Calibrating Permitting Data to Capital Expenditures Data

PRELIMINARY DRAFT FOR THE REAL ESTATE RESEARCH INSTITUTE

Tomasz Piskorski and Jacob S. Sagi

Abstract

Capital expenditures data is critical in accurately calculating commercial real estate (CRE) property return. For example, price indices and related benchmarks that only rely on transaction information may not accurately reflect price appreciation returns. Capital expenditure details are also important in understanding the benefits to investing in various property improvements, in predicting operational risk, and in assessing the impact of changes to a structure on neighboring properties as well as the local economy. Unfortunately, few data sources capture capital expenditures. We explore a statistical solution to these issues by studying the relationship between permitting data, acquired from BuildFax via county-level sources, and known outlays reported in the NCREIF property-level dataset. Our model is able to predict CapEx out of sample and captures significant time-series and cross-sectional variation. We demonstrate the model's utility by applying its out of sample predictions to correcting a repeat sales index which, in the absence of adjustment for capital investment, results in a 2% bias per year in true capital gains.

We gratefully acknowledge financial support from the Real Estate Research Institute and the Kenan Institute for Private Capital. We are also grateful for data from the NCREIF, BuildFax, and RCA, as well as for guidance from Jeff Fisher, Tim Riddiough, and Bob White. Pivotal research assistantship was provided by Susan Cherry, Huan Lian, Sam Liu, Wenting Ma, Jose Ernesto Aldana Vizcaino, and Tomasz Wisniewski. Special thanks in working with the data go to Nick Marcone & Drake Whitehurst at BuildFax, and especially to Willem Vlaming at RCA.
†Graduate School of Business, Columbia University, 3022 Broadway, Uris Hall, NY, NY 10027. Tel: 212.854.4655, email: tp2252@gsb.columbia.edu.
‡Kenan-Flager Business School, UNC Chapel Hill, CB #3490, Chapel Hill, NC 27599-3490. Phone: 919.962.8399, email: sagi@unc.edu.

JEL Classification: R30, G23, G31
Keywords: Commercial real estate, capital expenditure, investment, repeat sales.

1. Introduction

Commercial real estate (CRE) development in the United States amounts to roughly 1% of the existing stock.[1] Anecdotally, annual capital expenditures (CapEx) in CRE is 1%-2% of property market value, suggesting that investment in existing CRE through CapEx is comparable with the production of new inventory.[2] Correspondingly, if investment in *new* supply is deemed important to understanding local economics and the dynamics of urban growth, then it seems plausible to conjecture that investment in existing supply merits equal attention. At a more abstract level, the important connection between "local" investment and economic growth has been studied extensively, from the macro level where "local" refers to different countries (e.g., De Long, et al., 1991; Levine and Renelt,1992), to the city level (Glaeser, et al., 1992). Focusing on real estate, specifically, the Great Financial Crisis provided proof that the economic condition (e.g., neglect) of a single building can have significant externalities on neighboring structures (Harding, et al., 2009; Campbell, et al., 2011). More recently, Liu, et al. (2018), provide evidence that the structure of a building impacts the productivity and industrial organization of its tenants. CapEx, which corresponds to investment in the already-built environment, touches on all of these issues.

Despite their estimated magnitude (in comparison with investment in new supply), little scholarly work is available on capital expenditures. While commercial real estate (CRE) price information is widely available from public and private sources, the same is

[1] This figure is crudely estimated as follows: According to the Federal Reserve Bank (current Z.1 publication, Table H.2), commercial banks have $340B outstanding in commercial real estate construction and land development loans. Assuming that the average construction loan is offered at 65% loan-to-cost and the time to build averages two years, this represents roughly $250B of annual new construction. The Z.1 release estimates the market value of non-financial business real estate holdings at about $25T, thereby yielding the figure cited in the text.

[2] In the subset of properties that we study, average annual CapEx is 1.45% of estimated property value.

not true for capital expenditure data. This poses a major problem for both academics and practitioners who seek to understand the relevance of CapEx to a host of issues, both social and economic. For instance, without the ability to control for investment activity in a property during the time it was held, one cannot assess how well it performed on the dimensions of risk and/or return. Likewise, in viewing asset price indices as an indicator of demand (or a benchmark for performance), it is important to separate increase in value because of increased investment (i.e., supply) versus actual demand for the asset. Price appreciation indices, (e.g., the S&P CoreLogic Case-Shiller Home Price Indices) do not adjust for investment because information about investment is not widely available.

Our goal in this paper is to investigate the degree to which permitting information, which is widely available, can be used to fill the capital expenditures gap for CRE. To do this, we undertake a comparison of permitting data with known capital expenditures data. The former comes from BuildFax (BFX), and the latter from the National Commercial Real Estate Investment Fiduciaries (NCREIF).

There are several concerns to address in using aggregated permit valuations as a proxy for CapEx. First, permit valuation and description data is noisy. Many contractors neglect to provide valuation information on permitted work or fail to provide comprehensible descriptions of the work (enforcement of permit requirements and regulations varies across jurisdiction and time). Moreover, property tax assessments and construction use tax paid by contractors may be linked to the valuation reported on a permit. This provides an incentive for both owners contractor to understate the valuation reported on the permit. In other words, one expects the permit valuation data to be biased.

For instance, in the matched dataset that we use, roughly 55% of permitted work reports a work value of zero (or missing) and roughly one seventh provide no work description.

Although permitting data is notoriously noisy and cost details are likely subject to reporting bias, as long as a statistical relationship exists between the capital expenditures associated with a given property and the permit information submitted by contractors, bias can be corrected and noise will wash out at the portfolio level. Our goal is therefore to investigate whether there are any reliable statistical relationships between permit information submitted for work on NCREIF properties and the capital expenditures reported by NCREIF members for these properties.

We start by matching work permit records from BFX to NCREIF properties and aggregate permitted work into a quarterly panel. Because of sparseness of permit valuation data, we then estimate a “backfilling” model to replace work assigned a zero (or missing) value. To address the ambiguity in how permitted work and reported CapEx might be associated with cash outlays, we smooth quarterly total permit valuations and reported CapEx. Next, after deflating both smoothed permit valuations and CapEx, we assess the statistical relationship between them. The best results we achieve come from estimating a linear mixed model of CapEx that includes permit valuation and nested real estate asset category, location, and year random effects. We then conduct an out of sample analysis and establish that the model is able to fit the CapEx data well in practical applications (national, state-level, and large-fund portfolios).

Although our contribution is largely methodological, it has broad implications for research in a multi-trillion dollar asset class. To demonstrate this, we provide two examples of applications illustrating the promise of incorporating the methodology in both academic and industry research. The first application is to the repeat sales methodology (Calhoun, 1996) popularized by Case and Shiller (1987). Wallace and Meese (1997) and Stanton et al. (2018) criticize this methodology for neglecting property improvements, offering instead a methodology that adjusts for capital improvements via hedonic analysis. Related to this, we first demonstrate that ignoring CapEx leads to gross overestimation of price appreciation that increases at a rate of roughly 2% per year. Correcting for this using our out of sample model-implied CapEx (generated from permitting data) leads to an index that is slightly below the actual CapEx adjusted index, but the two indices are still statistically close. Because data on physical characteristics (needed for hedonic-based adjustments) is not comprehensive, our CapEx prediction methodology opens the possibility of greatly improving on existing CRE index methodologies.

Our second application links to the literature on energy efficient investment. The majority of work in this area typically looks at green investment classifications (e.g., LEED or ENERGY STAR) and price appreciation or income measures, missing the all-important cost required to earn the certification.[3] Our methodology holds promise for filling this gap, but also goes beyond. Specifically, although we do not have certification information, we can still establish the usefulness of the backfilled permit valuation data even without the additional step of tying it directly to known CapEx details or specific energy consumption data. By text-mining permit descriptions, BFX classifies permits into several categories.

[3] The literature on the subject is too vast to summarize here but is reviewed in Zhu et al. (2018). Notable references that are directly related to the application we highlight include Eichholtz, et al., 2010, and Pivo and Fisher, 2010. Contemporary work by Kontokosta, et al. (2019) is also highly relevant. In an example of how one might investigate the relationship between CapEx and property cash flow, Reher (2018) uses CapEx reported in securitized debt underwritings (from Trepp) to investigate the link between property improvements and rental growth rates.

We investigate whether investment in electrical work predicts utility expenditure savings. Even with limited data that does not separate between consumption and unit cost, we are able to estimate an econometric specification that (at the margin) suggests investments made by institutional holders of CRE in electrical work lead to significant savings.

Our methodology for predicting CapEx has implications for several other literature strands. Because the rate of CapEx determines the rate of obsolescence, our work provides a foundation for further study in the closely related literature on CRE depreciation (Bokhari and Geltner, 2018; Yoshida, 2016). Likewise, because CapEx decisions are endogenous, our work can provide an empirical basis for a broad investigation into the theory of optimal investments in CRE (Ambrose and Steiner, 2018). Finally, Sagi (2017) and Giacoletti (2019) demonstrate that controlling for CapEx is critical if one is to use transaction data to gain insights into the real estate price formation process.

The remainder of the paper is organized as follows: Section 2 outlines our approach to combining the two data sources – a non-trivial undertaking and one that will require standardization if the methodology we outline is to be replicated in other contexts. Section 3 summarizes the combined raw datasets and describes our approach to backfilling the permitting data. Section 4 outlines the development of our statistical methodology to predict CapEx using permit valuation data, and assesses the quality of the predictions. Section 5 presents our two applications, and Section 6 provides concluding thoughts.

2. Processing and Merging of the Data Sources

Permitting information comes from BuildFax (BFX), while property-level capital expenditures (CapEx) and other accounting information comes from the National Commercial Real Estate Investment Fiduciaries (NCREIF). We first provide a general description of the data sources and then outline the challenges in combining them as well as our approach to merging.

2.1 *BuildFax*

BFX collects permit information from municipal and county building departments in all 50 United States. Data is primarily covers large population centers, with comprehensive coverage available for every city with population over 50,000 and substantial coverage for smaller cities with population above 25,000. The data extends back, on average, for about two decades but some data may date back as much as 40 years. Roughly half of the BFX dataset is updated on a monthly schedule, and the remainder is updated annually. Although work requiring permits varies across jurisdictions, major changes to the built environment typically require permitting. Anecdotally, much residential renovation work goes unpermitted (in violation of local jurisdiction requirements). Because of large liability and career concerns, we expect unreported work to be much less of a problem in the case of commercial real estate managed and operated by professionals for the benefit of institutional clients.

The permit information collected by BFX importantly includes key dates (such as the application date and work close-date), contractor reported work valuation, and work description. BFX text-mines the permit descriptions to generate a classification of the work into various categories (these include: roof, remodeling, new construction, electrical, mechanical systems, plumbing). BFX uses this data to generate assessments of property condition for its clients (e.g., insurance companies) as well as a residential remodeling index.

2.2 *NCREIF*

NCREIF member firms report property-level accounting, appraisal, and transaction information. They also provide limited descriptive information such as size (square footage) and use (e.g., Office, Apartment, etc.), as well as ownership information (e.g., the managing fund type, joint venture status, and mortgage debt). Operating income/expenses and CapEx are divided into broad categories (see Appendix A). These categories are based on accounting rather than economic guidelines and may overlap. For instance, leasing commission expenditures do not physically change a property and should perhaps not count as true capital expenditures. Moreover, certain “Other expenses” may be better classified as capital expenditures (e.g., when tenant improvements paid by the owner are expensed for accounting purposes). Finally, at least some CapEx is reported on an accrual rather than cash basis. This is evident in cases where CapEx is negative, corresponding to instances where allocated CapEx is not actually invested. Thus even if the permit valuation data was error-free, the correspondence with NCREIF data would not be exact.

2.3 *Challenges to combining the datasets*

The raw data from NCREIF and BFX are not suitable for merging. NCREIF members report quarterly data for each PROP identifier which, while corresponding to a single "property" in the abstract, in practice references a single acquisition potentially containing a number of buildings. In the remainder of this paper, we employ the NCREIF's use of the term "property", which potentially refers to multiple structures. In merging the datasets, our goal is therefore to capture every permit in BFX that is associated with a PROP ID through the PROP's addresse(s). Moreover, because the primary use of NCREIF data is to produce aggregated statistics, there appear to be no strict standards for recording the precise location of structures. Appendix B documents the main difficulties we encountered and our approach to extracting complete address information for each NCREIF PROP identifier.

BFX addresses are relatively clean and it is straight forward to both normalize and geocode them. However, there are tens of millions of addresses and scores of permits associated with each (over 23 billion data points, as BFX currently touts on its website). Because we were not able to obtain the entire BFX dataset we proceeded as follows:

1. Normalizing and geocoding NCREIF addresses (see Appendix B).
2. NCREIF requires confidentiality for it property address and other granular data. To mask NCREIF property addresses, we requested and obtained all permit addresses in the BFX dataset for 5443 zip codes in which *both* NCREIF and Real Capital Analytics have data for more than five properties. This permit address

data consists of over 29 million records, the majority of which are residential.[4] These addresses were likewise normalized and geocoded.

3. The permit and NCREIF addresses were initially matched via an algorithmic sequence of search criteria. To validate our matching methodology, we sent a sanitized set of address data to BFX, which included four RCA properties for each NCREIF property. BFX matched these to their permit addresses and we then compared the results (see discussion at the end of Appendix B), which suggested that our matching algorithm performs well.
4. Addresses do not uniquely identify buildings, and we were concerned that in only matching on addresses we risked missing some permits. Such problems are prevalent in extended properties (e.g., garden style multi-family, office/industrial parks, and retail complexes). The discussion in Appendix C highlights the main difficulty and the solution we implemented. To summarize: Let g_{target} represent the geocode of some target address to be matched with permits. We look for all permits with geocodes within 110 meters of g_{target} whose geocodes pierce the same object (the object being either a legal parcel shape or a building footprint). To provide a sense of the impact this has consider that, when strictly applied, the parcel piercing methodology reduced 198,477 sanitized distinct geocodes (and addresses), to 162,707 distinct parcels.[5] In other words, matching only by address

[4] BFX employs a likelihood scoring measure, provided by Experian, to assess whether an address is residential or commercial. In an initial attempt to restrict the scope of the data request, we discovered that the Experian measure misclassified a significant number of properties.

