cambio en el resultado para la unidad i entre dos periodos consecutivos, y $\Delta D_{it} = D_{it} - D_{i,t-1}$ capta el cambio en la condición de tratamiento. El coeficiente β_{fd} busca estimar el efecto del tratamiento sobre las unidades que cambian su estatus de tratamiento —conocidas como *switchers*— entre dos periodos adyacentes. Esta especificación es particularmente útil cuando se dispone de pocos periodos, o cuando se busca controlar directamente por heterogeneidad no observable constante a lo largo del tiempo sin especificar efectos fijos por unidad.

Cabe señalar que, en paneles balanceados con solo dos periodos, los estimadores TWFE y de primeras diferencias son algebraicamente equivalentes. Sin embargo, esta equivalencia se pierde cuando se dispone de múltiples periodos o paneles desbalanceados (de Chaisemartin y D’Haultfoeuille (2020)).

Un ejemplo del uso empírico de esta estrategia lo proporciona Gentzkow et al. (2011), quienes analizan el efecto de la entrada y salida de periódicos en Estados Unidos sobre los resultados electorales, utilizando modelos en primeras diferencias para capturar cambios en la exposición mediática a lo largo del tiempo.

Descomposiciones del estimador TWFE y sus extensiones

Cuando el tratamiento es binario pero se implementa de forma escalonada entre diferentes unidades a lo largo del tiempo, el estimador TWFE ya no equivale al clásico estimador DiD 2x2. En presencia de efectos heterogéneos, este estimador puede mezclar comparaciones entre grupos tratados en diferentes momentos, generando estimaciones sesgadas o de difícil interpretación.

Para entender estas limitaciones, varios trabajos han propuesto reescrituras o descomposiciones del estimador TWFE desde distintos ángulos. Strezhnev (2018) lo expresa algebraicamente como un promedio no ponderado entre pares de observaciones con distinta exposición al tratamiento. Athey y Imbens (2022) lo interpretan como una combinación de efectos a distintos tiempos relativos al tratamiento, destacando los riesgos de interpretaciones equivocadas cuando hay dinámica en los efectos. Borusyak y Jaravel (2018), por su parte, muestran que los coeficientes TWFE pueden ser entendidos como promedios ponderados de efectos específicos por grupo y periodo, en los que algunos pesos pueden ser negativos, lo que complica su interpretación causal.

Entre estos enfoques, dos propuestas han sido particularmente influyentes por su claridad estructural y relevancia empírica. La descomposición de de Chaisemartin y D'Haultfoeuille (2020), basada en diferencias en celdas de grupo y periodo con cambios en el tratamiento, permite identificar claramente qué observaciones contribuyen a la estimación y con qué peso. Por otro lado, la descomposición alternativa de Goodman-Bacon (2021) interpreta el estimador TWFE como un promedio ponderado de estimadores DiD tipo 2x2 entre cohortes tratadas en distintos momentos o frente a grupos nunca tratados.

En la presente tesis el análisis se centrará en estas dos últimas estrategias, la de de Chaisemartin y D'Haultfoeuille (2020) y la de Goodman-Bacon (2021), por ser las más directamente aplicables al diseño de evaluaciones con tratamientos repetidos y adopción no simultánea, como el que se desarrolla en este trabajo.

Descomposición del estimador TWFE bajo efectos del tratamiento heterogéneos

El modelo TWFE descrito en la ecuación (1.0.1) ha sido ampliamente utilizado bajo el supuesto de que el efecto del tratamiento es constante a lo largo del tiempo y entre unidades. Sin embargo, este supuesto rara vez se cumple en la práctica. Cuando el efecto del tratamiento varía entre grupos y periodos, el estimador β_{fe} deja de identificar directamente el efecto promedio del tratamiento y, en cambio, puede alejarse considerablemente del parámetro de interés.

de Chaisemartin y D’Haultfoeuille (2020) demuestran que, bajo un diseño *sharp* (esto es, donde el tratamiento es binario y todos los miembros del grupo reciben tratamiento completo o ninguno), el estimador β_{fe} puede expresarse como una combinación ponderada de los efectos promedio del tratamiento por celda (g, t) . Específicamente:

$$\beta_{fe} = \sum_{(g,t) \in \mathcal{S}} w_{g,t} \cdot ATE_{g,t}, \quad (1.0.3)$$

donde \mathcal{S} denota el conjunto de celdas en las que ocurre un cambio en la condición de tratamiento, $ATE_{g,t}$ representa el efecto promedio del tratamiento para el grupo g en el periodo t , y $w_{g,t}$ es un peso que depende únicamente de la estructura del panel (distribución de tratamientos y número de unidades por celda), pero no del resultado observado.

Un hallazgo crucial de esta descomposición es que algunos pesos $w_{g,t}$ pueden ser negativos. Esto implica que, cuando los efectos del tratamiento son heterogéneos entre grupos y periodos, el estimador β_{fe} puede otorgar peso negativo a ciertos efectos positivos (o viceversa), alejándose del efecto promedio esperado. En aplicaciones empíricas revisadas por los autores, más del 50 % de los pesos asociados a las celdas tratadas eran negativos, lo que ha motivado un creciente interés en estudiar la composición interna del estimador TWFE.

Esta descomposición permite visualizar con claridad cómo el estimador TWFE combina información a lo largo del tiempo y entre grupos. A continuación se presenta una descomposición alternativa, basada en comparaciones tipo DiD 2x2 entre cohortes, propuesta por Goodman-Bacon (2021).

Descomposición alternativa tipo 2x2: enfoque de Goodman-Bacon

Una descomposición alternativa del estimador TWFE fue propuesta por Goodman-Bacon (2021), quien demuestra que, en contextos de adopción escalonada, dicho estimador puede interpretarse como un promedio ponderado de múltiples estimadores DiD tipo 2x2.

El punto de partida es reorganizar el panel con múltiples periodos y grupos tratados en distintos momentos en una serie de subconjuntos más simples, cada uno comparable mediante un diseño 2x2 clásico. Bajo esta lógica, el estimador TWFE puede escribirse como:

$$\hat{\beta}_{\text{TWFE}} = \sum_{\ell} \omega_{\ell} \cdot \widehat{\text{DiD}}^{(\ell)}, \quad (1.0.4)$$

donde $\widehat{\text{DiD}}^{(\ell)}$ es un estimador de diferencias en diferencias construido entre dos grupos definidos por su cohorte de tratamiento, y ω_{ℓ} es el peso asociado a esa comparación, el cual depende de la varianza del tratamiento y del tamaño relativo de los grupos involucrados.

Esta descomposición permite distinguir tres tipos principales de comparaciones:

- (1) Comparaciones entre unidades tratadas y unidades que nunca reciben tratamiento;
- (2) Comparaciones entre unidades tratadas temprano y unidades tratadas más tarde, antes de que estas últimas sean tratadas;
- (3) Comparaciones entre unidades tratadas más tarde y unidades tratadas temprano, después de que las segundas ya han sido tratadas.

Cada uno de estos subconjuntos genera un estimador 2x2 clásico, pero la combinación ponderada de todos ellos no siempre coincide con el efecto promedio del tratamiento. Un hallazgo clave de Goodman-Bacon (2021) es que algunos de los pesos ω_{ℓ} pueden ser negativos, particularmente cuando se comparan grupos tratados en diferentes momentos. Esto ocurre con mayor frecuencia cuando el efecto del tratamiento es dinámico o heterogéneo.

Aunque cada estimador 2x2 representa una comparación válida entre grupos y periodos específicos, la descomposición de Goodman-Bacon permite entender cómo el estimador TWFE sintetiza esta información mediante un promedio ponderado, cuya estructura depende de la dinámica de tratamiento

y la composición del panel. Esta perspectiva ha contribuido significativamente a clarificar la interpretación del TWFE en estudios con adopción escalonada.

Extensión al diseño *fuzzy*: tratamiento parcial y grupos con exposición intermedia

Las descomposiciones anteriores se han basado en un diseño *sharp*, en el cual cada unidad dentro de un grupo tratado recibe efectivamente el tratamiento en su totalidad. Sin embargo, muchas aplicaciones empíricas presentan escenarios *fuzzy*, donde la intensidad del tratamiento varía entre unidades de un mismo grupo o existen periodos en los que solo una fracción de la población es tratada.

En este contexto, la variable de tratamiento D_{it} puede asumir valores continuos entre 0 y 1, reflejando una dosificación o intensidad parcial. Por ejemplo, un grupo puede estar parcialmente expuesto a una política si solo algunas unidades la reciben, o si la magnitud del tratamiento difiere entre subpoblaciones.

Para estos casos, de Chaisemartin y D’Haultfoeuille (2020) extienden la descomposición del estimador TWFE a un nuevo estimador que identifican como:

$$\beta_{fe} = \sum_{(g,t) \in \mathcal{S}} w_{g,t}^{\text{fuzzy}} \cdot \text{ATT}_{g,t},$$

donde:

- $\text{ATT}_{g,t} = \mathbb{E}[Y_{gt}(1) - Y_{gt}(0) \mid \Delta D_{gt} \neq 0]$ representa el efecto promedio del tratamiento entre las unidades que cambian su nivel de exposición entre $t - 1$ y t ,
- $w_{g,t}^{\text{fuzzy}}$ son pesos construidos a partir de los cambios en el tratamiento promedio por grupo, definidos como $\Delta D_{gt} = D_{gt} - D_{g,t-1}$.

En este escenario, las celdas (g, t) activas son aquellas donde se observa un cambio (no necesariamente binario) en la proporción tratada, es decir, donde $\Delta D_{gt} \neq 0$. El estimador TWFE en contextos *fuzzy* puede entonces interpretarse como una media ponderada de efectos locales sobre los *switchers parciales*.

Este resultado es relevante para el análisis de políticas graduales, descentralizadas o de implementación desigual, donde los diseños experimentales o

cuasiexperimentales clásicos (binarios y homogéneos) no son aplicables. Sin embargo, persisten las advertencias ya señaladas para el caso sharp: los pesos $w_{g,t}^{\text{fuzzy}}$ también pueden ser negativos, lo que introduce riesgos de sesgo si los efectos del tratamiento varían entre grupos o a lo largo del tiempo.

La identificación en este caso depende del supuesto de tendencias paralelas en niveles de tratamiento continuo, lo cual exige precaución adicional al interpretar los resultados. En la práctica, se recomienda contrastar estimaciones TWFE con métodos robustos o específicos para diseños con dosificación, como los propuestos por Callaway et al. (2024) o estimadores basados en *switchers* individuales.

Limitaciones estructurales del modelo TWFE

La regresión con efectos fijos de dos vías (TWFE), representada en la ecuación (1.0.1), ha sido objeto de críticas en la literatura, debido a sus limitaciones en contextos donde no se cumplen los supuestos de homogeneidad del tratamiento ni de comparaciones válidas entre grupos.

Antes de que se formalizaran las descomposiciones algebraicas del estimador TWFE, algunos trabajos ya habían advertido sobre sus posibles limitaciones en aplicaciones empíricas específicas. Por ejemplo, Wolfers (2006) plantea que los estudios de eventos con tendencias unitarias pueden conducir a inferencias engañosas, especialmente cuando los efectos del tratamiento no son inmediatos ni homogéneos. De manera similar, Meer y West (2016) muestran que los efectos del salario mínimo sobre el empleo pueden estar sesgados en modelos con efectos fijos cuando existe dinámica de tratamiento y heterogeneidad temporal.

Estas advertencias empíricas se enmarcan en una preocupación más general expresada por Imbens (2015), quien señala que la adopción limitada de nuevos estimadores en la literatura aplicada se debe, en parte, a la falta de evidencia convincente sobre las fallas de métodos simples como la regresión lineal con efectos fijos. Este llamado motivó trabajos posteriores que proveen dicha evidencia, como el de Borusyak y Jaravel (2018).

Aunque Borusyak y Jaravel (2018) reconocen estos antecedentes, su trabajo constituye la primera crítica sistemática y formal al uso del modelo TWFE en contextos de adopción escalonada. Demuestran que, cuando los grupos adoptan el tratamiento en distintos momentos, el estimador TWFE puede utilizar como comparación a unidades que ya han sido tratadas, lo que rompe con el supuesto de un grupo de control válido. Esto implica que el estimador no

identifica un ATT claro, sino una combinación de efectos con interpretaciones ambiguas. Aunque no recurren a una descomposición algebraica explícita, sus argumentos destacan la invalidez de las comparaciones implícitas en la estimación estándar.

Sun y Abraham (2021) profundizan esta crítica mostrando que, incluso bajo supuestos estándar de tendencias paralelas, el estimador TWFE confunde los efectos contemporáneos con los efectos dinámicos del tratamiento. Específicamente, cuando se incluyen unidades tratadas anticipadamente como control para unidades tratadas más tarde, el estimador introduce un sesgo debido a la superposición de efectos temporales heterogéneos. Además, estos autores proponen un marco de descomposición alternativo que revela explícitamente las comparaciones problemáticas implícitas en el estimador TWFE.

Una línea complementaria de críticas al enfoque de diferencias en diferencias con efectos fijos de dos vías (TWFE) surge de trabajos que advierten sobre la ambigüedad de la estimación cuando el tratamiento no es binario, sino que varía en intensidad o se presenta en múltiples niveles. En estos casos, la comparación directa entre efectos locales puede inducir a errores de interpretación si no se cuenta con un marco teórico claro que justifique dicha comparación.

Por ejemplo, Oreopoulos (2006), Angrist y Fernández-Val (2013) y Mogstad et al. (2018) discuten los riesgos de comparar parámetros como $ATT(d | d')$, en los cuales el efecto del tratamiento en el nivel d se compara con el efecto en otro nivel d' . Estos estudios, desarrollados principalmente en el contexto de los efectos locales promedio (LATE), muestran que dicha comparación carece de sentido causal en ausencia de supuestos adicionales. El problema se agrava en contextos con variación continua del tratamiento, donde la interpretación de los estimandos depende críticamente de la forma funcional y la composición del grupo tratado.

En la misma línea, Frölich (2004) y Fricke (2017) evidencian que, en contextos con múltiples tratamientos discretos, la agregación de efectos puede no representar un parámetro bien definido. Es decir, el estimando resultante de un modelo TWFE puede ser una combinación arbitraria de efectos heterogéneos sin una interpretación clara, especialmente si las exposiciones al tratamiento no son homogéneas entre grupos ni constantes en el tiempo.

Estas advertencias, aunque previas a las descomposiciones algebraicas formales, sentaron las bases para cuestionar la validez causal del estimador TWFE en escenarios más complejos que el diseño clásico 2x2. Su relevancia es particularmente alta en aplicaciones empíricas con políticas dosificadas, sub-

sidios proporcionales o intervenciones con intensidad variable entre unidades.

Las críticas anteriores, aunque conceptualmente sólidas, no mostraban de forma explícita la mecánica interna por la cual el estimador TWFE produce estimaciones sesgadas o difíciles de interpretar. Este vacío fue subsanado con las descomposiciones algebraicas propuestas por de Chaisemartin y D’Haultfoeuille (2020) y Goodman-Bacon (2021), quienes aportan una caracterización estructural del estimador TWFE en contextos con heterogeneidad y tratamiento escalonado.

de Chaisemartin y D’Haultfoeuille (2020) demuestran que el TWFE no identifica un efecto promedio global, sino una media ponderada de efectos específicos por grupo y periodo ($ATT_{g,t}$), donde los pesos pueden ser negativos o desproporcionados si se utilizan como contrafactual unidades que ya han sido tratadas. Esta estructura puede distorsionar gravemente la estimación, incluso generando estimadores de signo opuesto al efecto real.

De forma complementaria, Goodman-Bacon (2021) descompone el TWFE en una combinación de estimadores 2x2, mostrando que muchas de estas comparaciones ocurren entre grupos tratados en distintos momentos. Si los efectos son dinámicos o heterogéneos, estas comparaciones cruzadas inducen sesgos sistemáticos, aún cuando se cumpla el supuesto clásico de tendencias paralelas.

Extensiones metodológicas recientes

Las descomposiciones recientes del estimador TWFE, particularmente las propuestas por Goodman-Bacon (2021) y de Chaisemartin y D’Haultfoeuille (2020), revelaron que, en contextos con tratamiento escalonado y efectos heterogéneos, el estimador convencional puede interpretarse como un promedio ponderado de comparaciones tipo DiD simples, muchas de las cuales son inválidas o incluso con pesos negativos. Estas contribuciones marcaron un punto de inflexión metodológico al demostrar que el sesgo no proviene solo de violaciones al supuesto de tendencias paralelas, sino de la forma en que el TWFE combina información entre grupos y periodos. En respuesta, emergieron múltiples extensiones que buscan evitar estas comparaciones espurias, ya sea restringiendo las comparaciones a pares válidos o reescribiendo el parámetro de interés como un conjunto estructurado de efectos promedio. En términos generales, los enfoques recientes pueden clasificarse según su foco: mientras de Chaisemartin y D’Haultfoeuille (2020) proponen estimadores que fijan los periodos y comparan grupos para construir $ATT_{g,t}$, Goodman-Bacon (2021)

descompone el TWFE como un promedio de comparaciones entre cohortes tratadas en distintos momentos, es decir, fijando unidades y comparando momentos. A partir de esta base, han surgido extensiones que permiten evaluar efectos dinámicos, múltiples tratamientos, intensidades continuas y esquemas fuzzy, como se describe a continuación.

Diseños con adopción escalonada y correcciones al TWFE

Las críticas metodológicas al estimador TWFE en contextos con adopción escalonada del tratamiento han motivado el desarrollo de estrategias que corrigen el sesgo introducido por comparaciones inválidas entre grupos en diferentes etapas de tratamiento. Estas extensiones buscan preservar la validez de la inferencia causal bajo efectos heterogéneos, evitando las ponderaciones problemáticas que afectan al estimador tradicional en presencia de dinámicas complejas.

Una primera línea de avance está representada por el enfoque de Callaway y Sant’Anna (2021), quienes proponen descomponer el efecto causal promedio en un conjunto de efectos específicos para cada cohorte de tratamiento y cada periodo ($ATT_{g,t}$). Para cada par (g, t) , se compara el grupo tratado g con un conjunto adecuado de controles que aún no han recibido el tratamiento en t , lo cual garantiza una comparación válida y evita las inferencias contaminadas por unidades ya tratadas. Los $ATT_{g,t}$ pueden luego agregarse usando distintos esquemas de ponderación (por ejemplo, por tamaño de grupo o calendario de tratamiento), lo que otorga flexibilidad interpretativa y mejora la trazabilidad del estimador.

En paralelo, Sun y Abraham (2021) desarrollan un enfoque alternativo mediante modelos de eventos (event studies) robustos a heterogeneidad del tratamiento. En lugar de imponer una dinámica común para todos los grupos, proponen especificaciones que permiten coeficientes diferenciados por cohorte, usando regresores de tiempo relativo (event-time) completamente interactuados. Esto evita la imposición de dinámicas paralelas implícitas y permite interpretar directamente la evolución del tratamiento por cohorte sin incurrir en los sesgos del TWFE tradicional.

Un enfoque metodológicamente distinto, pero complementario, es el propuesto por Borusyak et al. (2024), quienes desarrollan un método de imputación directa del contrafactual no tratado. En su marco, se estima primero el componente no tratado del resultado ($Y_{it}(0)$) usando unidades nunca tratadas, y luego se calcula el efecto del tratamiento como la diferencia entre el

resultado observado y el imputado. Esta estrategia evita completamente la necesidad de especificar una forma funcional para el efecto del tratamiento y es especialmente útil cuando la adopción es altamente escalonada o irregular.

Finalmente, Bellego et al. (2025) proponen una generalización conocida como *Chained DiD*, que permite estimar efectos promedio aún en paneles desbalanceados, con rotación muestral o estructuras incompletas. El método se basa en encadenar comparaciones válidas a lo largo del tiempo, manteniendo la lógica de identificación del $ATT_{g,t}$ pero extendiéndola a contextos donde las cohortes y los periodos no están uniformemente representados.

En conjunto, estos enfoques han desplazado al TWFE como estándar en escenarios de tratamiento escalonado, ofreciendo alternativas que permiten inferencias más sólidas, interpretables y compatibles con efectos dinámicos y heterogéneos.

Diseños con tratamiento múltiple o secuencial

En muchos contextos aplicados, las unidades de observación pueden estar expuestas a más de una intervención a lo largo del tiempo, ya sea de forma secuencial (eventos escalonados) o simultánea (tratamientos concurrentes). Estos escenarios representan un reto metodológico sustantivo para los modelos DiD, ya que la estimación de un efecto causal limpio requiere distinguir entre el impacto de cada tratamiento, evitar contaminación entre ellos y controlar adecuadamente por la heterogeneidad de efectos.

Una contribución reciente en este frente es la de Tsai (2024), quien propone un estimador de doble diferencia en diferencias (*Double DiD*) para casos con dos tratamientos secuenciales que pueden generar confusión causal. Su estrategia se desarrolla en dos etapas: en la primera, estima el efecto conjunto de ambos tratamientos usando como grupo de control a unidades no expuestas a ninguno; en la segunda, aísla el efecto del tratamiento de interés empleando como control a unidades que aún no han recibido el segundo tratamiento. El método requiere suponer la existencia de tendencias paralelas con respecto al tratamiento “confusor”, así como la ausencia de anticipación al segundo evento. El estimador está implementado en el paquete `fastdid` de R y muestra ventajas prácticas en contextos como reformas educativas o políticas económicas superpuestas.

Por su parte, de Chaisemartin y D’Haultfoeuille (2023) abordan el problema de identificar efectos causales cuando las unidades pueden recibir múltiples tratamientos distintos en diferentes momentos del tiempo. Muestran que, en

estos escenarios, los modelos TWFE pueden generar estimaciones sesgadas y muy difíciles de interpretar, ya que los coeficientes obtenidos combinan efectos de distintos tratamientos con ponderaciones potencialmente negativas o contaminadas. Como alternativa, proponen un nuevo estimador basado en diferencias en diferencias seleccionadas cuidadosamente (DID_M^f y DID_M^b), que permiten estimar el efecto del primer tratamiento manteniendo constantes los demás. Esta estimación se realiza únicamente sobre subconjuntos válidos de unidades donde el tratamiento de interés cambia, mientras los otros se mantienen fijos, y donde existe un grupo comparable. El efecto final se obtiene como un promedio ponderado sobre estos subconjuntos.

Ambos enfoques representan avances sustanciales para analizar escenarios donde las unidades reciben más de un tratamiento en distintos momentos o de forma simultánea. La literatura emergente en esta línea ha comenzado a sistematizar herramientas que permiten distinguir efectos específicos en presencia de secuencias de intervención o políticas correlacionadas, contribuyendo así al desarrollo de estimadores más robustos y aplicables a situaciones reales complejas.

Tratamientos de intensidad continua o dosificada

El enfoque DiD ha sido recientemente extendido para abordar situaciones en las que el tratamiento no es dicotómico, sino que presenta distintas intensidades o dosis. Este tipo de intervenciones es común en contextos donde la exposición a una política varía cuantitativamente entre unidades o a lo largo del tiempo, por ejemplo, subsidios diferenciados, impuestos progresivos o cobertura parcial de programas, lo cual desafía los supuestos tradicionales de identificación y requiere nuevos estimadores adaptados a la naturaleza continua del tratamiento.

Callaway et al. (2024) proponen una formulación general del DiD para tratamientos continuos, en la que definen un conjunto de parámetros análogos al ATT, incluyendo la respuesta causal promedio ante un cambio marginal en la intensidad del tratamiento. Su enfoque se basa en un supuesto de tendencias paralelas generalizadas: en ausencia de tratamiento, las trayectorias medias de los resultados deben ser paralelas entre unidades con niveles similares de dosis. Los autores advierten, sin embargo, que comparar efectos entre diferentes niveles de intensidad puede generar sesgos de selección no corregibles únicamente mediante este supuesto. Para superar tales limitaciones, proponen estrategias de estimación no paramétricas y discuten supuestos adicionales

como la monotonicidad en la respuesta que permiten interpretar diferencias de resultados entre dosis de manera causal. Además, ilustran cómo los modelos lineales con efectos fijos tradicionales pueden llevar a interpretaciones erróneas, al imponer una estructura funcional inadecuada cuando la intensidad varía.

De forma complementaria, de Chaisemartin et al. (2025) desarrollan una familia de estimadores DiD adaptados a contextos de tratamiento continuo con variación periódica. Su metodología distingue entre unidades que modifican su nivel de tratamiento a lo largo del tiempo (switchers) y aquellas que mantienen una dosis constante (stayers). Bajo un supuesto de tendencias paralelas condicionales al nivel inicial del tratamiento, el efecto marginal del tratamiento puede identificarse como un promedio ponderado de las pendientes (slopes) de los resultados potenciales para los switchers, utilizando como contrafactuales a los stayers con la misma dosis inicial. Esta estrategia permite una estimación más precisa del efecto de cambios graduales en la intensidad del tratamiento, incluso cuando dicho efecto es no lineal o varía en el tiempo. Asimismo, los autores extienden el marco al caso de diseños fuzzy con dosis continuas, permitiendo estimar efectos locales inducidos por variables instrumento. Sus estimadores cuentan con propiedades deseables tales como, doble robustez, consistencia \sqrt{n} y normalidad asintótica, y han sido aplicados empíricamente en estudios como la estimación de la elasticidad precio-consumo de la gasolina.

Ambos enfoques abren nuevas posibilidades metodológicas para el análisis causal en entornos con tratamientos dosificados, superando las limitaciones del DiD clásico e incorporando herramientas que permiten modelar con mayor fidelidad la complejidad de las políticas públicas reales.

Diseños *fuzzy* y tratamiento parcial

Los enfoques estándar de diferencia en diferencias (DiD) asumen que el tratamiento es asignado de manera dicotómica y completamente observada: una unidad está tratada o no lo está. Sin embargo, en muchas intervenciones reales, la implementación del tratamiento es incompleta o imperfecta, lo que da lugar a escenarios de *tratamiento parcial o cumplimiento imperfecto*. Estos casos pueden surgir, por ejemplo, cuando solo una fracción de las unidades elegibles accede efectivamente al programa, o cuando la intensidad del tratamiento depende de decisiones individuales no observadas. Para abordar esta complejidad, se han desarrollado diseños *fuzzy DiD* que combinan la lógica del DiD con herramientas de variables instrumentales (IV).

de Chaisemartin y D’Haultfoeuille (2018) introducen una estructura formal para el análisis de tratamientos imperfectos en contextos DiD, demostrando que el estimador tradicional tipo Wald-DID —cociente entre la diferencia-en-diferencias en el resultado y la diferencia-en-diferencias en la tasa de tratamiento— solo identifica un efecto causal válido bajo supuestos de homogeneidad muy restrictivos. Para relajar dichas condiciones, proponen estimandos que permiten identificar un efecto promedio local del tratamiento (LATE) para el subgrupo de *compliers*, es decir, las unidades cuya decisión de tratarse fue inducida por la intervención. Su marco requiere supuestos como: (i) monotonía en la adopción del tratamiento, (ii) ausencia de efectos directos del instrumento sobre el resultado, y (iii) tendencias paralelas condicionales para los *compliers*. Asimismo, los autores extienden el análisis a efectos de cuantiles locales (LQTE) y desarrollan una implementación computacional en Stata (`fuzzydid`), que admite múltiples periodos, grupos y tratamientos no binarios, y permite realizar análisis de identificación parcial en ausencia de supuestos fuertes.

Por su parte, Picchetti y Campos de Xavier Pinto (2022) proponen un enfoque alternativo basado en el marco de los efectos marginales de tratamiento (MTE) para casos donde existe selección no aleatoria en la participación. Argumentan que los individuos pueden decidir tratarse en función de sus ganancias esperadas, lo que introduce un sesgo de selección sobre beneficios no observables. Combinando supuestos del DiD y de variables instrumentales, derivan condiciones de identificación para la curva MTE en este contexto, permitiendo estimar cómo varían los efectos causales a lo largo de la distribución de propensión a tratarse. Proponen dos estimadores (uno paramétrico y otro semiparamétrico), así como una versión del LATE robusta ante especificación incorrecta del modelo MTE. A diferencia del enfoque de 2018, este modelo reconoce explícitamente la posible violación del supuesto de ignorabilidad, adoptando técnicas de identificación inspiradas en el trabajo de Heckman y Vytlačil (2007).

Ambos enfoques reflejan una evolución importante del diseño DiD hacia escenarios más realistas, donde la implementación de políticas es parcial o endógena. En tales casos, el uso de estimadores que incorporan variación inducida por instrumentos, junto con supuestos estructurales plausibles, permite recuperar parámetros causales relevantes incluso en presencia de incumplimiento, selección o heterogeneidad no observable.

Tratamiento de efectos dinámicos o heterogéneos

El análisis de efectos dinámicos —es decir, la evolución del impacto de una intervención a lo largo del tiempo— se ha vuelto una práctica estándar en estudios DiD con múltiples periodos. Los denominados *event studies* permiten trazar perfiles temporales del tratamiento, estimando el efecto causal en cada periodo relativo al momento de la intervención. Sin embargo, capturar correctamente dicha dinámica implica retos metodológicos que han sido abordados desde distintos ángulos en la literatura reciente.

Sun y Abraham (2021) fueron pioneros en señalar que los estimadores tradicionales de regresión con efectos fijos (TWFE) generan distorsiones cuando los efectos del tratamiento son heterogéneos entre cohortes. Su propuesta —actualmente implementada en paquetes como `sunab()` o `eventstudyinteract`— permite estimar efectos específicos por periodo, sin contaminación cruzada entre grupos tratados en momentos distintos. Esta aproximación facilita interpretar con precisión la trayectoria del tratamiento: por ejemplo, si su impacto se acumula, se disipa o persiste en el tiempo.

Sobre esta base, Roth et al. (2023) ofrecen una síntesis de buenas prácticas para el análisis de efectos dinámicos, destacando aspectos que van más allá del estimador en sí. En particular, recomiendan no depender exclusivamente de tests de pre-trends clásicos para validar el diseño, ya que una aparente violación podría reflejar genuina heterogeneidad temporal y no necesariamente invalidar la estrategia. Además, proponen usar bandas de confianza robustas a pequeñas desviaciones de las tendencias paralelas, como las desarrolladas por Rambachan y Roth (2023). Estos métodos de sensibilidad cuantifican cuán grandes deberían ser las diferencias en la evolución del grupo de control, en ausencia de tratamiento, para explicar los efectos estimados.

Otro aporte clave de esta línea de investigación es la inferencia ajustada por múltiples pruebas en contextos con numerosos *leads* y *lags*, así como el uso de intervalos de confianza ampliados que reflejan la incertidumbre sobre la duración o intensidad del efecto. Estas herramientas refuerzan la credibilidad empírica de los perfiles dinámicos estimados, especialmente cuando se utilizan en debates de política pública.

Aunque trabajos como de Chaisemartin y D’Haultfoeuille (2020) se centran en la descomposición del estimador TWFE, también aportan evidencia sobre cómo los pesos negativos y las interacciones complejas entre cohortes y momentos pueden sesgar la interpretación de los efectos dinámicos si no se emplean los estimadores adecuados.

En suma, esta literatura reciente no introduce estimadores radicalmente nuevos, sino que mejora la fiabilidad de los efectos dinámicos ya estimados. El foco está en la interpretación robusta, la validación empírica y la claridad comunicativa de los resultados, haciendo que los estudios DiD modernos sean metodológicamente más transparentes y creíbles.

Capítulo 2

Diseños DiD con tratamientos repetidos, asignación no simultánea y efectos de nivel sin ruptura de tendencia

Este capítulo aborda la identificación y estimación de efectos causales cuando una misma intervención puede aplicarse repetidamente sobre una unidad en momentos no simultáneos (adopción no simultánea), con un número de aplicaciones que puede variar entre unidades. Nos enfocamos en dos objetos: (i) el *efecto marginal* de una intervención adicional —el cambio asociado al paso de $h-1$ a h en el nivel acumulado H_{it} en $t=k$ — y (ii) el *efecto en un instante fijo* $t=k$, independientemente del nivel acumulado vigente.

A lo largo de este capítulo trabajaremos bajo el *supuesto de no ruptura de tendencia*: después de cada intervención, la trayectoria de y conserva el mismo *patrón de evolución* (tendencia o ritmo subyacente) que traía antes del evento. Puede existir o no un salto en el momento de la intervención (eso es precisamente lo que se medirá), pero *no* se admite un cambio del patrón de evolución a partir de dicho momento.

Seguimos la notación general establecida en la introducción. Trabajaremos bajo los Supuestos 1, 2 y 3.

A partir de aquí trabajamos con ATT puntuales mediante dos estrategias complementarias.

La primera estrategia (Sección 2.1) compara, en $t=k$, una unidad tratada m que pasa de $h-1$ a h con una unidad de comparación n que en k no cambia

y mantiene $H_{n,k-1} = H_{n,k} < h$. La comparación se implementa *entre tratadas* dentro de ventanas admisibles alrededor de k , promediando *en el tiempo* para cada par (m, n) :

$$\begin{aligned} \widehat{ATT}_{m,n}(k) &= \mathbb{E} \left[y_{m,\bar{t}} - y_{m,\underline{t}} \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right] \\ &\quad - \mathbb{E} \left[y_{n,\bar{t}} - y_{n,\underline{t}} \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right]. \end{aligned} \quad (2.0.1)$$

según la Def. 2.1.3.1, bajo tendencias paralelas por ventana (Sup. 2.2) y con identificación en la Prop. 2.2. De forma secundaria, se reporta como contraste el caso con nunca tratadas (caso particular con $\underline{h}=0$; cf. Def. 2.1.2.1).

La segunda estrategia (Sección 2.2) fija dos tiempos $t^- < k \leq t^+$ y compara el cambio de las unidades que pasan de $h-1$ a h en k con el de unidades cuyo H permanece constante y menor que h en la misma ventana. El efecto puntual por ventana, promediando *entre unidades*, es:

$$\begin{aligned} \widehat{ATT}^{t^-,t^+}(k, h) &= \mathbb{E}[y_{i,t^+} - y_{i,t^-} \mid i \in \mathcal{I}^{k,h}(t^-, t^+)] \\ &\quad - \mathbb{E}[y_{j,t^+} - y_{j,t^-} \mid j \in \mathcal{J}^{k,h}(t^-, t^+)], \end{aligned} \quad (2.0.2)$$

según la Def. 2.2.2.1, bajo tendencias paralelas por ventana (Sup. 2.5) y con identificación en la Prop. 2.4.

Cada estrategia produce estimadores por instante k (y por (k, h) en la segunda) que luego se agregan a un estimador global. Los pesos dan mayor relevancia a comparaciones más informativas (ventanas más largas, mejor balance y menor varianza), siguiendo la motivación de la descomposición de TWFE y desarrollos afines de la literatura (Callaway y Sant’Anna (2021), de Chaisemartin y D’Haultfoeuille (2020) y Goodman-Bacon (2021)).

La inferencia se realiza sobre los agregados por instante y sobre el efecto global, empleando procedimientos de remuestreo (p. ej., bootstrap) y criterios de validez comunes: exclusión de controles potencialmente afectados, evitación de siempre-tratadas como controles y verificación de estabilidad temporal en las ventanas consideradas.

Proponemos dos familias de chequeos teóricos para evaluar la plausibilidad de los supuestos del capítulo. (i) En la estrategia unidad–unidad, comparamos sistemáticamente los promedios por instante $\widehat{ATT}^\circ(k)$ (contrafactual puro; Def. 2.1.2.1) y $\widehat{ATT}^+(k)$ (entre tratadas; Def. 2.1.3.1). Su cercanía respalda la

plausibilidad de los supuestos; discrepancias sugieren diferencias estructurales o dinámicas acumulativas. (ii) En la estrategia tiempo–tiempo, implementamos *placebos por ventana*: uno previo $[t^{--}, t^-]$ y otro posterior $[t^+, t^{++}]$, con valor esperado nulo bajo el diseño (Ecs. (2.2.8) y (2.2.9)). Ambos conjuntos de chequeos son complementarios: el primero contrasta tipos de contrafactual dentro del mismo k ; el segundo verifica, localmente, comparabilidad previa y patrón de evolución inalterado posterior.

En las Secciones 2.1.6 y 2.2.7 mostramos, con el mismo conjunto de datos simulados, que ambas estrategias producen estimaciones muy próximas del efecto marginal bajo los supuestos del diseño, lo que respalda la coherencia interna del enfoque.

2.1. Comparación unidad–unidad (promedio temporal)

En un instante fijo $t=k$, emparejamos una unidad tratada m que pasa de $h-1$ a h con una unidad de comparación n que en k no cambia y mantiene $H_{n,k-1} = H_{n,k} < h$. En ventanas admisibles alrededor de k , el efecto puntual por par (m, n) es una DiD promediada *en el tiempo*; véanse la Def. 2.1.3.1 y la Prop. 2.2. Como contraste, se reporta el caso con nunca tratadas ($\underline{h}=0$; Def. 2.1.2.1).

2.1.1. Comparaciones válidas en diseños con múltiples intervenciones

El diseño basado en trayectorias con múltiples intervenciones acumulativas plantea el reto de construir comparaciones creíbles entre unidades con diferentes niveles de tratamiento. No toda comparación observable genera un estimador válido de efecto causal: se requiere que el grupo de control represente de manera fiel la trayectoria contrafactual que habría seguido el grupo tratado en ausencia de la intervención adicional.

Siguiendo la lógica de Goodman-Bacon (2021), una comparación es válida si satisface dos condiciones centrales: (i) el grupo de control no ha sido afectado por el tratamiento que se desea evaluar, y (ii) las trayectorias de ambos grupos habrían seguido tendencias paralelas en ausencia de dicha intervención. Cuando estas condiciones se violan —por ejemplo, al usar como grupo de comparación unidades que ya han sido tratadas en periodos previos— pueden surgir sesgos sustanciales, incluyendo el problema conocido como *pesos negativos*, que distorsiona la interpretación causal del estimador agregado.

Para minimizar estos riesgos, este trabajo se concentra en un conjunto reducido de comparaciones que cumplen de forma clara con los requisitos de identificación. En particular, se delimitan dos escenarios metodológicamente sólidos:

- Comparaciones entre unidades tratadas y unidades nunca tratadas, que actúan como contrafactual puro. Este escenario permite estimar el efecto de una intervención específica comparando con trayectorias completamente exentas de tratamiento, asumiendo que en ausencia del tratamiento adicional la tendencia se habría mantenido constante.

- Comparaciones entre unidades tratadas con diferente número acumulado de intervenciones en un mismo periodo calendario, es decir, pares (i, t) y (j, t) tales que $H_{it} > H_{jt}$. En este caso, la unidad con menor nivel acumulado H_{jt} actúa como grupo de control para evaluar el efecto marginal de una intervención adicional. Bajo la hipótesis de tendencia paralela posterior, este diseño permite identificar efectos causales acumulativos a partir de diferencias en la intensidad del tratamiento. Esta comparación se asemeja al escenario de “tratadas temprano vs. tratadas tarde” descrito en Goodman-Bacon (2021), pero adaptado a un contexto con múltiples intervenciones secuenciales. A diferencia de los diseños de adopción única, aquí se comparan niveles sucesivos del tratamiento acumulado, asumiendo que las trayectorias se mantienen paralelas tras cada nueva intervención.

Estos escenarios excluyen de manera deliberada las comparaciones entre unidades tratadas y siempre tratadas, así como aquellas que utilizan como grupo de control a unidades con mayor nivel acumulado de tratamiento. Aunque en este trabajo no se realiza una descomposición algebraica del estimador como en Goodman-Bacon (2021), los hallazgos de dicha literatura advierten que este tipo de comparaciones pueden inducir interpretaciones problemáticas —por ejemplo, a través de contrafactuales poco creíbles o ponderaciones no intuitivas. En línea con estas observaciones, se opta aquí por restringir el análisis a escenarios metodológicamente más sólidos, donde las comparaciones permiten una identificación clara del efecto marginal de nuevas intervenciones. A continuación, se presentan los estimadores correspondientes a cada uno de estos escenarios válidos.

Nota metodológica. El caso en que se utiliza como grupo de control a unidades nunca tratadas puede interpretarse como una situación particular del escenario general descrito, en el que el nivel acumulado de tratamiento del grupo de comparación es $H_{jt} = 0$. Esta observación refuerza la consistencia del diseño adoptado, ya que permite incluir naturalmente a las unidades nunca tratadas como contrafactuales válidos sin requerir una categoría analítica separada.

2.1.2. Análisis DiD entre Unidades Tratadas y Nunca Tratadas

Este escenario evalúa el efecto de una intervención sobre las unidades que la reciben en un periodo específico k , utilizando como grupo de comparación aquellas unidades que nunca fueron tratadas durante todo el análisis.

Para cada tiempo k , se define el conjunto \mathcal{I}_k como las unidades que experimentan una nueva intervención en $t = k$, y \mathcal{I}_0 como el conjunto de unidades que no reciben ninguna intervención en ningún momento del periodo de estudio.

$$\mathcal{I}_k = \{m \mid H_{m,k} - H_{m,k-1} = 1\} \quad (2.1.1)$$

$$\mathcal{I}_0 = \{n \mid H_{n,t} = 0 \ \forall t\} \quad (2.1.2)$$

El objetivo es estimar el efecto del tratamiento aplicado en el año k , contrastando los resultados de cada unidad tratada $m \in \mathcal{I}_k$ con los de una unidad nunca tratada $n \in \mathcal{I}_0$, durante un intervalo de tiempo razonable alrededor de la intervención.

Definimos el intervalo de comparación válido para cada par (m, n) como $[t_{m,n}^{\text{pre}}(k), t_{m,n}^{\text{pos}}(k)]$, donde:

$$t_{m,n}^{\text{pre}}(k) = \min \{\underline{t} \mid H_{m,k} - H_{m,\underline{t}} = 1\} \quad (2.1.3)$$

$$t_{m,n}^{\text{pos}}(k) = \max \{\bar{t} \mid H_{m,\bar{t}} - H_{m,k} = 0\} \quad (2.1.4)$$

2.1.2.1 Definición: Efecto del tratamiento con contrafactual puro

Para cada par (m, n) , donde $m \in \mathcal{I}_k$ representa una unidad tratada en el periodo k y $n \in \mathcal{I}_0$ una unidad nunca tratada, se define el efecto causal específico del tratamiento como:

$$ATT_{m,n}^{\circ}(k) = \mathbb{E} [Y_{m,\bar{t}}(h) - Y_{m,\bar{t}}(h-1) \mid k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k)] \quad (2.1.5)$$

Este efecto mide el cambio promedio en los resultados de la unidad tratada m al pasar de nivel de tratamiento $h-1$ a h , comparando los periodos antes y después del cambio, dentro de un intervalo válido centrado en k .

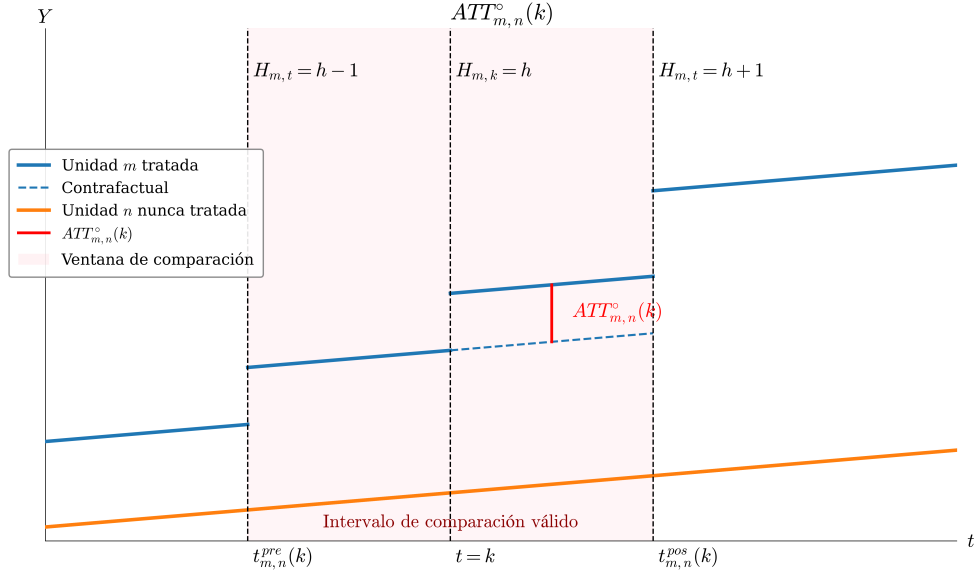


Figura 2.1: Estimación puntual del efecto del tratamiento para la unidad m en $t = k$: $ATT_{m,n}^{\circ}(k)$.

La Figura 2.1 ilustra la lógica del estimador $ATT_{m,n}^{\circ}(k)$, construido para evaluar el efecto del tratamiento sobre la unidad m en el periodo $t = k$, utilizando como grupo de comparación una unidad n que nunca ha sido tratada. La trayectoria azul representa los valores observados de $y_{m,t}$ en la unidad tratada, mientras que la línea naranja corresponde a la evolución de la unidad nunca tratada. La línea punteada azul representa la trayectoria contrafactual esperada para la unidad tratada, bajo el supuesto de tendencias paralelas. La diferencia vertical en rojo representa el valor puntual del efecto estimado. El sombreado indica el intervalo de comparación válido, desde $t_{m,n}^{pre}(k)$ hasta $t_{m,n}^{pos}(k)$, determinado por la ventana compartida entre ambas unidades.

Además del efecto en $t = k$, también pueden definirse efectos similares en otros momentos del historial de tratamiento de la unidad m , como \underline{k} y \bar{k} , en los cuales esta unidad también experimenta un cambio de nivel de tratamiento. Cada uno de estos efectos se estima en una ventana de comparación distinta, correspondiente a la posición temporal del evento y al nivel de tratamiento vigente en ese momento. En particular, $ATT_{m,n}^{\circ}(\underline{k})$ evalúa el cambio de $h-2$ a $h-1$, mientras que $ATT_{m,n}^{\circ}(\bar{k})$ captura el paso de h a $h+1$. Las Figuras A.1 y A.2, incluidas en el Anexo A.3, ilustran estos casos de forma análoga.

