A synthetic cap rate methodology

Andreas D. Christopoulos*, Joshua G. Barratt†and Daniel C. Ilut‡§

September 4, 2020

Abstract

The US commercial real estate (CRE) market has an inherent information content deficit. One aspect of this deficit is found in the disconnect between macroeconomic indicators and CRE risk. We address this issue in a new method that applies Euclidean distancing and least squares techniques to macroeconomic variables to project synthetic capitalization rate indices (SCXs). This method is driven by the new measure of economic proximity which is independent of asset specific characteristics and has no geographic restrictions, ex-ante. This study challenges the primacy of property geographic location which may be too coarse a characterization of CRE valuation. SCXs, driven by economic proximity, provide different, and better, forward guidance as to lifetime CRE default risk and valuation than are obtained with actual capitalization rates. As such, SCXs represent a new set of hard information content that bridges one aspect of the information content deficit in CRE.

Key Words: Credit Risk, Indexing, Information Content, Macroeconomics, Real Estate
JEL Code: C58, E10, G14, G17, G23, R30

*Yeshiva University, Sy Syms School of Business, New York, NY 10033, e-mail: andreas.christopoulos@yu.edu

†Barratt Consulting, Wildenbruchstrasse 84 12045 Berlin, Germany, joshua@bigend.io

‡Cornell University, Plant Breeding and Genetics Section, School of Integrative Plant Science, Ithaca, NY, 14853, dci1@cornell.edu

§The authors would like to thank the participants at the following conferences and seminars over the past year where this paper was presented in various stages: American Real Estate Society, American Real Estate and Urban Economics, Blackrock Global Risk, Eastern Economics Association, Ecole hoteliere de Lausanne, Stockholm Business School, UNC Chapel Hill CREDA, USC Symposium: Architecture, the City and Democratic Capitalism, and Yeshiva University's Sy Syms School of Business Finance Seminar. Their comments greatly improved earlier versions of this study. Christopoulos would also like to thank Yeshiva University's Sy Syms School of Business and the University of Texas at Austin's McCombs School of Business for their support of this project.

1 Introduction

The commercial real estate (CRE) market at ~$6 trillion is quite large and makes broad contributions to the US Economy. It is thus somewhat paradoxical that a market so large in size and scope should, nevertheless, be so informationally opaque. In CRE we observe substantially lower quantities of market pricing information than is observed in many other sectors. These comparative deficits of information stem from fewer asset sales in CRE compared with other markets, such as the residential property market, in part due to the comparatively large size and idiosyncratic character of CRE assets. The US Census and US Energy Information Administration, for example, report that residential properties outnumber CRE properties by at least a factor of 14x in the US. Additionally, the average securitized CRE mortgage size is 30-40x the average size of US residential mortgages.

CRE information opacity translates into asset illiquidity which is observed in its related debt and equity instruments in capital markets where we observe muted frequency of securities trading than observed in other markets. In the absence of frequent transactions, which often reveal prices, CRE asset values remain inherently opaque. There is no secondary market for building assets; they are not traded on public exchanges and they exhibit infrequent transactions which contribute to their highly illiquid profile. When comparing the issuance volumes of residential and CRE securities as described in Killian and Cox (2016), as well as their relative trading frequency, as seen in He and Mizrach (2017), we see substantially lower quantities of observable transaction information for CRE securities investors than are seen in the residential sector. These discrepancies in market driven information content are further confirmed in Marcato and Nanda (2016) who find the residential sector to be significantly more responsive to changes in investor sentiment as captured in prices, compared with the non-residential sector. Findings of opacity are also supported in the literature on the equity side focusing on a disconnect between investor expectations and REIT portfolio valuations as discussed in Pavlov, Steiner and Wachter (2018). This aggregation effect is also noted by Bond and Mitchell (2011) on the debt side of the capital structure for U.K. real estate derivatives. Misspecifcation of underlying risks also affects the liquidity of instruments as found Christopoulos (2017) who suggests a possible aggregation effect within

commercial mortgage backed securities (CMBS) index prices. At the loan level, Titman, Tompaidis and Tsyplakov (2005) find a considerable portion of CRE loan risk premia may be associated with illiquidity of the securing asset and a lack of integration between real estate and more developed corporate credit markets. While the recording of trading in CRE securitizations is improving the information content flow to enhance liquidity as discussed in Hollifield, Neklyudov and Spatt (2017), information deficits *within* such revealed prices in CRE transactions may still confound transparency. Asset size may be assessed through building and cashflow characteristics and those influences on valuation are observed directly through the channel of the numerator of net operating income (NOI) within the CRE valuation identity. But they are only implicitly endogenous to investor expectations.

In this paper we seek to make explicit one influence on investor expectations, the macroeconomy, and position our findings into the denominator of the CRE valuation identity, the capitalization rate (cap rate). To do this we introduce a new methodology, creating CRE cap rate indices from macroeconomic indicators independent of asset characteristics and valuations. We combine ordinary least squares (OLS) and Euclidean distancing to identify the economic proximity, $\text{dist}(k, j)_t$, between house price indices and unemployment rates observable at both metropolitan statistical areas (MSAs) and supra-geographic locations (supralocations) to project their influence onto the value of CRE through the new measure of economic proximity. The implementation results in the projection of *synthetic* capitalization rate indices (SCXs) for 402 MSAs over the period 1991 thru 2015 which do not exist in nature. This technique differs from the motivating indexing literature of Bailey, Muth and Nourse (1963), Case and Shiller (1987) and Fisher (2000) who do not use distancing techniques for indexing.

Our methodology is not, ex ante, geographically restricted. If an MSA and supralocation pair are geographically distant from one another (e.g. the MSA of Tallahassee, FL compared with the supralocation of the State of Alaska) but, nevertheless, exhibit similar variation in their macroeconomic signals, we hypothesize that such similarities should have similar influences on CRE property valuation, regardless of geographic distance. This insight is central to our method. Matches based upon economic proximity, which is temporal, may vary considerably from period to period. Indeed the richness of our approach is established

through our ability to methodically scan the Nation for signals of economic similarity between MSAs and supralocations and to then project MSA level SCXs based upon that similarity. In so doing, our approach inherently challenges the primacy of property geographic location in valuation as discussed in Grissom, Hartzell and Liu (1987) and Wheaton and Nechayev (2005). Our results suggest that location in the traditional sense matters much less than is ordinarily thought while macroeconomic influences appear to matter much more than previously considered in CRE property valuation. Earlier work in the real estate literature does not fully disentangle the geographic influence from the economic influence in CRE valuation. Hartzell, Shulman and Wurtzebach (1987) and Mueller (1993) identify economic diversification through reclassification of geographic regions based on industry representation. However, those reclassifications are inherently geographic and do not partition economic proximity from geographic proximity. In contrast, our work does disentangle these influences. Additionally, recent work by Chegut, Eicholtz and Rodrigues (2015), Hyun and Milcheva (2018) and Silver (2016) suggest geographic proximity is driving the idiosyncrasy in valuation of CRE assets. Our approach acknowledges and utilizes geographic distinctions but does so with a freer economic configuration that supersedes geographic proximity.

By projecting values from disparate macroeconomic sources and reconstituting them into the SCX benchmark measures, we create sector wide benchmark indices of CRE health. Post-estimation analysis of SCXs reveal greater insights into CRE risks than actual cap rates. SCXs appear to be comparatively more sensitive than actual cap rates to lifetime risks of CRE default in regions outside of central business districts (CBDs). This is noteworthy given that the majority of CRE loans in the US are underwritten in locations outside of CBDs. It also makes intuitive sense. SCXs are by their nature forward looking.

Our new methodology addresses a central research question in the literature as to whether or not 'location matters' in the valuation of CRE. Our results indicate that location does matter, but not in the ordinary geographical sense used in practice. The coarse term 'location' used to characterize inherent (but undisclosed) economic drivers of property value is, in our view, is an agglomeration of more precise macroeconomic influences which we make explicit with the introduction of our new methodology and which heretofore have not been articulated. Given our new methodology and results, we claim SCXs provide new and more

objective 'hard information' content, in the sense of Liberti and Peterson (2019), appropriate for risk evaluation and investment decisionmaking for the CRE sector.

The remainder of this article is organized as follows. Section 2 discusses the data used in our study. Section 3 describes the methodology for SCXs. Section 4 discusses the validation of our model. Section 5 validates the use of SCXs in lifetime default estimation of CRE loans. Section 6 summarizes with suggestions future research. The Online Appendix provides robustness checks and supplementary information.

2 Data

This section discusses the data we use in our study to project SCXs. To project these measures we use both public and private information. Both sets of data are broadly utilized by market actors. Wherever possible we preserve the integrity of the data in its raw form consistent with the way in which it is absorbed in common practice.

2.1 Public data

The public data we use to capture macroeconomic influences on CRE are: house price indices (HPI), unemployment rates (UE), the corporate bond credit slope (equal to the difference in yields to maturity for Baa and Aaa rated corporate bonds), the Freddie Mac (FHLMC) conforming 30-year mortgage rate, and the CRE chargeoff rate. HPI are reported by the FHFA. UE are reported by the Bureau of Labor Statistics (BLS). The credit slope, mortgage rate and CRE chargeoff rate are reported by the Federal Reserve Board as captured in the FRED Economic Database of the St. Louis Federal Reserve[1]. The selection of these specific data is purposeful. First, capturing public data influences on market actors decision-making requires that such information be readily observable and in the form in which it is transmitted. Each of the variables we have selected satisfy this criteria. Second, we want prior continuity with the literature[2] that utilize macroeconomic variables to describe the impact debt instruments. Third, for our distancing technique to be implemented, we require

[1]See https://fred.stlouisfed.org/

[2]See Yan, Shi and Wu (2006) and Peng (2008), and Figlewski, Frydman and Liang (2012).

that the macroeconomic variables be readily observable at both the supralocational levels and the MSA levels and used by market actors. HPI and UE satisfy those criteria and thus are selected for the distancing procedures described in Section 3.

We capture all data from 1990-2015 on a quarterly basis. HPI and UE are reported at National, State, and 402 MSA levels of granularity, while the credit slope, mortgage rate and CRE chargeoff rate are reported only at the National level. HPI values are converted into simple nominal returns. UE and CRE chargeoff rates are expressed as annual percentages. The credit slope between Baa and Aaa rated corporate bond yields to maturity is converted to basis points (bps). The mortgage rate is expressed as an annualized percent rate of interest. We also use quarterly observations of the Dow Jones Industrial Average (DJIA) which were captured from Yahoo! Finance. In Section 5 we utilize the VIX volatility index as reported by the Chicago Board Options Exchange.

2.2 Private data

The private data[3] is used in a variety of ways in projecting and validating SCXs. Since 1977, NCREIF has provided quarterly CRE asset valuation data reported by member institutional fiduciaries with property value estimations lagging by 1 quarter (3-months). NCREIF's regular reporting reflects institutional valuations on more than 35000 CRE properties totaling several trillions of US dollars.[4] The data is reported quarterly and reflects actual sale prices and marked-to-market valuations by the largest commercial real estate property holders in the United States including: pension funds, commercial banks, investment banks, and life insurance companies. NCREIF indices have been widely used in academic studies for more than twenty-years, as described in Diehl (1993) and others. However, problems with NCREIF's frequency of reporting and smoothing of such data as noted in Clayton, Geltner and Hamilton (2001) partially motivates our work. NCREIF returns are captured at the State and National levels for each of 8 mutually exclusive and exhaustive Regional partitions of the US: East North Central, Mideast, Mountain, Northeast, Pacific, Southeast, Southwest,

[3]Private data was provided by an anonymous institutional investor with ~$0.5 trillion AUM.
[4]See https://www.ncreif.org/.

and West North[5] and 6 property types: Industrial, Lodging, Multifamily, Office, Retail and Other.

Table 2.1: Summary statistics of CRE loans

	Totals		Averages			Totals by proptype					
Location	# loans	($mm)	occ (%)	age (yrs)	loan ($mm)	MF	LO	IN	OF	RT	OT
non-CBDs	19906	$256,758	89.2	9.2	$13	2463	1993	2952	3393	8419	686
Atlanta	173	$4,283	84.4	8.7	$25	17	30	17	57	45	7
Baltimore	81	$1,457	90	10.7	$18	10	7	13	24	18	9
Boston	54	$4,257	87.5	9.8	$79	0	5	6	35	5	3
Charlotte	157	$2,127	89.3	7.6	$14	16	22	25	42	45	7
Chicago	255	$8,688	88.8	11	$34	27	28	23	73	84	20
Cincinnati	80	$1,002	91.6	11.8	$13	16	2	6	25	29	2
Cleveland	80	$1,460	88.6	9.3	$18	11	7	5	20	36	1
Dallas	251	$4,343	89	11.2	$17	55	12	24	78	80	2
Dayton	42	$483	87.9	9.8	$12	6	3	8	7	18	0
Denver	136	$2,684	89.7	9.8	$20	5	7	18	49	50	7
Detroit	26	$376	84.9	14.4	$14	7	3	3	9	4	0
Fort Worth	89	$929	90.3	12	$10	16	8	18	13	34	0
Houston	489	$7,293	89.5	10.7	$15	111	38	54	135	145	6
Indianapolis	120	$1,796	86.2	9.6	$15	13	15	16	31	39	6
Jacksonville	109	$2,195	91.5	12.4	$20	17	6	27	19	23	17
Kansas City	67	$1,197	86.6	8.7	$18	7	9	7	18	25	1
Los Angeles	446	$8,732	91.9	12.9	$20	54	19	64	136	143	30
Miami	126	$2,622	87.6	10.8	$21	9	12	22	32	43	8
Milwaukee	58	$778	90.7	9.9	$13	5	9	6	23	14	1
Minneapolis	93	$2,325	88.5	10.6	$25	5	10	16	30	27	5
New Orleans	30	$707	81.9	7.2	$24	1	12	4	9	4	0
New York	776	$35,252	93.1	17.4	$45	251	54	38	243	101	89
Norfolk	35	$516	92.5	10.2	$15	2	5	2	15	10	1
Oakland	31	$227	83.7	7.9	$7	1	1	13	5	7	4
Oklahoma City	79	$1,078	87.6	7.8	$14	11	5	22	19	20	2
Philadelphia	121	$2,649	90.4	11.1	$22	18	8	9	33	42	11
Phoenix	191	$2,712	88.4	10.7	$14	18	10	40	41	78	4
Pittsburgh	70	$903	91.5	7.6	$13	10	8	7	18	19	8
Richmond	63	$949	86.5	7.2	$15	9	8	7	14	17	8
Saint Paul	64	$762	92.6	8.9	$12	1	4	9	25	23	2
San Diego	223	$4,672	89	11.5	$21	6	41	28	74	60	14
San Francisco	110	$3,106	93.7	12.8	$28	19	10	13	37	19	12
San Jose	47	$1,006	92	15.3	$21	4	5	9	14	12	3
Seattle	96	$3,647	89.3	10.5	$38	7	22	13	25	12	17
St. Louis	10	$242	87.6	19.5	$24	0	2	4	3	0	1
Tampa	105	$1,308	87.2	10.6	$12	15	14	9	31	34	2
Washington, DC	212	$11,892	90.6	9.7	$56	13	32	12	118	24	13
Totals (Avg *)	25101	$387,416	89.3*	9.7*	$15*	3256	2486	3569	4973	9808	1009

This table provides summary statistics for the 25,101 commercial mortgages in our sample grouped by CBD and non-CBD location classification. The header provides the total number of loans and aggregate loan amount at origination. Next, we provide averages for occupancy rate, property age, and loan size (all at the time of origination). The final columns provide counts for the number of loans in the sample corresponding to property types of multifamily/apartments (MF), lodging/hotel (LO), industrial (IN), office (OF), retail (RT), and mixed-use/other (OT). The rows at the bottom of the table provide sums for the totals and property types and arithmetic means for occupancy rate, property age, and loan size. Total loan balances by property type are provided in the final row. Source: Intex.

We also have loan level information (Table 2.1) from Intex reported for origination dates spanning 2000-2015 for 25101 properties and mortgages underlying 175 CMBS transactions. We use these data in comparative default and valuations assessments. The total balance of the private loan data at origination was $389 billion with property values at the point of origination totaling approximately $700 billion. The average loan size at origination was about $15mm. This sample represents 1/3 of the CMBS universe and 10% of all CRE properties in the US. Of these loans, 1013 totaling about $15 billion, defaulted over the sample period of 2007 and 2015. Defaults in our study are defined as instances where

[5]The partitioning of US states is shown in A.1.

loan payments are terminated and where a liquidation of the securing property occurred as indicated by a State of Real Estate Owned/Foreclosure. The private loan level data includes zip code level information and property cap rates at origination allowing for grouping by CBD and non-CBD location classification. Table 2.2 provides a summary of the defaulted loans.[6]

Table 2.2: Summary characteristics of Defaulted CRE Loans

Proptype	MF	OF	RT	LO	IN	OT	All
# of defaults	65	271	429	111	92	45	1013
ln(avg prop size)	16.1260	16.4524	16.0708	16.3669	16.1300	16.0358	16.2127
ln(avg uwnoi)	13.4080	13.7548	13.3799	13.8237	13.4480	13.3527	13.5356
avg (actual cap rate)	0.0729	0.0726	0.0716	0.0845	0.0733	0.0727	0.0736
avg (SCX)	0.0750	0.0757	0.0772	0.0757	0.0772	0.0777	0.0765
avg LTV (%)	78.46	76.35	75.91	72.96	77.95	76.57	76.09
occupancy (%)	86.45	87.92	88.03	65.14	86.61	85.00	85.13
loan spread (bps)	124.83	128.28	127.30	151.64	130.82	140.06	130.96

This table provides mean statistics for the 1,013 loans that defaulted within our total sample of 25,019 commercial mortgages. The header provides the property types of multifamily/apartments (MF), office (OF), retail (RT), lodging/hotel (LO), industrial (IN), and other (OT). Source: Intex.

3 Synthetic Capitalization Rate Methodology

This section presents the methodology to project SCXs.

3.1 Capitalization rates

Cap rates are applied throughout the CRE industry to determine a benchmark estimate of the value of a property, and ex-post to record the relationship between NOI and property value following a sale transaction and in ongoing monitoring of the value of the property through time. It is well known in the literature, as for example in Titman (2010), that cap rates may be interpreted as an application of the Gordon Growth model for dividend paying stocks, introduced by Gordon and Shaprio (1959). In the Gordon Growth framework, the observable price, P_0, of a dividend paying stock may be expressed as, $P_0 = \frac{D_0}{r-g}$ with D_0 the current dividend, r the rate of return on the stock and g the growth rate of dividends. In the CRE context NOI, O_0, is substituted for the dividend D_0 and the property value, V_0, is substituted for the stock price P_0. Following Brueggeman and Fisher (2016), the investor's rate of return on the property, r, and the expected growth rate of NOI, g, allows us to

[6]For greater detail see Table A.2 in the Online Appendix.

express the cap rate $C_0 = r - g$, such that

$$V_0 = \frac{O_0}{C_0} \ \therefore \ C_0 = \frac{O_0}{V_0} \tag{3.1}$$

The values for C are are imbued with the market's expectations of future real estate growth, but such expectations on future growth may be misspecified as noted in Sivitanides et al., (2001). Recent cap rate studies by Chervachidze, Costello and Wheaton (2009), Peng (2013), and Seagraves and Wiley (2015) provide innovations into cap rate risk composition but do not project cap rate indices which is our focus. Cap rates as a ratio measure in the real estate context are observed through indices (or calculated ex-post transaction sales) and both expected rates of return r and implied NOI growth rates, g, are subject to investor interpretation at any given time t. Our work uses a new quantitative approach which results in the projection of CRE valuation indices as 'all-in' cap rates at the MSA level, and thus does not investigate its components r and g, which is left to future research.

