Cities 52 (2016) 140–147

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

How website users segment a city: The geography of housing search in London

Alasdair Rae [a,*], Ebru Sener [b]

[a] Department of Urban Studies and Planning, University of Sheffield, Sheffield S10 2TN, United Kingdom
[b] KONDA Research and Consultancy, Yıldız Posta Caddesi, Çiğdem Apt. No:11/6, Gayrettepe, 34349 Istanbul, Turkey

ARTICLE INFO

Article history:
Received 1 May 2015
Received in revised form 16 December 2015
Accepted 19 December 2015
Available online 4 January 2016

Keywords:
Housing markets
Search extent
London
Housing
Spatial analysis
Big data

ABSTRACT

This paper explores spatial patterns of housing search in London, using data generated by users of the UK's most popular real estate portal. By focusing on the variable geographies of 'search extent', it attempts to make a contribution to a long line of studies focused on understanding the fragmented geography of metropolitan housing markets. It also builds upon more recent work in economics on the utility of user-generated search data. After introducing our approach, we discuss the background to housing search and the wider emergence of 'search' as an object of study. We then provide further details on the data and methodology before exploring the spatial and sectoral characteristics of search in London. The results suggest that there is much to be gained by incorporating search studies into housing market analysis and that there is significant potential for future work in this area.

1. Introduction

Housing markets are inherently spatial entities, so it is not surprising that their spatial structure has been a key concern for analysts over the years (e.g. Palm, 1978; Goodman & Thibodeau, 2007). Underlying the spatial composition of housing markets – and submarkets in particular – is the concept of 'search extent'. This relates to the mental and physical geographies associated with housing search at the very early stages of market interaction. It is a feature of the housing search process in which households search for substitutable properties within specific areas, based on preferences such as price and property characteristics. Some analysts have favoured a spatial approach to understanding housing search processes and others have explored the topic from a sectoral perspective but there remains no clear consensus on which approach is optimal (Watkins, 2001; GLA, 2004; 2013).

Rather than seek to provide a definitive approach to housing submarket definition, this paper combines sectoral and spatial approaches to develop an overview of the geography of search extent in London using a very large housing search dataset. It does so in the context of a London housing market in the midst of historic price inflation and lack of affordability (Reed, 2015) and at a time when online real estate portals are the first point of engagement for the vast majority of individuals entering the market (Dunning & Grayson, 2014). In addition to literature on housing submarkets, our approach also draws upon recent work on the subject of search extent (e.g. de Groot, Mulder, & Manting, 2011; Chen & Lin, 2012) and earlier studies focusing on search behaviour in a metropolitan context (Brown & Holmes, 1971; Huff, 1986).

The research is based on a user-generated dataset of more than 100,000 unique searches, sourced from rightmove.co.uk, the UK's leading housing market portal. It builds on previous work in the United Kingdom by Rae (2014), and emerging work in the United States by Beracha and Wintoki (2013) and Kroft and Pope (2014). A novel element of this research is that all search areas have been drawn by website users and the geographical structures we analyse are therefore not confined to existing administrative boundaries of the kind so commonly encountered in previous studies (e.g. Hincks & Baker, 2012). The focus in the paper is on residential properties for sale, and we use data from March 2013.

The next section of the paper performs two functions. First, following previous work, it makes the case for incorporating user-generated search data into housing market research (e.g. Wheaton, 1990; Maclennan, 1992). Second, it reflects upon emerging interest in the use of online search data and its potential to add value to existing knowledge (e.g. Choi & Varian, 2012; Wu & Brynjolfsson, 2009). After this initial positioning, the methods and data are explained in more detail. This is followed by an overview of housing search activity within London as a means to foreground the main exploratory spatial data analysis, which in turn looks at spatial and sectoral search, and the differences between where

* Corresponding author.
E-mail addresses: a.j.rae@sheffield.ac.uk (A. Rae), esener@konda.com.tr (E. Sener).

http://dx.doi.org/10.1016/j.cities.2015.12.002

people search and where properties are available. We conclude by discussing the implications and limitations of our findings and by suggesting opportunities for further research in this area.

2. Making the case for user-generated search data in housing market studies

This study builds upon a significant volume of previous research on the topic since, as Chen and Lin note, 'the extent of a search is one of the most studied aspects' of the housing search process (2012, p. 901). They identify several measures previously used to investigate spatial search extent, including the number of properties visited, the number of neighbourhoods searched, the mean distance among vacancies searched, and the total area searched (e.g. Brown & Holmes, 1971). What is most significant in the context of this study is that Chen and Lin proposed to understand the extent of search through the concept of mental maps and awareness of space, following Lynch's famous 'mental maps' approach in *The Image of the City* (1960).

The Chen and Lin method helps overcome some of the traditional limitations encountered in previous survey-based research (e.g. Maclennan, 1992) and situates conceptions of space at the first stage of the search process, as we do in this study and as Rae (2014) does in his model of housing search. Despite their innovative approach, however, the authors still had to solicit their 82 participants through newspaper advertisements, brokers, public relations offices and personal connections so unlike what we propose in this paper, it would be difficult to replicate and update at a national scale over time.

There are several other examples of seminal work on the importance of housing market search, including papers by Huff (1986) and Wheaton (1990). Making the case for incorporating search models, Huff states that 'search-based models of residential mobility are conceptually appealing, but they have proven to be difficult to operationalize' (1986, p. 224), partly as a result of data constraints. Wheaton in particular provides a useful departure point for the present study since his conception of 'search costs' was constructed in an era when information asymmetries were more strongly weighted against buyers in that 25 years ago it was much more difficult to identify suitable properties and obtain detailed information on them. This was also highlighted by Brown and Holmes (1971, p. 308) when they commented on the limited 'awareness space' of households in the search process.

With the advent of mass market real estate portals in many countries, it gives rise to the possibility that we might be able to develop a more accurate understanding, and build upon existing research, if we draw upon the vast data resources being generated by website users. Such 'big data' approaches to understanding housing market search extent could significantly enhance existing 'small data' approaches and simultaneously overcome the restrictions reported in other studies in relation to the use of administrative boundary data (e.g. Hincks & Baker, 2012). This is the first city-focused study of its kind in the British context, but previous studies in economics have taken a similar approach, as we describe below.