2.1 Supuesto: Tendencias paralelas con contrafactual puro

Para cada par (m, n) , donde $m \in \mathcal{I}_k$ es una unidad tratada en el periodo k y $n \in \mathcal{I}_0$ es una unidad nunca tratada, se asume que:

$$\begin{aligned} \mathbb{E} \left[Y_{m,\bar{t}}(h-1) - Y_{m,\underline{t}}(h-1) \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right] = \\ \mathbb{E} \left[Y_{n,\bar{t}}(0) - Y_{n,\underline{t}}(0) \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right] \end{aligned} \quad (2.1.6)$$

Este supuesto postula que, en ausencia del tratamiento de nivel h , la trayectoria de la unidad tratada m habría evolucionado en promedio igual que la trayectoria de una unidad nunca tratada n , al comparar los periodos antes y después del cambio. Es un supuesto exigente, pues implica que la intervención previa (nivel $h-1$) no altera la pendiente de la trayectoria con respecto a la de una unidad sin intervención alguna.

Este supuesto afirma que, de no haber recibido el tratamiento adicional en $t = k$, la unidad m —que ya había acumulado $h-1$ niveles— habría continuado su trayectoria como si nunca hubiese sido tratada. En otras palabras, las trayectorias de unidades tratadas parcialmente y de unidades nunca tratadas son paralelas en promedio, incluso después de varias rondas de intervención.

Se trata de un supuesto fuerte, ya que ignora cualquier posible efecto dinámico o acumulativo de las intervenciones previas sobre las tendencias. Sin este supuesto, no es posible identificar el efecto causal del cambio de tratamiento en $t = k$ usando unidades nunca tratadas como grupo de comparación. Bajo este supuesto, es posible interpretar el contraste entre $Y_{m,\bar{t}}$ y $Y_{m,\underline{t}}$ como una estimación válida del efecto $ATT_{m,n}^\circ(k)$.

Proposición 2.1 (Identificación del efecto $ATT_{m,n}^\circ(k)$ con contrafactual puro). *Bajo el Supuesto 2.1, el efecto del tratamiento con contrafactual puro definido en la Definición 2.1.2.1 puede ser identificado a partir de resultados observados como:*

$$\begin{aligned} \widehat{ATT}_{m,n}^\circ(k) = \mathbb{E} \left[y_{m,\bar{t}} - y_{m,\underline{t}} \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right] \\ - \mathbb{E} \left[y_{n,\bar{t}} - y_{n,\underline{t}} \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right] \end{aligned} \quad (2.1.7)$$

Una forma equivalente de expresar este efecto es:

$$\begin{aligned} \widehat{ATT}_{m,n}^{\circ}(k) &= \mathbb{E} [y_{m,\bar{t}} - y_{n,\bar{t}} \mid k \leq \bar{t} \leq t_{m,n}^{pos}(k)] \\ &\quad - \mathbb{E} [y_{m,\underline{t}} - y_{n,\underline{t}} \mid t_{m,n}^{pre}(k) \leq \underline{t} < k] \end{aligned} \quad (2.1.8)$$

Esta expresión resalta la comparación entre diferencias intergrupales en el periodo posterior y el periodo anterior al tratamiento.

La demostración de esta proposición se presenta en el Anexo A.1.

Nota 2.1. En esta expresión, la notación de esperanza condicional se utiliza como una forma compacta de representar promedios empíricos calculados exclusivamente a partir de resultados observados. En particular, el estimador $\widehat{ATT}_{m,n}^{\circ}(k)$ puede escribirse de manera explícita como:

$$\widehat{ATT}_{m,n}^{\circ}(k) = \frac{1}{|\mathcal{T}_{m,n}^{pos}(k)|} \sum_{\bar{t} \in \mathcal{T}_{m,n}^{pos}(k)} (y_{m,\bar{t}} - y_{n,\bar{t}}) - \frac{1}{|\mathcal{T}_{m,n}^{pre}(k)|} \sum_{\underline{t} \in \mathcal{T}_{m,n}^{pre}(k)} (y_{m,\underline{t}} - y_{n,\underline{t}}),$$

donde $\mathcal{T}_{m,n}^{pre}(k) = \{\underline{t} : t_{m,n}^{pre}(k) \leq \underline{t} < k\}$ y $\mathcal{T}_{m,n}^{pos}(k) = \{\bar{t} : k \leq \bar{t} \leq t_{m,n}^{pos}(k)\}$ denotan los conjuntos de periodos observados antes y después del tratamiento, respectivamente.

2.1.3. Análisis DiD entre Unidades con Diferente Nivel de Tratamiento Acumulado

Este escenario permite estimar el efecto marginal de recibir una intervención adicional en el periodo $t = k$, comparando unidades que en el periodo previo ($t = k-1$) tenían distintos niveles acumulados de tratamiento, todos inferiores a h . En particular, se considera como tratada a la unidad que tenía nivel $h-1$ en $t = k-1$ y que recibe una nueva intervención en $t = k$, alcanzando así el nivel h . Como grupo de control, se utilizan unidades que en $t = k-1$ tenían un nivel acumulado menor $\bar{h} < h$ y que en $t = k$ no reciben ninguna intervención adicional, permaneciendo en \bar{h} . Esto permite identificar el efecto marginal de avanzar desde un nivel de tratamiento menor hacia uno mayor, manteniendo constante el periodo de observación.

Se conserva la definición del conjunto de unidades tratadas presentada en la ecuación (2.1.1). Para cada unidad $m \in \mathcal{I}_k$, se definen dos posibles conjuntos de comparaciones, dependiendo del tipo de contrafactual que se desea construir:

- El conjunto restringido a unidades tratadas en algún otro momento del periodo de estudio (comparación entre tratadas):

$$\mathcal{I}_m^+(k) = \{n \mid H_{n,k-1} = H_{n,k} < H_{m,k}, n \notin \mathcal{I}_0\} \quad (2.1.9)$$

- El conjunto más general, que incluye también a las unidades nunca tratadas:

$$\mathcal{I}_m(k) = \{n \mid H_{n,k-1} = H_{n,k} < H_{m,k}\} \quad (2.1.10)$$

La primera definición permite construir comparaciones exclusivamente entre unidades que han sido tratadas en algún momento del periodo de estudio, facilitando estimaciones que evitan supuestos fuertes sobre la evolución de unidades estructuralmente distintas. La segunda amplía el conjunto de comparación para incluir también a las unidades nunca tratadas ($n \in \mathcal{I}_0$), lo cual puede ser útil al estimar efectos con contrafactual puro.

De manera similar al caso anterior, se define una ventana válida de comparación para cada par (m, n) , restringida a los periodos cercanos a la intervención en $t = k$. Esta ventana está delimitada por los extremos $t_{m,n}^{\text{pre}}(k)$ y $t_{m,n}^{\text{pos}}(k)$, definidos como:

$$t_{m,n}^{\text{pre}}(k) = \min \{ \underline{t} \mid H_{m,\underline{t}} = H_{m,k-1} = h - 1 \wedge \underline{h} = H_{n,\underline{t}} = H_{n,k} < H_{m,k} = h \} \quad (2.1.11)$$

$$t_{m,n}^{\text{pos}}(k) = \max \{ \bar{t} \mid H_{m,\bar{t}} = H_{m,k} = h \wedge \underline{h} = H_{n,\bar{t}} = H_{n,k} < H_{m,k} = h \} \quad (2.1.12)$$

2.1.3.1 Definición: Efecto marginal del tratamiento con comparación entre tratadas.

Para cada par (m, n) , donde $m \in \mathcal{I}_k$ y $n \in \mathcal{I}_m^+(k)$, se define el efecto causal marginal de la intervención como:

$$ATT_{m,n}^+(k) = \mathbb{E} [Y_{m,\bar{t}}(h) - Y_{m,\bar{t}}(h-1) \mid k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k)] \quad (2.1.13)$$

donde $h = H_{m,k}$ representa el nivel de tratamiento alcanzado por la unidad tratada m en el periodo $t = k$.

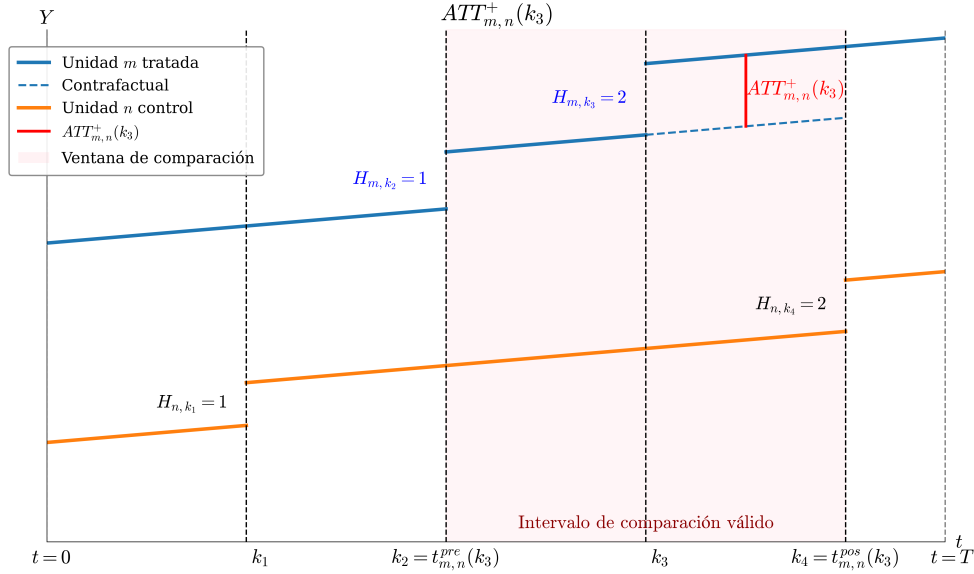


Figura 2.2: Estimación puntual del efecto marginal del tratamiento: $ATT_{m,n}^+(k_3)$.

La Figura 2.2 representa el estimador $ATT_{m,n}^+(k_3)$, construido a partir del salto observado en la unidad m cuando pasa de $H = 1$ a $H = 2$, usando como comparación a una unidad n que se mantiene en nivel $H = 1$ durante el mismo intervalo. El sombreado indica el rango temporal común que permite identificar el efecto.

El uso de n como grupo de control no es válido para estimar la primera intervención de m , ya que en ese momento ambas unidades comparten el mismo nivel acumulado. Este y otros casos se analizan con más detalle en el Anexo A.4.

Para que este efecto marginal pueda identificarse empíricamente a partir de datos observados, se requiere asumir que, en ausencia de la intervención adicional en $t = k$, las trayectorias de la unidad tratada m y de su correspondiente unidad de control n habrían evolucionado de forma paralela dentro de la ventana de comparación válida. Este supuesto de tendencias paralelas entre tratadas se presenta a continuación.

2.2 Supuesto: Tendencias paralelas entre tratadas.

Para cada par (m, n) , donde $m \in \mathcal{I}_k$ y $n \in \mathcal{I}_m^+(k)$, se cumple que:

$$\begin{aligned} \mathbb{E} \left[Y_{m,\bar{t}}(h-1) - Y_{m,\underline{t}}(h-1) \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right] = \\ \mathbb{E} \left[Y_{n,\bar{t}}(\underline{h}) - Y_{n,\underline{t}}(\underline{h}) \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right] \end{aligned} \quad (2.1.14)$$

donde $h = H_{m,k}$ y $\underline{h} < h$.

Nota. Este supuesto es más exigente que el utilizado en el caso de contrafactual puro, ya que requiere asumir tendencias paralelas entre unidades que pueden haber recibido distintos números de intervenciones previas. En efecto, el supuesto de contrafactual puro se obtiene como un caso particular cuando $\underline{h} = 0$.

Proposición 2.2 (Identificación del efecto $ATT_{m,n}(k)$ con comparación entre tratadas.). *Bajo el Supuesto 2.2, el efecto marginal definido en la Definición 2.1.3.1 puede ser identificado a partir de resultados observados como:*

$$\begin{aligned} \widehat{ATT}_{m,n}^+(k) = & \underbrace{\mathbb{E} \left[y_{m,\bar{t}} - y_{m,\underline{t}} \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right]}_{\text{cambio observado en la tratada } m} \\ & - \underbrace{\mathbb{E} \left[y_{n,\bar{t}} - y_{n,\underline{t}} \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right]}_{\text{cambio observado en la comparación } n} . \end{aligned} \quad (2.1.15)$$

Una forma equivalente de expresar este efecto es:

$$\begin{aligned} \widehat{ATT}_{m,n}^+(k) = & \underbrace{\mathbb{E} [y_{m,\bar{t}} - y_{n,\bar{t}} \mid k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k)]}_{\text{diferencia intergrupales en el periodo posterior } (\bar{t})} \\ & - \underbrace{\mathbb{E} [y_{m,\underline{t}} - y_{n,\underline{t}} \mid t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k]}_{\text{diferencia intergrupales en el periodo previo } (\underline{t})} . \end{aligned} \quad (2.1.16)$$

Esta expresión resalta la comparación entre diferencias intergrupales en el periodo posterior y el periodo anterior a la intervención adicional.

La demostración de esta proposición se presenta en el Anexo A.2.

Relación entre escenarios y validación empírica de supuestos

Los dos escenarios desarrollados anteriormente —comparación con unidades nunca tratadas y comparación con unidades tratadas de menor intensidad acumulada— comparten una estructura formal similar: en ambos casos se busca estimar el efecto marginal de una intervención adicional en el periodo $t = k$ sobre una unidad tratada m , utilizando como contrafactual otra unidad n que permanece en un nivel de tratamiento acumulado inferior.

La diferencia clave radica en la naturaleza del contrafactual: en el caso del *contrafactual puro*, la unidad de comparación $n \in \mathcal{I}_0$ nunca ha sido tratada, por lo que su nivel acumulado es $\underline{h} = 0$. En contraste, en el caso de *comparación entre tratadas*, la unidad $n \in \mathcal{I}_m^+(k)$ puede haber recibido una o más intervenciones previas, siempre que cumpla $\underline{h} < h$, donde h es el nivel alcanzado por m en $t = k$.

Desde esta perspectiva, el estimador con contrafactual puro $ATT_{m,n}^\circ(k)$ puede entenderse como un caso particular del estimador general entre tratadas $ATT_{m,n}^+(k)$, restringido al subconjunto de comparaciones en las que la unidad de control nunca ha sido intervenida. No obstante, se ha optado por mantener ambos enfoques de manera diferenciada —tanto en la formulación como en la implementación— por dos razones fundamentales.

Primero, los supuestos necesarios para la identificación causal difieren en su grado de exigencia. El Supuesto 2.1, requerido para identificar $ATT_{m,n}^\circ(k)$, exige que las unidades tratadas parcialmente sigan en promedio la misma trayectoria que las unidades nunca tratadas, en ausencia del tratamiento adicional. Por su parte, el Supuesto 2.2, asociado a $ATT_{m,n}^+(k)$, implica que distintas trayectorias de tratamiento acumulado —con historias de intervención distintas— evolucionarían paralelamente en ausencia del nuevo tratamiento. Esta última hipótesis es más restrictiva y sensible a dinámicas acumulativas no observadas.

Segundo, la separación metodológica permite utilizar ambos escenarios como mecanismo de validación cruzada de los supuestos. Si para una misma unidad tratada m se obtienen estimaciones similares de $ATT_{m,n}^\circ(k)$ y $ATT_{m,n}^+(k)$, ello sugiere que las hipótesis de evolución paralela podrían no estar siendo gravemente violadas. Aunque esta coincidencia no constituye una prueba definitiva de validez causal, sí aporta evidencia de consistencia interna en el diseño.

En consecuencia, se propone trabajar con ambos enfoques de manera

complementaria. Lejos de ser excluyentes, los escenarios de comparación con unidades nunca tratadas y con tratadas de menor intensidad permiten triangular la evidencia empírica, robustecer la interpretación de los efectos estimados y evaluar de forma más rigurosa la plausibilidad de los supuestos subyacentes.

2.1.4. Agregación de efectos y construcción de pesos

Para obtener estimadores agregados que sinteticen los efectos marginales calculados a lo largo del tiempo y entre distintas unidades, es necesario definir un sistema de pesos adecuado. La motivación para construir dichos pesos proviene de la descomposición clásica del estimador TWFE realizada por Goodman-Bacon (2021), donde se muestra que el estimador global puede expresarse como un promedio ponderado de múltiples comparaciones tipo 2×2 entre grupos tratados en diferentes momentos o frente a no tratados.

En dicha descomposición, cada peso asociado a una comparación se compone de cuatro elementos clave. El primero es el tamaño del subpanel, que representa la cantidad de observaciones que participan en la comparación específica; si una comparación involucra pocos datos, su contribución al estimador global será baja. El segundo elemento es la proporción relativa entre grupos, es decir, qué tan balanceado está el número de unidades en los dos grupos comparados (por ejemplo, tratados temprano y tratados tarde); comparaciones muy desbalanceadas tienen menor peso. El tercer componente es la varianza temporal del tratamiento, que refleja cuánta variación efectiva hay en el estado de tratamiento dentro del subconjunto comparado; si todos los grupos están igualmente tratados, el peso será prácticamente nulo. Por último, está la duración efectiva de la ventana de comparación, que aunque no se expresa de forma aislada en una fórmula, se encuentra implícita en las expresiones relacionadas con la duración del periodo pre o post tratamiento utilizado; comparaciones con ventanas muy cortas aportan menos información útil.

La estructura cuadrática de algunos de estos elementos —como el tamaño del subpanel— responde al principio estadístico de que la precisión de un estimador aumenta con el tamaño de la muestra, y lo hace de forma cuadrática bajo ciertos supuestos de varianza constante. Esta lógica es clave para entender por qué algunas comparaciones pesan más que otras, incluso si sus efectos estimados son similares en magnitud.

A partir de estos principios, en esta tesis se propone una estrategia de

agregación basada en pesos que, sin replicar exactamente la estructura TWFE, conserva la lógica esencial de ponderar por tamaño, duración y variabilidad. La siguiente sección detalla su formulación específica en dos niveles: primero para sintetizar efectos entre unidades con un mismo calendario de tratamiento, y luego para obtener un estimador agregado final.

Agregación de comparaciones por instante k

En un primer nivel, se busca agregar las distintas comparaciones entre unidades tratadas en un mismo instante k , a fin de obtener un estimador promedio $\widehat{ATT}(k)$. Cada comparación (m, n) recibe un peso proporcional a la información efectiva que aporta, considerando tanto la extensión temporal de la ventana de comparación como la variación del tratamiento en dicha ventana.

El peso base asignado a cada comparación se define como:

$$\omega_{m,n}(k) = (t_{m,n}^{pos}(k) - k + 1) \cdot (k - t_{m,n}^{pre}(k))$$

Este producto captura dos elementos clave: (1) la duración posterior al tratamiento dentro de la ventana móvil, y (2) la duración previa al tratamiento. La estructura cuadrática implícita favorece comparaciones con ventanas amplias y balanceadas, penalizando aquellas que sólo cubren pocos periodos o que están fuertemente sesgadas hacia un lado temporal.

La expresión anterior surge de multiplicar la longitud total de la ventana por la varianza temporal del estado de tratamiento, tal como se detalla en el Anexo A.5. Allí se demuestra que esta forma de ponderación sigue la misma lógica conceptual de los pesos en estimadores TWFE, garantizando mayor influencia a comparaciones más informativas.

Para fines de agregación, se utiliza la versión normalizada de estos pesos:

$$\widehat{\omega}_{m,n}(k) = \frac{\omega_{m,n}(k)}{\sum_{(m,n)} \omega_{m,n}(k)}$$

A partir de los efectos individuales válidos $\widehat{ATT}_{m,n}^{\circ}(k)$ o $\widehat{ATT}_{m,n}^{+}(k)$, calculados para cada par de unidades según el sentido de comparación permitido, se define el efecto promedio por instante como:

$$\widehat{ATT}(k) = \sum_{(m,n)} \widehat{\omega}_{m,n}(k) \cdot \widehat{ATT}_{m,n}(k)$$

donde $\widehat{ATT}_{m,n}(k)$ representa indistintamente $\widehat{ATT}_{m,n}^\circ(k)$ o $\widehat{ATT}_{m,n}^+(k)$, dependiendo de cuál comparación sea válida para ese par.

Nota 2.2. También es posible calcular por separado los efectos promedio obtenidos mediante comparaciones de tipo \widehat{ATT}° y de tipo \widehat{ATT}^+ . Para ello, se definen pesos normalizados independientes dentro de cada subconjunto de comparaciones válidas.

En cada caso, los pesos se normalizan sobre los pares válidos (m, n) tales que $m \in \mathcal{I}_k$ y $n \in \mathcal{I}_0$ o $n \in \mathcal{I}_m^+(k)$, para los cuales se puede calcular $\widehat{ATT}_{m,n}^\circ(k)$ o $\widehat{ATT}_{m,n}^+(k)$ respectivamente:

$$\widehat{\omega}_{m,n}^\circ(k) = \frac{\omega_{m,n}(k)}{\sum \omega_{m,n}(k)}, \quad \widehat{\omega}_{m,n}^+(k) = \frac{\omega_{m,n}(k)}{\sum \omega_{m,n}(k)}$$

A partir de ellos, los promedios se definen como:

$$\widehat{ATT}^\circ(k) = \sum \widehat{\omega}_{m,n}^\circ(k) \cdot \widehat{ATT}_{m,n}^\circ(k), \quad \widehat{ATT}^+(k) = \sum \widehat{\omega}_{m,n}^+(k) \cdot \widehat{ATT}_{m,n}^+(k)$$

Estos dos promedios no se emplean directamente para inferencia, pero su comparación puede ser útil como mecanismo de validación interna del diseño. Una alta similitud entre ambos sugiere (aunque no garantiza) que se cumplen los supuestos de tendencias paralelas entre unidades tratadas. Por el contrario, discrepancias marcadas podrían evidenciar problemas de validez en las comparaciones.

Construcción del estimador global a partir de efectos por instante k

En un segundo nivel de agregación, se busca combinar los efectos promedio por instante $\widehat{ATT}(k)$ en un estimador global, tomando en cuenta el poder informativo relativo de cada cohorte de tratamiento. Para ello, se asigna un peso específico a cada instante k , proporcional al número total de comparaciones efectivas realizadas en dicho instante.

El peso base asignado se define como:

$$\Omega(k) = \sum_{m \in \mathcal{I}_k} |\mathcal{I}_m(k)|$$

Esta expresión equivale al número total de comparaciones válidas realizadas para estimar $\widehat{ATT}(k)$, sumando las unidades que actúan como controles en

cada comparación individual. Por tanto, $\Omega(k)$ refleja el tamaño efectivo del subpanel utilizado en la estimación del efecto para el instante k .

Para fines de agregación global, se utiliza la versión normalizada de estos pesos:

$$\hat{\Omega}(k) = \frac{\Omega(k)}{\sum_k \Omega(k)}$$

El estimador final del efecto promedio del tratamiento acumulativo se construye como un promedio ponderado de los efectos por instante:

$$\widehat{ATT} = \sum_k \hat{\Omega}(k) \cdot \widehat{ATT}(k)$$

Este esquema de ponderación permite preservar la lógica de comparación cruzada entre unidades con distinto historial de tratamiento, respetando las diferencias en tamaño y calidad informativa de cada subpanel asociado a los distintos instantes k .

Una discusión más detallada sobre la justificación de este peso puede encontrarse en el Anexo A.6, donde se presenta una reinterpretación de los pesos propuestos por Goodman-Bacon (2021) bajo una lógica de doble agregación.

2.1.5. Análisis de inferencia

El análisis de inferencia constituye un componente esencial para interpretar adecuadamente los efectos estimados a lo largo del estudio. Su objetivo principal es evaluar si los resultados obtenidos reflejan diferencias sistemáticas atribuibles a las intervenciones acumuladas, o si podrían explicarse por variabilidad aleatoria en los datos observados. Para ello, se cuantifica la incertidumbre asociada a cada estimación mediante errores estándar, se construyen intervalos de confianza, y se identifican los efectos estadísticamente significativos bajo distintos niveles de significancia.

En el presente trabajo, los estimadores marginales $\widehat{ATT}_{m,n}(k)$ se construyen a partir de comparaciones puntuales entre pares de unidades tratadas y de control, utilizando ventanas móviles centradas en un instante de intervención $t = k$. Aunque cada uno de estos estimadores captura un efecto marginal válido en su contexto, no son adecuados como objeto individual de inferencia estadística. Esto se debe a que cada comparación involucra una sola pareja de unidades y un número limitado de periodos, lo cual impide

justificar supuestos asintóticos o calcular errores estándar robustos. Por tanto, los efectos $\widehat{ATT}_{m,n}(k)$ se entienden como componentes informativos que, al ser agregados, permiten estimar efectos promedio de interés sustantivo.

El análisis de inferencia se realiza sobre los estimadores agregados por instante k , definidos como:

$$\widehat{ATT}(k) = \sum_{(m,n)} \hat{\omega}_{m,n}(k) \cdot \widehat{ATT}_{m,n}(k)$$

donde $\hat{\omega}_{m,n}(k)$ son los pesos normalizados que reflejan la duración y el balance temporal de cada ventana de comparación. Bajo condiciones de representatividad y estabilidad, este estimador puede interpretarse como una medida del efecto medio de recibir una intervención adicional en el instante k , sobre la base de múltiples trayectorias observadas.

A partir de estos valores agregados, se estiman errores estándar y se construyen intervalos de confianza simétricos al 90 %, 95 % y 98 %. Además, se asigna a cada estimación un código de significancia convencional (uno, dos o tres asteriscos) en función del valor- p obtenido. Esta codificación permite identificar visualmente qué efectos son estadísticamente significativos a los niveles habituales, facilitando la interpretación de los hallazgos.

A continuación, se enuncian los supuestos que permiten justificar formalmente la validez del análisis de inferencia para $\widehat{ATT}(k)$, así como el uso de técnicas como el bootstrap para estimar su error estándar y construir intervalos de confianza.

2.3 Supuesto: Representatividad de las comparaciones marginales.

El conjunto de pares (m, n) empleados en la construcción de $\widehat{ATT}(k)$ constituye una muestra representativa del universo de posibles comparaciones válidas entre unidades que avanzan de nivel en el instante k y unidades que permanecen en niveles inferiores. Esto garantiza que el estimador agregado puede interpretarse como una aproximación del efecto promedio de la intervención adicional.

2.4 Supuesto: Estabilidad del efecto dentro de la ventana de comparación.

Para cada par (m, n) , se asume que el efecto de avanzar del nivel $h-1$ al nivel h en el instante k permanece constante a lo largo de la ventana definida

entre $t_{m,n}^{\text{pre}}(k)$ y $t_{m,n}^{\text{pos}}(k)$. Esto permite interpretar la diferencia observada como un contraste válido entre niveles de tratamiento acumulado.

Teorema 2.1 (Comportamiento asintótico del estimador $\widehat{ATT}(k)$). *Bajo los Supuestos 3, 2.3 y 2.4, el estimador agregado por instante k , definido como:*

$$\widehat{ATT}(k) = \sum_{(m,n)} \widehat{\omega}_{m,n}(k) \cdot \widehat{ATT}_{m,n}(k),$$

es consistente y asintóticamente normal. En particular, se cumple que:

$$\sqrt{N_k} (\widehat{ATT}(k) - ATT(k)) \xrightarrow{d} \mathcal{N}(0, \sigma_k^2),$$

donde N_k es el número total de comparaciones (m, n) utilizadas en el instante k , y la varianza asintótica está dada por:

$$\sigma_k^2 = \sum_{(m,n)} \widehat{\omega}_{m,n}^2(k) \cdot \text{Var} \left(\widehat{ATT}_{m,n}(k) \right).$$

Bajo esta aproximación asintótica, el intervalo de confianza de nivel $1 - \alpha$ para $ATT(k)$ se expresa como:

$$CI_{1-\alpha}^{(k)} = \left[\widehat{ATT}(k) \pm z_{1-\alpha/2} \cdot \widehat{\sigma}_k \right],$$

donde $\widehat{\sigma}_k$ denota una estimación consistente de σ_k y $z_{1-\alpha/2}$ es el cuantil de orden $1 - \alpha/2$ de la distribución normal estándar.

Aunque esta expresión analítica permite caracterizar la varianza de forma teórica, en la práctica resulta complejo estimar de manera precisa las varianzas individuales de cada $\widehat{ATT}_{m,n}(k)$, debido al tamaño reducido de las ventanas y a la heterogeneidad entre pares. Por esta razón, se emplea un método de remuestreo bootstrap que aproxima de forma empírica la distribución de $\widehat{ATT}(k)$, proporcionando estimaciones robustas del error estándar e intervalos de confianza asociados.

Demostración. Véase el Anexo A.7. El argumento aplica un CLT por clúster (independencia entre unidades), con representación asintóticamente lineal y consistencia del bootstrap por unidad. \square

Teorema 2.2 (Comportamiento asintótico del estimador global \widehat{ATT}). *Como se definió previamente, el estimador global del efecto promedio del tratamiento*

acumulativo se construye como un promedio ponderado de los efectos por instante:

$$\widehat{ATT} = \sum_k \widehat{\Omega}(k) \cdot \widehat{ATT}(k),$$

donde $\widehat{\Omega}(k)$ son pesos fijos y normalizados que reflejan la cantidad de comparaciones informativas disponibles en cada instante k .

Bajo los mismos supuestos del Teorema 2.1, y asumiendo además que los pesos $\widehat{\Omega}(k)$ son determinísticos o independientes del muestreo de los efectos individuales $\widehat{ATT}_{m,n}(k)$, se cumple que:

$$\sqrt{N} \left(\widehat{ATT} - ATT \right) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

donde $N = \sum_k N_k$ es el número total de comparaciones utilizadas en la construcción del estimador global, y la varianza asintótica está dada por:

$$\sigma^2 = \sum_k \widehat{\Omega}^2(k) \cdot \sigma_k^2,$$

siendo σ_k^2 la varianza asintótica del estimador $\widehat{ATT}(k)$ definida en el Teorema 2.1.

A partir de esta caracterización, es posible construir un intervalo de confianza de nivel $1 - \alpha$ para el efecto global, dado por:

$$CI_{1-\alpha}^{(global)} = \left[\widehat{ATT} \pm z_{1-\alpha/2} \cdot \widehat{\sigma} \right],$$

donde $\widehat{\sigma}$ es una estimación empírica de σ , obtenida ya sea a partir de la fórmula anterior (si se conocen o estiman las $\widehat{\sigma}_k$), o bien mediante procedimientos de remuestreo bootstrap. En este trabajo, se emplea la segunda opción como estrategia robusta para capturar la variabilidad agregada y construir inferencias válidas sobre el efecto promedio del tratamiento acumulativo.

Demostración. Véase el Anexo A.7. Se usa combinación lineal de los resultados por k (Crámer–Wold/Slutsky) y bootstrap por unidad para la varianza y los intervalos. \square

Nota 2.3 (Agregación alternativa por nivel de tratamiento h). Hasta ahora, la agregación de efectos individuales $\widehat{ATT}_{m,n}(k)$ se ha realizado por instante k , interpretando $\widehat{ATT}(k)$ como el efecto promedio del tratamiento sobre las

unidades que lo reciben en ese instante específico. Sin embargo, en ciertos contextos puede resultar más informativo estimar el efecto promedio asociado al nivel acumulado del tratamiento, es decir, al número de veces que una unidad ha sido intervenida hasta un punto dado.

Bajo este enfoque, se puede definir un nuevo estimador promedio agrupando todas las comparaciones (m, n) que permiten estimar el efecto asociado al nivel h , independientemente del año en que ocurrió la comparación.

Aunque esta agregación reorganiza los efectos por nivel de intervención en lugar de por año calendario, la lógica de los pesos asignados a cada comparación sigue siendo válida. Esto se debe a que $\omega_{m,n}(k)$ depende exclusivamente de la estructura temporal de la ventana de comparación entre m y n , capturando su capacidad informativa sin importar en qué nivel h se clasifique el tratamiento. En consecuencia, no es necesario redefinir una nueva función de ponderación base, sino únicamente reescalar los pesos dentro de cada grupo relevante.

Denotando este nuevo estimador como $\widehat{ATT}^{[h]}$, se tiene:

$$\widehat{ATT}^{[h]} = \sum_{(m,n): H_{m,k}=h} \hat{\omega}_{m,n}^{[h]}(k) \cdot \widehat{ATT}_{m,n}(k),$$

donde:

- $H_{m,k}$ representa el nivel acumulado del tratamiento para la unidad tratada m en el instante k ;
- $\widehat{ATT}_{m,n}(k)$ es el estimador individual del efecto en la comparación válida entre m y n construida en k ;
- $\hat{\omega}_{m,n}^{[h]}(k) = \frac{\omega_{m,n}(k)}{\sum_{(m',n'): H_{m',k'}=h} \omega_{m',n'}(k')}$ es el peso normalizado sobre el subconjunto de comparaciones que contribuyen al nivel h .

Esta agregación permite responder a preguntas del tipo: “¿cuál es el efecto promedio de haber recibido la primera intervención?”, sin importar en qué año se haya dado ni en qué cohorte temporal fue estimado.

La agregación alternativa por nivel de tratamiento h permite definir un estimador global que resuma todos los efectos promedio según el número de veces que una unidad ha sido intervenida, sin importar el año en que ocurrió la comparación. A diferencia del enfoque basado en instantes k , este estimador captura el efecto promedio asociado a cada nivel acumulado h ,

y puede interpretarse como una versión estructurada del impacto según la intensidad de la intervención.

Para construir este estimador, se define el conjunto total de comparaciones válidas que contribuyen al cálculo de $\widehat{ATT}^{[h]}$ como:

$$\mathcal{C}^{[h]} = \{(m, n, k) \in \mathcal{C} : H_{m,k} = h\},$$

donde \mathcal{C} representa el conjunto total de comparaciones posibles entre unidades tratadas y de referencia que cumplen los requisitos establecidos en el diseño (ventanas válidas, progresión del tratamiento, etc.).

El peso base asignado a cada nivel se define como la cantidad de comparaciones que lo componen:

$$\Omega^{[h]} = |\mathcal{C}^{[h]}|,$$

y su versión normalizada es:

$$\hat{\Omega}^{[h]} = \frac{\Omega^{[h]}}{\sum_{h'} \Omega^{[h']}}.$$

Con estos elementos, se construye el estimador global como promedio ponderado:

$$\widehat{ATT}_{global}^{[h]} = \sum_h \hat{\Omega}^{[h]} \cdot \widehat{ATT}^{[h]}.$$

Este estimador permite evaluar el efecto promedio del tratamiento acumulativo según su nivel de intensidad, y puede ser útil para identificar patrones de respuesta no lineales o umbrales en la efectividad de la política analizada.

Nota 2.4 (Comparación entre estimadores globales por instante y por nivel). En esta investigación se han propuesto dos formas alternativas de agregación global de los efectos estimados $\widehat{ATT}_{m,n}(k)$. La primera, denominada \widehat{ATT} , se obtiene como un promedio ponderado por instante k , y busca capturar el efecto promedio de recibir el tratamiento en un determinado momento. La segunda, denotada como $\widehat{ATT}^{[1]}$, se construye agrupando los efectos por nivel acumulado de tratamiento h , con el objetivo de estimar el impacto asociado a cada instancia sucesiva de intervención, sin importar el año en que tuvo lugar.

Ambos estimadores utilizan los mismos efectos individuales $\widehat{ATT}_{m,n}(k)$ y los mismos pesos normalizados $\hat{\omega}_{m,n}(k)$, diferenciándose únicamente en la

forma de agregación. Como consecuencia, si la distribución de los niveles h a lo largo del tiempo es más o menos estable, o si el efecto del tratamiento no varía mucho entre cohortes ni entre niveles, entonces los valores de \widehat{ATT} y $\widehat{ATT}^{[1]}$ tenderán a ser similares.

Sin embargo, diferencias sustantivas entre estos dos estimadores pueden ser indicio de variaciones importantes en el comportamiento del tratamiento a lo largo del tiempo o con el número de aplicaciones. Por ejemplo, podrían reflejar:

- variaciones en la efectividad del tratamiento según el momento en que se recibe;
- efectos crecientes, decrecientes o no lineales conforme se acumulan intervenciones;
- diferencias estructurales entre las cohortes que entran en tratamiento en distintos años;
- o incluso cambios en la forma como se implementa el tratamiento a lo largo del tiempo.

Por tanto, reportar ambos estimadores globales no solo enriquece la interpretación, sino que también permite identificar posibles diferencias sustantivas en el impacto del tratamiento según el momento o el número de veces que ha sido recibido. Si los valores resultan similares, ello refuerza la coherencia general del efecto estimado. Pero si difieren de manera importante, será necesario examinar en cada aplicación empírica cuáles son las fuentes plausibles de esa divergencia: ya sea por cohortes, intensidades, cambios en la implementación o características de las unidades tratadas.

2.1.6. Aplicación a datos simulados

Aplicamos el método sobre un panel simulado de 60 unidades y 20 periodos. La Figura 2.3 ilustra la evolución de y para las primeras ocho unidades; no se observan rupturas de tendencia, en consonancia con los supuestos del diseño. Este subpanel corresponde al escenario *sin ruptura* y *sin dosis*; la simulación completa del panel y el procedimiento de generación se describen en el Anexo B, Sección B.1 (véanse en particular §B.1.1 y §B.1.3).

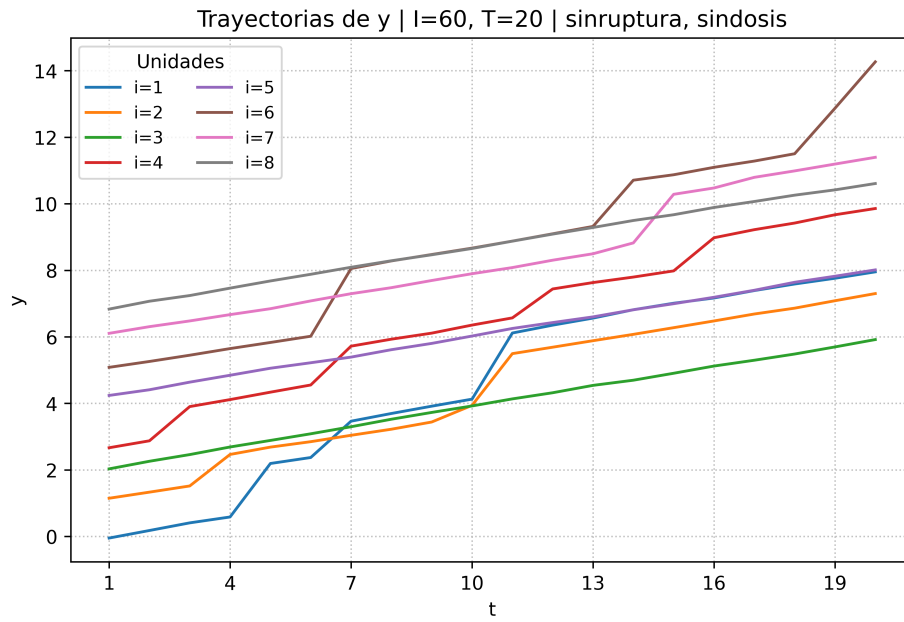


Figura 2.3: Evolución de y_t para las primeras ocho unidades.

Resultados por instante k

Cuadro 2.1: Inferencia por instante k

k	$\widehat{ATT}(k)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	w_k (%)	sig
3	0.7748	0.1487	0.4958	1.0795	9.13	***
4	1.0425	0.1496	0.7389	1.2928	4.55	***
5	0.6905	0.1662	0.3975	1.0522	6.88	***
6	0.4516	0.1163	0.2882	0.7745	4.69	***
7	1.5927	0.1903	1.1305	1.8668	7.13	***
8	1.5187	0.2984	0.8463	2.0048	5.74	***
9	0.9830	0.1766	0.6117	1.3141	7.86	***
10	0.7954	0.1047	0.6081	0.9904	5.79	***
11	1.3003	0.1852	0.8904	1.6103	7.45	***
12	0.8116	0.1753	0.4787	1.1436	5.76	***
13	1.2400	0.1968	0.8223	1.5878	5.68	***
14	1.2688	0.2102	0.8837	1.6270	3.39	***
15	0.9216	0.1694	0.6491	1.3113	3.95	***
16	0.9765	0.2233	0.5430	1.4071	5.20	***
17	1.0011	0.1643	0.6272	1.3128	4.28	***
18	0.8879	0.1004	0.6612	1.0524	5.06	***
19	1.1196	0.3503	0.5028	1.8031	4.34	***
20	0.7873	0.1258	0.5153	1.0305	3.11	***

Estimación global ponderada por $\hat{\Omega}(k)$: $\widehat{ATT}_{global} = 1.0161$, $SE_{boot} = 0.0444$, IC95 % = [0.9186, 1.0967] ***.

Resultados por nivel de intensidad h

Cuadro 2.2: Inferencia por nivel h

h	$\widehat{ATT}^{[h]}$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	w_h (%)	sig
1	1.0986	0.1193	0.8614	1.3197	24.92	***
2	1.0435	0.0939	0.8572	1.2216	23.64	***
3	1.1191	0.1085	0.9070	1.3201	21.53	***
4	0.7451	0.1358	0.5053	1.0271	16.74	***
5	0.9800	0.1997	0.6050	1.3622	10.02	***
6	1.1321	0.1824	0.8601	1.5062	3.14	***

Estimación global ponderada por $\hat{\Omega}^{[h]}$: $\widehat{ATT}_{\text{global}}^{[h]} = 1.0199$, $SE_{\text{boot}} = 0.0525$, $IC95\% = [0.9168, 1.1204]$ ***.

Comparación global y notas de validación

Cuadro 2.3: Comparación de estimadores globales

Estimador	Valor
$\widehat{ATT}_{\text{global}}$ (por k)	1.016144
$\widehat{ATT}_{\text{global}}^{[h]}$ (por h)	1.019954
DPS_{global} (%)	0.3742

- Nota 2.5.** (1) La comparación entre los promedios por instante, $\widehat{ATT}^{\circ}(k)$ (contrafactual puro) y $\widehat{ATT}^{+}(k)$ (entre tratadas), es consistente en orden de magnitud y, salvo algunos k , cercana; véase la Tabla B.4 del Anexo B.2. Esta concordancia respalda que, en estos datos, el supuesto de ausencia de rupturas de tendencia es razonable para el diseño aplicado.
- (2) Los dos estimadores globales —agregado por k y agregado por nivel h — son muy próximos ($DPS_{\text{global}}=0.3742\%$), lo que refuerza la coherencia entre ambos métodos de agregación y sugiere estabilidad del efecto promedio; véase la Tabla B.7 del Anexo B.2.

2.2. Comparación tiempo–tiempo (promedio entre unidades)

Esta estrategia estima el mismo efecto marginal que la anterior, pero fijando *dos tiempos* $t^- < k \leq t^+$ alrededor de k : uno previo y uno posterior (no se usa toda la franja intermedia). Se comparan los cambios de las unidades que en k pasan de $h-1$ a h con los de unidades cuyo H permanece constante y menor que h en $[t^-, t^+]$. El estimador puntual por ventana promedia *entre unidades* y se define en (2.0.2); su identificación sigue de la Def. 2.2.2.1 y la Prop. 2.4.

2.2.1. Comparaciones válidas centradas en cortes temporales

El enfoque basado en cortes temporales busca identificar efectos causales de intervenciones acumulativas a partir de saltos discretos observados en momentos específicos del tiempo. A diferencia de los diseños clásicos con tratamiento único o adopción escalonada, aquí una misma unidad puede recibir múltiples tratamientos en distintos periodos, lo cual obliga a redefinir cuidadosamente qué constituye una comparación válida.

Siguiendo la lógica de de Chaisemartin y D’Haultfoeuille (2020) y Callaway y Sant’Anna (2021), una comparación es válida si (i) el grupo de control no experimenta ningún cambio en su nivel acumulado de tratamiento durante el intervalo de análisis, y (ii) en ausencia de intervención, las trayectorias de ambos grupos habrían seguido tendencias paralelas. La validez de esta estrategia descansa, por tanto, en la correcta identificación de pares de tiempos y subconjuntos de unidades que cumplan dichas condiciones.

Este trabajo se concentra exclusivamente en comparaciones que cumplen rigurosamente con estos requisitos. En particular, se considera como tratadas a las unidades que reciben una intervención adicional en un periodo específico k , y como controles a aquellas unidades que no han sido tratadas en k , que mantienen constante su nivel de tratamiento entre dos tiempos (t^-, t^+) , y cuyo nivel acumulado es estrictamente menor que el alcanzado por las tratadas. Esta estructura garantiza que el cambio en el resultado observado entre t^- y t^+ pueda atribuirse al salto en k , bajo el supuesto de tendencia paralela posterior.

Este enfoque excluye de forma explícita comparaciones con unidades que

ya han sido tratadas antes del periodo k , así como aquellas con niveles de tratamiento iguales o superiores al de las unidades tratadas. La lógica detrás de esta restricción es evitar contrafactuales poco creíbles y minimizar sesgos derivados de comparaciones entre trayectorias estructuralmente distintas. A diferencia de otros enfoques que descomponen efectos agregados a lo largo del tiempo, aquí se busca identificar efectos puntuales bien definidos asociados a cada salto específico en el tratamiento acumulado. A continuación, se presentan los conjuntos formales de tratamiento y control, junto con el estimador propuesto para capturar estos efectos marginales.

2.2.2. Definición de los conjuntos relevantes para cada configuración (k, h)

Para cada periodo k y para cada nivel acumulado h alcanzado mediante una nueva intervención en dicho periodo, se buscan comparaciones válidas entre unidades tratadas y unidades de control con el fin de estimar el efecto causal asociado a ese salto puntual.

El conjunto de unidades *tratadas* en el periodo k y que alcanzan el nivel acumulado h se define como:

$$\mathcal{I}^{k,h} = \{i \mid H_{i,k} - H_{i,k-1} = 1, \quad H_{i,k} = h\}. \quad (2.2.1)$$

Las comparaciones admisibles deberán cumplir tres condiciones fundamentales sobre las unidades *de control* y la ventana temporal de contraste:

- No haber recibido tratamiento en el periodo k (esto es, no saltar en k),
- Mantener constante su nivel acumulado de tratamiento entre los tiempos de comparación,
- Tener un nivel de tratamiento estrictamente menor que h .

