SENTIMENTS AND SEMANTICS: EVALUATING THE BEHAVIOURAL ASPECTS OF REAL ESTATE USING CONTENT ANALYSIS

KIMBERLY WINSON-GEIDEMAN

University of Melbourne

ABSTRACT

Drilling into data to understand real estate as a financial investment is one of the primary considerations of property research. This paper is designed to shed light on some of the newer and more innovative studies by highlighting the use qualitative data and content analysis to enhance our understanding of real estate. Particular attention is placed on Latent Semantic Analysis, sentiment analysis and ocular tracking as stand alone methods as well as supplements to conventional pricing models.

Keywords: content analysis, semantics, sentiment analysis, Latent Semantic Analysis, ocular tracking, big data

INTRODUCTION

Quantitative analysis has been the backbone of real estate research since the publication of the first journal in 1973. This is not surprising given that real estate is, by its very nature, a subject dominated by financial theories focused on investigating the investment potential of properties. The capacity for property to improve the financial position of the owner-occupier or investor has led to a plethora of studies on everything from valuation to REITs to land contamination to real estate cycles. Most of these studies incorporate, at some level, the fundamental questions of pricing, value, and related behaviours such as when to buy, when to sell, when to increase (decrease) rents and to what level. Even studies focused on more qualitative measures are designed to investigate these relationships.

Advances in technology have provided ample opportunity to explore real estate from less quantitative perspectives, although most of the research is still firmly rooted in understanding real estate as a financial investment. Data types, storage, and access have changed markedly since the mid 1990s when advances in computing power, coupled with the development of new data sources, helped introduce 'big data' to the world. Large amounts of data that are being constantly created in both structured and unstructured forms challenges conventional modes of research, triggering the need for new methods of analysis. Recent studies have focussed on leveraging these new types of data and innovative methods to see how and what they add to the growing body of research in real estate.

This paper investigates some of the more innovative methods that have been used, focusing primarily on content analysis and exploring its potential in the era of big data. Recent improvements in the collection and evaluation of data coupled with the development of new sources have opened exciting avenues of research related to various forms of discourse including product descriptions, verbal interactions, formal documents, social media, and websites. Even though the real estate discipline is sometimes late to embrace new technologies, researchers that capitalise on these new sources and methods expand the body of knowledge by finding novel ways to answer interesting (and sometimes old) questions. We investigate some of these innovations that have been applied in the commercial and residential sectors, present them here, and elaborate on the potential for future research in real estate using content analysis and qualitative sources of data.

BACKGROUND

Inductive content analysis, the most basic variation of content analysis, has a long history in many subject areas, not excluding the real estate discipline. While it is not commonly used in the evaluation of the behavioural aspects of property, those studies that do (aside from literature reviews) are directed toward identifying the value-adding components that motivate investors.

In inductive content analysis, researchers use their expertise and knowledge of the field to analyse the content of textual documents to derive commonalities and develop an understanding of the relationships

within texts. This type of content analysis (after Miles and Huberman, 1994) relies on inductive reasoning, through the repeated review and classification of raw data. The data are organised through open coding, i.e. taking notes and adding headings as text is read, and is grouped based on similar themes. Repeated regrouping results in larger groups representing broader themes as the topics become more generalised.

Examples of the use of inductive content analysis in the real estate literature are primarily focused on the evaluation of open-ended surveys and literature reviews. Lindholm, Gibler and Levainen (2006), for instance, use inductive content analysis to extract themes from open-ended surveys in their study of the value-adding attributes of real estate to corporate wealth. The researchers were able to derive eight themes from 39 respondents, themes that include items such as decrease costs, increase productivity, and increase value of assets. When the themes were generalised further, two main factors emerged as relevant to increasing shareholder value, revenue growth and profitability.

