

SSUS WORKING PAPER SERIES

Space Is Not a Control Group: Quasi-Experimental Design in the Presence of Spatial Dependence

Elijah Knaap[1]

[1]University of California, Irvine, Urban Planning and Public Policy

Spatial Science and Urban Systems · UC Irvine 2026-08-01

ABSTRACT Difference-in-differences (DID) and fixed-effects (FE) designs are numerically identical for an unstaggered treatment in a balanced two-period panel, so applied work treats the choice between them as immaterial. The equivalence is an artifact of the aspatial case, and it conceals two problems with treatments defined by location. The first afflicts both designs. Because the spatial multiplier is a function of location, spatial variation in the period shock produces a differential trend correlated with treatment by construction. I give conditions sufficient for parallel trends to survive and show they are jointly untestable at two periods, so an explicit spatial model is required, whichever design a paper chose. The second separates them. The spatial parameter is identified from the residual variation that distinguishes unit-level from group-level conditioning, and the treatment effect is a function of that parameter. In a Monte Carlo the between-group design understates the effect by up to 68 percent, and where the outcome carries no spatial dependence at all it still reports a spatial parameter of 0.735, manufacturing dependence from the heterogeneity its own conditioning failed to remove. On Montgomery County rents and the Purple Line announcement it returns the larger parameter in all 124 comparisons, an ordering that holds conditionally and reverses where the omitted heterogeneity is spatially unstructured. Units sharing a building also share a coordinate, so co-location accounts for most apparent unit-level dependence. The announcement moved rents little, and that estimate is recoverable only from the fixed-effects design.

KEYWORDS spatial econometrics; hedonic model; causal inference; transit; land-value capture

1 Introduction

Writing in 2007, Getis (2007) observed that spatial autocorrelation was “not mentioned at all in 12 currently used econometrics textbooks,” and drew the consequence that “all economists being educated today must learn about dealing with spatial models and data outside of their

curricula." The diagnosis was already two decades old: Anselin & Griffith (1988) had shown that ignoring spatial dependence "may lead to serious errors in the interpretation of familiar regression diagnostics and misspecification tests," and located the obstacle in dissemination, not in theory. The methods existed; applied practice had not absorbed them.

The observation still holds. None of Angrist & Pischke (2008), Cunningham (2018) or Huntington-Klein (2022) devotes a paragraph to spatial effects, and all three supply the designs on which applied policy analysis runs.

That analysis relies overwhelmingly on two designs: difference-in-differences (DID), which draws its counterfactual from a comparison group, and fixed effects (FE), which draws it from each unit's own history. For an unstaggered treatment in a two-period panel these designs are provably equivalent (Wooldridge, 2021), and applied work accordingly treats the choice between them as a matter of presentation. Two consequences follow when the treatment is defined by location.

The first concerns identification, and it applies to both designs alike. When the outcome is spatially dependent, the untreated potential outcome carries a spatial multiplier that is itself a function of location. I derive the conditions under which parallel trends survives such a treatment: a single spatial regime, a normalization equalizing the multiplier's row sums, and a spatially constant period shock. Failure of any one produces a differential untreated trend correlated with treatment by construction. With two periods none of the three is testable. Because the two designs are numerically identical in the aspatial case, this bias is common to both: it condemns the aspatial estimator, whichever design produced it. An explicit spatial model is required.

The second concerns the estimator, and it separates the designs. The equivalence that makes the first bias common turns out to be an artifact of the aspatial case. Once a spatial model is fitted, the spatial parameter is identified from precisely the residual variation that distinguishes unit-level from group-level conditioning, and the treatment effect is a function of that parameter. The divergence is systematic and one-directional in these data: in every one of 124 paired comparisons, the DID design returns the larger spatial parameter. At the unit level it settles against the upper boundary of the stationary region, where the process is nearly a spatial unit root and regression against it is spurious (Müller & Watson, 2024). The design most closely associated with credible causal identification is the one that drives the estimated process there. The within-unit estimates are smaller and better behaved, and no safer for an aspatial analysis: Anselin & Griffith (1988) showed that spatial effects matter at ordinary magnitudes, and applied work has spent four decades ignoring that warning, not the one about unit roots.

The designs largely agree about where the effect lies; they differ in whether either can determine it. Because a treatment defined by location makes a unit's own exposure nearly collinear with its neighbors', the direct and spillover coefficients are not separately identified and only their sum is recoverable. The within-unit design recovers it; the between-group design returns a quantity that changes sign with the bandwidth.

The stakes are concrete. A growing literature isolates the causal effects of transit investment on property values, and the question is distributional: rising land values may displace residents. Measures such as land value capture, affordable housing preservation and community land trusts can relieve that pressure if enacted early. Montgomery County is weighing rent control measures sized against the anticipated land-value increment, so the magnitude carries equity implications directly. Since cumulative impacts propagate through $(I - \rho W)^{-1}$, which is convex in ρ and diverges as $\rho \to 1$, a gap in the spatial parameter of the size reported here implies land-value increments an order of magnitude apart.

The construction of the Purple Line, a light rail extension to the DC Metro system in suburban Maryland, supplies the case. The treatment is defined purely by location, the panel is balanced with a single unstaggered activation, and the two designs are therefore exactly equivalent absent spatial dependence, so any divergence between them is attributable to spatial structure alone. A policy-enforced residential rent survey collects unit-level price and location data annually. Rental units are "transacted" far more frequently than the repeat-sales data of the hedonic literature, and that density is what makes the within-unit arm of the comparison available at all. Figure 1 places the line and its stations within the regional rail network.

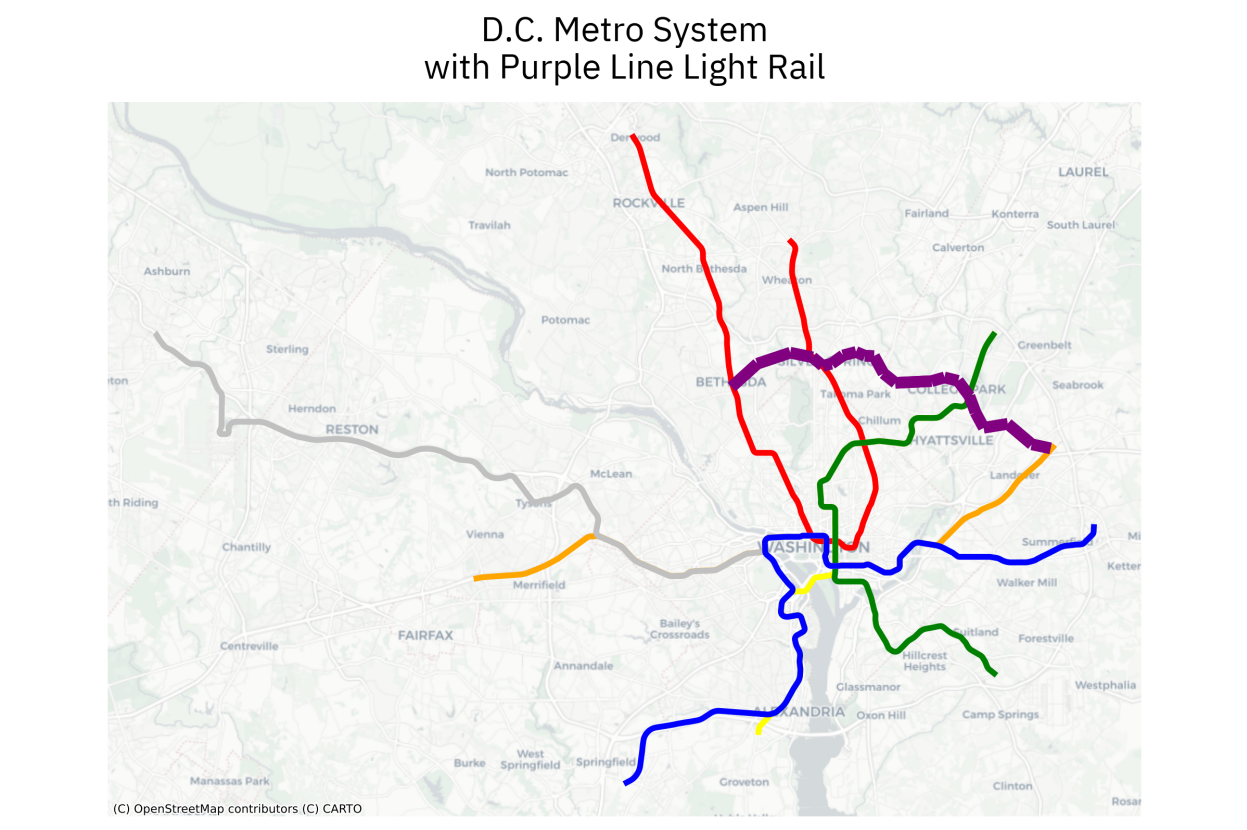


Figure 1: D.C. Metro System with the Purple Line

Misspecification is costly on two fronts. Omitting spatial effects biases coefficients and inferences even under quasi-experimental designs, and can attribute spatial dependence to the treatment (Bardaka et al., 2018; Delgado & Florax, 2015; Diao et al., 2017; Dubé et al., 2014). Clustering standard errors geographically does not address this (Wolf et al., 2021): under strong dependence both clustered and spatial HAC standard errors fail, and produce spurious significance, not merely imprecise inference (Müller & Watson, 2024). Geographic clustering corrects the wrong object, adjusting the variance of an estimator whose identifying variation is already contaminated.

Spillovers also mean the causal impact decomposes into direct and indirect components whose sum, the *cumulative* effect, is what policy depends on (Anselin, 2003; LeSage & Pace, 2014a; LeSage & Pace, 2009). Price increases reverberate beyond the corridor, so a design ignoring spatial interaction can misstate the effect in either direction: inflating it where dependence is mistaken for treatment (Bardaka et al., 2018; Diao et al., 2017), or understating

it where the cumulative effect goes unreported. Indeed, Dubé et al. (2018) show cumulative transit effects exceeding those from an aspatial model. Either error bears directly on *where* and *by how much* to provide anti-displacement subsidies.

The next section develops both results: the parallel-trends derivation, and the sense in which the FE and DID designs cease to be equivalent once the data are spatially dependent. I then review the literature merging quasi-experimental designs with spatial specifications, define the spatial panel models, report results, and close on the reliance of these models on theory.

2 Classic Quasi-Experimental Design

Traditional regression approaches fail when the process under study is spatially dependent (Anselin, 1989; Anselin et al., 1996; Anselin, 2002). The very term *spatial dependence* says that observations are no longer independent of one another; the Stable Unit Treatment Value Assumption (SUTVA) is therefore violated if the dependent variable is spatially autoregressive, or if treatment spills over onto nearby units (Delgado & Florax, 2015; Imbens & Rubin, 2015). A great deal of contemporary research nonetheless applies classic quasi-experimental methods where a treatment is defined in space, where the dependent or treatment variable plausibly follows an autoregressive process, or both. Akbari et al. (2023) and Kolak & Anselin (2020) survey these issues more generally.

2.1 The Two Estimators

The between-group estimator identifies the effect from variation between treated and untreated groups. Its critical assumption is that the groups would have moved in parallel absent treatment:

$$E[Y_0(1) - Y_0(0)|D = 1] = E[Y_0(1) - Y_0(0)|D = 0] \tag{1}$$

Under Equation 1 any differential change between the groups is attributable to the treatment, and the two-period, two-group specification is

$$Y_{it} = \alpha + \gamma D_i + \tau T + \delta DT + \epsilon_{it} \tag{2}$$

with D a binary treatment indicator, γ the pre-treatment difference between groups, T a period indicator, and δ the causal effect. Goodman-Bacon calls this the "two-by-two" design (Goodman-Bacon, 2021).

The within-unit estimator identifies the effect from variation within units over time, holding every time-invariant unit characteristic fixed. Adding a period effect gives the Two-Way Fixed Effects model,

$$Y_{it} = \delta D_{it} + \mu_i + \tau_t + \epsilon_{it} \tag{3}$$

where μ_i absorbs unobserved heterogeneity for unit i and τ_t for period t.

2.2 Counterfactuals with Treatments Defined in Space

The two designs use different sources of variation, between groups versus within units, to construct a counterfactual. Recent work questions DID under staggered treatments and dynamic effects (Callaway & Sant'Anna, 2021; de Chaisemartin & D'Haultfœuille, 2022). For an unstaggered treatment, though, the TWFE and DID estimators are agreed to be equivalent

(Wooldridge, 2021). Equation 2 assigns to the treatment any post-treatment difference between groups remaining after pre-treatment differences and a temporal trend; Equation 3 assigns to it any within-unit variation remaining after unit and period effects.

Neither reading holds once units interact with one another. That is the common case when a treatment is defined *in space*, as with a new transit line (Delmelle et al., 2020; Peng et al., 2023), affordable housing (Voith et al., 2022), the legacy of racist land policy (Aaronson et al., 2021), or the closing of abortion clinics (Quast et al., 2017), because treatment may spill onto nearby places (Anselin, 2003), leaving some units treated *indirectly* (Halloran & Struchiner, 1995; LeSage & Pace, 2014a; Sobel, 2006). The standard critique then applies to both designs equally. Coefficients and inferences are biased under spatial dependence, and a temporal design for removing confounders does nothing about it.

The between-group design carries a further assumption. Randomization is what licenses the claim that treated and untreated groups differ in nothing but treatment, and with a spatially defined treatment it is absent by construction: assignment is a deterministic function of location, so inferences will be invalid if anything systematically distinguishes units near stations (McMillen, 2018). Applied work usually leaves that claim as an assertion, and the rest of this section derives it.

2.3 When Parallel Trends Survives a Spatially Defined Treatment

The derivation that follows shows the claim to be smaller than the literature's harshest critics suppose, and larger than its practitioners generally assume.

Suppose the outcome in each period follows the spatial autoregressive process that motivates the rest of the paper,

$$y_t \;=\; \rho W y_t + X_t\beta + \tau_t + \delta d_t + \varepsilon_t \tag{4}$$

so that with the spatial multiplier $A = (I-\rho W)^{-1}$ the reduced form is $y_t = A(X_t\beta+\tau_t+\delta d_t+\varepsilon_t)$. Differencing over the two periods, the trend in the *untreated potential outcome* is

$$\Delta y(0) \;=\; A\,(\Delta X\beta + \Delta\tau + \Delta\varepsilon) \tag{5}$$

and Equation 1 requires the group means of Equation 5 to coincide. The covariate and disturbance terms raise the usual non-spatial questions. The time shock $\Delta\tau$ is where space enters, and its behavior depends on the configuration.

2.3.1 The benign case

Suppose there is a single spatial regime, W is row-standardized, and the period shock is a common scalar, $\Delta\tau = \Delta\tau\,\iota$. Row standardization gives $W\iota = \iota$, so $A\iota = \sum_{k\geq 0}\rho^k W^k\iota = \iota/(1-\rho)$, and the shock term equals $\Delta\tau/(1-\rho)$ for *every* unit, treated or not. The multiplier is common, it cancels in the group comparison, and parallel trends survives untouched. A spatially defined treatment therefore does not violate parallel trends *per se*.

2.3.2 A spatially varying shock

Now let $\Delta\tau$ vary over space: a county-wide reassessment that lands unevenly, a credit shock that reaches some submarkets first, any period effect with geography. Then $A\Delta\tau$ is unit-specific. Because A is a spatial smoother whose weights decay in graph distance, $(A\Delta\tau)_i$ is a local average of the shock field around i. Treatment defined in space means D is spatially clustered by construction, so the two groups sample *different parts of that field*, and

$$\text{mean}_{D=1}\,(A\Delta\tau) \;\neq\; \text{mean}_{D=0}\,(A\Delta\tau) \tag{6}$$

whenever $\Delta\tau$ is spatially non-constant. The multiplier amplifies this difference, since A weights nearby values most heavily and the groups are defined by nearness. This channel operates under row standardization, requires no regime story, and is difficult to rule out in most settings.

2.3.3 Two regimes

Suppose instead that the spillover parameter itself differs between the inside and the outside of the buffer, so the multiplier is $(I - \text{diag}(\rho_{g(i)})W)^{-1}$ (Anselin & Amaral, 2023; Durlauf & Johnson, 1995; Ertur et al., 2006). For a unit whose neighborhood lies wholly within its own regime the row sum is $1/(1-\rho_g)$, so the untreated trend is $\Delta\tau/(1-\rho_1)$ for treated units and $\Delta\tau/(1-\rho_0)$ for untreated ones: parallel trends holds if and only if $\rho_1 = \rho_0$. For units near the boundary the row sums are a blend of the two, and Figure 2 shows that such units are common, since many links cross the half-mile buffer. The blend is therefore unit-specific and varies with distance from the boundary, and so with treatment status.

2.3.4 A remark on connectivity

Unequal neighbor counts are sometimes said to break the comparison on their own. Under row standardization they do not. $W\iota = \iota$ closes that channel, and essentially all applied work, including this paper, row-standardizes. Two related facts remain. The diagonal of A is not constant unless W is regular, so *direct* effects remain unit-specific even under row standardization; and Ad is unit-specific by construction, so exposure is heterogeneous in its own right.

Heterogeneous exposure invites an obvious remedy. The remedy promises more than it delivers. Under a global specification, exposure is Ad and cannot be conditioned away by adding regressors. Under a local specification it is parameterized by WD, the share of each unit's neighbors that are treated (Delgado & Florax, 2015), and a zero-exposure comparison group exists. But WD carries no information about degree. Row standardized, it lies in $[0, 1]$, and units with four and with forty treated neighbors both return $WD = 1$. It also fails to repair parallel trends. Equation 1 concerns the untreated potential outcome, while WD is a treatment-side regressor. It redefines the estimand as own treatment holding neighborhood exposure fixed, plus the effect of exposure, and leaves the two cases above untouched. Because co-located units share their neighbor sets exactly, WD is constant within a building, so at unit level it is a building-level covariate and does nothing about the platial problem developed in Section 2.4.

