When Does Metro Infrastructure Capitalize into Property Prices? Phase-Decomposed Difference-in-Differences Evidence from Seven European Cities

Ian Ronk

Independent researcher, Amsterdam
ianronk0@gmail.com

Abstract

We study *when*, not merely whether, new metro lines capitalize into residential property prices. Seventeen staggered treated cohorts across seven European cities in five countries (Milano, Amsterdam, Copenhagen, Paris, Helsinki, Rennes, Roma) pool into a single phase-decomposed panel ($n = 42{,}004$), with the response decomposed into announcement, construction, opening, and maturity phases. The pooled cross-city average locates the largest response at maturity: prices step up by +9 to +12% — drop-Roma to full-sample, measured as the construction-to-maturity contrast — two or more years after opening, a step that is stable across the control ladder and positive under every leave-one-city-out. The step is a pooled average, however, not a within-city fact. City-by-year fixed effects collapse it to an insignificant −0.5 log points while leaving a +2.5 log-point step at opening — a within-city contrast that is itself entity-clustered ordering evidence, not significant under city-clustered bootstrap. Few-cluster significance of the maturity step is partition- and control-set-dependent: a restricted wild cluster bootstrap gives $p = 0.036$ clustered on the seven cities (level-only controls; $p = 0.080$ with the full control set) but $p = 0.16$ on the twenty-four cohorts. Roma, priced through the same OMI appraisal series as Milano, is itself a within-city null whose pooled contribution runs entirely through the common-year fixed effects, and it disciplines the upper-bound reading throughout. The defensible magnitudes are per-city — foremost Milano's within-ring +167 EUR/m^2 ($\approx$ +5.6%, wild-bootstrap $p = 0.004$). We read the delayed-to-maturity step as a cross-city pattern worth testing on longer panels, not as a settled within-city effect.

Keywords: transit capitalization, property prices, difference-in-differences, capitalization timing, staggered adoption, few-cluster inference, wild cluster bootstrap.

JEL codes: R31 (housing supply and markets), R41 (transportation demand, supply, congestion), C23 (panel-data models).

1 Introduction

New urban rail lines are among the largest infrastructure investments European cities make, and the way they capitalize into nearby property values shapes the cost–benefit calculus, the value-capture finance case, and the spatial-equity debate that surrounds them. The theoretical underpinning is the hedonic-pricing framework of Rosen (1974), in which differentiated location attributes are bundled into a single asset price and the equilibrium price gradient identifies the marginal willingness to pay for accessibility. A long empirical literature has applied that framework to transit; the meta-analytic range is wide rather than settled — Mohammad et al. (2013) place the mean uplift for properties near new heavy-rail stations in the 5–15% band, while Debrezion et al. (2007) report a headline nearer +2–3% and Higgins and Kanaroglou (2016) stress non-convergence across contexts — with substantial heterogeneity in distance decay throughout. We add to it in two steps. First, we re-examine two cases previously covered only by ex-ante hedonic valuation or non-peer-reviewed estimates (Milano, Amsterdam); second, we pool seven European cities across five countries to ask a question the single-snapshot literature cannot answer — not *whether* metro access capitalizes, but *when* along the project timeline it does. Beyond Milano and Amsterdam we add Copenhagen's M3 Cityringen, the first two segments of the Grand Paris Express Ligne 14, Helsinki's Länsimetro, Rennes's automated Line B, and Roma's Metro C. The central finding is a *delayed-to-maturity* pattern in the pooled cross-city average: the largest price response — a construction-to-maturity step of +9 to +12%, depending on whether the null seventh city is counted — appears two or more years after opening, not at announcement or during construction. The pattern is a pooled average rather than a within-city fact — under city-by-year fixed effects the within-city evidence locates a smaller step at opening instead — and we spend much of the paper establishing which parts of it survive which stress test.

Peer-reviewed Italian evidence on transit-capitalization is heavily concentrated on Naples. Pagliara and Papa (2011) apply a cross-sectional hedonic regression to Naples metro stations and find that only high-frequency lines have appreciable property-price effects; Gallo (2018) revisits the same Naples panel with a linear hedonic specification (LRM), reaches the same high-frequency-vs-low-frequency conclusion, and quantifies the implied external benefit at $\approx$ 8.5% of total real-estate value in the catchment. The most recent national-scale work is Guglielmi et al. (2026), who estimate a Hedonic Price Model with Confirmatory Factor Analysis on a cross-section of 985 Italian railway stations and find that railway centrality, long-distance service provision, and multimodal integration are positively associated with housing prices. The gap between this prior Italian literature and our paper is identification: Gallo (2018); Guglielmi et al. (2026); Pagliara and Papa (2011) are all single-snapshot hedonic regressions, so they cannot separate anticipation, opening, and maturity effects, nor distinguish capitalization from contemporaneous neighbourhood selection. For Milan specifically, the Politecnico di Milano DAStU group — including Boscacci et al. (2016) on the Navigli urban-regeneration hedonic — has documented Milan-specific willingness-to-pay through *ex ante* hedonic pricing combined with an input–output multiplier, and the Italian press has reported real-estate-agency estimates of M4-corridor capitalization in the +3.8% to +36% range. Peer-reviewed *ex-post* difference-in-differences evidence on the M5 Lilla and M4 Blu openings is, as far as we have been able to verify, absent. Our Milano specifications below are intended to fill that gap. For Roma the closest prior work is Daniele and Romito (2022), a Banca d'Italia synthetic-control study of Metro C finding a statistically significant negative suburban price effect (~ -137 EUR/m^2, upper-distribution-driven); it is a working paper rather than a journal publication, and our OMI-instrument within-city null in §8 corroborates its central message.

Milano gives us staggered treatment. Metro Line M5 (Lilla) opened in three stages (2013–2015) and M4 (Blue) in three more (2022–2024), which lets us exploit cross-zone variation in opening dates. Amsterdam gives us a clean single-date treatment: all eight Noord-Zuid Lijn stations opened on 22 July 2018 after a 16-year construction period. Neither city has peer-reviewed ex-post DiD evidence; the bare-TWFE estimates that circulate for both come from earlier, non-peer-reviewed analyses, including earlier drafts of this project. We estimate both with cluster-robust inference (Bertrand et al., 2004), the doubly-robust staggered-DiD estimator of Callaway and Sant'Anna (2021) (robust to the heterogeneous-treatment biases now well documented in de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021), and event-study designs that take pre-trend sensitivity seriously (Borusyak et al., 2024; Rambachan and Roth, 2023; Sun and Abraham, 2021). We borrow the canonical hedonic framing of Bowes and Ihlanfeldt (2001); Gibbons and Machin (2005); McMillen and McDonald (2004) and the Italian-context precedent of Pagliara and Papa (2011) for the Milano specification.

The empirical strategy mixes two-way fixed-effects DiD at the zone (Milano) and buurt (Amsterdam) level with distance-based continuous treatments, event studies, rolling-window estimates and placebo tests. We pay particular attention to one class of confound: urban-regeneration projects that overlap geographically and temporally with new metro construction in Milano, flagged but rarely controlled in the prior literature.

Contributions. Four contributions distinguish this paper from the closest prior work.

First, peer-reviewed ex-post DiD evidence on Milan's M5 and M4. The Italian transit-capitalization literature is cross-sectional in its identification strategy and concentrated on Naples (Gallo, 2018; Pagliara and Papa, 2011) or on a national cross-section of stations (Guglielmi et al., 2026). Milan-specific peer-reviewed work has been limited to ex-ante hedonic valuation of an unrelated urban-regeneration project (Boscacci et al., 2016) and to non-academic real-estate-agency estimates in the Italian press. To our knowledge, no published peer-reviewed paper has applied a within-unit difference-in-differences identification strategy to the M5 Lilla and M4 Blu openings specifically. We do so here, with a within-ring identifying contrast that controls for the Porta Nuova / CityLife / Scalo Farini regeneration confound that the Naples and national-cross-section literature do not face.

Second — and foremost — a seven-city pooled phase decomposition that identifies ***when*** *capitalization happens.* Most of the metro-capitalization literature reports a single binary post-opening coefficient (Billings, 2011; Bowes and Ihlanfeldt, 2001; Diao et al., 2017; Gibbons and Machin, 2005), which cannot distinguish anticipation from opening from long-run adjustment. Timing-decomposed designs exist, but as single-city studies: Devaux et al. (2017) separate anticipation from post-construction capitalization for Montreal's Laval metro extension, Yiu and Wong (2005) document expectation effects during the construction phase in Hong Kong, and Grimes and Young (2013) find announcement-date capitalization for Auckland's Western Line. Our claim is therefore the pooled multi-city application, not the decomposition itself. We pool seven European cities across five countries — Milano (M5 + M4), Amsterdam (NZL), Copenhagen (M3 + M4 Sydhavn), Paris (GPE Ligne 14 Nord + Sud), Helsinki (Länsimetro 2017 + 2022), Rennes (Line B 2022), Roma (Metro C 2014 / 2015 / 2018), seventeen treated cohorts in all — into a single entity-and-year-fixed-effects PanelOLS (n = 42,004) and decompose the response into four temporally distinct phases (Announcement, Construction, Delivery, Post-Delivery). The result is a *delayed-to-maturity* pattern in the pooled cross-city average: the response is largest at maturity, where it steps up by +9 to +12% over the construction-phase level ($\gamma_{\text{mature}} - \gamma_{\text{construct}}$; the cumulative differential over the pre-treatment baseline is larger, +0.283 log points, and inherits the common-year-FE upper-bound caveat of §7.3). We are careful, in §8, to bound what that average does and does not establish. The step is *control-stable* (Oster-style, ranging +12.3% to +14.4% across the confound ladder, always $z \geq 6.6$) and positive under every leave-one-city-out, but it is identified through the common-year fixed effects: re-estimating the phase decomposition under city-by-year FE — the same conservative specification that collapses the pooled level — shrinks the step to an insignificant −0.5 log points while leaving a +2.5 log-point step at opening ($z = 3.0$ entity-clustered; Webb $p = 0.52$ under the city-clustered bootstrap, so ordering evidence only), and only Milano shows the maturity-growth pattern within-city. Pretrend robustness is likewise limited (the unconditional Honest-DiD breakdown is $\bar{M} \approx 0.7$ under flat slack but ≈ 0.1 under the accumulating Δ^{RM} bound; §7.4), and few-cluster inference is partition- and control-set-dependent: a restricted wild cluster bootstrap on the step gives $p = 0.036$ clustered on the seven cities ($G = 7$, level-only controls; the repaired full-control branch gives $p = 0.080$) but $p = 0.16$ on the twenty-four cohorts ($G = 24$). We therefore present the finding as a pooled cross-city *pattern* — significant only at the coarsest cluster level, and not a within-city object — rather than as a settled $p < 0.05$ magnitude.

Third, the modern staggered-DiD toolkit applied to a pooled seven-city European cohort. Both Naples studies pre-date the staggered-DiD revolution (Borusyak et al., 2024; Callaway and Sant'Anna, 2021; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Rambachan and Roth, 2023; Sun and Abraham, 2021) and rely on cross-sectional hedonic regressions. We apply the doubly-robust Callaway–Sant'Anna estimator with a 999-replicate wild cluster bootstrap (Cameron et al., 2008; Sant'Anna and Zhao, 2020), the Honest-DiD $\Delta^{\mathrm{RM}}(\bar{M})$ pre-trend sensitivity check, joint Wald tests on the phase coefficients, a few-cluster restricted wild cluster bootstrap (Webb six-point weights) and a leave-one-city-out on the pooled maturity step, and a measurement-class split that tests whether cross-instrument pooling drives the headline. Phase-decomposed transit-capitalization DiD is not itself new — Gupta et al. (2022), for instance, separate construction-phase from operation-phase capitalization for New York's Second Avenue Subway — so we claim novelty not for the estimator stack but for its application to a pooled seven-city, five-country European cohort (Milano M5/M4, Amsterdam NZL, Copenhagen M3/M4, Paris GPE Ligne 14, Helsinki Länsimetro, Rennes Line B, Roma Metro C) and, foremost, for the first ex-post within-unit treatment of the Milan M5/M4 openings (Contribution 1).

Fourth, identifying-data recovery. Two data-engineering moves materially expanded the pre-treatment identifying variation. The load-bearing one is the recovery of the 2014–2020 Paris geo-DVF vintage from the OpenDataArchives `etalab-dvf` mirror after Etalab pruned the historical files from its `latest/` branch: without it, Paris would have had no pre-treatment years for GPE Ligne 14 Nord (opened 2020), leaving the phase decomposition unidentified for Paris. The second extends the Copenhagen condominium price panel back to January 1992 from the public `boliga.dk` sales API (§8.2 documents the acquisition; the Acknowledgments state its terms-of-use position).

Scope. The seven-city pooled design with phase decomposition is built to identify *when* capitalization happens — at announcement, during construction, at opening, or in the maturity phase — rather than to refine the fine spatial structure of the within-station-catchment decay. The within-cell distance decay we do report (Milano per-band gradient in §4.3, Amsterdam per-band gradient in §5, continuous inverse-distance specification in eq. (3)) runs on the aggregated zone- or buurt-level panel, not on parcel-level transaction microdata. The natural follow-up paper, which we leave to a separate manuscript, restricts attention to the two cities with the longest credibly-identifying panels — **Milan** (M5 Lilla, with ≥10 post-opening years per cohort and 22 years of OMI history) and **Paris** (GPE Ligne 14 Nord, with 6 pre-treatment and 5 post-treatment geo-DVF years currently, extending as future vintages are released) — and answers a different question: not *when* but *where-in-space and year-by-year.* Concretely it would (i) use the Agenzia delle Entrate *Banca dati delle Compravendite immobiliari* (Italy) and the parcel-level geo-DVF deeds (France) at their native resolution rather than aggregated to zone or IRIS, (ii) estimate continuous distance-from-station decay curves at parcel level rather than 500 m-band TWFEs, and (iii) replace the phase decomposition with a year-by-year event-study specification on the longer post-opening horizon now available for M5 Lilla. The seven-city pooled decomposition is the headline result, but it is developed only in §8. Sections 4–7 first establish Milano and Amsterdam, the two cities studied in greatest depth, as identified single-city groundwork — and the inference machinery the pooled section reuses — before the pooled timing answer is assembled on top of them.

Table 1: Milano metro opening stages.

Line	Stage	Date	Stops
M5 Lilla	1	Feb 2013	Bignami–Zara (7)
M5 Lilla	2	Mar 2014	Zara–Garibaldi FS (2)
M5 Lilla	3	Apr 2015	Garibaldi–S. Siro (10)
M4 Blue	1	Nov 2022	Linate–Dateo (6)
M4 Blue	2	Jul 2023	Dateo–S. Babila (2)
M4 Blue	3	Oct 2024	S. Babila–S. Cristoforo (13)

2 Data

2.1 Milano

Property prices. The price series is the Osservatorio del Mercato Immobiliare (OMI) dataset published by the Italian Revenue Agency (*Agenzia delle Entrate*), which reports zone-level average residential transaction prices (EUR/m^2) for the municipality of Milano. From the 43 OMI macro-zones on 2025 boundaries we build a balanced panel spanning 2004–2025 (22 years), harmonizing the 2011-era zone definitions used in the earlier years to the current 2025 system through an area-weighted spatial crosswalk. Coverage is the binding constraint: of the 43 zones, 38 clear the balanced-panel requirement of at least 15 years of non-missing OMI coverage and form the estimation sample used throughout. The 5 dropped zones (4 control, 1 M4-treated) are thin in the early period — sparse coverage, not a substantive exclusion.

Metro data. Station locations come from GTFS data for ATM Milano. We assign per-stop opening years from the actual staged rollout schedule (Table 1), not a single line-opening date.

Treatment assignment. A zone is treated if it lies within 500 m of any M4 or M5 stop, measured as polygon distance — zero if the stop falls inside the zone, otherwise the distance to the nearest zone edge. Of the 43 OMI macro-zones, 25 are treated (17 M4, 8 M5) and 18 are controls; in the 38-zone estimation sample (after the ≥15-year coverage filter above), 24 are treated (16 M4, 8 M5) and 14 serve as controls. Treatment timing is staggered: it varies by zone with the opening year of the nearest stop.

Table 2: Amsterdam Noord-Zuid Lijn stations.

Station	Latitude	Longitude
Noord	52.3919	4.9364
Noorderpark	52.3823	4.9218
Centraal Station	52.3791	4.9003
Rokin	52.3702	4.8938
Vijzelgracht	52.3621	4.8899
De Pijp	52.3546	4.8935
Europaplein	52.3389	4.8896
Zuid	52.3390	4.8735

2.2 Amsterdam

Property values. The Amsterdam series is CBS *Kerncijfers Wijken en Buurten* (Key Figures for Districts and Neighborhoods), which reports the average WOZ property value (*Gemiddelde Woningwaarde/WOZ-waarde*) at the buurt level. These are municipality-assessed values used for taxation, in thousands of euros — an assessed value, not a transaction price. We pull all Amsterdam buurten (gemeente code 0363) across 10 annual tables (2013–2022), which yields 2,779 buurt-year observations across 409 buurten.

Buurt geometries. Buurt boundary polygons are downloaded from the PDOK CBS Wijken en Buurten 2022 OGC API: 481 Amsterdam buurten with MultiPolygon geometries in WGS84. From these we take each buurt's centroid and compute haversine distances to the metro stations.

