EAI Endorsed Transactions
on Scalable Information Systems

Research Article EAI.EU

New Behavioural Big Data Methods for Predicting Housing Price

Jiaying Kou[1,*],Yashar Gedik[1]

[1]Institute for Sustainable Industries & Liveable Cities, Victoria University, Melbourne, Australia

Abstract

Housing market price prediction is a big challenge. The 2008 global recession strongly showed that even the most sophisticated traditional economic models failed to foresee the crisis. New developments of behavioural economic theory indicate that the information from micro-level's decision making will bring new solution to the age-old problem of economic forecasting. Additionally, the information revolution and big data methods have provided a new lens to study economic problems apart from traditional methodologies.

This research provides the theoretical link between irrationality and big data methods. Empirically, big data methods will be used in forecasting the housing market cycle in Australia. Specifically, Google trends is included as a new variable in a time series auto-regression model to forecast housing market cycles.

Received on 07 January 2019; accepted on 15 February 2019; published on 16 April 2019

Keywords: information systems

doi:10.4108/eai.13-7-2018.158418

1. Introduction

Housing market prediction is one of the biggest challenges in the economic cycle research. The global financial and economic crisis (GFEC) that happened in 2008 is strong evidence to show that most economists failed to predict this real estate cycle. US $13 trillion of property wealth was wiped off in the US real estate market after the crisis. Noticeably Case and Shiller [5] raised the question of housing bubble before the crisis happened, but it didn't get sufficient attention both from the academic and political world. If the economic world had strong and reliable forecasting system to raise the alarm long before the crisis happened, the great loss to the world could have been prevented.

Why housing price prediction is such a difficult task? The assumption is that the key element to stop the forecasting model to be functional is the omission of the valuable information from the micro-level human decision-making behaviour, especially the irrational behaviours. The economic forecasting model cannot be improved by a technically more and more sophisticated models, without investigation of newly available high-dimensional data [6] [7]. Business cycle forecasting or economic growth forecasting is still one of the most difficult questions in the economic field. Just like the Nobel laureate Murray Gell-Mann once said, "Think how hard physics would be if particles could think." This described exactly the challenge of forecasting housing market performance.

The question is: could we capture other information besides the traditional economic data to reach a better prediction of housing price? Specifically, on-line information has become a new rich mining to understand human behaviour. Can we capture new information from on-line mining to increase the economic data dimension for a better and quicker prediction? This challenge involves finding useful behavioural information, identifying new variables, using appropriate big data methods to improve prediction power. In summary, the challenge is two-fold:

(i) will the adoption of on-line data methods improve the understanding of housing market forecasting from economic perspective? And how?

*Corresponding author. Email: jiaying.kou@live.vu.edu.au

(ii) what kind of information processing methods would suit for economic forecasting problem? What kind of adaptation process needs to be done?

Luckily, there have been some fruitful results produced from both outside of economic world and within economic world in particular forecasting application. For example, information embedded in the Twitter stream can track the H1N1 disease levels in a timely manner and this is extremely important as it shows the potential to control the flu spreading in a faster and more efficient way [30]. Another example of using Twitter stream sentiment information to forecast stock market movement shows how to apply the behavioural economics theory to capture the public mood to support the prediction [36]. This is a very good example to show the interdependence of economic theory and computer science methods to foster a multi-disciplinary fruit. In summary, these examples shed lights on the possibility of improving forecasting ability in housing market movement through new information from big data.

Stock market prediction has a longer research history compared to the housing market, due to the available stock price data and significant amount of wealth involved. Academic finance has passed the excitement of efficient markets theory based on rational expectation. Excess volatility has continuously been found which cannot be explained by the efficient market hypothesis. People tended to look at this problem from a different angle and introduced behavioural finance. Psychologists found people consistently making irrational decisions in experimental environment[8].

