Online housing search and the geography of submarkets

Abstract

The importance of search behaviour has long been recognised in the study of housing markets but research in this area has frequently been hampered by a lack of data. In many nations, the vast majority of initial housing search queries are now conducted online data this generates could, in theory, provide us with better insights into how housing market search operates spatially, in addition to generating new knowledge on the geography of local housing submarkets. This paper seeks to explore these propositions by discussing existing conceptions of search before developing a framework for understanding housing search in the digital age. A large, user-generated housing market search dataset is then introduced and analysed with respect to area definition, submarket geography and search pressure locations. The results indicate that this kind of 'big data' approach to housing research could generate important new insights for housing market analysts.

1. Introduction

The question of how people find housing is a long-standing subject of study in housing market research (e.g. Maclennan, 1982). However, the existing literature has tended to focus on outcome indicators such as price changes and migration patterns, rather than on the search activity of potential buyers. This imbalance has often been attributed to a lack of data (e.g. Jones and Watkins, 2009; Sun and Manson, 2010) but in recent years the ways in which people begin to search for housing have fundamentally changed and this has created opportunities for much richer analyses of search activity. Whereas housing search previously began with a visit to an estate agent's office or by looking at the property pages of a local newspaper (Palm, 1976), the initial stages of housing search are now primarily an online activity via various market portals, including Trulia in the United States, Funda in The Netherlands, ImmobilienScout24 in Germany and Rightmove in the UK.

This situation is replicated in a multitude of national contexts worldwide but little is known about how this initial extensive search process operates spatially and how this might connect to existing knowledge on housing search and the definition of housing submarkets, which are typically defined in relation to housing characteristics, or discrete spatial units. Such matters are a key consideration in analysing, understanding and planning for housing at the sub-national level. Research by Watkins (2001), Jones (2002), Brown and Hincks (2008) and Royuela and Vargas (2009) has advanced knowledge in this area but, to date, approaches have tended to focus on what is referred to as 'effective' or 'revealed' demand; that is, where people actually move to rather than where they might want to move to. A notable exception to this is Maclennan and O'Sullivan's 2012 work on housing markets, signals and search which is based on a 1990 housing search survey in the Strathclyde region of Scotland. However, the authors note that with the establishment of widespread access to the internet, 'the ratio of 'passive' to 'active' search behaviour may have changed significantly over time' (Maclennan and O'Sullivan, 2012, p. 330).

The focus in this paper, instead of looking at 'revealed' demand or using surveys, is on understanding the spatial dimensions of search through user-generated data, sourced from rightmove.co.uk, the UK's most popular online property market website (Google Trends, 2013). This represents a first step towards integrating search data into our understanding of housing market search dynamics. Given the ease with which people can search online for housing, however, there is a need to carefully theorise 'search' within the digital context and to construct a conceptual framework within which it can be understood. In an attempt to properly position this research within its wider context, the next section of the paper reflects upon the existing literature on housing search and how it has been conceptualised to date. Following an examination of the existing literature, a theory of housing search in the digital age is proposed. After this initial theorisation, housing search data are explored in relation to how users define their search areas, the geography of submarkets, and search pressure locations. The paper concludes by reflecting on the potential benefits and challenges associated with using this kind of 'big data' in housing market research.

2. A theoretical basis for approaching online housing search

Research on housing market search activity is, in many respects, theory rich but data poor. The current study attempts to redress the balance in a small way by introducing and analysing a new dataset. However, it is first necessary to consider the conceptual and theoretical lineage of the topic so this section focuses on developing a more nuanced understanding of information acquisition in the housing market and on understanding the spatiality of search from a theoretical perspective. This follows Marsh and Gibb's recent contribution to the literature on housing market search, which emphasised the 'centrality of information acquisition in the process of choice' (2011, p. 218), and Maclennan and O'Sullivan's assertion that 'research on the spatial dimensions of housing market search has been especially slow to develop' (2012, p. 328).

2.1 Information acquisition in the housing market

In a 1976 paper on real estate agents and geographical information, Risa Palm noted that 'home buyers have a limited amount of time and resources which they are willing to expend in their search for a house' (Palm, 1976, p. 268). This fundamental time and resource pressure has remained, yet technology has advanced considerably in the intervening years. In 2001, as online sources of housing market information began to emerge, Palm and Danis (2001) concluded that, despite the apparent potential of the web as a source for housing information, the internet had little impact on search patterns. Revisiting the issue in 2013 this is no longer the case and in nations such as the US, The Netherlands, Germany and the United Kingdom, the internet is now the first port of call for most consumers of housing. Web analytics data from Google Trends for searches in the United Kingdom since 2004, for example, show a marked decline in searches for traditional residential property information channels (estate agents) compared to the two leading UK online property portals (Figure 1), with Rightmove particularly dominant.

Figure 1 – UK internet search terms, indexed to 100, for 'rightmove', 'zoopla' and 'estate agent'
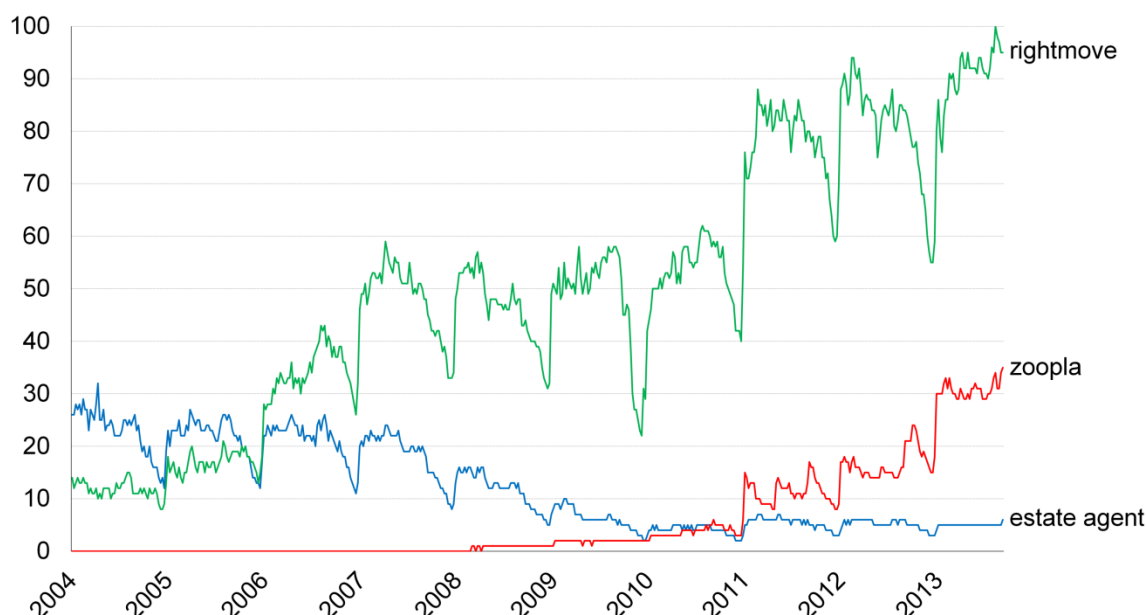
A question therefore arises as to whether this move towards web-based information acquisition necessitates a re-examination of housing market search theory and whether existing notions of 'search' need to be revised. A useful starting point in tackling this question is Clark's (1982) edited volume on modelling housing market search. Following similar studies in psychology, marketing, economics and geography, this represented a landmark contribution in attempts to understand individual decision making processes as they pertained to housing market search. From this volume, it is the conceptualisation of 'search' itself which is initially most intriguing; particularly in light of the dataset introduced below.