Para organizar la comparación en torno a k , comenzamos por definir los *tiempos previos admisibles* para (k, h) :

$$\mathcal{T}^-(k, h) = \{t^- < k : \exists i \in \mathcal{I}^{k,h} \text{ con } H_{i,t^-} = h - 1, \quad \exists j \text{ tal que } H_{j,t^-} = H_{j,k} < h\}.$$

en cada t^- admisible, al menos una tratada está en $h - 1$ y existe algún control que se mantiene en un nivel $h' < h$ hasta k .

Para cada $t^- \in \mathcal{T}^-(k, h)$, definimos los *tiempos posteriores admisibles* como

$$\mathcal{T}^+(k, h \mid t^-) = \left\{ t^+ \geq k \mid \begin{array}{l} \exists j : H_{j,t^-} = H_{j,t^+} < h, \\ \exists i \in \mathcal{I}^{k,h} : H_{i,t^+} - H_{i,t^-} = 1 \end{array} \right\}.$$

Nota 2.6. *La definición condicional de $\mathcal{T}^+(k, h \mid t^-)$ no es arbitraria: el conjunto de controles válidos depende de la ventana completa $[t^-, t^+]$. Al extender t^+ , algunos controles dejan de ser elegibles por presentar saltos antes de t^+ ; y al mover t^- hacia adelante, pueden incorporarse nuevos controles que cumplen estabilidad en la ventana. Por ello, los tiempos posteriores admisibles deben fijarse dado cada t^- . Un ejemplo ilustrativo se desarrolla en el Anexo A.8.*

Para cada pareja válida (t^-, t^+) alrededor de (k, h) , definimos los conjuntos de *controles efectivos* y *tratadas efectivas* sobre la ventana $[t^-, t^+]$ como sigue:

$$\mathcal{J}^{k,h}(t^-, t^+) = \{ j \mid H_{j,t^-} = H_{j,t^+} < h \}. \quad (2.2.2)$$

$$\mathcal{I}^{k,h}(t^-, t^+) = \left\{ i \in \mathcal{I}^{k,h} \mid H_{i,t^-} = h-1, \quad H_{i,t^+} = h \right\}. \quad (2.2.3)$$

Nota 2.7. *De la propia definición de los conjuntos de controles efectivos y tratadas efectivas se obtienen las siguientes propiedades de monotonicidad.*

Proposición 2.3. *Tomando (k, h) fijos y $t^- \in \mathcal{T}^-(k, h)$, si $t_1^+, t_2^+ \in \mathcal{T}^+(k, h \mid t^-)$ con $t_1^+ \leq t_2^+$, entonces*

$$(1) \quad \mathcal{J}^{k,h}(t^-, t_2^+) \subseteq \mathcal{J}^{k,h}(t^-, t_1^+).$$

$$(2) \quad \mathcal{I}^{k,h}(t^-, t_2^+) \subseteq \mathcal{I}^{k,h}(t^-, t_1^+).$$

Asimismo, fijemos t^+ tal que $t^+ \in \mathcal{T}^+(k, h \mid t^-)$ sea admisible para los tiempos previos considerados. Si $t_1^-, t_2^- \in \mathcal{T}^-(k, h)$ con $t_1^- < t_2^- < k$ y $t^+ \in \mathcal{T}^+(k, h \mid t_1^-) \cap \mathcal{T}^+(k, h \mid t_2^-)$, entonces

$$(3) \quad \mathcal{J}^{k,h}(t_1^-, t^+) \subseteq \mathcal{J}^{k,h}(t_2^-, t^+).$$

$$(4) \quad \mathcal{I}^{k,h}(t_1^-, t^+) \subseteq \mathcal{I}^{k,h}(t_2^-, t^+).$$

La demostración completa se presenta en el Anexo A.10.

Con los conjuntos $\mathcal{I}^{k,h}$, $\mathcal{T}^-(k, h)$, $\mathcal{T}^+(k, h | t^-)$ y $\mathcal{J}^{k,h}(t^-, t^+)$ ya definidos, podemos formalizar el efecto promedio del tratamiento en una ventana (t^-, t^+) .

2.2.2.1 Definición: Efecto del tratamiento entre los tiempos t^- y t^+

Para cada configuración (k, h) , y para todo par (t^-, t^+) tal que

$$t^- \in \mathcal{T}^-(k, h) \quad \text{y} \quad t^+ \in \mathcal{T}^+(k, h | t^-),$$

se define

$$ATT^{t^-, t^+}(k, h) = \mathbb{E}[Y_{i, t^+}(h) - Y_{i, t^+}(h-1) | i \in \mathcal{I}^{k,h}(t^-, t^+)]. \quad (2.2.4)$$

Este efecto compara, en el mismo tiempo t^+ , el resultado potencial bajo nivel h con el que habría prevalecido bajo $h-1$ para las unidades que saltan en k y alcanzan h . El tiempo t^- solo delimita la ventana válida y se usará luego en la identificación mediante el supuesto de tendencias paralelas.

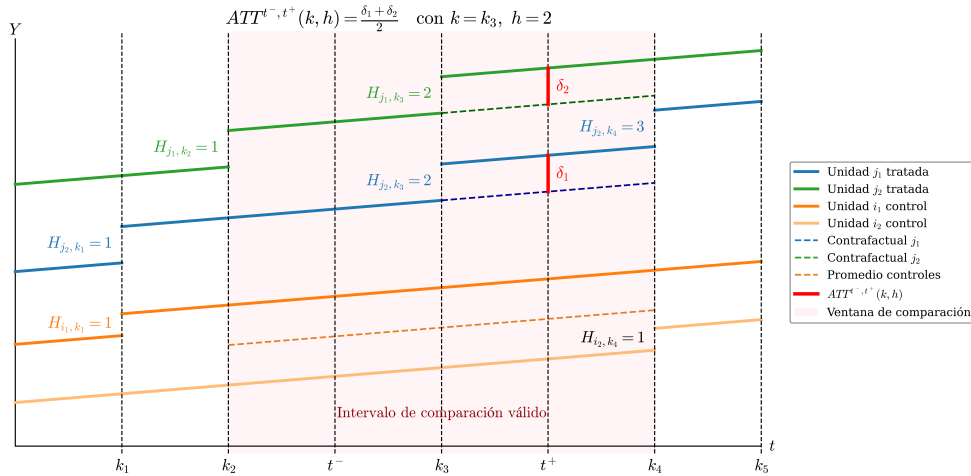


Figura 2.4: Ilustración esquemática de $ATT^{t^-, t^+}(k, h)$ para una ventana $[t^-, t^+]$ alrededor de k .

La Figura 2.4 ilustra esquemáticamente esta definición para una ventana $[t^-, t^+]$ alrededor de k .

Con $ATT^{t^-, t^+}(k, h)$ ya definido en la Definición 2.2.2.1, para identificarlo a partir de resultados observados necesitamos un supuesto de tendencias paralelas formulado a nivel de cada ventana (t^-, t^+) .

2.5 Supuesto: Tendencias paralelas por ventana

Para toda configuración (k, h) y todo par (t^-, t^+) tal que $t^- \in \mathcal{T}^-(k, h)$ y $t^+ \in \mathcal{T}^+(k, h \mid t^-)$, se cumple

$$\begin{aligned} & \mathbb{E}[Y_{i, t^+}(h-1) - Y_{i, t^-}(h-1) \mid i \in \mathcal{I}^{k, h}(t^-, t^+)] \\ &= \mathbb{E}[Y_{j, t^+}(H_{j, t^-}) - Y_{j, t^-}(H_{j, t^-}) \mid j \in \mathcal{J}^{k, h}(t^-, t^+)]. \end{aligned} \quad (2.2.5)$$

Proposición 2.4 (Identificación de $ATT^{t^-, t^+}(k, h)$ por ventana). *Bajo el Supuesto 2.5, para toda configuración (k, h) y todo par (t^-, t^+) tal que $t^- \in \mathcal{T}^-(k, h)$, $t^+ \in \mathcal{T}^+(k, h \mid t^-)$, se tiene que el efecto de la Definición 2.2.2.1 puede ser identificado por*

$$\begin{aligned} \widehat{ATT}^{t^-, t^+}(k, h) &= \underbrace{\mathbb{E}[y_{i, t^+} - y_{i, t^-} \mid i \in \mathcal{I}^{k, h}(t^-, t^+)]}_{\text{cambio observado tratadas}} \\ &\quad - \underbrace{\mathbb{E}[y_{j, t^+} - y_{j, t^-} \mid j \in \mathcal{J}^{k, h}(t^-, t^+)]}_{\text{cambio observado controles}}. \end{aligned} \quad (2.2.6)$$

o reescrito de forma equivalente como

$$\begin{aligned} \widehat{ATT}^{t^-, t^+}(k, h) &= \underbrace{\mathbb{E}[y_{i, t^+} - y_{j, t^+} \mid i \in \mathcal{I}^{k, h}(t^-, t^+), j \in \mathcal{J}^{k, h}(t^-, t^+)]}_{\text{diferencia intergrupala en } t^+} \\ &\quad - \underbrace{\mathbb{E}[y_{i, t^-} - y_{j, t^-} \mid i \in \mathcal{I}^{k, h}(t^-, t^+), j \in \mathcal{J}^{k, h}(t^-, t^+)]}_{\text{diferencia intergrupala en } t^-}. \end{aligned} \quad (2.2.7)$$

La demostración de la Proposición 2.4 se delega al Anexo A.9.

Nota 2.8 (Sobre el supuesto de tendencias paralelas). *El supuesto 2.5 empleado para la identificación tiene un carácter local por ventana: en cada intervalo válido $[t^-, t^+]$, la pendiente contrafactual de las unidades tratadas en el nivel $h-1$ se iguala con la pendiente observada de las unidades de control*

que permanecen en algún nivel $u < h$ a lo largo de la ventana. Este supuesto no exige que la pendiente sea idéntica para cada historial $u < h$ por separado.

Aunque en la discusión sustantiva del capítulo se considera el caso en el cual la tendencia se mantiene después de las intervenciones (efecto de nivel con pendiente estable en ventanas sin nuevos saltos), la formulación que verdaderamente impone esa invariancia por historial es la siguiente.

2.6 Supuesto: Paralelismo invariante por historia

Para toda configuración (k, h) y toda ventana válida (t^-, t^+) se cumple, para todo $u < h$, que

$$\begin{aligned} & \mathbb{E}[Y_{i,t^+}(h-1) - Y_{i,t^-}(h-1) \mid i \in \mathcal{I}^{k,h}(t^-, t^+)] \\ &= \mathbb{E}[Y_{j,t^+}(u) - Y_{j,t^-}(u) \mid j \in \mathcal{J}^{k,h}(t^-, t^+)]. \end{aligned}$$

En palabras: la *pendiente sin nuevo tratamiento* es común para las tratadas en $h-1$ y para las unidades de control en cualquier $u < h$ dentro de $[t^-, t^+]$.

El Supuesto 2.6 tiene un carácter global en el sentido de que exige igualdad de pendientes para todos los niveles $u < h$ de forma simultánea, haciendo la comparación invariante a la composición del grupo de control por historial. Además, dicho supuesto implica el supuesto local por ventana 2.5: al igualar la pendiente de las tratadas con la de los controles en cada $u < h$, la composición por niveles dentro de $\mathcal{J}^{k,h}(t^-, t^+)$ se vuelve irrelevante. En consecuencia, la identificación de $ATT^{t^-, t^+}(k, h)$ se mantiene.

Desde el punto de vista práctico, la versión reforzada simplifica el diseño (permite usar cualquier subconjunto de controles con $u < h$, incluso nunca tratados) y puede aumentar la potencia al combinar ventanas y controles; sin embargo, es más exigente: sería cuestionable incluso ante leves variaciones de la pendiente “sin nuevo tratamiento” con el historial u (sobre todo en niveles altos de intervención). Por ello se trabaja con el supuesto 2.5, que es menos restrictivo y mantiene la identificación con menor riesgo de sesgo cuando la pendiente puede depender del historial. Además, cuando los “nunca tratados” son escasos, el supuesto local resulta especialmente útil: permite usar controles con distintos $u < h$ siempre que no salten en $[t^-, t^+]$. Este enfoque local, además, puede evaluarse empíricamente de manera puntual mediante placebos previos por ventana, mientras que la verificación del reforzado (igualdad para todo u) resulta mucho más difícil. Finalmente, nótese que el uso del supuesto local no contradice el espíritu del capítulo: permite identificar $ATT^{t^-, t^+}(k, h)$

conservando la narrativa de efecto de nivel y pendiente estable en post, lo cual puede contrastarse con placebos posteriores adecuados.

2.2.3. Cheques placebo por ventana: antes y después

Con el objetivo de validar empíricamente las condiciones que sostienen el identificador por cortes (t^-, t^+) en (k, h) , proponemos dos *placebos* calculados en ventanas donde *no debería haber efecto*. El *placebo previo* evalúa la comparabilidad dinámica entre tratadas y controles *antes* del salto; el *placebo posterior* contrasta la hipótesis de *pendiente estable después* del salto (supuesto de este capítulo). La construcción por ventanas asegura que tanto tratadas como controles sean *efectivos* en cada intervalo, evitando las contaminaciones documentadas para TWFE con heterogeneidad de efectos (de Chaisemartin y D’Haultfoeulle (2020), Goodman-Bacon (2021) y Sun y Abraham (2021)) y es coherente con el uso de estimadores alternativos robustos (Callaway y Sant’Anna (2021)).

Placebos *antes* (pre). *¿Por qué?* Para verificar, de manera *local* en la ventana previa $[t^{--}, t^-]$, que las tratadas de la ventana (que saltarán en k) y sus controles efectivos seguían trayectorias comparables *antes* del salto y que no hubo anticipación. No pretende probar paralelismo global, sino diagnosticar la comparabilidad dinámica en la misma ventana que se usa para identificar $\widehat{ATT}^{t^-, t^+}(k, h)$; esta idea de placebo previo por ventana tiene respaldo formal (valor esperado cero bajo los supuestos) (de Chaisemartin y D’Haultfoeulle (2020)) y, además, la construcción por ventanas evita las contaminaciones propias de especificaciones TWFE con leads/lags bajo heterogeneidad (Sun y Abraham (2021)). *Pero este chequeo no alcanza por sí solo* (Rambachan y Roth (2023)).

¿Cómo? Elegimos (t^{--}, t^-) con $t^{--} < t^- < k$. Con base en la Proposición 2.3, usamos las *tratadas* y *controles* anclados en k : $i \in \mathcal{I}^{k, h}(t^{--}, k)$ y $j \in \mathcal{J}^{k, h}(t^{--}, k)$. Definimos el placebo como el mismo funcional “cambios tratadas menos cambios controles” en $[t^{--}, t^-]$:

$$\begin{aligned} \widehat{ATT}_{\text{pre}}^{t^{--}, t^-}(k, h) &= \underbrace{\mathbb{E}[y_{i, t^-} - y_{i, t^{--}} \mid i \in \mathcal{I}^{k, h}(t^{--}, k)]}_{\text{cambio obs. tratadas en pre}} \\ &\quad - \underbrace{\mathbb{E}[y_{j, t^-} - y_{j, t^{--}} \mid j \in \mathcal{J}^{k, h}(t^{--}, k)]}_{\text{cambio obs. controles en pre}}. \end{aligned} \tag{2.2.8}$$

¿Cómo leerlo? Si no hay anticipación y el *paralelismo local* en $[t^{--}, t^-]$ es plausible, entonces $\mathbb{E}[\widehat{ATT}_{\text{pre}}^{t^{--}, t^-}(k, h)] = 0$. Un valor distinto de cero sugiere pre-tendencias/anticipación. En la práctica, reportaremos IC y, preferiblemente, una *prueba de equivalencia* con margen sustantivo ε (para formalizar “ ≈ 0 ”), con ajuste por multiplicidad si se exploran varias ventanas, y además *intervalos de confianza/sensibilidad robustos* que acotan posibles violaciones post a partir de las pre-tendencias (Rambachan y Roth (2023)).

Placebos después (post). ¿Por qué? En este capítulo asumimos que tras el salto en k la *pendiente se mantiene*. Por tanto, en una ventana totalmente posterior, el “gap” tratado–control debe permanecer estable; este placebo verifica esa hipótesis *local* clave del capítulo. Además, al trabajar por ventanas evitamos las contaminaciones típicas de especificaciones TWFE con perfiles dinámicos (leads/lags) bajo heterogeneidad (Sun y Abraham (2021)). Como advierten Rambachan y Roth (2023), este chequeo *no alcanza por sí solo*.

¿Cómo? Elegimos (t^+, t^{++}) con $k < t^+ < t^{++}$. Con base en la Proposición 2.3, usamos *tratadas* y *controles* anclados en $k-1$: $i \in \mathcal{I}^{k,h}(k-1, t^{++})$ y $j \in \mathcal{J}^{k,h}(k-1, t^{++})$. Definimos el placebo post análogo al pre, evaluando cambios en $[t^+, t^{++}]$:

$$\begin{aligned} \widehat{ATT}_{\text{post}}^{t^+, t^{++}}(k, h) &= \underbrace{\mathbb{E}[y_{i,t^{++}} - y_{i,t^+} \mid i \in \mathcal{I}^{k,h}(k-1, t^{++})]}_{\text{cambio observado tratadas en post}} \\ &\quad - \underbrace{\mathbb{E}[y_{j,t^{++}} - y_{j,t^+} \mid j \in \mathcal{J}^{k,h}(k-1, t^{++})]}_{\text{cambio observado controles en post}}. \end{aligned} \tag{2.2.9}$$

¿Cómo leerlo? Bajo *pendiente estable* y ausencia de nuevos saltos/shocks diferenciales en $[t^+, t^{++}]$, se espera $\mathbb{E}[\widehat{ATT}_{\text{post}}^{t^+, t^{++}}(k, h)] = 0$. Un valor distinto de cero sugiere *dinámica del efecto* (p.ej. cambios de pendiente tras el salto); no invalida la identificación al cruce, pero indica que el comportamiento no es el del Cap. 2 y puede requerir un marco con tendencia que dependa del nivel (ver Cap. 3). Este diagnóstico es consistente con la lectura de perfiles dinámicos tipo event-study, pero evita la mezcla de cohortes y la contaminación de coeficientes documentada para TWFE (Sun y Abraham (2021)). Al igual que en el placebo pre, acompañaremos este chequeo con pruebas de equivalencia e *intervalos de confianza/sensibilidad robustos* que acotan violaciones post a partir de las pre-tendencias (Rambachan y Roth (2023)).

2.2.4. Agregación de efectos y construcción de pesos

Los estimadores $\widehat{ATT}^{t^-, t^+}(k, h)$ capturan el salto asociado al paso hacia el nivel h alrededor del instante k en una ventana específica (t^-, t^+) . Para reportar un *único* efecto representativo del corte (k, h) debemos agregar esas estimaciones puntuales. Cada comparación (t^-, t^+) recibe un peso que refleje la *información efectiva* que aporta: más observaciones útiles y mejor balance entre tratadas y controles implican, en principio, mayor precisión.

Este criterio es consistente con la descomposición clásica del estimador TWFE como promedio ponderado de comparaciones 2×2 mostrada por Goodman-Bacon (2021): allí, las comparaciones con mayor tamaño efectivo y variación relevante del tratamiento pesan más en el promedio global. Sin replicar esa estructura, aquí adoptamos una regla de ponderación mínima y transparente basada en el *tamaño muestral efectivo* de tratadas y controles dentro de cada ventana.

Sea $\mathcal{I}^{k,h}(t^-, t^+)$ el conjunto de tratadas efectivas y $\mathcal{J}^{k,h}(t^-, t^+)$ el de controles efectivos en la ventana. Definimos el *peso base* de cada comparación como

$$\omega(t^-, t^+ | k, h) = \frac{|\mathcal{I}^{k,h}(t^-, t^+)| |\mathcal{J}^{k,h}(t^-, t^+)|}{|\mathcal{I}^{k,h}(t^-, t^+)| + |\mathcal{J}^{k,h}(t^-, t^+)|} \quad (2.2.10)$$

el cual captura dos elementos clave: (i) balance entre tratadas y controles y (ii) tamaño efectivo de la comparación. Para una interpretación compacta en términos de proporciones y “varianza del tratamiento” dentro de la ventana, véase el Anexo A.11.

Para una agregación limpia, trabajamos con la versión normalizada de los pesos:

$$\widehat{\omega}(t^-, t^+ | k, h) = \frac{\omega(t^-, t^+ | k, h)}{\sum_{(u^-, u^+) \in \mathcal{W}(k, h)} \omega(u^-, u^+ | k, h)},$$

donde $\mathcal{W}(k, h)$ es el conjunto de parejas (t^-, t^+) admisibles para (k, h) .

Con ello, el efecto promedio por cada instante k y nivel h es

$$\widehat{ATT}(k, h) = \sum_{(t^-, t^+) \in \mathcal{W}(k, h)} \widehat{\omega}(t^-, t^+ | k, h) \cdot \widehat{ATT}^{t^-, t^+}(k, h).$$

Nota 2.9 (Sin distancia a k en los pesos). *Cada ventana usa sólo dos tiempos (t^-, t^+) ; por ello, la “cercanía/lejanía” a k no aporta información adicional dentro de la ventana. Además, bajo el supuesto de pendiente estable en post, el parámetro objetivo para (k, h) no debe depender de esa distancia.*

2.2.5. Agregación sobre tiempo k e intensidad h

Hasta aquí contamos con efectos por corte $\widehat{ATT}(k, h)$, ya agregados por ventanas. El siguiente paso es sintetizarlos en dos direcciones naturales: a lo largo del tiempo (k) y a lo largo de los niveles de intervención (h), y finalmente en un único promedio global. Para que la agregación sea transparente, usamos como “pesos” simples conteos de cuántas unidades efectivamente alcanzan cada nivel h en cada instante k ; al normalizarlos, esos conteos se pueden leer como probabilidades.

Como ya se definió en (2.2.1), el número de unidades que en k alcanzan por primera vez el nivel h es la *cardinalidad* $|\mathcal{I}^{k,h}|$.

A partir de esto, definimos las fracciones (probabilidades) que usaremos como pesos:

$$p(k, h) = \frac{|\mathcal{I}^{k,h}|}{\sum_u \sum_v |\mathcal{I}^{u,v}|}, \quad p(k) = \frac{\sum_v |\mathcal{I}^{k,v}|}{\sum_{u,v} |\mathcal{I}^{u,v}|}, \quad p(h) = \frac{\sum_u |\mathcal{I}^{u,h}|}{\sum_{u,v} |\mathcal{I}^{u,v}|},$$

$$p(h | k) = \frac{|\mathcal{I}^{k,h}|}{\sum_v |\mathcal{I}^{k,v}|}, \quad p(k | h) = \frac{|\mathcal{I}^{k,h}|}{\sum_u |\mathcal{I}^{u,h}|}.$$

Primero, fijamos un instante k . El siguiente promedio resume “lo que está pasando en k ” mezclando los cortes (k, h) según la composición observada de niveles en ese tiempo:

$$\widehat{ATT}_K(k) = \sum_h p(h | k) \widehat{ATT}(k, h). \quad (2.2.11)$$

Ahora, fijamos un nivel h . Este promedio recoge “el efecto a ese nivel” a lo largo del calendario, ponderando por cuándo ocurre ese nivel en los datos:

$$\widehat{ATT}_H(h) = \sum_k p(k | h) \widehat{ATT}(k, h). \quad (2.2.12)$$

Finalmente, si queremos un único número que resuma todo el diseño, promediamos directamente sobre todas las celdas (k, h) con su frecuencia relativa conjunta:

$$\widehat{ATT}_{\text{global}} = \sum_k \sum_h p(k, h) \widehat{ATT}(k, h). \quad (2.2.13)$$

En todos los casos, los pesos $p(\cdot)$ provienen de las cardinalidades $|\mathcal{I}^{k,h}|$ (definidas arriba) y se normalizan para sumar uno. Si para algún k (o h) no hay eventos, el correspondiente promedio condicional no se define y se omite.

Nota 2.10 (El global no depende del orden). Como $p(k)p(h | k) = p(k, h) = p(h)p(k | h)$ por construcción, se cumple

$$\sum_k p(k) \widehat{ATT}_K(k) = \sum_h p(h) \widehat{ATT}_H(h) = \widehat{ATT}_{global}.$$

Es decir, da lo mismo agregar “primero en k ” o “primero en h ”: el global es único. La comprobación detallada se presenta en el Anexo A.12.

2.2.6. Análisis de inferencia para los ATT por ventanas (t^-, t^+) en (k, h)

El objetivo es cuantificar la incertidumbre de los estimadores ya definidos a partir de ventanas (t^-, t^+) alrededor del salto (k, h) , sin introducir nueva notación. La inferencia descansa en los supuestos previamente establecidos: *representatividad* (Supuesto 2.3), *estabilidad del efecto en la ventana* (Supuesto 2.4) y *i.i.d. por unidad* (Supuesto 3).

(i) Agregado por (k, h) . Dado que el estimador agregado por (k, h) es un promedio ponderado de las ventanas válidas,

$$\widehat{ATT}(k, h) = \sum_{(t^-, t^+)} \hat{\omega}^{t^-, t^+}(k, h) \widehat{ATT}^{t^-, t^+}(k, h),$$

su comportamiento asintótico es normal siempre que el número de ventanas válidas crezca y los pesos no estén dominados por un subconjunto finito. En particular,

$$\sqrt{N_{k,h}}(\widehat{ATT}(k, h) - ATT(k, h)) \xrightarrow{d} \mathcal{N}(0, \sigma_{k,h}^2),$$

con varianza

$$\sigma_{k,h}^2 = \text{Var}\left(\sum_{(t^-, t^+)} \hat{\omega}^{t^-, t^+}(k, h) \widehat{ATT}^{t^-, t^+}(k, h)\right).$$

Caso especial (independencia entre comparaciones): si las ventanas empleadas en (k, h) fuesen independientes entre sí, entonces

$$\sigma_{k,h}^2 = \sum_{(t^-, t^+)} (\hat{\omega}^{t^-, t^+}(k, h))^2 \text{Var}(\widehat{ATT}^{t^-, t^+}(k, h)).$$

En presencia de reutilización de unidades, superposición de periodos o shocks comunes, la expresión anterior deja de ser válida y deben considerarse las covarianzas implícitas; en la práctica, las capturamos con remuestreo por unidades (bootstrap por bloques de unidad).¹

Un intervalo de confianza de nivel $1 - \alpha$ para $ATT(k, h)$ se reporta como

$$CI_{1-\alpha}^{(k,h)} = \left[\widehat{ATT}(k, h) \pm z_{1-\alpha/2} \widehat{\sigma}_{k,h} \right],$$

donde $\widehat{\sigma}_{k,h}$ se obtiene por bootstrap.

(ii) Marginal por instante y por nivel. Los marginales se definieron como combinaciones lineales del agregado por (k, h) :

$$\widehat{ATT}_K(k) = \sum_h p(h | k) \widehat{ATT}(k, h), \quad \widehat{ATT}_H(h) = \sum_k p(k | h) \widehat{ATT}(k, h).$$

Por linealidad, ambos son asintóticamente normales. Sus varianzas son, respectivamente,

$$\begin{aligned} \text{Var}(\widehat{ATT}_K(k)) &= \text{Var}\left(\sum_h p(h | k) \widehat{ATT}(k, h)\right) \\ \text{Var}(\widehat{ATT}_H(h)) &= \text{Var}\left(\sum_k p(k | h) \widehat{ATT}(k, h)\right) \end{aligned}$$

admitiendo covarianzas entre distintos niveles h (o instantes k) cuando éstas existan. Los intervalos de confianza se construyen de la misma forma que en (i), estimando las varianzas por bootstrap por unidades.

(iii) Estimador global. El estimador global se definió como

$$\widehat{ATT}_{\text{global}} = \sum_{k,h} p(k, h) \widehat{ATT}(k, h).$$

De nuevo, por linealidad y bajo los supuestos anteriores,

$$\sqrt{N}(\widehat{ATT}_{\text{global}} - ATT_{\text{global}}) \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{global}}^2),$$

con

$$\sigma_{\text{global}}^2 = \text{Var}\left(\sum_{k,h} p(k, h) \widehat{ATT}(k, h)\right),$$

¹El remuestreo por unidades re-muestra trayectorias completas y recalcula los estimadores, manteniendo los pesos normalizados como plug-in.

que incorpora, cuando existan, covarianzas entre diferentes (k, h) . La estimación práctica de σ_{global}^2 se realiza por bootstrap por unidades, lo que garantiza que la dependencia inducida por ventanas superpuestas o por shocks comunes quede reflejada en los errores estándar.

(iv) Reporte. Se informan errores estándar bootstrap, intervalos de confianza simétricos al 90 %, 95 % y 98 %, y códigos de significancia basados en valores- p bootstrap. Cuando se agregan múltiples instantes k o niveles h , el procedimiento captura automáticamente la covarianza entre componentes sin requerir supuestos adicionales.

2.2.7. Aplicación a datos simulados

Aplicamos el diseño *por cortes temporales* (sin ruptura de tendencia) sobre el mismo panel simulado de 60 unidades y 20 periodos descrito en el Anexo B (§B.1). La Figura 2.3 (presentada en la Sección 2.1.6) ilustra la evolución de y para las primeras ocho unidades y es también representativa del escenario aquí considerado. La implementación detallada (ventanas, conjuntos efectivos y placebos) se documenta en el Anexo B.3.

Resultados por instante k

De acuerdo con (2.2.11), los promedios por instante k se obtienen con $p(h | k)$. La tabla reporta punto, SE_{boot} , intervalo al 95 % y significancia (bootstrap por unidades).

Cuadro 2.4: Inferencia por instante k (diseño 1.2)

k	$\widehat{ATT}_K(k)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)
3	0.837022	0.159346	0.494569	1.090989
4	1.098082	0.172214	0.688616	1.316256
5	0.707535	0.178278	0.447589	1.164678
6	0.585633	0.125223	0.292961	0.787229
7	1.472518	0.145668	1.157020	1.685531
8	1.356586	0.218132	0.996401	1.805057
9	0.724828	0.115396	0.507158	0.963995
10	0.804357	0.122355	0.558590	1.110762
11	1.429169	0.249390	0.749239	1.613265
12	0.939340	0.154690	0.644160	1.237398
13	1.219886	0.200128	0.841826	1.583857
14	1.289593	0.214422	0.845700	1.629365
15	1.029119	0.213739	0.668369	1.576464
16	0.956727	0.191153	0.437519	1.109671
17	0.978248	0.240036	0.430431	1.316967

Resultados por nivel de intensidad h

Según (2.2.12), los promedios por nivel h usan $p(k | h)$. Incluimos el peso marginal $p(h)$ como w_h (en %).

Cuadro 2.5: Inferencia por nivel h (diseño 1.2)

h	$\widehat{ATT}_H(h)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	w_h (%)
1	1.029295	0.075466	0.848963	1.152194	27.12
2	1.005194	0.085944	0.850335	1.167594	25.99
3	1.131724	0.094928	0.896121	1.251719	21.47
4	0.768489	0.110978	0.515805	0.929746	15.25
5	1.018925	0.168675	0.720693	1.366656	7.91
6	1.177504	0.203099	0.720172	1.665167	2.26

Global y notas de validación

El promedio global, conforme a (2.2.13), es

$$\widehat{ATT}_{global} = 1.007767, \quad SE_{boot} = 0.050465, \quad IC95 \% = [0.887980, 1.080875].$$

Cuadro 2.6: Consistencia de agregación (vía k y vía h)

Promedio global (vía k)	1.007767
Promedio global (vía h)	1.007767

Nota 2.11. (1) Los diagnósticos placebo *PRE* (989) y *POST* (912) muestran alta estabilidad relativa: *VERDE* en 91.3 % y 89.5 % de los casos, respectivamente (véase Anexo B.3).

(2) La invarianza del orden de agregación se verifica numéricamente: $\sum_k p(k) \widehat{ATT}_K(k) = \sum_h p(h) \widehat{ATT}_H(h) = 1.007767$.

Capítulo 3

Diseños DiD con tratamientos repetidos, asignación no simultánea y efectos de nivel con ruptura de tendencia

Este capítulo aborda la identificación y estimación de efectos causales en contextos donde las trayectorias del resultado pueden experimentar *rupturas de tendencia* posteriores a las intervenciones. A diferencia del diseño previo —basado en la conservación del patrón de evolución—, aquí se admite que cada tratamiento puede modificar tanto el nivel como la *tendencia de evolución* del resultado. El propósito es extender los métodos del Capítulo 2 para capturar correctamente los *saltos* o *cambios en la trayectoria* que surgen tras la aplicación de un tratamiento repetido, manteniendo al mismo tiempo la comparabilidad entre unidades y ventanas temporales.

Trabajaremos bajo el *supuesto de ruptura de tendencia*: tras cada intervención, la trayectoria de y puede modificar su *tendencia* o *ritmo de evolución* respecto al periodo anterior. Este supuesto permite que los efectos del tratamiento se expresen ya sea como un *salto inmediato* o como una *variación gradual en la trayectoria posterior*, sin alterar el principio de comparabilidad entre grupos. Aun así, se mantiene la hipótesis de que, dentro de cada nivel acumulado h , las unidades tratadas y de comparación presentan *tendencias paralelas condicionales* a su historia y a las ventanas de observación empleadas.

Seguimos la notación general establecida en la introducción del documento. Se mantienen los Supuestos 1, 2 y 3, que garantizan la coherencia de los

niveles acumulativos H_{it} y la independencia entre unidades. Las versiones del supuesto de *tendencias paralelas* se formulan dentro de cada diseño y sección: en la comparación unidad–unidad con ruptura (Sección 3.1) y en la comparación tiempo–tiempo con ruptura (Sección 3.2).

A partir de aquí se desarrollan dos estrategias complementarias para estimar los efectos marginales bajo ruptura: una basada en comparaciones *entre unidades* (Sección 3.1) y otra en comparaciones *entre periodos de tiempos* (Sección 3.2). Ambas comparten la lógica de los ATT puntuales definidos en el Capítulo 2, pero adaptadas a la presencia de cambios estructurales.

La primera estrategia (Sección 3.1) mantiene la comparación entre unidades tratadas y de control en el mismo instante $t=k$, pero permitiendo que la *tendencia posterior* difiera entre ellas. El efecto puntual se obtiene a partir de los cambios promedio antes y después de la intervención para cada pareja (m, n) :

$$\begin{aligned} \widehat{ATT}_{m,n}^{\text{rup}}(k) &= \mathbb{E} \left[y_{m,\bar{t}} - y_{m,t} \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right] \\ &\quad - \mathbb{E} \left[y_{n,\bar{t}} - y_{n,t} \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right]. \end{aligned} \quad (3.0.1)$$

La identificación de este estimador se establece en la Prop. 3.1. Dicho estimador mide la diferencia en el cambio promedio de las trayectorias, capturando tanto los *saltos inmediatos* como las *variaciones posteriores en la tendencia*. De forma complementaria, la Sección 3.1.5 descompone este efecto en dos componentes: el *efecto instantáneo* —el salto en $t=k$ — y el *efecto total*, que incorpora la evolución posterior. El agregado por instante k , denotado $\widehat{ATT}_K^{\text{rup}}(k)$, resume los efectos entre todas las parejas válidas observadas en ese momento.

La segunda estrategia (Sección 3.2) fija dos momentos $t^- < k \leq t^+$ que enmarcan la ocurrencia de la intervención y compara la variación del resultado entre unidades tratadas y de control dentro de esa ventana. El efecto marginal por ventana se define como:

$$\begin{aligned} \widehat{ATT}_{\text{rup}}^{t^-,t^+}(k, h) &= \mathbb{E}[y_{i,t^+} - y_{i,t^-} \mid i \in \mathcal{I}^{k,h}(t^-, t^+)] \\ &\quad - \mathbb{E}[y_{j,t^+} - y_{j,t^-} \mid j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-, t^+)] , \end{aligned} \quad (3.0.2)$$

según la Def. 3.2.2.1 y la Prop. 3.2. El estimador $\widehat{ATT}_{\text{rup}}^{t^-,t^+}(k, h)$ capta la variación promedio del resultado entre periodos antes y después del evento,

ponderando las unidades tratadas y de control que permanecen comparables en la ventana.

Ambas estrategias generan estimadores por instante k y nivel acumulado h , que posteriormente se agregan en estimadores globales ponderados. Los pesos privilegian comparaciones informativas (ventanas amplias y bien balanceadas), siguiendo la motivación de la descomposición de TWFE y los criterios de eficiencia empleados en el Capítulo 2. La inferencia se basa en procedimientos de remuestreo (bootstrap por unidades) y en un esquema análogo al descrito en el capítulo anterior, manteniendo los mismos principios de independencia entre unidades y consistencia temporal.

Finalmente, la Sección 3.2.7 introduce una extensión con covariables X , que permite relajar la hipótesis de tendencias paralelas no condicionales y avanzar hacia un marco de *tendencias paralelas condicionales*. Allí se formulan tres representaciones equivalentes del efecto ajustado: (a) *Outcome Regression (OR)*, (b) *Inverse Probability Weighting (IPW)*, y (c) *Doubly Robust (DR)*. Esta última combina ambos mecanismos y mantiene consistencia siempre que al menos uno de los modelos de primer paso esté correctamente especificado. En esta extensión, el análisis de inferencia se realiza específicamente sobre el estimador *Doubly Robust*, considerado como el más eficiente y estable frente a posibles errores de especificación en los modelos auxiliares.

En las aplicaciones con datos simulados se muestra que los métodos propuestos detectan correctamente los saltos y los cambios de pendiente incorporados artificialmente en el proceso generador de datos. Las simulaciones se realizaron sin incluir covariables, de modo que los resultados corresponden al diseño básico con ruptura de tendencia. Aun así, las estimaciones permanecen estables ante variaciones del tamaño de ventana, lo que respalda la coherencia interna del enfoque y su capacidad para captar rupturas estructurales dentro del marco de tratamientos repetidos y asignación no simultánea.

3.1. Comparación unidad–unidad (promedio temporal)

Siguiendo la notación general fijada en la introducción, trabajamos con unidades $i \in \mathcal{I}$, periodos $t \in \mathcal{T}$, el instante k y el nivel acumulado H_{it} . Retomamos la comparación *unidad–unidad* presentada en la Sección 2.1, pero ahora permitiendo *ruptura de tendencia*. Por “ruptura” entenderemos lo siguiente: (i) antes de cualquier intervención ($h=0$) las trayectorias exhiben tendencias paralelas entre unidades; (ii) cuando una unidad pasa de $h-1$ a h en $t=k$, su tendencia puede cambiar; (iii) las tendencias son paralelas dentro de cada nivel h ; y (iv) la adopción no es simultánea entre unidades (el momento k puede diferir por unidad).

El objetivo aquí es comparar, en $t=k$, una unidad que pasa de $h-1$ a h con otra que en k no cambia (permanece en $h-1$), dentro de ventanas admisibles alrededor de k , y promediar *en el tiempo* para cada par. A partir de esta idea, precisaremos qué comparaciones son válidas bajo los supuestos del capítulo y cómo se construyen las esperanzas correspondientes.

3.1.1. Comparaciones válidas en diseños con múltiples intervenciones

En la comparación *unidad–unidad* con adopción no simultánea y *ruptura de tendencia*, una comparación es válida si el “control” reproduce la trayectoria contrafactual que habría seguido la unidad tratada en ausencia del *salto adicional* en $t=k$ y, alrededor de ese instante, las trayectorias mantienen *tendencias paralelas por tramos* (antes y después de k).

Para evaluar el salto de $h-1$ a h en $t=k$, se empareja una unidad tratada que *cambia* de $h-1$ a h en k con una unidad de comparación que, en ese mismo k , *no cambia y permanece exactamente en $h-1$* . La comparación se realiza en alguna vecindad alrededor de $t=k$ y se promedia *en el tiempo* para cada par, bajo el supuesto de tendencias paralelas por tramos en dicha vecindad.

A diferencia del caso sin ruptura, aquí el control debe estar exactamente en $h-1$ en k . No es válido usar como control: (i) unidades en niveles menores que $h-1$, (ii) unidades en el mismo nivel h , o (iii) unidades en niveles mayores que h . Tampoco es válido que el control cambie de nivel en k ni que presente anticipación (cambio dentro de la vecindad anterior a k).

Nota metodológica. El contrafactual “puro” con nunca tratadas ($H=0$) solo es pertinente para evaluar el salto hacia $h=1$. Para $h \geq 2$ deja de ser informativo respecto del salto específico en k ; por ello no se usa aquí como comparación principal.

3.1.2. Análisis DiD (unidad–unidad) con controles en $h-1$

Usaremos el conjunto de tratadas $\mathcal{I}^{k,h}$ definido en (2.2.1) (véase la Sección 2.2). Definimos como conjunto de comparación en k a las unidades que permanecen en $h-1$:

$$\mathcal{C}^{k,h} = \{n \mid H_{n,k-1} = H_{n,k} = h-1\}.$$

Es decir, $\mathcal{I}^{k,h}$ recoge las unidades que en $t=k$ saltan de $h-1$ a h , y $\mathcal{C}^{k,h}$ las que, en ese mismo k , no cambian y permanecen exactamente en $h-1$.

Para cada par (m, n) fijamos los extremos de la ventana de comparación alrededor de k mediante los tiempos “pre” y “post” del caso con ruptura. El extremo “pre” es el primer instante anterior a k en que ambas unidades están en $h-1$ (el menor $t < k$ tal que $H_{m,t} = H_{n,t} = h-1$); el “post” es el máximo instante, a partir de k , en el que la tratada está en h y la comparación permanece en $h-1$. Esta convención es análoga a la empleada en la Sección 2.1, pero se marca con rup para distinguir el diseño con ruptura.

$$t_{m,n}^{\text{pre, rup}}(k) = \min \{t^- : t^- < k, H_{m,t^-} = h-1, H_{n,t^-} = h-1\},$$

$$t_{m,n}^{\text{pos, rup}}(k) = \max \{t^+ : k \leq t^+, H_{m,t^+} = h, H_{n,t^+} = h-1\},$$

$$t_{m,n}^{\text{pre, rup}}(k) \leq t^- < k \leq t^+ \leq t_{m,n}^{\text{pos, rup}}(k).$$

Medimos el efecto marginal del *salto adicional* en $t=k$ para una unidad m que pasa de $h-1$ a h , tomando como referencia una unidad n que en k permanece en $h-1$. El efecto se resume promediando el tramo posterior de la ventana válida, delimitada por $t_{m,n}^{\text{pre, rup}}(k)$ y $t_{m,n}^{\text{pos, rup}}(k)$.

3.1.2.1 Definición: Efecto marginal con ruptura (comparación unidad–unidad)

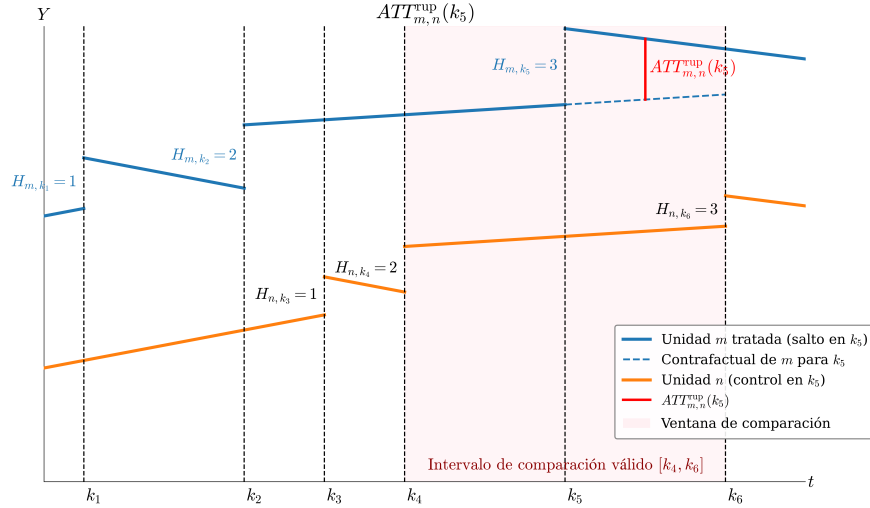


Figura 3.1: Ilustración de $ATT_{m,n}^{rup}(k_5)$ con control en $h-1$ en k_5 . La ventana de comparación es $[k_4, k_6] = [10, 18]$; el contrafactual extiende la tendencia en $h-1$ y el efecto puntual se marca en $t = 16$.

Sean $m \in \mathcal{I}^{k,h}$ y $n \in \mathcal{C}^{k,h}$. Definimos el efecto causal puntual del salto adicional en k como

$$ATT_{m,n}^{rup}(k) = \mathbb{E}[Y_{m,t}(h) - Y_{m,t}(h-1) \mid k \leq t \leq t_{m,n}^{\text{pos},rup}(k)]. \quad (3.1.1)$$

Aquí $h = H_{m,k}$ y la ventana posterior admisible para el par (m, n) está delimitada por $t_{m,n}^{\text{pre},rup}(k)$ y $t_{m,n}^{\text{pos},rup}(k)$.

Nota 3.1. Al fijar m y k queda determinado $h := H_{m,k}$; por economía de notación escribimos $ATT_{m,n}^{rup}(k)$ en lugar de $ATT_{m,n}^{rup}(k; h)$.

La Figura 3.1 ejemplifica el objeto de la Definición 3.1.2.1 para el salto en k_5 . La unidad m salta de $h-1$ a h en k_5 , mientras que n permanece en $h-1$ dentro de la ventana admisible $[k_4, k_6] = [10, 18]$ (sombreada). El contrafactual de m prolonga, para $t \geq k_5$, la tendencia del tramo anterior (nivel $h-1$), y la barra roja muestra el contraste puntual en $t = 16$ (mitad de $[k_5, k_6]$).