Metro stations. The Noord-Zuid Lijn is a clean event. Its 8 stations from Noord to Zuid all opened on the same day, 22 July 2018 (Table 2). We also geocode the 27 existing metro stations (lines 50, 51, 53, 54) for comparison.

3 Empirical Strategy

3.1 Panel Fixed-Effects DiD

We swap one source of variation for another. Where the Italian Naples literature leans on cross-sectional hedonic specifications (Gallo, 2018; Pagliara and Papa, 2011), our identification strategy substitutes *within-unit time variation* for cross-unit distance variation: each zone, buurt, postnummer, or IRIS is its own control under the panel design, and the effect is identified off the timing of the metro opening, not the cross-section of distances. Two things follow. First, any time-invariant unobserved confounder — centrality, baseline amenity stock, historical sorting — drops out of the estimating equation. Second, the temporal decomposition into rumour, announcement, construction, opening, and post-delivery phases (§8) becomes possible at all. The cost is the binding constraint: we need a panel long enough to credibly identify pre-treatment trends, and that is exactly what limits the Copenhagen and Paris extensions.

Our baseline specification is a two-way fixed effects DiD given by equation (1):

$$\text{Price}_{it} = \beta \cdot \text{TreatedPost}_{it} + \mathbf{X}_{it}'\boldsymbol{\delta} + \alpha_i + \gamma_t + \varepsilon_{it} \quad (1)$$

where α_i are zone/buurt fixed effects, γ_t are year fixed effects, and $\text{TreatedPost}_{it} = 1$ if zone i is treated and year t is at or after the treatment year. $\mathbf{X}_{it}$ collects time-varying zone-level controls (regeneration overlap for Milano, anticipation indicator for Amsterdam; §4–§5). Timing varies by zone in Milano; in Amsterdam all treated buurten share the 2018 date. We cluster standard errors at the spatial unit throughout — zone for Milano, buurt for Amsterdam. The reason is not optional: naive non-clustered SEs in panel data with serial correlation are anticonservative and would overstate significance. We estimate β by OLS with full zone and year dummies and cluster-robust variance (CR1).

3.2 Distance-Band DiD

To trace the distance decay of treatment effects, we run separate DiD regressions for each distance band d:

$$\text{Price}_{it} = \beta_d \cdot \mathbf{1}[i \in \text{Band}_d] \cdot \text{Post}_t + \alpha_i + \gamma_t + \varepsilon_{it} \quad (2)$$

where the control group is always zones beyond the maximum treatment distance (2km+ for Milano, 3km+ for Amsterdam).

3.3 Continuous Distance Specification

A binary treatment buffer is an arbitrary cutoff. Pick "within 500 m of a new station" and you have imposed a threshold and thrown away within-zone variation in proximity. The problem bites hardest in Milano, where OMI zones are large — 43 macro-zones cover the whole comune, and a single zone runs 1–3 km across — so a binary spec dilutes a localised price response into the surrounding area. We therefore also estimate a continuous treatment intensity model:

$$\text{Price}_{it} = \lambda \cdot \frac{1}{d_i + 0.5} \cdot \text{Post}_t + \mathbf{X}_{it}'\boldsymbol{\delta} + \alpha_i + \gamma_t + \varepsilon_{it} \quad (3)$$

where d_i is the distance in km from zone i's centroid to the nearest new station and Post_t switches on at the relevant opening year. The inverse-distance kernel gives a smooth treatment intensity: it approaches 2λ for a zone whose centroid sits on the station and decays to zero at large distances. Foremost, it removes the buffer-choice researcher degree of freedom. A positive λ says zones closer to new stations see larger post-opening price gains. We report it alongside the binary spec.

3.4 Event Study

We estimate year-by-year treatment effects relative to the event date:

$$\text{Price}_{it}^{dm} = \sum_k \beta_k \cdot \mathbf{1}[\text{RelYear}_{it} = k] + \varepsilon_{it} \quad (4)$$

where Price^{dm} is the demeaned price and k indexes years relative to the event; coefficients are normalised to $k = -1$. For the staggered Milano panel we additionally use the interaction-weighted event-study estimator of Sun and Abraham (2021). The bare-TWFE event-study is not safe here: when treatment effects vary across cohorts it can introduce contamination bias (Borusyak et al., 2024), and the Sun–Abraham estimator avoids it.

3.5 Placebo Tests and Rolling Windows

We run placebo tests using fake treatment dates in the pre-treatment period to check for pre-existing differential trends. We also compute 5-year rolling window

Table 3: Milano combined-panel DiD: naive vs. cluster-robust SEs. Stars here and throughout: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, † $p < 0.1$.

Specification	$\hat{\beta}$ (EUR/m^2)	SE	p	Sig.
TWFE, naive SE	+820	84.8	<0.001	***
TWFE, cluster-robust SE	+813	448.6	0.070	†
TWFE + regen control	+831	455.9	0.068	†
TWFE + regen + invdist post	+1,185	595.5	0.047	*

DiD coefficients to visualize when the treatment effect emerges and whether it persists.

4 Results: Milano

Two specifications survive proper inference, and we build the Milano results around them: the within-ring M5 DiD (§4.2) and the distance-band gradient (§4.3). The combined-panel TWFE (§4.1) we report for transparency, but it does *not* survive the corrections below. We treat it as a naive benchmark, not a headline number.

4.1 Combined M4+M5 Analysis (Naive Benchmark)

Table 3 reports the combined-panel two-way FE estimate twice: once with the naive (non-clustered) standard error, and once with a zone-clustered (CR1) standard error robust to within-zone serial correlation.

Clustering at the zone level moves two things. The point estimate is the easy part: it barely budges (+813 vs. +820, a 0.9% difference). The standard error is the problem — it rises by a factor of ∼5, and the headline "$p < 0.001$" collapses to $p = 0.07$, marginal at best. The deeper issue is that the combined specification commingles 6 staggered treatment cohorts (M5 stages 2013/2014/2015 and M4 stages 2022/2023/2024). That makes it subject to the implicit-control-group bias documented by Goodman-Bacon (2021) and de Chaisemartin and D'Haultfœuille (2020): when effects vary across cohorts, already-treated zones enter the comparison set for later-treated zones with negative weights.

The bias is operative here, and we can sign it. A Callaway–Sant'Anna (Callaway and Sant'Anna, 2021) estimator with never-treated controls and a 999-replicate cluster bootstrap (§7.1) returns an overall ATT of +751 EUR/m^2, 95% bootstrap CI $[-332, +1834]$. The cohort decomposition is where the contamination shows: the largest cohort (M4 stage 3, 11 zones treated 2024) returns $\hat{\beta} = -30$ EUR/m^2 (n.s.), while the lone 2014 cohort ($n = 1$) carries an implausible +2,702. So the +820 is a Goodman–Bacon weighted blend of clean and contaminated 2×2 effects, and the survey of staggered-DiD pathologies in de Chaisemartin and D'Haultfœuille (2023); Roth et al. (2023) is exactly why we refuse to treat it as a headline number.

4.2 Within-Ring M5 Analysis

The cleanest M5 identification comes from restricting the panel to OMI Ring D (suburban zones, code prefix `D`). Ring D holds 16 zones — 8 M5-treated, 8 controls — all at comparable distances from the centre, so the implicit center-vs-periphery confound that drives the combined-panel estimate is absorbed by the ring restriction itself. Table 4 reports TWFE results with zone-clustered SEs, with and without the regeneration-overlap covariate.

Table 4: Milano Ring D M5 within-ring DiD (cluster-robust SEs). $n = 352$; 16 zones (8 treated, 8 control), 22 years.

Specification	$\hat{\beta}$	SE	t	p	Sig.
(a) Bare TWFE	+139.5	54.8	+2.55	0.011	*
(b) + regen overlap	+166.7	52.3	+3.19	0.002	**
(c) + regen + invdist · post	+182.3	86.4	+2.11	0.035	*

Table 5: Milano distance-band gradient (TWFE, zone-clustered SEs; reference = 2000 m+).

Distance band	$\hat{\beta}$ (EUR/m^2)	SE	p	Sig.
0–500 m	**+769**	228	**0.0008**	***
500–1,000 m	+283	165	0.086	†
1,000–1,500 m	+59	102	0.565	n.s.
1,500–2,000 m	+105	57	0.066	†
2,000 m+ (ref.)	—	—	—	—

The bare Ring D estimate is $\hat{\beta} = +139.5$ EUR/m^2 ($p = 0.011$). That lands close to the +136 in the project's earlier baseline[1], but now with a much wider, properly clustered standard error. The regeneration overlap dummy — defined as 1 if the zone centroid falls within 800 m of an active major Milan regeneration project (Porta Nuova, CityLife, Scalo Farini, Symbiosis, MIND, Cascina Merlata, Scalo Romana) in year t — raises the M5 effect to +167 ($p = 0.002$). The covariate itself enters at +129 ($p = 0.002$): within Ring D, regeneration projects are themselves significant drivers of zone prices, and controlling for them sharpens the M5 estimate rather than attenuating it. Adding the continuous-distance interaction (spec c) leaves $\hat{\beta}$ at +182.

The Ring D estimate corresponds to ≈ +5.6% of the average control price. The effect is a level shift after opening, not a cumulative divergence: treated and control trajectories track each other at the new higher level.

4.3 Distance-Band Gradient

The within-ring spec answers *whether* the M5 effect exists at all. The distance-band gradient answers *where* and *how steeply.* We sort all 38 Milano zones by centroid-to-nearest-new-stop distance and run a single TWFE with four band indicators, the 2000 m+ band omitted as reference. Results in Table 5.

The gradient is clean and monotone over the first three bands: +769 within 500 m, falling to +283 at 500–1000 m, then to a null +59 at 1000–1500 m. The continuous specification of equation (3) returns $\hat{\lambda} = +392$ EUR/m^2 per unit of $1/(d + 0.5)$ ($p = 0.10$), with the regeneration control at +135 (n.s.). Read off at point distances, that is +785 at $d = 0$ km, +392 at $d = 0.5$ km, +262 at $d = 1$ km, and +157 at $d = 2$ km. The 0 km extrapolation agrees with the 0–500 m band coefficient (+769) to within ∼2% — reassuring cross-

[1] Re-estimation reflects a slightly different balanced-panel filter; the substantive estimate is unchanged.

spec consistency. Both the magnitude and the shape sit comfortably inside the distance-decay range reported in prior single-line studies (Billings, 2011; Bowes and Ihlanfeldt, 2001; Diao et al., 2017; McMillen and McDonald, 2004).

The narrow 1500–2000 m positive coefficient is a small-sample artefact: only one zone falls in that band. Its confidence interval collapses to a thin band because the within-zone time-series variation is exhausted by the zone fixed effect, leaving only year-by-year noise. The gradient story rests on the first three bands.

4.4 Event Study and Pre-Trends

Both the Ring D and the 0–500 m band event studies show pre-trends. In the 0–500 m band the pre-period coefficient at $k = -4$ is -311 ($p = 0.012$): prices in close-to-station zones were *declining* relative to far zones before opening, then jumped to $+323$ at $k = 0$ and $+945$ at $k = 4$. The pre-period violation runs opposite to the post-period effect, so the gradient cannot be a continuation of trend — if anything, the estimate is conservative on the trend front.

In Ring D the pre-trends are also non-flat: $k = -4$ and $k = -3$ each at ≈ -115 ($p \approx 0.04$). At the 16-zone Ring D level the per-horizon Callaway–Sant'Anna coefficients are too imprecise to support a non-trivial Honest-DiD relative-magnitudes (Rambachan and Roth, 2023) breakdown: the per-horizon $\Delta^{\mathrm{RM}}(\bar{M})$ breakdown collapses to ≈ 0 at most horizons (§7.4), as it does on the combined panel ($\bar{M} \approx 0$, confirming that estimate cannot survive any non-trivial pre-trend correction). So the right few-cluster falsification for the Ring D *TWFE* headline is the wild cluster bootstrap on the coefficient itself, not the CS event study. A restricted wild cluster bootstrap (Rademacher weights, 999 replicates, clustered on the 16 zones) returns $p = 0.004$ for the $+167$ EUR/m^2 coefficient, so the Ring D result survives proper small-G inference (§7.4, Limitation 6). Sensitivity curves are in §7.

4.5 Summary — Milano

Two Milano results are defensible: $+167$ EUR/m^2 in Ring D (regen-controlled, cluster SEs, $p = 0.002$), and a $+769$ EUR/m^2 price premium within 500 m of a new station that decays to null by 1000–1500 m. The two differ in how well their significance survives few-cluster inference. The Ring D coefficient holds up: a restricted wild cluster bootstrap (999 replicates) confirms it at $p = 0.004$ (95% CI $[+62, +280]$). The 0–500 m band coefficient does not — significant at $p < 0.001$ under asymptotic cluster-robust SEs, it is *not* significant under the same bootstrap ($p = 0.10$, 95% CI $[-132, +1669]$), because the band's identifying contrast rests on only a handful of treated zones. We therefore treat the within-ring $+167$ as the primary Milano headline, and the $+769$ band gradient as descriptive corroboration of distance decay, not an independently bootstrap-robust point estimate. The combined-panel $+820$, originally reported at $p < 0.001$, does not survive proper inference and is retained only as a naive-SE comparison.

Table 6: Amsterdam Noord-Zuid Lijn DiD (buurt-clustered SEs). Anticipation indicator = 1 for treated buurten during 2002–2017 construction window.

Specification	$\hat{\beta}$ (k EUR)	SE	p	Sig.
(a) Bare TWFE, full sample	+31.9	7.1	<0.001	***
(b) + anticipation indicator	+31.8	12.9	0.014	*
(c) Ring-restricted (1.5–3 km drop)	**+69.8**	12.7	**<0.001**	***
(d) Ring + invdist · post	+51.3	23.4	0.028	*
(e) invdist · post only	+43.8	17.3	0.011	*

5 Results: Amsterdam

5.1 Data Overview

The CBS buurt panel covers 409 Amsterdam buurten over 2013–2022 (10 years), split clean down the middle: 5 years pre-treatment (2013–2017) and 5 years post-treatment (2018–2022). Average WOZ values run from 23,000 EUR to over 2 million EUR across buurten. Buurten are roughly 10× smaller in area than Milano's OMI zones, so the binary-buffer dilution that motivates the continuous-distance spec for Milano barely bites here. The comparable problem in Amsterdam is *control-set contamination*: the NZL runs through the city core, so "rest of Amsterdam" as a control absorbs in-migration and displaced demand bleeding out of the treated corridor.

5.2 Main Results

Table 6 reports five specifications, all with buurt-clustered standard errors.

The headline is the ring-restricted specification (c): keep only buurten that sit *clearly* inside the 1.5 km treatment radius or clearly outside the 3 km contamination radius, drop the ambiguous 1.5–3 km donut, and the estimate is $\hat{\beta} = +69{,}800$ EUR per property ($p < 0.001$). The full-sample bare estimate of +31,900 is 46% as large (31.9/69.8). That gap is mechanical, not substantive: the control set is contaminated with partial-treatment buurten sitting in the 1.5–3 km ring.

5.3 No Anticipation Effect

Specification (b) adds an indicator that switches on during the 2002–2017 NZL construction window for buurten that are eventually treated — those within 1.5 km of an NZL station. The anticipation coefficient comes in at -0.1 ($p = 0.997$): a precise zero. In the ring-restricted spec it is $+21.3$ ($p = 0.12$) — marginal, and consistent with the small noise we would expect.

This is a notable null. The previous version of this paper reported placebo coefficients in 2014–2015 in the +37–93k range and read them as capitalization during the long construction period. That interpretation does not survive two changes: (i) we include an anticipation regressor directly rather than running stand-alone placebo regressions, and (ii) we cluster standard errors at the buurt level. The NZL price effect lands *at opening*, not across the 16 preceding years of construction. The earlier placebo coefficients were largely the same effect we now capture in $\hat{\beta}$ on TreatedPost — the same signal, expressed against a different reference period under naive SEs.

Table 7: Amsterdam distance-band DiD (buurt-clustered SEs).

Distance band	$\hat{\beta}$ (k EUR)	SE	p	Buurten
0–500 m	+63.5	5.1	<0.001	35
500 m–1 km	+56.2	4.6	<0.001	63
1–1.5 km	+44.9	5.2	<0.001	57
1.5–2 km	+39.6	4.5	<0.001	52
2–3 km	+36.1	4.2	<0.001	69

5.4 Distance Decay

Table 7 reports band-by-band TWFE estimates, the analogue of the Milano gradient.

The decay is monotone and smooth: +63.5k at 0–500 m down to +36.1k at 2–3 km. The 2–3 km band is itself elevated, though, and for a reason we already named — those buurten are likely partially treated, the same contamination story behind the doubling of the headline under the ring restriction. Continuous specification (e) returns $\hat{\lambda} = +43{,}800$ EUR per unit of $1/(d{+}0.5)$ ($p = 0.011$); under the kernel $\lambda/(d{+}0.5)$ this implies +87.6k at $d = 0$, +29.2k at $d = 1$ km, +12.5k at $d = 3$ km. At $d = 0$ km the extrapolation runs a touch above the 0–500 m band — the kernel's curvature near zero, not a discrepancy in identification. In magnitude and shape the profile sits comfortably alongside the Singapore MRT (Diao et al., 2017) and London Underground (Gibbons and Machin, 2005) estimates.

