The Commercial Real Estate Ecosystem

Ralph S.J. Koijen* Neel Shah† Stijn Van Nieuwerburgh‡

June 8, 2025

Abstract

We develop a new approach to understand the joint dynamics of transaction prices and trading volume in the market for commercial real estate. We start from a micro-founded model in which buyers and sellers differ in their private valuation of building characteristics, such as size, location, and quality. Consistent with the decentralized nature of the commercial real estate market, we model the probability that a seller meets a particular buyer, where the meeting probability depends on the characteristics of the buyer, the seller, and the building. In equilibrium, the mapping from building characteristics to observed transaction prices depends on the identity of the buyer and the seller, an important property missed by traditional hedonic valuation models. We estimate the model using granular data on commercial real estate transactions, which contain detailed information on the identity of buyers and sellers. Our central finding is that the identity of buyers and sellers has a first-order effect on both property valuation and the likelihood of trade. The importance of investor characteristics for valuations remains true, in fact is amplified, in a rich machine learning model that allows for non-linearities and interactions. We show how the model can be used for out-of-sample predictability and for counterfactual analyses on investment flows and prices. As a concrete example, we find that the Manhattan office market would have seen 7% lower valuations if it had not been for a large inflow of foreign buyers in 2013–2021. Our methodology extends to other private markets, including private equity, private credit, and infrastructure.

JEL: G11, G12, R30, R33

kywords: private and real assets, demand-based pricing, commercial real estate

*Booth School of Business, University of Chicago, ralph.koijen@chicagobooth.edu.

†Columbia Business School, ns3481@columbia.edu.

‡Columbia Business School, svnieuwe@gsb.columbia.edu.

The authors would like to thank Dayou Xi, Luofeng Zhou, Zi Wang, Laurenz De Rosa, and especially Michael Leahy for excellent research assistance. We thank the Fama-Miller Center at Chicago Booth and the Paul Milstein Center (PMC) at Columbia Business School for financial support to acquire the MSCI Real Capital Analytics data. The PMC also provided the Compstak data. We thank Jeff Fisher for providing the NCREIF data. The authors thank the participants at the Inquire Europe conference in Valencia for their helpful comments. Koijen thanks the Center for Research in Security Prices at the University of Chicago and the Fama Research Fund at the University of Chicago Booth School of Business.

1 Introduction

Financial markets have been undergoing a major shift from public to private markets over the past 25 years. For example, the number of publicly traded stocks in the U.S. decreased from around 8,000 in 1997 to around 4,000 in 2024. Over this same period, assets under management in the private equity market, broadly defined, have grown from a small base to $7.1 trillion in the U.S. and to $13 trillion globally.[1] Many large institutions have decreased the share of their investments in public equity and fixed income markets, and increased their allocation to alternatives.[2] The recent surge in private debt markets is another manifestation of this trend.[3] Using a broad definition of private and real assets that includes privately-held businesses and collectibles, Goetzmann et al. (2021) estimate a total asset base of $84 trillion in 2020.

This "great rotation" to private and real assets has important implications for finance research and practice. Private and real assets differ in important ways from public assets. They trade infrequently, often in bilateral search and matching markets. Prices are observed infrequently, which poses challenges for risk management and creates scope for volatility laundering. Assets are lumpy and heterogeneous, and unique features such as asset location can be important for valuations. Finally, there is an ecosystem of specialized investors. This paper provides a new methodology to understand the valuation of such assets.

Under the traditional rational asset pricing paradigm, the identities of buyers and sellers are irrelevant. If one buyer or one seller does not show up, another buyer or seller is always ready to take her place. There are no implications for asset prices, since the other buyer is assumed to have the same valuation for the asset in expectation.[4] Asset characteristics, "hedonics", determine asset value, not the composition of the investor base. There are many reasons to believe that real-world asset markets violate these assumptions, especially for private and real assets. This paper focuses on commercial real estate (CRE) assets, a large segment of the private and real asset space, and one for which we have excellent data on the identity of the buyers and sellers, as well as a diverse ecosystem with buyers and sellers of different types and sizes that changes significantly over time.

To motivate the value of considering investor composition, consider two CRE examples. A suburban

[1] McKinsey Global Private Markets Review 2024. Private asset markets include buyout, venture capital, growth, and other private equity, private debt, real estate, and infrastructure and natural resources, and assets under management are measured as of the end of June 2023. AUM in private markets is predicted to double by 2029.

[2] For example, public pension funds have increase the share in private and real assets from below 10% in the late 1990s to over 25% by 2020 (Mittal, 2024). Alternatives includes private equity, hedge funds, real estate, infrastructure, commodities, and miscellaneous alternatives.

[3] Private credit refers to lending outside of the traditional banking system, in which lenders work directly with borrowers to negotiate and originate privately held loans that are not traded in public markets. Strategies include direct lending, subordinated debt (mezzanine, second lien debt and preferred equity), distressed debt, and special situations. The private credit market has grown from $1 trillion to $1.5 trillion between 2020 and 2024. AUM is predicted to double to $2.8 trillion by 2028 (Morgan Stanley, Understanding Private Credit, June 2024).

[4] Even in heterogeneous agent asset pricing models where investors differ in their valuations, investors are highly sensitive to price fluctuations absent frictions, leading to excessively large flows and little price movement (Gabaix et al. (2025)).

shopping mall has seen its anchor tenant Sears go dark, triggering a major decline in foot traffic and in rents. This mall has been owned by a publicly-listed REIT that specializes in malls and is experiencing similar issues in other parts of its portfolio. This mall REIT is at a competitive disadvantage to reposition the property into an Amazon warehouse since its expertise is in owning and operating malls. In addition, the REIT is likely unable to raise equity capital to fund such a repositioning, since its stock price is trading below the NAV. An opportunistic real estate private equity (REPE) fund, like Blackstone, which is looking for creative ways to deploy its ample dry powder, with geographic and sector flexibility and generalist managerial talent, may be better suited to reposition the asset. The asset is worth more in the hands of the REPE fund than in the hands of the REIT. The value of the mall increases in the amount of capital of potential REPE buyers.

A second example considers the importance of large foreign institutional investors, mostly sovereign wealth funds and pension funds, for U.S. CRE markets. Portfolio diversification considerations and expectations of lower returns on bond or stock markets may lead them to increase allocations to alternative asset classes like U.S. CRE. Large institutional foreign capital is particularly fond of large, high-quality office and retail properties in "superstar cities," which they consider to be "bond-like" assets. How large an effect do these foreign buyers have on the prices of these properties when they buy billions of dollars' worth of such assets? What would the prices of these assets have looked like in the absence of the foreign investment surge these markets experienced in 2015 and 2018? What other assets did domestic buyers who were crowded out by the foreign investment surge actually buy, and how did those purchases affect valuations? Were the buildings of similar size and in similar locations? Answering these questions helps us understand the spillover effects of this foreign purchase activity on other locations and CRE sectors and substitution patterns across asset features. These questions have taken on renewed urgency as cross-border investors turned from net buyers into net sellers after 2018.

Our paper is the first to systematically document the composition of buyers and sellers in CRE and its evolution over time. We use data from MSCI Real Capital Analytics (henceforth RCA) that cover the near-universe of CRE transactions between the year 2001 and 2023. It contains about 470,000 transactions that combine for $10 trillion in traded value, covering the four traditional CRE sectors of apartment, office, retail, and industrial properties. RCA has spent enormous effort unraveling the true buyers and sellers involved in each transaction.[5] There are about 325,000 unique investors in our sample. Using the history of transactions, we build up the portfolios of these investors, and find that they account for about $8.6 trillion dollars in holdings at the end of 2023. We document new stylized facts on the composition of the

[5]This is a difficult task since CRE properties are usually bought using a separate LLC for each building, typically with a non-descriptive name.

investor base: who trades and holds what CRE assets. We organize investors into eight types, but also note important heterogeneity in terms of total portfolio size (and, related, the size of the typical asset traded), sector concentration, and geographic concentration. The resulting picture is one of partial segmentation, where smaller investors tend to specialize in smaller assets and fewer geographies, while larger investors tend to be national in scope and trade larger ticket sizes.

We develop a model that takes seriously the decentralized nature of trade of private assets. Each investor's valuation for each asset depends on a conditional mean, which is a (potentially highly non-linear) function of asset characteristics, and an idiosyncratic stochastic component that could capture liquidity or funding shocks hitting that investor. The conditional mean component reflects that investor's valuation of the characteristics of that asset. Different investors have different valuations for the property's attributes such as asset size, location, cash flow generation, etc.

The first main finding of the paper is that allowing for investor-specific feature valuation substantially increases the model's explanatory power for transaction prices. Compared to a model with asset hedonics, geography-specific macro variables, and time fixed effects, adding investor characteristics significantly improves the model's fit. Specifically, investor characteristics reduce the unexplained variation of the richest hedonic model by over 50%.

These results are for a Light Gradient Boosted Machine (LGBM) model. The LGBM is a tree-based model that allows for non-linearities and interaction effects of all variables. Compared to a standard linear hedonic model with a rich set of property characteristics, local macro variables, year fixed effects, and market fixed effects, the LGBM hedonic valuation model already delivers a substantial improvement in fit. The R^2 goes up by 13-24% points depending on the sector relative to the linear hedonic model.[6] Functional form flexibility substantially improves model accuracy.

Adding investor characteristics to the best hedonic LGBM model improves the R^2 by another 8-20% points, depending on the sector. Investor characteristics reduce the unexplained variation in the hedonic LGBM by 48-64%. The resulting valuation model has an in-sample R^2 between 87% and 91% points compared to 43% to 59% points for a linear hedonic valuation model. The increase in R^2 is $\approx$ 30% points for Apartments and Industrial, 44% points for Office and 34% points for Retail. We follow best practices in choosing hyper-parameters, mindful of overfitting concerns. Out-of-sample tests of the LGBM model with investor features deliver an OOS R^2 between 78% and 84% points. These R^2 constitute a reduction of 34-51% in the unexplained OOS variation of the hedonic LGBM.

[6] The richness of the hedonic LGBM model (without investor features) may well allow it to infer features related to investor characteristics from the property features, market features, and geographic and time fixed effects via *triangulation*, in a similar way that the random forest model in Fuster et al. (2022) can infer race from other borrower characteristics in a residential mortgage default prediction problem where the household features do not include race.

We then model the probability of listing a property for sale, meeting a buyer, and transacting with that buyer. The meeting probability lies at the core of the model. We model the likelihood that a buyer meets a building that is for sale as an increasing function of the size of the buyer, the similarity of the buyer and seller in terms of size, and the similarity of the asset for sale to the other assets the buyer has in her portfolio, where similarity is measured in terms of the asset's size, geography, sector, and asset quality. Conditional on a meeting, the transaction goes through if the value of the buyer exceeds that of the seller. Finally, the listing probability depends on time in a way that allows the model to match the time-varying transaction volume, given all other parameters.

A key insight that makes the estimation of the meeting probability computationally feasible is to use negative sampling, a concept used in the natural language processing computer science literature (Mikolov et al., 2013). Rather than assessing how a candidate buyer's meeting probability compares to that of *every* other potential buyer (B potential buyers) for every one of N buildings in every one of T periods, which involves computing $B \times N \times T$ numbers for every sector, we only need to sample a relative small number of $K << B$ potential buyers, $K-1$ of which did not buy the building and one who did. Maximizing the likelihood ratio of the actual buyer relative to the other $K-1$ non-buyers delivers consistent parameter estimates (Ma and Collins, 2018). The higher K, the more efficient the estimator. We show that this procedure works well for $K = 1,000$, with diminishing efficiency gains for larger values.

Having estimated the model parameters, we turn to the main counterfactuals of interest. An important object in the computation of counterfactuals is the *potential price distribution* of a property that is listed for sale. The seller could have met any one of a large number of buyers besides the one she transacted with. By sampling (with replacement) many potential buyers, in proportion to their probability of meeting the seller, and discarding the ones where there are no gains from trade, we can construct a distribution of prices for which the property could have traded. This potential price distribution is a useful object in its own right. It can be used to assess the seller's luck or the skill of her broker (by seeing where in the distribution the realized sales price lies), for setting a future sales price target, or for risk management purposes. More importantly, the potential price distribution is a useful input in three key applications.

The first application is the problem of building owners looking to sell their property in the next period. They want to predict the transaction price and the price risk. A key insight from the valuation model is that building, market, and investor characteristics all matter for the observed transaction prices. Our valuation model captures the nonlinear and interaction effects of those input variables on prices. Predicting the transaction price and estimating the uncertainty of the price amounts to predicting the values for the inputs in the next period. The most difficult inputs to predict are the features of the investor that will buy the property. This is where the estimated meeting model comes in, which draws potential buyers in proportion to their likelihood of meeting the building and executing a sale. We show that our benchmark model outperforms the linear model and the hedonic LGBM in producing an OOS prediction for the price that is closer to the actual transaction price more frequently and in terms of percent distance.

The second application highlights the importance of heterogeneity in valuations. We study an episode in 2007, just prior to the onset of the Global Financial Crisis, that saw several very large transactions between REITS sellers and REPE fund buyers. The REPE funds operated in a very strong fundraising environment and may have been compelled to put the money to work. We ask what prices for the traded assets would have looked like if REPE funds had had the same valuation for these properties as all other investors in 2007. We can also ask what prices would have looked like if REPE funds had not been in the market for these assets at all or to a reduced extent. How different would CRE prices have been and would they have spelled trouble brewing? To quantify the impact of a group of buyers on the price, we remove that group from the potential matching partners, and recompute the potential price distribution. The difference between the benchmark potential price distribution and the counterfactual price distribution that removes one investor group quantifies the impact of that investor group on prices. Naturally, the average price is a key moment as is the inter-quartile range, which is a measure of price risk.

The third application shows how the model can be used to analyze substitution patterns in deal flows across building size, sector, and location. This question is of interest to policymakers and CRE investors alike. Concretely, if a large number of foreign investors purchase high-end Manhattan office buildings, what is the impact on smaller office buildings in Manhattan, apartment buildings in Manhattan, and office buildings in other locations such as Chicago, Los Angeles, and San Francisco? We show that the model can be used as a lens to analyze transaction flows along these dimensions. The counterfactual price distribution spells out the impact of foreign investors on transaction prices and price risk.

Related Literature Our paper contributes to the literature on valuing private assets. Practitioner approaches often ignore risk, simply comparing total cash outflows from operations and asset disposition to the initial amount invested (TVPI) or taking into account the timing of the cash flows and time value of money (IRR). The finance literature has emphasized the importance of recognizing the systematic risk in the cash flows, with risk being measured in increasingly sophisticated ways. The public market equivalent (PME) approach Kaplan and Schoar (2005) considers the alternative investment in a stock market index fund with the same timing of cash flows. The generalized PME approach of Korteweg and Nagel (2016) takes into account that the private asset cash flows may have a risk exposure to the equity market different from one. The strip-by-strip approach of Gupta and Van Nieuwerburgh (2021) takes into account that private asset cash flows may be exposed to a richer menu of risk factors present in public markets, not just the equity index but several cross-sectional equity factors as well as interest rate risk priced in the cross-section of government bonds, and that exposures of operating cash flows may differ from those of disposition cash flows. [7] While Gupta et al. (2025) value CRE assets using a richer model for the cash flow dynamics, like the earlier approaches, they continue to assume that the Stochastic Discount Factor can be inferred from public markets data. This paper deviates from this tradition and does not assume that the market prices of risk for the private market-relevant risk exposures are determined in public securities markets. The valuation model allows for the investor composition in the private asset market to impact valuations.

We also contribute to the literature that studies asset prices using asset demand systems (Koijen and Yogo, 2019; Koijen et al., 2024). We share the same objective to jointly understand valuations, asset characteristics, portfolio holdings, and trades. This is the first paper to develop a demand system for private assets. Since these assets have fundamentally different features than public assets in that they trade in bilateral markets, the approach we pursue is very different. Rather than modeling portfolio shares for divisible shares of a public stock, we model the probability that an entire private asset changes hand in a meeting between one buyer and one seller. We provide a micro foundation of the valuation equation reflecting this unique feature of private assets. Our approach extends to other private assets or assets that trade in over-the-counter markets: corporate and municipal bonds, private credit, private equity, infrastructure and natural resource assets, residential real estate, collectibles (art), et cetera.

The standard hedonic regression approach for valuing private assets, which posits that the asset's value is a linear function of its physical characteristics, each with a fixed price attached to them, has a long tradition in economics (Lancaster, 1966; Griliches, 1971; Rosen, 1974; Witte et al., 1979; Wallace, 1996). Our model is richer in that the *composition of the investor base* affects the mapping from characteristics to transaction prices. Indeed, the transaction price reflects the average of the buyer's and seller's valuation, and thus their valuation of the characteristics. Since the investor composition changes over time and in the cross-section, the standard hedonic model's residual variance would change over time and in the cross-section. If hedonic coefficients on the building characteristics (the prices of the characteristics) are allowed to depend on time, then that time-variation may pick up valuation changes that are in fact due to a changing investor mix. For example, the characteristic "number of floors" may appear to have gained in importance (price) in the estimation of a standard hedonic model for Manhattan office. In reality, it could instead reflect the increased demand from foreign investors who have a strong preference for tall office buildings. Controlling for investor mix can "cleanse" the hedonic price index from the impact of (time-varying) investor composition.

Our approach accounts for the dynamic composition of the investor base in the context of a non-linear

[7] Other important contributions to the literature of measuring risk and return private equity performance literature include (Korteweg and Sørensen, 2010; Driessen et al., 2012; Ang et al., 2018).

machine learning model. This allows us to better capture the complex non-linearities and interaction effects inherent in the data, patterns that linear models may overlook. In the process, we also improve on the hedonic model without investor characteristics, by adding several important characteristics including an important variable that capture neighborhood quality, and by generalizing the linear regression approach to a rich nonlinear model with interaction effects. Investor characteristics add substantial explanatory power for valuations *even relative* to the rich hedonic LGBM model.

A large REIT literature uses standard empirical asset pricing tools to explain returns of publicly listed CRE companies (Van Nieuwerburgh, 2019, for a recent example). However, REITS represent only 12% of transaction value in our sample. As a large and growing share of CRE transactions moves from public to private markets, the task of understanding risk and return in private CRE becomes both more difficult and more important.[8] Peng (2016), Sagi (2021), and Giacoletti (2021) study risk and return at the individual property level and emphasize liquidity considerations. Our potential price distribution provides a complementary liquidity risk measure. Plazzi et al. (2008, 2010) study cross-sectional variation in expected returns and expected cash flow growth in CRE markets across U.S. cities.

Recent work has begun to analyze the importance of investor heterogeneity for CRE valuation. Ghent (2021) shows that institutional investors like REPE funds concentrate their CRE investments in markets with many other similar institutions. The liquidity generated by a large pool of buyers results in higher valuations for assets with otherwise identical cash-flows. Similarly, Cvijanović et al. (2022) finds that institutional investors in CRE are highly sensitive to liquidity considerations. Badrinza et al. (2022) emphasize the importance of same-nationality buyers and sellers in a CRE transaction as a mechanism to resolve the gravity puzzle. In earlier work, Badarinza and Ramadorai (2018) exploit differences in pre-existing concentration of foreigners in London neighborhoods to identify and estimate the effect of cross-border purchases on home prices.

We contribute a new estimation technique to the matching literature. We import insights from the Artifical Intelligence (AI) literature, which has estimated large language models using word embeddings. Particularly, using negative sampling (Mikolov et al., 2013; Ma and Collins, 2018) is necessary to overcome the computational challenges when each of many assets could trade each period and be purchased by one of many buyers. Fox (2018) proposes a maximum score estimator to overcome the curse of dimensionality in many-to-many matching settings. Gabaix et al. (2024) uses techniques from the LLM literature to define similar stocks using investor portfolio data. Badarinza et al. (2024) estimates the matching function in the housing market, using unique data on online search for properties and broker meeting requests. We do not

[8]Van Nieuwerburgh et al. (2015) contains a discussion of the challenges with measuring returns for infrequently traded CRE assets and the various approaches that have been proposed to make progress.

directly observe meetings and must estimate the meeting function.

The rest of the paper is organized as follows. Section 2 develops the theoretical framework, which contains a valuation and a matching model, bridging key methodological gaps in the current literature. Section 3 details the estimation procedure for the models. Section 4 describes the data and provides summary statistics. Section 5 shows results for the parameter estimates of the valuation model and Section 6 for the transaction model. Section 7 explores the three application outlined above. Section 8 concludes. The appendix contains additional details on the data and empirical methods.

2 Model

2.1 Notation

There are four building types, which we call sectors: office, retail, industrial, and apartments. We estimate separate models for each sector. Time is indexed by $t, t = 1, \ldots, T$. Buildings are indexed by n, $n = 1, \ldots, N$, and investors by $i, i = 1, \ldots, I$. If a building is sold, then we refer to the seller as investor s and the buyer as investor b. The transaction price is $P_t(n)$.

Buildings are differentiated by a vector of characteristics, x_{nt}. We use several asset characteristics. The first seven are measured at the property level in the RCA data: log asset size (measured in terms of square feet for office, retail, and industrial or number of apartment units for multifamily), a variable indicating whether the building is located in a superstar city, a variable indicating whether the property is located in the central business district (CBD), the number of floors, log building age, the log renovation-adjusted age, and a property sub-type dummy.

Next, we consider five local macro-economic variables, the first four of which vary at the market-level. We define 60 markets.[9] We fold into x_{nt} the following four market-specific macro variables: the log population (for apartments, from the BEA) or log employment (for the other sectors, BEA), log real personal income per capita (BEA), the occupancy rate (NCREIF), and the NOI growth rate (NCREIF).[10]

The next asset characteristic is the net effective rent (NER) per square foot for office, retail, and industrial or the net operating income (NOI) per apartment unit of the property or the block (small geographic area) where the property is located. For simplicity, we will refer to the NOI for multifamily as the NER, so we can use the same variable name for all four sectors. Appendix A.3 describes the details.

[9] Appendix A.2 shows our mapping from RCA markets to our 60 markets.

[10] The National Council for Real Estate Investment Fiduciaries (NCREIF) is a consortium of large institutional real estate investors that share data on their institutional-quality portfolios allowing for the construction of market-level cash flow and occupancy statistics. We thank NCREIF for making these data available to us. We build a bridge between the NCREIF markets and our 60 markets. Appendix A.2 provides the details.

All five market variable series vary over time. The first one is a measure of market size, the second one an indicator of the average income level in the market, the next two are key metrics of the overall health of the specific CRE sector in that market. These four market characteristics are all natural candidates to be included in the hedonic model, even though they not always are. The fifth one, the NER, measures differences in average cash flows per unit across fine geographies. We interpret it as a (time-varying) measure of asset or neighborhood quality. This NER variable, which is the most important of the five, is new to the hedonic literature.

The final characteristics that enters into x_{nt} is the transaction type categorical variable, whether the asset belongs to a distressed sale, an entity sale or a regular sale.

Investor heterogeneity is captured by a vector of seven characteristics, z_{it}. This vector includes information on the investor's type (e.g., REIT, PE fund or foreign investor), the log dollar value of the investor portfolio across all sectors, the share of the overall portfolio in superstar markets, the share of the portfolio in the same market as the asset in question, the share of the portfolio in the same sector, whether the investor transacted as a part of a Joint Venture (JV), and the relative size of the buyer and the seller measured as the log ratio of the buyer's portfolio size to the seller's portfolio size.

2.2 Investors' Private Valuations

We first provide the main empirical specification of an investor's private valuation in Section 2.2.1, for which we provide a micro foundation in Section 2.2.2.

2.2.1 A characteristics-based model of investors' private valuations

We assume that investor i's valuation of building n is given by

$$v_{it}(n) \quad \equiv \ln V_{it}(n) = h\left(z_{it}, x_{nt}; \gamma_t\right) + \epsilon_{it}(n), \tag{1}$$

where γ_t is a vector of parameters that may vary over time. The error term, $\epsilon_{it}(n)$, captures liquidity or funding constraints, unobserved quality characteristics or differences in investors' beliefs. We assume that the valuation shocks are uncorrelated across investors and buildings, conditional on all characteristics that include investor-type and geography fixed effects, and normally distributed, $\epsilon_{it}(n) \sim N\left(0, \sigma_t^2\right)$.

In our empirical specification, we use a non-linear model that flexibly captures interactions between investor and asset characteristics without imposing a specific functional form. Market fixed effects are included to capture local heterogeneity.

2.2.2 A micro foundation of the private valuation model

Before proceeding, we provide a micro foundation of the valuation equation in (1) and, in particular, how an investor's valuation depends on characteristics, $h_{it}(n) = \beta'_{x,i} x_{n,t}$, with a loading vector (price vector for the hedonics) that itself depends on investor characteristics $\beta_{x,i} = \beta_x z_{i,t}$. To explain the core economics, we focus on a linear model. The model can be extended to feature non-linearities.

We consider a two-period model, $t = 0, 1$. In period $t = 0$, the investor considers buying a building. In period $t = 1$, the investor receives the net cash flow and the resale value of the building. In mapping the model to the data, we think of $t = 1$ as the payoff of the building over a longer holding period.[11] The building in question may be part of a broader property portfolio or, in case of a pension fund or insurer, part of a broader portfolio that may include equities and fixed income assets. Without the new building, the broader portfolio generates a payoff D_{1i}. If the investor has no other assets, $D_{1i} = 0$.

At $t = 0$, the investor has access to cash C_{0i} that can be used to purchase the building. Any remaining cash earn a gross risk-free rate that is normalized to one. This implies that, without purchasing the new building, the investor's wealth in $t = 1$ is $A_{1i} = D_{1i} + C_{0i}$.

The new building generates a time $t = 1$ payoff of N_{1i}. We assume that $(D_{1i}, N_{1i}) \sim N(\mu_i, \Sigma_i)$. We denote elements of μ_i and Σ_i by their logical counterparts, e.g., μ_{Ni} and Σ_{NDi}. We allow investors to have heterogeneous beliefs about future payoffs and index all moments by i. If the investor adds the building to her portfolio, period $t = 1$ wealth equals $A^P_{1i} = D_{1i} + C_{0i} - P_0 + N_{1i}$, where P_0 is the purchase price of the property.[12] To complete the model, we assume that the investor has mean-variance preferences over terminal wealth:

$$\mathbb{E}_i[A_{1i}] - \gamma_i \text{Var}_i(A_{1i}),$$

where γ_i is risk aversion.