The newly development of behavioural economics can provide theoretical guidance of housing market prediction. The challenge is how to apply the theory into the empirical discovery stage. First we introduce some relevant behavioural economic concepts.

"Winner's curse" is a behavioural economic term to describe the situation that people tend to pay far above the expected price in a competitive pricing system. If people pay much higher than the expected value, this difference should gradually disappear because rational people will learn from their mistakes. If such anomaly consistently appears in the human decision making empirically, this is considered against the traditional economic assumption of rationality [1]. It is natural to assume such behaviour would happen in the housing market. When people are in the street auction, the winner of the auction pays much higher than the intrinsic value of the house.

There are other theories can also be applied in the housing market environment, such as prospect theory [2] , the endowment effect [3], or anchoring theory [4].

All these theories study the human decision behaviour when they are apart from the rational status. These theories can show evidence of people's irrational decision-making behaviour in a micro experimental environment.If these individual behaviours aggregate in a social context, it is reasonable to assume the housing price also reflects people's overreact and underreact behaviour. These behaviours are hard to be captured from the traditional economic data collection and model building.

This shows the gap between the macro-level economic forecasting modelling and the available established micro-level behavioural theories. In other words, how to use the advantages of the findings in the micro behavioural theories in the macro behavioural predictions, is a research problem that needs to be solved. The current macro-economic cycle study is based on the rational assumption, which means that the irrational behaviour is not considered in the housing price forecasting models. Even though lots of experimental evidence can be found in the micro economics, this cannot be transferred directly into the macro economics context.

The recent development in the big data analysis in business and e-commerce shed lights on the aggregation power of millions and billions of individual on-line information for a macro-understanding. The hidden information in the on-line search clicks, business transactions, on-line news, chats and Tweets is studied by researchers. For example, since Google has freely opened the Google Trend Index database in 2015, researchers have started to use these search engine data to improve economic indicator predictions [9]. Wu and Brynjolfsson [10] did a pioneer study of housing market cycle prediction in the US and demonstrated the forecasting had a significant improvement with the search engine indicator input.

This paper aims to improve the economic forecasting ability by using the new lens of online big data method, as a way to accumulate millions and billions of individual behavioural decision-making information. It is now possible to produce a more accurate, timely, low cost forecasting of housing market cycles with the support of new technology—big data. The detailed research aims are listed below:

(i) Investigate the forecasting ability of Google Trend Index in the Australian housing market.

(ii) Introduce a new variable—auction clearance rate in the performance model to simulate Australian housing market better. Test the interpretation ability of this variable.

(iii) Test how Google Trend Index interact with the auction clearance rate.

(iiii) Develop the forecasting model based on other big data method, such as indicator generated from text stream analysis from online news, Twitter and Facebook.

2. Contribution to Knowledge and Statement of Significance

2.1. Contribution to knowledge

This will be the pioneer study in Australia to use big data methods to forecast housing market cycles. Housing market performance indicators which have better interpretation for Australian housing market will be developed. Additionally, text stream data will be analysed in the housing market forecasting model. There is a gap to study housing market behaviour using text from Facebook or Twitter. Finally, this research provides an original illustration of the theoretical connection between micro-level behavioural economic theories and macro-level business cycle models.

2.2. Statement of significance

A more accurate forecasting model for the Australian housing market will benefit investors and policy makers to understand the market behaviour. The prediction of housing market cycle is extremely important as the recent global financial and economic crisis showed us the power of the housing market to the economy as a whole. This pioneering study will demonstrate how to implement the big data methods in the economic forecasting study. Hopefully more research can follow and improve the forecasting ability significantly.

The uniqueness of Australian housing market provides a special opportunity to study the human decision-making behaviour under a street auction environment and how such behaviour will influence the housing market as a whole. Australian housing market has been booming continuously since early 2000, only had short slow downs around 2010 and 2018. Financial institutions and government are closely monitoring the current downturn. It is a good timing to test the big data method forecasting ability. This improved prediction will serve as a watchdog of the potential risk or the notice of housing bubbles and help to provide more time for policy reaction for financial institutions and government policy makers. A potential crisis may be prevented with a timely solution.