> **Proposition.** For a treatment defined in space, (a) a single spatial regime, (b) a normalization that equalizes the row sums of the spatial multiplier, and (c) a spatially constant period shock are jointly *sufficient* for parallel trends. Absent them the untreated trend is a function of location, and treatment is a function of location, so the two are correlated by construction; parallel trends can then hold only by a coincidence between the shock field and the treatment geometry, which nothing in the setting supplies. With two periods none of the three conditions is testable, and neither is the coincidence.

The proposition does not distinguish the two designs. Equation 7 below establishes that $\hat{\delta}_{FE} = \hat{\delta}_{DID}$ exactly in the aspatial two-period case, so whatever bias a differential trend induces,

it is induced identically in both. The proposition condemns the *aspatial estimator common to both designs*, and it therefore reaches the applied literature whichever design a given paper chose. An explicit spatial model is a requirement of the setting. Separation *between* the designs is a distinct result, taken up in Section 2.4.

Unit effects are no remedy for the shock channel. The term $A\Delta\tau$ in Equation 5 is a *change*, not a level. It passes through the within transformation intact, so conditioning on μ_i removes nothing from it. A spatially varying period shock biases the within-unit design exactly as it biases the between-group design. Remedying it would require either more than two periods or a parameterization of $\Delta\tau$ as a low-dimensional spatial function. At $T = 2$ that function competes directly with the treatment, since both are functions of location. That parameterization is identified only if the shock field is smoother than the treatment geometry. This design cannot test that assumption, and I do not make it.

The exercise is limited accordingly, and the same fact protects the comparison. Because $A\Delta\tau$ enters Δy identically under both designs, it is common to both and cancels in the *difference* between them. The divergence documented below is therefore attributable to the μ channel, the one the designs treat differently, and the normative conclusion drawn from it should be read as applying to that channel alone.

The proposition also disposes of the standard escape route. Where a differential trend is suspected, the usual remedy is a triple-difference design, which adds a third contrast, such as an unaffected outcome or an unaffected subgroup, and identifies the effect from the difference between the two differential trends. That works when the contaminating trend is common to the two comparisons and cancels. Here it is not. The differential trend arises because the spatial multiplier is a function of location, so it belongs to the geography of the treatment rather than to the outcome or the subgroup. It therefore enters both arms of the triple difference in the same way, and a DDD design combining fixed effects with a between-group comparison inherits the violation instead of differencing it away.

The closest existing precedent is direct. Lacombe (2004) asked whether border counties drawn from adjacent states, a widely used quasi-experimental control-group design adopted specifically to handle spatial confounding, succeeded in removing spatial dependence, and found that it did not. That forecloses the natural counterexample, that a boundary design comparing locally similar units escapes the problem. Locally similar units are precisely the ones whose multiplier row sums blend across the boundary.

The underlying intuition is that, for treatments defined in space, the DID design always uses a *different place* as a counterfactual for *this place*, and one location is a poor proxy for another (Anselin, 1989; Paelinck, 2009). The literature on spatial heterogeneity that motivates geographically weighted regression and spatial regimes models is a sustained argument to that effect (Koschinsky, 2009; McMillen, 2018; Wolf et al., 2017). The groups differ systematically before treatment, along the same dimension that defines treatment.

The within-unit estimator instead pools across units that have larger or smaller deviations from their mean. The FE model assumes that the treatment variable is the only variation induced over time, and a unit at a fixed place cannot by definition change location, so the assumption transfers well to spatial data. The FE approach also avoids cross-regime comparison, since a unit cannot relocate into another regime's spatial process and the comparison is always made against the same place[1].

[1] This requires that there has been no shift in spatial regimes over time, which is a reasonable assumption in many cases, but also subject to scrutiny, depending on the research question.

2.4 Why the Designs Cease to Be Equivalent

The two designs coincide exactly in the aspatial case, and coincide again under a spatial lag once ρ is held fixed. They diverge because ρ itself is identified from residual variation the designs do not share. Concede the equivalence in full first.

Let $d_{it} = D_i \cdot \mathbb{1}[t = 2]$ denote the activated treatment in a balanced two-period panel. The FE design conditions on the space spanned by unit and time dummies, the DID design on the smaller space spanned by $\{1, D_i, T_t\}$. Writing M_S for the residual maker of a control space S, both residualize the treatment *identically*:

$$M_{\{\text{unit, time}\}}\, d \;=\; M_{\{1,\,D,\,T\}}\, d \;=\; \big(D_i - \bar{D}\big)\Big(\mathbb{1}[t = 2] - \tfrac{1}{2}\Big) \tag{7}$$

Everything the FE control space absorbs beyond the DID control space is time-invariant, and so orthogonal to the residualized treatment at $T = 2$; by Frisch–Waugh–Lovell, $\hat{\delta}_{FE} = \hat{\delta}_{DID}$ exactly (Wooldridge, 2021).

The equivalence extends to a spatial lag. Let $\mathbf{W} = I_T \otimes W_N$ and define the filtered outcome $\tilde{Y} = (I_{NT} - \rho\mathbf{W})Y$. Conditional on ρ the model is linear in $\tilde{Y}$, and neither the regressor nor either control space has changed, so Equation 7 still holds and

$$\hat{\delta}_{FE}(\rho) \;=\; \hat{\delta}_{DID}(\rho) \qquad \text{for every } \rho \tag{8}$$

The within transformation does not remove spatial dependence. The within operator $Q = Q_T \otimes I_N$ acts on time while $\mathbf{W} = I_T \otimes W_N$ acts on the cross-section, so the two commute and ρ is unchanged by demeaning. Thus a spatial panel model is still required after fixed effects are included, consistent with Anselin & Arribas-Bel (2013). Unit effects control for unit-level unobservables while remaining blind to interactions *between* units.

The designs can differ only through $\hat{\rho}$, and the direction of that difference follows from the control spaces being nested, since every column of the DID design matrix lies in the span of the unit and period dummies. Writing P_{FE} and P_{DID} for the two projections and $M = I - P$ for the corresponding annihilators,

$$M_{DID} \;=\; M_{FE} + G, \qquad G \;=\; P_{FE} - P_{DID} \tag{9}$$

G is itself a projection, orthogonal to M_{FE}, onto the space of time-invariant deviations of units from their group means, the $\mu_i - \bar{\mu}_{g(i)}$ direction that group-level conditioning fails to remove. The residual sums of squares therefore decompose exactly, as in Equation 10. With $y(\rho) = (I_{NT} - \rho\mathbf{W})y$ and $q(\rho) = n^{-1}\, y(\rho)' M\, y(\rho)$,

$$q_{DID}(\rho) \;=\; q_{FE}(\rho) + g(\rho), \qquad g(\rho) \;=\; n^{-1} y(\rho)' G\, y(\rho) \;\geq\; 0 \tag{10}$$

The concentrated log-likelihood in ρ is $L(\rho) = \ln|I - \rho\mathbf{W}| - \frac{n}{2}\ln q(\rho)$, and the Jacobian term depends only on $\mathbf{W}$, and is therefore identical under both designs. Thus the designs are separated by q alone. Evaluating the DID score at $\hat{\rho}_{FE}$, where the FE first-order condition sets $\ln|I - \rho\mathbf{W}|' = \frac{n}{2}\, q'_{FE}/q_{FE}$, gives

$$L'_{DID}(\hat{\rho}_{FE}) \;=\; \frac{n}{2} \cdot \frac{q'_{FE}\, g - g'\, q_{FE}}{q_{FE}\,(q_{FE} + g)} \tag{11}$$

The denominator is positive, so the sign of the score, and hence whether the DID likelihood is still rising where the FE likelihood has peaked, is the sign of $d \ln q_{FE}/d\rho - d \ln g/d\rho$. Since $q'/q = -2\, e'(MWy)/(e'e)$ for either design, the comparison has a direct reading:

$$\hat{\rho}_{DID} > \hat{\rho}_{FE} \quad \Longleftrightarrow \quad \frac{e_G' G W y}{e_G' e_G} > \frac{e_{FE}' M_{FE} W y}{e_{FE}' e_{FE}} \qquad \text{at } \rho = \hat{\rho}_{FE} \tag{12}$$

The inequality in Equation 12 says that the component the DID design leaves in the error is more spatially autocorrelated than the FE residual is. The right-hand side is positive. The SAR first-order condition sets $e' M W y$ equal to the Jacobian term, not to zero, so at $\rho = \hat{\rho}_{FE}$

$$\frac{e_{FE}' M_{FE} W y}{e_{FE}' e_{FE}} \;=\; n^{-1} \operatorname{tr}\!\left[W \left(I - \hat{\rho}_{FE} W\right)^{-1}\right] \;=\; n^{-1} \sum_{k \geq 1} \hat{\rho}_{FE}^{k} \operatorname{tr}\left(W^{k+1}\right) \tag{13}$$

which is strictly positive for $\hat{\rho}_{FE} > 0$ and grows without bound as $\hat{\rho}_{FE} \to 1$. The condition is therefore that the within-group heterogeneity $\mu_i - \bar{\mu}_{g(i)}$ carry residual spatial autocorrelation exceeding a positive threshold, not merely exceeding zero. Nothing guarantees that the ordering reported below meets it.

The threshold's magnitude is governed by the density of W. Since $\operatorname{tr}(W) = 0$ for a graph without self-loops and $\operatorname{tr}(W^2)$ is $O(1/\bar{k})$ in the mean degree under row standardization, Equation 13 is small for the dense graphs used here, and Section 5.2 reports its measured values. Because it is evaluated at the *FE* estimate rather than the DID estimate, treating it as negligible is defensible in this application. The defense is empirical, about these weights and not about weights in general, and it degrades exactly where $\hat{\rho}_{FE}$ approaches unity.

The result is an equivalence. The ordering holds exactly when that condition holds, and if unobserved unit heterogeneity is spatially unstructured the two designs agree. At the wider treatment buffer of Section 5.7 the one-bedroom subgroup returns $\hat{\rho}_{DID}$ below $\hat{\rho}_{FE}$, the ordering reversing where the condition fails to bite. The condition is compelling here because treatment is a *binary* function of location. A group dummy can absorb the mean difference between the inside and outside of a half-mile buffer, but not the finer spatial variation within either region, which passes through into G intact. The condition is also directly checkable, by computing a Moran statistic on the estimated $\hat{\mu}_i - \bar{\hat{\mu}}_{g(i)}$ from the FE fit. Section 5.1 sets the spatial structure of μ by design and confirms that the two estimators coincide when that structure is absent and separate monotonically as it is introduced. Section 5.2 evaluates the condition on the estimation sample itself, where it is satisfied in every cell and its margin predicts the size of the realized divergence.

Two qualifications attach to the derivation and are stated in the footnote[2].

Further, because $\hat{\delta}$ and $\hat{\rho}$ are determined jointly in a non-linear system, a divergence in $\hat{\rho}$ propagates into $\hat{\delta}$. The equivalence in Equation 8 is exact only at a common ρ; once the designs disagree about ρ, nothing in the argument bounds how far apart their $\hat{\delta}$ may fall.

More important for policy, the cumulative impacts propagate through $(I_N - \rho W_N)^{-1}$, which is convex in ρ and diverges as $\rho \to 1$. An upward bias in $\hat{\rho}$ therefore produces a disproportionate distortion in the indirect and total effects. It is the total effect, not δ, that determines the magnitude against which a land-value-capture or anti-displacement policy should be sized.

[2] Equation 11 evaluates the DID score at $\hat{\rho}_{FE}$ alone, so a positive score there locates the DID maximizer to its right only if L_{DID} is unimodal over the stationary region, which is not guaranteed for the concentrated SAR likelihood in general. Four dispersed chains mix to $\hat{R} = 1.0$ for ρ in every fit reported below, which well-separated modes would not generally produce; were unimodality to fail, Equation 12 would characterize a local ordering only. Separately, Equation 11 is derived for the concentrated likelihood and so characterizes the mode, while the models are estimated by Gibbs sampling; the argument transfers to the posterior under the flat prior on ρ over the stationary region used here.

Because the multiplier grows without bound as ρ approaches unity, two designs need not differ greatly in $\hat{\rho}$ to differ by an order of magnitude in what they imply.

The natural rejoinder, that one could simply include unit fixed effects in the DID specification, concedes the argument instead of defeating it. With unit fixed effects, G in Equation 9 vanishes, $\hat{\rho}$ coincides across designs, and the specification *is* the FE design. Meeting the objection means abandoning the between-group design.

A second, independent mechanism reinforces the distinction. The DID design carries the treated group through two regressors, D_i and $D_i \times T_t$, whereas the unit fixed effect in the FE design absorbs D_i and leaves a single treatment indicator. Aspatially this is a cosmetic difference. Under a spatial lag it is not. Sheng & LeSage (2021) show that multiplicative interaction terms in SAR and SDM specifications cannot be interpreted through the standard impact formulas, because the interaction cannot be perturbed independently of its constituent terms. Recovering direct and indirect effects from a spatial DID therefore requires the total derivative through both D and $D \times T$, and applying the conventional LeSage & Pace (2009) decomposition to the interaction coefficient alone misstates them. Both mechanisms trace to one act: absorbing D_i into μ_i. The scope of the derivation is narrower than that of the empirical exercise. Equation 7 through Equation 12 treat the SAR case, in which the treatment enters through a single lag; this second mechanism acts on the regressor list, and so on a different object from the residual maker of Equation 9, and I have not carried it through the same machinery. The Durbin results reported below are therefore empirical. The paired comparisons pool the two mechanisms, so the count in Section 5.3 measures their joint effect and separates neither.

None of this rescues the FE design from the SUTVA violation. If there is spillover in the treatment or the dependent variable, *all* units in the post period receive an indirect treatment through spatial interaction, so the "untreated" group is always treated indirectly under either design. But the violation has different consequences in the two designs. Under the between-group design, interference contaminates the *comparison group*, the very object on which identification rests, so the contamination enters where it cannot be modeled away. Under the within-unit design it contaminates the *estimand*, which an explicit spatial model then decomposes into direct, indirect and cumulative components. The same violation is fatal to identification in one case and a specification requirement in the other.

The two results compose, with the scope each carries. Parallel trends fails for a spatially defined treatment, identically under both designs, so an explicit spatial model is required; and once one is fitted, the estimate turns on a parameter that the between-group design recovers only when the omitted unit heterogeneity is spatially unstructured. Where that heterogeneity is spatially structured, as a treatment assigned by a distance buffer all but guarantees, the between-group design fails to recover the effect. The parameter governing the estimate is then measuring the design's own omission, not the outcome process. On that channel, the remedy for difference-in-differences is to stop differencing between groups. On the shock channel neither design is protected, and the remedy is more periods.

Figure 2 shows the issue for the Purple Line corridor. Each apartment building appears as a point, red circles mark the half-mile buffers around Purple Line stations, and a line joins two buildings when either is among the other's eight nearest neighbors. "Treated" units fall inside a red circle and "untreated" units outside. Many links cross the buffer, so units in different treatment groups stand in a spatial-interactive relationship to one another.

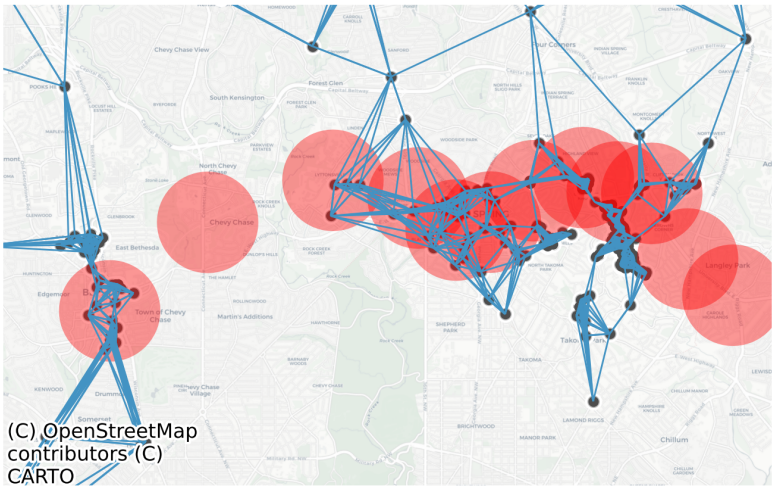


Figure 2: Eight Nearest-Neighbor Connectivity Graph for Apartment Buildings in the Purple Line Corridor

3 Spatial Models for Causal Policy Analysis

Cross-sectional data in a spatial specification do not provide a credible causal design, and Gibbons & Overman (2012) call for applied work that adds a temporal dimension and exogenous variation. A small but growing literature has responded.

3.1 Global and Local Spillovers

I use the taxonomy of Halleck Vega & Elhorst (2015), restricting attention to the spatial autoregressive (SAR), spatial error (SEM), spatial lag of X (SLX) and Durbin (SDM, SDEM) specifications; Akbari et al. (2023) give a graphical overview of the interaction structures each implies.

Three terms distinguish them. $\rho W Y$ (SAR, SDM) captures *global* spillovers in the dependent variable, $\theta W X$ (SLX, both Durbin variants) captures *local* spillovers in the exogenous variables, and $\lambda W u$ (SEM, SDEM) captures spatial correlation in the error. Estimating all three jointly yields the General Nesting Specification, which is under-identified in practice (Elhorst, 2010).

The distinction that matters here is between global and local. Global spillovers propagate without limit, reaching neighbors, then neighbors of neighbors, and onward. They therefore generate *feedback*, in which a change at one location returns to affect that location through the system. Local spillovers fall only on immediate neighbors and die out, so no feedback arises (LeSage & Pace, 2014a). The choice between them is theoretical, not empirical: practitioners "cannot simply estimate an SDM and SDEM model and draw meaningful conclusions about the correct model specification from point estimates of ρ or λ" (LeSage, 2014, p. 22). The error is asymmetric. Fitting a local model where a global process operates understates indirect and total effects, because the local model has no feedback to capture.

