X

x

Matt Larriva, CFA · **Peter Linneman, PhD.**

Abstract Purpose

To explain both national and market-level cap rate changes as simply an interaction between Gross Domestic Product and Consumer Price Index.

Design/Methodology/Approach

We use a binary logistic regression with binary independent variables, trained on a synthetic minority over-sampling set, to explain cap rate expansion and contraction at the national level and at 20 MSAs.

Results

Both the national model and the MSA models capture around 40% of all cap rate expansion periods with a robust confusion matrix.

Originality

This model contributes to the existing corpus by 1) establishing a statistically significant relationship between GDP, CPI, and cap rates which 2) holds explanatory power at the national and MSA levels, by 3) mapping the ground truth from scalar to binary.

Practical Implications

The accuracy of the model suggests a simpler and more robust explanation for understanding cap rates and navigating property markets in an environment of fluctuating interest rates.

Keywords Binary Logistic Regression · Cap Rates · US Real Estate Markets · Multifamily Cap Rate · United States · Apartment Cap Rate

Address(es) of author(s) should be given

1 Introduction

At 16 trillion dollars, the value of commercial real estate in the United States (NAREIT, 2019) represents half of a percent of the world's total wealth (Shorrocks et al., 2018). And at 65% home-ownership in the United States (The St. Louis Federal Reserve Bank, 2020), real estate represents a more invested-in asset class than equities–only 55% of Americans own stock (Saad, 2020). As such, the value of this asset class and the underlying determinants are of importance not only to owners and operators but also to the economy as a whole.

2 Literature Review

3 Data

For the national model there are three data series that are used: the Consumer Price Index, the Gross Domestic Product, and the nominal Apartment cap rate series.

The Gross Domestic Product is a quarterly, unadjusted (not de-seasoned) series from the St. Louis Fed's economic data page (FRED). It is important to use a nominal (versus real) series, as we are comparing this to the consumer price index, which is a close proxy to the interest rate series used to convert the GDP from nominal to real. That is to say, we would be double counting the impact of inflation.

The Consumer Price Index series is sourced to the Bureau of Labor Statistics, and it is also a non-seasonally adjusted series. In both cases, we did not seek to alter the series with seasonal adjustments because the raw data itself contains information which we do not wish to strip. Knowing if the economy grew faster than inflation, even if it was in 4Q, and even if it was due to holiday sales, is valuable information. Such a situation suggests a very different environment from its seasonally adjusted counterpart which may shows an economy in which 4Q growth is similar to say the prior quarter.

Finally, cap rate series at the national level are sourced to Green Street. Green Street defines its cap rate series as the next-twelve-months' NOI divided by the spot asset value.

MSA-level data follows the conventions of the national data. Non-seasonally adjusted series are selected both for MSA GDP and the MSA's CPI. The former is from FRED while the later is from the Bureau of Labor Statistics.

As national CPI is presented monthly, and national GDP is presented quarterly, we proceeded with quarterly data for the national model, pairing the CPI release most close to the GDP release.

More challenging was the MSA-level data wherein GDP is presented only annually and CPI is presented with varying frequencies depending on the region. Atlanta is published on even months. Boston is published on odd months, and New York is published monthly. Some are bimonthly. The Bureau of Labor statistics only offers CPI measures for 22 Core Based Statistical Areas. Two of these are "Urban Alaska" and "Urban Hawaii", but these have no equivalent GDP series, so they were discarded. The remaining 20 CBSAs have GDP series. But the GDP series are not necessarily for the exact same areas i.e, some GDPs are published for an MSA, some are published for a CBSA. Adding to this confusion: the BLS revised its reporting in 2018 and started offering CBSA measurements instead of MSA measurements, thus revising the coverage areas.

We attempt to reconcile this by matching MSA to CBSA where needed (between the BLS and Fed). And we average the measurements of CPI over a year before matching them to the GDP metrics. In this way,

we end up with 21 geographies: 20 CBSAs and 1 national geography. Each geography has a cap rate, CPI, and GDP.

4 Model Selection, Specification, Analysis

4.1 Model Selection

While most analyses forecasting cap rates allow the random variable to be generated from a normal distribution, we opted instead to represent the variable as generated from a Bernoulli distribution. While this choice might be unorthodox–modeling a continuous variable like percentage as a binary variable– we argue that it is well suited for two reasons.

First, it is practical. Real estate is not an equity, it is an alternative asset. As such, it is illiquid. This lack of liquidity means an investor cannot enter and exit the market readily. As such, a would-be buyer is somewhat ambivalent between caring if a cap rate will go up by 40 basis points next year and if it will simply go up. In either case, next year's prices will offer a better value. Similarly a would-be seller should be ambivalent between knowing if next year a cap rate will go up by 10 basis points or if it will go up by 100. In either situation, next year is a worse time to sell than this year. Contrast this with an equities trader who can easily place stop-losses or 'level-in' to a position. Such a trader cares very much whether his holdings will increase 10% next year, hence estimating liquid asset pricing is well suited by a continuous variable. But estimating illiquid asset pricing is, we argue, better done with a discrete variable.

Second, it maps to the binary nature of the motivation of the model: will GDP growth (contraction) exceed CPI growth (contraction)? It matters less the exact amount one exceede or does not exceede the other. We argue that the question is simply: can a landlord pass on rising expenses to tenants or not? In the case where GDP is rising faster than inflation, we argue the tenants will be wealthier than the growth in expenses, and thus able to absorb higher rents, creating an asset worth more. Conversely, when inflation outpaces the growth of the economy, we argue the tenant is not in a position to absorb higher costs, leaving the landlord to bear them, creating an asset worth less.