As the building is a discrete purchase (as opposed to buying shares in a company), we compute the investor's valuation as the $t = 0$ price $P_0 = V_{0i}$ that makes the investor indifferent between purchasing the building or not. This valuation then solves the following equation:[13]

$$\mathbb{E}_i[D_{1i} + C_{0i}] - \gamma_i \text{Var}_i(D_{1i}) = \mathbb{E}_i[D_{1i} + C_{0i} - V_{0i} + N_{1i}] - \gamma_i \text{Var}_i(D_{1i} + N_{1i}),$$

where the left-hand side is the investor's utility if she does not buy the building and the right-hand side is

[11] At the expense of some additional notation, this micro foundation can be extended to a multi-period model.

[12] We assume throughout that the investor has deep pockets and $C_{0i} > P_0$. The model can easily be extended so that the investor finances a fraction ϕ of the property with a mortgage at a rate $R > 1$, or the investor's valuation is bounded by C_{0i}, which caps the valuation at C_{0i}.

[13] To focus on the key economics, we abstract from transaction costs here.

the utility when she does buy the building. We use this equation to solve for the investor's private valuation:

$$V_{0i} = \mathbb{E}_i[N_{1i}] - \gamma_i \mathrm{Var}_i(N_{1i}) - 2\gamma_i \mathrm{Cov}_i(D_{1i}, N_{1i}), \tag{2}$$

which depends on the expected payoff of the property, $\mathbb{E}_i[N_{1i}]$, a discount for its variance, $\mathrm{Var}_i(N_{1i})$, and a further discount or premium that depends on the property's covariance with other assets in the investor's portfolio, $\mathrm{Cov}_i(D_{1i}, N_{1i})$.

The main valuation equation is (2). To obtain a characteristics-based model of investors' private valuations, we follow Koijen and Yogo (2019) and model the moments as functions of characteristics with investor-specific coefficients:

$$\begin{aligned}
\mathbb{E}_i[N_{1i}] &= \beta_{i0}' x_n, \\
\gamma_i \mathrm{Var}_i(N_{1i}) &= \beta_{i1}' x_n, \\
\gamma_i \mathrm{Cov}_i(D_{1i}, N_{1i}) &= \beta_{2i}' x_n,
\end{aligned}$$

Beliefs about first and second moments, risk aversion γ_i, and time-1 portfolio payoffs D_{1i} are all allowed to be heterogeneous across investors. We model this heterogeneity across investors as a function of the investor characteristics including investor type, size of the investor portfolio, share of assets in superstar cities, share of assets in Office, etc. :

$$\beta_{ki} = \beta_k' z_i,$$

for $k = 0, 1, 2$. This yields the valuation equation in (1).[14]

2.3 Transaction Prices

There is a proportional cost of $cP_t(n)$ if investor i and j transact. The bargaining power of the seller is measured by $\beta \in [0, 1]$, which is used to split the proportional transaction cost and in setting the price. Investor b is willing to buy the building when

$$V_{bt}(n) > P_t(n)(1 + \beta c),$$

or, equivalently,

$$v_{bt}(n) - \chi_b > p_t(n),$$

[14] The difference between a valuation in levels versus logs arises because we assume mean-variance preferences. Alternatively, we can assume CRRA preferences and log-linearize the model.

where $\chi_b \equiv \ln(1+\beta c) \simeq \beta c$. Analogously, investor s is willing to sell the building when

$$V_{st}(n) < P_t(n) - (1-\beta)cP_t(n),$$

or, equivalently,

$$v_{st}(n) - \chi_s < p_t(n),$$

where $\chi_s \equiv \ln(1-(1-\beta)c) \simeq -(1-\beta)c$.

We assume that the price is set by Nash bargaining (in logs),[15] implying

$$p_t(n) = \beta v_{bt}(n) + (1-\beta)v_{st}(n) - c.$$

This implies that trade takes place when

$$v_{bt}(n) - v_{st}(n) > c.$$

In our empirical analysis, we set $c = 0$ and $\beta = 0.5$.

2.4 Transaction Probabilities

We now discuss the model of the listing probability, the meeting probability, and the probability of transacting conditional on a meeting taking place between a buyer and a seller. At this stage, we assume that we know the parameters of the valuation model. We therefore have the component of valuations explained by characteristics, h_b and h_s, for the buyer and seller.

We model the probability of transaction as:

$$\pi(b,s) = \pi_{l,t}\pi_m(b,s)\pi_\tau(b,s),$$

where the first term, $\pi_{l,t}$, is the probability of listing, π_m the probability of meeting a listed building, and π_τ the probability that the valuation of the buyer exceeds the valuation of the seller (conditional on a meeting).

We assume that the listing probability is only a function of time and sector, $\pi_{l,t}$. More advanced models can model the listing decision as a function of the price dynamics in a certain market, for example to capture the importance of reference points. In this case, the listing probability would increase in last year's price in that market relative to a running maximum of the price index up to that point. Indeed, if prices declined

[15] Formally, the price is the solution to $p^* = \arg\max\left((p-c)-v_s\right)^\beta \left(v_b-(p+c)\right)^{1-\beta}$.

sharply last year, owners of buildings are less likely to list. Additionally, listing probabilities could reflect market flow dynamics, such as the entry or exit of large investors or shifts in the composition of investor types, which influence market liquidity and selling incentives.

The second term is the meeting probability, which is a reduced-form approach to modeling directed search. We assume that the likelihood of a buyer b meeting a building listed by seller s depends on investor characteristics, market frictions, and the degree to which the features of the property for sale aligns with the buyer's property portfolio:

$$\pi_m(b,s) = \begin{cases} 0 & \text{if } b = s \\ \frac{\exp\left(\lambda_1 S_b + \lambda_2 \Delta S_{b,s}^{-1} + \lambda_3' \delta_{b,s} + \lambda_4 N_b\right)}{\sum_{c \neq s} \exp\left(\lambda_1 S_c + \lambda_2 \Delta S_{c,s}^{-1} + \lambda_3' \delta_{c,s} + \lambda_4 N_c\right)} & \text{otherwise,} \end{cases} \tag{3}$$

where $\lambda_i \geq 0$, $i = 1, \ldots, 4$.

When $\lambda_i = 0$, $i = 1, \ldots, 4$, the matching probability is the inverse of the number of buyers and we have random matching. When $\lambda_1 > 0$, larger buyers are more likely to meet the asset, consistent with their higher trading activity and greater market presence.[16] When $\lambda_2 > 0$, matching probabilities increase when buyers and sellers have similar size (high $\Delta S_{b,s}^{-1}$). Investors may prefer to deal with investors of similar size to reduce deal execution risk.

The vector $\delta_{b,s}$ captures the buyer's consideration set. It measures the similarity between the characteristics of seller s's building and the portfolio characteristics of buyer b. Similarity has four dimensions. First, asset size alignment, the similarity between the building's log dollar value and the average log dollar value of the assets in the buyer's portfolio. Second, market alignment, the similarity between the building's geography and the share of the buyer's portfolio in that geography. Third, sector alignment, the similarity between the building's sector and the share of the buyer's portfolio in that sector. Fourth, quality alignment, proxied by the log NER in the building's neighborhood and the average log NER in the neighborhoods of the properties in buyer b's portfolio. Hence, λ_3 is a 4×1 vector. For a given characteristic x_s of the building and x_b, the average characteristic of the buyer's portfolio, similarity is derived from the distance $|x_b - x_s|$ by converting smaller distances into higher similarities. Specifically, we use 1 - $|x_b - x_s|$ for geography and sector alignment and $\frac{1}{1+|x_b - x_s|}$ for size and NER alignment.

Lastly, N_b is an indicator variable that equals one when the buyer b owns more than two buildings and zero otherwise.

The third term measures whether a transaction takes place conditional on a meeting: $\pi_\tau(b,s) = P\left(V_s < V_b\right)$.

[16] In reality, a seller may have bilateral meetings with multiple buyers, and choose among those making a bid for the asset based on multiple criteria ranging from offered price to certainty of execution. The buyer size as well as some of the other variables in (3) capture in reduced-form that larger buyers are more likely to emerge as winners from this process.

We have:

$$P\left(V_s < V_b\right) = P\left(\epsilon_s - \epsilon_b < h_b - h_s\right) = \Phi\left(\frac{h_b - h_s}{\sqrt{2}\sigma}\right).$$

under our assumption that $\epsilon_b, \epsilon_s \sim N(0, \sigma^2)$.

The probability that a transaction does not happen is given by

$$\pi_n(s) = (1 - \pi_l) + \pi_l \sum_b \pi_m(b, s)(1 - \pi_\tau(b, s)). \tag{4}$$

The first term, $1 - \pi_l$, is the probability that the seller does not list the building. The second term is the probability that the building lists, and thus a match will be formed, but the valuation of the buyer does not exceed the valuation of the seller, as captured by $1 - \pi_\tau(b, s)$. As $\sum_b \pi_m(b, s) = 1$, $\sum_b \pi_m(b, s)(1 - \pi_\tau(b, s)) \leq 1$ and $\pi_n(b, s) \leq 1$.

The probabilities that the building transacts with some buyer and that it does not transact always sum to one:

$$\pi_{l,t} \sum_b \pi_m(b, s)\pi_\tau(b, s) + \pi_n(b, s) = 1.$$

3 Estimation Procedure

The model is estimated sequentially, in two steps. In Section 3.1, we discuss the estimation of the valuation equation. We discuss the estimation of the listing and matching model in Section 3.2.

3.1 Valuation Model

We begin by estimating the valuation model, which is done separately for each sector (A, O, I, and R). We rely on observable transaction prices to estimate the other parameters. The log price is given by

$$p_t(n) \quad = \quad \frac{1}{2}(h_{bt}(n) + h_{st}(n)) + \frac{1}{2}(\epsilon_{bt}(n) + \epsilon_{st}(n)). \tag{5}$$

We note that prices are observed only when $v_b > v_s$, which is when $h_b - h_s > \epsilon_b - \epsilon_s$. The difference in shocks is uninformative about the sum of the shocks because

$$\mathbb{E}\left[\epsilon_b + \epsilon_s \mid \epsilon_b - \epsilon_s\right] = 0,$$

when $\epsilon_b, \epsilon_s \sim N(0, \sigma^2)$. This implies that estimating the valuation model based solely on the observed transactions does not create a bias.

The investor valuations h_{bt} and h_{st} contain $\dim(x) + \dim(z) + 2$ predictors, where the last two predictors refer to the RCA Market and year categorical variables. To flexibly capture the complex relationships between asset and investor characteristics, we use LightGBM (Guolin Ke, 2017), a gradient-boosted decision tree model. LightGBM is well-suited for this task as it efficiently handles high-dimensional data, categorical variables, captures non-linear interactions without requiring pre-specified terms, and incorporates built-in regularization to mitigate overfitting. We use a two-step "coordinate descent" procedure with two separate LightGBM regressors for the buyer and the seller. In each iteration, we first hold the seller-side model fixed and update the buyer-side function by setting the training target to $2p(n) - v_s(x_n; z_s; \gamma)$. Next, we hold the buyer-side fixed and update the seller-side function using $2p(n) - v_b(x_n; z_s; \gamma)$ as the target. Alternating these steps ensures that, at convergence, the final predictions satisfy $p(n) = 0.5 \cdot \big(v_b(x_n; z_s; \gamma) + v_s(x_n; z_s; \gamma)\big)$ in a flexible, non-linear manner. The algorithm converges quickly in a few iterations, due to the high correlation between v_b and v_s. We discuss the empirical implementation of the valuation model in Section 5.

3.2 Listing and Meeting Probabilities

Having estimated the valuation equation and obtained estimates of $h_{bt}(n)$ and $h_{st}(n)$, we turn to the estimation of the listing and meeting probabilities. In theory, we can use maximum likelihood. The contribution of the building of seller s to the likelihood is:

$$\ell_s = \sum_{b=1}^{B} y_{b,s} \ln \pi(b,s) + (1 - \sum_{b=1}^{B} y_{b,s}) \ln \pi_n(s),$$

where $y_{b,s}$ equals one when a transaction take place and zero otherwise.

In practice, computing this likelihood is a high-dimensional problem due to the large number of buildings (and thus sellers) and buyers and computationally infeasible. We instead estimate the model using negative sampling, inspired by applications in the natural language processing literature (Mikolov et al., 2013). Intuitively, instead of using the full likelihood, we consider a classification problem. For each transaction, we sample another $K - 1$ potential buyers c that did not purchase the building from seller s, $c \in \mathcal{N}_s$, $|\mathcal{N}_s| = K - 1$. We then compute the probability that the actual buyer b purchases the building, conditional on one of the K potential buyers purchasing the property:

$$\pi_r(b,s) = \frac{\xi_{b,s}}{\xi_{b,s} + \sum_{k \in \mathcal{N}_s} \xi_{k,s}}, \tag{6}$$

where $\xi_{b,s} = \exp\left(\lambda_1 S_b + \lambda_2 \Delta S_{b,s}^{-1} + \lambda_3' \delta_{b,s} + \lambda_4 N_b\right) \pi_\tau(b,s)$.[17] We then minimize the loss function over the observed transactions, i.e., (b,s) pairs: $-\sum_s \ln \pi_r(b,s)$. This estimator is the ranking estimator in Ma and Collins (2018), who also develop the asymptotic theory. In particular, they show that the estimator is consistent for $K \geq 1$, asymptotically normal, and it converges to the maximum-likelihood estimator as $K \to \infty$.

Lastly, we estimate the listing probability, $\pi_{l,t}$. The total number of buildings that transact in a given sector-year in the data is denoted by T_t. We estimate the listing probability such that T_t equals the expected number of transactions in the model in that same sector-year:

$$\sum_s \pi_{l,t} \sum_b \pi_m(b,s)\pi_\tau(b,s) = T_t, \tag{7}$$

which implies a sector-year listing probability:

$$\pi_{l,t} = \frac{T_t}{\sum_s \sum_b \pi_m(b,s)\pi_\tau(b,s)}. \tag{8}$$

3.3 Potential Price Distribution and Counterfactuals

For several applications discussed below, we are interested in computing the distribution of prices $f(p)$ at which a trade could have occurred for an asset that is listed for sale. Once we have this distribution, we can compute from it the expected price, $\mathbb{E}[p]$, and the inter-quartile range, $IQR(p)$, a measure of price risk. Once the parameters of the valuation and matching models have been estimated, we implement the following algorithm to retrieve the potential price distribution.

For each asset n that we see transact at time t, we know the transaction price ($p_t(n)$), the characteristics of the asset $x_{n,t}$, and the valuation of the buyer ($h_{b,t}$) and the seller ($h_{b,t}$). This allows us to back out $\epsilon_{p,t} = \frac{1}{2}(\epsilon_{s,t} + \epsilon_{b,t})$. We use Bayes rule to estimate the most likely $\epsilon_{s,t}$: $\widehat{\epsilon}_{s,t} = \mathbb{E}[\epsilon_{s,t} \mid \epsilon_{p,t}]$. This gives us $v_{s,t} = h_{s,t} + \widehat{\epsilon}_{s,t}$. Next, we compute the probabilities that this seller meets any possible buyer b', including the actual buyer who materialized: $\{\pi_m(b',s)\}_{b=1}^B$ from (3). We take C random draws with replacement from the distribution of candidate buyers, with sampling weights given by the distribution $\{\pi_m(b',s)\}_{b'=1}^B$. For each potential buyer that is drawn, we draw a latent demand shock $\epsilon_{b',t} \sim N(0, \sigma_t^2)$. We construct that candidate buyer's valuation $v_{b',t} = h_{b',t} + \epsilon_{b',t}$. For each candidate buyer, we check that $v_{b',t} > v_{s,t}$. If this condition is satisfied, we record the price as $p_t(n) = \frac{1}{2}(v_{b',t} + v_{s,t})$. If the valuation of the buyer is too low, $v_{b',t} < v_{s,t}$, we set the price to missing. We then compute the mean $\mathbb{E}[p]$ and inter-quartile range $IQR(p)$ of the $C \times 1$-dimensional price distribution across the non-missing values. We refer to (the non-missing part

[17] The common denominator, $\sum_{c \neq s} \exp\left(\lambda_1 S_c + \lambda_2 \Delta S_{c,s}^{-1} + \lambda_3' \delta_{c,s} + \lambda_4 N_c\right)$, drops out, simplifying the objective function.

of) this distribution as the potential price distribution.

The potential price distribution reflects the alternative buyers the seller could have met and the prices that would have resulted from those meetings, if indeed the deal was consummated. Seeing where in the potential price distribution the actual transaction price with the actual buyer b falls, provides a gauge of how lucky the seller and/or how skilled her broker was. The potential price distribution is a useful tool in private markets with bilateral price formation for price setting strategy and risk management.

The algorithm is also how we compute counterfactuals. To understand the importance of a subset G of investors, we compute the potential price distribution removing that group from the buyer pool. The sampling of C potential buyers is now from the set $B \backslash G$. We can compute the potential price distribution and its moments for every subset of buyers that makes up the universe B. Comparing means, we see the effect of each group of buyers on average prices, and comparing IQRs provides a metric of how the buyer pool affects the price risk of an asset.

Beyond examining the price impact of excluding certain buyer types from specific markets, we are also interested in studying spillover effects. When a particular investor type purchases a property, other potential buyers who might have acquired that asset are displaced, purchasing other properties instead. By identifying these displaced buyers, we can track which properties they actually ended up buying, providing insight into the reallocation of demand. This allows us to analyze the characteristics of the properties they purchased—such as sector, location, and asset quality—and assess how their alternative transactions shaped market outcomes. Comparing these patterns across different investor types and time periods helps us understand how competition for assets in one segment influences liquidity, pricing, and allocation dynamics across the broader market.

4 Data

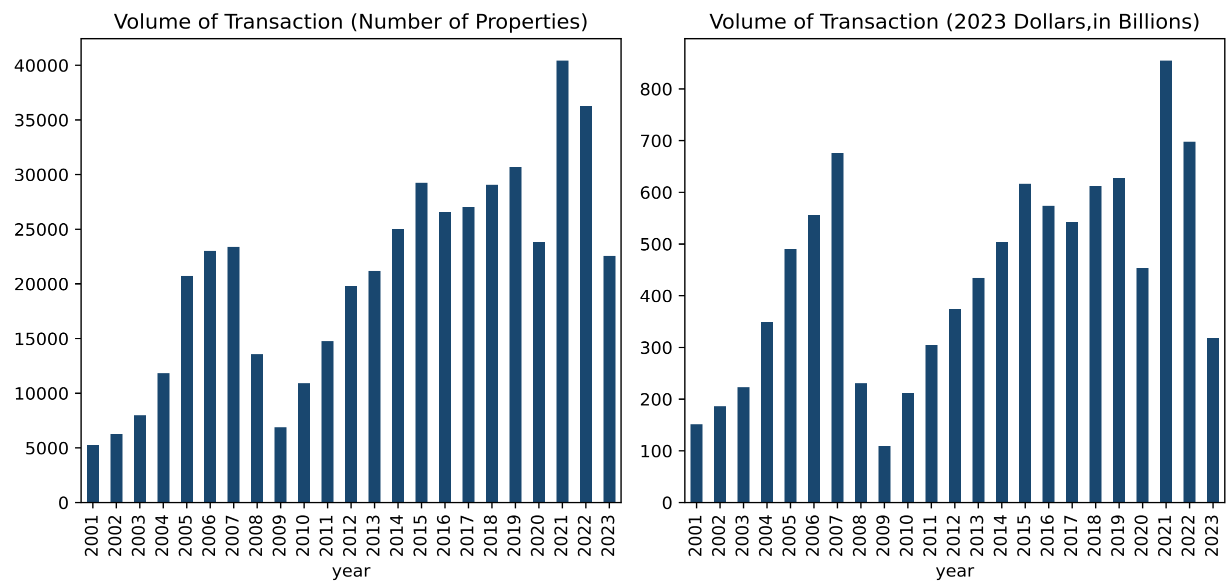
Our data consists of the universe of commercial real estate transactions over \$0.5 Million in the four main property sectors (apartments, office, retail, industrial) between 2001 and 2023, about 370,000 transactions representing \$10 trillion in transaction value. The data are from MSCI Real Capital Analytics (RCA). Its unique feature is that each transaction contains the name of the buyer(s) and seller(s), as well as an investor type classification.[18] Appendix A.1 shows our data cleaning algorithm that leads to the final data sample used in the estimation.

[18] Most property transactions take place using a property-specific LLC. The RCA team has spent years unraveling who is behind each LLC, often cross-checking each transaction using multiple sources.

4.1 Summary Statistics of Transactions

Figure 1 shows the number of transactions (left) and the dollar volume of transactions (right) in our final sample. All dollar values in the paper are expressed in 2023 real dollars. Henceforth, we use the word volume to describe the dollar value of transactions. The graph shows three boom periods, from 2001 until 2007, from 2009 until 2019, and in 2021-22. It also shows three busts, the Great Financial Crisis in 2008-09, the Covid-19 recession in 2020, and the interest rate hiking cycle in 2022-23. Appendix Table A.3 shows transaction volume by year in table format. There are additional transaction dynamics that are sector specific, such as the decline in retail due to the rise in e-commerce and the decline in office due to the rise in working-from-home in the post-2020 period (Gupta et al., 2025). Figures A.1 and A.2 show transaction volume by sector and property subtype, respectively.

Figure 1: Commercial Real Estate Transactions



Notes: The figure shows the aggregate number of transactions across all property sectors, i.e. Apartments, Office, Industrial and Retail, by year between 2001 and 2023. The left panel shows the transaction volume in terms of the number of properties traded while the right panel shows the total dollar value of properties traded, expressed in billions of 2023 U.S. dollars.

Property transactions are about evenly split between six size categories: >\$250M, \$100-250M, \$50-100M, \$25-50M, \$10-25M, < \$10M. Table A.4 shows the breakdown. Most investors typically stay within one of these size buckets.

Table A.5 reports transaction volume for the 16 largest markets, as well as all other 44 markets combined. Not surprisingly, Manhattan is the largest property market in the sample, followed by Los Angeles. There are large differences in the importance of each sector across markets. Office accounts for over 60% of volume in Manhattan, Washington DC, and San Francisco, but for only 34% of volume in Los Angeles and 24% in Dallas. About half of total transaction volume (and 40% of the number of transactions) is concentrated in the largest 16 markets.

About 75% of transactions accounting for 70% of volume are conventional sales, where one property changes hands. About 19% of transactions and 21% of volume are portfolio sales, where a group of properties is sold together. The remainder of the transactions are entity transactions, where an entire corporate entity changes hand, or distressed sales (foreclosures or debtor/trustee sales). In the case of portfolio and entity sales, RCA imputes the price of each property that is part of the transaction. We further improve on RCA's imputation as explained in Appendix A.1. Table A.6 shows the transaction breakdown by transaction type.

4.2 Summary Statistics By Investor Characteristics

Next, we turn to investor characteristics. Table A.7 shows how we map RCA's investor types to a smaller group of eight investor types. Our breakdown creates meaningful heterogeneity in terms of constraints, investment objectives, geographic reach, et cetera. Table 1 shows the transaction breakdown by investor type. About 600 Real Estate Private Equity funds (REPE) account for 10-12% of volume. About 3,400 Institutional investors, made up of investment managers, pension funds, insurance, banks and finance companies, and open-ended funds, make up about 14% of volume. The two largest investment categories are local and national Owner-Operator-Developers (OOD_L and OOD_N, respectively), accounting for about 15% and 30% of volume, respectively. These are typically smaller, more geographically-concentrated, private firms that are fully dedicated to commercial real estate. They tend to have substantial discretion in terms of investments, but are subject to tighter financial constraints. We define local OODs as those we see transact in only one market, and national OODs as those we see transact in more than one market. Individuals are made up of high-net worth buyers as well as non-traded REITs who serve high-net worth households, and account for about 3-4% of volume. REITS are publicly-listed commercial real estate vehicles and account for 12% of volume. We include real estate operating companies and listed funds in this category as well. As listed entities, REITs are subject to much stricter oversight from the regulatory and financial analyst communities, and face difficulties raising additional equity when their share price trades below net asset value. Foreign investors include sovereign wealth funds and foreign OODs, but also the foreign investors of all other investor types. They represent about 6-8% of volume.[19] Users consist of corporations, governments, non-profits, educational, and religious institutions and charities that own their own real estate. They account for about 5% of transactions. For about 1-2% of volume we do not know the investor type, and create an Unknown

[19]Table A.7 shows that there are some foreign entities among most investor types. For example, about 33% of the pension fund buys and 21% of pension fund sells are by foreign pension funds. Our foreign investor flag keeps track of this.

category.