That has a direct implication for treatment effects. Under a global specification, a treatment applied to some units propagates to all units in the system, so every observation receives an indirect treatment in the post period and the untreated group is untreated only nominally. Under a local specification, spillover reaches immediate neighbors and stops, so a unit may be directly treated, indirectly treated, both, or genuinely neither.

3.2 Interpreting the Parameters

For local specifications the model remains linear and coefficients read as usual, with δD the direct effect and θWD the spillover; reported t-statistics are valid (LeSage & Pace, 2014a). For global specifications they are not. The coefficients are not marginal effects, and interpretation requires computing impacts: the average direct effect from the mean diagonal of $(I_N - \rho W)^{-1}$, and the indirect effect from the mean row sums of its off-diagonal elements (Elhorst et al., 2012, p. 216).

In a causal framework these are the average direct and indirect treatments, and the causal effect of the intervention is their sum, the *cumulative* effect. Direct, indirect and cumulative effects all bear on policy. Because the counterfactual under a global specification is the absence of treatment anywhere, an analyst must report the direct and indirect components separately. Dispersion measures for them are given in LeSage & Pace (2009) and Elhorst (2010).

For the *local* specifications the decomposition is immediate and is reported in full: δ is the direct effect, θ the indirect, and $\delta + \theta$ the cumulative effect, each with a posterior standard deviation. For the *global* specifications it requires the multiplier $(I - \rho W)^{-1}$. There I report the coefficients and $\hat{\rho}$ separately, leaving the impacts to be computed from them, because the paper's central finding is that the two designs disagree about ρ. Impacts computed from a $\hat{\rho}$ that is absorbing the design's own omitted heterogeneity would present that omission as an estimate with a credible interval; reporting the ingredients separately leaves the multiplier's sensitivity visible.

3.3 Spatial Difference in Differences

A recent literature develops DID estimators for spatial models, motivated by the natural experiment framework that dominates applied policy analysis. For local spillovers, Delgado & Florax (2015) develop what amounts to an SLX-DID, which yields a clean counterfactual, since excluding global spillovers leaves observations genuinely untreated in the post period, at the cost of admitting no feedback. Comber & Arribas-Bel (2017) fit a SEM-DID to anticipatory effects of Crossrail openings, and Bardaka et al. (2018) extend Delgado et al. to an SDEM-DID, finding a small but significant direct effect of new stations and concluding that transit effects are overstated when spatial effects are ignored.

For global spillovers, Dubé et al. (2014) introduce a SAR-DID estimator (their "SDID"), reporting small direct and indirect effects of rail stations on property values. Dubé et al. (2018) apply it to find positive effects during construction but negative effects at announcement, and, because feedback is admitted, larger price increases under the spatial model than an aspatial analogue, a result unavailable to designs without global spillovers. Chagas et al. (2016) and Diao et al. (2017) apply the same specification to high-speed rail openings.

Two features of this literature bear on what follows. Reporting is uneven: several of these studies give regression t-values instead of the dispersion measures appropriate to a global model, and some give no direct or indirect effects at all. Because the coefficients are not marginal effects, that reticence limits what can be concluded from them (LeSage & Pace, 2014a; LeSage & Pace, 2009). More important for this paper, every one of these designs is a between-group design, and none carries unit effects, the property Section 2.4 concerns.

3.4 Spatial Panel Models

Spatial fixed- and random-effects panel models have a longer history (Anselin et al., 2008; Elhorst, 2003, 2014), with extensions to dynamic effects, unbalanced panels and multiple

weights matrices (Lee & Yu, 2012; Yu, 2007). In the two-period fixed-effects case the analogous model is

$$Y_{it} = \rho \sum_j w_{ij} Y_{jt} + \delta D_{it} + \tau_t + \mu_i + \epsilon_{it} \tag{14}$$

the Spatial Panel Fixed Effects model. Although it appears routinely in the methodological literature, I am not aware of applied transit research that adopts it, and the reason is practical. It requires repeat observations on the same unit, and homes transact too infrequently to supply them. Rental contracts escape that limit. Leases are short and a unit is "transacted" annually even when a tenant renews. That is what makes the design available here, and what puts the comparison in Section 2.4 on real data at all.

Table 1 sets the two observations side by side. No one has run the comparison this paper runs, and the omission has a cause: the design that would supply the within-unit arm requires panel data on individual dwellings, which property-transaction records do not provide. The table also forecloses the natural objection that unit effects are already standard practice, so that the between-group problem is a straw man. In this literature, they are not standard.

Table 1: Applied and methodological work combining a quasi-experimental design with a treatment defined in space. Every study identifies the effect from variation *between* treated and untreated groups, and none carries unit effects; the spatial panel fixed-effects model that would supply them appears in the methodological literature but not, so far as I am aware, in applied transit research, because repeat observations on the same dwelling are rarely available. The published analysis of the present data is aspatial, and is therefore reached by the parallel-trends result rather than by the result on the spatial parameter. Reporting practice for direct and indirect effects varies across these studies and is discussed in the text rather than tabulated here.

Study	Specification	Spillovers	Design	Unit effects
Peng et al. (2023)	Aspatial DID	None modeled	Between-group	No
Delgado & Florax (2015)	SLX–DID	Local	Between-group	No
Comber & Arribas-Bel (2017)	SEM–DID	Error only	Between-group	No
Bardaka et al. (2018)	SDEM–DID	Local	Between-group	No
Dubé et al. (2014)	SAR–DID	Global	Between-group	No
Chagas et al. (2016)	SAR–DID	Global	Between-group	No
Diao et al. (2017)	SAR–DID	Global	Between-group	No
Dubé et al. (2018)	SAR–DID	Global	Between-group	No
Spatial panel FE *(not adopted in applied transit work)*	SAR, SEM, SDM, SDEM	Either	Within-unit	Yes
This paper	SAR, SEM, SDM, SDEM	Both	Both	Yes

4 Methods

The hedonic price model decomposes a selling or rental price into its constituent parts. Rosen (1974) formulated it as a structural account of housing demand, and applied work now uses it chiefly to estimate marginal willingness to pay for features of the housing bundle: in practice analysts skip the second stage and read the coefficients as marginal changes in price (Freeman, 1979; Harrison & Rubinfeld, 1978). The urban economics literature has long assumed those prices are interdependent, because attributes are valued relative to other nearby units (Can & Megbolugbe, 1997; Steimetz, 2010).

4.1 Data

The dataset is taken from a survey of multifamily rental units in Montgomery County, Maryland. The county's Department of Housing and Community Development requires a lessor to apply for a rental permit before a unit may be rented in the county, and the survey is administered during the permitting process. The rents are landlord-reported, and the survey does not distinguish contract from asking rents. Peng et al. (2023) describe the dataset in detail[3]. Following Peng et al. (2023), I fit separate models for units of different sizes (i.e. studio, one-bedroom, two-bedroom, and three-bedroom), which may represent different submarkets[4]. Table 2 reports rent distributions by unit type and period. Median rents rise across every subgroup between 2015 and 2018, and the dispersion within each subgroup is large relative to the between-subgroup differences.

4.1.1 Treatment

The Purple Line has an unusually long history for a treatment of this kind. Proposals for an east–west transitway across the Maryland suburbs circulated for decades through several revivals that did not proceed, so that *a* line might eventually be built was not news to the market. Over the study window the project became certain enough to act on and its station locations became specific: at the most recent approval the general corridor was already known, while precise locations were settled only later[5]. No construction had begun within the study window.

The treatment is therefore best read as a resolution of uncertainty about location and timing and as a shock weaker than an unanticipated announcement would deliver. Small effects follow from that weakness. Because the corridor was known before the station points were fixed, any anticipatory capitalization prior to the study window would attach to the corridor, which is broader than the half-mile buffers used here. That too works against finding an effect on the treatment indicator.

Peng et al. (2023) give fuller institutional treatment of the survey and the corridor; their published analysis of these data reports a positive announcement effect and is the result to which this paper responds. The comparison requires only that both designs face the identical setting. Whatever that setting withholds, it withholds symmetrically, and cannot account for the divergence between them.

Table 2: Monthly contract rent by unit type and period, computed on the estimation sample. Median rents rise across every subgroup between 2015 and 2018, and dispersion within each subgroup is large relative to the differences between subgroups, which is what makes the within-unit comparison informative.

	count		mean		std		min		50%		max	
Units	2015	2018	2015	2018	2015	2018	2015	2018	2015	2018	2015	2018
Studio	1,668	1,668	1,313	1,359	293	289	555	572	1,351	1,400	1,900	2,102
1 BR	13,744	13,744	1,323	1,377	414	381	154	196	1,275	1,340	2,607	2,700
2 BR	13,815	13,815	1,544	1,607	358	348	477	476	1,468	1,546	2,714	2,891
3 BR	2,440	2,440	1,711	1,781	432	448	647	445	1,708	1,775	3,096	3,331

[3] For more information see https://montgomerycountymd.gov/dhca/housing/licensing/apply.html

[4] Although there are four models from different subsets of the data, the single specification applies to all subsets, so in this section I refer to "the model" for simplicity.

[5] Planning documents recording the alignment and station locations across successive stages of the project are collected in the Purple Line document library, https://www.purplelinemd.com/document-library/#plans.

4.2 Model Specification

A long tradition of theoretical and empirical work gives good reason to believe that home prices affect one another in a hedonic housing price model, which makes a variation of the SDM the appropriate model for this design (Can & Megbolugbe, 1997; Kim et al., 2020; Steimetz, 2010). I therefore define a treatment D as proximity to a station that will be served by the Purple Line, and allow for both global and local spillovers. After controlling for unit and temporal fixed effects, a fully specified model takes the form

$$Y_{it} = \rho W y_{jt} + \delta D_{it} + \theta W D_{jt} + \tau_t + \mu_i + \epsilon_{it} \tag{15}$$

which I term the Spatial Durbin Model with Two Way Fixed Effects (SDM-TWFE)[6]. This is essentially the same model preferred by Kolak & Anselin (2020). Here, Y_{it} is the rent for unit i at time t and $W y_{jt}$ the spatial lag of price. D_{it} is a binary treatment indicator equal to one when unit i lies within half a mile of a Purple Line station, and $W D_{jt}$ is its spatial lag, capturing local spillovers[7]. The period indicator takes the value one for rents measured in 2018 and zero for 2015, with coefficient τ_t; μ_i is a unit-level fixed effect, and ϵ_{it} is a random error term. This setup is conceptually similar to Elhorst & Fréret (2009) except that I consider only two time periods and a single spatial regime.

This specification lets prices in the study region interact with one another (i.e. global spillover in the dependent variable), and it admits a direct effect of the Purple Line and a local spillover effect of the Purple Line that are independent of each other. The direct effect could be negative, if anticipated construction noise and traffic depress values nearest to the station, while the *spillover effect* is positive for units outside the construction window but still proximate enough to receive an accessibility boost once the line opens.

These specifications carry no hedonic characteristics. The omission follows from the fixed effect. Unit attributes such as floor area, layout, bedroom count, on-site amenities, structure type and position relative to shops, employment and existing rail do not change over a three-year window, and μ_i absorbs every time-invariant characteristic of a unit whether observed or not. Including them would add collinear columns to a design matrix that has already differenced them away. The corresponding assumption is not that these attributes are irrelevant but that they are constant, which for structural characteristics is close to definitional and for neighborhood attributes is the identifying assumption stated below. With $T = 2$ the within transformation and first differencing coincide up to a scale factor, but I estimate the model in levels with unit and period effects, so the spatial lag enters contemporaneously as $W y_{jt}$, not as a lag of differences.

[6] Note that this specification has large N and small T, which raises the incidental parameters problem. Because the argument of Section 2.4 turns on a comparison of $\hat{\rho}$ across designs, it matters whether short T biases that parameter mechanically. It does not. Lee & Yu (2010) show that for the spatial panel model with *individual* effects, the direct and transformation approaches differ only in the variance parameter: $\hat{\sigma}^2_{\text{direct}} = \frac{T-1}{T}\hat{\sigma}^2$, a constant factor that leaves the concentrated likelihood's maximizer in the spatial parameter unchanged, so the two approaches return numerically identical $\hat{\beta}$ and $\hat{\rho}$. Their Theorem 1 displaces only the σ^2 element of the probability limit, and they state that the estimators of the remaining common parameters "will be consistent even when T is small." Contamination of the common parameters arises in their framework only when individual *and* time effects are both concentrated out; here $T = 2$, so the period effect enters as a single estimated coefficient, the case they treat as equivalent to an additional regressor. The direct estimate understates σ^2 by the factor $\frac{T-1}{T}$, so I report it multiplied by $\frac{T}{T-1}$; $\hat{\rho}$ requires no correction.

[7] Since the treatment is binary, the spatial lag term here represents the share of each unit's neighbors that were treated (Delgado & Florax, 2015)

This does raise a fair objection to the comparison itself. The between-group design absorbs only group means, so unit attributes are *not* differenced away there, and a DID specification of this form omits controls it arguably needs. One might therefore ask whether the divergence reported below simply reflects an under-specified DID, remediable by adding the hedonic covariates available in these data.

Covariates cannot close it. They narrow the omitted-variable problem without eliminating it, since what remains is by construction whatever was not measured; the within transformation faces no such limit. The asymmetry is categorical. A covariate set conditions on a list, while the fixed effect conditions on the unit. Spatial dependence then raises the cost of that asymmetry. In an aspatial regression the residual heterogeneity biases $\hat{\delta}$ and stops there; in a spatial model it is additionally available to be absorbed by ρ, and Section 2.4 shows how a bias in ρ propagates. I report the parsimonious specification because it isolates the conditioning sets themselves. The results report a pooled specification in which the DID does carry a control, and the gap does not close.

The model's critical assumption is that the announcement of the Purple Line is the only variable changing over time. All other time-varying factors must be orthogonal to treatment. Over a window as short as 2015–2018, substantial neighborhood change is unlikely, but this remains an identifying assumption, not a testable one. I also fit the standard (non-Durbin) SAR-TWFE model shown in Equation 14, as well as the Durbin and non-Durbin analogues for the spatial panel error model. Adopting the same notation used throughout, these are given by

$$\begin{aligned} Y_{it} &= \delta D_{it} + \theta W D_{jt} + \tau_t + \mu_i + u_{it} \\ u &= \lambda W u_{jt} + \epsilon_{it} \end{aligned} \tag{16}$$

for the Spatial Durbin Error Model with Two Way Fixed Effects (SDEM-TWFE), and

$$\begin{aligned} Y_{it} &= \delta D_{it} + \tau_t + \mu_i + u_{it} \\ u_{it} &= \lambda W u_{jt} + \epsilon_{it} \end{aligned} \tag{17}$$

for the spatial panel error model with two way fixed effects (SEM-TWFE). Together, Equation 15, Equation 16, Equation 17 and Equation 14 define the conceptual specifications under comparison, each of which I estimate under both the FE and the DID design and under each conception of the connectivity graph W described below. The argument of Section 2.4 concerns the divergence between designs, not the merits of any single specification, so I report every model for both. The quantity of interest is the comparison of $\hat{\rho}$ across designs, not its value in any one of them. I begin by fitting the classic aspatial designs from Equation 2 and Equation 3. For these specifications, the classic spatial econometric tests show clear evidence of spatial structure in the residuals, grounds for rejecting them in favor of models that include spillover effects.

4.3 Spatial Weights

Every model above requires a spatial connectivity graph W, and every standard conception of proximity assumes that observations occupy distinct locations. The apartment units in these data are *co-located*: many share a building and therefore share a coordinate pair exactly, a structure the classical literature, developed on polygon lattices, does not anticipate. Wolf et al. (2021) separate *platial* dependence, in which observations drawn from the same place are correlated, from *spatial* dependence, in which nearby observations are correlated regardless of grouping. The two are normally distinguishable because the grouping is administrative

while proximity is continuous, but that separation collapses when the group is a single point: within-group dependence is "only 'spatial' correlation if the groups themselves are spatially designed," and a building is spatially designed in the strongest possible sense. Units sharing one are at distance zero, which every conception of W treats as maximal proximity.

The consequence is that a single-level spatial model fitted to unit-level data cannot distinguish the two components. The lag identifying ρ mixes correlation between units in the same building with correlation between units in *different* buildings, and reports the sum as spatial dependence. The mixture is substantial. Depending on unit type and the choice of W, between 10% and 37% of each unit's spatial weight falls on its own building, while within-building rents are tightly clustered, at a median coefficient of variation of 0.08[8]. I return to this in the results, where estimating the same models on building-level aggregates isolates the genuinely spatial component and measures the share of $\hat{\rho}$ the platial component was carrying.

This mirrors Wolf et al. (2021), who show that omitting spatial dependence from a model with platial structure yields "more extreme estimates that have too narrow confidence intervals": when two forms of dependence are present and only one is modeled, the estimated parameter absorbs the other. The mechanism is distinct from the design problem of Section 2.4, though both inflate $\hat{\rho}$, and aggregating to buildings removes the platial source but not the design source, so I estimate both levels.

I use three conceptions of proximity. A distance band connects all units within a fixed radius, and a triangular kernel over the same support holds the band's neighbor set while letting the weight decay linearly to zero at the bandwidth. The kernel earns its place because a binary band weights a neighbor at 100 m identically to one at 3,999 m and then drops to zero, so any result sensitive to that discontinuity would measure the cutoff, not proximity. The kernel removes the discontinuity while leaving the support unchanged, and with it the sample.

The third conception, a k-nearest-neighbor graph, is available only where observations are distinct points[9]. Applied to co-located units, KNN is degenerate. A building containing more than k units connects only to itself, and the graph fragments into within-building components carrying no spatial information whatsoever. A *clique* adjustment, connecting all units in each of a building's k nearest *buildings*, so that connectivity is established across structures and not within them[10], repairs the fragmentation in principle, and the resulting graph, shown in Figure 3, is fully connected in the graph-theoretic sense[11]. In practice the co-located points

[8] Computed across units of the same bedroom count in the same building, the median coefficient of variation in rent is 0.077, so units in one building and bedroom class rent within roughly eight percent of one another. Even so, 23.5 percent of the variance in unit rents falls within building and bedroom class, not between classes. It carries no spatial information. Co-located units share a coordinate, so floor, aspect, and renovation history separate them, not location, and no spatial model at any level can attribute that variation to W. A multilevel specification (Dong & Harris, 2015) would represent it as a variance component instead of as spatial dependence, so aggregating costs $\hat{\rho}$ nothing spatial. The building-level models take the mean rent of the units of a given bedroom count in each building.

