Estimating Commercial Property Fundamentals from REIT data*

David Geltner[†] Anil Kumar[‡] Alex M. Van de Minne [§]

January 14, 2023

In this paper we propose a new methodology for the estimation of fundamental property-level investment real estate time series performance and operating data using real estate investment trust (REIT) data. The methodology is particularly useful to develop publicly accessible operating statistics, such as income or expenses per square foot. Commercial property operating statistics are relatively under-studied from an investment perspective. To demonstrate the methodology and its usefulness, we estimate the time series of property values, net operating income, cap rates, operating expenses and capital expenditures, per square foot of building area, by property type (sector) at a quarterly frequency for multiple specific geographic markets from 2004 through 2018. We show illustrative results for Los Angeles offices and Atlanta apartments. The methodology is essentially an extension and enhancement of the so-called "Pure Play" method introduced by Geltner and

*We thank Brad Case (AREUEA discussant), Louis Johner (ReCapNet discussant), and conference and seminar participants at the 2022 Real Estate Markets and Capital Markets (ReCapNet) Conference, 2022 REALPAC/Toronto Metropolitan University Research Symposium, 2021 AsRES/GCREC/AREUEA Joint Real Estate Conference and KTH Royal Institute of Technology for their helpful comments.

[†]Center for Real Estate, Massachusetts Institute of Technology (MIT). dgeltner@mit.edu.

[‡]Department of Economics and Business Economics, Aarhus University; Danish Finance Institute. akumar@econ.au.dk.

[§]Department of Finance, University of Connecticut. avdminne@uconn.edu.

Kluger (1998). It enables easy derivation of important basic data that should be useful for academic and industry practitioner analysts, derived from high quality stock market based information. The extensions and enhancements introduced here to the prior methodology allow estimation of actual quantity levels rather than just longitudinal relative values (index numbers). They also avoid the need for any data source other than published REIT data. Our methodology allows for an “additive” model structure that is more parsimonious to address the need for granular market segmentation. We also introduce a Bayesian framework that allows the estimation of reliable time series even in small markets.

Keywords: Real Estate Price Indices, Commercial Real Estate, REITs, Structural Time Series Modelling, Bayesian Inference, Real estate operating statistics, Capital expenditures, Operating income and expense

JEL classification: R30, C01, C11

I. Introduction

The primary purpose of this paper is to introduce, explicate, and demonstrate a new methodology for deriving operating and investment performance data for commercial investment property. A secondary purpose is to examine some salient features of the historical data that we derive to illustrate the methodology. Investors, policymakers, and economic analysts all benefit from having time series on market and operating fundamentals by real estate sector (such as property types) and/or geographic market. Fundamentals can include (among others): asset transaction prices, net operating income (NOI), capital expenditures (CAPEX), operating expenses (OPEX), and income yields.

Operating statistics in particular are relatively under-studied. Yet they are crucial for understanding commercial property asset and property management and the fundamental sources of property productivity and investment performance. But lack of transparency in the private real estate industry makes such time series hard to obtain. Thus, the methodology introduced here should be very valuable for industry and academic researchers and analysts.[1]

The present paper estimates market and operating fundamental time series from data provided in the S&P Global SNL Real Estate Database, which

[1]Two existing sources of commercial property operating data are noteworthy: the Building Owners & Managers Association (BOMA), and the National Council of Real Estate Investment Fiduciaries (NCREIF). While useful, these sources leave room for a new contribution. BOMA is not focused on the investment industry and lacks investment performance statistics. And the properties represented in both the NCREIF and BOMA data sources are somewhat special subsets of the broader population of medium-to-large scale (so-called "prime" or "Class A") investment property in the United States. For example, many of the properties covered by BOMA or NCREIF are owned or held by corporate users or private funds with deep pockets and unique objectives and constraints. Their optimal management may differ from that of other types of investors including in particular REITs, who tend to hold properties long term for maximization of total return.

tracks Real Estate Investment Trusts (henceforward "REITs") and provides detailed information about company level financials as well as property operating data for properties owned by REITs. The equity REITs we use are publicly listed firms that invest a large portion of their assets in, and derive a large portion of their income from, income-producing (i.e., investment) real estate. Tax law requires all REITs to have at least 75% of assets and income in/from real estate, and in fact in our data net (depreciated) real estate amounts to 89% of REITs' total book value of assets. (It is likely this percentage is even greater on a market value basis. See Feng and Liu (2021) for further discussion.)

Like other publicly registered companies, REITs must disclose financial information to investors and report on material business developments and risks on a timely basis. According to our data, total real estate value owned by REITs is approximately \$1.5T, or one-fifth of all investable real estate.[2]

With an eye to the statutory requirements of REIT status and the stock market's emphasis on earnings and dividends, listed equity REITs generally directly manage their owned investment properties to optimize cash flow through very long-term holding of the assets. This may result in somewhat different property operating performance characteristics than that of the NCREIF properties which are generally held by third-party investment management firms in a fund structure that may be more oriented toward initial yield and asset value retention through medium-term holding of as-

[2] In this paper we use the terms "commercial real estate" and "investable real estate" interchangeably and as defined by Real Capital Analytics Inc., a firm that is widely cited in industry analyses of commercial property transactions. The definition refers to properties that are owned and traded for investment purposes (including private rental apartments), excluding corporate and owner-occupied properties, and that are valued at at-least \$2.5M, thereby excluding "mom-and-pop investor" properties. Almost half of the total \$1.5T REIT holdings consists of the four major traditional investment property sectors that also dominate in the NCREIF portfolio: office, industrial (warehouse), retail, and apartments.

sets.[3]

Commercial property is physical capital, "bricks & mortar" is the vernacular phrase, a physical product whose real productivity underlies both its contribution to the economy and its performance as an investment, its nature as a financial product. Much of this productivity derives directly from the space markets in which the properties are located, the markets for the use and occupancy of built space. Therefore, data on operating statistics need to relate to specific types and locations of property in order to be most useful for analytical purposes. Yet, although REITs in the US do tend to specialize by property type, this specialization is not complete or "pure", and does not extend to the geographical dimension, where many REITs invest in multiple markets and locations. Thus, a downside of using REIT data directly is that their asset holdings typically do not exactly match one property type and/or geographical location, the type of granularity necessary for commercial property performance analysis. Finding fundamental time series for a *specific* sector and geographical market using such *firm* level data is thus not straightforward.

Geltner and Kluger (1998) (GK) were the first to use regression based techniques to produce "pure" sector level commercial property price indexes using REIT data. The essence of the GK methodology is simply to regress returns (r) of REITs (either total returns or capital returns) onto their holdings' shares (x) in different sectors, in a cross-sectional regression across all REITs, and to apply this regression in each period of time:

[3]Many NCREIF properties are held in closed-end funds that have finite lifetimes typically of 7 to 10 years, or in open-end funds where investors can cash out their units with relatively short notice. Most NCREIF properties are owned ultimately by tax-exempt institutions, whereas REIT-owned properties are ultimately subject to income tax paid on dividends and realized capital gains by taxable investors in REIT stocks.

$$r_{kt} = \sum_{i}^{I} \beta_{it} \times x_{ikt} + \epsilon_{ikt}, \tag{1}$$

for REIT k and sector i (which can be property type, location, or an interaction between the two) in index period t (which can be annual, quarterly, monthly, or even daily, based on stock market returns). Within each REIT and each period of time the x_{ikt} sum to unity across the sectors. The error term is denoted ϵ, and Eq. (1) can be estimated *period-by-period* in principle with Ordinary Least Squares (OLS). The estimated β_{it} coefficients are the estimates of the returns in period t to sector i.

In practice the REIT returns in the dependent variable have first been delevered (typically using a simple weighted average cost of capital – WACC – formula at the firm level). The holdings' shares x in principle are the fraction of each REIT's total assets value in each sector. This information is generally not directly available, but square footage of buildings in each sector typically is. Thus, the GK method is typically implemented by estimating the value shares by multiplying the square feet a REIT has in every sector with the average square foot price in that sector, and then dividing this by the sum of the value of all the REIT's properties combined, with the price/SF coming from a third-party data source based on private property transaction prices or appraised values.

While the GK regressions are cross-sectional, a longitudinal investment performance index is produced by chain-linking the results of the cross-sectional regressions across time. Thus for example, for an annual index, within each year (each cross-sectional regression), all the shares within each REIT (each observation of the regression) must sum to 1. Obviously, one of the sectors has to be dropped from Eq. (1) to circumvent the dummy trap.

The average square foot price in a specific sector is obtained from private market information, for example average private property transaction prices provided by a source such as Real Capital Analytics Inc. (RCA). The GK methodology was successfully used as described in Horrigan, Case, Geltner, and Pollakowski (2009) to create a commercial property investment return index product called the "PureProperty®" Index Series produced daily starting in 2012 by the FTSE/NAREIT joint venture.