[5] The number of addresses reported in this example represents a sanitized (i.e., inflated) set that includes both NCREIF building locations as well as (i) all BFX addresses within 110 meters of NCREIF building locations, (ii) a large set of randomly chosen buildings from RCA, and (iii) all BFX addresses within 110 meters of the latter. We sent this large list of addresses to our RAs and Willem Vlaming at RCA and asked them to group all the addresses by building footprint (RAs) or parcels (RCA).

would have resulted in a loss of roughly 22% of the permits (assuming each address in a structure is associated with the same number of permits).[6]

5. Finally, we worked back to the original set of permit addresses from BFX and requested all permit data associated with the sanitized set of addresses. Because of redundancies, the request amounted to over 700,000 permit addresses.[7]

3. Backfilling and Benchmarking Permit Work Value Data

In creating the merged NCREIF-BFX panel, we consider only permits whose "preferred date" is within one year of the dates in which the property appears in the NCREIF dataset. We use this expanded window to account for the ambiguity in the timing of permitted work and the reporting of CapEx. Our matching methodology results in 681,837 permit records matched to 14,025 NCREIF PROP identifiers (i.e., distinct NCREIF properties). Some permit records refer to the same permitted work, so we identify a unique permitted work with a unique combination of the NCREIF PROP identifier and the BFX permitNum. In addition, some redundant PROP-pertmitNum records report different permit valuation amounts.[8] In case of redundancy, we keep the highest recorded permit valuation amount. This results in 622,882 permit records corresponding to 13,993 PROP IDs.[9]

---

[6] We are grateful to Bob White for giving us access to RCA's resources, and to Willem Vlaming for developing and running the parcel parsing methodology on our behalf.

[7] BFX permit addresses are not all normalized, and often include suite or unit numbers as well as the occasional variation on address representation.

[8] This typically happens because multiple forms might be (re)filed in conjunction with a single permit application. In such cases, the contractors might neglect to enter the same valuation amount in each form.

[9] A small number of properties were associated with empty permitNum fields.

Table 3.1 provides summary statistics on the permitting dataset. The dominant property type in the NCREIF dataset corresponds to Industrial properties (I), followed by Office (O), Apartment (A), and Retail (R). Although there are a small number of Hotel (H), land development deals (L), and non-standard property types (X) in the dataset, they constitute less than 10% of the total. A significant number of properties in the dataset do not report a property type.[10]

The lion's share of permits appear to be associated with Office properties, with Retail properties coming at a distant second place. The most common type of permitted work is a remodel (also used to classify alterations), followed by electrical work and building permits. Mechanical (e.g., work on HVAC systems) and plumbing permits are also common. About a fifth of permits are unclassified. The damage repair, solar, and pool permits jointly comprise less than 1% of the total. Many of the permits carry multiple classifications of work (e.g., plumbing and electrical). The total number of classifications is therefore larger than the total number of permits (by about 36%).

Table 3.2 reports the valuation statistics for each permit work type. We treat zero and missing valuations similarly, and replace missing valuations with zero valuation. As mentioned earlier, a great number of permit valuations are listed as zero (or missing).[11] Missing data afflicts some types of work more than others. Plumbing, for instance, appears to have one of the highest rates of missing/zero valuation data.

---

[10] NCREIF members are only required to provide comprehensive details on qualifying properties that are incorporated into the NCREIF Property Index. According to the NCREIF Data Contributor Manual, NPI qualifying properties are stabilized, owned through a tax-exempt vehicle, and can be classified as one of Apartment, Hotel, Industrial, Office, or Retail. It is our understanding that less quality control in data entry is applied to non-qualifying properties.

[11] Either way, when aggregating the valuation data to form a property-quarter panel, missing and zero valuation data result in the same total valuation number.

Table 3.3 reports property-level permit statistics broken out by property type. In the Panel A, we calculate the annual utilization of permits by summing the total number of permits for a given property and dividing by the years that the property is in the panel. Among the major asset types, Office and Retail properties are the most intensive users of permits. The distribution of annual utilization is highly right-skewed which could result from occasional periods of intense investment. Panel B of Table 3.3 documents reported permit valuations, aggregated at the property level. According to the panel, Office leads the major asset classes with an average of $4.5M spent per property (while the property is held by the NCREIF member).[12] Panel C calculates a similar statistic to the second, but using CapEx as reported in the NCREIF dataset. Although it is not universally so, the CapEx valuation statistics are consistently higher than the permit valuation statistic. Panel C also confirms the statement made in Section 2.3 concerning accrual accounting practices and how they may lead to negative recorded values of CapEx. To provide a visual comparison of the two valuations, Figure 3.1 plots the natural logarithm of permit valuation, aggregated at the property level, against the corresponding CapEx calculation. The scatter plot demonstrates that there is a clear and strong relationship between these quantities. The least squares slope is about 0.69 (adjusted R^2 of 39%), confirming that permit valuations are an attenuated proxy for CapEx.

Table 3.4 reports repeats the calculations in Table 3.3, but applied to the type of fund ownership structure rather than the property type. Different fund vehicles have different incentives and may therefore target different quality of assets as well as take a

[12] The primary purpose of these tables is to make a simple comparison of actual recorded CapEx to permit valuations. We do not divide the total amount spent by the property’s time in the dataset because permit valuations data use an expanded window at the property level relative to CapEx data and dividing by years in data will distort the comparison (especially for properties held for only short durations).

different view on how CapEx drives future value. Panel A shows that, in terms of median intensity of permitted work, all fund types are the same. This is slightly less true at higher percentiles (i.e., higher intensity of investment). Major differences across funds show up in the amount spent on properties for permitted work or CapEx, but much of this is driven by the duration of the property in the dataset: Properties owned by open end funds (O) and (D) have a median duration in the dataset of about 3 years, while closed-end funds (C) and segregated funds (S) are roughly 1.5 and 2, respectively. On the other hand, properties held by closed-end funds tend to be slightly smaller than those held by open-end and segregated funds.[13]

To address the large missing number of permit valuations, we estimate a model of permit work valuation with the intention of "filling in" missing or zero valuation data. A sensible model should attempt to control for the fact that work costs per square foot vary by the type of work, the location of the work (due to different labor costs), the property type, and when the work was done (because of price inflation). In addition, the motivations of the property owner might also feed into the intensity of work done (per square foot). For instance, a closed-end private equity fund might be inclined to make more significant changes, measured on a per square foot basis, than other types of owners.

To control for the type of work, we note that BFX may assign a single permit to multiple permit categories (e.g., electrical *and* plumbing). Thus, we estimate work valuation at the level of a permit *combination*. Specifically, for every combination of

[13] Between 2000 and 2016 in our cleaned dataset, a single REIT reported property data.

permit categories containing at least 100 permits, we construct a permit combination group to be employ in our model.

Let *lnPerm* correspond to the natural logarithm of the permit valuation per square foot. We work with the natural logarithm because we expect costs to increase proportionately to some price deflator (which would increase exponentially). We winsorize *lnPerm* at the first and 99th percentiles within every quarter and permit combination group and estimate the following linear mixed-effects model:

$$lnPerm_{i,g} = \beta t + \gamma \boldsymbol{X}_f + \omega_g t + \nu_g \boldsymbol{X}_f + \theta_g + \epsilon_{i,g}. \quad (3.1)$$

In the equation above, t corresponds to the quarter (set to equal one in 1977q4), $\boldsymbol{X}_f$ is a dummy variable that proxies for the owning entity fund type (open-end, segregated, closed-end, etc.), while ω_g, ν_g, and θ_g correspond to group-level random effects. The group is a composite of permit combinations, NCREIF property types, and geographies (counties).

The model in Equation 3.1 estimates the average work cost per square foot associated with a permit combination, taking into account that costs vary across property types, jurisdictions, and time. The model estimates are then used to replace every instance of a zero (or missing) permit valuation for which there is county, property type, and fund type information. This reduces the ratio of zero valuation permits from 55% to 33%.

4. Statistical Modeling of CapEx with Permit Valuations

We collapse the data by property to form a quarterly panel, summing over all permit valuations and counting permits. There are some quarters in which there is no permit information on a property, and (because of the expanded permit window) there are quarters in which there is no NCREIF accounting data.[14]

4.1: Summary Statistics and Full-Sample Regressions

Property Level Summary Statistics

We begin by describing the properties present in the dataset. We initially restrict our analysis to panel observations that are not missing capital expenditure, permit valuation, state, age, property type, and fund type data, which results in nearly 50,000 observations from 7,061 unique properties. We also restrict the dataset to observations from the first quarter of 1995 and after. Data quality before 1995 is poor, with many missing observations. Table 4.1 provides summary statistics for these observations, with the outliers trimmed at the 1% level by county for the permit valuation and capital expenditure variables. The average building is 254,067 square feet and is slightly more than 20 years old. The average reported permit valuation is $62,301 ($0.21 per square foot), with a higher average capital expenditure of $147,638 ($0.55 per square foot). Note that the capital expenditures and permit valuations summarized here are the original values in the data, not the smoothed values that will be used in the following regressions. However, the large discrepancy in average values results from the prevalence of zero permit valuations.[15]

Property Level Linear Regressions

Next, we conduct linear regressions to investigate whether permit valuations from the Buildfax database contain information about CapEx, and what additional property and

[14] We extrapolate NCREIF descriptive information (e.g., SqFt, property type) into the expanded window.
[15] We treat work permits that do not report a work estimate the same way as those that report zero value.

regional information might help increase their predictive power. First, to account for the timing mismatch between permit and CapEx reporting, we smooth permit valuation and capital expenditure at the property level by taking the average of the current value, the next two quarters, and the previous two quarters. Then the smoothed values are deflated by the producer price index. Table 4.2 contains the results from these regressions, which are of the form:

$$CapEx\_SqFt_{i,t} = \alpha + \beta Permit\_Valuation\_SqFt_{i,t} + \mu X_{i,t} + \epsilon_{i,t} \qquad (4.1)$$

where $CapEx\_SqFt_{i,t}$ is the trimmed, smoothed, and deflated capital expenditure per square foot of property *i* in quarter *t*, $Permit\_Valuation\_SqFt_{i,t}$ is the trimmed, smoothed, and deflated permit valuation per square foot of property *i* in quarter *t*, and $X_{i,t}$ is a vector with additional property level and regional controls. To keep the comparison constant across regressions, the regression is performed using the restricted panel used to generate Table 4.1.

In Column (1) of Table 4.2, we simply regress capital expenditure per square foot on permit valuation per square foot. The coefficient on permit valuation is 0.783 and the R-squared is 0.205. In Column (2), we add the number of permits per quarter to the regression. The addition of this variable increases the R-squared slightly and decreases the coefficient on permit valuation. Next, in Column (3), we add building age to the regression. Age increases the R-squared from 0.205 to 0.225 and further decreases the coefficient on permit valuations to 0.674.

We add fixed effects in Columns (4), (5), and (6). The addition of year fixed effects in Column (4) increases the R-squared slightly and the addition of state fixed effects in

Column (5) increases it even more to 0.238. The coefficient on permit valuations increases slightly, but stays fairly stable. In Column (6), we add both property type and fund type fixed effects, leading to a decrease in the coefficient on permit valuations from 0.664 to 0.552 and a substantial increase in explanatory power. It is worth noting that almost all of the additional predictive power comes from adding the property type effects rather than the fund type effects. In Column (7), we add valuation amounts for permit types. These permit types include the valuations for repairs, construction, roofs, electrical, mechanical, plumbing, pools, solar, damage, and uncategorized permits. Adding permit details leaves the R-squared nearly unchanged, but increases the coefficient on permit valuations to 0.754. Finally, we add regional variables in Column (8). Zip code controls include auto sales, auto sales growth, house prices, house price growth, education levels, racial composition, percent married with children, and median age. County level controls include unemployment, unemployment change, mean income, and mean income growth. The addition of these regional controls increases both the R-squared and the coefficient on permit valuations slightly.

Taken together, the regressions in Table 4.2 show that adding additional property and regional variables to the regression of capital expenditure on permit valuations increases the R-Squared from roughly 20% to 30%. These variables also impact the coefficient on permit valuation somewhat, though this increase is not dramatic. Although the details are not reported, removing the permit variables from the specification in Column (8) substantially reduces the explanatory power. Overall this evidence indicates that permit valuation data contains meaningful information about the level of capital expenditures at the property level. Moreover, this information is best supplemented by adding state and property type fixed effects.

Property Level Mixed Model

The preceding analysis points to using parsimonious model that combines permit valuation data with geographic and property type influences. Following these insights, we explore the use of a mixed model to investigate the relationship between permit valuation and capital expenditure. Linear mixed models allow us to specify both random and fixed effects. Our main specification is a mixed model with three levels of clustering: property type, county, and year. This results in a model with three random effect equations – a random intercept at the property type level, random intercepts at the county level nested inside of property type, and random intercepts at the year level nested inside of property type – county level. Formally, the mixed model with random intercepts at the property type, county, and year levels takes the form:

$$CapEx\_SqFt_{it,p,c,y} = \beta Permit\_Valuation\_SqFt_{it,p,c,y} + \gamma_p + \omega_{p,c} + \nu_{p,c,y} + \epsilon_{it,p,c,y} \quad (4.2)$$

where $CapEx\_SqFt_{it,p,c,y}$ is the smoothed and deflated capital expenditure per square foot in quarter *t* of property *i* (which corresponds to a property of type *p* in county *c* and in year *y*), and $Permit\_Valuation\_SqFt_{it,p,c,y}$ is the smoothed and deflated permit valuation per square foot of the same property. γ_p is the random intercept of property type p, $\omega_{p,c}$ is the random effect of county *c* within property type *p*, and $\nu_{p,c,y}$ is the random effect of year *y*, within county *c*, within property type *p*.