There is a long tradition of studying search frictions in the labour market and it is in this field that most progress has been made (KVA, 2010). Although labour and housing markets exhibit many fundamental differences, several transferable concepts suggest that studying search behaviour in the housing market could yield potentially useful results, particularly in relation to search extent. For example, the changing dynamics of information asymmetry between buyers and sellers, mediated by online housing portals, is one example (Pope, 2008). Another is the way in which potential buyers can now search very easily an almost complete set of products across an entire country, in contrast to previous decades when search was restricted to local newspapers, real estate offices, or driving around individual neighbourhoods and looking for properties (Palm, 1976). This has undoubtedly lowered search frictions but, of course, has not necessarily led to better outcomes for buyers.

In 2001, when Palm and Danis looked at the impact of web-based information on housing market search, they found that little had changed from previous decades, but more recent research suggests that the situation is now fundamentally different (Rae, 2014; Dunning & Grayson, 2014). This mirrors existing work in economics pioneered by Choi and Varian in a series of contributions since 2009, where they have explored the potential of 'predicting the present' with Google search data (e.g. Choi & Varian, 2009, 2012). Other researchers, notably Wu and Brynjolfsson (2009), have attempted to apply search data analysis methods to the housing market to predict how Google searches might foreshadow housing prices and sales. More recently, Beracha and Wintoki (2013) used Google search data to claim that abnormal search intensity in a particular city could help predict future price volatility.

A more recent advance in the use of online search data in understanding segmentation of the housing market has come from Piazzesi, Schneider, and Stroebel (2015). Instead of using coarser, indexed Google Trends data, they source search information from Trulia, a major US real estate portal. They focused their study on 183 San Francisco Bay Area zip codes to explore factors such as turnover, time on the market, inventory and search queries. This research represents the cutting edge of search studies in the housing market, but to date nothing similar has been conducted in the UK. This is not surprising, given the difficulty in obtaining search data, but it is somewhat disappointing given the importance of the housing market to national economic performance, and the current crisis of affordability in London. In 2011 The Bank of England highlighted this issue in a study on using online search data as economic indicators (McLaren & Shanbhogue, 2011) but only now is work beginning to emerge which explores this potential in any depth (Rae, 2014). The next section therefore explains the data and methods used herein, before we proceed to an analysis of housing search extent from a user perspective.

3. Data and methods

This paper represents a tentative first step towards a greater understanding of housing market search extent for London, focused on residential properties for sale in March 2013. There is a strong exploratory spatial data emphasis in the study since this is an effective way of unleashing the explanatory power of the underlying data, which are very large and inherently spatial. It is also somewhat experimental since such a large user-generated search dataset has never before been used in the UK housing market context. The data are sourced from Rightmove plc, the nation's leading online housing market portal. Rightmove is similar to Trulia or Zillow in the United States or Funda in The Netherlands and within England includes more than 90% of all properties for sale. Users can search for any kind of geographical entity, such as a town, a postcode or a train station, using the search form on the website (Fig. 1). A list of available properties is returned so that users can then refine results based on property characteristics such as price and number of bedrooms. Users can also choose a map-based search where they draw an area on a Google map and then view properties for that area only, which they can then refine further, as above. This is Rightmove's 'Draw-a-Search' tool and has been active on their website since it was first developed in 2010.

Two types of data are used here. First, we have a set of almost 100,000 user-drawn search polygons from March 2013, covering the entirety of Greater London. For each user-defined search area, we have details on the price and property attributes they entered on the website (if any). We also have a time-stamped search session variable associated with each polygon so we are able to ascertain whether an individual website user performed multiple searches during a session. This could be used in future to understand the kinds of 'search journeys' users go through during online sessions. In order to preserve the confidentiality of website users, we were not provided with any personal identifying information such as name, street address or IP addresses. The second kind of data is for individual properties for sale. We have

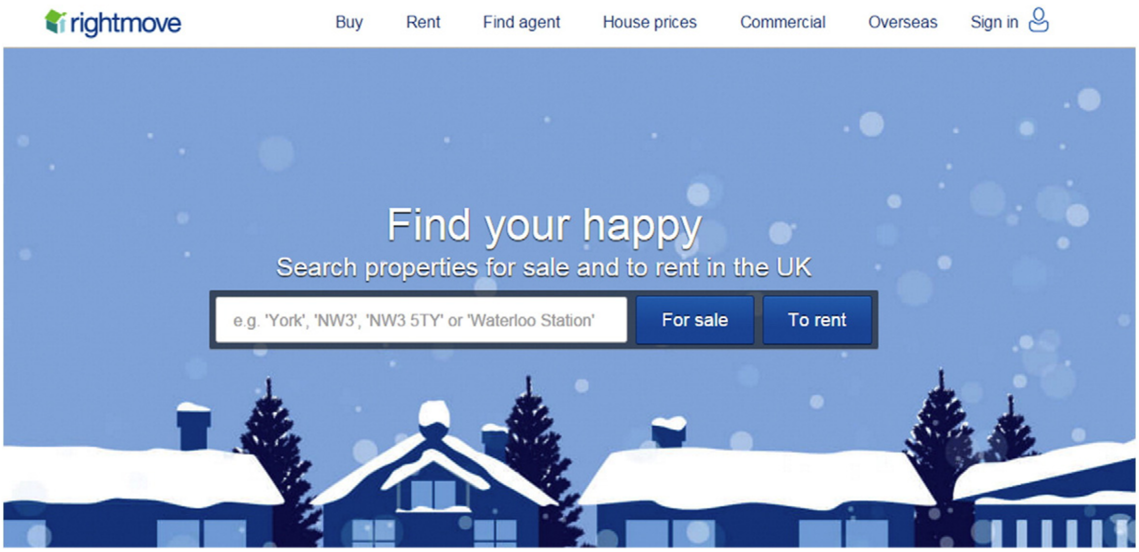

Fig. 1. rightmove.co.uk home page.

precise property locations for over 80,000 residential properties listed for sale during March 2013 on rightmove.co.uk. For each property, we have details of its size, listed price, type (such as 'flat', 'house') full postcode, and date of first listing.