Table 1: Transactions By Investor Type

	Buyer (#Trans)	Buyer ($ Vol)	Buyer (% Vol)	Seller (#Trans)	Seller ($ Vol)	Seller (% Vol)	Unique Investors
REPE	28,853	1241	12.30	22,058	1031	10.22	596
Institutional	38,066	1479	14.66	38,148	1371	13.59	3,435
OOD_L	147,030	1316	13.05	161,967	1656	16.41	238,140
OOD_N	150,678	3083	30.56	131,966	3081	30.54	25,385
Individual	19,453	406	4.02	19,399	285	2.82	15,811
REITS	33,518	1182	11.72	35,135	1254	12.43	389
Foreign	17,606	844	8.37	13,055	616	6.11	2,782
User	28,771	418	4.14	33,663	554	5.49	29,845
Unknown	12,044	119	1.18	20,627	241	2.39	7,802
Total	476,018	10,088	100	476,018	10,088	100	324,185

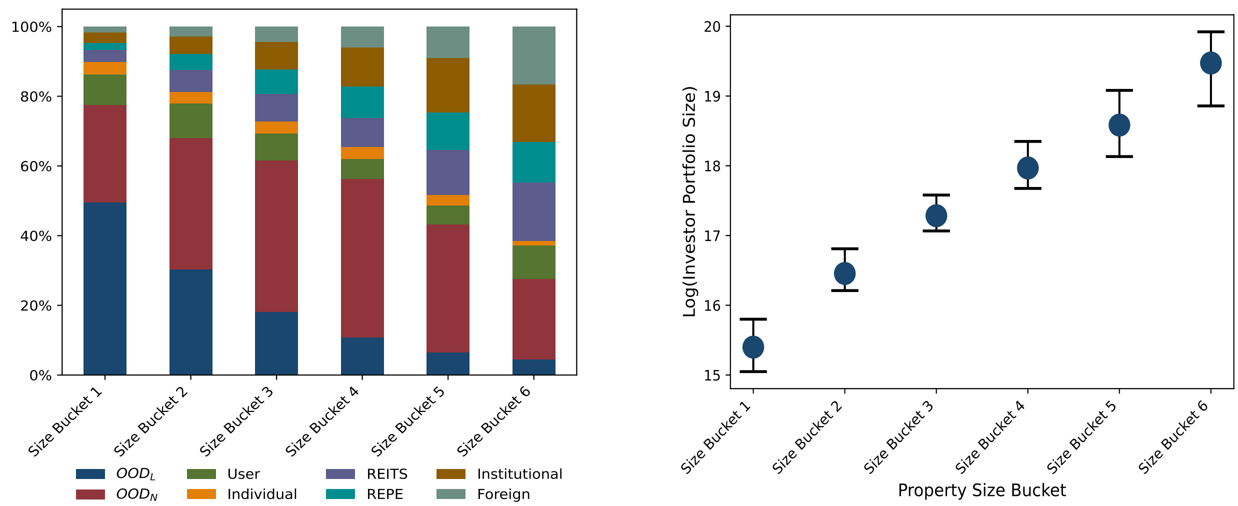
Notes: The table shows the number and dollar volume of buy and sell transactions for various investor types, i.e REPE (Real Estate Private Equity), Institutional Investors, OOD_L (Local Developers), OOD_N (National Developers), Individual investors, REITs, Foreign Investors and Users from 2001 to 2023. Volume is expressed in billions of 2023 U.S. dollars. The table also reports the the share of buy and sell volume, which sums to 100% across investor types.

Table A.8 shows that 87% of transactions by count and 73% by volume involve a single buyer and a single seller. A further 10% by transactions and 20% by volume involve one buyer and two sellers or two buyers and one seller. The remainder has multiple buyers and multiple sellers where we observe up to four buyers and up to four sellers. Table A.9 provides additional detail on these Joint Ventures, in terms of which investor types tend to collaborate with each other.

Investors are also heterogeneous in terms of their portfolio size, the size of the typical asset they trade, and their geographic, sector, and asset quality concentration. Figure A.3 shows the distribution of investor portfolio sizes at the end of 2023. The left panel of Figure 2 shows that OOD_L predominantly hold smaller properties. As property size increases, the share of OOD_L transactions diminishes, giving way to a more diverse mix of investor types in higher size buckets. In contrast, Foreign investors, Institutions, and REITs become increasingly prominent in the larger size buckets. This could indicate a preference or ability to invest in larger properties. The right panel of Figure 2 illustrates the average log of investor portfolio size by property size bucket, with error bars showing the inter-quartile range. A clear upward trend in log portfolio size is evident, suggesting that investors transacting larger buildings (by dollar value) tend to have larger portfolio sizes. This could reflect economies of scale or better access to capital for larger investors.

The left panel of Figure 3 summarizes the distribution of investor types by the number of markets the investor is active in. Investors of type OOD_L are active in a single market by definition. Many Users also only show up in a single market. REPE, REITs, and institutional investors, on the other hand, are more likely to hold geographically-diversified portfolios. The right panel shows a strong positive correlation

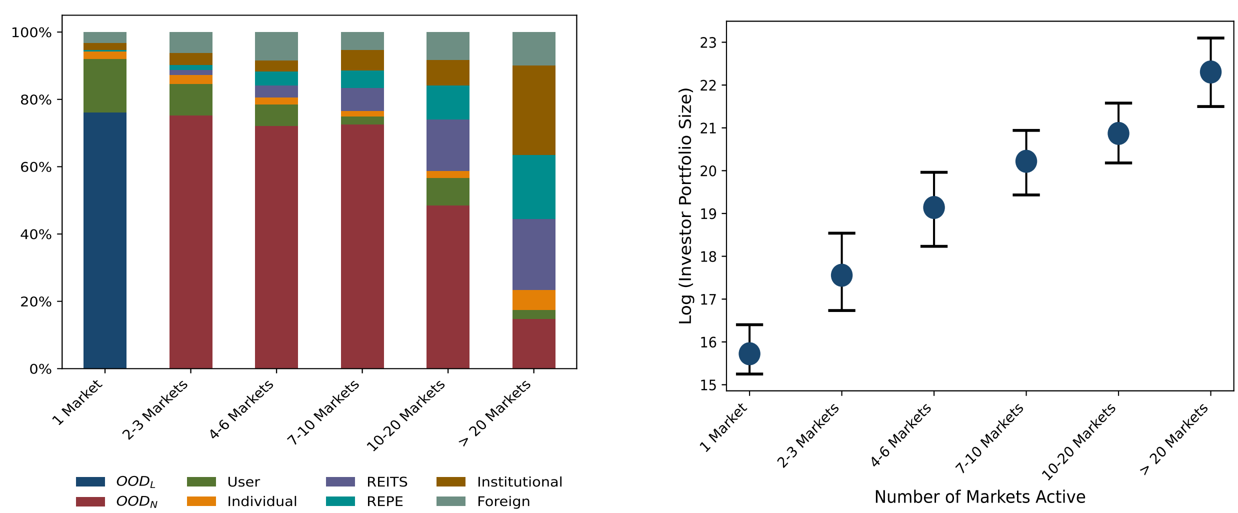
Figure 2: Size Concentration



Notes: This figure shows investor concentration by the aforementioned property size buckets at the end of 2023 for investors that own at least one property. The left figure shows the distribution of investor types by the size buckets while the right figure shows average log investor size (and inter-quartile range) of average log(investor portfolio size) for each property size bucket.

between investor size and geographic diversification. Overall, these facts suggest that there is a substantial amount of market segmentation among the investors in our sample.

Figure 3: Geographic Concentration



Notes: This figure shows investor concentration by geography at the end of 2023 for investors that own atleast one property[17]. The left figure shows the distribution of investor types by the number of aggregated RCA markets they are active in while the right figure shows average log investor size (and inter-quartile range) of average log(investor portfolio size) by the number of markets the investor is active in.

4.3 Flow Dynamics

Identifying buyers and sellers at the transaction level, and aggregating their buy and sell trades, allows us to analyze aggregate capital flows by investor type. Figure 4 shows net flows for select investor types by sector and year. It offers insights into how capital was allocated across CRE sectors over time.

Figure 4: Net Flows by Investor Types

Notes: This figure shows the net acquisitions of Real Estate Private Equity (REPE), REITs, Institutional, and Foreign investors, calculated as the dollar value of purchases minus sales (in 2023 dollars) for each sector and year. The measure captures transaction activity while excluding changes in property values.

One interesting trend is the large inflow of foreign investors from 2013 onwards. These inflows totaled \$317 billion, with a significant concentration in Office (\$123 billion, 39%) and Apartments (\$106 billion, 33%). This surge included landmark deals such as Canada's Brookfield Property Partners' \$18 billion acquisition of the REIT GGP in 2018 and its \$11.4 billion acquisition of the REIT Forest City that same year. Singapore's sovereign wealth fund GIC made a major push into the US industrial market by acquiring IndCor Properties from the REPE fund Blackstone for \$8.1 billion in 2015. Figure A.4 further reveals that most of this foreign investment in Offices and Apartments was concentrated in superstar cities, primarily Manhattan, Los Angeles, San Francisco, and Boston. These patterns provide a compelling motivation to study the price and portfolio effects of such foreign capital inflows in Section 7.

Another notable pattern is the high volatility in REIT flows, particularly in the lead-up to the 2008 financial crisis. Several REITs aggressively offloaded properties in 2006 and 2007. Notable examples include Sam Zell's $39 billion entity sale of his Equity Office Properties REIT to Blackstone in 2007 and the $22 billion purchase of Archstone-Smith REIT by Tishman Speyer and Lehman Brothers. These major transactions, occurring on the eve of the Great Financial Crisis, provide a natural setting for counterfactual analyses, helping us examine how market outcomes might have differed had these REITs not liquidated their holdings.

Finally, we document trading patterns among types in Table A.10. We see significant trade within and across investor types. Notably, a fourth of all transaction volume involves buyers and sellers of the same type, reflecting segmentation and shared investment strategies among similar investor groups. There is substantial trading activity between local and national *OOD*s. Additionally, REITs and Users exhibit a higher propensity to trade within their own investor type, indicating a degree of segmentation in trading preferences among these groups.

5 Estimation of the Valuation Model

We estimate the valuation model (5) from observed transaction prices. We impose that investors with the same features have the same valuation for an asset regardless of whether they are buyers or sellers.[20] The left-hand side variable is the log real price per square foot for sectors O, I, and R and the log price per unit for sector A.

5.1 The Hedonic Valuation Model

As a reference point, we first ignore the role of investor characteristics in valuation. We proceed in steps. We first estimate the valuation model with only asset characteristics: log asset size, log age, log renovation-adjusted age, subtype indicator, number of floors, CBD indicator,[21] superstar city indicator,[22] entity sale indicator, and distressed sale indicator.

In the second step, we add macroeconomic variables: log population for A and log employment for O, I, R, log real per capita personal income, NOI growth, and occupancy rate. These macro variables vary by market and year, and the latter two also by sector. We also add NER, which varies by neighborhood, year,

[20]Relaxing this assumption is possible but leads to parameter proliferation without much gain in explanatory power.

[21]Since the office subtypes are CBD and suburban office, the CBD indicator is colinear with the subtype category CBD, and we drop the CBD indicator from the model for Office.

[22]Superstar cities are defined as those markets with office prices above $350 per square foot in 2023. The superstar dummy is one for the following markets (for all sectors): Washington DC, DC VA burbs, DC MD burbs, Chicago, Los Angeles, Orange Co, Manhattan, NYC Boroughs, San Jose, San Francisco, East Bay, Seattle, Boston, Austin, Miami/Dade Co, and San Diego.

and sector.

In the third step, we add the year fixed effects. The fourth step adds market fixed effects. [23] [24]

This fourth model is the hedonic model against which we wish to assess the role of investor covariates. We note that this is a much richer hedonic model in terms of covariates than the traditional hedonic model. Market and time fixed effects, for example, could pick up liquidity fluctuations driven by investor composition. By including them in the hedonic model, we make it harder for investor characteristics to matter.

For reference, it is instructive to start with a hedonic valuation model that imposes that log prices are linear in characteristics, the standard assumption in the literature. Table 2 shows the results for this linear hedonic model of step 4. The many covariates, including market and time fixed effects, result in a good fit. The full-sample R^2 ranges from 46.4% for Office to 60.0% for Apartments. Nearly all property and macro characteristics come in with the predicted sign. The neighborhood quality variable NER, which we newly introduce to the hedonic literature, is positive and strongly significant. The last row of the table indicates the R^2 of a linear hedonic model that omits the NER.

Next, we allow for all hedonics, i.e., the property characteristics, market characteristics, year fixed effects, and market fixed effects to affect valuations in a nonlinear way. We also allow for all hedonics to interact with one another. We estimate a Light Gradient Boosted Machine (LGBM) model, which provides a flexible, data-driven and scalable approach compared to traditional parametric models. LGBM is well-suited for this estimation as it efficiently handles large datasets, captures complex non-linear relationships without requiring manual specification, and is robust to overfitting through built-in regularization.[25] We estimate the valuation model for each sector and model specification separately, calibrating hyper-parameters uniquely for each case. We standardize each continuous variable to facilitate comparison across the sectors.[26]

The first four rows of the top panel of Table 3 show the results. The first row shows the full-sample R^2 from the LGBM model estimated on just the property characteristics. The second row adds the macro variables including NER. The third row adds the year fixed effects, and the fourth row the market fixed effects. We refer to the model in row 4 as the hedonic LGBM model. The hedonic LGBM model produces a R^2 that ranges from 67.3% for Office to 80.5% for Apartments.

This R^2 in row 4 is directly comparable to the R^2 in Table 2. The difference between the two sets

[23] We make use of the flexibility of LightGBM to incorporate categorical variables as predictors directly. However, for the linear model, this is not possible, so we create i - 1 dummy variables to incorporate fixed effects.

[24] Market fixed effects render the superstarcities indicator redundant. Market and year fixed effects could already reflect some investor characteristics that differ by market or over time. By assigning them to the "hedonics," we are being conservative in assessing the importance of investor characteristics.

[25] The hyperparameters are optimized using 10-fold cross-validation to minimize out-of-sample mean squared error. The optimization procedure searches over key parameters, including tree depth, the number of leaves, feature and bagging fractions, and regularization terms to ensure that the model effectively balances flexibility and generalization. Details on the tuning process and parameter search space are provided in Appendix B.1.

[26] Standardizing variables is not a requirement for tree-based models such as LGBM which make splits based on the relative ordering of feature values rather than their absolute magnitudes.

Table 2: Linear Hedonic Model

	Apartment	Industrial	Office	Retail
CBD Indicator	0.153*	0.288***	0.095	0.269***
	(0.057)	(0.059)	(0.049)	(0.052)
Age	-0.075***	0.001	-0.036***	-0.006
	(0.010)	(0.007)	(0.008)	(0.006)
Renovation Adj Age	-0.032***	-0.093***	-0.081***	-0.105***
	(0.007)	(0.010)	(0.011)	(0.009)
Property Size	-0.091***	-0.269***	-0.226***	-0.373***
	(0.020)	(0.014)	(0.022)	(0.014)
Property Subtype	0.137***	0.129***		0.050*
	(0.026)	(0.020)		(0.025)
No. of Floors	0.116***	0.024	0.087***	0.055*
	(0.016)	(0.021)	(0.010)	(0.020)
Entity Sale	0.207*	0.117	0.152	0.050
	(0.090)	(0.093)	(0.093)	(0.110)
Transfer	-0.233***	-0.228***	-0.316***	-0.292***
	(0.028)	(0.032)	(0.032)	(0.040)
Market Occupancy	0.294	-0.082	0.404***	0.075
	(0.364)	(0.093)	(0.096)	(0.100)
NOI growth	0.078	0.010	0.035	-0.069
	(0.090)	(0.047)	(0.036)	(0.043)
Personal Income	0.568***	0.284***	0.352***	0.433***
	(0.072)	(0.066)	(0.054)	(0.031)
Population/Employment	0.022	0.019	-0.030*	0.038***
	(0.013)	(0.013)	(0.012)	(0.006)
NER	0.130***	0.230***	0.461***	0.164***
	(0.026)	(0.037)	(0.070)	(0.033)
Year FE	✓	✓	✓	✓
Market FE	✓	✓	✓	✓
Observations	141,135	116,737	96,139	114,223
Adj. R^2	59.94	58.46	46.35	57.96
Adj. R^2 (Excluding NER)	58.48	57.54	43.26	57.20

Notes: This table shows the full-sample coefficient estimates and goodness of fit statistics for a linear hedonic valuation model. Standard errors in bracket are clustered by RCA Market and Year.

of R^2 statistics tells us how important non-linearities and interactions are. The improvement from the added flexibility in the functional form between hedonics and valuations boosts the R^2 by 20% points for Apartments, 15% points for Industrial, 21% points for Office, and 16% points for Retail. The large improvement in fit is a first main empirical result, and is new in the context of CRE valuation.

5.2 The Main Valuation Model

We are now ready to add investor characteristics. Again, we proceed in steps. In the fifth row, we add investor type fixed effects, leaving out the indicator for the Unknown category: REPE, Institutional, OOD_L, OOD_N, Individual, REIT, Foreign, User. In the sixth and final row, we add the log portfolio value of the investor, the share of the total portfolio value that the investor has in superstar markets, in the same market as the building in question, in the same sector as the building, the joint venture dummy, and the relative size of buyer versus seller variable. These portfolio variables are computed just prior to the transaction for the sellers and just after the transaction for the buyers, so that the transacted asset is always included for both the seller and the buyer.

Table 3: LGBM Valuation Model

In-Sample performance

Sector	Apartment		Industrial		Office		Retail	
	R^2	R2C	R^2	R2C	R^2	R2C	R^2	R2C
Hedonic Model	53.36		53.39		46.99		61.93	
+ Macro Vars	73.95		68.95		64.53		70.39	
+ Year Fixed Effects	76.31		70.39		67.32		71.52	
+ Market Fixed Effects	80.45		73.76		67.25		73.98	
+ Investor Types	81.93	7.57***	79.20	20.73***	71.87	14.11***	79.39	20.79***
+ Portfolio Vars	89.76	47.62***	89.89	61.47***	87.20	60.92***	90.70	64.26***

Out-of-Sample performance

Sector	Apartment		Industrial		Office		Retail	
	R^2	R2C	R^2	R2C	R^2	R2C	R^2	R2C
Hedonic Model	49.06		50.07		44.53		59.14	
+ Macro Vars	68.97		63.74		58.25		65.66	
+ Year Fixed Effects	70.39		64.51		58.98		66.49	
+ Market Fixed Effects	72.72		66.34		59.97		67.62	
+ Investor Types	74.08	4.99***	69.77	10.19***	63.55	8.94***	70.98	10.38***
+ Portfolio Vars	81.98	33.94***	82.64	48.43***	78.34	45.89***	83.98	50.53***

Notes: This table summarizes the performance R^2 from valuation model estimations in equation (5) using LightGBM. The two panels summarize the performances of our non-linear valuation model by successively adding variables as described in 5. Each row chooses the optimal LightGBM hyperparameters using cross validation. The top panel shows in-sample R^2, and the bottom panel shows out-of-sample R^2.

We compute the comparative R-squared (R2C) for the models in rows 5 and 6, relative to the model without investor characteristics in row 4. The R2C is the ratio of the difference in R-squared values between the alternative and null models to the unexplained variation in the dependent variable by the null model:

$$R2C = \frac{R^2_{alt} - R^2_{null}}{1 - R^2_{null}}$$

Kim (2024) develops the asymptotic distribution of R2C and a test of the null hypothesis that R2C equals zero. The stars on the R2C statistic reflect the p-values of this test.

Row 5 of the top panel of Table 3 shows that the investor categories provide useful additional information for valuations. Relative to row 4, the R^2 goes up between 2 and 6% points, reducing the unexplained variation in model 4 by between 8 and 21% (R2C).

The gains from adding the other investor characteristics in row 6 are substantially larger. This model in row 6 is our main valuation model. The R^2 reach values between 87% points for Office and 90% points for the other sectors. These are high R^2 indeed. Comparing row 6 to row 4, all investor variables combined increase the R^2 by 9.3% points (A), 16.1% points (I), 20.0% points (O), and 16.7% points (R). They reduce the unexplained variation in model 4 by 47.6% for A, 61.5% for I, 60.9% for O, and 64.3% for R. These R2C gains are highly statistically significant in all four sectors and economically large.

In sum, investor characteristics are a crucial "new hedonic" in commercial real estate valuation models that strongly add to the valuation model's explanatory power. When different investors value asset characteristics differently, it matters for the transaction price which investors are in the market at a given point in time.

The second panel of Table 3 shows the out-of-sample (OOS) performance of the LGBM model. For the latter, we use a 10-fold cross-validation procedure where the R^2 is calculated as the average of the hold-out samples across all folds. Naturally, the main model in row 6 has a lower OOS than in-sample R^2. However, both the absolute level of the R^2 and the gain in explanatory power relative to the hedonic model in row 4 remain large. Investor characteristics explain an additional 9.3% points (A), 16.3% points (I), 18.4% points (O), and 16.4% points (R) of the variation in property transaction prices, reducing unexplained variation between 33.9% and 50.5%. This gain remains highly statistically significant in all four sectors.

The valuation model also allows us to compute the residual standard deviation for buyers and sellers in each year. We allow for time variation but impose: $\sigma^2_{b,t} = \sigma^2_{s,t}$.

5.3 Variable Importance Analysis

The previous analysis showed that (i) investor characteristics matter and (ii) that allowing for non-linearities and interaction effects matters. Our main model allows for traditional hedonics to be valued differently by different types or sizes of investors, for example. We systematically investigate the importance of each feature using SHAP values (Lundberg and Lee, 2017).[27]

Figure 5 highlights the most important features selected by the model, organized by category. Investor portfolio characteristics are the most important feature category, ahead of property characteristics and

[27] SHAP (Shapley Additive Explanations) values provide a game-theoretic approach to interpreting machine learning models by quantifying the marginal contribution of each predictor to the model's predictions. For a given observation, SHAP values measure how much each feature pushes the model's output away from the mean prediction. Mathematically, for a prediction function $f(x)$ and a set of features N, the SHAP value for feature j is calculated as:

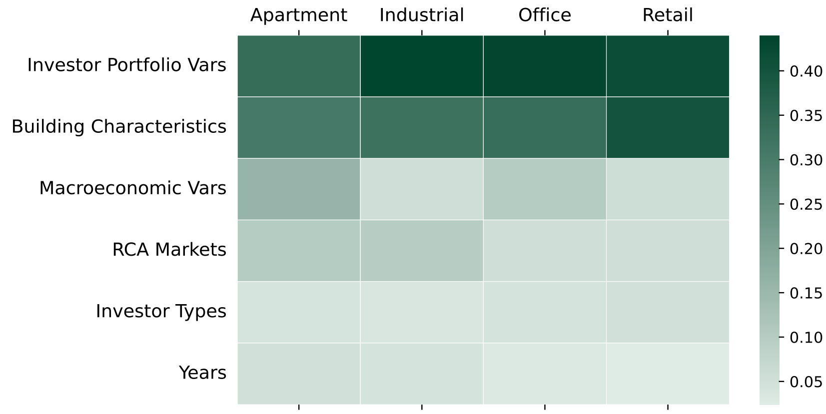
$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(M - |S| - 1)!}{M!} (f(S \cup \{j\}) - f(S)),$$

where M is the total number of features and $f(S)$ represents the model output when only the features in subset S are known. Feature Importance (I_j) is obtained by averaging its absolute SHAP value over all n observations.

$$I_j = \frac{1}{n} \sum_{i=1}^{n} \left| \phi_j^{(i)} \right|$$

SHAP values provide a more comprehensive importance measure by accounting for both direct and interaction effects, unlike traditional importance measures such as split frequency.

Figure 5: Valuation Model Feature Importance by Category



Notes: Figure shows the importance of features selected by the LightGBM Model. Feature importance is defined as the absolute SHAP (Shapley Additive Explanations) values for each feature summed across all observations. We aggregate the importance by feature types and normalize them such that the importance for each feature group sums to 1

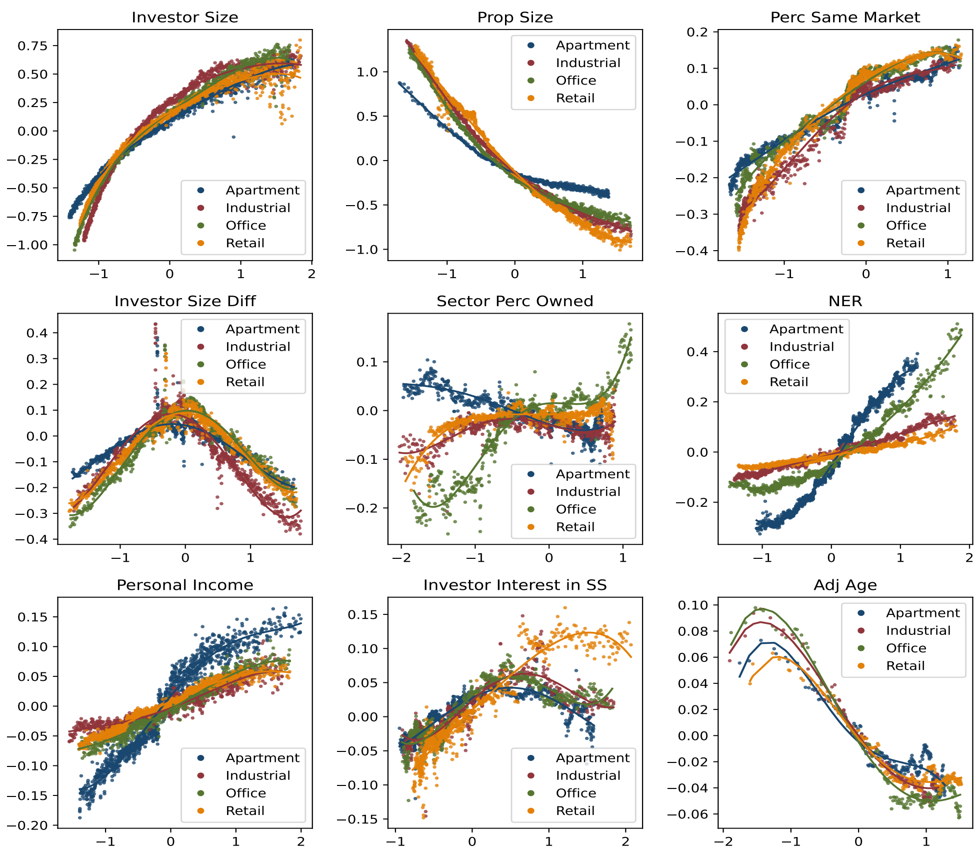
macro variables. Appendix Figure B.1 breaks out the features further. Among all variables, investor size and property size are the two most important features in all sectors. The log ratio of buyer-to-seller sizes, the share of properties the investor has in the same market, and the investor type are all among the ten most important features. NER is the most important macro variable and market fixed effects are important as well.

To analyze how transaction prices vary with predictors, we use SHAP dependence plots. For each feature, we compute the average SHAP value at different percentiles of the feature (from the 5th to the 95th percentile) and plot these against the corresponding feature values using a binned scatter plot. The X-axis represents the standardized feature value, while the Y-axis represents the standardized SHAP value averaged across the buyer and seller valuation models, ensuring comparability across sectors. Positive (negative) SHAP values indicate that a higher feature value increases (lowers) the price. Figure 6 shows the results for the nine most important continuous features, reported in the title of each panel.