At the outset, Clark and Flowerdew (1982, p. 9-11) propose five separate concepts of search: i) that search is goal-directed and undertaken by people who have some idea of what they want; ii) that search involves a complex process of information gathering from different channels; iii) that the searcher is never aware of the complete set of opportunities; iv) that there must be some way to determine when the search ends and a choice is made; and v) that search takes place subject to a set of constraints, including money, time and access to information. They also propose a temporal search sequence, which proceeds from newspapers, through agents, friends and then to relatives. These concepts still have considerable currency, but the ubiquity of online information sources in the housing market means reappraisal is necessary if we are to more fully understand allocative efficiency in the contemporary housing market.

In relation to Clark and Flowerdew's five-fold conceptualisation, search is often still goal-directed but with the proliferation of recreational search (so-called 'property porn', cf. Botterill, 2013) there also now exists a much larger volume of seemingly superficial search activity which should not be classed as 'search' in the same way that the initial 'extensive' search activity of market entrants is (cf. Maclennan, 1982). However, this recreational search could in fact have

the effect of modifying household aspirations in the short to medium term, or provide longer term aspirational targets, and is therefore not necessarily trivial. In fact, this process is similar to what Maclennan describes as 'initial aspiration formation' (1982, p. 71). The concept of search as a complex process of information gathering from different channels is still valid, though the mix has come to be dominated by online search (Dunning and Watkins, 2012). In the case of online property portals, market coverage in England is now so extensive that the searcher will in many areas be able to view a complete set of opportunities. For example, all English estate agents with 100 or more branches list their properties on Rightmove, and 18,270 estate agents advertise properties in this way (over 90% of the total; Experian Hitwise, 2013). Determining when search ends, however, remains a more intractable problem and the increase in information is perhaps less helpful in this respect. In relation to Clark and Flowerdew's fifth concept, money and time may remain constraints whereas *excess* information rather than *access* to it may now be a bigger problem.

Broadly speaking, then, Clark and Flowerdew's conceptualisation of search remains valid, even with the advent of highly efficient new information acquisition channels. However, it is worth questioning whether it is true that the advent of the internet, as Marsh and Gibb claim, (2011, p. 218), '*in no way* removes the fundamental problems facing potential purchasers' (emphasis added). It has been convincingly argued that certain distinctive features of the housing market make standard consumer choice theories problematic (Maclennan, 1982, p. 60-62) but with respect to imperfect information, the practical difficulties of evaluating spatially dispersed housing vacancies and the associated 'psychic costs of housing search' (Maclennan, 1982, p. 61) online housing search portals could be said to significantly reduce, and in certain instances eliminate, the impact of these problems. Most pertinently in the case of spatially dispersed housing vacancies, users of online housing search portals now have access to an almost complete national database of residential vacancies, with more than 1 million properties available to buy or rent via Rightmove at any one time. At the initial stages of search, therefore, a person living in Dundee can easily establish a list of relevant vacancies in Dover (over 500 miles away) in a way that simply was not possible when Maclennan conceived his seven features of the housing market and housing as a commodity in 1982. Information in the housing market is still imperfect but, one could argue, at the initial stage of search it is significantly less imperfect than it used to be. This has implications for how we should conceptualise online search in the contemporary housing market.

As with the fundamental conceptualisations of search, recent developments in the availability and volume of online housing market information also make a re-assessment of the *process* of search an important task. In Maclennan and Wood's (1982) 'chain of search stages', they state that the available evidence 'does not conclusively support the notion of "search-stages"' and that it is based upon their '*a priori* notion of the housing search process' (p. 157). Nonetheless, the process of search that they forward is both convincing and comprehensive and its recent simplification by Marsh and Gibb (2011) provides a useful classification with which to contemplate both the online information acquisition process and the large user-generated search dataset described later in this paper. Marsh and Gibb identify six components of the search process that users are likely to engage in once they have decided to enter the market for

owner-occupied housing. The first three (search strategy, area orientation, establish vacancies) are most relevant to the current study since they are directly linked to the online, extensive search phase. The latter three (personally visit vacancies, evaluate in detail, form and place bid) are less relevant here, though in most cases it is also now possible to evaluate potential properties in considerable detail using online housing portals.

Marsh and Gibb (2011, p. 218) explicitly state that 'households do not necessarily proceed through these stages sequentially' and that households will not always proceed to the bidding stage (cf. Maclennan and Wood, 1982, p. 157). The contention here is that, in the context of online search, these stages are not necessarily separate and that there is an overlap in particular between area orientation and establishing vacancies. This concept is developed further in the next section of the paper but the point is that the accessibility of housing market information via online property portals, in addition to the inherent spatiality of such platforms (based, as they are, on maps and geographical search queries) makes it possible for consumers of housing to acquire vast amounts of information at very low cost in a way that was not possible even a decade ago. However, with the rise of recreational search, there may be some additional complications to this standard choice process in that the decision to move, the assessment of available resources and the specification of housing requirements could also be part of an initial, non-active housing search process. These three elements are conceptualised by Marsh and Gibb (2011, p. 230) as being outside the housing search process but in a model of housing search which includes online portals this may need to be reconsidered.

The question raised at the beginning of this discussion was whether the dominance of online sources of housing market information necessitates a re-examination of housing market search theory and whether conceptions of 'search' need to be revised. The answer proposed here is 'yes' but in relatively subtle ways. This is the subject of the next main section of the paper. Prior to that, however, the spatiality of search is approached from a theoretical perspective. This is a particularly important issue since housing markets are inherently spatial and, as Blank and Winnick observed six decades ago, housing markets are 'local markets, distinct from one another, although it is sometimes not easy to draw the boundary lines with any measure of precision' (Blank and Winnick, 1953, p. 184).

2.2 The spatial dimensions of housing market search
Following Maclennan and Tu (1996), this study recognises that the definition of housing market areas, and submarket areas in particular, can be both sectoral and spatial. As and Watkins (2001) notes, however, robust submarket definition has historically been problematic and no single definition of a housing submarket exists. The focus here is on definitions of submarkets as discrete spatial entities and the data introduced in the paper represent an attempt to empirically test some of the assumptions surrounding how people actually define their housing search areas. The fundamental starting point for such a study must be a recognition of the spatial fixity of the housing stock (Maclennan, 1982; Meen, 2001) and the necessity for people to employ spatial search strategies when they wish to move. Therefore Huff's assertion that 'each dwelling has a location and the set of locations corresponding to the set of vacancies

seen constitutes a search pattern' (1982, p. 106) is particularly apposite. Huff then goes on to identify three specific issues relating to spatial search: i) length of search and the number of alternatives households consider; ii) the role of information in the search process; and iii) the spatial aspects of search – where will households look for a new residence? The latter two are most relevant to the current study, with iii) in particular being appropriate for further investigation in relation to the user-generated search dataset introduced below.