3. Literature Review

As this research is a cross-disciplinary research joining the economic business cycle forecasting with the big data forecasting methods, literature will be introduced in both areas.

3.1. Existing knowledge of macroeconomic housing market forecasting

The study of housing market is based on the market efficiency hypothesis (EMH). The concept of efficiency came from information theory. The level of efficiency of the stock market is defined on the degree of all available information being reflected in the current stock market prices. A highly efficient market absorbs all available information into its stock prices [11]. Millions and billions of micro-transactions in trading time form stock prices. Institutions and individual investors make trading decisions based on available information and analysis. The same idea works in the housing market. Each of the micro selling and buying decisions forms the housing market prices.

Researchers started to discover overreactions, anomalies in the stock market behaviour that questioning the rationality of stock investors [12] [13]. Interestingly, not many studies have been done on the housing market efficiency behaviour in the literature, comparing with the enormous work been done in the stock market. One of the earliest empirical tests of housing market efficiency for single-family home suggested the housing market is not efficient [14], which means the current housing market performance contains valuable information for the future prediction. Pollakowski and Ray [15] continued the study of housing market efficiency in depth of the interaction of different geographic locations.

3.2. Existing knowledge of microeconomic behavioral theories

Besides testing the market efficiency directly, many researchers turned to psychology and co-developed theories across economics and psychology disciplinary. New theories developed under empirical psychological experiments. These theories are guiding current behavioural research.

A recent World Bank report [16] on human decision making has developed three principles to direct behavioural studies. They are named as "thinking automatically", "thinking socially" and "thinking with mental models". These models are fundamental to the understanding of this area, and literature relating to them are reviewed in the subsections below.

Thinking automatically. People tend to make their decisions automatically, without careful thinking. This is also called System 1 [17] [18]. Making decisions with careful thinking is called thinking deliberatively, also called System 2. People tend to use very limited information to make analysis and make decisions based on very few alternatives. These decisions are easily biased based on the discovery of behavioural theories, such as prospective theory [2], heuristics in judgment and decision-making [19], anchoring and adjustment [4], intertemporal choice [20], the endowment effect [3]. These are the theoretical assumptions of how people make decisions in housing market. Applications have been done in the housing market, such as [21].

Thinking socially. People are highly influenced by the society they live in. Their personal beliefs, prestige, desires, the sense of belongings, motivations are highly influenced by the social norms, social preferences [22]. This will naturally influence the big decisions such as buying a house. Are we buying what we like or are we buying what other people may think of us? Would buy house in a rich suburb be an important factor of my decision-making process? Will people have peer pressure if their friends are buying houses? Are they searching locations where their friends and relatives live? These will be examples of questions to be designed if a survey is drafted.

Thinking with mental models. The term "mental models" means that people tend to make decisions under the acceptance of the "agreed" beliefs among people from the same community. "Culture", defined by the anthropologists and sociologists, is "the collection of mental models" [23] [24].

Australian housing market is an interesting research field to find out the influence of cultures. People from different ethnic groups contain their own culture in what it means of "home", or "house". This can influence their "tenure choice", the choice between renting and owning a property [25], when to buy houses, family structure and house preferences, etc. There may be decision making differentiate between people living in Australia for generations and new migrants. Even the housing locations can be influenced by ethnic backgrounds.

Online social media, such Facebook and Twitter may provide tremendous information about what people are thinking when they buy or sell houses, instead of the traditional survey approach.

3.3. The gap between behavioral theories and macro-level forecasting models, and big data

Historically, there has been a tension between traditional macroeconomic theories of forecasting and those theories arising from behavioural economics. Assumptions in traditional macroeconomics such as the EMH, introduced in Section 3.1, and rational economic actors ("homo economicus"), do not resolve well with the models used in behavioural economic analysis. In the 20th century, there is a gap between these two views of economics without being able to go much further.