Additional studies using variations of inductive content analysis include several literature surveys directed at topics in real estate literature. Johnson, Roulac, and Followill (1996) examined the content of papers presented during the first ten years of the American Real estate Society annual meetings, identifying the 16 most frequently researched and presented topics. Jud (1996) reviewed nine years of published research in the *Journal of Real Estate Research* and noted ten primary subject areas including investment, appraisal, corporate real estate, housing values, brokerage, REITs, housing markets, regional, environment, and mortgages. Dombrow and Turnbull (2004) extracted nine topics—appraisal, brokerage, housing, institutions, investments, mortgages, nonresidential, public policy, and other—after reviewing all the papers published in the *Journal of Real Estate Finance and Economics* and *Real Estate Economics* from 1988 to 2001, and Harrison and Manning (2008) used the *Journal of Real Estate Literature* classification system to identify the real estate-related topics published in 119 academic journals.

These early attempts at content analysis are characterised by either the relatively small number of observations that were analysed, the extended timeframe that was required to complete studies with larger volumes of data, and/or the lack of a reliably systematic method to analyse information. These and other issues are characteristic of inductive content analysis for a variety of reasons. Firstly, it is time consuming to read and re-read a large body of text, and what seemed reasonable, sensible and/or intuitive in one circumstance may not at a later time. While systematic to a degree, inductive content analysis relies on the intuition and biases of individuals with cognitive limitations that prevent the reliable interpretation of too many documents at once. Human cognition also lacks the depth of analysis computer-aided methods provide when evaluating large bodies of data.

ADVANCES IN CONTENT ANALYSIS

Content analysis has benefited greatly from the progression of technology as computer-aided methods have (almost) rendered moot some of the main deficiencies of less sophisticated methods. While the theory behind content analysis remains the same, the analytical methods have improved to such a degree that much larger amounts of data can be assessed, in much greater detail, in a fraction of the time. Content analysis is still primarily focused on semantics, but now covers much broader areas including mood or sentiment analysis as well as the examination of the effectiveness of unstructured data to convey meaning. Here we look at a number of different ways researchers have implemented semantic analysis, sentiment analysis and ocular tracking to analyse content in large bodies of text, social media and real estate websites.

Semantic Analysis

In many ways real estate academics have only scratched the proverbial surface of semantic analysis, research that analyses the meaning behind words, phrases, and/or texts. Modern semantic analysis moves beyond cognitive limitations by applying mathematical equations to extract the concepts embedded in text. A form of machine learning, computers are used to recognise patterns within documents to derive meaningful conclusions regarding the content of the documents.

A series of papers by Winson-Geideman and Evangelopoulos (2013a, 2013b, 2013c) provide an excellent example of the potential for semantic analysis in property research. Building on previous literature surveys, the researchers use Latent Semantic Analysis (LSA), a computer-aided method, to analyse abstracts from over 2500 articles printed it the *Journal of Real Estate Research*, *Real Estate Economics*, and the *Journal of Real Estate Finance and Economics* from 1973-2010, and then use the results to derive a set of topics that

best describe the discipline. After extracting as many as 50 themes or topics, the authors eventually settled on 25 as the number that best represented the overall content of the abstracts. Examples of the topics include Research Methods, REITs, Brokerage & Agency, Default & Foreclosure, Mortgage Termination and Homeownership, among others.

LSA and similar methods fill the void left by inductive content analysis as they systematically, quickly and intuitively categorise data based on the content of the text. LSA is similar to factor analysis in that it statistically analyses word usage patterns to produce a matrices of terms and documents that load on a particular factor or theme. As with any method, however, LSA does have limitations. Of the 2500 abstracts evaluated, approximately 20% failed to identify with the 25 themes that were extracted, leaving a relatively large number of uncategorised documents. This raises the question as to why the abstracts were not categorised, a question that may be best answered by reviewing each abstract individually, without the use of computer-aided methods.

While it is expected that academic literature convey information as clearly as possible, when carefully considered, linguistics can be used to obfuscate information relative to the intent of the deliverer. To thoroughly understand content, one must not only be cognizant about what is conveyed but also about how it is conveyed. Transparency and disclosure all play an important part in knowledge and understanding, and rules and regulations have been established to ensure the unambiguous delivery of information to protect consumers and investors. For example, in 1998 the Securities and Exchange Commission in the United States began advocating for publicly traded companies to use "plain English" in their financial disclosures, specifically in annual reports, under the dual assumptions that firms could hide unpleasant information in their reports and that information asymmetries between investors and management are partially attributable to complex financial language.

