Local Spillovers of Startup Innovation

Carlos Hurtado* Jaehee Song† Walter D'Lima‡

25th March, 2026

Abstract

This paper examines how innovative startups affect local real estate markets and business activity, exploiting the quasi-random assignment of patent examiners with varying leniency as an instrument for patent approval. Using nationwide data on patent applications, repeat-sale property transactions, and establishment-level business records, we find that patent-granted startups raise annualized real estate returns by 0.3 to 0.6 percentage points for residential properties and by 0.5 to 0.9 percentage points for commercial properties and vacant land within 10 miles of the innovating firm. The effects are largest for the first innovative startup and decline with additional entrants, indicating diminishing marginal externalities. Examining mechanisms, we find that patent approvals spur nearby patenting, increase local business formation in the same industry, and raise the share of non-individual buyers in the local housing market. These results show that innovative startups create significant local spillovers, with direct implications for place-based innovation policy.

Keywords: innovation, house prices, local effects, agglomeration.

*Robins School of Business, University of Richmond. Email: churtado@richmond.edu

†Leeds School of Business, University of Colorado Boulder. Email: jaehee.song@colorado.edu

‡College of Business, Florida International University. Email: wdlima@fiu.edu

1 Introduction

Startups are widely recognized as engines of technological progress and economic growth, playing a central role in driving innovation and job creation (Jaffe, Trajtenberg and Henderson, 1993; Carlino and Kerr, 2015). Yet, while large firms and industrial plants have been studied extensively for their transformative local impacts (Greenstone, Hornbeck and Moretti, 2010; Qian and Tan, 2021; Garin and Rothbaum, 2025), the localized spillovers generated by smaller innovative startups remain poorly understood. This paper shows that successful innovative startups causally raise nearby property values and business activity. We study U.S. startups that file patent applications and compare the evolution of local real estate prices and business formation around startups whose patents are granted versus rejected, using variation in patent-examiner leniency as an instrument for patent approval. We thereby provide new evidence on the spatial externalities of startup-driven innovation.

Our analysis combines nationwide micro-level data on patent filings, property transactions, and business establishments to measure the local economic impacts of startup innovation. We link USPTO patent application records to geocoded startup addresses, allowing us to pinpoint the timing and location of each innovation event. We then merge these data with repeat-sales property transactions and establishment-level counts to construct a panel that tracks real estate markets and business activity within 10 miles of each innovating startup. Outcomes are summarized as distance gradients around each firm, enabling us to compare how local housing, commercial real estate, and business formation evolve around startups whose patents are granted versus those whose applications are rejected.

To identify causal effects, we exploit the quasi-random assignment of patent examiners who differ in their propensities to grant patents, following the instrumental-variable approach pioneered by Sampat and Williams (2019) and further developed by Farre-Mensa, Hegde and Ljungqvist (2020). For each application, we construct an examiner leniency measure defined as

the examiner's historical grant rate on previously assigned cases and use this as an instrument for patent approval. Because examiners are effectively randomly assigned within technology fields, differences in leniency generate plausibly exogenous variation in the probability that otherwise similar startups obtain a patent, while remaining orthogonal to local shocks to property markets or business activity. Intuitively, we compare areas around startups assigned more lenient examiners, which are more likely to receive a patent, with areas around similar startups assigned stricter examiners, thereby isolating the causal impact of successful innovation on nearby economic outcomes.

Our main results show that patent approvals by innovative startups generate sizable positive spillovers in nearby real estate markets. Using repeat-sales data, we estimate that a patent grant increases annualized property returns within 10 miles by about 0.4 percentage points for single-family homes, 0.6 percentage points for multi-family residential properties, 0.5 percentage points for commercial properties (including office, retail, and industrial), and 0.9 percentage points for vacant land. These effects attenuate with distance from the innovating firm, indicating that the spillovers are highly localized. They are also strongest when a startup is the first successful innovator in an area; subsequent successful startups generate additional but smaller gains.[1]

The estimated real estate return premiums likely reflect not only the direct impacts of successful startups, but also changes in neighborhood composition and forward-looking expectations of local growth that are capitalized into property values. Consistent with this interpretation, we document that patent-granted startups significantly stimulate nearby patenting by other startups: following a patent approval, the number of patent applications by other startups within 10 miles rises by about ten, and the number of patent grants increases by roughly five. We also find increased business formation, with the number of establishments

1. These return premiums are partly, but not fully, explained by shorter holding periods around successful startups. Moreover, repeat-sales transactions completed entirely before the patent approval date exhibit no differential trends between treated and comparison areas, providing evidence against pre-existing differences and supporting the exclusion restriction underlying our empirical strategy.

within 10 miles in the same industry as the focal startup rising by four to six. Finally, the share of transactions in which non-individual buyers purchase from individual sellers increases by about 1.3 percentage points in the surrounding area, suggesting that institutional and other sophisticated investors enter these neighborhoods following startup innovation.

This paper contributes to several strands of literature. First, it contributes to research on the local economic impacts of large firm activity. A large literature shows that new manufacturing plants and government-led industrial expansions generate substantial local gains in productivity, wages, and employment, and that corporate relocations and big-box retail entry affect nearby housing markets and business outcomes.[2] Other work highlights heterogeneous or negative local effects: Qian and Tan (2021) document that high-skilled firm entry benefits high-skilled homeowners but harms nearby low-skilled renters, and Bernstein, Colonnelli, Giroud and Iverson (2019) find that liquidation, relative to reorganization, produces highly localized negative spillovers, particularly in service sectors. We contribute to this literature by shifting the focus from large plants and major corporate events to small, innovative startups, and by providing causal evidence that early-stage innovative firms generate sizable, highly localized spillovers in both real estate values and business activity.

Second, this paper contributes to the literature on the role of startups in local economic growth. Prior work shows that new firm formation, startups, and local knowledge spillovers are key drivers of employment growth and urban economic dynamism (Glaeser, Kallal, Scheinkman and Shleifer, 1992; Haltiwanger, Jarmin and Miranda, 2013; Adelino, Ma and Robinson, 2017; Carlino and Drautzburg, 2020). A related strand develops macro and spatial frameworks, together with reduced-form evidence, to study how shocks to startup activity shape aggregate output, productivity, and regional trajectories (Glaeser, Kerr and Kerr, 2015; Gourio, Messer and Siemer, 2016; Walsh, 2023). We add to this literature in two ways: first, by shifting the

2. See, among others, Greenstone et al. (2010); Monte, Redding and Rossi-Hansberg (2018); Garin and Rothbaum (2025) on manufacturing plants and industrial expansions, and Hu, Tsang and Wan (2025); Pope and Pope (2015) on corporate relocations and big-box retail entry.

focus from startup formation to the local spillover effects of successful innovation by early-stage firms; and second, by providing causal evidence on how these innovation shocks affect nearby real estate markets and business activity.

Third, this paper contributes to the innovation literature, particularly studies examining how firm-level innovation affects both the originating firm and the broader economy. Prior work using examiner leniency shows that patent approvals substantially increase firm growth and subsequent innovation (Farre-Mensa et al., 2020), and a related line of research documents how a firm's patenting activity shapes the innovation outcomes of others, through both delays in approval and competitive shocks (Hegde, Ljungqvist and Raj, 2022; Matray, 2021). At a more aggregate level, innovation measured by citation-weighted patent counts plays a central role in explaining medium-run economic growth (Kogan, Papanikolaou, Seru and Stoffman, 2017). We contribute to this literature in two ways: (i) we provide new micro-level evidence that successful startup innovation causally increases both patent applications and approvals among nearby startups; and (ii) we show that these innovation shocks also stimulate broader local business activity, amplifying the overall impact of startup innovation on the surrounding economy.

Finally, this paper contributes to the literature on how local economic activity and buyer demographics are capitalized into housing markets. Prior work shows that expectations about regional economic growth and various local shocks are reflected in housing prices, and that house prices are correlated with metro-level innovation measured by patent citations.[3] We expand this literature in two ways: (i) we provide causal evidence on the real estate market consequences of startup-driven innovation, showing that startup patent approvals generate return premiums across all major property types; and (ii) we document that these innovation

3. See Coulson, Liu and Villupuram (2013) on expectations about metro-level growth; Butler, Fauver and Spyridopoulos (2019); Hartman-Glaser, Thibodeau and Yoshida (2023) on IPO activity and nearby real estate values; Beracha, Kim, Wintoki and Xi (2025); Aiello, Kotter and Schubert (2022) on the effects of high-skilled migration and migrants with housing equity gains on local house prices; and Beracha, He, Wintoki and Xi (2022) on the relation between house prices and metro-level innovation.

shocks are accompanied by shifts in buyer composition, with a higher share of transactions involving non-individual (often institutional) buyers.

The rest of the paper proceeds as follows. Section 2 describes the institutional background, introduces the three datasets, and presents descriptive evidence on the relationship between patent approvals and nearby economic outcomes. Section 3 develops the instrumental variable strategy, reports the causal estimates of startup patent approvals on nearby property returns and business activity, and examines robustness, heterogeneity, and indirect mechanisms. Finally, Section 4 concludes, discusses limitations of the analysis, and suggests directions for future research.

2 Institutional Background and Data

In this section, we provide background on the U.S. patenting process and describe the data used in the analysis. We first outline the patent application process, then introduce the datasets used in the empirical analysis: the USPTO PatEx Patent Examination Research Dataset, the ATTOM property transactions data, and the Data Axle business registry.

2.1 The Patent Application Process

When a patent application is filed with the USPTO, it undergoes a preliminary review for completeness and is classified by technology area (Marco, Toole, Miller and Frumkin, 2017). The application is then forwarded to the appropriate Technology Center and assigned to a group art unit specializing in that field. Within the art unit, the application is placed in an examiner's queue, generally ordered by filing date. The specific assignment to an examiner is effectively random, conditional on the art unit and timing (Farre-Mensa et al., 2020). This institutional feature implies that assignment is orthogonal to the underlying quality of the invention or

applicant. Examiners themselves vary in their propensity to grant patents: some are more lenient while others are more stringent, introducing a degree of randomness in whether a given application is approved (Farre-Mensa et al., 2020).

Once an examiner takes up the application, they evaluate whether the submission meets all legal requirements. The examiner ensures that the claims are directed to patent-eligible subject matter, that the specification adequately describes and enables the invention, and that the claims distinctly define the invention. The examiner conducts a thorough prior-art search to assess novelty and non-obviousness. Based on this examination, the examiner issues an initial decision known as the *first action on the merits*. In the vast majority of cases, this first action is a non-final rejection outlining the examiner's objections; only about 10 to 15 percent of applications receive a first-action allowance without any rejections (Carley, Hegde and Marco, 2015). The USPTO communicates the first-action decision to the applicant via an official letter, at which point the applicant also learns the examiner's identity (Kogan et al., 2017).

The applicant then has an opportunity to respond, typically by amending claims or presenting arguments to address the examiner's concerns. Prosecution may involve multiple rounds: after the applicant's response, the examiner can either allow the application or issue a final rejection if issues remain. Even after a final rejection, the process can continue: the applicant may abandon the application at that point, or choose to file a Request for Continued Examination (RCE) to reopen prosecution, or appeal the decision to the Patent Trial and Appeal Board. This iterative process concludes only when the application is either granted as a patent or permanently abandoned by the applicant. Notably, the USPTO never conclusively "rejects" an application: any rejection can in principle be overcome or appealed, so a "rejected" application is one that the applicant elects to abandon after receiving an adverse decision (Kogan et al., 2017).

Patent examination is a lengthy process. It often takes several months (up to a year or more) for a newly filed application to be assigned to an examiner and receive an initial review (Kogan

et al., 2017). On average, in our context, an application waits about 0.7 years to be taken up by an examiner and roughly another year for the first-action decision. The final disposition (grant or abandonment) is typically reached only after multiple rounds of interaction. In fact, the final decision usually comes about three years after the initial filing on average.[4] As such, the first office action provides an early but meaningful signal about an application's prospects long before the ultimate outcome is decided. Following the prior literature (Kogan et al., 2017; Kline, Petkova, Williams and Zidar, 2019; Farre-Mensa et al., 2020), we use the first-action decision to define the treatment and its timing.

2.2 Patent Data and Startups

Our patent data come from the USPTO's Patent Examination Research Dataset (PatEx), which contains detailed records of patent applications and their prosecution histories. The PatEx data, derived from the USPTO's Patent Application Information Retrieval (PAIR) system, includes "research-ready" information on application filings, actions, and outcomes, covering unpublished and abandoned applications. We restrict our sample to applications filed on or after 2001, following the enactment of the American Inventors Protection Act of 1999, which mandated publication of patent applications. As a result, for applications filed in 2001 and later, the applicant's identity is observable regardless of the application's eventual outcome (Carley et al., 2015; Farre-Mensa et al., 2020). This allows us to include both granted and rejected (abandoned) applications in the analysis. We use all original (non-continuation) utility patent applications filed between 2001 and 2009 that received a final disposition by 2013.

We follow Farre-Mensa et al. (2020) in defining the set of startup firms in the patent data. Specifically, we focus on first-time patent applicants that are private, incorporated businesses based in the U.S. We exclude not-for-profit organizations (e.g., universities, government labs),

4. Strictly speaking, patent applications are never irrevocably "rejected" by the USPTO; rather, an application that fails to secure a patent is one that the applicant ultimately abandons following an examiner's final rejection. For expositional convenience, we refer to such outcomes as rejections.

as well as any applicant that is publicly traded or a subsidiary of an existing firm. We further restrict the sample to applicants that qualified for "small entity" status under the U.S. Small Business Act (i.e., those below the employee-size threshold for discounted filing fees). In addition, to ensure we are capturing new ventures, we require that each firm's first patent application in our data occurs on or after January 1, 2001, and that the firm had no patenting activity in the prior 25 years. Applying these filters yields a set of young, innovative companies that we treat as startups, consistent with the sample definition in Farre-Mensa et al. (2020).

For each startup in the sample, we identify the firm's address at the time of the patent application. We obtain the assignee address listed on the patent application (from the USPTO assignment records in PatEx) and use the Bing Maps API to geocode this location. We ensure that this address corresponds to the firm's headquarters or primary place of business, not the inventor's personal address. Our final dataset includes 29,487 startups that filed a first patent application between 2001 and 2009 and received a first-action decision by the end of 2009 (with a final patent grant or abandonment by the end of 2013). These startups are located across 8,504 ZIP codes, covering all 50 U.S. states.

A limitation of the USPTO PatEx dataset is that it does not record industry classifications for patent applications. Because we aim to estimate within-industry agglomeration effects, we would ideally observe the industry of each startup and nearby establishment. In practice, however, we only observe technology classifications: each application is assigned to an art unit and one or more classes in the US Patent Classification (USPC) system, which contains more than 400 classes describing specific technological domains. To bridge this gap, we rely on the probabilistic concordance developed by Goldschlag, Lybbert and Zolas (2020), which links USPC classes to 6-digit NAICS industries using an algorithm that matches patent abstracts to industry descriptions. This "algorithmic links with probabilities" approach assigns each technology class a vector of industry match probabilities rather than a single deterministic

sector. We use these probabilities to map applications to likely industries and assign startups to the sector with the highest implied match probability.

An important variable in our analysis is the leniency of the patent examiner assigned to each startup's first application. We define the *Examiner Leniency* for examiner j in art unit a at time t as the proportion of applications that examiner j has approved (allowed) out of all applications he or she has examined in that art unit, prior to examining startup i's application:

$$\text{Examiner Leniency}_{j,a,t} = \frac{n_{\text{granted},j,a,t}}{n_{\text{reviewed},j,a,t}},$$

where $n_{\text{granted},j,a,t}$ and $n_{\text{reviewed},j,a,t}$ are the total numbers of patent applications examiner j has granted and reviewed, respectively, in art unit a up to time t (the filing date of startup i's application). We calculate this rate using the examiner's prior examination history in the PatEx data, based on the full universe of patent filings rather than those from startups alone. To ensure the measure is reliable, we exclude observations where the assigned examiner had reviewed fewer than ten applications before the focal one.