It is important to recognize that in the GK methodology the index returns were reflective of stock market based valuations of the underlying commercial property sectors, as revealed by REIT share prices. Of course, commercial properties are traded directly in, and in that sense valued by, the private property asset market. But REIT shares indirectly reflect an alternative valuation of those same property assets, namely the valuation by investors and traders of REIT shares in the stock market. Thus, the GK/PureProperty indexes provided an alternate time series of investment property valuation (relatively speaking across time), different from private market based indexes including transaction price based indexes such as the RCA CPPI or CoStar CCRSI and appraisal-based indexes such as NCREIF or IPD/MSCI.[4] This alternate perspective was of great interest, because the stock market is in some respects more informationally efficient than the private property market, and the PureProperty indexes tended to lead the private market based indexes in time, at least regarding the major turning points in the market cycle.

Like the GK methodology, the enhanced methodology we propose in the

[4]RCA CPPI = Real Capital Analytics Commercial Property Price Index (based on the repeat-sale methodology described in Van de Minne, Francke, Geltner, and White, 2020), CCRSI = CoStar Commercial Repeat Sales Index, IPD = Investment Property Databank, owned by MSCI.

current paper retains the ability to reflect stock market valuations of commercial investment property. But the enhanced methodology goes farther by allowing not just longitudinal *relative* value changes to be derived as for an index number series, but actual average dollar values per building square foot, each period. And by including operating statistics such as income and expenses, the methodology is not limited to stock market valuations, as the operating information is based directly on the audited accounting data provided by the REITs.

In essence, we make several improvements to the GK methodology, apart from the greater parsimony of our specification. First, as noted, we focus on the money-valued *levels* of the fundamentals to be retrieved. That is, we plot the price or capital expenditures or other such values of interest in *dollars per building square foot* over time. Secondly, by focusing on levels per square foot (and yields for other metrics), we avoid the need for any other data source outside of the REIT data itself. As mentioned, with the GK method one needs to get an estimate of the property values held by REITs, which requires access to private property market valuation data (either transaction or appraisal based). This extra step, that might introduce noise and bias, is not needed in our proposed methodology. Third, the GK model was fully interactive (i.e. time $\times$ location $\times$ property type trends), our model also allows for fixed effects (for example: time + property type + location; or (time $\times$ property type) + location; or (time $\times$ property type) + (time $\times$ location)). Fourth, the GK methodology required bond market data for the de-leveraging process, whereas we avoid that necessity by using the firm's enterprise value (which includes both stock market capitalization plus book value of debt).

Finally and importantly, in a specific sub-specification of our approach we introduce a structural time series component (Francke and Van de Minne, 2017; Van de Minne et al., 2020). This Bayesian framework allows for more granular data series and higher frequency reporting, without suffering from excessive noise affecting the time series.

For the demonstration empirical analysis reported in this paper, we use quarterly data provided by S&P Global SNL Real Estate database covering the historical span from 2004 through 2018. The data and methodology allow us to produce the following fundamental time series in value levels per building square foot (SF): (1) asset values (proxied by enterprise value), (2) capital expenditures (CAPEX), (3) operating expenses (OPEX), and (4) net operating income yields (i.e. "cap rate", which can be multiplied times asset values to retrieve the underlying NOI level per SF). We compute these series using four different sub-specifications for two dissimilar markets: (1) offices in Los Angeles, and (2) apartments in Atlanta. The first market has plenty of observations, whereas the second lacks observations. We also show how our Bayesian structural time series enhancement enables quarterly time series for the same markets. Also, the specification does not impact the results as much as it did under the OLS framework.

The results are encouraging as we get realistic estimates at impressively granular levels. Thus, the promise of the methodology introduced here, to provide the ability to obtain a rich new source of vital commercial property performance and operating data for industry and academic researchers, seems to be realistic.

The remainder of this paper is organized as follows. Section II presents the methodology, followed by a discussion of the data in Section III. Section IV provides the demonstration of empirical results, and Section V concludes.

II. Methodology

As noted, the primary contribution of this paper is methodological. In this Section we will introduce and explicate the methodology we are proposing. Broadly, we develop four enhancements: (A) Focus on value levels rather than returns or relative changes; (B) Allowance for an "additive" (or "hierarchical") model structure that is more parsimonious to address the need for granular market segmentation; (C) Use of Bayesian techniques to further address noise in small-sample sizes; and (D) Application of a representative property fitting procedure to derive the final empirical ("pure price change") result. To introduce these enhancements, we will proceed one step at a time by describing four specifications that progressively elaborate, extend and enhance the basic perspective described in Eq.(1) in the previous Section. These steps are presented formally in Eqs.(2) through (5) in this Section.

A. A Levels Specification of the GK Methodology

Recall that, in contrast to the GK setup, which models REIT *returns*, we model the *levels* of REITs' assets and/or other variables of interest, normalized by, for example, square footage of built area. The general idea is that a REIT's assets in every sector sum up to the total assets of the REIT, where "sector" can be defined in various ways as allowed by the data. (Here, we for now will use the term "sector" and "market" interchangeably.) The coefficients - estimated from the data - on the different weights per sector (that sum up to 1) can be interpreted as the "values of interest" per sector. (Like square foot prices.)

More specifically, consider a panel of REITs $k = 1, \ldots, K$, and time $t = 1, \ldots, T$. For every REIT we observe P explained variables (values of interest) in $Y = \{y_1, \ldots, y_p\}$ (for example, enterprise value, net operating income, capex, and so forth) and covariates in X (in particular the shares invested in each sector). Take for example REITs' enterprise value (EV = debt + equity - cash). EVs should proxy for the underlying value of all the REITS' real estate assets. Assuming the square foot price of real estate times the square footage in every sector (i) adds up to the total enterprise value within every period, we have for any individual REIT k;

$$EV_k = \sum_{i}^{I} \text{Price per square foot}_i \times \text{Square foot}_{ik} + \epsilon_k, \tag{2}$$

ϵ is the residual term for REIT k, which is assumed to be normally distributed with mean zero and variance σ_ϵ^2. In this example, we would know the amount of square feet invested in every sector. Applying Eq.(2) in a cross-sectional regression across all K REITs within each period of time t allows us to estimate the square foot prices per sector, by Ordinary Least Squares regression (OLS), as the estimated coefficients on the sector shares. If we divide both the left and right hand sides of the equation by the total square footage of properties held by the REIT, then the RHS variables would be the *shares* of the REIT's holdings in each sector rather than the absolute square footage in each sector.

In words, the EV of a REIT in a specific year is the sumproduct (linear combination) of square footage of the REIT in every market (i) times the square foot price in the corresponding market. Across all REITs within period t, regress the REIT's enterprise value onto the square footage the REIT holds in every market. The estimated coefficients on the X variables (the square foot holdings) would thus provide the average (across the REITs) real estate values per square foot of built space for year t in each market or sector i represented by the X variables, as evaluated by the stock market.[5] Running the regression repeatedly in consecutive years, we could trace out the history of the square foot values over time, or we could normalize to an inception year and create a value change index across time for each sector i.

B. Normalization Per Square Foot & the Additive Model

A constraint of the Eq.(2) approach is that one cannot have *overlapping markets* in the same model (for example, the office sector aggregated across *all metros* as well as the aggregate of *all property type sectors* within the New York Metro). That is, we can't have an *additive* structure of geographical locations *and* property types (in statistical terminology, the model is limited to a single *"cluster"*). We either have to regress the square foot per location or per property type, or interact the two multiplicatively. If there are N types and M locations, we could have separate models for types and for locations, or we could have a single interacted model with N *times* M type-location combinations. However, previous literature has shown that the interaction typically results in a loss of too many degrees of freedom and - thus - results in noisy / unreliable indexes (Geltner and Kluger, 1998; Geltner and Ling, 2006). Yet the interesting analysis of commercial property markets is often at the level of both property type *and* geographic location. Of course, this criticism holds for the GK methodology as well.

As a next step, we therefore propose a more general approach that allows for more flexibility. First, divide the enterprise value by the amount of square feet (Sqft) per REIT per year. This gives us (by approximation) the real

[5] The EV largely reflects the stock market capitalization of the firm's equity, plus the value of the firm's debt, including mortgages. Although the debt is valued at book value, in general debt book value corresponds closely to market value of the debt.

estate value of REITs' per square foot of structure. Next, estimate the following *pooled* regression;

$$y_{kt} = \frac{\mathrm{EV}_{kt}}{\mathrm{Sqft}_{kt}} = \mu_t D^T_{kt} + \delta_i D^I_{kt} + \gamma_m D^M_{kt} + \beta X_{kt} + \epsilon_{kt}, \tag{3}$$

where D^T is a dummy (1/0) indicating the time period (t) of observation for REIT k, D^I is a matrix which contains the % of square footage the REIT has invested in property type i (at time t), and D^M is a matrix containing the % of square footage the REIT has invested in location m (at time t). Both D^I $(I \times N)$ and D^M $(M \times N)$ row sum to 1. The corresponding vector of coefficients are denoted μ, δ, and γ. Covariate matrix X contains any other explanatory variables that might impact the y variable that the analyst might want to control for (average age of properties, for example, if such data were available), with corresponding vector of parameters β. The stochastic term ϵ is assumed to be normally distributed with mean zero and standard deviation σ_ϵ.