Additional work in semantics evaluates the issue of how meaning is conveyed through language, particularly how readability and disclosure relate to REITs. A 2012 study by Dempsey, Harrison, Luchtenberg and Seiler examined the readability of REIT annual reports using the Flesch-Kincaid Grade Level metric to determine if the language embedded in the annual reports was related to firm performance. FKGL reports readability on a scale comparable to grade levels, e.g. a score of 10 is approximately the equivalent of what an average 10th grade student is able to understand. The results of the FKGL were then fed into a regression to determine if the opacity of the annual reports was related to excess returns. Findings show that transparency is closely linked with REIT returns, where the reports of poorly performing firms tend to be less transparent (i.e. written at a higher grade level) and the reports of well performing firms are more easily understood.

In a similar vein of research, Danielsen, Harrison, Van Ness and Warr (2014) assessed transparency and disclosure on financial market liquidity by analysing the linguistics of REIT annual reports. To assess readability the researchers used the Flesch Reading Ease Index, which incorporates measures of the average sentence length as well as the average number of syllables in words. Because of the anticipated complexity associated with the annual reports and indecision regarding whether that complexity was related to greater opacity (negative) or greater degrees of disclosure and therefore transparency (positive), the number of words in the document was added as an additional linguistic variable. Results show that as disclosure/transparency increases, so does liquidity as measured by lower bid-ask spreads. Firms with easier to read annual reports show larger bid-ask spreads, implying fewer disclosures, less information and greater opacity.

Market Sentiment

Market sentiment studies analyse different sources of information to assess the prevailing attitude or mood of investors toward a given market, making qualitative judgments that are used to predict directionality. Not to be confused with event studies that document the effect of a single incident, sentiment studies take a much more holistic view to develop an assessment of the general tone of the market. While many studies document the influence of individual types of information on behaviour, market sentiment studies are unique in that they look at a variety of information sources to assess the overall impact. News analytics (focused on news cycles), semantic analysis and other techniques are often used in the evaluation.

One of the most recognised and highly cited market sentiment studies involves the use of Twitter feeds to predict movements in the Dow Jones Industrial Average. Bollen, Mao and Zheng (2011) used the mood-tracking tools, OpinionFinder and Google-Profile of Mood States, to evaluate two general sentiments (positive and negative) and six different moods including calm, alert, sure, vital, kind and happy. Using a

Granger causality analysis and Self-Organising Fuzzy Neural Network, the researchers were able to predict, with 87.6% accuracy, the direction of the changes in the closing value of the DJIA within three to four days, specifically related to the mood 'calm'. Neither of the broad sentiment categories, i.e. 'positive' and 'negative' provided any explanatory result nor did any of the other moods.

Examples of sentiment studies in real estate are somewhat rare due to the nature of the asset. Fixed asset prices (i.e. real estate) tend to be sticky and resistant to short-term change. Conversely, moods such as those studied by Bollen et al can change relatively quickly. Furthermore, real estate suffers from much greater illiquidity than equities. Unless moods or general sentiments are sustained for a reasonable amount of time, it is difficult to determine how real estate prices will be affected.

That said, REITs provide an excellent opportunity to evaluate market sentiment in much the same manner as share markets. REIT prices have the potential to fluctuate quickly and sometimes dramatically, as do equities. They can be very sensitive to information and are liquid to a much great degree than fixed assets. While many older studies have purposely resisted making qualitative judgments and intuitively relied on price and volume movements to be indicative of market sentiment (if prices increase then investors must be feeling good), more recent studies look at capturing market sentiment and using it to predict price.[1] Research by Doran, Peterson, and Price (2012) assesses the impact of conference call dialogue on REIT prices. The authors focus on the overall content of the calls with particular emphasis on the question and answer portion, theorising that it is an insightful source of information regarding transactions that is not typically captured in quarterly earnings. The tone of the conversations between management and analysts are also analysed and relative measures of positive, negative, active, passive, strong, weak, overstated, and understated are applied. Results show that not only does the tone of the calls play an important part in generating abnormal returns for the REITs, a positive tone almost completely offsets any negative earning surprise. What this appears to reveal is that positive delivery in conference calls has the potential to negate the fundamentals of the security apparent in share price, at least in the short term.