The basic characteristics of the two datasets are in themselves quite revealing. For example, the search polygon dataset contains 47,638 entries where the user did not specify the number of bedrooms required. This compares to 7059 searches for one bedroom properties, 21,983 for two bedrooms, 13,001 for three bedrooms, 2398 for four bedrooms and 4265 for five bedrooms. The available properties point dataset contains 2826 properties with no bedrooms ('studio' flats in most cases), 14,093 with one bedroom, 27,694 with two bedrooms, 21,949 with three bedrooms, 9609 with four bedrooms, 4266 with five bedrooms and 2117 with six or more. The differences between what is available (supply) and what people are searching for (demand) are, of course, a fundamental concern in studies of search frictions so these simple descriptive statistics alone can potentially provide an insight into the ways in which London's housing submarkets might be spatially and sectorally mismatched.

The next part of the paper explores how users of the Draw-a-Search tool define their search extents in a way that was not possible only a few years ago. Website users have a completely free hand in deciding the shape and size of their search area – since it involves drawing an area of any size or shape on the computer – so the shapes of areas reflect the spatial preferences of users and not pre-determined administrative units. We then move beyond individual drawn searches to explore the differences between where people look and where properties are actually available. This spatial search mismatch could be a potentially important part of raising awareness of localised market pressures in London in relation to identifying areas of unmet need, or particularly high demand. At the very least, we hope it will provide new information on the unique pressures of the wider London market. Following Anselin (1998), the analysis presented takes an exploratory spatial data approach, since our aim is primarily to demonstrate the potential of a user-generated search extent approach in this initial foray into the data.

4. Variable geographies of search extent in London

In the first phase of the analysis, our goal was to understand how website users undertook their searches in relation to spatial and sectoral segmentation. This is based on an analysis of queries for March 2013; a time of year when housing search is particularly vibrant. We were interested in what price users specified, the size of the areas they identified, and the size of properties in relation to number of bedrooms; all very common property search attributes. Conceptually, this follows the Watkinsian view that submarkets (and the search processes which underpin them) can be simultaneously sectoral and spatial (Watkins, 1999).

British properties are not typically marketed using area metrics such as square metres so in this case number of bedrooms serves as a proxy for size. We are well aware that not all search is 'active' in the sense that people are actively engaged in housing search with a view to buy (see Rae, 2014). However, previous studies using aggregate and disaggregate Google search data have shown that even when the ultimate intentions of searchers are not known the results of such studies correlate with future market activity (e.g. McLaren & Shanbhogue, 2011; Choi & Varian, 2012). Furthermore, even recreational search has an important role to play in aspiration formation in the housing market, particularly as it pertains to identifying desirable areas. We clearly need to be circumspect in our interpretation of results based on such 'big data' sources but the evidence below suggests that there are important sectoral and spatial differences in the way people search and this could play an important role in the functioning of the housing market at a metropolitan level. The pertinent question now is what online housing search looks like on the ground. This is explored below, after we provide an overview of the dataset's characteristics.

From a London-wide perspective, the average maximum property value that Draw-a-Search users specified is £535,000, which is higher than the current average house price in London (c. £464,000; HM Land Registry, 2015). The average minimum price users specify is £418,000. The mean property size Rightmove website users specify in London is 2.2 bedrooms, although just over half the time (51.5%) users do not specify an initial bedroom number. From a more technical perspective, the median number of vertices that people use to draw their search area is 8 and the modal value is 5, which indicates that most users do not draw very intricate housing search areas.

Of more interest in the context of search extent is the distribution of search areas by geographical size. The total area of Greater London is 1572 km^2 (607 miles^2) but, as one might expect, the vast majority of search in London is at a very local scale. For example, 23.8% of drawn searches are 1 km^2 or less, and just over 50% are 5 km^2 or less. The distribution has a long tail and searches up to 30 km^2 account for 86.5% of the total. This is shown in Fig. 2, aggregated to 5 sq. km bands. To put this into some kind of perspective, the central London Borough of Camden is just under 20 km^2 in size so the search horizons of the majority of users are significantly smaller than the administrative geography of London Boroughs. From a geographical perspective, this confirms what is already known in theory about the first stage of housing market

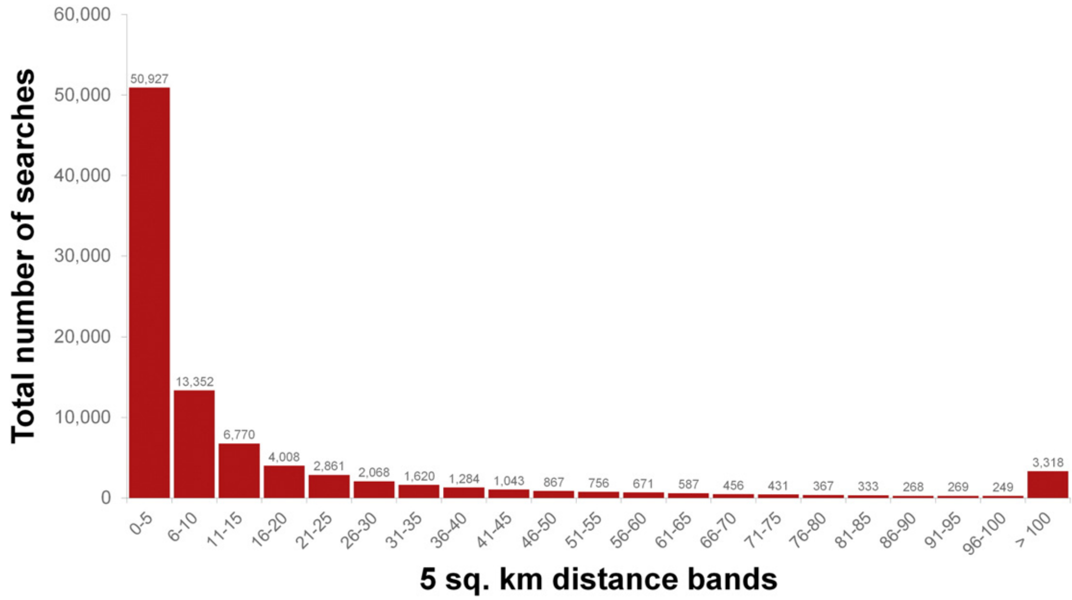

Fig. 2. Histogram of the geographical scale of housing search in London.

engagement, but it provides new data on the precise nature of search extent for a large metropolitan area.

