When Treatment is Defined by Distance: Challenges and Approaches

Nathan Schiff[1]
Shanghai University of Finance and Economics

SPACE Summer School, Jiaotong University
August 12th, 2024

[1] For comments or to fix mistakes, please email me at nschiff@gmail.com. These slides and other notes on urban economics are available on my website https://nathanschiff.com.

Distance-based Treatments

Many economic phenomena and policies may vary in intensity with geographic distance to a location

A few examples:

- Distance to other firms and productivity spillovers (Rosenthal and Strange 2003, Arzaghi and Henderson 2008, Greenstone, Hornbeck, and Moretti 2010, Baum-Snow, Gendron-Carrier, and Pavan 2024)
- Distance to firms and competition: (Ellickson and Grieco 2013, Busso and Galiani 2019, Arcidiacono, Ellickson, Mela, and Singleton 2020, Schiff, Cosman, and Dai 2023)
- Distance to pollution: (Currie, Greenstone, and Moretti 2011, Currie, Davis, Greenstone, and Walker 2015)
- Distance to transit: (Billings 2011, Gupta, Van Nieuwerburgh, and Kontokosta 2022, Jerch, Barwick, Li, and Wu 2024)
- Distance to foreclosures: (Campbell, Giglio, and Pathak 2011, Anenberg and Kung 2014, Gerardi, Rosenblatt, Willen, and Yao 2015)

Identifying distance-based treatment effects

One strategy to identify the treatment effect is to compare regions closer to the treatment location to regions further away

However, this raises several related questions:

1. How close is "close enough" to be treated? How far is "far enough" to be a control?
2. How does the effect of treatment vary (ex: decay) with distance? Linearly? Exponentially? Concave? Convex?
3. The set of potential untreated units (controls) may be much larger than treated units (the area of a circle increases quadratically with distance to center). Should we only use a subset of the controls? How to select these?

What if treatment varies with distance, but nearby units are *not* good controls?

This talk

In my own research and when advising my students, I've struggled with all of the questions.

Today I will discuss some of these issues from my own experience and discuss possible solutions, a few from my work, but mostly from the literature.

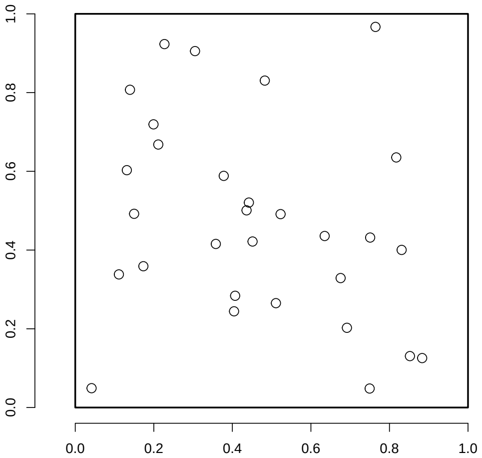
I'm focusing on cases where we think the treatment effect varies continuously with distance from a location (a point, or small area); for discussion of Spatial RDD see (Keele and Titiunik 2015)

My emphasis is on identification strategies and I won't talk at all about inference (ex: spatially correlated errors (Conley 1999))

My working example will be the case of point data but the intuition applies to areal data (regions: counties, census tracts, 街道)

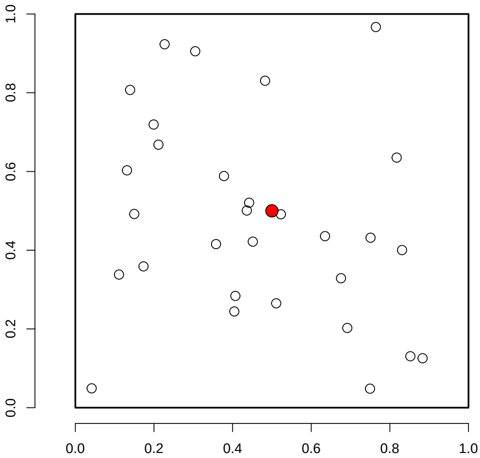
Working example: incumbent restaurants' response to entry

- A number of restaurants exists in a small town in an equilibrium state



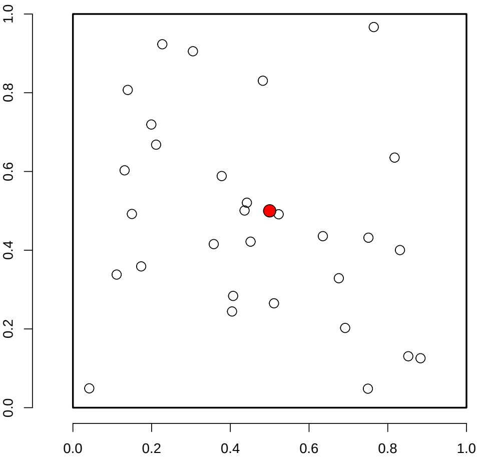
Working example: incumbent restaurants' response to entry

- A number of restaurants exists in a small town in an equilibrium state
- Suddenly, a new restaurant enters the market



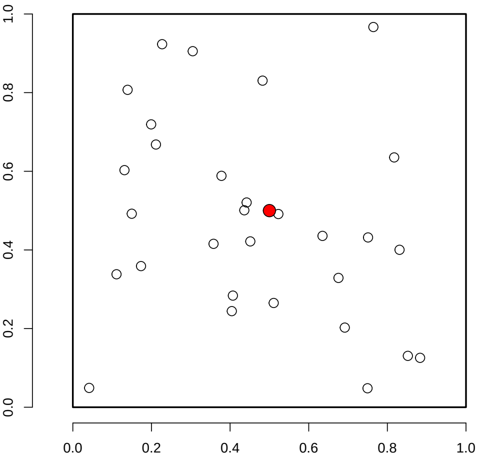
Working example: incumbent restaurants' response to entry

- A number of restaurants exists in a small town in an equilibrium state
- Suddenly, a new restaurant enters the market
- How will the existing restaurants respond? Will they change prices? Will they change products?



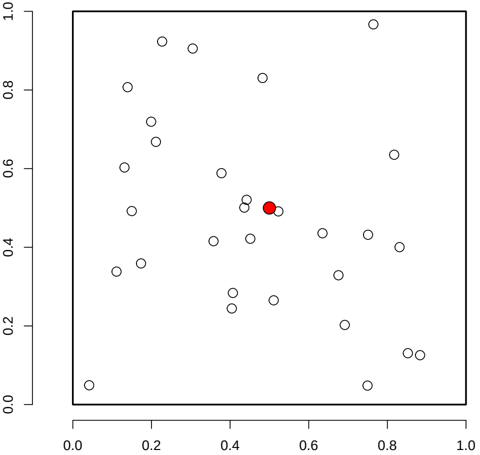
Working example: incumbent restaurants' response to entry

- A number of restaurants exists in a small town in an equilibrium state
- Suddenly, a new restaurant enters the market
- How will the existing restaurants respond? Will they change prices? Will they change products?
- How can we estimate the entry effect on the existing (incumbent) restaurants?



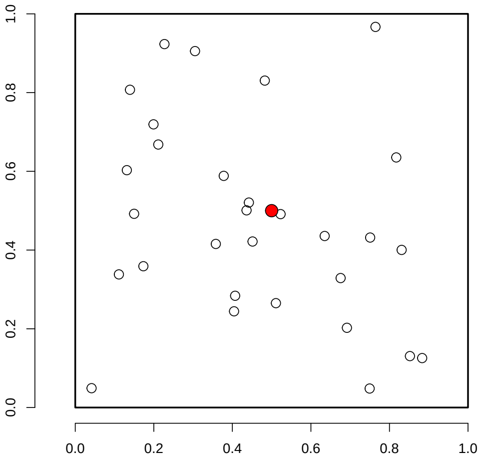
Estimating the response to entry

- Customers have to travel to restaurants (ignore 外卖)



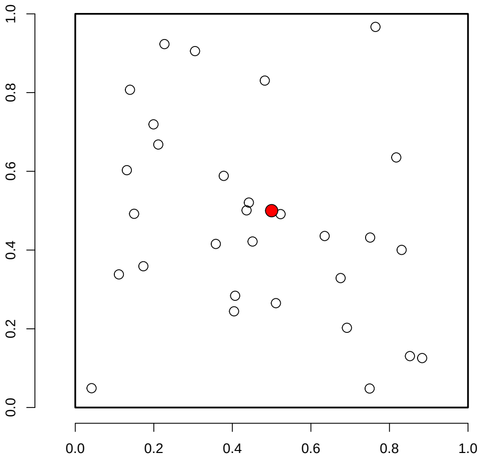
Estimating the response to entry

- Customers have to travel to restaurants (ignore 外卖)
- This travel cost may allow for local market power in the restaurant market and...