[9] A nearest-neighbor graph is ambiguous when candidate neighbors are equidistant, and the tie breaks differently across runs. Refitting the KNN specifications reproduces $\hat{\rho}$ to within 0.006 but moves the between-group treatment coefficient by as much as thirteen dollars, under half of its own posterior standard deviation. The distance-band and kernel results reproduce exactly. The instability is therefore confined to the quantity this paper reports as unidentified, and the KNN treatment coefficients should be read to the nearest ten dollars.

[10] Technically, Figure 3 demonstrates the clique approach, because it connects all eight nearest *buildings*, (i.e. forcing connections across buildings instead of allowing the eight nearest neighbors to reside in the same building.)

[11] Arguably, the KNN in general (and the clique approach) in particular are most consistent with classic regional science W matrices, which are typically based on polygon contiguity. These graphs are fully connected, in the

defeat the construction at unit level, so I report the nearest-neighbor results at the building level, where each structure contributes a single observation and the difficulty does not arise.

The distance-based conceptions match how the transit literature reasons about accessibility, where absolute distance, not relative rank, determines whether an amenity is available, and they let density matter. The cost is visible in Figure 4: the graph is not connected, resolving into dense local clusters separated by empty space, and a few buildings in the low-density north of the county have no neighbors at all. Such observations carry an empty row in W, the default handling in spatial econometric software, which leaves their spatial lag at zero. I exclude them instead, so that every specification at a given bandwidth is estimated on an identical sample and the comparison across weighting schemes is a comparison of weights, not of samples. There are four such units at the four-kilometer bandwidth and seven at two kilometers.

The graphs encode genuinely different theories of how rents interact, and none is obviously correct. LeSage & Pace (2014b) argue that results should be insensitive to the specification of W, and in Section 2.4 I have argued that the spatial structure is doing identifying work. Both claims cannot hold, and the results below show which of them these data support.

The treatment's geography governs the reading of the exogenous spillover term $\theta W D_{jt}$. Figure 5 plots the share of each unit's neighbors that fall inside a station buffer. Under both conceptions of W the distribution is sharply bimodal. The overwhelming majority of units have either no treated neighbors or almost exclusively treated neighbors. Defining treatment by proximity produces this directly. Treated units cluster, so their neighborhoods are largely treated as well. The bimodality means that $W D_{jt}$ carries little independent information beyond D_{it} for most observations, and that θ should be read with corresponding caution.

5 Results

I begin with the aspatial designs of Equation 2 and Equation 3. With the panel balanced and the treatment activated once, the two return identical estimates to four decimal places, exactly as Equation 7 requires: \$19.00 for studios, −\$24.01 for one-bedroom units, \$36.80 for two-bedroom units and \$50.65 for three-bedroom units. Nothing distinguishes the designs here, and on the strength of these models alone the distinction is immaterial. Spatial diagnostics on the residuals of both reject the null of no spatial dependence in every subgroup[12]. Table 3 reports both designs.

Table 3: Aspatial treatment effect under both designs on the estimation sample. The between-group and within-unit estimators coincide to four decimal places, as the equivalence result requires. Standard errors differ because the designs absorb different numbers of parameters; the point estimates do not.

			DID		FE	
Units	N	Obs.	$\hat{\delta}$	s.e.	$\hat{\delta}$	s.e.
Studio	1,668	3,336	18.9975	20.02	18.9975	5.47
1 BR	13,744	27,488	-24.0100	10.08	-24.0100	4.37

sense that each observation can traverse the graph to reach every other observation, consistent with the notion of global spillovers.

[12] All spatial models are estimated with `bayespecon`, which reports posterior summaries, impacts, and posterior-based specification diagnostics (Rey et al., 2021). Convergence is satisfactory throughout: $\hat{R} = 1.0$ for every parameter in every model reported.

2 BR	13,815	27,630	36.7952	10.04	36.7952	3.19
3 BR	2,440	4,880	50.6521	28.45	50.6521	8.95

5.1 Verifying the Mechanism

Equation 12 states an equivalence, so I check it directly, controlling the quantity it concerns and setting known values of δ and ρ_y that the application cannot supply.

I generate unit heterogeneity as a SAR process, $\mu \sim (I - \rho_\mu W)^{-1}\varepsilon_\mu$, on a 22×22 lattice with rook contiguity. Treatment is assigned to a contiguous block covering a quarter of the units, so that, as in the application, a binary group indicator can absorb the mean difference between inside and outside but not the spatial variation within either region. The outcome is a SAR process with parameter ρ_y and a known treatment effect $\delta = 10$, and both designs are fitted to identical replicates. Table 4 and Figure 6 report five hundred replicates in each of seven cells.

At $\rho_\mu = 0$ the mean gap is -0.009 and the DID estimate exceeds the FE estimate in 43% of replicates, a coin flip. As ρ_μ rises the gap opens monotonically, reaching 0.416 at $\rho_\mu = 0.8$, and the DID estimate exceeds the FE estimate in every replicate at every positive value. Moran's I on the within-group component $\mu_i - \bar{\mu}_{g(i)}$, the statistic Equation 12 says *is* the condition, correlates with the realized gap at 0.974.

The asymmetry between the designs is complete. The FE estimator recovers $\rho_y = 0.5$ at every value of ρ_μ, with a bias never exceeding 0.005; the DID estimate climbs from 0.489 to 0.911. The FE estimator also recovers $\delta = 10$ at every value of ρ_μ, with a bias never exceeding 0.08, while the DID estimate falls from 10.20 to 3.18, an understatement of 68% where the FE design is accurate to within one percent. Because Equation 8 fixes $\hat{\delta}_{FE}(\rho) = \hat{\delta}_{DID}(\rho)$ pointwise, that entire divergence in $\hat{\delta}$ is attributable to the divergence in $\hat{\rho}$. The observational analysis shows that the designs disagree. This one shows which is wrong.

Table 4: Monte Carlo, 500 replicates per cell on a 22×22 lattice with rook contiguity. Unit heterogeneity is drawn as a SAR process with parameter ρ_μ; the outcome is a SAR process with parameter ρ_y; treatment is a contiguous block, and the true effect is $\delta = 10$ throughout. Moran's I is computed on the within-group component of μ, which is the quantity the equivalence condition concerns. The within-unit design recovers both parameters in every cell. The between-group design recovers them only where μ is spatially unstructured, and its error grows with ρ_μ: at $\rho_\mu = 0.8$ it understates δ by 6.8. The final block sets $\rho_y = 0$, removing spatial dependence from the outcome entirely: the between-group design nonetheless reports $\hat{\rho} = 0.735$, so what it is measuring is the omitted heterogeneity rather than the outcome process. At the null the mean gap is -0.0088 and the between-group estimate exceeds the within-unit estimate in 43% of replicates.

		$\hat{\rho}$				$\hat{\delta}$	
ρ_y	ρ_μ	FE	DID	Gap	Moran I	FE	DID
0.5	0.0	0.498	0.489	-0.009	-0.005	10.03	10.20
0.5	0.2	0.496	0.596	+0.100	+0.103	10.07	8.43
0.5	0.4	0.498	0.700	+0.202	+0.218	10.04	6.70
0.5	0.6	0.498	0.807	+0.309	+0.361	10.03	4.91
0.5	0.8	0.495	0.911	+0.416	+0.555	10.08	3.18
0.0	0.0	0.000	-0.007	-0.007	-0.004	10.01	10.09
0.0	0.8	-0.000	0.735	+0.735	+0.559	10.00	3.56

The last two rows of Table 4 answer the most serious alternative explanation. A reader may grant everything above and still hold that $\hat{\rho}_{DID} \approx 0.96$ in the application is simply what

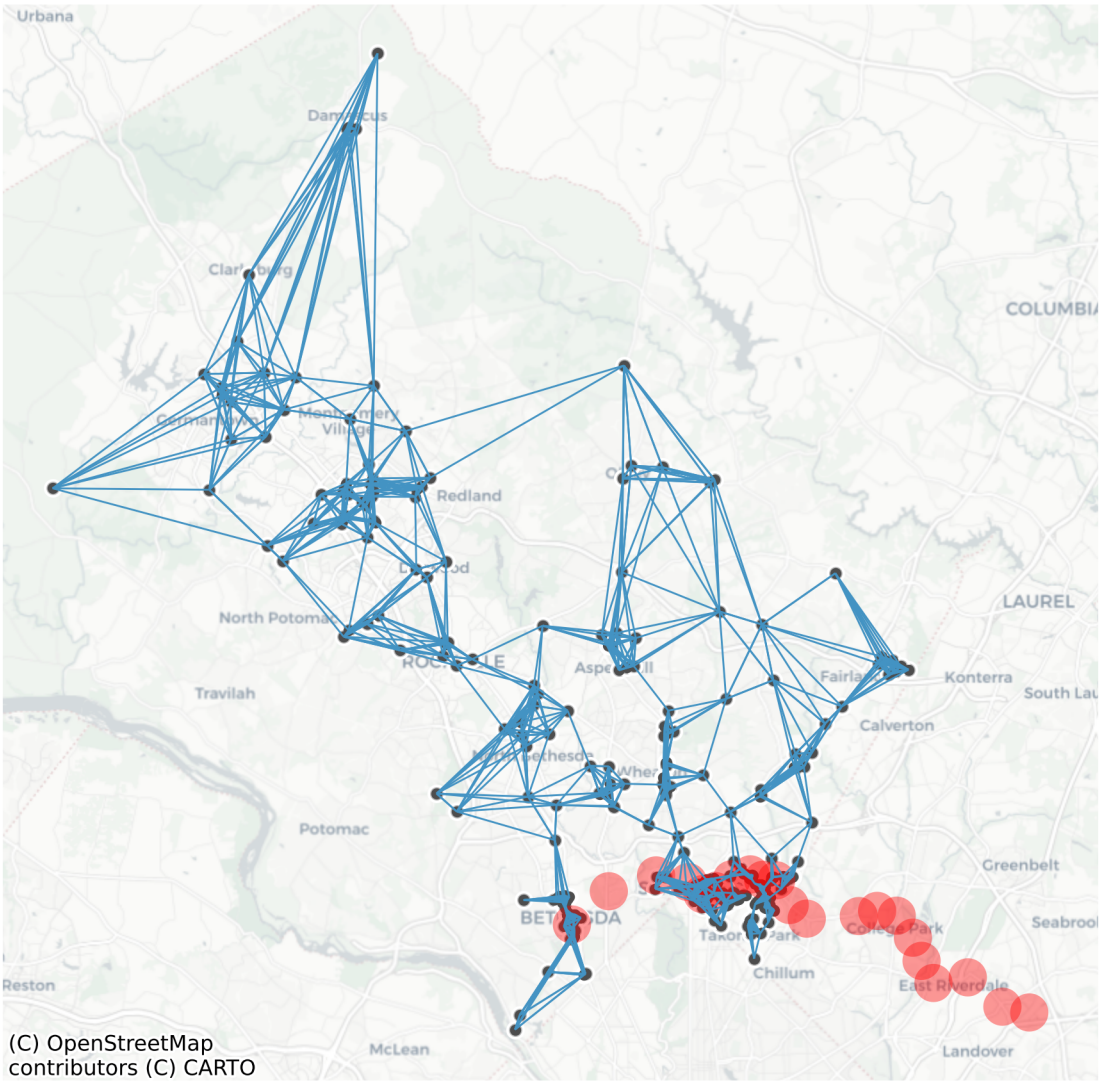


Figure 3: Nearest-neighbor connectivity graph for apartment buildings, Montgomery County. Purple Line station buffers shown in red.

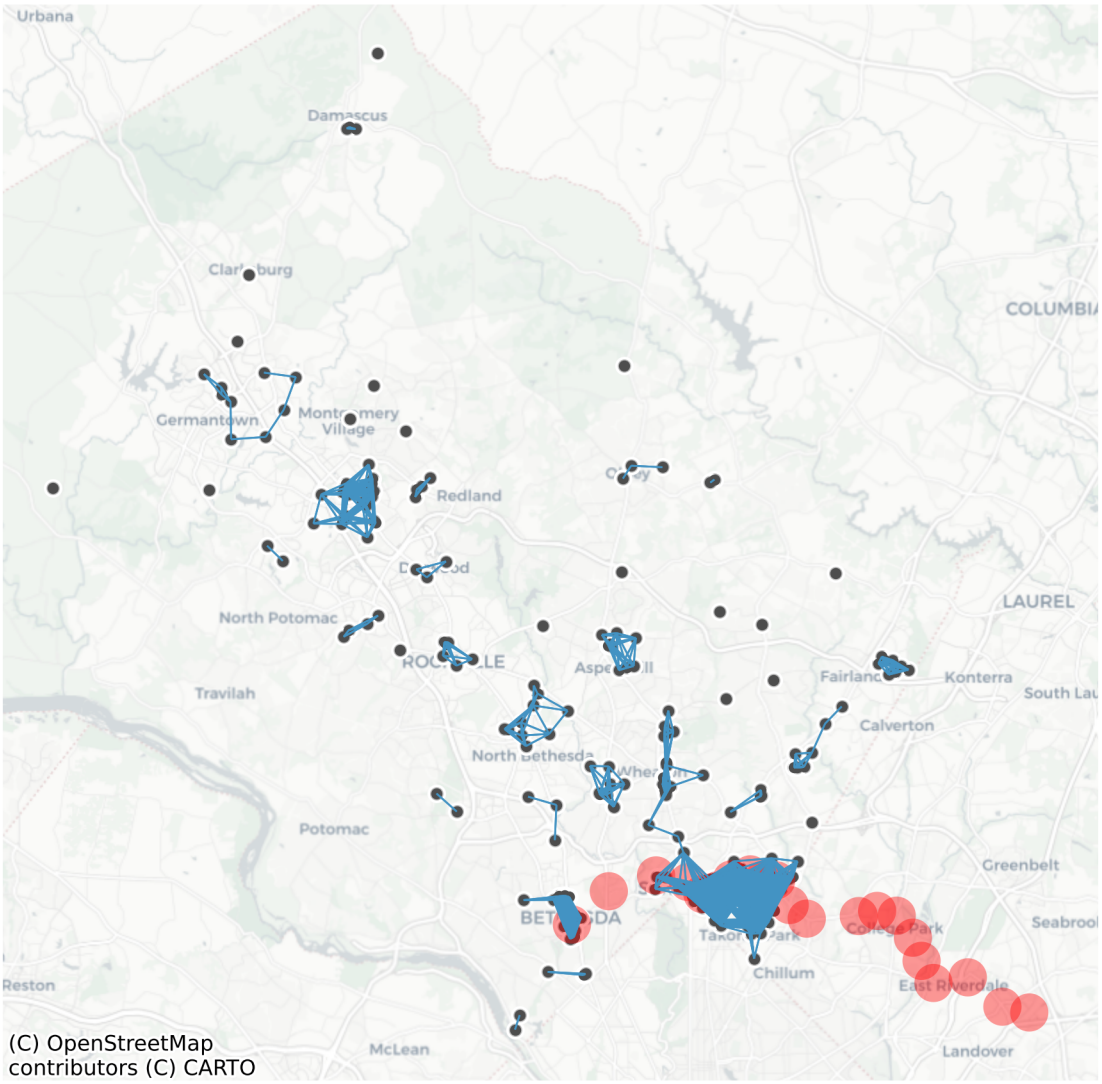


Figure 4: Distance-band connectivity graph for the same buildings. Compared with Figure 3, the graph fragments into dense local clusters and leaves isolated observations in low-density areas.

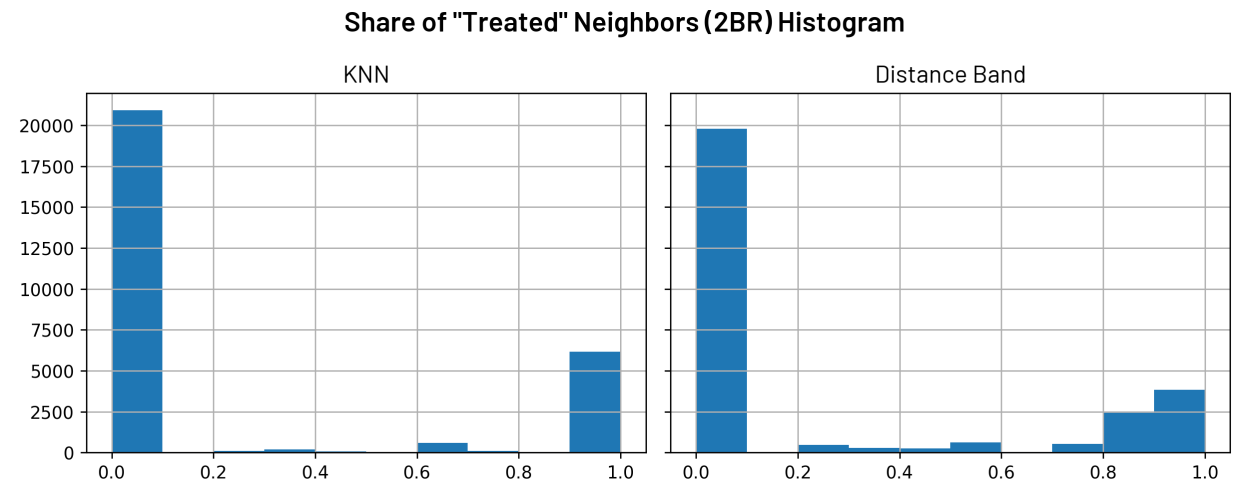


Figure 5: Share of “treated” neighbors under each connectivity graph, two-bedroom units. Both distributions are strongly bimodal.