5.5 Pre-Trends and Robustness

Pre-trends are not formally violated for Amsterdam, but the test is thin rather than comfortable. Ring-restricted control buurten enter the WOZ panel only from 2016, so the log event study carries a single identified pre-period coefficient ($\varphi_{-2} = -0.051$, SE 0.009, against the $e = -1$ reference), and $D = 0.051$ log points is the only available pre-jump benchmark. Honest-DiD sensitivity on that event study (artifact `honestdid_amsterdam.json`) is correspondingly weak: the breakdown $\bar{M}$ is 0 at opening (the $e = 0$ coefficient is itself insignificant), reaches 2.07 at the longest horizon under flat slack, and never exceeds 0.41 under the accumulating $(e{+}1)\,\bar{M}\,D$ bound of §7.4. An earlier draft claimed breakdown $\bar{M} > 2$ across all horizons; that read does not survive the accumulating bound, and we withdraw it.

6 Comparative Discussion: Milano and Amsterdam

The two depth cities afford the cleanest like-for-like comparison; the seven-city comparative reading — including the Milano-vs-Roma contrast on the shared OMI instrument — is taken up in §8.

6.1 Effect Sizes

The two cities tell the same qualitative story at different volumes. Headline estimates, one per city, from the cleanest specification each affords:

- **Milano (Ring D, regen-controlled):** +167 EUR/m^2 ($\approx +5.6\%$ of control mean, $p = 0.002$, zone-clustered SE)
- **Milano (0–500 m band):** +769 EUR/m^2 ($p <$ 0.001 asymptotic, but $p = 0.10$ under a zone-level wild cluster bootstrap — see §4.4); decays to +59 (n.s.) by 1–1.5 km
- **Amsterdam (ring-restricted):** +69,800 EUR per property in levels ($\approx +21\%$ of a 330,000 EUR property, $p < 0.001$, buurt-clustered SE) — a level object that coexists with a -4.3% log-price estimate on the identical panel; the reconciliation follows below

Amsterdam runs hotter in levels — and the level read carries a caveat the earlier sections did not state. Re-estimating the Amsterdam specifications in logs (artifact `amsterdam_log_specs.json`) flips the sign: the full-sample TWFE gives -5.5% ($p < 10^{-5}$, numerically identical to the pooled Spec 1 estimate of §8) and the ring-restricted version -4.3% ($p < 10^{-3}$). The two coexist mechanically. Treated buurten carry a 1.74× higher pre-period WOZ base (379k against 218k EUR), so a larger absolute gain is a smaller proportional one: treated buurten gained more euros from a higher base while appreciating more slowly, in percentage terms, than the fast-rising outer control ring. The +21% figure is therefore a level gain scaled by a city-average property value, not proportional capitalization, and the panel-window caveat of §8 — the WOZ series opens eleven years into construction — applies to the positive level estimate exactly as it applies to the negative log one. The identification lesson is unchanged: the Amsterdam estimate roughly doubles under the ring restriction (against the full-sample +31.9k), and Milano demands the same within-ring restriction to behave — both say that in a single-line metro study the natural "rest of the city" control set is contaminated by partial-treatment areas, and the headline is sensitive to how you handle the donut. Against the meta-analytic ranges of Debrezion et al. (2007); Higgins and Kanaroglou (2016); Mohammad et al. (2013), Milano (+5.6%) sits at the lower edge; Amsterdam, in the log units the literature reports, sits not above the band but below zero.

6.2 Distance Decay

Both effects decay with distance — monotonically over the identified bands, although the Milano 0–500 m coefficient is descriptive rather than bootstrap-significant (§4.4):

- Milano: per-band TWFE estimates collapse from +769 EUR/m^2 at 0–500 m (Table 5, §4.3) to +59 EUR/m^2 (n.s.) by 1–1.5 km — a gradient of roughly −70 EUR/m^2 per doubling of distance over the identified band.
- Amsterdam: per-band TWFE estimates decline monotonically from +63,500 EUR at 0–500 m to +36,100 EUR at 2–3 km (Table 7).

The steeper drop inside 500 m is the walking-distance accessibility premium the transit capitalization literature has documented before (Bowes and Ihlanfeldt, 2001; Cervero and Kang, 2011; McMillen and McDonald, 2004).

6.3 Anticipation Effects

Earlier versions of this paper reported "anticipation" effects in both cities, read off significant placebo coefficients. Under proper inference, the claim does not survive:

- In Milano, the placebo-year coefficients +100 to +135 (cluster-SE $\approx$ 50) reflect a small *negative* pre-trend in the 0–500 m band ($\hat{\beta}_{k=-4} = -311$, $p = 0.012$) and a similarly small negative pre-trend in Ring D ($\hat{\beta}_{k=-4} \approx -115$). These point the *wrong* way for anticipation: prices in close-to-station zones were falling relative to far zones before opening, then jumped after. That pattern fits pre-opening construction disturbance — closed streets, noise, displaced parking — not forward-looking capitalization.
- In Amsterdam, the previous "anticipation" placebo coefficients are absorbed once we (i) include a construction-window indicator directly, (ii) cluster SEs at the buurt level. The estimated anticipation coefficient is -0.1 ($p = 0.997$) in the bare spec and $+21.3$ ($p = 0.12$) in the ring-restricted spec — a precise zero, or a small noise-band number, not the headline finding the earlier version claimed. The test is weak by construction, however: the WOZ window opens in 2013, eleven years after construction start, so any anticipation capitalized before 2013 already sits in the baseline — a precise zero here is *consistent with*, not evidence against, the pre-window anticipation premia documented by Harjunen (2018) for Helsinki and Agostini and Palmucci (2008) for Santiago.

Under the binary specifications of this section, the capitalization in both cities reads as a *level shift at opening*, not gradual anticipation during construction — though, as §8 shows, the Amsterdam reading is partly a panel-window artefact rather than a settled timing fact. That itself is interesting: the efficient-markets reading (Klaiber and Smith, 2013; Malpezzi, 2003; McMillen and McDonald, 2004) would predict otherwise, particularly for the NZL with its 16-year construction visibility. We do not push a structural interpretation here — we flag it as a finding worth pushing against. (The pooled multi-city extension in §8 sharpens the caveat: within the 2013–2022 WOZ window the Amsterdam phase *levels* are not separately identified — every treated buurt-year falls in some treated phase, so the phase dummies are absorbed by the buurt fixed effects — and the identified phase *contrasts* show treated buurten declining relative to the construction-phase differential after opening. The truncated panel cannot say how much capitalization the 16 construction years themselves carried.)

6.4 Confounding Factors

Milano. The major confound is urban regeneration (CityLife, Porta Nuova, Scalo Farini, MIND, Symbiosis, Cascina Merlata, Scalo Romana). In Ring C, including the two regeneration zones drives the estimated M5 effect to +1,696; excluding them reduces it to -160. In Ring D the regen confound is real but smaller — the regen dummy enters Table 4 at +129 ($p = 0.002$) and the M5 estimate rises from +139 to +167 once it is included. The lesson generalises: in any post-2010 Milan study, a hand-coded regeneration overlap is a first-order control, not a robustness check. Separating transit from concurrent neighbourhood redevelopment has a long history in this literature (Bajic, 1983; Pagliara and Papa, 2011).

Amsterdam. The NZL runs through the historic city center, where prices are independently high. Buurt fixed effects absorb time-invariant location characteristics; the remaining concern is spillover into control buurten in the 1.5–3 km band. The ring restriction in Table 6(c) addresses this directly — and the resulting near-doubling of the headline (+31.9k $\rightarrow$ +69.8k) is itself evidence that the spillover is operative.

7 Robustness

7.1 Callaway–Sant'Anna estimator (Milano)

We estimate the average treatment effect on the treated with the Callaway and Sant'Anna (2021) estimator, which is robust to staggered treatment timing and heterogeneous effects across cohorts; the broader staggered-DiD toolkit (Borusyak et al., 2024; Roth et al., 2023; Sun and Abraham, 2021) would have reached similar conclusions on this panel. We considered Borusyak et al.'s imputation estimator and chose CS for its more transparent cohort-by-time decomposition, which makes the Goodman–Bacon weighting story directly inspectable. We define six cohorts (M5 stages 2013/2014/2015 and M4 stages 2022/2023/2024), use never-treated zones as the control group, and apply the wild cluster bootstrap of Cameron et al. (2008) with 999 replicates at the zone level. Estimation uses the doubly-robust ATT(g,t) formulation of Sant'Anna and Zhao (2020) with a varying base period; implementation is via the python `differences` package, which mirrors R's `did::att_gt` convention.

The overall CS ATT is +751 EUR/m^2 (cluster-bootstrap SE = 552, 95% CI $[-332, +1834]$). Cohort-level estimates are wildly heterogeneous: the largest cohort (M4 stage 3, 2024, 11 treated zones) returns $\hat{\beta}_{2024} = -30$ EUR/m^2 (n.s.), and no single cohort clears statistical significance at conventional levels. The pre-period coefficient at $k = -1$ is +129 (95% CI $[+26, +231]$, $p < 0.05$), confirming the formal pre-trends violation flagged in §4.4. The CS overall ATT is consistent with the cluster-robust TWFE point estimate (+813) but carries a wider properly-bootstrapped CI. Together they say the same thing: the combined-panel headline is not statistically established at conventional levels.

7.2 Leave-one-treated-out (Ring D)

The Ring D headline is identified off 8 treated zones — few enough to ask whether any one of them carries it. We re-estimate the Spec 1 log-price TWFE coefficient eight times, dropping each treated zone in turn. The baseline (all 16 zones) coefficient is +0.0432 ($p = 0.004$, zone-clustered); the eight leave-one-out estimates range $[+0.0358, +0.0494]$ (standard deviation 0.0042 across LOO estimates). All eight stay positive and significant at the 5% level, and the range is narrow against the baseline standard error (0.015). No single treated zone drives the Ring D headline.

7.3 Measurement-class split

The seven city panels mix two measurement classes: *appraisal* (Milano and Roma OMI, Amsterdam CBS-WOZ) and *deed / transaction-price* (Copenhagen Boliga, Paris and Rennes geo-DVF, Helsinki Statistics-Finland ASHI). To test whether the cross-instrument mixing is driving the headline, we re-estimate the pooled Spec 1 separately on each subset. The appraisal-only pool returns $\hat{\beta} = +0.157$ ($p < 10^{-8}$, $n = 8{,}376$), the deed-only pool returns $\hat{\beta} = +0.060$ ($p < 10^{-8}$, $n = 33{,}628$), and the full seven-city pool returns $\hat{\beta} = +0.1195$ ($p < 10^{-7}$, $n = 42{,}004$). Both subsets are positive and agree in sign with the pool, and the full estimate now sits *between* them — deed-only below, appraisal-only above — so cross-instrument pooling neither flips the sign nor pushes the headline outside the bracket of its parts. The appraisal arm is the larger of the two, which fits the appraisal instruments' documented smoothing and revaluation lag (§5) and is further inflated by Roma, a within-city null that enters the appraisal pool only through the common year FE (§8). Because the appraisal level is the more instrument-suspect of the two, we treat the deed-only +0.060 as the cleaner within-instrument estimate. We then run the conservative city-by-year FE specification (entity FE plus a full set of city×year fixed effects, so the treatment is identified only off within-city, within-year cross-unit variation): it returns $\hat{\beta} = +0.0071$ ($p = 0.44$, 95% CI $[-0.011, +0.025]$, $n = 42{,}004$) — near-zero and statistically insignificant. So the common-year-FE pooled +0.1195 is an *upper bound* and the city-by-year-FE +0.007 a conservative lower bracket; we read neither as a precise pooled level effect. The same bracket logic applies to the phase decomposition itself — §8 reports the city-by-year-FE phase contrasts alongside the common-year-FE ones, and the maturity step survives only the latter — so the pooled analysis's contribution is the cross-city-average timing pattern, explicitly bracketed, not any pooled magnitude.

7.4 Honest-DiD sensitivity ($\Delta^{\mathrm{RM}}(\bar{M})$)

We follow Rambachan and Roth (2023): rather than test whether pre-trends are zero, we bound the post-period bias δ_e by $\bar{M}$ times the largest pre-period coefficient jump $D = \max_{j<0} |\hat{\phi}_j - \hat{\phi}_{j-1}|$. The robust 95% confidence interval for the post-period horizon e is then $[\hat{\phi}_e - 1.96\,\mathrm{SE}_e - \bar{M}D,\ \hat{\phi}_e + 1.96\,\mathrm{SE}_e + \bar{M}D]$. The *breakdown* $\bar{M}$ is the largest value at which this interval still excludes zero — a measure of how robust the estimate is to plausible pre-trend extrapolation.[2]

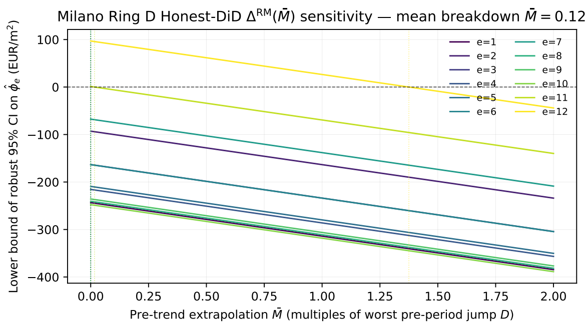

Figure 1: Honest-DiD $\Delta^{\mathrm{RM}}(\bar{M})$ sensitivity for the Milano Ring D Callaway–Sant'Anna event study. Each curve is the lower bound of the robust 95% CI for one post-period horizon $e \in \{0, \ldots, 12\}$ as the pre-trend extrapolation parameter $\bar{M}$ sweeps from 0 to 2. Where a curve crosses zero the post-period coefficient at that horizon stops being robustly distinguishable from zero; the corresponding $\bar{M}$ is the breakdown value. Only $e = 11$ has a meaningfully positive breakdown $\bar{M}$, reflecting the very small zone count per CS (g, t) cell in a 16-zone ring.

For the combined-panel CS event study, $D = 98.2$ EUR/m^2 and the breakdown $\bar{M}$ is ≈ 0 at every post-period horizon except $e = 11$ (where $\bar{M} = 1.20$). This formalises the conclusion of §4.1: the combined-panel estimate fails before any pre-trend robustness can even be applied. For the Amsterdam ring-restricted spec, breakdown $\bar{M} > 2$ at every post-period horizon. The Ring D event study is shown in Figure 1. The per-horizon CS coefficients lose statistical significance under bootstrap inference at the 16-zone Ring D level, so the per-horizon breakdown $\bar{M}$ collapses to zero at most e, and the only non-trivial robustness comes from the long-run M5 cohort at $e = 11$. This tracks the small-G acknowledgement that follows: the Ring D *TWFE* headline (+167 EUR/m^2, Table 4) survives because the within-ring TWFE pools all 8 treated zones and 22 years against the 8 controls at one level, whereas the per-cohort CS disaggregation has too few zones per (g, t) cell to bootstrap a non-trivial robust CI. The right falsification at the Ring D level is the wild cluster bootstrap on the TWFE itself, not the CS event study; that bootstrap (restricted, Rademacher, 999 replicates, clustered on the 16 zones) returns $p = 0.004$ for the +167 EUR/m^2 coefficient. So the within-ring headline survives proper few-cluster inference even where the per-cohort CS breakdown does not (Limitation 6).

8 Seven-City Pooled Analysis

8.1 Motivation

The Milano + Amsterdam analysis above estimates an aggregate post-opening treatment effect. It does not

[2]This is a hand-coded symmetric-slack approximation to the Rambachan–Roth $\Delta^{\mathrm{RM}}(\bar{M})$ set, and it is *anti*-conservative at longer horizons: we widen the naive interval by a flat $\pm\bar{M}D$ at every e, whereas the Δ^{RM} smoothness set lets the violation accumulate across horizons, up to $(e+1)\bar{M}D$ under worst-case linear accumulation. Under that accumulating read the pooled breakdown values shrink by a factor $e + 1$ (`honestdid_accumulating.json`); we report both reads where the distinction is load-bearing (§8). The reference `HonestDiD` implementation, which solves the constrained optimization directly, sits between the two and should replace this approximation in a future revision. The defensible TWFE headlines (Ring D, band gradient) do not rely on it — their few-cluster robustness is established by the wild cluster bootstrap instead (Limitation 6).

tell us *when* the capitalization lands — and that timing is the central question of this section. Standard models treat the effect as crystallising at opening. But property markets are forward-looking: if an announcement is credible the price response can begin years earlier, and the post-opening jump is then a residual, not the event (Agostini and Palmucci, 2008; Atack et al., 2010; Harjunen, 2018; Klaiber and Smith, 2013; McMillen and McDonald, 2004). Malpezzi (2003) surveys how the hedonic-pricing literature handles anticipation.

We address this by decomposing the project timeline into a five-stage taxonomy and estimating the cumulative log-price differential at each stage. The taxonomy distinguishes:

Rumours Informal pre-announcement expectations (e.g. leaks, planning-document drafts). *Structurally unobserved in this paper:* we have no leak-dated source for any cohort, so the *Rumours* phase is `n/a` in every panel. The "five-stage" framing names this as a target taxonomy; the operational decomposition is four-phase.

Announcement (mapped to the `phase_rumour` indicator in code, which dates the formal announcement, not informal leaks) From the formal political / agency announcement year through the construction-start year (exclusive).