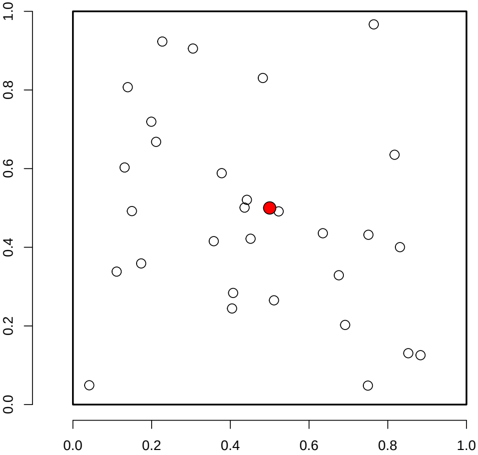
Estimating the response to entry

- Customers have to travel to restaurants (ignore 外卖)
- This travel cost may allow for local market power in the restaurant market and...
- implies that the effect of new competition may *vary with distance*



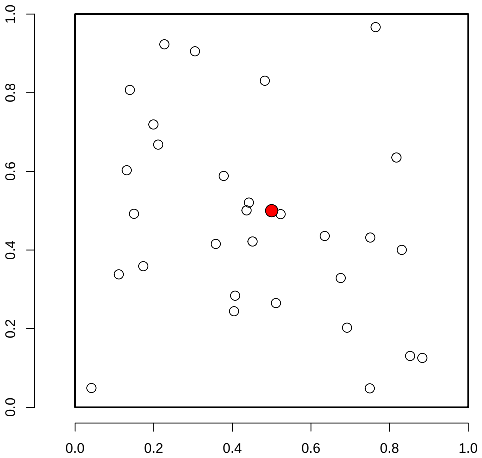
Estimating the response to entry

- Customers have to travel to restaurants (ignore 外卖)
- This travel cost may allow for local market power in the restaurant market and...
- implies that the effect of new competition may *vary with distance*
- Therefore, we can compare restaurants close to the entrant with those further away to estimate the entry response



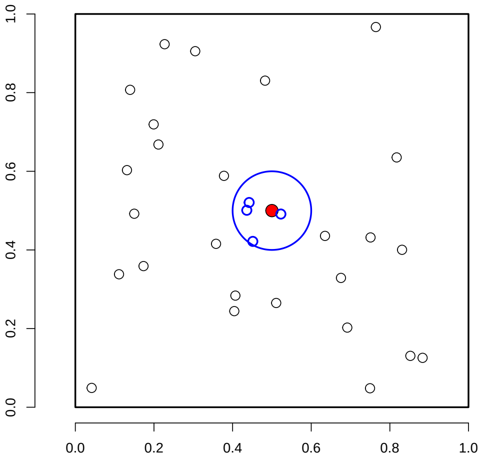
Distance buffers

- One common estimation strategy is to assume a treatment buffer: all restaurants within distance *d* are treated



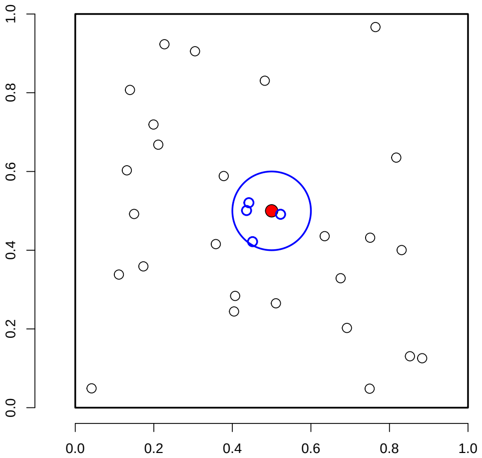
Distance buffers

- One common estimation strategy is to assume a treatment buffer: all restaurants within distance d are treated
- For example, treatment occurs if the distance d_i between incumbent i and the entrant is less than d; here $d = 0.1$



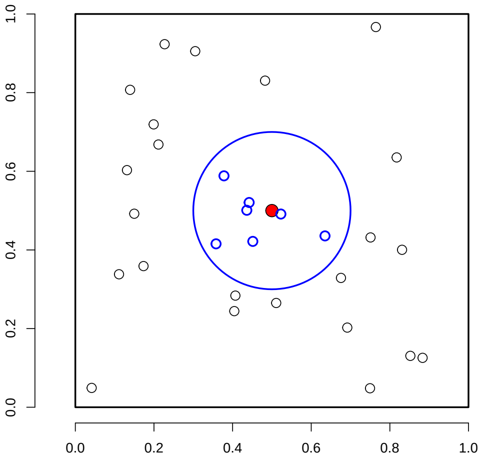
Distance buffers

- One common estimation strategy is to assume a treatment buffer: all restaurants within distance *d* are treated
- For example, treatment occurs if the distance d_i between incumbent *i* and the entrant is less than *d*; here $d = 0.1$
- But how do we know $d = 0.1$?



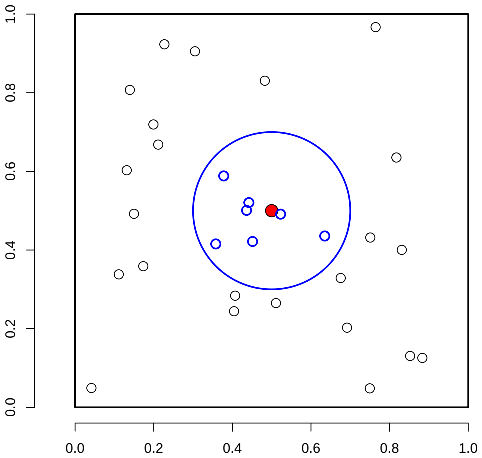
Distance buffers

- One common estimation strategy is to assume a treatment buffer: all restaurants within distance d are treated
- For example, treatment occurs if the distance d_i between incumbent i and the entrant is less than d; here $d = 0.1$
- But how do we know $d = 0.1$?
- Why not $d = 0.2$?



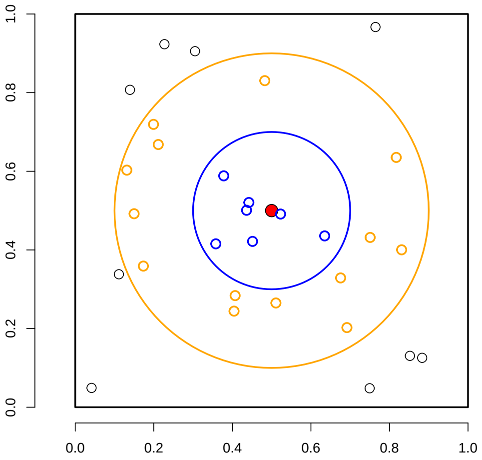
Which units are controls?

- Setting a distance buffer implies that all restaurants outside the buffer are non-treated, but are these all controls?



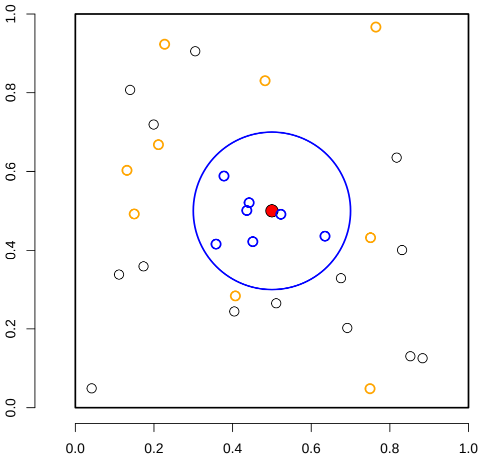
Which units are controls?