Since the 2008 Global Financial and Economic Crisis, a few economists started to work on the improvement of business cycle model by introducing more realistic factors. For example, Kiyotaki [26] has brought financial credit constraints in the interaction of firms and households. He argued that the financial credit constraints can create the economic fluctuation in reality. But these are the first attempts in the theoretical basis, there is still a gap to transfer these new models in the empirical testing models. In addition, these theoretical modellings are in the beginning stage of inputting more realistic factors, with very limited attempts and development. It is far from complete, so these theories have very limited impact on the real forecasting modelling.

However, as we shall describe in the next section, the advent of social data from Internet sources such as Google Trends and Twitter, provides for the first time an empirical basis upon which we may compare predictions arising from these two different views of economics, in particular with respect to the housing market. In section 4, we provide the theoretical framework and methodology for using big data to analyse the housing market. Such type of research is still at its early stages globally, with only a small number of studies being done . There is plenty of room for original research in this area, especially for the housing market in Australia.

4. Methodology and Theoretical Framework

4.1. Current literature of big data methods for housing market forecasting

A few decades ago, both Leontief and Simon pointed out that a better business cycle forecasting cannot be reached by applying more and more sophisticated econometric models, the solution should come from a higher level of new information input, which means a painful empirical micro-level data collection is the way to lead to a higher solution [6] [7]. This emphasizes the necessity to study individual decision making, which contributes to the business cycle.

Leontief predicted that the future economic research will be based on surveys in larger scales and in multiple dimensions. Case et al. [27] have done lots of pioneering work in this area. They have kept sending thousands of surveys to homebuyers yearly for three decades and try to understand their economic reasoning for home buying behaviours. As we are coming to the big data era, some researchers have realized that the online big data is a very powerful new survey tool. Leontief's prediction is gradually realized.

Some pioneer studies used google trend index as an indicator to do the economic indicators' forecasting. [10] studied the real estate cycle in the US. [9] did a similar research in the forecasting of other economic indicators, such as automobile sales, unemployment claims in the US. But the relationship between the behavioural economic theories and the big data methods in forecasting is not discussed in both papers. This discussion is essential for the theoretical framework, because this is to explain why big data method should be effective in the forecasting modelling. This is another gap in the theoretical framework area when analysing the current literature.

Other researchers worked on the forecasting ability of online news, Twitter and Facebook text streams analysis. Soo [28] studied 34 cities news about housing market sentiment and found out that this information has a predictive power about the future housing prices. Sun et al. [29] combined the information from online news articles and search engine to predict the real estate cycle in China.

There hasn't been any literature on using Twitter or Facebook text streams of public sentiment to predict the housing market cycle. This is another gap. One of the relevant inspirational research is done by [30]. Their research demonstrated the forecasting ability of public sentiment from Twitter text streams on the H1N1 disease levels. Social network analysis methods in computer science provides advanced theoretical framework, algorithms that can be adapted in the housing market studies using Twitter or Facebook text streams [31] [32].

In the following sections, methodologies of Google Trend Index, online news and text streams from Facebook and Twitter will be discussed.

4.2. Forecasting model using Google Trend Index

Background of Google Trend Index. In 2015, Google company made real time Google Trends data available publicly. There are trillions of search clicks every year through Google search engine. This data provides a unique opportunity to find out people's searching interest globally or granulated down to the city level in a specific period.

Google Trends index is a normalized data to look at the search interest about a topic relatively comparing to the total clicks in the same place and time. After selecting the time period and location, the highest search interest is counted as 100. Other search interests in the same period will adjust to compare with the maximum figure and generate a ratio.

Selection of housing market performance indicators. The purpose of this model is to test the interpretation power of Google trend index in forecasting the future housing market performance.