We estimate the model on the unrestricted sample. Table 4.3 shows the results from estimating various mixed models on the full dataset. Column (1) contains the estimate from regressing capital expenditure on permit valuation with no levels, which is equivalent to simple linear regression. Column (2) adds property type and county levels, with county nested inside of property types. The coefficient on permit valuation decreases slightly from 0.794 to 0.645. Finally, Column (3) adds year levels, which are nested inside of county. The coefficient on permit valuations does not change substantially. We also report the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for the three models. The AIC and BIC are both penalized-likelihood criteria that are used to evaluate models, with lower AIC and BIC preferred. Table 4.3 shows that both the AIC and BIC decrease as we add levels, with the lowest values occurring in Column (3). We also report the root mean squared error (RMSE) of the difference between the actual capital expenditure values and those predicted by the mixed models. Like the AIC and BIC, the RMSE decreases across columns, with the lowest RMSE corresponding to the model with three levels.

The specification in Column (3) with three levels of clustering is our preferred specification and the coefficient on permit valuation is similar to the coefficients reported from the simple linear regressions (on a restricted subsample) in Table 4.2.

4.2: Out of Sample Predictions

Next, we analyze how well our regression models predict capital expenditure out of sample. First, we randomly split the dataset in half, using half of the data to estimate the model and the remaining half to predict capital expenditure using those estimates. We use the mixed model described in Equation 4.2 to generate the capital expenditure predictions.

As above, variables are trimmed at the 1% level by county and smoothed at the property/quarterly level.

Property Level Out of Sample Predictions

First, we examine the out of sample information content of our model. Table 4.4 reports regressions of actual capital expenditure per square foot on predicted capital expenditure and actual permit valuation per square foot. The coefficient in the regression reflects the sensitivity of actual CapEx to the predictor, while the regression R-squared (or regression RMSE) is a measure of how informative the predictor is about CapEx. Column (1) shows that the coefficient on predicted capital expenditure is 1.012 (i.e., they move one-for-one). The coefficient on permit valuation is much lower at 0.664. The regression R-squared is also higher in Column (1) (0.309 versus 0.241). From these regressions we conclude that model-predicted CapEx (which uses an adjusted permit value as well as non-permit data) is more informative about actual CapEx than the raw permit valuation.

In applications, one would proxy for CapEx using either the model prediction or the unadjusted permit valuation. The "Predictive Accuracy RMSE" row in Table 4 reports the root mean square error for using such proxies. For instance, if one used the model prediction to proxy for CapEx per square foot, then the RMSE of the *difference* between actual capital expenditure and predicted capital expenditure would be 0.882 (Column 1). If the proxy used were permit valuations then the RMSE would increase by 22% to 1.073 (Column 2). Note that the predictive accuracy in Column (1) is close to the in-sample RMSE from Column (3) of Table 4.3. This provides further evidence that model-predicted capital expenditures are better proxies for actual capital expenditure than unadjusted permit valuations.

State Level Out of Sample Predictions

Next, we look at whether our findings hold once we aggregate to the state level. We average actual capital expenditure per square foot, predicted capital expenditure per square foot, and permit valuation per square foot to the state level. This results in one observation per state each quarter. We repeat the analysis in Table 4.4 using the state-level averages. The results are reported in Table 4.5, which shows that actual CapEx is close to but less sensitive to predicted capital expenditure than it was at the property level. The coefficient on permit valuations is also slightly lower than it was in Table 4, but is very similar. The R-squared is higher in Table 4.5 than it was in Table 4.4 for both columns, but now the R-squared is nearly identical across columns. In other words, upon averaging at the state level, a lot of the noise in individual property permits washes out and there is comparable information content in the model predictions and the raw permit data.

As mentioned previously, in applications one would simply use the model prediction or the raw permit valuation as a proxy. The last row in Table 4.5 demonstrates that the RMSE associated with using the model prediction as a proxy for CapEx is substantially lower than the RMSE associated with using the raw permit valuation as a proxy. This further reinforces our prior inference that our out of sample model predictions better reflect actual capital expenditure than relying only on permit valuation data.

In order to visualize these results, we plot actual versus model-predicted capital expenditure per square foot for the most represented states in our data (in terms of number of properties). Figure 4.1 shows these plots. Actual and predicted capital expenditure match quite well for most states, particularly Georgia, Texas, and Colorado. Predictions also tend to be more accurate in recent years, especially from 2005 onwards. It is worth noting that predictions appear more accurate in some states than in others. In particular, Figure 1 shows that predictions in Arizona and Virginia are less accurate than other states.[16] Overall though, the model seems capable of capturing the different cyclical patterns

State-Property Type Level Out of Sample Predictions

Next, we aggregate to the state-property type level instead of the state level. We average from the property level to the state-property type level so that each state has an observation for each property type per quarter. Table 4.6 shows the results. Here we find that, once again, the mixed-model predictions contain more information than the valuation data alone. Consistent with that, the last row of Table 4.6 demonstrates that using the model predictions as a proxy for CapEx is superior to using only the permit valuation.

4.3: Fund Level Out of Sample Predictions

Finally, we test whether our model performs well when applied to fund portfolios. We investigate performance on the four funds with the largest numbers of properties in our dataset. For each fund, we remove all properties that belong to that fund and estimate a model to predict capital expenditure using all other funds. Then, we use the estimated

[16] "Virginia" includes some properties that are in the DC MSA.

model to predict capital expenditure per square foot for the properties that belong to that specific fund.

Two models are used to predict capital expenditure per square foot out of sample: a simple linear regression and the mixed model used earlier (Equation 4.2). The linear regression is of the form

$$CapEx\_SqFt_{i,t} = \alpha + \beta Permit\_Valuation\_SqFt_{i,t} + \mu X_{i,t} + \epsilon_{i,t} \qquad (4.3)$$

where $X_{i,t}$ is a vector of controls that includes building age, time since acquisition, state fixed effects, and time fixed effects.

Table 4.7 examines the information content in each predictor (via regression analysis), as before, and assesses its predictive accuracy as a proxy for CapEx. The results vary widely across funds. In terms of information content, measured using the regression R-squared or regression RMSE, the mixed model outperforms all of the others. As a proxy for CapEx, the mixed model predictions are the most accurate for the largest three funds (Panels A-C) and are very close to being the most accurate for the fourth largest fund (Panel D). Moreover, in *every* case, relying only on permit valuation as a proxy for CapEx yields poor results.

To visualize these results, Figure 4.2 plots actual versus predicted capital expenditure per square foot for the four funds, where the predicted capital expenditure is from the mixed model. The accuracy of the predictions varies substantially across funds. Panel B shows that the predictions appear quite accurate for the second largest fund, but less accurate for the other funds. The figure also shows that predictions tend to become more accurate over time, with higher accuracy from 2005 onwards.

Taken together, Table 4.7 and Figure 4.2 show that our model does not always perform well when applied to fund portfolios. Still, it performs better than the simpler alternatives and appears to capture some of the major cyclical fund-specific trends.

5. Applications

5.1 Application to Repeat Sales Price Appreciation Index Construction

There are not enough transactions each quarter in the NCREIF dataset to create a reasonable repeat-price index. However, because the NCREIF dataset provides market value appraisals at the quarterly level, one can create synthetic non-overlapping holding periods that can then be treated as "repeat sales". To construct these, for each property time-series we first look for all contiguous "spells" where data needed for the experiment is available (i.e., non-outlier market values, permit valuations, and capital expenditures). Further, we only consider: properties with a non-negative total CapEx aggregated across their history in the NCREIF dataset; properties with age information, consistent County information, and held for at least four quarters. This results in about 8,000 contiguous spells of quarterly data, each of which we use to create a synthetic repeat sale pair (a purchase and a sale pair). The average spell is 4.5 years in duration (median of 3.25 years).

Next, we split randomly the sample of spells into two parts and estimate the CapEx imputation model on one part, and predict CapEx (yhat) for the other. We then calculate the total discounted value of all CapEx expenditures for a given spell, using the compounded 3m TBill as the discount rate. We do the same for the out of sample (OOS) imputed CapEx. Figure 5.1 shows a log-log scatter plot.

The next step is to calculate the adjusted capital appreciation for the OOS properties, using both actual and imputed CapEx. This is calculated as follows for each property:

$$v = \ln\left(\frac{lastMV}{firstMV + discTotCapEx}\right) \tag{5.4}$$

A Kolmogorov-Smirnov test confirms that the two distributions cannot be distinguished (p=0.452). If capital expenditures are ignored, the corresponding "unadjusted distribution" is significantly different (see Table 5.1 and Figure 5.2). Table 5.1 also indicates that the capital gains calculated by using actual versus model-imputed CapEx are very close (a correlation of 99.6% between the second and third columns of the bottom table). The correlations between the adjusted and unadjusted returns, by contrast, are close to 70%. Finally, Figure 5.3 provides a scatter plot of the OOS predicted property gains versus actual gains, confirming that the two are similar.

Next, using the Calhoun (1996) four-stage OLS methodology, we calculate a quarterly price appreciation index for the OOS properties. Figure 5.4 plots the index using actual CapEx versus the index unadjusted for CapEx. Standard deviations error bars (not 95% confidence intervals) are shown for each point. It is clear that failing to adjust substantially distorts the index, and the distortion becomes significant after roughly ten years.

The corresponding graph in Figure 5.5 compares the index with actual CapEx versus the index with model imputed CapEx. The return correlation between the actual versus imputed CapEx price gain indices is 93%. After 2000, the indices are never more than seven points apart and at no point are the two indices further apart than their combined error bars. The index formed using actual CapEx is nearly always larger than the one formed using imputed CapEx. This is potentially because the distribution of actual CapEx is "fatter" at the extremities. Note that the equation for capital gains is convex in CapEx so, similar to the case with Jensen's inequality, averaging will lead to larger gains.

4.2 Application to estimating the impact of CapEx on future cash flow

An important question for CRE professional and policy makers is the extent to which capital investment impacts energy consumption. We cannot provide a comprehensive analysis using the data we have, but we are able to demonstrate that one can, in principle, use permitting valuation data to address interesting questions.

From the permit valuation data, we focus on permits flagged as "electrical". Many permits are flagged for multiple purposes (e.g., electrical and plumbing). For each such permit, we consider the work valuation to be equally divided among the different purposes (e.g., For a $1M work value permit flagged as both electrical and plumbing, we associate $0.5M with electrical work). We then create a smoothed version of electrical CapEx (per square foot) by quarter and property, eliminating outliers (as usual). For each property, we create a corresponding smoothed quarterly variable from the NCREIF's quarterly utility expenditures, and divide by square footage (likewise trimmed). Finally, temperature data for each of the largest NCREIF MSAs (by number of properties) is used to create an extreme temperature indicator. Specifically, temperature is considered extreme if the average for a given quarter and MSA is above 80 or below 40 degrees Fahrenheit (it is at these temperature extremes that we expect energy would be used most intensively).[17] Table

[17] Using mechanical permit valuations (often associated with HVAC work) instead of or together with electrical permit valuations leads to qualitatively similar results.

5.3 reports the results of various regressions exploring the relationship between reported electrical work value and utility expenditures (at the property level). The variables are defined as follows:

UtilExp: Smoothed utility expenditures (per square foot) for a property-quarter combination.

TShock: Extreme temperature indicator described above.

L.ElVal: Smoothed electrical work valuation (per square foot) from permit data, lagged one year.

CRent: An indicator variable equal to one whenever current rents equal to rents from the previous year.

AvgUtil: The average of UtilExp within an MSA for that quarter

The TShock, CRent, and AvgUtil variables are de-meaned in order to clarify stand-alone versus interaction effects in the regression. The first regression indicates that, unconditionally, temperature extremes are associated with higher utility expenses. The second and third regressions indicate that, contrary to intuition, utility expenditures weakly increase with investment and create no savings during periods of temperature extremes. This can be interpreted as evidence that serious endogeneity issues abound. Some potential sources of endogeneity are: (i) Capital investment may anticipate increasing costs (the “anticipation effect”); (ii) Investing firms may elect to consume more energy, with the difference made up by higher rental revenues (e.g., more tenants). It may be straight forward to control for expectations of rising utility costs, but a fundamental driver behind the second type of endogeneity is the fact that utility expenditures correspond to unit cost multiplied by consumption. Consumption, unlike unit cost, is a choice variable (hence the source of endogeneity). In the ideal empirical experiment, one would have access to energy consumption data as well and be able to separate this effect. The closest we can come to that is to look for instances where there is less motivation by an owner to change consumption, holding constant the level of temperature. For instance, if an owner invests in electrical updates to attract new tenants, one might anticipate seeing higher energy consumption (and therefore utility expenditures) in the following year. To control for this, one can single out properties whose rental revenues are fixed. Under such situation, as the example suggests, consumption might be more likely to depend only on exogenous factors such as temperature.

Column (4) of Table 5.3 includes AvgUtil and its interaction with TShock as an explanatory variable to control for overall increases in utility costs. Because, presumably, everyone is able to anticipate the rise in costs, this controls for the anticipation effect. In addition to the interaction between L.ElVal and TShock, the explanatory variables also include an additional level of interaction with CRent. As suspected, the average effects introduced by the control variable AvgUtil soak up a great deal of explanatory power. Because AvgUtil is de-meaned, TShock and L.ElVal maintain the same impact on utility expenditures as in Columns (1) through (3), although L.ElVal is no longer significant. The latter suggests that the new controls soak up some of the factors leading to endogeneity effects. After removing the average effect, which (hopefully) controls for anticipated investments ahead of rising utility costs, it now seems that the interaction between L.ElVal and TShock is significant and negative (as might be expected). Properties with unchanging revenues exhibit lower utility expenditures than other properties, confirming the intuition that such properties are associated with less increase in consumption. Moreover, constant-rent properties do not spend significantly more than at other times during periods of temperature extremes. The largest (and most significant) impact of L.ElVal is expressed through the interaction with temperature extremes and constant rent properties (highlighted). Consistent with the narrative that they do not increase consumption beyond what they would if they did not invest, constant rent properties appear to benefit enormously from investment in electrical work during temperature extremes. According to the estimated model in Column (4), a one standard deviation in such investment (about five cents per square foot) leads to relative savings of anywhere from \$0.01 to \$0.035 per square foot in the following year if extreme temperatures are realized (roughly, a 20% chance).