- Setting a distance buffer implies that all restaurants outside the buffer are non-treated, but are these all controls?
- Units far away may no longer be "similar," requires restricting controls to within a given distance of treatment location



Which units are controls?

- Setting a distance buffer implies that all restaurants outside the buffer are non-treated, but are these all controls?
- Units far away may no longer be “similar,” requires restricting controls to within a given distance of treatment location
- What if good control units are not a function of distance to the treated location?



Using Nearby Units as Controls: Framework and Method from Butt JUE Insight 2023

Butts 2023: Spatial DiD

A nice paper from Butts (2023) considers these issues for the case when control units are distributed as a function of distance to the treatment location

Following Butts, consider a unit i located at (x_i, y_i) that is treated by an event (ex: entry) located at $(\bar{x}, \bar{y})$ which occurs between periods $t = 0$ and $t = 1$.

The outcome of i in period t is: $Y_{it} = \mu_i + \tau(Dist_i)\mathbf{1}_{t=1} + \lambda(Dist_i)\mathbf{1}_{t=1} + \epsilon_{it}$

The $\tau(Dist_i)$ term is the treatment effect while the $\lambda(Dist_i)$ term represents non-treatment shocks, and is the source of endogeneity

Notice that $\lambda(Dist_i)$ is interacted with post ($\mathbf{1}_{t=1}$), so the endogeneity is a time-*varying* effect (μ_i captures all invariant effects)

Butts defines average treatment effect as $\bar{\tau} = E[\tau_i|\tau(Dist_i) > 0]$, where τ_i is the treatment effect for unit i; both treatment effect and other shocks depend on distance to treatment event

Identifying assumption: local parallel trends

Local parallel trends: for a distance $\bar{d}$, local parallel trends hold if for all $d, d' \leq \bar{d}$ it's true that $\lambda(d) = \lambda(d')$

In other words, up to $\bar{d}$ the non-treatment shock is constant

Strong assumption, but may be reasonable for small distances (more later)

Butts points out that "local parallel trends" implies "average parallel trends" (our usual assumption for DiD):

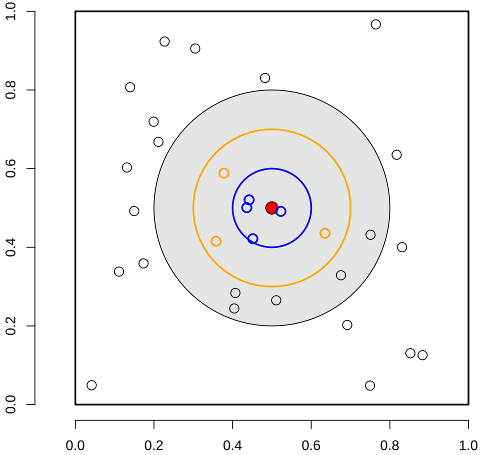
Average parallel trends: for a pair of distances d_t and d_c, it is true that $E[\lambda_d | 0 \leq d \leq d_t] = [\lambda_d | d_t \leq d \leq d_c]$

Note: average parallel trends allows for λ_d to vary with distance, as long as the average value is equal for treated and control units.

Local parallel trends requires it be constant: $\lambda_d = \lambda, \ \forall d < \bar{d}$

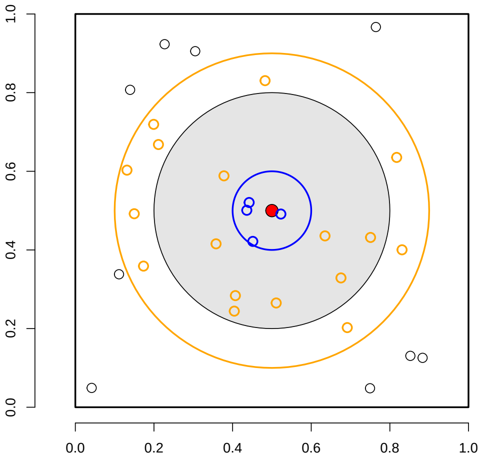
Importance of treatment and control distances

- If both treated and control units are within $\bar{d}$—the area where lambda is constant–then the endogeneity simply differences out



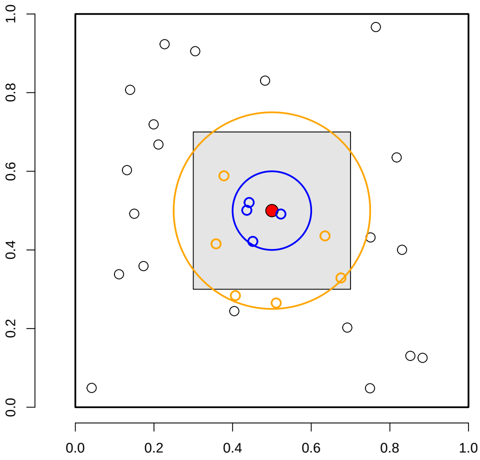
Importance of treatment and control distances

- If both treated and control units are within $\bar{d}$—the area where lambda is constant–then the endogeneity simply differences out
- Notice that it's important controls aren't too far away: it must be that $d_c < \bar{d}$



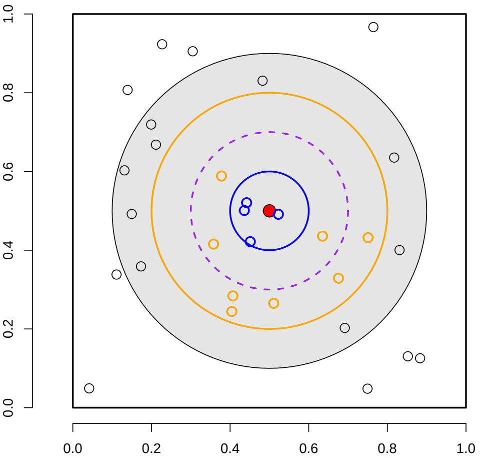
Importance of treatment and control distances

- If both treated and control units are within $\bar{d}$—the area where lambda is constant–then the endogeneity simply differences out
- Notice that it's important controls aren't too far away: it must be that $d_c < \bar{d}$
- And important that λ is a simple function of distance



Importance of treatment and control distances

- If both treated and control units are within $\bar{d}$—the area where lambda is constant–then the endogeneity simply differences out
- Notice that it's important controls aren't too far away: it must be that $d_c < \bar{d}$
- And important that λ is a simple function of distance
- But what if you don't know the true d_t (solid blue) and used a different distance (dashed purple)?



Unknown treatment distance d_t

If you set the distance d_t to be too large then you assign some control units to the treatment group, thus diluting (attenuating) the effect of treatment (since these units have zero effect by assumption)

If you set distance d_t to be too small, then some control units are actually treated, also leading to bias

- Butts notes that if treatment is decreasing in distance, this can lead to an *upward* bias since the remaining treated units have a stronger treatment (closer to treatment location)
- If treatment is constant within distance (or varies non-monotonically) this could lead to downward bias

Assuming *average* parallel trends requires knowing the treatment distance.

What to do when treatment distance is unknown?