By construction, the computation of examiner leniency for a given application excludes that application itself, avoiding any mechanical correlation between the instrument and the startup's outcome. At the firm level, we assign each startup the leniency rate of the examiner who reviewed its first patent application. This examiner leniency serves as the baseline instrumental variable for our analysis, as it captures quasi-random variation in the stringency of the patent review process faced by different startups.

Table 1 reports summary statistics comparing startups whose first patent applications were ultimately approved (granted) to those whose applications were ultimately rejected (not granted). Panel A shows that at the time of the patent filing, the median startup in our sample is about two years old, has eight employees, and about $0.8 million in annual sales. Startups with

unsuccessful applications exhibit slightly lower pre-filing employment and sales growth than those with successful applications, but these differences are not statistically significant.

The geographic distribution of these startups is uneven, though certain counties stand out. Consistent with expectations, counties traditionally viewed as startup hubs, such as Santa Clara, Los Angeles, San Diego, and Orange County, contain the largest number of firms in our analysis sample. Nevertheless, as shown in Appendix Figure A.1, a substantial number of observations come from other parts of the country, indicating that our analysis is not confined to established innovation clusters. In Section 3.2.2, we show how treatment effect estimates vary with prior exposure to innovators.

2.3 ATTOM Property Tax and Deed Data

We assemble property transaction data from ATTOM's deed records and tax assessor files, which we use to track real estate market outcomes around each startup. The ATTOM data provide property-level information on transaction dates and prices, buyer and seller names, and property characteristics and assessed values from tax records. For each recorded transaction, we observe the exact property address and detailed use codes that we map into four broad property categories: single-family residential, multi-family residential, commercial, and vacant land.[5] Using these records, we construct a repeat-sales dataset of properties that transact at least twice. Each repeat-sale observation consists of a purchase (buy) and a subsequent sale of the same property. For every such pair, we compute the annualized unlevered return between the two transactions:

$$R = \left(\frac{\text{sell price}}{\text{buy price}}\right)^{1/\text{holding period in years}} - 1,$$

5. See Appendix Table A.1 for the mapping of detailed property use codes into these four categories.

and use this as our primary outcome variable. All transaction prices are converted to real 2010 dollars using the Consumer Price Index.

Additionally, we incorporate shifts in buyer characteristics observed in the ATTOM data as part of the mechanism analysis in Section 3.4. Specifically, the dataset indicates whether a buyer is an individual or a non-individual entity, such as a corporation or LLC. It also includes buyer mailing addresses, which we use to infer whether the buyer is in-state or out-of-state relative to the property's location. Using this information, we construct indicator variables for ownership transitions between individual and non-individual owners, as well as between in-state and out-of-state sellers and buyers.

We apply several filters to ensure data quality. First, we drop any repeat-sale observations with implausible outlier values (in either buy price, sell price, or resulting return), defined using the interquartile range (IQR) rule calculated at the county-by-property-type level. Next, we link each repeat sale to the property's tax assessment history to detect major property improvements. In particular, we compare the property's effective year built recorded in the tax assessor data after the second transaction to the timing of the first transaction. We exclude any repeat sale where the first purchase date precedes a recorded renovation or rebuild, so that our measured price appreciation is not conflated with substantial new construction.[6] We also extract key property characteristics from the tax records (e.g., lot size for all properties; building square footage and year built for commercial; number of bedrooms and bathrooms for residential) to use as control variables in the property-level analysis (Section 3.3).

Using the geocoded property coordinates, we identify which repeat-sale transactions occurred in the vicinity of each startup. Our baseline definition of a "local" transaction is one within a 10-mile radius of the startup's location. A 10-mile radius provides a broad local housing market around each firm, balancing coverage and tractability, and roughly corresponds

6. We do not apply this restriction for vacant land transactions, since land can appreciate through development even if the land itself was unimproved at the time of sale.

to typical commuting distances in the U.S. (pre-pandemic).[7] Within 10 miles of the 29,487 startups, we identify a total of 15,234,534 single-family, 4,471,333 multi-family, 252,061 commercial, and 141,654 vacant land repeat-sale transactions over our sample period.

For the firm-level analysis in Section 3.2, we aggregate the repeat-sale data to construct measures of local real estate market performance for each startup. In particular, we group transactions by startup, property category, county, buy year, sell year, and distance band (in miles from the startup), and we compute the median values of the key variables within each group: the unlevered return R, the initial purchase price, and the subsequent sale price. We then consider three distinct samples of these aggregated repeat-sale observations, defined by the timing of the transactions relative to the startup's first-action date: (i) a *Baseline sample*, consisting of cases where the buy date is before and the sell date is after the startup's first-action date (i.e., straddling the patent decision); (ii) a *Post-only sample*, where both the buy and sell dates occur after the first-action date; and (iii) a *Pre-only (placebo) sample*, where both transactions occurred before the first-action date. These samples allow us to compare property market dynamics around the time of the patent decision (baseline) to purely post-treatment trends and a pre-treatment counterfactual. Summary statistics for the firm-level repeat-sales sample are presented in Appendix Table A.2.

2.4 Data Axle Business Registry

We use establishment-level business data from Data Axle (formerly known as infoUSA) to measure local economic activity surrounding each startup. Data Axle compiles information on virtually all U.S. businesses from a variety of sources, including telephone directories, business filings, surveys, and web mining. For each establishment, the dataset provides attributes such as name, address, industry classification (NAICS/SIC), and the year the business was established.

7. According to the Census Bureau's American Community Survey, the average one-way commute in 2006 was about 25 minutes. Proprietary data from Gusto suggest that the mean commute distance had risen to approximately 10 miles by 2019.

Comparisons with official statistics (e.g., the Census Bureau's Zip Code Business Patterns) indicate that Data Axle's coverage of total establishments is very close to the government counts, while offering much finer geographic detail (exact addresses) and higher frequency (annual updates).

One limitation of the raw Data Axle data, however, is a tendency to overcount establishments by listing multiple records for what is operationally a single business. For example, each individual lawyer in a law firm might appear as a separate "business". To address this issue, we preprocess the data by merging records that likely refer to the same physical business: specifically, we collapse any entries that share either a common Data Axle unique identifier or the same business name and ZIP code. We also drop any records with only a P.O. Box address, since those do not correspond to a brick-and-mortar establishment.

Using historical Data Axle archives dating back to 1997, we construct measures of local business density around each startup over time. For each startup, we identify all establishments in the Data Axle records located within a 10-mile radius of the startup's address, based on the geocoded coordinates described above. We then aggregate these establishments by year and by industry supersector, defined by two-digit NAICS codes, to create a panel of local business counts. Thus, for each startup, we observe the number of nearby establishments in each broad sector for every year before and after the startup's patent decision.

Reassuringly, we find that business counts from Data Axle yield very similar patterns to Zip Code Business Patterns at the aggregate level (see Appendix B). The resulting panel of local business counts (by year and sector, for each startup) forms the basis for our analysis of agglomeration effects in Section 3.4.2. In Table A.3, we report summary statistics for two key variables from this Data Axle panel: the number of establishments within 10 miles of the startup in the year prior to the first-action decision, and the ten-year change in the number of nearby establishments following the first-action date.

2.5 Descriptive Analysis

We present descriptive regressions using the firm-level ATTOM repeat-sales sample and the Data Axle establishment sample. Specifically, we examine how key outcome variables, including unlevered returns, buy and sell prices in local real estate markets, and the number of nearby establishments, vary depending on whether the startup received a first-action patent grant.

First, for the repeat-sales sample aggregated at the firm level as described in Section 2.3, we estimate the following weighted least squares regression:

$$y_{i,k,c,t_0,d} = \beta_0 + \beta_1 \mathbf{1}(\text{Approved})_i + \lambda_{gau_i \times appl_year_i} + \lambda_k + \lambda_{c \times t_0} + \lambda_d + \epsilon_{it}, \tag{1}$$

where the unit of observation is startup i, disaggregated by property type k, county c, initial purchase year t_0, and distance bin d (measured in miles from the startup). Outcome variables y include the median unlevered returns, the log of the median sell price, and the log of the median buy price. The indicator $\mathbf{1}(\text{Approved})_i$ denotes whether the firm received a first-action patent grant. The specification includes fixed effects for art unit-by-application year ($\lambda_{gau_i \times appl_year_i}$), property type ($\lambda_k$, as originally coded by ATTOM), county-by-purchase year ($\lambda_{c \times t_0}$), and distance bin (λ_d). The estimation sample comprises repeat-sale transactions located within 10 miles of each startup (i.e., $d \in \{1, \ldots, 10\}$). We weight regressions by the number of repeat-sale transactions in each i-k-c-t_0-d cell.

The primary parameter of interest, β_1, captures the average difference in local real estate outcomes between patent-granted and non-granted startups. We evaluate four subsamples based on property use category: single-family residential, multi-family residential, commercial, and vacant land. Figure 1a plots the estimated β_1 coefficients and their 95 percent confidence intervals. We find consistent patterns across all four property use categories. In the baseline sample, defined as transactions where the property's buy date precedes the startup's first-action

decision and the sell date follows it, both buy and sell prices are higher near patent-granted startups. Importantly, the differential is larger for sell prices, leading to higher unlevered returns in local markets around patent-granted startups relative to non-granted startups.

The placebo sample, which includes transactions where both the buy and sell dates occur before the startup's first-action decision date, shows a different pattern. Although point estimates indicate slightly higher buy and sell prices around patent-granted startups for single-family, multi-family, and commercial properties, none of these coefficients is statistically different from zero at the 5 percent level. Moreover, the sell-price differentials are generally smaller than the corresponding buy-price differentials, yielding negative point estimates for unlevered returns. This pattern suggests relatively lower returns near patent-granted startups in the pre-treatment period. Vacant land exhibits the opposite pattern: both buy and sell prices are slightly lower while unlevered returns are slightly higher around patent-granted startups. However, none of these estimates is statistically significant.

Finally, the post-treatment sample, comprising transactions where both the buy and sell dates occur after the startup's first-action decision, shows patterns consistent with the baseline sample but with substantially smaller magnitudes. Here, buy and sell prices are still higher near patent-granted startups, but unlevered returns are only slightly higher and not statistically significant.

For the establishment sample, we estimate the following OLS regression:

$$y_{i,k,c,t} = \beta_0 + \beta_1 \mathbf{1}(\text{Approved})_i + \lambda_{gau_i \times appl_year_i} + \lambda_{c \times appl_year_i} + \epsilon_{ikct}, \tag{2}$$

where $y_{i,k,c,t}$ denotes either the number of establishments or the change in establishments in sector k in county c within 10 miles of startup i in year t, measured relative to the first-action date of startup i. This specification includes art unit-by-application year fixed effects ($\lambda_{gau_i \times appl_year_i}$) and county-by-application year fixed effects ($\lambda_{c \times appl_year_i}$).

Figure 1b shows the estimated β_1 coefficients and their 95 percent confidence intervals. We find that prior to receiving patent grants, patent-granted startups tend to be located in areas with fewer local establishments. However, in the 10 years following their first-action dates, these areas experience substantial increases in the number of nearby businesses, both in absolute and percentage terms.

Overall, the correlation analysis suggests that local housing markets and business activity initially show weaker performance around startups that are eventually granted patents but experience substantial improvement after the patent signal is revealed. This pattern is especially evident for real estate transactions purchased before the first-action decision date and for local business formation in the years following innovation events.

3 Main Empirical Analysis

This section presents the empirical framework and main results. We begin by outlining the instrumental variable strategy that leverages patent examiner leniency. We then report the baseline IV estimates of the impact of patent-granted startups on local real estate returns, followed by placebo tests, heterogeneity analyses, and estimates of the cumulative effects of multiple firms. We close by examining the mechanisms behind these effects: innovative startups increase nearby patenting activity and establishment counts, and shift the composition of homebuyers.

3.1 Empirical Strategy: Patent Examiner Leniency Instrument

The main empirical concern is that startups endogenously select into specific locations. For example, startups may choose lower-cost areas to reduce expenses or gravitate toward high-growth, amenity-rich locations to attract skilled workers. These locational preferences

may correlate with both startup success and real estate dynamics, raising endogeneity concerns. In particular, if inherently more innovative startups have systematically different preferences, such that locational characteristics and patenting outcomes are correlated, the OLS estimates in Section 2.5 are biased.

To causally identify the effects of innovative startups, we exploit variation in patent examiner leniency as an instrumental variable, as defined in Section 2.2. This strategy leverages the quasi-random assignment of applications to patent examiners, who systematically vary in their approval tendencies. The strategy relies on two institutional features: (i) applications are randomly assigned to examiners within a USPTO art unit, conditional on the technology field, and (ii) examiners differ in their average grant rates. Our use of this instrument follows Sampat and Williams (2019), and we further draw on evidence from Farre-Mensa et al. (2020), who show that examiner leniency is uncorrelated with applicant characteristics. Additionally, we present placebo test results in Section 3.2.1 and first-stage regression results in Appendix Table A.4 in support of these assumptions.

3.2 Firm-Level Analysis: Real Estate Market Externalities

As our main IV strategy, we adopt patent examiner leniency as an instrument for $\mathbf{1}(\text{Approved})_i$ in Equation 1. Specifically, the first-stage regression equation for the two-stage least squares estimation is:

$$\mathbf{1}(\text{Approved})_i = \gamma_0 + \gamma_1 \cdot (\text{Examiner Leniency})_i + \tilde{\lambda}_{gau_i \times appl_year_i} + \tilde{\lambda}_k + \tilde{\lambda}_{c \times t_0} + \tilde{\lambda}_d + \tilde{\epsilon}_{it}, \quad (3)$$

where $(\text{Examiner Leniency})_i$ is defined as described in Section 2.2. The second-stage equation is:

$$y_{i,k,c,t_0,d} = \beta_0 + \beta_1 \mathbf{1}(\widehat{\text{Approved}})_i + \lambda_{gau_i \times appl_year_i} + \lambda_k + \lambda_{c \times t_0} + \lambda_d + \epsilon_{it}, \quad (4)$$

where $\mathbf{1}(\widehat{\text{Approved}})_i$ is the predicted value from Equation 3. We weight regressions by the number of repeat-sale transactions in each i-k-c-t_0-d cell. The primary outcome variable y is the median annualized unlevered return.[8] We also examine related outcomes including log buy price, log sell price, and holding duration.

3.2.1 Baseline Effects: Innovation Increases Nearby Real Estate Returns

We first examine the baseline sample, consisting of repeat-sale transactions where the property was purchased before and sold after the first-action date of the associated patent application. Panel A of Table 2 reports IV estimates from Equation 4 by property use type. All specifications include the full set of fixed effects and cluster standard errors at the art unit level. Panel A of Appendix Table A.4 reports the corresponding first-stage results, with examiner leniency coefficients ranging from 0.66 to 0.67 and standard errors between 0.022 and 0.028.

Across all four property uses, the IV estimates in Columns (1), (3), (5), and (7) of Table 2 show that patent-granted startups increase local real estate returns by 0.38 percentage points for single-family homes, 0.64 percentage points for multi-family properties, 0.50 percentage points for commercial real estate, and 0.91 percentage points for vacant land. All estimates are statistically significant at the 1 percent level. The return premiums are larger for multi-family and commercial properties than for single-family homes, suggesting that the effects of innovation are particularly pronounced in denser, likely downtown, neighborhoods. Vacant land exhibits the largest premium, consistent with the idea that expectations of future development are capitalized strongly into land values.