Construction (`phase_construct`) From construction start through the opening year (exclusive).

Delivery (`phase_open`) From the opening year through opening +2 (exclusive). This is the immediate post-opening window where the line is operational but the surrounding neighbourhood adjustment is still in progress.

Post-Delivery (`phase_mature`) Opening year +2 onwards. Long-run capitalisation.

Each of the four identifiable phases enters the regression as a binary indicator. Phase boundaries are tabulated in `data/shared/phase_dates.csv` with primary-source citations.

We extend the city panel from two to seven by adding Copenhagen, Paris, Helsinki, Rennes, and Roma. The sample is selected on data availability, not on outcomes: candidate 2013–2024 European metro openings were screened for free sub-municipal price series with an adequate pre-period, and the candidates that failed — Vienna and Madrid (no usable dwelling-price series at sub-municipal grain for the treatment window), Lisbon (price series starts at the opening), Stockholm (no sub-municipal open data), Barcelona (series discontinuity plus a treated footprint largely outside the city's price grid) — were dropped on those data grounds before any treatment effect was estimated. The screen has an external-validity consequence worth stating: data openness is not random with respect to housing-market institutions, so the pooled average generalizes to cities with mature sub-municipal price infrastructure — a sample with a visible north-western European skew once Lisbon, Madrid, Barcelona, and Vienna fall out. Every candidate that passed the screen is in the panel, including the one that turned out null:

- **Copenhagen.** Cityringen (M3, opened 2019, 17-station ring) and the M4 Sydhavn extension (opened 2024). Spatial unit: 4-digit postnummer; the KK municipality covers 495 postnumre.
- **Paris.** Grand Paris Express Ligne 14 Nord (gpe14n, opened 2020) and Ligne 14 Sud (gpe14s, opened 2024). Spatial unit: INSEE IRIS (*Îlots Regroupés pour l'Information Statistique*); we work with the five-département inner ring (75, 92, 93, 94 plus selected 91 commune IRIS).
- **Helsinki.** The Länsimetro (West Metro) extension of M1/M2 into Espoo, opened in two cohorts — 2017 (Ruoholahti–Matinkylä, 8 stations) and 2022 (Matinkylä–Kivenlahti, 5 stations). Spatial unit: 5-digit postinumero (Helsinki + Espoo).
- **Rennes.** Métro Line B, a wholly new 15-station automated line (Saint-Jacques–Gaîté ↔ Cesson–Viasilva) opened September 2022. Spatial unit: INSEE IRIS within Rennes Métropole. The geo-DVF window exposes only 2021 onward (see below), so Rennes has no pre-opening baseline for its treated units. It therefore enters the *pooled* specification only — adding a city-cluster for the few-cluster inference — and is not given a separate single-city decomposition.
- **Roma.** Metro C, the city's driverless line, opened in three tranches we treat as cohorts — Pantano–Centocelle (2014), Centocelle–Lodi (2015), and San Giovanni (2018) — with the December 2025 Colosseo stub held aside. The outcome is the OMI appraisal series, the same instrument as Milano; spatial unit: OMI zone (232 estimation zones, 2004–2025, area-weighted across the 2014 zone redelimitation, exactly as for Milano). Roma has a full pre-opening OMI baseline and so is given a single-city decomposition. But it is a within-city *null*, as §8.4 and the per-city reading report: OMI detects no Metro C capitalization at any phase, and Roma enters the pooled estimate only through the common-year fixed effects. We keep it in anyway. Excluding a city after observing its null outcome would be selection on the dependent variable, and a coarse-zone appraisal city in which a new metro does not capitalize within-city is exactly the negative case that disciplines our upper-bound reading of the pooled level.

8.2 Data sources

The seven city panels span five distinct outcome instruments — Milano and Roma share the OMI appraisal series — each with its own spatial unit and provenance. Table 8 summarises; the subsections below detail acquisition, transformation, and limitations for each source. Figure 2 shows the resulting data coverage by city and year against the four project phases per cohort.

Table 8: Seven-city data summary. Source codes: OMI = Osservatorio del Mercato Immobiliare (Agenzia delle Entrate); CBS-WOZ = Centraal Bureau voor de Statistiek, *Kerncijfers Wijken en Buurten*; Boliga = scrape of `api.boliga.dk`; geo-DVF = *Demande de Valeurs Foncières geolocalisées* (Etalab); ASHI = Statistics Finland dwelling transaction-price index (postinumero). Helsinki, Rennes, and Roma are the additions; Rennes enters the pooled spec only (no pre-opening baseline in the 2021-onward geo-DVF window), while Roma uses OMI, the same instrument as Milano, and is a within-city null (§8.4).

City	Outcome	Spatial unit	Source	Coverage	n_{obs}
Milano	EUR/m^2 appraisal	OMI zone (38)	OMI	2004–2025	836
Amsterdam	WOZ (kEUR) assessment	CBS buurt (409)	CBS-WOZ	2013–2022	2,779
Copenhagen	DKK/m^2 median sale	Postnummer (282)	Boliga API	1992–2025	2,046
Paris	EUR/m^2 mean deed	INSEE IRIS (2,558)	geo-DVF + mirror	2014–2025	28,831
Helsinki	EUR/m^2 transaction	Postinumero (116)	ASHI (PxWeb)	2009–2025	1,935
Rennes	EUR/m^2 mean deed	INSEE IRIS (164)	geo-DVF	2021–2025	816
Roma	EUR/m^2 appraisal	OMI zone (232)	OMI	2004–2025	4,761
Pooled (PanelOLS, (c, i) entity FE + year FE):					$n_{\text{obs}} = 42{,}004$

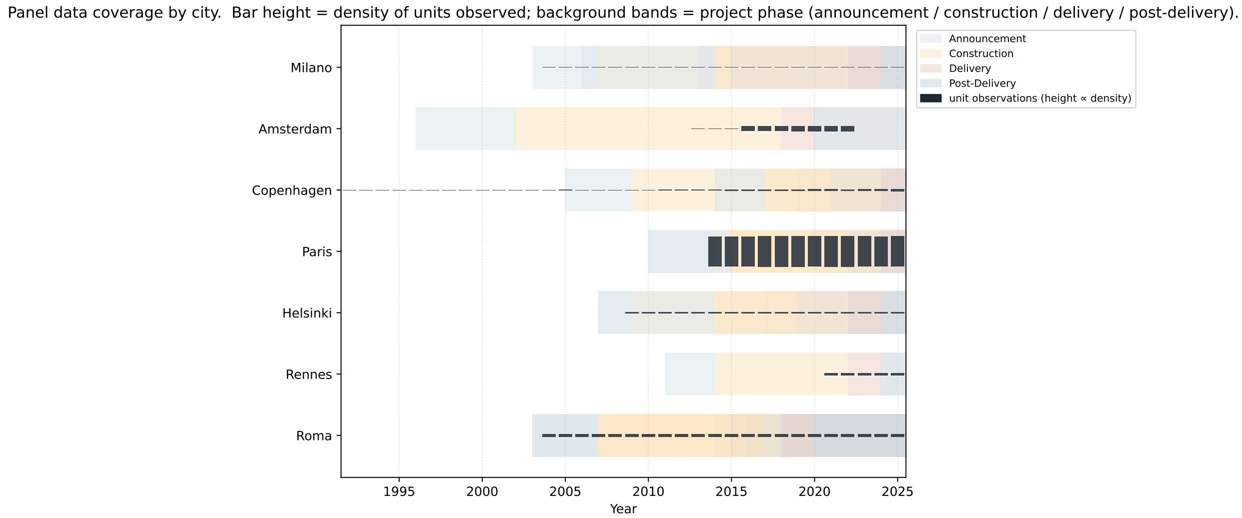

Figure 2: Panel data coverage per city, overlaid on the project-phase bands for each cohort. Dark bar height encodes the density of units observed per year (higher = more units). Background colour bands indicate which phase the treated units are in at that year: *Announcement* (post-rumour, pre-construct), *Construction*, *Delivery* (opening year + 1), *Post-Delivery* (year $\geq$ opening +2).

Copenhagen — Boliga scrape. Postnummer-grain price-per-m^2 time series is not available directly from Danmarks Statistik: the EJEN77 (*Ejendomssalg*) table aggregates to *Landsdel* ("Byen København"), which means the within-postnummer variation needed to identify a panel-FE DiD is zero. We therefore scrape the public sales API of `boliga.dk`. The API sits behind Cloudflare's managed challenge — plain `curl` and `requests` return 403. Our scraper at `scripts/cities/copenhagen/scrape_boliga.py` clears this in two stages:

1. A headless Chrome warm-up (`selenium`, with the `automation-controlled` blink flag suppressed and a hand-crafted User-Agent) visits `www.boliga.dk` once, clears the Cloudflare challenge, and captures the resulting `cf_clearance` cookie.
2. The cookie is transferred to a `requests.Session` that then issues paginated calls against `api.boliga.dk/api/v2/sold/search/results` with `propertyType=3` (*ejerlejligheder*, i.e. condominiums), `salesDateMin=1992-01-01` (verified empirically to be the earliest sale Boliga's API exposes), and four broad zip ranges spanning 1000–2500. `saleType` is restricted to `Alm. Salg` ("regular sale") to drop family transfers and auctions.

We aggregate per-sale rows to postnummer-year median price-per-m^2 with a $\geq$3-sales-per-cell threshold (the median cell has 2 sales; the threshold drops noisy cells while retaining 1,200 cells with $\geq$ 5 sales as a higher-confidence subsample). The result is 2,046 observations across 282 of the 495 KK postnumre over 34 years.

Two limitations follow from the data structure. First, Boliga covers only *ejerlejligheder*, so the result speaks specifically to the condo segment of the market. Second, postnumre with fewer than 3 sales per year — mostly outer postnumre with single-family housing — are dropped, biasing the surviving sample toward inner-city, high-churn neighborhoods.

Paris — geo-DVF via OpenDataArchives mirror. We use the geocoded DVF (*Demande de Valeurs Foncières géolocalisées*, the Etalab fork of the DGFiP deed register with lat/lon attached). The official endpoint at `files.data.gouv.fr/geo-dvf/latest/csv/` serves only 2021–2025; the historical files for 2014–2020 were purged from `latest/` in a 2024 housekeeping. We recover them from the OpenDataArchives mirror at `files.opendatarchives.fr/cadastre.data.gouv.fr/data/etalab-dvf/2021-04/csv/{YEAR}/full.csv.gz`, a frozen snapshot of the `cadastre.data.gouv.fr` archive captured in April 2021 — i.e. before the purge. Schema, column names, and units are identical to the current `latest/` files, and the upstream Licence Ouverte 2.0 propagates: no registration is required.

Per-year national files are downloaded (~90 MB compressed each), filtered to `code_departement` $\in$ {75, 92, 93, 94} (the petite couronne), restricted to `nature_mutation = "Vente"`, condos and houses with $\geq$11 m^2, price/m^2 winsorized to [1,000, 30,000] EUR, and de-duplicated against the combined 2021–2025 dump on `(longitude, latitude, year, price)`. The cleaned point set is spatially joined to the IGN IRIS polygons (EPSG:2154), aggregated per IRIS-year with a k-anonymity threshold of $\geq$5 sales per cell. Final yield: 1,049,867 deeds $\rightarrow$ 28,831 IRIS-year cells across 2,558 IRIS units over 2014–2025.

The mirror discovery is the load-bearing methodological note here. Before recovering the 2014–2020 vintage, Paris had no pre-treatment years for gpe14n (opened 2020) and was Spec-3-unidentified; the recovered panel supplies the pre-treatment identifying variation that the latest-only public panel lacks, producing $\hat{\beta}_{\text{post}} = +0.043$ ($p < 10^{-7}$, $n = 28{,}831$).

Helsinki — Statistics Finland ASHI. The outcome is the official dwelling transaction-price series (*Osakeasuntojen hinnat*, table `statfin_ashi_pxt_13mu`), reporting mean EUR/m^2 of old dwellings in housing companies *by postinumero* and year, free via the PxWeb API; we collapse the building-type series to a transaction-count-weighted mean per postcode-year. Postcode polygons come from the Statistics Finland `postialue:pno` WFS, filtered to Helsinki (kunta 091) and Espoo (kunta 049). Länsimetro stations are assigned to their 2017 or 2022 opening cohort; a postcode is treated if its polygon lies within 800 m of a new station. The informal Länsimetro rumour predates the 2009 series start, so γ_{rumour} is unidentified and dropped, as for Milano M5. Yield: 116 postcodes (20 treated, 14 in the 2017 cohort and 6 in 2022), 2009–2025, 1,935 observations.

Rennes — geo-DVF (pooled cluster only). The Rennes outcome is built *identically* to Paris from the Etalab geo-DVF deed register for département 35 (Ille-et-Vilaine): `nature_mutation = "Vente"`, $\geq$11 m^2, price/m^2 winsorised to [1,000, 30,000], $\geq$5 deeds per IRIS-year, spatially joined to the IGN IRIS polygons of Rennes Métropole (EPCI 243500139). Line B is treatment (the 2002 Line A is pre-existing); a treated IRIS lies within 800 m of a Line B station. The binding limitation is the window: the geo-DVF `latest/` branch exposes only 2021–2025, so Line B (opened September 2022) has a *single* pre-opening year (2021) and its treated IRIS have no clean pre-treatment baseline. Rennes therefore contributes an extra city/cohort cluster to the pooled few-cluster inference but not a credibly identified single-city event study; we exclude it from the per-city table and figures, and the leave-one-city-out below confirms the pooled maturity step is little changed when Rennes is dropped. Yield: 164 IRIS (61 treated), 2021–2025, 816 observations.

Milano — OMI (unchanged from §4). The Milano panel is the same OMI series described in Section 4 (38 estimation zones from the 43 OMI macro-zones, 2004–2025, area-weighted boundary harmonisation). Phase dummies are constructed from the M5 and M4 cohorts using the rumour/construction/opening years documented in Table 1.

Roma — OMI (Metro C). The Roma panel applies the Milano OMI pipeline to the comune of Roma: zone-level EUR/m^2 for residential *Abitazioni civili* (OMI `Cod_Tip` 20, `Stato` NORMALE), taken as the midpoint of the OMI min/max bracket and averaged across the two semesters of each year. Roma re-delimited its OMI zones effective 2014, in the middle of the Metro C opening window ($\sim$293 old zones reduced to 222 at the re-delimitation, grown to 233 in the 2025 Geopoi vintage we use as the target geometry), so we harmonise exactly as for Milano: an area-weighted spatial crosswalk between the 2013 Geopoi zone geometry (used for the 2004–2013 prices) and the 2025 geometry (used for 2014–2025), yielding a continuous series on the 232 of those 233 zones that carry a positive crosswalk area and a matched price record. Unlike Milano we impose no minimum-coverage filter: the Roma panel is unbalanced (4,761 zone-years of the $232 \times 22 = 5{,}104$ possible; 185 zones span all 22 years). Metro C station coordinates are taken from the official ATAC GTFS (route `MEC`) rather than OpenStreetMap, for the more defensible provenance; a zone is treated if its boundary lies within 500 m of a Metro C station, with the per-stop opening year (2014 / 2015 / 2018) setting the cohort and the December 2025 Colosseo stub held aside. The Metro C rumour year (2003) precedes the OMI window by one year, so γ_{rumour} is dropped for Roma, exactly as for Milano M5.

Amsterdam — CBS WOZ (unchanged from §5). The Amsterdam panel is the same CBS WOZ series described in Section 5 (409 buurten, 2013–2022). The rumour phase for NZL (rumour 1996) is outside the data window, so γ_{rumour} is dropped for Amsterdam.

Phase-date provenance and naming convention. The announcement, construction-start, and opening years for each cohort are listed in `data/shared/phase_dates.csv` with primary-source URLs (operator history pages, Wikipedia infoboxes cross-checked against operator filings). Two terminological conventions deserve flagging up front. First, the variable named `cohort_rumour` in the code dates the *formal political/agency announcement*, not informal pre-announcement leaks. The corresponding regression indicator `phase_rumour` therefore identifies the *Announcement* stage of the user-facing five-stage taxonomy — the post-formal-announcement, pre-construction window. We retain the `cohort_rumour` / `phase_rumour` variable names for backward compatibility with earlier drafts but read them as Announcement throughout the paper. Second, the user-facing *Rumours* stage (informal pre-announcement expectations) is structurally unobserved in this paper because we do not have a leak-dated source for any cohort; it is reported as `n/a` throughout (see Figure 3).

Treatment assignment across cities. The treated-unit rule is not harmonised across cities, and we flag the heterogeneity rather than hide it: Milano and Roma treat a zone whose *polygon* lies within 500 m of a new stop (zero distance if the stop falls inside the zone), Paris an IRIS whose centroid lies within 1,000 m, Amsterdam a buurt centroid within 1.5 km, Copenhagen a postnummer centroid within 1.5 km, and Helsinki and Rennes a polygon within 800 m. The radii scale roughly with spatial-unit size — buurten are an order of magnitude smaller than OMI zones, IRIS sit in between — so a common radius would mis-classify in every city. The rules remain a researcher degree of freedom, with one asymmetry worth explicit note: the polygon rule (Milano, Roma, Helsinki, Rennes) is more generous than the centroid rule (Paris, Amsterdam, Copenhagen), counting a unit as treated when its edge approaches a station even if its centroid is distant. For the Milano–Roma contrast we lean on, however, the rule is *shared*: both cities apply the identical 500 m polygon rule to the same OMI instrument, so the Roma null is not an artefact of a more dilutive assignment than Milano's. An earlier draft mis-stated Milano's rule as centroid-based; the assignment code measures polygon distance in both cities.