Ocular Tracking and Unstructured Data

One of the more challenging issues facing researchers involves the proliferation of unstructured data. Unstructured data is data that fails to conform to a predetermined model. While most languages tend to be well defined and suitable for pattern recognition in semantic analysis, unstructured data not only contains text, but may also contain numbers and symbols (website URLs), photos, GIFs, and videos. This type of data is used regularly in real estate on websites such as Domain, RealEstate.com.au, Realtor.com, and Zillow.

While variety and inexactness make extracting the meaning embedded in non-conforming data challenging, the questions associated with it are potentially less so. When one considers that the primary purpose for unstructured data such as photos and videos on real estate websites is to garner interest in a real estate that effectuates a sale or lease transaction, then analysing the effectiveness of that data becomes much more straightforward, as exemplified in Seiler, Madhavan, and Liechty (2012). Using ocular tracking technology, the researchers assess the opinions of homebuyers looking for properties on the Internet, focussing on the length of time spent looking at a photo of a real estate, a list of the real estate features and the agent's description of the real estate. Using a sample of 20 recent homebuyers and a control of 25 university students, the research shows that homebuyers spend the majority of their time looking at the photo of the real estate followed by the quantitative description. Agent comments receive substantially less attention. Furthermore, the first photo is viewed the longest, providing clues as to how an agent or marketer can optimally portray a home online to capture the interest of potential buyers.

---

[1] Downs, Guner, Hartzell and Torres (2001) studied the impact of market commentary on REIT prices. While not a content analysis study per se, the researchers were interested in unexpected price and volume movements on days when a well-respected Barron's column, The Ground Floor, released REIT-specific information. The objective of the paper was to determine if the commentary played a significant enough role in market perceptions that it influenced investment behaviour. The interesting component relative to this paper is that the authors specifically sought to remove qualitative assumptions about the directionality of the effect, assuming that any movements would result in a rebalancing of portfolios. This is in contrast to recent studies on sentiment analysis that make those qualitative assumptions in an attempt to predict the direction of market movements.

THOUGHTS ON THE USE OF CONTENT ANALYSIS IN REAL ESTATE RESEARCH

The introduction of new forms of data and research methods strengthens our ability to provide greater insight into the problems real estate researchers have addressed over the years. Innovation has changed some of the what and how, but the underlying motivation of exploring the investment potential of real estate remains the same. Regardless of the source or method, researchers are subject to some of the same constraints and subjective decision-making they have been in the past and are still beset with issues of bias and objectivity. Searching for more meaningful and explanatory results when new methods and types of data are applied to old problems may change the way we think about real estate. On the other hand, it may serve to reinforce long-standing theories and ideas.

In general, when we think about how to fit new data and methods to the questions we have in real estate, we need to firstly consider what value, if any, it will add to the research. Data should not be used just because they are there; they should be used because they are the best available data to evaluate the issue at hand.

Similar considerations need to be made when selecting a research methodology. Granted, content analysis is in no way a paradigm shift in real estate research; it is nothing more than tool in an already rich supply of methodologies. What is does provide, however, is a different way of addressing real estate questions, which is particularly important in an era of data-driven research. In some cases it stands on its own as a methodology (Winson-Geideman and Evangelopoulus, 2013a, 2013b, and 2013c), in others it supplements more conventional approaches such as REIT pricing models (Dempsey et al, 2012; Danielson et al, 2014). If evaluating semantics will provide a more holistic understanding of an issue, then the researcher should search out data and analyse them. In some cases the data may be readily available (e.g. REIT annual reports, literature abstracts). In other cases data may be collected over a period of time (e.g. Twitter feeds) or initiated and developed by the researcher (e.g. open-ended survey questions).