μ_t is estimated as the money-valued (per square foot) vector of coefficients on a vector of time "fixed-effects", or time dummy-variables. Thus, the parameter μ_t traces out the common trend across time in *all* of the sector values, in money values per square foot. Note that we do not include a constant on the RHS, meaning that the μ_t values directly reflect the money-value per square foot in the common trend. As the other parameters do not vary across time, they therefore reflect only the inception period difference between each sector's square foot price and the common component represented in μ. Thus, Eq.(3) constrains all the sectors values to move longitudinally in lock-step. (Similar to Francke and van de Minne, 2017). Parameter δ (γ) therefore effectively give a separate constant per property type (location), of which we do need to omit one category to avoid the dummy trap.

One contribution of the Eq.(3) specification over Eq.(2) is that the covariates do not have to be limited to the square footage percentages. For example, an interesting second type of variable in X can be the average age of the buildings in the REIT's real estate portfolio (e.g., in years) to control for depreciation (Francke and van de Minne, 2017; Bokhari and Geltner, 2019). Age could be inserted linearly, or as a polynomial, or by splines, or using age cohorts (i.e. dummies), or semi-parametrically (for example a random walk). With such an enhancement, the interpretation of μ_t changes to be that of the age reference group in the market reference group, all relative to the base period.

Another advantage of Eq.(3) over Eq.(2) is that we conserve degrees of freedom, arguably resulting in less volatile indexes. We have, essentially, $T + I + M$ instead of $T \times I \times M$ variables. However, this comes at the previously noted price that Eq.(3) constrains all the property types and locations to move in lock-step. Commercial property space markets are highly segmented, especially by property type and location. Many interesting and important statistics characterizing commercial property investment and operating performance are best analyzed at a granular level, by specific market segment. To remedy this caveat we interact the time dummies (D^T) with property types (D^I) and/or location (D^M). For this paper we consider four different specifications of the baseline model (Eq. (3)), ranging from pure additive to pure interactive with two hybrid specifications in between:

time + pt + loc: $y_{kt} = \mu_t D^T_{kt} + \delta_i D^I_{kt} + \gamma_m D^M_{kt} + \beta X_{kt} + \epsilon_{kt};$

time x pt + loc: $y_{kt} = \mu_{it}(D^T_{kt} \times D^I_{kt}) + \gamma_m D^M_{kt} + \beta X_{kt} + \epsilon_{kt};$

time x pt + time x loc: $y_{kt} = \mu_{it}(D^T_{kt} \times D^I_{kt}) + \mu_{mt}(D^T_{kt} \times D^M_{kt}) + \beta X_{kt} + \epsilon_{kt};$

time x pt x loc: $y_{kt} = \mu_{imt}(D^T_{kt} \times D^I_{kt} \times D^M_{kt}) + \beta X_{kt} + \epsilon_{kt}.$

The names of the (sub)specifications are given in bold. The first specification ("time + pt + loc") is similar to Eq. (3). In the second specification ("time x pt + loc") we interact the time dummies with property type holdings but treat the locations as additive. The third specification ("time x pt + time x loc") interacts the time dummies with property holdings and additively interacts the time dummies with the location holdings. In the final specification ("time x pt x loc") we interact all our variables of interest.

In our example empirical analysis in the present paper we show results for a specification based on the four main property types and the 15 largest metro regions (in terms of population). As a result, in the first specification we are estimating only one trend (the sectors differ only by a constant). In the second specification we are estimating 4 trends by property type with location differences only as constants (no common trend). In the third specification we are estimating 4 property type + 15 metro areas = 19 trends. And in the last specification we are estimating 4 property type times 15 metro areas = 60 trends. The common trend μ_t is only estimated in the first specification ("time + pt + loc").

Going from the first specification to the last we lose degrees of freedom with the advantage of gaining more flexible/unique trends. However, less degrees of freedom would likely give more noisy results for many market segments of interest (Geltner and Ling, 2006). We make no upfront assumption of what specification is the "best", i.e. what the best balance between flexibility and degrees of freedom is. As explained in Van de Minne et al. (2020) and Guo, Zheng, Geltner, and Liu (2014), finding the "best" index is not straightforward, and depends mostly on personal preference.[6] Thus, in keeping with the purpose of the present paper, to introduce the methodology, we limit ourselves here to documenting how for the various specifications the resulting indexes look for a selection of cities.

C. A Bayesian Extension for Subtrends

While the Eq.(3) model and subsequent interactive modifications offer interesting possibilities, the need to balance conservation of degrees of freedom versus production of more granular and versatile indexes (Geltner and Ling, 2006) can be essentially avoided by introducing a Bayesian component into the model (Francke and Van de Minne, 2017).

More specifically, we follow recent literature on advances made in structural time series models, by adding (1) a Bayesian common trend, and (2) a structural time series component to all time series variables. To start with the first, we add a (common) trend that goes through all observations. The main benefit of using such a Bayesian common trend is that the sub trends will gravitate towards this common trend if there is not enough evidence that the sub trends really deviate from the common trend. This happens if the observations within a sub trend within a given period pro-

[6] The issue is that most measures of fit (like the R^2 and RMSE) are essentially cross-sectional in nature and not longitudinal. One way to compare index fit is to look at index revisions (Van de Minne et al., 2020). However, in our case we have no such revisions to begin with.

duce an indecisive signal. We designate the common trend κ_t which enters in all (sub)specifications.[7] (See Appendix A for details on how the Equations change with this addition.) This has proven to be especially powerful in small/noisy markets, see Francke and Van de Minne (2017) for example. Secondly, introducing a structure to the time series, we assume all the (sub)trends follow random walks;

$$\Delta\theta^k \sim \mathcal{N}(0, \sigma_k^2), \tag{4}$$

with $\theta^k = \{\mu_t, \kappa_t\}$. We estimate all our (sub)specifications using the Bayesian extensions.

We estimate the Bayesian model using the No-U-Turn-Sampler (NUTS) developed by Hoffman and Gelman (2014). For more details on the estimation and specification, see Appendix A.

Note that such "hierarchical" models have been employed in real estate literature before, in both hedonic (Francke and DeVos, 2000) and repeat sales Francke and Van de Minne (2017) settings. The Bayesian enhancement is very helpful for dealing with the small sample size challenge posed by commercial property market segments.

D. *Retrieving Trend Levels: Fitting the Representative Property*

In this paper we are interested in the *levels* of our explained variables per sector i per period t. Parameters μ_t, δ, γ and β_{it} in Eq. (3) give, respectively, the common trend and the deviation from that trend. But they do not directly in themselves give a value series for a well-defined property

[7] The only exception is in our additive specification called "time + pt + loc" where parameter μ is already a common trend in itself. Thus, in this case κ replaces μ.

sector, holding other things constant through time. That is, we want to produce what economic statisticians refer to as pure (constant-quantity or constant-quality) "price" levels across time. In order to calculate such pure price-changes in levels for every sector we must first define and construct a *"representative property"* (de Haan and Diewert, 2011).

Constructing a representative property, and subsequent "chain" parameters to retrieve price or constant-quality value levels across time, is a well established field in economic statistics literature (McMillen, 2008; de Haan and Diewert, 2011). In this paper, the representative property will "simply" be the average property as held by all the REITs over the entire 2004-18 period (see Section III below for more details on the average property). For example, in our application the property is aged approximately 20 years old, and it reflects a weighted average location which happens to be 8.7% in the New York metro area for example. (Note that the age of the representative property does not advance across the history of price estimates, hence, the price changes do not reflect depreciation.) If we want to estimate a price series of office properties, we set the office holding (D^I) at 1, and the other property type sector dummies all to zero. This is similar to the setup in Geltner and Van de Minne (2017).