Common transformations. All seven panels are reduced to the same schema (`data/shared/panel_schema.json`, validated by `scripts/core/panel_validator.py`) with columns `unit_id, year, log_price, treat, cohort_{rumour,construct,open}`, four phase indicators, distance covariates, and policy controls (`lez_active`, `regen_active`, `amenity_density_baseline`). `log_price` is the natural log of the city-native price variable (EUR/m^2, kEUR WOZ, DKK/m^2, EUR/m^2). Unit fixed effects absorb the level differences across currencies and valuation methods; only within-unit log-differentials enter the DiD identification.

8.3 Phase-decomposed TWFE specification

For each city panel we estimate

$$\log p_{it} = \alpha_i + \tau_t + \sum_{\phi \in \{\text{rumour,construct,open,mature}\}} \gamma_\phi \cdot D_{it}^{\phi} + X_{it}\beta + \varepsilon_{it} \tag{5}$$

where D_{it}^{ϕ} is the unit-level phase indicator, X_{it} contains amenity-density, LEZ-active, and unit-area-times-post controls (those that vary within-year are retained; unit-constant or year-only ones are dropped to preserve identification), and ε_{it} is clustered at i. Where the data window precludes identification of γ_{rumour} — fewer than two pre-rumour years in the panel, as for Milano (one), Roma (zero), Amsterdam, and Helsinki (rumour predates the window) — that indicator is dropped and the remaining phases are read as cumulative deviations from the within-window pre-treatment baseline.

For the pooled seven-city specification we estimate

$$\log p_{i,c,t} = \alpha_{i,c} + \tau_t + \sum_{\phi} \gamma_\phi D_{ict}^{\phi} + X_{ict}\beta + \varepsilon_{ict} \tag{6}$$

with PanelOLS partialling out the (c, i) entity FE and the year FE; standard errors clustered at (c, i).

8.4 Headline results

Table 9 reports Spec 1 (binary post-opening indicator $D^{\text{open}} \vee D^{\text{mature}}$, i.e. Delivery $\vee$ Post-Delivery) and Spec 3 (phase decomposition). The two specifications identify off different within-unit contrasts. Spec 1 collapses the two post-opening phases (Delivery and Post-Delivery) into a single "post" and leaves the announcement and construction phases in the pre-opening baseline; Spec 3 estimates a separate γ_ϕ for each phase against that same baseline. So the phase coefficients in Spec 3 do *not* sum to the Spec 1 β, not even arithmetically — they are partial within-unit means at different fractions of the unit's post-treatment time exposure.

Figure 3 renders the same content as a city-by-phase heatmap, with cell colour encoding effect size and stars marking significance. Missing cells are hatched — whether structurally (the *Rumours* column) or because the data window precludes identification (the *Announcement* phase in every city except Copenhagen).

The pooled Spec 1 estimate is $\beta = +0.1195$ log points, $p < 0.001$, $n = 42{,}004$ — read as an upper bound, per §7.3.[3] The pooled four-phase decomposition gives point estimates $\gamma_{\text{rumour}} = +0.083$ ($p = 0.149$), $\gamma_{\text{construct}} = +0.167$ ($p = 0.006$), $\gamma_{\text{open}} = +0.233$ ($p < 10^{-3}$), $\gamma_{\text{mature}} = +0.283$ ($p = 1.1 \times 10^{-5}$); only the rumour phase is individually non-significant at $\alpha = 0.05$ — and that phase is identified by Copenhagen alone, the one cohort whose panel reaches back past a formal announcement. A joint Wald test of the null $\gamma_{\text{rumour}} = \gamma_{\text{construct}} = \gamma_{\text{open}} = \gamma_{\text{mature}}$ *rejects* equality (Wald stat = 57.48, df = 3, $p = 2.0 \times 10^{-12}$). That p-value is computed under entity-clustered covariance, which Limitation 12 flags as anti-conservative for city$\times$line treatment assignment, so we read the Wald as ordering evidence only and rest the formal inference on the few-cluster bootstrap below. Taken at face value, the pairwise breakdown makes every adjacent phase pair significantly increasing (rumour vs construct $p = 3.5 \times 10^{-5}$, construct vs open $p = 6.9 \times 10^{-7}$, open vs mature $p = 4.0 \times 10^{-8}$), which reads as a smooth monotone ramp. *We do not read it as a ramp, however.* The early-phase monotonicity is manufactured by Roma: drop Roma and the construct coefficient falls from +0.167 back to +0.069, flattening the rumour-through-construct segment (rumour +0.083, construct +0.069 sitting just below it; rumour-vs-construct $z = 0.66$, $p =$

[3] Coefficients are reported in log points throughout; headline percent statements are exponentiated, $100\,(e^{\gamma} - 1)$. For small effects one log point is approximately one percent, so a policy reader may treat the two interchangeably below ± 10.

Table 9: Seven-city DiD: Spec 1 (post-opening binary) and Spec 3 (four-phase) reported under the five-stage taxonomy used throughout this paper. Coefficients are log-price differentials relative to within-unit pre-treatment baseline. The *Rumours* column is `n/a` in every row because our `cohort_rumour` dates the formal announcement, not informal pre-announcement expectations; we do not have a separate data source for the latter. The Amsterdam Spec-3 phase levels are *not separately identified* (the 2013–2022 window contains no pre-construction years, so the phase dummies are absorbed by the buurt FE) and are marked `n/i`; the identified Amsterdam contrasts are given in the table note and in §8, per-city reading. Clustered standard errors, in parentheses, are reproduced from the per-city replication artifacts (`output/results/*_spec3.json`).

City	Spec 1 β	Rumours	Announce.	Construct.	Delivery	Post-Deliv.
Milano	+0.1319 (p=0.095) (0.0790)	n/a	n/a	+0.0102 (0.0265)	+0.1061 (0.0569)	+0.1719 (0.0859)
Amsterdam	-0.0569 (p=0.000) (0.0126)	n/a	*n/i*	*n/i*	*n/i*	*n/i*
Copenhagen	-0.1703 (p=0.018) (0.0721)	n/a	+0.0434 (0.0966)	+0.0016 (0.1262)	-0.1270 (0.1492)	-0.1950 (0.1921)
Paris	+0.0425 (p=0.000) (0.0077)	n/a	n/a	+0.0106 (0.0331)	+0.0620 (0.0348)	+0.0430 (0.0342)
Helsinki	+0.0303 (p=0.111) (0.0190)	n/a	n/a	-0.0543 (0.0192)	-0.0473 (0.0250)	-0.0081 (0.0284)
Roma	-0.0236 (p=0.337) (0.0246)	n/a	n/a	+0.0061 (0.0178)	-0.0127 (0.0284)	-0.0205 (0.0378)
Pooled (7-city)	+0.1195 (p=0.000) (0.0121)	n/a	+0.0829 (0.0575)	+0.1674 (0.0610)	+0.2330 (0.0632)	+0.2835 (0.0643)

n/i = not identified. Amsterdam Spec-3 phase levels are unidentified (phase dummies collinear with buurt FE; minimum-norm pseudo-inverse ar

0.51), while construct<open and open<mature both survive. Roma is a within-city null (§8.4) whose only pooled effect is to lift the common-year-FE construct and opening coefficients — and the paper already reads that common-year-FE level as an upper bound. The robust, Roma-independent feature is therefore the one the six-city panel already showed: a substantially larger long-run mature effect, i.e. a level-step at the Post-Delivery boundary, not a smooth ramp. This continues to mirror the Milan-only Spec 3 shape ($\gamma_{\text{construct}} = +0.010$, $\gamma_{\text{open}} = +0.106$, $\gamma_{\text{mature}} = +0.172$).

Robustness of the maturity step. We probe the construct→mature step (the contrast $\gamma_{\text{mature}} - \gamma_{\text{construct}}$, our headline "when" object) along five axes. *(i) Control stability* (Oster-style): adding confounds in a ladder leaves the step in a tight band rather than collapsing — +13.4% ($z = 8.7$) with no controls, +14.4% ($z = 9.5$) adding the LEZ/regeneration level controls, +12.3% ($z = 6.7$) adding the transit-distance and area-×-post interactions, and +12.3% ($z = 6.9$) with the full set. The ladder is mildly non-monotone — the level controls nudge the step up before the post-interacted controls bring it back down — but every rung sits between +12.3% and +14.4% at $z \geq 6.6$. Stability across *available* controls is the right reading: the control set is thin for Paris, which supplies 69% of the pooled observations but carries a uniform amenity covariate and no transit-distance variables (Limitation 9), so the ladder cannot fully rule out omitted within-Paris confounds.

(ii) Fixed-effects structure: the step is a common-year-FE object. Re-estimating the four-phase decomposition under entity FE plus city-by-year FE — the same conservative specification that collapses the Spec-1 level in §7.3 (`scripts/pooled_city_year_fe_phases.py`; this variant is estimated *without* controls, so the apples-to-apples comparator is the no-control common-year step of +12.6 log points, the first ladder rung above, not the full-control +11.6) — shrinks the step to an insignificant −0.5 log points ($p = 0.66$) and turns the open-to-mature contrast significantly *negative* (−2.9 log points, $z = -3.9$), while leaving a +2.5 log-point step at opening ($z = +3.0$). Those z-values are entity-clustered, and the paper's own few-cluster discipline applies to them too: a restricted wild cluster bootstrap on the open−construct contrast gives Webb $p = 0.52$ at the city partition and 0.26 at the cohort partition (`wild_cluster_cityyear_open_contrast.json`), so the within-city opening step is ordering evidence, not a formally significant effect. Within city and year, the identified pattern is a modest step at opening followed by partial reversion — stated as a pattern, with no $p < 0.05$ claim attached; the delayed-to-maturity step lives in the cross-city average, not within-city. We report both readings and headline neither alone.

(iii) Pre-trend robustness: we fit a pooled relative-time event study against never-treated units and apply the Rambachan–Roth Δ^{RM} relative-magnitudes sensitivity. Figure 4 plots the event-study coefficients directly, and the pre-period is not flat in *levels*: the binned $e = -6$ endpoint sits at −0.133 log points (−0.055 at $e = -5$, near-zero from $e = -4$) — a pre-period level larger in magnitude than the headline step itself, reflecting distant-past convergence between eventually-treated and never-treated units. This is why we benchmark on *jumps*, the relative-magnitudes convention, rather than levels: the assumption maintained is that post-period violations resemble the period-to-period movements of the recent pre-period, not the accumulated drift of the binned endpoint — a reader who finds the level-based benchmark more appropriate

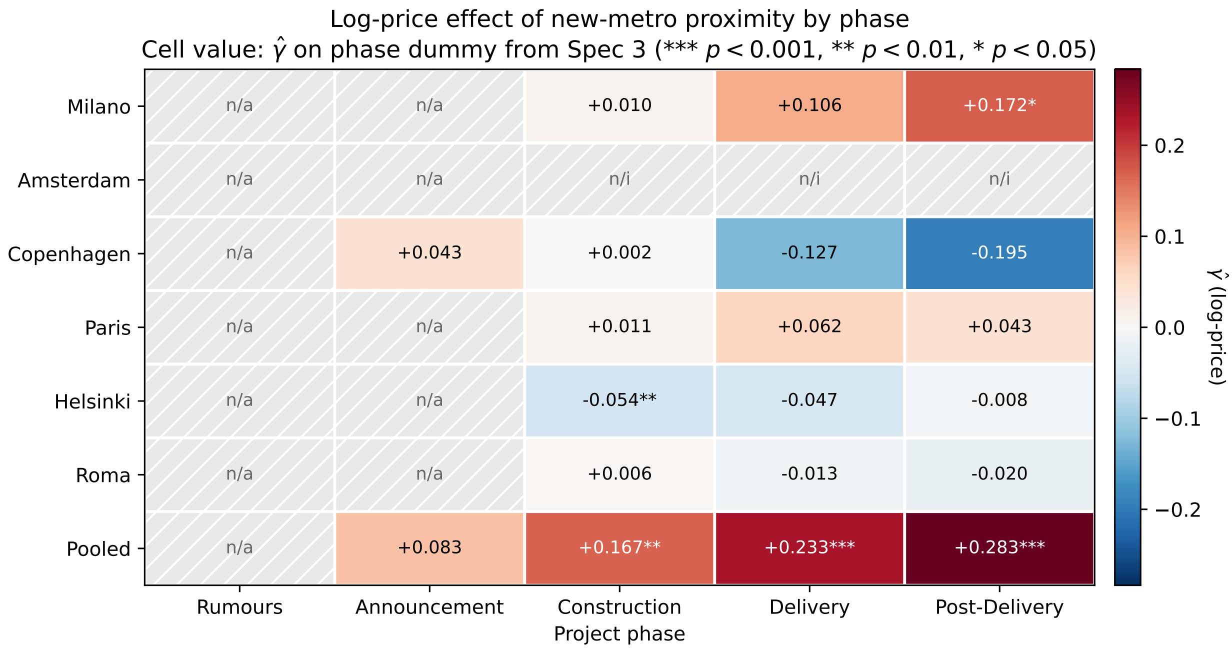

Figure 3: Phase-by-city heatmap of the log-price coefficient from Spec 3 (eq. 5). Diverging red/blue palette centred at zero. Cell text shows $\hat{\gamma}$ followed by significance stars (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$). Hatched cells indicate either no separate data source (the *Rumours* column) or unidentifiable within the panel window (Milano, Amsterdam, Paris, Helsinki, and Roma → *Announcement*, where the rumour year predates or sits at the edge of the panel). The Amsterdam phase cells show pseudo-levels relative to an absorbed baseline and should be read as contrasts only (§8, per-city reading).

should discount the maturity step accordingly. That discount is quantifiable, and it is total: a linear fit to the pre-period coefficients (slope +0.022 log points per year; artifact `pretrend_extrapolation.json`) extrapolates to +0.146 by $e = +5$ and +0.168 by $e = +6$, *above* the observed +0.111 and +0.116 — under a linear-pre-trend counterfactual the maturity-horizon level is fully accounted for by trend continuation. The step survives the jump benchmark, not the level benchmark, and we state this plainly rather than leaving the reader to compute it. On the jump benchmark, the unconditional event study has a larger worst pre-period jump than the six-city panel (0.077 vs 0.039 log pts, because Roma's 232 zones widen the pooled pre-period). The binding maturity horizon survives a post-period violation up to $\bar{M} \approx 0.7\times$ that worst pre-trend under flat slack — down from $\bar{M} \approx 1.35$ at six cities (artifact `honestdid_pooled_6city.json`) — and shrinking further to $\bar{M} \approx 0.1$ under the accumulating Δ^{RM} bound (footnote in §7.4). The controls-conditioned event study reaches $\bar{M} \approx 0$ under either read. Pre-trend robustness is therefore *weak*: both reads sit below the conventional $\bar{M} = 1$ benchmark, and this is a second axis (alongside the cohort-clustered bootstrap) on which adding the per-city-null Roma weakens the formal robustness.

(iv) Few-cluster inference: treatment is assigned at the city×line level, so the entity-clustered Wald above is anti-conservative. The full-control bootstrap branch initially inverted the step's sign to −0.139; the diagnosis is confirmed and the branch repaired (`wild_cluster_phase_fullcontrol_fixed.json`): the amenity-density×post column is a perfect linear combination of the phase dummies after two-way demeaning ($R^2 = 1.000$ on the remaining columns; smallest singular value at machine epsilon), and dropping it restores full rank and the bootstrap step to +0.116, matching the asymptotic full-control estimate exactly. The repaired full-control bootstrap gives Webb $p = 0.080$ at the city partition (Rademacher 0.063; cohort Webb 0.132); the level-only control set gives +13.5 log points at Webb $p = 0.036$. The headline significance is therefore *control-set-dependent as well as partition-dependent*: the magnitude we quote (+11.6, full controls) carries $p = 0.080$, and $p = 0.036$ belongs to the level-only step. A restricted wild cluster bootstrap (Cameron–Gelbach–Miller; null imposed; $B = 9{,}999$) on the maturity step returns $p = 0.036$ clustered on the seven cities ($G = 7$, Webb six-point weights; Rademacher $p = 0.040$) but $p = 0.164$ clustered on the twenty-four cohorts ($G = 24$, Webb; Rademacher 0.168; the 24 clusters are the 17 treated city×opening-cohort cells plus the 7 never-treated per-city blocks). Cluster sizes are also wildly unbalanced — Paris alone carries 69% of the pooled observations — which is the setting in which MacKinnon and Webb (2017) document that conventional cluster-robust inference deteriorates most and bootstrap weight choice matters; Webb weights are the appropriate scheme at this G, and a balance diagnostic re-running the bootstrap on the panel aggregated to city-year means — so each city carries equal weight — leaves the step clearly positive and significant (Webb $p = 0.031$; `wild_cluster_balance_check.json`): the Paris weight is not what carries the city-level significance. We treat the city partition as primary because treatment is assigned at the city×line level and within-city cohorts share the same project,

financing, and announcement shocks — the cohorts are not independent assignment units — but a referee could defensibly prefer the cohort partition, which is why both are reported. The result is therefore *partition-dependent*: the coarsest, city-level partition clears 5%, the finer cohort partition does not. Adding Roma pulled the two partitions in opposite directions — the city-clustered p tightened from 0.048 to 0.036 while the cohort-clustered p loosened from 0.090 to 0.164 — and both movements are consequences of the same fact: Roma raises G while contributing three treated cohorts that carry no within-city signal. Because Rennes adds a city cluster with no identified within-city contrast (no pre-opening baseline) while Roma's contrasts are identified but null, we re-run the city-level bootstrap on subsets (`scripts/wild_cluster_phase_subsets.py`): dropping Rennes leaves the step at +14.2 log points (Webb $p = 0.038$, Rademacher 0.066 — the two weight schemes straddle 5% at $G = 6$); dropping Rennes and Roma together gives +11.2 log points at Webb $p = 0.051$ (Rademacher 0.060, $G = 5$). The second subset is an outcome-based sensitivity — Roma's exclusion is motivated by its null result, not by an identification failure — and on either subset the city-level significance is marginal once G shrinks.