The ocular tracking study by Seiler et al (2012) exemplifies a possible way in which the data extracted through content analysis can be combined with more traditional forms of data to enhance our understanding of website marketing. The researchers note that potential homebuyers view the first photo the longest followed by the quantitative description and then the agent's comments. Additional research could evaluate the semantics of the agent's comments to determine if any reoccurring themes translate into longer viewing times and/or willingness to view additional photos/information. Assuming that the underlying motivation of posting a photo and description of a real estate on a real estate website is to facilitate a real estate transaction, data regarding themes and viewing time can be fed into a traditional regression model to measure effects on foot traffic (physical viewing), rental rates, time on market and even price.

While Twitter sentiment may be difficult to apply to fixed-asset pricing, it may track with other economic variables such as interest rates and rate expectations, and thus be used to enhance our understanding of mortgage choice. Tweets can be geocoded, evaluated and the results combined with more conventional data sources to track and possibly predict speculative acquisitions within markets and sub-markets. Additional research opportunities lie in other sources of non-traditional, qualitative data such as real estate descriptions in Ebay, news sources, and company reports.

REFERENCES

Bollen, J, H Mao, and X J Zeng, 2011. Twitter mood predicts the stock market, *Journal of Computational Science*, 2, 1, 1-8.

Danielsen, B R, D M Harrison, R A Van Ness and R S Warr, 2014. Liquidity, Accounting Transparency, and the Cost of Capital: Evidence from Real Estate Investment Trusts, *Journal of Real Estate Research*, 36, 2, 221-251.

Dempsey, S J, D M Harrison, K F Luchtenberg, M J Seiler, 2012. Financial Opacity and Firm Performance: The Readability of REIT Annual Reports, *Journal of Real Estate Finance and Economics*, 45, 450-470.

Dombrow, J and G Turnbull, 2004. Trends in Real Estate Research, 1988-2001: What's Hot and What's Not, *Journal of Real Estate Finance and Economics*, 29, 1, 47-70.

Doran, J S, D R Peterson, S M Price, 2012. Earnings Conference Call Content and Stock Price: The Case of REITs, *Journal of Real Estate Finance and Economics*, 45, 402-434.

Downs, D H, N Z Guner, D J Hartzell, M A Torres, 2001. Why do REIT prices change? The information content of Barrons "The Ground Floor". *Journal of Real Estate Finance and Economics*, 45, 63-80.

Harrison, D M and C Manning, 2008. Characteristics of Recent Real Estate Research: 2000-2006, *Journal of Real Estate Practice and Education*, 11, 2, 109-26.

Johnson, L L, S E Roulac, and R A Followill, 1996. The First Decade (1985-94). *Journal of Real Estate Research*, 12, 2, 195-207.

Jud, G D, 1996. The Journal of Real Estate Research: A Ten-Year Review, *Journal of Real Estate Research*, 12, 2, 249-313.

Lindholm, A L, K M Gibler and K I Levainen, 2006. Modeling the Value-Adding Attributes of Real Estate to the Wealth Maximization of the Firm, *Journal of Real Estate Research*, 28, 4, 445-475.

Miles, MC and AM Huberman, Qualitative Data Analysis: An Expanded Source Book, Thousand Oaks, CA: Sage Publications, 1994.

Seiler, M J, P Madhavan and M Liechty, 2012. Toward and Understanding of Real Estate Homebuyer Internet Search Behavior: An Application of Ocular Tracking Technology, *Journal of Real Estate Research*, 34, 2, 211-241.

Winson-Geideman, K and N Evangelopoulus, 2013a. Topics in Real Estate Research 1973-2010: A Latent Semantic Analysis, *Journal of Real Estate Literature*, 21, 1, 59-80.

Winson-Geideman, K and N Evangelopoulus, 2013b. Research in Real Estate 1973-2010: A Three Journal Comparison, *Journal of Real Estate Literature*, 21, 2, 255-267.

Winson-Geideman, K and N Evangelopoulus, 2013c. Reading Lists for PhD Seminars in Real Estate, *Journal of Real Estate Practice and Education*, 16, 1, 41-82.

Email contact: kimberly.winson@unimelb.edu.au