One ad-hoc approach: an exclusion buffer

Sometimes we don't know how far the treatment distance extends, but we do know an area that should be treated: we know a d_{b1} such that $d_{b1} < d_t$

If we are further willing to state a distance d_{b2} past which all units must be controls, $d_c < d_{b2}$, then we can use an "exclusion buffer" (my term)

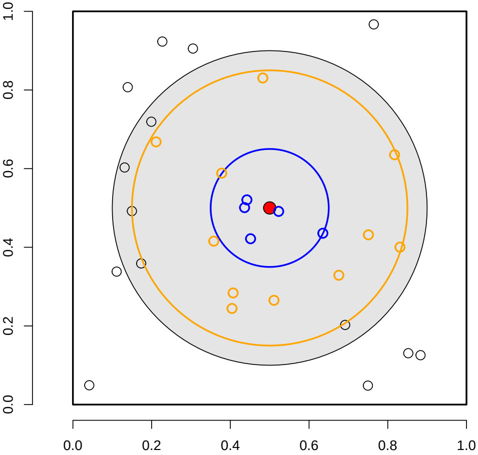
We then compare treated units where $d < d_{b1}$ to control units where $d_{b2} < d$ and exclude all units between d_{b1} and d_{b2} from the regression

The estimated effect is no longer the effect for all treated units—it's a LATE—but it is an unbiased (under these assumptions) estimate of treatment for units within d_{b1}

We used a version of this in Schiff, Cosman, Dai (2023) to look at incumbent restaurant responses to entry

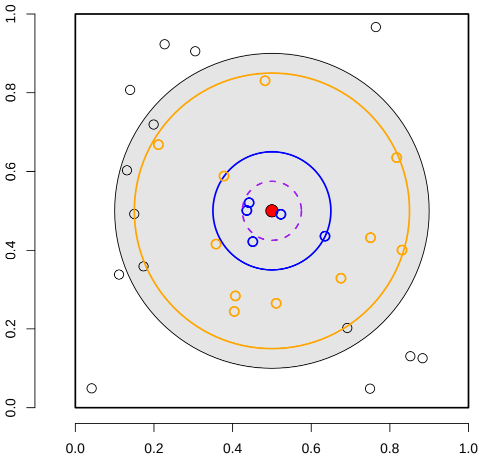
Exclusion buffer illustration

- We don't know the true distance d_t (blue)



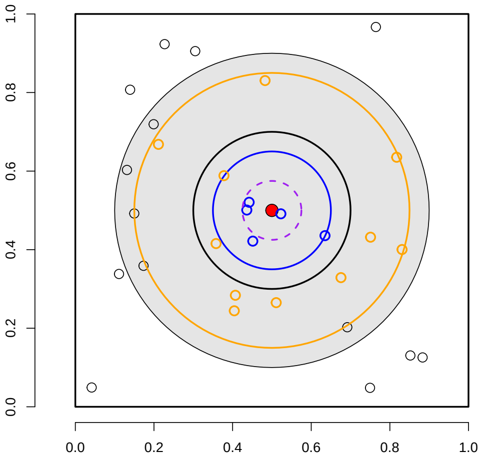
Exclusion buffer illustration

- We don't know the true distance d_t (blue)
- We assume that within d_{b1} units must be treated (dashed purple)



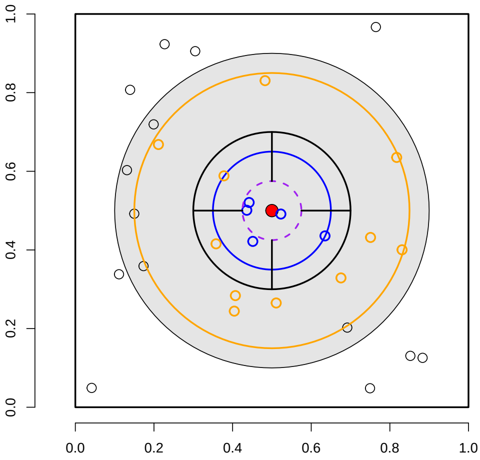
Exclusion buffer illustration

- We don't know the true distance d_t (blue)
- We assume that within d_{b1} units must be treated (dashed purple)
- We specify another distance d_{b2} after which units must be controls (**black solid line**)



Exclusion buffer illustration

- We don't know the true distance d_t (blue)
- We assume that within d_{b1} units must be treated (dashed purple)
- We specify another distance d_{b2} after which units must be controls (**black solid line**)
- Then we exclude all units between d_{b1} and d_{b2} from the regression



Multiple ring approach

A more systematic approach is to use multiple rings to estimate treatment effects at different distances, which can be interpreted as a non-parametric estimate of $\tau(Dist)$

This is common in literature, but the number of rings and the maximum distance from the treatment location can be ad-hoc

Butts shows that as long as the true (and unknown) treatment distance t_d is less than the control distance d_c specified by the researcher, then it's possible to identify the treatment effect *under local parallel trends*

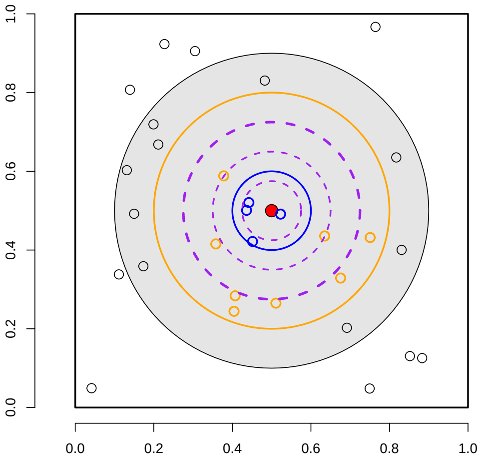
Basic idea: estimate L equal width rings between the treatment location and control distance d_c

Within the last ring where $d_t < d_L < d_c$, the treatment effect must be zero, leaving only λ (given local parallel trends).

Thus by comparing the treatment effect of all previous $L - 1$ rings to the last ring L, λ is differenced out and the true treatment effect can be estimated within each ring

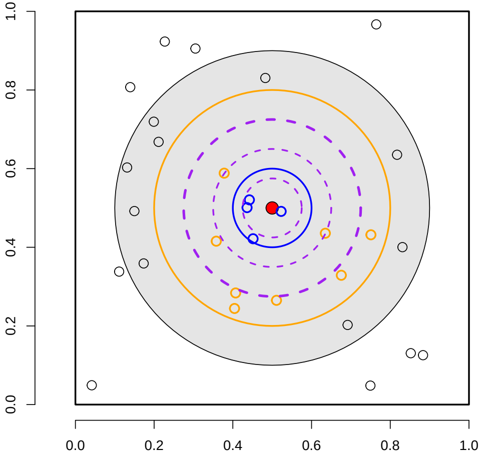
Identifying treatment effect with last ring

- In the last ring (last purple dashed line to orange d_c) the treatment effect is zero, thus only λ remains



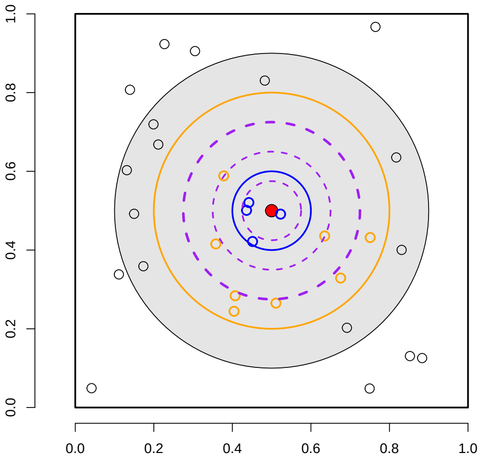
Identifying treatment effect with last ring

- In the last ring (last purple dashed line to orange d_c) the treatment effect is zero, thus only λ remains
- If we compare the first ring to the last, we isolate the treatment effect at a close distance



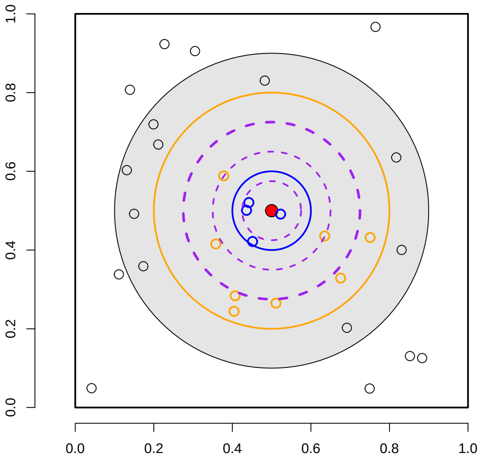
Identifying treatment effect with last ring

- In the last ring (last purple dashed line to orange d_c) the treatment effect is zero, thus only λ remains
- If we compare the first ring to the last, we isolate the treatment effect at a close distance
- The second ring mixes treated and control units and will give a smaller estimate (both because treatment declines from the first ring, and some units are not treated)



Identifying treatment effect with last ring

- In the last ring (last purple dashed line to orange d_c) the treatment effect is zero, thus only λ remains
- If we compare the first ring to the last, we isolate the treatment effect at a close distance
- The second ring mixes treated and control units and will give a smaller estimate (both because treatment declines from the first ring, and some units are not treated)
- Within the third ring there are only control units, thus the estimated effect should be zero



Linden and Rockoff 2008

In the US, a federal law known as “Megan’s Law” require sex offenders (people convicted of a sex crime) to register their location in a public database (even after serving jail time)