Let's denote the vector of characteristics for this average property as $\bar{Z}$. Note that $\bar{Z}$ has no subscripts as this average property does not change per sector or per period (nor per REIT). By fixing the representative property, the resulting time series reflect "pure" price changes.[8] To retrieve an office

[8]Economic statisticians draw a distinction between "value" and "price" in which value equals price times quantity and can change over time due either to change in price or change in quantity, whereas price reflects a constant quantity. In the case of a complex good like real estate, "quantity" effectively includes many quality dimensions. For example, the "value" of a building that sold at two points in time could change in part to reflect the aging of the building, reflecting a depreciation or reduction in the "quantity"

price series for Los Angeles (in dollar levels per period) we thusly compute:

$$\hat{y}_{i=\text{office},m=\text{Los Angeles},t} = \hat{\mu}_t D^T + \hat{\delta} D^I_{i=\text{office},t} + \hat{\gamma} D^M_{m=\text{Los Angeles},t} + \hat{\beta}\bar{Z}. \quad (5)$$

We can redo these calculations for any representative property we desire. For example, to create a series for Los Angeles office properties, one could tailor the representative property characteristics in $\bar{Z}$ to reflect those of Los Angeles office properties. In the present paper, due to data availability constraints and to keep the presentation simple, we keep $\bar{Z}$ always just reflecting the average REIT-held property over our 2004-18 history.

III. Data and Descriptive Statistics

To demonstrate the estimation of market and operating fundamentals of commercial real estate properties in the U.S., we examine all publicly traded equity REITs available from 2004 through 2018. We use REIT level financial data from the S&P Global SNL REIT Financial database merged with REITs' properties holding data from the S&P Global Real Estate Properties database, at both annual and quarterly frequency. This database provides data for each commercial property held by a listed equity REIT. We start our sample from 2004 because that is when asset level information at the quarterly frequency begins in that database.

To define our property holdings shares variables (D_{kt} of Eq.(3)), we

of building, and such change in value in itself would not represent a "price" change in the economic statistics terminology. The "price" versus "value" distinction is different in real estate, where "price" refers to a transaction of exchange of ownership, while "value" may refer to a "valuation" estimated or indicated by some source, such as in our case, the stock market share prices of REITs. In this paper we use the two terms somewhat interchangeably, though in the present section we are attempting to introduce the distinction.

calculate, for each REIT, the percentage of its property portfolio based on building square footage, by both geographical location (metropolitan statistical areas – MSA) and property types. For this purpose we choose the top 15 MSAs by 2014 population.[9] All the remaining geographical locations outside those 15 metros are lumped into one "Other" category of location. We also construct asset holdings by the 4 major core investment property types: office, retail, industrial, apartments, plus a catch-all "other" category. We define the percentage of the REIT's portfolio in each MSA as the ratio of a REIT's holding (in SqFt) in a given MSA-period divided by total REIT holdings (in SqFt) in that period. Similarly, we define REITs percentage holding shares by property type as the ratio of the REIT's holding (in SqFt) by property type in a given period divided by that REIT's total holdings (in SqFt) in that period.

Enterprise value (EV) is defined as the stock market capitalization of ongoing operations, including common equity share capitalization at market value and all non-common equity, debt, and mezzanine at book value, less cash and cash equivalents at book value. The income yield is defined as the firm's net operating income (NOI) for all real estate operations divided by the firm's enterprise value. Yields based on annual net operating income are also referred to as "capitalization rates" (or "cap rates" for short) and are a widely used and published metric in the real estate investment industry.[10]

[9] Based on the population data from the U.S. Census website, these MSAs are: New York-Newark-Jersey City, Los Angeles-Long Beach-Anaheim, Chicago-Naperville-Elgin, Dallas-Fort Worth-Arlington, Houston-The Woodlands-Sugar Land, Philadelphia-Camden-Wilmington, Washington-Arlington-Alexandria, Miami-Fort Lauderdale-Pompano Beach, Atlanta-Sandy Springs-Alpharetta, Boston-Cambridge-Newton, San Francisco-Oakland-Berkeley, Phoenix-Mesa-Chandler, Riverside-San Bernardino-Ontario, Detroit-Warren-Dearborn, and Seattle-Tacoma-Bellevue.

[10] In practice, "forward-looking" cap rates are most common, based on Year "t+1" NOI and end of Year "t" property asset valuation, though "backward-looking" cap rates (based on Year "t" NOI and end of Year "t" property asset valuation) are also employed. In our

CAPEX is defined as the SNL variable labeled "non-revenue generating capital expenditures incurred". Operating expenses (OPEX) is the SNL variable reflecting total expenses (as distinct from "expenditures") resulting from operating and maintaining all real estate assets. Both CAPEX and OPEX are also expressed as a rate, by dividing said variables by enterprise value. Property age is defined as the number of years from the "property built year".

After compiling panel data of REITs' holdings per quarter between 2004 and 2018 we end up with $6,585$ observations between 2004 and 2018 that have data on *at least one* of our 4 main target variables: (1) enterprise value per square foot (EV/SqFt), (2) net operating income yield (NOI/EV), (3) capital expenditures divided by enterprise value (CAPEX/EV), and (4) operating expenses divided by enterprise value (OPEX/EV). We express all our yields and percentage rates on an annualized basis, as that is common practice in industry, by multiplying the quarterly "flow" variables by 4. Flow values in Period t are divided by EV as of end of Period t (where the period is either quarter or year).

The number of observations per quarter (for which we have at least 1 of our main target variables) is given in Figure 1, and some descriptive statistics of the main dependent variables are given in Table I. Note that the panel is quite unbalanced in multiple ways. First, the number of observations gradually increases over the years. Also, we have more observations in Q4 for every year, with on average 30% more. And some of our dependent variables are more often reported than others. (See the bottom row of Table I.) For example, for almost the full sample we observe net operating income based yields of the REITs ($5,628$ observations in total). In contrast, enterprise

results, effectively, our yields are backward-looking.

value per square foot (EV/SqFt) is available for only $4,067$ observations where the main limiting variable is the amount of square feet of real estate in a REIT's portfolio.

[Place Figure 1 about here]

[Place Table I about here]

We observe in Table I that enterprise value (our indicator of the stock market's valuation of property asset value) per square foot (EV/SqFt) has been trending upward, from \$158 in 2004 to \$ 384 in 2018, more than doubling. The Global Financial Crisis (GFC) is also clearly visible, with EV/SqFt dropping from \$209 in 2007 to \$ 161 two years later. Of course this history is "apples versus oranges", just the raw material from which to produce our "pure price change" data series and indexes. We also see a spike in 2009 in the net operating income yields (NOI/EV), reflect the jump in cap rates as asset values plummeted in the GFC. Capital expenditure (CAPEX/EV) and operating expense (OPEX/EV) rates have generally trended slightly downward (no doubt reflecting the denominator effect of asset value growth). Note that the average age of REITs' portfolios in our sample went up by 6 years in a 14 year period (2004 – 2018), meaning that *some* new(er) properties entered the REITs' portfolios over the period (and/or older ones exited).

In Table II we provide some descriptive statistics of the covariates that define the "sectors" in our data set, based on building square-footage of floor area.[11] Table II highlights our five property type sector categorization

[11]We use the SNL variable labeled "Property Size", which is defined as the total interior area of the building or buildings in square feet.

(office, industrial, retail, apartment, and remaining property types grouped into "other") in Panel A, and our 16 geographic location sectors (15 largest MSAs by population plus "other") in Panel B. Note that the fractions *within* each panel sum to 1. Retail is the largest property type (29%), followed by office (26%). The smallest share of investment by square footage is in the apartment category, with 11.2% across all the REITs. Of the 15 largest MSAs, REITs have the most square footage in New York with almost 9% across all REITs over the sample period on average.

The final row of both Panels gives a Herfindahl-Hirschman index (HHI) of the average REIT in terms of the property types and the MSAs. Within each REIT, each period, we sum the square of the REITs' holdings shares per year (within property types and within MSAs separately). The resulting HHI says something about concentration; an HHI of 1 would mean that every REIT (in every year) only invests in one of the categories. Thus, the average HHI of around 0.9 in Panel A indicates that the REITs are mostly concentrated (or specialized) in one property type, while the HHI of 0.5 in Panel B indicates that REITs tend to be more diversified by location.

[Place Table II about here]

Finally, consider Figure 2. Every dot on the Figure represents geographical location of a property owned by REITs in our data as of 2018. In this figure, we present geographical distribution of the four main investment property types: apartment, retail, office, and industrial in Panel (a), (b), (c), and (d), respectively. (The authors will provide a map for the "other" category upon request.) The geographic scope of REITs' property holdings across the US is notable in this Figure. In particular, REITs' property ownership is not limited to the major cities but it is much diversified as shown in Figure 2 and also indicated by the HHI number in panel B of Table II (though in terms of dollar value, we can expect a greater concentration in high-price cities).

[Place Figure 2 about here]

IV. Results

While the main contribution of this paper is the proposed methodology as described in Section II, it is of interest to show some example empirical results using that methodology, both to demonstrate how the methodology works, and because the illustrative results may hold some interest in their own right. The results shown in this section will be presented roughly in the same step-wise progression by which the methodological enhancements were presented in Section II.