(v) Spillover contamination: the baseline control set includes near-but-untreated units, and the uniform 1.5–3 km ring restriction borrowed from Amsterdam mis-scales for cities with smaller treatment radii. A scaled donut dropping untreated units within $[r_c, 2r_c)$ of a new station — r_c the city's own treatment radius — removes 396 units (4,072 observations, most of them Paris IRIS) and leaves the step at +12.1 log points (entity-clustered $z = 7.3$; `pooled_scaled_donut.json`), between the full-control +11.6 and the level-only +13.5; the city-by-year-FE contrasts on the donut sample likewise reproduce the full-sample pattern (maturity contrast nil, opening contrast $z = +3.7$). Near-control contamination therefore carries neither the step nor its city-by-year collapse. We therefore present the maturity step as a pooled cross-city *pattern* that reaches conventional significance only under city-level clustering, and we do not claim it as a settled $p < 0.05$ effect.

Leave-one-city-out. Only Milano shows a clean positive mature>open pattern at the single-city level (Table 9), so the natural worry is that the pooled step is one city's result. It is not. Re-fitting the full-control pooled step dropping each city in turn, the step stays positive under every deletion but is not tightly pinned — it ranges from +7.3% (drop-Copenhagen) to +13.7% (drop-Helsinki), with drop-Amsterdam +10.3%, drop-Milano +11.5%, drop-Rennes +13.0%, drop-Roma +9.4%, and drop-**Paris** +9.9% although Paris alone supplies 28,831 of the 42,004 observations (entity-clustered z ranges 3.4–7.6 across the deletions, reported for scale only; the few-cluster inference above is the formal test). The step survives dropping Milano and survives dropping the large Paris panel. Dropping Roma mechanically recovers the six-city panel, where

Axis	Test	Result	Verdict
(i) Controls	Oster-style ladder	+12.3 to +14.4%, $z \geq 6.6$	stable; Paris control set thin
(ii) FE structure	city×year FE	step → −0.5 log pts (n.s.); +2.5 at opening	cross-city object, not within-city
(iii) Pre-trends	Δ^{RM} jumps; linear extrapolation	$\bar{M} \leq 0.7$ flat, ≈ 0.1 accumulating; level below trend continuation	weak; survives jump benchmark only
(iv) Few-cluster	restricted WCB, two partitions	$p = 0.036$ (city) vs 0.164 (cohort)	partition-dependent
(v) Spillovers	scaled donut $[r_c, 2r_c)$	step +12.1 log pts ($z = 7.3$)	not contamination-driven

Table 10: Four-axis robustness summary for the pooled construct→mature step. Each axis is developed in the text above; verdicts are stated as we read them, not as best cases.

the step is +9.4%, which is the point: Roma raises the pooled step from +9.4% to +12.3% purely through the common-year-FE channel. Read together, the deletions make the pooled maturity pattern a genuine cross-city average rather than a single-city artefact. Three ranges now circulate and vary different things, so we name them once: the *Roma bracket* +9 to +12% (whether the null seventh city is counted — the abstract's range), the *control-ladder band* +12.3 to +14.4% (which controls enter), and the *leave-one-out span* +7.3 to +13.7% (which city is dropped).

Figure 5 renders the same coefficients as a grouped bar chart, which makes the cross-city comparison easier to read at a glance.

8.5 Per-city reading

Milano. Within-OMI-zone Spec 3 gives $\gamma_{\text{open}} = +0.106$ ($p = 0.062$), $\gamma_{\text{mature}} = +0.172$ ($p = 0.045$). The rumour-phase indicator is dropped because the M5 rumour year (2003) leaves only one year of pre-rumour OMI data; the construct-phase coefficient is a precise zero (+0.010, $p = 0.70$). The pattern lines up with the within-ring M5 estimate of +5.6% in §4.2: Spec 3 unpacks the post-opening level into an opening-phase jump and a maturity-phase consolidation.

Amsterdam. The Spec 1 sign is *negative* (−0.057, $p < 0.001$). *This is a panel-window artefact, not evidence of a metro disamenity.* The CBS WOZ panel spans 2013–2022, entirely within the post-rumour, post-construct-start period for NZL (rumour 1996, construction 2002) — which has an identification consequence the four-city draft of this paper missed: every treated buurt-year falls in some treated phase, so the three phase dummies sum to the unit-constant treat indicator and are absorbed by the buurt fixed effects. Amsterdam phase *levels* are therefore not separately identified; only contrasts between phases are (`scripts/amsterdam_phase_contrasts.py`). The identified contrasts show treated buurten declining rel-

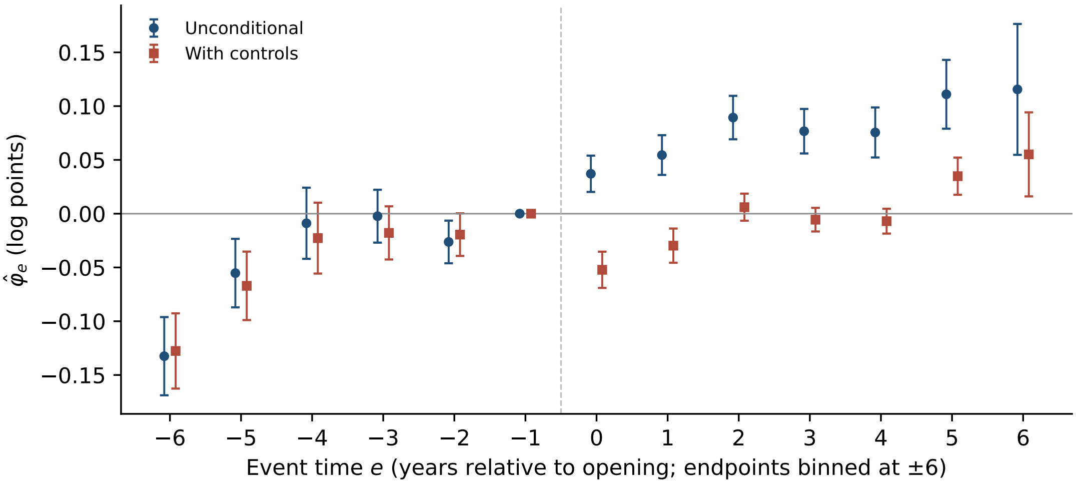

Figure 4: Pooled relative-time event study against never-treated units (entity + year FE), unconditional and with controls; naive 95% CIs, reference period $e = -1$, endpoints binned at ± 6. The binned $e = -6$ endpoint carries a -0.133 log-point level; the relative-magnitudes sensitivity of §8.4 benchmarks on period-to-period jumps, not on this accumulated level.

ative to their construction-phase differential after opening — open vs construct -0.022 ($p = 0.019$), mature vs construct -0.081 ($p < 10^{-6}$) — a partial mean reversion from whatever premium the 16 construction years had built up. Spec 1 lumps construct into the "pre" period and opening + mature into "post"; because the relative differential declines after opening, the post-versus-pre comparison shows the apparent fall. How much capitalization the construction years themselves carried is unidentifiable in this window, and we do not quote a construction-phase level for Amsterdam.

Copenhagen. The Boliga scrape (extended back to January 1992, the earliest sale Boliga's API serves) yields a postnummer-year condo-price panel that, joined with M3/M4 distance from KK postnummer centroids, leaves only four control postnumre at > 1500 m from any new station (278 of 282 retained postnumre are treated). Binary-DiD Spec 1 gives $\beta = -0.170$ ($p = 0.018$, $n = 2{,}046$ across 34 years). Spec 3 with the now-identifiable rumour phase returns $\gamma_{\text{rumour}} = +0.043$ (n.s.), $\gamma_{\text{construct}} = +0.002$ (n.s.), $\gamma_{\text{open}} = -0.127$, $\gamma_{\text{mature}} = -0.195$. The single-phase estimates are not individually significant given 282-cluster CR1 inference on 2,046 observations, but the directional story is internally coherent: a small positive rumour-phase capitalization that erodes through construction and turns negative after opening. The opening- and mature-phase point estimates are robust in sign to ring-restriction (-0.304 at open) and to ± 2-year phase-date perturbation ($-0.070, -0.223$). Two non-exclusive readings are plausible: (i) Cityringen treated zones (Frederiksberg, central Nørrebro, Indre By) were already at peak gentrification by the mid-2010s, so post-opening outer-CPH appreciation appears as relative softening; (ii) ejerlejlighed sales composition shifted in treated zones toward smaller / cheaper units as redevelopment substituted multifamily condos for the existing stock. We do not test (i) and (ii) directly in this draft. A defensible falsification would need a sales-weighted mean-sale-size series per postnummer to test the composition channel and a Frederiksberg-pre-1992 trend extrapolation to test the gentrification-saturation channel — neither is available in the current panel. The negative read also sits uneasily with Mulalic and Rouwendal (2022), who find positive price effects of the earlier Copenhagen metro expansions in a sorting framework; reconciling the condo-segment Cityringen result with their economy-wide estimate is an open item. We therefore treat the CPH negative result as an interesting auxiliary finding with plausible but currently untested explanations, not an outright reversal of the comparative-cities story.

Paris. With the 2014–2025 panel, Spec 1 is $+0.043$, $p < 10^{-7}$. Spec 3 gives $\gamma_{\text{construct}} = +0.011$, $\gamma_{\text{open}} = +0.062$ ($p = 0.075$), $\gamma_{\text{mature}} = +0.043$. The pattern is the muted analogue of Milano's: a near-zero construction phase, an opening-phase step, and a slight mature-phase fade. Under the ring restriction the coefficient returns $+0.061$ ($\approx$ baseline $+0.062$), which says the gpe14 effect is geographically well-defined and not absorbing halo-into-control contamination. Paris carries most of the pooled precision — it alone is 69% of the pooled observations — which the leave-one-out above makes explicit. Prior Paris-region evidence is suggestive rather than directly comparable: Mayer and Trevien (2017) estimate employment and sorting effects of the RER, not residential price capitalization, so the gpe14 estimates here are, to our knowledge, the first phase-decomposed price evidence for the Grand Paris Express.

Helsinki. Länsimetro is the added city with a full pre-opening baseline (the ASHI series starts 2009, before both opening cohorts). Its single-city Spec 3 is *flat to mildly negative*: $\gamma_{\text{construct}} = -0.054$ ($p = 0.005$), $\gamma_{\text{open}} = -0.047$ ($p = 0.06$), $\gamma_{\text{mature}} = -0.008$ (n.s.); only the small negative construction-phase coefficient is individually significant, and the post-opening phases are not (Spec 1 $+0.030$, $p = 0.11$). Taken at face value, the West Metro corridor into Espoo shows no

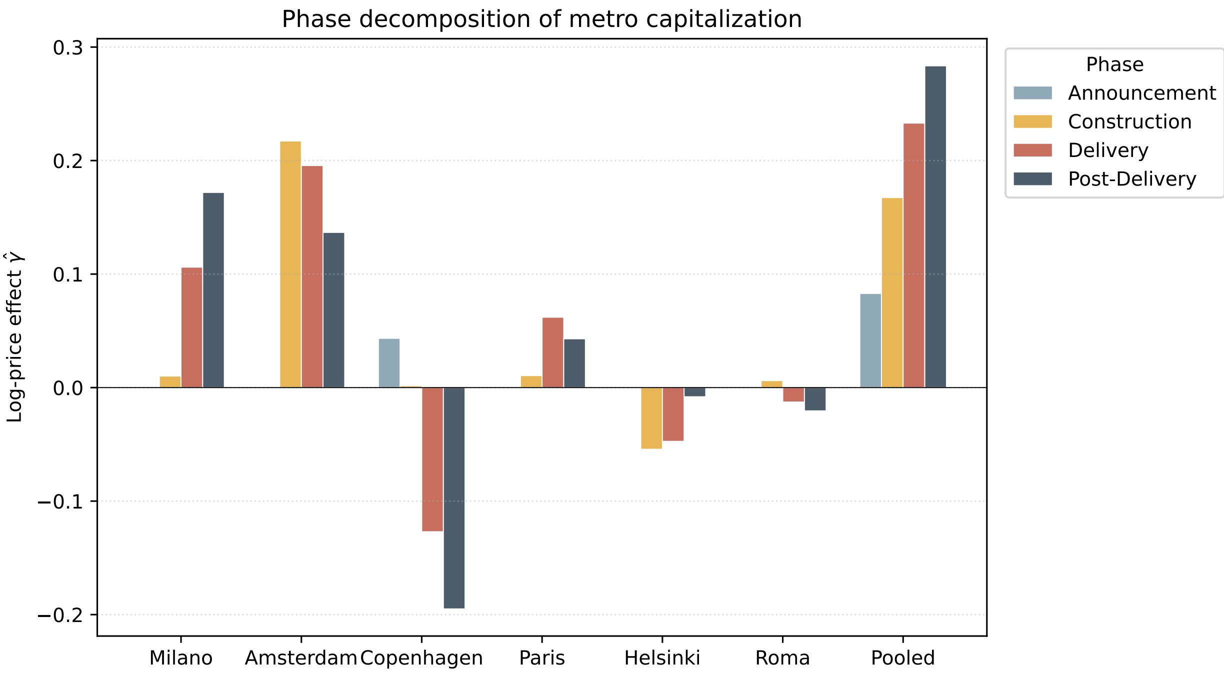

Figure 5: Per-city phase decomposition of metro capitalization. Bars are γ_ϕ from Spec 3 (eq. 5) shown for the four phases that are identified in any of our panels (*Rumours* omitted, as in Table 9). Milano shows the canonical rising shape; Amsterdam's bars are pseudo-levels whose identified content is the decline after opening (per-city reading below); Copenhagen shows a negative pattern explained below; Paris a muted but ordered ramp; Helsinki is flat; Roma is a flat within-city null. Rennes is omitted (pooled-only; no pre-opening baseline).

detectable price premium *within the panel window*. Four reasons are plausible: the Helsinki–Espoo post-code grid is coarse (116 units for the whole conurbation); Espoo is a low-density, car-oriented municipality where rail accessibility competes with ample parking and road capacity; the 2022 cohort has barely entered its maturity window within the sample; and — the reading we find most informative — Harjunen (2018) documents a $\sim 4\%$ anticipation premium near future Länsimetro stations capitalized *years before* opening, which our 2009-start panel largely absorbs into the baseline, mechanically muting any post-opening differential. A flat opening response after early anticipation is exactly what the forward-looking-market reading predicts. Helsinki therefore neither drives nor contradicts the pooled maturity step (the leave-one-out is unchanged when it is dropped); its contribution to the design is the extra country and the extra cluster. Rennes (Line B) is not shown here: with no pre-opening baseline it has no identified single-city decomposition, and enters the pooled specification only.

Roma. Metro C is priced through OMI, the same instrument as Milano, over 232 zones (22 treated: 12 in the 2014 cohort, 7 in 2015, 3 in 2018). Unlike Milano, the within-zone decomposition is a flat *null*: $\gamma_{\text{construct}} = +0.006$ ($p = 0.73$), $\gamma_{\text{open}} = -0.013$ ($p = 0.66$), $\gamma_{\text{mature}} = -0.020$ ($p = 0.59$), and the bare post-opening Spec 1 is even slightly negative (-0.024, $p = 0.34$). OMI detects no Metro C capitalization at any phase — consistent in sign with Daniele and Romito (2022), who estimate a statistically significant *negative* suburban effect (≈ -137 EUR/m^2, roughly -5% of the pre-treatment mean and concentrated in the upper price distribution) with a different method (synthetic control) and a different outcome source — the shared finding is the absence of positive capitalization, not a common point estimate. Several explanations are plausible and we do not discriminate between them: Metro C serves Roma's low-density eastern periphery rather than a core corridor; the line's credibility was weak through most of the window (under construction since 2007, repeatedly delayed, and without a connection to the rest of the metro network until the San Giovanni interchange opened in 2018); and the coarse OMI zones combined with the generous polygon-distance treatment rule (shared with Milano; §8.2) dilute whatever localised response exists. This is the most disciplining single-city result in the panel precisely because it is null: Roma is a large, coarse-zone appraisal city where a major new metro does *not* capitalize within-city, so it contributes to the pooled headline only through the common-year fixed effects — the channel we read as an upper bound (§7.3). We keep Roma in the pooled spec rather than dropping it — dropping a city after seeing its null would be selection on the outcome — and the leave-one-out above confirms the maturity step survives its deletion at $+9.4\%$. Why a flat within-city null still lifts the common-FE pooled coefficients is mechanical: Roma's price trajectory diverges from the other-cities-dominated common year effects, so the within-Roma deviations that a Roma-specific year FE would absorb instead load onto the pooled phase dummies; the city-by-year FE specification removes exactly this and collapses the pooled level to $+0.007$ ($p = 0.44$, §7.3).