Larger investors have larger valuations but the effects are concave in investor size. When buyers and sellers have similar size, the impact on valuations is small, but when one party is much larger than the other in either direction, that lowers valuations. Investors who specialize in the office sector have higher valuations for office, while sector focus matters much less for valuations in other sectors. A high NER (expensive neighborhood with high-quality assets) affects office and apartment valuations disproportionately. The graph illustrates that many of the effects are nonlinear in ways that would be difficult to uncover without the help of a machine learning model like LGBM.

To study time trends in valuations, we plot the SHAP values for the year fixed effects learned by the

Figure 6: SHAP Dependence: Non-linearities



Notes: This figure shows SHAP dependence plots for the nine most important continuous features in our valuation models. Each plot is a binned scatter plot with a best-fit polynomial trend line. The X-axis represents the standardized feature value, while the Y-axis represents the standardized SHAP value, averaged across the buyer and seller valuation models, to allow comparability across sectors.

valuation model in Figure B.2. Valuations declined through the 2008-recession and recovered steadily post 2012 with industrial and apartments showing the strongest recovery. Retail valuations begin falling post 2016—coinciding with the retail apocalypse and rise of E-Commerce, while industrial valuations continued to rise sharply.

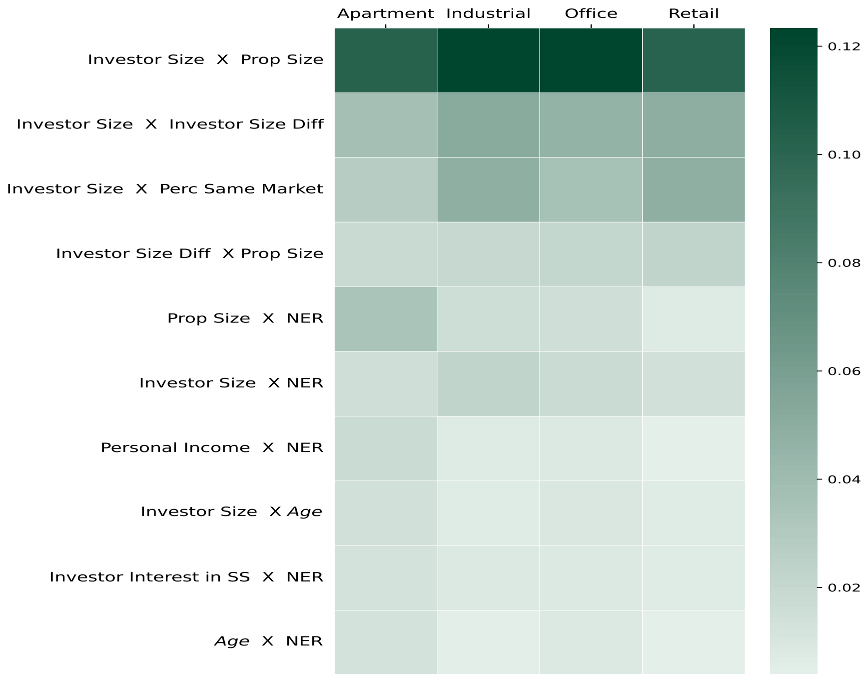
SHAP values capture both direct and interaction effects, making them ideal for analyzing feature interactions. Beyond attributing predictions to individual features, SHAP can be extended to decompose contributions into pairwise interactions. Following (Lundberg et al., 2020), we compute SHAP interaction values for all pairwise interactions.[28]

[28] The SHAP interaction values quantify how a feature pair jointly influences predictions beyond their individual effects. For a pair of features i and j these values are defined as:

$$\phi_{ij} = \sum_{S \subseteq N \setminus \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M-1)!} \left[f(S \cup \{i,j\}) - f(S \cup \{i\}) - f(S \cup \{j\}) + f(S) \right]$$

where the term inside the square brackets captures the marginal effect of including both features i and j together compared to including them individually.

Figure 7: Top Ten Interaction effects



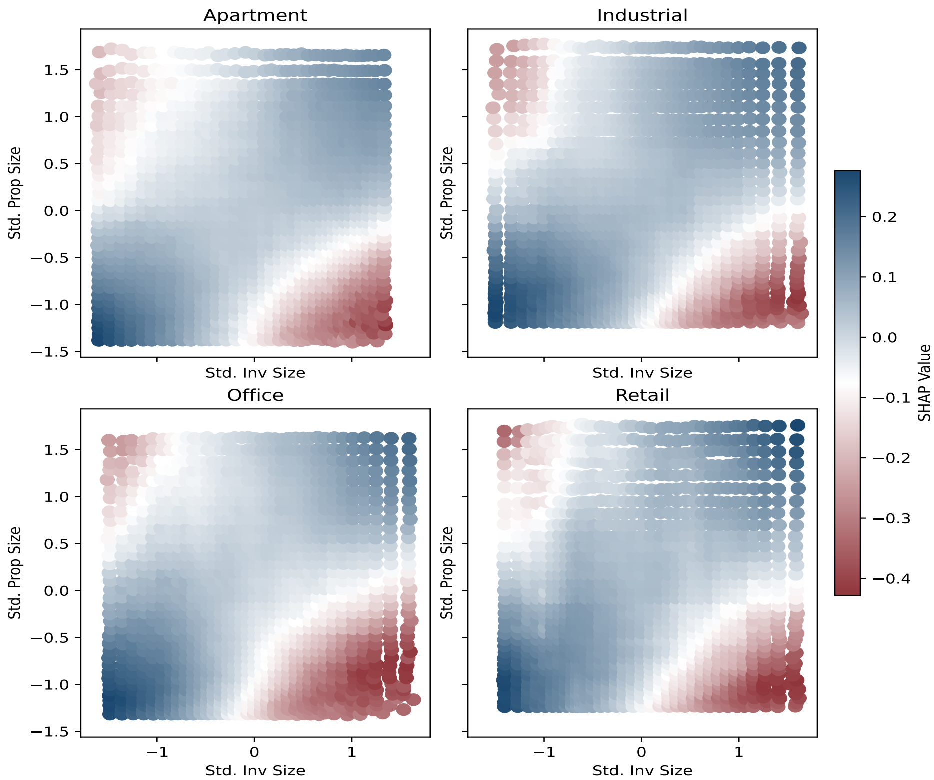
Notes: This figure presents the ten most important interaction effects, measured by averaging the absolute SHAP interaction values across all observations for each feature pair within a sector, averaged across buyer and seller valuation models. To facilitate comparability, values are normalized so that the sum across all feature pairs equals 1 for each sector.

Figure 7 shows the 10 most important interaction effects for each sector. Interactions between investor characteristics, notably investor portfolio size, and building features such as property size, NER, and building age are among the most influential. There are also several interactions of two investor characteristics in the top ten. This result reinforces the critical role of investor characteristics in the valuation model, as well as an important mechanism: different investors value different hedonics differentially so it matters which investors are in the market at a given point in time.

Figure 8 studies the interaction effects between investor size and property size, the most important interaction effect identified in Figure 7. For the pair, we compute the average interaction SHAP value at different percentiles of the feature (from the 5th to the 95th percentile) and plot these against the corresponding feature values using a binned scatter plot. The X-axis represents the standardized investor size, while the Y-axis represents the standardized property size. The color intensity reflects the magnitude and direction of the interaction SHAP values, where deep red indicates a negative interaction effect and blue represents a positive interaction effect.[29] A key observation is that large investors have lower valuations for small properties and small investors have dampened valuations for large properties in every sector.

[29]Unlike the main SHAP values, which include both direct and interaction effects, this plot isolates only the interaction component, which measures how the effect of one variable changes depending on the level of the other variable.

Figure 8: SHAP Interaction Effects



Notes: This figure shows a binscatter plot to understand the interaction effects between investor size and property size, capturing how their combined influence deviates from the sum of their individual contributions. The color intensity reflects the magnitude and direction of the interaction SHAP values, where deep red indicates a negative interaction effect, while blue represents a positive interaction effect.

To further examine interaction effects, we use SHAP values at the individual observation level to assess how the marginal effect of one feature varies across different levels of another. Since SHAP values decompose model predictions into contributions from each feature, we can analyze how the effect of a given feature depends on another feature. Appendix Figure B.3 analyzes how the impact of investor size on valuations depends on other features such as asset size, investor type, or market. For example, the higher valuation that large investors have for office and retail properties is most pronounced for the largest assets, less pronounced for REITS, and is much stronger in superstar cities than in tertiary markets.

Figure B.4 explores interactions of NER with investor characteristics. Higher neighborhood quality (NER) is associated with higher valuations on average. The quality preference in office and retail markets is particularly strong for foreign, large, sector-focused, and geographically-focused investors.

6 Estimation of the Listing and Meeting Model

6.1 Point Estimates in the Meeting Model

Next, we estimate the parameters of the meeting model in (3). The variables that determine whether a potential buyer b meets a building that has been listed for sale by a seller s are: the buyer's log portfolio size, the log difference between the buyer and the seller's portfolio sizes, four consideration set variables in $\delta_{b,s}$ which measure how similar the building for sale is to all the assets in the buyer's portfolio (in terms of property size, sector, location, and neighborhood quality), and an indicator N_b, which is one if the buyer owns more than two buildings. We estimate one model for each sector-year using all transactions from that year with maximum likelihood. Table 4 reports the coefficient point estimates $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ that minimize the loss function $-\sum_s \ln \pi_r(b, s)$, with π_r as defined in (6), and their standard errors. The parameter estimates $\hat{\lambda}_t$ are averaged over time $\hat{\lambda} = (T)^{-1} \sum_{t=1}^{T} \hat{\lambda}_t$. The standard errors for $\hat{\lambda}_t$ are obtained using bootstrap and then also averaged over time.

With the coefficient estimates for the valuation model in hand, we can compute the private valuation gap between buyers and sellers $\Delta^h_{b,s}$ and obtain π_τ, the probability whether a transaction takes place conditional on a meeting. We estimate the models for each sector separately on an annual basis, using all transactions from that year.

Since most investors in our sample transact infrequently, including (a random subset of cardinality K of) all investors in the negative sample would insert a substantial number of inactive investors, adding substantial noise to the estimation. Therefore, we restrict the negative sample to those investors who have transacted at least once in the preceding five years.[30] This still leaves open the possibility that some potential buyers arrive in the transaction year with zero assets. In those cases, we assume that they own median number of properties (which is 1) with characteristics equal to the average portfolio in that year.

Since the asset listed for sale is included in the portfolio of the actual buyer, we also add it to the portfolio of all other potential buyers to avoid a bias. Otherwise, it would be artificially easy for the algorithm to select the true buyer from the set of all potential buyers.

As explained above, rather than including all potential buyers in the likelihood estimation, we use an important insight from the NLP literature showing it is sufficient to consider a negative set with small cardinality K. We explore values of K between 50 and 1,500 and choose $K = 1,000$ for our main results, since parameter estimates stabilize around that value. We discuss robustness to values for K in Appendix B.3.

The first column of Table 4 shows that the size of the buyer has a large positive effect on the meeting

[30] We explore robustness to this assumption in Appendix B.4.

Table 4: Meeting Model Parameter Estimates

	λ_1	λ_2	$\lambda_{3,1}$	$\lambda_{3,2}$	$\lambda_{3,3}$	$\lambda_{3,4}$	λ_4
Apartment	1.55	2.69	8.27	5.58	4.12	7.31	2.21
	(0.03)	(0.17)	(0.25)	(0.14)	(0.15)	(0.33)	(0.08)
Industrial	1.66	3.01	8.53	5.83	4.02	9.71	2.13
	(0.04)	(0.19)	(0.32)	(0.17)	(0.15)	(0.50)	(0.10)
Office	1.58	2.76	8.13	5.66	3.30	10.35	2.37
	(0.04)	(0.19)	(0.3)	(0.17)	(0.16)	(0.54)	(0.09)
Retail	1.54	2.76	8.22	5.54	3.85	7.6	2.19
	(0.04)	(0.18)	(0.29)	(0.17)	(0.14)	(0.38)	(0.09)

Notes: This table reports the average estimated values of the meeting model parameters along with their standard errors in bracket. The point estimates $\hat{\lambda}_t$ are obtained by maximum likelihood for each year, and averaged across years. Standard errors are calculated using bootstrap (with 1,000 samples) for each year, and averaged across years. The maximum likelihood estimation minimizes the loss function in (6). λ_1 and λ_2 are coefficients of investor size and buyer-seller size difference respectively. $\lambda_{3,1}$ through $\lambda_{3,4}$ are the coefficients of asset, market, sector and NER alignment an λ_4 is the coefficient of the indicator that equals one when the buyer owns more than 2 buildings.

probability. A potential buyer which is 1% larger than another has between 1.54% (R) and 1.66% (I) higher likelihood of meeting the seller. This parameter is identified from the fact that large investors, for example large REITS or REPE funds, are engaged in many more transactions than small investors.

The positive an significant λ_2 estimate in the second column shows that buyers are more likely to meet sellers with a similar portfolio size. This is consistent with the positive assortative matching we showed in the right panel of Figure 2 between investor size and asset size.

The consideration set variables all have the expected positive sign, and are large and significant (the vector λ_3 in columns 3–6). Buyers are much more likely to meet assets that have similar size, location, sector, and neighborhood quality (NER) to the assets they already own. Property size alignment ($\lambda_{3,1}$) and asset quality alignment ($\lambda_{3,4}$) are the most important similarity measures. To understand the coefficient magnitudes, a potential buyers of apartments that is 1% more invested in the same market is 5.6% more likely to meet the seller. A 1% increase in NER alignment increases the odds of a buyer-seller meeting by a factor of about 7.3% for Apartments and 10.3% for Offices.

Finally, λ_4 is positive and significant, indicating that owning more than two assets makes a buyer about 2.2% points more likely to meet the property for sale.

6.2 Variable Importance in the Meeting Model

To better understand the importance of each of the variables in the meeting model, Table 5 successively leaves out each of the variables, reporting the resulting coefficient estimates and log likelihood. The first row of each panel is the baseline model, repeated for convenience. The eighth column shows $1 - \pi_{m_0}\prime/\pi_{m_0}$,

Table 5: Meeting Model Analysis

Apartments

	λ_1	λ_2	$\lambda_{3,1}$	$\lambda_{3,2}$	$\lambda_{3,3}$	$\lambda_{3,4}$	λ_4	1 - $\pi_{m_0\prime}/\pi_{m_0}$	Avg LL
All λ	1.55	2.69	8.27	5.58	4.12	7.31	2.21		-3.35
Excluding λ_1		2.52	5.31	2.96	4.02	6.8	3.62	61.6%	-4.37
Excluding λ_2	1.48		8.88	5.47	4.04	7.43	2.06	18.7%	-3.49
Excluding $\lambda_{3,1}$	1.09	3.92		5.08	4.15	7.85	2.26	44.6%	-3.99
Excluding $\lambda_{3,2}$	1.00	2.44	7.31		4.61	8.6	1.25	61.9%	-4.09
Excluding $\lambda_{3,3}$	1.36	2.60	7.87	6.04		7.76	2.10	37.7%	-3.85
Excluding $\lambda_{3,4}$	1.54	2.87	8.58	6.05	4.32		1.92	22.7%	-3.63
Excluding λ_4	1.88	2.52	8.85	5.19	4.37	6.67		16.2%	-3.58

Industrial

	λ_1	λ_2	$\lambda_{3,1}$	$\lambda_{3,2}$	$\lambda_{3,3}$	$\lambda_{3,4}$	λ_4	1 - $\pi_{m_0\prime}/\pi_{m_0}$	Avg LL
All λ	1.66	3.01	8.53	5.83	4.02	9.71	2.13		-3.07
Excluding λ_1		3.24	3.0	2.81	3.71	9.32	3.45	68.0%	-4.45
Excluding λ_2	1.62		9.21	5.75	3.98	9.97	1.84	27.8%	-3.26
Excluding $\lambda_{3,1}$	1.16	4.06		5.46	4.46	10.08	2.08	36.9%	-3.61
Excluding $\lambda_{3,2}$	1.24	3.05	8.03		4.67	12.22	0.92	50.5%	-3.73
Excluding $\lambda_{3,3}$	1.63	3.05	8.86	6.31		10.37	1.32	33.4%	-3.52
Excluding $\lambda_{3,4}$	1.66	3.31	8.81	6.59	4.32		1.88	26.2%	-3.41
Excluding λ_4	1.91	2.79	8.93	5.37	3.7	9.42		11.5%	-3.24

Office

	λ_1	λ_2	$\lambda_{3,1}$	$\lambda_{3,2}$	$\lambda_{3,3}$	$\lambda_{3,4}$	λ_4	1 - $\pi_{m_0\prime}/\pi_{m_0}$	Avg LL
All	1.58	2.76	8.13	5.66	3.30	10.35	2.37		-3.57
Excluding λ_1		2.89	4.27	2.95	2.17	10.50	3.46	63.4%	-4.66
Excluding λ_2	1.53		8.81	5.56	3.3	10.54	2.20	24.9%	-3.73
Excluding $\lambda_{3,1}$	1.06	3.86		5.17	2.86	11.42	2.16	46.6%	-4.17
Excluding $\lambda_{3,2}$	1.09	2.62	7.23		3.68	12.43	1.28	58.3%	-4.25
Excluding $\lambda_{3,3}$	1.41	2.78	7.66	5.99		10.98	1.81	31.2%	-3.85
Excluding $\lambda_{3,4}$	1.58	2.97	8.65	6.28	3.59		2.15	31.3%	-3.93
Excluding λ_4	1.85	2.63	8.33	5.18	2.81	10.04		19.4%	-3.82

Retail

	λ_1	λ_2	$\lambda_{3,1}$	$\lambda_{3,2}$	$\lambda_{3,3}$	$\lambda_{3,4}$	λ_4	1 - $\pi_{m_0\prime}/\pi_{m_0}$	Avg LL
All	1.54	2.76	8.22	5.54	3.85	7.6	2.19		-3.25
Excluding λ_1		3.12	3.66	2.71	3.74	6.98	3.43	66.4%	-4.40
Excluding λ_2	1.5		8.76	5.47	3.79	7.74	1.98	24.8%	-3.41
Excluding $\lambda_{3,1}$	1.05	3.53		5.14	4.29	7.94	2.07	42.5%	-3.79
Excluding $\lambda_{3,2}$	1.12	2.74	7.72		4.29	9.48	1.01	45.8%	-3.85
Excluding $\lambda_{3,3}$	1.49	2.65	8.88	5.75		7.82	1.51	40.1%	-3.73
Excluding $\lambda_{3,4}$	1.52	3.01	8.5	6.18	4.05		1.94	23.8%	-3.55
Excluding λ_4	1.79	2.61	8.55	4.93	3.54	7.23		11.0%	-3.45

Notes: Figure shows the point estimates of the meeting model for each of the four sectors. The last column reports the log-likelihood. The first row in each panel is the baseline model. The successive rows leave out one of the variables in the meeting model, each of which is a special case of the baseline model.

the percentage reduction in the probability that the true buyer meets the seller when we exclude a specific variable from the estimation.

We find that buyer size is far and away the most important determinant of the meeting probability.

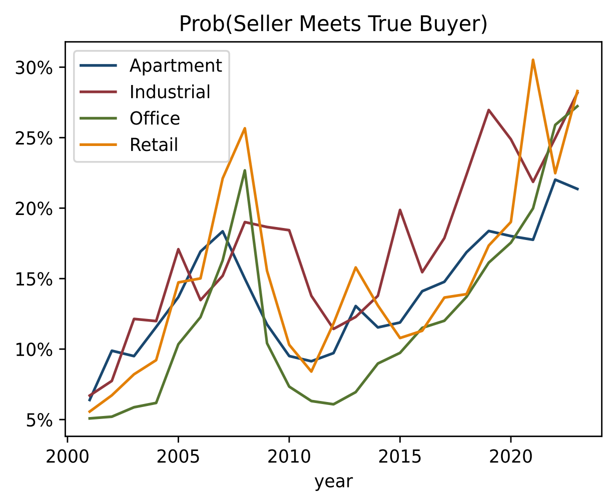
Excluding it from the estimation results in a more than 60% drop in the probability that the algorithm identifies the correct buyer in every sector. The likelihood of the restricted model is 100-130 log points lower than that of the baseline model across all four sectors. Excluding buyer size leads to large changes (biases) in the remaining coefficient estimates.

Size similarity between buyers and sellers is relatively less important. Excluding it lowers the probability of identifying the correct buyer by about 18–28%.

The consideration set variables are all quantitatively important, with some variation across sectors. The most important one in every sector is asset location similarity ($\lambda_{3,2}$). Excluding it lowers the likelihood by 60–75% and the probability of identifying the correct buyer by 45–62%. Asset size similarity is the second most important consideration set variable in every section, with a 37-47% reduction in likelihood from excluding it.

Finally, owning at least two properties increases the probability of identifying the right buyer by 11-19%.

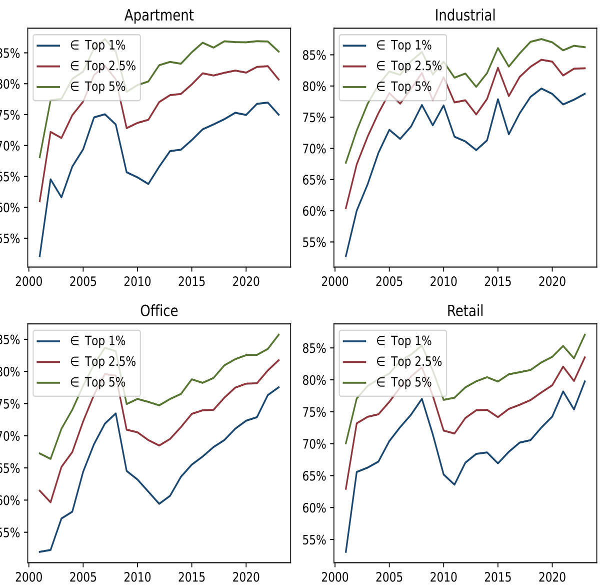
Figure 9: Meeting Probabilities



Notes: This figure shows the probability of meeting the actual buyer from among a set of $K = 1,000$ potential buyers (negative set) and the actual buyer, as implied by the benchmark model estimation.

Figure 9 displays the likelihood that the model identifies the true buyer from among the set of 1,000 potential buyers and the true buyer. That likelihood is consistently high, averaging over 20%. This underscores the model's ability to capture the key determinants of buyer-seller matches and highlights the model's accuracy in capturing the directed search behavior inherent in CRE markets. For comparison, random search would result in a 0.1% probability of identifying the true buyer. Directed search therefore delivers a 200x gain over random search. A drop in meeting probabilities during the financial crisis is evident across all sectors. This may reflect the heightened uncertainty and reduced liquidity in the market, disrupting the typical alignment between buyer preferences and assets available for sale. The graph also shows that the model's ability to identify the correct buyer rises from about 10% at the start of the sample to nearly 30% at the end of the sample. While the estimation is done year-by-year, the number of annual transactions in the data set is growing over time. Also, with every passing year of additional transaction data, we obtain a clearer picture of the investors' portfolios as holdings have to be built from past transactions.

Figure 10: Meeting Probabilities: True Buyer in Top Predicted Matches



Notes: This figure reports the probability that the true buyer ranks among the top 1%, 2.5%, or 5% of the 1,000 negative sample buyers—i.e., among the top 10, 25, or 50 buyers respectively—based on model-implied meeting probabilities.

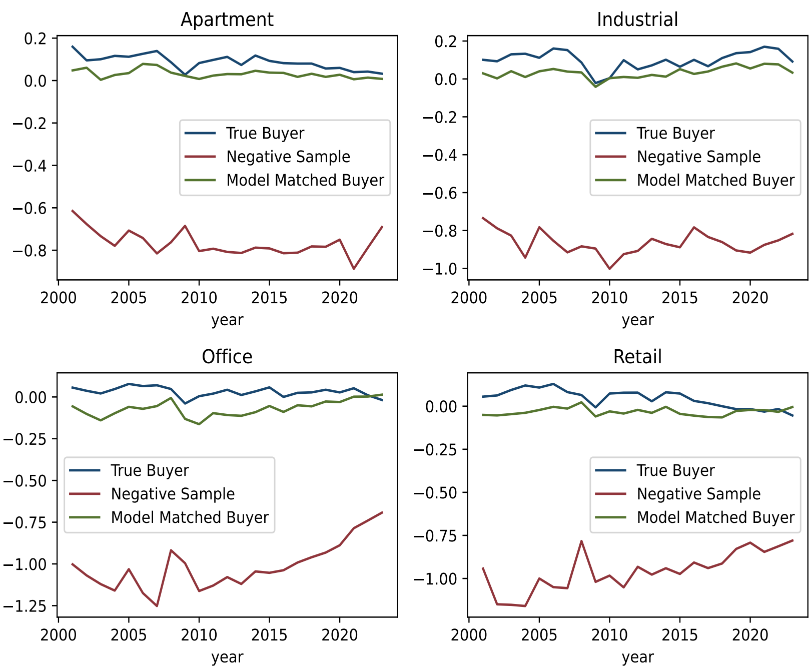
As shown in figure 9, the meeting model correctly identifies the true buyer in over 20% cases. However, this benchmark is stringent: in a pool of 1,000 potential buyers, there are often several buyers with characteristics similar to the true buyer, making exact identification difficult. What matters in practice is whether the model ranks the true buyer near the top. Figure 10 shows that the true buyer appears among the top 1%, 2.5%, or 5% of predicted matches with high probability ($\sim$ 70% - 80%), highlighting the model's ability to focus the search on a small and relevant subset of potential buyers.

6.3 Transaction Probability

Having estimated the probability of a meeting, we next turn to the probability that a transaction takes place conditional on a meeting, π_τ. The latter depends on the valuation gap between the buyer and the seller, and hence combine the estimates from the meeting and valuation models.

Figure 11 plots the valuation gap $\Delta_{b,s}$ for the positive set (the actual buyer) and the negative set (the average among the K potential buyers). Across all sectors, the true buyers exhibit consistently higher valuation gaps than the other potential buyers. This reflects not only the model's ability to select the correct buyer, as noted above, but also reinforces the model's assumption that matches occur when buyers perceive gains from trade. True buyers have about a 10–20% higher valuation for the asset than the seller. For the negative sample (the non-buyers), the valuation gap is large and negative (around -70% to -100% on a log scale).