A. Comparing Level Specification with GK Methodology

In this Section we show the results for our base "level" specification (Eq. (2)), which is estimated on a year-by-year basis. Given that the model is similar in spirit to the FTSE/Nareit PureProperty Index, we also plot the that index which was commercially produced historically.[12] Note that our methodology produces square foot prices (black line, left-axis), whereas the PureProperty is an index (red dotted line, right-axis). The PureProperty data series ends during 2018. The resulting trends are plotted in Figure 3.

[12]Keep in mind that FTSE/Nareit PureProperty used a different REIT property dataset, and also used an additional dataset - from Real Capital Analytics Inc transaction prices - to compute the square foot holdings per REIT. Hence, it is not expected that the indexes will be exactly the same.

[Figure 3 about here]

On average we find that office had the highest price per square foot (\$280), and industrial the lowest price per square foot (\$40). This is in line with expectations. In all models the Great Financial Crisis (GFC) is clearly visible, with the exception of apartment (Figure 3a). The recovery from the GFC is estimated to be a year later compared to the PureProperty method. The correlation between the returns per property type for the two index methodologies is highest for retail with +0.6, and even negative for office, with -0.1. (Although note that we only have 13 observations for this exercise.) Industrial property prices per square foot also look more volatile than what seems reasonable (Figure 3d).

All in all, the resulting indexes from our base "level" specification are a bit of a mixed bag. The apartment index (Figure 3a) hardly shows any effect of the GFC, and reveals a large 40% price decline in the final year. Industrial prices seem too volatile (Figure 3d). On the other hand, the office and retail indexes arguably "look" better compared to the original FTSE PurePlay indexes. Although as noted earlier, it hard is to say which index is "best." It is probably not a surprise that office and retail are the two best populated property types in our dataset, revisit Table II. Period-by-period regression models are known to give volatile and/or noise-driven results in scarce data environments (de Haan and Diewert, 2011). We therefore turn our attention to the "pooled" regression models next.

B. Results for our Ordinary Least Squares Methodology

In the previous Subsection, we have only estimated a model with a single cluster, namely property type at the national level. By giving national aggregate results, this model is useful to get a comprehensive picture of the nature and reasonableness of the estimation of the empirical performance of commercial properties held by U.S. REITs. But we have previously noted the value for commercial property of more granular sectors identified by both property type *and* geographic location. We noted in Subsection II.B that an additive specification can be useful to enable such granularity. Now we present an illustration of this using our pooled regression models.

To start with, we show results of the additive model using OLS methods. Figure 4 presents these results for the office property type in the Los Angeles metro, which provides time series for Enterprise Value per square foot, Net Operating Income/Enterprise Value, Capital Expenditure/Enterprise Value, and Operating Expenses/Enterprise Value in panels (a), (b), (c), and (d), respectively. For each of these time series, we use 4 different specifications as discussed in Subsection II.B: (1) time + pt + loc; (2) time x pt + loc; (3) (time x pt) + (time x loc); and (4) time x pt x loc; where pt is Offices and loc is Los Angeles. As noted, as we move from specifications (1) to (4), we loose degrees of freedom as (4) is a fully interacted model. This is evident in figure 4, where the time series from specification (4) is most volatile, while it is not so volatile as we consider the additive model. This shows the advantage of employing the additive model over a fully interactive model. Given our limited number of observations, and the fact that fractions within certain markets are already quite low, the fully interactive specification using the OLS method is not advised. In some cases we even find negative values, which should not be possible in reality.

[Figure 4 about here]

For comparison purposes, in Figure 5, we also plot the same time series as in Figure 4 but for Apartment property types in Atlanta. This is a much smaller market segment than Los Angeles offices (in terms of REIT property holdings), and therefore provides a sort of lower bound picture of what the methodology can handle. As Table II shows, among property type sectors, apartments present lowest share and Atlanta is not well populated either, therefore it is not surprising that the OLS method generates very noisy time series (especially from the fully interactive specification). The time series from other additive models are reasonably better as compared to the fully interactive model, but compared to a bigger market (e.g. offices in Los Angeles as presented in Figure 4) they are more volatile. For example, the (time x pt) + (time x loc) model still gives reasonable results for offices in Los Angeles, but less so for our smaller Atlanta apartment market.

The model accurately estimates higher property values (i.e. enterprise value) for LA offices than Atlanta apartments, although prices in Atlanta did increase more in later years. "Cap rates" (NOI / EV) are higher on average (6%) for offices in Los Angeles compared to apartments in Atlanta (3.5%).[13] OPEX spending shows a reverse relationship. The average OPEX spending as a % of EV for apartments in Atlanta is 5%, whereas for office in Los Angeles it is only 3%. CAPEX spending as a fraction of EV is around 1% per year for both markets.

There are also some interesting insights in the cyclicality of OPEX and CAPEX which is understudied in the literature. For example, we find that during the GFC, CAPEX spending went down (for offices in Los Angeles most notably), but OPEX spending went up as a fraction of EV. The latter is explainable by realizing that many expenses still have to be paid even

[13] These averages are based on the time + pt + msa and time x pt + msa models only, as these give the most realistic looking results.

if property values decline, such as utilities, insurance and property taxes. CAPEX spending, on the other hand, is discretionary and affected by new tenant signings. (Via tenant improvements and brokerage fees.) Thus, when demand for space decreases, so does CAPEX spending, partly mitigating the hit in rents. Although note that CAPEX spending for apartments in Atlanta as a fraction of property values hardly changed.

Overall, results in Figure 4 and 5 show that even using the OLS method, additive models perform much better than fully a interacted specification for levels of granularity at the metro level by property type, for large metros and major property types. However, when one tries to estimate these time series for smaller market segments in smaller metros, even the additive models present volatile indexes. Therefore, next we show how using a Bayesian estimation technique, one can generate reliable indexes for smaller markets.

[Figure 5 about here]

C. Results for our Bayesian Methodology

In this Subsection, we use the Bayesian method described in Subsection II.C to reproduce our main time series (i.e. enterprise value per square foot, net operating income yield, capital expenditure/EV, and operating expenses/EV). Figure 6 presents these results for office properties in Los Angeles while Figure 7 plots these series for apartment properties in Atlanta. To demonstrate the strength of such Bayesian models even further, we also estimate the times series on a quarterly frequency, instead of a yearly frequency.

[Figure 6 about here]

[Figure 7 about here]

As these results show, using the Bayesian approach, we are able to generate time series which are less volatile even on a quarterly frequency. As discussed in the Methodology Section, there is a trade-off between using different specifications, as with a more interacted model, versus a loss of too many degrees of freedom which results in more volatile series. But as shown in Figure 6 and 7, the Bayesian approach mitigates these concerns to a great extent, and when combined with an additive model, one can often obtain a best set of empirical time series for many applications. We also do not end up with negative values for any of the models. The fully interactive model (time x pt x msa) is not the most volatile model anymore on average. The technical insight is that when there is not enough evidence in the data (i.e. noisy observations) to prove otherwise, the series estimations will revert to the common trend, see Appendix A.

Finally, note that most trends remain roughly the same, even after using more flexible (Bayesian) specifications. Some differences in the timing of the movements might change, but the overall level of the estimates remain similar. One notable exception is the CAPEX/EV for apartments in Atlanta. In fact, we find an explosion in CAPEX spending when using more flexible specifications, most notably for the (time x pt) + (time x msa) model in Figure 7c. This was partly visible in the OLS model as well, see the green line in Figure 5c. Thus, when prices fell in this market, CAPEX remained same or went up slightly, resulting in a high CAPEX/EV.

V. Concluding Remarks

In this paper we propose a new methodology that allows the estimation of fundamental real estate time series using REITs' financial and asset holding data. This is the central purpose of this methodology and main contribution of this paper: to enable and promote a new and rich source of performance and operating data for commercial property in the US by the use of this stock market based asset performance retrieval methodology. The methodology "*purifies*" the data in an important sense, enabling representation of specific property types and/or geographic locations.

In the present paper, we examine all the commercial properties owned by the universe of publicly traded U.S. equity REITs during 2004 – 2018. Based on that data, we estimate time series of property values, cap rates, operating expenses and capital expenditures, per square foot of building area, by property type (sector) at the quarterly frequency and for specific geographic markets. We present illustrative results for Los Angeles offices and Atlanta apartments. More exploration of the technique may yield interesting specific findings, but our initial analysis indicates that granularity at the level of four property types within geographic areas of sizes down to individual metro areas at least for a dozen or so of the largest metros, or other such clusterings, can produce good results. The key point is that, while REITs' property portfolios are not limited to only one property type and/or geographical location, the proposed methodology constructs essentially "pure" fundamental time series for a specific property type and/or geographical location.