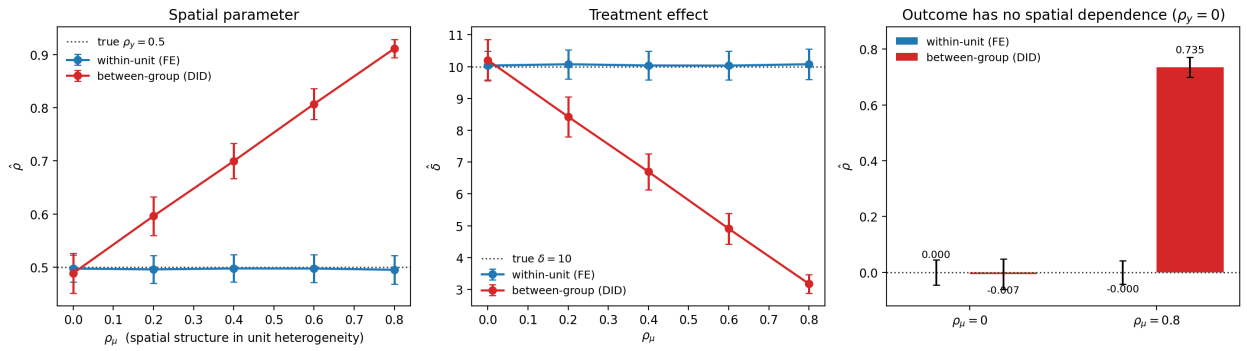


Figure 6: Monte Carlo, five hundred replicates per cell. Left: the estimated spatial parameter against the spatial structure of unit heterogeneity, with the true ρ_y marked. Center: the estimated treatment effect against the true $\delta = 10$. Right: the same estimators where the outcome carries no spatial dependence at all, $\rho_y = 0$; the within-unit estimate is zero to three decimal places while the between-group estimate is 0.735. Bars are one standard deviation across replicates.

a spatial autoregression returns when fitted to a persistent spatial field in levels, as Müller & Watson (2024) describe, and not evidence that the design is absorbing its own omission. Setting $\rho_y = 0$ removes the field. No spatial dependence remains in the outcome to detect, and the within-unit design correctly reports none. $\hat{\rho}_{FE}$ is zero to three decimal places. The between-group design reports 0.735, and recovers a treatment effect of 3.56 against a truth of 10. With no outcome process to measure, that parameter is measuring the heterogeneity its own conditioning failed to remove.

5.2 Testing the Condition on the Estimation Sample

The simulation switches the mechanism on and off by construction. The same condition can be checked directly on the data from which the estimates come.

Equation 12 concerns the within-group component of unit heterogeneity, $\hat{\mu}_i - \bar{\hat{\mu}}_{g(i)}$. For a balanced two-period panel in which treatment is constant within group, that quantity has a convenient closed form: the unit fixed effect is $\bar{y}_{i\cdot}$ less a term proportional to D_i, and because D_i is constant within treatment group the second term is annihilated by the group demeaning. The diagnostic therefore reduces to the unit means demeaned within treatment group, requires no fit, and does not depend on $\hat{\delta}$, on $\hat{\rho}$, or on any spatial model. Table 5 reports it, together with both sides of Equation 12 evaluated exactly at $\hat{\rho}_{FE}$.

Table 5: Testing the identifying condition on the estimation sample. Moran's I is computed on the within-group deviations of the estimated unit effects, which for a balanced two-period panel with a group-constant treatment reduce to the unit means demeaned within treatment group. LHS and RHS are the two sides of the condition evaluated exactly at the within-unit estimate. The condition predicts a larger between-group estimate where LHS exceeds RHS; it does so in all 16 of 16 cells. The margin is informative beyond its sign: its correlation with the realized gap is 0.934 in logs, and the smallest predicted margin (0.0006) falls in the same cell as the smallest realized gap (0.001).

	Band-			Condition			
Units	width (m)	W	Moran I	LHS	RHS	Margin	$\hat{\rho}_{DID} - \hat{\rho}_{FE}$
Studio	2000	Band	0.527	0.4338	0.0120	0.4218	0.345
Studio	2000	Kernel	0.600	0.3910	0.0255	0.3656	0.207
Studio	4000	Band	0.447	0.4081	0.0020	0.4062	0.592
Studio	4000	Kernel	0.484	0.2024	0.0195	0.1828	0.122
1 BR	2000	Band	0.342	0.0935	0.0207	0.0729	0.058
1 BR	2000	Kernel	0.440	0.1214	0.0430	0.0784	0.027
1 BR	4000	Band	0.219	0.0125	0.0119	0.0006	0.001
1 BR	4000	Kernel	0.277	0.0905	0.0217	0.0689	0.025
2 BR	2000	Band	0.374	0.1531	0.0081	0.1450	0.176
2 BR	2000	Kernel	0.469	0.2241	0.0148	0.2093	0.128
2 BR	4000	Band	0.206	0.0771	0.0024	0.0747	0.203
2 BR	4000	Kernel	0.302	0.1260	0.0069	0.1191	0.103
3 BR	2000	Band	0.518	0.4173	0.0148	0.4026	0.361
3 BR	2000	Kernel	0.624	0.5899	0.0209	0.5690	0.350
3 BR	4000	Band	0.286	0.1590	0.0077	0.1513	0.308
3 BR	4000	Kernel	0.446	0.3070	0.0158	0.2913	0.255

The within-group deviations are strongly spatially autocorrelated in every cell, with Moran's I between 0.206 and 0.624. A treatment defined by a spatial buffer produces exactly this. The

group indicator absorbs the mean difference between the inside and the outside of the buffer, and the finer spatial variation within each region passes through untouched. The right-hand side of Equation 12 is the positive threshold of Equation 13. It runs from 0.002 to 0.043, small but nonzero.

The condition is satisfied in all sixteen cells, so the derivation anticipates the ordering reported in the next section. The margin says more than the sign. The correlation between LHS−RHS and the realized gap $\hat{\rho}_{DID} - \hat{\rho}_{FE}$ is 0.79, and 0.93 in logs. Put differently, the condition predicts not only *that* the designs will diverge but *by how much*. The smallest predicted margin in the table falls in the one-bedroom subgroup at a four-kilometer bandwidth under the distance band, 0.0006, and the realized gap is correspondingly the smallest, 0.001, the single comparison in the entire analysis close to a tie.

5.3 The Designs Diverge in One Direction

Once the spatial lag enters, the equivalence breaks. Table 6 reports the spatial parameter for both designs across every combination of bandwidth, connectivity graph, spatial specification, level of aggregation and bedroom subgroup, two hundred and forty-eight fitted models in all, yielding one hundred and twenty-four paired comparisons.

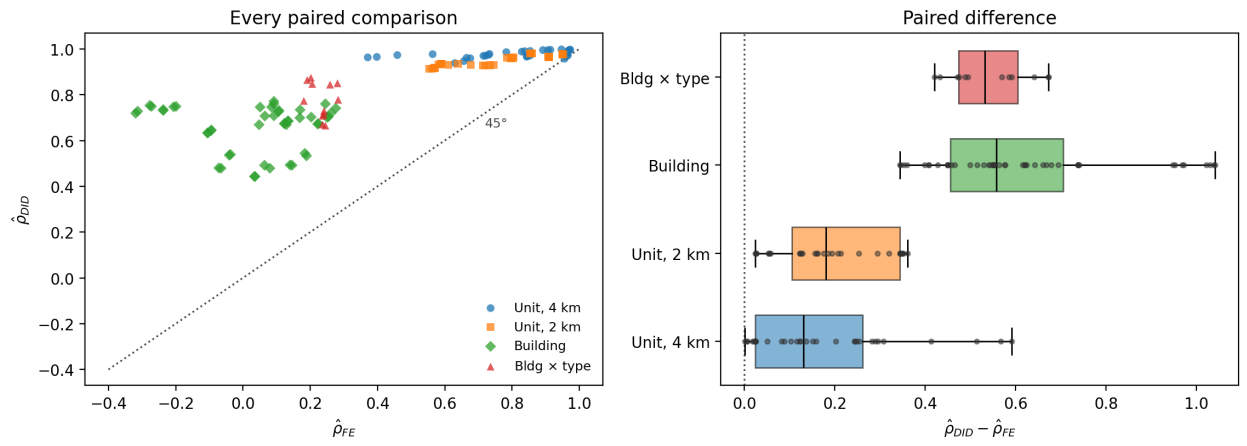


Figure 7: Paired comparisons of the spatial parameter. Left: each of the 124 comparisons, the within-unit estimate against the between-group estimate; every point lies above the 45-degree line. Right: the distribution of the paired difference by level of aggregation. Full estimates are in Table 6.

The DID design returns the larger spatial parameter in one hundred and twenty-four of one hundred and twenty-four comparisons. The ordering holds at both bandwidths, under a binary distance band, a triangular decay kernel and a nearest-neighbor graph, at all three levels of aggregation, across all four spatial specifications, for the spatial lag parameter ρ and the spatial error parameter λ alike, and in every bedroom subgroup. Wide margins carry the count. Table 6 reports posterior standard deviations alongside the means, and the gap exceeds twice the pooled posterior standard deviation in one hundred and twenty of the one hundred and twenty-four comparisons, at a median of 5.2 pooled standard deviations. The four exceptions share a structure. In each the within-unit estimate is itself pinned near unity, between 0.956 and 0.966, so the gap has no room to open; three are the one-bedroom subgroup at a four-kilometer bandwidth and the fourth a studio fit under the kernel. Where the FE design escapes the boundary the separation is unambiguous; where it sits against the boundary, neither design delivers an estimate conventional inference can use.

The magnitude matters as much as the direction. At unit level the DID estimates occupy a narrow band immediately below unity, 0.938 to 0.998 at a four-kilometer bandwidth, 0.914 to 0.981 at two kilometers. The FE estimates range far more widely, from 0.371 to 0.973. The DID design places the estimate at the boundary of the stationary region regardless of specification. Müller & Watson (2024) would call a parameter that close to unity a spatial unit root: the regression is spurious, and neither clustered nor spatial HAC standard errors repair it. The DID estimates fail that test and most of the FE estimates pass it. The design most closely associated with credible causal identification is the one that drives the estimated process there.

The older objection is broader, and the FE estimates fail it too. Anselin & Griffith (1988) asked whether spatial effects really matter in regression analysis and answered that they matter at ordinary magnitudes, not only at extreme ones. A lag of 0.371 still routes a shock through the system and returns it; an error parameter of that size still delivers standard errors confident about the wrong quantity. The smallest estimate in the table therefore condemns an aspatial specification on these data as surely as the largest does. Applied economics has spent four decades ignoring that warning and now receives the unit-root result as news.

Equation 9 gives the mechanism. The DID specification conditions on group means, leaving the within-group component of unit heterogeneity in the composite error; where treatment is assigned by proximity that remainder is spatially structured by construction, and the spatial lag absorbs it. The spatial parameter the DID design reports is largely that confounder.

5.4 The Consequence for Policy

Because Equation 8 holds only at a common ρ, a divergence of this size disturbs the treatment effect. The disagreement is not, however, about where the effect lies. Under SAR and SEM the two designs put it in nearly the same place. The posterior means agree in sign in sixty of sixty-two comparisons and differ by less than one DID posterior standard deviation in sixty-one. Table 7 reports the estimates.

They disagree instead about whether the effect can be determined at all. The DID posterior is a median 4.1 times wider than the FE posterior, and never narrower; the ratio ranges from 2.7 to 9.5. The consequence is a reversal of what the two designs can support. The FE posterior excludes zero in thirty-one of sixty-two comparisons, the DID posterior in five. This is Equation 9 read as a statement about variance instead of bias. The precision the within-unit design spends on the treatment effect, the between-group design spends on ρ.

The Durbin specifications, on which the paper's argument is built, carry the point, because there the policy-relevant quantity is not a single coefficient. Where treatment is defined by location, a unit's own exposure D and its neighbors' exposure WD are nearly the same vector, their correlation running from 0.68 to 0.98 across the graphs used here, a variance inflation factor as high as 31. At that degree of collinearity δ and θ are not separately identified. Their sum, the total local effect of the announcement on a treated unit and its surroundings, is the quantity the data identify. Table 8 gives the decomposition.

Under the within-unit design the coefficient sum behaves. Across the sixteen unit-level SDM specifications, spanning two bandwidths, two conceptions of proximity and four bedroom subgroups, it averages \$7.63 per month with a standard deviation of \$6.14, is positive in fifteen, and never exceeds \$21.18. The two terms composing it, meanwhile, range over \$120 and swing sign with the bandwidth. Under the between-group design it does not. It averages −\$6.49 with a standard deviation of \$31.98, ranges from −\$81.41 to \$45.19, and is positive in

eight of sixteen, an even split. Put differently, the design is uninformative about the sign of the quantity it was chosen to identify.

Under a global specification, however, the coefficient sum is not the effect. For a row-standardized W, $(I - \rho W)^{-1}\iota = \iota/(1-\rho)$ and $W\iota = \iota$, so the average cumulative impact under SDM is exactly $(\delta + \theta)/(1-\rho)$, and the multiplier inherits every bit of $\hat{\rho}$'s sensitivity. The same sixteen unit-level specifications whose coefficient sum sits within a six-dollar band imply cumulative impacts averaging $47.6 with a standard deviation of $46.7. The apparent precision belongs to a quantity other than the one against which policy is sized.

The substantive answer has to be read at the building level, for a reason internal to the argument. Co-location inflates $\hat{\rho}$ at unit level, so unit-level multipliers inherit the platial contamination along with it. Once each building contributes a single observation, $\hat{\rho}_{FE}$ falls to approximately zero, the multiplier falls to between 0.76 and 1.29, and coefficient sum and cumulative impact nearly coincide. The within-unit design puts the impact at $11.07 with a standard deviation of $20.24 across twenty-four specifications, a range from −$28.8 to $37.9 that comfortably includes zero. The between-group design puts it at $48.67 with a standard deviation of $176.96 and a range from −$359 to $365.

Thus the design gap is widest precisely in the quantity that matters. The cumulative impact propagates through $(I - \rho W)^{-1}$, which is convex in ρ. At the median unit-level fixed-effects estimate of $\hat{\rho} = 0.78$ the implied multiplier is below five; at the median between-group estimate of 0.96 it is twenty-eight, and at the largest it is two hundred. An analyst choosing the DID design would infer a land-value increment an order of magnitude larger than one choosing FE, from the same data, the same treatment definition and the same weights matrix, without either model appearing pathological on conventional diagnostics.

5.5 Separating Spatial Dependence from Co-location

The unit-level spatial parameters are large under both designs, and the design gap of Section 2.4 accounts for only part of that magnitude. The platial structure described in the methods accounts for the rest. Units sharing a building share a coordinate, and every conception of W treats them as maximally proximate. Estimating the same models on building-level aggregates removes that component, since each building contributes a single observation.

Aggregation changes the picture. Under the FE design the spatial parameter collapses, from 0.371 to 0.129 for studios, from 0.783 to 0.092 for two-bedroom units, and from 0.956 to −0.313 for one-bedroom units. Once co-located units are removed, the FE design finds little dependence between *buildings*, and for one-bedroom units a small negative value. Most of the dependence a unit-level spatial model reports is correlation among units within a single building.

The DID design behaves differently. Its spatial parameter falls but remains large, at 0.44 to 0.76 across subgroups and weighting schemes. On the reading of Equation 9, it continues to report between-building heterogeneity. That reading takes $\hat{\rho}_{FE}$ as the benchmark, and the data alone leave open which of the two estimates is displaced. An alternative account, taken up below, is that the FE estimate is attenuated and the DID estimate not inflated. Either way, the two designs differ by 0.4 to 0.8 on the same observations with the same weights.

The building-level comparison also disposes of the most natural objection to the unit-level results, that the divergence is an artifact of a weights matrix that connects near-duplicates. Aggregation removes that artifact. The divergence persists: approximately zero against approximately 0.7. Two distinct sources inflate the spatial parameter at unit level, one platial

and one arising from the design, and the two-level comparison switches the first off while leaving the second untouched.

At building level the KNN graph becomes available, since buildings are distinct points and the co-location problem that makes a nearest-neighbor graph degenerate at unit level does not arise. Its mean degree is six, against eighty to one hundred for the banded graphs, an order of magnitude sparser. It returns the same answer: DID near 0.7, FE near zero. Under row standardization $\hat{\rho}$ need not move with graph density, and its insensitivity to degree is a documented property of the SAR model. The *gap between the designs* persists across that change, and across the changes in aggregation and in the conception of proximity. Whatever the DID parameter is measuring, every weights specification finds it.

One substantive alternative to that reading remains. With two periods, $\hat{\rho}_{FE}$ is identified from spatial correlation in rent *changes*, while $\hat{\rho}_{DID}$ is identified from correlation in levels net of group and period means. These are different objects, and rent changes are the noisier of the two, since lease timing, concessions, rounding and administratively set rents all attenuate them. The FE estimates may therefore be attenuated and the DID estimates not inflated, and the negative building-level values under FE (−0.313, −0.237 and −0.108 for one-bedroom units) are consistent with attenuation. The design gap stands either way. It does qualify the inference in the preceding section that most unit-level dependence is co-location: that conclusion reads $\hat{\rho}_{FE}$ at building level as the benchmark, and should be held with the corresponding caution.

5.6 Pooling the Submarkets, and Controlling the Between-Group Design

Fitting separate models by bedroom type assumes away spillover between types: as specified above, a two-bedroom unit's neighbors are only other two-bedroom units, and a one-bedroom unit across the hall is not a neighbor at all. Relaxing that assumption also permits a fairer test of the between-group design. Pooling the unit types into a single model at the building × bedroom-type level admits cross-type co-location, recovers the four-bedroom units that segmentation discarded, and, because the unit fixed effect absorbs bedroom count while group means do not, requires the DID specification to carry an explicit bedroom-count control. The between-group design comes closer here than anywhere else to having what it needs.