Linden and Rockoff estimate the effect of distance to a known sex offender on housing prices

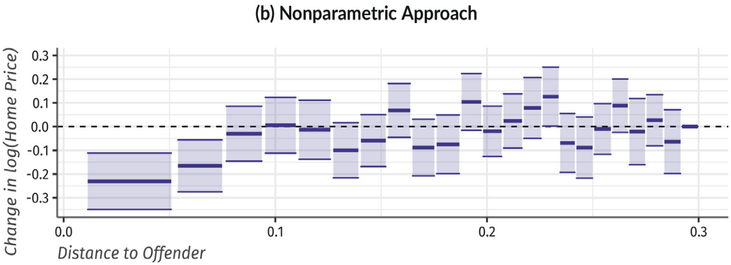
They define treated houses as those within 0.1 miles of the offender and control houses as 0.1 to 0.3 miles

Butts re-examines their results and suggests that the DiD approach misses substantial heterogeneity, with close houses having a much larger price decline than houses further away, but still within 0.1 miles

To show this, he non-parametrically estimates the treatment effect at different distance bins (rings)

Butts approach to Linden and Rockoff 2008

Butts estimation of effect from Linden and Rockoff, AER (2008)



Importantly, the ring coefficients level off to zero; otherwise hard to argue that $d_t < d_c$

Further, if they level off it suggests (but does not prove) that trends within each ring are roughly similar across distance ($\lambda_d = \lambda$)

Note: this approach does *not* estimate the ATE, but rather shows the full curve; one could then choose a given distance with which to estimate an average effect

Choosing the number of rings

A necessary condition is that the number of rings L has to be large enough so that the treatment effect is zero within the last ring

However, given that this is true, the number of rings chosen could still have a non-trivial effect on the estimated shape of the distance effect

Butts notes that choosing the optimal number of rings is similar to choosing the optimal number of bins in a binscatter diagram, a question addressed in the recent work by Cattaneo et al. AER (2024)

As L increases, rings become smaller, thus reducing bias of estimate within distance ring; however, smaller rings also have fewer observations, thus increasing the variance (bias-variance trade-off)

Cattaneo et al. suggest choosing bin (ring) count to minimize error; can use their software package *binsreg* (Stata, R, Python) to choose optimal L

Alternative: estimating changes in a distance gradient

$Y_{it} = \mu_i + \tau(Dist_i)\mathbf{1}_{t=1} + \lambda_i(Dist_i)\mathbf{1}_{t=1} + \epsilon_{it}$

Sometimes researchers want to estimate how the distance effect has changed after an event: $\tau(Dist_i)\mathbf{1}_{t=1} - \tau(Dist_j)\mathbf{1}_{t=1}$ for $i \neq j$

Consider the case where we have local parallel trends $\lambda_i(Dist_i) = \lambda$ for $d < \bar{d}$ and we want to know how an outcome has changed within a radius less than $\bar{d}$

We estimate the following *without controls*:
$Y_{it} = \mu_i + \tau(Dist_i)\mathbf{1}_{t=1} + \lambda\mathbf{1}_{t=1} + \epsilon_{it}$

We can estimate how the outcomes changes with distance in the post period—the change in gradient—even though the overall treatment effect isn't identified

Consider two points i, j at different distances and take the first difference (over time) to remove the individual FE.

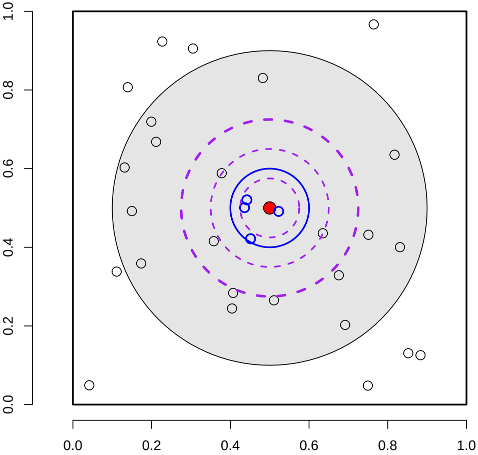
Then the post period difference between i and j is:
$\Delta Y_i - \Delta Y_j = \tau(Dist_i) + \lambda - (\tau(Dist_j) + \lambda) = \tau(Dist_i) - \tau(Dist_j)$.

Distance gradients

- We can estimate the change in the distance gradient for a specific functional form by specifying $f(Dist)$, such as linear $f(Dist) = \beta * Dist$ or quadratic $f(Dist) = \beta * Dist^{-2}$

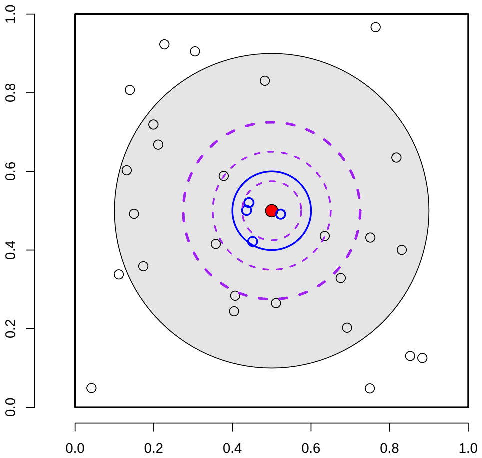
Distance gradients

- We can estimate the change in the distance gradient for a specific functional form by specifying $f(Dist)$, such as linear $f(Dist) = \beta * Dist$ or quadratic $f(Dist) = \beta * Dist^{-2}$
- Alternatively, we could use the ring method to estimate the shape non-parametrically



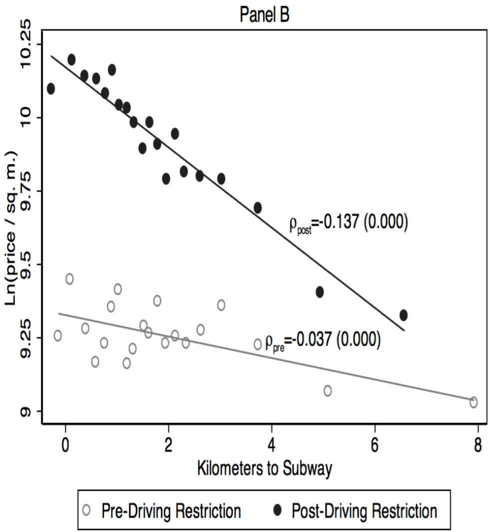
Distance gradients

- We can estimate the change in the distance gradient for a specific functional form by specifying $f(Dist)$, such as linear $f(Dist) = \beta * Dist$ or quadratic $f(Dist) = \beta * Dist^{-2}$
- Alternatively, we could use the ring method to estimate the shape non-parametrically
- Notice that with the ring method, only $L - 1$ rings are identified and that all rings must be within the λ area (gray) defined by $\bar{d}$



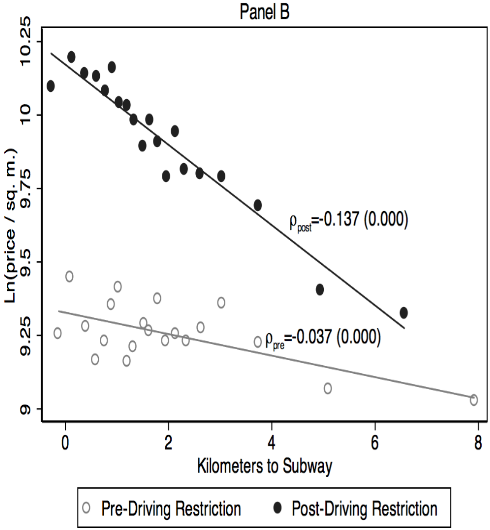
Beijing Road Rationing: Price and Income Gradients

- A good example of gradient change estimation is Jerch, Barwick, Li, and Wu (2024), who examine how prices and income change with distance to a subway station after Beijing introduces driving restrictions