We accomplish this by proposing several important enhancements to the pure play method introduced by Geltner and Kluger (1998) (GK). First, in addition to generating indexes (as in GK), which are effectively limited to total returns and capital returns data, our method also allows us to estimate money-valued levels and ratios of the fundamentals, and can be applied to any performance or operating variable that is available in the REIT database, ranging from enterprise value (property asset value) to net operating income, CAPEX, and OPEX. Second, unlike the GK approach, which had to combine REIT data with private property market valuation and other external data sources, and was therefore more prone to noise or error, we do not need any additional data sources besides the REIT data itself. Third, while the GK model was limited to only interactive (single-cluster) specifications, our model also allows for fixed effects and multiple-cluster additive specifications which are more parsimonious and effective for smaller datasets. Finally, we introduce a Bayesian estimation framework that allows us to produce reliable time series even in smaller markets.

We present results using both a one-level cluster (i.e. property type only) and a two-level cluster (i.e. property type in a particular MSA) specification. Though not presented here, it is clear that one can also produce derivative data series, such as effective gross income, total return, and a free cashflow-based yield time series. (Results available from authors upon request.)

While most of the existing literature in the investment industry is focused on return based indexes, our paper contributes by providing a methodology that generates both operating and investment related time series. Such operating statistics are important for understanding not only commercial property management but also the fundamental sources of property productivity and investment performance.

REFERENCES

Betancourt, M., and M. Girolami, 2015, Hamiltonian Monte Carlo for hierarchical models, *Current trends in Bayesian methodology with applications* 79, 30.

Bokhari, S., and D. Geltner, 2019, Commercial buildings capital consumption and the United States national accounts, *Review of Income and Wealth* 65, 561–591.

de Haan, J., and W. E. Diewert, 2011, *Handbook on residential property price indexes*, chapter Hedonic regression methods, 50–64 (Eurostat Methodologies & Working papers).

Feng, Zifeng, and Peng Liu, 2021, Introducing "Focused Firms": Implications from REIT prime operating revenue, *The Journal of Real Estate Finance and Economics, Forthcoming* .

Francke, M. K., and A. M. Van de Minne, 2017, The hierarchical repeat sales model for granular commercial real estate and residential price indices, *Journal of Real Estate Finance and Economics* 55, 511–532.

Francke, M. K., and A. M. van de Minne, 2017, Land, structure and depreciation, *Real Estate Economics* 45, 415–451.

Francke, M.K., and A.F. DeVos, 2000, Efficient computation of hierarchical trends, *Journal of Business and Economic Statistics* 18, 51–57.

Gelman, A., 2006, Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper), *Bayesian analysis* 1, 515–534.

Geltner, D., and B. Kluger, 1998, REIT-based pure-play portfolios: The case of property types, *Real Estate Economics* 26, 581–612.

Geltner, D., and D. Ling, 2006, Considerations in the design and construction of investment real estate research indices, *Journal of Real Estate Research* 28, 411–444.

Geltner, D., and A. M. Van de Minne, 2017, Do different price points exhibit different investment risk and return commercial real estate, *Journal of Portfolio Management* 43, 105–119.

Guo, X., S. Zheng, D. M. Geltner, and H. Liu, 2014, A new approach for constructing home price indices: The pseudo repeat sales model and its application in China, *Journal of Housing Economics* 25, 20–38.

Hoffman, M. D., and A. Gelman, 2014, The No-U-turn sampler: adaptively setting path lengths in hamiltonian monte carlo., *Journal of Machine Learning Research* 15, 1593–1623.

Horrigan, H., B. Case, D. Geltner, and H. Pollakowski, 2009, REIT-based property return indices: a new way to track and trade commercial real estate, *The Journal of Portfolio Management* 35, 80–91.

McMillen, D. P., 2008, Changes in the distribution of house prices over time: Structural characteristics, neighborhood, or coefficients?, *Journal of Urban Economics* 64, 573–589.

Van de Minne, A. M., M. K. Francke, D. M. Geltner, and Robert White, 2020, Using revisions as a measure of price index quality in repeat-sales models, *Journal of Real Estate Finance and Economics* 60, 514–553.

Tables

Table I. Averages of Main Variables per Year. Notes: *EV:* enterprise value (mkt cap + debt - cash) per square foot, *NOI:* Net Operating Income / Enterprise Value (i.e. "cap rate"), *CAPEX:* capital expenditures / Enterprise Value, *OPEX:* operational expenses / Enterprise Value, and age: is the average age of properties within REITs' portfolios. The yield (NOI) and percentage rates (CAPEX and OPEX) are expressed yearly (by multiplying the quarterly flow variables by 4 if necessary.)

year	**N**	**EV**	**NOI**	**CAPEX**	**OPEX**	**age**
2004	365	$ 158.42	6.44%	0.93%	3.27%	21
2005	391	$ 161.55	6.46%	0.65%	3.20%	19
2006	371	$ 194.64	5.72%	0.57%	2.90%	20
2007	345	$ 209.05	5.72%	0.43%	2.83%	21
2008	375	$ 190.15	6.81%	0.51%	3.34%	21
2009	395	$ 161.83	7.86%	0.57%	4.04%	23
2010	396	$ 190.37	6.68%	0.67%	3.54%	23
2011	432	$ 204.88	6.49%	0.72%	3.41%	24
2012	456	$ 253.09	6.37%	0.68%	3.20%	24
2013	471	$ 272.81	6.14%	0.71%	3.06%	24
2014	499	$ 307.61	6.11%	0.65%	3.02%	25
2015	520	$ 355.17	6.11%	0.59%	2.96%	26
2016	525	$ 387.52	5.82%	0.53%	2.65%	26
2017	526	$ 405.98	5.78%	0.54%	2.60%	27
2018	518	$ 384.22	6.14%	0.61%	2.81%	27
N	6,585	4,067	5,628	3,966	5,606	6,066

Table II. Sector-wise Descriptive Statistics. This table reports sector-wise (by property type in Panel A and by geographical location in Panel B) descriptive statistics based on the building square-footage of floor area owned by REITs in our sample for the period 2004 – 2018. HHI is the Herfindahl-Hirschman Index, measured by summing the sectors squared per REIT per year. A higher HHI indicates more concentrated investments.

Panel A: Property Types					
	mean	**sd**		**mean**	**sd**
Office	0.264	0.381	Apartment	0.112	0.296
Industrial	0.133	0.289	Other	0.199	0.284
Retail	0.293	0.421			
HHI			0.901		
Panel B: Metropolitan Statistical Areas					
	mean	**sd**		**mean**	**sd**
New York-Newark-Jersey City	0.087	0.209	Atlanta-Sandy Springs-Alpharetta	0.041	0.105
Los Angeles-Long Beach-Anaheim	0.048	0.138	Boston-Cambridge-Newton	0.020	0.067
Chicago-Naperville-Elgin	0.041	0.110	San Francisco-Oakland-Berkeley	0.024	0.068
Dallas-Fort Worth-Arlington	0.037	0.085	Phoenix-Mesa-Chandler	0.020	0.055
Houston-The Woodlands-Sugar Land	0.036	0.100	Riverside-San Bernardino-Ontario	0.009	0.042
Philadelphia-Camden-Wilmington	0.031	0.096	Detroit-Warren-Dearborn	0.013	0.069
Washington-Arlington-Alexandria	0.070	0.165	Seattle-Tacoma-Bellevue	0.013	0.045
Miami-Fort Lauderdale-Pompano Beach	0.026	0.087	Other	0.484	0.299
HHI			0.516		

Figures

Figure 1. Observations per Quarter. This figure displays the number of quarter-wise observations in our data, where atleast one of our main target variables (enterprise value per square foot, net operating income yield, capital expenditure divided by enterprise value, and operating expenses divided by enterprise value) is non-missing for the period 2004 – 2018.