Of course, it is possible that such savings cannot be realized by anyone that invests. Investment is itself a choice, and it is sensible to assume that it is undertaken by those who will realize the greatest payoffs from investment. To drill into this question better, one must obtain data in cases where investment was mandatory (due, for instance, to changing codes or regulations). Again, it is important to stress that the analysis here is only illustrative of what can be done using permit data.

6. Conclusions

The preceding analysis provides robust evidence that it is possible to link permitting information to actual CapEx. The methodology, while involved, appears to do a good job out of sample and can be used in a variety of applications (e.g., adjusting price appreciation indices for investment). At this point, the study provides a “proof of concept”. To proceed towards a standardized approach, additional work remains. Some obvious next steps include: (1) Check to see whether closer attention to partial sales in the data improves accuracy of methodology (there roughly 800 partial sales in the dataset). (2) Compare the accuracy of the methodology with an application restricting attention to single structure properties, and across various measures of “matching confidence” that we produced in the course of matching the permit data to NCREIF property data. (3) Experiment with various definitions of CapEx, in light of the discussion in Section 2.2. (4) Experiment with an alternative CapEx prediction model that employs a "rolling" methodology to deal with variation in model accuracy through time. (5) Testing the model performance across different measures of property “quality”.

Appendix A: NCREIF Income/Expense/CapEx Categories

Table A.1 shows operating income, expense, and CapEx categories, as well as short explanations for the categories.

Table A.1: NCREIF Income/Expense/CapEx Categories

Income Category	Explanation	Expense Category	Explanation	CapEx Category	Explanation
Inc_BRent	Base Rent	Exp_Admin	General and Administrative	CapEx_AddAcqCost	Acquisition costs (paid post-acquisition)
Inc_Cntgnt	Contingent Rents	Exp_Mktng	Marketing/Advertising	CapEx_LeaseComm	Leasing Commissions
Inc_Reimb	Expense Reimbursements	Exp_Util	Utilities	CapEx_TI	Tenant Imrovements
Inc_Other	Other Income	Exp_Mnt	Repair and Maintenance	CapEx_BldImp	Building Improvements
		Exp_Insur	Insurance	CapEx_BldExpan	Building Expansions
		Exp_MgtFee	Property Management Fees	CapEx_Other	Other Improvements
		Exp_Tax	Taxes		
		Exp_Other	Other expenses		
		Exp_Payrol	Payroll and benefits		
		Exp_ProFee	Professional fees		

Appendix B: Parsing NCREIF Address Records

NOTE: This document has been "sanitized". All addresses and identifiers are fictitious.

NCREIF data is structured around the "PROP" identifier. Each prop ID is used to track a "property" owned by one of the NCREIF members, but only while that member owns that property. A property could be a single building or a portfolio of buildings. Each quarter, member firms enter accounting, economic, and attribute data for each of their properties.

Once a property is sold, it is no longer tracked under the same prop ID, and will only continue to be tracked (under a different prop ID) if it is sold to another NCREIF member. In particular, there is no direct way to tell that a property has been sold to another member other than by a (potentially fuzzy) matching of address and other characteristics.

The following briefly describes the strategy for matching NCREIF property data with BuildFax data.

1. Each NCREIF prop ID is associated with a
    a. A list of address ranges (a range could be a singleton), and
    b. A range of dates, corresponding to the period of time during which the property was owned by some specific NCREIF member.

   Some addresses are accurate, some are not. In particular, it is possible that partial sales will change the list corresponding to a property's addresses.
2. A BuildFax (BFX) permit will be associated with a single address and a date (the date will reflect a best guess for when the work was done).
3. A permit should be matched with a prop if the BFX address of that permit is in one of the ranges on the prop's list of ranges, and the BFX date is within the prop date range, plus/minus one year.

Implementation of the strategy above first requires that, for every prop ID, we generate a valid address range and date range where possible. Generating the date range is easy and corresponds to the start- and end-dates of the property in the dataset. Generating the address range is more challenging. This is because property identifying attributes (like address fields) are entered every quarter and are sometimes missing or not entered consistently (i.e., they can vary from quarter to quarter). Moreover, in relatively rare cases, the property attributes actually change (e.g., a building from an address range is sold), which should trigger a change in addresses associated with the property.

Description of NCREIF Address Data Fields

The table below documents key location and date fields in the raw dataset supplied by NCREIF (extracted to include data up to and including 2017Q2). For each field, the table lists the number of properties for which there is data for that field, and the number of properties for which there is conflicting information about that field across time. Missing data in a given quarter is not counted when determining whether there are "conflicts" with non-missing data.

Table B.1: Location and Date Fields

This table documents key location and date fields in the NCREIF dataset. The number of properties and the number of properties with conflicting information over time are listed for each variable.

Variable	Number of Properties	Number of inconsistent Properties
prop	36236	
address1	31926	4675
address2	1715	126
propertyname	33515	6063
propertytype	36236	1069
city	36236	41
zip	36236	85
cbsa	36150	18
state	36236	12
division	36236	9
county	36236	27
msa	35809	17
startdate	36236	NA
enddate	36236	NA

Below are detailed descriptions of each field and (partial) documentations of the source of inconsistencies.

prop

This is the identifier described earlier. Each prop ID is linked to a series of separate records, and each record corresponds to a calendar quarter during which the property is owned by the NCREIF member.

address1

This is the primary street address field. It is missing entirely for some properties. For others, it is missing key information like a street number (e.g., an intersection such as "Kansas & North 17th Avenue"). Moreover, it can feature the addresses of more than one building. Some examples follow (the actual prop IDs and addresses have been altered):

prop	address1
224968	4335 Noclue Circle
841133	2354-2394 Blue Ave.
200608	1400-08 Strive Drive & 1305 Mostly Highway
744	100 and 180 Pinball Way
527540	1799-1813, 1827-1855 Northside Drive
263851	18, 24 and 26 Maryjane Road
72382	1714 Deer Valley Trail/12801 Pine Woods
417514	10850-10862/10876-10880 Delicate Court
812292	5909-5929-5959-6101-6121 Painters Road

Inconsistencies in address1 within the records of a single prop ID can arise from spelling variations/mistakes, variable uses of white spaces, and multiple location referencing. Below are some examples of address1 records grouped by prop ID and corresponding to different inconsistencies.

prop	address1
501103	1650-1750 S. Babcock Avenue
501103	1650 - 1750 S. Babcock Avenue
193439	1417 Seventh Avenue
193439	1501 Seventh Avenue
584876	Bienvenue Road
584876	1610-1650 Bienvenue Road
669416	200 East Main Street
669416	170 East Main St
208317	1100 South Paw Street
208317	110 South Paw Street
213707	3280-3294 E.29th Street
213707	3250-3294 E 29TH ST
213707	3250-3294 East 29th Street
215665	12836 Aloevera Blvd.
215665	See Notes below
519372	1800 Mill Park East
519372	1800 - 1840 Mill Park East
291749	8400 8500 and 8550 Barb Blvd
291749	8550 Barb Boulevard
291749	8400, 8500 & 8550 Barb Blvd
845120	Riverside Parkway
845120	Riverside Parkwat
845120	Riverside Parwat
521337	308 SW Monty
521337	RIVER SQUARE (II)
521337	0308 SW Monty St.
521337	308 SW Monty St.
521337	2083 SW Riverside Drive

631535	100 Crembo Town Center
631535	770 Tamarind Drive

In the case of prop 669416 and prop 631535, the two references are to the same location (e.g., a block-sized building might be associated with multiple addresses on the same block, and a corner building can have two different street addresses). In the case of prop 193439, the locations are close but correspond to buildings in distinct blocks. In case of prop 215665, the entry “See Notes below” was likely a data-entry mistake. Entry mistakes were also likely with props 208317, 845120, and 521337. Finally, in the case of props 213707, 519372 and 291749, the address1 field referred to different ranges of addresses inconsistently.

address2
This is a secondary address field and it is sparsely populated. When it is populated, the additional information it contains typically adds nothing to what is already in address1 or propertyname. There are some exceptions, though. On a rare occasion, this is the only field containing a complete street address. In addition, this field often contains alternative street addresses that are not part of the street address information in the address1 or propertyname. Some examples:

prop	propertyname	address1	address2
233033	OVC San Juan LLC	3800 Richland Avenue	3651 Westeros Road
184915	Orange County Industrial Portfolio	"847-853 South Lamb Drive, 179-195 National	"250-260, 268-278, 310 Windy Pipe"

Thus, address2 may be used to expand an address range for a property.

propertyname
This field is much better populated than address2, but it also features the greatest number of inconsistencies (i.e., exhibits the most variation within prop IDs). As with address2, it can sometimes be used to supplement missing information in address1 or expand an address range. In addition, the information in this field can potentially be fed to a geocoder together with an address to help pinpoint an address (e.g., Empire State Building). Some examples of its usefulness:

prop	propertyname	address1	address2
253424	Bigshot Tower	Main & 11th Street	
373676	Zealot Center	I-65 & Alabama	
184264	2630 Comp Industrial		
121979	BS Mall, 7667 Ridi Blvd.	4200 Brown Boulevard	
632744	6250 - 6270 Cow Blvd.	6270 Cow Blvd.	

Some caveats to using this field: In the fourth example of the table above, the propertyname entry looks like a street address but it does not correspond to any known address in the vicinity of address1. In addition, it is important to note that the names of buildings and landmarks change (sometimes with ownership), thus using the propertyname together with the street address from address1 (or address2) in a geocoder might actually “confuse” the geocoder algorithm and lead to a null response. There may therefore be some value in submitting the geocoding request both with *and* without the propertyname (at least in some cases).

The propertyname field occasionally contains parenthetical remarks like “(wholly-owned)”, “(Partnership)”, “(Sold)”, and others (we found 393 distinct examples). This parenthetical information is not always useful, so a case could be made for editing it out before usage.

propertytype
A property-type field (e.g., Office, Retail, Apartment, etc.). In principle, property classification can change (about 1% of cases) because of dual or alternative use, or because it reflects a blend of buildings in a portfolio (which makes the property type classification ambiguous).

city and **zip**
These are the city and zip code fields, respectively. There are a very small number of properties that inconsistently refer to the property’s city or zip code. The inconsistencies in the zip field likely arise (for the most part) from changes to city and zip code boundary definitions. The same for the other spatial categories documented below.

cbsa – The name of the city-state in which the property is located, using the US government standardization.
state – Name of the state in which the property is located.
Division - A US government numerical code for the “census division” in which the property is located.
county – Name of the county in which the property is located.
enddate – The date of the last record for that prop in the dataset.
Msa - A numerical code for the municipal statistical area of the property’s location.
startdate – The date of the first record for that prop in the dataset.

Finding Valid Address Ranges
This is a necessary element of the matching strategy. Because from quarter to quarter, a given property may be identified differently, we collapse the data to reflect *unique* combinations of the following fields for each property: address1, address2, propertyname, zip, and city. Each of these combinations is henceforth referred to as a “combo”, and the calculated field “combos” contains the number of unique combo instances per property:

Table B.2: Number of Combinations for Finding Valid Address Ranges

This table shows the number of properties that have a certain number of combos. The left column contains the number of combos and the right column provides the number of properties that have that number of combos.

Number of Combos	Frequency
1	28,057
2	11,136
3	5,241
4	1,920
5	1,285
6	516
7	154
8	72
9	45
11	33
12	12
14	14

There are 28,057 properties with only one combo (i.e., the combo for each of these properties is unique). The largest number of combinations per property is 14. The following additional fields have been calculated:

For each combo, we extract a range of addresses containing a starting number, an ending number, and a street name. For example, consider the following combination:

prop	propertyname	address1	address2
184915	Orange County Industrial Portfolio	"847-853 South Lamb Drive, 179-195 National	"250-260, 268-278, 310 Windy Pipe"

From which we extract the following distinct ranges:

start_of_range	end_of_range	streetname
847	853	South Lamb Drive
179	195	National
250	260	Windy Pipe
268	278	Windy Pipe
310	310	Windy Pipe

Note that each distinct combo is associated with unique city and zip fields. In the above case, this might be Charlotte, TN 12345 (not a real place). To successfully extract the addresses, one has to judiciously parse the string using the positions of forward slashes, dashes, ampersands, the word “and”, and commas. In addition, we note that periods are used either for abbreviations or decimals when referring to land (e.g., 1.04 acres). Address extraction was mostly performed by algorithm (after training on a smaller set that included “anomalies”) but also required some hand correction.
Each of the addresses in start_of_range & end_of_range are then validated and standardized using an address parser (LibPostal) and then geocoded to prepare it for matching. Geocoding is done using ArcGIS and, when the address cannot be found, using Google Maps. Overall, the method yields 52,565 combos with normalized, parsed, and geocoded address ranges. In turn, these correspond to 29,377 prop IDs (out of a possible 36,236).

Validating the matching methodology

As explained in the text, BFX undertook to compare their matching methodology with ours using a set of addresses containing four RCA properties for every NCREIF property. In both cases, the match rate was about 73% (the difference was a mere 0.0024%). Interestingly, there were complementarities between the two algorithms. In other words, BFX found matches that we did not, and we found address matches that BFX did not. Specifically, we agreed on 88% of the matches, while each uniquely contributed 6%. About half of the complementary matches found by BFX (i.e., 3% of the total matches) corresponded to BFX addresses that were *missing* from the address extract of 29M BFX originally sent us, suggesting a glitch in the extraction process. Had we had those addresses, the agreement rate between the methodologies would have been 91% of total matches, and our methodology would have contributed 6% of the remainder.