There are a few indicators to evaluate the housing market performance, such as sales quantity, auction clearance rate, housing price index (HPI). Sales Quantity and HPI were used by [10] for housing market indicators. Auction Clearance Rate (ACR) is an original indicator proposed in this paper.

ACR is included as an indicator because the clearance rate indicates the percentage of sellers are satisfied with the auction price. If a house auction is passed in, it means the auction price hasn't reached the vendor's reserve price.

Street auction is quite unique and popular in Australia, which is rare to see in other countries. Large number of houses are sold in the auction, especially the high value properties in Australia. Both in UK and US, the auction of houses is not the main selling method. And the auction is not held in front of the selling houses, but gathered in an auction room, or in front of some court areas.

The assumption is that people in Australia have more chance to observe or experience a house auction than people in most other countries. Based on the endowment effect, we assume people are more attached to the property they are willing to bid if they are standing in front of the property for auction. The feeling of losing an auction is more real and stronger comparing to quietly sitting in an auction room. Therefore, the clearance rate provides a unique opportunity to understand the behaviour of Australian housing market.

Google Trend Index forecasting model design. The model design follows [10]'s model structure. A simple seasonal autoregressive time series model is chosen to test the interpretation power of search indices for the housing market performance. The major difference from [10]'s model is the newly introduced indicator—ACR.

The basic auto-regression of home sales quantity model is shown as equation (1), then Google trend data—Search Frequency is introduced into the equation to find out if Google Trend Index improves the prediction of the current Home Sales figure. The current and one lag behind search frequency are introduced in the equation. All the performance indicators data are collected in a seasonal basis.

$$\begin{aligned} HomeSales_{it} = \alpha &+ \beta_1 HomeSales_{i,t-1} + \beta_2 HPI_{i,t-1} \\ &+ \beta_3 ACR_{i,t-1} + \beta_4 Population_{it} \\ &+ \sum S_i + \sum R_j + \sum T_t + \varepsilon_{it} \end{aligned} \quad (1)$$

$$\begin{aligned} HomeSales_{it} = \alpha &+ \beta_1 HomeSales_{i,t-1} + \beta_2 HPI_{i,t-1} \\ &+ \beta_3 ACR_{i,t-1} + \beta_4 SearchFreq_{it} \\ &+ \beta_5 SearchFreq_{i,t-1} + \beta_6 Population_{it} \\ &+ \sum S_i + \sum R_j + \sum T_t + \varepsilon_{it} \end{aligned} \quad (2)$$

The logic to choose a simple linear regression structure. First, the main focus of this paper is to test the interpretation power of Google Trend Index in Australia. Therefore, we use a simple linear regression to distinguish this variable.

Second, based on [14] and [10], past housing market indicators have the power to predict the future. Therefore, a simple auto regression is appropriate.

Third, Wu and Brynjolfsson [10] have demonstrated the interpretation power of search indices using US data. Their results were even better than the predictions from the National Association of Realtors.

Fourth, this paper also aims to test if Auction Clearance Rate is a good indicator for housing market performance.

Fifth, they [10] have also found that simple linear regression had even better results than more sophisticated nonlinear models. This finding is coherent with the theoretical assumptions made in the previous section, that a higher level of information input by adding the search information is the solution of current forecasting problems, not how sophisticated the econometric model is.

Future prediction model. The next step is to test the forecasting ability of the search data for the future home sales quantity. We follow [10]'s model, the current, one-period and two-period lags search frequencies in the model. We could get the current search frequency for the future prediction. Two-period lags cover nine months of searching period prior house purchasing. We assume that this is a practical maximum searching period. The difference from [10]'s model is that we introduce the auction clearance rate into the model.