We then investigate the sources of these return premia by re-estimating the model with log buy price, log sell price, and holding term as alternative outcome variables. Table 3 shows that the higher returns are primarily driven by lower buy prices and shorter holding periods in treated

8. Outcome variables y include the median annualized unlevered return, the log of the median sell price, and the log of the median buy price. We also use the average annualized unlevered return as an alternative outcome and find that the results remain consistent and statistically significant.

areas. Specifically, IV estimates for log median sell prices are not statistically significant at the 10 percent level, suggesting similar exit prices across treated and control areas. By contrast, the IV estimates for log median buy prices are negative and generally larger in absolute value, indicating that properties in treated areas are purchased at lower prices, which likely reflects differences in quality or size. Even when these estimates are not statistically significant at conventional levels, they are meaningfully larger than those for sell prices. Additionally, IV estimates for holding term are negative and statistically significant, showing that properties in treated areas tend to be re-sold more quickly, further contributing to higher annualized returns.

To further assess the role of the holding-term channel, we augment Equation 4 and the corresponding first-stage regressions by including fixed effects for county-by-holding term (in years), and report the corresponding IV estimates in columns (2), (4), (6), and (8) of Table 2. The coefficient estimates remain positive and statistically significant: 0.20 percentage points for single-family homes (significant at the 1 percent level), 0.39 percentage points for multi-family properties (1 percent level), 0.23 percentage points for commercial properties (1 percent level), and 0.58 percentage points for vacant land (10 percent level). The reductions in the coefficient estimates relative to the baseline are somewhat comparable in magnitude across property types, with the largest reduction for commercial properties (a decline from 0.50 to 0.23 percentage points, or about 54 percent) and the smallest reduction for vacant land (from 0.91 to 0.58 percentage points, about a 36 percent reduction). In summary, we find that the shorter holding-term in the treated area plays an important role, but it does not fully account for the estimated return premia.

Next, we examine the placebo and post-treatment samples. Panel B of Table 2 reports results for placebo transactions in which both the purchase and sale occurred before the innovation signal. That is, prior to the first-action date of any associated startup patent application. In this pre-treatment sample, we find that the IV coefficient estimates are not statistically significant across all four property types at the 10 percent significance level. This null result provides

empirical support for the validity of our identification strategy. Specifically, it suggests that the estimated return premiums in the baseline sample are not driven by pre-existing trends in real estate returns that happen to coincide with later innovation activity. The absence of a treatment effect in this placebo setting helps to rule out the possibility that our instrument is spuriously capturing upward price trends or unobserved local shocks that precede innovation.

Panel C of Table 2 presents results for post-treatment transactions in which both the purchase and sale occurred after the innovation signal was revealed. We find little evidence that these later buyers capture the return premium: estimates are statistically insignificant across all property types at the 10 percent level. These null results persist when holding-term fixed effects are included in Columns (2), (4), (6), and (8). The only exception is commercial real estate returns in Column (6), where the coefficient estimate of a 0.07 percentage point increase, though statistically significant at the 10 percent level, is substantially smaller than the corresponding baseline estimate of 0.23 percentage points. This finding suggests that most of the appreciation is realized by earlier investors who purchased prior to the innovation signal.

Figure 2 further characterizes the spatial and temporal patterns of the return premiums. We first stratify treated observations by distance bin $d \in \{1, \ldots, 10\}$ and re-estimate the two-stage least squares separately for each bin. Figure 2a displays the IV estimates of β_1 with 95 percent confidence intervals. The results reveal a consistent pattern of declining effects with increasing distance: return premiums are consistently higher for properties located closer to the innovating startup across all four property use categories. Within one mile, the estimated effects, without controlling for holding years, are especially pronounced: 7.26 percentage points for single-family homes (t-statistic = 2.84), 3.61 percentage points for multi-family properties (t = 2.41), 4.04 percentage points for commercial properties (t = 2.84), and 13.14 percentage points for vacant land (t = 1.84). In contrast, at distances of 9 to 10 miles, the return premiums are much smaller: 0.88 percentage points for single-family homes (t-statistic = 3.57), 1.30 percentage points for multi-family (t = 3.37), 1.28 percentage points for commercial properties (t = 4.65),

and 2.36 percentage points for vacant land (t = 2.34). The right panel of the figure confirms this pattern, even after including holding-term fixed effects.

Figure 2b presents the cumulative effects on real estate returns by sale timing relative to treatment (τ). To construct these estimates, we re-estimate β_1 from the IV specification in Equation 4, including county-by-holding-term fixed effects, using subsamples defined by the sale year relative to the first-action grant decision. This yields an estimate denoted $\beta_{1,\tau}$ for each event time τ. For $\tau \leq 0$, we rely on the placebo sample (transactions sold before treatment), restricting the treated group to properties sold $|\tau|$ years prior to treatment. For $\tau > 0$, we estimate $\beta_{1,\tau}$ using the combined baseline and post-treatment samples, restricting the treated group to transactions sold τ years after treatment while allowing the control group to include all relevant observations. We then plot the cumulative effect $\sum_{t=0}^{\tau} \beta_{1,t}$ to characterize the overall price impact over time.

The estimates for $\tau < 0$ in Figure 2b show little evidence of systematic price differentials or pre-trends between treated and control areas. Over the five years prior to treatment, properties near treated startups experienced modest cumulative declines of roughly 0.6 percent for single-family, multi-family, and commercial properties, while vacant land values increased by about 0.6 percent. These magnitudes are economically small and not indicative of differential pre-treatment dynamics, consistent with the pooled estimate shown earlier in Panel B of Table 2.

In contrast, we observe a substantial and immediate increase in property values following the innovation signal. Within three years, cumulative price effects reach approximately 4.8 percent for single-family homes, 5.7 percent for multi-family properties, 5.1 percent for commercial properties, and 8.9 percent for vacant land. These short-run price responses likely reflect not only the direct effects of startup growth, including firm expansion and local labor market spillovers documented by Farre-Mensa et al. (2020), but also the capitalization of forward-looking expectations about future neighborhood economic activity, which we examine further in Section 3.4.

The estimated premia then moderate and decline through roughly year eight. This mid-run adjustment may partly reflect shifts in transaction composition, as properties sold immediately after innovation in treated areas may disproportionately capture larger return premia. It may also reflect gradual supply responses that take time to materialize in local real estate markets. Beginning around year nine, the estimates indicate a renewed upward trend in prices, consistent with longer-run agglomeration forces and increased local economic activity reinforcing property value appreciation. We discuss how the timing of real economic responses aligns with this temporal pattern of return premia in Section 3.4.

3.2.2 Robustness of Real Estate Return Premia

We begin by examining heterogeneity in return premia by historical exposure to innovation and by technology type. First, we use PatentsView data to count the total number of granted patents in each startup's city during the 20-year period preceding the startup's first-action date. We then re-estimate the return premia across four treatment subgroups defined by quartiles of this historical patenting intensity: "Q1" (lowest prior patenting activity), "Q2," "Q3," and "Q4" (highest prior patenting activity). Figure 3a reports the results. Although the coefficients across quartiles are not always statistically distinguishable, the point estimates suggest that areas with the least prior patenting activity tend to exhibit the largest return premia. This pattern is consistent with the interpretation that early innovation in a location generates the largest marginal externality.

Two points deserve attention. Even in the highest-exposure quartile, successful startups continue to generate statistically significant return premia at the 5 percent level, indicating that later-stage innovators still confer positive spillovers, albeit of smaller magnitude. Moreover, the declining pattern does not necessarily contradict agglomeration theories: the first successful innovator may exert a disproportionate impact not only directly, but also by attracting subsequent firms and contributing to cluster formation. We explore this channel further in Section 3.4.1.

Figure 3b presents heterogeneity in return premia by technology category. We classify patents using the NBER technology classification developed by Hall, Jaffe and Trajtenberg (2001), which aggregates U.S. Patent Classification (USPC) codes into six broad groups. In the baseline specification, all technology categories display positive and statistically significant return premia at the 5 percent level. Most categories remain positive after controlling for holding-term fixed effects, though magnitudes and statistical precision generally decline. Patents in the Chemical and Electrical & Electronics categories exhibit consistently larger point estimates across specifications and property types, albeit with wider standard errors. Overall, we do not find statistically meaningful differences in return premia across historical exposure levels or technology types.

Additionally, we present results from alternative specifications as robustness checks in Figure 3c. The first specification ("Unweighted") replicates the baseline model but omits the weights based on the number of transactions in each county–firm–buy year–sell year–distance bin, and thus reports unweighted 2SLS estimates. Relative to the baseline, this approach assigns equal weight to areas regardless of transaction volume in a given period. The second specification ("Excl. overlap") accounts for properties located near multiple patent-applying firms. To isolate cleaner exposure, we restrict the sample to repeat-sale transactions for which, between the property's buy and sell dates, only one firm within a 10-mile radius receives a first-action patent decision. The third specification ("ZIP level") maintains the baseline structure but constructs the sample at the ZIP code–firm–buy year–sell year–distance bin level instead of the county level. Accordingly, we replace county-by-buy year fixed effects with ZIP code-by-buy year fixed effects and, when applicable, substitute ZIP code-by-holding year fixed effects in place of county-by-holding year fixed effects. This provides more granular fixed effects as controls but relies on more limited within-ZIP code variation. Finally, the fourth specification ("Excl. recession") excludes repeat-sale transactions during the financial crisis (2008–2009).

Figure 3c shows that the coefficient estimates overall remain statistically indistinguishable from the baseline estimates at the 10 percent level, with the exception of the commercial and vacant land return premia in the ZIP-level analysis, where more granular ZIP-by-buy year and ZIP-by-holding year fixed effects are included. One potential explanation for the smaller estimated effects is that the ZIP-level analysis may exclude most exposure to nearby startups, because such firms tend to affect most properties within a given ZIP code. As a result, this specification primarily captures exposure to more distant innovative startups, whose effects are smaller in magnitude, as shown in Figure 2a.

3.3 Property-Level Analysis: Effects of Multiple Innovative Startups

The previous section provides robust evidence that patent-granted startups significantly increase nearby real estate returns compared to those that are not granted. However, an important caveat applies to the interpretation of the coefficient magnitudes. Due to the spatial clustering of startups, some repeat-sale transactions classified as controls in the firm-level analysis may also be indirectly treated by their proximity to other patent-granted startups. In such cases, the estimated treatment effects may be attenuated, leading to conservative firm-level estimates. To address this concern and to further examine the cumulative effects of multiple innovative startups, we complement the firm-level analysis with a property-level approach.[9]

We conduct a property-level analysis that incorporates both the number of nearby patent-applying firms and the number of successful patent grants occurring between the property's buy and sell dates. This approach allows us to examine how effects vary across areas with different densities of innovative firms. The property-level estimation equation takes the following form:

$$r_{i,c,k,t_0,T} = \beta_0 + \beta_1 \mathbf{1}(N_{1,i} > 0) + \beta_2 N_i + \lambda_k + \lambda_{c\times t_0} + \lambda_{c\times T} + \epsilon_{ickt_0T}, \tag{5}$$

9. Unlike the firm-level analysis, this specification does not allow us to control for firm- and application-level characteristics, such as art unit-by-application-year fixed effects, as in Kline et al. (2019); Farre-Mensa et al. (2020). However, we provide a placebo analysis that supports the validity of the instrumental variable in this setting.

where i indexes a repeat-sale transaction in county c, property type k, buy year t_0, and holding term (in years) T, and r denotes the annualized unlevered return of the transaction. $N_{1,i}$ denotes the number of startups located within 10 miles of the property that received patent approval (first-action grant) between the property's buy and sell dates. The treatment indicator $\mathbf{1}(N_{1,i} > 0)$ equals one if at least one such startup received an approval during this period. While the treatment variable is binary, we assess treatment heterogeneity by examining $N_{1,i}$ to quantify the cumulative effects of multiple innovative startups without imposing a strong parametric assumption. N_i indicates the total number of nearby startups, including both patent-granted and non-granted firms. λ_k represents property type fixed effects, and $\lambda_{c\times t_0}$ denotes county-by-buy-year fixed effects. We also include county-by-holding-term fixed effects $\lambda_{c\times T}$ in the property-level analysis. We cluster all standard errors at the county level.

We instrument the treatment indicator $\mathbf{1}(N_{1,i} > 0)$ using the average leniency of the assigned patent examiners for all applications within 10 miles of the property that received first-action decisions between the buy and sell dates. The first-stage regression results are presented in Appendix Table A.4. Although the first stage is weaker than that for the firm-level regressions, since we aggregate multiple patent applications into a single binary treatment variable while averaging their patent examiner leniency, it remains strongly statistically significant across subsamples and specifications. Additionally, the exclusion restriction in the property-level analysis differs from that in the firm-level approach, which more closely follows existing work such as Farre-Mensa et al. (2020). Specifically, the exclusion restriction is $E[\epsilon_{ickt_0T} \cdot (\text{avg. leniency})_i \mid N_i, \lambda_k, \lambda_{c\times t_0}, \lambda_{c\times T}] = 0$. That is, conditional on the total number of startups within 10 miles of the properties, property type, county-by-buy-year, and county-by-holding-year fixed effects, the average leniency of patent examiners assigned to nearby patent applications is as good as random. In support of this assumption, we present placebo test results, as in the firm-level analysis.

3.3.1 Cumulative but Diminishing Returns to Nearby Innovative Startups

Table 4 presents the results from the property-level IV regressions. Panel A includes all repeat-sale transactions in which at least one patent application, whether approved or rejected, has a first-action date that falls between the property's buy and sell dates. Panel B examines heterogeneity by N_1, the number of approved applications, by estimating five separate regressions for treated transactions grouped into bins defined by $N_1 = 1$ and the quartiles of the remaining distribution (denoted by $Q_{N_1,1}$, $Q_{N_1,2}$, $Q_{N_1,3}$, and $Q_{N_1,4}$). Panel C uses repeat-sale transactions from the post-treatment period, where the first-action date of the associated patent application precedes both the property's buy and sell dates. Panel D uses repeat-sale transactions from the pre-treatment period, where both the buy and sell dates occur before the first-action date of the associated application. Columns (1), (3), (5), and (7) report baseline IV estimates for the four property use categories, while Columns (2), (4), (6), and (8) add controls for observable property characteristics. All standard errors are clustered at the county level.

Columns (1), (3), (5), and (7) in Panel A show a positive and statistically significant return premium across all property types at the 1 percent level. The estimated effects correspond to a 5.60 percentage point increase in annualized returns for single-family properties, 17.06 percentage points for multi-family, 7.57 percentage points for commercial, and 10.10 percentage points for vacant land. These magnitudes reflect the cumulative impact of multiple nearby innovative startups. For example, the average number of patent-granted startups within 10 miles of treated single-family transactions is 17.87 (see conditional means of N_1 in Table 4). Scaled by exposure, the average effect per startup is approximately 0.31 percentage points for single-family, 0.57 for multi-family, 0.32 for commercial, and 0.96 for vacant land. These per-startup estimates are consistent with the firm-level results and indicate particularly strong returns for multi-family and vacant properties. Columns (2), (4), (6), and (8) demonstrate the

robustness of these estimates to controlling for building-level characteristics, as described in Section 2.3.