Additionally, for stabilized properties with existing history of NOI and tenancy, the Office of the Comptroller of the Currency (2017, OCC) advise the use of current NOIs instead of forecasted NOIs. The entirety of our sample from both NCREIF and Intex are based on stabilized properties. As such, we follow the recommendation of the OCC, the Commercial Real Estate Finance Council (2010), and Trepp (2018) and use *current* NOIs in Eq. (3.1) to calculate NCREIF indexed cap rates and property specific cap rates from the Intex data which we compare to our SCX cap rate measures.

For NCREIF data, we use values for O and V as reported in aggregate form across all loans in the fiduciary portfolios for each time t. These aggregates are partitioned by supralocation, property-type and time in the sample data. Consistent with NCREIF (and others), we too compute an *all property-type* cap rate index.

We adjust Eq. (3.1) with the following notation. Let $j \in \begin{cases} 1...51 & = \text{Fifty-one (51) total State regions} \\ 52 & = \text{One (1) total US National region} \\ 53...60 & = \text{Eight (8) total NCREIF regions} \end{cases}$ represent the distinct geographic supralocations; let $t \in [1, 95]$ represent the consecutive

quarters from Q4 1991 to Q2 2015; and let $p \in \begin{cases} 1 = & \text{Industrial (IN)} \\ 2 = & \text{Lodging/Hotel (LO)} \\ 3 = & \text{Multifamily (MF)} \\ 4 = & \text{Office (OF)} \\ 5 = & \text{Other (OT)} \\ 6 = & \text{Retail (RT)} \end{cases}$ represent the distinct CRE property types. Then, for each location j, at each time t, across *all* property types p, each NCREIF indexed NOI, O, is given by

$$O_{j,t} = \sum_{p=1}^{6} O_{j,t,p} \tag{3.2}$$

and each observed NCREIF indexed property value, V, is given by

$$V_{j,t} = \sum_{p=1}^{6} V_{j,t,p} \tag{3.3}$$

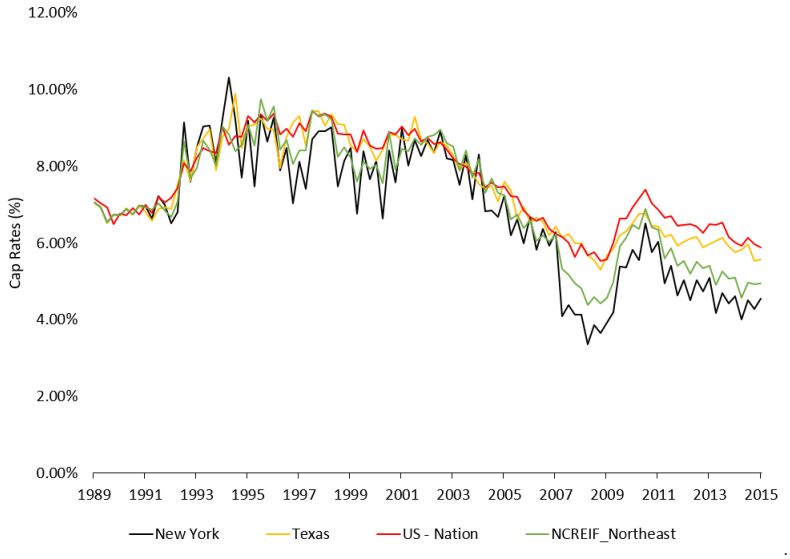
giving the NCREIF indexed cap rate for the time t, j-th supralocation as

$$C_{j,t} = \frac{O_{j,t}}{V_{j,t}} \tag{3.4}$$

We compute $C_{j,t}$ for each $j \in [1, 60]$ geographic supralocation for all times t resulting in 60 distinct NCREIF cap rate time series. Figure 3.1 depicts 4 of 60 NCREIF cap rate time series from Eq. (3.4) for: $j = 35$ (State of New York), $j = 44$ (State of Texas), $j = 52$ (US-Nation), and $j = 55$ (NCREIF Region of Northeast).[7] We use these 60 series of $C_{j,t}$ as dependent variables in OLS Eqs. (3.5) and (3.6) which are required to project SCXs at the MSA level. In the lifetime default study comparison we use Intex data for more than 25101 loans and apply Eq. (3.1) to that data to compute 25101 actual cap rates at origination. For SCXs, we do not use any property specific values for O and V and project the left hand side of Eq. (3.1) independent of its right hand side.

[7] Table A.1 in the Online Appendix provides statistics for all 60 supralocation cap rate series.

Figure 3.1: NCREIF cap rates



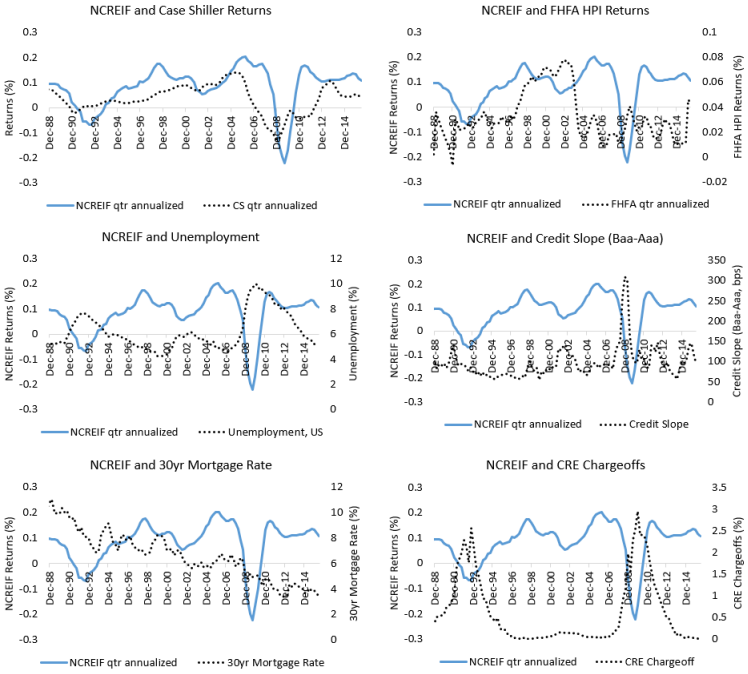
This figures shows 4 of 60 cap rate indices constructed from NCREIF data as described in Eqs (1) thru (4) from aggregated NOI and property values aggregated across property types within geographic locations. Depicted are the time series for US - Nation, US States of New York and Texas, and NCREIF region of the Northeast. The cap rates are computed quarterly. Source: NCREIF.

3.2 Correlation and ordinary least squares

CRE lags the macroeconomy[8]. We observe supporting evidence for this in the relationships in Figure 3.2. The plots show CRE property returns lagging housing price returns and other macroeconomic indicators. We exploit these relationships at the National level to form estimates at more granular geographic levels. In Table 3.1 we report the Pearson correlations for CRE returns (as measured by NCREIF National index returns) with the public macroeconomic data. The relationships of the macroeconomic variables with CRE returns over 25 years of quarterly observations are intuitive. CRE returns are positively correlated with housing returns (0.1694) and negatively correlated with both UE (-0.4671) and chargeoff rates on CRE properties (-0.7644). This suggests that when the economy is healthy CRE returns are higher. Additionally, we observe through National lending influences that CRE returns are negatively correlated with both mortgage rates (-0.1690) and the Baa-Aaa credit slope (-0.3093). These facts are also intuitive. Higher costs of borrowing articulated through higher mortgage rates put pressure on the economics of leveraged CRE

[8]See National Association of Realtors (2006), among others.

Figure 3.2: Macroeconomic factors and NCREIF returns



These figures show macroeconomic variables at the US National plotted alongside the National level NCREIF commercial real estate property return index over the period Q1 1987 through Q2 2016. From left to right, the top two figures depict NCREIF compared to the Case Shiller home price index (left) and then against the FHFA national home price index (right). In the middle row the figures depict NCREIF to US Unemployment (left) and the corporate bond credit slope (Baa - Aaa, right). In the bottom row the figures depict NCREIF compared to 30 year FHLMC conforming mortgage rate (left) and commercial real estate chargeoff rate (right). Sources: NCREIF, S&P, FHFA, BLS, FRED-St. Louis.

Table 3.1: Correlations of Macroeconomic factors with NCREIF returns

	NCREIF
NCREIF	1.0000
HPI	0.1694
UE	-0.4671
Mortgage rate	-0.1690
Chargeoff	-0.7644
Credit slope	-0.3093

This table shows the Pearson correlation coefficients determined from observations of quarterly data between 1987 and 2015. The variables are at the US National level of granularity. The variables include the NCREIF commercial real estate property return index (NCREIF), returns on the home price index (HPI), US Unemployment (UE), the 30 year FHLMC conforming mortgage rate (Mortgage rate), the commercial real estate charge off rate (Chargeoff) and the corporate bond credit slope (Credit slope)/ Sources: NCREIF, FHFA, BLS, FRED-St. Louis.

Table 3.2: Cross-correlations of macroeconomic leading indicators with NCREIF

Lead NCR by	hpi	ue	creditslope	mtgrate	chgoff
0Q	0.06	-0.61	-0.30	-0.15	-0.77
1Q	0.09	-0.53	-0.46	-0.18	-0.71
2Q	0.12	-0.43	-0.57	-0.22	-0.64
3Q	0.16	-0.30	-0.61	-0.25	-0.55
4Q	0.20	-0.16	-0.58	-0.29	-0.44

This table summarizes the cross-correlations between the NCREIF national index and each of the FHFA house price index (hpi), the US unemployment rate (ue), the corporate bond credit slope (creditslope, Baa-Aaa), the FHLMC 30 year conforming mortgage rate (mtgrate), and the commercial real estate charge off rate (chgoff). The leftmost column indicates the correlations between NCREIF and the macroeconomic variables for different quarterly lead times from 0 to 4 quarters. Sources: NCREIF, FHFA, BLS, FRED-St. Louis.

property values. This is especially so when broader market credit deterioration indicated by a steepening credit curve requires increasing risk premia for riskier assets (Baa) than safer assets (Aaa), negatively impacting the economic feasibility of new loans for acquisition or refinancing. In Table 3.2 we show the cross-correlations between NCREIF and each of the macroeconomic independent variables. The OLS regression with NCREIF cap rates as the dependent variable is given by

$$\begin{aligned} C_{US,\,t} = \quad & \alpha + \beta_1 \mathrm{HPI}_{US,\,t-4} + \beta_2 \mathrm{UE}_{US,\,t-4} + \beta_3 \mathrm{CreditSlope}_{t-4} \\ & + \beta_4 \mathrm{MtgRate}_{t-4} + \beta_5 \mathrm{CREchgoff}_{t-4} + \epsilon_t \end{aligned} \tag{3.5}$$

In Eq. (3.5) the dependent variable is observed at time t while all independent variables are observed at time $t-4$. All variables are observed initially at the US National level.

Table 3.3: OLS National Level with Newey-West standard errors

ncr caprate52	coeff	Newey-West std error	t-stat	p>|t|	95% Conf.	Interval
crechgoff4qlead	-0.00992***	0.00099	-10.02000	0.00000	-0.01189	-0.00795
mtgratelead4q	0.00674***	0.00040	16.72000	0.00000	0.00593	0.00754
creditslope4qlead	0.00006***	0.00002	4.15000	0.00000	0.00003	0.00010
hpi52lead4q	-0.11217***	0.02984	-3.76000	0.00000	-0.17149	-0.05286
ue52lead4q	0.00142***	0.00045	3.15000	0.00200	0.00053	0.00232
_cons	0.02584***	0.00512	5.04000	0.00000	0.01566	0.03602

This table summarizes the results for the National cap rate ordinary least squares regression of the form $C_{US,\,t} = \alpha + \beta_1 \mathrm{HPI}_{US,\,t-4} + \beta_2 \mathrm{UE}_{US,\,t-4} + \beta_3 \mathrm{CreditSlope}_{t-4} + \beta_4 \mathrm{MtgRate}_{t-4} + \beta_5 \mathrm{CREchgoff}_{t-4} + \epsilon_t$. The results correct for autocorrelation using Newey-West corrected standard errors. All independent variables lead the dependent variable by 4 quarters

In Table 3.3 we report the regression results at the National level, corrected for serial correlation in the residuals with Newey-West standard errors. We note that the signs of the coefficients are consistent with intuition: Higher CRE cap rates (lower CRE property values) in one year correspond with i.) lower HPI today; ii.) higher UE today; iii.) wider/steeper

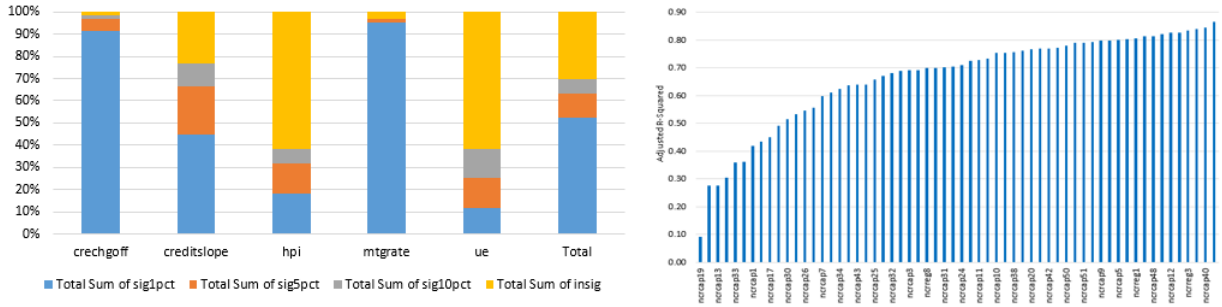
credit-slope today; iv.) higher mortgage rates today; and v.) lower CRE chargeoffs today.The regression and each of the independent variables are highly significant at the 1% significance level with an adjusted R-squared of 0.82.We generalize the OLS in Eq. (3.5) with notation such that

$$\begin{aligned} C_{j,\,t} = \quad & \alpha_j + \beta_{1,j}\mathrm{HPI}_{j,\,t-4} + \beta_{2,j}\mathrm{UE}_{j,\,t-4} + \beta_{3,j}\mathrm{CreditSlope}_{j=52,t-4} \\ & + \beta_{4,j}\mathrm{MtgRate}_{j=52,t-4} + \beta_{5,j}\mathrm{CREchgoff}_{j=52,t-4} + \epsilon_{j,t} \end{aligned} \tag{3.6}$$

with $j \in \begin{cases} 1...51 & = \text{Fifty-one (51) total State regions} \\ 52 & = \text{One (1) total US National region} \\ 53...60 & = \text{Eight (8) total NCREIF regions} \end{cases}$. Since we compute NCREIF cap rates for all $j \in [1,\, 60]$ geographic supralocations. we are able to compute 60 total OLS. We note that in Eq. (3.6) we restrict the index for credit slope, 30 year mortgage rate and CRE chargeoff rates to $j = 52$ which corresponds to the National (US) supralocation. This is because those variables are only reported at the National level, while HPI and UE are available for all j. We compute Eq. (3.6) for all supralocations $j \in [1,\, 60]$. As we are determining a different set of coefficients for each Region, the α, β and ϵ values must also carry the index associated with the j-th supralocation. The indexed estimated constants and coefficients are used in the projection of SCXs. All regressions were significant as evidenced by the F-test. Figure 3.3 summarizes the statistical profile graphically. Figure 3.3a (on the left) shows the composition of significance for each of the independent variables within all OLS regressions. For example, HPI and UE exhibit varying levels of significance in about 40% of the cases while CRE chargeoffs are significant in virtually all cases. Figure 3.3b (on the right) shows the adjusted R-squared values for all OLS regressions.[9]

[9] Table A.3 in the Online Appendix provides summary results for the 60 regressions.

Figure 3.3: Statistical summaries for 60 NCREIF cap rate OLS



This figure summarizes the results for the 60 OLS regressions corresponding of the form $C_{j,t} = \alpha + \beta_1 \text{HPI}_{j,t-4} + \beta_2 \text{UE}_{j,t-4} + \beta_3 \text{CreditSlope}_{j=1,t-4} + \beta_4 \text{MtgRate}_{j=1,t-4} + \beta_5 \text{CREchgoff}_{j=1,t-4} + \epsilon_t$ with j =1:60 supralocaitons (1:51 states, 1 US nation, 8 NCREIF regions. The dependent variable $C_{j,t}$ is the cap rate index for the j-th location. The first two independent variables, are j-th locational values for the house price index returns and the unemployment rate. The last three independent variables are the corporate bond credit slope, the FHLMC 30 year conforming mortgage rate, and the CRE charge off rate which are all captured at the national level ($j = 1$). The plot on the left shows the composition of the significance of each of the independent variables. The plot on the right shows the R-squared values. All regressions were significant as measured by the F-test.

3.3 Robustness checks

We perform many standard robustness tests on the OLS model at National, Regional and State levels corresponding to the OLS models described in Eqs. (3.5) and (3.6) detailed in Section A.2 of the Online Appendix including tests of: i.) Omitted variables; ii.) Multicollinearity amongst the explanatory variables; iii.) Serial correlation; iv.) Endogeneity between the model regressors and the error term; and v.) Cap rate deltas as explanatory variables. The results from these tests do not require us to alter the OLS in Eqs. (3.5) and (3.6).

3.4 Euclidean distancing method

Distance matrices are a well established tool used in scientific research. Fields such as genomics, chemistry, economics, and financial risk management apply distance matrices to evaluate differences between object pairs based on some common characteristic.[10] In biology, for example, Jiang et al., (2016), use similarity matrices derived from nucleotide sequences and structural similarity between genes in different organisms to assign putative biological functions to genes for which no experimental functional validation is available. The particular application of economic distancing we develop in this paper is new to the literature focusing on CRE. Economic parameters of observed HPI and UE create a reasonable backdrop for

[10]See Dokmanic et al., (2015).

characterization of economic health. Since publicly available time series of cap rate indices at the MSA level do not exist to our knowledge, we resolve to project them as follows. For each time t, we define the economic proximity, $\text{dist}(k,\, j)_t$, between each of the 402 MSA levels and the 60 supralocational levels as

$$\text{dist}(k,\, j)_t = \sqrt{\left(\text{HPI}_k - \text{HPI}_j\right)^2 + \left(\text{UE}_k - \text{UE}_j\right)^2} \tag{3.7}$$

with $k = [1,\, 402]$ MSAs and $j = [1,\, 60]$ supralocational levels of granularity above the MSA level. FHFA's HPIs and BLS's UEs are reported at the MSA level which allows us to use Euclidean distancing to determine the SCXs from these observable primitives. We exclude CRE chargeoff rate, credit slope, and mortgage rate variables from Eq. (3.7) since they are reported only at the National level. We normalize HPI and UE at each time t by the range of their values in the sample to ensure neither value dominates the other in the distancing.