$$\begin{aligned} HomeSales_{it} = \alpha &+ \beta_1 HomeSales_{i,t-1} + \beta_2 HPI_{i,t-1} \\ &+ \beta_3 ACR_{i,t-1} + \beta_4 SearchFreq_{it} \\ &+ \beta_5 SearchFreq_{i,t-1} \\ &+ \beta_6 SearchFreq_{i,t-2} + \beta_7 Population_{it} \\ &+ \sum S_i + \sum R_j + \sum T_t + \varepsilon_{it} \end{aligned} \tag{3}$$

Same modelling process with other performance indicators. Following the same structure, we can predict the current and future HPI. Then, we can also predict the current and future Auction Clearance Rate (ACR). Examples are shown in equation (4) and (5).

$$\begin{aligned} HPI_{it} = \alpha &+ \beta_1 HPI_{i,t-1} + \beta_2 Homesales_{i,t-1} \\ &+ \beta_3 ACR_{i,t-1} + \beta_4 SearchFreq_{it} \\ &+ \beta_5 SearchFreq_{i,t-1} + \beta_6 Population_{it} \\ &+ \sum S_i + \sum R_j + \sum T_t + \varepsilon_{it} \end{aligned} \tag{4}$$

$$\begin{aligned} ACR_{it} = \alpha &+ \beta_1 ACR_{i,t-1} + \beta_2 Homesales_{i,t-1} \\ &+ \beta_3 HPI_{i,t-1} + \beta_4 SearchFreq_{it} \\ &+ \beta_5 SearchFreq_{i,t-1} + \beta_6 Population_{it} \\ &+ \sum S_i + \sum R_j + \sum T_t + \varepsilon_{it} \end{aligned} \tag{5}$$

4.3. Forecasting model with online news

The key element to predict housing market cycle with the information from online news is to build the relationship through "sentiment".

Barberis et al. [33]studied stock market investors sentiment on the stock performance news, such as earnings announcements. And he generalized the sentiment as "overreaction" and "underreaction" with empirical evidence.

Tetlock [34] also quantified the interactions between Wall Street Journal column and the stock market performance. A pessimistic view can drive down the market prices as an overreaction, which will follow a reversion to fundamentals.

With the development in the stock market sentiment study, the housing market study follows the trend. Soo [28] did a pioneer study of American housing market sentiment study by analysing the newspaper tones as "positive" or "negative".

The textual analysis is used to quantify the tone of financial documents. The standard dictionary-based method is used to count the raw frequency of positive and negative words in a text. Soo [28] prepared a housing dictionary and presented the calculation of the overall tone of housing market news sentiment by:

$$S = \frac{\#pos - \#neg}{\#totalwords} \tag{6}$$

Kou et al. [35] followed this study and calculated the media sentiment index with two different dictionary methods at a suburb-level in Australia. This approach can be continued by extending the suburb study to a macro-level.

4.4. Forecasting with text streams from Facebook and Twitter

Articles directly using Facebook or Twitter text streams to predict housing market cycles are not found in the literature. But there are quite a few research papers using the idea of mood or sentiment analysis to predict stock market behaviour [36] [37] [38]. We could reasonably assume that the nature of the prediction algorithm for housing market will be similar to the stock market.

How to find emerging topics is a challenge in social media analysis. Early detection can improve the understanding of people's behaviour towards market movement. Novel tracking method can be found in [39]. Big data analysis could be another challenge.Some technique and solutions include [40] [41].

We assume that the selection of Twitter feeds need to address the location difference, unlike the stock market prediction, because the housing market cycle shows a strong trend difference world widely. But Shiller [42] has shown the big glamorous cities experienced massive boom within similar timeframe (1999-2014). Australian cities, Sydney and Melbourne are recognised as the glamorous cities and following the global boom trends. Therefore, we may test the Tweets in Australia and world-wide for the location hypothesis.

4.5. Data source

Google Trend Index is a free publicly available database. The index figure can be downloaded at state level in Australia with monthly interval. Most other housing market indicator data can be found through AURIN data (The Australian Urban Research Infrastructure Network). Online news and samples of text streams from Twitter and Facebook can be captured by Crawler Programs.