Panel B provides further evidence of cumulative but diminishing marginal effects. As the number of nearby innovative startups increases, the total return premium rises, but the additional effect per startup declines. For single-family properties, the return premium increases from 1.49 percentage points for exposure to a single startup to 11.13 percentage points in the top quartile ($Q_{N_1,4}$), where the average number of startups is 62.71. In contrast, the per-startup effects are larger in lower exposure bins: 0.75 in $Q_{N_1,1}$ (calculated as 2.11/2.81), 0.64 in $Q_{N_1,2}$ (4.31/6.72), 0.44 in $Q_{N_1,3}$ (6.68/15.15), and only 0.18 in $Q_{N_1,4}$ (11.13/62.71). A similar pattern holds for multi-family properties, where the average per-startup effect declines from 2.19 percentage points for a single startup and 0.58 in $Q_{N_1,1}$ to 0.32 in $Q_{N_1,4}$. For commercial properties, the effect falls from 2.01 and 0.63 to 0.18, and for vacant land, from 3.82 and 2.46 to 0.69. These results further support the presence of diminishing marginal effects as the number of nearby innovative startups increases, consistent with earlier results in Section 3.2.2. Coefficient estimates controlling for building-level characteristics in Columns (2), (4), (6), and (8) show only minor changes.

Panels C and D yield results consistent with the timing analysis presented in Section 3.2.1. Panel C, which focuses on the pre-treatment period (when both purchase and sale occurred before any associated patent activity), shows return premiums that are not statistically significant across all property types, lending further support to the validity of the instrumental variable strategy. Additionally, Panel D shows that return premiums are statistically insignificant at the 10 percent level during the post-treatment period, when properties are purchased after the patent application's first-action date, except for a minor increase in the annualized return for single-family homes of 0.94 percentage points in Column (2).

Overall, the property-level analysis reveals strong and economically meaningful cumulative effects of innovative startups on local real estate returns. We observe these effects across all

property types, with particularly pronounced impacts for multi-family and vacant land. While the total return premium increases with the number of nearby innovative firms, the marginal effect per additional startup tends to decline, indicating diminishing marginal externalities. These findings complement the firm-level results and provide further evidence that local innovation activity generates substantial, nonlinear spillovers in the real estate market.

3.4 Mechanism: Channels Linking Innovation to Real Estate Returns

To understand the economic forces behind the observed real estate return premiums near innovative startups, we investigate the underlying mechanisms through which patent-granted startups generate localized spillovers. We focus on three complementary channels motivated by knowledge spillovers, theories of agglomeration, and demographic shifts in local housing demand. First, we study whether successful innovators catalyze additional patenting activity by nearby startups. Second, we examine how innovation affects the number and composition of nearby business establishments. Third, we analyze whether innovation changes the composition of homebuyers in local housing markets. Together, these channels help clarify the mechanisms through which patent-granted startups translate into greater real estate returns in nearby markets.

3.4.1 Localized Innovation Diffusion

We first examine whether a startup's patent grant stimulates nearby innovative activity. For each startup i, we measure changes in patent applications and approvals by other startups located within 10 miles over the three years following startup i's first-action date, relative to the preceding year.

To do so, we estimate the following regression equation:

$$y_{i,k,c} = \beta_0 + \beta_1 \mathbf{1}(\text{Approved})_i + \lambda_{gau_i \times appl_year_i} + \lambda_{c \times appl_year_i} + \epsilon_{ikc}, \tag{6}$$

where $y_{i,k,c}$ denotes the change in the number of patent applications (or approvals) by other startups within a 10-mile distance of startup i and within a 3-year window after the first-action date of startup i. The change is relative to the year prior to a startup's first action date. The variable $\mathbf{1}(\text{Approved})_i$ is binary and indicates whether the startup received a first-action grant. This binary variable is instrumented by examiner leniency as done previously. The specification includes art unit-by-application year fixed effects ($\lambda_{gau_i \times appl_year_i}$) and county-by-application year fixed effects ($\lambda_{c \times appl_year_i}$). We cluster standard errors at the art unit level. Since the sample comprises startups that have a first action date from 2001 to 2009, we focus on the time frame from the years 2002 to 2006 to allow a year of prior activity and to enable a three-year window within which spillovers may occur.

Panel A of Table 5 presents results for the change in the number of patent applications within a 10-mile distance of the focal granted patent and within a 3-year window after the first-action date of the granted patent. All specifications include art unit-by-application year and county-by-application year fixed effects. The explanatory variable of interest, $\mathbf{1}$(Approved), is binary and indicates whether the focal patent application was approved and is instrumented by examiner leniency. Columns (1) and (2) report results based on the change in the number of patent applications and the change in the number of approved patent applications, irrespective of the sector of the focal patent. We find a positive and statistically significant coefficient on $\mathbf{1}$(Approved). This finding implies that a patent approval has a positive effect on patent applications and approvals by startups nearby. Columns (3) and (4) report results based on the change in the number of patent applications and the change in the number of approved patent applications by startups that have the same sector as the focal patent. Here too, the coefficient of $\mathbf{1}$(Approved) is statistically significant, suggesting a positive spillover.

3.4.2 Entry of Local Businesses

We next evaluate the effects of innovative startups on local business activity. As mentioned in Section 2.4, we construct annual counts of establishments located within a 10-mile radius of each startup, disaggregated by 2-digit NAICS industry. For each startup i, industry n, and year t relative to the startup's first-action year (with $t = 0$ denoting the year of the first action), we observe the number of nearby establishments, $N_{i,n,t}$, and construct changes relative to the pre-innovation baseline $N_{i,n,-1}$ and use them as outcomes: $D_{i,n,t} = N_{i,n,t} - N_{i,n,-1}$ which we interpret as the difference in levels and percentage change in the number of nearby establishments relative to the year before the first-action decision.[10]

We estimate the following system of 2SLS regressions at the startup-by-industry-by-relative-year level, for $t = -5, \ldots, 10$:

$$\begin{aligned}\mathbf{1}(\text{Approved})_i = {} & \gamma_0 + \gamma_1(\text{Leniency})_i + \gamma_2\mathbf{1}(\text{Within})_{i,n} + \gamma_3(\text{Leniency})_i \times \mathbf{1}(\text{Within})_{i,n} \\ & + \tilde{\lambda}_{gau_i \times appl_year_i} + \tilde{\lambda}_{c_i \times appl_year_i} + \tilde{\lambda}_{n \times t_{i,t}} + \tilde{\epsilon}_i, \end{aligned} \tag{7}$$

$$\begin{aligned}D_{i,n,t} = {} & \beta_{0,t} + \beta_{1,t}\mathbf{1}(\widehat{\text{Approved}})_i + \beta_{2,t}\mathbf{1}(\text{Within})_{i,n} + \beta_{3,t}\mathbf{1}(\widehat{\text{Approved}})_i \times \mathbf{1}(\text{Within})_{i,n} \\ & + \lambda_{gau_i \times appl_year_i} + \lambda_{c_i \times appl_year_i} + \lambda_{n \times \tau_{i,t}} + \epsilon_{i,n,t}, \end{aligned} \tag{8}$$

where $\mathbf{1}(\text{Approved})_i$ denotes whether startup i was granted a patent, $\mathbf{1}(\text{Within})_{i,n}$ indicates whether industry n matches the startup's own industry, and $\tau_{i,t}$ denotes the calendar year corresponding to relative year t (i.e., the application year of startup i plus t). The fixed effects $\lambda_{gau_i \times appl_year_i}$ absorb art-unit-by-application-year shocks, $\lambda_{c_i \times appl_year_i}$ control for county-by-

10. In practice, to avoid sample selection in the proportional changes,$P_{i,n,t}$, by removing from the regressions the few observations that have $N_{i,n,-1} = 0$, for those, observations we use the average of the number of establishments in the previous 5 years. Also, we winsorize extreme values of $N_{i,n,t}$, $D_{i,n,t}$, and $P_{i,n,t}$ at the 0.1st and 99.9th percentiles.

application-year shocks, and $\lambda_{n\times\tau_{i,t}}$ capture national shocks specific to each two-digit NAICS industry and calendar year. We cluster standard errors at the art-unit level.[11]

We first estimate a restricted version of the system in Equations 7 and 8 that omits the interaction terms $(\text{Leniency})_i \times \mathbf{1}(\text{Within})_{i,n}$ and $\mathbf{1}(\widehat{\text{Approved}})_i \times \mathbf{1}(\text{Within})_{i,n}$. In this specification, β_1 captures the effect of innovative startups on local business activity, averaged across all sectors. Panel (a) of Figure 4 reports the corresponding IV estimates. We find an almost immediate increase in the number of nearby establishments following the patent decision, and this increase is sustained through the end of our analysis window, ten years after the patent grant.

We next estimate the full specification, including the interaction terms, and report the results in Panel (b) of Figure 4 to decompose the effects on establishments in the same industry versus other industries. The results indicate that the overall effect is largely driven by growth in establishments within the same sector as the focal startup.

In both specifications, the placebo coefficients for the pre-innovation periods ($t = -5, \ldots, 0$) are not statistically significant at the 10 percent level, providing additional empirical support for the validity of our identification strategy.

3.4.3 Change in Homebuyer Characteristics

Finally, we examine whether and to what extent innovative startups lead to changes in homebuyer characteristics. To do so, we re-estimate the property-level IV regression in Equation 5 using alternative outcome variables that capture buyer transitions in repeat-sale transactions, focusing on single-family homes. Specifically, we analyze shifts in ownership between individuals and non-individual entities, as well as between in-state and out-of-state buyers.

11. Our results for changes in the number of establishments are equivalent to those obtained from specifications that additionally control for the pre-innovation level $N_{i,n,-1}$.

Panel B of Table 5 presents the results. In Columns (1) and (2), the outcome variable is defined as $\mathbf{1}$(Buyer = Non-individual) − $\mathbf{1}$(Seller = Non-individual), which equals 1 if a non-individual buyer purchases the property from an individual seller, -1 if an individual purchases from a non-individual, and 0 if the transaction occurs between two individuals or two non-individuals. Regardless of whether property characteristics are excluded (Column 1) or included (Column 2), the results are consistent: properties near patent-granted startups are more likely to transition from individual to non-individual ownership. Specifically, we find an increase of approximately 1.33 percentage points in the likelihood of such transitions.

Columns (3) and (4) use an analogous outcome variable to capture in-state versus out-of-state buyer status: $\mathbf{1}$(Buyer = Out-of-state) − $\mathbf{1}$(Seller = Out-of-state), which equals 1 if an out-of-state buyer purchases the property from an in-state seller, -1 if the reverse occurs, and 0 otherwise. Although the estimated coefficients are positive in both specifications, they are not statistically significant.

Several caveats apply to this analysis. First, non-individual buyers may include institutional investors as well as individuals purchasing through legal entities such as LLCs. Nonetheless, the results suggest that the presence of patent-granted startups leads to an increase in institutional and other sophisticated buyers in the local housing market. Second, we determine out-of-state buyer status by comparing the buyer's mailing address with the property's address. However, many buyers list the property address as their mailing address, even if relocating from another state, resulting in measurement error and a likely undercount of true out-of-state purchasers. Overall, these findings suggest that patent-granted startups shift buyer composition in the local housing market, particularly toward non-individual ownership.

4 Conclusion

This paper provides causal evidence on how innovative startups shape local real estate markets and business activity. Using nationwide micro-level data on patent applications, repeat-sale property transactions, and business establishment counts, and exploiting the quasi-random assignment of patent examiners as an instrument for patent approval, we show that patent-granted startups generate economically substantial and highly localized spillovers. In our baseline firm-level analysis, a startup patent approval raises annualized real estate returns within 10 miles by 0.38 percentage points for single-family homes, 0.64 percentage points for multi-family properties, 0.50 percentage points for commercial real estate, and 0.91 percentage points for vacant land. These return premiums attenuate with distance from the innovating startup, are largest for the first successful innovator in an area, and are not driven by pre-existing trends in local property values. A property-level analysis that accounts for exposure to multiple nearby innovators confirms the pattern: the total return premium rises with the number of nearby patent-granted startups, but the marginal effect per additional startup declines, indicating diminishing externalities. Examining mechanisms, we find that patent approvals stimulate nearby patenting activity, spur business formation particularly within the same industry, and shift the composition of homebuyers toward non-individual ownership. Together, these findings underscore the broader role of innovative startups in reshaping local real estate markets and neighborhood composition, and they highlight innovation policy as a driver of place-based economic development.

A large prior literature documents how large manufacturing plants, government-led industrial expansions, and major corporate events affect local economies (Greenstone et al., 2010; Garin and Rothbaum, 2025; Monte et al., 2018). Relative to this work, we shift the focus to smaller, early-stage innovative firms and show that even a single patent-granted startup generates meaningful spillovers at the neighborhood level. Our findings complement evidence

that startup formation and local knowledge spillovers drive regional economic dynamism (Glaeser et al., 1992; Haltiwanger et al., 2013; Adelino et al., 2017), by documenting that the causal effect of successful innovation — rather than the mere presence of startups — produces measurable local externalities in real estate, business activity, and buyer composition. Prior work using patent examiner leniency has established that patent grants raise the growth and investment of the focal firm (Farre-Mensa et al., 2020; Kline et al., 2019); we extend this line of inquiry by showing that the economic effects of startup patent approval reach well beyond the firm itself. Finally, we add to the literature on the capitalization of local economic shocks into housing markets (Coulson et al., 2013; Butler et al., 2019; Beracha et al., 2022) by providing causal evidence that startup-driven innovation generates return premiums across all major property types and that these effects are accompanied by shifts in buyer composition toward institutional and other sophisticated buyers.

Several limitations qualify these findings. First, our sample is restricted to startups that file patent applications, which represent a particular type of innovative activity. Innovative firms that do not seek patents — because of the nature of their innovations or strategic disclosure concerns — lie outside our analysis, and the estimated spillovers may not extend to the full population of startups. Second, the instrumental variable design leverages the quasi-random assignment of examiners within technology art units, which provides the sharpest variation around close-call applications. The estimates may therefore reflect the effects of marginal inventions rather than unambiguously high-quality patents, and they may not fully capture the spillovers generated by clearly transformative innovations. Third, our establishment count data cover approximately one decade following each patent decision, which limits our ability to examine business formation dynamics beyond the ten-year horizon. Fourth, we do not directly observe neighborhood-level labor market outcomes such as wages or employment; our mechanism evidence therefore relies on establishment counts and buyer composition as proxies for the broader economic changes that underlie the observed real estate return premiums.

These limitations point to several directions for future research. First, linking our setting to administrative labor market records — such as Quarterly Workforce Indicators or Longitudinal Employer-Household Dynamics data — would allow direct measurement of wage and employment consequences of startup patent approvals at the neighborhood level, connecting the real estate effects we document to household income and welfare. Second, extending the analysis to non-patent forms of innovative success, such as venture capital investment events or receipt of Small Business Innovation Research grants, would help assess whether the spillovers we document are specific to patented innovation or reflect a broader phenomenon of startup success. Third, examining distributional consequences — specifically, whether renters and lower-income households bear net costs from rising property values near innovative startups — would connect our findings to the broader literature on the local costs and benefits of innovation clusters (Qian and Tan, 2021; Beracha et al., 2025). Finally, our finding that the first innovative startup in an area generates the largest marginal return premium, with diminishing effects for subsequent entrants, has direct implications for place-based innovation policies such as innovation districts, opportunity zones, and startup subsidy programs; evaluating these policies through the lens of spatial externalities and the first-mover dynamics we document represents a promising avenue for future work.