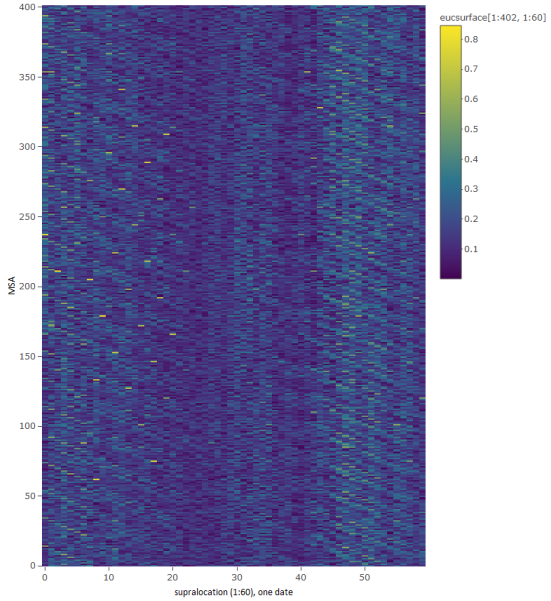
3.5 Projecting synthetic capitalization rate indices

For each of $t \in [1,\, 95]$ consecutive quarters from Q4 1991 to Q2 2015 a matrix with dimensions 402 rows (corresponding to the $k \in [1,\, 402]$ MSAs) and 60 columns (corresponding to the $j \in [1,\, 60]$ geographic supralocations of Nation, Region and State) is formed with the distances calculated in Eq. (3.7). These distances are elements of the matrix $D_{k,\,j,\,t}$ where

$$D_{k,\,j,\,t} = \begin{bmatrix} \text{dist}(1,\, 1)_t & \text{dist}(1,\, 2)_t & \cdots & \text{dist}(1,\, 60)_t \\ \text{dist}(2,\, 1)_t & \text{dist}(2,\, 2)_t & \cdots & \text{dist}(2,\, 60)_t \\ \vdots & \vdots & \vdots & \vdots \\ \text{dist}(401,\, 1)_t & \text{dist}(401,\, 2)_t & \cdots & \text{dist}(401,\, 60)_t \\ \text{dist}(402,\, 1)_t & \text{dist}(402,\, 2)_t & \cdots & \text{dist}(402,\, 60)_t \end{bmatrix}, \forall\ t \in [1,\, 95] \tag{3.8}$$

There are 95 such matrices, one corresponding to each quarterly date, t. Figure 3.4, a heatmap, depicts one such matrix, $D_{k,\,j,\,t=93}$, corresponding to *one* time, $t = 93$. The elements of each matrix consist of 24120 economic proximities $\text{dist}(k,\, j)_{t=93}$ as defined in Eq. (3.7). These economic proximities, differ from quarter to quarter and are used to calculate the SCXs. We are 'searching' for the supralocation that is most economically similar to the MSA by using the HPI and UE dimensions in each period t. Thus, for a given time period t, the supralocation that is most economically similar to the MSA will be the one corresponding

Figure 3.4: Economic proximity Euclidean distances for MSAs x supralocations



This figure summarizes the distribution of Euclidean distanced economic proximities calculated on one date (quarter 93) between 402 MSAs and 60 supralocations. The x-axis corresponds to the 60 supralocations and the y-axis to the 402 MSAs.

to the smallest value for the economic proximity, $\text{dist}(k,\, j)_t$ as defined in Eq. (3.7). For each time t and on each row k, the minimum economic proximity is calculated as

$$d^*_{k,t} = \min_j \left[D_{k,\,j,\,t}\right] \tag{3.9}$$

In other words, $d^*_{k,t}$ is the distance between k-th MSA and the j-th nearest neighboring supralocation among the 60 supralocations at time t. These minimizing values, $d^*_{k,t}$, will reflect the closest (most similar) economic proximity values between the k-th MSA and the minimizing j-th supralocation. Necessarily, we must also capture, for any time t, the corresponding j-th supralocation for the k-th MSA to be a function of the minimization search expressed in Eq. (3.9), such that

$$j^*_{k,t} = \underset{j}{\text{argmin}} \left[D_{k,\,j,\,t}\right] \tag{3.10}$$

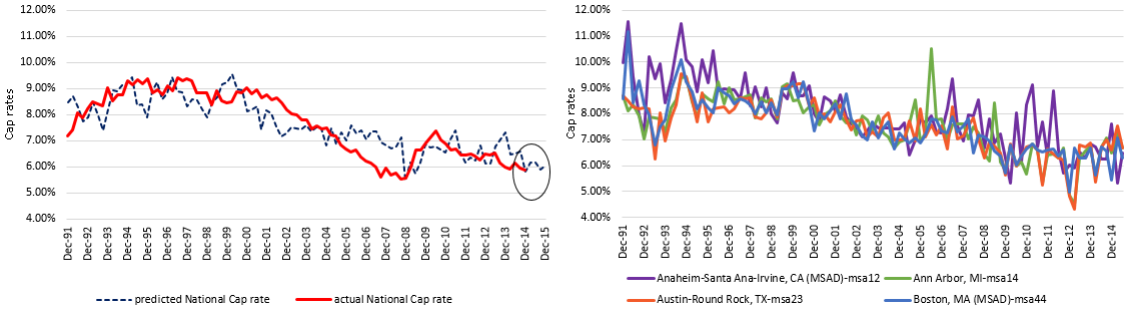
where $j^*_{k,t}$ is the index of the supralocation nearest to MSA k at time t. The best economic proximity value at time t will have the smallest value $d^*_{k,t}$ as determined in Eq. (3.9) and its corresponding supralocation, $j^*_{k,t}$, as determined in Eq. (3.10).

Since we do not observe public MSA level cap rate indices in nature, we project them with the following selection method. The minimum economic proximity, $d^*_{k,t}$, is calculated, and the corresponding $j^*_{k,t}$ supralocation for the k-th MSA is recorded. We then select the constants (α_j's) and coefficients (β_j's) from the regressions estimated in Eq. (3.6), where such selections correspond to the recorded $j^*_{k,t}$ supralocation for the k-th MSA. Using data for each time t for each of the independent variables we interact those reported independent variable data with the j-th selected OLS constants (α_j's) and coefficients (β_j's) to compute $k \in [1,\ 402]$ MSA level SCXs as

$$\begin{aligned}\tilde{\mathrm{C}}_{k,t} = \quad & \alpha_j + \beta_{1,j}\mathrm{HPI}_{k,\,t-4} + \beta_{2,j}\mathrm{UE}_{k,\,t-4} + \beta_{3,j}\mathrm{CreditSlope}_{j=52,\,t-4} \\ & + \beta_{4,j}\mathrm{MtgRate}_{j=52,\,t-4} + \beta_{5,j}\mathrm{CREchgoff}_{j=52,\,t-4} + \epsilon_t\end{aligned} \tag{3.11}$$

Eq. (3.11) restricts the index for the values of the credit slope, 30 year mortgage rate and CRE chargeoff rates to $j = 52$ which corresponds to the National (US) supralocation. This is because those variables are only reported at the National level consistent with Eqs. (3.5) and (3.6). We emphasize that the dependent variable values in Eq. (3.11) are not observed in nature. Rather, they are projected by us through the selection of constants (α_j's) and coefficients (β_j's) for independent variables and resulting calculations described above. We compute 38190 values for the left-hand side $\left(\tilde{\mathrm{C}}_{k,t}\right)$ of Eq. (3.11) with selected values on the right-hand side of Eq. (3.11) corresponding to all MSAs $k \in [1,\ 402]$ for all times $t \in [1,\ 95]$ shown in Figure 3.5.

Figure 3.5: Actual and predicted synthetic cap rates



This figure provides plots reflecting synthetic cap rate calculations. The plot on the left shows the estimated cap rate index ($\hat{C}_{j=1}$) determined from the NCREIF index in the first OLS compared with the actual cap rate index ($C_{j=1}$) at the national level. The second plot, on the right, shows synthetic cap rate indices ($\tilde{\mathrm{C}}$) for 4 of 402 total MSAs.

Figure 3.5a (to the left) shows the predicted cap rate index ($\hat{C}_{j=1}$) determined from the NCREIF index in Eq. (3.5) versus the actual NCREIF cap rate index ($C_{j=1}$) at the National level. Figure 3.5b (to the right) shows SCXs ($\tilde{\mathrm{C}}_{k,t}$) for 4 of the 402 total MSAs as specified in Eq. (3.11) governed by the distance equation Eq. (3.7) and the selection procedure as previously described. The procedure yields a total of 402 separate quarterly time series of SCXs from Q4 1991 to Q2 2015.[11] Importantly, the values $\tilde{\mathrm{C}}_{k,t}$ have neither ex ante geographic restrictions nor information directly identifying industry concentration, population density, construction, or other local factors. In our approach, economic conditions are captured solely in the parameters HPI and UE which govern the MSA level to supralocational level economic proximity used for coefficient selection in each period for each MSA. As such, the j-th minimizing supralocations determined in each time t for each of the $k \in [1, 402]$ MSAs are often different for different times t. This is purposeful. Table 3.4 shows the economic proximities for MSA $k = 1$ for two periods: period $1 = Q41991$ and period $80 = Q42011$. In period $t = 1$, the minimal value of $d_{k,t} = d^*_{1,1} = 0.3876$ which corresponds to supralocation $j_{k,t} = j^*_{1,1} = 21$, the US State of Hawaii. In period $t = 80$, the minimal value of $d_{k,t} = d^*_{1,80} = 0.1911$ which corresponds to supralocation $j_{k,t} = j^*_{1,80} = 17$, the US State of Delaware.

Table 3.4: Distance examples for period 1 and period 80 for MSA=k=1

Period,t=1,k=MSA=1						Period,t=80,k=MSA=1					
UE_k	HPI_k	j∈[1,60]	UE_j	HPI_j	d_{kt}	UE_k	HPI_k	j∈[1,60]	UE_j	HPI_j	d_{kt}
5.4	-0.01601	1	10.6000	0.0179	2.3284	6.6	-0.0124	1	6.8000	-0.0247	0.9924
		2	8.1000	0.0027	1.2700			2	10.4000	-0.0172	0.6947
		3	8.6000	0.0109	1.7834			3	8.7000	0.0104	1.8692
		4	7.9000	-0.0065	0.7541			4	10.4000	0.0071	1.6747
		5	9.2000	-0.0104	0.7875			5	12.2000	-0.0074	0.9394
		6	6.5000	0.0328	3.0536			6	8.1000	-0.0239	0.9549
		7	8.1000	-0.0218	0.6178			7	9.1000	-0.0175	0.5580
		8	8.9000	0.0045	1.4351			8	10.7000	0.0369	4.0232
		9	6.3000	-0.0076	0.5523			9	7.6000	0.0358	3.8876
		10	8.7000	-0.0028	1.0283			10	10.7000	-0.0035	0.9474
		11	5.5000	-0.0042	0.7360			11	10.7000	-0.0039	0.9269
		12	3.4000	-0.0528	2.3301			12	6.9000	-0.0380	2.0642
		13	5.8000	0.0278	2.7393			13	5.2000	-0.0123	0.2123
		14	7.8000	0.0293	2.8629			14	7.8000	-0.0250	1.0323
		15	9.3000	0.0155	2.0941			15	10.4000	-0.0370	2.0619
		16	6.9000	0.0227	2.4362			16	9.4000	-0.0030	0.8673
		17	4.8000	0.0038	1.2448			17	6.8000	-0.0101	0.1911
		18	7.9000	0.0012	1.1681						
		19	8.2000	0.0199	2.2991						
		20	9.4000	-0.0201	0.7834						
		21	7.4000	-0.0142	0.3876						

This table shows distance calculations $d_{k,t}$ for periods $t = 1$ and $t = 80$ for $k = 1$.

[11] Long form tables providing statistics for all 402 SCXs are available upon request.

3.6 Discussion

When predictive models perform well, there is some suggestion that the variables under consideration are important as in our model choices. Our approach is more subtle. We believe the effects of HPI and UE are different under different regimes, and thus we apply different effects to those and all other variables relative to each other based on the most similar observable regime. HPI and UE will be prominent in some regime models, less so in others, and we will choose the model that is observably closest to the property in question based on all available information in the model. All models consider HPI and UE in addition to all the global variables. The selection of a model supralocation is as dependent on HPI as it is on UE, since both values are normalized prior to calculation of the distance measurement. Thus both have the same variance, and over all time periods and locations will have the same aggregate impact on choice of the nearest model. Only the selection of the model parameterization is determined by economic distance. The minimizing distance measure governs the selection of the j-th supralocation level coefficients used in Eq. (3.11) to calculate the SCX at each time t. In this way the economic distancing selects coefficients based upon economic similarity without regard to geographic proximity, ex ante. SCXs thus differ from existing index methods which, through use of repeat sales indices or appraisal based indices, as described in Bailey and Muth (1963), Case and Shiller (1987), and Pagliari et al., (2002), all take information on underlying valuations as a given. SCXs in contrast do not. Instead we create projections of implied cap rates which may be thought of as commercial real health indices distilled from disparate information that is bound together through our methodology.

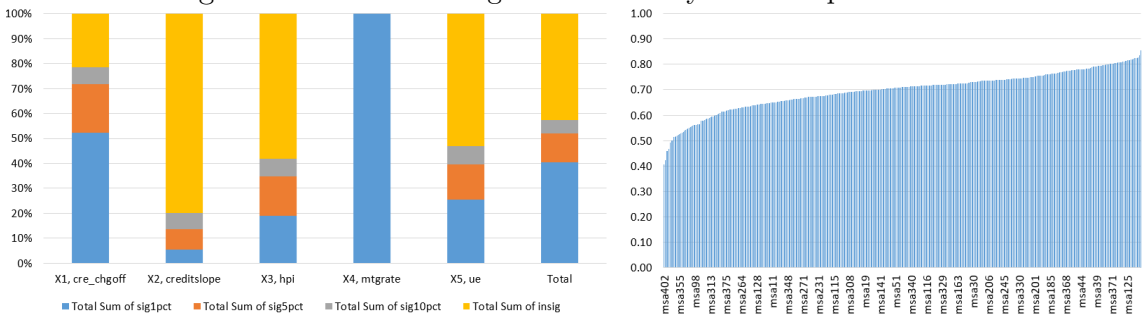
4 Model validation

Since we captured viable SCXs at the MSA level with the procedure above, we first test the statistical significance of the now existent SCXs. We revisit the initial model parameters and re-estimate the constants and coefficients for the model computing 402 new OLS of the form

$$\begin{aligned}\tilde{\mathrm{C}}_{k,t} = \quad & \alpha + \beta_1 \mathrm{HPI}_{k,\,t-4} + \beta_2 \mathrm{UE}_{k,\,t-4} + \beta_3 \mathrm{CreditSlope}_{j=52,\,t-4} \\ & + \beta_4 \mathrm{MtgRate}_{j=52,\,t-4} + \beta_5 \mathrm{CREchgoff}_{j=52,\,t-4} + \epsilon_t \end{aligned} \tag{4.1}$$

Figure 4.1 summarizes the statistical results.

Figure 4.1: Statistical significance 402 synthetic cap rate OLS



This figure summarizes the results for the 402 OLS regressions corresponding of the form $\tilde{\mathrm{C}}_k = \alpha + \beta_1 \mathrm{HPI}_{k,\,t-4} + \beta_2 \mathrm{UE}_{k,\,t-4} + \beta_3 \mathrm{CreditSlope}_{j=1,\,t-4} + \beta_4 \mathrm{MtgRate}_{j=1,\,t-4} + \beta_5 \mathrm{CREchgoff}_{j=1,\,t-4} + \epsilon_t$ with $j \in [1, 60]$ supralocations and $k \in [1, 402]$ MSAs. The plot on the left shows the composition of the significance of each of the independent variables normalized across all 402 OLS regressions. The plot on the right shows the R-squared values for each of the 402 OLS regressions. All regressions were significant as measured by the F-test.

As before, all regressions were significant as evidenced by the *F*-test. The figure on the left shows the composition of significance for each of the independent variables across all regressions while the figure on the right shows their adjusted R-squareds. HPI and UE at the MSA level exhibit some level of significance in 40% to 50% of the cases with the mortgage rate statistically significant for all MSAs. The adjusted R-squared values are somewhat improved compared with the original supralocation results produced from Eqs. (3.5) and (3.6). The SCX approach should produce cap rate values that are intrinsically stable. Because we use OLS, there may be some instability in the coefficients exclusively between the National, Regional, and Statewide variables and MSAs due to multicollinearity. However, this is a straight result of interactions between those National, Regional, and Statewide variable averages which are by their nature multicollinear. The initial regressions from which the coefficients are 'farmed' for SCX projection do not exhibit multicollinearity in testing. The estimations therefore should remain the same regardless of shifts amongst National, Regional, and Statewide coefficient weights.

While explanatory variables may exhibit multicollinearity it will not reduce the predictive power of the model. We are 'switching' from one set of coefficients to another based upon the

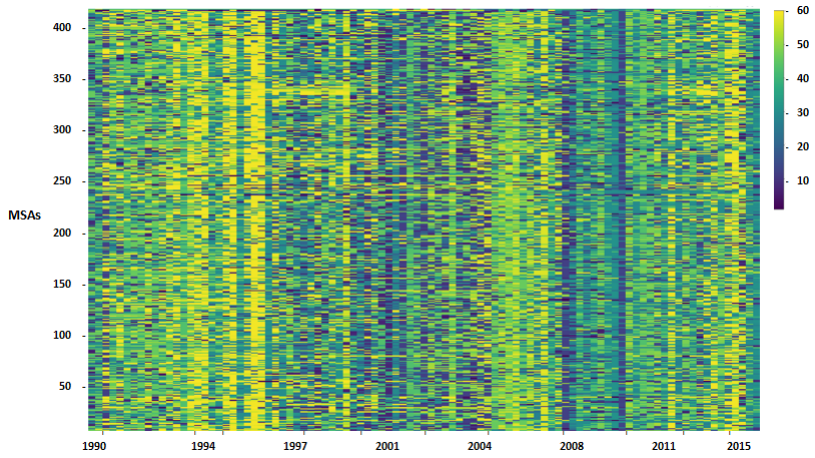
distance $d^*_{k,t}$ in each period. As such, concerns related to instability of coefficient estimation for small changes in data should be muted. To investigate, we note that the economic distance, $d^*_{k,t}$, in our model will not necessarily produce the same $\tilde{\mathrm{C}}_{k,t}$ for a given period even if the aggregated values $\mathrm{dist}(k,\, j)_t$ for two or more different k locations are equal. Only in cases where HPI and UE from two different locations are identical to one another will the model produce identical $\tilde{\mathrm{C}}_{k,t}$ values for the period. However, in this study where we calculate 38190 SCX values between 1991 and 2015, our method produces only 110 instances of non-unique (identical) values (or 0.288% of all calculated SCXs). This bolsters our claims as to the precision and stability of SCXs determined using our methodology, especially given the geographic and historical breadth of the study. If we simply found MSA x supralocation pairs that were consistently reliable proxies from period to period, one would question the contribution of our approach.

To get a sense for different levels of stability corresponding to different MSAs across time we note that the variables of interest are the MSAs and the supralocations across time. The matrix $J_{k,t}$ with 402 rows (corresponding to the $k \in [1,\, 402]$ MSAs) and 95 columns (corresponding to the $t \in [1,\, 95]$ successive quarters) has as its elements the collection of 38190 supralocations, $j^*_{k,t}$, determined from Eq. (3.10) such that

$$J_{k,\,t} = \begin{bmatrix} j^*_{1,1} & j^*_{1,2} & \cdots & j^*_{1,95} \\ j^*_{2,1} & j^*_{2,2} & \cdots & j^*_{2,95} \\ \vdots & \vdots & \vdots & \vdots \\ j^*_{401,1} & j^*_{401,2} & \cdots & j^*_{401,95} \\ j^*_{402,1} & j^*_{402,2} & \cdots & j^*_{402,95} \end{bmatrix} \tag{4.2}$$

Figure 4.2, a heatmap, corresponding to Eq. (4.2) reveals the distribution of all 38190 supralocation selections (with values 1 thru 60, as indicated in the legend) for all 402 MSAs across all 95 quarters in the sample period. We see the selection procedure of mapping coefficients from the supralocation level to the MSA level varies geographically over time. There are instances of lighter and darker clusterings corresponding to supralocation selection for given dates. There does not appear to be a regular pattern. This suggests that the procedure is 'using' the entire Nation divided amongst the supralocations to find the best

Figure 4.2: Supralocations corresponding to minimum economic proximity



This figure summarizes the distribution of supralocations selected from the distancing procedure (1:60 as indicated in the legend) and mapped to each of the 402 MSAs in each quarter in the sample period (95 quarters). The x-axis corresponds to the quarters and the y-axis to the MSAs.

economic proximity for a given period and MSA.