Figure 11: Valuation Gap

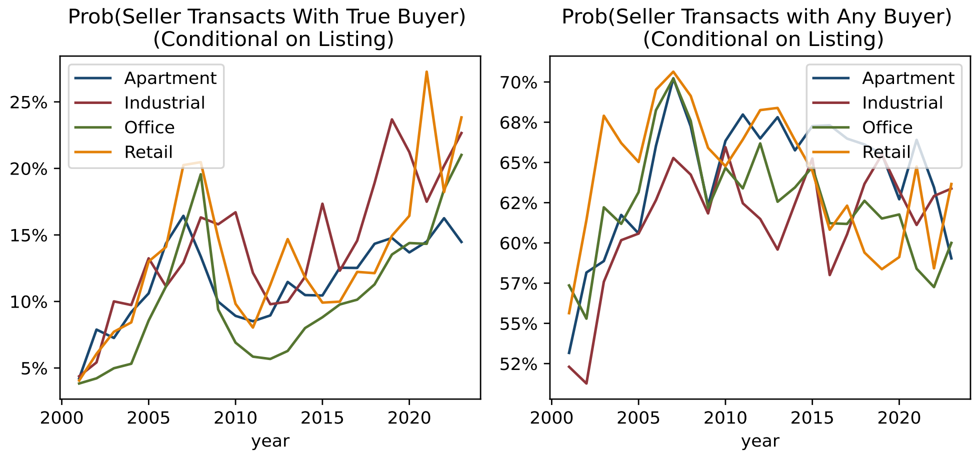


Notes: This figure shows the valuation gaps, i.e. difference between the log valuation of the buyer and the log valuation of the seller for the positive sample (the actual buyer) and the negative sample (the set of K = 1,000 potential buyers)

The left panel of Figure 12 shows the model-implied probability that the actual buyer transacts with the seller. This probability is consistently high, reflecting the model's ability to accurately capture directed search (identify the correct buyer) as well as the relative valuations of the buyer and the seller for the asset (the valuation model delivers a positive probability of sale conditional on a meeting).

The right panel shows the model-implied probability of a transaction occurring with any buyer. These probabilities are 60-70% after the first few years. They show that the model has a high likelihood of predicting a trade for assets that actually traded.

Figure 12: Transaction Probabilities



Notes: This figure shows the model-implied probabilities of transactions. The left panel displays the probability that the seller transacts with the actual buyer, combining the meeting and transaction probabilities. The right panel shows the probability that the seller transacts with any buyer, reflecting aggregate transaction likelihoods across all potential buyers.

6.4 Listing Probability

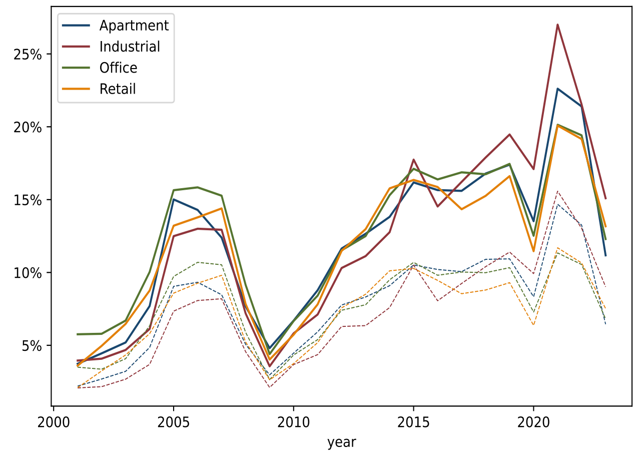
Finally, Figure 13 shows the listing probability, defined in equation (8). The listing probability reconciles the model-implied transaction probabilities with the observed transaction volumes in the data for each sector-year. For all sectors, the estimation results in a boom-bust pattern in listing probabilities in the 2001–2009 period and a rise in listing probabilities in the 2010–2019 period. Listing probabilities dip during the Covid year 2020 and again in 2022-23 when the Fed embarks on its interest rate hiking cycle. The model generates realistic average listing probabilities of around 15-20% per year, as inferred from actual transaction probabilities that around 7-10% per year (dotted lines).[31]

7 Applications: Prediction and Counterfactuals

Having estimated the valuation and matching models, we now turn to three applications of the model. In Section 7.1, we use the model can be used to predict the transaction price of a building, out-of-sample. In Section 7.2, we show how the model can be used to understand the dynamics of CRE prices, focusing in particular on the period leading up to the 2008-09 financial crisis. We conclude in Section 7.3 with an application where we use the model to uncover investors' substitution patterns across building size, sector, and location. This analysis is an important input to understanding the impact on the CRE ecosystem of, for instance, an inflow of foreign buyers in the high-quality segment of the Manhattan office market. All

[31] In future versions, we plan to further micro-found the listing probabilities based on lagged transaction volumes and motivated by reference dependence and loss aversion (Genesove and Mayer, 2001; Andersen et al., 2022).

Figure 13: Listing Probabilities



Notes: This figure shows the listing probabilities π_t for the benchmark model over time across all sectors (Apartment, Industrial, Office, and Retail), as defined in (8). Dotted lines, show for each sector the fraction of available properties that did transact in that year. The gap between the two lines being the likelihood that the seller listed the property, but failed to find a suitable buyer.

three applications make use of the potential price distribution concept explained in Section 3.3, which in turn relies on both the valuation model and the transaction model.

7.1 Predicting transaction prices

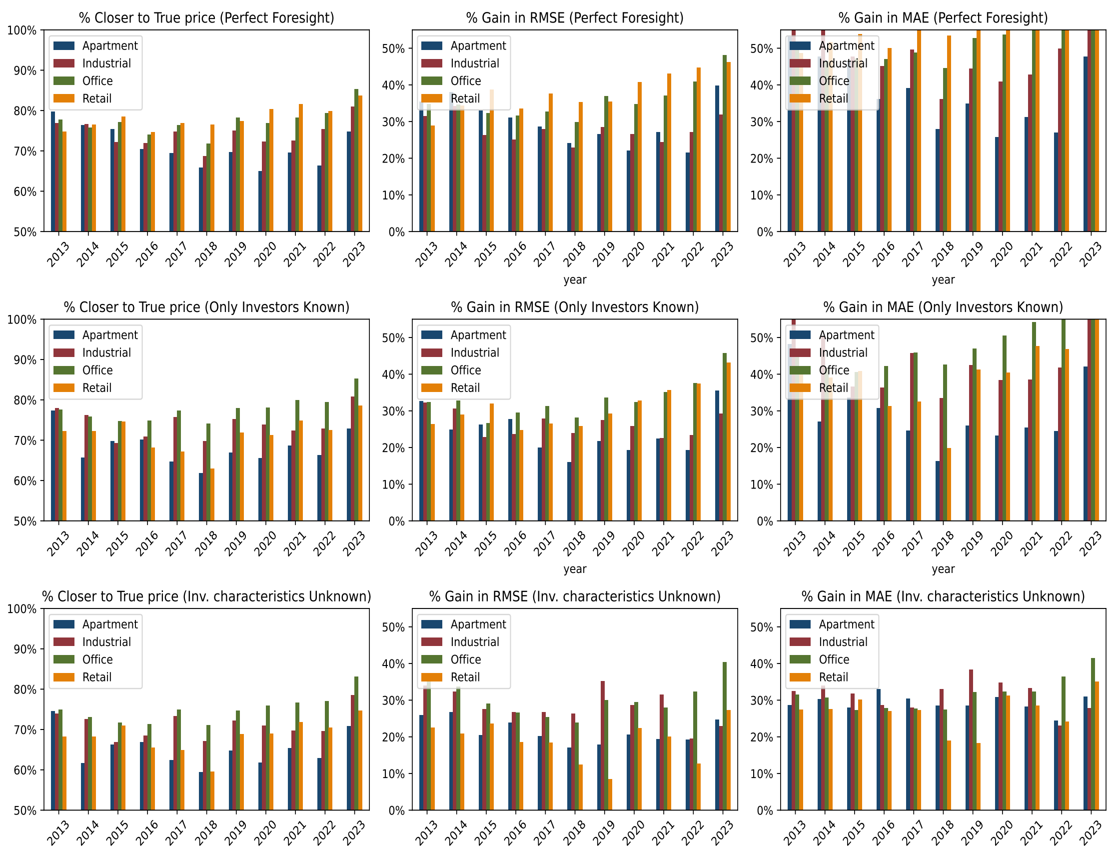
For an owner looking to sell their property in the next period, a key issue is to predict the transaction price and assess the price risk. Our CRE model is perfectly positioned to answer both questions.

The two main insights from the valuation model are that building, market, and investor characteristics all matter for observed transaction prices, and that it is important for accurate valuations to capture the nonlinear and interaction effects of those variables on prices. We now explore the model's ability to predict transaction prices out-of-sample, and how important these different features are.

Standing at time t, predicting prices and estimating the uncertainty of the transaction price amounts to predicting the inputs in the valuation model. To understand which inputs are most important, we proceed in three steps.

In the first step, we assume that the building, market, and investor characteristics can be predicted with perfect foresight. We estimate the model using information up to time t and predict the prices in $t+1$ using the observed covariates (x_{t+1}, z_{t+1}). This is the OOS prediction analysis discussed before, with R^2 values reported in the second panel of Table 3. Figure 14 reports gains in OOS predictability of our baseline LGBM model relative to the linear hedonic model reported in Table 2, which reflects the state of affairs in CRE valuation. The left panel shows the fraction of transactions for which the LGBM is closer to the true transaction price than the linear hedonic model. The middle panel shows the percent gains in terms of median absolute pricing error (MAE), and the right panel reports the gains in terms of root mean squared error (RMSE). The top row shows that the LGBM strongly outperforms the linear hedonic model out-of-sample. The LGBM model's forecast is closer to the true observed price in 70-80% of transactions across years and sectors. In terms of the magnitude of the pricing error, the LGBM model's predicted price is typically 30-40% closer to the true price than the linear hedonic model's prediction (with MAE gains considerably higher than RMSE gains).

Figure 14: Comparing Out of Sample performance



Notes: This figure shows the relative out-of-sample performance of the LGBM model relative to the linear hedonic model (LHM). The first column reports the fraction of transactions for which the LGBM prediction is closer to the true transaction price than the LHM prediction. The second column reports the percent gain in root mean squared prediction error of the LGBM relative to the LHM model. The third column reports the percent gain in median absolute prediction error of the LGBM relative to the LHM model. While the prediction model is always estimated with information up until time t in order to predict the price at $t+1$, the rows differ in how we determine the covariates (x_{t+1}, z_{t+1}) that enter in the prediction of the time-$t+1$ price. The first row assumes that (x_{t+1}, z_{t+1}) are known. The second row assumes that z_{t+1} is known and that x_{t+1} must be predicted. The third row assumes that (x_{t+1}, z_{t+1}) must both be predicted.

In the second row of Figure 14, we assume that the time-$(t+1)$ building and market characteristics and time fixed effects are unknown. We use their time-t values as estimates; i.e., we assume that the macro variables follow a random walk: $E_t[x_{t+1}] = x_t$. We continue to assume that the buyer characteristics z_{t+1} are known. The LGBM model continues to produce predictions that are closer to the true transaction price than the linear hedonic model's predictions for 70-80% of transactions. The gains in RMSE and MAE are lower by 3-6% on average when hedonics have to be predicted than when they are known.

Third, we assume that all of the building, market, and investor characteristics need to be predicted. For the building and market characteristics, we continue to use the time-t values. For the buyer information, we take advantage of our listing model. Specifically, for each property that transacts, the forecaster who stands at time t knows the characteristics of the seller, but not the characteristics of the buyer who materializes at time $t+1$, $z_{b,t+1}$. To deal with this, we draw $C = 1,000$ potential buyers in proportion with their likelihood according to the time-t meeting model. For each one, we compute the price based on that buyer's valuation and a shock, and check whether the buyer's value exceeds that of the seller. We obtain a distribution of potential prices which we average to estimate the predicted transaction price. The third row of Figure 14 shows that the LGBM generates better out-of-sample predictions than the linear hedonic model in about 60-70% of transactions. This outperformance metric shows smaller values than in rows 1 and 2. The average reduction in pricing error is 20-30%, with more variation across years and sectors.

The exercise shows that the LGBM model with investor characteristics displays superior out-of-sample predictive ability. The results also suggest that the benefit to CRE investors of better prediction models for building and market characteristics are modest; the reduction in prediction error from step 1 to step 2 is small. The gains from having a better matching model are larger; as shown by the larger difference between steps 2 and 3.

7.2 Understanding CRE price dynamics

A key insight of our analysis is that investors have heterogeneous valuations for properties. We now explore an application of the CRE model in which we quantify the role of heterogeneous valuations.

Figure 4 showed that a large number of offices and apartments changed hands from REITS to REPE funds and Institutions in the lead-up to the Great Financial Crisis. As our model indicates, some of these transactions may have reflected differences in beliefs between these types of investors about expected returns and differences in risk tolerance. But some may have arisen because REPE funds had strong fundraising years in 2005–2007 and may have been compelled to execute deals. To understand the impact of this heterogeneity across investors on Office and Apartment prices in 2006 and 2007, we compute the counterfactual experiment in which REITS still list their properties, but the potential buyer pool is different. Specifically, we explore what would have happened had REPE and Institutional investors not been around to buy these properties.

We follow the procedure outlined in Section 3.3 to compute potential price distributions. For every office and apartment building sold by a REIT in 2006 and 2007, we simulate a potential buyer from the meeting probability distribution. If the buyer's valuation exceeds the seller's the building transacts and we record the price. We compute the average price among the simulated potential buyers who transact.

Table 6 presents the results. In the first row of each panel, we show the average price per unit (square feet for office, units for apartments) that actually occurred in the data. In the second row, we show the average price from the potential price distribution when all investors are present in the potential buyer pool. As discussed before, this baseline model does a good job generating a high fraction of transactions (second column) and predicts an average price among transactions that is close to the observed one (third column). It also gets the composition of buyer types about right (fourth column).

Table 6: CRE Price Dynamics (REIT sell-off in 2006-2007)

2006 Offices (719 Buildings)

Model	% Trans.	Avg ppsf	Major Buyers (% Value of properties bought)
Truth	100.00	159.4	[REPE: 26.47, Institutional: 25.6, REITS: 23.07, OOD_N: 13.84]
None	85.45	170.9	[REPE: 22.39, Institutional: 22.24, OOD_N: 20.73, REITS: 14.89]
Exclude REPE and Institutional	74.61	159.7	[REITS: 35.46, OOD_N: 30.64, OOD_L: 17.40]
Exclude REPE, Institutional and REITS	68.90	138.2	[OOD_N: 48.03, OOD_L: 26.71, User: 11.7]

2007 Offices (532 Buildings)

Model	% Trans.	Avg ppsf	Major Buyers (% Value of properties bought)
Truth	100.00	208.3	[REPE: 41.73, Institutional: 19.87, :OOD_N 17.42, REITS: 9.92]
None	78.21	194.1	[REPE: 23.9, Institutional: 19.42, OOD_N: 17.11, REITS: 16.51]
Exclude REPE and Institutional	67.31	168.1	[OOD_N: 31.97, REITS: 27.42, OOD_L: 20.4]
Exclude REPE, Institutional and REITS	62.39	149.5	[OOD_N: 41.91, OOD_L: 26.99, 'User': 11.72]

2006 Apartments (670 Buildings)

Model	% Trans.	Avg ppu	Major Buyers (% Value of properties bought)
Truth	100.00	68,955	[OOD_N: 30.87, Institutional: 23.34, REITS: 19.6, REPE: 18.24]
None	84.14	70,058	[REITS: 33.32, OOD_N: 32.77, Institutional: 11.9, REPE: 11.63]
Exclude REPE and Institutional	75.94	66,255	[OOD_N: 40.03, REITS' 34.14, OOD_L: 15.84]
Exclude REPE, Institutional and REITS	69.12	57,380	[OOD_N: 58.05, OOD_L: 29.07]

2007 Apartments (701 Buildings)

Model	% Trans.	Avg ppu	Major Buyers (% Value of properties bought)
Truth	100.00	113,831	[OOD_N: 28.62, Institutional: 28.39, REPE: 26.35]
None	78.38	113,254	[OOD_N: 29.16, REITS: 20.88, REPE: 19.55, Institutional: 17.49]
Exclude REPE and Institutional	68.23	96,827	[OOD_N: 36.91, REITS: 28.51, OOD_L: 22.04]
Exclude REPE, Institutional and REITS	63.53	83,711	[OOD_N: 54.14, OOD_L: 28.85]

Notes: Figure shows the CRE Price dynamics for office and apartment buildings sold by REITs during 2006-2007. % Trans. is the percentage of buildings that transacted in that model. The first row in every table refers to the actual data.

The third row simulates a model where we exclude REPE funds and Institutions from the set of potential buyers. The main finding is that the prices of offices sold by REITS in 2006 would have been 7% lower and the prices of offices sold in 2007 13% lower. The average price of apartments sold in 2006 would have been 5% lower and the average price of apartments sold in 2007 15% lower. These assets would have also seen about a 10% drop in the likelihood of a sale.

The reason for these price drops is that the model identifies REPE and Institutional investors to have particularly high valuations for these properties relative to other investors in those years. Specifically, the last column shows that national owner-operator-developers assume a much higher share of transactions in the counter-factual. These investors can be considered the "value investors" of this market, looking for attractively priced assets. There are many of them; Table 1 showed that OOD_N is the largest investor category by dollar volume, accounting for 30% of buys and sells in our data set.

We note that in this counterfactual, REITs also emerge as potential buyers of the assets sold by (other) REITs. As an additional counterfactual, we ask what would have happened if all REITs had stayed out of the market for these 2,622 office and apartment properties. The fourth row of Table 6 shows that the average price for the offices sold in 2006 and 2007 would have been 19% and 23% lower, respectively than in the baseline model of row 2. The average price for the apartments sold in 2006 and 2007 would have been 18% and 26% lower. In this counterfactual, end-users emerge as alternative buyers of offices and local owner-operator-developers step into the fray in both office and apartment markets. These investors have even lower valuations for these properties, possibly because of a size, location, or quality mismatch with their existing holdings or because of tighter financial constraints. This counterfactual illustrates that the presence of REITs as buyers lifted values by 11%-14% (comparing rows 3 to 4 in each panel). The model produces substantial heterogeneity in valuations, even within investor types such as REITs.

This exercise suggests that REPE funds and Institutions propped up prices just prior to the GFC. Many of the 2,622 properties investigated here were trophy assets and the deals received a lot of media coverage.[32] Their high sale prices, relative to the counterfactual valuation, may have blurred an early-warning signal of the impending financial crisis and led to larger losses for the buyers.

7.3 Substitution Patterns across Size, Sector, and Location

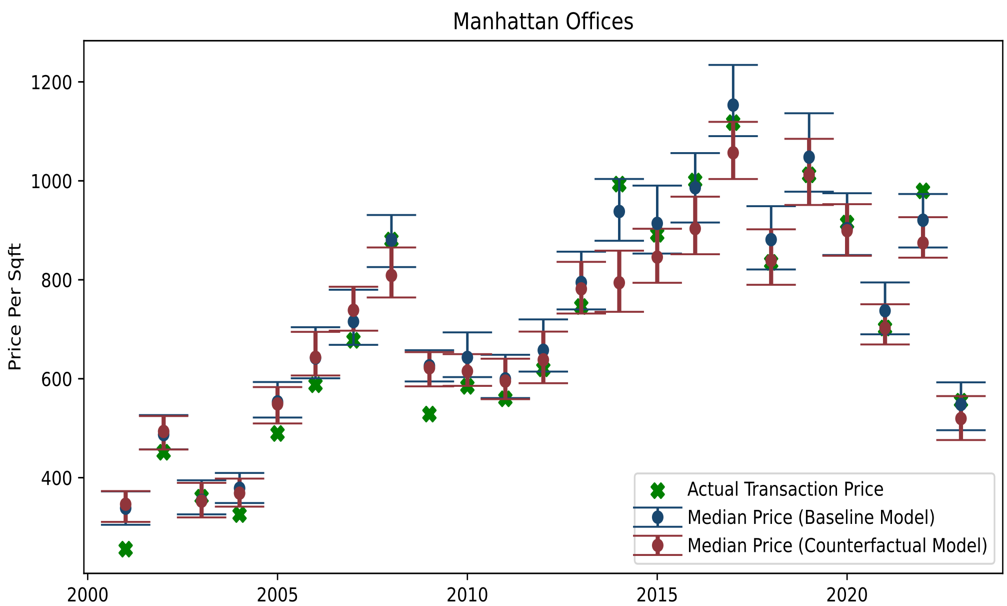
In our third application, we show how our model can be used to analyze substitution patterns in deal flows across building size, sector, and location. This is of interest to policymakers and CRE investors alike.

Concretely, we focus on the role of foreign buyers who became large net buyers of U.S CRE in 2015–2018

[32] For example, the REIT office sales in 2007 contained the entity sale of the REIT Equity Office Properties by legendary CRE investor Sam Zell to the REPE fund Blackstone. That deal contained 500 office buildings with a combined transaction value of $39 billion.

and again in 2021, as shown in Figures 4 and A.5. These foreign buyers, such as Middle Eastern sovereign wealth funds and large Canadian pension plans, had a strong preference for large, high-end properties in superstar cities. In addition to understanding the price impact on those types of office assets, we also use the model to study substitution patterns. If a large number of foreign investors purchase large, high-end Manhattan office buildings, what is the impact on transaction volume and prices of smaller, lower-quality office buildings in Manhattan, on retail and apartment buildings in Manhattan, or on office buildings in other cities such as Chicago, Los Angeles, and San Francisco?

Figure 15: Manhattan Office Market: Counterfactuals



Notes: This figure shows the potential price distribution for Manhattan office in the baseline model (In Blue) and in a counterfactual model without foreign investors (In Red) in each year between 2001 and 2023. The top (bottom) line segment indicates the 75th (25th) percentile of potential price distribution, and the green cross the median. The actual transaction price is given by the red dot.

We start from the high-end office buildings in Manhattan bought by foreign investors in a given year. Following the algorithm described in Section 3.3, we assess the impact of foreign buyers by computing a counterfactual economy in which foreign investors did not enter the Manhattan office market. We draw $C = 1,000$ potential buyers for the assets that were bought by foreigners from the set of all potential buyers that excludes foreigners. As always, these alternative buyers are drawn in proportion to their likelihood of buying the specific asset under consideration.

The blue line segments in figure 15 show the potential price distribution for the Manhattan office properties bought by foreigners in each year between 2001 and 2023 according to the baseline model that includes

Foreign investors in the buyer pool. The top and bottom whiskers indicate the 75th and 25th percentiles and the blue dot in between the median of the potential price distribution. The green cross is the actual transaction price. The red whiskers show the potential price distribution for the same assets in a counterfactual world where we remove all Foreign investors. The assets that in reality were bought by Foreigners are now bought by other types of investors, sampled in proportion to their model-implied transaction probabilities. We see that there are several years, such as 2015, 2016 and 2017, where the counterfactual price distribution shifts down meaningfully from the benchmark one. These were years with large foreign buyer activity, both in terms of dollars and share of purchases. Removing foreign buyers lowers average Manhattan office prices by 4.7% over the full sample period 2001–2023, and by 7.5% during the period 2013–2022. For trophy assets, defined as the top 50% of Manhattan office properties by price per square foot, removing foreign buyers lowers prices by 6.2% over the full sample and by 8.1% during 2013—2022. Put differently, the foreign purchase activity meaningfully pushed up the prices of Manhattan office assets, and especially at the top end of the market.

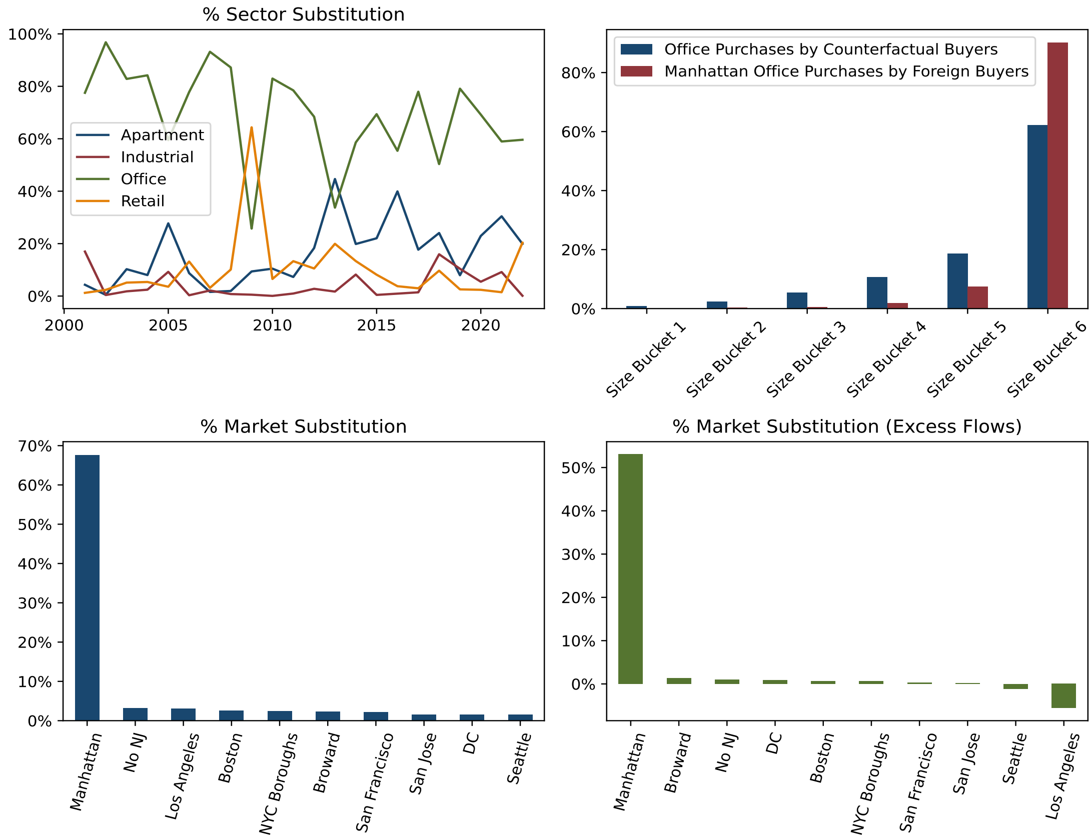
The counterfactual exercise gives us a list of potential buyers that would have bought the properties in question in the absence of foreign investors. With this list in hand, we explore whether those buyers bought another property in the same year or in year thereafter. By analyzing what kinds of properties they bought instead, we obtain an estimate of the substitution patterns across property sizes, sectors, and locations.