A problem with the way existing models of spatial housing market search deal with information in an era where online search now dominates the initial, extensive search phase is succinctly captured by a quote from Wheaton's seminal housing market matching model, in the section on 'search technology'. He states that 'search is necessary because there is imperfect information, possibly about which units are for sale and certainly about a unit's type' and that 'in effect, advertising is assumed to be either non-existent or imperfect' (Wheaton, 1990, p. 1278). Later, Meen (2001, p. 49) comments that search costs 'limit the spatial range of the search'; a statement which is certainly true of in-person intensive search but no longer the case for the millions of people who regularly use online property portals. As with some of the original assumptions about the early stages of information acquisition in the housing market, then, it is worth revisiting conceptions of spatial search in the housing market.

In one sense, spatial questions relating to housing market search are no different now than they were in 1976 when Palm looked at the role of real estate agents as information mediators in Minneapolis and San Francisco. At this time, as now, housing markets were spatially and sectorally segmented and potential buyers would search for varying periods of time, based on a range of information sources, in specific locations. Now, however, the practicalities of spatial search have changed. Whereas, in 1982 (p. 109), Huff wrote that 'the early stages of search are often characterised by a greater dependency on newspaper advertisements and driving around', the initial stages of search are today substantially different as new ways of interacting with housing search options have emerged. For example, online information acquisition channels only increase the ability of searchers to construct their own non-contiguous search area within wider metropolitan context. This phenomenon was recognised previously by Goodman and Thibodeau (2007) but the search tools offered by online housing portals increase the likelihood that spaces for the consumption of housing are being constructed in ways that were inconceivable when the predominant models of housing search were conceived.

In light of the subject of the current study, therefore, two observations in particular are worth emphasising. First, the role of spatial information in the housing search process is now qualitatively and quantitatively different than it was even a decade ago. This does not render current models of search obsolete, but it does implore us to revisit the conceptual foundations of housing market search. The second point is that the rise of online housing market portals has created a potentially unlimited spatial canvas upon which people can draw their own housing search area boundaries. The data presented later in the paper allow this proposition to be tested for the first time and reveal that, in fact, the freedom to define search areas merely helps validate some of the key assumptions in relation to the spatial segmentation of housing market

search (Watkins, 2001). Prior to the empirical contribution, a framework for understanding housing search in the digital age is now proposed.

3. A framework for understanding housing search in the digital age

The contention here is that the staged approach to understanding housing market search is still valid, but that the nature of information acquisition has changed so significantly in the last decade that some revisions to accepted housing search models are necessary. These changes have been described in more detail above so the intention now is to forward a modified framework within which the data introduced in the next part of the paper makes conceptual and empirical sense. This builds, in particular, upon Maclennan's account of staged search (1982) and Marsh and Gibb's (2011) six stage search process. The conclusion is that there are some important substantive differences between traditional information channels and contemporary search methods.

One of the most obvious issues that must be dealt with before proceeding is the meaning of 'search' in the digital domain. This issue was highlighted more generally in a Bank of England *Quarterly Bulletin* of 2011 which discussed using internet search data as an economic indicator. In this piece, McLaren and Shanbhogue (2011, p. 135) note that 'a lot of searches will be purely out of curiosity' and that 'there is often significant noise in the search data'. Such concerns do not render online housing search data meaningless but it does mean we have to approach it carefully. It has already been established that the focus here is on the initial stage of engagement with the housing search process (i.e. 'extensive' search) but equating web-based housing search queries with empirically meaningful housing search behaviour is of course rather more challenging.

For example, how does one know whether a web-based search for a three bedroom semi-detached house within a particular area represents the initial preferences of a potential buyer rather than the aspirations of a long-term renter? Similarly, we might ask whether most people searching for seven bedroom detached houses in very expensive areas across the country are engaging in meaningful 'search' activity or simply indulging themselves in a form of recreational search activity of a kind which was impossible before the advent of online housing portals. No definitive answers are provided now but from a practical point of view a distinction can be made between five different types of online housing search. Understanding these different kinds of search activity is a necessary first step prior to analysing online housing search data in a more meaningful way.

- **Recreational search:** searchers in this category have no immediate intention of transacting in the housing market and are typically focused on looking at properties positioned at the very apex of the market in terms of price, size and location. This could in theory form a part of Maclennan's initial aspiration formation process but it seems more likely that the vast majority of search activity in this area would be frivolous. *Example:* a searcher on an average income may decide to look at all properties in an

area which cost £2m or more and then share details of these properties with friends via e-mail or social media such as Facebook and Twitter.

- **Inquisitive search: s**earchers in this group have no intention of buying and are motivated by inquisitiveness in relation to the features and dimensions of individual properties and a desire to 'see what's out there'. With online search, this activity is likely to have expanded significantly and allows individuals to make a remote inspection of any property they are curious about from their computer or mobile device. One could postulate that this kind of search merely represents a form of snooping or nosiness but as with recreational search it could also feed into the aspiration formation process and, more significantly in this case, give potential buyers a greater understanding of the housing market. *Example:* a searcher sees a number of properties in their street with a 'For Sale' sign up and decides to go online and look at the listings and view the property photos, just to see what they look like from the inside.

- **Aspirational search:** searchers in this category are not currently looking to buy a property but are instead engaging in aspiration formation with a view to either entering the housing market for the first time, or moving to a more expensive property. A searcher would typically use an online housing portal to make a judgement on the gap between their currently available housing finance and some kind of 'ideal' property which they cannot yet afford. They could quickly form or alter their aspirations based on a series of searches with slightly modified search criteria. *Example*: a searcher currently lives in a two bedroom terraced house with a market value of £125,000 and they wish to establish what kind of property they could buy if the finance were in place for properties in the region of £200,000. This kind of search is neither recreational nor simply inquisitive but it is not yet functionally active.

- **Active search**: this category applies to those searchers who have decided to actively look for new housing because they actually wish to move and have the financial means to do so. Only people in this category would be considered to be engaging in 'search' activity in the traditional sense. Searchers in this category therefore typically turn to online housing market portals at the very first stage of extensive search and use the portal's location and property-attribute tools to refine their search. *Example*: a couple with a combined annual income of £70,000 and a £30,000 deposit decide that the time is right to purchase their first property. They register their details with the online property portal and begin looking at three bedroom semi-detached properties in the £200,000 to £220,000 price band. They subsequently continue to refine their search based on area and property characteristics. After identifying a number of suitable properties, they then send a series of e-mails direct to estate agents via the property portal website.

- **Professional search:** this category of searchers is populated not by potential buyers but by property professionals engaging in search for a number of different reasons. For example, estate agents are able to efficiently and effectively compare their listings to those of local competitors, surveyors are able to view property details instantly for a wide

range of areas and property types and developers are able to very quickly assess the state of the market in terms of asking prices and the volume of stock in particular locations.. *Example*: a local, independent estate agent wishes to view all current listings in Nottingham held by a rival estate agent. They can do this at the click of a mouse via the Rightmove website.

These search categories suggest, of course, that one needs to be careful when interpreting data on housing 'search' generated by online housing portals. In this regard, two particularly important issues are of relevance here. First, there is clearly a need to find some way to filter online search data so that any analysis of 'search' is as representative as possible of active online search which might eventually lead to a transaction in the housing market and a physical move of one or more individuals. Such a filtering process could also make possible more meaningful studies of other kinds of search activity which, while seemingly frivolous in some cases, may actually play an important role in long-term aspiration formation. The second issue is the extent to which spatial perceptions of areas may remain the same between different search categories. The purpose of search may be qualitatively different between categories, but at the same time searchers might realistically be identifying areas using similar geographical identifiers (e.g. postcode sectors) or by drawing areas on a map on the housing search portal, regardless of their underlying search intentions.