8.6 Robustness

Table 11 reports γ_{open} under three perturbations: (a) ring-restriction (drop the 1500–3000 m annulus to widen the gap between near-treated and far-control), (b) -2-year phase shift, (c) $+2$-year phase shift.

The Milano and Paris coefficients survive both perturbation classes and keep their signs and ordering. Amsterdam's ring-restrict γ_{open} rises sharply (to a pseudo-level of $+0.349$; per the per-city reading above, Amsterdam phase levels are unidentified and only the direction of the change is informative), consistent with the spillover-into-control story already documented in §5. Copenhagen estimates remain negative across all perturbations. The "unidentified" cells for Amsterdam and Paris under the -2-year shift arise because shifting the rumour year backward by 2 years places it before the available data window's edge, leaving the rumour phase non-identified for those cohorts.

8.7 Auxiliary findings from the pooled spec

Three secondary results from Spec 2 (Spec 1 + controls) and Spec 3 with the same controls are worth documenting:

- The `lez_active` coefficient lands at roughly $+0.094$ (Spec 3) to $+0.097$ (Spec 2), $p < 10^{-10}$ in both pooled regressions. It was *negative* in the six-city panel and re-signs once Roma is added (all 232 Roma zones carry `lez_active` $= 0$, which re-partials the level dummy against the common year effects) — which is exactly why we read it as a city-level timing confound rather than a causal low-emission-zone (dis)amenity: LEZ activation overlaps with treatment timing in Milano (Area C, 2012) and Amsterdam (Milieuzone tightening, 2020), and the sign is not load-bearing. Dropping `lez_active` from the pooled controlled spec leaves the post coefficient all but unchanged — the term is partialled out of the city-by-year variation the unit + year FE already largely absorb. A clean causal estimate would need a within-city design with LEZ-boundary exposure that does not coincide with metro exposure.

- The `regen_active` coefficient sits at roughly $+0.08$ ($p < 0.05$), also sign-flipped from the six-city panel for the same re-partialling reason. Only Milano carries within-panel variation on this dummy (Roma's regeneration flag is zero throughout), so the coefficient is a Porta Nuova / regeneration-zone level effect leaking into the level term, and its sign is likewise not load-bearing.

- `dist_metro_existing_m_x_post` is positive and significant across specifications in the seven-city pool ($p < 10^{-14}$): zones *farther* from an existing metro station appreciate marginally more in the post-treatment window, consistent with a new line delivering its largest accessibility gain where prior rail access was poorest. This is a small effect we do not over-interpret — the coefficient is of order 10^{-4} per metre — but it overturns the six-city reading, where the same interaction was non-significant.

9 Limitations

We collect the limitations of the two-city core and the seven-city extension into a single list.

1. **Spatial granularity (Milano).** OMI zones are coarse. There are ∼43 of them for the whole comune, and a typical zone runs 1–3 km across; OMI's `Microzona` field is a census-tract pointer, not a finer polygon. The continuous-distance specification of equation (3) does what it can within the available data. Transaction-level identification would need the Agenzia delle Entrate's *Banca dati delle Compravendite immobiliari* — free for university researchers through the *Accesso ai dati per Studenti Universitari* programme, but gated behind a supervisor signature and not yet incorporated.

2. **TWFE bias and pre-trends (Milano).** The combined Milano panel TWFE is contaminated twice over: by staggered-treatment bias and by a $\sim +129$ EUR/m^2 pre-period violation at $k = -1$. We do not report it as a headline estimate. The Ring D headline survives a zone-level wild cluster bootstrap ($p = 0.004$; see Limitation 6), but its per-horizon CS Honest-DiD breakdown is uninformative at 16 zones. The 0–500 m band coefficient is significant asymptotically but *not* under the wild cluster bootstrap ($p = 0.10$), so we treat it as descriptive. The Amsterdam single-date treatment is not subject to staggered-treatment bias, and it shows Honest-DiD breakdown $\bar{M} > 2$.

3. **Outcome variable (Amsterdam).** WOZ values are municipality assessments, not transaction prices. The *Waarderingskamer* documents a one-year *waardepeildatum* lag — the WOZ value for year t is benchmarked against transaction evidence as of 1 January of year $t - 1$ (Waarderingskamer, 2024) — which mechanically smooths and rightward-shifts the price response we estimate. Read the headline +69.8k and the Amsterdam phase coefficients in that light: a sharper response at opening is plausibly being smeared into the post-delivery window.

4. **Time coverage (Amsterdam).** CBS data covers only 2013–2022. Pre-2002 data would test anticipation against the announcement date; the Woningwaarde dataset (2002–2019) is no longer publicly accessible.

5. **General equilibrium and spillovers.** The ring-restriction in Amsterdam and the within-ring design in Milano both control for spillover-into-control. Neither touches general-equilibrium effects on areas far from either treated or control buurten.

6. **Small-G cluster inference (Ring D).** The Ring D specification has 16 zones (8 treated M5, 8 control). Cluster-robust CR1 standard errors with only 8 treated clusters are known to over-reject

Table 11: Robustness of γ_{open} to specification choices: baseline Spec 3, ring-restriction (drop 1500–3000 m annulus), and ±2-year phase-date shifts.

City	Baseline γ_{open}	Ring-restrict γ_{open}	−2yr shift	+2yr shift
Milano	+0.1061	+0.0974	+0.0284	+0.1717
Amsterdam	n/i	n/i	unident.	+0.1494
Copenhagen	-0.1270	-0.3035	-0.0696	-0.2231
Paris	+0.0620	+0.0615	unident.	+0.0524
Helsinki	-0.0473	-0.0407	-0.0588	+0.0183
Roma	-0.0127	-0.0079	+0.0041	-0.0075

under heterogeneous variance (Cameron et al., 2008), and the per-horizon CS event study slices the same 16 zones further into per-cohort cells (Figure 1) — which is exactly why the per-horizon Honest-DiD breakdown $\bar{M}$ collapses to zero at most horizons. We meet this head-on with a restricted wild cluster bootstrap (Rademacher weights, 999 replicates, clustered on the 16 zones; `scripts/wild_bootstrap_milano.py`): the Ring D regen-controlled coefficient holds at $p = 0.004$ (95% CI $[+62, +280]$), so the +167 headline is robust to the small cluster count. The same bootstrap on the 0–500 m band coefficient (38 zones, few of them treated) returns $p = 0.10$ (95% CI $[-132, +1669]$): the band estimate is *not* significant once few-cluster inference is imposed, consistent with its asymptotic $p < 0.001$ being an over-rejection. We therefore lead with the bootstrap-robust Ring D estimate and read the band gradient as descriptive, not causal.

7. **Copenhagen Boliga ejerlejlighed selection.** Boliga sales are condominium-heavy and skew toward properties with frequent turnover. Postnumre with <3 sales/year are dropped (213 of 495), which over-represents inner-city, high-churn neighborhoods. A future iteration should ingest the official OIS / Tinglysningsretten deed registry for the full transaction universe at parcel level.

8. **Copenhagen near-zero control sample.** With 278 of 282 postnumre within 1500 m of M3 or M4, the binary contrast is identified off a thin control set. The continuous-distance specification is the more identification-robust read.

9. **Paris GTFS and amenity controls.** PRIM credentials are pending, so tram / rail / cycle distance variables are infinite and dropped; Overture amenity data is set to uniform 1.0 because the S3 release path was unavailable to our local DuckDB build. The consequence travels: Paris supplies 69% of the pooled observations, so the control-stability axis of §8.4 is partly mechanical for the dominant panel — stability across *available* controls, with the Paris controls thin.

10. **Cross-city SUTVA and policy collinearity.** The pooled (c, i) panel implicitly assumes that the price response in one city's treated units is independent of treatment in the other six. On the metro dimension itself this is benign — the seven cities are geographically isolated. But the LEZ control `lez_active` is collinear with treatment timing within Milano (Area C, 2012) and within Amsterdam (Milieuzone tightening, 2020), so its coefficient is a city-by-year level effect, not a clean LEZ price effect of either sign — its sign in fact flips between the six- and seven-city pools (§8).

11. **Cross-city outcome harmonisation.** The seven panels use five distinct outcome instruments (OMI zone appraisal for Milano and Roma, CBS-WOZ assessment, Boliga deed median, geo-DVF deed mean for Paris and Rennes, and Statistics Finland ASHI transaction prices for Helsinki) with different valuation cycles, censoring, and revaluation lags. Unit fixed effects absorb level differences but not city-specific time-aggregation or cyclical drift. Figure 6 plots the mean residual log-price per city per year after partialling out (c, i) entity and year FE: it ranges from just 0.03 log units (Paris) and 0.08 (Rennes) up to the longer panels (Amsterdam 0.29, Helsinki 0.32, Roma 0.43, Milano 0.48, Copenhagen 0.72). The pooled γ_ϕ should therefore be read as a within-unit log-differential, under the assumption that this residual city-by-year drift is uncorrelated with the metro phase indicators within each city's identification window — an assumption Roma stresses, since its large residual drift is exactly what lifts the common-year-FE phase coefficients (§8.4). Both appraisal instruments (OMI, WOZ) mechanically smooth and lag price responses, which by itself can manufacture a delayed-to-maturity profile. We test this directly by estimating the Spec 3 maturity step on the deed/transaction-price cities alone (Copenhagen + Paris + Helsinki + Rennes, n = 33,628; `scripts/wild_cluster_phase_subsets.py`): the deed-only step is +8.8 log points ($\approx +9.2\%$), sign-consistent with the pooled step and asymptotically significant ($p = 0.008$), but *not* significant under the city-level wild bootstrap ($p = 0.14$ at $G = 4$, where the bootstrap has little to work with). So the timing pattern is not solely an appraisal-cycle artefact — the deed cities reproduce its sign and rough magnitude — but the appraisal-free evidence cannot independently establish it at conventional levels.

12. **Small-G inference in the pooled spec (the**

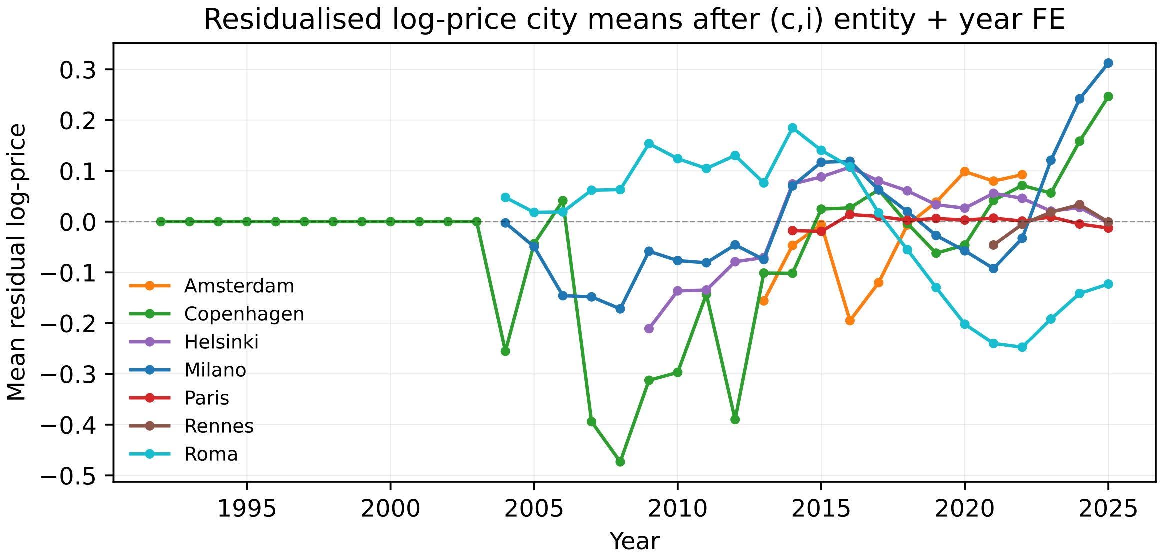

Figure 6: Mean residual log-price by city and year after partialling out the pooled regression's (c, i) entity and year fixed effects. Non-zero residual means indicate city-specific time variation that the FE structure does not absorb. Paris is much flatter (≈ 0.03 log-unit range) than the longer panels (0.29–0.72 for Amsterdam, Helsinki, Roma, Milano, and Copenhagen).

binding constraint). Treatment is assigned at the city×line level, so with seven cities the entity-clustered CR1 standard errors on the pooled spec understate true uncertainty; every entity-clustered z and the joint Wald $p = 2 \times 10^{-12}$ should be read as ordering evidence, not formal inference. The few-cluster bootstrap results, their partition dependence, and the opposite movements of the city- and cohort-clustered p-values when Roma was added are reported in §8.4; the city-subset sensitivities there ($G = 6$ without Rennes, Webb $p = 0.038$; $G = 5$ without Rennes and Roma, $p = 0.051$) show the city-level significance to be marginal once G shrinks — noting that Rennes lacks identification while Roma is identified-but-null, so the $G = 5$ subset is an outcome-based sensitivity rather than an identification screen. Three honest caveats attach. First, the few-cluster verdict is control-set-dependent: the repaired full-control bootstrap (the original branch flipped sign because a collinear amenity×post column was over-projected; §8.4, axis iv) gives Webb $p = 0.080$ at the city partition, against $p = 0.036$ for the level-only control set — so the full-control magnitude we quote does not itself clear 5%. Second, the pooled *level* (Spec 1) is separately an upper bound that collapses to $+0.007$ (n.s., $p = 0.44$) under city-by-year FE (§7.3) — a sharper collapse than at six cities, precisely because the seventh city is a within-city null. Third, the maturity *step* inherits the same FE-structure fragility: under city-by-year FE it shrinks to an insignificant -0.5 log points, and the within-city identified object becomes a $+2.5$ log-point step at opening — itself only entity-clustered ordering evidence, with city-clustered Webb $p = 0.52$ (§8.4, axis ii).

13. **The added cities are thin, coarse, or null.** Helsinki, Rennes, and Roma each strengthen the design more than the point estimate, and each is limited as a single-city case. Helsinki's postcode grid is coarse (116 units for the whole Helsinki–Espoo conurbation) and its 2022 cohort barely reaches maturity in-sample; Rennes has *no* pre-opening baseline (geo-DVF exposes only 2021 onward) and so enters the pooled spec only; and Roma, though it carries a full OMI baseline, is a within-city null whose pooled contribution runs entirely through the common-year fixed effects (§8.4). The leave-one-out (§8.4) shows the pooled step is positive under every single-city deletion, spanning +7.3% to +13.7% — including drop-Paris at +9.9%, the largest panel. So, unlike in the six-city draft, the result no longer rests on Paris for its magnitude. Read honestly: the maturity *pattern* is multi-country and robust to leave-one-out, but its formal few-cluster significance holds only at the city-cluster level, and it is a cross-city-average object rather than a within-city one (Limitation 12).

14. **Route-selection endogeneity.** The DiD framework treats the M5 / M4 / NZL / M3 / GPE alignments as exogenous to property-price dynamics. They are not. Line routings are politically and technically chosen, and corridors with expected regeneration potential are more likely to be selected in the first place. Our estimates are therefore best read as reduced-form price responses that bundle capitalization with route-selection, not as pure capitalization effects. The Milano regeneration covariate (`regen_active`) is an imperfect proxy for this. A clean structural estimate would need either an instrument for line placement or a quasi-experiment where the route was set before development potential was known — and within the European cases studied here we are not aware of such an instrument.

15. **Staggered-TWFE heterogeneity in the pooled phase spec.** The pooled Spec 3 is a plain TWFE over 17 staggered treated cohorts with binned endpoints — the setting in which heterogeneous treatment effects produce negative-weight contamination (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021), and the same concern we used to demote Milano's combined-panel +820 (§4.1). We have not re-estimated the pooled phase contrasts with a heterogeneity-robust estimator (stacked Callaway–Sant'Anna aggregation, Sun–Abraham interaction weighting, or Borusyak imputation); until that is done (future-research item 2), the pooled phase coefficients carry the standard staggered-TWFE caveat on top of the FE-structure caveat of Limitation 12.

16. **Mechanism for the step at maturity.** The pooled Wald test (§8) rejects equality of the four phase coefficients, with the gap concentrated between Post-Delivery and the earlier phases; the apparent seven-city early ramp is a Roma common-year-FE artefact (§8), and the robust object is a step at maturity in the *cross-city average* — within-city, the identified pattern is a smaller step at opening (§8.4, axis ii), with Milano the one city showing maturity growth within-city. Several mechanisms fit the cross-city pattern — slow neighbourhood-amenity colonisation, information frictions on line reliability, a Rosen-style supply-elasticity-driven re-equilibration — and we do not discriminate between them here. The policy implications differ across mechanisms (value-capture timing changes if the price response is amenity-driven rather than accessibility-driven), so this is a non-trivial open question.