5 Tables and Figures

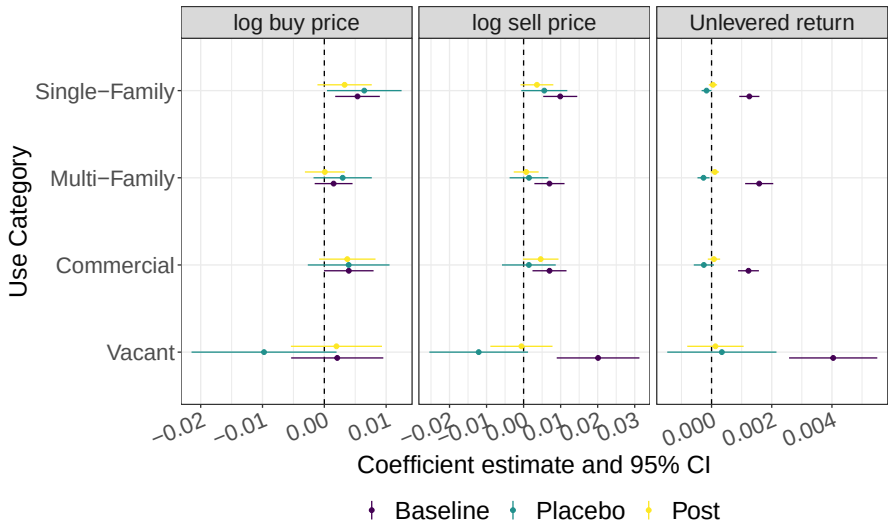
Table 1 — Summary Statistics of Startups

	By First Patent Outcome	
	Approved	Rejected
Panel A. # Observations		
No. of firms	18,923	10,564
% of firms	64.2	35.8
Panel B. Firm Characteristics		
Age at first patent filing (years), Median	2	2
Normalized HPI at first patent filing, Median	100	100
Employees at Filing Date		
Mean	29.5	29.4
Median	8	7
SD	61.7	62.1
Sales at Filing Date ($ million)		
Mean	4.3	4.3
Median	0.8	0.7
SD	9.7	10.0
Total Establishments in ZIP Code at Filing Date		
Mean	976.3	1,070.4
Median	845	946
SD	683.5	714.6
Pre-Patent-Filing Employment Growth (%)		
Mean	15.2	14.7
SD	64.0	62.8
Pre-Patent-Filing Sales Growth (%)		
Mean	19.3	17.4
SD	84.2	80.4
Pre-Patent-Filing Establishment Growth (%)		
Mean	1.9	1.9
SD	4.9	4.8
Pre-Patent-Filing HPI Growth (%)		
Mean	7.3	7.7
SD	7.0	7.6

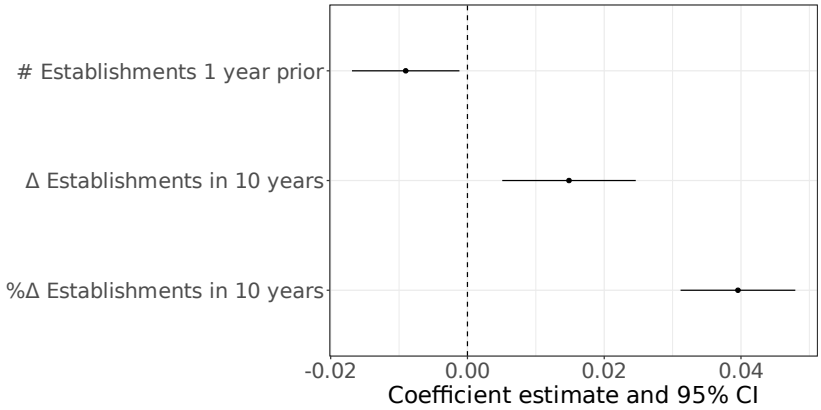
Note. This table reports summary statistics on startups whose first patent applications were ultimately approved (granted) and those whose applications were ultimately rejected (not granted).

Figure 1 — Descriptive Regression Results

(a) Repeat Sales Sample



(b) Establishment Sample



Note. The figure presents descriptive regression results. Panel (a) shows coefficient estimates of **1**(Approved) and their 95% confidence intervals from Equation 1, across three outcome variables in each facet: log buy price, log sell price, and unlevered return. All regressions in Panel (a) include art-unit-by-application-year fixed effects, property type fixed effects, county-by-buy-year fixed effects, and distance-bin fixed effects. Each facet reports results for three samples (baseline, placebo, and post) and four property use categories (single-family, multi-family, commercial, and vacant). Panel (b) displays coefficient estimates of **1**(Approved) and their 95% confidence intervals from Equation 2, using the Data Axle sample, with three outcome variables shown on the y-axis: the number of establishments one year prior to the associated startup's first-action date (serving as the pre-innovation baseline), the change in the number of establishments ten years after the first-action date relative to the baseline, and the percent change. All regressions in Panel (b) include art-unit-by-application-year fixed effects and county-by-application-year fixed effects. The estimates in Panel (b) are normalized by the standard deviation of each outcome variable. Standard errors in Panel (a) are clustered at the art-unit level, and those in Panel (b) are clustered at the county level.

Table 2 — Firm-Level Analysis: Effects on Real Estate Returns by Property Use

	Outcome: median annualized unlevered return							
	Single-Family		Multi-Family		Commercial		Vacant	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. baseline sample — properties bought before and sold after treatment								
1(Approved)	0.0038***	0.0020***	0.0064***	0.0039***	0.0050***	0.0023***	0.0091***	0.0058***
	(0.0007)	(0.0005)	(0.0011)	(0.0007)	(0.0008)	(0.0005)	(0.0027)	(0.0021)
Observations	38,101,658	38,101,658	35,986,905	35,986,905	5,237,993	5,237,993	1,595,151	1,595,151
R^2								
Panel B. placebo sample — properties bought and sold before treatment								
1(Approved)	5.78×10^{-5}	0.0001	0.0003	0.0001	0.0002	0.0002	-0.0034	-0.0008
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0007)	(0.0007)	(0.0036)	(0.0035)
Observations	10,331,945	10,331,945	12,662,816	12,662,816	1,781,337	1,781,337	377,622	377,622
R^2	0.39545	0.60911	0.21621	0.39758	0.09709	0.16814	0.09861	0.21499
Panel C. post sample — properties bought and sold after treatment								
1(Approved)	6.48×10^{-5}	6.49×10^{-5}	0.0002	0.0002	0.0007	0.0007*	0.0002	0.0005
	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0004)	(0.0004)	(0.0021)	(0.0019)
Observations	31,409,014	31,409,014	28,623,627	28,623,627	4,172,772	4,172,772	2,201,112	2,201,112
R^2	0.48032	0.53665	0.40528	0.46348	0.15266	0.20987	0.21155	0.30356
Holding term FE		✓		✓		✓		✓

Notes: Standard errors clustered at the art unit level in parentheses.
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Note. This table reports coefficient estimates of β_1 from IV regressions in Equation 4, where the outcome variable is the median annualized unlevered return. Panel A uses the baseline sample, Panel B uses the placebo sample, and Panel C uses the post sample. In each panel, four subsamples are analyzed by property use categories: single-family residential, multi-family residential, commercial, and vacant real estate. Columns (1), (3), (5), and (7) report the baseline estimates including art-unit-by-application-year fixed effects, property type fixed effects, county-by-buy-year fixed effects, and distance-bin fixed effects.Columns (2), (4), (6), and (8) additionally include holding-term (in years) fixed effects.

Table 3 — Firm-Level Analysis: Effects on Prices and Holding Terms

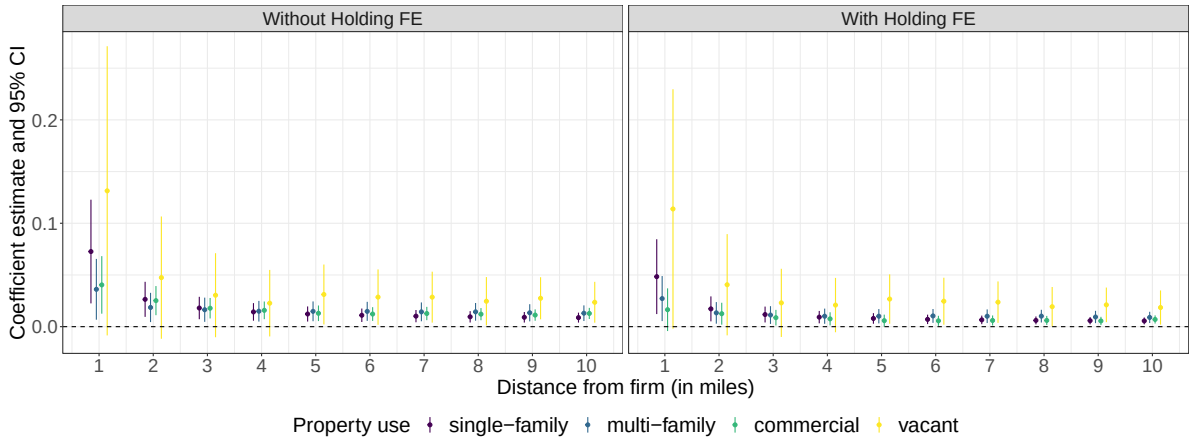
	Single-Family		Multi-Family		Commercial		Vacant	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: outcome — log median buy price								
1(Approved)	-0.0060	0.0011	-0.0135**	-0.0067	-0.0152*	-0.0083	-0.0228	-0.0194
	(0.0067)	(0.0070)	(0.0057)	(0.0056)	(0.0084)	(0.0094)	(0.0158)	(0.0148)
Panel B: outcome — log median sell price								
1(Approved)	0.0043	0.0097	0.0046	0.0097	-0.0073	0.0009	0.0022	-0.0026
	(0.0084)	(0.0086)	(0.0076)	(0.0075)	(0.0090)	(0.0102)	(0.0230)	(0.0214)
Panel C: outcome — holding term (in years)								
1(Approved)	-0.4591***		-0.4995***		-0.5186***		-0.5827***	
	(0.0610)		(0.0737)		(0.0634)		(0.0784)	
County-by-Holding FE		✓		✓		✓		✓
Observations	38,101,658	38,101,658	35,986,905	35,986,905	5,237,993	5,237,993	1,595,151	1,595,151

Notes: Standard errors clustered at the art unit level in parentheses.
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

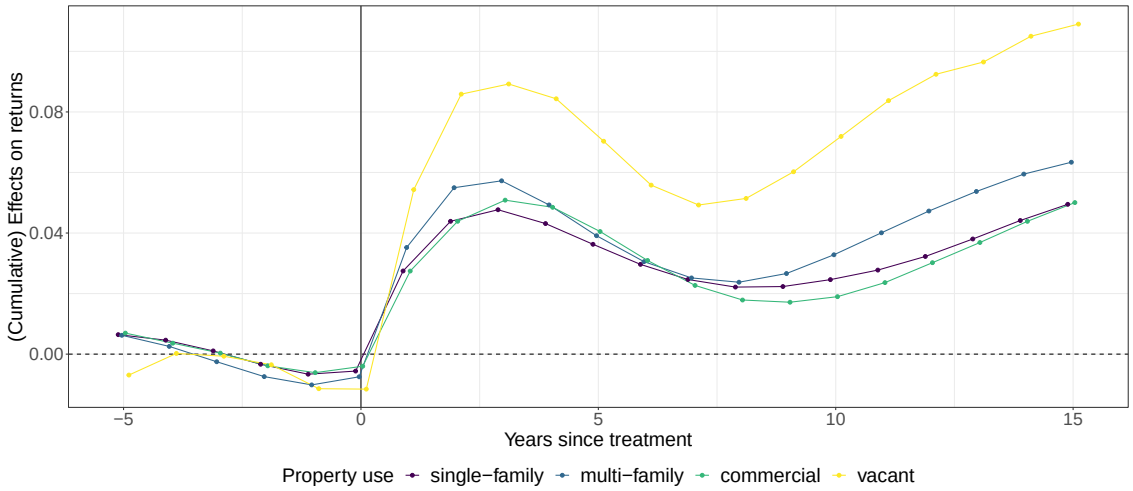
Note. This table reports coefficient estimates of β_1 from IV regressions in Equation 4, using the baseline repeat-sales sample for three different outcome variables: log median buy price in Panel A, log median sell price in Panel B, and holding term in Panel C. In each panel, four subsamples are analyzed by property use categories: single-family residential, multi-family residential, commercial, and vacant real estate. Columns (1), (3), (5), and (7) report the baseline estimates including art-unit-by-application-year fixed effects, property type fixed effects, county-by-buy-year fixed effects, and distance-bin fixed effects. In Panels A, B, and C, Columns (2), (4), (6), and (8) additionally include county-by-holding term (in years) fixed effects.

Figure 2 — Return Premiums by Distance and Years Since Innovation

(a) By Distance



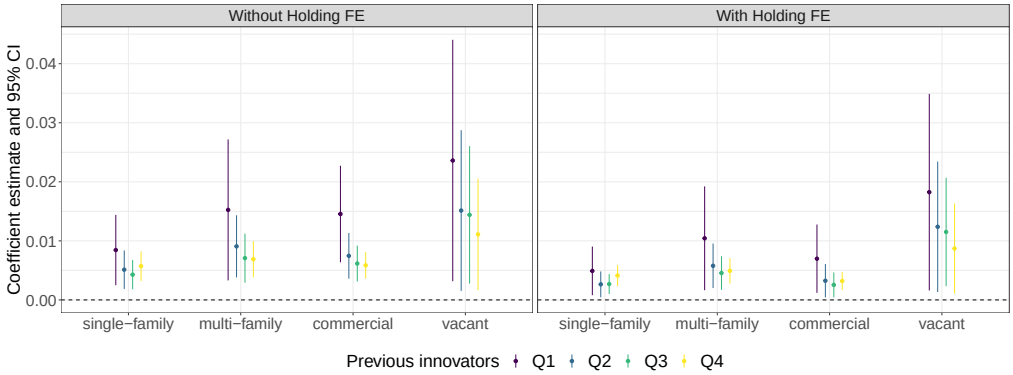
(b) By Years Since



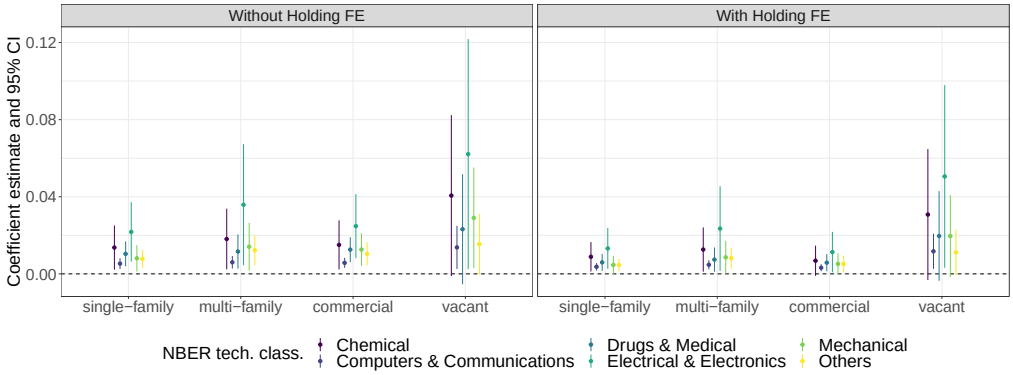
Note. This figure displays the estimated return premiums by distance and years since treatment (first-action grant). Estimates are shown separately by property use type and color-coded: single-family residential (purple), multi-family (blue), commercial (green), and vacant land (yellow). Panel (a) plots IV estimates of β_1 from Equation 4 by distance to the startup, using ten one-mile bins: 0–1 mile, 1–2 miles, …, 9–10 miles. The outcome is the median annualized unlevered return. The left subpanel reports the baseline specification, and the right subpanel adds holding-term fixed effects. Standard errors are clustered at the art-unit level. Panel (b) reports event-time patterns in returns by the sale year relative to the first-action decision date, where τ denotes years since treatment. For $\tau > 0$, we estimate $\beta_{1,\tau}$ using all repeat-sales transactions sold τ years after treatment (combining baseline and post-treatment subsamples) with county–by–holding term fixed effects and plot the cumulative sum $\sum_{t=1}^{\tau} \beta_{1,t}$. Similarly, for $\tau \leq 0$, we estimate $\beta_{1,\tau}$ using repeat-sales sold $|\tau|$ years before treatment (placebo subsample) and plot $\sum_{t=0}^{\tau} \beta_{1,t}$.