Appendix C: Multi-addressed properties

As mentioned in the text, some buildings are associated with more than one address, but NCREIF members rarely report more than one address per building. We have adopted a strategy for dealing with this problem. To make things concrete, consider the generic example (not something confidential) of 1050 17th ST NW, Washington, DC 20036. The latter is the "premise address" in the deeds and assessors' records. However, there are multiple other addresses associated with that property: 1020, 1030, and 1040. In Figure C.1 below, the orange bull's eye corresponds to the premise address, while 1020 17th ST NW corresponds to a unit presumably leased to "The Hair Shoppe". Now, suppose that this was a NCREIF property (it is not), meaning that it would most likely be identified by its premise address (1050). As long as all BFX permits associated with the subject property were identified with 1050 17th ST NW there would be no problem. Unfortunately, there are permit activities associated with 1040 17th ST NW. Thus only matching to 1050 would miss permitted work on other parts of the building. This is a more general problem with matching property records across platforms: One platform may refer to 1040 17TH while another to 1030 17th even though both correspond to the same "premise address" of 1050 17th.

Figure C.1: Example of Multi-Address Buildings

This figure illustrates an example of a multi-address building. The building in question is at the top right part of the block and its address in the municipal records is documented as 1050 17th St. NW, Washington DC.

To tackle this problem, we employ a multi-pronged strategy:

1. We attempt to identify not only NCREIF property addresses, but also all properties in a given block containing a NCREIF property. This allows us to easily identify a subsample of Census Blocks for which BFX associates only one address, and that address matches the NCREIF address. Results obtained using this much smaller dataset for which we have “maximal trust” can then be compared (for validation purposes) to those described below.
2. OpenStreetMap provides free access to a repository of distinct building profile shape information linked to geocodes. All address geocodes that “pierce” the same building shape can be associated with the same building. The main challenge with this methodology is the fact that OpenStreetMap contains building shape information for only about one half of the addresses in the inflated dataset (i.e., the NCREIF address set, supplemented by the RCA address set, and all addresses in BFX within 110 meters of NCREIF addresses).
3. In the majority of cases, the owner of a parcel and the owner of a structure on the parcel coincide. RCA has parcel shape file information obtained through a third party. They graciously agreed to implement a “parcel piercing” methodology on the inflated dataset (similar to the methodology used with OpenStreetMap). Here, two algorithms were implemented. A strict one requiring a geocode to land inside the parcel for an association to hold, and a more liberal algorithm that associates a geocode with the nearest parcel if the geocode lands outside a parcel.

References

Ambrose, B. W., & Steiner, E. (2018). Economic Fundamentals, Capital Expenditures and Asset Dispositions. *Real Estate Research Institute Working Paper (May 15, 2018)*.

Bokhari, S., & Geltner, D. (2018). Characteristics of depreciation in commercial and multifamily property: An investment perspective. *Real Estate Economics*, *46*(4), 745-782.

Calhoun, C. A. (1996). OFHEO house price indexes: HPI technical description. *Office of Federal Housing Enterprise Oversight*, *20552*, 1-15.

Campbell, J. Y., Giglio, S., & Pathak, P. (2011). Forced sales and house prices. *American Economic Review*, *101*(5), 2108-31.

Case, K. E., & Shiller, R. J. (1987). Prices of single family homes since 1970: New indexes for four cities.

De Long, J. B., & Summers, L. H. (1991). Equipment investment and economic growth. *The Quarterly Journal of Economics*, *106*(2), 445-502.

Eichholtz, P., Kok, N., & Quigley, J. M. (2010). Doing well by doing good? Green office buildings. *American Economic Review*, *100*(5), 2492-2509.

Giacoletti, M. (2019). Idiosyncratic risk in housing markets. *Available at SSRN 2995323.*

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of political economy*, *100*(6), 1126-1152.

Harding, J. P., Rosenblatt, E., & Yao, V. W. (2009). The contagion effect of foreclosed properties. *Journal of Urban Economics*, *66*(3), 164-178.

Kontokosta, C., Fuerst, H., Pivo, G., & Sagi, J. (2019), “Estimating Office and Residential Building Energy Retrofit Hurdle Rates” *DOE-RERI Working Paper*.

Levine, R., & Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *The American economic review*, 942-963.

Liu, C. H., Rosenthal, S. S., & Strange, W. C. (2018). The vertical city: Rent gradients, spatial structure, and agglomeration economies. *Journal of Urban Economics*, *106*, 101-122.

Pivo, G., & Fisher, J. (2010). Income, value, and returns in socially responsible office properties. *Journal of Real Estate Research*, *32*(3), 243-270.

Reher, M. (2018) The Renovation Rebalance: How Financial Intermediaries Affect Renter Housing Costs. *Mimeo.*

Sagi, J. S. (2017). Asset-level risk and return in real estate investments. *Available at SSRN 2596156*.

Stanton, R., Strickland, C., & Wallace, N. (2018) A New Dynamic House-Price Index for Mortgage Valuation and Stress Testing. *Mimeo.*

Wallace, N. E., & Meese, R. A. (1997). The construction of residential housing price indices: a comparison of repeat-sales, hedonic-regression, and hybrid approaches. *The Journal of Real Estate Finance and Economics*, *14*(1-2), 51-73.

Yoshida, J. (2016). Structure depreciation and the production of real estate services. *Available at SSRN 2734555*.

Zhu, C., White, A., Mathew, P., Deason, J., & Coleman, P. (2018). Raising the Rent Premium: Moving Green Building Research Beyond Certifications and Rent

Table 3.1: Permit Type Summary Statistics

This table reports summary statistics for BuildFax permit records matched to NCREIF property locations. The permits are broken down by type of work, as classified by BuildFax from permit text data. The columns correspond to different NCREIF property types: (A) Apartment, (H) Hotel, (I) Industrial, (L) Land, (O) Office, (R) Retail, (U) Unclassified, (X) Specialty (including healthcare, manufactured housing, parking, self-storage, and senior living). NCREIF data of matched properties is from 1982Q1 to 2017Q2. BuildFax permits are collected for the holding period of a NCREIF property as well up to one year prior and one year after. *Sources*: BuildFax data (property types are from NCREIF).

	A	H	I	L	O	R	U	X	Total
Properties	2,299	271	4,059	249	3,384	1,378	1,895	458	13,993
All Permits	64,910	8,741	65,056	8,764	348,964	101,435	11,640	13,372	622,882
Roof	3,797	294	1,660	106	3,265	1,425	387	241	11,175
Remodel	21,091	3,436	21,935	2,471	144,586	33,957	3,228	4,319	235,023
Damage	712	16	103	9	248	169	104	40	1,401
Fire Damage	289	1	47	3	50	104	17	6	517
New Construction	5,601	218	2,559	1,717	7,149	3,914	1,331	10,42	23,531
Building	12,943	1,368	13,950	1,614	68,617	21,395	2,260	2,510	124,657
Electrical	16,063	2,030	16,026	2,584	105,179	26,601	3,237	3,412	175,132
Mechanical	6,950	936	6,335	831	34,733	10,475	1,568	1,455	63,283
Plumbing	10,057	1,019	5,468	1,036	34,179	10,163	1,283	1,744	64,949
Pool	1,068	135	78	74	82	140	109	37	1,723
Repair	7,271	715	3,365	202	19,744	5,754	765	833	38,649
Solar	28	10	112	5	88	64	15	20	342
Water Damage	289	14	14	2	130	34	33	22	538
Wind Damage	1	0	2	1	2	1	1	0	8
Unclassified	11,018	1,793	15,986	1,741	61,202	23,838	2,211	2,571	120,360

Table 3.2: Permit Type Cost Statistics

This table reports summary cost statistics for BuildFax permit records matched to NCREIF property locations. The permits are broken down by type of work, as classified by BuildFax from permit text data. NCREIF data of matched properties is from 1982Q1 to 2017Q2. BuildFax permits are collected for the holding period of a NCREIF property as well up to one year prior and one year after. Column (1) provides the means and Column (3) shows the standard deviation of costs. Columns (4) through (7) show the 25th, 50th, 75th, 90th, and 99th percentiles. All quantities are in $U.S. *Sources*: BuildFax data (property types are from NCREIF).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Permit Type	Mean	Std. Dev.	p25	p50	p75	p90	p99
Roof	195,078	2,263,103	0	10,000	65,000	265,000	1,887,231
Remodel	72,241	553,296	0	1,920	25,000	120,000	1,124,800
Damage	67,240	265,143	0	5,000	31,000	101,523	1,356,150
Fire Damage	55,899	226,561	0	3,200	26,700	100,000	1,200,000
New Construction	746,753	6,046,214	0	0	100,000	600,000	14887475
Building	231,448	2,553,396	0	10,863	75,000	270,000	3,000,000
Electrical	23,852	409,182	0	0	1,500	16,200	322,000
Mechanical	72,544	3,452,155	0	0	6,500	45,000	715,000
Plumbing	22,175	573,026	0	0	1	8,000	285,000
Pool	318,068	3,765,818	0	1	25,000	92,200	2,315,244
Repair	72,681	485,826	0	3,000	28,000	130,000	1,058,648
Solar	598,567	3,747,527	0	1,131	60,000	594,480	8,717,200
Water Damage	54,889	244,606	0	3,400	20,000	88,500	907,358
Wind Damage	93,779	189,939	0	14,430	87,000	547,369	547,369
Unclassified	64,342	1,023,909	0	0	3,000	37,909	835,869

Table 3.3: Permit Statistics by Property Type

This table reports permit statistics by property types for BuildFax permit records matched to NCREIF property locations. NCREIF property types are coded as follows: (A) Apartment, (H) Hotel, (I) Industrial, (L) Land, (O) Office, (R) Retail, (X) Specialty (including healthcare, manufactured housing, parking, self-storage, and senior living). We do not report statistics for properties for which there was no reported classification. Panel A reports statistics on the annualized number of permits per property, calculated as the total number of permits matched divided by the holding period of the property. Panel B reports statistics on aggregated permit valuations per property. Panel C reports statistics on the aggregated CapEx per property (as reported by NCREIF). Column (1) contains the number of properties, Column (2) contains means, and Column (3) reports standard deviations. Columns (4) through (8) report percentiles. *Sources:* BuildFax and NCREIF data.

Panel A: Number of Permits per Property per Year						Observations: 12,098		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Count	Mean	Std. Dev.	p1	p25	p50	p75	p99
A	2,299	7.15	14.77	0.36	1.57	3.51	6.80	78.67
H	271	5.20	7.29	0.48	2.00	3.14	5.60	46.26
I	4,059	4.14	4.58	0.41	1.65	3.00	4.73	22.00
L	249	11.22	19.18	0.71	3.33	5.76	11.27	128.46
O	3,384	14.82	20.58	0.71	4.00	8.44	18.50	89.54
R	1,378	10.11	14.87	0.48	2.80	5.21	10.86	77.53
X	458	7.71	20.65	0.29	1.63	4.00	8.00	89.67
Panel B: Permit Valuations ($1000's) per Property								
A	2,299	2,534	17,238	0	0	20	198	57,224
H	271	3,285	19,788	0	5	295	908	89,012
I	4,059	932	3,603	0	1	85	485	16,297
L	249	18,135	42,988	0	19	1,011	15,661	243,935
O	3,384	4,518	16,053	0	43	716	2,782	68,706
R	1,378	2,202	7,351	0	6	265	1,622	29,919
X	458	1,631	10,553	0	0	12	264	28,300
Panel C: CapEx ($1000's) per Property								
A	2,299	4,274	13,816	-251	68	359	1,377	75,945
H	271	8,230	41,097	-14	282	1,155	3,576	70,964
I	4,059	915	3,831	-606	3	122	590	14,939
L	249	20,036	39,699	-3,357	153	4,816	23,858	166,083
O	3,384	8,859	23,159	-1,040	266	1,882	7,371	120,413
R	1,378	4,729	42,393	-756	23	357	1,955	52,646
X	458	2,851	18,488	-1,716	11	99	614	45,750

Table 3.4: Permit Statistics by Fund Type

This table reports permit statistics by owning fund type for BuildFax permit records matched to NCREIF property locations. NCREIF private equity fund types are coded as follows: (C) Closed-end, (D) Open-end diversified core equity (ODCE), (N) Not elsewhere classified, (O) Non-ODCE open-end, (R) REIT, (S) Separate account. We exclude properties for which there was no reported fund classification. Panel A reports statistics on the annualized number of permits per fund property, calculated as the total number of permits matched divided by the holding period of the property. Panel B reports statistics on aggregated permit valuations per fund property. Panel C reports statistics on the aggregated CapEx per fund property (as reported by NCREIF). Column (1) contains the number of properties, Column (2) contains means, and Column (3) reports standard deviations. Columns (4) through (8) report percentiles. *Sources:* BuildFax and NCREIF data.

Panel A: Number of Permits per Property per Year						Observations 11,940		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Count	Mean	Std. Dev.	p1	p25	p50	p75	p99
C	1,947	9.30	22.40	0.50	2.15	4.00	8.80	80.00
D	2,354	8.92	14.48	0.42	2.00	4.00	9.60	72.70
N	772	8.38	14.08	0.30	1.88	4.00	8.38	85.25
O	2,154	8.04	13.21	0.48	2.00	4.00	8.25	66.91
R	367	10.61	14.98	0.55	2.32	4.40	12.97	82.49
S	4,346	8.37	12.68	0.44	2.24	4.00	8.80	65.17
Panel B: Permit Valuations ($1000's) per Property								
C	1,947	1,883	8,505	0	0	60	635	43,438
D	2,354	3,823	15,920	0	1	191	1,451	71,636
N	772	3,234	17,850	0	8	217	1,172	68,517
O	2,154	3,210	16,981	0	2	170	1,151	50,527
R	367	2,170	8,376	0	15	407	1,558	18,192
S	4,346	2,554	13,093	0	1	114	909	50,928
Panel C: Permit Valuations ($1000's) per Property								
C	1,947	2,891	9,438	-438	25	240	1,589	46,980
D	2,354	6,891	23,942	-371	70	573	3,278	107,890
N	772	3,471	14,144	-2,630	9	279	1,846	56,204
O	2,154	5,255	34,346	-244	60	439	2,505	79,844
R	367	9,445	28,799	-9,486	45	902	5,969	168,019
S	4,346	4,327	18,017	-645	21	326	1,854	72,245

Table 4.1: Property Level Summary Statistics

This table provides summary statistics for quarterly property level data from 1995 Q1 to 2016 Q4. Column (1) contains the means for these variables and Column (2) contains the standard deviations. Permit valuation and capital expenditure variables have been trimmed at the 1% level by county. The observations have been restricted to the subset of properties without missing data among the following variables: capital expenditure, permit valuation, state, age, property type, and fund type. *Sources:* Permit valuations and number of permits come from BuildFax. Capital expenditure and other variables come from NCREIF.