While it is tempting to look at the variation of nearest neighboring regions over the 95 quarters in terms of the variance in their indices in the [1, 60] domain (i.e., $E\left[{(x-\mu)^2}/{\sigma^2}\right]$) it is important to recognize that any such variance estimate is strongly dependent on the arbitrary choice of order of the index. Thus there is no non-arbitrary measure of the variance of a non-ordinal categorical variable. However, there is a clear difference between a categorical variable which is typically observed to take on only a small subset of its possible values, and one that is observed to take on all of its possible values at similar rates. The corresponding measure of variance of such a non-ordinal variable is its Shannon Entropy[12] which, like the scalar variance, takes into account the number and frequency of different values it takes on, without the scalar's requirement of a mean, which is meaningless to categoricals. We compute Shannon's Entropy score, $H_j(x_k)$, where

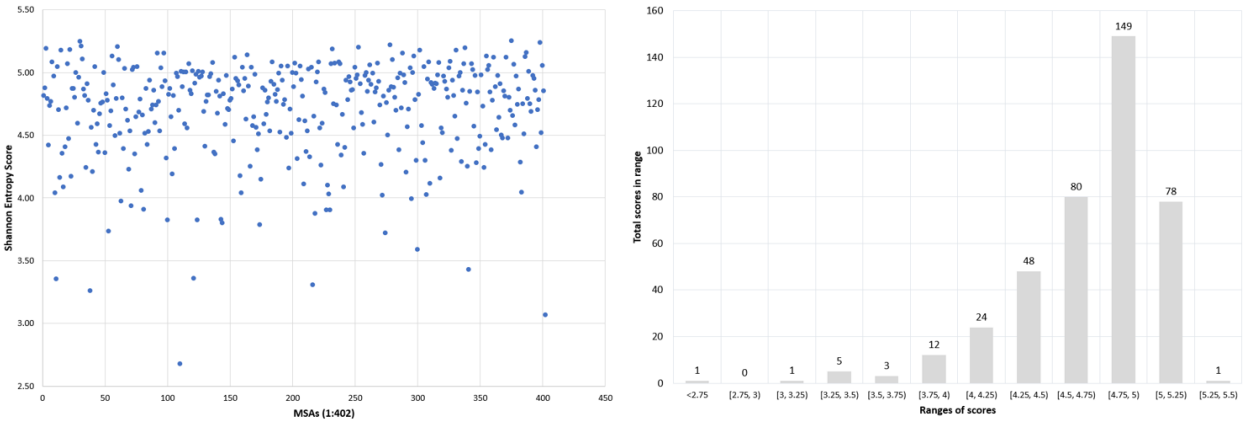
$$H_j(x_k) = -\sum_{x_k} p(x_k) log_2 p(x_k) \tag{4.3}$$

which is the negative logarithm of base 2 of the probability mass function for categorical variables. In this context, the categorical variables are the MSAs and supralocations and $p(x_k)$ represents the percentage of quarters of the j-th MSA in the k-th supralocation

[12] As introduced in Shannon (1948).

observed over 95 quarters. Shannon's Entropy scores which span from 0 (certain) to infinity provide an additive measure of uncertainty for categorical variables. Computing $H_j(x_k)$, allows us to create a single value reflecting the stability of the distancing procedure for each MSA determined from the set of supralocations. The score allows us to then rank order based on supralocation selection variability as depicted in Figure 4.2. Figure 4.3 summarizes the Shannon Entropy score analysis.

Figure 4.3: Shannon's Entropy scores for 402 MSAs



This figure provides the Shannon Entropy score for 402 MSAs. The score is defined as the negative logarithm of base 2 of the probability mass function for categorical variables. In this context, the categorical variables are the MSAs and supralocations and the probability represents the percentage of quarters of the j-th MSA in the k-th supralocation observed over 95 quarters. The plot on the left shows the scores x MSA locations. The plot on the right aggregates the frequency of scores in bins of 0.25 from 2.75 to 5.

Figure 4.3a shows the scatterplot of the Shannon Entropy scores for MSAs which range from 2.6763 to 5.2524. Figure 4.3b groups the scores in partitions with increments of 0.25, with most scores falling between 4.75 and 5.00 and mean and median of 4.7 and 4.8, respectively.[13] The Shannon Entropy scores we calculate support the perspective that our distancing procedure, while utilizing many supralocations across the US, appears to be consistent across most MSAs. The scoring results do not suggest serious uncertainty based on the additivity principle, nor does the scoring indicate levels of certainty as might be seen in perfect substitutes. The distancing measure appears to be sensitive to changes in the HPI and UE drivers across the Nation and reasonably stable with respect to selection uncertainty across MSAs.[14]

[13]Shannon Entropy scores for all 402 MSA locations are available upon request.

[14]In Section A.3 of Online Appendix we summarize the results of the LOOCV approach as described in Kohavi (1995).

5 Lifetime CRE default estimation

While the VAR and IRF results are good, they do not identify the explicit benefit of SCXs to CRE risk assessment. To verify SCX contribution to CRE risk assessment we turn to 25101 CRE properties underlying 125 CMBS transactions in our sample.

5.1 Probabilistic analysis

We conducted *t*-tests summarized in Table 5.1 for paired samples for non-defaulted and defaulted subsets. These results allow us to reject the null hypothesis that the means of the differences between the sample are zero at the 1% level of significance.

Table 5.1: Paired t-test and summary statistics

	Total Loans		Non-Defaulted Loans		Defaulted Loans	
	actualcap	SCX	actualcap	SCX	actualcap	SCX
Mean	0.07007	0.07378	0.06994	0.07367	0.07307	0.07654
Variance	0.00045	0.00006	0.00044	0.00006	0.00088	0.00004
Observations	25,083	25,083	24,071	24,071	1,012	1,012
Pearson Correlation	-0.05965		-0.06334		-0.04198	
Hypothesized Mean Difference	0.00000		0.00000		0.00000	
df	25,082		24,070		1,011	
t Stat	-25.35452		-25.30470		-3.60033	
P(T<=t) one-tail	0.00000		0.00000		0.00017	
t Critical one-tail	1.64491		1.64492		1.64636	
P(T<=t) two-tail	0.00000		0.00000		0.00033	
t Critical two-tail	1.96006		1.96006		1.96231	

This table provides the summary results for the paired t-test applied to the actual and synthetic cap rates corresponding to the origination date on 25,101 CRE transactions. The reported statistics are for the non-defaulted and defaulted subsets. Cap rates greater than 50% were excluded.
.

We quantify the relative propensity to lend at or below cap rate upper bounds in probabilistic terms. The probability of lending at, or below, a given cap rate C is given by

$$F(\mathrm{C}) - F(0) = \Pr(0 < X \leq \mathrm{C}) = \int_0^{\mathrm{C}} f(x)\, dx \tag{5.1}$$

Table 5.2 provides the cumulative normal probability distributions corresponding to the mean and variances previously provided in Table 5.1.

Table 5.3: Valuation statistics

	properties		
	all	non defaulted	defaulted
number of properties	25083	24071	1012
synthetic value	$628,295,272,395	$610,603,768,427	$17,691,503,967
actual UW value	$706,124,906,978	$685,863,114,677	$20,261,792,301
difference	$(77,829,634,583)	$(75,259,346,250)	$(2,570,288,334)
ratio (wtd mean) synthetic/actual	0.88978	0.89027	0.87315
nominal mean synth/actual	0.96266	0.96269	0.96198
nominal min synth/actual	0.20617	0.20617	0.20617
nominal max synth/actual	5.69317	5.69317	4.23399
nominal median synth/actual	0.92843	0.93052	0.89363
nominal kurtosis synth/actual	24.96369	25.99173	12.13431
nominal stdev synth/actual	0.31301	0.30835	0.40873

This table provides valuation statistics for the private loan data using actual cap rates and corresponding SCXs at the point of origination. Source: Intex.

Table 5.2: Cumulative lending probabilities at or below cap rate upper boundaries

	Total Loans		Non Defaulted Loans		Defaulted Loans	
Cap Rate	actualcap	SCX	actualcap	SCX	actualcap	SCX
7.00%	0.4987	0.3167	0.5011	0.3229	0.4588	0.1439
6.00%	0.3183	0.0411	0.3170	0.0433	0.3300	0.0036
5.00%	0.1732	0.0014	0.1698	0.0015	0.2188	0.0000
4.00%	0.0792	0.0000	0.0758	0.0000	0.1329	0.0000
3.00%	0.0301	0.0000	0.0279	0.0000	0.0736	0.0000

This table provides the results for the paired t-test applied to the actual and synthetic cap rates corresponding to the origination date on 25101 CRE transactions. The statistics are for non-defaulted and defaulted subsets. We exclude actual cap rates greater than 50%.

Looking at the top row for all loans, the probability of lending on a property at or below a 7% cap rate is 49.87% using the lending practices associated with the actual cap rate history. In contrast, using SCXs for all loans, the probability of lending at or below 7% (on the same properties) is 31.67%. The economic interpretation is that lending according to actual practices is about 57% more likely to result in a loan at a cap rate at or below 7%, evidencing a more conservative CRE valuation with SCXs, $\tilde{\mathrm{C}}_{k,t}$, than with actual cap rates.[15] This relationship increases non-linearly. The effect is most pronounced in the default subsample. Loans that ultimately defaulted with values at origination corresponding to actual cap rates were about 3x as likely to occur (0.4588/0.1439) at or below the 7% upper bound than for evaluations informed by SCXs.

5.2 Valuation with SCXs

Table 5.3 provides summary statistics conducting property valuation with SCXs and comparing to actual values reported. The actual underwritten valuation of all properties (25083) in the

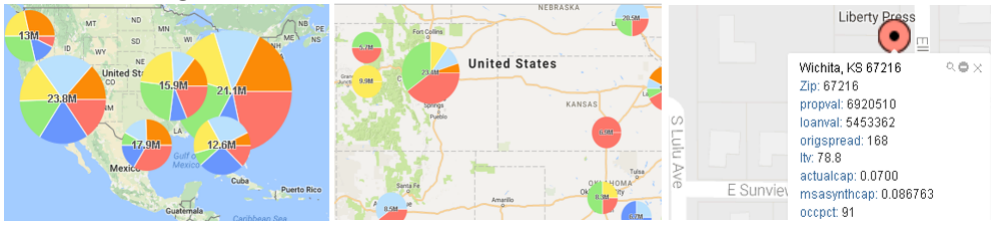
[15]With NOI, $O_{\iota,t}$, held constant at origination date t, for $\kappa_{\iota,t} < \tilde{\mathrm{C}}_{k,t}$, $V_{\iota,t,\kappa} = \frac{O_{\iota,t}}{\kappa_{\iota,t}} > V_{\iota,t,\tilde{\mathrm{C}}} = \frac{O_{\iota,t}}{\tilde{\mathrm{C}}_{k,t}}$.

sample was $706.1 billion of which $685.9 billion (24071) did not not default, while $20.2 billion (1012) did default on loan obligations over the sample period. For this sample, we then computed the implied valuation of the the property (V) using the NOI (O) at origination and the corresponding SCX (C) corresponding to the quarter in which the loan was originated and within the MSA in which the loan's collateralizing property zip code is found. We then use that SCX (C) in the CRE valuation identity, Eq. (3.1). We find aggregate property valuations to be lower by about 11%, with about 13% lower property valuation for those that defaulted on loan obligations. Lower valuations with SCXs are not categorically the case. For example, in many instances the valuations were considerably higher than indicated at the point of origination in the max ratios. This valuation could be refined further with simulated sensitivity analysis of upper and lower bounds over a confidence interval for SCXs, but that is left to future research, Overall, the valuation statistics are consistent with the cap rate distributional discussion above[16]. Together, these analyses support questioning as to whether lending practices and corresponding CRE valuations associated with actual cap rates were overly optimistic relative to underlying credit risks.

5.3 Default example

Figure 5.1, for example, shows the geographic distribution of 1013 defaulted loans in our sample at the US (Figure 5.1a); approximate geographic center of the continental US (Figure 5.1b); and neighborhood in Wichita, KS (Figure 5.1c).

Figure 5.1: Locational mapping of CRE loan defaults



This figure depicts the increasing granularity geographically contained in our data set which permits us to identify at the zip-code, CBD, MSA, State, Regional and National levels. Source: Intex.

This granular parsing allows us to observe that a defaulted $5.4mm loan on a $6.9mm

[16]We excluded 18 loans that exhibited very large cap rates at origination from this section.

property carried with it a SCX in the corresponding MSA at origination of 0.087, while the corresponding actual cap rate at origination of 0.070 was 187 bps tighter. Thus, for this one loan, the underwritten CRE valuation was substantially higher than would have been the case with a CRE property valuation conducted with the MSA level SCX.

5.4 Logistic regressions of default

To quantify whether lending practices which used actual cap rates (vs SCXs) accurately reflected the impounding of macroeconomic influences contained in SCXs we use the standard method of multivariate logistic regression of the form

$$Pr(\text{Def}) = F(x) = \frac{1}{1 + e^{-\beta_0 + \sum_{i=1}^{n} \beta_i x_i}} \tag{5.2}$$

where $F(x)$ represents the probability of default. The dependent variable is the default indicator variable equal to 1 if the loan defaulted at any point over the sample period and 0, otherwise. The parameter estimates, β_i, are determined numerically with maximum likelihood estimation. Our actual cap rates in this portion of the study are not NCREIF cap rate estimates. Rather we use actual cap rates from Intex data using Eq. (3.1) at the point of loan origination, as previously described. We compare them to the corresponding SCX for the loan in the quarter in which the loan was originated and within the MSA in which the loan's collateralizing property zip code is found. Because we have actual cap rates at origination, and corresponding matched SCXs, we may investigate into the comparative strengths of actual cap rates compared with corresponding SCXs in the estimation of lifetime default likelihoods. As such, we implicitly test market expectations of future real estate values with a 'fair value' measure independent of market prices.

In testing our SCXs against 'going-in' actual cap rates, $C_{t,i\in[1,\,25101]}$, we use the SCX, $\tilde{\text{C}}_{k,t}$, defined in Eq. (3.11), but select the SCX computed 1-year earlier, or 4 quarterly periods prior to the origination date of the loan to ensure we are not comparing an actual cap rate at origination to a 'forecasted' cap rate. A robustness check was conducted comparing natural '1-year forecast SCXs' $\tilde{\text{C}}$ to going-in actual cap rates $C_{t,i\in[1,\,25101]}$ where time t origination time corresponds to time t calculation time of SCXs. Those results are reported in Section

A.7 of the Online Appendix and consistent with main findings in the text.

5.4.1 *First Treatment*

In the first treatment we consider only one explanatory variable, x_1= cap rate (actual or synthetic), to evaluate differences between defaults that took place in CBDs compared with those in non-CBDs. The results are summarized in Table 5.4.

Table 5.4: Logistic regression - Treatment 1

	Panel A: All		Panel B: CBD only		Panel C: Non CBD only	
	(i)	(ii)	(i)	(ii)	(i)	(ii)
cons	-3.228507***	-6.571107***	-3.9056897***	-7.15312***	-3.128669***	-6.455659***
	(0.0441904)	(0.3387108)	(.1558658)	(0.874274)	(0.0491975)	(0.3685678)
caprate (actual)	.8331781**		4.894396***		0.5957658	
	(.4177213)		(1.807289)		(0.4770939)	
caprate (SCX)		46.32364***		48.85778***		45.88012***
		(4.526286)		(11.65481)		(4.927489)
Adj-Pseudo R-sq	0.0002	0.0041	0.0015	0.0031	0.0001	0.0043
Prob chi2	0.0461	0.0000	0.0068	0.0000	0.2118	0.0000
Obs	25101	25101	5195	5195	19906	19906

This table shows the results of logistic regressions with 0,1 indicator of loan default for the sample. The independent variables are:actual cap rate in columns (i) msa synthetic cap rate (SCX) in columns (ii). Panel A captures the entire sample; Panel B the CBD properties; and Panel C the non-CBD properties. Estimates are provided along with standard errors in parentheses. ***/**/*/' indicate 0.1%, 1%, 5% and 10% significance for Wald z-values

The significance of the estimates is indicated by the z-values produced by the Wald test. The significance of the regression overall is determined with Chi-squared test and we use McFadden's adjusted pseudo R-squared which measures 1 minus the ratio of the estimated log likelihood for the model with parameters (adjusted for the number of parameters) with that of the null model. The results show higher cap rates are associated with higher probability of default at the 1% significance level with better Chi-Squared and Adjusted Pseudo-R-squared statistics. Interestingly, the results show that SCXs have greater sensitivity to lifetime default risk than actual cap rates for the entire sample and in each of the CBD and non-CBD sub-samples. These disparities in sensitivity are most pronounced for loans categorized as non-CBD loans where SCXs are highly significant and actual cap rates are insignificantly different from zero. This result supports claims of misspecification of CRE risk when lending outside of CBDs.

5.4.2 *Second Treatment*

In the second treatment we consider five independent variables x_1= cap rate (actual or synthetic); x_2= occupancy rate; x_3= property age; x_4= CBD dummy with 1 indicating property location within a CBD and 0 indicating property location outside of a CBD; and x_5= property type fixed effects which exclude the 1009 mixed-use/other properties (OT). The results in Table 5.5 are in-line with Table 5.4.

Table 5.5: Logistic regression - Treatment 2

	Panel A: All		Panel B: CBD only		Panel C: Non CBD only	
	(i)	(ii)	(i)	(ii)	(i)	(ii)
_cons	-2.135964***	-5.445156***	-2.51904***	-5.755882***	-1.642249***	-4.926696***
	(0.2221568)	(0.4036153)	(0.5220886)	(1.002803)	(0.2321479)	(0.4326184)
caprate (actual)	0.6873697		4.242379***		0.5254299	
	(0.4414735)		(1.846048)		(0.5005424)	
caprate (SCX)		46.00407***		48.65101***		45.53765***
		(4.573642)		(11.72515)		(4.968723)
occpct	-0.0129654***	-0.0133184***	-0.0128622***	-0.0134254***	-0.0129051***	-0.0133027***
	(0.0015615)	(0.0015635)	(0.0040888)	(0.0040364)	(0.0016923)	(0.001696)
propage	-0.0137843***	-0.0116875***	-0.0125176*	-0.0124223*	-0.0139661***	-0.0114937***
	(0.0033768)	(0.003344)	(0.007349)	(0.0072263)	(0.0037987)	(0.0037801)
cbd	0.4487141***	0.4505224***				
	(0.0920274)	(0.092203)				
prp_fxd	-0.2044051	-0.2146195	-0.0843447	-0.0759683	-0.2427588	-0.2522155
	(0.1567765)	(0.1571089)	(0.3503293)	(0.350645)	(0.1753219)	(0.1757145)
Adj-Pseudo R-sq	0.0047	0.0085	0.0034	0.0061	0.0038	0.0080
Prob chi2	0.0000	0.0000	0.0003	0.0000	0.0000	0.0000
Obs	25101	25101	5195	5195	19906	19906

This table shows the results of logistic regressions with 0,1 indicator of loan default for the sample. The independent variables are:actual cap rate in columns (i) msa synthetic cap rate (SCX) in columns (ii). Panel A captures the entire sample; Panel B the CBD properties; and Panel C the non-CBD properties. Estimates are provided along with standard errors in parentheses. ***/**/*/' indicate 0.1%, 1%, 5% and 10% significance for Wald z-values

Occupancy rate and property age are highly significant as expected. Higher likelihoods of default are associated with lower occupancy rates at origination and older properties (possibly due to more stable tenancy and usage). Presence within CBDs is highly significant and positive suggesting higher likelihoods of default in CBDs. Property fixed effects are insignificant. Consistent with the first treatment, SCXs are highly significant predictors of default across all groupings. In particular, non-CBD results again indicate that actual cap rates provide no statistically significant insights into the lifetime risk of CRE default. This contrasts with the significant assessment of CRE risk articulated by SCXs.