Nota 3.2. El contraste causal dentro del corchete involucra únicamente potenciales de m ; sin embargo, el promedio depende de n porque la ventana admisible es propia del par (m, n) . Conceptualmente, existe una ventana

máxima que depende solo de m : el último instante previo a k en que m aún está en $h-1$ y el último instante posterior (a partir de k) en que m permanece en h antes de un nuevo cambio de nivel. Si se promediara sobre esa ventana máxima, el objeto no dependería de n ; no obstante, tal objeto no es identificable con observables sin especificar una unidad de comparación que provea el contrafactual y garantice soporte común. Por ello, en el cuerpo principal trabajamos con $ATT_{m,n}^{\text{rup}}(k)$, definido sobre la ventana admisible del par (m, n) ; distintos controles $n_1, n_2 \in \mathcal{C}^{k,h}$ pueden inducir ventanas de distinta longitud y, en consecuencia, valores distintos del estimador para el mismo salto de m en k . La formalización de la ventana máxima de m y su relación con las truncaciones inducidas por n se presenta en el Anexo A.13.

Nota 3.3 (Comparación con el caso sin ruptura). *En la Sección 2.1 (diseño sin ruptura) la ventana máxima también es perfectamente definible; no se enfatizó porque, bajo el contrafactual puro con nunca tratadas (cf. Supuesto 2.1), la ventana admisible coincide con la máxima y queda plenamente identificada: las nunca tratadas sostienen el contrafactual en toda la vecindad relevante para cualquier salto. En el presente diseño con ruptura, en cambio, la admisible puede ser un subconjunto estricto de la máxima porque el control debe estar exactamente en $h-1$ en $t=k$ y no cambiar/anticiparse en la vecindad; por ejemplo, con sólo las dos unidades de la figura 3.1, la ventana máxima (post) en k_5 excede la admisible (post), que termina en k_6 .*

Para identificar empíricamente $ATT_{m,n}^{\text{rup}}(k)$ con datos observados, fijamos el par (m, n) y promediamos en el tiempo dentro de la ventana válida. En lo que sigue, las esperanzas condicionan exclusivamente los tiempos: se promedia sobre todas las parejas (t^-, t^+) con $t_{m,n}^{\text{pre,rup}}(k) \leq t^- \leq k-1$ y $k \leq t^+ \leq t_{m,n}^{\text{pos,rup}}(k)$. Bajo un supuesto local de *tendencias paralelas por nivel* en $h-1$, la evolución contrafactual de m (si no hubiese saltado en k) coincide, en promedio temporal, con la de n , que permanece en $h-1$. Este supuesto se enuncia a continuación.

3.1 Supuesto: Tendencias paralelas por nivel $h-1$ (ventanas pre y post)

Fija $m \in \mathcal{I}^{k,h}$ y $n \in \mathcal{C}^{k,h}$. Entonces:

$$\begin{aligned} \mathbb{E} \left[Y_{m,t^+}(h-1) - Y_{m,t^-}(h-1) \mid \begin{array}{l} t_{m,n}^{\text{pre,rup}}(k) \leq t^- \leq k-1 \\ k \leq t^+ \leq t_{m,n}^{\text{pos,rup}}(k) \end{array} \right] = \\ \mathbb{E} \left[Y_{n,t^+}(h-1) - Y_{n,t^-}(h-1) \mid \begin{array}{l} t_{m,n}^{\text{pre,rup}}(k) \leq t^- \leq k-1 \\ k \leq t^+ \leq t_{m,n}^{\text{pos,rup}}(k) \end{array} \right] = \end{aligned} \quad (3.1.2)$$

Equivalente:

$$\begin{aligned} \mathbb{E} \left[Y_{m,t^+}(h-1) - Y_{n,t^+}(h-1) \mid \begin{array}{l} t_{m,n}^{\text{pre,rup}}(k) \leq t^- \leq k-1 \\ k \leq t^+ \leq t_{m,n}^{\text{pos,rup}}(k) \end{array} \right] = \\ \mathbb{E} \left[Y_{m,t^-}(h-1) - Y_{n,t^-}(h-1) \mid \begin{array}{l} t_{m,n}^{\text{pre,rup}}(k) \leq t^- \leq k-1 \\ k \leq t^+ \leq t_{m,n}^{\text{pos,rup}}(k) \end{array} \right] = \end{aligned} \quad (3.1.3)$$

Nota 3.4 (Elección del control y contraste con 2.1). *En presencia de ruptura de tendencia el control debe estar **exactamente en** $h-1$ en $t=k$. En particular: (i) no es válido usar controles en niveles $< h-1$, en h o $> h$; (ii) el control no debe cambiar de nivel en k ni anticiparse dentro de la ventana; y (iii) el paralelismo se exige por nivel y de forma local en la ventana.*

Esto contrasta con la sección 2.1: el contrafactual puro con nunca tratadas (Supuesto 2.1) sólo es admisible para el primer salto ($h=1$); para $h \geq 2$ deja de identificar el efecto marginal en k . A su vez, el supuesto entre tratadas (Supuesto 2.2) imponía paralelismo entre niveles distintos $\underline{h} < h$; aquí se restringe al nivel $h-1$ del par (m, n) .

Implicación práctica. Aunque el requisito de paralelismo es más local (sólo en $h-1$), empíricamente es más exigente en términos de soporte: reduce el conjunto de controles elegibles a quienes están en $h-1$ justo en k , invalida pares con cambios/anticipaciones en esa vecindad y puede acortar la ventana si m vuelve a saltar pronto. En bases con adopciones dispersas, esto puede dejar pocos pares válidos y elevar la varianza de los estimadores.

Proposición 3.1 (Identificación del efecto $ATT_{m,n}^{\text{rup}}(k)$ (unidad–unidad, control en $h-1$)). *Bajo el Supuesto 3.1, el efecto de la Definición 3.1.2.1 se identifica como:*

$$\begin{aligned}
\widehat{ATT}_{m,n}^{\text{rup}}(k) &= \underbrace{\mathbb{E}\left[y_{m,t^+} - y_{m,t^-} \mid \begin{array}{l} t_{m,n}^{\text{pre,rup}}(k) \leq t^- \leq k-1 \\ k \leq t^+ \leq t_{m,n}^{\text{pos,rup}}(k) \end{array} \right]}_{\text{cambio observado en la tratada (m)}} \\
&\quad - \underbrace{\mathbb{E}\left[y_{n,t^+} - y_{n,t^-} \mid \begin{array}{l} t_{m,n}^{\text{pre,rup}}(k) \leq t^- \leq k-1 \\ k \leq t^+ \leq t_{m,n}^{\text{pos,rup}}(k) \end{array} \right]}_{\text{cambio observado en la comparación (n)}}.
\end{aligned} \tag{3.1.4}$$

Forma equivalente (DiD intergrupal):

$$\begin{aligned}
\widehat{ATT}_{m,n}^{\text{rup}}(k) &= \underbrace{\mathbb{E}\left[y_{m,t^+} - y_{n,t^+} \mid k \leq t^+ \leq t_{m,n}^{\text{pos,rup}}(k) \right]}_{\text{diferencia intergrupal en el periodo posterior (t^+)}} \\
&\quad - \underbrace{\mathbb{E}\left[y_{m,t^-} - y_{n,t^-} \mid t_{m,n}^{\text{pre,rup}}(k) \leq t^- \leq k-1 \right]}_{\text{diferencia intergrupal en el periodo previo (t^-)}}.
\end{aligned} \tag{3.1.5}$$

La demostración es análoga a la de la Proposición 2.2 y se sigue paso a paso del Anexo A.2, reemplazando el supuesto 2.2 por 3.1, los tiempos (\underline{t}, \bar{t}) por (t^-, t^+) , $Y_{i,\cdot}(\underline{h})$ por $Y_{i,\cdot}(h-1)$, y el conjunto de control por $\mathcal{C}^{k,h}$.

3.1.3. Agregación de efectos

Siguiendo la misma idea y justificación expuesta en la Sección 2.1.4 —esto es, la motivación por la descomposición TWFE de Goodman-Bacon (2021)— podemos definir pesos de forma análoga, justificados siguiendo las mismas líneas del Anexo A.5.

Para cada celda (k, h) y para cada comparación válida, con $m \in \mathcal{I}^{k,h}$ y $n \in \mathcal{C}^{k,h}$, asignamos un peso proporcional a la información efectiva de su ventana admisible alrededor de k . El *peso base* es el mismo producto “pre \times post” utilizado en la Sección 2.1.4:

$$\omega_{m,n}^{\text{rup}}(k, h) = (t_{m,n}^{\text{pos,rup}}(k) - k + 1) (k - t_{m,n}^{\text{pre,rup}}(k)).$$

Su versión normalizada es:

$$\widehat{\omega}_{m,n}^{\text{rup}}(k, h) = \frac{\omega_{m,n}^{\text{rup}}(k, h)}{\sum_{(m',n')} \omega_{m',n'}^{\text{rup}}(k, h)}.$$

Con estos pesos, el efecto agregado por nivel de intervención h en el instante k se define como:

$$\widehat{ATT}^{\text{rup}}(k, h) = \sum_{(m,n)} \widehat{\omega}_{m,n}^{\text{rup}}(k, h) \widehat{ATT}_{m,n}^{\text{rup}}(k).$$

Nota 3.5. En la agregación fijamos la celda (k, h) , por eso h aparece explícito en $\widehat{ATT}^{\text{rup}}(k, h)$. En los sumandos, h está implícito: para cada par (m, n) y k se toma $h := H_{m,k}$.

Nota 3.6. Aunque la forma funcional de los pesos es la misma que en la Sección 2.1.4, sus valores no tienen por qué coincidir. La diferencia proviene del conjunto de controles admisibles: en 2.1.4 las comparaciones se forman con $\mathcal{I}_m(k)$, mientras que aquí se restringen a $\mathcal{C}^{k,h}$ (controles que permanecen en $h-1$ en k). En consecuencia, las ventanas admisibles ($t_{m,n}^{\text{pre,rup}}(k)$, $t_{m,n}^{\text{pos,rup}}(k)$) y, por ende, las ponderaciones normalizadas $\widehat{\omega}_{m,n}^{\text{rup}}(k, h)$ pueden diferir respecto del caso sin ruptura. Esta discrepancia es esperable y metodológicamente deseable, pues refleja la elegibilidad específica del control en el diseño con ruptura.

Nota 3.7. En la Sección 2.1.4 (diseño sin ruptura), el objeto agregado $\widehat{ATT}(k)$ se construye únicamente por instante k , ya que las tendencias paralelas son globales y no requieren distinguir entre niveles h : los distintos saltos que ocurren en el mismo k se colapsan en un único promedio. En cambio, en la Sección 3.1.3 (diseño con ruptura), las tendencias sólo son paralelas dentro de cada nivel de intervención h . Por ello, el objeto homogéneo de interés es $\widehat{ATT}^{\text{rup}}(k, h)$, definido a nivel de celda (k, h) , y no basta con agregar únicamente sobre k .

3.1.4. Agregación entre celdas y estimadores globales

Partiendo de los estimadores de celda $\widehat{ATT}^{\text{rup}}(k, h)$ (obtenidos agregando comparaciones (m, n) con los pesos par-par $\widehat{\omega}_{m,n}^{\text{rup}}(k, h)$ dentro de cada (k, h)), definimos pesos *entre celdas* que dependen del *soporte efectivo* de tratadas y controles, sin reutilizar los $\omega_{m,n}$.

Peso entre celdas: tamaño efectivo balanceado

Sea $T(k, h) := |\mathcal{I}^{k,h}|$ el número de unidades que en $t=k$ saltan de $h-1$ a h , y $C(k, h) := |\mathcal{C}^{k,h}|$ el número de controles elegibles que permanecen en $h-1$

en k . Definimos el *peso de celda* como

$$W(k, h) = \frac{T(k, h) C(k, h)}{T(k, h) + C(k, h)}. \quad (3.1.6)$$

Este peso crece con el tamaño y penaliza el desbalance entre grupos. Nótese que $W(k, h)$ es *distinto* de los pesos par-par $\omega_{m,n}^{\text{rup}}(k, h)$ utilizados en la agregación *intra-celda*.

Nota 3.8. *Alternativamente puede usarse $W(k, h) = \min\{T(k, h), C(k, h)\}$ para priorizar aún más el balance. En esta tesis adoptamos (3.1.6) por su conexión con precisión en comparaciones de medias.*

Agregación sobre h (resumen por instante k)

Para sintetizar en un único estimador por instante, mezclamos las celdas (k, h) con pesos relativos proporcionales a $W(k, h)$:

$$\Psi(k, h) = \frac{W(k, h)}{\sum_{h'} W(k, h')}, \quad \widehat{ATT}_K^{\text{rup}}(k) = \sum_h \Psi(k, h) \widehat{ATT}^{\text{rup}}(k, h). \quad (3.1.7)$$

Agregación sobre k (resumen por nivel h)

Análogamente, para resumir a través del calendario para un nivel dado h :

$$\Phi(h, k) = \frac{W(k, h)}{\sum_{k'} W(k', h)}, \quad \widehat{ATT}_H^{\text{rup}}(h) = \sum_k \Phi(h, k) \widehat{ATT}^{\text{rup}}(k, h). \quad (3.1.8)$$

Estimador global e invarianza del orden

Definimos el promedio global como el promedio ponderado de todas las celdas (k, h) con $W(k, h)$:

$$\widehat{ATT}_{\text{global}}^{\text{rup}} = \frac{\sum_k \sum_h W(k, h) \widehat{ATT}^{\text{rup}}(k, h)}{\sum_k \sum_h W(k, h)}. \quad (3.1.9)$$

Sea

$$\pi(k) := \frac{\sum_h W(k, h)}{\sum_{u,v} W(u, v)}, \quad \rho(h) := \frac{\sum_k W(k, h)}{\sum_{u,v} W(u, v)}.$$

Por construcción,

$$\sum_k \pi(k) \widehat{ATT}_K^{\text{rup}}(k) = \sum_h \rho(h) \widehat{ATT}_H^{\text{rup}}(h) = \widehat{ATT}_{\text{global}}^{\text{rup}},$$

esto es, el estimador global no depende del orden de agregación (primero en k o primero en h).

Nota 3.9. La elección de los pesos $W(k, h)$ responde a la propia estructura del diseño con ruptura. Dado que los efectos $\widehat{ATT}^{\text{rup}}(k, h)$ se definen a nivel de celdas (k, h) , resulta más natural ponderarlos según el tamaño efectivo y el balance de tratadas y controles en cada celda. En este sentido, la lógica de agregación se acerca más a la empleada en la Sección 2.2.4, donde también se pondera por tamaños efectivos, que a la de la Sección 2.1.4, basada únicamente en ventanas temporales. De esta manera, se mantiene la coherencia con el principio baconiano de dar mayor peso a las comparaciones más informativas, adaptado aquí a la doble indexación en k y h .

3.1.5. Efecto instantáneo y comparación con el efecto total

En el diseño con ruptura, además del estimador ya definido sobre la ventana posterior admisible $\widehat{ATT}_{m,n}^{\text{rup}}(k)$, resulta útil descomponer el efecto en dos piezas: (i) el *efecto instantáneo* (salto en $t=k$) y (ii) la *parte no instantánea* asociada a la evolución posterior. Esta descomposición no altera la estructura de identificación ni los pesos de agregación; únicamente *restringe* la ventana POST cuando se calcula el efecto instantáneo.

Efecto instantáneo (salto en k). Para cada par (m, n) con $m \in \mathcal{I}^{k,h}$ y $n \in \mathcal{C}^{k,h}$, fijamos

$$t_{m,n}^{\text{pos,rup}}(k) = k, \quad \text{manteniendo } t_{m,n}^{\text{pre,rup}}(k) \text{ inalterado.}$$

El estimador instantáneo (salto de nivel en k) se define como

$$\widehat{ATT}_{m,n}^{\text{inst}}(k) = (y_{m,k} - y_{n,k}) - \mathbb{E}[y_{m,t^-} - y_{n,t^-} \mid t_{m,n}^{\text{pre,rup}}(k) \leq t^- \leq k-1]. \quad (3.1.10)$$

Agregación y comparación. Para construir los análogos agregados $\widehat{ATT}^{\text{inst}}(k, h)$, $\widehat{ATT}_K^{\text{inst}}(k)$, $\widehat{ATT}_H^{\text{inst}}(h)$ y $\widehat{ATT}_{\text{global}}^{\text{inst}}$, se sustituyen $\widehat{ATT}_{m,n}^{\text{rup}}(k)$ por $\widehat{ATT}_{m,n}^{\text{nivel}}(k)$ en las expresiones de agregación ya dadas en §3.1.3–3.1.4, manteniendo *exactamente* los mismos pesos par–par $\widehat{\omega}_{m,n}^{\text{rup}}(k, h)$ y pesos entre celdas $W(k, h)$.

Para cualquier nivel de reporte (celda (k, h) , colapsos en K o H , o global), definimos la diferencia

$$\Delta^{\text{post}} = \widehat{ATT}^{\text{rup}} - \widehat{ATT}^{\text{inst}}, \quad (3.1.11)$$

que recoge la parte *no instantánea* del efecto (evolución posterior al salto).

- Si $\widehat{ATT}^{\text{rup}}$ y $\widehat{ATT}^{\text{inst}}$ *coinciden* en magnitud y signo ($\Delta^{\text{post}} \approx 0$), la intervención se manifiesta principalmente como un *cambio de nivel estable* (poca ruptura de tendencia).
- Si $\Delta^{\text{post}} > 0$, hay *acumulación/fortalecimiento* tras el salto.
- Si $\Delta^{\text{post}} < 0$, el salto se *atenúa o revierte* en promedio en el POST.
- Si difieren en el *signo*, hay *reversión*: el salto y el promedio posterior van en sentidos opuestos.

Nota 3.10 (Lectura rápida de la diferencia Δ^{post}). *Cuando la diferencia $\Delta^{\text{post}} = \widehat{ATT}^{\text{rup}} - \widehat{ATT}^{\text{inst}}$ es pequeña en relación con $\widehat{ATT}^{\text{rup}}$, la lectura es que casi todo el efecto se explica por el salto instantáneo; la ruptura de tendencia inducida por la intervención es mínima.*

En cambio, si Δ^{post} es grande en relación con $\widehat{ATT}^{\text{rup}}$, una parte sustantiva del efecto proviene de la dinámica posterior, lo que indica una ruptura marcada en la tendencia.

3.1.6. Análisis de inferencia

El esquema de inferencia en el diseño con ruptura sigue las mismas ideas de la Sección 2.1.5. Al igual que allí, los estimadores marginales por pareja $\widehat{ATT}_{m,n}^{\text{rup}}(k)$ no son objeto individual de inferencia (por involucrar una única pareja y pocas observaciones temporales), de modo que la inferencia válida se realiza sobre agregados construidos con pesos:

$$\widehat{ATT}^{\text{rup}}(k, h), \quad \widehat{ATT}_K^{\text{rup}}(k), \quad \widehat{ATT}_H^{\text{rup}}(h), \quad \widehat{ATT}_{\text{global}}^{\text{rup}},$$

definidos en las Secciones 3.1.3–3.1.4.

Bajo el Supuesto 3 (muestreo i.i.d. por unidad y no interferencia), las comparaciones $\widehat{ATT}_{m,n}^{\text{rup}}(k)$ pueden exhibir dependencia intraclúster por traslapes (en unidades y en tiempo). Por ello, como en la Sección 2.1.5, la varianza de los agregados no se obtiene en general como suma de varianzas sin covarianzas. La inferencia efectiva (errores estándar, intervalos de confianza y valores p) se obtiene mediante *bootstrap por unidad* y, cuando procede, *block bootstrap temporal*, reconstruyendo en cada réplica ventanas, pesos y agregados.

El rol del Supuesto 2.4 en la Sección 2.1.5 es aquí desempeñado por el paralelismo por nivel en $h-1$ (Supuesto 3.1), que garantiza la validez del contraste dentro de la ventana admisible $[t_{m,n}^{\text{pre,rup}}(k), t_{m,n}^{\text{pos,rup}}(k)]$. Asimismo, mantenemos el supuesto de *representatividad* de las comparaciones utilizado en la Sección 2.1.5 (cf. Supuesto 2.3).

Teorema 3.1 (Comportamiento asintótico de $\widehat{ATT}^{\text{rup}}(k, h)$). *Bajo los Supuestos 3, 3.1 y 2.3, y suponiendo que el número de comparaciones válidas en la celda (k, h) , $N_{k,h}$, crece de modo que los pesos normalizados permanezcan acotados, se tiene*

$$\sqrt{N_{k,h}}(\widehat{ATT}^{\text{rup}}(k, h) - ATT^{\text{rup}}(k, h)) \xrightarrow{d} \mathcal{N}(0, \sigma_{k,h}^2),$$

para alguna varianza finita $\sigma_{k,h}^2$ que se estima consistentemente mediante *bootstrap por unidad* (y, cuando aplica, *block bootstrap temporal*) reconstruyendo comparaciones, pesos y agregados en cada réplica. Los intervalos de confianza de nivel $1-\alpha$ se obtienen por cuantiles percentil o BCa del *bootstrap*, o en versión normal usando la *desviación estándar bootstrap*.

Demostración. Véase el Anexo A.14. El razonamiento de la sección 2.1.5 se adapta a celdas (k, h) bajo paralelismo local en $h-1$. \square

Teorema 3.2 (Agregados en k y en h). *Sean*

$$\widehat{ATT}_K^{\text{rup}}(k) = \sum_h \Psi(k, h) \widehat{ATT}^{\text{rup}}(k, h), \quad \widehat{ATT}_H^{\text{rup}}(h) = \sum_k \Phi(h, k) \widehat{ATT}^{\text{rup}}(k, h),$$

con

$$\Psi(k, h) \geq 0, \quad \sum_h \Psi(k, h) = 1, \quad \Phi(h, k) \geq 0, \quad \sum_k \Phi(h, k) = 1,$$

donde Ψ y Φ se definen a partir de $W(k, h)$ (cf. (3.1.6)–(3.1.8)). Bajo las condiciones del Teorema 3.1 y acotación de los pesos, se tiene asintótica

normalidad para $\widehat{ATT}_K^{\text{rup}}(k)$ y $\widehat{ATT}_H^{\text{rup}}(h)$, con varianzas estimables por el mismo esquema de bootstrap.

Demostración. Véase el Anexo A.14. Se combinan $\widehat{ATT}^{\text{rup}}(k, h)$ con pesos Ψ y Φ ; normalidad por linealidad y bootstrap por unidad para varianzas. \square

Teorema 3.3 (Promedio global e invarianza del orden). Para $\widehat{ATT}_{\text{global}}^{\text{rup}}$ definido en (3.1.9), bajo las condiciones previas,

$$\sqrt{N}(\widehat{ATT}_{\text{global}}^{\text{rup}} - ATT_{\text{global}}^{\text{rup}}) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad N = \sum_{k,h} N_{k,h},$$

y σ^2 se estima por bootstrap reproduciendo todo el proceso de agregación entre celdas. Los ICs se reportan a niveles 90 %, 95 % y 98 %, y la significancia se codifica con el mismo criterio que en la Sección 2.1.5.

Demostración (remisión). Véase el Anexo A.14. La agregación total ponderada preserva la normalidad; la incertidumbre se estima por bootstrap. \square

3.1.7. Aplicación a datos simulados (con ruptura)

Aplicamos el método sobre un panel simulado de 60 unidades y 20 periodos bajo el escenario con ruptura y sin dosis. La Figura 3.2 muestra las trayectorias de y para las primeras ocho unidades; se observan saltos de nivel localizados, consistentes con el mecanismo de ruptura. La simulación y la implementación completa del diseño (ventanas, agregación y bootstrap) se documentan en el Anexo B.4.

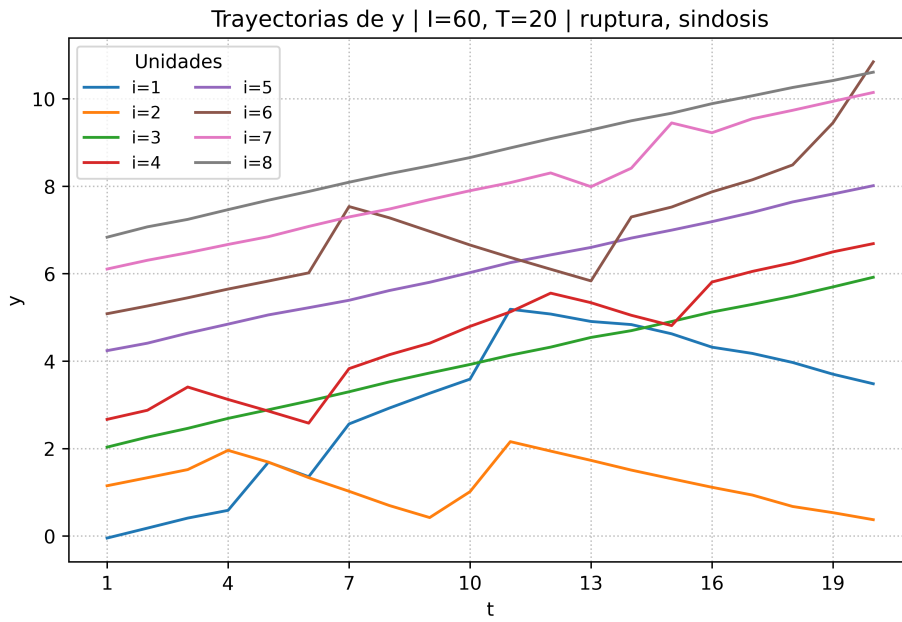


Figura 3.2: Trayectorias de y_t para las primeras ocho unidades (escenario con ruptura, sin dosis).

Resultados por instante k

Cuadro 3.1: Inferencia por instante k (efecto total)

k	$\widehat{ATT}_K^{rup}(k)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	sig
3	-0.6029	0.1280	-0.8384	-0.3437	***
4	-0.4940	0.2344	-0.7863	0.0613	
5	0.5019	0.4569	-0.5449	1.3195	
6	0.1627	0.5286	-0.8878	1.1764	
7	1.5527	0.4454	0.7328	2.4933	***
8	0.9684	0.4047	0.3408	1.7619	***
9	-0.0018	0.3318	-0.6099	0.6348	
10	0.8790	0.4485	0.0494	1.7406	***
11	0.6718	0.4188	-0.1470	1.5142	
12	0.6183	0.5320	-0.4402	1.5714	
13	1.2525	0.3882	0.4078	1.9477	***
14	2.1374	0.3278	1.4154	2.7085	***
15	1.2453	0.3792	0.4953	2.0377	***
16	1.4457	0.3487	0.6706	1.9913	***
17	1.1546	0.4554	0.2705	2.0208	***
18	0.5380	0.2273	0.0176	0.9315	**
19	0.9595	0.3231	0.4196	1.6333	***
20	0.5738	0.1876	0.3598	1.0660	***

Estimación global ponderada: $\widehat{ATT}_{global}^{rup} = 0.7241$, $SE_{boot} = 0.0618$, IC95 % = [0.5662, 0.8048] ***.

Resultados por nivel de intensidad h

Cuadro 3.2: Inferencia por nivel h (efecto total)

h	$\widehat{ATT}_H^{rup}(h)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	sig
1	-0.4617	0.1027	-0.6466	-0.2416	***
2	2.2695	0.1422	1.9307	2.4764	***
3	0.0927	0.1028	-0.0777	0.3294	
4	1.6574	0.1564	1.3247	1.8989	***
5	-0.0743	0.2021	-0.4024	0.3738	
6	2.6139	0.5395	1.6604	3.7762	***

Comparación global: total vs. instantáneo

Cuadro 3.3: Comparación global y lectura

Estimador	Valor
$\widehat{ATT}_{\text{global}}^{\text{rap}}(\text{total})$	0.724063
$\widehat{ATT}_{\text{global}}^{\text{inst}}$	0.876665
$\Delta^{\text{post}}(\text{total} - \text{instantáneo})$	-0.152602

Lectura: cuando la diferencia entre el total y el instantáneo es pequeña, el efecto proviene sobre todo del salto en k (poca ruptura de tendencia posterior). En cambio, diferencias grandes reflejan que la dinámica posterior contribuye de forma importante; si cambian de signo, hay reversión. Las tablas detalladas por (k, h) y la implementación se presentan en el Anexo B.4.

3.2. Comparación tiempo–tiempo con ruptura (promedio entre unidades)

Esta sección estima el mismo efecto marginal que en la comparación unidad–unidad con ruptura (Sección 3.1), pero fijando dos tiempos $t^- < k < t^+$ alrededor de k : uno previo y uno posterior. En el mismo espíritu de la comparación tiempo–tiempo *sin* ruptura (Sección 2.2), se contrastan los cambios de las unidades que en k pasan de $h-1$ a h con los de unidades cuyo H permanece *constante en* $h-1$ en $[t^-, t^+]$. A diferencia del caso sin ruptura, aquí se permite un quiebre de *tendencia* en k : el contrafactual “sin nuevo tratamiento” para las tratadas es la evolución propia del nivel $h-1$ (observada en los controles), mientras que su trayectoria *observada* tras k sigue la tendencia asociada al nivel h . Este estimador puntual por ventana promedia entre unidades; su identificación y notación se presentan a continuación.

3.2.1. Comparaciones válidas centradas en cortes temporales

En coherencia con la Sección 3.1, aquí permitimos *ruptura de tendencia* en el instante k y construimos comparaciones por *ventanas* $[t^-, t^+]$ con $t^- < k < t^+$, promediando *entre unidades*. El objetivo es aislar el salto de $h-1$ a h ocurrido en k mediante una diferencia de cambios entre tratadas y controles dentro de la misma ventana, bajo un supuesto de tendencias paralelas *por tramos* alrededor de k .

Una comparación es válida cuando, en la ventana $[t^-, t^+]$, las unidades tratadas experimentan en k el paso de $h-1$ a h y se encuentran en $h-1$ en t^- y en h en t^+ ; simultáneamente, las unidades de control no presentan salto ni en k ni dentro de la ventana y permanecen *exactamente* en $h-1$ tanto en t^- como en t^+ . No se admiten como controles unidades con niveles $< h-1$ o $\geq h$, ni unidades que cambien de nivel en k .

No se utilizan como controles unidades con niveles $< h-1$ o $\geq h$ dentro de la ventana, ni unidades que cambien de nivel en k .

Esta restricción ancla a ambos grupos en el mismo nivel preintervención ($h-1$) en t^- . Así, el cambio observado de los controles entre t^- y t^+ sirve como trayectoria contrafactual de las tratadas si no hubieran recibido el salto en k , mientras que el cambio de las tratadas incorpora la posible ruptura después de k . La diferencia de estos cambios dentro de la misma ventana

identifica el efecto marginal del paso de $h-1$ a h en k sin mezclar historiales incompatibles.

3.2.2. Definición de los conjuntos relevantes para cada configuración (k, h)

En el diseño *tiempo-tiempo con ruptura* reutilizamos el conjunto de tratadas $\mathcal{I}^{k,h}$ tal como está definido en (2.2.1). El único ajuste respecto de la Sección 2.2 es que, para asegurar comparabilidad con ruptura, los controles deben permanecer *exactamente* en $h-1$ dentro de la ventana $[t^-, t^+]$. Bajo este criterio, los extremos de las ventanas admisibles se definen como:

$$\mathcal{T}_{\text{rup}}^-(k, h) = \left\{ t^- < k \mid \exists i \in \mathcal{I}^{k,h} : H_{i,t^-} = h-1, \exists j : H_{j,t^-} = H_{j,k} = h-1 \right\}. \quad (3.2.1)$$

$$\mathcal{T}_{\text{rup}}^+(k, h \mid t^-) = \left\{ t^+ \geq k \mid \begin{array}{l} \exists j : H_{j,t^-} = H_{j,t^+} = h-1, \\ \exists i \in \mathcal{I}^{k,h} : H_{i,t^-} = h-1, H_{i,t^+} = h \end{array} \right\}. \quad (3.2.2)$$

Con estos extremos, los *controles efectivos* y las *tratadas efectivas* en la ventana $[t^-, t^+]$ quedan dados por:

$$\begin{aligned} \mathcal{J}_{\text{rup}}^{k,h}(t^-, t^+) &= \{j \mid H_{j,t^-} = H_{j,t^+} = h-1\}, \\ \mathcal{I}^{k,h}(t^-, t^+) &= \{i \in \mathcal{I}^{k,h} \mid H_{i,t^-} = h-1, H_{i,t^+} = h\}. \end{aligned} \quad (3.2.3)$$

Nota 3.11. *En comparación con $\mathcal{J}^{k,h}(t^-, t^+)$ de la Sección 2.2, aquí se sustituye la condición $H_{j,\cdot} < h$ por $H_{j,\cdot} = h-1$ dentro de la ventana. Las propiedades de monotonicidad por encaje al ampliar t^+ o desplazar t^- (véase Prop. 2.3) se mantienen de forma análoga y no se repiten.*

3.2.2.1 Definición: Efecto del tratamiento entre los tiempos t^- y t^+ con ruptura

Para cada configuración (k, h) , y para todo par (t^-, t^+) tal que

$$t^- \in \mathcal{T}_{\text{rup}}^-(k, h) \quad \text{y} \quad t^+ \in \mathcal{T}_{\text{rup}}^+(k, h \mid t^-),$$

definimos

$$ATT_{\text{rup}}^{t^-, t^+}(k, h) = \mathbb{E}[Y_{i,t^+}(h) - Y_{i,t^+}(h-1) \mid i \in \mathcal{I}^{k,h}(t^-, t^+)]. \quad (3.2.4)$$

Este parámetro compara, en el mismo tiempo t^+ , el resultado potencial bajo nivel h con el que habría prevalecido bajo $h-1$ para las unidades que saltan en k y alcanzan h . La diferencia frente a la Sección 2.2 radica en la *admisibilidad de la ventana*: aquí viene dada por $\mathcal{T}_{\text{rup}}^-(k, h)$ y $\mathcal{T}_{\text{rup}}^+(k, h | t^-)$ (véanse (3.2.1) y (3.2.2)), que exigen controles en $h-1$ dentro de $[t^-, t^+]$.

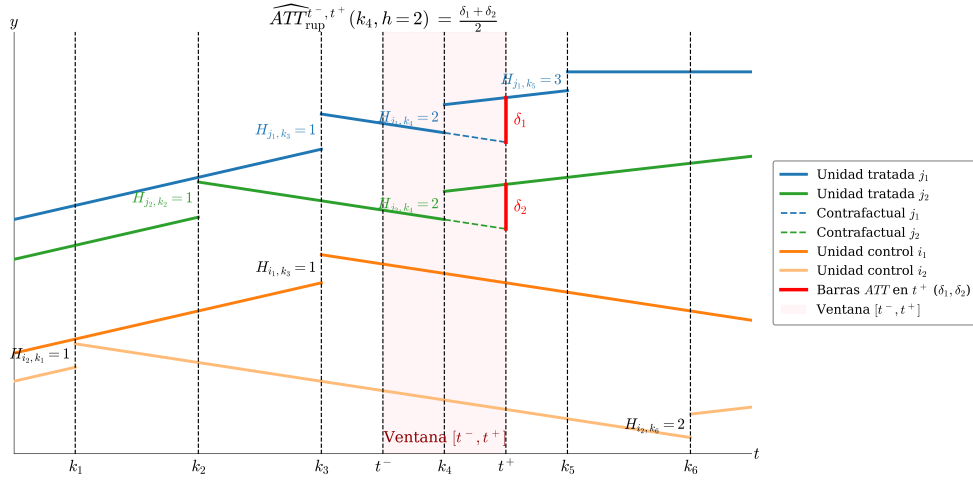


Figura 3.3: Ilustración esquemática de $ATT_{\text{rup}}^{t^-, t^+}(k, h)$ para una ventana $[t^-, t^+]$ alrededor de k en el caso con ruptura.

Nota 3.12. La Figura 3.3 ilustra el caso $(k, h) = (k_4, 2)$ con una ventana fija $[t^-, t^+]$. Se muestran dos unidades tratadas (j_1, j_2) que alcanzan h en k_4 y dos unidades de control (i_1, i_2) que permanecen exactamente en $h-1$ dentro de la ventana. Las líneas punteadas representan, para cada tratada, la trayectoria contrafactual al nivel $h-1$ con la tendencia local; en t^+ se marcan dos segmentos rojos δ_1 y δ_2 . El estimador por ventana mostrado es el promedio entre unidades:

$$\widehat{ATT}_{\text{rup}}^{t^-, t^+}(k_4, 2) = \frac{\delta_1 + \delta_2}{2}.$$

Con $ATT_{\text{rup}}^{t^-, t^+}(k, h)$ ya definido en la Definición 3.2.2.1, la identificación por ventanas requiere un supuesto de *tendencias paralelas locales* formulado para cada (t^-, t^+) , anclado en el nivel $h-1$ (común a tratadas y controles en t^-).

3.2 Supuesto: Tendencias paralelas por ventana (con ruptura)

Para toda configuración (k, h) y todo par (t^-, t^+) tal que $t^- \in \mathcal{T}_{\text{rup}}^-(k, h)$ y $t^+ \in \mathcal{T}_{\text{rup}}^+(k, h | t^-)$, se cumple

$$\begin{aligned} & \mathbb{E}[Y_{i,t^+}(h-1) - Y_{i,t^-}(h-1) | i \in \mathcal{I}^{k,h}(t^-, t^+)] \\ &= \mathbb{E}[Y_{j,t^+}(h-1) - Y_{j,t^-}(h-1) | j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-, t^+)]. \end{aligned} \quad (3.2.5)$$

Dentro de la ventana $[t^-, t^+]$, si no hubiera salto en k , las tratadas seguirían la misma *tendencia* (al nivel $h-1$) que exhiben los controles que permanecen en $h-1$ durante toda la ventana.

Con el parámetro definido en la Definición 3.2.2.1 y bajo el Supuesto 3.2, es posible identificarlo a partir de datos observados. La idea es comparar los cambios de tratadas y controles dentro de la misma ventana, anclados en el nivel $h-1$. La siguiente proposición establece formalmente la identificación.

Proposición 3.2 (Identificación de $ATT_{\text{rup}}^{t^-, t^+}(k, h)$ por ventana). *Bajo el Supuesto 3.2, para toda configuración (k, h) y todo par (t^-, t^+) tal que $t^- \in \mathcal{T}_{\text{rup}}^-(k, h)$ y $t^+ \in \mathcal{T}_{\text{rup}}^+(k, h | t^-)$, el efecto de la Definición 3.2.2.1 se identifica como*

$$\begin{aligned} \widehat{ATT}_{\text{rup}}^{t^-, t^+}(k, h) &= \underbrace{\mathbb{E}[y_{i,t^+} - y_{i,t^-} | i \in \mathcal{I}^{k,h}(t^-, t^+)]}_{\text{cambio observado tratadas}} \\ &\quad - \underbrace{\mathbb{E}[y_{j,t^+} - y_{j,t^-} | j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-, t^+)]}_{\text{cambio observado controles}}. \end{aligned} \quad (3.2.6)$$

De manera equivalente, este mismo efecto puede reescribirse como la diferencia entre las brechas intergrupales en t^+ y en t^- :

$$\begin{aligned} \widehat{ATT}_{\text{rup}}^{t^-, t^+}(k, h) &= \underbrace{\mathbb{E}[y_{i,t^+} - y_{j,t^+} | i \in \mathcal{I}^{k,h}(t^-, t^+), j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-, t^+)]}_{\text{diferencia intergrupal en } t^+} \\ &\quad - \underbrace{\mathbb{E}[y_{i,t^-} - y_{j,t^-} | i \in \mathcal{I}^{k,h}(t^-, t^+), j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-, t^+)]}_{\text{diferencia intergrupal en } t^-}. \end{aligned} \quad (3.2.7)$$

La demostración de la Proposición 3.2 se delega al Anexo A.15, donde se muestran los pasos algebraicos que vinculan la definición y el supuesto con estas expresiones empíricas.

Nota 3.13 (Sobre los chequeos placebo). *En el marco con ruptura no es necesario introducir un placebo posterior, pues la estabilidad de la pendiente después del salto deja de ser un supuesto central. Lo que sí conserva utilidad es el placebo previo, que puede replicarse exactamente como en la Sección 2.2.3 (véase la construcción formal en (2.2.8)), para verificar empíricamente el supuesto de tendencias paralelas locales antes del salto en k .*

Nota 3.14 (Implicaciones prácticas). *Si el placebo previo no valida el paralelismo local en la ventana de referencia, la metodología con ruptura pierde sustento bajo la formulación básica. En ese caso, la alternativa es enriquecer la especificación incorporando covariables u otros mecanismos de control, lo cual se desarrollará más adelante.*

3.2.3. Agregación de efectos y construcción de pesos (caso con ruptura)

Los estimadores $\widehat{ATT}_{\text{rup}}^{t^-, t^+}(k, h)$ capturan el salto asociado al paso hacia el nivel h alrededor del instante k en una ventana específica (t^-, t^+) . Para reportar un *único* efecto representativo del corte (k, h) agregamos esas estimaciones puntuales. Cada comparación (t^-, t^+) recibe un peso que refleje la *información efectiva* que aporta: más observaciones útiles y mejor balance entre tratadas y controles implican, en principio, mayor precisión.

Sea $\mathcal{I}^{k, h}(t^-, t^+)$ el conjunto de tratadas efectivas y $\mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+)$ el de controles efectivos (ambos definidos para el caso con ruptura). Definimos el *peso base* de cada comparación como

$$\omega_{\text{rup}}(t^-, t^+ | k, h) = \frac{|\mathcal{I}^{k, h}(t^-, t^+)| |\mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+)|}{|\mathcal{I}^{k, h}(t^-, t^+)| + |\mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+)|} \quad (3.2.8)$$

el cual recoge (i) el balance entre tratadas y controles y (ii) el tamaño efectivo de la comparación.

Para la agregación, normalizamos los pesos sobre el conjunto de ventanas admisibles en el caso con ruptura, $\mathcal{W}_{\text{rup}}(k, h)$:

$$\widehat{\omega}_{\text{rup}}(t^-, t^+ | k, h) = \frac{\omega_{\text{rup}}(t^-, t^+ | k, h)}{\sum_{(u^-, u^+) \in \mathcal{W}_{\text{rup}}(k, h)} \omega_{\text{rup}}(u^-, u^+ | k, h)}.$$

Con ello, el efecto promedio por corte (k, h) queda

$$\widehat{ATT}_{\text{rup}}(k, h) = \sum_{(t^-, t^+) \in \mathcal{W}_{\text{rup}}(k, h)} \widehat{\omega}_{\text{rup}}(t^-, t^+ | k, h) \cdot \widehat{ATT}_{\text{rup}}^{t^-, t^+}(k, h).$$

Obsérvese que la distancia temporal a k no entra en los pesos: la ventana usa solo (t^-, t^+) y la información relevante ya queda capturada por (3.2.8).

Nota 3.15 (Ventanas admisibles y comparaciones disponibles). *En el caso con ruptura los controles deben permanecer exactamente en $h-1$ dentro de la ventana. Por ello, $\mathcal{W}_{\text{rup}}(k, h)$ puede ser un subconjunto estricto del conjunto usado en la Sección 2.2.4. En la práctica, algunas comparaciones (t^-, t^+) válidas “sin ruptura” dejan de estar disponibles “con ruptura”; si para algún (k, h) no hay ventanas admisibles, $\widehat{ATT}_{\text{rup}}(k, h)$ no se define y se omite del reporte.*

3.2.4. Agregación sobre tiempo k e intensidad h (caso con ruptura)

Con $\widehat{ATT}_{\text{rup}}(k, h)$ ya agregados por ventanas, sintetizamos en dos direcciones naturales (a lo largo del calendario k y de los niveles h), y en un promedio global. Usamos como “pesos” conteos de cuántas unidades efectivamente alcanzan cada nivel h en cada k (cf. $\mathcal{I}^{k, h}$ en (2.2.1)), normalizados para sumar uno:

$$p(k, h) = \frac{|\mathcal{I}^{k, h}|}{\sum_u \sum_v |\mathcal{I}^{u, v}|}, \quad p(k) = \frac{\sum_v |\mathcal{I}^{k, v}|}{\sum_{u, v} |\mathcal{I}^{u, v}|}, \quad p(h) = \frac{\sum_u |\mathcal{I}^{u, h}|}{\sum_{u, v} |\mathcal{I}^{u, v}|},$$

$$p(h | k) = \frac{|\mathcal{I}^{k, h}|}{\sum_v |\mathcal{I}^{k, v}|}, \quad p(k | h) = \frac{|\mathcal{I}^{k, h}|}{\sum_u |\mathcal{I}^{u, h}|}.$$

Promedios marginales:

$$\widehat{ATT}_{K, \text{rup}}(k) = \sum_h p(h | k) \widehat{ATT}_{\text{rup}}(k, h). \quad (3.2.9)$$

$$\widehat{ATT}_{H, \text{rup}}(h) = \sum_k p(k | h) \widehat{ATT}_{\text{rup}}(k, h). \quad (3.2.10)$$

Promedio global:

$$\widehat{ATT}_{\text{global, rup}} = \sum_k \sum_h p(k, h) \widehat{ATT}_{\text{rup}}(k, h). \quad (3.2.11)$$

Nota 3.16. *Las fórmulas de agregación coinciden con las de la Sección 2.2.4; lo que cambia es el dominio de celdas (k, h) disponibles, porque ahora algunas comparaciones por ventana pueden no existir al exigir controles en $h-1$. En consecuencia, los promedios marginales y el global se calculan sobre las celdas con $\widehat{ATT}_{\text{rup}}(k, h)$ definidas.*

Nota 3.17 (Invarianza del orden). *Como $p(k)p(h | k) = p(k, h) = p(h)p(k | h)$ por construcción, se mantiene*

$$\sum_k p(k) \widehat{ATT}_{K,\text{rup}}(k) = \sum_h p(h) \widehat{ATT}_{H,\text{rup}}(h) = \widehat{ATT}_{\text{global},\text{rup}}.$$

La comprobación detallada —análogo al caso de la Sección 2.2.4— se presenta en el Anexo A.12.