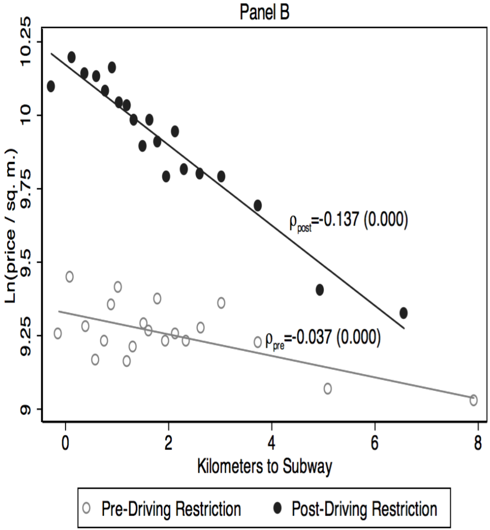
Beijing Road Rationing: Price and Income Gradients

- A good example of gradient change estimation is Jerch, Barwick, Li, and Wu (2024), who examine how prices and income change with distance to a subway station after Beijing introduces driving restrictions
- They note that if “there are no other contemporaneous shocks that affect the price (or income) gradient” then the change is identified



Beijing Road Rationing: Price and Income Gradients

- A good example of gradient change estimation is Jerch, Barwick, Li, and Wu (2024), who examine how prices and income change with distance to a subway station after Beijing introduces driving restrictions
- They note that if "there are no other contemporaneous shocks that affect the price (or income) gradient" then the change is identified
- They find a sharp increase in the *slope* (intercept not identified): housing prices drop more rapidly moving away from a subway station after the driving restrictions



Model-based controls

Why are nearby units good controls?

General argument is that within small area (ex: neighborhood), variation in exact location of treatment is random

Ex: stores choose a retail area but exact address depends on vacancies; house-buyers choose neighborhood but exact house depends on stock

Idea (first?) used with areal data in Bayer, Ross, and Topa (2008), "BRT", to study job referrals

Authors estimating whether two people, *i* and *j* are more likely to work in the same location ($W_{ij}^b = 1$), if they live in the same block ($R_{ij}^b = 1$):

$$W_{ij}^b = \rho_g + \alpha_0 R_{ij}^b + \epsilon_{ij}$$

The key is the reference group (ex: block *group*) fixed effect ρ_g, which captures all location specific factors that could affect work location

Notice that this is essentially an areal version of the local parallel trends: all shocks are common within a small area g

Baum-Snow et al. AER 2024

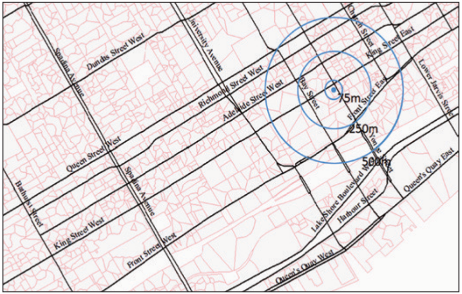


Figure 1. Map of Downtown Toronto

Notes: Postal codes are outlined by thin lines. Major streets are in black. All postal codes with centroids within the indicated central 75-meter-radius circle are included in the indicated example peer group area.

Authors focus on 75m ring, control for 500m ring

Evidence in support of a small area strategy

The identification strategy from BRT 2008 has been used by multiple high quality papers

- Ex: Currie, Greenstone, and Moretti (2011), Campbell, Giglio, and Pathak (2011),Anenberg and Kung (2014), Bayer, Mangum, and Roberts (2021),Baum-Snow, Gendron-Carrier, and Pavan (2024)

A nice thing about the method is one can show how correlation of key *observables* is reduced by inclusion of the reference group fixed effect

This not a test of the identification strategy, but does help to persuade readers of validity, just like a balance test for matching

Note: in spatial DiD framework, units within $\bar{d}$ can differ in covariate *levels*, since these are differenced out. However, if one can show that the levels are similar, that provides support that trends should also be similar

Evidence for Bayer Ross Topa 2008 strategy

TABLE 3
Correlation between Individual and Average Characteristics of Neighbors Residing on Same Block

	Sample: Blocks with Five+ Workers in Sample		
	Unconditional (1)	Conditional on Census Block Group (2)	Conditional on 10 Closest Blocks Reference Group (3)
High school graduate	.182	.040	.021
College graduate	.294	.060	.030
Age 45–59	.051	.008	−.020
Age 35–44	.017	−.004	−.031
Age 25–34	.098	.027	−.005
Single female	.110	.033	.014
Single male	.094	.027	.004
Married female	.080	.005	−.015
Married male	.088	.026	.011
Children	.142	.046	.006
Children 0–5	.046	.019	−.007
Children 6–12	.058	.017	−.017
Children 13–17	.048	.015	−.025
Children 18–24	.064	.022	−.014
Black	.593	.054	.017
Asian/Hispanic	.275	.084	.049

Note.—The table reports unbiased estimates of correlation between a series of individual characteristics and the

Col 1 shows correlation between characteristics of individuals *i* and *j* living in same block; col 2 shows correlation after condition on census block group; col 3 shows correlation after conditioning on 10 closest blocks (alternative reference group)

When are nearby units *not* good controls?

An important decision in using nearby controls is defining "nearby" ($\bar{d}$)

With areal data (ex: census blocks, block groups), there is often not much flexibility since spatial units are predefined

With point data, we can choose $\bar{d}$, but even small distances may still be important in some contexts

For restaurant location choice, distances under 1km still matter: think about popular restaurant streets, or locating near a square or park

Moreover, fundamentals may vary with distance from that point (ex: commercial rent, distance to transit)

Rather than local parallel trends, it could be that $\tau(Dist)$ and $\lambda(Dist)$ are strongly correlated

Two ends of 大学路: 800 meters apart, line 10 entrance

West end 大学路

East end of 大学路

Modeling (reduced form) treatment assignment

Thinking about how treatment is assigned may yield a clearer and more defensible identification strategy

By modeling/predicting treatment, one can then compare two locations with equal likelihood of treatment, but one received treatment due to purely idiosyncratic factors

The strong assumption one has to make is that all factors determining treatment are controlled for; conditional on modeled treatment likelihood, treatment assignment is random (conditional mean independence)

However, (in my opinion) local parallel trends is equally strong but less explicit because the source of endogeneity is not stated—what are the components of λ?

Note: Even when using local parallel trends, modeling treatment can help one to select appropriate covariates for BRT-type balance test

ATE or Average Treatment on the Treated?

Another benefit of modeling treatment assignment is it forces the researcher to consider what is being estimated

In spatial DiD framework, we defined the ATE as $\bar{\tau} = E[\tau_i | \tau(Dist_i) > 0]$

However, treatment location is rarely random and often we only see specific types of locations receiving treatment

Ex: In Ellickson and Grieco (2013) they estimate effect of Walmart entry on supermarkets using a ring approach

Authors note that since they only know locations Walmart chose, estimated effect is average treatment on treated (ATET)

"Without modeling (as opposed to controlling for) the selection of locations, we cannot predict the impact of Wal-Mart's entry in locations that are vastly different from the ones it has entered so far (e.g. a Wal-Mart in downtown Manhattan or San Francisco proper)."

Restaurant responses to entry; Schiff, Cosman, Dai, JUE 2023

In "Delivery in the city" (2023) we were interested in evaluating theories of competition among firms with differentiated products

Classic spatial competition models (ex: Hotelling) suggest firms compete strategically, but only with rivals whose products are most similar (which included geographic distance)

On the other hand, monopolistic competition models (CES and others) suggest firms are always differentiated enough that there is no strategic competition, but rather firms compete for market share with the rest of the market (aggregate)

To assess these competing theories, we assembled a panel (unbalanced) of 550,000 restaurant *menus* from New York City restaurants over 68 consecutive weeks

Our basic strategy was to compare the prices and products on a restaurant's menu, before and after it received a new competitor within a "close" distance

Restaurant entry is not random

The identification challenge is that entry is not random and therefore we could not simply compare restaurants with new competition to those without

(a) Menu attributes

	Menu stats	
	t-tests	N
item count	-14.43***	88115
mean item price	0.17**	88115
median item price	0.21***	88115
p25 item price	0.13***	88115
p95 item price	0.00	88115
stars	0.06*	85202
review count	40.87***	79003
order rating	0.71***	86252
food rating	0.44*	86252
delivery rating	1.23***	86252

Tests difference between treated and control.