	Mean	Standard Deviation
Permit Valuation	62,301	318,642
Permit Valuation (per SqFt)	0.21	0.78
Capital Expenditure	147,638	488,742
Capital Expenditure (per SqFt)	0.55	1.24
Number of Permits per Quarter	1.30	4.01
Building Age	20.17	12.22
Square Footage	254,067	263,099
Number of Properties	5,628	

Table 4.2: Property Level Regressions of Capital Expenditure on Permit Valuations

This table reports results from property level regressions of capital expenditure on permit valuations. Data is as described in Table 4.1. Column (1) regresses quarterly capital expenditure per square foot on quarterly permit valuations per square foot. Column (2) adds the number of permits per quarter to the regression and Column (3) adds building age. In Columns (4) and (5), year and state fixed effects are included. Column (6) adds property and fund type fixed effects. Column (7) adds permit type valuations to the explanatory variables in Column (6). These permit types include the valuations for repairs, construction, roofs, electrical, mechanical, plumbing, pools, solar, damage, and uncategorized permits. Finally, Column (8) adds regional variables at both the zip code and county levels. Zip code controls include auto sales, auto sales growth, house prices, house price growth, education levels, racial composition, percent married with children, and median age. County level controls include unemployment, unemployment change, mean income, and mean income growth. Capital expenditure and permit valuations have been trimmed at the 1% level by county and smoothed by taking an average of the current quarter, the previous two quarters, and the following two quarters. They are also deflated by the producer price index. Standard errors are reported in parentheses. *Sources:* Permit valuations and number of permits come from BuildFax. Capital expenditure, state, property type, and fund type come from NCREIF. House price data comes from Zillow. Auto sales data comes from Polk. Remaining zip code variables come from the U.S. Census Bureau American Community Survey 5-Year Estimates. County unemployment and income come from the U.S. Census Bureau, Small Area Income and Poverty Estimates.

	(1) CapEx Per SqFt	(2) CapEx Per SqFt	(3) CapEx Per SqFt	(4) CapEx Per SqFt	(5) CapEx Per SqFt	(6) CapEx Per SqFt	(7) CapEx Per SqFt	(8) CapEx Per SqFt
Permit Valuation Per SqFt	0.783***	0.681***	0.674***	0.673***	0.664***	0.552***	0.754***	0.744***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.018)	(0.018)
Number of Permits		0.0274***	0.0260***	0.0258***	0.0245***	0.0190***	0.0190***	0.0176***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Building Age			0.00796***	0.00794***	0.00787***	0.00769***	0.00771***	0.00687***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year FE	No	No	No	Yes	Yes	Yes	Yes	Yes
State FE	No	No	No	No	Yes	Yes	Yes	Yes
Property and Fund Type FE	No	No	No	No	No	Yes	Yes	Yes
Permit Type Valuations per SqFft	No	No	No	No	No	No	Yes	Yes
Regional Controls	No	No	No	No	No	No	No	Yes
Observations	130,669	130,669	130,669	130,669	130,669	130,669	130,669	130,669
R-squared	0.205	0.215	0.225	0.226	0.238	0.307	0.310	0.318
Regression RMSE	0.871	0.866	0.861	0.859	0.853	0.813	0.812	0.807

Table 4.3: Property Level Linear Mixed Model

This table contains the results from various linear mixed models. Column (1) contains no levels and is equivalent to simply regression capital expenditure per square foot on permit valuation per square foot. Column (2) contains two levels: both property type and county levels. Column (3) contains property type, county, and year levels. The model in Column (3) is of the form $CapEx\_SqFt_{i,p,c,y} = \beta Permit\_Valuation\_SqFt_{i,p,c,y} + \gamma_p + \omega_{p,c} + \nu_{p,c,y} + \epsilon_{i,p,c,y}$, where γ_p is the random intercept of property type p, $\omega_{p,c}$ is the random effect of county c within property type p, and $\nu_{p,c,y}$ is the random effect of year y, within county c, within property type p. Capital expenditure and permit valuations have been trimmed at the 1% level by county and smoothed by taking an average of the current quarter, the previous two quarters, and the following two quarters. They are also deflated by the producer price index. Standard errors are reported in parentheses. *Sources:* Permit valuations come from BuildFax. Capital expenditure and other variables come from NCREIF.

	(1)	(2)	(3)
	No Levels	Property Type and County Levels	Property Type, County, and Year Levels
	CapEx per SqFt	CapEx per SqFt	CapEx per SqFt
Permit Valuation per SqFt	0.794	0.645	0.645
	(0.004)	(0.004)	(0.004)
AIC	539,913	510,959	503,297
BIC	539,943	511,009	503,358
Log Likelihood	-269,953	-255,474	-251,642
Regression RMSE	0.976	0.897	0.845
Number of Observations	193,548	193,548	193,548

Table 4.4: Property Level Out of Sample Regressions

This table reports regressions of actual capital expenditure from our dataset on both permit valuations and model-imputed capital expenditure in order to assess the information content in various proxies for CapEx. Permit valuations and capital expenditure are trimmed at the 1% level and smoothed by taking the average of the current quarter, the previous two quarters, and the following two quarters. They are then deflated by the producer price index. The capital expenditure predictions are from a mixed model with property, county, and year levels. To generate the predictions, the dataset is randomly split in half, where half of the data is used to estimate the mixed model and the other half is used for out of sample prediction of capital expenditure. The regressions in this table are estimated using only the observations that were randomly assigned for prediction. Column (1) regresses actual capital expenditure per square foot on out-of-sample model-imputed capital expenditure per square foot, while Column (2) regresses actual capital expenditure per square foot on permit valuations per square foot. In the row, “Predictive Accuracy RMSE” we calculate the RMSE of errors corresponding to a simple *difference* between the dependent and independent variables. Standard errors are in parentheses. *Sources:* Permit valuations and number of permits come from BuildFax. Capital expenditure and other variables come from NCREIF.

Panel A: Regression Results

	(1) CapEx per SqFt	(2) CapEx per SqFt
Model-imputed CapEx per SqFt	1.012***	
	(0.005)	
Permit Valuation Per SqFt		0.664***
		(0.006)
Observations	94,758	96,464
R-squared	0.309	0.241
Regression RMSE	0.882	0.928
Predictive Accuracy RMSE	0.882	1.073

Table 4.5: State Level Out of Sample Regressions

This table reports regressions of actual capital expenditure from our dataset on both permit valuations and model-imputed capital expenditure in order to assess the information content in various proxies for CapEx. All variables are averaged by state. Permit valuations and capital expenditure are trimmed at the 1% level and smoothed by taking the average of the current quarter, the previous two quarters, and the following two quarters. They are then deflated by the producer price index. The capital expenditure predictions are from a mixed model with property, county, and year levels. To generate the predictions, the dataset is randomly split in half, where half of the data is used to estimate the mixed model and the other half is used for out of sample prediction of capital expenditure. The regressions in this table are estimated using only the observations that were randomly assigned for prediction. Then, the property level observations are aggregated through averaging to the state level, so there is one observation per state each quarter. Column (1) regresses actual state-level averaged capital expenditure per square foot on out-of-sample model-imputed capital expenditure per square foot averaged to the state level, while Column (2) regresses actual state-level aggregated capital expenditure per square foot on aggregated permit valuations per square foot. In the row, "Predictive Accuracy RMSE" we calculate the RMSE of errors corresponding to a simple *difference* between the dependent and independent variables. Standard errors are in parentheses. *Sources:* Permit valuations and number of permits come from BuildFax. Capital expenditure and other variables come from NCREIF.

	(1) CapEx per SqFt	(2) CapEx per SqFt
Model-imputed CapEx per SqFt	0.932***	
	(0.020)	
Permit Valuation Per SqFt		0.623***
		(0.036)
Observations	3202	3271
R-squared	0.415	0.423
Regression RMSE	0.430	0.429
Predictive Accuracy RMSE	0.432	0.701

Table 4.6: State-Property Type Level Out of Sample Regressions

This table reports regressions of actual capital expenditure from our dataset on both permit valuations and model-imputed capital expenditure in order to assess the information content in various proxies for CapEx. All variables are averaged by state and property type. Permit valuations and capital expenditure are trimmed at the 1% level and smoothed by taking the average of the current quarter, the previous two quarters, and the following two quarters. They are then deflated by the producer price index. The capital expenditure predictions are from a mixed model with property, county, and year levels. To generate the predictions, the dataset is randomly split in half, where half of the data is used to estimate the mixed model and the other half is used for out of sample prediction of capital expenditure. The regressions in this table are estimated using only the observations that were randomly assigned for prediction. Then, the property level observations are aggregated through averaging to the state-property type level, so each state has an observation for each property type per quarter. Column (1) regresses actual aggregated capital expenditure per square foot on out-of-sample model-imputed capital expenditure per square foot averaged to the state-property type level, while Column (2) regresses actual aggregated capital expenditure per square foot on aggregated permit valuations per square foot. In the row, "Predictive Accuracy RMSE" we calculate the RMSE of errors corresponding to a simple *difference* between the dependent and independent variables. Standard errors are in parentheses. *Sources:* Permit valuations and number of permits come from BuildFax. Capital expenditure and other variables come from NCREIF.

	(1) CapEx per SqFt	(2) CapEx per SqFt
Model-imputed CapEx per SqFt	1.049***	
	(0.011)	
Permit Valuation Per SqFt		0.562***
		(0.024)
Observations	10,202	10,696
R-squared	0.462	0.293
Regression RMSE	0.430	0.711
Predictive Accuracy RMSE	0.591	0.938

Table 4.7: Fund Level Out of Sample Regressions

This table reports regressions of actual capital expenditure from our dataset on both permit valuations and model-imputed capital expenditure in order to assess the information content in various proxies for CapEx. Each panel reports results for averages within a specified fund. Permit valuations and capital expenditure are trimmed at the 1% level and smoothed by taking the average of the current quarter, the previous two quarters, and the following two quarters. They are then deflated by the producer price index. To carry out this analysis, we look at the four funds with the largest number of properties in our dataset. For each fund, we first remove all of the properties belonging to that fund. Then, we estimate both a simple linear regression and a mixed model using all other properties in order to predict capital expenditure. The linear regression regresses capital expenditure per square foot on state, property type, age, time since acquisition, and permit valuation per square foot. The mixed model is estimated with property type, county, and year levels. Using the estimated regression and mixed model, we generate quarterly capital expenditure per square foot predictions using only the properties that belong to that fund. Finally, we investigate the accuracy of these predictions through the following regressions. Columns (1) regress actual capital expenditure per square foot on the model-imputed capital expenditure per square foot from the linear regression. Columns (2) regress actual capital expenditure per square foot on the model-imputed capital expenditure per square foot from the mixed model. Columns (3) regress actual capital expenditure per square foot on permit valuations per square foot. In the row, "Predictive Accuracy RMSE" we calculate the RMSE of errors corresponding to a simple *difference* between the dependent and independent variables. Standard errors are in parentheses. *Sources:* Permit valuations and number of permits come from BuildFax. Capital expenditure and other variables come from NCREIF.

Panel A: Largest Fund

	(1) CapEx per SqFt	(2) CapEx per SqFt	(3) CapEx per SqFt
Model-imputed CapEx per SqFt v1	1.574***		
	(0.341)		
Model-imputed CapEx per SqFt v2		2.412***	
		(0.204)	
Permit Valuation Per SqFt			-1.493
			(1.010)
Observations	88	88	88
R-squared	0.199	0.620	0.205
Regression RMSE	0.432	0.298	0.433
Predictive Accuracy RMSE	0.494	0.448	0.859

Panel B: Second Largest Fund

	(1) CapEx per SqFt	(2) CapEx per SqFt	(3) CapEx per SqFt
Model-imputed CapEx per SqFt v1	1.610***		
	(0.253)		
Model-imputed CapEx per SqFt v2		1.385***	
		(0.161)	
Permit Valuation Per SqFt			1.521***
			(0.232)
Observations	88	88	88
R-squared	0.320	0.463	0.376
Regression RMSE	0.18573	0.165	0.179

Predictive Accuracy RMSE	0.206	0.173	0.453

Panel C: Third Largest Fund

	(1) CapEx per SqFt	(2) CapEx per SqFt	(3) CapEx per SqFt
Model-imputed CapEx per SqFt v1	0.503***		
	(0.106)		
Model-imputed CapEx per SqFt v2		0.562***	
		(0.079)	
Permit Valuation Per SqFt			0.694***
			(0.176)
Observations	88	88	88
R-squared	0.207	0.367	0.280
Regression RMSE	0.109	0.097	0.104
Predictive Accuracy RMSE	0.188	0.155	0.525

Panel D: Fourth Largest Fund

	(1) CapEx per SqFt	(2) CapEx per SqFt	(3) CapEx per SqFt
Model-imputed CapEx per SqFt v1	1.066***		
	(0.205)		
Model-imputed CapEx per SqFt v2		0.824***	
		(0.118)	
Permit Valuation Per SqFt			0.696***
			(0.200)
Observations	88	88	88
R-squared	0.239	0.361	0.166
Regression RMSE	0.102	0.094	0.108
Predictive Accuracy RMSE	0.106	0.111	0.419

Table 5.1: Summary of Unadjusted and Adjusted Capital Appreciation

The top table reports summary statistics for the adjusted capital appreciation resulting from actual and model-imputed capital expenditures, as well as the unadjusted capital appreciation (i.e., without adjusting for CapEx). Column (2) contains the number of observations, Column (3) reports the means of the variables, and Column (4) reports the standard deviations. The statistics are across all holding period horizons. The table at the bottom breaks out the mean returns across different holding horizons (in years). *Sources:* NCREIF data.