The gap persists. Across the three connectivity graphs and all four specifications the fixed-effects spatial parameter is 0.181 to 0.283, and the between-group parameter 0.665 to 0.874, a divergence of 0.42 to 0.68, as large as anywhere in the analysis. Supplying an observable control left the gap unchanged, as Section 2.4 anticipates: no covariate list reaches the residual the spatial lag absorbs.

The treatment effects isolate the precision result, because here the two designs are estimating the same small number. Under the FE design the estimate is \$5.77, \$6.50 and \$4.43 for the distance band, triangular kernel and KNN graph under SAR, against posterior standard deviations of \$6.5 to \$6.6. Under the DID design, on the same observations, the same treatment coding, the same weights and with the bedroom control it requires, the estimates are \$1.34, \$2.07 and −\$1.53, the same neighborhood, but the posterior standard deviations are \$28.0 to \$31.4. Under SEM the contrast is larger. FE runs \$5.21 to \$6.40 against standard deviations of \$7.3 to \$8.4, DID \$0.58 to \$7.50 against standard deviations of \$40.4 to \$72.8.

Neither design finds an announcement effect distinguishable from zero in the pooled specification, but only one is in a position to say so. The DID posterior is roughly four to nine times wider across the twelve pooled comparisons, the difference between reporting a small null result and reporting nothing at all. That gap arises with the observable control supplied, the sample identical and the weights identical. Only their conditioning sets differ.

5.7 Robustness to the Treatment Buffer

The half-mile buffer is held fixed everywhere above, so the divergence could in principle be a property of where the line was drawn, not of assigning treatment by distance at all. Redefining the treatment as a cumulative one-mile buffer tests that directly. Sample, graph, sampler and seed are held fixed, the SAR and SEM specifications are refit under both designs at both buffers, and units whose treatment status flips under the wider definition are dropped from both, so the two buffers are fit on identical observations. These are thirty-two further comparisons on a slightly reduced sample, reported separately from the hundred and twenty-four of Section 5.3. Widening the buffer moves 1,933 units from control to treatment, raising the treated share from 29 to 35 percent.

The divergence holds at every width. The median gap is 0.130 at half a mile and 0.135 at one mile, and the between-group design returns the larger spatial parameter in sixteen of sixteen comparisons at half a mile and fifteen of sixteen at one mile. Treatment is a function of location, and that dependence generates the structured omitted heterogeneity at any cut point.

The single reversal is instructive. For one-bedroom units under the distance band, the one-mile fit returns $\hat{\rho}_{FE} = 0.955$ against $\hat{\rho}_{DID} = 0.949$, a gap of -0.006 and less than half a pooled posterior standard deviation from zero. That cell also supplied the closest call at half a mile, and for the same reason: both designs sit against the upper boundary, where neither estimate is separated from the other or from unity. Equation 12 states a condition, not a theorem, and here the condition fails to bite.

5.8 Sensitivity to W Is a Property of the Design

LeSage & Pace (2014b) argue that estimates should be insensitive to the specification of W. The results here qualify that claim in a specific way. Under the FE design the treatment effect is stable across weighting schemes. For two-bedroom units at building level, the estimates are 18.00, 17.95 and 16.99 under the distance band, the triangular kernel and KNN respectively. Under the DID design the same quantity moves with the graph. The Durbin decomposition in Table 8 gives the range, where the total local effect $\delta + \theta$, which the within-unit design holds within a six-dollar band across all sixteen unit-level specifications, swings from $-\$81$ to \$45 under the between-group design and changes sign with the bandwidth.

The apparent sensitivity to W at unit level has the same source. A triangular kernel assigns maximal weight at distance zero and therefore concentrates weight on co-located units, raising the within-building share from roughly 19% to 29% for studios. The resulting movement in $\hat{\rho}$, from 0.371 to 0.844 for studios, measures the platial correlation the weights admit, not instability in the spatial estimator. At building level, where no co-located pairs remain, that same gap falls to 0.074.

Sensitivity to W in these data therefore has two identifiable causes: the treatment of co-located observations, and the DID design. It is not a general property of spatial estimation, and once both causes are addressed the FE estimates are stable across every conception of proximity considered.

5.9 What These Data Do and Do Not Identify

Three quantities remain, determined to very different degrees.

The split between a unit's own exposure and its neighbors' is not identified under either design. The treatment produces that collinearity, and no estimator can undo it. When treatment is assigned by location, D and WD are collinear to the degree measured here, 0.68 to 0.98

across these graphs, and at that degree the data do not separate δ from θ. Where a distance buffer produces collinearity of this order, a confidently signed direct effect reported alongside a confidently signed spillover is an arbitrary split of a sum.

The sum is identified, but only by the within-unit design. Its *coefficient* is small and stable at unit level, about eight dollars, within a six-dollar band across two bandwidths, two conceptions of proximity and four bedroom subgroups. That stability is a property of the coefficient and not of the effect, since the cumulative impact multiplies it by $1/(1-\hat{\rho})$ and $\hat{\rho}$ at unit level is inflated by co-location. Read where that inflation is removed, at the building level, the cumulative impact is \$11.07 with a standard deviation of \$20.24 and a range spanning zero. The announcement effect is small, and too imprecisely determined to size a policy.

The between-group design identifies a third thing. Its treatment estimate is too imprecise to sign, and its spatial parameter is pinned near unity by residual heterogeneity. The transit literature typically reports something stronger, and the preceding sections suggest why. Applied to a treatment defined in space and to observations that share locations, the between-group design returns a large and confidently signed spatial parameter, together with a Durbin decomposition that appears to separate direct from indirect effects when it has done nothing of the kind. Both are artifacts of the design and the data structure.

6 Discussion

The results carry an uncomfortable implication for the literature that motivated this paper. The applied studies combining spatial econometrics with a quasi-experimental design, among them Dubé et al. (2014), Diao et al. (2017), Chagas et al. (2016), Comber & Arribas-Bel (2017) and Bardaka et al. (2018), adopt between-group designs. If the mechanism in Section 2.4 operates generally, the spatial parameters those studies report are inflated by the group heterogeneity their designs leave in the error, and the indirect and total effects computed from them are inflated more than proportionately.

The claim is bounded. The magnitude of any inflation depends on the spatial structure of the omitted heterogeneity in each setting, which cannot be assessed from outside. Inside a given study it is a testable implication, and two cheap tests reach it. Refitting with unit fixed effects is the direct route: under the argument advanced here $\hat{\rho}$ should fall, and the direct effect should move with it. Moran's I on the unit means demeaned within treatment group is cheaper still and requires no refitting at all, as in Section 5.2. That statistic is available from the data alone, Equation 12 names it as decisive, and in these data its margin over the threshold predicts the size of the divergence at a correlation of 0.93 in logs. Both diagnostics require panel data.

The published analysis of these data by Peng et al. (2023) is the immediate case, and it is instructive because the authors note the difficulty of reconciling the result with theory. Reporting a positive announcement effect, they observe that one would expect none "before service begins because there would be no transportation cost savings or gains in household utility before service begins," and add that "one could even argue that disruption brought by construction would cause utility and willingness to pay to fall in the pre-service period" (Peng et al., 2023, p. 17). The results here supply a candidate explanation for the discrepancy they identify, but only one of the two results reaches them. Their specification is an aspatial difference-in-differences, so the mechanism of Section 2.4, a spatial parameter absorbing the design's omitted heterogeneity, does not describe their estimate, which carries no spatial parameter to absorb anything. Section 2.3 reaches them instead. Treatment is assigned by

a distance buffer, and the diagnostic of Section 5.2 shows the within-group component of unit heterogeneity in these very data to be strongly spatially autocorrelated, with Moran's I between 0.206 and 0.624 across every subgroup and weighting scheme. A between-group design cannot condition on it. The untreated trend therefore differs across the buffer by construction, and the aspatial estimator attributes that difference to the announcement.

The comparison rests on direct evidence, since both designs are fit here on the same sample, the same treatment coding and the same weights. The announcement effect the between-group design returns on these data fails under the within-unit design. On the evidence available, their estimate and the artifact documented here remain indistinguishable. A design that conditions on the unit instead of the group would distinguish them, and the comparison above supplies one.

The choice of W qualifies a standing position in the field. LeSage & Pace (2014b) argue that the weights matrix is the "biggest myth" in spatial econometrics because estimates are largely insensitive to it. Their position holds here, but conditionally: for well-specified designs on data without a platial hierarchy. The conditional form is the more useful one, since it tells an analyst when to expect exceptions.

The field offers least guidance on that hierarchy. Classical spatial econometrics developed on polygon lattices, where each observation occupies a distinct areal unit and co-location does not arise. Modern administrative microdata routinely violate that premise: apartment units, business establishments, repeat transactions at one address, individuals in the same household. The distinction Wolf et al. (2021) draw between platial and spatial dependence then collapses at the level of W, since co-located observations sit at distance zero under any conception of proximity, and the connectivity graphs required differ from the polygon-contiguity structures on which the field was built (Corrado & Fingleton, 2012). Multilevel spatial models (Dong & Harris, 2015) are the obvious direction, and the magnitudes here suggest the problem is serious: between a tenth and a third of each unit's spatial weight falls inside its own building, and removing that component moves the estimated spatial parameter from as high as 0.97 to approximately zero.

Several limitations bound these conclusions, and the first is about what kind of result this is. The evidence is a demonstration, not a proof of impossibility. The Monte Carlo covers one lattice, one weights matrix and two values of ρ_y; the application covers one county and two periods, and its one hundred and twenty-four comparisons are nested on overlapping samples, so they are not one hundred and twenty-four independent tests. The exercise establishes that the mechanism operates, that it operates in one direction, and that its magnitude tracks a statistic computable in advance. It leaves open whether some between-group design might recover a spatially defined treatment effect.

The panel has two periods, which makes the FE/DID equivalence exact and therefore makes the divergence attributable to spatial structure alone. That same brevity leaves no pre-treatment period against which to test parallel trends, and it blocks the shock channel of Section 2.3 entirely. The balanced panel excludes units entering or leaving the rental stock, and if entry and exit are themselves responses to the announcement, the estimated effect is conditioned on survival. Building-level estimation removes the co-location confound at the cost of sample size, leaving the treatment effect imprecisely determined. This design cannot distinguish the substance of the small building-level effect from the power surrendered to aggregation. A multilevel spatial panel would separate them, and building one for a two-period design is beyond what this paper attempts. The SUTVA violation that spillovers impose stands under either design.

7 Conclusion

This paper began from a design question. For an unstaggered treatment in a balanced two-period panel, the difference-in-differences and fixed-effects designs return identical estimates, and applied work reasonably treats the choice between them as immaterial. I have shown that this equivalence is a property of the aspatial case. Once the outcome is spatially dependent the two designs separate, because the spatial parameter is identified from exactly the residual variation that distinguishes conditioning on unit effects from conditioning on group means. Across one hundred and twenty-four comparisons, spanning two bandwidths, three conceptions of proximity, three levels of aggregation, every bedroom subgroup, and all four specifications (SAR, SEM, SDM, SDEM), the between-group design returns the larger spatial parameter every time. At the unit level it settles against the upper boundary of the stationary region, where the process is nearly a spatial unit root and regression against it is spurious (Müller & Watson, 2024). Because cumulative impacts propagate through a matrix inverse that is convex in that parameter, the gap in $\hat{\rho}$ implies land-value increments differing by an order of magnitude.

A second result concerns the data structure. Apartment units sharing a building share a coordinate, so every conception of W treats them as maximally proximate, and a single-level spatial model cannot separate dependence among units in the same place from dependence among units in different places. Estimating at the building level separates the two: the fixed-effects spatial parameter falls from as high as 0.97 to approximately zero, and in one subgroup turns negative. The divergence between designs persists through this aggregation, so it belongs to the design itself and not to a weights matrix connecting near-duplicates.

The substantive question that motivated the analysis is answered, and the answer displaces the published one. Once both problems are addressed the announcement effect is small: a cumulative impact of roughly eleven dollars a month at the building level, with a spread that includes zero. Two features of that finding matter more than its magnitude. Only the within-unit design recovers it: the between-group design, given the same data, the same weights and an additional observable control, returns a quantity too imprecise to sign. It is also recoverable only as a sum. Because the treatment is a distance buffer, a unit's own exposure and its neighbors' exposure are nearly the same regressor, and the direct-versus-spillover decomposition that spatial Durbin models are usually fitted to produce is, for treatments of this kind, an arbitrary partition of a quantity identified only in total. A between-group design applied to a treatment defined in space, and to observations that share locations, returns exactly the large and confidently signed announcement effects the transit literature reports. They are products of the design and the data structure, and both fail under correction.

It may also simply be too early. Dubé et al. (2018) find that effects appear during construction, with the announcement itself passing unpriced, because the market "needs tangible proof that public policies will effectively be undertaken before the impact can be perceived." These models speak only to the announcement phase over a three-year window and imply nothing about the Purple Line once operational. Price effects are likely to differ substantially across the announcement, construction and operational phases, and I will extend the analysis as the line is built out and the rental survey accumulates. The absence of an anticipatory effect also says nothing about the construction period, when Dubé et al. (2018) would lead one to expect effects.

Gibbons & Overman (2012) urge applied researchers toward the "experimentalist paradigm," and a growing body of work has responded by merging causal-inference designs with spatial specifications. That merger requires more forethought than it is usually given:

the definition of the reference group, and the assumptions made about the nature and extent of spillovers, determine what the resulting estimate means, and in many cases only theory can guide those decisions.

None of this is new. Anselin & Griffith (1988) wrote that ignoring spatial dependence "may lead to serious errors in the interpretation of familiar regression diagnostics and misspecification tests," located the obstacle in dissemination, not in theory, and called for an end to subfields developing "in parallel, rather than being complementary." The diagnostics that call produced have been available since: Lagrange Multiplier tests for spatial dependence in regression residuals, in their original and robust forms, and their more recent Bayesian analogues (Anselin, 1988; Anselin et al., 1996), all of which this paper uses. Nearly four decades later, Müller & Watson (2024) derive the corresponding apparatus for strongly persistent spatial processes from inside econometrics, and the result is received as (award winning) news[13]. The second literature reached the narrower prescription. Anselin & Griffith (1988) asked whether spatial effects matter at ordinary magnitudes and answered that they do; the unit-root result concerns the case where the process is barely estimable at all. The broader claim traveled so little that the narrower one arrived as news. Applied work proceeded in the interval on designs whose failure modes were documented before those designs became canonical. This paper is one more instance of that cost, not an exception to it.

Data are always located in both time and space. Analysts who apply a classic quasi-experimental design to geo-located data without considering spatial effects should proceed with caution, particularly in housing studies and in work concerned with spillover effects. Three decades of spatial econometric research establish that such designs may attribute spillover to a treatment, and that they cannot recover the cumulative impact of a shock, which is composed of direct and indirect effects accruing through spatial interaction. The present analysis adds the converse: a spatial model alone leaves the problem in place, because the causal design determines what the spatial parameter measures, and the structure of the data determines whether that parameter is measuring space at all. In the context of transit investment, where the estimated increment is what land-value-capture and anti-displacement policy would be sized against, either error is expensive. Bridging causal inference and spatial econometrics requires attention to both.

8 References

Aaronson, D., Faber, J., Hartley, D., Mazumder, B., & Sharkey, P. (2021). The long-run effects of the 1930s HOLC "redlining" maps on place-based measures of economic opportunity and socioeconomic success. *Regional Science and Urban Economics*, *86*, 103622. https://doi.org/10.1016/j.regsciurbeco.2020.103622

Akbari, K., Winter, S., & Tomko, M. (2023). Spatial Causality: A Systematic Review on Spatial Causal Inference. *Geographical Analysis*, *55*(1), 56–89. https://doi.org/10.1111/gean.12312

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*.