The first stage of our more localised spatial-sectoral analysis examined drawn search across Greater London in relation to *both* property size and property value. We identified all searches where bedroom number was not specified, in addition to 1, 2, 3 and 4 or more bedrooms. At the same time, we also divided searches into four categories of price, with the mid-point of £500,000 being slightly higher than the London average house price. The minimum price specified by users is £50,000, since this is the lowest value available on rightmove.co.uk. The results of the analysis are shown in a series of 'small multiples' in Fig. 3 and immediately provide evidence of London's differentiated housing search, though not necessarily its stock. The maps are shaded with semi-transparent search polygons so that darker colours represent a higher volume of search than lighter areas. We tested this approach by segmenting the data geographically using different search extent cut-offs (1 km, 2 km, 5 km and 10 km) but the results were not significantly different in terms of the locations of the most intense search. This is most likely related to the fact that the vast majority of search is actually very localised, as shown in Fig. 2.

Even at this overview level we can see that a significant number of people search for properties in the lower price band without specifying the number of bedrooms (top left image) and that there are a very small number of people searching rather optimistically for 4 bedroom properties up to £250,000. Further, if we compare the images in a single column – based on price – we can observe an obvious differentiation. This is particularly clear in properties of £750,000 or above where the geography of search extends outwards to Outer London as property size increases. A similar phenomenon occurs when we compare images in a single row – based on number of bedrooms. For two bedroom properties, for example, we can see a different pattern emerging for each price band, and a much narrower geography of search for the most expensive category. What is particularly interesting is an observable east/west London divide in search for two bedroom properties in the lowest price band. This reflects spatial demand on the one hand but also price variation within London.

In order to provide a larger scale, more detailed view of aggregated search extents – potential submarket structure – in London, Fig. 4 shows all drawn search for three bedroom properties in the £250,000–500,000 price range. Here we can see several areas where the core of distinct 'submarkets' appear to emerge, including in Havering to the north east, Sutton to the south, Enfield in the north and Hillingdon in the west. The extent to which these areas *actually* serve as functional submarkets is a matter for further research but the point here is that from a user perspective these patterns are likely to represent some form of spatial demand. The question now is the extent to which this kind of 'demand' matches the reality on the ground; that is, where such properties are actually available (cf. Brown & Holmes, 1971). The next section of the paper explores this question in more detail.

5. Search extent and market geography: is there a spatial mismatch?

The next stage of our analysis involved examining the differences between where people search for housing and where properties are available. We do this in an attempt to fill a gap in the literature around the spatiality of demand and supply and in relation to the geography of 'latent' vs 'revealed' demand in particular. Put simply, 'revealed demand' (Jones & Watkins, 2009) represents where movers actually relocate to and is traditionally examined in housing market studies using migration data. Here we attempt to take one step back from the migration process and examine the issue of 'latent demand' (Jones, Leishman, & Watkins, 2005) in relation to identifying the basic relationship between search and property availability. Understanding the differences between spatial demand and supply could ultimately allow us to determine the extent to which the housing market is subject to sub-optimal 'satisficing' moves (Flowerdew, 1976), though this analysis only represents a first step towards such a goal.

We once again segment our data using simple property characteristic and price bands but this time we add in spatial data on the location, asking price and type of houses actually on the market. In order to simplify the analysis and remove uncertainty, here we exclude searches where the price or property size is not specified by searchers. The resulting small multiple map series is displayed in Fig. 5. The spatial patterning of available properties (black points) and search intensity (shaded as in previous maps) allows us to explore visually the relationship between housing market search on the one hand, and available products on the other.

Starting at the bottom left of Fig. 5, we can again see that very few users search for properties with four or more bedrooms in the cheapest price band. This would appear to reflect the perception of London's wider housing market. In fact, there were many properties listed in this price band in March 2013, though the vast majority were in very poor condition. Probably the most obvious examples of latent demand in London's housing market on the above matrix are visible for two