5.4.3 *Third Treatment*

The purpose of SCXs is to isolate the time-agnostic aspect of the model. As such, in the third treatment, we control for distinct lending regime time fixed effects with an additional

independent variable x_6. We divide the sample into origination sub-periods corresponding to: i.) pre-911 (34 loans originated prior to September 2001); ii.) pre-crisis (11192 loans originated between October 2001 and December 2006); iii.) crisis (8736 loans originated between January 2007 and December 2009); and iv.) recovery (5139 loans originated between January 2010 and December 2014). The mean loan-to-value ratios (LTVs) of 58%, 68%, 74%, and 63% corresponding to these regimes map to intuition. Following the September 11th terrorist attack, lending restrictions in the US eased (higher LTVs than pre-911). Lending became freer still in the crisis (with higher leverage) in carry over funding obligations until origination ceased in mid-2009. In the recovery period, when lending resumed in 2011, a new lending paradigm characterized by lower leverage took hold with lower LTVs than observed in both pre-crisis and crisis periods. Because of distinct phase shifts in lending regimes, we project our time fixed effect with dummies from pre-911, pre-crisis, and recovery periods, 'bookending' the crisis period. Additionally, we consider the publicly observable VIX volatility index as another covariate, x_7, to capture exogenous market volatility. We map end of day closing values for VIX values to corresponding origination dates for each loan in our sample. The results found in Table A.9 echo key observations from the two prior treatments. Property fixed effects are insignificant while occupancy, property age, and CBD location remain highly significant. The lending regime time fixed-effects and VIX exhibit significance with little distinction between sub-samples. The main results persist: SCXs exhibit more significant assessment of lifetime property default risk than actual cap rates outside of CBDs, with equivalent assessments within CBDs.

6 Summary

The CRE market in the US faces an inherent informational deficit. This deficit is indicated by infrequent transactions which can be ascribed to idiosyncratic property characteristics and large asset size. Our methodology discloses important links between the macroeconomy and CRE valuation not before seen in the literature in SCXs projected from public information. SCXs appear to provide insights into CRE property valuation and lifetime default estimations that improve upon those made with actual cap rates. Given the comparative accuracy that

SCXs provide, legitimate questions are raised as to the validity of a geographic centric lending paradigm in CRE and the sensitivity of that market to macroeconomic influences. We provide some answers in this study. The results suggest that misspecified reliance upon location of a property may have been too coarse a benchmark that did not discern the public information content influence on property values. This, in turn, erroneously influenced actual valuation estimates and lending practices at the time of loan origination. Since SCXs by their nature challenge the primacy of geographic location in valuation, we demonstrate the additional contribution of providing new hard information content which may assist in better valuation estimates of illiquid CRE assets in an informationally constrained market.

In future work, expansions of this new methodology can be made to property-type levels and more granular county level. These extensions of our work might improve sensitivity with additional factors that satisfy our methodological requirements which may be of interest to regulatory policy in stress test evaluation and bank capital adequacy requirements as well as, more tactically, in the context of risk management, origination and investment. As shown in the valuation section, SCX estimates may provide considerably different valuations of CRE properties than actual market consensus cap rates. This is as expected given the improved accuracy in lifetime default estimation provided by SCXs, particularly outside of CBDs and well-documented irrational exuberance in underwriting practices during the Great Financial Crisis embedded within our sample period. This study leaves opens the question as to the origins of the apparent misplaced reliance upon geographic proximity and coarser notions of location, rather than economic proximity, which is left to future work. SCXs represent new hard information derived from public information content based on economic proximity that may be used to bridge part of the information content deficit in CRE.

7 Compliance with Ethical Standards

Andreas D. Christopoulos declares that he has no conflict of interest. Joshua G. Barratt declares that he has no conflict of interest. Daniel C. Ilut declares that he has no conflict of interest.

This article does not contain any studies with human participants or animals performed

by any of the authors.

References

[1] Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse. "A regression method for real estate price index construction." Journal of the American Statistical Association 58 (1963), 933-942.

[2] Bond, Shaun A., and Paul Mitchell. "The Information Content of Real Estate Derivative Prices." The Journal of Portfolio Management 37 (2011), 170-181.

[3] Brueggeman, William B., and Jeffrey D. Fisher. Real estate finance and investments - 16th Edition. New York, NY, McGraw-Hill Irwin (2019).

[4] Case, Karl E., and Robert J. Shiller. "Prices of single-family homes since 1970: new indexes for four cities." New England Economic Review Sep (1987), 45-56.

[5] Chegut, Andrea M., Piet MA Eichholtz, and Paulo JM Rodrigues. "Spatial dependence in international office markets." The Journal of Real Estate Finance and Economics 51 (2015), 317-350.

[6] Chervachidze, Serguei, James Costello, and William C. Wheaton. "The secular and cyclic determinants of capitalization rates: the role of property fundamentals, macroeconomic factors, and structural changes." The Journal of Portfolio Management 35 (2009), 50-69.

[7] Christopoulos, Andreas D. "The composition of CMBS risk." Journal of Banking & Finance 76 (2017), 215-239.

[8] Clayton, Jim, David Geltner, and Stanley W. Hamilton. "Smoothing in commercial property valuations: Evidence from individual appraisals." Real Estate Economics 29 (2001), 337-360.

[9] Commercial Real Estate Finance Council, CREFC Principles-Based Underwriting Framework (2018).

[10] Diehl, Joseph. "The Russell-NCREIF Property Indices: Institutional Real Estate Performance Benchmarks." Journal of Real Estate Literature 1.1 (1993), 94-103.

[11] Dokmanic, Ivan, Reza Parhizkar, Juri Ranieri, and Martin Vetterli. "Euclidean distance matrices: essential theory, algorithms, and applications." IEEE Signal Processing Magazine 32 (2015), 12-30.

[12] Figlewski, Stephen, Halina Frydman, and Weijian Liang. "Modeling the effect of macroeconomic factors on corporate default and credit rating transitions." International Review of Economics & Finance 21 (2012), 87-105.

[13] Fisher, Jeffrey D. "A Repeat Sales Index For Commercial Real Estate-Using Sold Properties in the NCREIF Database." Real Estate Finance 17.2 (2000): 66-71.

[14] Gordon, Myron J., and Eli Shapiro. "Capital equipment analysis: the required rate of profit." Management science 3.1 (1956): 102-110.

[15] Grissom, Terry V., David Hartzell, and Crocker H. Liu. "An approach to industrial real estate market segmentation and valuation using the arbitrage pricing paradigm." Real Estate Economics 15 (1987), 199-219.

[16] Halcoussis, Dennis. Understanding Econometrics. Thomson-South Western, Australia (2005).

[17] Hartzell, David, David Shulman, and Charles Wurtzebach. "Refining the analysis of Regional diversification for income-producing real estate." Journal of Real Estate Research 2 (1987), 85-95.

[18] He, An and Bruce Mizrach. "Analysis of securitized asset liquidity - Research Note" FINRA Office of the Chief Economist (2017).

[19] Hollifield, Burton, Artem Neklyudov, and Chester Spatt. "Bid-ask spreads, trading networks, and the pricing of securitizations." The Review of Financial Studies 30 (2017), 3048-3085.

[20] Hyun, Dongwoo, and Stanimira Milcheva. "Spatial dependence in apartment transaction prices during boom and bust." Regional Science and Urban Economics 68 (2018), 36-45.

[21] Jiang, Yuxiang, Tal Ronnen Oron, Wyatt T. Clark, Asma R. Bankapur, Daniel D'Andrea, Rosalba Lepore, Christopher S. Funk et al. "An expanded evaluation of protein function prediction methods shows an improvement in accuracy." Genome biology 17 (2016), 184.

[22] Kohavi, Ron. "A study of cross-validation and bootstrap for accuracy estimation and model selection." International Joint Conference on Artificial Intelligence 14 (1995), 1137-1145.

[23] Kramer, Walter, and Andre Guttler. "On comparing the accuracy of default predictions in the rating industry." Empirical Economics 34 (2008), 343-356.

[24] Killian, Chris. and Joseph Cox. "US Securitization First Half 2016" SIFMA Securitization Report, SIFMA Securitization Group (2016).

[25] Liberti, Jose Mar and Mitchell A. Petersen. "Information: Hard and soft." Review of Corporate Finance Studies 8.1 (2018): 1-41.

[26] Marcato, Gianluca, and Anupam Nanda. "Information content and forecasting ability of sentiment indicators: case of real estate market." Journal of Real Estate Research 38 (2016), 165-203.

[27] Mueller, Glenn. "Refining economic diversification strategies for real estate portfolios." Journal of Real Estate Research 8 (1993), 55-68.

[28] National Association of Realtors, “NAR Commercial Leading Indicator for Brokerage Activity”, (https://www.nar.realtor/sites/default/files/reports/2006/description-methodology-background-2006-05.pdf” (2006).

[29] Office of the Comptroller of the Currency. “Commercial real estate lending” 1 (2017).

[30] Pagliari, Joseph L., Frederich Lieblich, Mark Schaner, and James R. Webb. "Twenty years of the NCREIF property index." Real Estate Economics 29 (2001), 1-27.

[31] Pavlov, Andrey, Eva Steiner, and Susan Wachter. "The consequences of REIT index membership for return patterns." Real Estate Economics 46 (2018), 210-250.

[32] Peng, Liang. "How integrated is the commercial real estate asset market? Evidence from transaction cap rates." University of Colorado at Boulder working paper (2013).

[33] Seagraves, Philip A., and Jonathan A. Wiley. "The cap rate spread: A new metric for commercial underwriting." Real Estate Economics 44 (2016), 490-520.

[34] Shannon, Claude Elwood. "A mathematical theory of communication." Bell system technical journal 27 (1948), 379-423.

[35] Silver, Mick. "Real-estate price indexes. Availability, importance, and new developments." Reality, Data and Space. International Journal of Statistics and Geography 7 (2016), 4-25.

[36] Sivitanides, Petros, Jon Southard, Raymond G. Torto, and William C. Wheaton. "The determinants of appraisal-based capitalization rates." Real Estate Finance 18 (2001), 27-38.

[37] Titman, Sheridan. Valuation. Pearson Higher Ed, 2014.

[38] Titman, Sheridan, Stathis Tompaidis, and Sergey Tsyplakov. "Determinants of credit spreads in commercial mortgages." Real Estate Economics 33 (2005), 711-738.

[39] Trepp, LLC. “How the CMBS market thawed out from its ice age”. Trepp CMBS Research, (2018), 1-6.

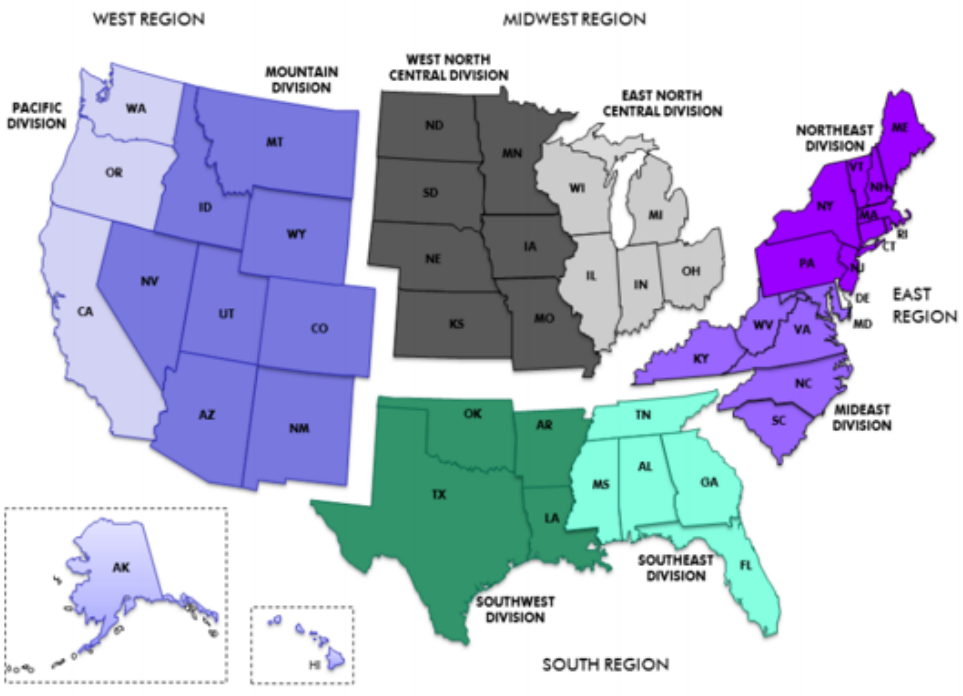
[40] Yan, Alice Xie, Jian Shi, and Chunchi Wu. "Do macroeconomic variables matter for pricing default risk?." International Review of Economics & Finance 17 (2008), 279-291.

[41] Wheaton, William C., and Gleb Nechayev. "Does location matter?." The Journal of Portfolio Management 31 (2005), 100-108.

A Online Appendix

A.1 Support Figures and Tables

Figure A.1: NCREIF regional partitions



This figure shows NCREIF's partitioning of the US into 8 distinct regions. Source: NCREIF.

Table A.1: Summary statistics for 60 NCREIF cap rate time series

Numeric Code, j	Supralocations	min	max	mean	median	kurtosis	variance	stdev
1	Alaska	0.0384	0.1013	0.0777	0.0802	0.1260	0.0002	0.0127
2	Alabama	0.0462	0.1489	0.0811	0.0738	0.3366	0.0006	0.0252
3	Arkansas	0.0552	0.0942	0.0769	0.0757	-1.1632	0.0001	0.0108
4	Arizona	0.0491	0.1148	0.0742	0.0709	-0.6291	0.0002	0.0146
5	California	0.0474	0.0970	0.0707	0.0695	-1.1543	0.0002	0.0137
6	Colorado	0.0514	0.1031	0.0734	0.0697	-1.1840	0.0002	0.0138
7	Connecticut	0.0529	0.1195	0.0773	0.0755	0.1544	0.0002	0.0150
8	Washington, DC	0.0447	0.1089	0.0703	0.0686	-1.1732	0.0003	0.0174
9	Delaware	0.0518	0.0942	0.0751	0.0738	-1.2720	0.0001	0.0121
10	Florida	0.0511	0.0924	0.0714	0.0697	-1.1173	0.0001	0.0113
11	Georgia	0.0524	0.0988	0.0741	0.0710	-1.2624	0.0002	0.0123
12	Hawaii	0.0242	0.1023	0.0708	0.0733	-0.5226	0.0003	0.0165
13	Iowa	0.0536	0.1822	0.0787	0.0782	13.6162	0.0003	0.0167
14	Idaho	0.0554	0.0962	0.0747	0.0726	-1.4095	0.0001	0.0118
15	Illinois	0.0493	0.0931	0.0714	0.0696	-1.0914	0.0001	0.0114
16	Indiana	0.0530	0.0995	0.0789	0.0794	-0.7841	0.0001	0.0110
17	Kansas	0.0611	0.1265	0.0812	0.0780	1.0358	0.0002	0.0133
18	Kentucky	0.0570	0.1256	0.0870	0.0842	-0.9619	0.0002	0.0156
19	Louisiana	0.0165	0.1986	0.0846	0.0855	5.1645	0.0007	0.0255
20	Massachusetts	0.0453	0.1021	0.0724	0.0709	-1.0300	0.0002	0.0153
21	Maryland	0.0528	0.0992	0.0752	0.0721	-1.3218	0.0002	0.0135
22	Maine	0.0528	0.0992	0.0752	0.0721	-1.3218	0.0002	0.0135
23	Michigan	0.0567	0.1060	0.0787	0.0789	-0.2329	0.0001	0.0097
24	Minnesota	0.0547	0.1112	0.0792	0.0781	-1.2290	0.0002	0.0151
25	Missouri	0.0604	0.1033	0.0783	0.0734	-1.1689	0.0001	0.0121
26	Mississippi	0.0575	0.0942	0.0787	0.0798	-1.0544	0.0001	0.0100
27	Montana	0.0554	0.0962	0.0747	0.0726	-1.4095	0.0001	0.0118
28	North Carolina	0.0563	0.0987	0.0769	0.0748	-1.1124	0.0001	0.0114
29	North Dakota	0.0566	0.0967	0.0773	0.0750	-1.3071	0.0001	0.0114
30	Nebraska	0.0471	0.0942	0.0776	0.0787	-0.3542	0.0001	0.0111
31	New Hampshire	0.0549	0.0986	0.0758	0.0740	-1.4335	0.0002	0.0123
32	New Jersey	0.0513	0.1130	0.0730	0.0710	-0.0963	0.0002	0.0134
33	New Mexico	0.0279	0.1090	0.0723	0.0721	0.6978	0.0002	0.0134
34	Nevada	0.0456	0.1005	0.0726	0.0697	-0.2566	0.0001	0.0112
35	New York	0.0335	0.1031	0.0681	0.0689	-0.9773	0.0003	0.0169
36	Ohio	0.0575	0.1072	0.0769	0.0742	-1.0016	0.0002	0.0123
37	Oklahoma	0.0538	0.0948	0.0799	0.0805	-0.8837	0.0001	0.0091
38	Oregon	0.0531	0.1043	0.0745	0.0729	-1.0545	0.0002	0.0127
39	Pennsylvania	0.0522	0.1067	0.0758	0.0713	-1.1170	0.0002	0.0140
40	Rhode Island	0.0536	0.0951	0.0746	0.0730	-1.3922	0.0002	0.0123
41	South Carolina	0.0494	0.1738	0.0820	0.0726	2.3430	0.0005	0.0223
42	South Dakota	0.0566	0.0967	0.0773	0.0750	-1.3071	0.0001	0.0114
43	Tennessee	0.0586	0.1075	0.0787	0.0756	-0.9193	0.0002	0.0123
44	Texas	0.0530	0.0989	0.0742	0.0708	-1.4108	0.0002	0.0125
45	Utah	0.0447	0.1327	0.0787	0.0765	-0.3247	0.0003	0.0177
46	Virginia	0.0503	0.0931	0.0718	0.0709	-1.4969	0.0002	0.0138
47	Vermont	0.0536	0.0951	0.0746	0.0730	-1.3922	0.0002	0.0123
48	Washington	0.0500	0.1005	0.0728	0.0740	-1.0636	0.0002	0.0134
49	Wisconsin	0.0552	0.1166	0.0815	0.0813	-0.6430	0.0002	0.0133
50	West Virginia	0.0567	0.0959	0.0760	0.0732	-1.4461	0.0002	0.0123
51	Wyoming	0.0554	0.0962	0.0747	0.0726	-1.4095	0.0001	0.0118
52	US - Nation	0.0552	0.0942	0.0755	0.0738	-1.3863	0.0001	0.0116
53	NCREIF_East North Central	0.0528	0.0966	0.0734	0.0706	-1.1551	0.0001	0.0111
54	NCREIF_Mideast	0.0515	0.0948	0.0730	0.0702	-1.5098	0.0002	0.0139
55	NCREIF_Northeast	0.0438	0.0975	0.0712	0.0704	-1.1514	0.0002	0.0147
56	NCREIF_Southeast	0.0526	0.0945	0.0731	0.0705	-1.2853	0.0001	0.0116
57	NCREIF_Southwest	0.0530	0.0989	0.0744	0.0708	-1.3968	0.0002	0.0124
58	NCREIF_Midwest	0.0446	0.1001	0.0732	0.0710	-1.1555	0.0002	0.0130
59	NCREIF_West North Central	0.0573	0.1042	0.0788	0.0765	-1.4151	0.0002	0.0128
60	NCREIF_Western Pacific	0.0473	0.0882	0.0679	0.0682	-1.2436	0.0001	0.0107
	Averages All Supralocations	0.0505	0.1068	0.0757	0.0738	-0.5520	0.0002	0.0135

This table shows the summary statistics over the period Q1 1989 to Q2 2015 for 60 NCREIF cap rates time series determined from the procedures described in this paper.