Anselin, L. (1988). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*, *20*(1), 1–17. https://doi.org/10.1111/j.1538-4632.1988.tb00159.x

[13] https://www.econometricsociety.org/publications/econometrica/2024/09/01/Spatial-Unit-Roots-and-Spurious-Regression

Anselin, L. (1989). What is special about spatial data?: {Alternative} perspectives on spatial data analysis. *Santa Barbara, CA, NCGIA Report*, 84–89. https://escholarship.org/uc/item/3ph5k0d4

Anselin, L. (2002). Under the hood Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, *27*(3), 247–267. https://doi.org/10.1016/S0169-5150(02)00077-4

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, *26*(2), 156–166. https://doi.org/10.1177/0160017602250972

Anselin, L., & Amaral, P. (2023). Endogenous spatial regimes. *Journal of Geographical Systems*. https://doi.org/10.1007/s10109-023-00411-2

Anselin, L., & Arribas-Bel, D. (2013). Spatial fixed effects and spatial dependence in a single cross-section: Spatial fixed effects and spatial dependence. *Papers in Regional Science*, *92*(1), 3–17. https://doi.org/10.1111/j.1435-5957.2012.00480.x

Anselin, L., Bera, A., Florax, R. J. G. M., & Yoon, M. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, *26*, 77–104. https://doi.org/10.1016/0166-0462(95)02111-6

Anselin, L., Gallo, J. L., & Jayet, H. (2008). Spatial Panel Econometrics. In *The Econometrics of Panel Data* (Vol. 46, pp. 625–660). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75892-1_19

Anselin, L., & Griffith, D. A. (1988). Do spatial effects really matter in regression analysis? *Papers Regional Science*, *65*, 11–34. https://doi.org/10.1111/j.1435-5597.1988.tb01155.x

Bardaka, E., Delgado, M. S., & Florax, R. J. G. M. (2018). Causal identification of transit-induced gentrification and spatial spillover effects: The case of the Denver light rail. *Journal of Transport Geography*, *71*, 15–31. https://doi.org/10.1016/j.jtrangeo.2018.06.025

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Can, A., & Megbolugbe, I. (1997). Spatial dependence and house price index construction. *The Journal of Real Estate Finance and Economics*, *14*(1–2), 203–222. https://doi.org/10.1023/A:1007744706720

Chagas, A. L. S., Azzoni, C. R., & Almeida, A. N. (2016). A spatial difference-in-differences analysis of the impact of sugarcane production on respiratory diseases. *Regional Science and Urban Economics*, *59*, 24–36. https://doi.org/10.1016/j.regsciurbeco.2016.04.002

Comber, S., & Arribas-Bel, D. (2017). "Waiting on the train": The anticipatory (causal) effects of Crossrail in Ealing. *Journal of Transport Geography*, *64*, 13–22. https://doi.org/10.1016/j.jtrangeo.2017.08.004

Corrado, L., & Fingleton, B. (2012). WHERE IS THE ECONOMICS IN SPATIAL ECONOMETRICS?*. *Journal of Regional Science*, *52*(2), 210–239. https://doi.org/10.1111/j.1467-9787.2011.00726.x

Cunningham, S. (2018). *Causal Inference: The Mixtape (v1.7)*. http://www.apache.org/licenses/

de Chaisemartin, C., & D'Haultfœuille, X. (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, utac017. https://doi.org/10.1093/ectj/utac017

Delgado, M. S., & Florax, R. J. G. M. (2015). Difference-In-Differences Techniques for Spatial Data: Local Autocorrelation and Spatial Interaction. *Economics Letters*, *137*, 123–126. https://doi.org/10.1016/j.econlet.2015.10.035

Delmelle, E. C., Nilsson, I., & Bryant, A. (2020). Investigating Transit-Induced Displacement Using Eviction Data. *Housing Policy Debate*, *00*(00), 1–16. https://doi.org/10.1080/10511482.2020.1815071

Diao, M., Leonard, D., & Sing, T. F. (2017). 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. https://doi.org/10.1016/j.regsciurbeco.2017.08.006

Dong, G., & Harris, R. (2015). Spatial Autoregressive Models for Geographically Hierarchical Data Structures: Hierarchical Spatial Autoregressive Models. *Geographical Analysis*, *47*(2), 173–191. https://doi.org/10.1111/gean.12049

Dubé, J., Legros, D., & Devaux, N. (2018). From bus to tramway: Is there an economic impact of substituting a rapid mass transit system? An empirical investigation accounting for anticipation effect. *Transportation Research Part A: Policy and Practice*, *110*, 73–87. https://doi.org/10.1016/j.tra.2018.02.007

Dubé, J., Legros, D., Thériault, M., & Des Rosiers, F. (2014). A spatial Difference-in-Differences estimator to evaluate the effect of change in public mass transit systems on house prices. *Transportation Research Part B: Methodological*, *64*, 24–40. https://doi.org/10.1016/j.trb.2014.02.007

Durlauf, S. N., & Johnson, P. A. (1995). Multiple regimes and cross-country growth behavior. *Journal of Applied Econometrics*, *10*, 365–384. https://doi.org/10.1002/jae.3950100404

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, *26*(3), 244–268. https://doi.org/10.1177/0160017603253791

Elhorst, J. P. (2010). Applied Spatial Econometrics: Raising the Bar. *Spatial Economic Analysis*, *5*(1), 9–28. https://doi.org/10.1080/17421770903541772

Elhorst, J. P. (2014). Spatial Panel Models. In *Handbook of Regional Science* (pp. 1637–1652). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-23430-9_86

Elhorst, J. P., & Fréret, S. (2009). EVIDENCE OF POLITICAL YARDSTICK COMPETITION IN FRANCE USING A TWO-REGIME SPATIAL DURBIN MODEL WITH FIXED EFFECTS. *Journal of Regional Science*, *49*(5), 931–951. https://doi.org/10.1111/j.1467-9787.2009.00613.x

Elhorst, J. P., Lacombe, D. J., & Piras, G. (2012). On model specification and parameter space definitions in higher order spatial econometric models. *Regional Science and Urban Economics*, *42*(1–2), 211–220. https://doi.org/10.1016/j.regsciurbeco.2011.09.003

Ertur, C., Le Gallo, J., & Baumont, C. (2006). The European Regional Convergence Process, 1980-1995: Do Spatial Regimes and Spatial Dependence Matter? *International Regional Science Review*, *29*(1), 3–34. https://doi.org/10.1177/0160017605279453

Freeman, A. M. (1979). Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues. *The Scandinavian Journal of Economics*, *81*(2), 154. https://doi.org/10.2307/3439957

Getis, A. (2007). Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, *37*(4), 491–496. https://doi.org/10.1016/j.regsciurbeco.2007.04.005

Gibbons, S., & Overman, H. G. (2012). Mostly Pointless Spatial Econometrics? *Journal of Regional Science*, *52*(2), 172–191. https://doi.org/10.1111/j.1467-9787.2012.00760.x

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

Halleck Vega, S., & Elhorst, J. P. (2015). THE SLX MODEL. *Journal of Regional Science*, *55*(3), 339–363. https://doi.org/10.1111/jors.12188

Halloran, M. E., & Struchiner, C. J. (1995). Causal Inference in Infectious Diseases: *Epidemiology*, *6*(2), 142–151. https://doi.org/10.1097/00001648-199503000-00010

Harrison, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, *5*(1), 81–102. https://doi.org/10.1016/0095-0696(78)90006-2

Huntington-Klein, N. (2022). *The effect: An introduction to research design and causality*. CRC Press, Taylor & Francis Group.

Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

Kim, G., Schieffer, J., & Mark, T. (2020). Do superfund sites affect local property values? Evidence from a spatial hedonic approach. *Economic Analysis and Policy*, *67*, 15–28. https://doi.org/10.1016/j.eap.2020.05.007

Kolak, M., & Anselin, L. (2020). A Spatial Perspective on the Econometrics of Program Evaluation. *International Regional Science Review*, *43*(1–2), 128–153. https://doi.org/10.1177/0160017619869781

Koschinsky, J. (2009). Spatial Heterogeneity in Spillover Effects of Assisted and Unassisted Rental Housing. *Journal of Urban Affairs*, *31*(3), 319–347. https://doi.org/10.1111/j.1467-9906.2009.00462.x

Lacombe, D. J. (2004). Does Econometric Methodology Matter? An Analysis of Public Policy Using Spatial Econometric Techniques. *Geographical Analysis*, *36*(2), 105–118. https://doi.org/10.1111/j.1538-4632.2004.tb01128.x

Lee, L., & Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, *154*(2), 165–185. https://doi.org/10.1016/j.jeconom.2009.08.001

Lee, L., & Yu, J. (2012). QML Estimation of Spatial Dynamic Panel Data Models with Time Varying Spatial Weights Matrices. *Spatial Economic Analysis*, *7*(1), 31–74. https://doi.org/10.1080/17421772.2011.647057

LeSage, J. P. (2014). What Regional Scientists Need to Know About Spatial Econometrics. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2420725

LeSage, J. P., & Pace, R. K. (2014a). Interpreting Spatial Econometric Models. In *Handbook of Regional Science* (pp. 1535–1552). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-23430-9_91

LeSage, J., & Pace, R. (2014b). The Biggest Myth in Spatial Econometrics. *Econometrics*, *2*(4), 217–249. https://doi.org/10.3390/econometrics2040217

LeSage, J., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. CRC Press.

McMillen, D. (2018). *Spatial Variation in House Prices and the Opening of Chicago's Orange Line*. Lincoln Institute for Land Policy. https://www.lincolninst.edu/sites/default/files/sources/events/mcmillen_spatial_variation_in_house_prices_and_the_opening_of_chicagos_orange_line.pdf

Müller, U. K., & Watson, M. W. (2024). Spatial Unit Roots and Spurious Regression. *Econometrica*, *92*(5), 1661–1695. https://doi.org/10.3982/ECTA21654

Paelinck, J. H. P. (2009). Rook or Queen, But How About Some Other Royal Court Suite Members?: Rook or Queen, But How About Some Other Royal Court Suite Members? *Geographical Analysis*, *41*(4), 425–429. https://doi.org/10.1111/j.1538-4632.2009.00771.x

Peng, Q., Knaap, G., & Finio, N. (2023). Do Multifamily unit Rents Increase in Response to Light Rail in the Pre-service Period? *International Regional Science Review*, 016001762311625. https://doi.org/10.1177/01600176231162563

Quast, T., Gonzalez, F., & Ziemba, R. (2017). Abortion facility closings and abortion rates in texas. *Inquiry (United States)*, *54*. https://doi.org/10.1177/0046958017700944

Rey, S. J., Anselin, L., Amaral, P., Arribas-Bel, D., Cortes, R. X., Gaboardi, J. D., Kang, W., Knaap, E., Li, Z., Lumnitz, S., Oshan, T. M., Shao, H., & Wolf, L. J. (2021). The PySAL Ecosystem: Philosophy and Implementation. *Geographical Analysis*, gean.12276. https://doi.org/10.1111/gean.12276

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, *82*(1), 34–55. https://doi.org/10.1086/260169

Sheng, Y., & LeSage, J. P. (2021). Interpreting spatial regression models with multiplicative interaction explanatory variables. *Journal of Geographical Systems*, *23*(3), 333–360. https://doi.org/10.1007/s10109-021-00356-4

Sobel, M. E. (2006). What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference. *Journal of the American Statistical Association*, *101*(476), 1398–1407. https://doi.org/10.1198/016214506000000636

Steimetz, S. S. C. (2010). SPATIAL MULTIPLIERS IN HEDONIC ANALYSIS: A COMMENT ON "SPATIAL HEDONIC MODELS OF AIRPORT NOISE, PROXIMITY, AND HOUSING PRICES"*. *Journal of Regional Science*, *50*(5), 995–998. https://doi.org/10.1111/j.1467-9787.2010.00679.x

Voith, R., Liu, J., Zielenbach, S., Jakabovics, A., An, B., Rodnyansky, S., Orlando, A. W., & Bostic, R. W. (2022). Effects of concentrated LIHTC development on surrounding house prices. *Journal of Housing Economics*, *56*, 101838. https://doi.org/10.1016/j.jhe.2022.101838

Wolf, L. J., Anselin, L., Arribas-Bel, D., & Mobley, L. R. (2021). On Spatial and Platial Dependence: Examining Shrinkage in Spatially Dependent Multilevel Models. *Annals of the American Association of Geographers*, 1–13. https://doi.org/10.1080/24694452.2020.1841602

Wolf, L. J., Oshan, T. M., & Fotheringham, A. S. (2017). Single and Multiscale Models of Process Spatial Heterogeneity. *Geographical Analysis*, 1–24. https://doi.org/10.1111/gean.12147

Wooldridge, J. M. (2021). Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3906345

Yu, L. L. and J. (2007). A Spatial Dynamic Panel Data Model with Both Time and Individual Fixed Effects. *Econometric Theory*.

9 Appendix

9.1 Full Estimation Results

Per-model posterior summaries for all two hundred and forty-eight fits, covering both designs, four spatial specifications, three levels of aggregation, three conceptions of proximity and two bandwidths, are distributed with the replication archive. Each row carries the treatment coefficient and its posterior standard deviation, the Durbin lag coefficient where the specification includes one, the spatial parameter with its own posterior standard deviation and highest-density interval, and the maximum $\hat{R}$ across all parameters in the fit.

9.2 Full Comparison Tables

The three tables below report every fitted model. At 124 paired comparisons they run to several pages, which is why they appear here rather than in the body; Figure 7 summarizes the first of them.

Table 6: Spatial parameter by causal design. ρ for the lag specifications (SAR, SDM), λ for the error specifications (SEM, SDEM). The between-group (DID) estimate exceeds the within-unit (FE) estimate in all 124 of 124 comparisons, across every specification, bandwidth and conception of proximity. Posterior standard deviations in parentheses. The gap exceeds twice the pooled posterior standard deviation in 120 of 124, so the count is not carried by close calls.

		Band-		SAR		SEM		SDM		SDEM	
Units	Level	width (m)	W	FE	DID	FE	DID	FE	DID	FE	DID
Studio	Unit	4000	Band	0.371 (0.123)	0.963 (0.012)	0.459 (0.116)	0.973 (0.011)	0.398 (0.122)	0.965 (0.012)	0.564 (0.198)	0.977 (0.009)
Studio	Unit	4000	Kernel	0.844 (0.043)	0.966 (0.010)	0.854 (0.042)	0.971 (0.009)	0.844 (0.043)	0.968 (0.009)	0.966 (0.011)	0.971 (0.009)
Studio	Unit	2000	Band	0.589 (0.046)	0.934 (0.010)	0.639 (0.044)	0.934 (0.010)	0.610 (0.045)	0.930 (0.011)	0.678 (0.050)	0.931 (0.011)
Studio	Unit	2000	Kernel	0.720 (0.035)	0.927 (0.010)	0.735 (0.034)	0.929 (0.010)	0.714 (0.035)	0.927 (0.010)	0.745 (0.037)	0.930 (0.010)
Studio	Building	4000	Band	0.129 (0.204)	0.669 (0.085)	0.262 (0.191)	0.720 (0.084)	0.048 (0.217)	0.669 (0.085)	0.170 (0.215)	0.734 (0.083)
Studio	Building	4000	Kernel	0.203 (0.174)	0.702 (0.073)	0.276 (0.163)	0.741 (0.071)	0.169 (0.180)	0.699 (0.074)	0.245 (0.183)	0.760 (0.070)
Studio	Building	4000	KNN	0.225 (0.112)	0.673 (0.059)	0.250 (0.114)	0.700 (0.058)	0.221 (0.113)	0.672 (0.058)	0.253 (0.114)	0.704 (0.058)
1 BR	Unit	4000	Band	0.956 (0.010)	0.957 (0.008)	0.964 (0.009)	0.982 (0.006)	0.960 (0.009)	0.967 (0.008)	0.966 (0.009)	0.988 (0.004)
1 BR	Unit	4000	Kernel	0.968 (0.007)	0.993 (0.002)	0.971 (0.007)	0.995 (0.002)	0.969 (0.007)	0.993 (0.002)	0.973 (0.007)	0.997 (0.001)
1 BR	Unit	2000	Band	0.908 (0.009)	0.966 (0.004)	0.910 (0.009)	0.966 (0.004)	0.909 (0.009)	0.964 (0.004)	0.910 (0.009)	0.964 (0.004)
1 BR	Unit	2000	Kernel	0.950 (0.006)	0.977 (0.003)	0.950 (0.006)	0.976 (0.003)	0.950 (0.006)	0.977 (0.003)	0.952 (0.006)	0.976 (0.003)
1 BR	Building	4000	Band	-0.313 (0.167)	0.728 (0.047)	-0.277 (0.168)	0.753 (0.045)	-0.319 (0.169)	0.719 (0.048)	-0.272 (0.171)	0.749 (0.047)
1 BR	Building	4000	Kernel	-0.237 (0.147)	0.735 (0.042)	-0.206 (0.147)	0.747 (0.041)	-0.237 (0.147)	0.731 (0.043)	-0.199 (0.149)	0.749 (0.042)
1 BR	Building	4000	KNN	-0.108 (0.078)	0.634 (0.036)	-0.098 (0.079)	0.644 (0.036)	-0.108 (0.078)	0.634 (0.037)	-0.098 (0.079)	0.647 (0.037)
2 BR	Unit	4000	Band	0.783 (0.030)	0.986 (0.004)	0.854 (0.026)	0.991 (0.004)	0.829 (0.028)	0.988 (0.004)	0.838 (0.028)	0.990 (0.004)
2 BR	Unit	4000	Kernel	0.892 (0.020)	0.995 (0.002)	0.914 (0.017)	0.996 (0.002)	0.907 (0.018)	0.995 (0.002)	0.947 (0.017)	0.998 (0.001)
2 BR	Unit	2000	Band	0.785 (0.020)	0.961 (0.004)	0.802 (0.019)	0.959 (0.005)	0.797 (0.019)	0.960 (0.005)	0.804 (0.020)	0.964 (0.004)
2 BR	Unit	2000	Kernel	0.853 (0.015)	0.981 (0.002)	0.857 (0.014)	0.981 (0.002)	0.855 (0.014)	0.981 (0.002)	0.859 (0.014)	0.981 (0.002)
2 BR	Building	4000	Band	0.092 (0.136)	0.708 (0.049)	0.106 (0.141)	0.726 (0.049)	0.065 (0.141)	0.707 (0.049)	0.107 (0.143)	0.731 (0.049)
2 BR	Building	4000	Kernel	0.085 (0.121)	0.746 (0.042)	0.092 (0.124)	0.760 (0.041)	0.051 (0.125)	0.746 (0.042)	0.092 (0.125)	0.771 (0.042)
2 BR	Building	4000	KNN	0.118 (0.071)	0.672 (0.032)	0.128 (0.071)	0.683 (0.032)	0.118 (0.071)	0.672 (0.032)	0.130 (0.071)	0.683 (0.032)
3 BR	Unit	4000	Band	0.630 (0.049)	0.938 (0.014)	0.665 (0.048)	0.961 (0.011)	0.657 (0.049)	0.947 (0.012)	0.676 (0.049)	0.959 (0.011)
3 BR	Unit	4000	Kernel	0.715 (0.041)	0.970 (0.007)	0.730 (0.040)	0.974 (0.007)	0.721 (0.041)	0.970 (0.007)	0.733 (0.041)	0.978 (0.006)
3 BR	Unit	2000	Band	0.553 (0.038)	0.914 (0.009)	0.568 (0.039)	0.917 (0.009)	0.562 (0.038)	0.914 (0.009)	0.572 (0.038)	0.917 (0.009)
3 BR	Unit	2000	Kernel	0.582 (0.033)	0.932 (0.007)	0.589 (0.033)	0.934 (0.007)	0.586 (0.032)	0.932 (0.007)	0.591 (0.033)	0.935 (0.007)
3 BR	Building	4000	Band	0.140 (0.129)	0.493 (0.087)	0.189 (0.129)	0.533 (0.087)	0.146 (0.129)	0.492 (0.088)	0.184 (0.131)	0.544 (0.086)
3 BR	Building	4000	Kernel	0.034 (0.125)	0.443 (0.080)	0.080 (0.124)	0.479 (0.081)	0.036 (0.125)	0.443 (0.080)	0.064 (0.125)	0.492 (0.081)
3 BR	Building	4000	KNN	-0.072 (0.125)	0.480 (0.082)	-0.037 (0.126)	0.539 (0.082)	-0.065 (0.126)	0.479 (0.081)	-0.041 (0.128)	0.536 (0.083)
Pooled	Bldg × type	4000	Band	0.259 (0.110)	0.844 (0.029)	0.281 (0.111)	0.850 (0.029)	0.245 (0.113)	0.665 (0.052)	0.283 (0.113)	0.778 (0.048)
Pooled	Bldg × type	4000	Kernel	0.191 (0.092)	0.864 (0.024)	0.202 (0.093)	0.874 (0.023)	0.181 (0.093)	0.773 (0.036)	0.206 (0.093)	0.847 (0.031)
Pooled	Bldg × type	4000	KNN	0.241 (0.039)	0.671 (0.020)	0.245 (0.039)	0.730 (0.018)	0.242 (0.039)	0.707 (0.018)	0.245 (0.040)	0.717 (0.018)

Table 7: Treatment coefficient by causal design: posterior mean in dollars of monthly rent, posterior standard deviation in parentheses. Under SAR and SEM the two designs agree closely on where the effect lies – the gap is under one DID posterior standard deviation in 61 of 62 comparisons – and disagree sharply on how precisely it is determined, the DID posterior being a median 4.1 times wider. Under SDM and SDEM the reported coefficient is the direct term alone, one half of a near-collinear pair, and is not separately identified; the Durbin decomposition is reported separately.