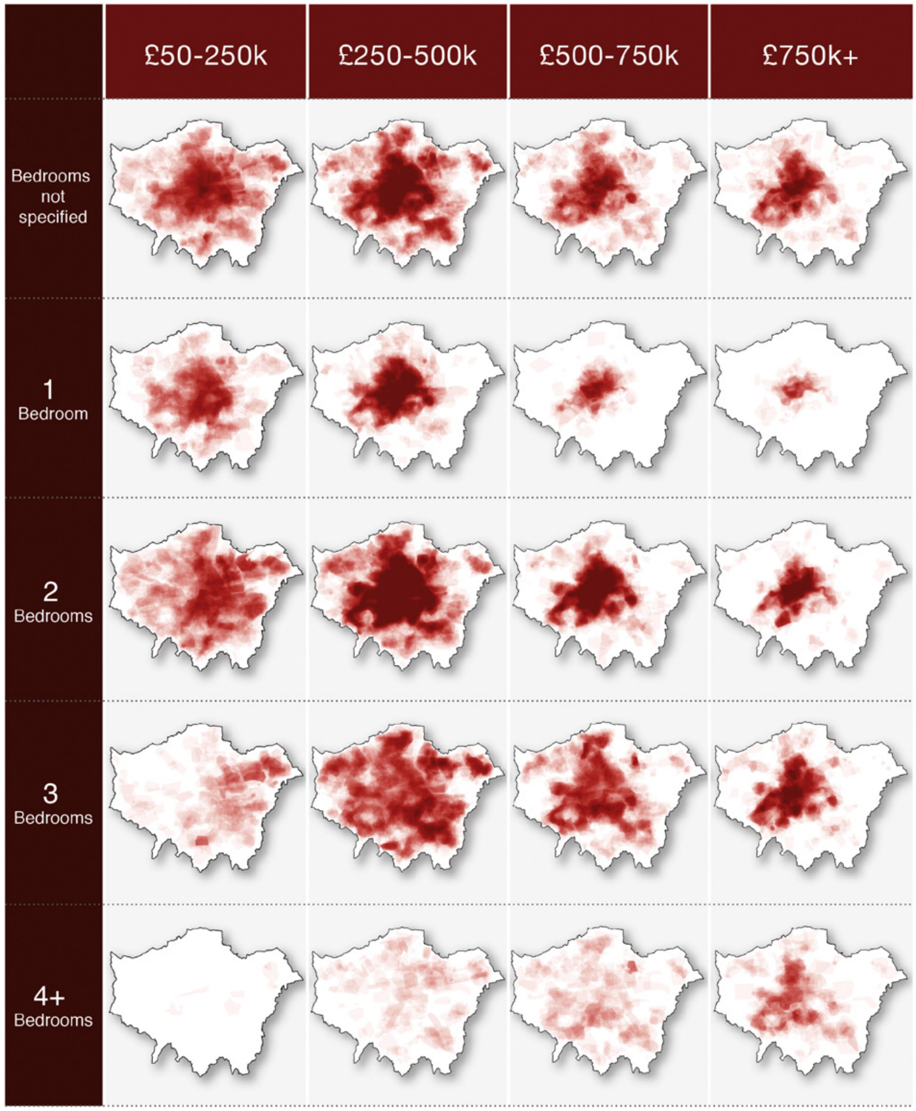

Fig. 3. Spatial and sectoral disaggregation of London housing search.

bedroom properties in the £50–250 k and £250–500 k price bands in central London. Very few such properties exist in these areas yet the intensity of search is highest here.

In most other cases there is a close match between the location of available properties and search intensity, which suggests that people searching for housing have a reasonable understanding of market geography. One exception to this is in the lowest price category for one bedroom properties (top left of Fig. 5). In this case, there are actually many available properties in areas where there are low levels of search. On further inspection, many of these properties are also in poor condition but this alone does not explain the mismatch and apparent low demand. To understand more about this, it is useful to look more closely at the data underlying the maps.

Taking two bedroomed properties as an example, Table 1 illustrates the relationship between drawn searches and available properties within London. In the lower price band, there is relatively little drawn search in relation to the available listings but as explained above many of these listings are not market-ready dwellings, so this is a somewhat anomalous situation. Nonetheless, the average number of drawn searches a property received in this category was over 90 during March 2013. At the other end of the spectrum, there was a much closer match between drawn searches and listings in the £250,000–£500,000 price band, with 407 drawn searches per property, on average. This helps identify which market segment is under most pressure in terms of demand at the initial stage of search.

The spatial distribution of this pattern is shown in Fig. 6, for two bedroom properties in the £250,000–£500,000 price band. In this map, which is divided into 100 m cells, only areas with properties are shown. There is an obvious ring of high intensity search closer to the centre of London, which then dissipates as distance from the centre increases. This matches the spatial price distribution of the wider London market and is therefore not surprising but viewed in this way we are able to understand much more, at a very fine spatial resolution, about the relationship between search extent and property availability for specific market segments.

The key point in this section is that, yes, there is clearly a spatial mismatch between search extent and housing stock characteristics. But, crucially, this varies across space and by market sector. The example of two bedroom properties in Table 1 demonstrates that in some price bands properties are the focus of quite intense levels of search and others much less so. In Fig. 6, we are then able to view the spatial distribution of these relationships.

This initial exploratory analysis between search extent and property availability has revealed two key features of the wider London housing market. First, we can see that on the whole people appear to be searching in the right places based on sectoral and spatial factors.