3.2.5. Análisis de inferencia para los ATT por ventanas (t^-, t^+) en (k, h) (caso con ruptura)

El esquema de inferencia es análogo al de la Sección 2.2.6: agregación lineal por ventanas con pesos plug-in, varianzas obtenidas por *bootstrap* por unidades (re-muestreo de trayectorias completas), y construcción de intervalos de confianza y promedios marginales (k) , (h) y global. La única diferencia es el *dominio* de comparación, pues aquí las ventanas admisibles exigen controles que permanecen en $h-1$ (véase §3.2.3). En particular, denotando por $\widehat{ATT}_{\text{rup}}^{t^-, t^+}(k, h)$ las estimaciones puntuales, se aplica la misma lógica de agregación y remuestreo descrita en §2.2.6, pero sobre el conjunto $\mathcal{W}_{\text{rup}}(k, h)$ y con $\mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+)$.

Nota 3.18. *Esta subsección no incorpora covariables; el análisis con covariables y sus implicaciones para la inferencia se presenta más adelante.*

3.2.6. Aplicación a datos simulados (con ruptura)

Aplicamos el diseño *por cortes temporales con ruptura de tendencia* sobre el mismo panel simulado de 60 unidades y 20 periodos descrito en el Anexo B (§B.1). La Figura 2.3 (presentada en la Sección 2.1.6) ilustra la evolución de y para las primeras ocho unidades y es también representativa del escenario aquí considerado. La implementación detallada (ventanas, conjuntos efectivos y placebos) se documenta en el Anexo B.5.

Resultados por instante k

De acuerdo con (3.2.9), los promedios por instante k se obtienen con $p(h | k)$. La tabla reporta punto, SE_{boot} , intervalo al 95 % (bootstrap por unidades).

Cuadro 3.4: Inferencia por instante k (diseño 3.2, con ruptura)

k	$\widehat{ATT}_{K,\text{rup}}(k)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)
3	-0.902761	0.122365	-1.006180	-0.535394
4	-0.889165	0.248656	-0.982363	-0.018508
5	0.686220	0.263281	0.113457	1.073718
6	0.209948	0.418760	-0.863511	1.372141
7	1.948508	0.304174	1.528092	2.784463
8	0.941323	0.301190	0.261557	1.619498
9	-0.126252	0.230476	-0.333502	0.354157
10	1.181488	0.436221	0.210262	1.823722
11	0.231810	0.306272	-0.375718	0.856761
12	0.556348	0.445342	-0.357263	1.279464
13	1.250582	0.312780	0.449252	1.589705
14	2.491059	0.337920	1.698276	3.032584
15	1.374593	0.420374	0.978997	2.431135
16	1.707728	0.335222	0.871643	2.141617
17	1.256912	0.449663	0.201874	2.264239

Resultados por nivel de intensidad h

Según (3.2.10), los promedios por nivel h usan $p(k | h)$. Incluimos el peso marginal $p(h)$ como w_h (en %).

Cuadro 3.5: Inferencia por nivel h (diseño 3.2, con ruptura)

h	$\widehat{ATT}_{H,\text{rup}}(h)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	w_h (%)
1	-0.714385	0.086805	-0.760509	-0.406389	27.12
2	2.649489	0.152390	2.160188	2.741677	25.99
3	-0.116961	0.132112	-0.241649	0.263201	21.47
4	1.580645	0.164279	1.367685	1.985410	15.25
5	-0.204491	0.227381	-0.496251	0.372400	7.91
6	2.745644	0.514096	1.657555	3.645682	2.26

Global y notas de validación

El promedio global, conforme a (3.2.11), es

$$\widehat{ATT}_{\text{global, rup}} = 0.756715, \quad SE_{\text{boot}} = 0.069937, \quad \text{IC95 \%} = [0.605288, 0.868130].$$

Cuadro 3.6: Consistencia de agregación (vía k y vía h)

Promedio global (vía k)	0.756715
Promedio global (vía h)	0.756715

Nota 3.19. (1) *Los diagnósticos placebo PRE (874) muestran alta estabilidad relativa: VERDE en 83.5 %, AMARILLO en 14.6 % y ROJO en 1.8 % (véase Anexo B.5).*

(2) *La invarianza del orden de agregación se verifica numéricamente: $\sum_k p(k) \widehat{ATT}_{K, \text{rup}}(k) = \sum_h p(h) \widehat{ATT}_{H, \text{rup}}(h) = 0.756715$.*

3.2.7. Extensión con covariables

Hasta aquí hemos desarrollado la identificación del efecto $ATT_{\text{rup}}^{t^-, t^+}(k, h)$ bajo el supuesto de *tendencias paralelas no condicionales* (Supuesto 3.2). Dicho supuesto establece que, en ausencia de la intervención, las unidades tratadas y de control habrían seguido en promedio la misma evolución dentro de la ventana $[t^-, t^+]$.

En la práctica, esta hipótesis puede ser demasiado restrictiva, pues la evolución de los resultados suele depender de características observables que difieren entre grupos. Para relajar esta exigencia introducimos un marco de *tendencias paralelas condicionales*, donde los resultados se comparan dentro de subgrupos homogéneos definidos por un vector de covariables X . De este modo, las covariables actúan como un mecanismo de ajuste que mejora la validez de las comparaciones, sin alterar el objeto de interés: el efecto promedio $ATT_{\text{rup}}^{t^-, t^+}(k, h)$.

3.3 Supuesto: Tendencias paralelas condicionales por ventana (con ruptura)

Sea X_i el vector de covariables observables asociado a la unidad i . Para toda configuración (k, h) y todo par (t^-, t^+) tal que $t^- \in \mathcal{T}_{\text{rup}}^-(k, h)$ y $t^+ \in \mathcal{T}_{\text{rup}}^+(k, h | t^-)$, se cumple que

$$\begin{aligned} & \mathbb{E}[Y_{i, t^+}(h-1) - Y_{i, t^-}(h-1) \mid i \in \mathcal{I}^{k, h}(t^-, t^+), X_i = x] \\ &= \mathbb{E}[Y_{j, t^+}(h-1) - Y_{j, t^-}(h-1) \mid j \in \mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+), X_j = x], \end{aligned} \quad (3.2.12)$$

para todo valor posible x del vector de covariables.

Nota 3.20. *El vector X_i representa las características observables de la unidad i . La expresión “ $X_i = x$ ” debe entenderse como que la comparación se realiza entre unidades que comparten el mismo valor de x .*

Bajo este supuesto, el contrafactual de las tratadas se identifica a partir de controles con covariables comparables en la misma ventana. El objeto de interés sigue siendo el efecto de la Definición 3.2.2.1, pero su validez descansa ahora en el paralelismo condicional.

Proposición 3.3 (Identificación condicional por covariables). *Bajo el Supuesto 3.3, para toda configuración (k, h) y toda ventana admisible (t^-, t^+) ,*

el efecto condicional en $X = x$ se identifica como

$$\begin{aligned} ATT_{\text{rup}}^{t^-, t^+}(k, h | x) &= \mathbb{E}[y_{i, t^+} - y_{i, t^-} \mid i \in \mathcal{I}^{k, h}(t^-, t^+), X_i = x] \\ &\quad - \mathbb{E}[y_{j, t^+} - y_{j, t^-} \mid j \in \mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+), X_j = x]. \end{aligned} \quad (3.2.13)$$

La demostración es idéntica a la de la Proposición 3.2, con la única diferencia de que todas las comparaciones se realizan dentro de los estratos homogéneos definidos por X . En consecuencia, los conjuntos de tratadas y controles efectivos no son los mismos de la ventana completa, sino los subconjuntos que comparten el valor $X = x$, de modo que el paralelismo se evalúa únicamente entre unidades comparables en covariables.

El objeto de interés, sin embargo, sigue siendo el efecto de la Definición 3.2.2.1, es decir, el ATT sin condicionar en X . Dicho efecto puede representarse como un promedio ponderado de los efectos condicionales anteriores.

Proposición 3.4 (Descomposición del ATT como promedio de efectos condicionales). *Para toda configuración (k, h) y ventana admisible (t^-, t^+) , se cumple*

$$ATT_{\text{rup}}^{t^-, t^+}(k, h) = \sum_x ATT_{\text{rup}}^{t^-, t^+}(k, h | x) \Pr(X = x \mid i \in \mathcal{I}^{k, h}(t^-, t^+)), \quad (3.2.14)$$

cuando X toma un número finito de valores.

Nota 3.21 (ATT sobre el soporte común). *La descomposición del ATT en la Proposición 3.4 es válida únicamente si existe soporte común en las covariables. Esto significa que, para cada valor x considerado en la ventana, debe cumplirse que*

$$\Pr(i \in \mathcal{I}^{k, h}(t^-, t^+) \mid X = x) > 0 \quad \text{y} \quad \Pr(j \in \mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+) \mid X = x) > 0. \quad (3.2.15)$$

En ausencia de tratadas o controles en algún estrato x , el efecto condicional correspondiente no está definido y, en consecuencia, el ATT original de la Proposición 3.4 tampoco resulta plenamente identificable.

En tales casos, lo que se identifica y se reporta en la práctica es el ATT sobre el soporte común, definido como

$$ATT_{\text{rup}, SC}^{t^-, t^+}(k, h) = \sum_{x \in \mathcal{S}_{k, h}(t^-, t^+)} ATT_{\text{rup}}^{t^-, t^+}(k, h | x) \Pr(X = x \mid i \in \mathcal{I}^{k, h}(t^-, t^+)).$$

Aquí, $\mathcal{S}_{k,h}(t^-, t^+)$ denota el conjunto de valores de X para los cuales existen tanto tratadas efectivas como controles efectivos.

En el desarrollo teórico que sigue se mantiene el análisis sobre el $ATT_{\text{rup}}^{t^-, t^+}(k, h)$ completo, pero debe tenerse presente que en las aplicaciones empíricas el estimando efectivo corresponde a su versión restringida al soporte común.

Nota 3.22. Si el vector de covariables X es continuo, la expresión (3.2.14) se generaliza a una integral de Riemann–Stieltjes:

$$ATT_{\text{rup}}^{t^-, t^+}(k, h) = \int ATT_{\text{rup}}^{t^-, t^+}(k, h | x) dF_{X|\mathcal{I}^{k,h}(t^-, t^+)}(x),$$

donde $F_{X|\mathcal{I}^{k,h}(t^-, t^+)}$ es la distribución de X entre las unidades tratadas efectivas en la ventana.

3.2.8. Tres formas equivalentes de identificar el efecto

Para cada ventana $[t^-, t^+]$ y configuración (k, h) , definimos el indicador de pertenencia al grupo tratado:

$$D_u^{k,h,t^-, t^+} = \mathbf{1}\{u \in \mathcal{I}^{k,h}(t^-, t^+)\}, \quad u \in \mathcal{I}^{k,h}(t^-, t^+) \cup \mathcal{J}_{\text{rup}}^{k,h}(t^-, t^+). \quad (3.2.16)$$

En otras palabras, $D_u^{k,h,t^-, t^+} = 1$ si la unidad u es tratada efectiva en la ventana, y $D_u^{k,h,t^-, t^+} = 0$ si pertenece al grupo de controles efectivos.

El *puntaje de propensión* específico de la ventana se define como la función

$$p_{k,h,t^-, t^+}(x) = \Pr(D^{k,h,t^-, t^+} = 1 | X = x), \quad (3.2.17)$$

que asigna a cada valor posible de covariables x la probabilidad de recibir el salto de $h-1$ a h en el instante k dentro de la ventana. En la práctica, este puntaje se evalúa para cada unidad u en su vector de covariables observado X_u , es decir, $p_{k,h,t^-, t^+}(X_u)$.

Definimos además la *evolución esperada de los controles* como la función

$$m_{k,h,t^-, t^+}^{\text{ctrl}}(x) = \mathbb{E}\left[y_{j,t^+} - y_{j,t^-} \mid j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-, t^+), X_j = x\right], \quad (3.2.18)$$

que asigna a cada vector de covariables x el valor esperado del cambio en el resultado de los controles efectivos dentro de la ventana $[t^-, t^+]$. En la práctica, se evalúa en el vector observado de cada unidad u , es decir $m_{k,h,t^-, t^+}^{\text{ctrl}}(X_u)$.

Con estas definiciones, el efecto $ATT_{\text{rup}}^{t^-,t^+}(k, h)$ admite tres representaciones algebraicamente equivalentes bajo el Supuesto 3.3 y la condición de solapamiento

$$0 < p_{k,h,t^-,t^+}(X) < 1,$$

la cual es simplemente la forma resumida del requisito de soporte común establecido en (3.2.15). Cada representación enfatiza un modelo de primer paso distinto: regresión de resultados (OR), reponderación (IPW) o una combinación de ambos (DR). Esta tríada ha sido ampliamente estudiada en la literatura sobre estimadores semiparamétricos (Abadie (2005), Bonhomme y Sauder (2011), Graham et al. (2012), Sant'Anna y Zhao (2020) y J. Wooldridge (2007)).

1. Representación OR (Outcome Regression).

$$ATT_{\text{rup-OR}}^{t^-,t^+}(k, h) = \mathbb{E} \left[(y_{i,t^+} - y_{i,t^-}) - m_{k,h,t^-,t^+}^{\text{ctrl}}(X_i) \mid i \in \mathcal{I}^{k,h}(t^-, t^+) \right]. \quad (3.2.19)$$

En esta representación, el contrafactual de cada tratada se obtiene imputando la evolución esperada de los controles con las mismas covariables. El estimador OR consiste, por tanto, en comparar el cambio observado en las tratadas con la predicción generada a partir de un modelo de resultados estimado únicamente sobre los controles efectivos de la ventana. Su validez descansa en que dicho modelo esté correctamente especificado y capture las diferencias relevantes en X .

2. Representación IPW (Inverse Probability Weighting).

$$ATT_{\text{rup-IPW}}^{t^-,t^+}(k, h) = \mathbb{E} \left[\frac{D_u^{k,h,t^-,t^+}}{\mathbb{E}[D_u^{k,h,t^-,t^+}]} (y_{u,t^+} - y_{u,t^-}) \right] - \mathbb{E} \left[\frac{(1 - D_u^{k,h,t^-,t^+}) \frac{p_{k,h,t^-,t^+}(X_u)}{1 - p_{k,h,t^-,t^+}(X_u)}}{\mathbb{E} \left[(1 - D_u^{k,h,t^-,t^+}) \frac{p_{k,h,t^-,t^+}(X_u)}{1 - p_{k,h,t^-,t^+}(X_u)} \right]} (y_{u,t^+} - y_{u,t^-}) \right]. \quad (3.2.20)$$

En esta representación, el contrafactual se construye mediante *reponderación por probabilidades inversas*. La idea es asignar a cada unidad de control un peso proporcional a $\frac{p(X_u)}{1-p(X_u)}$, de forma que la distribución de covariables de los controles *replique* la de las tratadas. El primer término promedia el cambio

observado en las tratadas (normalizado por su tamaño esperado), mientras que el segundo corrige el sesgo utilizando los controles reponderados. La validez de este enfoque depende críticamente de que el modelo del puntaje $p(\cdot)$ esté bien especificado y de que exista soporte común ($0 < p(X_u) < 1$ en el dominio relevante).

3. Representación DR (Doubly Robust). La representación *doblemente robusta (DR)* combina las estrategias de regresión de resultados (OR) y de reponderación por puntajes (IPW). Su principal ventaja es que garantiza consistencia si *al menos uno* de los dos modelos de primer paso está correctamente especificado: el modelo de resultados de controles $m^{\text{ctrl}}(\cdot)$ o el modelo del puntaje de propensión $p(\cdot)$. Esto la hace menos sensible a errores de modelación y especialmente útil en contextos con múltiples periodos, heterogeneidad del tratamiento o dinámicas complejas, tal como destacan Callaway y Sant’Anna (2021). En términos prácticos, el estimador DR corrige simultáneamente la sensibilidad de OR a una mala especificación funcional y la de IPW a valores extremos del puntaje, lo cual mejora el desempeño finito y la estabilidad de los estimadores. A lo largo de esta sección, las covariables X se entienden como *pre-salto* (no afectadas por el cambio en k).

$$\begin{aligned}
ATT_{\text{rup-DR}}^{t^-, t^+}(k, h) = & \mathbb{E} \left[\frac{D_u^{k, h, t^-, t^+}}{\mathbb{E}[D_u^{k, h, t^-, t^+}]} \left(y_{u, t^+} - y_{u, t^-} - m_{k, h, t^-, t^+}^{\text{ctrl}}(X_u) \right) \right] \\
& - \mathbb{E} \left[\frac{(1 - D_u^{k, h, t^-, t^+}) \frac{p_{k, h, t^-, t^+}(X_u)}{1 - p_{k, h, t^-, t^+}(X_u)}}{\mathbb{E} \left[(1 - D_u^{k, h, t^-, t^+}) \frac{p_{k, h, t^-, t^+}(X_u)}{1 - p_{k, h, t^-, t^+}(X_u)} \right]} \left(y_{u, t^+} - y_{u, t^-} - m_{k, h, t^-, t^+}^{\text{ctrl}}(X_u) \right) \right].
\end{aligned} \tag{3.2.21}$$

Nota 3.23. Forma muestral (implementación práctica). Sea $n_T = |\mathcal{I}^{k, h}(t^-, t^+)|$. Para cada control $j \in \mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+)$, defina el peso

$$a_j = \frac{\hat{p}_{k, h, t^-, t^+}(X_j)}{1 - \hat{p}_{k, h, t^-, t^+}(X_j)}.$$

Entonces,

$$\begin{aligned} \widehat{ATT}_{\text{rup-DR}}^{t^-, t^+}(k, h) &= \frac{1}{n_T} \sum_{i \in \mathcal{I}^{k, h}(t^-, t^+)} \left(y_{i, t^+} - y_{i, t^-} - \hat{m}_{k, h, t^-, t^+}^{\text{ctrl}}(X_i) \right) \\ &\quad - \frac{1}{\sum_{j \in \mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+)} a_j} \sum_{j \in \mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+)} a_j \left(y_{j, t^+} - y_{j, t^-} - \hat{m}_{k, h, t^-, t^+}^{\text{ctrl}}(X_j) \right). \end{aligned} \quad (3.2.22)$$

Aquí $\hat{p}_{k, h, t^-, t^+}(\cdot)$ es el puntaje estimado (por logit, probit, ML o métodos penalizados), y $\hat{m}_{k, h, t^-, t^+}^{\text{ctrl}}(\cdot)$ es el modelo de resultados estimado únicamente con controles válidos. Las normalizaciones n_T y $\sum a_j$ garantizan que cada término promedie sobre su grupo. En la práctica, cuando el puntaje estimado se aproxima demasiado a los extremos ($\hat{p}_{k, h, t^-, t^+}(X_u) \approx 0$ o $\hat{p}_{k, h, t^-, t^+}(X_u) \approx 1$), se considera que falla el solapamiento y el correspondiente $\widehat{ATT}_{\text{rup-DR}}^{t^-, t^+}(k, h)$ no se reporta. Las ventanas afectadas se excluyen de los promedios agregados y los pesos se renormalizan sobre el dominio restante.

Nota 3.24. Caso ideal con X discreta y positividad completa. En el caso ideal donde las covariables X toman un número finito de valores y se cumple positividad por celda (es decir, existen tratadas y controles efectivos para cada valor $x \in \mathcal{X}$), la implementación del estimador DR se simplifica considerablemente. Suponga entonces que X toma un número finito de valores \mathcal{X} y existe positividad por celda: para todo $x \in \mathcal{X}$,

$$n_T(x) \equiv |\mathcal{I}^{k, h}(t^-, t^+) \cap \{u : X_u = x\}| > 0, \quad n_C(x) \equiv |\mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+) \cap \{u : X_u = x\}| > 0.$$

En tal caso, el puntaje condicional puede aproximarse por frecuencias:

$$\tilde{p}(x) = \frac{n_T(x)}{n_T(x) + n_C(x)}, \quad \tilde{a}(x) = \frac{\tilde{p}(x)}{1 - \tilde{p}(x)} = \frac{n_T(x)}{n_C(x)}.$$

Y el modelo de resultados en controles se reduce al promedio por celda:

$$\tilde{m}^{\text{ctrl}}(x) = \frac{1}{n_C(x)} \sum_{j \in \mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+) \cap \{X_j = x\}} (y_{j, t^+} - y_{j, t^-}).$$

Con estas definiciones, la versión muestral DR se simplifica a

$$\widehat{ATT}_{\text{rup-DR}}^{t^-, t^+}(k, h) = \frac{1}{n_T} \sum_{i \in \mathcal{I}^{k, h}(t^-, t^+)} \left(y_{i, t^+} - y_{i, t^-} - \tilde{m}^{\text{ctrl}}(X_i) \right), \quad (3.2.23)$$

pues dentro de cada celda el segundo término se anula por construcción. En este escenario ideal, la representación DR colapsa a una OR estratificada por X , ya que el ajuste por propensiones se vuelve exacto y la comparación se realiza dentro de cada celda homogénea.

Las expresiones anteriores muestran que el efecto promedio del tratamiento puede obtenerse a partir de tres enfoques conceptualmente distintos pero coherentes entre sí: la imputación de resultados esperados (OR), la reponderación por puntajes (IPW) y la combinación de ambos (DR). Cada uno depende de un modelo de primer paso diferente, pero todos buscan identificar el mismo objeto causal. Establecer su equivalencia formal permite confirmar que las tres estrategias son versiones algebraicamente consistentes del mismo estimando y, por tanto, pueden usarse de manera intercambiable bajo los supuestos adecuados.

Proposición 3.5 (Equivalencia de las representaciones OR, IPW y DR). *Bajo el Supuesto 3.3 y la condición de solapamiento $0 < p_{k,h,t^-,t^+}(X) < 1$, el efecto medio del tratamiento definido en la Definición 3.2.2.1 es algebraicamente equivalente a sus tres representaciones condicionales:*

$$ATT_{\text{rup}}^{t^-,t^+}(k, h) = ATT_{\text{rup-OR}}^{t^-,t^+}(k, h) = ATT_{\text{rup-IPW}}^{t^-,t^+}(k, h) = ATT_{\text{rup-DR}}^{t^-,t^+}(k, h).$$

La demostración se basa en las identidades de esperanza condicional que vinculan la imputación de resultados ($m^{\text{ctrl}}(\cdot)$) con la reponderación por probabilidades inversas ($p(\cdot)$), y se presenta en el Anexo A.16.

3.2.9. Análisis de inferencia

A diferencia del caso sin covariables desarrollado en la Sección 2.2.6, el análisis de inferencia aquí no puede basarse en las mismas herramientas. Allí los estimadores se construían como simples promedios de diferencias entre tratadas y controles dentro de cada ventana, lo que permitía derivar propiedades asintóticas directas a partir de la ley de los grandes números y la normalidad muestral. En ese contexto, bastaba con evaluar la dispersión de los $\widehat{ATT}^{t^-,t^+}(k, h)$ mediante varianzas plug-in o bootstrap por unidades.

En el presente caso la situación es distinta. Al incorporar covariables X en la identificación del efecto y permitir representaciones alternativas (OR, IPW y DR), el estimador deja de ser lineal en los datos y pasa a depender de modelos

de primer paso: el modelo de resultados $m^{\text{ctrl}}(\cdot)$ y el modelo del puntaje de propensión $p(\cdot)$. Esto introduce nuevas fuentes de variabilidad —errores de predicción y de estimación del puntaje— que afectan la varianza total del estimador. Por tanto, las fórmulas analíticas usadas en la Sección 2.2.6 ya no resultan válidas, y la inferencia debe seguir un enfoque más general, como el propuesto por Callaway y Sant’Anna (2021), basado en funciones de influencia y remuestreo agrupado por unidad.

La inferencia para los estimadores con covariables se construye a partir de su representación como *funcionales de influencia lineal* (IFL). Cada representación (OR, IPW o DR) puede escribirse como un promedio de una función $\psi(W_i; \theta)$, donde $W_i = (Y_{i,t^-}, Y_{i,t^+}, D_i, X_i)$ es la información de la unidad y θ el conjunto de parámetros poblacionales. La validez de la inferencia asintótica depende de que el estimador tenga una expansión del tipo:

$$\sqrt{N}(\widehat{ATT} - ATT) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \psi(W_i) + o_p(1), \quad (3.2.24)$$

con $\mathbb{E}[\psi(W_i)] = 0$ y $\text{Var}[\psi(W_i)] < \infty$. Bajo condiciones regulares, el promedio muestral de $\psi(W_i)$ satisface un teorema del límite central, lo que garantiza normalidad asintótica.

Teorema 3.4 (Normalidad asintótica de los estimadores OR–IPW–DR). *Supóngase que los supuestos de identificación (3.3 y condición de solapamiento) se cumplen y que los estimadores de los modelos de primer paso $m^{\text{ctrl}}(\cdot)$ y $p(\cdot)$ son consistentes y suficientemente suaves. Entonces, para cada ventana válida (t^-, t^+) y configuración (k, h) ,*

$$\sqrt{N} \left(\widehat{ATT}_{\text{rup}}^{t^-, t^+}(k, h) - ATT_{\text{rup}}^{t^-, t^+}(k, h) \right) \xrightarrow{d} \mathcal{N}(0, \sigma_{k,h}^2),$$

donde $\sigma_{k,h}^2 = \text{Var}(\psi_{k,h}(W_i))$ y $\psi_{k,h}(\cdot)$ es la función de influencia asociada a la representación utilizada (OR, IPW o DR). La demostración formal se presenta en el Anexo A.17.1.

Teorema 3.5 (Consistencia del estimador DR). *El estimador doblemente robusto $\widehat{ATT}_{\text{rup-DR}}^{t^-, t^+}(k, h)$ es consistente para $ATT_{\text{rup}}^{t^-, t^+}(k, h)$ si se cumple al menos una de las dos condiciones siguientes:*

- (1) *El modelo de resultados $m_{k,h,t^-,t^+}^{\text{ctrl}}(X)$ está correctamente especificado.*

(2) El modelo del puntaje $p_{k,h,t^-,t^+}(X)$ está correctamente especificado.

Si ambos modelos están bien especificados, el estimador DR alcanza la varianza mínima asintótica entre los estimadores semiparamétricos equivalentes.

Teorema 3.6 (Inferencia por remuestreo agrupado). *Bajo los mismos supuestos del Teorema 3.4, la distribución asintótica de $\widehat{ATT}_{\text{rup-DR}}^{t^-,t^+}(k, h)$ puede ser aproximada mediante un remuestreo bootstrap por unidades, definido como*

$$\widehat{\sigma}_{k,h}^{*2} = \text{Var}^* \left(\widehat{ATT}_{\text{rup-DR}}^{t^-,t^+*}(k, h) \right),$$

donde $\text{Var}^*(\cdot)$ denota la varianza empírica bajo re-muestreo y \widehat{ATT}^{t^-,t^+*} se calcula sobre las unidades seleccionadas con reemplazo. Este método preserva la dependencia temporal dentro de cada unidad y proporciona intervalos de confianza válidos para los agregados $\widehat{ATT}(k, h)$, $\widehat{ATT}_K(k)$, $\widehat{ATT}_H(h)$ y $\widehat{ATT}_{\text{global}}$.

Demostración. La demostración formal del Teorema 3.6, incluyendo la justificación de la validez del remuestreo agrupado en presencia de dependencias temporales intra-unidad, se presenta en el Anexo A.17.2. \square

En síntesis, la inferencia en el caso con covariables sigue el mismo esquema conceptual de las secciones anteriores —normalidad asintótica y varianza obtenida por bootstrap agrupado—, pero se extiende al marco semiparamétrico en el cual los componentes OR e IPW se estiman en pasos previos. El estimador DR es el más robusto y eficiente entre ellos, por lo que será el que se implemente en la aplicación empírica posterior.

Capítulo 4

Aplicación

Para ilustrar la metodología presentada en esta tesis se implementa una aplicación empírica sobre el efecto de la des-regulación bancaria en la participación del ingreso laboral dentro del valor agregado a nivel estatal en Estados Unidos, tomando como artículo de referencia a Leblebicioğlu y Weinberger (2020).

Hacia finales de los años setenta, varios estados iniciaron la eliminación de barreras regulatorias que, por un lado, impedían a los bancos estatales operar en otros estados y, por otro, limitaban su expansión mediante nuevas sucursales. Al primer hito se le conoce como des-regulación bancaria interestatal, que habilitó a los bancos estatales a operar fuera de su estado de origen; al segundo, des-regulación bancaria intraestatal, que permitió la apertura de nuevas sucursales dentro del estado.

Leblebicioğlu y Weinberger (2020) analizan el efecto de estos dos hitos sobre la participación laboral del valor agregado (variable de interés) y encuentran que la des-regulación inter reduce dicha participación, mientras que para la des-regulación intra no hallan efectos significativos en sus especificaciones. Su estimación principal usa una regresión de efectos fijos en dos vías (TWFE). Sin embargo, trabajos recientes como de Chaisemartin y D’Haultfoeuille (2020), Sant’Anna y Zhao (2020), Goodman-Bacon (2021), Sun y Abraham (2021), Borusyak et al. (2024) muestran que, con adopción escalonada, TWFE puede sufrir pesos negativos y comparaciones “sucias” entre tratados en distintos momentos, sesgando los promedios.

En términos de magnitud, Leblebicioğlu y Weinberger (2020) estiman que la des-regulación inter reduce la participación laboral en torno a 0.8–1.2 puntos porcentuales. El mecanismo propuesto es un abaratamiento del crédito

vía mayor competencia bancaria que impulsa inversión y productividad: el valor agregado crece con más fuerza que la compensación a asalariados y, por tanto, el cociente cae. Esta explicación viene dada en gran medida por la forma en que se construye la variable de interés. La participación laboral es definida como el ratio entre las compensación de los trabajadores, es decir la suma entre salarios y los apartes del empleador a seguridad social y el valor agregado estatal o PIB estatal, medido como la suma entre la compensación de los trabajadores mas impuestos a la producción menos los subsidios. Los datos son tomados de la Bureau of Economic Analysis (BEA por sus siglas en inglés) una agencia del departamento de comercio de Estados Unidos y la ventana de tiempo de análisis se define desde 1970 hasta 1996 en 48 estados.

En contraste con Leblebicioğlu y Weinberger (2020), donde inter e intra se modelan como intervenciones distintas, en esta tesis tratamos la des-regulación bancaria como una única política cuyo mecanismo subyacente es común: la liberalización de barreras a la expansión y entrada bancaria y, por tanto, a la competencia financiera. Este enfoque es coherente con nuestro marco de **misma política**, cada hito de des-regulación añade una misma unidad de exposición (h), de modo que el paso de $h = 0$ a $h = 1$ captura el primer alivio regulatorio y el paso de $h = 1$ a $h = 2$ captura un incremento de la misma política, no una política esencialmente diferente. Para mantener una interpretación nítida de los niveles de exposición y, a la vez, comparar implícitamente los márgenes intra e inter, restringimos la muestra a los estados que primero des-regularon intraestatal y luego interestatal; con esta restricción, $h = 1$ se alinea con el “tratamiento intraestatal” y $h = 2$ con el “tratamiento interestatal” sobre una base homogénea (el mismo estado avanzando en la misma política). Además, esta elección tiene una ventaja estadística: el orden “intra→inter” abarca 29 estados (frente a 19 en “inter→intra”), abarcando mayor comparación de unidades. Con esta restricción de orden, se pueden comparar de forma implícita ambos hitos de liberación bancaria: el primer nivel de intervención capta el efecto de intra, mientras que el segundo nivel capta el efecto incremental de inter sobre una unidad que ya recibió intra. Como se mostró en la estructura metodológica, las agregaciones son plausibles en los diferentes niveles de intervención, lo cual permite que se capture el efecto intra e inter aislados, manteniendo una heterogeneidad de intervención para poder comparar los resultados propios con los encontrados por Leblebicioğlu y Weinberger (2020).

Por último, mientras Leblebicioğlu y Weinberger (2020) incluyen covariables (crecimiento del PIB estatal, población, desempleo, etc.), la metodología

propuesta en esta tesis prescinde de controles para enfatizar la identificación a partir de comparaciones limpias y la estructura de exposición común. Por tanto, se reconoce que el uso de las mismas podrían cambiar las estimaciones realizadas.

Con lo anterior, los resultados a partir de las estimaciones se muestran a continuación.

4.1. Comparación unidad–unidad, sin ruptura

En este conjunto de datos, todos los estados fueron eventualmente tratados, de modo que el grupo de control de nunca tratados (\mathcal{I}_0) resultó vacío. En consecuencia, los efectos se estimaron exclusivamente entre estados tratados, comparando aquellos que incrementaron su nivel de tratamiento en cada año k frente a otros que aún permanecían en niveles inferiores.

El procedimiento siguió la formulación teórica de la Sección 2.1. Para cada ventana temporal admisible, se calcularon efectos marginales $\widehat{ATT}_{m,n}(k)$, ponderados por sus respectivos pesos. Los resultados se agregaron primero por instante de tratamiento (k) y luego por nivel acumulado de tratamiento (h).

Cuadro 4.1: Efectos promedio agregados $\widehat{ATT}(k)$

k	Año	$\widehat{ATT}(k)$	SE	IC _{95%}	Sig.
6	1975	0.2380	0.0000	[0.2380, 0.2380]	***
7	1976	-2.8114	0.0000	[-2.8114, -2.8114]	***
8	1977	-3.0384	0.0000	[-3.0384, -3.0384]	***
9	1978	-5.4894	0.2940	[-5.9161, -5.0995]	***
10	1979	-0.9849	0.0000	[-0.9849, -0.9849]	***
11	1980	-12.5816	0.0000	[-12.5816, -12.5816]	***
12	1981	-20.3726	4.1402	[-27.5156, -15.9894]	***
13	1982	-6.1666	7.7175	[-14.5847, 6.5073]	
14	1983	-25.0444	1.0672	[-26.6133, -23.5610]	***
15	1984	-6.8010	4.2093	[-8.9791, 2.2066]	
16	1985	-5.9923	4.1354	[-12.7856, 2.8912]	
17	1986	-10.5442	2.4513	[-16.3156, -6.4213]	***
18	1987	-10.3320	6.4792	[-19.4855, 4.1918]	
19	1988	-12.5745	3.8961	[-17.6650, -3.0734]	**
21	1990	-4.2429	0.0000	[-4.2429, -4.2429]	***
22	1991	-3.6839	0.0000	[-3.6839, -3.6839]	***
23	1992	5.5641	0.0000	[5.5641, 5.5641]	***

Notas: Los errores estándar (SE) se obtienen mediante bootstrap con re-muestreo por unidad tratada (999 réplicas). En algunos años el SE aparece como cero porque sólo existía una unidad tratada efectiva en ese periodo. Al no haber más unidades con las cuales generar variabilidad, todas las réplicas del bootstrap producen el mismo valor del estimador, por lo que el algoritmo no puede calcular una desviación estándar distinta de cero. Estos casos deben interpretarse con cautela, ya que reflejan escasa información disponible en esa ventana temporal.

Cuadro 4.2: Efecto global (Agregación por k)

Estimador	Valor	SE	IC _{95%}	Sig.
$\widehat{ATT}_{\text{global}}$	-8.6101	1.3060	[-10.65, -5.68]	***

Cuadro 4.3: Efectos promedio agregados $\widehat{ATT}^{[h]}$

h	$\widehat{ATT}^{[h]}$	SE	IC _{95%}	Sig.
1	-7.7938	2.4407	[-12.98, -3.47]	***
2	-8.1822	2.3558	[-12.23, -3.25]	***

Cuadro 4.4: Efecto global (Agregación por h)

Estimador	Valor	SE	IC _{95%}	Sig.
$\widehat{ATT}_{\text{global}}^{[h]}$	-8.0610	1.7918	[-11.06, -4.11]	***

Cuadro 4.5: Comparación global entre agregaciones

Agregación	Estimador	Diferencia porcentual (DPS _{global})
Por k	-8.6101	6.59 %
Por h	-8.0610	6.59 %

Los resultados muestran un efecto negativo y significativo de la desregulación bancaria sobre la participación laboral, especialmente entre 1981 y 1983, años que concentran los mayores efectos de $\widehat{ATT}(k)$. El promedio global por instantes (-8.61) y por niveles (-8.06) son coherentes, con una diferencia relativa de solo 6.6 %, lo que sugiere una alta estabilidad interna del estimador frente a la forma de agregación.

En la formulación teórica se planteó un contraste adicional entre el efecto promedio sobre tratadas y nunca tratadas (ATT°) y el efecto promedio entre tratadas (ATT^+). Sin embargo, dado que todos los estados del panel recibieron tratamiento en algún momento, no fue posible estimar el componente ATT° . El análisis empírico corresponde, por tanto, únicamente al componente ATT^+ , cuyo resultado global fue consistente bajo ambas formas de agregación (k y h). Dados los resultados de la agregación por nivel de intervención h se observa que para $h = 1$ el efecto es de -7.79 puntos porcentuales mientras que para $h = 2$ es de -8.18. Si bien, se concibe el manejo de la intervención como

misma política, también se reconoce que gracias a la construcción del orden de intervención los dos hitos de des-regulación tienen efectos significativos.

4.2. Comparación tiempo–tiempo, sin ruptura

Siguiendo la Sección 2.2. Para cada ventana $[t^-, t^+]$, se construyeron los estimadores $\widehat{ATT}^{t^-, t^+}(k, h)$ y se agregaron mediante pesos proporcionales de las comparaciones. A continuación se presentan los resultados

Cuadro 4.6: Efectos promedio agregados $\widehat{ATT}(k)$

k	Año	$\widehat{ATT}(k)$	SE	IC _{95%}	Sig.
6	1975	0.1061	1.0672	[-2.1520, 2.0726]	
7	1976	-3.4934	1.3327	[-5.9680, -0.6793]	**
8	1977	-5.5965	2.3963	[-10.1387, -0.7391]	**
9	1978	-8.1670	2.7407	[-13.0058, -2.0675]	***
10	1979	-1.5274	2.5451	[-6.3996, 3.6107]	
11	1980	-13.1839	2.7121	[-18.1149, -7.8017]	***
12	1981	-21.1685	5.4356	[-32.3915, -12.4842]	***
13	1982	-6.1960	9.2653	[-22.8021, 10.6676]	
14	1983	-27.6993	3.6444	[-33.3063, -19.5148]	***
15	1984	-11.5482	8.6135	[-20.4265, 10.8193]	
16	1985	-9.0016	6.8158	[-18.2622, 8.1809]	
17	1986	-13.0457	7.0633	[-21.9222, 4.2062]	
18	1987	-13.8422	8.8282	[-19.8722, 13.7141]	
19	1988	-17.1002	11.4129	[-24.0329, 10.0137]	
21	1990	-2.3202	1.8679	[-5.6645, -1.0595]	***

Cuadro 4.7: Efecto global (Agregación por k)

Estimador	Valor	SE	Sig.
$\widehat{ATT}_{\text{global}}$	-10.9214	4.8070	*

Cuadro 4.8: Efectos promedio agregados $\widehat{ATT}^{[h]}$

h	$\widehat{ATT}^{[h]}$	SE	IC _{95%}	Sig.
1	-9.9145	2.4110	[-14.6684, -4.5877]	***
2	-11.5464	6.6089	[-16.8193, 5.8442]	

Cuadro 4.9: Comparación global entre agregaciones

Agregación	Estimador	Diferencia porcentual (DPS _{global})
Por k	-10.9214	0.00 %
Por h	-10.9214	0.00 %

Los resultados muestran efectos negativos y significativos de la des-regulación bancaria sobre la participación laboral. Al igual que los resultados anteriores, en los años 1981 y 1983 se concentran los mayores efectos. Por otra parte, dadas las agregaciones tanto en k como en h el efecto es negativo representando una disminución de 10.92 puntos porcentuales de participación laboral debido a los hitos de la des-regulación bancaria.

A diferencia de las estimaciones presentadas en la sección anterior, donde se observó que cada hito tuvo efectos estadísticamente significativos, bajo esta familia de estimadores se evidencia que únicamente el primer hito resulta significativo, con una disminución estimada de -9.91 puntos porcentuales. En contraste, el segundo hito, aunque mantiene el signo esperado del coeficiente, no presenta significancia estadística en ningún nivel convencional.

Cheques placebo por ventana y consistencia de la estimación

Cuadro 4.10: Resumen de cheques placebo (%)

	ROJO (>25 %)	AMARILLO (10–25 %)	VERDE (≤ 10 %)
PRE	58.0	22.7	19.3
POST	75.9	16.6	7.5

Los placebos *PRE* (anteriores a t^-) evalúan si las unidades tratadas y de control presentaban trayectorias paralelas antes de la intervención. Aunque un 58 % de las ventanas aparece en rojo, la distribución entre categorías indica que la hipótesis de paralelismo local no se viola de forma generalizada, siendo defendible en un contexto empírico con datos reales.

En contraste, los placebos *POST* (posteriores a t^+) muestran un 76 % de ventanas en rojo, lo que sugiere que las trayectorias tienden a divergir inmediatamente después de las intervenciones. Este resultado indica que el supuesto de tendencia común se debilita tras el cambio de política, lo cual apunta a la necesidad de un enfoque que permita **rupturas de pendiente** o cambios estructurales en la dinámica, como los que se desarrollan en las Secciones 4.3 y 4.4.

Cabe mencionar que una posible mejora de los chequeos *PRE* consiste en incorporar covariables de control que atenúen la heterogeneidad estructural entre grupos, es decir, las diferencias observables que pueden explicar divergencias iniciales en las trayectorias. De esta forma, el contraste se realizaría bajo el supuesto de *tendencias paralelas condicionales*, reduciendo la influencia de factores preexistentes ajenos al tratamiento.

4.3. Comparación unidad–unidad, con ruptura

Siguiendo la metodología establecida en la sección 3.1 y definiendo ventanas temporales admisibles en la cual se puedan realizar las comparaciones válidas; se construyeron efectos marginales y se agregaron por instante k y por nivel acumulado h . Los resultados se presentan a continuación.

Cuadro 4.11: Efectos promedio agregados $\widehat{ATT}_K^{\text{rup}}(k)$

k	Año	\widehat{ATT}	SE	IC _{95%}	Sig.
6	1975	0.2380	0.9792	[-1.6171, 1.8271]	
7	1976	-2.8114	1.4124	[-5.2417, -0.2277]	**
8	1977	-3.0384	2.3063	[-6.5617, 1.2676]	
9	1978	-3.5039	2.6732	[-8.1713, 2.1304]	
10	1979	-0.9849	2.5804	[-5.4537, 4.5496]	
11	1980	-12.5816	2.5905	[-17.4489, -8.0421]	***
12	1981	-20.3726	5.1820	[-30.5178, -12.4384]	***
13	1982	-0.9598	5.9736	[-10.7367, 11.3590]	
14	1983	-21.7847	3.9278	[-29.8843, -14.5355]	***
15	1984	1.7293	4.9582	[-7.4499, 11.1912]	
16	1985	0.5066	5.7949	[-8.9763, 14.7974]	
17	1986	-2.0885	5.2112	[-10.7343, 10.3185]	
18	1987	-3.6032	6.2414	[-11.7944, 12.8650]	
19	1988	5.1162	3.2489	[-1.5779, 10.1401]	
21	1990	-4.2429	1.9295	[-5.8347, -1.0595]	***
22	1991	-3.6839	0.8615	[-4.3888, -2.2741]	***
23	1992	5.5641	0.0000	[5.5641, 5.5641]	***

Cuadro 4.12: Efecto global (Agregación por k)

Estimador	Valor	SE	IC _{95%}	Sig.
$\widehat{ATT}_{\text{global}}^{\text{rup}}$	-3.5261	2.3575	[-7.176, 1.451]	

Cuadro 4.13: Efectos promedio agregado $\widehat{ATT}_H^{\text{rup}}(h)$

h	\widehat{ATT}	SE	IC _{95%}	Sig.
1	-9.5351	2.5328	[-14.1006, -4.4652]	***
2	0.8797	3.3335	[-4.7549, 7.5331]	

Los resultados muestran efectos negativos pronunciados alrededor de los episodios de ruptura (especialmente entre 1980 y 1983), con significancia estadística en varios instantes k . El agregado global es negativo, aunque con un intervalo que incluye el cero. Por niveles, el efecto es claramente negativo en $h = 1$ y no significativo en $h = 2$, lo que sugiere que el primer episodio concentra la mayor parte del impacto.

Efecto instantáneo versus efecto total

Se contrastaron los resultados del efecto instantáneo, estimado en el momento de la ruptura, con el efecto total acumulado posterior. El efecto instantáneo captura el cambio inmediato en la variable de resultado, en este caso la participación laboral, mientras que el efecto total incorpora la persistencia o corrección de tendencia en los periodos siguientes.

Efectos instantáneos y totales estimados

	$\widehat{ATT}_{\text{instantáneo}}$	$\widehat{ATT}_{\text{total}}$
Global	-1.94	-3.53
Nivel $h = 1$	-4.69	-9.53
Nivel $h = 2$	0.07	0.88

A nivel global, el impacto total (-3.53) duplica aproximadamente el efecto inmediato (-1.94), indicando que una fracción importante del ajuste ocurre de forma inmediata tras la ruptura, y el resto se consolida con el cambio de pendiente posterior. Por niveles, el tratamiento inicial ($h = 1$) concentra casi todo el efecto total, mientras que los niveles posteriores presentan variaciones pequeñas y no significativas. En conjunto, los resultados confirman que el mayor impacto se produce en el momento del quiebre, con ajustes marginales en los periodos subsiguientes.

4.4. Comparación tiempo–tiempo, con ruptura

Por último, y siguiendo la metodología establecida en la sección 3.2, se presentan a continuación los resultados empíricos del diseño *tiempo–tiempo* con ruptura, en los que se estiman efectos promedio por instante k , por nivel acumulado h y de forma global.