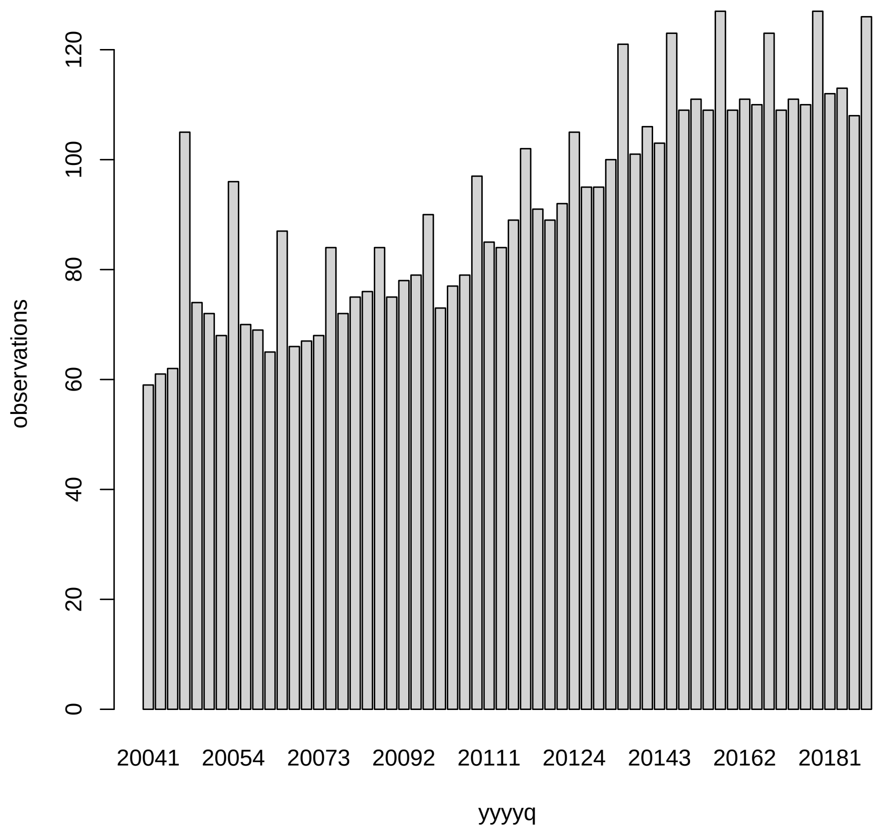
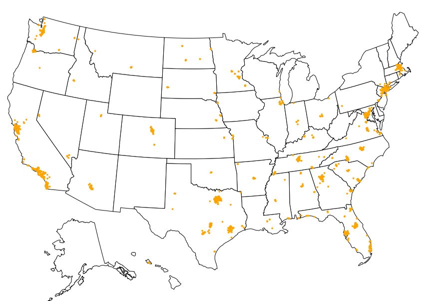
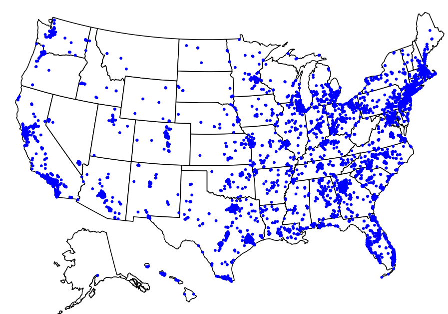


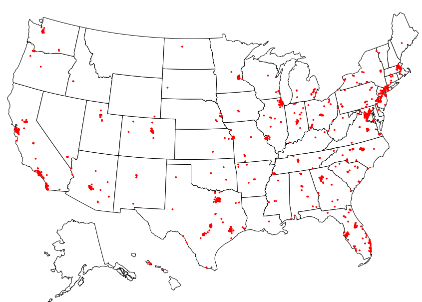
Figure 2. Location of REITs' Assets. This figure displays geographical location of REITs' assets in our data as of 2018 for main four property types (apartment in Panel (a), retail in Panel (b), office in Panel (c), and industrial in Panel (d)). Every dot represents one property owned by a REIT in our sample.



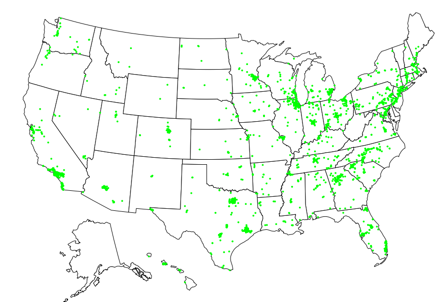
(a) Apartment



(b) Retail



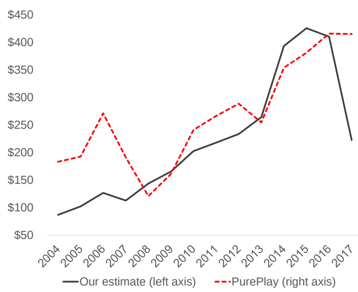
(c) Office



(d) Industrial

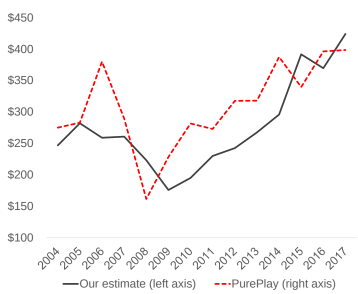
Figure 3. Enterprise Value per Square Foot using Levels Specification. This figure plots national level estimates for enterprise value per square foot using our base level specification (Eq. (2)) (left axis) and compare them with the official FTSE NAREIT index (right axis). Panel (a), (b), (c), and (d) plot the estimates for property type: apartment, retail, office, and industrial, respectively. Sample period for this figure is 2004 – 2017 as the FTSE NAREIT index was discontinued in early 2018. The FTSE NAREIT PurePlay indexes on right axis are fixed to be 100 in the base year 2004. Left axis presents $ value per square foot as estimated using Eq. (2).

(a) Apartment



(b) Retail

(c) Office



(d) Industrial

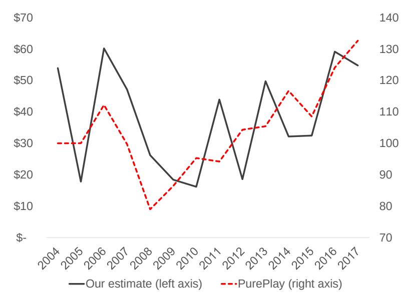
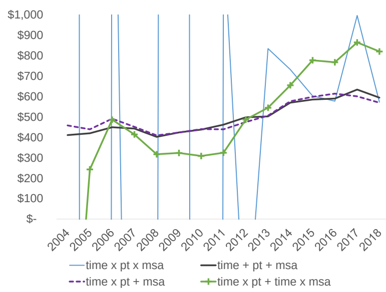
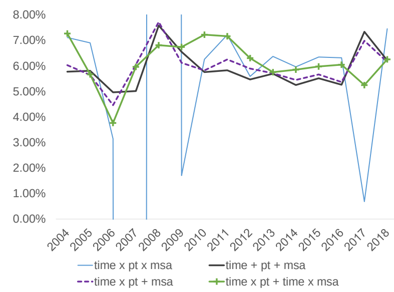


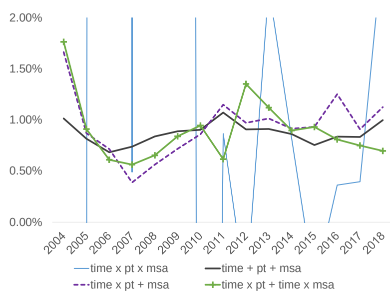
Figure 4. Results for Offices in Los Angeles using OLS Methods. This figure presents results for office property types in Los Angeles obtained from the additive structure model using OLS methods, see equation (3) and related 4 sub-specifications with different additive structure. Panel (a), (b), (c), and (d) display time-series for enterprise value per square foot, net operating income/enterprise value, capital expenditure/enterprise value, and operating expenses/enterprise value, respectively. Vertical axis represents the corresponding values (either $ per square foot or percentage over enterprise value), while horizontal axis shows different years for the period 2004 – 2018.



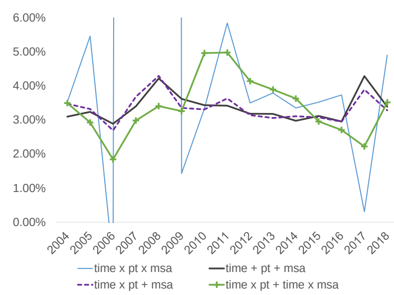
(a) Enterprise Value / SF



(b) NOI / EV



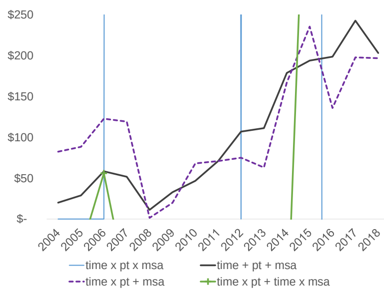
(c) Capex / EV



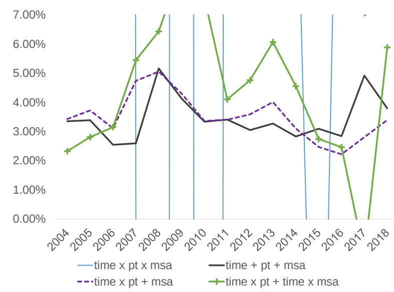
(d) Opex / EV

Figure 5. Results for Apartments in Atlanta using OLS Methods. This figure presents results for apartment property types in Atlanta obtained from the additive structure model using OLS methods, see equation (3) and related 4 sub-specifications with different additive structure. Panel (a), (b), (c), and (d) display time-series for enterprise value per square foot, net operating income/enterprise value, capital expenditure/enterprise value, and operating expenses/enterprise value, respectively. Vertical axis represents the corresponding values (either $ per square foot or percentage over enterprise value), while horizontal axis shows different years for the period 2004 – 2018.