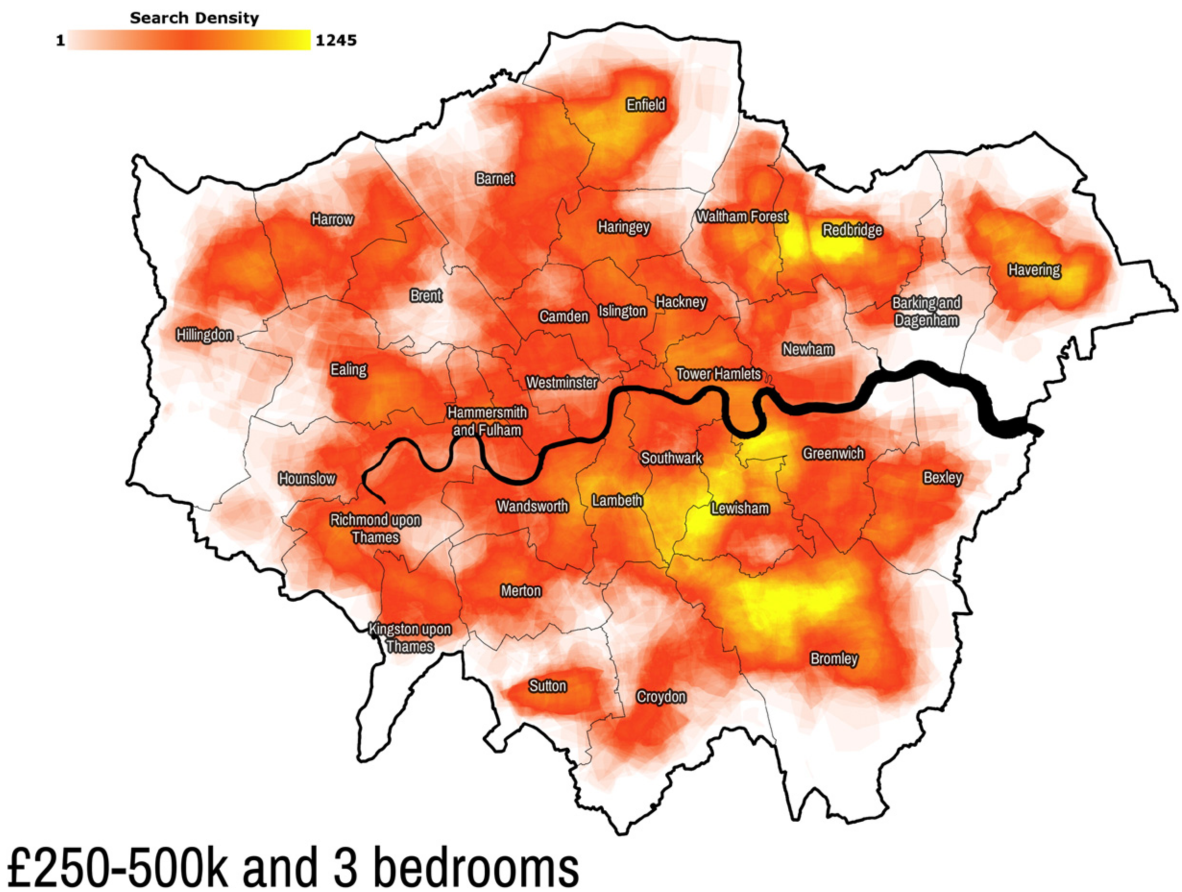

Fig. 4. Search intensity for 3 bedroom properties from £250,000 to £500,000.

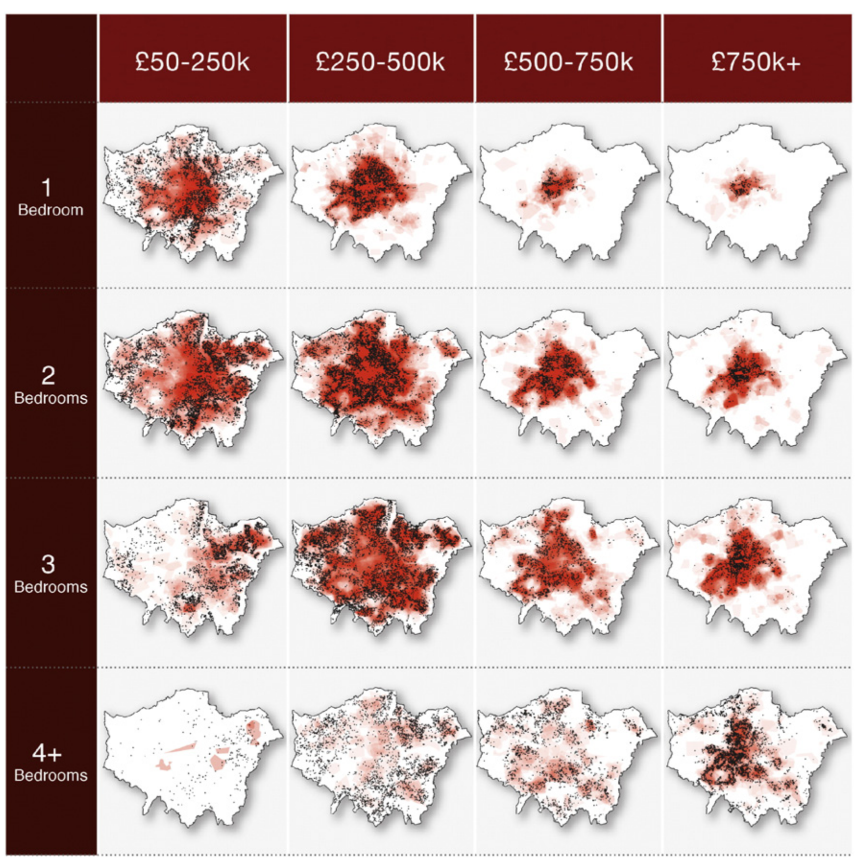

Fig. 5. London search vs location of available properties.

Table 1
Drawn search vs property availability (2 bedroom properties).

Price (£000)	Drawn searches	Listings	Drawn searches per property	Range
50–250	3168	10,783	91.6	0–525
250–500	10,562	11,302	407.1	0–1082
500–750	2950	2944	148.4	0–301
750 plus	2262	2655	222.6	0–402

Second, there are clear differences between search patterns associated with different price bands and property sizes. Further research is needed on other variables to test the extent to which we might call these 'submarkets' (Goodman, 1981) but it would appear reasonable to suggest that London is home to multiple, overlapping, spatially and sectorally differentiated submarkets (cf Watkins, 2001).

6. The utility of search data: learning points and caveats

The issue of what we can learn from studying search extent based on a large data sample generated by website users is a critical one and we are mindful of the important conceptual underpinnings provided in particular by Brown and Holmes (1971); Huff (1986); Hincks and Baker (2012) and Chen and Lin (2012). Here we identify three main lessons from which we hope further research in this field can be taken forward.

First, the longstanding question of whether a spatial or sectoral approach to understanding market geography is best (Watkins, 2001) appears to find an answer in this study owing to the way the technology now shapes early stage search behaviour. Varian's (2014) point about the mediating influence of computer technology in everyday transactions is particularly pertinent in the context of housing market analysis since the vast majority of first stage search is now conducted online and it is no longer necessary to decide upon what kind of property you want and *then* physically search different neighbourhoods. This can all be done at very low cost within the confines of users' homes. This means that in relation to search extent users are pushed to simultaneously think spatially *and* sectorally since online real estate portals lead users to specify multiple attributes on geography and property characteristics. This is qualitatively different from the way search fits into the established conceptual framework of housing search developed by Maclennan and Wood (1982) and later adapted by Marsh and Gibb (2011). Hence, the incorporation of search extent here is based upon Rae's updated model of housing search in the digital age (Rae, 2014).