Cuadro 4.14: Inferencia por año $t = k$: efectos promedio $\widehat{ATT}_K^{\text{rup}}(k)$

Año	$\widehat{ATT}_K(k)$	SE	IC _{95%} (lo)	IC _{95%} (hi)	p_{boot}	Sig.
1975	0.1061	1.0672	-2.1520	2.0726	0.9119	
1976	-3.4934	1.3327	-5.9680	-0.6793	0.0187	**
1977	-5.5965	2.3963	-10.1387	-0.7391	0.0321	**
1978	-4.3556	3.4912	-10.3545	2.9547	0.2380	
1979	-1.5274	2.5451	-6.3996	3.6107	0.4740	
1980	-13.1839	2.7121	-18.1149	-7.8017	0.0000	***
1981	-21.1685	5.4356	-32.3915	-12.4842	0.0000	***
1982	-0.8454	6.1329	-12.2659	10.6676	0.8445	
1983	-22.5441	3.9862	-31.0124	-15.0557	0.0000	***
1984	0.9164	6.8231	-9.0164	14.6569	0.7385	
1985	0.6856	6.0461	-8.6340	13.7404	0.8016	
1986	-0.5654	6.0169	-9.2669	12.0544	0.8970	
1987	-5.0354	6.4399	-11.7051	13.8208	0.5732	
1988	5.4600	3.5382	-3.2895	10.3264	0.2012	
1989	-2.3202	1.8679	-5.6645	-1.0595	0.0000	***

Cuadro 4.15: Inferencia global (ponderación $p(k, h)$, diseño con ruptura)

Estimador	Valor	SE (bootstrap)	p_{boot}
$\widehat{ATT}_{\text{global}}^{\text{rup}}$	-2.6031	2.6818	0.5852

Los resultados muestran nuevamente un patrón claro de efectos negativos concentrados entre 1980 y 1983, periodo en el que las diferencias entre tratadas y controles son estadísticamente significativas y de mayor magnitud. Este comportamiento coincide con el escenario de ruptura de tendencia introducido

en el diseño, donde los ajustes de pendiente posterior refuerzan el impacto inicial del cambio estructural.

Cuadro 4.16: Inferencia por nivel acumulado h : $\widehat{ATT}^{\text{rup},[h]}$

Nivel h	$\widehat{ATT}^{[h]}$	SE	IC _{95%} (lo)	IC _{95%} (hi)	p_{boot}	Sig.
1	-9.9145	2.4110	-14.6684	-4.5877	0.0000	***
2	1.9350	3.7863	-4.1025	10.4359	0.4609	

Los resultados sugieren que el primer hito de des-regulación ($h = 1$) concentra el efecto más importante del proceso de ruptura, mientras que el segundo hito presenta variaciones menores y no significativas. El impacto promedio global es negativo, aunque con intervalo que incluye el cero, lo cual indica que la pérdida de tendencia observada en los episodios iniciales no se compensa plenamente con los ajustes subsiguientes.

4.5. Síntesis comparativa de resultados entre diseños

En esta sección se presentan de manera sintética los resultados obtenidos en las cuatro aplicaciones empíricas desarrolladas en las secciones 4.1, 4.2, 4.3 y 4.4. Cada una corresponde a un diseño de comparación distinto —unidad–unidad o tiempo–tiempo— y a un supuesto diferente sobre la existencia o no de ruptura de tendencia en el resultado. El objetivo de esta síntesis es contrastar los valores estimados de los efectos promedio globales y por niveles acumulados (h), mostrando su consistencia y variación entre escenarios.

1. Comparaciones sin ruptura (4.1 y 4.2)

En los diseños sin ruptura, las pendientes de las trayectorias tratadas y de control se consideran paralelas antes y después de las intervenciones. Los resultados de ambas estrategias (unidad–unidad y tiempo–tiempo) son consistentes en magnitud y signo, con efectos negativos y estadísticamente significativos en los estimadores globales y en los niveles de intervención.

Cuadro 4.17: Efectos globales estimados en diseños sin ruptura

Diseño	Agregación	Estimador	Valor	Sig.
Unidad–unidad (4.1)	Global (k)	$\widehat{ATT}_{\text{global}}$	-8.6101	***
Unidad–unidad (4.1)	Global (h)	$\widehat{ATT}_{\text{global}}^{[h]}$	-8.0610	***
Tiempo–tiempo (4.2)	Global (k)	$\widehat{ATT}_{\text{global}}$	-10.9214	*

Cuadro 4.18: Efectos promedio por nivel acumulado en diseños sin ruptura

Diseño	Nivel h	$\widehat{ATT}^{[h]}$	Sig.
Unidad–unidad (4.1)	$h = 1$	-7.7938	***
Unidad–unidad (4.1)	$h = 2$	-8.1822	***
Tiempo–tiempo (4.2)	$h = 1$	-9.9145	***
Tiempo–tiempo (4.2)	$h = 2$	-11.5464	

En conjunto, ambos métodos sin ruptura producen estimaciones similares, confirmando que la elección del tipo de comparación (unidad–unidad o tiempo–tiempo) no altera sustancialmente la magnitud del efecto global. No obstante, los chequeos placebo POST evidenciaron indicios de pérdida de paralelismo, lo cual motivó la extensión hacia los escenarios con ruptura.

2. Comparaciones con ruptura (4.3 y 4.4)

Los diseños con ruptura permiten que las pendientes de las trayectorias cambien tras la intervención, capturando tanto el efecto instantáneo como los ajustes posteriores. Los resultados confirman una reducción en la magnitud del efecto global respecto a los casos sin ruptura, mientras que los efectos por nivel muestran que la mayor parte del impacto se concentra en el primer episodio de intervención ($h = 1$).

Cuadro 4.19: Efectos globales estimados en diseños con ruptura

Diseño	Agregación	Estimador	Valor	Sig.
Unidad–unidad (4.3)	Global (k)	$\widehat{ATT}_{\text{global}}^{\text{rup}}$	-3.5261	
Tiempo–tiempo (4.4)	Global (k, h)	$\widehat{ATT}_{\text{global}}^{\text{rup}}$	-2.6031	

Cuadro 4.20: Efectos promedio por nivel acumulado en diseños con ruptura

Diseño	Nivel h	$\widehat{ATT}^{[h]}$	Sig.
Unidad–unidad (4.3)	$h = 1$	-9.5351	***
Unidad–unidad (4.3)	$h = 2$	0.8797	
Tiempo–tiempo (4.4)	$h = 1$	-9.9145	***
Tiempo–tiempo (4.4)	$h = 2$	1.9350	

Finalmente, dentro del caso con ruptura se diferenció además entre el *efecto instantáneo* y el *efecto total acumulado*, obtenidos en la comparación unidad–unidad, mostrando la siguiente relación:

Cuadro 4.21: Efectos instantáneos y totales estimados (4.3)

	$\widehat{ATT}_{\text{instantáneo}}$	$\widehat{ATT}_{\text{total}}$
Global	-1.94	-3.53
Nivel $h = 1$	-4.69	-9.53
Nivel $h = 2$	0.07	0.88

Los valores de los diseños con ruptura son cercanos entre sí y coherentes en su estructura: el primer nivel de intervención concentra el efecto principal, mientras que los niveles posteriores no resultan significativos. Los estimadores globales, aunque más pequeños en magnitud que en los modelos sin ruptura, reflejan una medición más precisa del cambio de pendiente asociado al evento de intervención.

4.6. Análisis empírico comparativo con la literatura previa

Los modelos con ruptura de tendencia implementados en esta tesis muestran un efecto negativo pronunciado en el nivel inicial de intervención, es decir, en el primer hito de des-regulación. En particular, el efecto estimado acumulado para $h = 1$ ronda -9.5 puntos porcentuales, estadísticamente significativo, mientras que para $h = 2$ el coeficiente es cercano a cero y no significativo. Este patrón sugiere que el primer hito regulatorio contribuye la mayor parte de la disminución en la participación laboral. Los estimadores globales con ruptura muestran magnitudes menores que los modelos sin ruptura, pero reflejan con mayor precisión el cambio de pendiente inducido por la reforma. En todo caso, la dirección del efecto negativo es consistente y robusta a las especificaciones utilizadas.

Estos hallazgos están encaminados con los resultados reportados en Leblebicioğlu y Weinberger (2020). En ese estudio se estima que la liberalización interestatal de la banca redujo la participación laboral en aproximadamente 0.8 puntos porcentuales. Leblebicioğlu y Weinberger (2020) separan por su parte los impactos de la liberalización intraestatal e interestatal, encontrando un efecto significativo sólo para esta última y coeficientes prácticamente nulos para la desregulación intraestatal. La diferencia en magnitud (-0.8 vs. ≈ -9.5) puede atribuirse en buena medida al uso de controles y a la

especificación metodológica distinta en aquel trabajo. Sin embargo, ambos estudios coinciden en que la des-regulación bancaria tiende a desplazar el ingreso hacia el capital (es decir, disminuye el *labour share*).

Metodológicamente, y como ya se mencionó, la literatura reciente muestra que los estimadores TWFE convencionales pueden ser problemáticos en contextos de adopción escalonada. En efecto, el regresor TWFE tradicional equivale a un promedio ponderado de múltiples comparaciones 2×2 entre estados tratados en diferentes años, algunas de las cuales pueden involucrar unidades ya tratadas y generar pesos negativos. Este sesgo implícito podría afectar la estimación original de Leblebicioğlu y Weinberger (2020). Estudios posteriores con estimadores más robustos (por ejemplo, LP-DiD de Dube et al. (2023)) reexaminan el mismo caso: su aplicación confirma el efecto negativo de la liberalización interestatal e, incluso, muestran que la liberalización intraestatal pasa a mostrar un impacto negativo de magnitud similar cuando se excluyen las comparaciones “sucias”. En otras palabras, al corregir los problemas de TWFE los dos hitos regulatorios resultan ambos con efectos negativos para la participación laboral.

Otro aspecto metodológico relevante es la forma de modelar la acumulación de políticas. Mientras los estudios previos tratan la liberalización intraestatal e interestatal por separado, esta tesis las interpreta como fases de una misma política acumulativa. Para ello se restringe la muestra a los estados que primero liberalizaron intraestatal y luego interestatal, de modo que $h = 1$ corresponde al efecto de la des-regulación intraestatal inicial y $h = 2$ al efecto incremental de la des-regulación interestatal sobre un mismo estado ya tratado. Esta dinámica permite comparar ambos efectos en un único marco estimativo, conservando la heterogeneidad de secuencia observada en los datos (29 estados con secuencia intra→inter). De este modo, la tesis captura el impacto de cada hito por niveles de exposición acumulada, en lugar de estimar regresiones separadas, facilitando la comparación directa con los resultados de Leblebicioğlu y Weinberger (2020).

Finalmente, conviene destacar la ausencia de controles adicionales en las estimaciones de esta tesis. A diferencia de Leblebicioğlu y Weinberger (2020), que incluyen covariables como crecimiento del PIB estatal, población, desempleo, entre otras, la metodología aquí aplicada enfatiza la identificación a partir de comparaciones limpias y la estructura de tratamiento común. La literatura señala que omitir controles exógenos tiende a aumentar la magnitud puntual del coeficiente estimado, lo que explica en parte el mayor efecto

observado (-9.5 vs. -0.8) bajo esta especificación. No obstante, el signo negativo persiste sin controles, lo que refuerza la consistencia del hallazgo empírico: la des-regulación bancaria reduce la participación laboral en la economía.

Capítulo 5

Conclusiones

La presente tesis amplía de manera sustantiva el marco metodológico de la metodología más usada para evaluar el efecto de intervenciones, *Diferencia en Diferencias* (DiD), al demostrar que este enfoque puede extenderse con rigurosidad teórica a contextos donde las unidades enfrentan intervenciones repetidas, rupturas de tendencia e intensidades variables de tratamiento. Mientras la literatura reciente se ha centrado en corregir las limitaciones del estimador *two-way fixed effects* (TWFE) bajo adopción escalonada, este trabajo reconfigura el problema desde su raíz, redefine las comparaciones válidas, formaliza los contrafactuales entre unidades con distintos historiales de intervención y propone nuevos estimadores que preservan la interpretación causal aun cuando el tratamiento se aplica de forma secuencial.

Los capítulos metodológicos desarrollados en esta tesis responden de forma precisa a los objetivos específicos formulados en la introducción. El Capítulo 2 aborda el caso base de tratamientos repetidos que inducen saltos de nivel sin alterar la tendencia, y formaliza una estrategia identificadora para estimar el efecto marginal de cada intervención dentro de este escenario (cumpliendo el *Objetivo Específico 1*). El Capítulo 3 extiende dicho enfoque para capturar no sólo cambios discretos en el nivel, sino también modificaciones sostenidas en la pendiente del resultado —rupturas de trayectoria—, permitiendo descomponer con claridad el impacto instantáneo y la variación dinámica posterior a cada evento (correspondiente al *Objetivo Específico 2*). De este modo, la tesis despliega un marco general capaz de recuperar efectos causales bajo múltiples formas de repetición, intensidad y dinámica, alineado rigurosamente con los objetivos planteados.

El aporte no radica en ajustar modelos preexistentes, sino en construir un

marco unificado para estimar efectos acumulativos del tratamiento en escenarios donde los métodos tradicionales no pueden llegar. La tesis desarrolla estrategias complementarias —comparaciones unidad–unidad y tiempo–tiempo— que permiten distinguir entre cambios de nivel y de tendencia, e incorporan de forma opcional covariables mediante representaciones equivalentes (OR, IPW y DR). Este uso coincide con la lógica propuesta por Sant’Anna y Zhao (2020), quienes desarrollan estimadores DiD doblemente robustos apoyados en estas mismas representaciones. El presente trabajo extiende dicha estructura al contexto de intervenciones acumulativas, garantizando consistencia aun si sólo uno de los modelos auxiliares está correctamente especificado.

En esta tesis, el conjunto de contrafactuales válidos se define de manera específica para cada escenario analizado, siguiendo una estructura común en todas las secciones. En primer lugar, se plantea con precisión el *efecto puntual del tratamiento* que se busca identificar; en segundo lugar, se formula el supuesto de tendencias paralelas apropiado para el contexto —ya sea entre unidades o entre periodos de tiempo—; en tercer lugar, se construyen los contrafactuales válidos a partir de dicho supuesto; y, finalmente, se establece la condición de identificación formal. Este procedimiento se aplica sistemáticamente a los cuatro diseños desarrollados (2.1, 2.2, 3.1, 3.2), que difieren en la presencia o ausencia de ruptura de tendencia, en el tipo de comparación (unidad–unidad o tiempo–tiempo) y en la naturaleza discreta o acumulativa de los tratamientos. Así, el método DiD se generaliza a una amplia gama de configuraciones, sin restringirse a un único tipo de contrafactual ni de estructura de tratamiento.

Los cuatro diseños desarrollados en la tesis ilustran la amplitud y consistencia interna del marco general. En los escenarios sin ruptura (2.1 y 2.2), las comparaciones se centran en identificar efectos de nivel bajo supuestos estables de tendencia paralela entre unidades o entre periodos de tiempo. En los casos con ruptura (3.1 y 3.2), el método se amplía para capturar tanto los cambios de nivel como las variaciones en la tendencia posterior a cada intervención, manteniendo la validez causal dentro de cada tramo. Estos escenarios muestran que el marco propuesto no constituye un modelo único, sino una familia coherente de diseños DiD aplicables a distintos patrones de intervención y estructuras de datos.

Los chequeos y extensiones son específicos de cada diseño. En la Sección 2.1 (sin ruptura, comparación unidad–unidad), se estiman dos versiones del ATT: una frente a contrafactual “puro” (nunca tratadas en ese instante específico) y otra entre tratadas. La proximidad de ambos sirve como verificación interna

de que los datos se comportan como un caso sin ruptura; si divergen de manera sistemática, ello sugiere que el supuesto de evolución común podría no sostenerse en esa ventana.

En la Sección 2.2 (sin ruptura, comparación tiempo–tiempo), además del esquema por ventanas, se implementan placebos pre para evaluar homogeneidad previa y, de manera adicional, placebos post; este último sólo tiene sentido en el escenario sin ruptura y, bajo un ajuste adecuado del diseño, debe arrojar estimaciones cercanas a cero, funcionando como un chequeo complementario de la plausibilidad del supuesto.

En la Sección 3.1 (con ruptura, comparación unidad–unidad), al admitirse cambios de trayectoria después de la intervención, se fija la estructura temporal para separar un efecto instantáneo (salto en el momento del tratamiento) y un efecto dinámico o rezagado (ajuste posterior). Esa descomposición sólo es pertinente en este diseño con ruptura y permite interpretar por separado el impacto inmediato y la modificación de la tendencia.

En la Sección 3.2 (con ruptura, comparación por cohortes con tiempos fijados), no se introduce un chequeo análogo a los anteriores; aquí el valor añadido es la incorporación formal de covariables para la identificación e inferencia mediante representaciones equivalentes (OR, IPW y DR).

Las metodologías propuestas tienen aplicaciones directas en estudios de impacto donde las unidades enfrentan múltiples programas o fases de intervención. En la práctica, al pasar del desarrollo teórico a los datos reales no se sabe cuál de los diseños propuestos se ajusta mejor al comportamiento del fenómeno. En algunos casos la política puede generar cambios de nivel sin alterar la tendencia, mientras que en otros produce rupturas en la misma. Por esta razón, el marco planteado no desarrolla un único diseño, sino que permite aplicar cada uno de los cinco y utilizar los chequeos internos incorporados en ellos para diagnosticar la estructura de los datos.

Cada diseño ofrece su propia señal de validación: en la Sección 2.1, la comparación entre el ATT con contrafactual puro y el ATT entre tratadas permite verificar si los datos corresponden a un escenario sin ruptura; en la 2.2, los placebos pre y post sirven para confirmar la estabilidad de las trayectorias; en la 3.1, la descomposición entre efecto instantáneo y dinámico revela si existe ruptura en la tendencia; en la 3.2, la inclusión de covariables muestra si la heterogeneidad previa es relevante. De esta manera, el propio proceso de estimación se convierte en un mecanismo de diagnóstico que orienta cuál de los modelos refleja mejor la realidad observada.

Al igual que los desarrollos recientes de la literatura, esta tesis también

logra articular una unidad conceptual entre los distintos diseños de Diferencia en Diferencias. Todos los escenarios planteados —sin ruptura y con ruptura— comparten una misma estructura analítica basada en comparaciones válidas entre unidades y periodos, lo que permite expresar los efectos causales como combinaciones coherentes de promedios ponderados en el tiempo.

A diferencia de los enfoques tradicionales centrados en estimadores TWFE, el esquema propuesto define explícitamente las condiciones bajo las cuales cada comparación es válida y las integra dentro de un sistema común de identificación. De esta manera, los cinco diseños no representan versiones aisladas, sino expresiones particulares de una misma familia metodológica capaz de adaptarse a distintos patrones de intervención. En ese sentido, la tesis ofrece una extensión natural del enfoque grupo-tiempo desarrollado por Callaway y Sant’Anna (2021) y Goodman-Bacon (2021), trasladándolo a escenarios con tratamientos repetidos.

El marco de identificación sigue exigiendo supuestos fuertes de paralelismo extendido. Es decir, aún requerimos que, en ausencia de cada tratamiento relevante, los grupos (tratados vs. “no tratados por ninguna intervención”) evolucionarían bajo la misma tendencia. En la medida en que existan choques comunes no observados o interacciones complejas entre políticas, el supuesto puede fallar, lo cual limita la validez de los resultados.

Para estimar los múltiples efectos de manera confiable se necesita un panel suficientemente grande en número de unidades y periodos, donde existan grupos con las distintas combinaciones de tratamiento. Si sólo unas pocas unidades reciben ciertas combinaciones, la estimación puede ser inestable. Asimismo, se requiere información clara sobre el tiempo en cual se realiza cada tratamiento o intervención.

La tesis abre varias rutas para continuar el desarrollo de métodos en contextos complejos: Extensión a tratamientos continuos. Como apuntamos, sería útil generalizar el marco para manejar tratamientos de intensidad continua o con varias dimensiones. Esto permitiría abordar, por ejemplo, efectos lineales o no lineales en montos de subsidio, dosis de exposición ambiental, u otras variables cuantitativas.

Integración más amplia de covariables. Incorporar covariables no sólo como controles adicionales sino dentro de un esquema más complejo, cuya finalidad sea capturar de manera verídica el efecto.

Fusión con métodos de aprendizaje automático. Otro camino es explorar estimadores de machine learning (redes neuronales, bosques aleatorios) para predecir contrafactuales heterogéneos en setups de múltiples tratamientos.

Aplicación en redes, spillovers o contextos dinámicos complejos. Finalmente, una gran extensión sería considerar efectos de tratamientos que se propagan entre unidades (spillovers) o efectos temporales retardados más allá de los modelados aquí. También se podría generalizar para estudiar mecanismos causales en redes sociales o económicas donde el tratamiento de un nodo afecta a sus vecinos. Esto sería una valiosa expansión de la investigación.

El marco desarrollado en esta tesis se sitúa en la misma línea de evolución metodológica que los modelos recientes de Diferencia en Diferencias, pero aborda un ámbito complementario al de Tsai (2024), pues mientras allí se amplía el análisis hacia tratamientos simultáneos o de distinta naturaleza, el presente trabajo se concentra en una dimensión temporal dentro de una misma política de intervención, donde las unidades pueden recibir intervenciones repetidas, escalonadas o de intensidad creciente. En ese sentido, ambos enfoques comparten la búsqueda de una formulación general del DiD que preserve la interpretación causal bajo estructuras complejas de tratamiento.

En sintonía con los aportes recientes de Dube et al. (2023), el enfoque desarrollado en esta tesis también privilegia la transparencia, la claridad y la posibilidad de replicación frente a la complejidad de las dinámicas de intervención. Al igual que su aproximación de proyecciones locales aplicadas al DiD, el presente marco resalta la definición explícita de grupos de tratamiento y control, evita los problemas de ponderación espuria y mantiene la trazabilidad de los efectos estimados. Estas convergencias metodológicas muestran que las nuevas generaciones de modelos DiD —incluido el diseño secuencial aquí propuesto— avanzan hacia un enfoque de búsqueda de metodologías que le hagan frente a las dinámicas de las intervenciones.

Por su parte, de Chaisemartin y D’Haultfoeuille (2023) examinan los riesgos de interpretar regresiones TWFE en presencia de múltiples tratamientos diferentes, mostrando que los coeficientes combinan efectos cruzados entre intervenciones. En contraste, y como se ha venido mencionando, esta tesis se concentra en una única política aplicada secuencialmente a lo largo del tiempo, lo que permite construir contrafactuales específicos para cada repetición y mantener una interpretación causal clara en cada etapa. Así, el marco propuesto aborda un vacío no cubierto por los diseños de adopción escalonada o cohortes múltiples —como los de Goodman-Bacon (2021) y Callaway y Sant’Anna (2021)— al ofrecer herramientas para estudiar impactos acumulativos en contextos donde la intervención se repite o intensifica dentro de las mismas unidades.

Bibliografía

- Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The review of economic studies*, *72*(1), 1-19.
- Angrist, J. D., & Fernández-Val, I. (2013). ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework. En D. Acemoglu, M. Arellano & E. Deikel (Eds.), *Advances in Economics and Econometrics: Tenth World Congress* (pp. 401-434). Cambridge University Press.
- Angrist, J. D., & Krueger, A. B. (1999). Empirical Strategies in Labor Economics. En O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (pp. 1277-1366, Vol. 3). Elsevier. [https://doi.org/10.1016/S1573-4463\(99\)03004-7](https://doi.org/10.1016/S1573-4463(99)03004-7)
- Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, *24*(2), 3-30. <https://doi.org/10.1257/jep.24.2.3>
- Athey, S., & Imbens, G. W. (2022). Design-based analysis in Difference-In-Differences settings with staggered adoption. *Journal of Econometrics*, *226*(1), 62-79. <https://doi.org/10.1016/j.jeconom.2020.10.012>
- Bellego, C., Benatia, D., & Dortet-Bernadet, V. (2025). The Chained Difference-in-Differences. *Journal of Econometrics*, *248*(1), 105783. <https://doi.org/10.1016/j.jeconom.2024.105783>
- Bonhomme, S., & Sauder, U. (2011). Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling. *Review of Economics and Statistics*, *93*(2), 479-494.
- Borusyak, K., & Jaravel, X. (2018). *Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume* [Working paper, April 16, 2018. Version retrieved from Harvard University website], Harvard University y London School of Economics.

- https://scholar.harvard.edu/files/borusyak/files/borusyak_jaravel_event_studies.pdf
- Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting Event Study Designs: Robust and Efficient Estimation. *Review of Economic Studies*, 91(6), 3253-3285. <https://doi.org/10.1093/restud/rdae007>
- Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. C. (2024). *Difference in Differences with a Continuous Treatment* (NBER Working Paper N.º 32117). National Bureau of Economic Research. <https://doi.org/10.3386/w32117>
- Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225(2), 200-230. <https://doi.org/10.1016/j.jeconom.2020.12.001>
- Currie, J., & Gruber, J. (1996). Health Insurance Eligibility, Utilization of Medical Care, and Child Health. *Quarterly Journal of Economics*, 111(2), 431-466. <https://doi.org/10.2307/2946684>
- de Chaisemartin, C., & D'Haultfoeuille, X. (2018). Fuzzy Differences-in-Differences. *The Review of Economic Studies*, 85(2), 999-1028. <https://doi.org/10.1093/restud/rdx049>
- de Chaisemartin, C., & D'Haultfoeuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9), 2964-2996. <https://doi.org/10.1257/aer.20181169>
- de Chaisemartin, C., & D'Haultfoeuille, X. (2023). Two-Way Fixed Effects and Differences-in-Differences Estimators with Several Treatments. *Journal of Econometrics*, 236(2), 105480. <https://doi.org/10.1016/j.jeconom.2023.105480>
- de Chaisemartin, C., D'Haultfoeuille, X., Pasquier, F., Sow, D., & Vazquez-Bare, G. (2025). *Difference-in-Differences for Continuous Treatments and Instruments with Stayers* [arXiv preprint arXiv:2201.06898, version 5]. <https://arxiv.org/abs/2201.06898>
- Dube, A., Girardi, D., Jordà, Ò., & Taylor, A. M. (2023, abril). *A Local Projections Approach to Difference-in-Differences* (NBER Working Paper N.º 31184) (Revision Date May 2025). National Bureau of Economic Research. <https://doi.org/10.3386/w31184>
- Fricke, H. (2017). Identification Based on Difference-in-Differences Approaches with Multiple Treatments. *Oxford Bulletin of Economics and Statistics*, 79(5), 789-812. <https://doi.org/10.1111/obes.12178>

- Frölich, M. (2004). Programme Evaluation with Multiple Treatments. *Journal of Economic Surveys*, 18(2), 181-224. <https://doi.org/10.1111/j.0950-0804.2004.00001.x>
- Gentzkow, M., Shapiro, J. M., & Sinkinson, M. (2011). The Effect of Newspaper Entry and Exit on Electoral Politics. *American Economic Review*, 101(7), 2980-3018. <https://doi.org/10.1257/aer.101.7.2980>
- Goodman-Bacon, A. (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics*, 225(2), 254-277. <https://doi.org/10.1016/j.jeconom.2021.03.014>
- Graham, B. S., de Xavier Pinto, C. C., & Egel, D. (2012). Inverse probability tilting for moment condition models with missing data. *The Review of Economic Studies*, 79(3), 1053-1079.
- Heckman, J. J., & Vytlacil, E. J. (2007). Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation. En J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics* (pp. 4779-4874, Vol. 6B). Elsevier. [https://doi.org/10.1016/S1573-4412\(07\)06070-9](https://doi.org/10.1016/S1573-4412(07)06070-9)
- Hoynes, H. W., Miller, D., & Simon, D. (2015). Income, the Earned Income Tax Credit, and Infant Health. *American Economic Journal: Economic Policy*, 7(1), 172-211. <https://doi.org/10.1257/pol.20120179>
- Imbens, G. W. (2015). Matching Methods in Practice: Three Examples. *The Journal of Human Resources*, 50(2), 373-419. <https://www.jstor.org/stable/24735990>
- Leblebicioğlu, A., & Weinberger, A. (2020). Credit and the labour share: Evidence from US states. *The Economic Journal*, 130(630), 1782-1816.
- Meer, J., & West, J. (2016). Effects of the Minimum Wage on Employment Dynamics. *The Journal of Human Resources*, 51(2), 500-522. <http://www.jstor.org/stable/24736030>
- Meyer, B. D. (1995). Natural and Quasi-Experiments in Economics. *Journal of Business & Economic Statistics*, 13(2), 151-161. <https://doi.org/10.1080/07350015.1995.10524589>
- Mogstad, M., Santos, A., & Torgovitsky, A. (2018). Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters. *Econometrica*, 86(5), 1589-1619. <https://doi.org/10.3982/ECTA15463>
- Oreopoulos, P. (2006). Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter. *American Economic Review*, 96(1), 152-175. <https://doi.org/10.1257/000282806776157641>

- Picchetti, P., & Campos de Xavier Pinto, C. (2022). *Marginal Treatment Effects in Difference-in-Differences* [Working paper]. <https://doi.org/10.2139/ssrn.4110160>
- Rambachan, A., & Roth, J. (2023). A More Credible Approach to Parallel Trends. *Review of Economic Studies*, *90*(5), 2555-2591. <https://doi.org/10.1093/restud/rdad018>
- Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *Journal of Econometrics*, *235*, 2218-2244. <https://doi.org/10.1016/j.jeconom.2023.03.008>
- Sant'Anna, P. H. C., & Zhao, J. (2020). Doubly Robust Difference in Differences Estimators. *Journal of Econometrics*, *219*(1), 101-122. <https://doi.org/10.1016/j.jeconom.2020.06.003>
- Sant'Anna, P. H., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101-122.
- Snow, J. (1849). *On the Mode of Communication of Cholera* [Public domain edition. 31 pages. National Library of Medicine ID: 0050707]. J. Churchill. <http://resource.nlm.nih.gov/0050707>
- Strezhnev, A. (2018). *Semiparametric weighting estimators for multi-period difference-in-differences designs* [Paper presented at the Annual Meeting of the American Political Science Association, Boston, MA]. https://static1.squarespace.com/static/5931baca440243906ef65ca3/t/5b807c5d0ebbe8b9b8c75878/1535147102290/generalized_did.pdf
- Sun, L., & Abraham, S. (2021). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*, *225*(2), 175-199. <https://doi.org/10.1016/j.jeconom.2020.09.006>
- Tsai, L. (2024). Difference-in-Differences with Multiple Events [Working paper]. <https://doi.org/10.48550/arXiv.2409.05184>
- Wolfers, J. (2006). Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results. *American Economic Review*, *96*(5), 1802-1820. <https://doi.org/10.1257/aer.96.5.1802>
- Wooldridge, J. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, *141*(2), 1281-1301. <https://EconPapers.repec.org/RePEc:eee:econom:v:141:y:2007:i:2:p:1281-1301>

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2.^a ed.). MIT Press. <https://mitpress.mit.edu/9780262232586/econometric-analysis-of-cross-section-and-panel-data/>

Apéndice A

Demostraciones de resultados principales

A.1. Demostración de la Proposición 2.1

Demostración. Partimos de la definición del efecto con contrafactual puro para la unidad tratada m en el tiempo k :

$$ATT_{m,n}(k) = \mathbb{E} \left[Y_{m,\bar{t}}(h) - Y_{m,\underline{t}}(h-1) \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right] \quad (\text{A.1.1})$$

Sumamos y restamos el término contrafactual correspondiente al nivel previo de tratamiento, evaluado en la unidad tratada m :

$$ATT_{m,n}(k) = \mathbb{E} [Y_{m,\bar{t}}(h) - Y_{m,\bar{t}}(h-1)] + \mathbb{E} [Y_{m,\bar{t}}(h-1) - Y_{m,\underline{t}}(h-1)] \quad (\text{A.1.2})$$

Aplicando el Supuesto 2.1, sustituimos la evolución contrafactual de m por la observada en la unidad nunca tratada n :

$$\mathbb{E}[Y_{m,\bar{t}}(h-1) - Y_{m,\underline{t}}(h-1)] = \mathbb{E}[Y_{n,\bar{t}}(0) - Y_{n,\underline{t}}(0)] \quad (\text{A.1.3})$$

Reorganizando, obtenemos una expresión identificable empíricamente a partir de resultados observados:

$$\widehat{ATT}_{m,n}^{\circ}(k) = \mathbb{E}[y_{m,\bar{t}} - y_{n,\bar{t}}] - \mathbb{E}[y_{m,\underline{t}} - y_{n,\underline{t}}] \quad (\text{A.1.4})$$

que también puede expresarse como:

$$\widehat{ATT}_{m,n}^{\circ}(k) = \mathbb{E}[y_{m,\bar{t}} - y_{m,\underline{t}}] - \mathbb{E}[y_{n,\bar{t}} - y_{n,\underline{t}}] \quad (\text{A.1.5})$$

Lo cual concluye la demostración. \square

A.2. Demostración de la Proposición 2.2

Demostración. Partimos de la definición del efecto marginal entre tratadas para la unidad tratada $m \in \mathcal{I}_k$ y su correspondiente unidad de comparación $n \in \mathcal{I}_m^+(k)$:

$$ATT_{m,n}^+(k) = \mathbb{E} \left[Y_{m,\bar{t}}(h) - Y_{m,\underline{t}}(h-1) \mid \begin{array}{l} t_{m,n}^{\text{pre}}(k) \leq \underline{t} < k \\ k \leq \bar{t} \leq t_{m,n}^{\text{pos}}(k) \end{array} \right] \quad (\text{A.2.1})$$

donde $h = H_{m,k}$ es el nivel alcanzado por la unidad tratada en $t = k$, y $\underline{h} = H_{n,k} < h$ es el nivel constante de la unidad de control.

Sumamos y restamos el término contrafactual correspondiente al nivel previo de tratamiento de m :

$$ATT_{m,n}^+(k) = \mathbb{E} [Y_{m,\bar{t}}(h) - Y_{m,\bar{t}}(h-1)] + \mathbb{E} [Y_{m,\bar{t}}(h-1) - Y_{m,\underline{t}}(h-1)] \quad (\text{A.2.2})$$

Aplicando el Supuesto 2.2, sustituimos la diferencia contrafactual de la unidad tratada por la trayectoria observada de la unidad de control:

$$\mathbb{E}[Y_{m,\bar{t}}(h-1) - Y_{m,\underline{t}}(h-1)] = \mathbb{E}[Y_{n,\bar{t}}(\underline{h}) - Y_{n,\underline{t}}(\underline{h})] \quad (\text{A.2.3})$$

Reorganizando los términos, se obtiene una expresión identificada empíricamente:

$$\widehat{ATT}_{m,n}^+(k) = \mathbb{E}[y_{m,\bar{t}} - y_{m,\underline{t}}] - \mathbb{E}[y_{n,\bar{t}} - y_{n,\underline{t}}] \quad (\text{A.2.4})$$

que también puede escribirse como:

$$\widehat{ATT}_{m,n}^+(k) = \mathbb{E}[y_{m,\bar{t}} - y_{n,\bar{t}}] - \mathbb{E}[y_{m,\underline{t}} - y_{n,\underline{t}}] \quad (\text{A.2.5})$$

Lo cual concluye la demostración. \square

A.3. Representaciones gráficas del estimador $ATT_{m,n}^{\circ}(t)$

Este anexo complementa la Figura 2.1 presentada en el cuerpo principal del documento, la cual ilustraba el estimador $ATT_{m,n}^{\circ}(k)$ en el momento $t = k$, cuando la unidad m experimenta un cambio de nivel de tratamiento de $h-1$ a h . Sin embargo, como se discutió, la misma unidad puede haber sido tratada en otros momentos, generando efectos comparables bajo condiciones análogas.

Las siguientes figuras muestran dos casos adicionales, correspondientes a momentos distintos del historial de tratamiento de la unidad m , donde también es posible estimar un efecto con contrafactual puro válido.

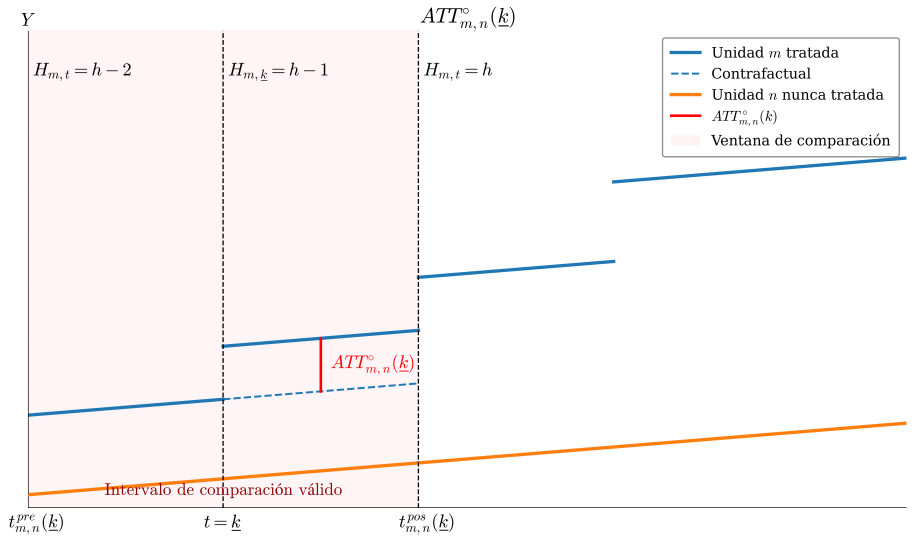


Figura A.1: Estimación de $ATT_{m,n}^{\circ}(k)$, correspondiente al cambio de nivel de tratamiento de $h-2$ a $h-1$ en la unidad m . Se representa la trayectoria contrafactual esperada bajo el supuesto de tendencias paralelas.

La Figura A.1 muestra el caso en que la unidad tratada experimenta un cambio de nivel en $t = k$, dos periodos antes de k . El efecto estimado compara la trayectoria observada de la unidad m con la de una unidad n nunca tratada, dentro de la ventana válida definida por el historial de ambas.

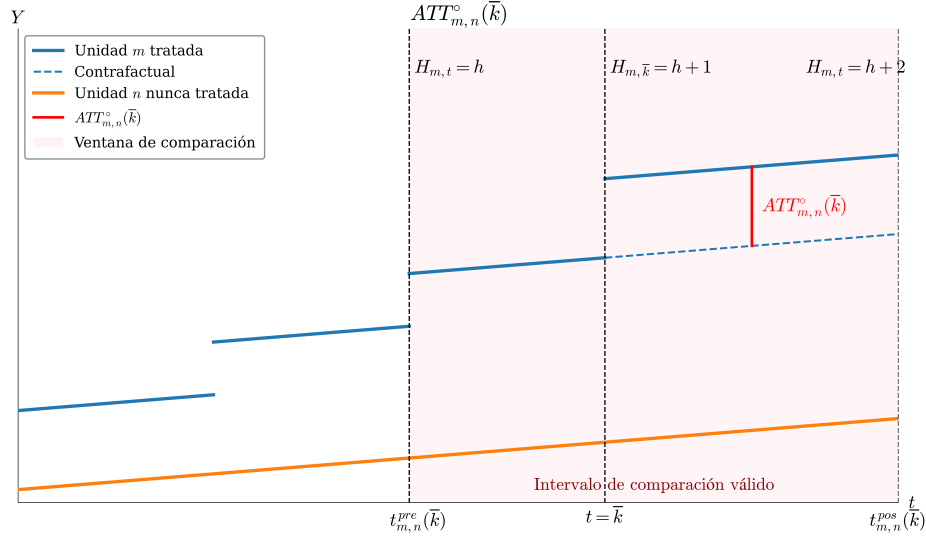


Figura A.2: Estimación de $ATT_{m,n}^{\circ}(\bar{k})$, correspondiente al cambio de nivel de tratamiento de h a $h+1$ en la unidad m .

Por su parte, la Figura A.2 representa el caso en que el cambio de nivel ocurre en $t = \bar{k}$, un periodo después de k , cuando la unidad m asciende de h a $h+1$. La estructura gráfica y la lógica de comparación son equivalentes al caso base, pero adaptadas a la nueva posición temporal y al nuevo nivel de tratamiento.

Estas representaciones permiten visualizar que el estimador $ATT_{m,n}^{\circ}(t)$ puede definirse en múltiples momentos del historial de una misma unidad tratada, siempre que se respete la validez de la ventana de comparación y exista un cambio observable en el nivel del tratamiento.

A.4. Limitaciones de comparación entre unidades tratadas

Este anexo complementa la discusión presentada en la Figura 2.2, donde se ilustró un caso en el que el nivel de intervención de la unidad de control es menor al de la tratada, permitiendo una estimación válida de $ATT_{i,j}^{+}(k)$. En contraste, la Figura A.3 muestra un caso donde no todos los efectos marginales pueden ser identificados recíprocamente entre dos unidades tratadas.

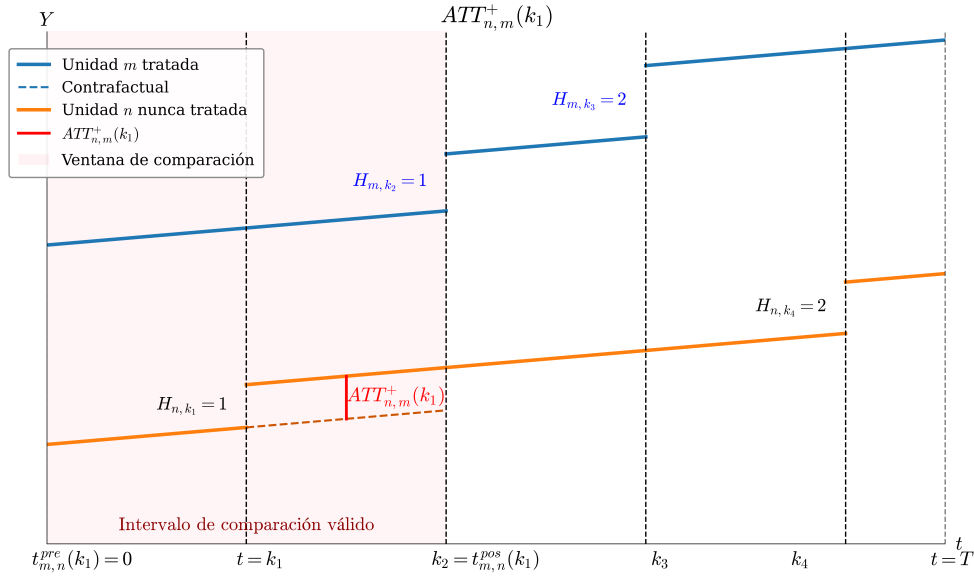


Figura A.3: Estimación válida de $ATT_{n,m}^+(k_1)$ y limitaciones para otros efectos.

En este escenario, la unidad n recibe su primera intervención en $t = k_1$, mientras que la unidad m aún no ha sido tratada. Por tanto, es posible estimar el efecto marginal $ATT_{n,m}^+(k_1)$, usando a m como grupo de comparación válido, ya que su nivel acumulado de tratamiento es menor en ese momento.

Sin embargo, no es posible usar a la unidad n como control para estimar el efecto del salto de m en $t = k_2$, pues para ese periodo ambas unidades ya han sido tratadas y presentan el mismo nivel de intervención ($H = 1$). La misma limitación aplica para el segundo salto de la unidad n en $t = k_4$, ya que m también ha aumentado su nivel de tratamiento y ya no cumple con la condición de tener un historial estrictamente menor.

Este ejemplo resalta que la reciprocidad en la comparación entre unidades tratadas no siempre es válida. El uso de una unidad tratada como contrafactual requiere que, en el momento del análisis, su nivel acumulado de intervención sea estrictamente menor al de la unidad focal.

A.5. Justificación del peso utilizado en la agregación de efectos

Siguiendo la lógica expuesta por Goodman-Bacon (2021), todo esquema de agregación tipo TWFE puede representarse como una suma ponderada de efectos 2x2, donde los pesos asignados dependen de dos grandes factores: (i) el tamaño efectivo de cada subpanel de comparación, y (ii) la varianza del tratamiento dentro de ese subpanel.

En nuestro caso, al construir estimadores $ATT_{m,n}^{\circ}(k)$ y $ATT_{m,n}^{+}(k)$ a partir de ventanas móviles individuales para cada par (m, n) , aplicamos esta misma lógica. Para cada comparación válida, definimos:

- La longitud total de la ventana de comparación como:

$$L_{m,n}(k) = t_{m,n}^{pos}(k) - t_{m,n}^{pre}(k) + 1$$

- La varianza temporal del tratamiento como:

$$Var(D_{m,n}(k)) = \frac{(t_{m,n}^{pos}(k) - k + 1)(k - t_{m,n}^{pre}(k))}{(t_{m,n}^{pos}(k) - t_{m,n}^{pre}(k) + 1)^2}$$

Este segundo componente representa la varianza de una variable ficticia D_t que toma valor 0 para todos los periodos antes del tratamiento $t < k$, y valor 1 desde el periodo de tratamiento en adelante $t \geq k$, dentro de la ventana específica definida por los límites $t_{m,n}^{pre}(k)$ y $t_{m,n}^{pos}(k)$. En otras palabras, se construye la siguiente secuencia binaria de longitud $L_{m,n}(k)$:

$$\underbrace{0, 0, \dots, 0}_{k - t_{m,n}^{pre}(k) \text{ veces}}, \quad \underbrace{1, 1, \dots, 1}_{t_{m,n}^{pos}(k) - k + 1 \text{ veces}}$$

La varianza de esta secuencia D_t se calcula como:

$$Var(D_t) = \bar{D}_t(1 - \bar{D}_t)$$

donde \bar{D}_t es el promedio de la secuencia, es decir, el número de unos dividido entre la longitud total:

$$\bar{D}_t = \frac{t_{m,n}^{pos}(k) - k + 1}{L_{m,n}(k)}$$

Entonces, desarrollando la expresión:

$$Var(D_t) = \frac{t_{m,n}^{pos}(k) - k + 1}{L_{m,n}(k)} \cdot \left(1 - \frac{t_{m,n}^{pos}(k) - k + 1}{L_{m,n}(k)} \right) = \frac{(t_{m,n}^{pos}(k) - k + 1)(k - t_{m,n}^{pre}(k))}{L_{m,n}(k)^2}$$

Este resultado refleja la idea intuitiva de que la varianza temporal del tratamiento es máxima cuando la ventana está balanceada entre ceros y unos, y tiende a cero cuando la secuencia está dominada por uno de los dos valores.