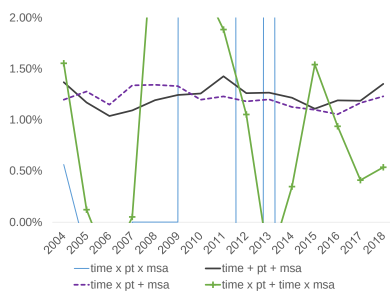
(a) Enterprise Value / SF



(b) NOI / EV



(c) Capex / EV



(d) Opex / EV

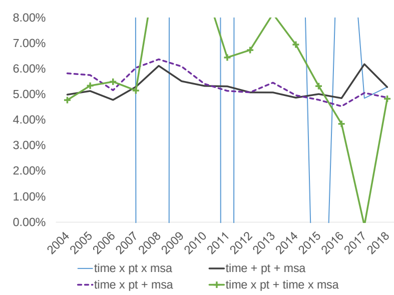
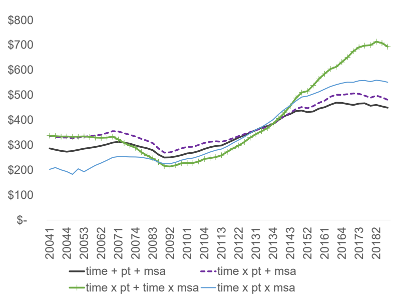
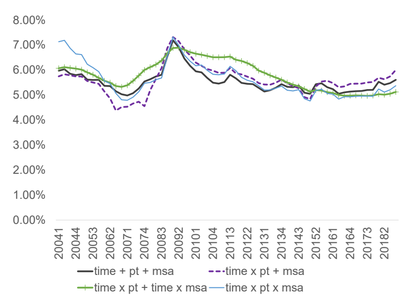


Figure 6. Results for Offices in Los Angeles using Bayesian Approach. This figure presents results for office property types in Los Angeles obtained from the additive structure model using Bayesian methods, see Eqs. (3) – (4) and related 4 sub-specifications with different additive structure. Panel (a), (b), (c), and (d) display time-series for enterprise value per square foot, net operating income/enterprise value, capital expenditure/enterprise value, and operating expenses/enterprise value, respectively. Vertical axis represents the corresponding values (either $ per square foot or percentage over enterprise value), while horizontal axis shows different years for the period 2004 – 2018.

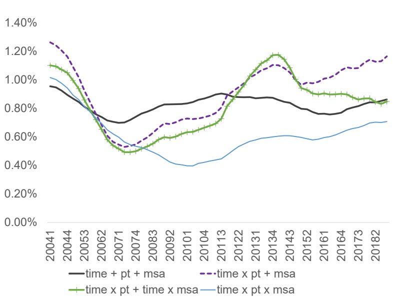
(a) Enterprise Value / SF



(b) NOI / EV



(c) Capex / EV



(d) Opex / EV

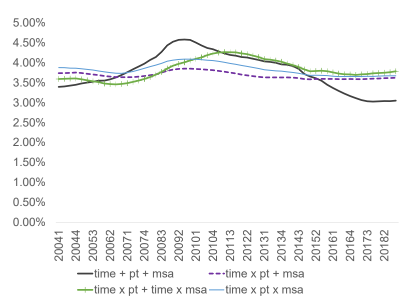
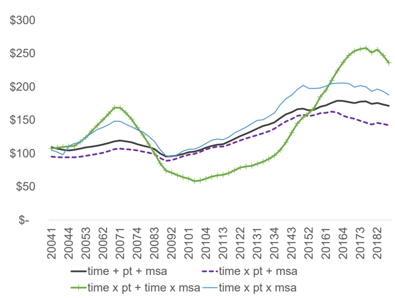
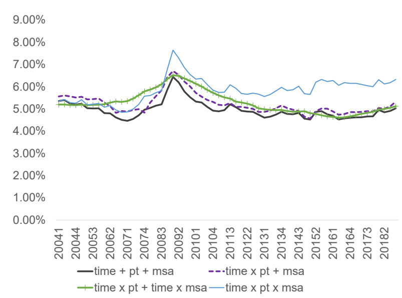


Figure 7. Results for Apartments in Atlanta using Bayesian Approach. This figure presents results for apartment property types in Atlanta obtained from the additive structure model using Bayesian methods, see Eqs. (3) – (4) and related 4 sub-specifications with different additive structure. Panel (a), (b), (c), and (d) display time-series for enterprise value per square foot, net operating income/enterprise value, capital expenditure/enterprise value, and operating expenses/enterprise value, respectively. Vertical axis represents the corresponding values (either $ per square foot or percentage over enterprise value), while horizontal axis shows different years for the period 2004 – 2018.

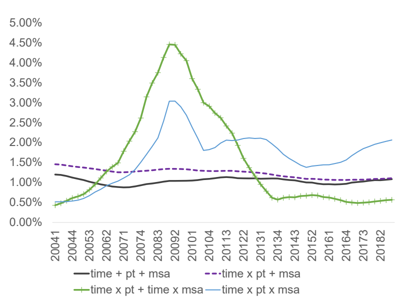
(a) Enterprise Value / SF



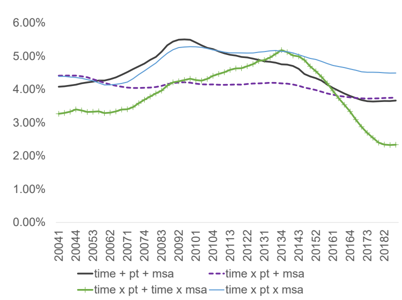
(b) NOI / EV



(c) Capex / EV



(d) Opex / EV



Appendix A. Bayesian Model Details

In this Appendix we will detail the Bayesian model estimation. In order to estimate the model efficiently (i.e. with fast convergence) we log transform the left-hand side variable in all models before running the No-U-Turn-Sampler (NUTS). Logging the dependent variable ensures that the estimated parameters are closer to normally distributed without any large values. To ensure that we do not lose observations due to log transforming zero entries, we add a very small (1E-6) number to our yields.[14] Unfortunately, this does mean that the estimated parameters cannot be easily interpreted. (Note that we do *not* log transform the right-hand side.) However, in this study, we do not focus on said parameters, but on the fitted values of our representative property. The trends are calculated by looking at (the exponentiated) fitted values over time of a "representative property". As explained in Section II, we also add a common trend to the models. Log transforming the left-hand side, and adding a common trend (κ) results in the following Bayesian models;

time + pt + loc: $y_{kt} = \kappa_t D^T_{kt} + \delta_i D^I_{kt} + \gamma_m D^M_{kt} + \beta X_{kt} + \epsilon_{kt};$

time x pt + loc: $y_{kt} = \kappa_t + \mu_{it}(D^T_{kt} \times D^I_{kt}) + \gamma_m D^M_{kt} + \beta X_{kt} + \epsilon_{kt};$

time x pt + time x loc: $y_{kt} = \kappa_t + \mu_{it}(D^T_{kt} \times D^I_{kt}) + \mu_{mt}(D^T_{kt} \times D^M_{kt}) + \beta X_{kt} + \epsilon_{kt};$

time x pt x loc: $y_{kt} = \kappa_t + \mu_{imt}(D^T_{kt} \times D^I_{kt} \times D^M_{kt}) + \beta X_{kt} + \epsilon_{kt};$

[14] The variable where this would happen the most is our OPEX variable. In total we observe 120 NNN leases, where the tenant pays the OPEX.

where;

$$\Delta\kappa_t \sim \mathcal{N}(0, \sigma_\kappa^2);$$
$$\Delta\mu_t \sim \mathcal{N}(0, \sigma_\mu^2).$$

Within each cluster the estimate 1 variance parameter. Another pro of estimating the model with the log on the left-hand side, is that we can keep the same priors irrespective of the left-hand side variable. Indeed, the average values (and thus priors) are vastly different for enterprise value per square foot, compared to for example cap rates in levels, but not so much in logs. We thus draw parameter $\kappa_{t=1}$ from a very uninformative prior (i.e. $\mathcal{N}$(0,10)), whereas the other parameters are largely uninformative ($\mathcal{N}$(0,1)) (Gelman, 2006). The random walks in state Eq. (4), are modelled in first differences, in line with Betancourt and Girolami (2015). The innovations of the random walks have a prior of $\mathcal{N}$(0,1).

We have 2,000 iterations, of which the first half are used as warm-up, over 3 chains. All convergence statistics ($\bar{R}$ and effective sample sizes, not presented here, but available upon request) are satisfactory.