Table A.2: Summary statistics of defaulted loans

Summary		Averages for defaulted loans				# defaults by proptype					
CBD/non-CBD	Defaults	caprate (actual)	SCX	occupancy (%)	propage (yrs)	MF	LO	IN	OF	RT	OT
non-CBDs	870	7.35%	7.64%	85.05	7.85	53	100	74	218	389	36
Atlanta	10	9.00%	7.47%	78.63	11.80	0	1	2	6	1	0
Baltimore	4	6.50%	8.20%	86.55	5.00	1	1	0	2	0	0
Boston	0	NA	NA	NA	NA	0	0	0	0	0	0
Charlotte	7	9.14%	7.73%	88.14	9.57	2	0	1	2	2	0
Chicago	3	10.00%	7.58%	78.97	3.00	0	1	0	2	0	0
Cincinnati	5	7.40%	7.92%	92.60	12.20	0	0	0	2	3	0
Cleveland	2	7.00%	7.58%	78.89	2.50	0	1	0	0	1	0
Dallas	4	9.00%	7.70%	92.50	5.75	0	0	0	1	3	0
Dayton	2	7.50%	7.67%	88.70	4.50	0	0	1	1	0	0
Denver	8	6.75%	8.17%	86.00	15.00	0	0	2	3	3	0
Detroit	4	6.25%	8.70%	70.30	12.25	1	1	0	1	1	0
Fort Worth	3	4.67%	7.12%	77.33	2.67	0	0	1	0	2	0
Houston	14	6.93%	7.72%	85.50	11.14	1	0	0	9	4	0
Indianapolis	6	9.00%	7.43%	85.33	16.50	0	1	0	2	2	1
Jacksonville	3	12.00%	7.33%	99.00	10.00	0	0	1	1	0	1
Kansas City	5	7.20%	8.30%	82.32	8.60	0	1	0	2	2	0
Los Angeles	3	6.67%	9.56%	92.67	6.33	0	0	0	0	3	0
Miami	1	8.00%	6.20%	95.00	1.00	1	0	0	0	0	0
Milwaukee	2	7.00%	7.71%	92.24	10.00	0	0	1	1	0	0
Minneapolis	4	7.50%	9.05%	93.00	12.25	0	0	2	1	1	0
New Orleans	1	6.00%	7.66%	100.00	1.00	0	0	0	1	0	0
New York	2	5.00%	7.58%	98.50	1.00	1	0	0	0	1	0
Norfolk	0	NA	NA	NA	NA	0	0	0	0	0	0
Oakland	3	6.00%	8.42%	78.00	5.33	0	0	2	0	1	0
Oklahoma City	4	8.00%	7.80%	70.80	9.00	1	0	0	1	0	2
Philadelphia	6	5.83%	7.74%	84.50	8.33	0	0	1	1	3	1
Phoenix	9	6.78%	7.25%	83.52	7.78	1	1	1	3	3	0
Pittsburgh	1	9.00%	8.43%	100.00	1.00	0	0	0	1	0	0
Richmond	2	4.50%	7.18%	81.50	4.50	0	1	0	1	0	0
Saint Paul	6	7.50%	7.18%	85.34	9.83	0	0	2	3	1	0
San Diego	7	6.71%	7.42%	84.39	14.29	0	1	1	3	1	1
San Francisco	2	5.00%	7.49%	97.35	7.00	2	0	0	0	0	0
San Jose	1	7.00%	7.37%	100.00	7.00	0	0	0	0	0	1
Seattle	1	6.00%	6.96%	100.00	11.00	0	0	0	0	0	1
St. Louis	0	NA	NA	NA	NA	0	0	0	0	0	0
Tampa	4	9.00%	7.18%	87.00	8.25	0	0	0	2	1	1
Washington, D.C.	4	6.25%	7.29%	87.75	3.00	1	1	0	1	1	0
Totals & Avgs	1013	7.36%	7.65%	85.13	8.05	65	111	92	271	429	45

This table provides summary statistics for the 1,013 loans in our sample of 25,101 CRE mortgages that defaulted over the sample period. The defaulted loans are grouped in rows that correspond to their CBD or non-CBD locations. The total number of loans which defaulted within the CBD/non-CBD location classification is provided immediately following the first column. The next four columns report averages for the defaulted loans at the time of origination. These averages include the actual caprate, followed by the synthetic cap rate (SCX) corresponding to the MSA within which the property is located as well as occupancy rate and property age. Finally, the last six columns show the counts of defaulted loans within the CBD/non-CBD locations distributed across the property types of multifamily/apartments (MF), lodging/hotel (LO), industrial (IN), office (OF), retail (RT), and mixed-use/other (OT). The bottom row provides totals and averages. Source: Intex.

Table A.3: OLS for 60 supralocations

Numeric Code, j	Supralocations	_cons	crechargeoff	creditslope	hpi	mortgagerate	ue	Adj R-sq
1	Alaska	0.0569***	-0.0375***	-0.0006***	0.405**	0.0798***	-0.0298**	0.626
2	Alabama	0.3047***	-0.0391**	-0.0007	-0.5896**	-0.0186***	-0.0026**	0.441
3	Arkansas	0.1775***	0.0546***	-0.0005***	0.4682*	-0.0126***	0.0006*	0.245
4	Arizona	0.2881***	-0.0752***	-0.0008*	0.5189	-0.0140***	0.0008***	0.759
5	California	0.5511***	-0.0523***	-0.0011***	-0.1535*	-0.0332***	-0.0143*	0.742
6	Colorado	0.3303***	-0.0580***	-0.0010***	-0.5307*	-0.0189***	0.0015	0.649
7	Connecticut	0.2765***	-0.0722***	-0.0006***	-0.2203**	-0.0077***	-0.007	0.629
8	DC	0.52***	-0.0219***	-0.0009***	0.2565	-0.0267***	-0.0218	0.548
9	Delaware	0.0339***	1.4308***	-0.0046***	-2.9421***	0.247***	-0.1141	0.983
10	Florida	0.3856***	-0.0152***	-0.0008***	0.8226***	-0.0240***	-0.0115	0.761
11	Georgia	0.1364***	-0.0769***	-0.0010***	-0.5202	0.0005***	0.0136***	0.821
12	Hawaii	0.3829***	-0.0274***	-0.0009	-0.1171	-0.013***	-0.0215	0.623
13	Iowa	0.0202***	-0.0515***	0***	0.4004	0.026	0.0026	0.568
14	Idaho	0.3315***	-0.0625***	-0.0009***	-0.2528***	-0.0184***	-0.0012***	0.737
15	Illinois	0.4298***	-0.0383***	-0.0008***	-0.0401***	-0.0288***	-0.0113	0.767
16	Indiana	0.2773***	-0.0400***	-0.0007***	-0.1236	-0.0134***	-0.005***	0.741
17	Kansas	0.2395***	-0.0307***	-0.0009	-0.0785	-0.0053***	-0.0054***	0.774
18	Kentucky	0.2954***	-0.0413***	-0.0006**	-0.3258	-0.0139***	-0.0065	0.515
19	Louisiana	0.4238**	-0.0379**	-0.0015	0.5519	-0.0330	0.0052	0.784
20	Massachusetts	0.5507***	-0.0404***	-0.0012***	-0.3993	-0.0278***	-0.0269***	0.8
21	Maryland	0.4053***	-0.0307***	-0.0007***	0.7276***	-0.0229***	-0.0179***	0.712
22	Maine	0.3362***	-0.0576***	-0.0009***	0.0456	-0.0191***	-0.0022**	0.733
23	Michigan	0.2652***	-0.0386***	-0.0008***	0.2468	-0.0158***	0.0002	0.724
24	Minnesota	0.2325***	-0.0552***	-0.0007***	-0.4272	-0.0110***	0.0026***	0.725
25	Missouri	0.1342***	-0.0563***	-0.0004**	-0.2168	-0.0002***	0.0027	0.805
26	Mississippi	0.4247***	-0.0208***	-0.0008***	0.3057	-0.0251***	-0.0154	0.849
27	Montana	0.3288***	-0.0614***	-0.0009***	-0.3168***	-0.0179***	-0.001***	0.736
28	North Carolina	0.2397***	-0.0623***	-0.0007*	-0.3741	-0.0104***	0.0024	0.682
29	North Dakota	0.3320***	-0.0604***	-0.0009***	-0.1334	-0.0184***	-0.0034	0.734
30	Nebraska	0.1743***	-0.0690***	-0.0008**	-1.2551	-0.017***	0.0348	0.711
31	New Hampshire	0.3002***	-0.0452***	-0.0006***	-0.0097	-0.0115***	-0.0088	0.557
32	New Jersey	0.3882***	-0.0401***	-0.0009***	0.4753	-0.0249***	-0.0074	0.717
33	New Mexico	0.1924***	-0.0502***	0***	0.7064***	-0.0146***	-0.0017	0.398
34	Nevada	0.3192***	-0.0148***	-0.0007	0.8142***	-0.0164***	-0.0102***	0.746
35	New York	0.6344***	-0.0643***	-0.0013***	0.1806	-0.0374***	-0.0239	0.719
36	Ohio	0.2349***	-0.0433***	-0.0008***	-0.2859***	-0.0166***	0.0045*	0.811
37	Oklahoma	0.2700***	-0.0651***	-0.0010**	-0.2651	-0.0179***	0.0103	0.841
38	Oregon	0.4040***	-0.0393***	-0.0005***	0.2554***	-0.0224***	-0.0154	0.733
39	Pennsylvania	0.4250***	-0.0318***	-0.0007***	0.4223***	-0.0263***	-0.0173***	0.726
40	Rhode Island	0.4475***	-0.0450***	-0.0010***	-0.2847***	-0.0191***	-0.0083***	0.752
41	South Carolina	0.2412***	-0.0519***	-0.0015	-0.1773**	0.011***	-0.0001**	0.781
42	South Dakota	0.2268***	-0.0439***	-0.0011***	-0.5193	-0.0054***	0.0054	0.622
43	Tennessee	0.2815***	-0.0675***	-0.0009**	-0.5484	-0.0159***	0.0107	0.742
44	Texas	0.2415***	-0.0546***	-0.0008***	0.0745*	-0.0147***	0.0053	0.716
45	Utah	0.4293***	-0.0315***	-0.0008	-0.2909	-0.0246***	-0.0153***	0.737
46	Virginia	0.2527***	-0.0339***	-0.0003***	0.9157***	-0.0142***	-0.009	0.545
47	Vermont	0.4106***	-0.0215***	-0.0008***	0.6612***	-0.0213***	-0.0220	0.634
48	Washington	0.3407**	-0.0552***	-0.0009***	0.0811	-0.0188***	-0.0052	0.735
49	Wisconsin	0.4150***	-0.0355***	-0.0010***	-0.0209*	-0.0180***	-0.0134**	0.658
50	West Virginia	0.3853***	-0.0371***	-0.0009***	0.199	-0.0288***	-0.007	0.629
51	Wyoming	0.0894***	-0.0108***	-0.0003***	0.0125*	-0.0051***	-0.0004***	0.582
52	US - Nation	0.3521***	-0.055***	-0.0009***	0.1391***	-0.0192***	-0.0068***	0.82
53	East NorthCentral	0.3031***	-0.0499***	-0.0009***	-0.0025***	-0.0158***	-0.0025	0.817
54	Mideast	0.3405***	-0.0351***	-0.0008***	0.1205***	-0.0184***	-0.008***	0.731
55	Northeast	0.4692***	-0.0535***	-0.0009***	-0.1186	-0.0251***	-0.0175	0.76
56	Southeast	0.2953***	-0.0452***	-0.0008	-0.0062***	-0.0164***	-0.0011**	0.763
57	Southwest	0.4227***	-0.0456***	-0.0011***	-0.4465*	-0.0240***	-0.009	0.786
58	Midwest	0.2082***	-0.0579***	-0.0006	0.1292**	-0.0106***	0.0042	0.729
59	West NorthCentral	0.2447***	-0.0637***	-0.0008***	-0.5812	-0.0046***	-0.0016	0.853
60	Western Pacific	0.4676***	-0.0610***	-0.0010	-1.0840	-0.0150***	-0.0160***	0.821

This table summarizes the estimates and R-squared values for the 60 supralocation OLS. ***/**/* indicate 1%, 5% and 10% significance levels.

A.2 Discussion of OLS robustness checks

We perform many standard econometric robustness checks on the model at national, regional and state levels corresponding to the OLS models described in the main body of the paper. The purpose of these initial tests is to ensure that the initial model choices we have made

are valid and the significance of the parameter estimates associated with those choices are accurate. The tests we executed are summarized below with supporting results.

A.2.1 Omitted variables test

To test for omitted variables we use the Ramsey RESET test. The test indicates, in about 80% of the cases, the presence of omitted variables. This is not uncommon. Based on our evaluation of the performance and stability of the modeling results post-estimation we do not find a need add additional variables to the model. These results are available upon request.

A.2.2 Multicollinearity

Next, we consider the issue of multicollinearity amongst the explanatory variables. The mean variance inflation factor, $VIF_j(\mu)$, is calculated as the average across all $i \in [1, 5]$ parameters where $VIF_j(\hat{\beta}_i) = 1/(1 - R_i^2)$ values for the $j - th$ location. The Condition Number is the value of largest Condition Index $= \frac{\sqrt{(\lambda_{max})}}{\sqrt{(\lambda_{min})}}$ representing the ratio of largest to smallest eigenvalue of $X'X$. Both the VIF and Condition Index Number represent overall measures for model multicollinearity. The thresholds which indicate the presence of multicollinearity are 10 and 30 for each of these measures, respectively. Summary results for the Condition Index are shown in Table A.4. Based upon these thresholds we do not see substantive evidence of multicollinearity in our OLS models other than for 1 supralocation and thus safely dismiss these concerns.

A.2.3 Serial correlation

Additionally, we consider whether there is serial correlation in the residual of the OLS models by using Durbin-Watson ("DW") test statistic. After adjusting for 6 degrees of freedom and 93 observations, the two critical values at the 5% significance level for Durbin-Watson are: $d_{Lower} = 1.529$ and $d_{Upper} = 1.800$. The test results are summarized in Table A.5 for all 60 locations. In general the DW test results indicate the presence of serial correlation in the OLS models. The presence of autocorrelation has no impact on the parameter estimates, goodness of fit, or overall significance of the regression as measured by the F-stat. As we are primarily interested in projection of SCXs using the parameter estimates, we simply use the

Table A.4: Multicollinearity test statistics

Numeric Code, j	Supralocations	Condition Index Number
1	Alaska	23.89
2	Alabama	20.26
3	Arkansas	23.62
4	Arizona	23.44
5	California	22.31
6	Colorado	25.72
7	Connecticut	22.54
8	Washington, DC	26.07
9	Delaware	21.09
10	Florida	19.90
11	Georgia	25.10
12	Hawaii	21.07
13	Iowa	24.71
14	Idaho	19.34
15	Illinois	23.09
16	Indiana	23.51
17	Kansas	27.78
18	Kentucky	24.29
19	Louisiana	21.10
20	Massachusetts	18.92
21	Maryland	21.97
22	Maine	19.61
23	Michigan	22.11
24	Minnesota	22.01
25	Missouri	23.30
26	Mississippi	24.67
27	Montana	18.95
28	North Carolina	25.83
29	North Dakota	18.28
30	Nebraska	31.94
31	New Hampshire	17.83
32	New Jersey	22.44
33	New Mexico	21.86
34	Nevada	23.11
35	New York	22.71
36	Ohio	24.00
37	Oklahoma	21.01
38	Oregon	25.02
39	Pennsylvania	22.30
40	Rhode Island	25.07
41	South Carolina	20.52
42	South Dakota	25.93
43	Tennessee	24.91
44	Texas	23.89
45	Utah	24.24
46	Virginia	22.38
47	Vermont	19.70
48	Washington	18.33
49	Wisconsin	23.77
50	West Virginia	26.10
51	Wyoming	17.46
52	US - Nation	19.55
53	NCREIF_East North Central	23.17
54	NCREIF_Mideast	28.99
55	NCREIF_Northeast	21.98
56	NCREIF_Southeast	26.87
57	NCREIF_Southwest	27.79
58	NCREIF_Midwest	26.06
59	NCREIF_West North Central	22.41
60	NCREIF_Western Pacific	20.92
min	1.32	17.46
max	2.75	31.94
mean	1.69	22.95

This table summarizes test results for multicollinearity amongst the regressors in for the ordinary least squares regressions of the form $C_{j,\,t} = \alpha + \beta_1 \text{HPI}_{j,\,t-4} + \beta_2 \text{UE}_{j,\,t-4} + \beta_3 \text{CreditSlope}_{j=1,t-4} + \beta_4 \text{MtgRate}_{j=1,t-4} + \beta_5 \text{CREchgoff}_{j=1,t-4} + \epsilon_t$. The dependent variable $C_{j,\,t}$ is the j-th cap rate index calculated from the NCREIF data with $j = 1...60$ geographic locations. Both HPI and UE independent variables also span all j geographic locations while each of the credit slope, mortgage rate and CRE charge off rates are only reported at national ($j = 1$) levels. The mean variance inflation factor,$VIF(\mu)$, is calculated as the average across all $i = 5$ parameter $VIF(\hat{\beta}_i) = 1/(1 - R_i^2)$ values for the $j - th$ location. The Condition Number is the value of largest Condition Index $= \sqrt{(\lambda_{max})}/\sqrt{(\lambda_{min})}$. Both the VIF and Condition Index Number represent overall measures for model mulicollinearity with thresholds indicating multicollinearity of 10 and 30, respectively.

common procedure of using Newey-West standard errors which will impact the significance of parameters. As previously discussed in the main text, we provide OLS results at the national level corrected for Newey-West standard errors. Compared with uncorrected OLS findings previously discussed, we see slight changes in standard errors and p-values, as expected. We use Newey-West corrected standard errors for the remainder of the paper.

A.2.4 Contemporaneous correlation

Finally, we test whether contemporaneous correlation between the model regressors and the error term (endogeneity) exists by executing a set of 60 auxiliary regressions for each of the 5 independent variables (300 total tests). As described in Halcoussis (2005), in the single-equation Hausman tests we first regress each of the independent variables acting as a dependent variable against the four other independent variables. We then save the residuals and then let them serve as a new independent variables in each of the new OLS calculations. If the parameter estimate of the residual in the new OLS is significant, the implication is that the independent variable in the first stage is endogenous with the error term. For the 300 individual Hausman tests we find the parameter estimate for the residuals to indicate endogeneity in about 40% of the cases with little exhibited difference of incidences of endogeneity among the 5 independent variables. The results are summarized in Table A.6.