This research is principally concerned with the group of searchers identified above as being 'active', and the framework of search identified below relates to this group. It is based on an adaptation of previous conceptualisations by Maclennan and Wood (1982) and Marsh and Gibb (2011), extensive discussions with Rightmove housing market experts, and in-depth analysis of the search data upon which this paper is based. The framework set out in Figure 2 recognises that prior to housing search or a decision to move there is often some kind of external trigger which motivates people to search. However, it is also possible that the process of online housing search itself could act as a trigger.

The relationship between the various components is explained in Figure 2, but an important feature of this new conceptualisation is that the process of online housing search is seen as being comprised of three linked elements which were previously identified as separate components of the search process by Maclennan and Wood (1982, p. 136) and Marsh and Gibb (2011, p. 230). Housing portals such as Funda, Trulia and Rightmove allow users to simultaneously select areas, establish vacancies and specify housing requirements during the search process. This is not always necessary in order to perform a search but with online search in particular the vast majority of users must make these choices to avoid the problem of the search engine returning too many properties to view. The following scenarios, shown in Figure 2 and described below, provide examples of how the online search process may operate in practice.

Figure 2 – A framework for understanding online housing search
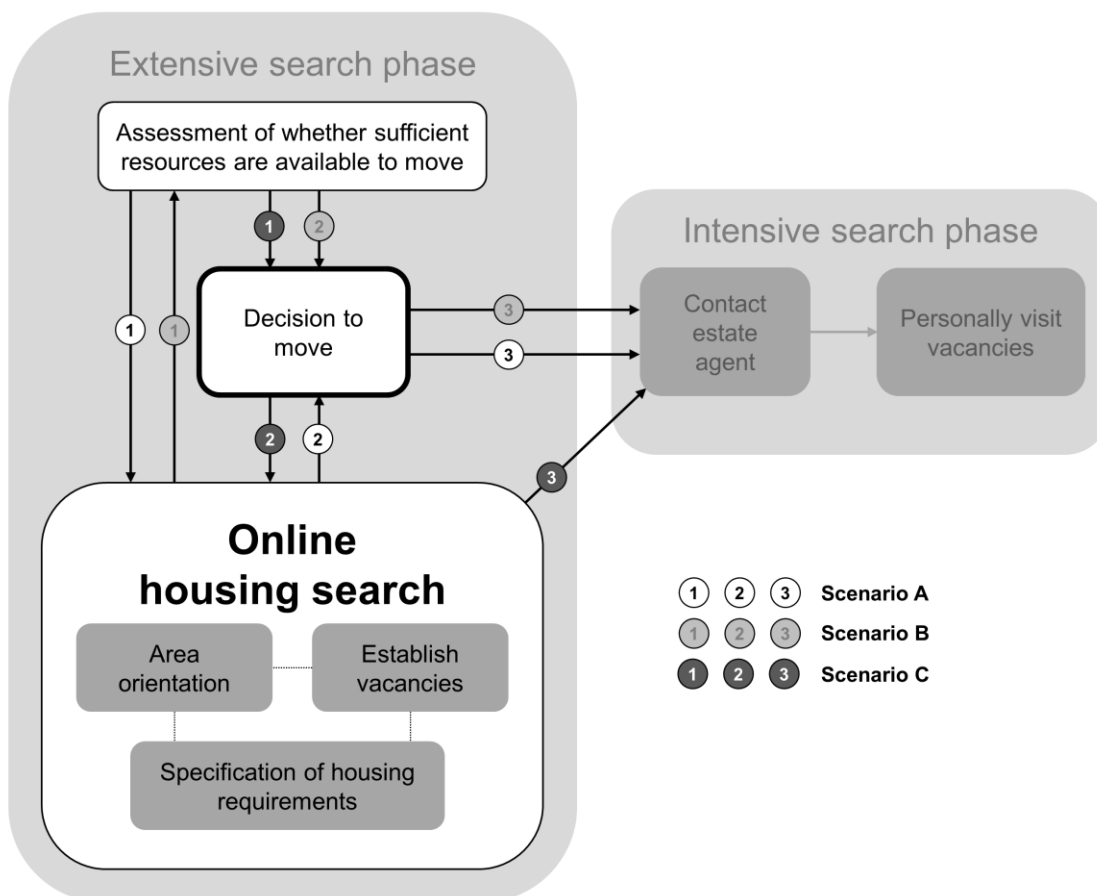
- **Scenario A:** A household assesses whether they have sufficient resources to be able to move. They do not yet make the decision to move but instead seek information from an online housing portal in order to find suitable areas, establish vacancies and specify housing features. This then leads to a decision to move, which in turn leads to the household contacting an estate agent and personally visiting properties.

- **Scenario B:** On the other hand, a person searching in a more aspirational manner may – after a period of extensive online search – subsequently assess whether they actually do have enough resources to be able to move. This could then lead to a decision to move and contact with an estate agent with a view to visiting properties.

- **Scenario C:** A third possibility is that a household makes an initial assessment of whether they have sufficient resources to move, then takes a decision to move before proceeding to online search and then contacting an estate agent and visiting properties.

Having established an underlying framework for understanding housing search in the digital age, the focus now shifts to the analysis of a large user-generated dataset sourced from Rightmove plc. These data represent the outcomes of an extensive search process in which area orientation, the specification of housing requirements, and the establishment of vacancies happen almost simultaneously online. The question of whether such searches result in a move to the intensive search phase is not addressed here but will be the subject of a future study based on the same data source.

4. The geography of online housing search

This paper draws upon search data sourced from Rightmove's 'Draw-a-Search' tool. This allows rightmove.co.uk users to draw their own housing search area on top of a Google map within the rightmove.co.uk website and then view available properties from that area which meet their search criteria. There are no restrictions on the shape or size of areas that users can draw. In the example screenshot in Figure 3, a user has drawn a search area within Coventry to find three bed properties in the price range from £200,000 to £250,000. A total of 58 search results are returned and the user can then explore these options. Each time a user draws a search area on the website, a geographic polygon is saved to Rightmove's web servers. This paper is based on the analysis of 808,870 such polygons, generated between 17 February and 16 March 2013. The time period of this data extract coincides with a high volume of search queries at a point in the annual cycle when many potential buyers begin to engage with the housing market. The rest of this section provides more detail of how users engage with 'Draw-a-Search', how this might be useful in identifying housing submarkets and where the areas of greatest search pressure currently are.

Figure 3 – Drawn search area example from rightmove.co.uk

## 4.1 How do users define their search areas?

During the time frame covered by the data sample, an average of 28,888 polygons were drawn each day. The complete freedom users have with Draw-a-Search means that search areas can be as detailed or as general as an individual wishes. In technical terms, the mean number of vertices people draw for each polygon is 16, with a median value of 11. In terms of geographic scale, the largest single category of search area (13.2%) is less than one square mile in size. Figure 4 provides an illustrative example of six drawn areas ranging from less than one square mile (a, b), two to five square miles (c, d), and ten to thirty square miles (e, f). These examples are typical of the search dataset as a whole in that they demonstrate the high level of spatial specificity employed by searchers when drawing areas and the generally smaller spatial scales people engage with during the area orientation phase of online search.