Finalmente, el peso total utilizado para ponderar cada efecto parcial es el producto entre la longitud total al cuadrado (componente de tamaño) y la varianza de tratamiento:

$$\omega_{m,n}(k) = L_{m,n}(k)^2 \cdot Var(D_t) = (t_{m,n}^{pos}(k) - k + 1)(k - t_{m,n}^{pre}(k))$$

Este resultado no solo recupera la lógica estructural de los pesos en TWFE, sino que también garantiza que las comparaciones con mayor simetría y mayor información efectiva reciban mayor peso, mientras que aquellas con ventanas sesgadas o cortas reciben peso nulo o cercano a cero.

A.6. Justificación del peso utilizado en la agregación global

La expresión utilizada para asignar un peso a cada instante k es:

$$\Omega(k) = \sum_{m \in \mathcal{I}_k} |\mathcal{I}_m(k)|$$

Esta se justifica conceptualmente a partir de la descomposición algebraica del estimador TWFE propuesta por Goodman-Bacon (2021). En particular, para el caso de comparaciones entre unidades tratadas tempranamente (k) y unidades tratadas más tarde (ℓ), el peso que acompaña al efecto en su descomposición toma la siguiente forma (ver Ecuación (10f), página 5):

$$s_{k,\ell}^k \propto ((n_k + n_\ell)(1 - \bar{D}_\ell))^2 \cdot n_{k\ell}(1 - n_{k\ell}) \cdot \frac{\bar{D}_k - \bar{D}_\ell}{1 - \bar{D}_\ell} \cdot \frac{1 - \bar{D}_k}{1 - \bar{D}_\ell} \quad (\text{A.6.1})$$

Reescribiendo esta expresión, se puede descomponer el peso como el producto de tres factores interpretables:

$$s_{k,\ell}^k \propto n_k n_\ell \cdot (1 - \bar{D}_\ell)^2 \cdot \frac{\bar{D}_k - \bar{D}_\ell}{1 - \bar{D}_\ell} \cdot \frac{1 - \bar{D}_k}{1 - \bar{D}_\ell} \quad (\text{A.6.2})$$

- El primer término, $n_k n_\ell$, representa el número total de comparaciones posibles entre unidades tratadas en k y unidades tratadas en ℓ . Este factor corresponde al tamaño efectivo del subpanel de comparación cruzada.
- El segundo término, $(1 - \bar{D}_\ell)^2$, refleja la longitud total del intervalo temporal relevante para la comparación.
- El tercer término, $\frac{\bar{D}_k - \bar{D}_\ell}{1 - \bar{D}_\ell} \cdot \frac{1 - \bar{D}_k}{1 - \bar{D}_\ell}$, captura la varianza del tratamiento entre grupos, es decir, la magnitud de la diferencia de exposición al tratamiento entre el grupo tratado y el grupo de control.

En nuestro enfoque, al construir estimadores $\widehat{ATT}(k)$ mediante una lógica de doble agregación, estas tres componentes deben ser distribuidas en etapas distintas:

- (1) En la primera etapa, se agregan comparaciones individuales (m, n) dentro de cada instante k . En dicha etapa ya se incorporan los elementos relacionados con la varianza del tratamiento y la longitud temporal del intervalo, a través del peso $\omega_{m,n}(k)$ basado en la ventana móvil.
- (2) En la segunda etapa, se combinan los efectos por instante $\widehat{ATT}(k)$. Para esta agregación, lo que queda por considerar es únicamente el número total de comparaciones válidas utilizadas en la construcción de $\widehat{ATT}(k)$. Este valor corresponde exactamente a:

$$\Omega(k) = \sum_{m \in \mathcal{I}_k} |\mathcal{I}_m(k)|$$

lo cual es análogo al término $n_k n_\ell$ en el peso TWFE original.

Así, en lugar de tomar directamente los pesos resultantes del TWFE, nuestra propuesta reconstruye su lógica subyacente y la redistribuye de forma coherente entre ambas etapas de agregación. Esto permite mantener la interpretabilidad del proceso y adaptar la ponderación a la estructura específica de nuestro diseño.

A.7. Demostraciones de los resultados asintóticos (diseño sin ruptura)

Marco y notación

Sea I el número de unidades. En el instante k , el estimador agregado es $\widehat{ATT}(k) = \sum_{(m,n)} \widehat{\omega}_{m,n}(k) \widehat{ATT}_{m,n}(k)$, con pesos normalizados acotados y supuestos globales: irreversibilidad, incremento unitario y muestreo i.i.d. por unidad sin interferencia (Sup. 1–3), además de representatividad y estabilidad en ventana (Sup. 2.3, 2.4).

Representación por clúster

Existe una descomposición

$$\widehat{ATT}(k) - ATT(k) = \frac{1}{I} \sum_{i=1}^I \psi_{i,k} + r_{I,k}, \quad \mathbb{E}[\psi_{i,k}] = 0, \quad \mathbb{E}[\psi_{i,k}^2] < \infty, \quad r_{I,k} = o_{\mathbb{P}}(I^{-1/2}),$$

donde $\psi_{i,k}$ agrega todas las contribuciones en que participa la unidad i (los traslapes se absorben en $r_{I,k}$). La acotación de pesos y momentos finitos se obtiene de los supuestos citados.

CLT por clúster y varianza

Por independencia entre unidades (Sup. 3) y condición de Lindeberg,

$$\sqrt{I} (\widehat{ATT}(k) - ATT(k)) \xrightarrow{d} \mathcal{N}(0, \sigma_k^2).$$

Como N_k es proporcional al tamaño efectivo de comparaciones en k , el reescalamiento por $\sqrt{N_k}$ es equivalente. La fórmula “sin covarianzas” del texto es expositiva; para inferencia se usa bootstrap por unidad.

Consistencia del bootstrap por unidad

Re-muestreando unidades completas con reemplazo y reconstruyendo el promedio ponderado (renormalizando los pesos dentro de cada réplica), el bootstrap es consistente para la ley límite de $\widehat{ATT}(k)$ y para combinaciones lineales finitas en k .

Agregación global

Para $\widehat{ATT} = \sum_k \widehat{\Omega}(k) \widehat{ATT}(k)$ con $\sum_k \widehat{\Omega}(k) = 1$ y pesos acotados/determinísticos, Crámer–Wold y Slutsky dan

$$\sqrt{N} (\widehat{ATT} - ATT) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad N = \sum_k N_k,$$

y el bootstrap por unidad (reproduciendo la agregación entre k) es consistente para varianza, ICs y p -valores.

A.8. Ventanas admisibles y controles efectivos

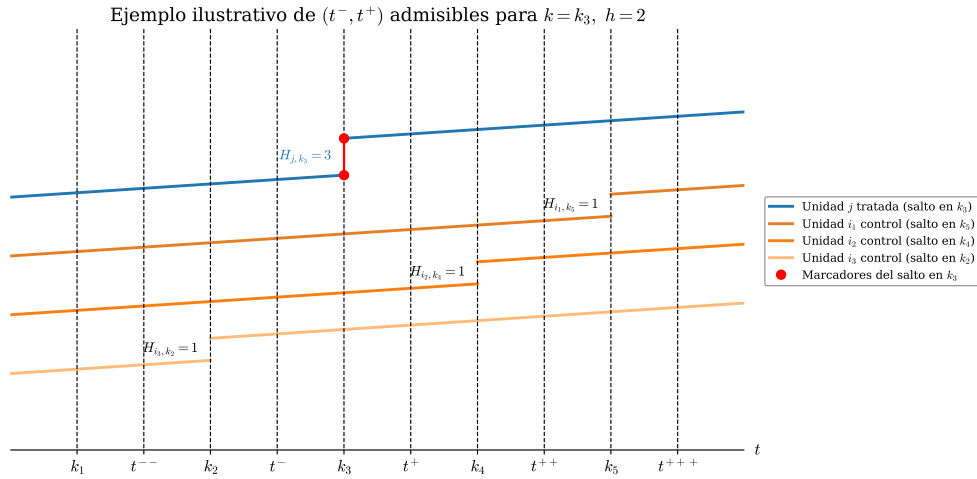


Figura A.4: Configuración ilustrativa de tiempos intercalados y niveles de intervención.

En lo que sigue nos centraremos en la unidad tratada j , que en k_3 alcanza el nivel $H_{j,k_3} = 3$. Las demás unidades i_1, i_2, i_3 actúan como potenciales controles y presentan niveles de intervención estrictamente menores que 3.

En esta configuración, el conjunto de tratadas para $(k, h) = (k_3, 3)$ es

$$\mathcal{I}^{k_3,3} = \{j\}.$$

Denotamos por $\mathcal{T}^-(k_3, 3)$ el conjunto de *tiempos previos admisibles* asociado a $(k_3, 3)$ (según la definición dada en el cuerpo del texto). En particular, en

la Figura A.4 los instantes t^{--} y t^- satisfacen

$$t^{--}, t^- \in \mathcal{T}^-(k_3, 3), \quad \text{con} \quad k_1 < t^{--} < k_2 < t^- < k_3.$$

Los *tiempos posteriores admisibles* no forman un conjunto único: dependen del t^- elegido. En este ejemplo, bajo la regla de elegibilidad $H_{j,t^-} = H_{j,t^+} < 3$,

$$\mathcal{T}^+(k_3, 3 \mid t^-) = \{t^+, t^{++}, t^{+++}\}, \quad \mathcal{T}^+(k_3, 3 \mid t^{--}) = \{t^+, t^{++}\},$$

de modo que $t^{+++} \notin \mathcal{T}^+(k_3, 3 \mid t^{--})$.

Como complemento, la Tabla A.1 resume, para cada par (t^-, t^+) admisible en torno a k_3 , qué unidades i_r pueden actuar como controles bajo la regla $H_{j,t^-} = H_{j,t^+} < 3$. Obsérvese que el conjunto resultante depende de t^- , tal como se anticipó arriba y se ilustra en la Figura A.4.

	t^+	t^{++}	t^{+++}
t^{--}	$\{i_1, i_2\}$	$\{i_1\}$	\emptyset
t^-	$\{i_1, i_2, i_3\}$	$\{i_1, i_3\}$	$\{i_3\}$

Cuadro A.1: Controles válidos por par (t^-, t^+) bajo la regla $H_{j,t^-} = H_{j,t^+} < 3$.

A.9. Demostración de la Proposición 2.4

Demostración. Partimos de la Definición 2.2.2.1 (condicionando en tratadas efectivas por ventana):

$$ATT^{t^-, t^+}(k, h) = \mathbb{E}[Y_{i,t^+}(h) - Y_{i,t^+}(h-1) \mid i \in \mathcal{I}^{k,h}(t^-, t^+)]. \quad (\text{A.9.1})$$

Sumamos y restamos $Y_{i,t^-}(h-1)$ dentro de la esperanza:

$$ATT^{t^-, t^+}(k, h) = \mathbb{E}\left[(Y_{i,t^+}(h) - Y_{i,t^-}(h-1)) - (Y_{i,t^+}(h-1) - Y_{i,t^-}(h-1)) \mid i \in \mathcal{I}^{k,h}(t^-, t^+) \right]. \quad (\text{A.9.2})$$

Aplicamos el Supuesto 2.5 (cf. ec. (2.2.5)) para reemplazar el segundo término por el promedio correspondiente en controles $j \in \mathcal{J}^{k,h}(t^-, t^+)$.

Además, de la construcción de los conjuntos $\mathcal{I}^{k,h}(t^-, t^+)$ y $\mathcal{J}^{k,h}(t^-, t^+)$, se tiene que en la ventana $[t^-, t^+]$:

$$\begin{aligned} y_{i,t^-} &= Y_{i,t^-}(h-1), & y_{i,t^+} &= Y_{i,t^+}(h) \quad \text{para todo } i \in \mathcal{I}^{k,h}(t^-, t^+), \\ y_{j,t^-} &= Y_{j,t^-}(H_{j,t^-}), & y_{j,t^+} &= Y_{j,t^+}(H_{j,t^-}) \quad \text{para todo } j \in \mathcal{J}^{k,h}(t^-, t^+). \end{aligned}$$

Sustituyendo en (A.9.2) llegamos a:

$$\begin{aligned} ATT^{t^-, t^+}(k, h) &= \mathbb{E}[y_{i,t^+} - y_{i,t^-} \mid i \in \mathcal{I}^{k,h}(t^-, t^+)] \\ &\quad - \mathbb{E}[y_{j,t^+} - y_{j,t^-} \mid j \in \mathcal{J}^{k,h}(t^-, t^+)] \\ &=: \widehat{ATT}^{t^-, t^+}(k, h). \end{aligned} \tag{A.9.3}$$

Con esto queda probado que el funcional en observables $\widehat{ATT}^{t^-, t^+}(k, h)$ identifica $ATT^{t^-, t^+}(k, h)$ bajo el Supuesto 2.5. \square

A.10. Demostración de la Proposición 2.3

Demostración. (1) Sean $t_1^+, t_2^+ \in \mathcal{T}^+(k, h \mid t^-)$ con $t_1^+ \leq t_2^+$. Sea $j \in \mathcal{J}^{k,h}(t^-, t_2^+)$; por definición, $H_{j,t^-} = H_{j,t_2^+} < h$. Como $H_{j,\cdot}$ es no decreciente (Supuesto 1), la igualdad en los extremos implica constancia en $[t^-, t_2^+]$. En particular, $H_{j,t_1^+} = H_{j,t^-} < h$, por lo que $j \in \mathcal{J}^{k,h}(t^-, t_1^+)$. Queda probado $\mathcal{J}^{k,h}(t^-, t_2^+) \subseteq \mathcal{J}^{k,h}(t^-, t_1^+)$.

(2) Sean $t_1^+, t_2^+ \in \mathcal{T}^+(k, h \mid t^-)$ con $t_1^+ \leq t_2^+$. Sea $i \in \mathcal{I}^{k,h}(t^-, t_2^+)$; entonces $H_{i,t^-} = h-1$ y, por pertenecer i a $\mathcal{I}^{k,h}$, en k se produce el salto unitario a h (Supuesto 2), manteniéndose $H_{i,s} = h$ para todo $s \in [k, t_2^+]$ por irreversibilidad. En particular, $H_{i,t_1^+} = h$, de modo que $i \in \mathcal{I}^{k,h}(t^-, t_1^+)$. Luego $\mathcal{I}^{k,h}(t^-, t_2^+) \subseteq \mathcal{I}^{k,h}(t^-, t_1^+)$.

(3) Sea t^+ tal que $t^+ \in \mathcal{T}^+(k, h \mid t^-)$ sea admisible para los tiempos previos considerados. Sean $t_1^-, t_2^- \in \mathcal{T}^-(k, h)$ con $t_1^- < t_2^- < k$ y $t^+ \in \mathcal{T}^+(k, h \mid t_1^-) \cap \mathcal{T}^+(k, h \mid t_2^-)$. Sea $j \in \mathcal{J}^{k,h}(t_1^-, t^+)$; entonces $H_{j,t_1^-} = H_{j,t^+} < h$. Por no decrecimiento, la igualdad en los extremos fuerza constancia en $[t_1^-, t^+]$, y en particular en $[t_2^-, t^+]$. Así, $H_{j,t_2^-} = H_{j,t^+} < h$ y $j \in \mathcal{J}^{k,h}(t_2^-, t^+)$. Se concluye $\mathcal{J}^{k,h}(t_1^-, t^+) \subseteq \mathcal{J}^{k,h}(t_2^-, t^+)$.

- (4) Sea $i \in \mathcal{I}^{k,h}(t_1^-, t^+)$. Entonces $H_{i,t_1^-} = h-1$ y, por irreversibilidad e incremento unitario, no hay salto en (t_1^-, k) y el salto a h ocurre en k , manteniéndose $H_{i,s} = h$ para todo $s \in [k, t^+]$. Dado que $t_2^- > t_1^-$, se tiene $H_{i,t_2^-} = h-1$ y $H_{i,t^+} = h$, por lo que $i \in \mathcal{I}^{k,h}(t_2^-, t^+)$. Así, $\mathcal{I}^{k,h}(t_1^-, t^+) \subseteq \mathcal{I}^{k,h}(t_2^-, t^+)$. □

A.11. Intuición del peso 2.2.10

Sea $N = n_T + n_C$ el total de unidades en la ventana y defina $p_T = \frac{n_T}{N}$ y $p_C = \frac{n_C}{N} = 1 - p_T$. Considere el indicador de tratamiento por unidad $Z_i = \mathbf{1}\{i \in \text{tratadas}\}$. Entonces

$$\mathbb{E}[Z_i] = p_T, \quad \text{Var}(Z_i) = p_T(1 - p_T) = p_T p_C.$$

Es decir, el producto de proporciones $p_T p_C$ es la “varianza del tratamiento” dentro de la ventana y mide el balance entre grupos. Al escalar por el tamaño N se obtiene el peso

$$\omega = N \cdot p_T p_C = \frac{n_T n_C}{n_T + n_C}.$$

A.12. Anexo: Invarianza del orden de agregación

Por definición de las probabilidades a partir de las cardinalidades $|\mathcal{I}^{k,h}|$,

$$p(k, h) = \frac{|\mathcal{I}^{k,h}|}{\sum_{u,v} |\mathcal{I}^{u,v}|}, \quad p(k) = \frac{\sum_v |\mathcal{I}^{k,v}|}{\sum_{u,v} |\mathcal{I}^{u,v}|}, \quad p(h | k) = \frac{|\mathcal{I}^{k,h}|}{\sum_v |\mathcal{I}^{k,v}|},$$

y análogamente $p(h) = \frac{\sum_u |\mathcal{I}^{u,h}|}{\sum_{u,v} |\mathcal{I}^{u,v}|}$ y $p(k | h) = \frac{|\mathcal{I}^{k,h}|}{\sum_u |\mathcal{I}^{u,h}|}$. Entonces,

$$p(k) p(h | k) = \frac{\sum_v |\mathcal{I}^{k,v}|}{\sum_{u,v} |\mathcal{I}^{u,v}|} \cdot \frac{|\mathcal{I}^{k,h}|}{\sum_v |\mathcal{I}^{k,v}|} = \frac{|\mathcal{I}^{k,h}|}{\sum_{u,v} |\mathcal{I}^{u,v}|} = p(k, h),$$

y de forma análoga $p(h) p(k | h) = p(k, h)$.

Por tanto,

$$\begin{aligned} \sum_k p(k) \widehat{ATT}_K(k) &= \sum_k p(k) \sum_h p(h | k) \widehat{ATT}(k, h) \\ &= \sum_{k,h} p(k, h) \widehat{ATT}(k, h) = \widehat{ATT}_{\text{global}}. \end{aligned} \quad (\text{A.12.1})$$

Del mismo modo se obtiene

$$\begin{aligned} \sum_h p(h) \widehat{ATT}_H(h) &= \sum_h p(h) \sum_k p(k | h) \widehat{ATT}(k, h) \\ &= \sum_{k,h} p(h) p(k | h) \widehat{ATT}(k, h) = \sum_{k,h} p(k, h) \widehat{ATT}(k, h) \\ &= \widehat{ATT}_{\text{global}}. \end{aligned} \quad (\text{A.12.2})$$

□

A.13. Ventanas máxima y admisible

En el diseño con *ruptura de tendencia*, la comparación unidad–unidad en el instante k requiere distinguir entre la *ventana máxima* propia de la unidad tratada m y la *ventana admisible* que resulta al emparejar m con una unidad de comparación n que permanece en $h-1$ en $t=k$ y no cambia de nivel en ese instante.

A.13.0.1 Definición: Ventanas máxima y admisible en el salto k

Sea $m \in \mathcal{I}^{k,h}$ (esto es, $H_{m,k-1} = h-1$ y $H_{m,k} = h$). Definimos los extremos de la **ventana máxima** de m en k como los bordes del último tramo previo en $h-1$ y del primer tramo posterior en h :

$$\begin{aligned} t_m^{\text{pre,max}}(k) &= \min\{t^- \leq k-1 : H_{m,t^-} = h-1\}, \\ t_m^{\text{pos,max}}(k) &= \max\{t^+ \geq k : H_{m,t^+} = h\}. \end{aligned}$$

Si emparejamos con n elegible en k (esto es, $H_{n,k} = h-1$ y sin cambio en k), la **ventana admisible** del par (m, n) viene dada por

$$\begin{aligned} t_{m|n}^{\text{pre,adm}}(k) &= \min\{t^- \leq k-1 : H_{m,t^-} = h-1, H_{n,t^-} = h-1\}, \\ t_{m|n}^{\text{pos,adm}}(k) &= \max\{t^+ \geq k : H_{m,t^+} = h, H_{n,t^+} = h-1\}. \end{aligned}$$

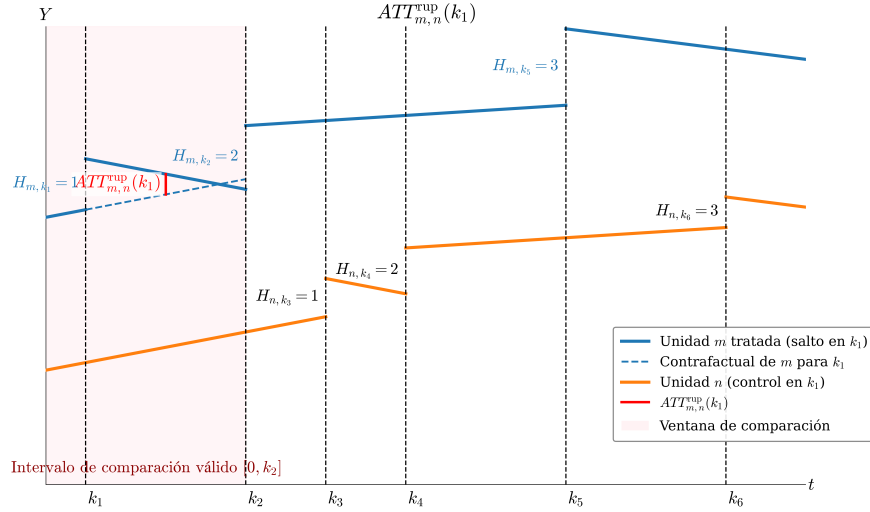


Figura A.5: Primer salto de m en k_1 : la ventana admisible (post) coincide con la ventana máxima (post), esto es, $[k_1, k_2]$.

Por construcción,

$$[t_{m|n}^{\text{pre,adm}}(k), t_{m|n}^{\text{pos,adm}}(k)] \subseteq [t_m^{\text{pre,max}}(k), t_m^{\text{pos,max}}(k)],$$

con igualdad únicamente si n permanece en $h-1$ durante todo el tramo previo y posterior relevantes. El *promedio del efecto* $ATT_{m,n}^{\text{rup}}(k)$ se toma **exclusivamente en el tramo posterior** $[k, \cdot]$; el tramo previo $[\cdot, k)$ se usa para el ajuste DiD (identificación), no para promediar el efecto.

La Figura A.5 ilustra el primer salto de m en k_1 . En este caso, n es elegible en k_1 y su permanencia en $h-1$ no recorta el tramo posterior de m . En consecuencia,

$$\underbrace{[k_1, t_{m|n}^{\text{pos,adm}}(k_1)]}_{\text{admisible (post)}} = \underbrace{[k_1, t_m^{\text{pos,max}}(k_1)]}_{\text{máxima (post)}} = [k_1, k_2].$$

En la figura se muestra el contrafactual que extiende la tendencia en $h-1$ y un marcador puntual del $ATT_{m,n}^{\text{rup}}(k_1)$ dentro de la ventana.

La Figura A.6 contrasta, para el salto de m en k_5 , la *ventana máxima (post)* $[k_5, T]$ con la *ventana admisible (post)* con n , que se trunca en k_6 :

$$[k_5, t_{m|n}^{\text{pos,adm}}(k_5)] = [k_5, k_6] \subset [k_5, T] = [k_5, t_m^{\text{pos,max}}(k_5)].$$

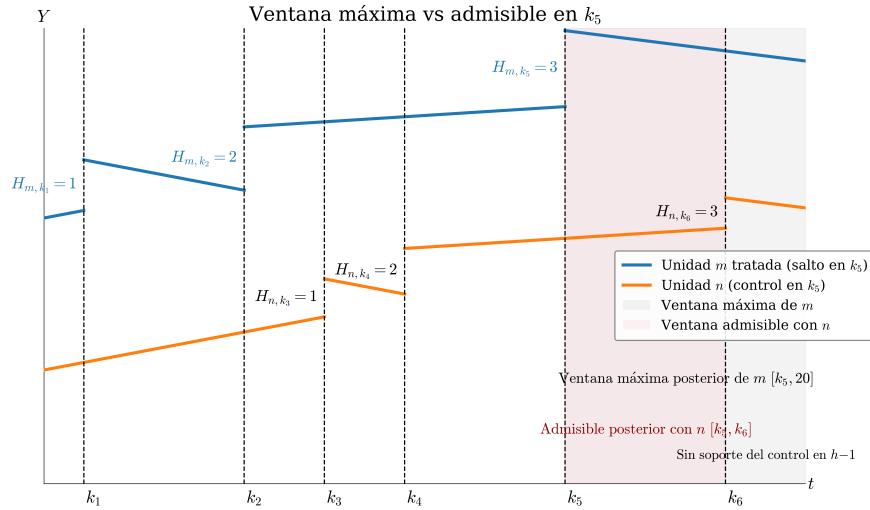


Figura A.6: Tercer salto de m en k_5 : la ventana admisible (post) $[k_5, k_6]$ es un subconjunto de la ventana máxima (post) $[k_5, T]$. El tramo $[k_6, T]$ no es identificable con este control por falta de soporte en $h-1$.

El tramo $[k_6, T]$ carece de soporte del control en $h-1$ y, por ende, el contrafactual no es identificable allí con esta comparación; por eso la figura resalta únicamente las ventanas (post) y no dibuja un ATT puntual más allá de k_6 .

Mapa de comparaciones posibles en la figura (justificación)

Usando los instantes $k_1, k_2, k_3, k_4, k_5, k_6$ y los niveles de m y n mostrados:

- **Salto de m en k_1 ($0 \rightarrow 1$): válido.** En $t=k_1$, n está **exactamente** en $h-1=0$ y no cambia en k_1 . La ventana admisible (post) coincide con la máxima (post): $[k_1, k_2]$.
- **Salto de m en k_2 ($1 \rightarrow 2$): no válido.** En $t=k_2$, el control requerido debe cumplir $H_{n,k_2} = h-1=1$. Sin embargo, n no está en 1 en k_2 . Falla la condición de elegibilidad $H_{n,k} = h-1$.
- **Salto de m en k_5 ($2 \rightarrow 3$): válido con truncación.** En $t=k_5$, n sí cumple $H_{n,k_5} = h-1=2$ y no cambia en k_5 , por lo que la comparación es válida. No obstante, la ventana admisible (post) se acorta a $[k_5, k_6]$

porque n deja de estar en $h-1$ a partir de k_6 ; la ventana máxima (post) de m llega hasta T .

▪ **Inversión de roles (tratar a n y usar a m de control): no hay saltos medibles de n con m como control.**

- En k_3 ($0 \rightarrow 1$ de n) se requiere $H_{m,k_3} = 0$, pero m ya está en un nivel superior.
- En k_4 ($1 \rightarrow 2$ de n) se requiere $H_{m,k_4} = 1$, pero m ya está en un nivel superior.
- En k_6 ($2 \rightarrow 3$ de n) se requiere $H_{m,k_6} = 2$, pero m ya está en un nivel superior.

En los tres casos falla la condición $H_{\text{control},k} = h-1$, por lo que no existe ventana admisible con m como control para saltos de n .

Nota sobre pre y post. Las figuras destacan únicamente los *tramos posteriores* porque el objeto de interés $ATT_{m,n}^{\text{rup}}(k)$ promedia valores en $[k, \cdot]$. Para la *identificación* (DiD) se emplea además el tramo *previo* común en $h-1$ (cuando existe): por ejemplo, para el salto k_5 el pre común de (m, n) está en $[k_4, k_5)$, mientras que el post admisible es $[k_5, k_6]$. Esta distinción entre definición (post) e identificación (pre & post) explica por qué en la Figura A.6 no se dibuja un ATT puntual más allá de k_6 : allí no hay soporte del control en $h-1$.

A.14. Demostraciones de los resultados asintóticos (diseño con ruptura)

Celdas (k, h) y supuestos

Trabajamos con $\widehat{ATT}^{\text{rup}}(k, h)$ definido en §3.1.3, bajo muestreo i.i.d. por unidad (Sup. 3), representatividad y paralelismo *local* por nivel $h-1$ (Sup. 3.1).

Representación por clúster en celdas

Para cada celda (k, h) ,

$$\widehat{ATT}^{\text{rup}}(k, h) - ATT^{\text{rup}}(k, h) = \frac{1}{I} \sum_{i=1}^I \psi_{i,k,h}^{\text{rup}} + r_{I,k,h}^{\text{rup}},$$

con momentos finitos y resto $o_{\mathbb{P}}(I^{-1/2})$.

CLT y varianza en celdas

Bajo independencia entre unidades y pesos acotados,

$$\sqrt{N_{k,h}} (\widehat{ATT}^{\text{rup}}(k, h) - ATT^{\text{rup}}(k, h)) \xrightarrow{d} \mathcal{N}(0, \sigma_{k,h}^2).$$

La inferencia operativa usa bootstrap por unidad (y, si aplica, bloques temporales) reconstruyendo pesos y agregados dentro de cada réplica.

Agregados en k y en h

Para $\widehat{ATT}_K^{\text{rup}}(k) = \sum_h \Psi(k, h) \widehat{ATT}^{\text{rup}}(k, h)$ y $\widehat{ATT}_H^{\text{rup}}(h) = \sum_k \Phi(h, k) \widehat{ATT}^{\text{rup}}(k, h)$, la normalidad asintótica se preserva por linealidad (Crámer–Wold), y el bootstrap por unidad es consistente para sus varianzas.

Promedio global

Para $\widehat{ATT}_{\text{global}}^{\text{rup}}$ (eq. (3.1.9)),

$$\sqrt{N} (\widehat{ATT}_{\text{global}}^{\text{rup}} - ATT_{\text{global}}^{\text{rup}}) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad N = \sum_{k,h} N_{k,h},$$

y el bootstrap reproduce la agregación entre celdas.

A.15. Demostración de la Proposición 3.2

Demostración. Partimos de la Definición 3.2.2.1 (condicionando en tratadas efectivas por ventana):

$$ATT_{\text{rup}}^{t^-, t^+}(k, h) = \mathbb{E}[Y_{i,t^+}(h) - Y_{i,t^-}(h-1) \mid i \in \mathcal{I}^{k,h}(t^-, t^+)]. \quad (\text{A.15.1})$$

Sumamos y restamos $Y_{i,t^-}(h-1)$ dentro de la esperanza:

$$\begin{aligned} ATT_{\text{rup}}^{t^-,t^+}(k,h) &= \mathbb{E}\left[(Y_{i,t^+}(h) - Y_{i,t^-}(h-1)) \right. \\ &\quad \left. - (Y_{i,t^+}(h-1) - Y_{i,t^-}(h-1)) \mid i \in \mathcal{I}^{k,h}(t^-,t^+) \right]. \end{aligned} \tag{A.15.2}$$

Aplicamos el Supuesto 3.2 (ec. (3.2.5)) para reemplazar el segundo término por el promedio correspondiente en los controles que permanecen en $h-1$ en la ventana, esto es, $j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-,t^+)$.

Además, por la construcción de los conjuntos efectivos en $[t^-,t^+]$,

$$\begin{aligned} y_{i,t^-} &= Y_{i,t^-}(h-1), & y_{i,t^+} &= Y_{i,t^+}(h) \quad \text{para todo } i \in \mathcal{I}^{k,h}(t^-,t^+), \\ y_{j,t^-} &= Y_{j,t^-}(h-1), & y_{j,t^+} &= Y_{j,t^+}(h-1) \quad \text{para todo } j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-,t^+). \end{aligned}$$

Sustituyendo en (A.15.2), obtenemos:

$$\begin{aligned} ATT_{\text{rup}}^{t^-,t^+}(k,h) &= \mathbb{E}[y_{i,t^+} - y_{i,t^-} \mid i \in \mathcal{I}^{k,h}(t^-,t^+)] \\ &\quad - \mathbb{E}[y_{j,t^+} - y_{j,t^-} \mid j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-,t^+)] \\ &=: \widehat{ATT}_{\text{rup}}^{t^-,t^+}(k,h), \end{aligned} \tag{A.15.3}$$

que coincide con la expresión en (3.2.6).

Finalmente, sumando y restando $\mathbb{E}[y_{i,t^-} - y_{j,t^-} \mid i \in \mathcal{I}^{k,h}(t^-,t^+), j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-,t^+)]$, la forma en *niveles* se obtiene como

$$\begin{aligned} \widehat{ATT}_{\text{rup}}^{t^-,t^+}(k,h) &= \underbrace{\mathbb{E}[y_{i,t^+} - y_{j,t^+} \mid i \in \mathcal{I}^{k,h}(t^-,t^+), j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-,t^+)]}_{\text{diferencia intergrupala en } t^+} \\ &\quad - \underbrace{\mathbb{E}[y_{i,t^-} - y_{j,t^-} \mid i \in \mathcal{I}^{k,h}(t^-,t^+), j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-,t^+)]}_{\text{diferencia intergrupala en } t^-}, \end{aligned} \tag{A.15.4}$$

que reproduce (3.2.7). Con ello queda demostrado que el funcional en observables identifica $ATT_{\text{rup}}^{t^-,t^+}(k,h)$ bajo el Supuesto 3.2. \square

Nota A.1. *La demostración replica, mutatis mutandis, la del caso sin ruptura (Anexo A.9), con la única diferencia de que aquí los controles permanecen en $h-1$ en toda la ventana y, por tanto, los términos observables para controles se evalúan siempre al nivel $h-1$.*

A.16. Demostración de la equivalencia entre las representaciones OR, IPW y DR

En este anexo se demuestra la equivalencia algebraica entre las tres representaciones del estimador $ATT_{\text{rup}}^{t^-,t^+}(k, h)$: Outcome Regression (OR), Inverse Probability Weighting (IPW) y Doubly Robust (DR). La demostración se desarrolla en tres partes: primero se establece la equivalencia del ATT con la forma OR, luego con la forma IPW y, finalmente, con la forma DR.

A.16.1. Equivalencia con la representación OR

En esta subsección se demuestra que el estimador propuesto $ATT_{\text{rup}}^{t^-,t^+}(k, h)$ es algebraicamente equivalente a su representación basada en regresión de resultados ($ATT_{\text{rup-OR}}^{t^-,t^+}(k, h)$).

Demostración. Partimos de la definición general del efecto promedio en la ventana:

$$ATT_{\text{rup}}^{t^-,t^+}(k, h) = \mathbb{E}[y_{t^+} - y_{t^-} \mid i \in \mathcal{I}^{k,h}(t^-, t^+)] - \mathbb{E}[y_{t^+} - y_{t^-} \mid j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-, t^+)].$$

Bajo el supuesto de paralelismo condicional con covariables X , el segundo término puede expresarse como

$$\mathbb{E}[y_{t^+} - y_{t^-} \mid j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-, t^+)] = \mathbb{E}[m_{k,h,t^-,t^+}^{\text{ctrl}}(X)],$$

donde $m_{k,h,t^-,t^+}^{\text{ctrl}}(X) = \mathbb{E}[y_{t^+} - y_{t^-} \mid j \in \mathcal{J}_{\text{rup}}^{k,h}(t^-, t^+), X_j = X]$ es el modelo de resultados condicional estimado sobre controles válidos.

Sustituyendo esta expresión, el ATT se reescribe como:

$$ATT_{\text{rup}}^{t^-,t^+}(k, h) = \mathbb{E}[(y_{t^+} - y_{t^-}) - m_{k,h,t^-,t^+}^{\text{ctrl}}(X) \mid i \in \mathcal{I}^{k,h}(t^-, t^+)].$$

Multiplicando y dividiendo por $\mathbb{E}[D]$, con $D \equiv D_u^{k,h,t^-,t^+}$, se obtiene una expresión global en términos de esperanza incondicional:

$$ATT_{\text{rup}}^{t^-,t^+}(k, h) = \mathbb{E}\left[\frac{D}{\mathbb{E}[D]} \left(y_{t^+} - y_{t^-} - m_{k,h,t^-,t^+}^{\text{ctrl}}(X)\right)\right],$$

que coincide exactamente con la representación $ATT_{\text{rup-OR}}^{t^-,t^+}(k, h)$ definida en la Sección 3.2.8. Por tanto,

$$ATT_{\text{rup}}^{t^-,t^+}(k, h) = ATT_{\text{rup-OR}}^{t^-,t^+}(k, h).$$

□

A.16.2. Equivalencia con la representación IPW

En esta subsección se demuestra que el estimador propuesto $ATT_{\text{rup}}^{t^-, t^+}(k, h)$ es algebraicamente equivalente a su representación por ponderación inversa de probabilidad ($ATT_{\text{rup-IPW}}^{t^-, t^+}(k, h)$).

Demostración. Partimos de la definición general del efecto promedio en la ventana:

$$ATT_{\text{rup}}^{t^-, t^+}(k, h) = \mathbb{E}[y_{t^+} - y_{t^-} \mid i \in \mathcal{I}^{k, h}(t^-, t^+)] - \mathbb{E}[y_{t^+} - y_{t^-} \mid j \in \mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+)].$$

El primer término puede expresarse como

$$\mathbb{E}[y_{t^+} - y_{t^-} \mid D = 1] = \frac{\mathbb{E}[D(y_{t^+} - y_{t^-})]}{\mathbb{E}[D]},$$

donde $D \equiv D_u^{k, h, t^-, t^+}$ es el indicador de tratamiento efectivo en la ventana.

Para el grupo de control, se aplica la reponderación mediante el puntaje de propensión $p(X)$, de modo que

$$\mathbb{E}[y_{t^+} - y_{t^-} \mid D = 0] = \frac{\mathbb{E}\left[(1 - D) \frac{p(X)}{1 - p(X)} (y_{t^+} - y_{t^-})\right]}{\mathbb{E}\left[(1 - D) \frac{p(X)}{1 - p(X)}\right]}.$$

Sustituyendo ambas expresiones, se obtiene:

$$ATT_{\text{rup}}^{t^-, t^+}(k, h) = \frac{\mathbb{E}[D(y_{t^+} - y_{t^-})]}{\mathbb{E}[D]} - \frac{\mathbb{E}\left[(1 - D) \frac{p(X)}{1 - p(X)} (y_{t^+} - y_{t^-})\right]}{\mathbb{E}\left[(1 - D) \frac{p(X)}{1 - p(X)}\right]}.$$

Esta expresión coincide exactamente con la definición formal del estimador $ATT_{\text{rup-IPW}}^{t^-, t^+}(k, h)$ presentada en la Sección 3.2.8. Por tanto,

$$ATT_{\text{rup}}^{t^-, t^+}(k, h) = ATT_{\text{rup-IPW}}^{t^-, t^+}(k, h).$$

□

A.16.3. Equivalencia con la representación DR

Finalmente, se demuestra la equivalencia entre el estimador base $ATT_{\text{rup}}^{t^-,t^+}(k, h)$ y su forma *doblemente robusta* ($ATT_{\text{rup-DR}}^{t^-,t^+}(k, h)$). El objetivo es mostrar que la inclusión simultánea del modelo de resultados $m^{\text{ctrl}}(\cdot)$ y del puntaje de propensión $p(\cdot)$ no altera el valor esperado del estimador, sino que lo estabiliza frente a una posible mala especificación de cualquiera de los dos modelos.

Demostración. Partimos de la definición general del estimador DR:

$$ATT_{\text{rup-DR}}^{t^-,t^+}(k, h) = \mathbb{E} \left[\frac{D}{\mathbb{E}[D]} \left(y_{t^+} - y_{t^-} - m^{\text{ctrl}}(X) \right) \right] - \mathbb{E} \left[\frac{(1-D) \frac{p(X)}{1-p(X)}}{\mathbb{E} \left[(1-D) \frac{p(X)}{1-p(X)} \right]} \left(y_{t^+} - y_{t^-} - m^{\text{ctrl}}(X) \right) \right]$$

Aplicando la linealidad de la esperanza y agrupando términos, puede expresarse como:

$$ATT_{\text{rup-DR}}^{t^-,t^+}(k, h) = ATT_{\text{rup-IPW}}^{t^-,t^+}(k, h) - \mathbb{E} \left[\left(\frac{D}{\mathbb{E}[D]} - \frac{(1-D) \frac{p(X)}{1-p(X)}}{\mathbb{E} \left[(1-D) \frac{p(X)}{1-p(X)} \right]} \right) m^{\text{ctrl}}(X) \right].$$

El segundo término se anula porque, por la identidad de normalización establecida en la demostración anterior ($\mathbb{E}[D] = \mathbb{E} \left[(1-D) \frac{p(X)}{1-p(X)} \right]$), se cumple que

$$\mathbb{E} \left[\frac{D}{\mathbb{E}[D]} - \frac{(1-D) \frac{p(X)}{1-p(X)}}{\mathbb{E} \left[(1-D) \frac{p(X)}{1-p(X)} \right]} \right] = 0.$$

Por tanto, el término asociado a $m^{\text{ctrl}}(X)$ desaparece y se obtiene

$$ATT_{\text{rup-DR}}^{t^-,t^+}(k, h) = ATT_{\text{rup-IPW}}^{t^-,t^+}(k, h) = ATT_{\text{rup}}^{t^-,t^+}(k, h).$$

□

A.17. Demostración de los teoremas de inferencia

En este anexo se presentan las demostraciones correspondientes a los Teoremas 3.4 y 3.6, relativos al análisis de inferencia para los estimadores $ATT_{\text{rup-DR}}^{t^-,t^+}(k, h)$. Ambas pruebas se desarrollan bajo los supuestos de independencia por unidad, regularidad de momentos y existencia de soporte común en las covariables.

A.17.1. Demostración del Teorema 3.4

El objetivo es establecer la convergencia en distribución:

$$\sqrt{N_{k,h}} \left(\widehat{ATT}_{\text{rup-DR}}^{t^-,t^+}(k, h) - ATT_{\text{rup-DR}}^{t^-,t^+}(k, h) \right) \xrightarrow{d} \mathcal{N}(0, \sigma_{k,h}^2).$$

Paso 1. Representación en promedio de influencias. Bajo los supuestos de independencia entre unidades y regularidad de los estimadores de primer paso, el estimador DR admite una expansión tipo von Mises:

$$\widehat{ATT}_{\text{rup-DR}}^{t^-,t^+}(k, h) - ATT_{\text{rup-DR}}^{t^-,t^+}(k, h) = \frac{1}{N_{k,h}} \sum_{u=1}^{N_{k,h}} \psi_{k,h}(Y_u, X_u, D_u) + o_p(N_{k,h}^{-1/2}),$$

donde $\psi_{k,h}(\cdot)$ es la función de influencia individual del estimador. La forma cerrada de $\psi_{k,h}$ se obtiene al derivar la expresión poblacional del estimador DR con respecto a la medida empírica de probabilidad.

Paso 2. Aplicación del teorema del límite central. Dado que las observaciones (Y_u, X_u, D_u) son i.i.d. por unidad, la media de $\psi_{k,h}$ converge a cero y su varianza es finita, por lo que el teorema del límite central se aplica directamente:

$$\sqrt{N_{k,h}} \overline{\psi}_{k,h} \xrightarrow{d} \mathcal{N}(0, \sigma_{k,h}^2), \quad \text{con} \quad \sigma_{k,h}^2 = \text{Var}(\psi_{k,h}(Y, X, D)).$$

Por continuidad asintótica, el término residual $o_p(N_{k,h}^{-1/2})$ no altera la distribución límite.

Paso 3. Varianza agregada y estabilidad. Cuando los estimadores por ventanas se combinan en los agregados $\widehat{ATT}(k, h)$, $\widehat{ATT}_K(k)$, $\widehat{ATT}_H(h)$ y $\widehat{ATT}_{\text{global}}$, la linealidad de los promedios garantiza que todos los agregados heredan la misma normalidad asintótica, con varianza compuesta definida como combinación cuadrática de las varianzas y covarianzas entre ventanas.

□

A.17.2. Demostración del Teorema 3.6

Esta demostración sigue las ideas desarrolladas por Callaway y Sant’Anna (2021) para los estimadores de efectos promedio por grupo y periodo, extendidas aquí al caso con covariables y ventanas dinámicas (t^-, t^+) . El objetivo es justificar que el remuestreo por unidades proporciona una aproximación válida de la distribución asintótica del estimador DR.

Idea general. Dado que las observaciones son dependientes en el tiempo dentro de cada unidad pero independientes entre unidades, el bootstrap debe preservar la estructura temporal de cada trayectoria. Por ello, se utiliza un *bootstrap agrupado* (cluster bootstrap) en el que las unidades son remuestreadas con reemplazo y cada trayectoria completa se mantiene intacta.

Paso 1. Re-muestreo de unidades. Sea \mathcal{U}^* el conjunto de índices remuestreados con reemplazo a partir de las N unidades originales. Para cada muestra bootstrap, se calcula nuevamente el estimador $\widehat{ATT}_{\text{rup-DR}}^{t^-, t^+*}(k, h)$ usando exactamente los mismos pesos, modelos y normalizaciones que en la muestra original.

Paso 2. Varianza empírica bootstrap. La varianza bootstrap se define como

$$\hat{\sigma}_{k,h}^{*2} = \text{Var}^* \left(\widehat{ATT}_{\text{rup-DR}}^{t^-, t^+*}(k, h) \right),$$

donde $\text{Var}^*(\cdot)$ es la varianza bajo las réplicas bootstrap. Bajo condiciones de regularidad estándar (consistencia de primer paso, momento finito de orden dos y convergencia de la función de influencia), esta varianza converge en probabilidad a la varianza asintótica $\sigma_{k,h}^2$.

Paso 3. Validez de los intervalos. Como resultado, los intervalos de confianza construidos a partir del bootstrap agrupado son asintóticamente válidos:

$$\Pr\left(ATT_{\text{rup-DR}}^{t^-,t^+}(k, h) \in \left[\widehat{ATT}_{\text{rup-DR}}^{t^-,t^+}(k, h) \pm z_{1-\alpha/2} \widehat{\sigma}_{k,h}^*\right]\right) \rightarrow 1 - \alpha.$$

Este resultado extiende de manera directa los argumentos de Callaway y Sant'Anna (2021), mostrando que el remuestreo por unidad captura la dependencia intratemporal y las covarianzas inducidas por ventanas superpuestas.