A second learning point from the above analysis is the ability to identify spatial mismatches between search extent and the geographical availability of properties. This kind of information could be of practical use to a wide constituency of users, including local real estate agents, house builders, lenders, local authorities, and national government. For local real estate agents, the ability to identify listings which are in areas with low volumes of search could help with targeted property marketing and improve consumer outcomes in relation to property views. Conversely, it could also help identify high volume search areas where strong marketing is less critical. For house builders, this information could provide a spatially-specific, fine-grained view of where latent demand exists vis-à-vis where they might currently have available land. On the other hand, this could be particularly useful for local and national government in identifying any emerging areas of low demand, of the kind experienced in the north of England in the 1990s (Cameron, 2006). Lenders may also benefit from the availability of such intelligence in relation to current and future property valuation and the assessment of risk. In short, there are several possible beneficiary groups and each could profit in different ways from understanding where spatial mismatches exist.

A third learning point relates to the validity of using aggregate search data in housing market studies. Several authors have challenged the concept of using search data in determining wider housing market areas, since each household is likely to have a separate search area based on their existing location (e.g. Jones, 2002; Jones & Watkins, 2009; Hincks & Baker, 2012). The Jones and Watkins (2009) view was that the geography of search was unlikely to coincide with a housing market area. Nonetheless, at the relatively high spatial resolution of search areas shown here, we have demonstrated that the micro-level search patterns of nearly 100,000 website users combines to create spatially and sectorally consistent search geographies across London

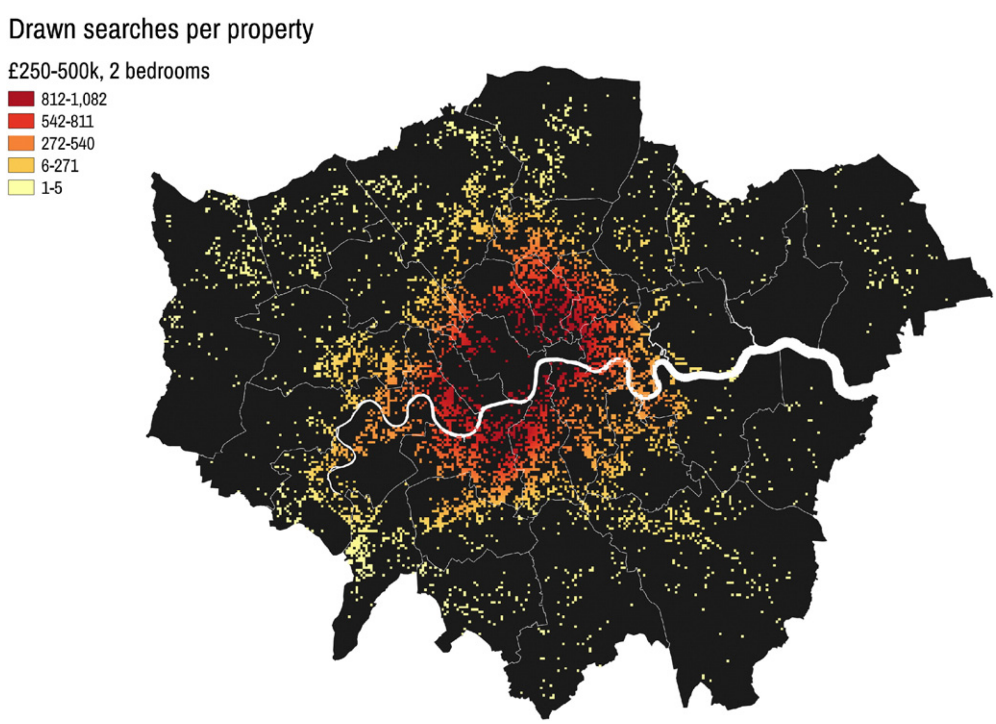

Fig. 6. Search density for 2 bedroom properties from £250,000–500,000.

(e.g. Fig. 3). It is certainly true that the boundaries of such areas, as Blank and Winnick (1953) argued, cannot be easily drawn with precision, so the emphasis here is on highlighting the fact that the geography of search extent in London, whilst messy, allows us to identify important spatial patterns associated with different price and property criteria. Although submarket identification is beyond the scope of this paper, we can also reasonably assume that there may be numerous, geographically overlapping submarkets in London and that this forms a multi-layered patchwork quilt, rather than a single identifiable set of areas or sectors.

In order to make sense of the contribution of this paper, we also need to make clear some important caveats. The first is that user-generated search data – and 'big data' in general – is inherently 'noisy'. The most obvious source of noise is that we don't know the intention of searchers and what proportion might simply be engaging in recreational search. However, we are particularly encouraged by previous studies (e.g. Choi & Varian, 2009; Wu & Brynjolfsson, 2009; McLaren & Shanbhogue, 2011) which have shown that even when using aggregate search data from the much less precise Google Trends, the results strongly suggest that search does foreshadow market activity. The recent work of Piazzesi et al. (2015) has reinforced the validity of such approaches, but we should of course remain cautious about the use of such data without further validation by local people such as real estate agents and potential buyers. This is now being undertaken in a series of interviews as part of our ongoing housing search project.

Another important caveat relates to the fact that this analysis has been undertaken using a single month of data from March 2013. The additional explanatory power which we could generate from taking a time-series approach would of course be valuable, and will be applied in future work in partnership with the data provider. Furthermore, we also acknowledge that price and property size are only two, albeit important, elements of the characteristics of individual properties. Future studies based on search data of this kind could usefully include other attributes (such as property sub-types) in order to increase the explanatory value of the results.

7. Future research

Our future work in this area will develop further the exploratory spatial data approach presented above, and aim to develop detailed metrics on search extent, with a view to deriving submarket areas. We also aim to focus on taking a time-series approach to understand how search volume and location varies over time, in London and the rest of the UK. We will attempt to define submarkets and larger housing market areas based on user-generated data and then compare these to existing knowledge on the geography of housing markets (e.g. Brown & Hincks, 2008). We also plan to examine in more detail the locations where search and property availability are mismatched. This could provide important policy and market intelligence of a kind which has never before been available. Finally, we aim to extend the economic aspects of the analysis to determine the nature of the relationship between search activity, sales volumes, price formation and local mortgage lending. The latter three of these are all now available as open data in the UK and offer exciting new possibilities for housing market researchers.