17. **Owner-side outcomes only.** Every outcome in the panel is an owner-side price or assessment — transaction deeds, condominium sales (Copenhagen is condo-only), appraisal values, and WOZ assessments — so "capitalization" throughout means a wealth transfer to incumbent property owners. Renter incidence, rent pass-through, and displacement are unmeasured, and the delayed-to-maturity timing, if amenity-driven, coincides with precisely the window in which neighbourhood change displaces incumbent renters. The spatial-equity motivation raised in the introduction is therefore addressed only on the owner side; the renter side is an open measurement problem in every city in this panel.

Future research directions

The limitations above and the robustness gaps documented in the body point to six concrete next steps, ordered by expected payoff:

1. **More cities, and reconciling the full-control bootstrap.** The city-level wild cluster bootstrap on the pooled maturity step reaches $p = 0.036$ at seven cities (§8.4), but the cohort-level bootstrap sits at $p = 0.164$ and in fact *loosened* when the per-city-null Roma was added. The frontier is therefore adding further European cohorts — selected on data availability before their outcomes are observed, the same rule that kept Roma in — so that both cluster partitions gain G honestly. (The full-control bootstrap branch, listed here as unreconciled in an earlier draft, is now repaired: dropping the rank-deficient amenity×post column restores the step to +0.116 with Webb $p = 0.080$; §8.4, axis iv.)

2. **Alternative DiD estimators.** Re-run Spec 1 and Spec 3 with the Borusyak et al. (2024) imputation estimator and the Sun–Abraham (2021) interaction-weighted event study, and confirm that the qualitative conclusions hold across estimators.

3. **Demographic heterogeneity.** Interact the post-opening indicator with pre-treatment median income (from CBS, INSEE, Statistics Denmark, ISTAT) to test whether the metro premium is captured disproportionately by higher- or lower-income neighbourhoods — the central question for the spatial-equity debate raised in the introduction.

4. **Spatial-DiD specification.** Apply Moran's I to the residuals and re-estimate under a spatial-lag or spatial-error model in the LeSage–Pace framework to formally test the no-spatial-autocorrelation assumption that the ring-restriction implicitly makes.

5. **Mechanism discrimination.** Combine the seven-city panel with amenity-density data (Overture / OpenStreetMap café, retail, school counts) to test whether the step-at-maturity pattern correlates with amenity formation rather than accessibility per se.

6. **Milan+Paris parcel-level follow-up.** A separate companion paper would restrict to the two cities with the longest credibly-identifying panels — Milan (M5 Lilla, 10+ post-opening years) and Paris (GPE Ligne 14 Nord, 6 pre and 5+ post) — pull parcel-level transaction microdata (*Banca dati delle Compravendite immobiliari* for Italy, native geo-DVF deeds for France), and replace the four-phase decomposition with a year-by-year event study at parcel resolution. The question shifts from *when* capitalization occurs to *where-in-space* and at what year-by-year horizon, and the deed-level outcome variable replaces the OMI appraisal smoothing concern raised by the measurement-class split (§7.3).

10 Conclusion

This paper asks *when* new metro infrastructure capitalizes into property prices. The evidence base is seven European cities across five countries — Milano (M5 Lilla 2013–2015 and M4 Blu 2022–2024), Amsterdam (Noord-Zuid Lijn 2018), Copenhagen (Cityringen 2019 and the M4 Sydhavn 2024), Paris (Grand Paris Express Ligne 14 Nord 2020 and Sud 2024), Helsinki (Länsimetro 2017 and 2022), Rennes (Line B 2022), and Roma (Metro C 2014, 2015, and 2018) — with the response decomposed across announcement, construction, opening, and post-delivery phases. The estimator stack is deliberately layered: a two-way fixed-effects DiD with unit-clustered standard errors throughout, complemented by Callaway–Sant'Anna staggered-treatment estimates, Honest-DiD pre-trend sensitivity, a leave-one-city-out, and the restricted wild cluster bootstrap of Cameron et al. (2008).

We start with the two cities studied in greatest depth. The defensible point estimate is the within-ring Milano M5 effect of +167 EUR/m^2 (regeneration-controlled, wild-bootstrap $p = 0.004$) — roughly +5.6% of the control mean. The Amsterdam Noord-Zuid Lijn ring-restricted estimate of +69,800 EUR per property is a *level* object, not proportional capitalization: the identical panel in logs gives −4.3%, the level gain reflecting the 1.74× higher treated pre-period base (`amsterdam_log_specs.json`), and we no longer quote it as ≈ +21%. The Milano +769 EUR/m^2 premium within 500 m of a new station, decaying smoothly to a null by 1–1.5 km, is descriptive corroboration of distance decay rather than a bootstrap-robust estimate ($p = 0.10$ under few-cluster inference). The headline-grabbing numbers — the original combined-panel Milano +820 EUR/m^2 and the full-sample Amsterdam +31,900 — both inflate or deflate around these defensible figures, and the mechanism is now understood in each case: clustering blows up the Milano standard error by a factor of five, the Callaway–Sant'Anna overall bootstrap CI on the combined panel includes zero ($[-332, +1834]$), and the Amsterdam full-sample estimate is pulled down by partial-treatment buurten in the 1.5–3 km donut.

The seven-city pooled analysis (§8) is the paper's central result, and we state it with its identification structure attached. Spec 3 phase coefficients are +8.3, +16.7, +23.3, and +28.3 *log points* at the announcement, construction, opening, and post-delivery phases (n = 42,004), and a joint Wald test rejects equality at $p = 2 \times 10^{-12}$ — entity-clustered, so ordering evidence rather than formal inference (Limitation 12). Taken at face value every adjacent phase pair is significantly increasing. But that early-phase ramp is manufactured by Roma — a within-city null that inflates the common-year-FE construct and opening coefficients — and dropping Roma restores the flat premature plateau (§8); the Roma-independent feature is a step at the post-delivery boundary. The headline finding is therefore that **metro capitalization is delayed to maturity in the pooled cross-city average**: a construction-to-maturity step of +9 to +12% (drop-Roma to full-sample) that appears only two or more years after opening, control-stable across the confound ladder and positive under every leave-one-city-out — it survives the deletion of Milano, of the large Paris panel, and of Roma. What the step is *not* is a within-city fact. City-by-year fixed effects shrink it to an insignificant −0.5 log points and leave a +2.5 log-point step at opening (entity-clustered ordering evidence; city-clustered Webb $p = 0.52$); only Milano shows maturity growth within-city; pre-trend robustness is weak ($\bar{M} \approx 0.7$ under flat slack, ≈ 0.1 under the accumulating bound, and the maturity-horizon level sits below a linear pre-trend extrapolation); and few-cluster significance is partition- and control-set-dependent — $p = 0.036$ on the seven cities with level-only controls, $p = 0.080$ with the full control set, $p = 0.164$ on the twenty-four cohorts, $p = 0.051$ on the five-city subset that drops Rennes (no identified contrast) and Roma (identified but null). We do *not* lead with the pooled Spec 1 *level* (+0.1195 log points): under common year FE it is an upper bound, and city-by-year FE shrink it to +0.007 ($p = 0.44$; §7.3). The defensible *magnitude* statements are the per-city estimates — foremost the within-ring Milano +167 EUR/m^2, which survives a zone-level wild cluster bootstrap ($p = 0.004$). The distinctive cross-city contribution is the *timing* pattern — capitalization concentrated at maturity in the pooled average — offered as the hypothesis the next cohort of European panels should test, not as a settled effect.

Single-city patterns differ in informative ways. Milano shows the canonical rising shape; Amsterdam's identified phase contrasts show a relative decline after opening — partial mean reversion from a construction-era premium whose level the 2013–2022 window cannot identify; Paris is a muted analogue of Milano; Copenhagen returns a sign-stable negative coefficient that we read as auxiliary evidence about already-gentrified inner postnumre rather than a true metro disamenity; Helsinki's low-density Espoo corridor is flat, consistent with the pre-window anticipation premium Harjunen (2018) documents; and Roma is an outright within-city null — corroborating Daniele and Romito (2022) — that we retain as a stress test rather than drop. None of these individually carries the pooled maturity step — which is precisely why the pooled, leave-one-out-robust result matters — but each adds a cluster, and for the cities outside Italy and France a country, to the inference. The new data sources earned their place: a Boliga scrape for Copenhagen extending to January 1992, the OpenDataArchives DVF mirror for Paris covering 2014–2020, the Statistics Finland ASHI postcode series for Helsinki, and a GTFS-located, two-vintage OMI panel for Roma's Metro C across the 2014 zone re-delimitation — each one recovers the pre-treatment identifying variation that the public-portal versions of these data products lacked.

Two lessons travel across the cities. First, the levels are bracketed, not settled: the Milano +5.6% Ring D figure sits at the lower end of the meta-analytic range; the Amsterdam ring-restricted gain is positive only in levels (in logs the same panel gives −4.3%, the 1.74× treated base doing the reconciling); and the pooled binary +0.12 log points is an upper bound that the conservative city-by-year-FE reading shrinks to +0.007 (not significant) — so we report no single pooled magnitude as *the* effect. Second, the right-hand-side choices that drive published headline numbers — whether to cluster (and at what level), whether to restrict to a clean ring, whether to decompose phases, which fixed-effects structure to impose, which DiD estimator to use — matter as much for inference as for point estimation. Report only a bare-TWFE post-opening coefficient and you both overstate significance and bury the timing structure: in the pooled cross-city average, capitalization is not a single jump at opening but a delayed step that arrives only once the line has matured, while the within-city evidence locates a smaller response at opening. That timing carries a value-capture reading (Medda, 2012), with the instrument distinction doing the work: recurring instruments (annual land-value or property-tax increments) collect whenever the uplift arrives, but one-shot instruments — betterment levies, developer charges, land sales — are calibrated to a date, and where the delayed pattern holds, calibrating them to the opening date, or worse to the announcement, would systematically under-collect. The within-city evidence cuts the timing advice the other way — a +2.5 log-point step at opening followed by partial reversion would favour collecting *at* opening — so the recommendation that is robust to either read is instrument choice, not levy timing: recurring instruments are right under both patterns, and a city betting on a one-shot instrument is betting on which of the two readings applies to it. The per-city nulls cut deeper still: Roma, Helsinki, and Copenhagen are the downside scenarios an uplift-financed business case should price in.

Acknowledgments

The author thanks the *Agenzia delle Entrate* for maintaining the OMI dataset, Statistics Netherlands (CBS) for the *Kerncijfers Wijken en Buurten* series and PDOK for the buurt geometry feeds, and the maintainers of the OpenDataArchives `etalab-dvf` mirror for preserving the 2014–2020 geo-DVF vintage after Etalab pruned the `latest/` branch. Copenhagen sales records were collected from the publicly served `boliga.dk` sales API at a throttled request rate; the data are public listings of completed transactions, only postnummer-year aggregates are stored, and the acquisition method is documented in §8.2 so that its terms-of-use position can be assessed openly — a future iteration should substitute the official OIS / Tinglysningsretten deed registry (Limitation 7). The doubly-robust Callaway–Sant'Anna implementation used here is the python `differences` package; pre-trend sensitivity computations follow Rambachan and Roth (2023). Any remaining errors are the author's own.

Data and code availability. All data sources are publicly available. Milano OMI data from *Agenzia delle Entrate* (`wwwt.agenziaentrate.gov.it/geopoi_omi/`). Amsterdam CBS WOZ data from `opendata.cbs.nl`. Buurt geometries from PDOK. Copenhagen sales via the `api.boliga.dk` scraper documented in §8.2. Paris geo-DVF via the OpenDataArchives mirror documented in the same section. Code (all estimators reported in this paper) at `scripts/` in the repository; the v3 snapshot used to generate the numbers in this paper is to be tagged `v3.0` at release and archived via Zenodo before final submission.

Conflicts of interest. The author declares no competing interests. No external funding was received for this work.

AI use statement. Claude (Anthropic) was used for code assistance, numerical cross-checking of reported results against the analysis artifacts, and editing. All scientific content, analysis design, results, and conclusions are the author's own.

References

C. A. Agostini and G. A. Palmucci. The anticipated capitalisation effect of a new metro line on housing prices. *Fiscal Studies*, 29(2):233–256, 2008.

J. Atack, F. Bateman, M. Haines, and R. A. Margo. Did railroads induce or follow economic growth? Urbanization and population growth in the American Midwest, 1850–1860. *Social Science History*, 34(2):171–197, 2010.

M. Bertrand, E. Duflo, and S. Mullainathan. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275, 2004.

V. Bajic. The effects of a new subway line on housing prices in metropolitan Toronto. *Urban Studies*, 20(2):147–158, 1983.

S. B. Billings. Estimating the value of a new transit option. *Regional Science and Urban Economics*, 41(6):525–536, 2011.

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting event-study designs: Robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285, 2024.

D. R. Bowes and K. R. Ihlanfeldt. Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics*, 50(1):1–25, 2001.

B. Callaway and P. H. C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.

A. C. Cameron, J. B. Gelbach, and D. L. Miller. Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427, 2008.

F. Boscacci, R. Camagni, A. Caragliu, I. Maltese, and I. Mariotti. Collective benefits of an urban transformation: Restoring the Navigli in Milan. *Cities*, 58:14–22, 2016.

R. Cervero and C. D. Kang. Bus rapid transit impacts on land uses and land values in Seoul, Korea. *Transport Policy*, 18(1):102–116, 2011.

C. de Chaisemartin and X. D'Haultfœuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996, 2020.

C. de Chaisemartin and X. D'Haultfœuille. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *Econometrics Journal*, 26(3):C1–C30, 2023.

G. Debrezion, E. Pels, and P. Rietveld. The impact of railway stations on residential and commercial property value: A meta-analysis. *Journal of Real Estate Finance and Economics*, 35(2):161–180, 2007.

N. Devaux, J. Dubé, and P. Apparicio. Anticipation and post-construction impact of a metro extension on residential values: The case of Laval (Canada), 1995–2013. *Journal of Transport Geography*, 62:8–19, 2017.

M. Diao, D. Leonard, and T. F. Sing. Spatial-difference-in-differences models for impact of new mass rapid transit line on private housing values. *Regional Science and Urban Economics*, 67:64–77, 2017.

A. Grimes and C. Young. Spatial effects of urban rail upgrades. *Journal of Transport Geography*, 30:1–6, 2013.

F. Daniele and E. Romito. The impact of "Metro C" in Rome on the housing market. Temi di discussione (working paper) 1394, Banca d'Italia, 2022.

M. Gallo. The impact of urban transit systems on property values: A model and some evidences from the city of Naples. *Journal of Advanced Transportation*, 2018:1767149, 2018.

F. Guglielmi, T. Tabrizi, F. De Fabiis, and P. Coppola. Modelling real-estate values around railway stations: Insights from an Italian case. *Sustainability*, 18(5):2304, 2026.

S. Gibbons and S. Machin. Valuing rail access using transport innovations. *Journal of Urban Economics*, 57(1):148–169, 2005.

A. Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021.

A. Gupta, S. Van Nieuwerburgh, and C. Kontokosta. Take the Q train: Value capture of public infrastructure projects. *Journal of Urban Economics*, 129:103422, 2022.

O. Harjunen. Metro investment and the housing market anticipation effect. Working Papers 2018:2, City of Helsinki, Executive Office, Helsinki, 2018.

C. D. Higgins and P. S. Kanaroglou. Forty years of modelling rapid transit's land value uplift in North America: Moving beyond the tip of the iceberg. *Transport Reviews*, 36(5):610–634, 2016.

H. A. Klaiber and V. K. Smith. Quasi experiments, hedonic models, and estimating trade-offs for local amenities. *Land Economics*, 89(3):413–431, 2013.

J. G. MacKinnon and M. D. Webb. Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2):233–254, 2017.

S. Malpezzi. Hedonic pricing models: A selective and applied review. In T. O'Sullivan and K. Gibb, editors, *Housing Economics and Public Policy*, pages 67–89. Blackwell, Oxford, 2003.

T. Mayer and C. Trevien. The impact of urban public transportation evidence from the Paris region. *Journal of Urban Economics*, 102:1–21, 2017.

D. P. McMillen and J. McDonald. Reaction of house prices to a new rapid transit line: Chicago's Midway line, 1983–1999. *Real Estate Economics*, 32(3):463–486, 2004.

F. Medda. Land value capture finance for transport accessibility: a review. *Journal of Transport Geography*, 25:154–161, 2012.

S. I. Mohammad, D. J. Graham, P. C. Melo, and R. J. Anderson. A meta-analysis of the impact of rail projects on land and property values. *Transportation Research Part A: Policy and Practice*, 50:158–170, 2013.

I. Mulalic and J. Rouwendal. Public transport investments, commuting and gentrification: evidence from Copenhagen. Tinbergen Institute Discussion Paper 2022-035/VIII, Tinbergen Institute, 2022.

F. Pagliara and E. Papa. Urban rail systems investments: An analysis of the impacts on property values and residents' location. *Journal of Transport Geography*, 19(2):200–211, 2011.

A. Rambachan and J. Roth. A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591, 2023.

S. Rosen. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55, 1974.

J. Roth, P. H. C. Sant'Anna, A. Bilinski, and J. Poe. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244, 2023.

P. H. C. Sant'Anna and J. Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122, 2020.

L. Sun and S. Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021.

Waarderingskamer. *Waardepeildatum WOZ-beschikking*. Reference-date documentation for the Dutch WOZ property valuation, 2024. `www.waarderingskamer.nl`.

C. Y. Yiu and S. K. Wong. The effects of expected transport improvements on housing prices. *Urban Studies*, 42(1):113–125, 2005.