Figure 4 – User-generated search areas from across the UK

a) Lenzie, near Glasgow
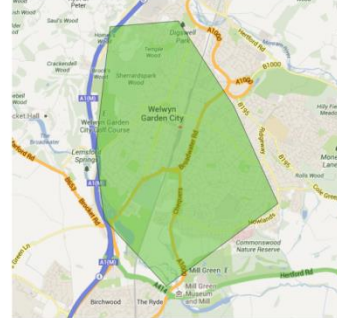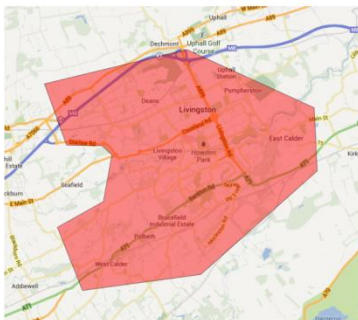
b) Halfway, Sheffield
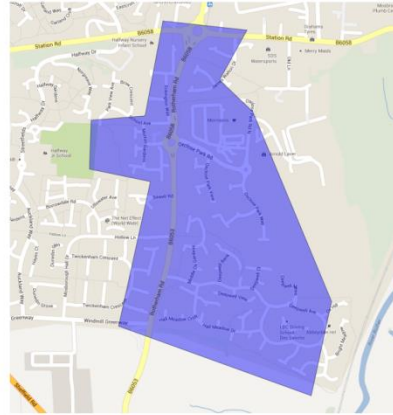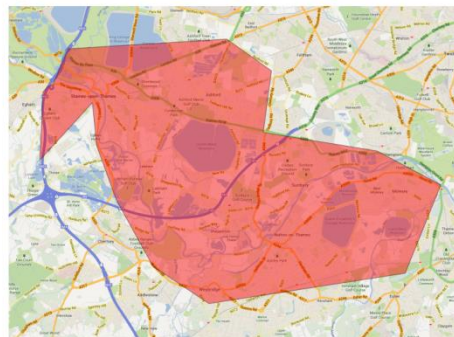
c) Prestatyn, North Wales

d) Welwyn Garden City

e) Livingston

f) Staines, Surrey

In addition to being able to accurately define their own search areas, users can also specify price and property details when using Draw-a-Search. Consistent with the conceptualisation of online search presented in Figure 2, these occur at the same time as area orientation. Of the

808,870 search areas in the data sample, 37% of users entered no maximum price threshold, but the distribution of those who did is illustrated in Figure 5 below. There are clear price break points at £100,000, £150,000, £200,000, £250,000 and £300,000 and out of those who identified a maximum price, £250,000 represents the modal value.

Figure 5 – Price criteria entered by Draw-a-Search users
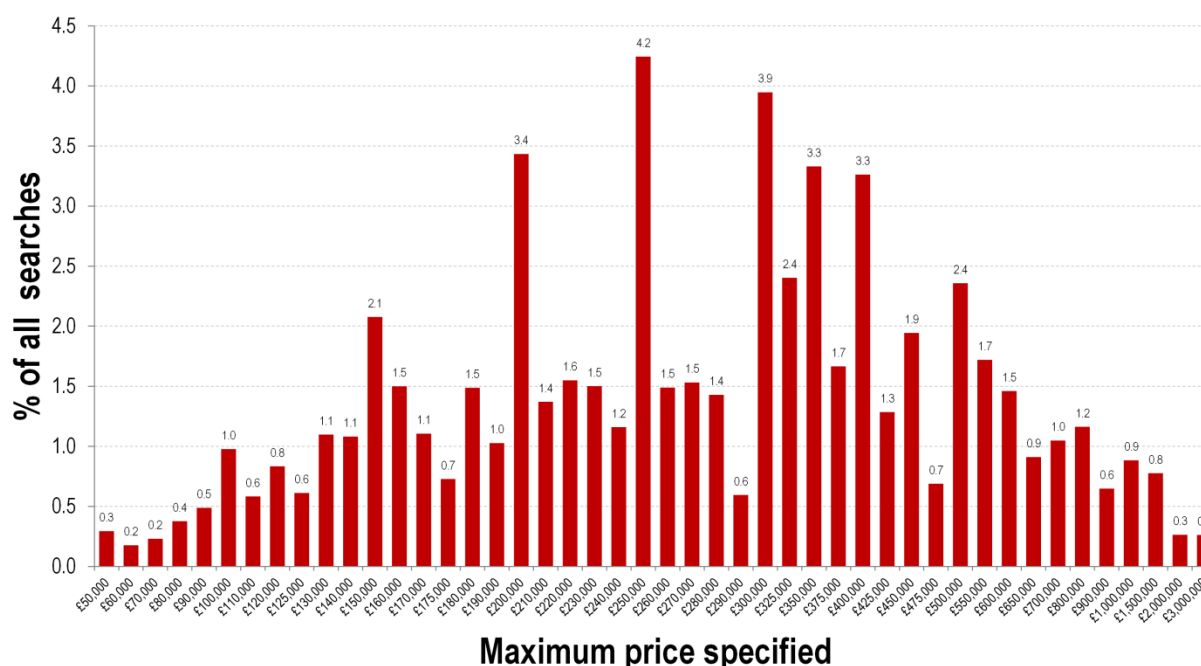
In terms of property characteristics, the minimum number of bedrooms is the next most common attribute entered by website users. A total of 51% of searchers did not specify a minimum number of bedrooms but of those who did, two and three bedroom properties are identified most often, at 17.4% and 22.1% respectively. Table 1 illustrates the differing characteristics of search, based on the minimum number of bedrooms specified by Draw-a-Search users. The median search area for all drawn searches was 26.3 square miles, but this differs according to property size, whereby the higher the number of bedrooms specified, the larger the search area becomes. For one bedroom properties, for example, the median search area covers just less than 10 square miles, but for four bedroom properties it is over 70 square miles. It is only at the five bedroom level that the search area (253 square miles) extends beyond the spatial scale of a typical town or city (e.g. the Birmingham City Council area covers 103 square miles).

These summary statistics on housing characteristics and area orientation could prove a particularly useful supplement to recent work on housing market definition in the UK and on the identification of housing submarkets in particular (e.g. Jones et al. 2012). Indeed, the data in Table 1 reinforce the idea that the spatiality of housing search can best be expressed through a

multi-layered, non-nested approach to housing submarkets where there is a relationship between income constraints, property types and spatial search patterns (Watkins, 2001; Hwang and Thill, 2009). This proposition is explored in more detail in the next section.

Table 1 – Summary statistics for Draw-a-Search queries

Bedrooms	% Total	Av Max Price (£)	Av Min Price (£)	Range (£)	Median Area (SQM)
Not specified	51.1	134,840	49,658	85,181	23.1
1	2.6	274,789	97,527	177,261	9.9
2	17.4	280,832	95,790	185,041	23.2
3	22.1	327,126	133,802	193,324	30.9
4	5.9	404,530	192,904	211,626	70.7
5	0.9	418,094	362,004	56,089	253.6

4.2 User-generated search areas as submarkets

The housing search data presented here allows us to begin to test for the existence of housing submarkets, as opposed to larger, more all-encompassing housing market areas of the kind identified in previous studies using administrative geographical units (Goodman and Thibodeau, 1998; Bourassa et al., 2003). In the past, submarkets have been segmented spatially using pre-existing statistical or administrative boundaries (e.g. Adair et al., 1996) or sectorally using an information-based approach which takes into account such factors as estate agent knowledge and property characteristics (e.g. Palm, 1978; Allen et al., 1995). The dataset presented here could provide us with new insights in relation to both the sectoral and spatial delineation of the market from a more bottom-up perspective. This proposition is examined below in relation to the geography of housing market search in London by comparing search patterns to an existing delineation of localised housing market areas delineated by a previous study (Jones et al, 2012).