Figure 3 — Heterogeneity and Robustness Checks

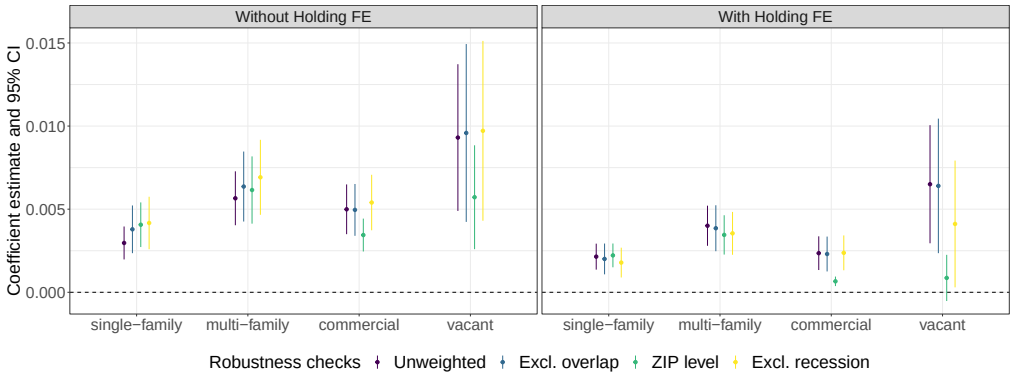
(a) By Historical Patent Activity in the City (20-Year Window)



(b) By Technology Classification



(c) Robustness Checks



Note. The figure reports coefficient estimates of β_1 from IV regressions in Equation 4 across subsamples and alternative specifications. Panel (a) plots estimates for four treatment subgroups defined by prior exposure to innovators, measured as the number of patents granted in the same city over the past 20 years. Subgroups are defined by quartiles of this historical patenting intensity: Q1 (purple), Q2 (blue), Q3 (green), and Q4 (yellow). Panel (b) plots estimates for treatment subgroups defined by NBER technology classification. Panel (c) presents robustness checks using alternative specifications described in Section 3.2.2. In each panel, estimates are shown separately by property use type. The left subpanel reports the baseline IV specification, and the right subpanel adds holding-term fixed effects. Standard errors are clustered at the art-unit level throughout.

Table 4 — Property-Level Analysis: Multiple Firm Effects

	Outcome: annualized unlevered return							
	Single-Family		Multi-Family		Commercial		Vacant	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: baseline repeat sales — properties bought before and sold after treatment								
$\mathbf{1}(N_1 > 0)$	0.0560***	0.0522***	0.1706***	0.1651***	0.0757***	0.0648***	0.1010***	0.0999***
	(0.0064)	(0.0061)	(0.0258)	(0.0254)	(0.0110)	(0.0100)	(0.0231)	(0.0231)
Mean of N_1 cond. $N_1 > 0$	17.87		29.89		23.42		10.48	
Observations	11,696,097	11,696,097	4,142,706	4,142,706	222,234	222,234	163,678	163,678
R^2	0.32941	0.34358	0.34575	0.35630	0.25528	0.26579	0.28645	0.28676
Panel B: baseline repeat sales & heterogeneity by the number of successful startups nearby (N_1)								
$\mathbf{1}(N_1 = 1)$	0.0149***	0.0133***	0.0219***	0.0213***	0.0201***	0.0183***	0.0382**	0.0379**
	(0.0026)	(0.0024)	(0.0059)	(0.0056)	(0.0073)	(0.0070)	(0.0182)	(0.0182)
$\mathbf{1}(N_1 \in Q_{N_1,1})$	0.0231***	0.0211***	0.0482***	0.0460***	0.0314***	0.0291***	0.0589***	0.0582***
	(0.0026)	(0.0026)	(0.0069)	(0.0065)	(0.0062)	(0.0060)	(0.0187)	(0.0186)
$\mathbf{1}(N_1 \in Q_{N_1,2})$	0.0431***	0.0391***	0.0978***	0.0935***	0.0634***	0.0563***	0.1003***	0.1002***
	(0.0046)	(0.0042)	(0.0134)	(0.0132)	(0.0100)	(0.0096)	(0.0285)	(0.0283)
$\mathbf{1}(N_1 \in Q_{N_1,3})$	0.0668***	0.0626***	0.1633***	0.1587***	0.0946***	0.0828***	0.2142***	0.2119***
	(0.0080)	(0.0074)	(0.0224)	(0.0226)	(0.0137)	(0.0132)	(0.0456)	(0.0456)
$\mathbf{1}(N_1 \in Q_{N_1,4})$	0.1113***	0.1082***	0.3090***	0.3090***	0.1435***	0.1264***	0.2900***	0.2843***
	(0.0183)	(0.0182)	(0.0747)	(0.0741)	(0.0279)	(0.0259)	(0.0768)	(0.0742)
$Q_{N_1,1}$ [Mean of N_1 cond. $N_1 > 0$]	2–4 [2.81]		2–6 [3.68]		2–5 [3.22]		2–3 [2.39]	
$Q_{N_1,2}$ [Mean of N_1 cond. $N_1 > 0$]	5–9 [6.72]		7–14 [10.06]		6–7 [8.19]		4–6 [4.84]	
$Q_{N_1,3}$ [Mean of N_1 cond. $N_1 > 0$]	10–23 [15.15]		15–37 [23.68]		12–28 [18.46]		7–15 [10.09]	
$Q_{N_1,4}$ [Mean of N_1 cond. $N_1 > 0$]	24–842 [62.71]		38–844 [97.21]		29–834 [81.04]		16–791 [41.81]	
Panel C: placebo repeat sales — properties bought and sold before treatment								
$\mathbf{1}(N_1 > 0)$	-0.0021	0.0047	0.0264	0.0301	-0.0046	-0.0018	-0.0038	-0.0052
	(0.0065)	(0.0062)	(0.0199)	(0.0209)	(0.0114)	(0.0115)	(0.0183)	(0.0182)
Observations	9,152,691	9,152,691	3,092,237	3,092,237	139,884	139,884	111,828	111,828
R^2	0.29679	0.30678	0.33128	0.33429	0.22235	0.22428	0.27840	0.27844
Panel D: post repeat sales — properties bought and sold after treatment								
$\mathbf{1}(N_1 > 0)$	0.0042	0.0094*	0.0483	0.0590	-0.0333	-0.0298	-0.0026	-0.0005
	(0.0066)	(0.0056)	(0.0475)	(0.0500)	(0.0319)	(0.0304)	(0.0387)	(0.0389)
Observations	5,464,306	5,464,306	1,517,271	1,517,271	85,421	85,421	159,973	159,973
R^2	0.15747	0.18129	0.17401	0.17912	0.28308	0.28528	0.25245	0.25257
Property Char's Controlled		✓		✓		✓		✓

Notes: Standard errors clustered at the county level in parentheses.
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Note. This table reports coefficient estimates of β_1 from IV regressions in Equation 5, where the outcome variable is the annualized unlevered return and the treatment variable indicates whether any startups within a 10-mile radius received a patent grant. Panel A uses the baseline repeat-sale transactions, Panel B examines treatment heterogeneity by the number of patent grants using the baseline sample, Panel C uses the pre-treatment (placebo) sample, and Panel D uses the post-treatment sample. In each panel, four property types are examined: single-family, multi-family, commercial, and vacant. Columns (1), (3), (5), and (7) show the baseline estimates, which include the total number of nearby startups, property type fixed effects, county-by-buy-year fixed effects, and county-by-holding-term fixed effects, while Columns (2), (4), (6), and (8) additionally include property-level control variables. These controls include land area for all property types; construction year and building square footage for commercial properties; and the number of bedrooms and bathrooms for single-family and multi-family properties.

Table 5 — Mechanisms: Knowledge Spillovers and Demographic Shifts

Panel A: Knowledge spillovers — subsequent patent applications

	All Sectors		Same Sector	
	(1)	(2)	(3)	(4)
1(Approved)	12.1884***	6.0948***	10.1443***	5.6835***
	(3.5452)	(2.0584)	(2.5781)	(1.6560)
Adjusted-R^2	-0.15	-0.14	-0.16	-0.15
N	15,031	15,031	15,031	15,031
Approvals Only		✓		✓

Notes: Standard errors clustered at the art unit level in parentheses.
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Panel B: Homebuyer composition shifts — non-individual and out-of-state buyers

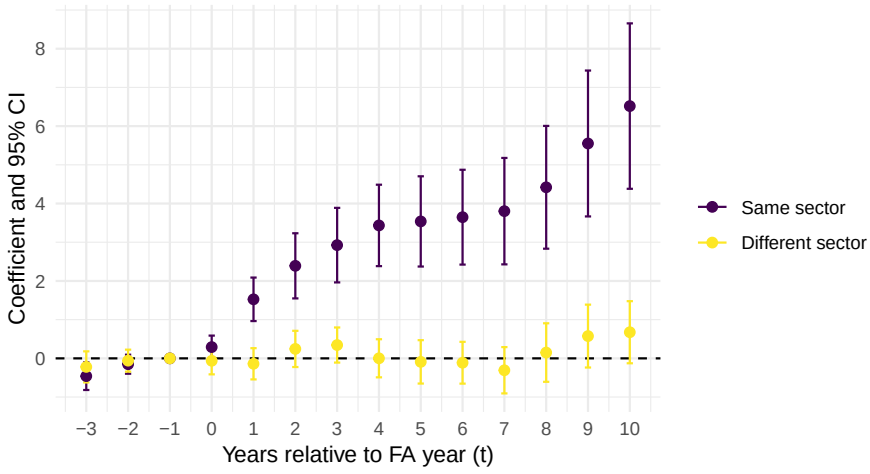
	Non-Individual		Out-of-State	
	(1)	(2)	(3)	(4)
1(Approved)	0.0133**	0.0126**	0.0061	0.0016
	(0.0059)	(0.0059)	(0.0077)	(0.0068)
Observations	11,095,559	11,095,559	9,035,617	9,035,617
Property Char's Controlled		✓		✓

Notes: Standard errors clustered at the county level in parentheses.
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

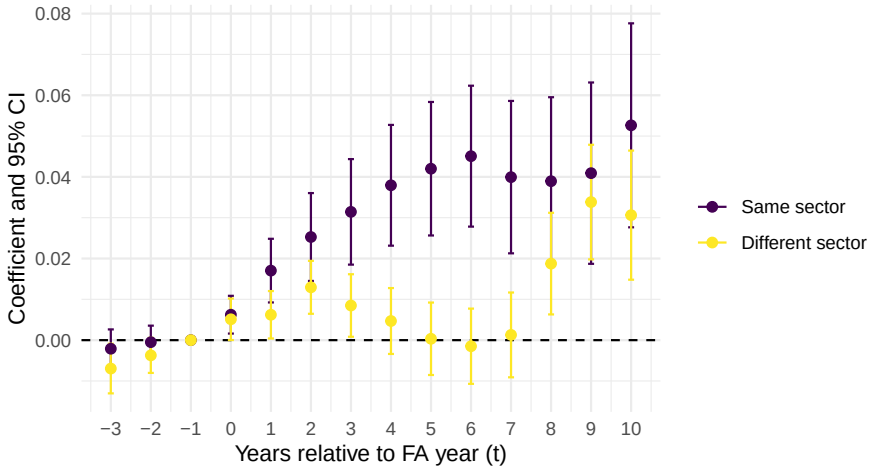
Note. Panel A presents the regression results that involve the change in the number of patent applications (or approvals) within a 10-mile distance of the focal granted patent and within a 3-year window after the first-action date of the granted patent. Panel B examines the transitions in homebuyer characteristics. In Columns (1) and (2) of Panel B, the outcome variable is defined as **1**(Buyer = Non-individual) − **1**(Seller = Non-individual), which equals 1 if a non-individual buyer purchases the property from an individual seller, -1 if an individual purchases from a non-individual, and 0 if the transaction occurs between two individuals or two non-individuals. In Columns (3) and (4) the outcome variable is defined as **1**(Buyer = Out-of-state) − **1**(Seller = Out-of-state), which equals 1 if an out-of-state buyer purchases the property from an in-state seller, -1 if the reverse occurs, and 0 otherwise.

Figure 4 — Mechanisms: Effects on Nearby Business

(a) Outcome: Change in the Number of Establishments ($D_{i,n,t}$)



(b) Outcome: Percentage Change in the Number of Establishments



Note.

References

Adelino, M., Ma, S., Robinson, D. 2017. Firm age, investment opportunities, and job creation. The Journal of Finance, 72, 999–1038.

Aiello, D., Kotter, J. D., Schubert, G. 2022. When money moves in: The consequences of housing wealth. Available at SSRN 4280776.

Beracha, E., He, Z., Wintoki, M. B., Xi, Y. 2022. On the relation between innovation and housing prices–a metro level analysis of the us market. The Journal of Real Estate Finance and Economics, 1–27.

Beracha, E., Kim, E., Wintoki, M. B., Xi, Y. 2025. Highly skilled immigrants and local housing prices: Evidence from the h-1b visa lottery. Journal of Real Estate Research.

Bernstein, S., Colonnelli, E., Giroud, X., Iverson, B. 2019. Bankruptcy spillovers. Journal of Financial Economics, 133, 608–633.

Butler, A. W., Fauver, L., Spyridopoulos, I. 2019. Local economic spillover effects of stock market listings. Journal of Financial and Quantitative Analysis, 54, 1025–1050.

Carley, M., Hegde, D., Marco, A. 2015. What is the probability of receiving a us patent? Yale Journal of Law and Technology, 17, p. 203.

Carlino, G., Drautzburg, T. 2020. The role of startups for local labor markets. Journal of Applied Econometrics, 35, 751–775.

Carlino, G., Kerr, W. R. 2015. Agglomeration and innovation. Handbook of regional and urban economics, 5, 349–404.

Coulson, N. E., Liu, C. H., Villupuram, S. V. 2013. Urban economic base as a catalyst for movements in real estate prices. Regional Science and Urban Economics, 43, 1023–1040.

Farre-Mensa, J., Hegde, D., Ljungqvist, A. 2020. What is a patent worth? evidence from the us patent "lottery". The Journal of Finance, 75, 639–682.

Garin, A., Rothbaum, J. 2025. The long-run impacts of public industrial investment on local development and economic mobility: Evidence from world war ii. The Quarterly Journal of Economics, 140, 459–520.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., Shleifer, A. 1992. Growth in cities. Journal of political economy, 100, 1126–1152.

Glaeser, E. L., Kerr, S. P., Kerr, W. R. 2015. Entrepreneurship and urban growth: An empirical assessment with historical mines. Review of Economics and Statistics, 97, 498–520.

Goldschlag, N., Lybbert, T. J., Zolas, N. J. 2020. Tracking the technological composition of industries with algorithmic patent concordances. Economics of Innovation and New Technology, 29, 582–602.

Gourio, F., Messer, T., Siemer, M. 2016. Firm entry and macroeconomic dynamics: A state-level analysis. American Economic Review, 106, 214–218.

Greenstone, M., Hornbeck, R., Moretti, E. 2010. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. Journal of political economy, 118, 536–598.