While the presence of endogeneity will not impact the goodness of fit of the model, OLS that exhibit endogeneity between the regressors and the error term will not be unbiased. The estimate of the dependent variable is unbiased; the estimates of the coefficients are biased. One way to address the issue of endogeneity is with instrumental variable (IV) and two-stage least squares (2SLS) techniques. We conduct 60 IV 2SLS regressions in which we instrument the mortgage rate independent variable with the annualized returns for NCREIF, both at the national level. NCREIF annualized returns which exhibit a correlation of -0.285 and also have an indirect relationship with the capitalization rate (dependent variable), but only through the mortgage rate. As there are many instances of endogeneity as previously discussed, for this test, we are primarily interested in determining post-estimation whether mortgage rates are indeed endogenous and, further, whether NCREIF annualized returns

Table A.5: Durbin-Watson autocorrelation test results

Numeric Code, j	Supralocations	rho	d_{Lower}	d_{Upper}	DW test result
1	Alaska	0.36363	1.52900	1.80000	rej Null>>Pos Autocorrel
2	Alabama	0.73223	1.52900	1.80000	rej Null>>Pos Autocorrel
3	Arkansas	0.66108	1.52900	1.80000	rej Null>>Pos Autocorrel
4	Arizona	0.48023	1.52900	1.80000	rej Null>>Pos Autocorrel
5	California	0.74964	1.52900	1.80000	rej Null>>Pos Autocorrel
6	Colorado	0.72204	1.52900	1.80000	rej Null>>Pos Autocorrel
7	Connecticut	0.49775	1.52900	1.80000	rej Null>>Pos Autocorrel
8	Washington, DC	0.33644	1.52900	1.80000	rej Null>>Pos Autocorrel
9	Delaware	0.70628	1.52900	1.80000	rej Null>>Pos Autocorrel
10	Florida	0.62385	1.52900	1.80000	rej Null>>Pos Autocorrel
11	Georgia	0.74710	1.52900	1.80000	rej Null>>Pos Autocorrel
12	Hawaii	0.36745	1.52900	1.80000	rej Null>>Pos Autocorrel
13	Iowa	0.35599	1.52900	1.80000	rej Null>>Pos Autocorrel
14	Idaho	0.72523	1.52900	1.80000	rej Null>>Pos Autocorrel
15	Illinois	0.60485	1.52900	1.80000	rej Null>>Pos Autocorrel
16	Indiana	0.36308	1.52900	1.80000	rej Null>>Pos Autocorrel
17	Kansas	0.44221	1.52900	1.80000	rej Null>>Pos Autocorrel
18	Kentucky	0.64260	1.52900	1.80000	rej Null>>Pos Autocorrel
19	Louisiana	0.11355	1.52900	1.80000	Inconclusive
20	Massachusetts	0.58220	1.52900	1.80000	rej Null>>Pos Autocorrel
21	Maryland	0.49304	1.52900	1.80000	rej Null>>Pos Autocorrel
22	Maine	0.60941	1.52900	1.80000	rej Null>>Pos Autocorrel
23	Michigan	0.47394	1.52900	1.80000	rej Null>>Pos Autocorrel
24	Minnesota	0.71279	1.52900	1.80000	rej Null>>Pos Autocorrel
25	Missouri	0.70942	1.52900	1.80000	rej Null>>Pos Autocorrel
26	Mississippi	0.69526	1.52900	1.80000	rej Null>>Pos Autocorrel
27	Montana	0.70899	1.52900	1.80000	rej Null>>Pos Autocorrel
28	North Carolina	0.61337	1.52900	1.80000	rej Null>>Pos Autocorrel
29	North Dakota	0.77661	1.52900	1.80000	rej Null>>Pos Autocorrel
30	Nebraska	0.57523	1.52900	1.80000	rej Null>>Pos Autocorrel
31	New Hampshire	0.53198	1.52900	1.80000	rej Null>>Pos Autocorrel
32	New Jersey	0.58603	1.52900	1.80000	rej Null>>Pos Autocorrel
33	New Mexico	0.20195	1.52900	1.80000	Inconclusive
34	Nevada	0.47961	1.52900	1.80000	rej Null>>Pos Autocorrel
35	New York	0.27552	1.52900	1.80000	rej Null>>Pos Autocorrel
36	Ohio	0.53937	1.52900	1.80000	rej Null>>Pos Autocorrel
37	Oklahoma	0.62375	1.52900	1.80000	rej Null>>Pos Autocorrel
38	Oregon	0.36684	1.52900	1.80000	rej Null>>Pos Autocorrel
39	Pennsylvania	0.57955	1.52900	1.80000	rej Null>>Pos Autocorrel
40	Rhode Island	0.79791	1.52900	1.80000	rej Null>>Pos Autocorrel
41	South Carolina	0.07694	1.52900	1.80000	No AutoCorrel
42	South Dakota	0.69944	1.52900	1.80000	rej Null>>Pos Autocorrel
43	Tennessee	0.61871	1.52900	1.80000	rej Null>>Pos Autocorrel
44	Texas	0.61422	1.52900	1.80000	rej Null>>Pos Autocorrel
45	Utah	0.46769	1.52900	1.80000	rej Null>>Pos Autocorrel
46	Virginia	0.66607	1.52900	1.80000	rej Null>>Pos Autocorrel
47	Vermont	0.72936	1.52900	1.80000	rej Null>>Pos Autocorrel
48	Washington	0.67701	1.52900	1.80000	rej Null>>Pos Autocorrel
49	Wisconsin	0.39005	1.52900	1.80000	rej Null>>Pos Autocorrel
50	West Virginia	0.78890	1.52900	1.80000	rej Null>>Pos Autocorrel
51	Wyoming	0.78521	1.52900	1.80000	rej Null>>Pos Autocorrel
52	US - Nation	0.64231	1.52900	1.80000	rej Null>>Pos Autocorrel
53	NCREIF_East North Central	0.51471	1.52900	1.80000	rej Null>>Pos Autocorrel
54	NCREIF_Mideast	0.69378	1.52900	1.80000	rej Null>>Pos Autocorrel
55	NCREIF_Northeast	0.50765	1.52900	1.80000	rej Null>>Pos Autocorrel
56	NCREIF_Southeast	0.63441	1.52900	1.80000	rej Null>>Pos Autocorrel
57	NCREIF_Southwest	0.59860	1.52900	1.80000	rej Null>>Pos Autocorrel
58	NCREIF_Midwest	0.69299	1.52900	1.80000	rej Null>>Pos Autocorrel
59	NCREIF_West North Central	0.71613	1.52900	1.80000	rej Null>>Pos Autocorrel
60	NCREIF_Western Pacific	0.73699	1.52900	1.80000	rej Null>>Pos Autocorrel

This table summarizes the results of the Durbin-Watson tests for first-order serial correlation which compares the DW test statistic to the two critical values $d_{Lower} = 1.529$, $d_{Upper} = 1.800$ at the 5% significance level reflecting 6 degrees of freedom and n=93 observations. ***/**/* indicate 1%, 5% and 10% significance levels.

Table A.6: Single-equation Hausman test results

Numeric Code, j	Supralocations	mtg endog	chgoff endog	credit endog	hpi endog	ue endog
1	Alaska	0.0210	0.0210	0.0210	0.0210	0.0210
2	Alabama	0.0820	0.0820	0.0820	0.0000	0.0820
3	Arkansas	0.1470	0.1470	0.0000	0.1470	0.0000
4	Arizona	0.7710	0.7710	0.7710	0.7710	0.7710
5	California	0.0330	0.0330	0.0330	0.0330	0.0330
6	Colorado	0.1290	0.1290	0.1290	0.1290	0.1290
7	Connecticut	0.0520	0.0520	0.0520	0.0520	0.0520
8	Washington, DC	0.5400	0.5400	0.5400	0.5400	0.5400
9	Delaware	0.0060	0.0060	0.0060	0.0000	0.0060
10	Florida	0.0000	0.0000	0.0000	0.0000	0.0000
11	Georgia	0.5110	0.5110	0.5110	0.0000	0.5110
12	Hawaii	0.2170	0.2170	0.0000	0.2170	0.2170
13	Iowa	0.2360	0.2360	0.2360	0.0000	0.2360
14	Idaho	0.0010	0.0010	0.0010	0.0010	0.0010
15	Illinois	0.0020	0.0020	0.0020	0.0020	0.0020
16	Indiana	0.9530	0.9530	0.9530	0.0000	0.9530
17	Kansas	0.7920	0.7920	0.7920	0.7920	0.0000
18	Kentucky	0.9660	0.9660	0.9660	0.0000	0.9660
19	Louisiana	0.1350	0.1350	0.1350	0.1350	0.1350
20	Massachusetts	0.6760	0.6760	0.6760	0.0000	0.6760
21	Maryland	0.0060	0.0060	0.0060	0.0060	0.0060
22	Maine	0.2540	0.2540	0.2540	0.2540	0.2540
23	Michigan	0.8450	0.8450	0.8450	0.0000	0.0000
24	Minnesota	0.3470	0.3470	0.3470	0.3470	0.3470
25	Missouri	0.4260	0.4260	0.4260	0.0000	0.4260
26	Mississippi	0.4040	0.4040	0.4040	0.4040	0.4040
27	Montana	0.0090	0.0090	0.0090	0.0000	0.0090
28	North Carolina	0.0760	0.0760	0.0760	0.0000	0.0760
29	North Dakota	0.5770	0.5770	0.5770	0.5770	0.5770
30	Nebraska	0.9380	0.9380	0.9380	0.0000	0.9380
31	New Hampshire	0.7910	0.7910	0.7910	0.0000	0.7910
32	New Jersey	0.7280	0.7280	0.7280	0.7280	0.7280
33	New Mexico	0.0080	0.0080	0.0080	0.0080	0.0080
34	Nevada	0.0030	0.0030	0.0030	0.0030	0.0030
35	New York	0.9750	0.9750	0.9750	0.0000	0.9750
36	Ohio	0.0910	0.0910	0.0000	0.0910	0.0910
37	Oklahoma	0.4480	0.4480	0.4480	0.0000	0.4480
38	Oregon	0.0000	0.0000	0.0000	0.0000	0.0000
39	Pennsylvania	0.0030	0.0030	0.0030	0.0000	0.0030
40	Rhode Island	0.0000	0.0000	0.0000	0.0000	0.0000
41	South Carolina	0.0310	0.0310	0.0310	0.0000	0.0310
42	South Dakota	0.1440	0.1440	0.1440	0.1440	0.1440
43	Tennessee	0.2570	0.2570	0.2570	0.2570	0.2570
44	Texas	0.1690	0.1690	0.1690	0.0000	0.1690
45	Utah	0.9670	0.9670	0.9670	0.0000	0.9670
46	Virginia	0.0000	0.0000	0.0000	0.0000	0.0000
47	Vermont	0.0130	0.0130	0.0130	0.0130	0.0130
48	Washington	0.0000	0.0610	0.0610	0.0610	0.0610
49	Wisconsin	0.1180	0.1180	0.1180	0.0000	0.1180
50	West Virginia	0.4670	0.4670	0.4670	0.4670	0.4670
51	Wyoming	0.1340	0.1340	0.1340	0.0000	0.1340
52	US - Nation	0.0000	0.0000	0.0000	0.0000	0.0000
53	NCREIF_East North Central	0.0010	0.0010	0.0010	0.0010	0.0010
54	NCREIF_Mideast	0.0000	0.0000	0.0000	0.0000	0.0000
55	NCREIF_Northeast	0.6130	0.6130	0.6130	0.0000	0.6130
56	NCREIF_Southeast	0.0000	0.0000	0.0000	0.0000	0.0000
57	NCREIF_Southwest	0.0000	0.0820	0.0820	0.0820	0.0820
58	NCREIF_Midwest	0.0010	0.0010	0.0010	0.0000	0.0010
59	NCREIF_West North Central	0.2440	0.2440	0.2440	0.0000	0.2440
60	NCREIF_Western Pacific	0.5300	0.5300	0.5300	0.5300	0.5300
	total endogenous	23	21	24	40	24
	total exogenous	37	39	36	20	36
	% endogenous	38.33%	35.00%	40.00%	66.67%	40.00%

This table summarizes the results of the single equation Hausman tests of endogeneity for $i = 1 - 60$ locations. The 5 columns to the right of the location correspond to each of the independent variables. The elements of each of the columns are the p-values which correspond to the residual parameter estimate which is included as independent variable in the second stage. The final 3 rows provide summary statistics for each of the columns where p-values<= .05 indicate endogeneity between the original regressor and the error term.

are a strong instrument for mortgage rates. Post-estimation analysis of the validity of the NCREIF returns as instrument reiterate the findings of endogeneity. In Table A.7, we summarize the post-estimation results. The first panel shows the p-values for Durbin and Wu-Hausman test statistics and indicate in 60% of the cases, the mortgage rate is endogenous and thus estimation would benefit from the IV 2SLS approach. The second panel shows the post-estimation results which test for weak instrumental values with the F-test. The F-stats indicate that NCREIF annualized returns are a strong instrument for mortgage rates at the national level. While it is always the case in a successful OLS model that we expect *some* linear relationship between the dependent variable and independent variable, selecting a proxy independent variable with these properties may be useful. Nevertheless, determining *valid* instruments for *all* variables is often a confounding exercise, if not impossible. We have addressed fully one such instrument which works quite well. These post-estimation findings regarding endogeneity do not require a change in the coefficients established in the initial OLS.

A.2.5 Cap rate delta check

It is worth reiterating that our study's focus is on projection of cap rate levels where such cap rate indices are not publicly observable. In this study the projection of cap rate changes (or deltas) on the left hand side of the OLS equation, serves no purpose since we not have a baseline actual cap rate from which an estimated delta could be applied to project independent cap rates at the MSA level. Nominal cap rate values matter considerably in the determination of property value estimates. Estimating a level of cap rates with changes in levels of explanatory variables obscures our objective in the projection technique. In contrast, in Section 6 of this study we must take normalized first differences. For example, we substitute $\Delta_{UE,t} = \frac{UE_t - UE_{t-1}}{UE_{t-1}}$ for UE_t to establish the stationarity of independent and dependent variables for use in VARs.

We test our intuition by modifying the OLS replacing each of the independent variables with their normalized one-period percentage change. The results confirm our intuition. They demonstrate that the use of periodic changes instead of levels results in consistently worse parameter estimates and goodness of fits than are achieved with the standard level variables.

Table A.7: Instrumental variable 2SLS post-estimation test results

Numeric Code, j	Supralocations	p-value	Wu-Hausman	p-value	F(1, 87)	Adj Rsq	Partial Rsq	F-stat
1	Alaska	0.4112	0.6290	0.4299	36.6107	0.4101	0.2962	0
2	Alabama	0.3830	0.7096	0.4019	45.215	0.5045	0.342	0
3	Arkansas	0.0000	18.5425	0.0000	44.7157	0.5296	0.3395	0
4	Arizona	0.0004	13.2395	0.0005	43.8842	0.5787	0.3353	0
5	California	0.0998	2.5801	0.1119	48.4883	0.5534	0.3579	0
6	Colorado	0.0048	8.0357	0.0057	33.3945	0.6307	0.2774	0
7	Connecticut	0.1428	2.0325	0.1576	51.6436	0.5865	0.3725	0
8	Washington, DC	0.1072	2.4690	0.1198	45.9914	0.5144	0.3458	0
9	Delaware	0.0700	3.1480	0.0796	45.2933	0.5243	0.3424	0
10	Florida	0.4480	0.5357	0.4662	52.718	0.5293	0.3773	0
11	Georgia	0.0002	14.9379	0.0002	40.2498	0.6441	0.3163	0
12	Hawaii	0.3250	0.9051	0.3441	36.8782	0.4616	0.2977	0
13	Iowa	0.0048	8.0381	0.0057	26.6303	0.5544	0.2344	0
14	Idaho	0.0000	20.1163	0.0000	44.0143	0.4669	0.336	0
15	Illinois	0.0976	2.6153	0.1095	51.9657	0.6204	0.3739	0
16	Indiana	0.2778	1.1033	0.2965	30.942	0.6053	0.2623	0
17	Kansas	0.0000	27.7141	0.0000	24.5742	0.5558	0.2203	0
18	Kentucky	0.0343	4.3498	0.0400	34.4911	0.5668	0.2839	0
19	Louisiana	0.2938	1.0315	0.3127	42.7636	0.442	0.3295	0
20	Massachusetts	0.0000	21.1856	0.0000	46.7757	0.4725	0.3497	0
21	Maryland	0.0896	2.7501	0.1009	51.1434	0.5211	0.3702	0
22	Maine	0.0324	4.4538	0.0377	48.2007	0.47	0.3565	0
23	Michigan	0.0056	7.7264	0.0067	32.7353	0.56	0.2734	0
24	Minnesota	0.1052	2.4985	0.1176	35.6736	0.534	0.2908	0
25	Missouri	0.0015	10.4336	0.0018	35.7721	0.556	0.2914	0
26	Mississippi	0.8011	0.0587	0.8091	39.1862	0.5524	0.3105	0
27	Montana	0.0001	18.3899	0.0000	39.244	0.4261	0.3109	0
28	North Carolina	0.0430	3.9611	0.0497	32.3811	0.6513	0.2712	0
29	North Dakota	0.0000	19.0129	0.0000	32.8497	0.4303	0.2741	0
30	Nebraska	0.0003	13.8261	0.0004	11.4162	0.6153	0.116	0.0011
31	New Hampshire	0.0018	10.0939	0.0021	47.5589	0.4554	0.3534	0
32	New Jersey	0.1307	2.1659	0.1448	55.7616	0.5588	0.3906	0
33	New Mexico	0.0074	7.1941	0.0088	43.4649	0.4468	0.3332	0
34	Nevada	0.2104	1.4755	0.2278	45.2657	0.6426	0.3422	0
35	New York	0.0004	13.6206	0.0004	55.5755	0.5149	0.3898	0
36	Ohio	0.0017	10.1289	0.0020	29.2916	0.5646	0.2519	0
37	Oklahoma	0.0939	2.6765	0.1055	42.5038	0.4486	0.3282	0
38	Oregon	0.0725	3.0888	0.0824	32.1904	0.5652	0.2701	0
39	Pennsylvania	0.1625	1.8425	0.1782	55.7257	0.534	0.3904	0
40	Rhode Island	0.0047	8.0876	0.0056	39.6458	0.4597	0.313	0
41	South Carolina	0.0249	4.9159	0.0292	59.1119	0.5728	0.4046	0
42	South Dakota	0.0001	16.4373	0.0001	25.258	0.5967	0.225	0
43	Tennessee	0.0537	3.5864	0.0616	25.6819	0.4962	0.2279	0
44	Texas	0.0028	9.1273	0.0033	37.603	0.5705	0.3018	0
45	Utah	0.0026	9.2625	0.0031	43.7345	0.4677	0.3345	0
46	Virginia	0.0164	5.6738	0.0194	33.7834	0.5052	0.2797	0
47	Vermont	0.0004	13.2065	0.0005	50.9958	0.4979	0.3695	0
48	Washington	0.1133	2.3825	0.1264	45.5195	0.4414	0.3435	0
49	Wisconsin	0.2854	1.0685	0.3042	42.503	0.5249	0.3282	0
50	West Virginia	0.0001	17.9949	0.0001	23.3894	0.6268	0.2119	0
51	Wyoming	0.0014	10.6253	0.0016	31.5881	0.4115	0.2664	0
52	US - Nation	0.0053	7.8465	0.0063	42.0405	0.4358	0.3258	0
53	NCREIF_East North Central	0.0049	8.0126	0.0058	58.7869	0.6499	0.406	0
54	NCREIF_Mideast	0.0209	5.2349	0.0246	54.753	0.6058	0.389	0
55	NCREIF_Northeast	0.0025	9.3540	0.0030	66.8686	0.5351	0.4374	0
56	NCREIF_Southeast	0.0616	3.3567	0.0704	65.0542	0.6397	0.4307	0
57	NCREIF_Southwest	0.0037	8.5842	0.0044	47.4706	0.4584	0.3557	0
58	NCREIF_Midwest	0.0014	10.6550	0.0016	70.9576	0.6388	0.4521	0
59	NCREIF_West North Central	0.0007	12.0556	0.0008	43.676	0.515	0.3368	0
60	NCREIF_Western Pacific	0.0498	3.7096	0.0574	51.4538	0.5291	0.3743	0

This table summarizes the post-estimation results of $i = 1 - 60$ instrument variable two-stage least squares regressions which instrument the mortgage rate independent variable with the annualized returns for NCREIF, both at the national level. The first panel shows the p-values for Durbin and Wu-Hausman test statistics and indicate in 60% of the cases, the mortgage rate is endogenous and thus estimation would benefit from the instrumental variable 2SLS approach. The second panel shows the post-estimation results which test for weak instrumental values with the F-test. The F-stats indicate that annualized NCREIF returns are a strong instrument for mortgage rates at the national level.

Of the 60 OLS using deltas of independent variables, only 3 were statistically significant as measured by the *F*-test. In terms of goodness of fit, the R-squared values ranged from a minimum of 0.008 to a maximum of 0.1942. Finally, 288 (96%) of the coefficients were statistically insignificant from zero at the 10% confidence level, with only 12 coefficients (4%) exhibiting statistical significance at or better than the 10% confidence level. These results confirm our intuition and are not used in the projection technique. These results are available upon request.