(b) Demographic attributes

	Demographics	
	t-tests	N
age2529	0.023***	88476
age3039	0.027***	88476
age7079	-0.003***	88476
racewhite	0.087***	88476
raceblack	-0.053***	88476
hhfamily	-0.086***	88476
hhmarried	-0.038***	88476
educdegree	0.118***	88476
poverty	-0.023***	88476
income100150	0.008***	88476
income150200	0.008***	88476
unitdetached	-0.064***	88476
competitors 500m	16.107***	81877

Modeling entry

Our main identification strategy was to match existing restaurants with equal likelihood of receiving new competition, and then compare these matched restaurants before and after one received entry (matching diff-in-diff)

But how to model entry? Is it a choice made by a fixed number of entrants? Or is it a more of a statistical process where the count varies?

We modeled entry as a Poisson process (statistical) where the count of entrants varied with the underlying characteristics of small neighborhoods (measured with both areal—census tracts—and continuous variables, such as distance to a subway)

We then matched restaurants based on the predicted count of entrants within 500m (baseline) over a given time period

Thinking about location-based treatments

When treatment is made by one or a small number of agents deciding on a location, then modeling treatment as a choice problem may be appropriate

- Ex: Walmart locations (Jia 2008, Holmes 2011), new factory location (Greenstone, Hornbeck, and Moretti 2010)

For cases with many agents making independent choices, or when treatment isn't the outcome of a choice, then a statistical model makes more sense

- Ex: large numbers of entrants, home foreclosures, crime, pollution sources, Covid cases

An important difference between these two approaches is that with choice models the number of treatments is fixed: the researcher is modeling which location is treated, not whether there is a treatment

With the statistical approach the total number of treated locations is also a random variable

Aside: discrete choice models, Poisson process

Count data is usually modeled Poisson while choice data uses discrete choice models (ex: conditional logit)

Great article by Schmidheiny and Brülhart (2011) showing relation between these models and important differences

Conditional logit models and Poisson models have the same likelihood function (when regressors are only location specific, not characteristics of choice-maker)

Thus, we can actually estimate logit models with a large number of choices quite quickly using Poisson (Guimaraes, Figueirdo, and Woodward 2003)

However, models are quite different in predictions: discrete choice models predict exact number of choices made, but Poisson predicted counts are variable (and can differ from actual observed)

Note: usually model treatment discretely at the region level because most covariates come from areal data (ex: census tracts); it is possible to model treatment in continuous space (points), but unusual to have continuous predictors

Back to restaurants: using our entry model

With our poisson entry model, we then matched followed the propensity score literature to trim the sample and run balance tests of restaurant and area characteristics

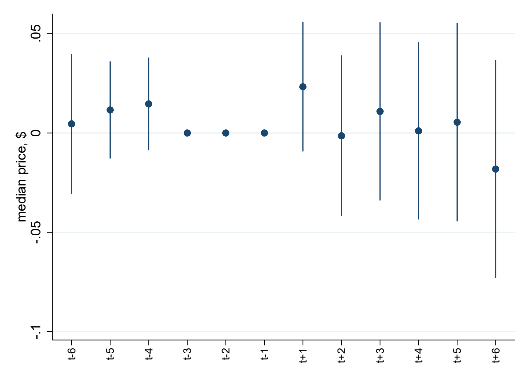
We then ran event studies and two-way fixed effect models comparing treated and matched control restaurants, before and after treated faced entry

In our baseline specification, we looked at entry within 500m, but we also tested many other entry distances

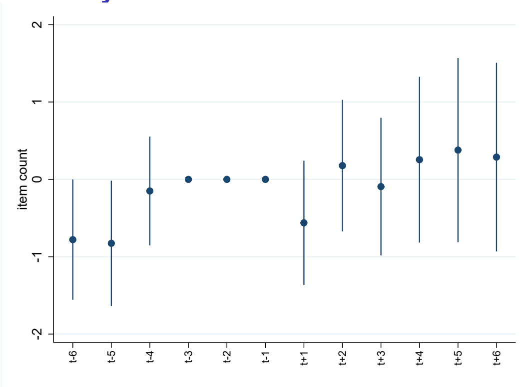
Note that in our study, changing distance changes treatment—restaurants facing a single new entrant within 500m are quite different than those facing one within 1500m—and so we matched separately for each distance

Result: **no effect** of entry on prices or products! Coefficients all very close to zero, statistically insignificant.

Results 1: Event Study Plots



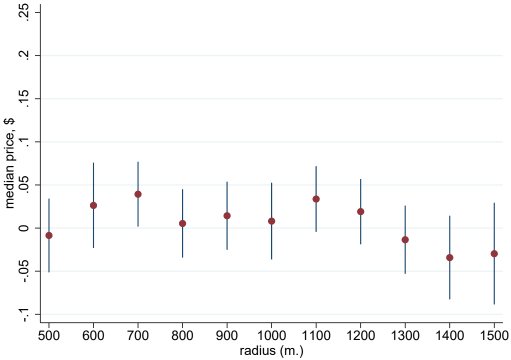
Median price



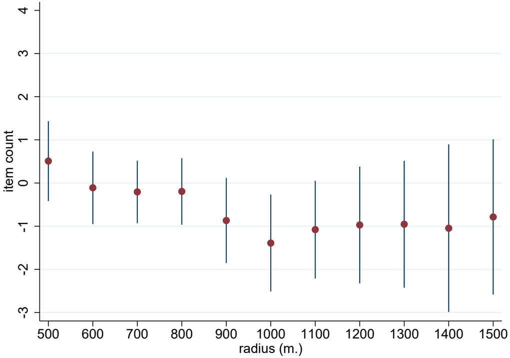
Item count

Plots show results for restaurants facing a single entrant over a 24 week period. We exclude three periods before entry to avoid including any anticipatory responses. Average median item price is $8.62, average item count is 124.4.

Results 2: Different spatial distances



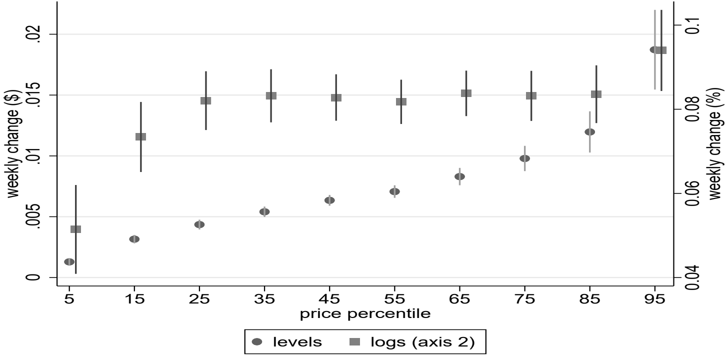
Median price



Item count

Plots show TWFE estimates results for different treatment distances (max distance to entrant) for restaurants facing a single entrant over a 24 week period.

Results 3: Prices do change, but in same way for treated and control



Plot shows within restaurant (r) estimates of $Y_{rt} = \beta * weeks_{rt} + \eta_r + \epsilon_{rt}$, where Y_{rt} is a price percentile (ex: 10th percentile of menu) t weeks after first observation of r.

Alternative explanation: strategic entry?

Finding no response was surprising and despite a large battery of robustness checks, people were still skeptical that there was no effect

Our entry model predicts the count of entrants, but does *not* differentiate between types of entrants (cuisine, price, etc...)

We often received the comment that our results could be explained by strategic entry: entrants intentionally chose locations where existing restaurants had different menus and would not be competitors (ex: Italian restaurants locate away from other Italian restaurants)

How to test this?

Could we simply compare the share of nearby restaurants that have the same cuisine (ex: Italian-Italian) to the share of different cuisine restaurants to that cuisine (ex: non-Italian to Italian)?

Measuring location patterns

There is a very developed literature in spatial statistics focused on concentration measures for areal data (Getis and Ord 1992, Ord and Getis 1995) and point data (Diggle 2013); also see online notes by Tony Smith (2014)

- For examples in urban economics, see (Ellison and Glaeser 1997, Duranton and Overman 2005, Billings and Johnson 2012, Dai and Schiff 2023)

Key messages from literature is importance of null distribution—to which spatial distribution should one compare?