Hall, B. H., Jaffe, A. B., Trajtenberg, M. 2001. The NBER patent citation data file: Lessons, insights and methodological tools.

Haltiwanger, J., Jarmin, R. S., Miranda, J. 2013. Who creates jobs? small versus large versus young. Review of Economics and Statistics, 95, 347–361.

Hartman-Glaser, B., Thibodeau, M., Yoshida, J. 2023. Cash to spend: Ipo wealth and house prices. Real Estate Economics, 51, 68–102.

Hegde, D., Ljungqvist, A., Raj, M. 2022. Quick or broad patents? evidence from us startups. The Review of Financial Studies, 35, 2705–2742.

Hu, M., Tsang, D., Wan, W. X. 2025. Corporate relocation and housing market spillovers. Management Science, 71, 4344–4376.

Jaffe, A. B., Trajtenberg, M., Henderson, R. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. the Quarterly journal of Economics, 108, 577–598.

Kline, P., Petkova, N., Williams, H., Zidar, O. 2019. Who profits from patents? rent-sharing at innovative firms. The quarterly journal of economics, 134, 1343–1404.

Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N. 2017. Technological innovation, resource allocation, and growth. The Quarterly Journal of Economics, 132, 665–712.

Marco, A. C., Toole, A. A., Miller, R. D., Frumkin, J. P. 2017. Uspto patent prosecution and examiner performance appraisal.Technical report, USPTO Economic Working Paper Series.

Matray, A. 2021. The local innovation spillovers of listed firms. Journal of Financial Economics, 141, 395–412.

Monte, F., Redding, S. J., Rossi-Hansberg, E. 2018. Commuting, migration, and local employment elasticities. American Economic Review, 108, 3855–3890.

Pope, D. G., Pope, J. C. 2015. When walmart comes to town: Always low housing prices? always? Journal of Urban Economics, 87, 1–13.

Qian, F., Tan, R. 2021. The effects of high-skilled firm entry on incumbent residents. Stanford Institute for Economic Policy Research (SIEPR) Working Paper, 21–039.

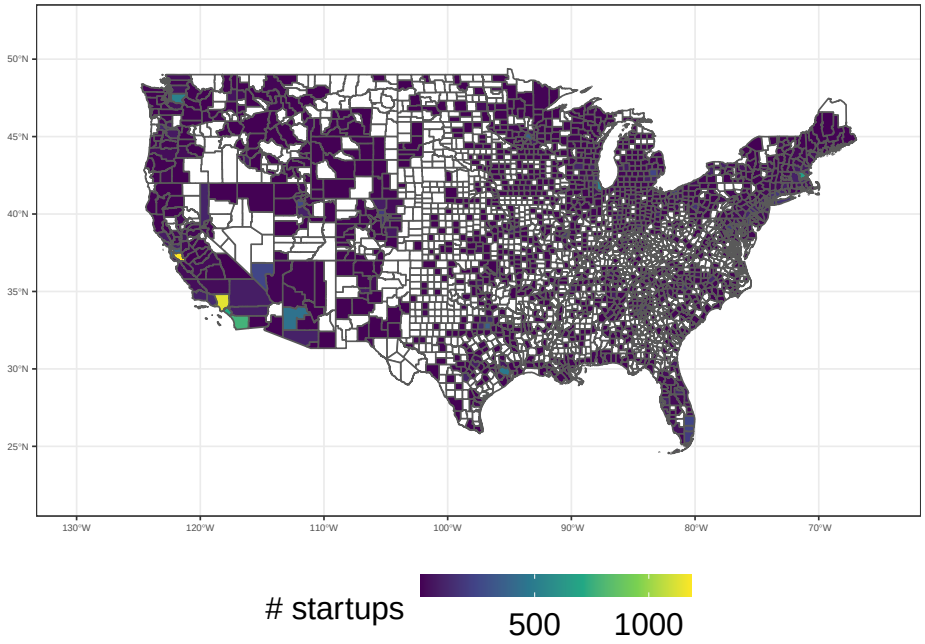
Sampat, B., Williams, H. L. 2019. How do patents affect follow-on innovation? evidence from the human genome. American Economic Review, 109, 203–236.

Walsh, C. 2023. The startup multiplier. Available at SSRN.

Appendix

A Supplementary Tables and Figures

Figure A.1 — Spatial Distribution of Startups in the Analysis Sample



Note. This figure displays the spatial distribution of startups in our analysis sample, showing the number of startups located in each county across the continental United States. An additional 29 startups in the sample are located in Alaska and Hawaii.

Table A.1 — Property Types by Use Category

Use Category	Property Type (as originally coded in ATTOM)	# Repeat Sales in Sample
Single-family	Single family residence	14,046,388
	Townhouse	880,924
	Single family residential	102,984
	Row house	101,188
	Zero lot line (residential)	59,843
	Residential (general/single)	33,895
	Bungalow (residential)	9,312
Multi-family	Condominium	3,552,058
	Duplex (2 units, any combination)	398,742
	Apartments (generic)	124,486
	Triplex (3 units, any combination)	94,263
	Multi-family dwellings (generic, any combination)	82,236
	Quadplex (4 units, any combination)	61,174
	Apartment house (5+ units)	56,352
	Cluster home	52,649
	Residential income (general/multi-family)	40,623
	Garden apt, court apt (5+ units)	8,750
Commercial	Retail stores (personal services, photography, travel)	64,612
	Commercial (general)	51,621
	Office building	49,149
	Commercial office/residential (mixed use)	34,680
	Restaurant	19,706
	Condominium offices	19,410
	Commercial office (general)	12,883
Vacant	Residential - vacant land	94,999
	Vacant land	23,393
	Vacant commercial	23,262

Table A.2 — Summary Statistics of Firm-level Repeat Sales Sample

Sample	1(Approved)	Annualized Unlevered Return (%)					Buy Price ($ 1000s)					Sell Price ($ 1000s)				
		Q1	Q2	Q3	Avg	SD	Q1	Q2	Q3	Avg	SD	Q1	Q2	Q3	Avg	SD
Panel A. Single-Family																
Baseline	0	-2.11	0.18	2.81	0.43	5.30	192.79	283.42	432.41	350.42	391.36	189.31	303.85	491.22	392.96	535.82
Baseline	1	-1.73	0.63	3.49	1.07	5.42	183.74	269.94	413.12	332.25	344.84	186.83	300.07	491.41	385.72	469.11
Placebo	0	-2.51	0.80	3.91	1.00	5.47	188.67	283.57	397.62	312.06	173.13	206.71	302.84	440.97	353.30	218.75
Placebo	1	-2.60	0.68	3.84	0.95	5.57	186.29	281.05	394.59	308.53	170.29	203.35	299.38	434.44	347.07	213.13
Post	0	-1.72	1.14	4.70	1.53	6.17	171.12	275.07	430.47	343.70	446.51	183.77	292.40	471.79	390.03	601.77
Post	1	-2.31	0.29	3.79	0.76	6.22	171.99	271.55	425.48	335.95	385.38	173.22	273.08	444.10	364.81	513.65
Panel B. Multi-Family																
Baseline	0	-2.76	0.38	3.97	0.80	8.02	174.29	287.10	460.99	408.98	1120.56	170.66	317.45	544.96	484.21	1488.83
Baseline	1	-2.33	0.96	4.91	1.63	8.22	162.81	267.25	436.59	382.11	986.04	167.30	313.44	536.92	467.43	1315.37
Placebo	0	-4.32	-0.29	4.13	0.17	9.78	189.49	268.66	394.60	341.75	333.60	178.09	267.59	424.59	362.93	425.99
Placebo	1	-4.47	-0.47	3.94	0.02	9.76	190.15	267.71	392.82	340.78	328.51	175.14	263.19	417.97	356.75	408.76
Post	0	-2.36	1.42	6.41	2.18	9.72	160.52	291.34	477.57	442.04	1465.19	164.39	307.06	544.87	525.08	1885.11
Post	1	-3.17	0.38	5.25	1.16	9.65	160.29	284.28	470.96	429.28	1362.87	151.16	282.27	513.61	486.35	1734.61
Panel C. Commercial																
Baseline	0	-2.60	1.53	6.86	2.47	10.03	223.99	433.00	820.95	909.00	3364.53	260.53	537.27	1053.11	1147.98	4028.36
Baseline	1	-2.31	2.10	7.81	3.16	10.24	209.74	408.57	795.60	815.53	2700.67	256.28	529.04	1051.91	1057.62	3276.55
Placebo	0	-5.94	-0.35	6.66	0.54	12.61	212.28	388.92	767.28	680.50	1051.21	208.73	391.36	798.43	716.08	1299.12
Placebo	1	-6.17	-0.59	6.44	0.35	12.64	214.29	394.57	780.59	686.57	1042.91	207.01	391.40	802.12	709.10	1243.33
Post	0	-3.14	1.47	8.62	2.84	12.04	227.57	478.28	901.64	1126.08	4688.36	235.67	524.93	1086.11	1360.74	5590.35
Post	1	-3.70	0.66	7.29	2.04	11.77	226.90	471.17	897.90	1003.14	3793.54	222.18	493.89	1029.30	1172.32	4522.81
Panel D. Vacant																
Baseline	0	-11.19	-1.00	7.39	-1.68	23.29	53.24	162.69	401.29	375.27	1072.36	27.60	145.46	444.76	559.17	2669.18
Baseline	1	-9.53	-0.03	8.96	-0.05	23.39	48.83	144.56	360.88	330.97	910.36	29.01	143.24	429.30	493.90	2246.12
Placebo	0	-8.28	-0.33	7.77	0.08	25.33	91.00	198.30	396.23	304.87	334.30	78.13	193.86	397.99	323.69	426.29
Placebo	1	-8.01	-0.42	7.72	0.23	25.33	90.90	201.66	396.00	302.83	329.49	77.76	190.66	390.04	313.94	407.35
Post	0	-13.10	0.67	14.20	0.80	32.14	28.55	121.27	355.04	358.27	1282.14	23.60	117.20	376.10	506.91	2754.51
Post	1	-14.37	-0.61	12.07	-0.54	31.51	29.03	115.96	340.26	326.80	1137.57	20.77	101.26	337.64	427.00	2320.37

Note. This table reports the summary statistics of the firm-level repeat-sale sample described in Section 2.3. Each panel presents a subsample by property use category: single-family in Panel A, multi-family in Panel B, commercial in Panel C, and vacant in Panel D. Within each panel, three sample periods are shown: the baseline period, which includes transactions bought before and sold after the first-action date of the associated patent application; the placebo period, which includes transactions bought and sold before; and the post period, which includes transactions bought and sold after. For each sample period, the table reports summary statistics for annualized unlevered returns, buy prices (inflation-adjusted to a 2010 baseline), and sell prices (inflation-adjusted) by first-grant status (**1**(Approved)).

Table A.3 — Summary Statistics of Local Business Sample

2-digit NAICS	1(Approved)= 0					1(Approved) = 1				
	Q1	Q2	Q3	Avg	SD	Q1	Q2	Q3	Avg	SD
Panel A. # Establishments within 10 miles (Pre-innovation)										
11	11.00	18.00	25.00	19.51	12.17	9.00	16.00	23.00	17.23	11.40
21	10.00	18.00	30.00	24.77	30.11	8.00	16.00	26.00	22.07	28.06
22	5.00	10.00	16.00	13.87	14.52	4.00	9.00	14.00	11.53	12.48
23	370.00	682.00	1022.00	777.67	609.56	286.00	586.00	903.00	659.30	535.09
31	34.00	79.00	162.00	166.12	307.97	26.00	64.00	135.00	129.51	254.86
32	79.00	180.00	326.00	250.78	269.35	58.00	149.00	294.00	213.36	236.10
33	162.00	369.00	653.00	494.24	481.98	122.00	313.00	583.00	431.65	438.12
42	306.00	731.00	1322.00	1164.01	1689.14	235.00	607.00	1155.00	949.70	1404.49
44	486.00	1018.00	1768.00	1433.08	1713.00	374.00	865.00	1526.00	1171.95	1423.50
45	224.00	449.00	776.00	633.60	766.16	171.00	383.00	666.00	516.82	633.83
48	79.00	165.00	286.00	227.59	251.59	60.00	136.00	239.00	186.03	211.21
49	21.00	47.00	78.00	57.01	48.77	17.00	39.00	69.00	48.40	43.09
51	56.00	146.00	304.00	266.31	433.54	38.00	110.00	247.00	205.21	351.45
52	212.00	488.00	868.00	660.20	703.92	155.00	388.00	713.00	527.62	588.54
53	130.00	285.00	552.00	429.55	505.56	95.00	235.00	444.00	345.26	421.51
54	295.00	693.00	1327.00	1033.04	1281.73	201.00	552.00	1101.00	824.91	1063.34
55	1.00	3.00	6.00	5.94	10.57	1.00	2.00	5.00	4.57	8.64
56	138.00	296.00	521.00	397.55	441.32	101.00	248.00	449.00	325.98	370.35
61	72.00	163.00	293.00	239.07	307.84	53.00	135.00	248.00	193.26	256.33
62	250.00	540.00	990.00	791.42	963.10	179.00	448.00	819.00	632.87	797.96
71	55.00	115.00	204.00	177.24	240.15	41.00	95.00	170.00	140.64	196.58
72	278.00	590.00	1099.00	872.81	1071.88	208.00	488.00	929.00	704.12	884.28
81	425.00	895.00	1588.25	1282.04	1489.02	327.00	752.00	1339.25	1048.55	1247.31
92	85.00	166.00	271.00	220.22	216.57	68.00	134.00	230.00	182.15	185.18
99	68.00	177.00	353.00	261.64	298.53	45.00	131.00	280.00	203.52	246.50
Panel B. Δ Establishments within 10 miles (Pre-innovation to 10 Years After)										
11	3.00	7.00	13.00	9.14	8.51	3.00	7.00	13.00	9.28	8.44
21	1.00	4.00	8.00	7.33	14.37	1.00	4.00	9.00	7.91	15.53
22	0.00	3.00	7.00	5.69	12.33	1.00	3.00	7.00	6.05	11.16
23	67.00	214.00	451.00	340.36	452.68	69.00	226.00	472.00	334.43	402.98
31	0.00	8.00	20.00	1.21	63.08	1.00	9.00	22.00	7.40	49.92
32	-47.00	-15.00	-1.00	-31.86	54.45	-31.00	-6.00	2.00	-20.45	45.06
33	-59.00	-12.00	5.00	-38.02	87.09	-34.00	-2.00	13.00	-19.87	72.46
42	-55.00	8.00	139.50	71.15	401.98	-7.00	47.00	175.00	106.52	331.91
44	-78.00	9.00	312.00	201.47	729.74	-16.00	93.00	380.00	258.38	645.60
45	-104.00	-3.00	164.00	77.48	458.61	-40.00	44.00	203.00	118.61	389.33
48	18.00	50.00	112.00	94.30	158.28	17.00	51.00	116.00	90.39	138.45
49	3.00	10.00	22.00	17.87	27.51	3.00	11.00	22.00	17.43	24.80
51	2.00	23.00	87.00	85.46	218.75	4.00	32.00	106.00	87.81	190.65
52	21.00	87.00	249.00	208.31	369.41	25.00	105.00	280.00	215.22	336.79
53	12.00	65.00	188.00	181.35	398.51	15.00	68.00	181.00	162.77	331.81
54	10.00	88.00	362.00	299.60	626.36	18.00	128.00	416.00	309.60	555.07
55	0.00	2.00	4.00	3.92	11.83	0.00	2.00	4.00	3.76	10.16
56	11.00	52.00	155.00	127.31	240.91	15.00	66.00	171.00	129.48	211.07
61	2.00	25.00	75.00	56.55	104.14	4.00	31.00	84.00	59.09	92.06
62	93.00	259.00	528.00	462.67	729.62	82.00	249.00	525.00	420.30	617.79
71	6.00	26.00	74.00	68.20	145.07	7.00	30.00	78.00	66.33	126.49
72	-11.00	35.00	191.00	148.89	454.37	2.00	73.00	229.00	171.92	401.86
81	26.00	140.00	354.00	315.21	609.67	33.00	155.00	382.00	306.63	526.61
92	10.00	31.00	66.00	48.84	65.59	14.00	35.00	74.00	51.88	59.66
99	-83.00	-1.00	80.00	8.94	206.92	-22.00	22.00	114.00	48.35	184.84

Note. This table reports the summary statistics of the firm-level establishment sample described in Section 2.4. Panel A shows the number of establishments within 10 miles of the startup one year prior to the startup's first-action grant date, by industry sector defined as the first two digits of the NAICS code. Panel B shows the change in the number of establishments from one year prior to ten years after the first-action date. The summary statistics are provided by the startup's first-grant status (**1**(Approved)).