A.3 Discussion of additional methodological robustness check

Finally, one additional method to validate the model methodological approach is to implement the standard leave one out cross-validation (LOOCV) approach. For each supralocation $j \in [1, 60]$ we assume that a supralocation k is our target MSA, and that our pool of supralocations now consists of the remaining 59. Then, reproducing our methodology with this change, for each of $t \in [1, 95]$ consecutive quarters from Q4 1991 to Q2 2015 a matrix of dimensions 60 rows (corresponding to each of the artificially 'missing' $k \in [1, 60]$ supralocations of Nation, Region and State) and 59 columns (corresponding to the $j \in [1, 59]$ geographic locations of existing Nation, Region and State) is formed with the distances, $\text{dist}(k, j)_t$ representing the elements of the matrix $D_{k,j,t}$.There are 95 such k x j matrices, one corresponding to each quarterly date, t.

Table A.8: Mean square errors from LOOCV test

Numeric Code, $j \in [1, 60]$	Supralocations	Mean Square Error
40	Rhode Island	0.35%
60	NCREIF_Western Pacific	0.39%
57	NCREIF_Southwest	0.45%
30	Nebraska	0.47%
53	NCREIF_East North Central	0.47%
59	NCREIF_West North Central	0.49%
6	Colorado	0.51%
13	Iowa	0.52%
17	Kansas	0.55%
18	Kentucky	0.55%
56	NCREIF_Southeast	0.56%
27	Montana	0.57%
43	Tennessee	0.61%
16	Indiana	0.63%
31	New Hampshire	0.64%
29	North Dakota	0.65%
14	Idaho	0.66%
55	NCREIF_Northeast	0.68%
24	Minnesota	0.68%
2	Alabama	0.69%
35	New York	0.73%
1	Alaska	0.76%
58	NCREIF_Midwest	0.81%
9	Delaware	0.84%
46	Virginia	0.88%
48	Washington	0.89%
37	Oklahoma	0.91%
28	North Carolina	0.95%
47	Vermont	0.99%
50	West Virginia	1.14%
51	Wyoming	1.27%
23	Michigan	1.46%
19	Louisiana	1.54%
45	Utah	1.76%
8	Washington, DC	2.25%
12	Hawaii	2.45%
41	South Carolina	2.51%
39	Pennsylvania	3.30%
22	Maine	3.40%
20	Massachusetts	3.62%
52	US - nation	4.65%
15	Illinois	4.73%
10	Florida	5.13%
25	Missouri	6.09%
38	Oregon	7.24%
5	California	7.34%
7	Connecticut	7.72%
11	Georgia	8.25%
21	Maryland	8.67%
26	Mississippi	8.88%
3	Arkansas	11.67%
42	South Dakota	12.95%
4	Arizona	14.13%
34	Nevada	14.43%
33	New Mexico	19.06%
44	Texas	23.19%
54	NCREIF_Mideast	24.25%
49	Wisconsin	27.04%
36	Ohio	27.18%
32	New Jersey	31.65%
	MSE - All	5.30%

This table summarizes the mean square errors (MSE) computed using the Leave One Amount Cross-Validation (LOOCV) technique. The columns provide the the numerical code, the name of the j-th supralocation and the MSE from the LOOCV test, respectively. The average MSE was 5.30% as reported in the final row of the table.

For each row, k, the minimum distance, $d^*_{k,t} = \min_j [D_{k,j,t}]$, as in Eq. (3.9), is calculated giving 60 such distances, one for each supralocation substitute for MSA. We compare the 60 time series comparing for each time t, the actual cap rates at the supralocations, $C_{j,t}$, to

the 60 LOOCV projected cap rates, $\widehat{C}_{j,t}$ by calculating their mean square errors (MSEs) as

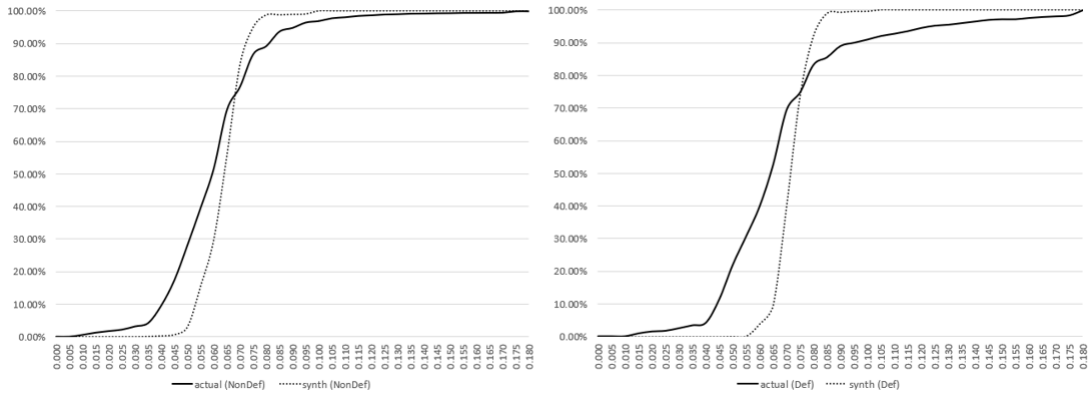
$$MSE = \frac{1}{T}\sum_{t=1}^{T=95}\left(C_{j,t} - \widehat{C}_{j,t}\right)^2 \tag{A.1}$$

The results are summarized in Table A.8 and are quite good. The overall MSE for all observations was 5.30% which compares well with the accuracy found in credit ratings literature as described in Kramer and Guttler (2008) and others. Of the 60 supralocations, 50 exhibited MSEs<9% with the remaining 10 exhibiting MSEs>11%. The MSE for the US (4.65%) was close to the overall MSE of 5.30%. Good results in NCREIF Northeast Region, NY and Rhode Island with MSEs of less than 0.70% were offset by extreme MSEs in NJ of more than 30%. Poor MSEs in TX and NCREIF Mideast Region were offset by excellent performance in other NCREIF Regional composites, NY, and CA. In conjunction, these results suggest that the SCX values are stable and fully utilizing the geographic breadth of the Nation through economic proximity.

A.4 Distributional analysis of SCXs and actual cap rates

Each of the properties in the private loan data are independent of properties included in the NCREIF portfolio and have an origination date, and corresponding property value, NOI, and cap rate information at origination. The cap rates associated with such properties are also independent of NCREIF cap rates. This allows us to conduct an out of sample comparative analysis between SCXs and values $\kappa_{\iota,t}$ representing the actual cap rate for the ι-th property with $\iota \in [1, 25101]$. For each property we find the SCX, $\tilde{\mathrm{C}}_{k,t}$, that corresponds to the MSA identifier of the property at the loan origination date. We split the sample between all properties and only those properties that defaulted on their loan obligation secured by the property. In Figure A.2a and A.2b we plot cumulative cap rate distributions for the non-defaulted and defaulted subsets. While actual cap rates and SCXs exhibit differentiating capabilities between safer and more risky assets, the assessment of such risks made with SCXs appears to be more conservative than were made by actual cap rates. SCXs are higher than actual cap rates and differences between defaulted and non-defaulted cumulative distributions are wider for SCXs than for actual cap rates.

Figure A.2: Cumulative cap rate distributions at origination



This figure depicts the cumulative distributions of actual cap rates at origination and corresponding synthetic MSA level cap rates at origination partitioned by the default indicator. The solid lines reflects the cumulative distribution for the actual cap rates at origination, while the dotted lines reflect the synthetic cap rates at origination that corresponded the MSA level in which the property is located. The plot on the left shows the comparison between actual and synthetic for the 24,089 loans that did not default. The plot on the right shows the comparison for the the 1,013 loans in our sample that defaulted over the CMBS loan sample period Q1 2000 to Q2 2015. The x-axis shows cap rates in increments of 0.005; the y-axis shows the proportion of loan counts; and the lines reflect the cumulative sum of the proportional count of loans as a percent of total loans in the category with corresponding cap rates at origination in the interval.

A.5 Logistic regression for Treatment 3

Table A.9: Logistic regression - Treatment 3

	Panel A: All				Panel B: CBD only				Panel C: non-CBD only			
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
_cons	-1.795667***	-4.039471***	-1.097695***	-3.498225***	-2.164676***	-4.154393***	-1.460711**	-3.67379***	-1.326086***	-3.561278***	-0.6271674**	-3.011455***
	(0.2235718)	(0.4660155)	(0.2716481)	(0.4817134)	(0.5249973)	(1.179183)	(0.6564918)	(1.229433)	(0.2332104)	(0.4994363)	(0.2873467)	(0.5161302)
caprate (actual)	0.9138623***		0.9155617***		5.006876***		5.081549***		0.7532925*		0.7503497*	
	(0.4140675)		(0.417609)		(1.934122)		(1.912539)		(0.4397526)		(0.4408342)	
caprate (SCX)		30.05074***		33.4522***		30.45716**		34.55862**		29.94889***		33.26209***
		(5.325379)		(5.439026)		(13.75627)		(14.31072)		(5.777469)		(5.883821)
occpct	-0.0131264***	-0.0133039***	-0.0131963***	-0.0134009***	-0.0126872***	-0.0133576***	-0.0126349***	-0.0132959***	-0.013111***	-0.0132988***	-0.0132014***	-0.0134228***
	(0.001555)	(0.0015575)	(0.0015576)	(0.0015605)	(0.0040816)	(0.0040248)	(0.00408)	(0.0040245)	(0.0016855)	(0.0016894)	(0.001689)	(0.0016935)
propage	-0.011782***	-0.0110729***	-0.0117838***	-0.0110209***	-0.0121205	-0.0124782*	-0.0120805	-0.0124056*	-0.0115546***	-0.010705***	-0.0115501***	-0.0106502***
	(0.0033626)	(0.0033447)	(0.003364)	(0.0033448)	(0.0074021)	(0.0073026)	(0.0074042)	(0.0072983)	(0.0037859)	(0.0037778)	(0.003787)	(0.0037777)
cbd	0.4340318***	0.4428436***	0.4347555***	0.444102***								
	(0.0922478)	(0.0922991)	(0.0922891)	(0.0923434)								
prp_fxd	-0.1468917	-0.1669184	-0.1539132	-0.1778631	-0.0392335	-0.0317022	-0.0507974	-0.0425155	-0.1816708	-0.2032251	-0.1873637	-0.2141602
	(0.1573038)	(0.1574485)	(0.1574304)	(0.1575736)	(0.3512744)	(0.3512873)	(0.3515278)	(0.3515253)	(0.1759596)	(0.1761186)	(0.1761099)	(0.176264)
lend_fxd	-0.7003789***	-0.4872252***	-0.8027001***	-0.5955145***	-0.8012297***	-0.534565***	-0.894481***	-0.6266869***	-0.687417***	-0.4796798***	-0.7917612***	-0.5903175***
	(0.0647368)	(0.0734939)	(0.068053)	(0.075003)	(0.1717053)	(0.197639)	(0.1784187)	(0.199032)	(0.0699762)	(0.0791817)	(0.0737219)	(0.0810013)
vix			-0.0434109***	-0.0493803***			-0.0449567*	-0.0503346*			-0.0433049***	-0.0491935***
			(0.0097565)	(0.0098353)			(0.0257139)	(0.0259446)			(0.0105545)	(0.0106327)
Pseudo R-sq	0.0096	0.0107	0.0105	0.0113	0.0079	0.0068	0.0084	0.0072	0.0089	0.0099	0.0098	0.0108
Prob chi2	0..000	0..000	0..000	0..000	0..000	0..000	0..000	0..000	0..000	0..000	0..000	0..000
Obs	25101	25101	25101	25101	5195	5195	5195	5195	19906	19906	19906	19906

This table shows the results of logistic regressions with 0,1 indicator of loan default for the sample. The independent variables are:actual cap rate in columns (i) msa synthetic cap rate (SCX) in columns (ii). Panel A captures the entire sample; Panel B the CBD properties; and Panel C the non-CBD properties. Estimates are provided along with standard errors in parentheses. ***/**/*/' indicate 0.1%, 1%, 5% and 10% significance for Wald z-values

Table A.10: Logistic regression, SCX at time of Origination - Treatment 1

	Panel A: All		Panel B: CBD only		Panel C: Non CBD only	
	(i)	(ii)	(i)	(ii)	(i)	(ii)
cons	-3.228507***	-6.28515***	-3.9056897***	-7.174764***	-3.128669***	-6.101395***
	(0.0441904)	(0.2816949)	(.1558658)	(.6582128)	(0.0491975)	(0.3140811)
caprate (actual)	.8331781**		4.894396***		0.5957658	
	(.4177213)		(1.807289)		(0.4770939)	
SCX		41.53196***		48.02963***		40.19381***
		(3.665853)		(8.469149)		(4.095152)
Adj-Pseudo R-sq	0.0002	0.0050	0.0015	0.0059	0.0001	0.0047
Prob chi2	0.0461	0.0000	0.0068	0.0000	0.2118	0.0000
Obs	25101	25101	5195	5195	19906	19906

This table shows the results of logistic regressions with 0,1 indicator of loan default for the sample. The independent variables are:actual cap rate in columns (i) msa synthetic cap rate (SCX) in columns (ii). Panel A captures the entire sample; Panel B the CBD properties; and Panel C the non-CBD properties. Estimates are provided along with standard errors in parentheses. ***/**/*/' indicate 0.1%, 1%, 5% and 10% significance for Wald z-values

A.6 Origination time Logistic regression check.

Finally, due to the nature of our methodology, SCXs are necessarily one year forecasts of the economy. This is supported by the VAR and IRF analyses of Section 5 as well as in the outpeformance of SCXs compared with actual cap rates as discussed in Section 6. However as a further check, we executed each of the three treatments in Section 6 and instead compared natural '1-year forecast SCXs' $\tilde{\mathrm{C}}$ to going-in actual cap rates $C_{t,i\in[1,25101]}$ where time t origination time corresponds to time t calculation time of SCXs. The results in this Section show similar, to somewhat improved results, as those provided in the main text.

Table A.11: Logistic regression, SCX at time of Origination - Treatment 2

	Panel A: All		Panel B: CBD only		Panel C: Non CBD only	
	(i)	(ii)	(i)	(ii)	(i)	(ii)
_cons	-2.135964***	-5.239663***	-2.51904***	-5.745108***	-1.642249***	-4.665585***
	(0.2221568)	(0.3624483)	(0.5220886)	(0.8046272)	(0.2321479)	(0.3943714)
caprate (actual)	0.6873697		4.242379***		0.5254299	
	(0.4414735)		(1.846048)		(0.5005424)	
SCX		42.00785***		48.75087***		40.45194***
		(3.76168)		(8.607216)		(4.17255)
occpct	-0.0129654***	-0.0134835***	-0.0128622***	-0.0143199***	-0.0129051***	-0.0133521***
	(0.0015615)	(0.0015591)	(0.0040888)	(0.004058)	(0.0016923)	(0.0016883)
propage	-0.0137843***	-0.0115128***	-0.0125176*	-0.0110648	-0.0139661***	-0.0116155***
	(0.0033768)	(0.0033513)	(0.007349)	(0.0072969)	(0.0037987)	(0.0037728)
cbd	0.4487141***	0.4464554***				
	(0.0920274)	(0.0922067)				
prp_fxd	-0.2044051	-0.1763371	-0.0843447	-0.1171699	-0.2427588	-0.1965155
	(0.1567765)	(0.1575886)	(0.3503293)	(0.3516687)	(0.1753219)	(0.176345)
Adj-Pseudo R-sq	0.0047	0.0094	0.0034	0.0080	0.0038	0.0083
Prob chi2	0.0000	0.0000	0.0003	0.0000	0.0000	0.0000
Obs	25101	25101	5195	5195	19906	19906

This table shows the results of logistic regressions with 0,1 indicator of loan default for the sample. The independent variables are:actual cap rate in columns (i) msa synthetic cap rate (SCX) in columns (ii). Panel A captures the entire sample; Panel B the CBD properties; and Panel C the non-CBD properties. Estimates are provided along with standard errors in parentheses. ***/**/*/' indicate 0.1%, 1%, 5% and 10% significance for Wald z-values

Table A.12: Logistic regression, SCX at time of Origination- Treatment 3

	Panel A: All				Panel B: CBD only				Panel C: non-CBD only			
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
_cons	-1.795667***	-4.244688***	-1.097695***	-3.694789***	-2.164676***	-4.807408***	-1.460711**	-4.243827***	-1.326086***	-3.664441***	-0.6271674**	-3.119542***
	(0.2235718)	(0.3902237)	(0.2716481)	(0.4038485)	(0.5249973)	(0.8812139)	(0.6564918)	(0.9181807)	(0.2332104)	(0.4246999)	(0.2873467)	(0.4395522)
caprate (actual)	0.9138623***		0.9155617***		5.006876***		5.081549***		0.7532925*		0.7503497*	
	(0.4140675)		(0.417609)		(1.934122)		(1.912539)		(0.4397526)		(0.4408342)	
caprate (SCX)		32.18799***		37.17372***		39.29561***		45.45055***		30.53571***		35.35839***
		(4.068254)		(4.225773)		(9.406807)		(9.941869)		(4.507228)		(4.669981)
occpct	-0.0131264***	-0.0134907***	-0.0131963***	-0.0135881***	-0.0126872***	-0.014073***	-0.0126349***	-0.0139964***	-0.013111***	-0.0134053***	-0.0132014***	-0.0135286***
	(0.001555)	(0.0015543)	(0.0015576)	(0.0015576)	(0.0040816)	(0.0040423)	(0.00408)	(0.0040429)	(0.0016855)	(0.0016838)	(0.001689)	(0.001688)
propage	-0.011782***	-0.0108282***	-0.0117838***	-0.0107469***	-0.0121205	-0.0112662	-0.0120805	-0.0109417	-0.0115546***	-0.0106737***	-0.0115501***	-0.010624***
	(0.0033626)	(0.0033499)	(0.003364)	(0.003352)	(0.0074021)	(0.0073448)	(0.0074042)	(0.0073336)	(0.0037859)	(0.0037719)	(0.003787)	(0.003774)
cbd	0.4340318***	0.4398174***	0.4347555***	0.4408713***								
	(0.0922478)	(0.0923284)	(0.0922891)	(0.0923795)								
prp_fxd	-0.1468917	-0.1371119	-0.1539132	-0.1488416	-0.0392335	-0.0701948	-0.0507974	-0.0952125	-0.1816708	-0.1593835	-0.1873637	-0.168075
	(0.1573038)	(0.1577904)	(0.1574304)	(0.1578952)	(0.3512744)	(0.3522653)	(0.3515278)	(0.3527562)	(0.1759596)	(0.1765577)	(0.1761099)	(0.1766565)
lend_fxd	-0.7003789***	-0.5147625***	-0.8027001***	-0.6506351***	-0.8012297***	-0.4979931***	-0.894481***	-0.6315574***	-0.687417***	-0.5178427***	-0.7917612***	-0.653256***
	(0.0647368)	(0.0684704)	(0.068053)	(0.0707694)	(0.1717053)	(0.1836873)	(0.1784187)	(0.1870089)	(0.0699762)	(0.0738501)	(0.0737219)	(0.0765036)
vix			-0.0434109***	-0.0575783***			-0.0449567*	-0.0652094**			-0.0433049***	-0.0563401***
			(0.0097565)	(0.010075)			(0.0257139)	(0.026953)			(0.0105545)	(0.0108709)
Pseudo R-sq	0.0096	0.0117	0.0105	0.0125	0.0079	0.0093	0.0084	0.0098	0.0089	0.0108	0.0098	0.0118
Prob chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Obs	25101	25101	25101	25101	5195	5195	5195	5195	19906	19906	19906	19906

This table shows the results of logistic regressions with 0,1 indicator of loan default for the sample. The independent variables are:actual cap rate in columns (i) msa synthetic cap rate (SCX) in columns (ii). Panel A captures the entire sample; Panel B the CBD properties; and Panel C the non-CBD properties. Estimates are provided along with standard errors in parentheses. ***/**/*/' indicate 0.1%, 1%, 5% and 10% significance for Wald z-values