Thus a logistic regression with a single binary input lent itself well to our analysis, with cap rate expansion as the response variable, and GDP change ¿ CPI change as the independent variable.

4.2 Model Specification

We preprocessed the GDP and CPI series differently for the national versus MSA model because the national data was more frequent. In both cases, the preprocessing included differencing the series and then creating a simple trailing average. The goal in smoothing the series was to prevent an overly sensitive signal if, say, one period had inflation that quickly reversed.

CPI and GDP percentage changes at t_0 were compared to trailing average historical percentage changes. A signal was generated when 1) the t_0 GDP change was less than its trailing average changes and 2) the t_0 CPI change was greater than its trailing averages. The signal suggests cap rate expansion in the next period.

The cap rate series were simply differenced and their sign taken to be a boolean variable such that cap rate expansion is TRUE and cap rate non-expansion (and contraction) is FALSE.

To establish the validity of the model for predictive power, we trained five separate models (Year = 2015, 2016....2020) which were given historical data and then asked to predict the next years' (Years 2016-2021, 2017-2021,...2021) cap rate expansions and contractions based on the signal series outlined above. This was done once for the national model and once for the MSA model. All MSA models are trained and specified the same. The model was not given the MSA name as a dummy variable.

The ground truth of cap rate expansion does not present a balanced set, as cap rate compressions are far more frequent than cap rate expansions (roughly two to three times more years see compressions than expansions). We overcame the imbalance through a Synthetic Minority Over-sampling Technique as created by (?). This balanced the frequency of cap rate expansions and contractions to prevent a the logistic regression from estimating all compressions (which would produce a low-variance high-bias estimator).

4.3 Model Goodness of Fit

In both the National and the MSA models, the logistic regressions displayed consistent coefficients, inline with our hypothesis. The coefficients had low standard errors relative to the coefficients' magnitude, and the 2.5% to 97.5% bounds did not include zero in any case.

Where a year is listed in the Slit-Year column, the data was segmented into training and test sets. The training periods were all prior to the split year, and the test set was the split year through 2021. When "all-time" is listed under the Split-Year, the model was trained on all data and evaluated on all avalailable data.

The columns, "Coef.", "Std.Err.", "Z", "$P > |z|$", and the 2.5% to 97.5% bounds are all populated with the parameters from the trained models using the training data. The columns to the right, "accuracy" through "specificity" are based on the test data.

National models were created every two years from 2000, as data is both more frequent and longer lived than MSA-level data. MSA models were created every year from 2015. It is worth noting that the coefficient decreased consistently as time advances, suggesting that perhaps the signal is not as robust as it once was. However the standard error decreases correspondingly, producing a Z score which is as high as the earlier periods. Still, though the coefficient is decreasing more recently, all of the train/test models had higher coefficients for lower standard errors than did the models trained on all data. This could be due to the difficulty of trying to incorporate the GFC and the unorthodox interest rate behavior of the COVID-19 era.

4.3.1 National Model

Year	Coef.	Std.Err.	Z	$P > |z|$	[0.025	0.975]
2000	2.2	1.05	2.08	0.0371	0.13	4.26
2004	1.7	0.77	2.22	0.0266	0.2	3.21
2008	1.39	0.56	2.48	0.0131	0.29	2.48
2012	1.3	0.46	2.82	0.0048	0.4	2.2
2016	1.35	0.42	3.18	0.0015	0.52	2.18
all-time	1.1	0.37	3.01	0.0026	0.38	1.81

4.3.2 MSA Model

Year	Coef.	Std.Err.	Z	$P > |z|$	[0.025	0.975]
2015	1.1	0.26	4.15	0.0000	0.58	1.62
2016	1.15	0.26	4.37	0.0000	0.63	1.67
2017	1.07	0.25	4.21	0.0000	0.57	1.56
2018	0.8	0.22	3.7	0.0002	0.38	1.22
2019	0.85	0.21	4.07	0.0000	0.44	1.26
all-time	0.58	0.18	3.14	0.0017	0.22	0.94

5 Results

5.0.1 National Model

Year	accuracy	precision	recall	F-score	specificity
2000	0.76	0.55	0.5	0.52	0.85
2004	0.75	0.53	0.5	0.51	0.84
2008	0.77	0.54	0.54	0.54	0.85
2012	0.78	0.56	0.56	0.56	0.85
2016	0.75	0.5	0.6	0.55	0.8
all-time	0.75	0.5	0.6	0.55	0.8

5.0.2 MSA Model

Year	accuracy	precision	recall	F-score	specificity
2015	0.65	0.16	0.25	0.19	0.73
2016	0.63	0.16	0.33	0.21	0.68
2017	0.61	0.11	0.33	0.16	0.65
2018	0.67	0.06	0.17	0.09	0.72
2019	0.62	0.07	0.33	0.12	0.65
all-time	0.62	0.07	0.33	0.12	0.65

5.1 Analysis of Results

6 Conclusion

Declarations

6.1 Funding

This research is a part of one author's role as VP of Research and Data Analytics at a Real Estate Private Equity firm.

6.2 Conflict of interest

One author works for a Real Estate Private Equity firm which has ownership interest in many office and multi-family assets throughout the US.

6.3 Availability of data and material

Data available upon request.

6.4 Code availability

Code available upon request.

6.5 Authors' contributions

Each of the authors confirms that this manuscript has not been previously published and is not currently under consideration by any other journal.

References

NAREIT (2019). Estimating the size of the commercial real estate market. *Nareit Research*.
Saad, L. (2020). What percentage of americans owns stock?
Shorrocks, Davies, and Lluberas (2018). Global wealth report 2018. ".
The St. Louis Federal Reserve Bank (2020). Real gross domestic product.