Acknowledgements

We wish to thank Alex Solomon and Iain Millar at Rightmove plc, the British Academy (Grant Reference SG120063) and the Erasmus Staff Mobility programme making this research possible.

References

Anselin, L. (1998). Exploratory spatial data analysis in a geocomputational environment. *Geocomputation, a primer* (pp. 77–94). New York: Wiley.
Beracha, E., & Wintoki, M. B. (2013). Forecasting residential real estate price changes from online search activity. *Journal of Real Estate Research*, *35*(3), 283–312.
Blank, D. M., & Winnick, L. (1953). The structure of the housing market. *The Quarterly Journal of Economics*, *67*(2), 181–208.
Brown, L. A., & Holmes, J. (1971). Search behaviour in intra-urban migration context: A spatial perspective. *Environment and Planning*, *3*(3), 307–326.
Brown, P. J., & Hincks, S. (2008). A framework for housing market area delineation: Principles and application. *Urban Studies*, *45*(11), 2225–2247.
Cameron, S. (2006). From low demand to rising aspirations: Housing market renewal within regional and neighbourhood regeneration policy. *Housing Studies*, *21*(1), 3–16.
Chen, C., & Lin, H. (2012). How far do people search for housing? Analysing the roles of housing supply, intra-household dynamics, and the use of information channels. *Housing Studies*, *25*(7), 898–914.
Choi, H., & Varian, H. (2009). *Predicting initial claims for unemployment benefits*. Google Inc.
Choi, H., & Varian, H. (2012). Predicting the present with Google trends. *Economic Record*, *88*(s1), 2–9.
Dunning, R., & Grayson, A. (2014). Homebuyers and the representation of spatial markets by information providers. *International Journal of Housing Markets and Analysis*, *7*(3), 292–306.
Flowerdew, R. (1976). Search strategies and stopping rules in residential mobility. *Transactions of the Institute of British Geographers*, *47-57*.
GLA (2004). *London's housing submarkets*. London: Greater London Authority.
GLA (2013). *The 2013 London strategic housing market assessment*. London: Greater London Authority.
Goodman, A. C. (1981). Housing submarkets within urban areas: Definitions and evidence*. *Journal of Regional Science*, *21*(2), 175–185.
Goodman, A. C., & Thibodeau, T. G. (2007). The spatial proximity of metropolitan area housing submarkets. *Real Estate Economics*, *35*(2), 209–232.
de Groot, C., Mulder, C. H., & Manting, D. (2011). Intentions to move and actual moving behaviour in The Netherlands. *Housing Studies*, *26*(3), 307–328.
Hincks, S., & Baker, M. (2012). A critical reflection on housing market area definition in England. *Housing Studies*, *27*(7), 873–897.
Huff, J. O. (1986). Geographic regularities in residential search behaviour. *Annals of the Association of American Geographers*, *76*(2), 208–227.
Jones, C. (2002). The definition of housing market areas and strategic planning. *Urban Studies*, *39*, 549–564.
Jones, C., & Watkins, C. (2009). *Housing markets and planning policy*. London: John Wiley & Sons.
Jones, C., Leishman, C., & Watkins, C. (2005). Housing market processes, urban housing submarkets and planning policy. *Town Planning Review*, *76*(2), 215–233.
Kroft, K., & Pope, D. G. (2014). Does online search crowd out traditional search and improve matching efficiency? Evidence from Craigslist. *Journal of Labor Economics*, *32*(2), 259–303.
KVA (2010). Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2010. *Markets with Search Frictions, compiled by the Economic Sciences Prize Committee of the Royal Swedish Academy of Sciences*.
Land Registry, H. M. (2015). *House price index*. Coventry: Land Registry.
Lynch, K. (1960). *The image of the city*. Cambridge, MA: MIT Press.
Maclennan, D. (1992). Housing search and choice in a regional housing system: New housing in Strathclyde. *A Report to the Housing Research Foundation, Glasgow*.
Maclennan, D., & Wood, G. (1982). Information acquisition: Patterns and strategies. *Modelling housing market search* (pp. 134–159).
Marsh, A., & Gibb, K. (2011). Uncertainty, expectations and behavioural aspects of housing market choices. *Housing, Theory and Society*, *28*(3), 215–235.
McLaren, N., & Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*, *2011*, Q2.
Palm, R. (1976). Real estate agents and geographical information. *Geographical Review*, 266–280.
Palm, R. (1978). Spatial segmentation of the urban housing market. *Economic Geography*, 210–221.
Palm, R., & Danis, M. A. (2001). Residential mobility: The impacts of web-based information on the search process and spatial housing choice patterns. *Urban Geography*, *22*(7), 641–655.
Piazzesi, M., Schneider, M., & Stroebel, J. (2015). *Segmented housing search*. National Bureau of Economic Research.
Pope, J. C. (2008). Buyer information and the hedonic: The impact of a seller disclosure on the implicit price for airport noise. *Journal of Urban Economics*, *63*(2), 498–516.
Rae, A. (2014). Online housing search and the geography of submarkets. *Housing Studies*, 1–20 (OnlineFirst).
Reed, R. (2015). Examining current trends in global house prices. *International Journal of Housing Markets and Analysis*, *8*(3).
Varian, H. R. (2014). Big data: New tricks for econometrics. *The Journal of Economic Perspectives*, *3-27*.
Watkins, C. (1999). Property valuation and the structure of urban housing markets. *Journal of Property Investment & Finance*, *17*(2), 157–175.
Watkins, C. A. (2001). The definition and identification of housing submarkets*33*, 2235–2253.
Wheaton, W. C. (1990). Vacancy, search, and prices in a housing market matching model. *Journal of Political Economy*, *98*(6), 1270–1292.
Wu, L., & Brynjolfsson, E. (2009). The future of prediction: How Google searches foreshadow housing prices and quantities. *ICIS 2009 Proceedings, 147*.