		Band-		SAR		SEM		SDM		SDEM	
Units	Level	width (m)	*W*	FE	DID	FE	DID	FE	DID	FE	DID
Studio	Unit	4000	Band	16.9 (4.0)	14.1 (14.6)	21.9 (4.8)	23.2 (19.9)	24.1 (5.3)	22.9 (17.6)	21.8 (5.3)	23.1 (19.9)
Studio	Unit	4000	Kernel	11.8 (3.8)	11.1 (13.9)	17.6 (5.6)	16.5 (20.7)	16.6 (5.6)	31.2 (17.8)	-13.2 (7.3)	23.4 (23.4)
Studio	Unit	2000	Band	18.3 (3.8)	18.3 (13.4)	33.6 (5.2)	39.7 (18.5)	37.6 (5.2)	-8.8 (16.3)	37.8 (5.2)	56.3 (18.9)
Studio	Unit	2000	Kernel	2.7 (3.8)	-1.7 (12.7)	-4.2 (7.3)	-13.2 (27.9)	-14.9 (8.4)	-0.5 (22.7)	-37.0 (10.7)	5.7 (36.1)
Studio	Building	4000	Band	-25.1 (13.5)	-27.1 (77.8)	-31.4 (15.1)	-39.9 (88.5)	-40.4 (15.2)	-24.6 (84.5)	-40.6 (15.1)	-39.7 (89.1)
Studio	Building	4000	Kernel	-25.3 (13.4)	-25.8 (74.2)	-29.9 (14.7)	-34.9 (87.3)	-36.2 (15.8)	-49.1 (82.3)	-35.8 (15.8)	-40.9 (87.4)
Studio	Building	4000	KNN	-21.2 (13.3)	-13.5 (68.0)	-24.0 (15.6)	-23.3 (101.3)	-18.9 (20.2)	-40.9 (89.5)	-19.7 (20.3)	-24.3 (103.4)
1 BR	Unit	4000	Band	-14.1 (2.9)	-13.9 (8.8)	-44.0 (4.5)	-44.2 (13.7)	-44.5 (4.5)	84.9 (11.7)	-42.0 (4.5)	-55.9 (13.7)
1 BR	Unit	4000	Kernel	-5.0 (2.9)	-4.4 (8.2)	-25.2 (5.0)	-26.1 (14.5)	-25.6 (5.0)	36.5 (11.9)	-20.3 (5.6)	-110.7 (15.6)
1 BR	Unit	2000	Band	-3.0 (2.8)	-1.5 (7.9)	-11.5 (5.3)	-11.4 (15.0)	-11.9 (5.3)	-77.4 (12.2)	-11.2 (5.4)	-18.6 (15.0)
1 BR	Unit	2000	Kernel	2.2 (2.7)	3.1 (7.3)	2.9 (6.7)	3.4 (18.0)	2.0 (6.7)	-45.7 (14.0)	31.6 (9.1)	15.6 (22.0)
1 BR	Building	4000	Band	-8.0 (11.7)	-10.6 (45.0)	-6.8 (10.6)	-12.0 (55.8)	-11.9 (14.7)	46.7 (51.6)	-11.8 (14.8)	-3.3 (56.0)
1 BR	Building	4000	Kernel	-8.1 (11.7)	-11.3 (43.6)	-7.4 (10.8)	-9.9 (56.9)	-9.3 (15.7)	25.1 (52.3)	-9.6 (15.7)	-8.1 (58.5)
1 BR	Building	4000	KNN	-10.2 (11.8)	-2.0 (42.9)	-9.5 (10.9)	-1.5 (85.4)	3.7 (30.2)	-37.6 (86.0)	4.4 (29.8)	-52.7 (103.9)
2 BR	Unit	4000	Band	19.0 (2.4)	14.6 (8.8)	61.8 (4.7)	64.8 (18.8)	65.0 (4.8)	133.3 (14.9)	64.1 (4.8)	63.8 (18.8)
2 BR	Unit	4000	Kernel	12.8 (2.3)	10.2 (8.0)	56.3 (5.3)	56.3 (19.3)	57.1 (5.3)	96.8 (15.0)	47.9 (6.2)	-28.2 (20.9)
2 BR	Unit	2000	Band	13.0 (2.3)	7.8 (7.8)	47.2 (5.3)	49.5 (20.1)	50.7 (5.7)	-33.5 (15.5)	46.8 (5.9)	15.0 (20.5)
2 BR	Unit	2000	Kernel	6.5 (2.3)	2.1 (7.0)	28.9 (7.2)	26.8 (26.0)	28.8 (8.1)	-6.6 (19.5)	16.7 (10.0)	-34.0 (30.4)
2 BR	Building	4000	Band	18.0 (9.0)	11.7 (53.1)	18.2 (9.4)	13.7 (67.1)	12.3 (11.4)	23.4 (61.8)	12.3 (11.4)	15.2 (67.5)
2 BR	Building	4000	Kernel	18.0 (9.0)	10.0 (50.7)	18.0 (9.4)	10.5 (68.5)	7.4 (12.3)	-3.5 (62.4)	7.2 (12.4)	-8.3 (71.0)
2 BR	Building	4000	KNN	17.0 (8.9)	8.0 (46.8)	18.6 (9.9)	13.4 (93.7)	12.6 (21.3)	-94.3 (88.4)	11.4 (21.4)	-15.6 (105.1)
3 BR	Unit	4000	Band	27.5 (6.5)	16.7 (23.7)	54.5 (10.8)	59.6 (45.7)	57.0 (12.2)	250.2 (37.0)	59.1 (11.8)	43.4 (46.3)
3 BR	Unit	4000	Kernel	17.3 (6.5)	5.9 (20.2)	40.8 (13.7)	38.1 (54.4)	38.0 (16.8)	154.7 (42.0)	33.1 (19.5)	-129.1 (70.2)
3 BR	Unit	2000	Band	26.9 (6.5)	11.7 (19.7)	56.8 (12.3)	75.8 (63.3)	71.7 (20.9)	-2.7 (49.9)	84.1 (21.8)	39.5 (66.8)
3 BR	Unit	2000	Kernel	20.2 (6.4)	2.2 (17.3)	47.1 (13.7)	64.6 (85.7)	75.0 (34.1)	-44.0 (70.3)	79.4 (39.3)	-81.4 (105.9)
3 BR	Building	4000	Band	42.9 (22.0)	43.8 (124.4)	49.4 (24.5)	62.0 (158.0)	64.9 (30.2)	5.6 (151.9)	63.9 (30.0)	41.3 (164.1)
3 BR	Building	4000	Kernel	43.7 (22.1)	45.4 (124.4)	47.0 (23.7)	65.9 (162.8)	75.8 (34.0)	-45.0 (164.5)	75.4 (34.0)	-18.4 (179.5)
3 BR	Building	4000	KNN	45.9 (22.4)	33.2 (123.7)	43.6 (21.8)	60.2 (208.1)	70.5 (52.5)	-291.5 (230.2)	71.9 (52.7)	-142.8 (250.2)
Pooled	Bldg × type	4000	Band	5.8 (6.6)	1.3 (31.4)	5.2 (7.4)	0.6 (40.4)	0.9 (8.5)	0.1 (40.7)	0.6 (8.5)	-0.2 (40.0)
Pooled	Bldg × type	4000	Kernel	6.5 (6.6)	2.1 (29.7)	6.3 (7.3)	0.9 (40.8)	0.1 (9.2)	-0.2 (41.7)	-0.3 (9.3)	-2.1 (43.3)
Pooled	Bldg × type	4000	KNN	4.4 (6.5)	-1.5 (28.0)	6.4 (8.4)	7.5 (72.8)	32.0 (25.3)	32.0 (102.8)	29.9 (25.1)	28.2 (99.1)

Table 8: Durbin decomposition by causal design, in dollars of monthly rent: direct coefficient δ, lag coefficient θ, their sum, and the average cumulative impact. Because the treatment is defined by location, D and WD are nearly the same vector and the two coefficients are not separately identified: they take large offsetting values that swing across specifications, while under SDM their sum does not, at a standard deviation of 12.6 across the 31 within-unit specifications against 37 and 33 for the terms taken separately. The sum is not the effect, however. Under SDM the cumulative impact is $(\delta + \theta)/(1 - \rho)$ for a row-standardized W, and the multiplier inherits the sensitivity of $\hat{\rho}$: the apparent stability of the coefficient sum coexists with an impact averaging 30 dollars with a standard deviation of 40. SDEM carries no global multiplier, so there the sum is the total effect.

		Band-		Within-unit (FE)				Between-group (DID)			
Units	Level	width (m)	W	δ	θ	$\delta + \theta$	Impact	δ	θ	$\delta + \theta$	Impact
SDM											
Studio	Unit	4000	Band	24.1	-16.2	7.9	13.1	22.9	-19.4	3.4	98.3
Studio	Unit	4000	Kernel	16.6	-9.3	7.4	47.2	31.2	-38.3	-7.0	-220.2
Studio	Unit	2000	Band	37.6	-38.4	-0.8	-2.0	-8.8	54.0	45.2	645.6
Studio	Unit	2000	Kernel	-14.9	23.8	8.9	31.0	-0.5	-1.6	-2.1	-29.1
Studio	Building	4000	Band	-40.4	57.5	17.1	18.0	-24.6	-8.8	-33.4	-101.0
Studio	Building	4000	Kernel	-36.2	35.4	-0.8	-1.0	-49.1	73.7	24.6	81.8
Studio	Building	4000	KNN	-18.9	-3.5	-22.5	-28.8	-40.9	41.7	0.8	2.5
1 BR	Unit	4000	Band	-44.5	51.2	6.7	167.8	84.9	-166.4	-81.4	-2467.0
1 BR	Unit	4000	Kernel	-25.6	30.3	4.7	153.2	36.5	-60.3	-23.8	-3403.4
1 BR	Unit	2000	Band	-11.9	12.3	0.4	4.7	-77.4	104.7	27.3	757.1
1 BR	Unit	2000	Kernel	2.0	0.3	2.3	45.6	-45.7	60.5	14.7	641.1
1 BR	Building	4000	Band	-11.9	10.0	-1.9	-1.4	46.7	-147.5	-100.9	-359.0
1 BR	Building	4000	Kernel	-9.3	2.8	-6.5	-5.3	25.1	-83.6	-58.5	-217.5
1 BR	Building	4000	KNN	3.7	-16.2	-12.5	-11.3	-37.6	41.3	3.7	10.2
2 BR	Unit	4000	Band	65.0	-60.7	4.3	25.0	133.3	-153.4	-20.1	-1675.3
2 BR	Unit	4000	Kernel	57.1	-54.0	3.1	33.4	96.8	-104.6	-7.8	-1564.2
2 BR	Unit	2000	Band	50.7	-44.3	6.4	31.4	-33.5	48.1	14.6	365.7
2 BR	Unit	2000	Kernel	28.8	-24.5	4.2	29.3	-6.6	9.6	3.0	155.8
2 BR	Building	4000	Band	12.3	14.7	26.9	28.8	23.4	-28.3	-4.8	-16.5
2 BR	Building	4000	Kernel	7.4	23.5	31.0	32.6	-3.5	28.8	25.3	99.5
2 BR	Building	4000	KNN	12.6	5.2	17.8	20.2	-94.3	121.5	27.2	82.8
3 BR	Unit	4000	Band	57.0	-41.1	15.9	46.2	250.2	-316.2	-66.0	-1244.7
3 BR	Unit	4000	Kernel	38.0	-24.9	13.1	46.9	154.7	-177.1	-22.4	-745.1
3 BR	Unit	2000	Band	71.7	-50.5	21.2	48.4	-2.7	16.1	13.4	156.1
3 BR	Unit	2000	Kernel	75.0	-58.5	16.4	39.7	-44.0	49.3	5.2	76.9
3 BR	Building	4000	Band	64.9	-43.7	21.1	24.8	5.6	75.9	81.5	160.5
3 BR	Building	4000	Kernel	75.8	-55.9	19.9	20.6	-45.0	157.6	112.6	202.2
3 BR	Building	4000	KNN	70.5	-30.1	40.3	37.9	-291.5	395.2	103.7	199.0
Pooled	Bldg × type	4000	Band	0.9	12.0	12.8	17.0	0.1	5.6	5.7	17.0
Pooled	Bldg × type	4000	Kernel	0.1	13.6	13.8	16.8	-0.2	5.7	5.5	24.1
Pooled	Bldg × type	4000	KNN	32.0	-29.2	2.7	3.6	32.0	-36.2	-4.2	-14.2

continued on next page

		Band-		Within-unit (FE)				Between-group (DID)			
Units	Level	width (m)	W	δ	θ	$\delta + \theta$	Impact	δ	θ	$\delta + \theta$	Impact
SDEM											
Studio	Unit	4000	Band	21.8	26.5	48.3	48.3	23.1	-298.3	-275.2	-275.2
Studio	Unit	4000	Kernel	-13.2	396.8	383.6	383.6	23.4	-92.7	-69.3	-69.3
Studio	Unit	2000	Band	37.8	-70.5	-32.7	-32.7	56.3	315.2	371.5	371.5
Studio	Unit	2000	Kernel	-37.0	92.5	55.5	55.5	5.7	-83.0	-77.3	-77.3
Studio	Building	4000	Band	-40.6	57.2	16.6	16.6	-39.7	25.7	-14.0	-14.0
Studio	Building	4000	Kernel	-35.8	28.4	-7.4	-7.4	-40.9	405.7	364.8	364.8
Studio	Building	4000	KNN	-19.7	-8.4	-28.2	-28.2	-24.3	0.3	-23.9	-23.9
1 BR	Unit	4000	Band	-42.0	180.4	138.3	138.3	-55.9	-785.5	-841.5	-841.5
1 BR	Unit	4000	Kernel	-20.3	-71.3	-91.6	-91.6	-110.7	1232.2	1121.4	1121.4
1 BR	Unit	2000	Band	-11.2	-9.6	-20.8	-20.8	-18.6	670.7	652.1	652.1
1 BR	Unit	2000	Kernel	31.6	-150.9	-119.3	-119.3	15.6	-66.7	-51.1	-51.1
1 BR	Building	4000	Band	-11.8	9.7	-2.1	-2.1	-3.3	-301.5	-304.9	-304.9
1 BR	Building	4000	Kernel	-9.6	3.7	-5.9	-5.9	-8.1	-16.0	-24.1	-24.1
1 BR	Building	4000	KNN	4.4	-15.8	-11.4	-11.4	-52.7	106.3	53.6	53.6
2 BR	Unit	4000	Band	64.1	-27.8	36.3	36.3	63.8	-374.7	-310.8	-310.8
2 BR	Unit	4000	Kernel	47.9	115.5	163.4	163.4	-28.2	1284.6	1256.4	1256.4
2 BR	Unit	2000	Band	46.8	2.5	49.3	49.3	15.0	630.9	645.9	645.9
2 BR	Unit	2000	Kernel	16.7	32.4	49.1	49.1	-34.0	262.6	228.7	228.7
2 BR	Building	4000	Band	12.3	16.1	28.4	28.4	15.2	-41.3	-26.1	-26.1
2 BR	Building	4000	Kernel	7.2	25.7	32.9	32.9	-8.3	220.0	211.8	211.8
2 BR	Building	4000	KNN	11.4	9.0	20.4	20.4	-15.6	76.4	60.8	60.8
3 BR	Unit	4000	Band	59.1	-23.1	36.0	36.0	43.4	-298.9	-255.4	-255.4
3 BR	Unit	4000	Kernel	33.1	19.1	52.2	52.2	-129.1	976.1	847.1	847.1
3 BR	Unit	2000	Band	84.1	-40.7	43.4	43.4	39.5	217.3	256.8	256.8
3 BR	Unit	2000	Kernel	79.4	-38.9	40.6	40.6	-81.4	402.5	321.0	321.0
3 BR	Building	4000	Band	63.9	-36.3	27.6	27.6	41.3	144.4	185.7	185.7
3 BR	Building	4000	Kernel	75.4	-53.3	22.1	22.1	-18.4	309.7	291.2	291.2
3 BR	Building	4000	KNN	71.9	-34.2	37.7	37.7	-142.8	391.5	248.8	248.8
Pooled	Bldg × type	4000	Band	0.6	17.1	17.7	17.7	-0.2	17.8	17.6	17.6
Pooled	Bldg × type	4000	Kernel	-0.3	17.5	17.2	17.2	-2.1	40.0	37.9	37.9
Pooled	Bldg × type	4000	KNN	29.9	-25.9	4.0	4.0	28.2	-38.2	-10.0	-10.0