A fundamental question, of course, is how one might define 'search' in the context of a low cost, high volume information channel like Draw-a-Search. In an attempt to identify the preferences of active searchers, the analysis in this section is based on a subset of 173,059 search polygons drawn by Rightmove.co.uk users who were registered with the website and who provided personal details, including their full postcode. Whilst imperfect, such a categorisation was considered a best-proxy for 'active search' after consultation with Rightmove analysts and further analysis of search informatics studies (Dunning and Watkins, 2012). This sub-group of searchers register with Rightmove in order to save previous search areas and queries, to get property alerts and to receive property news. As such they are significantly more likely to be engaged in meaningful housing 'search' activity, as per the 'active search' group identified above. The difference in activity and search attributes between this sub-group and the data

sample as a whole could also provide useful insights in relation to the level of 'noise' in the dataset and provide an indication of the different spatial perceptions of active versus non-active searchers.

In Table 2, we can see clear differences in search attributes between property types. Once again, just over half of all searchers do not specify a bedroom size but of those who do the highest percentages are for two and three bedroom properties, at 15.3% and 24.1% respectively. Search areas also increase according to property size, as they did for the full, unfiltered sample used in Table 1. With the exception of 5 bedroom properties these search areas are again considerably smaller than most local authority areas (e.g. the City of Edinburgh covers 102 square miles). However, these data will undoubtedly vary across the country, given the differing local and regional housing markets context of the UK (Ferrari and Rae, 2011). Therefore, it makes sense to conceptualise submarkets as being regionally rather than nationally specific.

Table 2 – Summary statistics for Draw-a-Search queries (registered users only)

Bedrooms	% Total	Av Max Price (£)	Av Min Price (£)	Range (£)	Median Area (SQM)
Not specified	52.0	118,891	43,124	75,767	41.2
1	1.8	304,530	12,773	191,757	12.3
2	15.3	287,097	103,758	183,339	29.6
3	24.1	341,646	140,286	201,360	34.4
4	6.0	419,166	187,431	231,736	71.5
5	0.7	487,504	264,370	223,134	115.9

As an illustrative example of the importance of sub-regional differences in housing submarkets, Table 3 presents price-stratified search data for two bedroom properties in London. Unlike the UK as a whole, the largest single property size specified by searchers is for two rather than three bedroom properties, with a total of 26.9% of London searches. Of this group, just over half search for properties in the £250,000 to £500,000 price band and the average maximum price specified is £381,542. This figure is more than £200,000 higher than the national average for this property type. In addition to this striking price differential, the search areas specified by users are significantly smaller than the national average, with median values of between 2.8 and 5.5 square miles.

This initial evidence suggests that, in contrast to the five London housing submarkets identified by Varma (2004) on a structural basis, the composition of submarkets in London may be considerably more complex, both sectorally and spatially. More work in this area is needed in order to fully test the implications of search data in relation to the identification of housing

submarkets. However, the data here do suggest that, following Watkins (2001), the dimensions of housing submarkets are influenced by both spatial and structural factors; particularly when users engage with a search information channel where spatial and structural characteristics are defined simultaneously. The crucial point here is that during initial engagement with the housing market at the extensive search phase, the preferences of searchers are often highly spatially and sectorally specific.

Table 3 – Summary search statistics for 2 bedroom properties in London (registered users only)

2 Bedroom Prices	% Total	Av Max Price (£)	Av Min Price (£)	Range (£)	Median Area (SQM)
Up to £250k	22.5	220,579	210,119	89,540	4.3
£250-500k	50.1	381,542	263,754	117,788	5.5
£500-750k	15.1	617,468	508,609	108,859	3.6
Over £750k	12.3	1,187,457	804,230	383,227	2.8

In order to demonstrate the sectoral and spatial specificity of submarket geography, Figures 6 and 7 illustrate Rightmove user-generated search patterns in London overlaid with the most local level of market boundaries derived for a recent National Housing and Planning Advice Unit (NHPAU) study (Jones et al., 2012). In the first example, search for three bedroom properties in the range from £250,000 to £300,000 demonstrate that this market segment is heavily skewed towards the eastern parts of London and that in some cases the existing NHPAU boundaries represent a good fit for delineating areas; for example, in the London (Outer East) market area.

In many other areas, however, the geography of search is more fragmented. When we examine the geography of search in relation to more expensive properties, as in Figure 7, we observe a much more spatially concentrated search area that does not adhere to the boundaries identified by NHPAU. In this case, search for two bedroom properties in the £750,000 to £1,000,000 range is highly concentrated in west central London, with some more localised satellite search areas to the south of the River Thames. Such patterns may therefore support the concept, as previously highlighted by Maclennan and Tu (1996) and Watkins (2001), of simultaneous spatial and sectoral submarkets. Further empirical testing over an extended time period could help validate such an assumption, and this task is to be the focus of future work in this area. Before concluding, however, the final analytical section of the paper examines the links between search activity on the one hand, and available properties on the other.