CapEx Adjustment	Observations	Mean	Std. Dev.
Actual	2,724	0.0245	0.3047
Model-imputed	2,724	0.0193	0.3122
Unadjusted	2,724	0.1017	0.3185

Holding Period (years)	Actual (mean return)	Model-imputed (mean return)	Unadjusted (mean return)	Number of Observations
1	-0.012	-0.010	0.015	316
2	0.013	0.017	0.054	516
3	0.026	0.022	0.078	375
4	0.014	0.006	0.084	316
5	0.034	0.025	0.109	253
6	0.036	0.029	0.131	222
7	-0.116	-0.121	0.009	142
8	-0.102	-0.113	0.019	115
9	-0.052	-0.064	0.065	115
10	-0.004	-0.006	0.127	92

Table 5.2: Relationship Between Reported Electrical Work Value and Utility Expenditures

This table reports the results of regressions that investigate the relationship between reported electrical work value and utility expenditures. If a permit is flagged for multiple purposes, we consider the work valuation divided among the different purposes. We then smooth electrical capital expenditure per square foot by quarter and property, removing outliers. We also generate a smoothed, quarterly utility expenditure per square foot. We create an indicator variable, "Temp. Shock", that takes on the value of one if the temperature is above 80 or below 40 degrees Fahrenheit in a given quarter and 0 otherwise. We also create an indicator variable, "Current Rent", that is equal to one whenever current rents equal the rents from the previous year. Temp. Shock, Current Rent, and Avg. Util. are demeaned. Column (1) regresses utility expenditures per square foot on Temp. Shock. Column (2) regresses utility expenditure on smoothed electrical work valuation per square foot, lagged by one year. Column (3) regresses utility expenditure on temperature shock, lagged electrical valuation and the interaction of lagged electrical valuation and temperature shock. Finally, Column (4) adds Current Rent, the interaction of Current Rent and Temp Shock, the interaction of Current Rent, Temp Shock, and lagged electrical valuation, average utilities within the MSA that quarter, and the interaction of temperature shock and average utilities. All regressions include year and property type fixed effects. *Sources:* Permit valuations and permit types come from BuildFax. Capital expenditure and other building come from NCREIF. Temperatures come from the National Oceanic and Atmospheric Administration.

	(1) Utility Exp.	(2) Utility Exp.	(3) Utility Exp.	(4) Utility Exp.
Temp. Shock	0.00963***		0.00997***	0.0129***
	(5.10)		(4.94)	(6.07)
Lagged Elect. Val.		0.0116*	0.0112*	0.0135
		(2.37)	(2.29)	(1.43)
Temp. Shock * Lagged Elect. Val.			0.0104	-0.0809*
			(0.70)	(-2.53)
Current Rent				-0.0168***
				(-14.60)
Current Rent * Temp. Shock				0.00677
				(1.82)
Current Rent Lagged Elect. Val.				0.0223
				(0.56)
Current Rent * Temp. Shock * Lagged Elect. Val.				-0.455**
				(-3.26)
Avg. Util.				0.175***
				(34.71)
Temp. Shock * Avg. Util.				0.184***
				(11.55)
Year & Property Type FEs	Yes	Yes	Yes	Yes
Constant	0.248***	0.259***	0.259***	0.243***
	(124.94)	(109.91)	(109.84)	(102.19)
Observations	141,847	123,922	123,922	123,922
Adj. R^2	0.0000	0.0004	0.0001	0.0688

Figure 3.1: Permit Valuation versus CapEx, Aggregated at the Property Level

This figure plots the natural logarithm of permit valuation, aggregated at the property level, against the natural logarithm of CapEx, also aggregated at the property level. *Sources:* Permit valuations and number of permits come from BuildFax. Capital expenditure come from NCREIF.

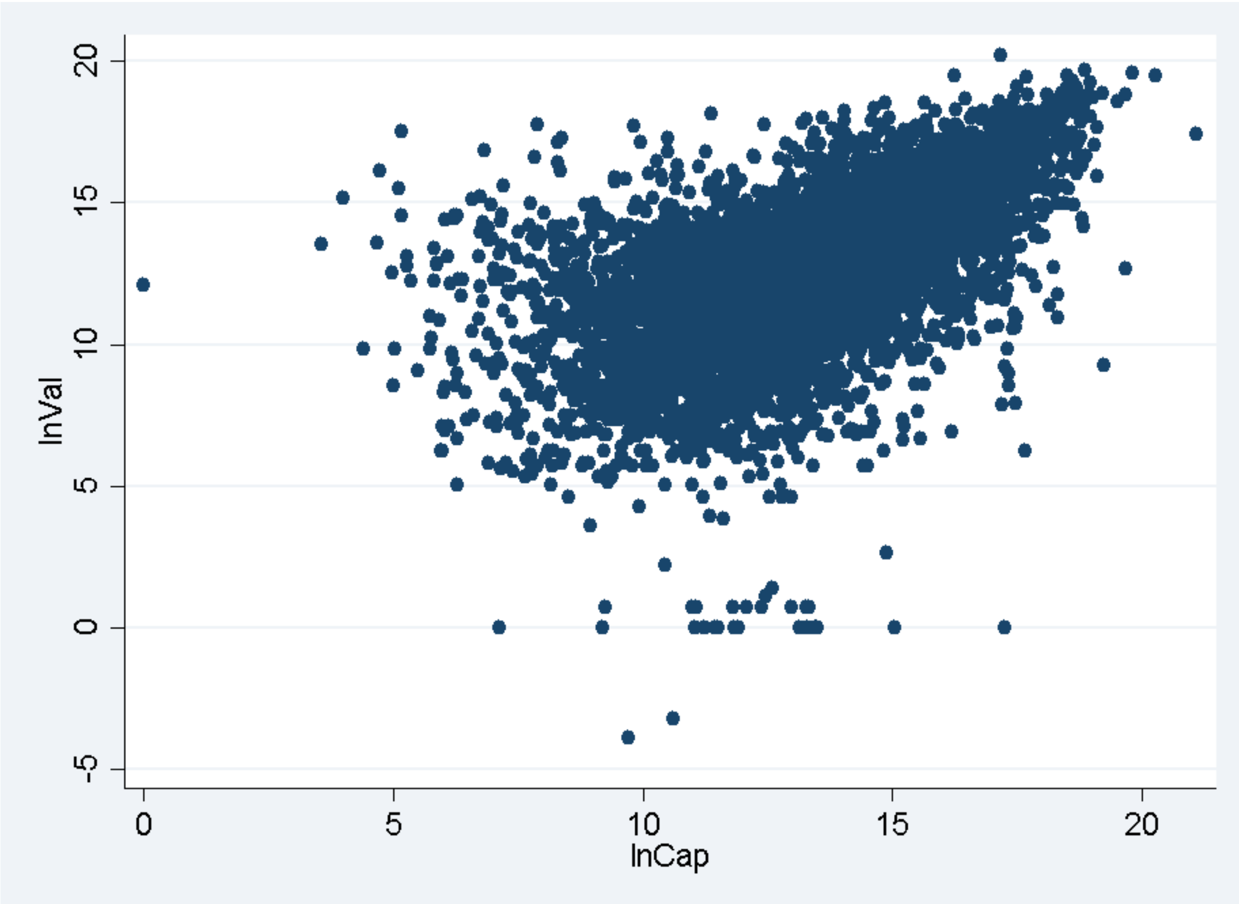
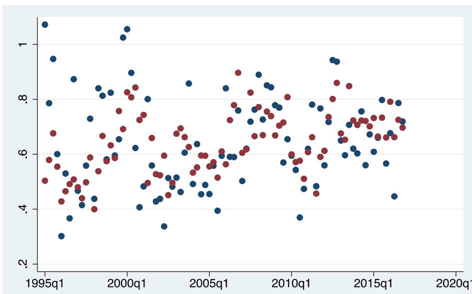


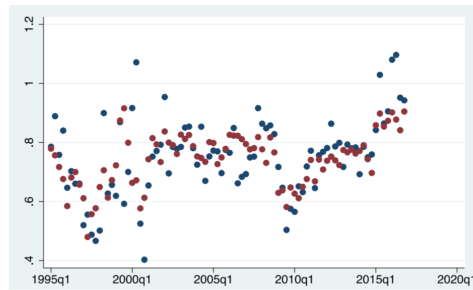
Figure 4.1: Actual Versus Model-imputed Capital Expenditure for Major States

These figures plot actual versus model-imputed capital expenditure per square foot for the major states in our dataset. Permit valuations and capital expenditure are trimmed at the 1% level and smoothed by taking the average of the current quarter, the previous two quarters, and the following two quarters. They are then deflated by the producer price index. The capital expenditure predictions are from a mixed model with property, county, and year levels. To generate the predictions, the dataset is randomly split in half, where half of the data is used to estimate the mixed model and the other half is used for out of sample prediction of capital expenditure. The figures below are created using only the observations that were randomly assigned for prediction. *Sources:* Permit valuations and number of permits come from BuildFax. Capital expenditure and other variables come from NCREIF.

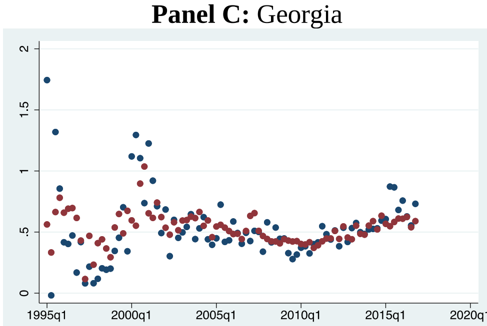
Panel A: Arizona



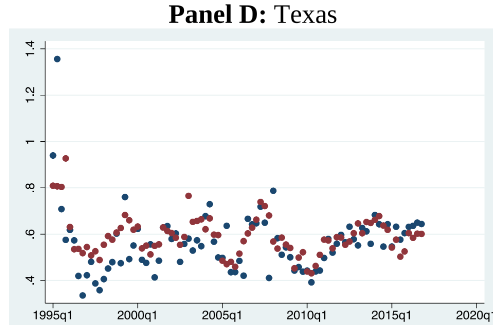
Panel B: California



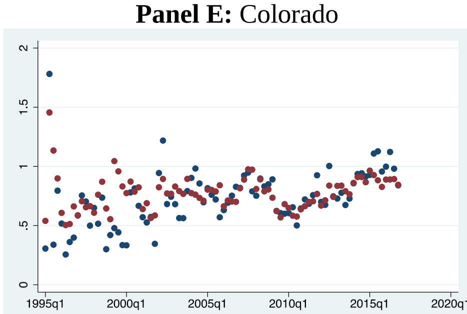
Panel C: Georgia



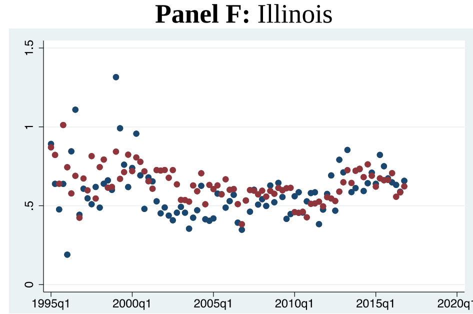
Panel D: Texas



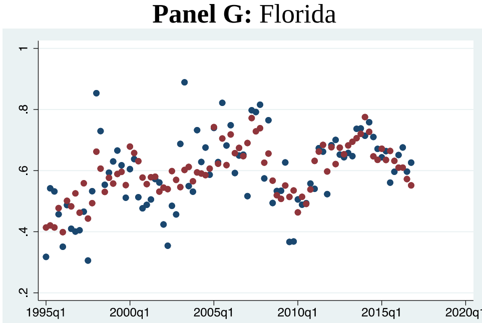
Panel E: Colorado



Panel F: Illinois



Panel G: Florida



Panel H: Virginia

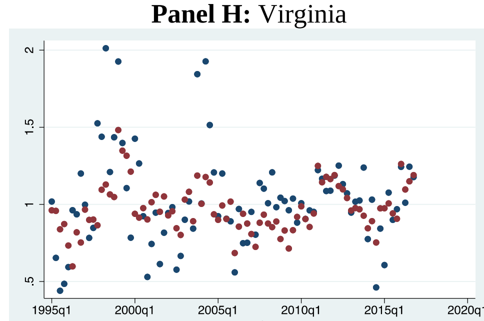
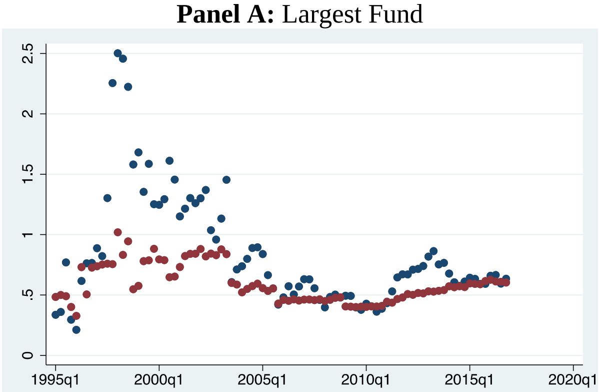


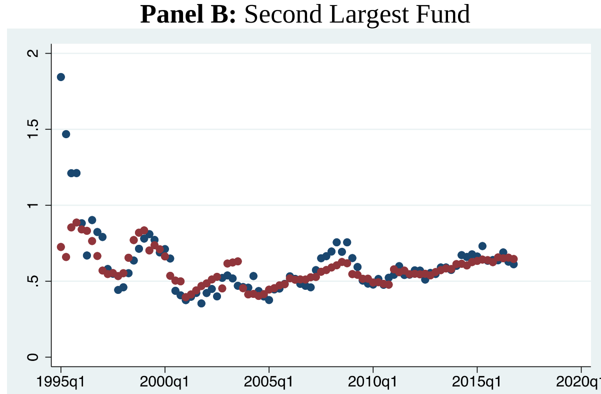
Figure 4.2: Actual Versus Model-imputed Capital Expenditure for Largest Funds

These figures plot actual versus model-imputed capital expenditure per square foot for the four largest funds in our dataset. Permit valuations and capital expenditure are trimmed at the 1% level and smoothed by taking the average of the current quarter, the previous two quarters, and the following two quarters. They are then deflated by the producer price index. The capital expenditure predictions are from a mixed model with property, county, and year levels. First for each fund, we remove all of the properties belonging to that fund. Then, we estimate the mixed model using all other properties in order to predict capital expenditure. Using the estimated mixed model, we generate quarterly capital expenditure per square foot predictions using only the properties that belong to that fund. *Sources:* Permit valuations and number of permits come from BuildFax. Capital expenditure and other variables come from NCREIF.