□

Apéndice B

Anexo: Simulaciones

B.1. Generación de datos simulados (PGD)

Trabajamos con un panel sencillo: 60 unidades ($i = 1, \dots, 60$) y 20 periodos ($t = 1, \dots, 20$). El panel trae:

- y : la variable de interés (resultado),
- H : la intensidad acumulada del tratamiento,
- D : la dosis cuando aplique.

De forma general, en cada periodo actualizamos la trayectoria sumando una pendiente y un efecto:

$$y_{it} = y_{i,t-1} + m_{it} + E_{it},$$

donde m_{it} es la pendiente en (i, t) y E_{it} es el efecto simulado en cada intervención.

Dónde están las intervenciones: generación de H

Lo primero es fijar dónde ocurren las intervenciones a lo largo del tiempo. Para esto, trabajamos con una regla simple y homogénea para todas las unidades:

- Una fracción de las unidades se deja como *nunca tratadas* (aquí 20%): en ellas $H_{it} = 0$ para todo t .

- En las unidades tratables, a cada unidad i se le fija una única probabilidad $p_i \in [0.20, 0.25]$. En cada periodo elegible $t \in \{T_{\min}, \dots, T\}$ se realiza un sorteo independiente con probabilidad p_i de intervención. Si hay intervención y aún $H_{i,t-1} < H_{\max}$, entonces $H_{it} = H_{i,t-1} + 1$; en caso contrario (si no hay intervención o ya se alcanzó H_{\max}), $H_{it} = H_{i,t-1}$. El mismo p_i se usa para todos los periodos elegibles de esa unidad.
- Por comodidad de las gráficas, las intervenciones sólo se permiten a partir de $t = 3$ ($T_{\min} = 3$). Esto es una decisión práctica; teóricamente podría permitirse desde $t = 1$.
- Para no dejar unidades tratables sin eventos, si al finalizar la ventana ninguna intervención ocurrió, se asegura una en un periodo elegible.
- La intensidad acumulada H arranca en cero y aumenta de uno en uno cada vez que ocurre una intervención, con tope $H_{\max} = 6$.

Con esto queda determinado H : para cada (i, t) se reconoce si hubo intervención (cuando $H_{it} > H_{i,t-1}$) y cuántas acumula cada unidad en el panel.

Valores iniciales de y en $t = 1$

Para arrancar, a cada unidad i se le asigna:

$$y_{i1} = a_0 + a_1(i - 1) + \nu_i, \quad \nu_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{y0}^2),$$

con $a_0 = 0.0$, $a_1 = 1.0$ y $\sigma_{y0} = 0.2$. El escalón a_1 es solo para que las gráficas se vean ordenadas; el ruido ν_i se incluye para darle realismo a los datos (evitar separaciones demasiado perfectas).

B.1.1. Sin ruptura

En el caso *sin ruptura*, la pendiente no depende de H . Usamos la misma regla para todo (i, t) y le añadimos un ruidito pequeño para que las trayectorias no queden perfectas. En concreto, la pendiente es

$$m_{it} \stackrel{iid}{\sim} \mathcal{N}(0.2, 0.02^2).$$

En un esquema con intervalos de tiempo generales, la contribución sería $m_{it} \cdot \Delta t$; como aquí los periodos avanzan de uno en uno ($\Delta t = 1$), se omite ese factor.

B.1.2. Con ruptura

Aquí la pendiente cambia cuando aumenta la intensidad acumulada H : antes de la primera intervención ($H = 0$) vale una pendiente; entre la primera y la segunda ($H = 1$) toma otra; entre la segunda y la tercera ($H = 2$) otra, y así sucesivamente. Para que la ruptura se vea clara en las gráficas, alternamos signos de forma deliberadamente marcada (es sólo un recurso visual).

Regla usada (con ruido pequeño en cada nivel):

$$\begin{aligned} H_{it} = 0 : & \quad m_{it} \sim \mathcal{N}(0.2, 0.02^2), \\ H_{it} = 1 : & \quad m_{it} \sim \mathcal{N}(-0.3, 0.03^2), \\ H_{it} = 2 : & \quad m_{it} \sim \mathcal{N}(0.3, 0.04^2), \\ H_{it} = 3 : & \quad m_{it} \sim \mathcal{N}(-0.2, 0.05^2). \end{aligned}$$

A partir de ahí, los niveles se repiten de forma cíclica según H_{it} mód 4. Si los intervalos de tiempo fuesen generales, la contribución sería $m_{it} \cdot \Delta t$; aquí $\Delta t = 1$, así que se omite.

B.1.3. Sin dosis

El efecto simulado entra solo cuando hay intervención (cuando $H_{it} > H_{i,t-1}$). Para capturar variación realista usamos dos capas de heterogeneidad:

- *Entre unidades:* cada unidad tiene una media propia $\mu_i = \mu_0 + \eta_i$, con $\eta_i \sim \mathcal{N}(0, \sigma_{\text{unit}}^2)$. En esta corrida: $\mu_0 = 1$ y $\sigma_{\text{unit}} = 0.3$.
- *Dentro de la unidad y en el tiempo:* en cada (i, t) se sortean $e_{it} \sim \mathcal{N}(\mu_i, \sigma_{\text{jump}}^2)$, con $\sigma_{\text{jump}} = 0.5$. Para evitar valores negativos, se activó la opción de tomar valor absoluto.

De forma equivalente, el efecto latente se descompone como

$$e_{it} = \mu_0 + \eta_i + \xi_{it}, \quad \eta_i \sim \mathcal{N}(0, \sigma_{\text{unit}}^2), \quad \xi_{it} \sim \mathcal{N}(0, \sigma_{\text{jump}}^2).$$

La activación del efecto es puntual en el periodo de intervención:

$$E_{it} = \begin{cases} e_{it}, & \text{si } H_{it} - H_{i,t-1}, \\ 0, & \text{en otro caso.} \end{cases}$$

B.1.4. Con dosis

Cuando hay intervención (es decir, cuando $H_{it} > H_{i,t-1}$), además del efecto E_{it} sorteamos una *dosis* para ese evento. Usamos un soporte discreto

$$D_{it} \in \{0.5, 1.0, 1.5, 2.0\},$$

con probabilidades proporcionales a $\{1, 3, 2, 1\}$ (se normalizan a 1). En periodos sin intervención tomamos $D_{it} = 0$. Las dosis se sortean de manera independiente en cada evento.

En las variantes *con dosis*, el efecto que entra a la trayectoria se pondera por la dosis:

$$y_{it} = y_{i,t-1} + m_{it} + E_{it} D_{it}.$$

B.2. Implementación del diseño sin ruptura de tendencia y de comparación por cohortes de unidades

En esta sección implemento, sobre el panel simulado de la Sección 2.1, el diseño *sin ruptura de tendencia* para identificar el efecto marginal de pasar de $h-1$ a h . Trabajo con un panel largo con columnas i , t , y y H , generado y explicado en el Anexo B.1 (60 unidades y 20 periodos).

Llamo *cohorte* (k, h) al conjunto de unidades que alcanzan el nivel h en el instante k mediante un salto $+1$. Para cada cohorte, construyo ventanas $[t^-, t^+]$ y comparo cambios en y entre: (a) tratadas efectivas —acumulan neto $+1$ en la ventana y alcanzan h en k — y (b) controles estables —mantienen H constante y menor que h en la misma ventana—. Sobre esta base, más adelante agrego resultados por (k, h) , por k , por h y a nivel global.

Conjuntos I_0 e I_k Recordemos que I_0 agrupa las unidades que nunca reciben tratamiento, mientras que $I_k = \{i : H_{i,k} - H_{i,k-1} = 1\}$ reúne las unidades que presentan *algún* tratamiento en $t = k$. A continuación se listan los elementos de I_0 y de cada I_k :

I_0 (nunca tratadas) — $ I_0 = 12$: {3, 5, 8, 18, 20, 23, 33, 47, 49, 53, 55, 60}

Cuadro B.1: Unidades en I_k por instante k

k	$ I_k $	Unidades en I_k
1	0	\emptyset
2	0	\emptyset
3	12	{4, 10, 11, 12, 13, 21, 24, 26, 27, 29, 39, 59}
4	7	{2, 9, 19, 25, 35, 51, 54}
5	11	{1, 10, 13, 17, 22, 25, 35, 36, 37, 48, 50}
6	7	{9, 13, 21, 22, 40, 45, 46}
7	11	{1, 4, 6, 11, 12, 14, 19, 21, 29, 39, 59}
8	10	{9, 13, 17, 26, 28, 38, 39, 41, 48, 52}
9	15	{9, 10, 11, 12, 21, 28, 30, 31, 32, 37, 39, 44, 45, 54, 58}
10	11	{2, 11, 15, 25, 26, 27, 32, 34, 36, 50, 59}
11	11	{1, 2, 11, 21, 22, 25, 28, 36, 40, 45, 50}
12	11	{4, 9, 17, 22, 29, 34, 38, 42, 50, 54, 57}
13	11	{7, 11, 19, 28, 32, 36, 37, 45, 48, 52, 56}
14	9	{6, 7, 12, 16, 17, 24, 43, 46, 56}
15	8	{7, 12, 16, 42, 43, 50, 51, 59}
16	10	{4, 10, 25, 26, 31, 37, 38, 44, 50, 57}
17	7	{7, 21, 32, 36, 37, 41, 57}
18	11	{15, 19, 26, 28, 30, 31, 39, 40, 42, 51, 57}
19	8	{6, 15, 19, 26, 29, 52, 54, 59}
20	7	{6, 22, 24, 30, 35, 40, 42}

Ventanas y contrafactual puro (I_0)

De acuerdo con la Proposición 2.1, el efecto por par $\widehat{ATT}_{m,n}^\circ(k)$ se estima usando controles nunca tratados (I_0). En esta corrida se obtuvieron **2124 pares válidos** (m, n, k). A modo ilustrativo, se muestran dos ejemplos (en ambos, t_{pre} y t_{pos} son los límites de la ventana alrededor de k):

Cuadro B.2: Ejemplos de $\widehat{ATT}_{m,n}^\circ(k)$ con contrafactual puro

k	m	n	$[t_{\text{pre}}, t_{\text{pos}}]$	$\widehat{ATT}_{m,n}^\circ(k)$
3	4	3	[1, 6]	0.821304
3	4	5	[1, 6]	0.839839

Comparaciones entre tratadas ($ATT_{m,n}^+(k)$)

Siguiendo la identificación para comparaciones *entre tratadas*, el efecto por par $\widehat{ATT}_{m,n}^+(k)$ se calcula contrastando la unidad m que salta en k con una unidad n que también está tratada pero no cambia de nivel al cruzar k . En esta corrida se obtuvieron **4 183 pares válidos** (m, n, k). A modo ilustrativo:

Cuadro B.3: Ejemplos de $\widehat{ATT}_{m,n}^+(k)$ (entre tratadas)

k	m	n	$[t_{\text{pre}}, t_{\text{pos}}]$	$\widehat{ATT}_{m,n}^+(k)$
3	4	1	[1, 4]	0.806424
3	4	2	[1, 3]	0.855209

Agregados por k : contrafactual puro vs. entre tratadas. Como chequeo de coherencia, comparo los promedios por instante k obtenidos con contrafactual puro, $\widehat{ATT}^\circ(k)$, frente a los obtenidos entre tratadas, $\widehat{ATT}^+(k)$. La Tabla B.4 resume, para cada k , el número de pares válidos en cada esquema y los estimadores agregados; también reporto la distancia porcentual simétrica (DPS_k) entre ambos. En general, los valores son del mismo orden; por ejemplo, en $k = 13$ casi coinciden ($\text{DPS} \approx 0.33\%$), mientras que en $k = 9$ y $k = 14$ las diferencias son mayores ($\text{DPS} \approx 32\%$ y 19%).

Cuadro B.4: Agregados por k : $\widehat{ATT}^\circ(k)$ vs. $\widehat{ATT}^+(k)$

k	N_k°	N_k^+	$\widehat{ATT}^\circ(k)$	$\widehat{ATT}^+(k)$	DPS_k (%)
3	144	432	0.8135	0.7553	7.4228
4	84	203	1.0912	1.0094	7.7880
5	132	302	0.6662	0.7068	5.9185
6	84	212	0.4270	0.4723	10.0747
7	132	318	1.7208	1.4500	17.0816
8	120	242	1.6085	1.3944	14.2612
9	180	316	1.0800	0.7813	32.0915
10	132	233	0.8104	0.7662	5.6066
11	132	338	1.3740	1.2255	11.4215
12	132	231	0.8256	0.7893	4.5024
13	132	226	1.2383	1.2424	0.3329
14	108	106	1.3259	1.0919	19.3552
15	96	153	0.9397	0.8965	4.7003
16	120	208	0.9613	0.9947	3.4204
17	84	186	1.0311	0.9719	5.9131
18	132	187	0.9173	0.8550	7.0274
19	96	178	1.1259	1.1141	1.0530
20	84	112	0.7690	0.8091	5.0800

Ejemplos. (i) En $k = 13$: $\widehat{ATT}^\circ(13) = 1.2383$ y $\widehat{ATT}^+(13) = 1.2424$ (DPS = 0.33%). (ii) En $k = 9$: $\widehat{ATT}^\circ(9) = 1.0800$ y $\widehat{ATT}^+(9) = 0.7813$ (DPS = 32.09%).

ATT por k y estimador global.

A partir de todas las comparaciones válidas (tanto con contrafactual puro como entre tratadas) obtengo, para cada instante k , un único $\widehat{ATT}(k)$. El promedio global se calcula ponderando por $\widehat{\Omega}(k)$, proporcionales al número de comparaciones efectivas en cada k .

Cuadro B.5: Inferencia por instante k

k	$\widehat{ATT}(k)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	w_k (%)	sig
3	0.7748	0.1487	0.4958	1.0795	9.13	***
4	1.0425	0.1496	0.7389	1.2928	4.55	***
5	0.6905	0.1662	0.3975	1.0522	6.88	***
6	0.4516	0.1163	0.2882	0.7745	4.69	***
7	1.5927	0.1903	1.1305	1.8668	7.13	***
8	1.5187	0.2984	0.8463	2.0048	5.74	***
9	0.9830	0.1766	0.6117	1.3141	7.86	***
10	0.7954	0.1047	0.6081	0.9904	5.79	***
11	1.3003	0.1852	0.8904	1.6103	7.45	***
12	0.8116	0.1753	0.4787	1.1436	5.76	***
13	1.2400	0.1968	0.8223	1.5878	5.68	***
14	1.2688	0.2102	0.8837	1.6270	3.39	***
15	0.9216	0.1694	0.6491	1.3113	3.95	***
16	0.9765	0.2233	0.5430	1.4071	5.20	***
17	1.0011	0.1643	0.6272	1.3128	4.28	***
18	0.8879	0.1004	0.6612	1.0524	5.06	***
19	1.1196	0.3503	0.5028	1.8031	4.34	***
20	0.7873	0.1258	0.5153	1.0305	3.11	***

Estimación global ponderada: $\widehat{ATT}_{global} = 1.0161$, $SE_{boot} = 0.0444$, IC95 % = [0.9186, 1.0967] ***.

Agregación por nivel de intensidad (h).

Se promedia el efecto dentro de cada nivel de tratamiento h , con pesos normalizados $\widehat{\Omega}^{[h]}$.

Cuadro B.6: Inferencia por nivel h

h	$\widehat{ATT}^{[h]}$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	w_h (%)	sig
1	1.0986	0.1193	0.8614	1.3197	24.92	***
2	1.0435	0.0939	0.8572	1.2216	23.64	***
3	1.1191	0.1085	0.9070	1.3201	21.53	***
4	0.7451	0.1358	0.5053	1.0271	16.74	***
5	0.9800	0.1997	0.6050	1.3622	10.02	***
6	1.1321	0.1824	0.8601	1.5062	3.14	***

Estimación global ponderada por $\hat{\Omega}^{[h]}$: $\widehat{ATT}_{\text{global}}^{[h]} = 1.0199$, $\text{SE}_{\text{boot}} = 0.0525$, $\text{IC95 \%} = [0.9168, 1.1204]$ ***.

Cuadro B.7: Comparación de estimadores globales

Estimador	Valor
$\widehat{ATT}_{\text{global}}$ (por k)	1.016144
$\widehat{ATT}_{\text{global}}^{[h]}$ (por h)	1.019954
$\text{DPS}_{\text{global}}$ (%)	0.3742

B.3. Implementación del diseño por cortes temporales (sin ruptura de tendencia)

En esta sección implemento el diseño *por cortes temporales* para identificar el efecto marginal de pasar de $h-1$ a h en el instante k , bajo el supuesto *sin ruptura de tendencia*. Trabajo sobre el *mismo* panel simulado descrito en el Anexo B.1 y utilizado en el anexo anterior (Sección B.2), de modo que cualquier diferencia en resultados provenga del *diseño* y no de los datos. Para facilitar la lectura, recuerdo que el panel es largo con columnas \mathbf{i} , \mathbf{t} , \mathbf{y} y \mathbf{H} , y contiene $I = 60$ unidades y $T = 20$ periodos.

Defino, para cada par candidato (k, h) , conjuntos efectivos de tratadas y controles según las reglas de ventanas de la Sección 2.2: para cada $t^- \in T^-(k, h)$ se construye $t^+ \in T^+(k, h | t^-)$ y se estima el efecto por cambios en y entre (a) tratadas efectivas y (b) controles estables. Más adelante agrego resultados por (k, h) , por k , por h y a nivel global.

Conjuntos y ventanas.

En total se identificaron 1 551 ventanas válidas $[t^-, t^+]$. El conteo agregado de tratadas a lo largo de todas las ventanas es

$$\sum_{\text{ventanas válidas } (k, h, t^-, t^+)} |I^{k, h}(t^-, t^+)| = 2\,770,$$

mientras que el conteo agregado de controles es

$$\sum_{\text{ventanas válidas } (k, h, t^-, t^+)} |J^{k, h}(t^-, t^+)| = 31\,090.$$

Nota. Una misma unidad puede aparecer en múltiples ventanas; los conteos son agregados sobre todas las ventanas válidas.

En la Tabla B.8 se muestran algunos ejemplos representativos de ventanas válidas y sus tamaños efectivos.

Cuadro B.8: Ejemplos de ventanas válidas y tamaños efectivos por (k, h, t^-, t^+)

k	h	t^-	t^+	$ \mathbf{I}^{\mathbf{k},\mathbf{h}}(\mathbf{t}^-, \mathbf{t}^+) $	$ \mathbf{J}^{\mathbf{k},\mathbf{h}}(\mathbf{t}^-, \mathbf{t}^+) $
3	1	1	3	12	48
3	1	1	4	12	41
3	1	1	5	10	34
3	1	1	6	9	31
3	1	1	7	3	29
3	1	1	8	2	25
3	1	1	9	2	20
3	1	1	10	1	18
3	1	1	11	1	18
3	1	1	12	1	16
3	1	1	13	1	14
3	1	2	3	12	48
3	1	2	4	12	41
3	1	2	5	10	34
3	1	2	6	9	31

Estimación elemental por ventana

De acuerdo con la Proposición 2.4 (bajo el Supuesto 2.5), para toda ventana válida (k, h, t^-, t^+) el estimando $\widehat{ATT}^{t^-, t^+}(k, h)$ se identifica tal como se establece en (2.2.6) (y de forma equivalente en (2.2.7)). Aplicamos dicha identificación a cada ventana construida en la subsección previa.

A continuación se muestran algunos ejemplos representativos con sus tamaños efectivos de tratadas n_T y controles n_C .

Cuadro B.9: Ejemplos de $\widehat{ATT}^{t^-,t^+}(k, h)$ por ventana

Ventana				Tamaños efectivos		Estimador
k	h	t^-	t^+	n_T	n_C	$\widehat{ATT}^{t^-,t^+}(k, h)$
3	1	1	3	12	48	0.733986
3	1	1	4	12	41	0.734633
3	1	1	5	10	34	0.670574
3	1	1	6	9	31	0.709707
3	1	1	7	3	29	0.939501
3	1	1	8	2	25	1.211210
3	1	1	9	2	20	1.205640
3	1	1	10	1	18	1.534698
4	1	1	4	7	41	0.935772
4	1	1	5	5	34	0.977023
4	1	1	6	4	31	1.075596
4	1	1	7	3	29	1.069924
4	1	2	4	7	41	0.933328
4	1	2	5	5	34	0.970501
4	1	2	6	4	31	1.069669
4	1	2	7	3	29	1.069984

Diagnóstico PRE (placebos)

Para cada par (k, h) y cada anclaje previo t^{--} anterior a t^- , se calcula el placebo $\widehat{ATT}_{t^{--},t^-}^{\text{pre}}(k, h)$ reusando los conjuntos efectivos definidos para la ventana principal.

Se obtuvieron 989 placebos PRE.

Cuadro B.10: Ejemplos de placebos PRE por ventana (k, h, t^{--}, t^-)

Ventana PRE				Tamaños efectivos		Estimador placebo
k	h	t^{--}	t^-	n_T^{pre}	n_C^{pre}	$\widehat{ATT}_{t^{--}, t^-}^{\text{pre}}(k, h)$
3	1	1	2	12	48	0.003809
4	1	1	2	7	41	0.002444
4	1	1	3	7	41	0.004849
4	1	2	3	7	41	0.002406
5	1	1	2	7	34	0.014085
8	2	6	7	3	36	-0.006917
8	3	6	7	1	40	0.012709
8	4	6	7	1	40	-0.004326
9	1	1	2	5	20	0.009990
9	1	1	3	5	20	0.011288

Para evaluar magnitud relativa, se considera el cociente

$$\frac{|\widehat{ATT}_{t^{--}, t^-}^{\text{pre}}(k, h)|}{\mathbb{E}[|\Delta y_C|]} \quad (\text{B.3.1})$$

donde $\mathbb{E}[|\Delta y_C|]$ es el cambio típico en controles de la misma ventana y se clasifica en un “semáforo”: VERDE ($\leq 10\%$), AMARILLO (10–25%), ROJO ($> 25\%$).

Cuadro B.11: Magnitud relativa y semáforo de placebos PRE

Ventana PRE				R (%)	Semáforo
k	h	t^{--}	t^-		
3	1	1	2	1.91	VERDE ($\leq 10\%$)
4	1	1	2	1.23	VERDE ($\leq 10\%$)
4	1	1	3	1.21	VERDE ($\leq 10\%$)
4	1	2	3	1.20	VERDE ($\leq 10\%$)
5	1	1	2	7.17	VERDE ($\leq 10\%$)
5	1	1	3	3.62	VERDE ($\leq 10\%$)
5	1	2	3	0.16	VERDE ($\leq 10\%$)
5	1	1	4	0.13	VERDE ($\leq 10\%$)
5	1	2	4	3.69	VERDE ($\leq 10\%$)
5	1	3	4	7.53	VERDE ($\leq 10\%$)
6	1	2	3	11.42	AMARILLO (10–25%)

Cuadro B.12: Distribución del semáforo (PRE)

Categoría	Porcentaje
VERDE ($\leq 10\%$)	91.3 %
AMARILLO (10-25 %)	8.5 %
ROJO ($> 25\%$)	0.2 %

Diagnóstico POST (placebos)

Para cada par (k, h) y cada anclaje posterior t^{++} subsiguiente a t^+ , se calcula el placebo $\widehat{ATT}_{t^+, t^{++}}^{\text{post}}(k, h)$ reusando los conjuntos efectivos de la ventana principal. Se obtuvieron 912 placebos POST (sin duplicados).

Cuadro B.13: Ejemplos de placebos POST por ventana (k, h, t^+, t^{++})

Ventana POST				Tamaños efectivos		Estimador placebo
k	h	t^+	t^{++}	n_T^{post}	n_C^{post}	$\widehat{ATT}_{t^+, t^{++}}^{\text{post}}(k, h)$
3	1	4	5	10	34	0.008551
3	1	4	6	9	31	0.010413
3	1	4	7	3	29	-0.017441
3	1	4	8	2	25	-0.021384
3	1	4	9	2	20	-0.028562
3	1	4	10	1	18	-0.033275
3	1	4	11	1	18	-0.070107
3	1	4	12	1	16	-0.020119
3	1	4	13	1	14	-0.008117
3	1	5	6	9	31	0.004049

Para evaluar magnitud relativa, se usa $R = |\widehat{ATT}_{t^+, t^{++}}^{\text{post}}(k, h)| / \mathbb{E}[|\Delta y_C|]$, donde $\mathbb{E}[|\Delta y_C|]$ es el cambio típico de controles en la misma ventana POST; se clasifica en: VERDE ($\leq 10\%$), AMARILLO (10-25 %), ROJO ($> 25\%$).

Cuadro B.14: Magnitud relativa y semáforo de placebos POST

Ventana POST				R (%)	Semáforo
k	h	t^+	t^{++}		
3	1	4	5	4.26	VERDE ($\leq 10\%$)
3	1	4	6	2.58	VERDE ($\leq 10\%$)
3	1	4	7	2.91	VERDE ($\leq 10\%$)
3	1	4	8	2.68	VERDE ($\leq 10\%$)
3	1	4	9	2.87	VERDE ($\leq 10\%$)
3	1	4	10	2.77	VERDE ($\leq 10\%$)
3	1	4	11	4.98	VERDE ($\leq 10\%$)
3	1	4	12	1.26	VERDE ($\leq 10\%$)
3	1	4	13	0.45	VERDE ($\leq 10\%$)
3	1	5	6	2.00	VERDE ($\leq 10\%$)

Cuadro B.15: Distribución del semáforo (POST)

Categoría	Porcentaje
VERDE ($\leq 10\%$)	89.5 %
AMARILLO (10-25 %)	8.6 %
ROJO ($> 25\%$)	2.0 %

Agregación por (k, h)

Aplicando la regla de ponderación de (2.2.10) y su versión normalizada dentro de cada (k, h) , agrego las estimaciones por ventana $\widehat{ATT}^{t^-, t^+}(k, h)$ para obtener $\widehat{ATT}(k, h)$. La tabla reporta, además, el número de ventanas agregadas y la suma $\sum \omega$ previa a la normalización (solo a modo informativo).

Cuadro B.16: Efecto agregado por par (k, h) con pesos de (2.2.10)

Par		Acumulados		Estimación
k	h	# ventanas	$\sum \omega$	$\widehat{ATT}(k, h)$
3	1	22	87.486644	0.837022
4	1	33	77.366475	1.098082
5	1	20	78.373457	0.576396
5	2	19	28.003107	0.937030
6	1	40	72.863378	0.384045
6	2	8	11.481758	0.503173
6	3	2	1.958754	1.437773
7	1	84	114.781906	2.040161
7	2	23	83.289309	1.298448
7	3	2	1.955202	1.729794
8	1	63	116.761194	1.995619
8	2	19	34.108725	1.291943
8	3	2	2.897178	0.772901
8	4	26	24.590129	0.161753
9	1	96	232.434943	1.245279

Lectura. (i) $\omega(t^-, t^+ | k, h) = \frac{|\mathcal{I}^{k,h}(t^-, t^+)| |\mathcal{J}^{k,h}(t^-, t^+)|}{|\mathcal{I}^{k,h}(t^-, t^+)| + |\mathcal{J}^{k,h}(t^-, t^+)|}$ como en (2.2.10); (ii) la agregación usa $\hat{\omega} = \omega / \sum_{(u^-, u^+) \in \mathcal{W}(k, h)} \omega$ dentro de (k, h) ; (iii) los promedios por k y por h , y el global, se construirán más adelante con los pesos $p(\cdot)$ de (2.2.11)–(2.2.13).

Agregados por instante k (inferencia)

De acuerdo con (2.2.11), los promedios por instante k se obtienen con los pesos $p(h | k)$. La tabla reporta punto, SE_{boot} , intervalo al 95 % y código de significancia (bootstrap por unidades).

Cuadro B.17: Inferencia por instante k

k	$\widehat{ATT}_K(k)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	sig
3	0.837022	0.159346	0.494569	1.090989	***
4	1.098082	0.172214	0.688616	1.316256	***
5	0.707535	0.178278	0.447589	1.164678	***
6	0.585633	0.125223	0.292961	0.787229	***
7	1.472518	0.145668	1.157020	1.685531	***
8	1.356586	0.218132	0.996401	1.805057	***
9	0.724828	0.115396	0.507158	0.963995	***
10	0.804357	0.122355	0.558590	1.110762	***
11	1.429169	0.249390	0.749239	1.613265	***
12	0.939340	0.154690	0.644160	1.237398	***
13	1.219886	0.200128	0.841826	1.583857	***
14	1.289593	0.214422	0.845700	1.629365	***
15	1.029119	0.213739	0.668369	1.576464	***
16	0.956727	0.191153	0.437519	1.109671	***
17	0.978248	0.240036	0.430431	1.316967	***

Agregados por nivel h (inferencia)

Según (2.2.12), los promedios por nivel h usan $p(k | h)$. Incluyo los pesos marginales $p(h)$ como w_h en porcentaje.

Cuadro B.18: Inferencia por nivel h

h	$\widehat{ATT}_H(h)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	w_h (%)	sig
1	1.029295	0.075466	0.848963	1.152194	27.12	***
2	1.005194	0.085944	0.850335	1.167594	25.99	***
3	1.131724	0.094928	0.896121	1.251719	21.47	***
4	0.768489	0.110978	0.515805	0.929746	15.25	***
5	1.018925	0.168675	0.720693	1.366656	7.91	***
6	1.177504	0.203099	0.720172	1.665167	2.26	***

Estimación global

Con (2.2.13), el promedio ponderado conjunto entrega:

$$\widehat{ATT}_{global} = 1.007767, \quad SE_{boot} = 0.050465, \quad IC95 \% = [0.887980, 1.080875] ***.$$

Chequeo de invarianza del orden: $\sum_k p(k) \widehat{ATT}_K(k) = 1.007767$ y $\sum_h p(h) \widehat{ATT}_H(h) = 1.007767$.

B.4. Implementación del diseño con ruptura y comparación unidad–unidad

Nota B.1. Ejecución en Colab, SEED=20250826, rutas fijadas y carpetas de salida creadas. Archivo: `panel_I60_T20_ruptura_sindosis.csv`. Panel ordenado y tipificado con columnas (i, t, y, H) sin NA.

Eventos, controles y ventanas con ruptura

Sea $\mathcal{I}^{k,h}$ el conjunto de unidades que pasan de $h-1$ a h en $t=k$, y $\mathcal{C}^{k,h}$ los controles que en ese mismo k se mantienen en $h-1$. Para cada pareja (m, n) con $m \in \mathcal{I}^{k,h}$, $n \in \mathcal{C}^{k,h}$, se identifican los extremos contiguos de la ventana: $t_{m,n}^{\text{pre},\text{rup}}(k)$ (ambos en $h-1$ hacia la izquierda) y $t_{m,n}^{\text{pos},\text{rup}}(k)$ (desde k hacia la derecha, m en h y n en $h-1$). En esta corrida: 177 eventos y 2784 parejas válidas.

Cuadro B.19: Ejemplos de parejas y ventanas [PRE, POST]

m	n	k	h	t^{pre}	t^{pos}	pares en celda
4	1	3	1	1	4	576 en (3, 1)
4	2	3	1	1	3	287 en (4, 1)
4	3	3	1	1	6	238 en (5, 1)

Efecto por pareja y agregados

Para cada (m, n) :

$$\widehat{ATT}_{m,n}^{\text{rup}}(k) = \mathbb{E}_{t^+ \in [k, t^{\text{pos}}]}[y_{m,t^+} - y_{n,t^+}] - \mathbb{E}_{t^- \in [t^{\text{pre}}, k-1]}[y_{m,t^-} - y_{n,t^-}].$$

Agregación en (k, h) con pesos $\omega_{m,n} = (t^{\text{pos}} - k + 1)(k - t^{\text{pre}})$ (normalizados en la celda). Entre celdas, se usa $W(k, h) = \frac{T(k,h)C(k,h)}{T(k,h)+C(k,h)}$ para obtener $\widehat{ATT}_K(k)$, $\widehat{ATT}_H(h)$ y el global. Inferencia por bootstrap por unidad (200 réplicas).

Cuadro B.20: $\widehat{ATT}^{\text{rup}}(k, h)$ e inferencia (muestra)

k	h	#parejas	T	C	W	$\widehat{ATT}^{\text{rup}}$	SE_{boot}	IC95 %
3	1	576	12	48	9.6000	-0.6029	0.1280	[-0.8384, -0.3437]
5	2	60	4	15	3.1579	2.4433	0.3542	[1.8374, 3.1854]
6	1	93	3	31	2.7353	-1.1367	0.3660	[-1.5963, -0.2397]

Cuadro B.21: Colapsos por k y por h (muestra)

k	$\widehat{ATT}_K(k)$	SE	IC95 %	h	$\widehat{ATT}_H(h)$	SE	IC95 %
3	-0.6029	0.1280	[-0.8384, -0.3437]	1	-0.4617	0.1027	[-0.6466, -0.2416]
7	1.5527	0.4454	[0.7328, 2.4933]	2	2.2695	0.1422	[1.9307, 2.4764]
8	0.9684	0.4047	[0.3408, 1.7619]	3	0.0927	0.1028	[-0.0777, 0.3294]

Promedio global ponderado: $\widehat{ATT}_{\text{global}} = 0.7241$, $SE_{\text{boot}} = 0.0618$, IC95 % = [0.5662, 0.8048].

Efecto instantáneo (salto en k)

Para cada pareja, el instantáneo se define como en (3.1.10) restringiendo el POST a $t^+ = k$ y manteniendo la misma ventana PRE:

$$\widehat{ATT}_{m,n}^{\text{nivel}}(k) = (y_{m,k} - y_{n,k}) - \mathbb{E}_{t^- \in [t^{\text{pre}}, k-1]}[y_{m,t^-} - y_{n,t^-}].$$

Agregación en (k, h) con pesos $(k - t^{\text{pre}})$ y colapsos usando los mismos $W(k, h)$. Bootstrap por unidad (200 réplicas).

Cuadro B.22: $\widehat{ATT}^{\text{nivel}}(k, h)$ e inferencia (muestra)

k	h	T	C	W	$\widehat{ATT}^{\text{nivel}}$	SE_{boot}	IC95 %
3	1	12	48	9.6000	0.2355	0.1254	[-0.0237, 0.4798]
5	2	4	15	3.1579	1.7018	0.2647	[1.2767, 2.5334]
6	2	3	19	2.5909	0.9946	0.1697	[0.7191, 1.2467]

Cuadro B.23: Colapsos por k y por h (muestra, instantáneo)

k	$\widehat{ATT}_K^{\text{nivel}}(k)$	SE	IC95 %	h	$\widehat{ATT}_H^{\text{nivel}}(h)$	SE	IC95 %
3	0.2355	0.1254	[-0.0237, 0.4798]	1	0.4803	0.0811	[0.2849, 0.6131]
7	1.6059	0.1523	[1.3150, 1.9310]	2	1.5254	0.0744	[1.3723, 1.6651]
8	1.2287	0.2424	[0.8055, 1.6758]	3	0.5189	0.0664	[0.3759, 0.6285]

Promedio global ponderado: $\widehat{ATT}_{\text{global}}^{\text{nivel}} = 0.8767$, $SE_{\text{boot}} = 0.0490$, IC95 % = [0.7516, 0.9478].

Comparación total vs. instantáneo

Definimos $\Delta^{\text{post}} = \widehat{ATT}^{\text{rup}} - \widehat{ATT}^{\text{nivel}}$. Diferencias pequeñas indican que el efecto se explica sobre todo por el salto en k ; diferencias grandes apuntan a una contribución importante de la dinámica posterior. Si difieren en signo, hay reversión.

Cuadro B.24: Ejemplos de comparación en (k, h) y promedios globales

(k, h)	$\widehat{ATT}^{\text{rup}}$	$\widehat{ATT}^{\text{nivel}}$	Δ^{post}
(5,2)	2.4433	1.7018	0.7414
(6,1)	-1.1367	-0.0845	-1.0522
Global	0.7241	0.8767	-0.1526

B.5. Implementación del diseño por cortes temporales (con ruptura de tendencia)

En esta sección implemento el diseño *por cortes temporales* para identificar el efecto marginal de pasar de $h-1$ a h en el instante k , bajo el supuesto *con ruptura de tendencia*. Trabajo sobre el *mismo* panel simulado utilizado en los anexos previos, con columnas i , t , y , H y D , y dimensiones $I = 60$ y $T = 20$. El programa fijó semilla global en 42, cargó el panel (`shape = 1200 × 5`), confirmó balance (sin celdas faltantes) y reportó $\sum S = 177$ (saltos acumulados +1).

Conjuntos y ventanas

Para cada par candidato (k, h) , se construyen ventanas válidas $[t^-, t^+]$ que respetan las reglas del diseño con ruptura en k : las *tratadas efectivas* son unidades que saltan de $h-1$ a h en k y están observadas en t^- y t^+ ; los *controles efectivos* son unidades que permanecen **exactamente en $h-1$** dentro de la ventana $[t^-, t^+]$ (sin salto en k), conforme a la Sección 3.2. El total de ventanas válidas fue **1 420**.

Cuadro B.25: Ejemplos de ventanas válidas y tamaños efectivos por (k, h, t^-, t^+)

k	h	t^-	t^+	n_T	n_C
3	1	1	3	12	48
3	1	1	4	12	41
3	1	1	5	10	34
3	1	1	6	9	31
3	1	1	7	3	29
3	1	1	8	2	25
3	1	1	9	2	20
3	1	1	10	1	18
3	1	1	11	1	18
3	1	1	12	1	16
3	1	1	13	1	14
3	1	2	3	12	48
3	1	2	4	12	41
3	1	2	5	10	34
3	1	2	6	9	31

Estimación elemental por ventana

De acuerdo con la Proposición 3.2 (bajo el Supuesto 3.2), para toda ventana válida (k, h, t^-, t^+) el estimando $\widehat{ATT}_{\text{rup}}^{t^-, t^+}(k, h)$ se identifica como en (3.2.6) (y de forma equivalente en (3.2.7)). En términos de datos observados:

$$\widehat{ATT}_{\text{rup}}^{t^-, t^+}(k, h) = \mathbb{E}[y_{i, t^+} - y_{i, t^-} \mid i \in \mathcal{I}^{k, h}(t^-, t^+)] - \mathbb{E}[y_{j, t^+} - y_{j, t^-} \mid j \in \mathcal{J}_{\text{rup}}^{k, h}(t^-, t^+)].$$

A continuación se muestran algunos ejemplos representativos con tamaños efectivos n_T, n_C :

Cuadro B.26: Ejemplos de $\widehat{ATT}_{\text{rup}}^{t^-, t^+}(k, h)$ por ventana

Ventana				Tamaños efectivos		Estimador
k	h	t^-	t^+	n_T	n_C	$\widehat{ATT}_{\text{rup}}^{t^-, t^+}(k, h)$
3	1	1	3	12	48	0.237427
3	1	1	4	12	41	-0.262616
3	1	1	5	10	34	-0.825082
3	1	1	6	9	31	-1.283990
3	1	1	7	3	29	-1.561499
3	1	1	8	2	25	-1.785293
3	1	1	9	2	20	-2.295583
3	1	1	10	1	18	-2.468712
4	1	1	4	7	41	0.436230
4	1	1	5	5	34	-0.020282
4	1	1	6	4	31	-0.425832
4	1	1	7	3	29	-0.933132
4	1	2	4	7	41	0.433786
4	1	2	5	5	34	-0.026804
4	1	2	6	4	31	-0.431758
4	1	2	7	3	29	-0.933072

Diagnóstico PRE (placebos)

Para cada par (k, h) y cada anclaje previo $t^{--} < t^-$, se calculó el placebo $\widehat{ATT}_{t^{--}, t^-}^{\text{pre}}(k, h)$ reutilizando los conjuntos efectivos definidos para la ventana principal. Se obtuvieron **874** placebos PRE (sin duplicados).

Cuadro B.27: Ejemplos de placebos PRE por ventana (k, h, t^{--}, t^-)

Ventana PRE				Tamaños efectivos		Estimador placebo
k	h	t^{--}	t^-	n_T^{pre}	n_C^{pre}	$\widehat{ATT}_{t^{--}, t^-}^{\text{pre}}(k, h)$
3	1	1	2	12	48	0.003809
4	1	1	2	7	41	0.002444
4	1	1	3	7	41	0.004849
4	1	2	3	7	41	0.002406
5	1	1	2	7	34	0.014085
8	2	6	7	3	11	-0.025759
8	3	6	7	1	4	-0.008670
9	1	1	2	5	20	0.009990
9	1	1	3	5	20	0.011288
9	1	2	3	5	20	0.001298

Para evaluar magnitud relativa, se consideró $R = |\widehat{ATT}_{t^-, t^-}^{\text{pre}}(k, h)| / \mathbb{E}[|\Delta y_C|]$ (cambio típico de controles en la misma ventana) y se clasificó: VERDE ($\leq 10\%$), AMARILLO (10–25%), ROJO ($> 25\%$).

Cuadro B.28: Magnitud relativa y semáforo de placebos PRE

Ventana PRE				R (%)	Semáforo
k	h	t^-	t^-		
3	1	1	2	1.91	VERDE ($\leq 10\%$)
4	1	1	2	1.23	VERDE ($\leq 10\%$)
4	1	1	3	1.21	VERDE ($\leq 10\%$)
4	1	2	3	1.20	VERDE ($\leq 10\%$)
5	1	1	2	7.17	VERDE ($\leq 10\%$)
5	1	1	3	3.62	VERDE ($\leq 10\%$)
5	1	2	3	0.16	VERDE ($\leq 10\%$)
5	1	1	4	0.13	VERDE ($\leq 10\%$)
5	1	2	4	3.69	VERDE ($\leq 10\%$)
5	1	3	4	7.53	VERDE ($\leq 10\%$)
6	1	2	3	11.42	AMARILLO (10–25%)

Cuadro B.29: Distribución del semáforo (PRE)

Categoría	Porcentaje
VERDE ($\leq 10\%$)	83.5%
AMARILLO (10–25%)	14.6%
ROJO ($> 25\%$)	1.8%

Agregación por (k, h)

Aplicando la regla de ponderación de (3.2.8) y su versión normalizada dentro de cada (k, h) , agrego las estimaciones por ventana $\widehat{ATT}_{\text{rup}}^{t^-, t^+}(k, h)$ para obtener $\widehat{ATT}_{\text{rup}}(k, h)$. La tabla reporta, además, el número de ventanas agregadas y la suma $\sum \omega$ previa a la normalización (solo informativo).

Cuadro B.30: Efecto agregado por par (k, h) (con ruptura)

Par		Acumulados		Estimación
k	h	# ventanas	$\sum \omega$	$\widehat{ATT}_{\text{rup}}(k, h)$
3	1	22	87.486644	-0.902761
4	1	33	77.366475	-0.889165
5	1	20	78.373457	-0.665403
5	2	14	17.953728	3.051560
6	1	40	72.863378	-1.399311
6	2	8	10.186391	1.657134
6	3	2	1.500000	0.696171
7	1	84	114.781906	-1.130228
7	2	23	47.974408	2.839630
7	3	2	1.633333	0.976998
8	1	63	116.761194	0.085437
8	2	19	26.144543	2.692391
8	3	2	2.492308	0.234005
8	4	1	0.500000	0.526302
9	1	96	232.434943	-1.282343

Agregados por instante k (inferencia)

De acuerdo con (3.2.9), los promedios por instante k se obtienen con los pesos $p(h | k)$. Se reportan el punto, SE_{boot} , intervalo al 95 %, p_{boot} y código de significancia (bootstrap por unidades).

Cuadro B.31: Inferencia por instante k (con ruptura)

k	$\widehat{ATT}_{K,rupt}(k)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	p_{boot}	sig
3	-0.902761	0.122365	-1.006180	-0.535394	0.000000	***
4	-0.889165	0.248656	-0.982363	-0.018508	0.036072	**
5	0.686220	0.263281	0.113457	1.073718	0.036072	**
6	0.209948	0.418760	-0.863511	1.372141	0.350101	
7	1.948508	0.304174	1.528092	2.784463	0.000000	***
8	0.941323	0.301190	0.261557	1.619498	0.000000	***
9	-0.126252	0.230476	-0.333502	0.354157	0.681363	
10	1.181488	0.436221	0.210262	1.823722	0.024048	**
11	0.231810	0.306272	-0.375718	0.856761	0.432866	
12	0.556348	0.445342	-0.357263	1.279464	0.200401	
13	1.250582	0.312780	0.449252	1.589705	0.012024	**
14	2.491059	0.337920	1.698276	3.032584	0.000000	***
15	1.374593	0.420374	0.978997	2.431135	0.008016	***
16	1.707728	0.335222	0.871643	2.141617	0.000000	***
17	1.256912	0.449663	0.201874	2.264239	0.032129	**

Agregados por nivel h (inferencia)

Según (3.2.10), los promedios por nivel h usan $p(k | h)$. Incluyo punto, SE_{boot} , IC95 %, p_{boot} y significancia.

Cuadro B.32: Inferencia por nivel h (con ruptura)

h	$\widehat{ATT}_{H,rupt}(h)$	SE_{boot}	IC95 % (lo)	IC95 % (hi)	p_{boot}	sig
1	-0.714385	0.086805	-0.760509	-0.406389	0.000000	***
2	2.649489	0.152390	2.160188	2.741677	0.000000	***
3	-0.116961	0.132112	-0.241649	0.263201	0.733467	
4	1.580645	0.164279	1.367685	1.985410	0.000000	***
5	-0.204491	0.227381	-0.496251	0.372400	0.601202	
6	2.745644	0.514096	1.657555	3.645682	0.000000	***

Estimación global

Con (3.2.11), el promedio ponderado conjunto entrega:

$$\widehat{ATT}_{global,rupt} = 0.756715, \quad SE_{boot} = 0.069937, \quad IC95 \% = [0.605288, 0.868130] ***.$$

Chequeo de invarianza del orden: $\sum_k p(k) \widehat{ATT}_{K,rupt}(k) = 0.756715$ y $\sum_h p(h) \widehat{ATT}_{H,rupt}(h) = 0.756715$.