The results are in Figure 16. The top left panel shows the sectoral composition of purchase volume by the counterfactual buyers. The majority of their purchases are in the office sector, about 60% on average but with a declining trend over time. There is substitution to apartments, which represent about 30% of purchases by the counterfactual buyers in the years 2015–18 of heavy foreign office purchase activity. There is also some substitution to the Industrial sector which was in-vogue in 2019–21, and to Retail in 2008 and 2023.

The top right panel studies substitution across property sizes. For reference, the red bars plot the size distribution of the actual foreign purchases of Manhattan office. The blue bars show the size distribution of actual office purchases made by the counterfactual buyers in those same years in any market. We find that foreigners predominantly purchased very large offices and pushed other office buyers towards smaller office purchases.

The bottom left panel shows the locations of the purchases made by the counterfactual buyers across all sectors. There is limited substitution across space. About 65% of actual purchase activity by this group took place in Manhattan. Foreign buyers may have crowded them out of specific Manhattan office assets, but it did not fully crowd them out of Manhattan. The counterfactual buyers may themselves have strong preferences for (the types of assets available in) Manhattan, and have shifted to Manhattan apartments

Figure 16: Manhattan Office Market: Substitution Patterns



Notes: This figure examines the purchasing behavior of the potential buyers who would have bought Manhattan office properties had it not been for the Foreign investors who actually bought them. The top left panel plots the share of transaction volume invested by these potential buyers across the four sectors. The top right panel shows the property size distribution of office purchases across all markets by the counterfactual buyers (blue bars) and compares them to the size distribution of Foreign purchases of Manhattan office properties (red bars). The bottom left panel plots the top ten markets by average percentage of purchase volume by the counterfactual buyers. The bottom right panel subtracts from the bottom left panel the average percentage of purchase volume by all manhattan office buyers.

rather than to San Francisco offices. To the extent that there is geographic substitution, it is fairly evenly directed towards other superstar markets such as San Francisco, Boston, Los Angeles, and DC. There also is some activity in the Northern New Jersey and NYC suburbs.

The bottom right panel of Figure 16 subtracts from the geographic portfolio shares of the counterfactual buyers, shown in the bottom left panel, the geographic portfolio shares of all investors who bought an office in Manhattan in the same period. It shows that the counterfactual buyer pool is much more heavily concentrated in Manhattan and somewhat less in Los Angeles, relative to the average Manhattan office buyer. Foreigners crowded out Manhattan office specialists.

8 Conclusion

We provide a novel approach to determine the joint evolution of prices and volume in decentralized markets for private and real assets. We apply our approach to a large and granular data set of commercial real estate transactions that contain detailed information not only on property characteristics but also on the identities of buyers and sellers. We find that using the investor composition in the commercial real estate ecosystem is critical to understanding asset prices. The meeting model, inspired by directed search and matching theory, demonstrates that buyer-seller interactions are shaped by factors such as investor size, portfolio similarity between buyers and seller, and asset feature alignment in terms of property size, location, sector, and quality. These insights underscore the role of matching frictions and heterogeneous preferences in shaping transaction dynamics. The counterfactual analyses highlight the pivotal role of investor composition in determining price outcomes. For example, foreign buyers increased Manhattan office prices by more than 7% in the decade 2013–22. REPE fund purchases of large REIT sales in the office and apartment sectors in 2006 and 2007 increased property valuations by more than 10%. The commercial real estate price collapse in the Great Financial Crisis would have been less severe, emphasizing how specific investor groups influence market equilibria.

These insights extend beyond real estate. Our paper offers a powerful new methodology for understanding price formation and transaction dynamics in other private asset classes, such as private equity, private credit, infrastructure, and residential real estate, where buyer-seller interactions are similarly crucial.

Furthermore, our new estimation technique for matching functions, which leverages insights in the computer science literature, should have broad applicability to other search and matching settings such as labor and marriage markets.

References

Andersen, Steffen, Badarinza, Cristian, Liu, Lu, Marx, Julie and Ramadorai, Tarun. (2022). 'Reference Dependence in the Housing Market', *American Economic Review* 112(10), 3398–3440.

Ang, Andrew, Chen, Bingxu, Goetzmann, William N. and Phalippou, Ludovic. (2018). 'Estimating Private Equity Returns from Limited Partner Cash Flows', *Journal of Finance* 73(4), 1751–1783.

Arora, Abhishek and Dell, Melissa. (2023). 'LinkTransformer: A Unified Package for Record Linkage with Transformer Language Models'.

Badarinza, Cristian, Balasubramaniam, Vimal and Ramadorai, Tarun. (2024), In Search of the Matching Function in the Housing Market, Technical report, SSRN.

Badarinza, Cristian and Ramadorai, Tarun. (2018). 'Home Away from Home? Foreign Demand and London House Prices', *Journal of Financial Economics* 130(3), 532–555.

Badrinza, Cristian, Ramadorai, Tarun and Shimizu, Chihiro. (2022). 'Gravity, Counterparties and Foreign Investment', *Journal of Financial Economics* 145(2), 132–152.

Cvijanović, Dragana, Milcheva, Stanimira and van de Minne, Alex. (2022). 'Preferences of Institutional Investors in Commercial Real Estate', *The Journal of Real Estate Finance and Economics* 65, 321–359.

Driessen, Joost, Lin, Tse-Chun and Phalippou, Ludovic. (2012). 'A new method to estimate risk and return of nontraded assets from cash flows: The case of private equity funds', *Journal of Financial and Quantitative Analysis* 47(3), 511–535.

Fox, Jeremy T. (2018). 'Estimating Matching Games with Transfers', *Quantitative Economics* 9(1), 1–38.

Fuster, Andreas, Goldsmith-Pinkham, Paul, Ramadorai, Tarun and Walther, Ansgar. (2022). 'Predictably Unequal? The Effects of Machine Learning on Credit Markets', *Journal of Finance* 77(1), 5–47.

Gabaix, Xavier, Koijen, Ralph S.J., Mainardi, Federico, Oh, Sangmin Simon and Yogo, Motohiro. (2025), Limited Risk Transfer between Investors: A New Benchmark for Macro-Finance Models, Technical report, NBER Working Paper 33336.

Gabaix, Xavier, Koijen, Ralph S.J., Richmond, Robert and Yogo, Motohiro. (2024), Asset Embeddings, Technical report, SSRN Working Paper Series.

Genesove, David and Mayer, Christopher. (2001). 'Loss Aversion and Seller Behavior: Evidence from the Housing Market', *Quarterly Journal of Economics* 116, 1233–1260.

Ghent, Andra C. (2021). 'What's wrong with Pittsburgh? Delegated investors and liquidity concentration', *Journal of Financial Economics* 139(2), 337–358.

Giacoletti, Marco. (2021). 'Idiosyncratic Risk in Housing Markets', *The Review of Financial Studies* 34(8), 3695–3741.

Goetzmann, William N, Spaenjers, Christophe and Van Nieuwerburgh, Stijn. (2021). 'Real and private-value assets', *The Review of Financial Studies* 34(8), 3497–3526.

Griliches, Zvi. (1971), Introduction: Hedonic Price Indexes Revisited, *in* **Zvi Griliches**., ed., 'Price Indexes and Quality Change: Studies in New Methods of Measurement', Harvard University Press, Cambridge, MA, pp. 3–15.

Guolin Ke, Qi Meng, Thomas Finley Taifeng Wang Wei Chen Weidong Ma Qiwei Ye Tie-Yan Liu. (2017). 'LightGBM: A Highly Efficient Gradient Boosting Decision Tree', *Advances in Neural Information Processing Systems* 30, 3149–3157.

Gupta, Arpit, Mittal, Vrinda and Van Nieuwerburgh, Stijn. (2025). 'Work From Home and the Office Real Estate Apocalypse', *American economic Review* forthc.

Gupta, Arpit and Van Nieuwerburgh, Stijn. (2021). 'Valuing private equity investments strip by strip', *The Journal of Finance* 76(6), 3255–3307.

Hug, Nicolas. (2020), 'Surprise: A Python library for recommender systems'.

Kaplan, Steven N. and Schoar, Antoinette. (2005). 'Private equity performance: Returns, persistence, and capital flows', *Journal of Finance* 60(4), 1791–1823.

Kim, Min. (2024), Comparative R-Squared. SSRN Working Paper No. 3790200.

Koijen, Ralph S.J., Richmond, Robert J. and Yogo, Motohiro. (2024). 'Which Investors Matter for Equity Valuations and Expected Returns', *Review of Economic Studies* 91(4), 2387–2424.

Koijen, Ralph S.J. and Yogo, Motohiro. (2019). 'A Demand System Approach to Asset Pricing', *Journal of Political Economy* 127(4), 1475–1515.

Korteweg, Arthur and Nagel, Stefan. (2016). 'Risk-Adjusting the Returns to Venture Capital', *Journal of Finance* 71(3), 1437–1470.

Korteweg, Arthur and Sørensen, Morten. (2010). 'Risk and Return Characteristics of Venture Capital-Backed Entrepreneurial Companies', *Review of Financial Studies* 23, 3738–3772.

Lancaster, Kevin. (1966). 'A New Approach to Consumer Theory', *Journal of Political Economy* 74, 132–157.

Lundberg, Scott M, Erion, Gabriel, Chen, Hugh, DeGrave, Avery, Prutkin, Jordan M, Nair, Bharat, Katz, Ronit, Himmelfarb, Jonathan, Bansal, Nisha and Lee, Su-In. (2020). 'From local explanations to global understanding with explainable AI for trees', *Nature Machine Intelligence* 2(1), 252–259.

Lundberg, Scott M and Lee, Su-In. (2017), A unified approach to interpreting model predictions, *in* 'Advances in Neural Information Processing Systems', Vol. 30.

Ma, Zhuang and Collins, Michael. (2018). 'Noise Contrastive Estimation and Negative Sampling for Conditional Models: Consistency and Statistical Efficiency', *CoRR* abs/1809.01812.

Mikolov, Tomás, Sutskever, Ilya, Chen, Kai, Corrado, Greg and Dean, Jeffrey. (2013). 'Distributed Representations of Words and Phrases and their Compositionality', *CoRR* abs/1310.4546.

Mittal, Vrinda. (2024), Desperate Capital Breeds Productivity Loss: Evidence from Public Pension Investments in Private Equity. SSRN Working Paper No. 4283853.

Peng, Liang. (2016). 'The risk and return of commercial real estate: A property level analysis', *Real Estate Economics* 44(3), 555–583.

Plazzi, Alberto, Torous, Walter and Valkanov, Ross. (2010). 'Expected Returns and the Expected Growth in Rents of Commercial Real Estate', *Review of Financial Studies* 23, 3469–3519.

Plazzi, Alberto, Torous, Walter and Valkanov, Rossen. (2008). 'The Cross-Sectional Dispersion of Commercial Real Estate Returns and Rent Growth: Time Variation and Economic Fluctuations', *Real Estate Economics* 36(3), 403–439.

Rosen, Sherwin. (1974). 'Hedonic Prices and Implicit Markets: Production Differentiation in Pure Competition', *Journal of Political Economy* 82, 34–55.

Sagi, Jacob S. (2021). 'Asset-Level Risk and Return in Real Estate Investments', *Review of Financial Studies* 34(8), 3647–3686.

Takuya Akiba, Shotaro Sano, Toshihiko Yanase Takeru Ohta and Koyam, Masanori. (2019). 'Optuna: A Next-generation Hyperparameter Optimization Framework', *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining* pp. 2623–2631.

Van Nieuwerburgh, Stijn. (2019). 'Why Are REITs Currently So Expensive?', *Real Estate Economics* 47(1), 18–65.

Van Nieuwerburgh, Stijn, Stanton, Richard and de Bever, Leo. (2015), 'A Review of Real Estate and Infrastructure Investments by the Norwegian Government Pension Fund Global (GPFG)'.

Wallace, Nancy E. (1996). 'Hedonic-based price indexes for housing: Theory, estimation, and index construction', *Economic Review* pp. 34–48.

Witte, Ann D., Sumka, Howard J. and Erekson, Homer. (1979). 'An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets', *Econometrica* 47(5), 1151–1173.

A Data Appendix

A.1 Data Cleaning Process

This appendix describes the process used to clean and filter the raw transaction data from RCA. The dataset includes transaction-level data for four core sectors—apartments, industrial, office, and retail—from 2001 to 2023.

We begin by removing transactions that are unlikely to be useful or may introduce noise. Using the transaction type variable, we retain only "Sale," "Construction," "Entity Level," and "Transfer" transactions, excluding land sales, air rights, and leases. Additionally, we drop incomplete transactions (e.g., contracts not completed) and those valued under \$500,000. These filters eliminate approximately 5% of all transactions.

For each transaction, RCA provides up to four buyers and sellers, though less than 1% involve all four. We consolidate investor names with inconsistent spellings or formats using the method in Arora and Dell (2023). Specifically, we use two independent models to generate 20 closest matches per name, apply a 90% similarity threshold, manually discard false positives, and replace names with the most frequent version. This standardization allows us to track investors over time, aiding in two key tasks: (1) identifying property owners for each year (critical for the seller role in our meeting model), and (2) constructing investor-level portfolios to study asset evolution. The latter supports variables like investor transaction activity, size, and consideration set characteristics.

Transactions suspected of being non-arms-length (e.g., multiple deals involving the same property within 30 days) are consolidated. We merge such transactions into a single record, using the price and characteristics of the final transaction and assigning buyers and sellers from the first and last transactions, respectively.

Some properties change sectors during the sample period due to conversions. To ensure accurate analysis, we assign separate property IDs for periods before and after such conversions.

Transactions where the buyer and seller are the same entity are removed. For cases where one party transacts on behalf of another, we retain the primary investor and drop intermediaries (e.g., investment managers). If there are multiple buyers or sellers, we allocate equal ownership shares unless specific ownership structures are known (e.g., in Joint Ventures, where we adjust shares accordingly).

We map RCA's raw investor types ($\sim$ 20 groups) into eight core investor categories, as shown in Table A.7, to capture meaningful heterogeneity in investment objectives, size, and geographic focus

We track property ownership over time to construct investor portfolios. Challenges such as missing or incorrect transactions are addressed by inferring missing links from prior and subsequent transactions. For example, if investor A sells to B, and C later sells the same property to D, we estimate an intermediate transaction where B sells to C. We adjust buyer and seller shares to ensure total ownership sums to 100% without negative values. Approximately 20% of properties require such adjustments during the sample period. These corrections are used solely for portfolio construction and are excluded from valuation and meeting model calibrations, which rely strictly on actual transaction data.

For entity sales, where RCA equally distributes transaction value across properties, we adjust property-level prices by assigning per-square-foot values proportional to RCA's sector-market price indices. This provides a more accurate estimate than assuming uniform prices across locations.

We construct property price panels for each quarter/year by combining RCA price indices with unadjusted CPI prices and adjusting them to 2023 dollars. CPI adjustments use the seasonally adjusted Consumer Price Index (CPIAUCSL) from FRED.

Market occupancy data is primarily sourced from NCREIF at the CBSA-sector level. For missing values, we calculate volume-weighted occupancy rates using RCA data, provided at least 15 properties in the sector-year contain valid occupancy information. Net Operating Income (NOI) data, also from NCREIF, is CPI-adjusted and annualized. NOI growth is calculated logarithmically if both the current and previous NOI values are positive; otherwise, we use a smoothed growth rate formula. Extreme outliers (top/bottom 1%) are replaced with median values.

Geographic variables (e.g., zip codes, FIPS codes) are filled where possible. For markets with intermittent missing data (e.g., at least 50% coverage), we use the Simple Python Recommendation System Engine (SURPRISE) Hug (2020) to impute values. Missing data is more prevalent for remote areas and for retail and industrial properties, which are often further from city centers. For geographies with no information, we impute values using sector-level national averages, though such cases are rare.

A.2 Mapping Geographies

The RCA dataset provides geo-codes for every property, which we leverage to map properties to various geographic indicators. This is essential because the macroeconomic time series data we use comes from multiple sources, each often organized by different geographic units. Using the 3-digit rounded latitude and longitude data provided by RCA, we generate ZIP codes for each property. This is done using standard geolocation algorithms that match coordinates to corresponding ZIP codes.

We use the county-level Federal Information Processing Standards (FIPS) codes to map properties to Core-Based Statistical Areas (CBSAs). This mapping is based on a crosswalk file obtained from the National Bureau of Economic Research (NBER). RCA provides its own market classification system, which consists of 165 individual markets. To align with other datasets and enhance interpretability, we aggregate RCA's 165 markets into 61 broader geographic categories based on logical groupings and economic characteristics. The aggregated RCA markets enable direct mapping of properties to RCA price index files and cap rate data, both of which are reported at the level of these 61 aggregated markets. This standardization allows for consistent integration of property-level data with macroeconomic indicators. Table A.1 shows the mapping between RCA markets and our 60 geographic markets.

A.3 Net Effective Rent

This appendix describes how we construct net effective rent statistics of the property or the neighborhood of each property that we see transact in the RCA data. Net effective rents are observed for office, retail, and industrial properties at the lease level in the Compstak database. We first merge the RCA and the CompStak databases at the property level using the following procedure: for each RCA property, we consider CompStak buildings within a 1 km radius of its geo point. We assign each CompStak building a matching score based on the geographical distance, the difference between the years the properties were built, the difference in log building size, and the fuzzy matching score on property names. Specifically, the matching score equals the sum of: (1) $0.3 \times \max\{1 - \frac{\text{geo distance (km)}}{0.3}, 0\}$, (2) $0.3 \times \max\{1 - \frac{|\Delta\text{year built}|}{5}, 0\}$, (3) $0.3 \times \max\{1 - \frac{|\Delta\text{log building size}|}{5\%}, -1\}$, and (4) $0.4 \times \max\{\frac{\text{name similarity}-70}{30}, 0\}$, where name similarity is the highest partial Levenshtein similarity between elements in the RCA property name list and the CompStak name list (including all street addresses and building names of that property). Additionally, if the similarity score is exactly 100, we assign this part of the score as 0.4×2 instead of 0.4×1. We keep the CompStak building with the highest matching score. If the maximum score ≤ 0.3, we skip the match.

We manually check the matching result for the top 100 RCA office buildings transacted after 2015 in

Manhattan: the correct matching rate is 92%, wrong matching rate is 3%, and the missing rate is 5%. Wrong matches occur when very similar and adjacent buildings have characteristics closer than the true match, while the missing matches occur because CompStak does not provide coverage for all RCA properties. The whole-sample building-level matching rates are 55.1%, 39.3%, and 48.3%, respectively, for office, retail, and industrial properties in the RCA data.

For RCA properties not matched with a CompStak building, or matched with a CompStak building lacking sufficient recent lease data to compute measures, we create variables in the lease-level data at a fine level of geography (call it a "block"), and match each property we see transact in RCA to its block-level variables.

The finest level of geography we consider uses the first 2 digits of the latitude and longitude. For example, for a property with latitude and longitude (40.754932,-73.984016), we define a 2d area as [40.75,40.76;-73.99,-73.98]. This corresponds to an area of roughly 0.6 miles by 0.6 miles (1km by 1km). We collect all the leases in this block to calculate the average rent measures. Given that leasing activity is sparse, we use all leases signed in the past three years to calculate the block-year level average rent. We weigh leases by their transaction sqft, and assign a lower weight for leases signed in the more distant past. In the baseline, we assign weights of 1, 1/2, and 1/3 to leases signed one, two, and three years before, and then renormalize to make the weights sum to one.

When there are fewer than 5 buildings in the block, or when there are no leases in the past three years, we use the average rent calculated in a larger block. The next block we consider is defined based on the first decimal place of the latitude and longitude. For example, a property at latitude and longitude (40.754932,-73.984016) belongs to block [40.7,40.8;-74.0,-73.9].

When the average rent is still missing in this roughly 6 miles by 6 miles block, we fill it with the value in the block defined based on the unit digit of the latitude and longitude. In our example, this would be the area given by [40,41;-73,-74], an area of roughly 60 miles by 60 miles.

For the remaining missing values, we sequentially fill them with the average rent at the RCA market, average rent at our own market definition (which combines RCA markets, see appendix A.2), and if necessary, average rent measured at the national level.

For apartment properties, we observe the net operating income per apartment unit at the loan level in the Fannie Mae's multifamily mortgage origination data. We follow a similar approach to calculate the NOI (referred to as NER below for simplicity). Given that loan origination activity is also sparse, we use all originations in the past three years to calculate the average NER. We first use the same matching algorithm to conduct a property-level matching, obtaining a matching rate of 45.8%, and then use the block-year-level NER to fill in missing values. We start with the 2-decimal place blocks, use the same weighting scheme explained above, and fill missing values sequentially with values computed on increasingly coarse geographic levels.

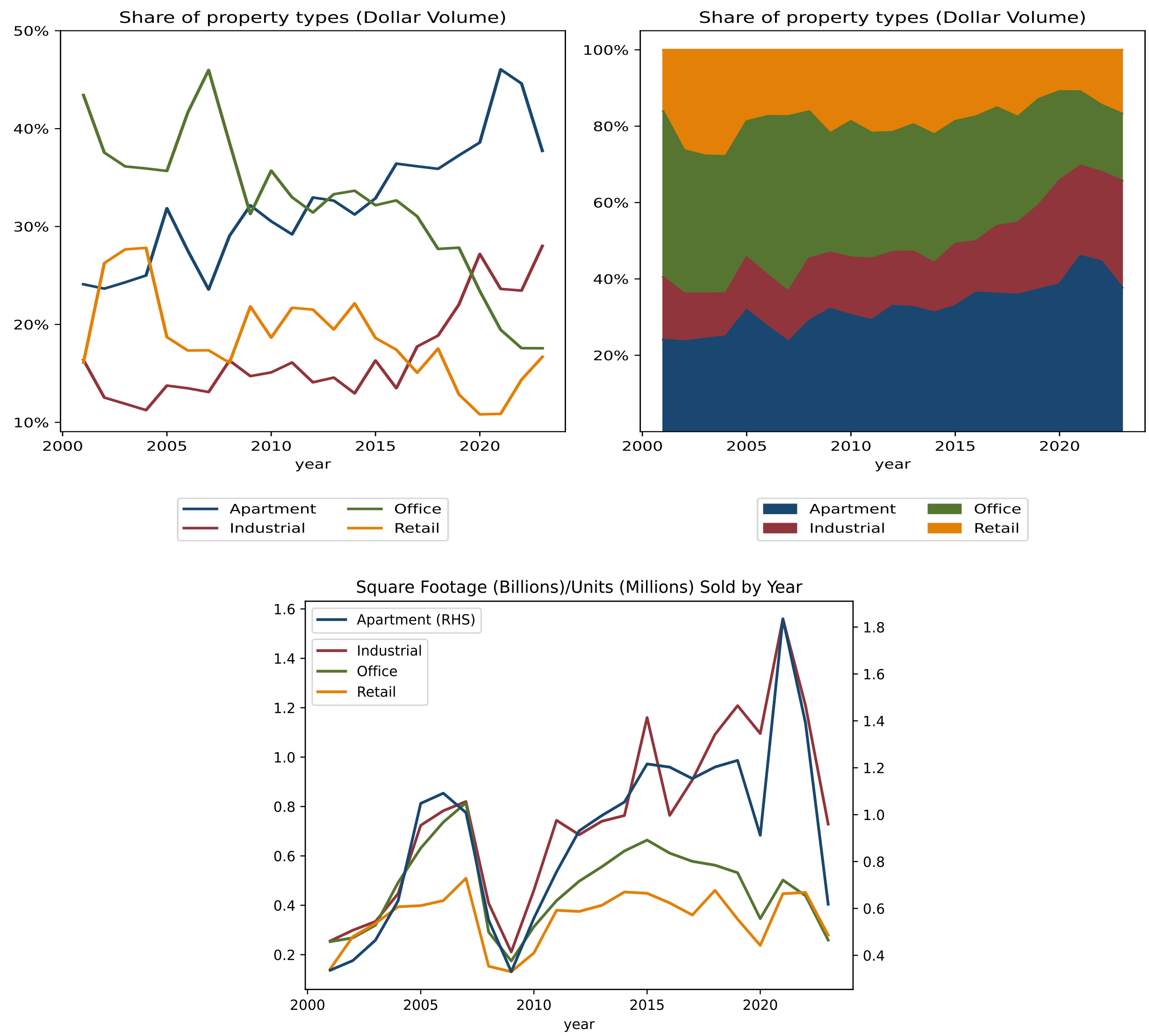
Table A.2 reports the proportion of the final NER observations for each sector (columns) that come from each level of geography (rows). Most of the NER are measured at the 1-decimal level (6 miles by 6 miles). For office, we have nearly 42% of observations measured at the even finer 2-decimal level (0.6 miles by 0.6 miles).

To understand how well the block-level NER measures captures property-level variation in NER, we estimate a linear regression of property-level NER in the Compstak/Fannie Mae data on the block-level NER variable, where we use the most spatially granular variable available for that property. The R^2 of those property-level regressions is reported in the last row of Table A.2.

We have explored several more neighborhood-level variables available in the Compstak data, such as occupancy, quality grade (class A, B, C), remaining lease duration, tenant size (number of employees of the tenant), tenant concentration (Herfindahl index computed form the leases in a building), but none of these variables add explanatory power to the valuation model, once the NER is included.