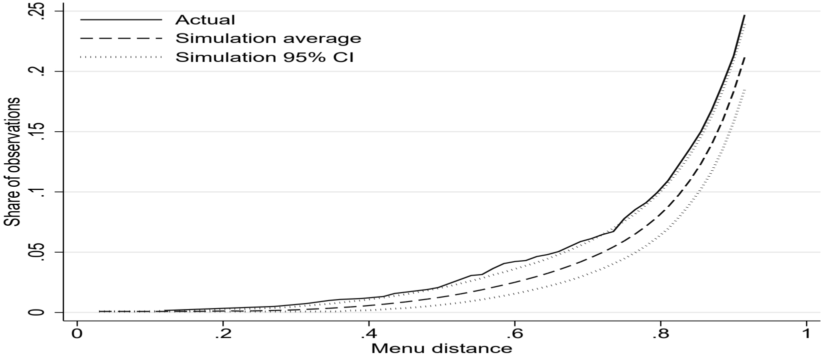
We used the observed location of entrants as our null and then randomly reshuffled which entrants are assigned to which location (“permutation test”–permute the restaurant identities)

This captures the idea that restaurant location choices are limited by many factors and thus using actual entrant locations ensures plausible counterfactuals

We then compared the similarity of entrants menus with nearby incumbents to our counterfactual location distributions (10,000 random permutations)

Restaurants locate *closer* to similar restaurants

We found that similar restaurants are actually more likely to co-locate than if they were simply assigned random locations from the set of entrant locations



Greater menu distance means menus are less alike

Concluding thoughts

Summary

Availability of spatial data has increased tremendously and researchers increasingly use geography-based identification strategies

The standard of rigor for these strategies is also increasing and we can improve upon vague statements about “close controls” or “decaying distance effects”

Instead, we can often non-parametrically estimate and plot the shape of distance effects

We can formally state the identification assumption, and by thinking about treatment assignment, show evidence in support of this assumption (ex: BRT-type balance tests)

Moreover, often we can model location-based treatments using choice models (ex: logit) or statistical processes (ex: poisson), and use model-based treatment assignment for identification

Final thought

In (Schiff, Cosman, and Dai 2023) we focused on cases where an existing restaurant received one (or a few) new entrants within a fixed time period (minimum of 16 weeks)

This allowed us to compare these restaurants with others that had no entry over the same period, thus cleanly separating treated and control

However, in New York City there are many places where existing restaurants face new entry *every month*!

There is no way use a treatment/control framework for these cases, yet these are important parts of the New York City restaurant market.

Instead, to estimate competitive effects one needs to explicitly model when and how restaurants adjust menus and then use the data to test this model.

How can we estimate a structural model of competition where space is important?
Ask 王子 and 黃子彬!

Thanks!

References I

ANENBERG, E., AND E. KUNG (2014): "Estimates of the size and source of price declines due to nearby foreclosures," *American Economic Review*, 104(8), 2527–2551.

ARCIDIACONO, P., P. B. ELLICKSON, C. F. MELA, AND J. D. SINGLETON (2020): "The competitive effects of entry: Evidence from supercenter expansion," *American Economic Journal: Applied Economics*, 12(3), 175–206.

ARZAGHI, M., AND J. V. HENDERSON (2008): "Networking off madison avenue," *The Review of Economic Studies*, 75(4), 1011–1038.

BAUM-SNOW, N., N. GENDRON-CARRIER, AND R. PAVAN (2024): "Local productivity spillovers," *American Economic Review*, 114(4), 1030–1069.

BAYER, P., K. MANGUM, AND J. W. ROBERTS (2021): "Speculative Fever: Investor Contagion in the Housing Bubble," *American Economic Review*.

References II

BAYER, P., S. L. ROSS, AND G. TOPA (2008): "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*.

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

BILLINGS, S. B., AND E. B. JOHNSON (2012): "A non-parametric test for industrial specialization," *Journal of Urban Economics*, 71(3), 312–331.

BUSSO, M., AND S. GALIANI (2019): "The causal effect of competition on prices and quality: Evidence from a field experiment," *American Economic Journal: Applied Economics*, 11(1), 33–56.

BUTTS, K. (2023): "JUE insight: Difference-in-differences with geocoded microdata," *Journal of Urban Economics*, 133, 103493.

CAMPBELL, J. Y., S. GIGLIO, AND P. PATHAK (2011): "Forced sales and house prices," *American Economic Review*, 101(5), 2108–2131.

References III

CATTANEO, M. D., R. K. CRUMP, M. H. FARRELL, AND Y. FENG (2024): "On Binscatter," *American Economic Review*.

CONLEY, T. G. (1999): "GMM estimation with cross sectional dependence," *Journal of econometrics*, 92(1), 1–45.

CURRIE, J., L. DAVIS, M. GREENSTONE, AND R. WALKER (2015): "Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings," *American Economic Review*, 105(2), 678–709.

CURRIE, J., M. GREENSTONE, AND E. MORETTI (2011): "Superfund cleanups and infant health," *American Economic Review*, 101(3), 435–441.

DAI, T., AND N. SCHIFF (2023): "The structure and growth of ethnic neighborhoods," *Journal of Urban Economics*, 137, 103570.

DIGGLE, P. J. (2013): *Statistical analysis of spatial and spatio-temporal point patterns*. CRC press.

References IV

DURANTON, G., AND H. OVERMAN (2005): "Testing for localization using micro-geographic data," *Review of Economic Studies*, pp. 1077–1106.

ELLICKSON, P. B., AND P. L. GRIECO (2013): "Wal-Mart and the geography of grocery retailing," *Journal of Urban Economics*, 75, 1–14.

ELLISON, G., AND E. GLAESER (1997): "Geographic Concentration in US Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105(5).

GERARDI, K., E. ROSENBLATT, P. S. WILLEN, AND V. YAO (2015): "Foreclosure externalities: New evidence," *Journal of Urban Economics*, 87, 42–56.

GETIS, A., AND J. K. ORD (1992): "The analysis of spatial association by use of distance statistics," *Geographical analysis*, 24(3), 189–206.

GREENSTONE, M., R. HORNBECK, AND E. MORETTI (2010): "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings," *Journal of political economy*, 118(3), 536–598.

References V

GUIMARAES, P., O. FIGUEIRDO, AND D. WOODWARD (2003): "A tractable approach to the firm location decision problem," *Review of Economics and Statistics*, 85(1), 201–204.

GUPTA, A., S. VAN NIEUWERBURGH, AND C. KONTOKOSTA (2022): "Take the Q train: Value capture of public infrastructure projects," *Journal of Urban Economics*, 129, 103422.

HOLMES, T. J. (2011): "The diffusion of Wal-Mart and economies of density," *Econometrica*, 79(1), 253–302.

JERCH, R., P. J. BARWICK, S. LI, AND J. WU (2024): "The impact of road rationing on housing demand and sorting," *Journal of Urban Economics*, 140, 103642.

JIA, P. (2008): "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 76(6), 1263–1316.

KEELE, L. J., AND R. TITIUNIK (2015): "Geographic boundaries as regression discontinuities," *Political Analysis*, 23(1), 127–155.

References VI

LINDEN, L., AND J. E. ROCKOFF (2008): "Estimates of the impact of crime risk on property values from Megan's laws," *American Economic Review*, 98(3), 1103–1127.

ORD, J. K., AND A. GETIS (1995): "Local spatial autocorrelation statistics: distributional issues and an application," *Geographical analysis*, 27(4), 286–306.

ROSENTHAL, S. S., AND W. C. STRANGE (2003): "Geography, industrial organization, and agglomeration," *review of Economics and Statistics*, 85(2), 377–393.

SCHIFF, N., J. COSMAN, AND T. DAI (2023): "Delivery in the city: Differentiated products competition among New York restaurants," *Journal of Urban Economics*, 134, 103509.

SCHMIDHEINY, K., AND M. BRÜLHART (2011): "On the equivalence of location choice models: Conditional logit, nested logit and Poisson," *Journal of Urban Economics*, 69(2), 214–222.

References VII

SMITH, T. (2014): "Notebook on Spatial Data Analysis [online]," Discussion paper, https://www.seas.upenn.edu/ tesmith/NOTEBOOK/index.html.