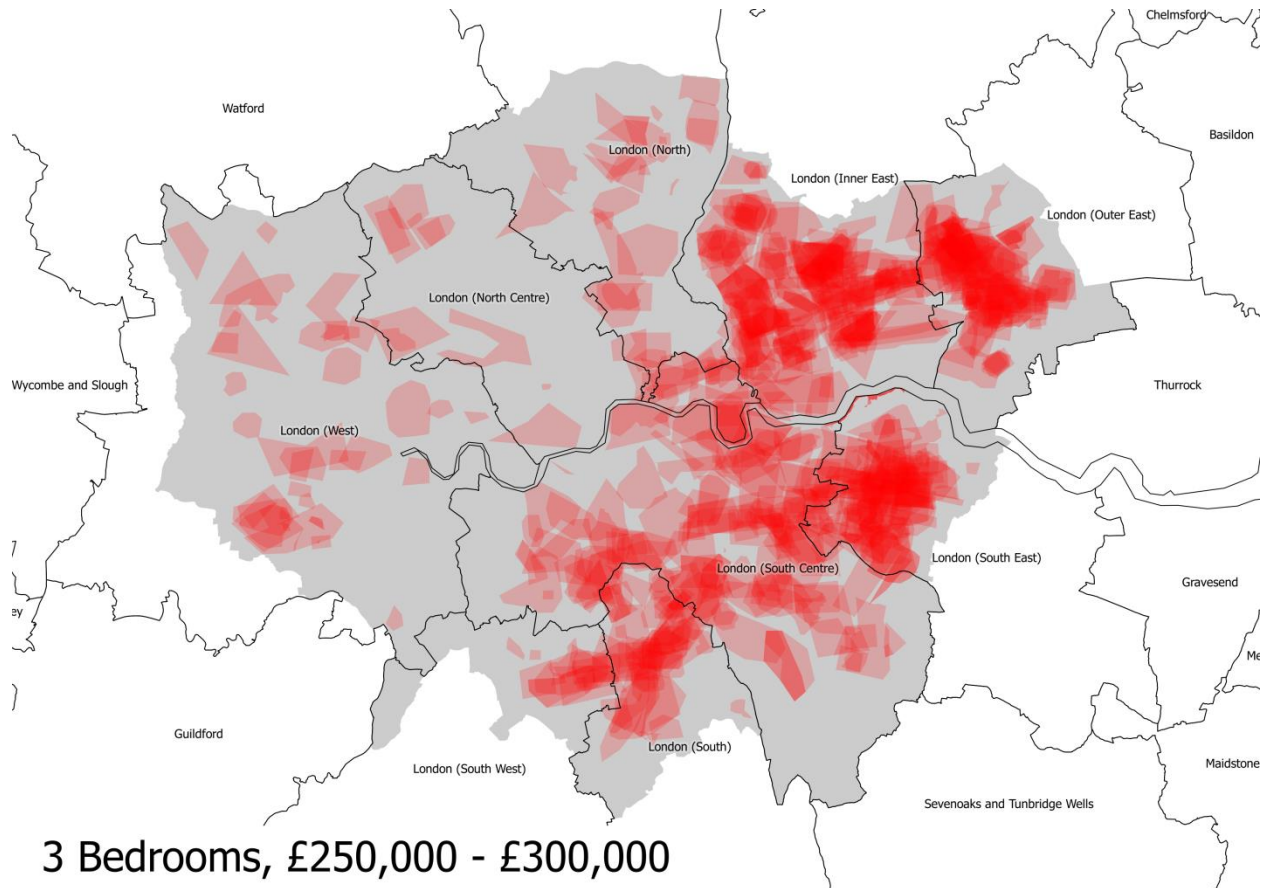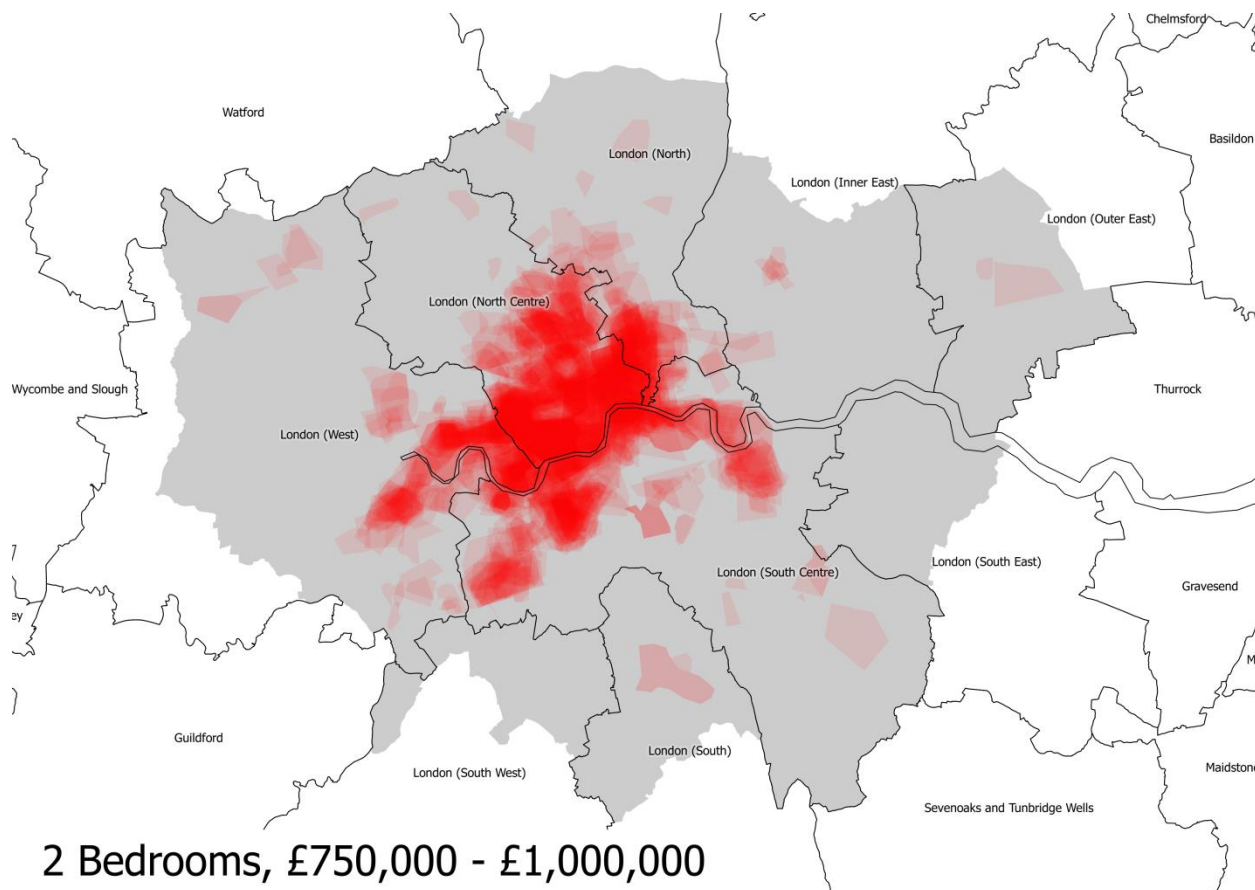**Figure 6 – Rightmove search patterns (3 bed, £250-300k) vs. NHPAU local market boundaries**

3 Bedrooms, £250,000 - £300,000

Figure 7 – Rightmove search patterns (3 bed, £750-1,000k) vs. NHPAU local market boundaries