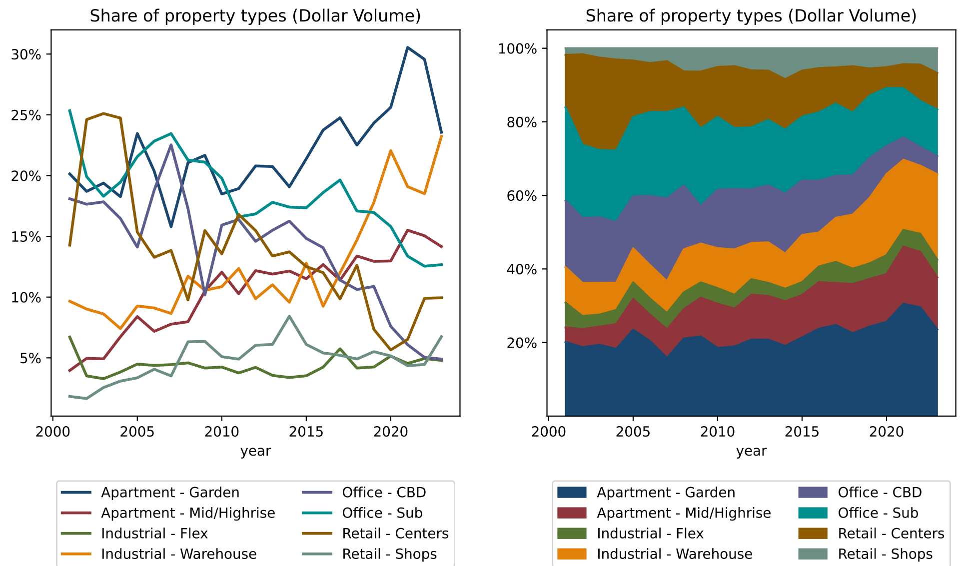
A.4 Transaction Summary Statistics

Figure A.1: Transaction Volume by Property Type



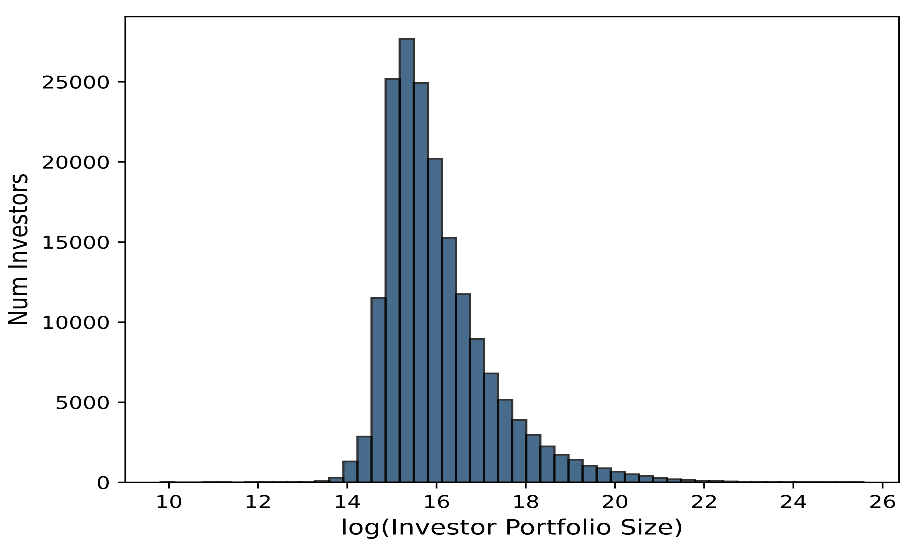
Notes: The top panel shows share of properties transacted by property types, i.e. Apartments, Office, Industrial and Retail, from 2001 to 2023. Value is defined in terms of 2023 USD. The bottom panel shows the total square feet transacted in billions of Industrial, Office and Retail space, and the number of Apartment units bought in millions.

Figure A.2: Transaction Volume by Property Subtype



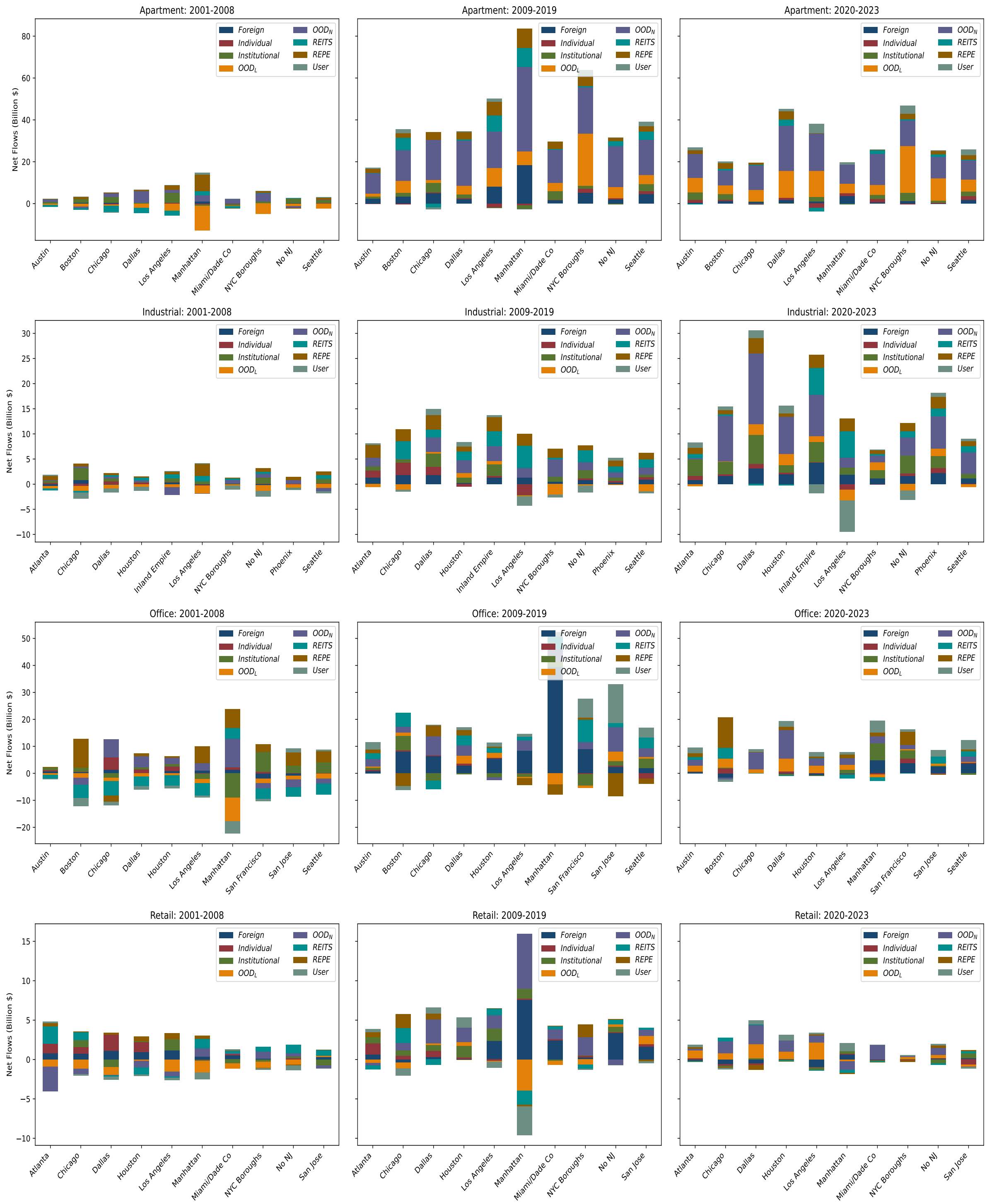
Notes: The top panel shows share of properties transacted by property subtypes., i.e. Apartments, Office, Industrial and Retail, from 2001 to 2023. Value is defined in terms of 2023 USD. The bottom Panel shows the total square footage bought (in billions) of Industrial, Office and Retail space, and the number of Apartment units bought (in millions) from 2001 to 2023.

Figure A.3: Investor Size Distribution



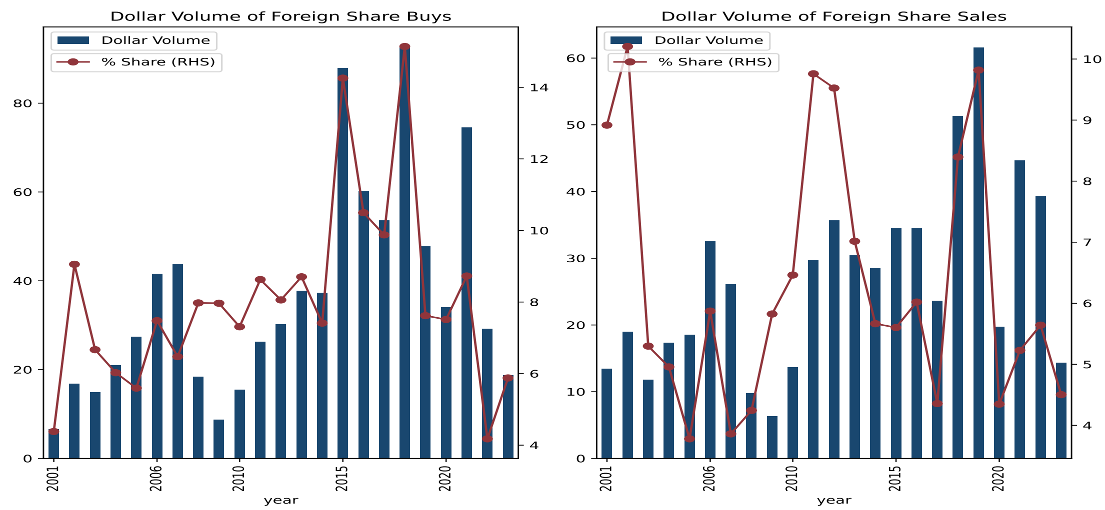
Notes: Figure shows a histogram of log(investor portfolio size) at the end of 2023 for investors that own atleast one property. Log investor size varies from 11.8 (∼ \$135K) to 25.5 (∼ \$119 bi). Not that 16.2 = \$10 mi, 17.7 = \$50 mi, 20.7 = \$1 bi.

Figure A.4: Net Flows by Investor Types



Notes: This figure shows the net acquisitions of properites by investor types and sector for during 2001-2008, 2009-2019 and 2020-2023, calculated as the dollar value of purchases minus sales (in 2023 dollars) for each sector.

Figure A.5: Foreign Investor Share



Notes: Figures show foreign investor activity in US Commercial Real Estate Ecosystem during 2001-2023. The blue bars show the dollar value (in Billions, 2023 USD) traded, aggregated by property type while the red lines show the share of foreign investors among all investors. The left panel summarizes foreign investor buys while the right panel summarizes foreign investor sales.

Table A.1: Mapping RCA Markets to Our Markets

Our Market	**RCA Markets**	**Our Market**	**RCA Markets**
Atlanta	Atlanta	Manhattan	Manhattan
Austin	Austin	Memphis	Memphis
Baltimore	Baltimore	Miami/Dade Co	Miami/Dade Co
Birmingham	Birmingham (AL)	Minneapolis	Minneapolis
Boston	Boston	Nashville	Nashville
Broward	Broward	No NJ	No NJ
Charlotte	Charlotte	NYC Boroughs	NYC Boroughs
Chicago	Chicago	Orange Co	Orange Co
Cincinnati	Cincinnati	Orlando	Orlando
Cleveland	Cleveland	Palm Beach Co	Palm Beach Co
Columbus	Columbus	Philadelphia	Philadelphia
Dallas	Dallas	Phoenix	Phoenix
DC	DC	Pittsburgh	Pittsburgh
DC MD burbs	DC MD burbs	Portland	Portland
DC VA burbs	DC VA burbs	Raleigh/Durham	Raleigh/Durham
Denver	Denver	Richmond/Norfolk	Norfolk, Richmond
Detroit	Detroit	Sacramento	Sacramento
East Bay	East Bay	Salt Lake City	Salt Lake City
Hartford	Hartford	San Antonio	San Antonio
Houston	Houston	San Diego	San Diego
Indianapolis	Indianapolis	San Francisco	San Francisco
Inland Empire	Inland Empire	San Jose	San Jose
Jacksonville	Jacksonville	Seattle	Seattle
Kansas City	Kansas City	St Louis	St Louis
Las Vegas	Las Vegas	Stamford	Stamford
Long Island	Long Island	Tampa	Tampa
Los Angeles	Los Angeles	Westchester	Westchester

Our Market	RCA Markets
Tertiary Mid-Atlantic	All Others - DE,MD,VA,WV; Ann Arbor, Albany, All Others - NJ, All Others - NY, All Others - PA, Allentown, Buffalo, Camden, NJ, Eastern PA, Harrisburg, Rochester, Syracuse, Trenton, NJ
Tertiary Midwest	Lexington, Louisville, Akron, All Others - IL, IN, OH, Dayton, Kankakee, IL, Lansing Toledo, All Others - Plains States, Des Moines, Omaha, Wichita, All Others - MI,WI, Flint, MI, Grand Rapids, Madison, Milwaukee, Monroe, Racine, All Others - MN
Tertiary Northeast	All Others - CT, All Others - New England, New Haven, Portland, ME, Providence, Southern NH, Worcester
Tertiary Southeast	Polk Co, All Others - FL, Brevard Co, Daytona Beach, Florida Panhandle, Fort Myers, Gainesville, Martin/Saint Lucie, Sarasota, Tallahassee, All Others - GA,NC,SC,TN,KY, Charleston, Chattanooga, Columbia, Greensboro, Greenville, Knoxville, Myrtle Beach, Savannah, Winston-Salem, All Others - AR,LA, All Others - MS,AL, Baton Rouge, Huntsville, AL, Jackson, MS, Little Rock, New Orleans
Tertiary Southwest	Boulder, Greeley, Albuquerque, All Others - Southwest, Tucson, All Others - OK,TX, Corpus Christi, El Paso, McAllen, Oklahoma City, Tulsa, Co Springs, Fort Collins, Provo
Tertiary West	Santa Barbara, Ventura Co, All Others - Northwest, Spokane, Bakersfield All Others - Southern California, Reno, Fresno, Madera, CA, Modesto, Salinas, Vallejo-Fairfield, All Others - Northern California, Santa Rosa, Boise, Napa, All Others - ID,MT,WY; Hawaii, Honolulu

Notes: This table provides the mapping we use from the raw RCA markets to the aggregated RCA markets we use.

Table A.2: NER Information Source

Level	**Office Sector**	**Industrial Sector**	**Retail Sector**	**Apartment Sector**
2-decimal lat/lon	41.9%	35.6%	21.2%	38.7%
1-decimal lat/lon	44.5%	43.8%	55.0%	51.0%
unit-digit lat/lon	9.8%	14.4%	18.3%	9.5%
RCA Market	2.2%	2.9%	3.8%	0.8%
Market	1.3%	2.4%	1.5%	0.0%
National	0.2%	1.0%	0.2%	0.0%
Regression R^2	74.5%	69.2%	45.3%	76.8%

Table A.3: Statistics by Transaction Year

Year	# Trans	% Trans	Cum. % Trans	Volume	% Vol	Cum. % Vol	% Foreclosures	% EntitySales
2001	5,272	1.11	1.11	150.71	1.49	1.49	0.14	6.82
2002	6,287	1.32	2.43	186.11	1.84	3.34	0.06	2.19
2003	7,977	1.68	4.10	223.00	2.21	5.55	0.21	2.96
2004	11,821	2.48	6.59	348.93	3.46	9.00	0.18	7.44
2005	20,725	4.35	10.94	490.05	4.86	13.86	0.15	3.14
2006	23,025	4.84	15.78	555.20	5.50	19.36	0.13	9.39
2007	23,385	4.91	20.69	675.31	6.69	26.05	0.34	21.01
2008	13,531	2.84	23.53	230.60	2.28	28.34	2.70	2.83
2009	6,884	1.45	24.98	109.09	1.08	29.42	17.25	0.38
2010	10,873	2.28	27.26	211.82	2.10	31.52	11.88	0.33
2011	14,749	3.10	30.36	304.41	3.02	34.53	6.64	8.58
2012	19,785	4.16	34.52	374.61	3.71	38.24	4.94	5.74
2013	21,200	4.45	38.97	434.17	4.30	42.54	2.30	5.71
2014	24,974	5.25	44.22	503.24	4.99	47.53	2.27	5.90
2015	29,252	6.15	50.36	616.83	6.11	53.64	0.48	10.14
2016	26,540	5.58	55.94	574.22	5.69	59.33	0.71	4.87
2017	26,996	5.67	61.61	542.27	5.37	64.70	0.73	4.33
2018	29,044	6.10	67.71	611.31	6.06	70.76	0.66	12.52
2019	30,679	6.44	74.16	627.41	6.22	76.98	0.36	1.26
2020	23,793	5.00	79.15	452.95	4.49	81.46	0.48	4.37
2021	40,410	8.49	87.64	854.42	8.47	89.93	0.55	4.27
2022	36,266	7.62	95.26	697.79	6.91	96.84	0.86	8.10
2023	22,554	4.74	100.00	318.54	3.16	100.00	1.68	2.95

Notes: This table shows the number of transactions, the share and cumulative share of total transactions, the volume (dollar value), share and cumulative share of volume by transaction year. The last two columns displays the share of transactions that are foreclosures and the share of transactions that are entity sales. Volume is defined in billions of 2023 U.S. dollars.

Table A.4: Transactions By Transaction Size

	# Trans	% Trans	Cum. % Trans	$ Vol	% Vol	Cum. % Vol
Above 1 Bil	269	0.06	0.06	327	3.24	3.24
500 Mil - 1 Bil	701	0.15	0.20	374	3.71	6.95
250-500 Mil	2,368	0.50	0.70	704	6.97	13.92
100-250 Mil	12,525	2.63	3.33	1,726	17.10	31.02
75-100 Mil	9,301	1.95	5.29	772	7.65	38.68
50-75 Mil	19,926	4.19	9.47	1,181	11.71	50.38
25-50 Mil	52,693	11.07	20.54	1,814	17.97	68.35
20-25 Mil	22,517	4.73	25.27	496	4.91	73.26
15-20 Mil	33,779	7.10	32.37	578	5.72	78.99
10-15 Mil	57,414	12.06	44.43	695	6.89	85.87
5-10 Mil	135,100	28.38	72.81	951	9.42	95.30
Below 5 Mil	129,429	27.19	100.00	474	4.70	100.00

Notes: This table shows the number of transactions, the % of transactions, Value and % of Value for transactions by various transaction size buckets. Dollar volume is defined in terms of Billions of 2023 US Dollars.

Table A.5: Transactions By Major Market

	# Trans	% Trans	$ Vol	% Vol	%A	%I	%O	%R
Manhattan	12,617	2.65	733.41	7.27	26.15	0.85	63.82	9.18
Los Angeles	30,892	6.49	578.30	5.73	29.03	20.25	33.95	16.77
Dallas	18,720	3.93	448.09	4.44	44.79	18.30	24.20	12.72
Chicago	19,060	4.00	405.98	4.02	21.55	24.14	36.34	17.97
Atlanta	15,828	3.33	372.71	3.69	43.72	17.27	24.37	14.64
Houston	12,937	2.72	303.41	3.01	42.57	14.15	28.35	14.92
Boston	8,268	1.74	303.20	3.00	20.36	12.95	57.92	8.78
Seattle	10,744	2.26	279.32	2.77	34.78	14.64	39.27	11.30
Phoenix	13,512	2.84	277.81	2.75	46.14	16.21	22.14	15.51
San Francisco	7,561	1.59	242.48	2.40	21.18	8.48	60.49	9.85
DC VA burbs	5,051	1.06	236.12	2.34	36.26	10.96	42.38	10.40
Northern NJ	10,114	2.12	205.36	2.03	24.83	28.42	32.81	13.94
San Diego	9,332	1.96	199.01	1.97	31.34	19.69	33.75	15.22
San Jose	6,280	1.32	197.36	1.96	17.02	23.26	50.56	9.15
Washington DC	2,395	0.50	147.88	1.47	16.20	1.18	78.14	4.48
Miami	7,239	1.52	142.94	1.42	30.99	19.78	27.49	21.74
All Others	285,472	59.97	5,019.61	49.73	36.85	20.13	21.08	21.94

Notes: This table shows the number of transactions, the % of transactions,volume and share of volume for transactions for the largest 16 markets, ranked by volume, as well as in the remaining 44 markets (last row). The last four columns shows the percentage breakdown of volume by sector (Apartments, Industrial, Office, and Retail) within each market. Volume is expressed in billions of 2023 U.S. dollars. The cities in bold are 11 of the 16 superstar cities. The other five superstar cities that are not explicitly listed in the top of this table are: East Bay, DC MD burbs, Orange Co, NYC Boroughs, and Austin.

Table A.6: Transactions By Transaction Type

Type	Subtype	# Trans	% Trans	Vol	% Vol
Conventional Sale	Ind. Asset	354,712	74.53	7068.52	70.06
	Portfolio	89,256	18.75	2087.41	20.69
Entity Sale	Private Merger	1,852	0.39	61.10	0.61
	Privatization	5,832	1.23	289.44	2.87
	Public Merger	7,549	1.59	274.93	2.73
	Publicization	2,098	0.44	61.28	0.61
Foreclosure	Ind. Asset	7,895	1.66	128.55	1.27
	Portfolio	1,867	0.39	22.48	0.22
Debtor/Trustee Sale	Ind. Asset	3,486	0.73	74.64	0.74
	Portfolio	1,415	0.30	20.50	0.20

Notes: This table shows the number of transactions, the share of transactions,volume and share of volume by transaction type and subtype. Volume is expressed in billions of 2023 U.S. dollars.

Table A.7: Mapping Investor Types

RCA Group	RCA Type	Investor Type	%Trans Buy	% Vol Buy	%For Buy	%Trans Sell	% Vol Sell	%For Sell
CMBS	CMBS	Institutional	0.90	0.71	0.68	0.93	0.62	0.43
Equity Fund	Equity Fund	REPE	6.79	12.54	2.03	5.34	10.43	1.89
Institutional	Bank	Institutional	1.25	1.83	26.65	1.75	2.15	25.37
	Endowment	Institutional	0.09	0.12	0.00	0.09	0.12	0.00
	Finance	Institutional	0.33	0.32	22.93	0.32	0.30	13.16
	Insurance	Institutional	0.84	1.88	25.77	0.97	2.13	14.98
	Investment Manager	Institutional	5.90	11.36	19.58	5.00	8.85	17.31
	Open-Ended Fund	Institutional	0.01	0.08	98.74	0.00	0.02	94.23
	Pension Fund	Institutional	1.12	2.73	33.09	1.01	2.29	21.31
	Sovereign Wealth Fund	Foreign	0.57	0.89	100	0.22	0.30	100
Private	Developer/Owner/Operator	Foreign	0.67	0.98	100	0.45	0.68	100
		DOO_L	29.40	12.96	0.02	32.62	16.33	0.02
		DOO_N	32.07	30.43	0.00	28.24	30.44	0.00
	High Net Worth	Individual	1.81	1.20	22.71	2.97	1.41	7.14
	Non Traded REIT	Individual	2.37	3.19	0.06	1.22	1.64	5.09
Public	Listed Funds	REITS	0.01	0.02	89.01	0.01	0.03	78.36
	REIT	REITS	7.09	12.27	6.66	7.34	12.93	5.60
	REOC	REITS	0.61	0.94	72.49	0.49	1.05	64.98
Unknown	Unknown	Unknown	2.39	1.22	0.83	4.12	2.44	0.27
User/Other	Cooperative	User	0.01	0.01	3.68	0.03	0.04	0.20
	Corporate	User	4.17	3.04	6.49	5.84	4.85	6.59
	Educational	User	0.31	0.34	0.10	0.12	0.16	0.13
	Government	User	0.62	0.49	1.12	0.38	0.34	0.73
	Non-Profit	User	0.48	0.31	1.69	0.39	0.30	7.96
	Other	User	0.06	0.06	4.33	0.03	0.03	0.00
	Religious	User	0.14	0.07	2.10	0.11	0.10	2.54

Notes: The top table shows the mapping from RCA investor group (column 1) and RCA investor type (column 2) into our investor type (column 3). We show the share of buy transactions and buy volume for each type in columns 4 and 5 and the share of sell transactions and sell volume for each type in columns 7 and 8. Columns 6 and 9 show the share of buy and sell transactions that include a foreign entity. Volume is expressed in billions of 2023 U.S. dollars.

Table A.8: Transactions by Number of Investors

	1 Seller	2 Sellers	3 Sellers	4 Sellers
1 Buyer	87.30	5.723	0.584	0.154
2 Buyers	4.73	0.654	0.099	0.020
3 Buyers	0.50	0.076	0.004	0.003
4 Buyers	0.14	0.022	0.001	0.001

	1 Seller	2 Sellers	3 Sellers	4 Sellers
1 Buyer	73.34	10.033	1.270	0.367
2 Buyers	10.31	1.914	0.422	0.118
3 Buyers	1.38	0.424	0.026	0.014
4 Buyers	0.29	0.071	0.003	0.018

Notes: The left panel splits the number of transactions into sixteen groups, by the number of buyers and the number of sellers involved in each transactions. The numbers in the table add up to 100%. The right panel does the same based on transaction volume, expressed in billions of 2023 U.S. dollars.

Table A.9: Joint Ventures

	REPE	Institutional	OOD_L	OOD_N	Individual	REITS	Foreign	User
REPE	2.35							
Institutional	5.03	2.51						
OOD_L	1.22	1.12	5.2					
OOD_N	13.31	14.06	7.45	13.63				
Individual	0.66	1.07	1.0	1.96	0.46			
REITS	1.46	2.75	0.23	2.37	1.09	0.07		
Foreign	1.62	3.41	0.64	6.26	0.72	1.83	4.49	
User	0.09	0.28	0.35	1.01	0.04	0.04	0.04	0.15

	REPE	Institutional	OOD_L	OOD_N	Individual	REITS	Foreign	User
REPE	9.12	16.63	7.10	22.16	9.40	14.84	8.54	4.68
Institutional	19.53	8.29	6.52	23.41	15.31	27.94	17.96	14.16
OOD_L	4.74	3.71	30.24	12.41	14.32	2.28	3.36	17.36
OOD_N	51.70	46.50	43.29	22.70	28.07	24.10	32.94	50.25
Individual	2.55	3.54	5.82	3.27	6.54	11.07	3.77	1.85
REITS	5.68	9.10	1.31	3.95	15.59	0.73	9.62	2.22
Foreign	6.31	11.29	3.71	10.43	10.25	18.58	23.59	1.97
User	0.36	0.94	2.03	1.68	0.53	0.45	0.21	7.51

	REPE	Institutional	OOD_L	OOD_N	Individual	REITS	Foreign	User
% Total Buys	27.34	24.77	4.43	13.99	11.66	10.73	34.16	2.42

Notes: This first panel breaks down the total number of buy transactions involving multiple buyers into who forms a JV with whom. The second panel shows this number as a share of JV property purchases by each investor type; every column sums to 100%. The third panel table shows the share of properties bought in JV with every other investor type, as a share of each investor's total purchases.