Table A.4 — IV First Stage Regression Results

	Single-Family		Multi-Family		Commercial		Vacant	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: firm-level								
Leniency	0.6576***	0.6570***	0.6726***	0.6722***	0.6584***	0.6579***	0.6534***	0.6525***
	(0.0236)	(0.0237)	(0.0326)	(0.0326)	(0.0292)	(0.0293)	(0.0273)	(0.0274)
Holding year FE		✓		✓		✓		✓
Observations	38,101,658	38,101,658	35,986,905	35,986,905	5,237,993	5,237,993	1,595,151	1,595,151
R^2	0.33567	0.33634	0.37976	0.38014	0.36444	0.36523	0.37841	0.38084
Within R^2	0.06094	0.06086	0.06719	0.06709	0.06340	0.06332	0.06151	0.06133
Notes: Standard errors clustered at the art unit level in parentheses.								
Panel A: property-level								
Avg. leniency	0.5335***	0.5375***	0.3904***	0.3919***	0.5027***	0.5048***	0.6722***	0.6748***
	(0.0341)	(0.0339)	(0.0471)	(0.0471)	(0.0400)	(0.0400)	(0.0807)	(0.0805)
N	0.0003***	0.0003***	0.0001***	0.0001***	0.0002***	0.0002***	0.0007***	0.0007***
	(8.19×10^{-5})	(7.78×10^{-5})	(3.26×10^{-5})	(3.22×10^{-5})	(4.87×10^{-5})	(4.86×10^{-5})	(0.0001)	(0.0001)
Property Char's controlled		✓		✓		✓		✓
Observations	11,696,097	11,696,097	4,142,706	4,142,706	222,234	222,234	163,678	163,678
R^2	0.36687	0.36812	0.38878	0.38904	0.49798	0.49818	0.50995	0.51105
Within R^2	0.06871	0.07056	0.04272	0.04313	0.06057	0.06093	0.10419	0.10619

Notes: Standard errors clustered at the county level in parentheses.
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Note. This table reports the first-stage regression results from the IV regressions in Equation 4 (Panel A) and Equation 5 (Panel B). Columns (1), (3), (5), and (7) use the baseline specifications, while Columns (2), (4), (6), and (8) include holding-term (in years) fixed effects in Panel A and property-level control variables in Panel B.

B Comparing Data Axle and Zip Code Business Patterns

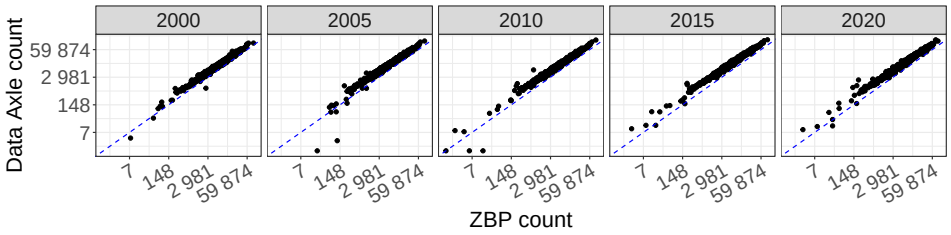
We find Data Axle provides a reliable measure of local business activity and aligns closely with statistics from the Census Bureau's Zip Code Business Patterns (ZBP) data. At the 3-digit ZIP code level, the correlation between the number of establishments reported by Data Axle and ZBP is remarkably high (0.966), as shown in Panel (a) of Figure A.2. Although Data Axle generally reports a higher number of establishments, this is partly due to its broader coverage. Unlike ZBP, which includes only payroll-based establishments, Data Axle captures a wider range of entities such as farms, railroads, postal services, notaries, and government offices without payroll employees—segments excluded from ZBP. Moreover, we notice that Data Axle assigns business identifiers at the individual professional level (e.g., each lawyer within a firm) that further explains the discrepancy.

The differences between the two datasets become more pronounced among larger establishments, as illustrated in Figures A.2 and A.3, presumably due to the disaggregation mentioned above. Even so, correlations remain strong across size categories, with values of 0.81 for firms with 250–499 employees, 0.68 for those with 500–999 employees, and 0.65 for those with 1,000 or more employees.

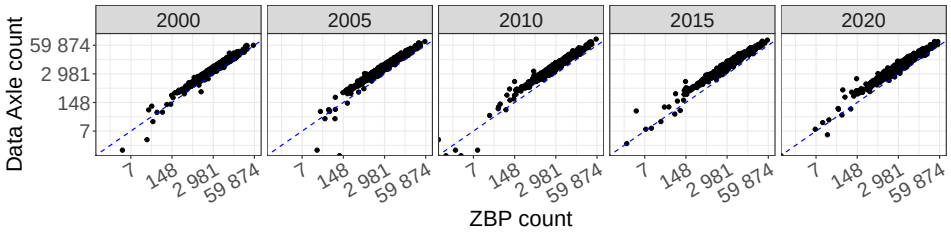
A notable limitation of Data Axle lies in its industry coding. As shown in Figure A.4, some records are missing NAICS codes, and the correlation between industry-level establishment counts in Data Axle and ZBP varies considerably, as detailed in Table A.5. These discrepancies are relatively minor in service-oriented sectors but are more pronounced in industries such as Federal Government and Agriculture, Forestry, Fishing, and Hunting. Although caution is warranted when conducting industry-specific analyses, we find Data Axle to be a consistent and comprehensive measure of local business activity, particularly useful for high-resolution spatial and longitudinal analyses.

Figure A.2 — Data Axle vs. Zip Code Business Patterns

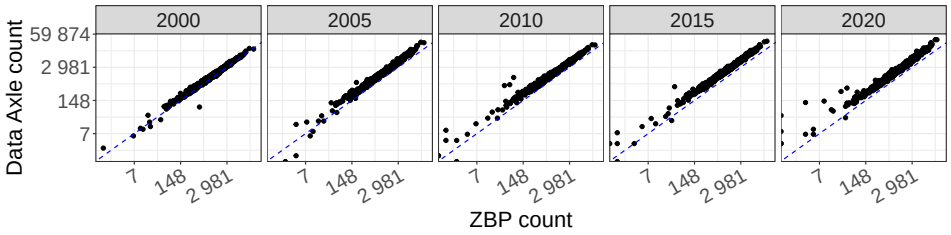
(a) All establishments



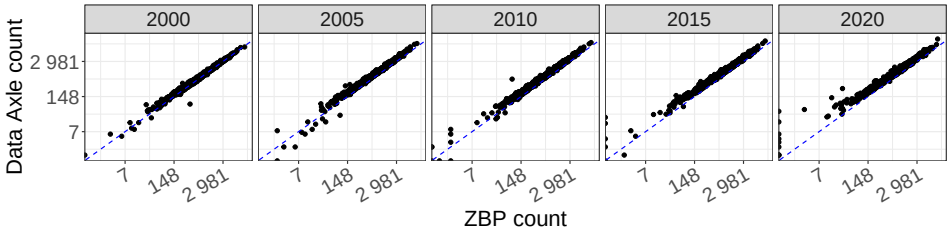
(b) Establishments with 1–4 employees



(c) Establishments with 5–9 employees



(d) Establishments with 10–19 employees



(e) Establishments with 20–49 employees

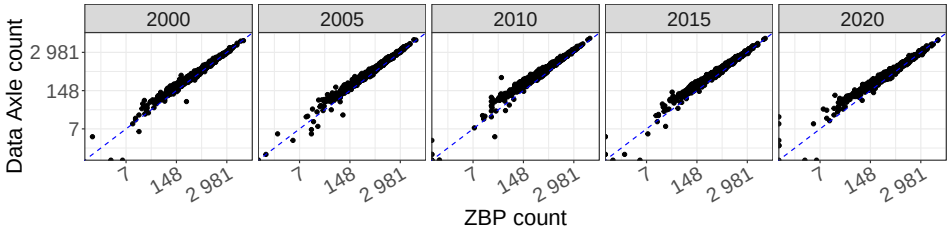
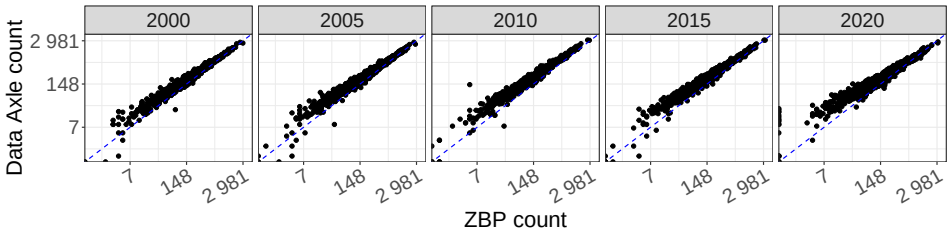
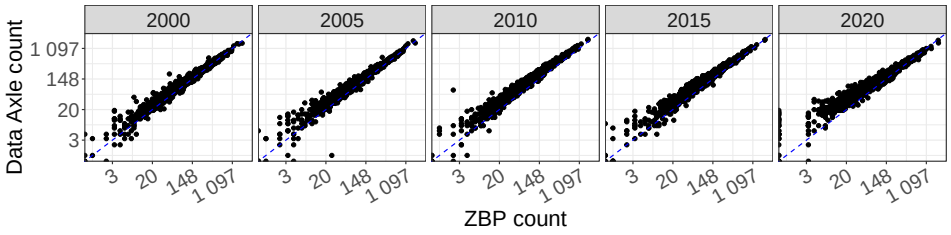


Figure A.3 — Data Axle vs. Zip Code Business Patterns (Cont'd)

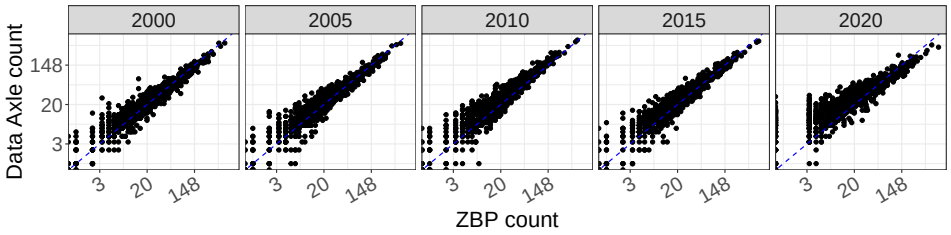
(a) Establishments with 50–99 employees



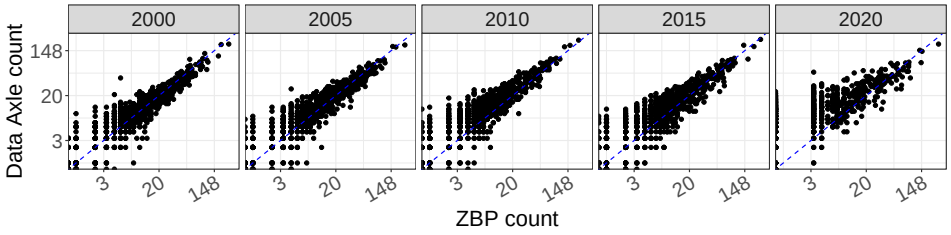
(b) Establishments with 100–249 employees



(c) Establishments with 250–499 employees



(d) Establishments with 500–999 employees



(e) Establishments with ≥ 1000 employees

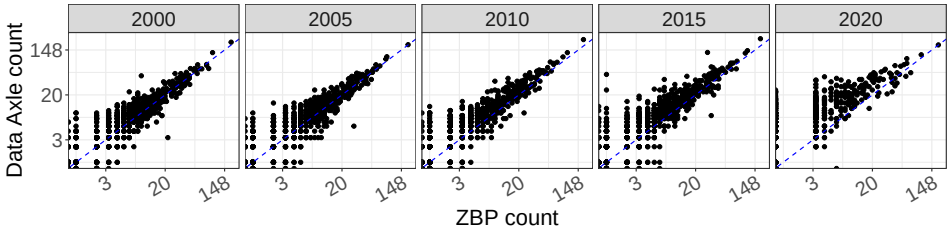


Figure A.4 — Data Axle vs. Zip Code Business Patterns by NAICS filled

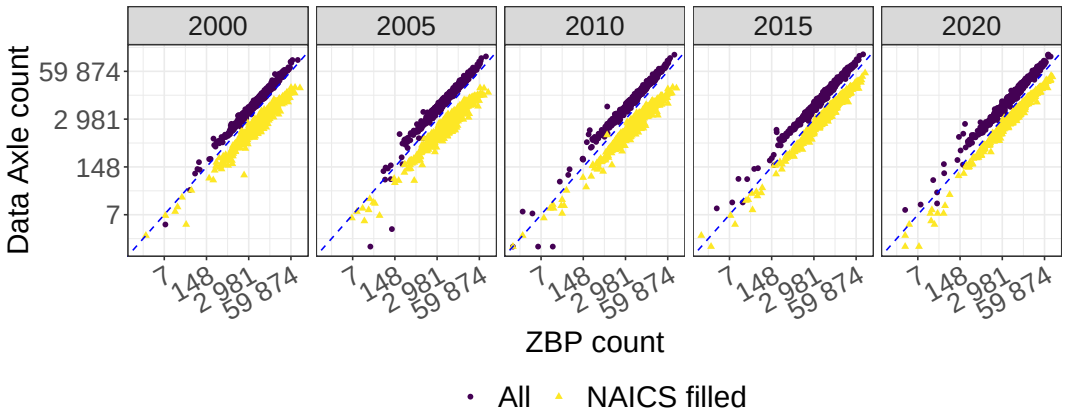


Table A.5 — Zip Code-Year Level Correlation between ZBP and Data Axle by Sector

2-Digit NAICS	Industry Description	Correlation
72—-	Accommodation and Food Services	0.8738
81—-	Other Services (except Public Admin.)	0.8626
42—-	Wholesale Trade	0.8468
56—-	Administrative and Support Services	0.7985
44—-	Retail Trade	0.7981
54—-	Professional, Scientific, and Technical Services	0.7959
62—-	Health Care and Social Assistance	0.7933
52—-	Finance and Insurance	0.7736
23—-	Construction	0.7703
61—-	Educational Services	0.7138
21—-	Mining, Quarrying, and Oil and Gas Extraction	0.7109
51—-	Information	0.7063
53—-	Real Estate and Rental and Leasing	0.6966
31—-	Manufacturing	0.6777
48—-	Transportation and Warehousing	0.5459
71—-	Arts, Entertainment, and Recreation	0.4455
22—-	Utilities	0.3605
55—-	Management of Companies and Enterprises	0.3241
11—-	Agriculture, Forestry, Fishing and Hunting	0.1759
99—-	Unclassified Establishments	0.0692
95—-	Federal Government	0.0055