The auto-regression time series forecasting model can be run through R or other statistical software. We will compare the results using different variables and find out the forecasting ability and compare the forecasting results of different big data methods.

Data privacy issue also needs to be considered when using text streams from Twitter or Facebook. Certain ways of privacy protection can be applied to avoid personally identifiable information [43][44].

5. Conclusion

This study uses the latest available technology, data and research method to analyse and forecast the Australian housing market. It will provide new insights in economic cycles and in the relationship between behavioural microeconomics and traditional macroeconomics. These methods will provide a significant improvement in the quality, cost-effectiveness and timeliness of Australian housing market forecasts, and become a valuable tool for investors, bankers and policy makers. Predictions from these methods may prove to be the earliest predictors of economic downturn and upturns. Hope this research may also inspire other research to develop more sophisticated methods of using big data in economics.

References

[1] Thaler, R.H. (1988) Anomalies: The winner's curse. *Journal of Economic Perspectives* **2**(1): 191–202.

[2] Kahneman, D. and Tversky, A. (2013) Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (World Scientific), 99–127.

[3] Kahneman, D., Knetsch, J.L. and Thaler, R.H. (1991) Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives* **5**(1): 193–206.

[4] Tversky, A. and Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases. *Science* **185**(4157): 1124–1131.

[5] Case, K.E. and Shiller, R.J. (2003) Is there a bubble in the housing market? *Brookings papers on economic activity* **2003**(2): 299–342.

[6] Leontief, W. (1971) Theoretical Assumptions and Nonobserved Facts. *American Economic Review* **61**(1): 1–7.

[7] Simon, H.A. (1984) On the behavioral and rational foundations of economic dynamics. *Journal of Economic Behavior and Organization* **5**(1): 35–55.

[8] Shiller, R.J. (2003) From efficient markets theory to behavioral finance.*Journal of economic perspectives* **17**(1): 83–104.

[9] Choi, H. and Varian, H. (2012) Predicting the present with google trends. *Economic Record* **88**(s1): 2–9.

[10] Wu, L. and Brynjolfsson, E. (2015) The future of prediction: How google searches foreshadow housing prices and sales. In *Economic Analysis of the Digital Economy* University of Chicago Press, 89–118.

[11] Malkiel, B.G. and Fama, E.F. (1970) Efficient capital markets: A review of theory and empirical work. *The journal of Finance* **25**(2): 383–417.

[12] De Bondt, W.F.M. and Thaler, R.H. (1989) Anomalies: A mean-reverting walk down wall street. *Journal of Economic Perspectives* **3**(1): 189–202.

[13] Jegadeesh, N. and Titman, S. (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* **48**(1): 65–91.

[14] Case, K.E. and Shiller, R.J. (1989), The efficiency of the market for single-family homes. *American Economic Review* **79**: 125–37.

[15] Pollakowski, H.O. and Ray, T.S. (1997) Housing price diffusion patterns at different aggregation levels: an examination of housing market efficiency. *Journal of Housing Research* : 107–124.

[16] World Bank (2015) World development report 2015: Mind, society, and behavior (World Bank Group).

[17] Evans, J.S.B. and Stanovich, K.E. (2013) Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science* **8**(3): 223–241.

[18] Kahneman, D. (2003) Maps of bounded rationality: Psychology for behavioral economics. *American economic review* **93**(5): 1449–1475.

[19] Kahneman, D. and Tversky, A. (1981) On the study of statistical intuitions. *Cognition* **11**(2): 123–141.

[20] Loewenstein, G. and Prelec, D. (1992) Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics* **107**(2): 573–597.

[21] Northcraft, G.B. and Neale, M.A. (1987) Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational behavior and human decision processes* **39**(1): 84–97.

[22] Tomasello, M. (2014) *A natural history of human thinking* Harvard University Press.