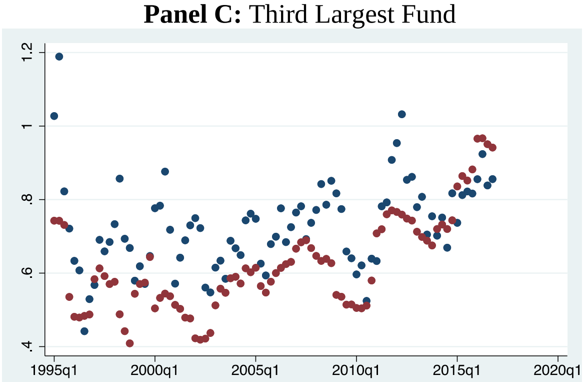
Panel A: Largest Fund



Panel B: Second Largest Fund



Panel C: Third Largest Fund



Panel D: Fourth Largest Fund

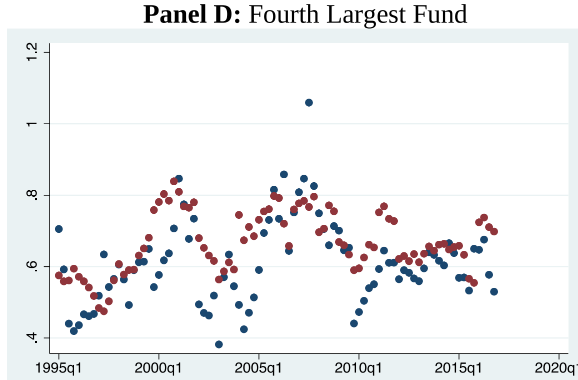


Figure 5.1: Log – Log Scatter plot of Actual versus Model-imputed Total Discounted CapExp

This figure plots log actual total discounted capital expenditure versus log model-imputed (out of sample) total discounted capital expenditure. Quarterly capital expenditures are trimmed at the 1% level by county. The quarterly capital expenditure predictions are from a mixed model with property, county, and year levels. To generate the predictions, the dataset is randomly split in half, where half of the data is used to estimate the mixed model and the other half is used for out of sample prediction of capital expenditure. Total discounted CapEx (actual and model-imputed) corresponds to discounting each quarterly CapEx to an initial date and then summing. The compounded prevailing 3m TBill is used to calculate the discount rate. *Sources:* NCREIF data.

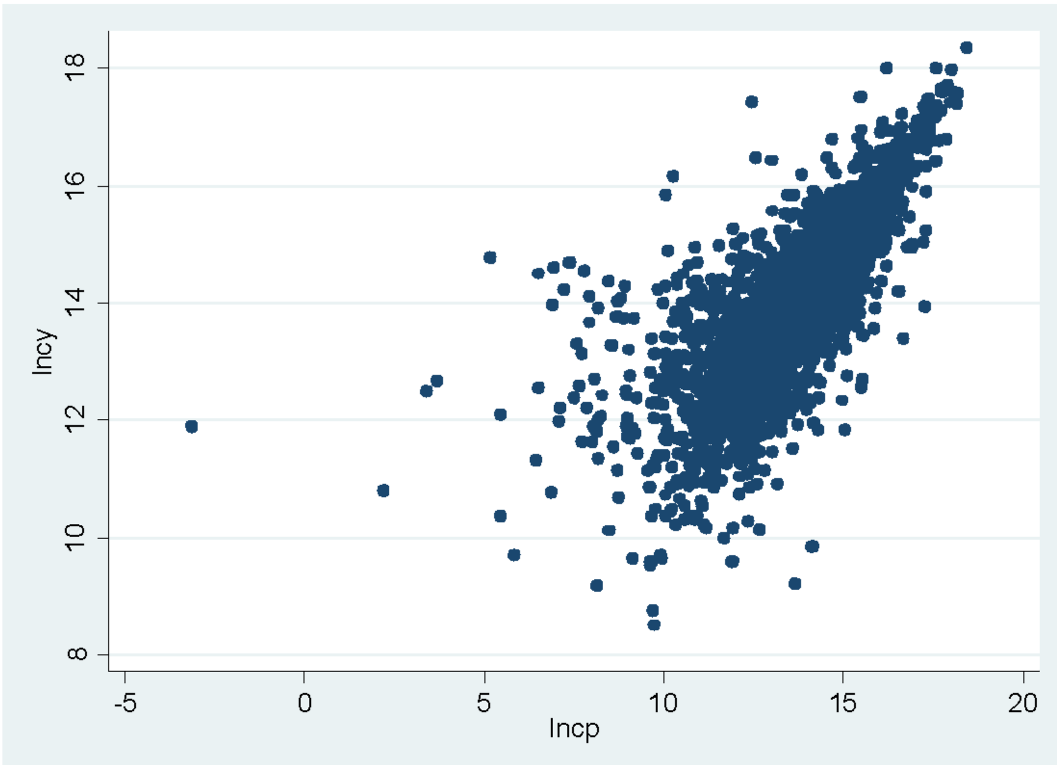
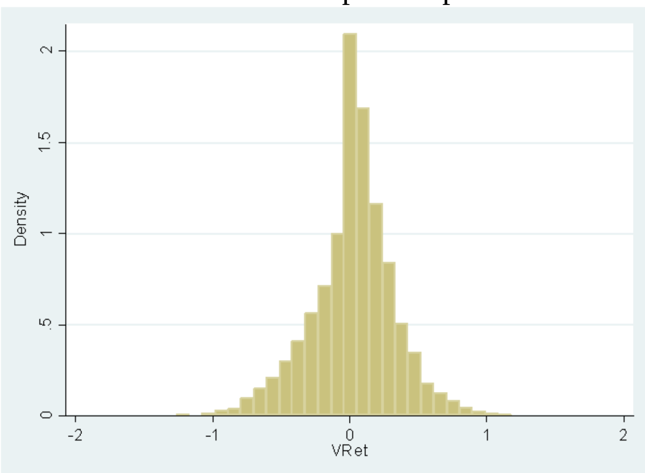


Figure 5.2: Distribution of Adjusted Capital Appreciation

This figure shows the distribution of the adjusted capital appreciation using actual (Panel A) and model-imputed (Panel B) capital expenditure. *Sources:* NCREIF data.

Panel A: Actual Capital Expenditure



Panel B: Model-imputed Capital Expenditure

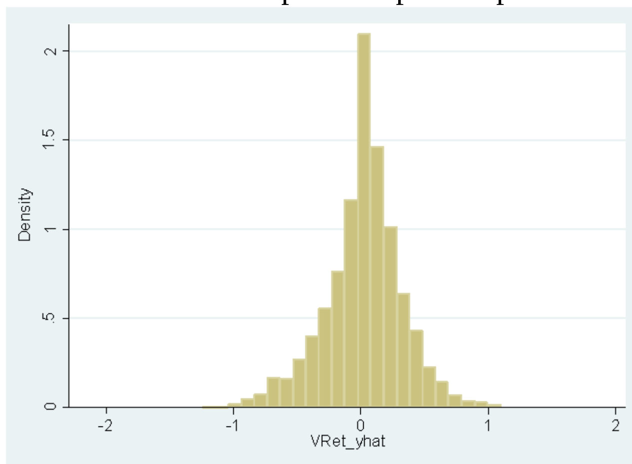


Figure 5.3: Property Gains Corrected for Actual Versus Model-imputed CapEx

This figure plots out of sample model-imputed property gains versus actual property gains. *Sources:* NCREIF data

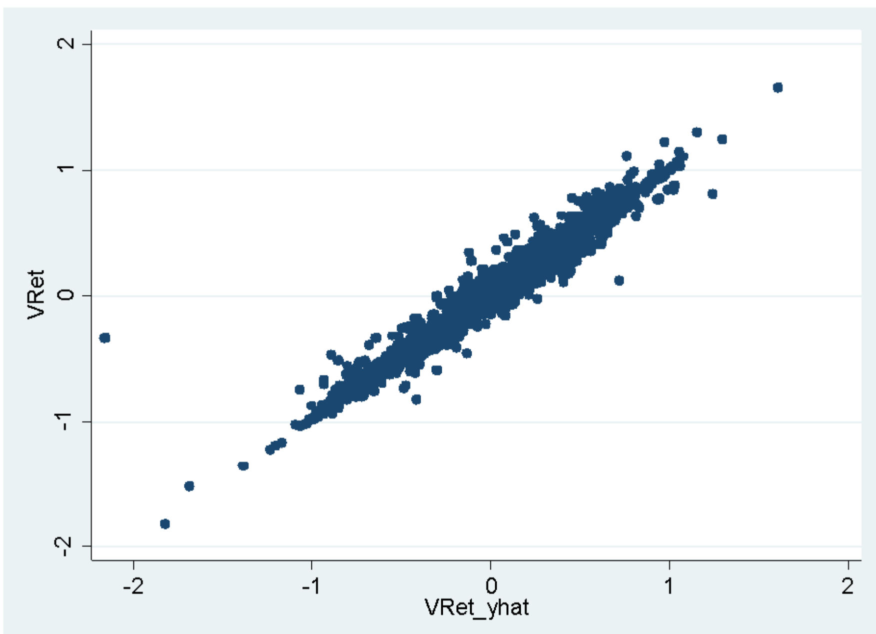


Figure 5.4: Repeat Sales Capital Gains Index for OOS Properties: Unadjusted for CapEx versus Adjusted for Actual CapEx

This figure shows the price appreciation index for out of sample properties. The blue line shows the index unadjusted for capital expenditure while the orange line shows the index adjusted for actual capital expenditure. Standard deviation error bars (not 95% confidence intervals) are shown. *Sources:* NCREIF data.

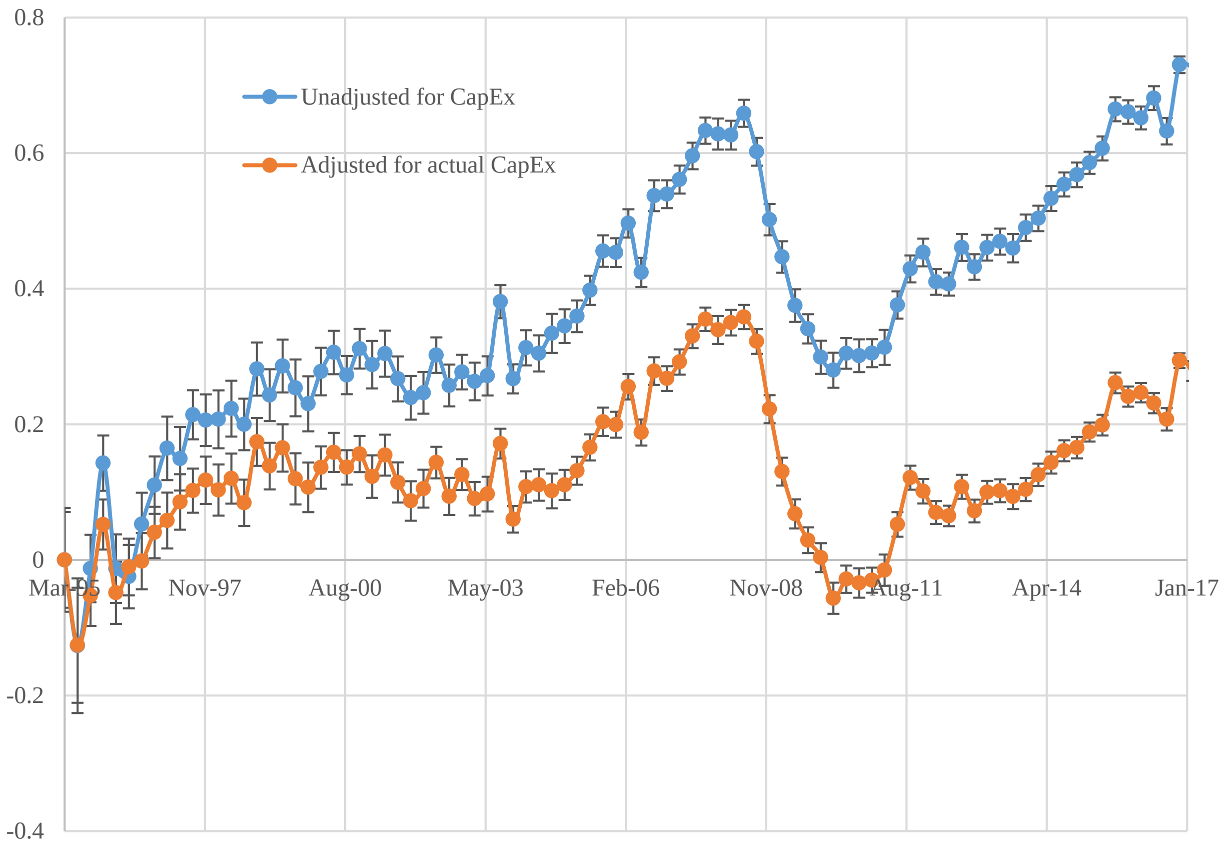


Figure 5.5: Repeat Sales Capital Gains Index for OOS Properties: Index Using Actual CapEx versus Index Using Imputed CapEx

This figure shows the price appreciation index for out of sample properties. The blue line shows the index using actual capital expenditure while the orange line shows the index using (out of sample) model model-imputed capital expenditure. Standard deviation error bars (not 95% confidence intervals) are shown. *Sources:* NCREIF data.