Table A.10: Trade Networks

	REPE	Institutional	OOD_L	OOD_N	Individual	REITS	Foreign	User	Unknown
REPE	205.87	241.76	55.18	334.05	44.70	132.03	159.68	28.33	4.71
Institutional	243.07	321.74	82.39	461.55	56.27	192.81	176.05	42.65	7.44
OOD_L	102.32	137.32	590.45	610.10	52.38	83.67	55.94	96.05	27.51
OOD_N	368.63	536.94	395.71	1263.38	152.55	309.66	235.54	111.02	29.29
Individual	27.06	36.34	52.15	82.97	17.98	62.32	31.71	9.40	2.23
REITS	318.93	243.37	39.15	278.62	51.98	324.63	186.06	30.75	8.51
Foreign	145.44	131.80	29.09	164.04	32.83	71.97	123.05	25.37	1.93
User	68.04	81.48	88.76	182.84	22.79	80.48	34.79	74.60	3.90
Unknown	9.67	23.89	48.21	76.17	9.65	27.93	8.43	12.82	37.86

	REPE	Institutional	OOD_L	OOD_N	Individual	REITS	Foreign	User	Unknown
REPE	1.81	2.13	0.49	2.94	0.39	1.16	1.40	0.25	0.04
Institutional	2.14	2.83	0.72	4.06	0.49	1.70	1.55	0.38	0.07
OOD_L	0.90	1.21	5.19	5.37	0.46	0.74	0.49	0.84	0.24
OOD_N	3.24	4.72	3.48	11.11	1.34	2.72	2.07	0.98	0.26
Individual	0.24	0.32	0.46	0.73	0.16	0.55	0.28	0.08	0.02
REITS	2.80	2.14	0.34	2.45	0.46	2.85	1.64	0.27	0.07
Foreign	1.28	1.16	0.26	1.44	0.29	0.63	1.08	0.22	0.02
User	0.60	0.72	0.78	1.61	0.20	0.71	0.31	0.66	0.03
Unknown	0.09	0.21	0.42	0.67	0.08	0.25	0.07	0.11	0.33

Notes: This table summarizes the trade networks among various investor types. We show buyers along the columns and sellers along the rows. Top Panel reports the $ Volume of transaction between different investor types, expressed in billions of 2023 U.S. dollars. If a transaction involves multiple buyers (sellers) of different types, transaction value is split among them proportionately. The bottom table reports the share of overall transaction $ Volume. Values in the entire table sum up to 100%

B Calibration Appendix

B.1 Light GBM Calibration

LightGBM (Guolin Ke, 2017) is a gradient-boosted decision tree algorithm designed for efficiency and scalability. Unlike traditional tree-based models, it uses histogram-based learning to speed up training and leaf-wise tree growth to capture complex patterns in the data. It can capture complex, non-linear relationships without requiring explicit interaction terms, automatically select relevant features and interactions, improving model interpretability and regularize effectively to prevent overfitting.

To optimize LightGBM's performance, we fine-tune hyperparameters separately for each model specfciation using Optuna (Takuya Akiba and Koyam, 2019), a Bayesian optimization framework for hyperparameter tuning. We use 10-fold cross-validation to select the best-performing set of hyperparameters, minimizing mean squared error (MSE) on validation folds. Each model specification undergoes 50 Optuna trials, selecting the best hyperparameters based on out-of-sample performance.

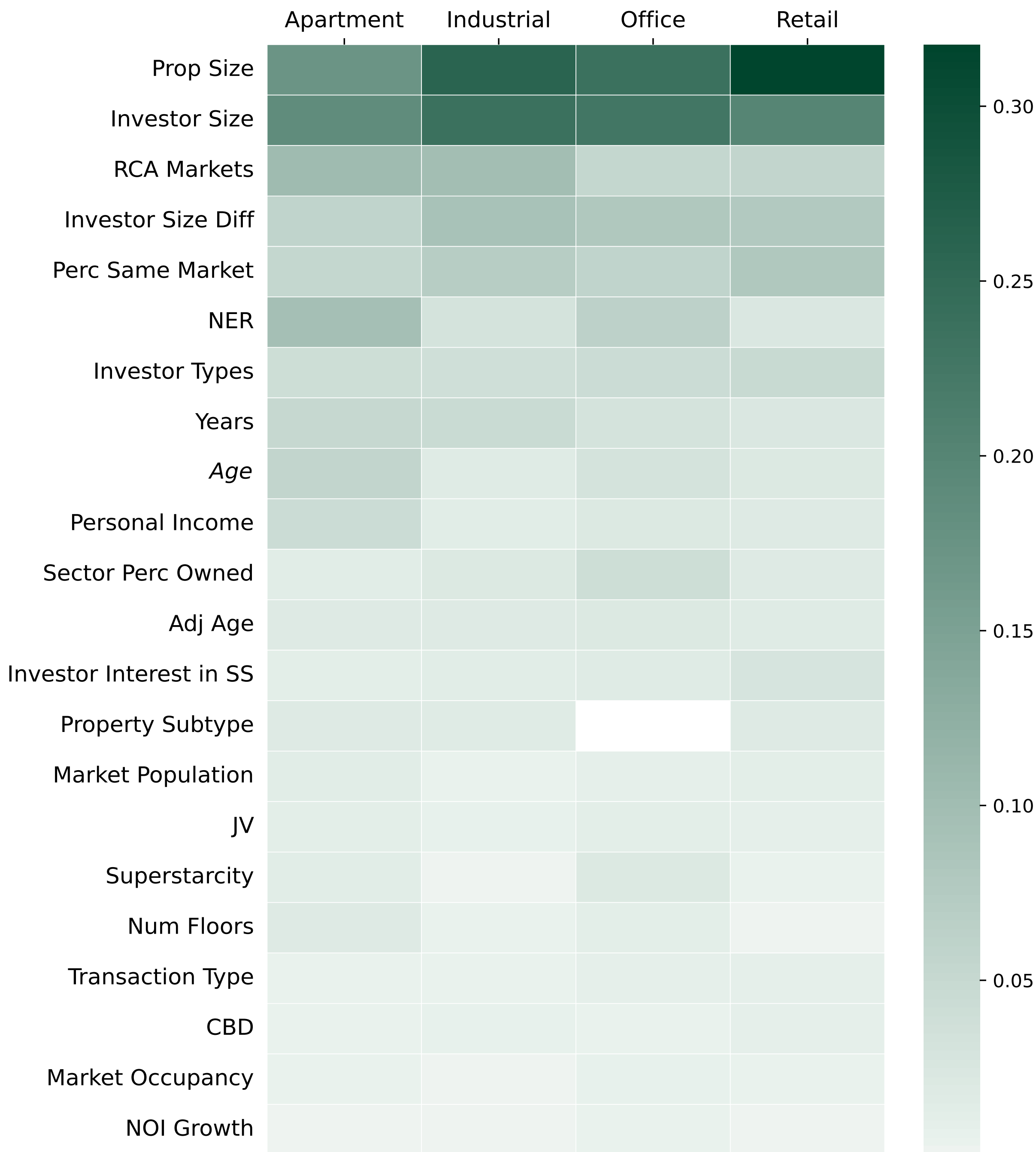
Table B.1 presents the hyperparameter search spaces and optimized values for the final LightGBM models, i.e. those with investor characteristics.

Table B.1: LightGBM Hyperparameters

	Search Space	**Apartment**	**Industrial**	**Office**	**Retail**
Number of Estimators	500 - 1000	960	812	791	525
Maximum Tree Depth	5 - 12	11	8	10	9
Number of Leaves	5 - min($2^{\text{Max Depth}}$, 255)	148	107	164	147
Minimum Child Samples	0.5% - 2% of Obs	1380	673	1158	594
Feature Fraction	0.5 - 1.0	0.641	0.806	0.553	0.617
Bagging Fraction	0.5 - 1.0	0.873	0.826	0.653	0.678
L1 Regularization	0.0 - 5.0	1.950	2.990	1.260	5.000
L2 Regularization	0.0 - 5.0	4.450	3.720	2.060	3.960
Minimum Split Gain	0.0 - 1.0	0.001	0.030	0.022	0.017
Number of Observations		141135	116737	96139	114223
Number of Predictors		29	29	28	29

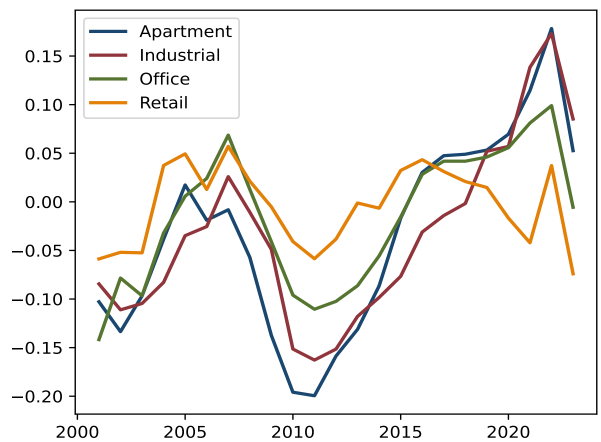
B.2 Valuation Model

Figure B.1: LightGBM Feature Importance



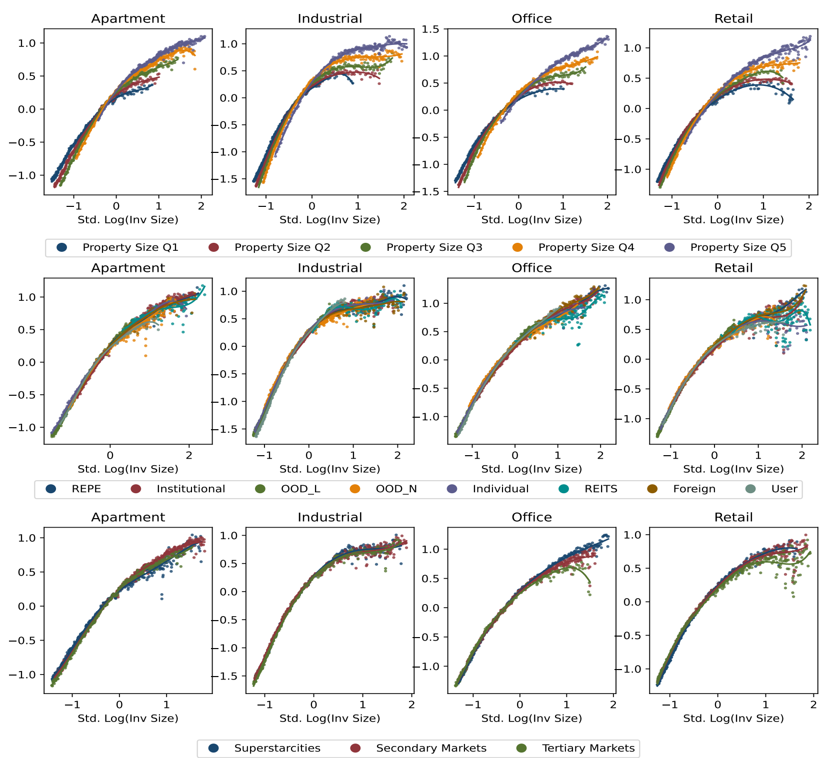
Notes: Figure shows the importance of features selected by the LightGBM Model. Feature importance is defined as the absolute SHAP (Shapley Additive Explanations) values for each feature summed across all observations and is normalized such that the importance for each feature sums to 1. Features are sorted by the average importance across sectors. Investor types are treated as a single group.

Figure B.2: Year Fixed Effects in the Valuation Model



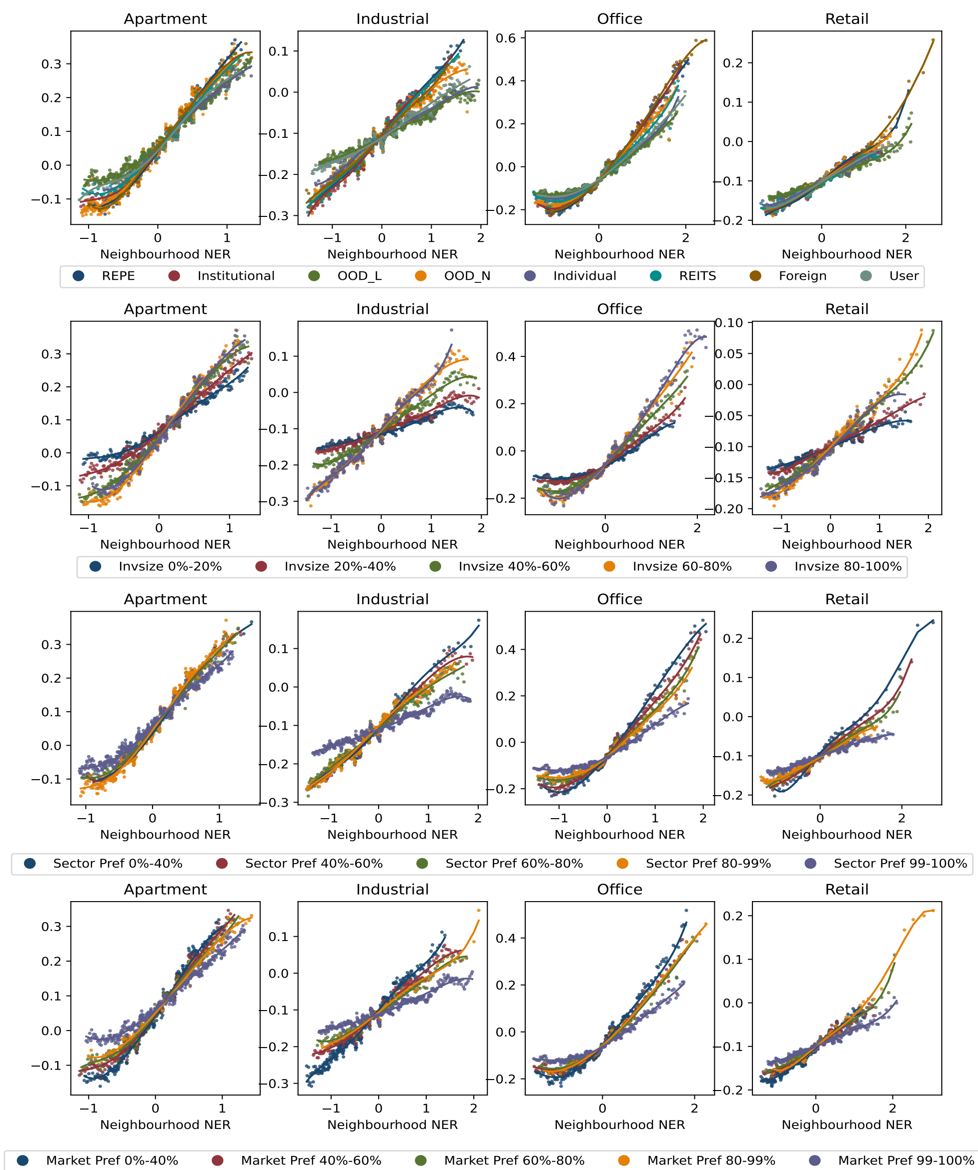
Notes: This figure This figure plots the average SHAP values for the year categorical variable. SHAP values are averaged across buyer and seller models and standardized to ensure comparability across sectors.

Figure B.3: SHAP Dependence: Interactions effects with Investor Size



Notes: This figure shows how the effect of investor size on property values depends on property size (top row), investor type (second row), or market (third row).

Figure B.4: SHAP Dependence: Interaction effects with Net Effective Rent

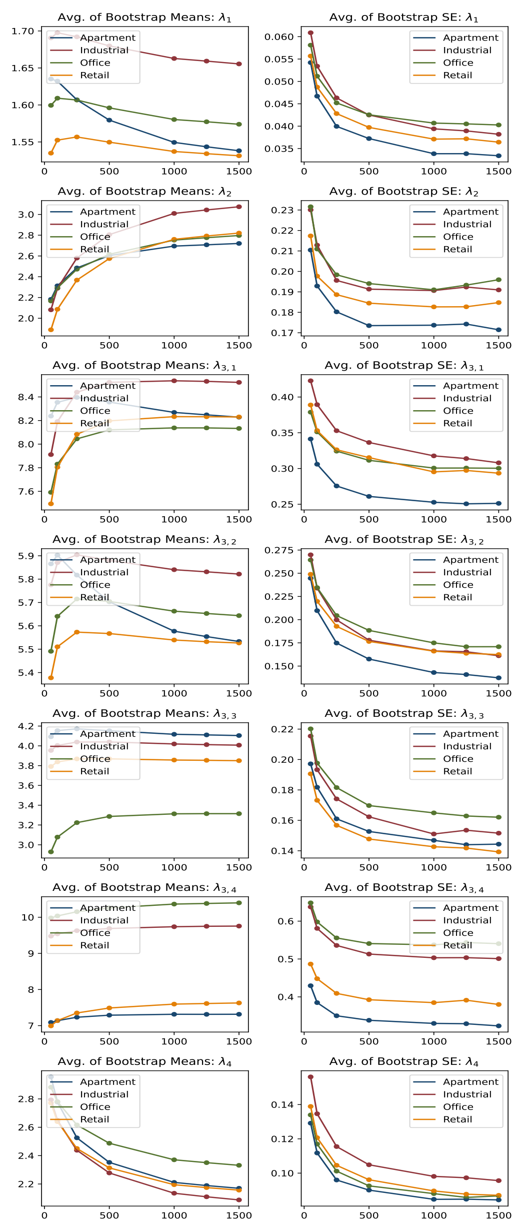


Notes: This figure shows how the effect of NER depends on investor type (top row) or investor size (second row), investor sector concentration (third row) and investor market concentration (fourth row).

B.3 Effect of the Size of Negative Sample

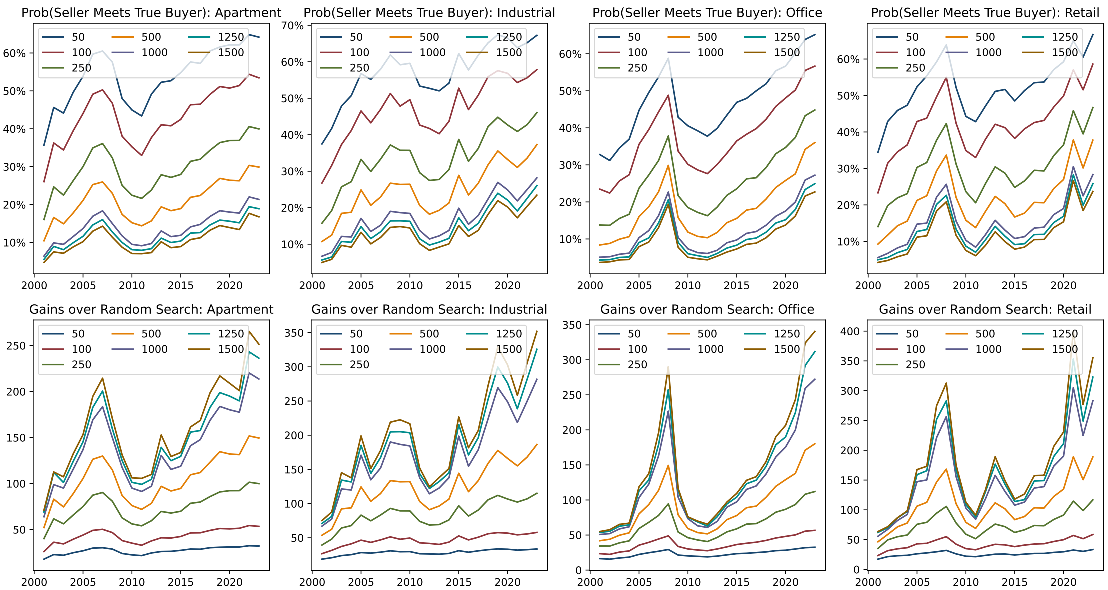
As discussed in Section 3.2, rather than including all potential buyers in the likelihood estimation, we use an important insight from the NLP literature showing it is sufficient to consider a negative set with small cardinality K. In this section, we verify that the parameter values and standard errors stabilize for K = 1,000. We test this by using bootstrap with 1,000 samples for various values of K between 50 and 1,500.

Figure B.5: Bootstrap Parameter Means and Standard Errors



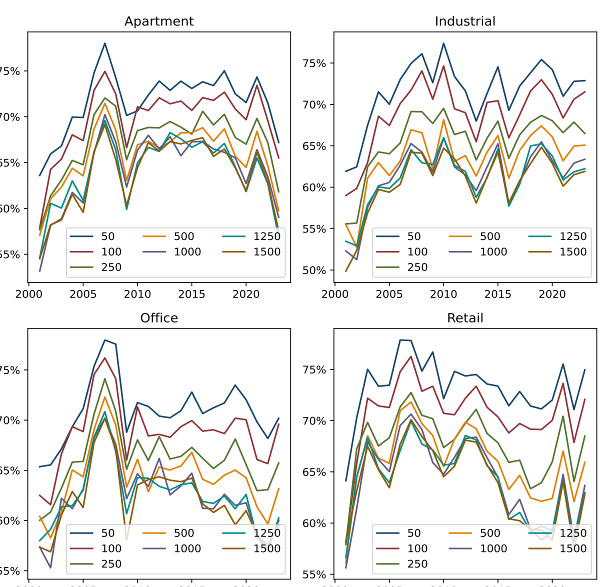
Notes: This figure shows, for each parameter in the matching model, the point estimates (left column) and the standard errors (right column) for varying values of K, the size of the negative sample. The point estimates $\hat{\lambda}_t$ and standard errors are calculated using bootstrap (with 1,000 samples) for each year, and averaged across years. The MLE minimizes the loss function in (6). λ_1 and λ_2 are coefficients of investor size and buyer-seller size difference respectively. $\lambda_{3,1}$ through $\lambda_{3,4}$ are the coefficients of asset, market, sector and NER alignment an λ_3 is the coefficient of the indicator that equals one when the buyer owns > 2 buildings.

Figure B.6: Probability of Meeting the True Buyer by K



Notes: This figure shows, the likelihood that the model identifies the true buyer from among the set of K potential buyers and the true buyer. The raw probabilities for each sector and K are shown in the top Panel. Since the probability of meeting the true buyer under random search varies with the size of K, in the bottom panel, we normalize the raw probability by the probability under random matching. The corresponding ratio thus gives the gains from directed search for various sizes of the negative sample. The incremental gains from increasing K beyond 1,000 are small.

Figure B.7: Probability of Transaction (Conditional on Listing) by K



Notes: This figure shows, the model-implied probability of a transaction occurring with any buyer as we vary the size of the negative sample. This stabilizes around K = 1,000.

B.4 Meeting Model with all investors

While calibrating the meeting model in Section 6, we restrict the negative sample to those investors who have transacted at least once in the previous five years to contain the number if inactive investors. In this robustness check, we relax that assumption.

Table B.2 presents the updated parameter estimates. The coefficients retain the same signs and exhibit comparable magnitudes as in Table 4, reinforcing the robustness of our results to the negative sample restriction.

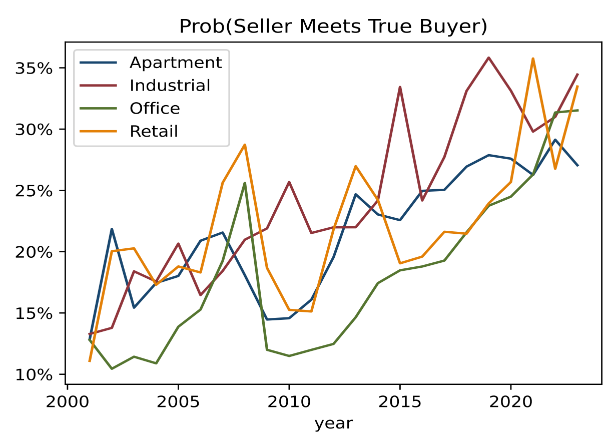
Table B.2: Meeting Model Parameter Estimates

	λ_1	λ_2	$\lambda_{3,1}$	$\lambda_{3,2}$	$\lambda_{3,3}$	$\lambda_{3,4}$	λ_4
Apartment	1.84	2.76	9.27	5.98	4.48	7.88	2.77
	(0.04)	(0.19)	(0.3)	(0.17)	(0.17)	(0.37)	(0.09)
Industrial	1.94	3.07	9.45	6.24	4.15	10.61	2.67
	(0.05)	(0.21)	(0.37)	(0.19)	(0.18)	(0.56)	(0.11)
Office	1.83	2.82	8.97	6.01	3.49	11.15	2.99
	(0.05)	(0.21)	(0.34)	(0.21)	(0.18)	(0.58)	(0.09)
Retail	1.79	2.78	9.02	5.88	4.08	8.2	2.76
	(0.04)	(0.21)	(0.34)	(0.2)	(0.16)	(0.43)	(0.1)

Notes: This table reports the average estimated values of the meeting model parameters along with their standard errors in bracket. The maximum likelihood estimation minimizes the loss function in (6). The parameter interpretation is as in Table 4

Figure B.8 illustrates the model's ability to identify the true buyer from the set of potential buyers under this alternative specification. Since the relaxed negative sample includes a larger share of infrequent traders—who are less likely to transact—the model performs better at matching sellers with actual buyers. The probability of correctly identifying the buyer rises to 25-30%, compared to a little over 20% in the base model, as shown in figure 9).

Figure B.8: Meeting Probabilities



Notes: This figure shows the probability of meeting the actual buyer from among a set of $K = 1,000$ potential buyers (negative set) and the actual buyer, as implied by the meeting model described in Section 6 with the change that the negative sample now includes all investors and not just those that transacted in the previous five years.