[23] Swidler, A. (1986) Culture in action: Symbols and strategies. *American sociological review* : 273–286.

[24] DiMaggio, P. (1997) Culture and cognition. *Annual review of sociology* **23**(1): 263–287.

[25] Van Ham, M. and Manley, D. (2009) The effect of neighbourhood housing tenure mix on labour market outcomes: a longitudinal investigation of neighbourhood effects. *Journal of Economic Geography* **10**(2): 257–282.

[26] Kiyotaki, N. (2011) A perspective on modern business cycle theory. *FRB Richmond Economic Quarterly* **97**(3):195–208.

[27] Case, K.E., Shiller, R.J. and Thompson, A. (2012) What have they been thinking? Home buyer behavior in hot and cold markets. *Brookings Papers on Economic Activity* **2012**(2): 265–98.

[28] Soo, C.K. (2018) Quantifying sentiment with news media across local housing markets. *The Review of Financial Studies* : hhy036

[29] Sun, D., Du, Y., Xu, W., Zuo, M., Zhang, C. and Zhou, J. (2015) Combining online news articles and web search to predict the fluctuation of real estate market in big data context. *Pacific Asia Journal of the Association for Information Systems* **6**(4).

[30] Signorini, A., Segre, A.M. and Polgreen, P.M. (2011) The use of twitter to track levels of disease activity and public concern in the us during the influenza a h1n1 pandemic. *PloS one* **6**(5): e19467.

[31] Peng, M, Zhu, J, Li, X, Huang, J, Wang, H and Zhang, Y (2015) Central topic model for event-oriented topics mining in microblog stream. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 1611–1620.

[32] Wang, G, Wang, H, Tao, X and Zhang, J (2013) A self-stabilizing algorithm for finding a minimal positive influence dominating set in social networks. *Proceedings of the Twenty-Fourth Australasian Database Conference*, **137**:93–99.

[33] Barberis, N., Shleifer, A. and Vishny, R. (1998) A model of investor sentiment. *Journal of financial economics* **49**(3): 307–343.

[34] Tetlock, P.C. (2007) Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* **62**(3): 1139–1168.

[35] Kou, J., Fu, X., Du, J., Wang, H. and Zhang, G.Z. (2018) Understanding Housing Market Behaviour from a Microscopic Perspective. *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, **:**1–9. IEEE

[36] Bollen, J., Mao, H. and Zeng, X. (2011) Twitter mood predicts the stock market. *Journal of Computational Science* **2**(1): 1 – 8.

[37] Mittal, A. and Goel, A. (2012) Stock prediction using twitter sentiment analysis. *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)* **15**.

[38] Zhang, X., Fuehres, H. and Gloor, P.A. (2011) Predicting stock market indicators through twitter “I hope it is not as bad as I fear”. *Procedia - Social and Behavioral Sciences* **26**: 55– 62.

[39] Huang, J., Peng, M., Wang, H., Cao, J., Gao, W., and Zhang, X. (2017) A probabilistic method for emerging topic tracking in microblog stream. *World Wide Web* **20**(2):325–350.

[40] Li, H, Wang, Y, Wang, H, and Zhou, B (2017) Multi-window based ensemble learning for classification of imbalanced streaming data. *World Wide Web* **20**(6):1507–1525.

[41] Zhang, J,Tao, X,and Wang, H (2014) Outlier detection from large distributed databases, *World Wide Web* **17**(4):539–568.

[42] Shiller, R.J. Irrational exuberance: Revised and expanded third edition (2015) *Princeton university press*.

[43] Wang, H. and Sun, L. (2010) Trust-involved access control in collaborative open social networks. *2010 Fourth International Conference on Network and System Security*, **239–246**, IEEE

[44] Wang, H. Jiang, X. and Kambourakis, G. (2015) Special issue on Security, Privacy and Trust in network-based Big Data. *Information Sciences*,**318**:48–50, Elsevier