2 Bedrooms, £750,000 - £1,000,000
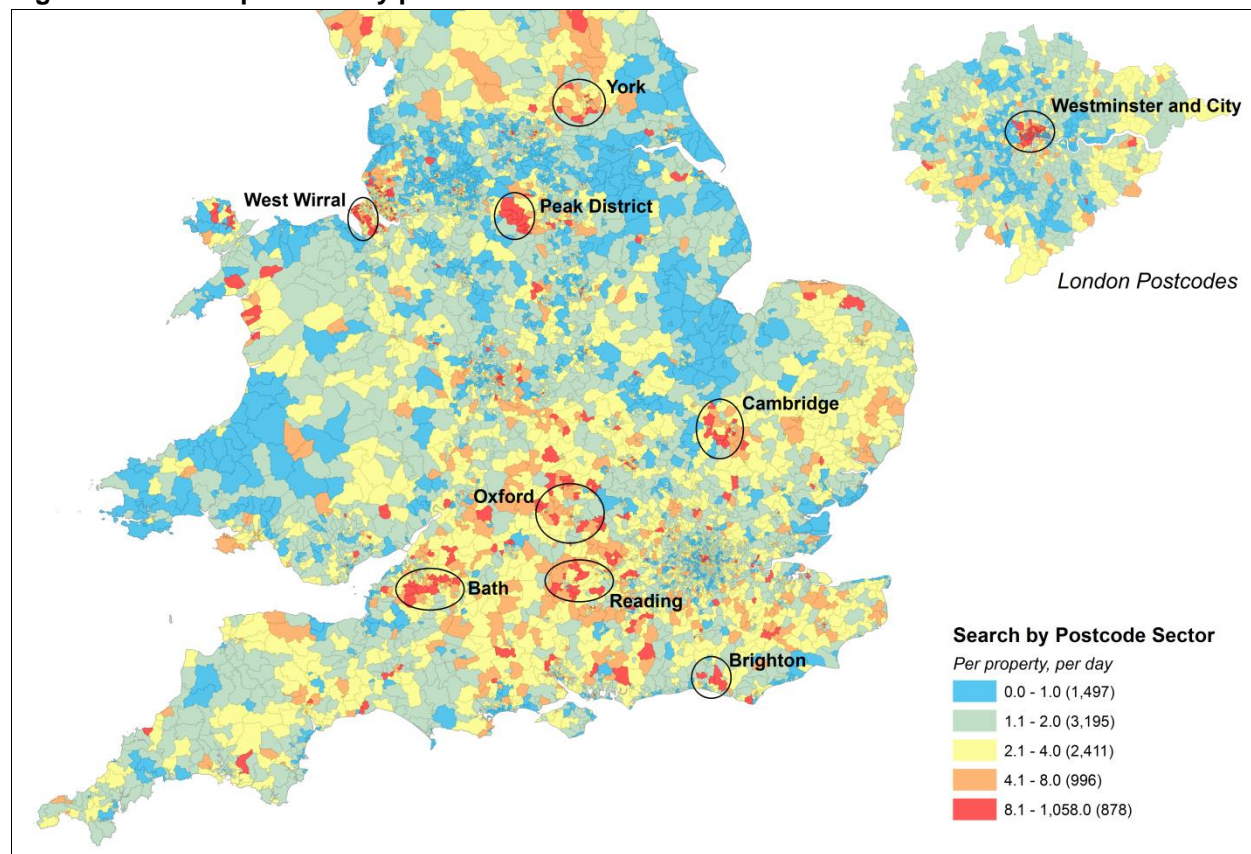
4.3 Search pressure areas

In addition to questions about submarket spatiality, another important question about search activity is how it differs from the available supply of properties on the market. If the searches – and subsequent online views – generated by registered users of rightmove.co.uk can be seen to represent a form of demand, then this must be compared to the available supply if meaningful comparisons are to be made between locations. Differences between the level of search activity, on the one hand, and the level of available properties, on the other, could be a useful indicator of locations where excess demand exists (Levin and Pryce, 2009). Conversely, this approach could also help highlight areas where low demand remains problematic, as it has been in several areas in the north of England during the past three decades (Bramley and Pawson, 2002).

Up to date spatial intelligence of this kind could play a potentially important role in relation to modelling demand, planning for future housing needs and even in helping to 'predict the present' in the housing market, as a recent Google study suggested (Choi and Varian, 2012). This section therefore introduces the concept of 'search pressure areas' by comparing search activity to the number of properties for sale at the time searches were conducted by registered users of rightmove.co.uk. Data are aggregated to the c. 10,000 postcode sectors of Great Britain in order to present the results at a sufficiently low spatial resolution.

The results of this spatially segmented search pressure analysis are shown in Figure 6 for much of England and Wales. This map shows how many views each property received, on average, as a direct result of user searches during the period covered by the data sample. The values range from just above zero views per property, per day to an outlying value of 1,058 per property per day. The median number of searchers per property, per day was 1.87. The number of postcode sectors in each class is shown in the map key, but it is particularly revealing that 1,497 postcode sectors (16% of the total) were areas of relatively low search pressure.

The vast majority of areas did, however, generate a higher volume of search activity and 878 postcodes received eight or more views per property per day during the data sample period. These areas were mostly situated in the south of England in locations such as Bath, Brighton, Cambridge, Oxford and Reading, in addition to central London, but there are notable clusters of high search pressure in the Peak District National Park area, the west of the Wirral peninsula, and York. Other areas of particularly high search pressure are to be found in north west Wales, southern Cumbria and the South Downs – generally considered to be among the most picturesque areas of Britain. However, the spatial patterns indicated by this search pressure analysis do not always correlate with levels of environmental amenity or natural beauty.

Figure 8 – Search pressure by postcode sectors

If search pressure is seen as a foreshadower of actual demand, as Wu and Brynjolfsson (2013) have suggested, then this kind of analysis could play an important role in our understanding of the spatiality, functioning and dynamics of housing markets in future. This could then play an important role in, *inter alia*, improved planning for housing, the development of more accurate demand models, predicting wider market volatility, and in helping identify local areas where excessively high or low demand might occur. This is not yet the case, but with live, user-generated search data the potential is certainly there.

5. Conclusions

This paper has attempted to re-frame our understanding of housing market search in light of the fundamental shift in information acquisition that has taken place in the past decade. However, it has not sought to undermine existing conceptions of how the housing market itself operates, how people think about housing search or about the functioning of the market itself. Instead, it has attempted to build upon longstanding theoretical constructs to develop a conceptual framework for interpreting housing search in an era where the internet is now the first port of call for the majority of consumers. From an information acquisition point of view, things are now significantly different and we do need new ways of understanding the mechanisms through which people interact with the housing market at their first point of engagement. Established models of housing search appear dated when faced with the changes in information technology that have taken place over the past decade in particular. The framework presented in Figure 2 is therefore a contribution to the development of understanding housing search activity in the digital age.

A continuation of the research presented here could also make a valuable contribution to understanding the way in which housing submarkets are segmented sectorally and spatially. If, as Maclennan and Tu (1996) postulate, both can occur simultaneously, then the stratified search data analysed above could prove particularly useful. Indeed, Figures 6 and 7 in particular suggest that such a proposition, as discussed at length by Watkins (2001), is viable and would be a logical next step to the analysis conducted here. If, as seems plausible, housing submarkets can be simultaneously spatial and sectoral in the same locations, then housing search data of the kind used here could be particularly valuable in helping identify them. An important aspect of understanding this, and one not attempted here, is exploring the temporal fluctuations which the housing market is subject to. Whereas this study is based on a limited time period, future studies should where possible examine longitudinal search datasets.

The exploration of search pressure areas in this paper represent an attempt to incorporate some form of supply function which can help us to make sense of the large volume of searches. The results clearly demonstrate that search activity is not purely a function of population density or the availability of properties. However, what has not been attempted here, is to link search pressure to longer term outcomes. This is to be the subject of future work in this area and will

focus on identifying links between search pressure in particular areas and price volatility. For example, it is hoped that future studies in this area can help us understand whether high volumes of search activity lead to price increases or, conversely, whether low search activity leads to stagnating or declining submarkets. In short, a more detailed investigation of the links between housing search pressure and price formation is necessary if we are to truly understand the significance of online search in the operation of the housing market at a national level.

In conclusion, then, this paper represents a first step towards incorporating user-generated search data into housing market analysis. In future, more work is needed in relation to understanding the individual characteristics of searchers, the links between spatial and sectoral submarkets and the relationship between search and price formation. This research could also make a wider contribution to our understanding of how internet search is temporally connected to market movements in different locations (e.g. Choi and Varian, 2012; Wu and Brynjolfsson, 2013) and how 'big data' can play a role in the development of better policy. In the first instance, however, it is hoped that this research will stimulate new thinking on a topic which continues to play a prominent role in the national economic debate in many nations and which remains integrally important to the functioning of the wider economy.

References

Adair, A. S., Berry, J. N., & McGreal, W. S. (1996). Hedonic modelling, housing submarkets and residential valuation. *Journal of Property Research*, 13(1), 67-83.

Allen, M. T., Springer, T. M., & Waller, N. G. (1995). Implicit pricing across residential rental submarkets. *The Journal of Real Estate Finance and Economics*, 11(2), 137-151.

Blank, D. M., & Winnick, L. (1953). The structure of the housing market. *The Quarterly Journal of Economics*, *67*(2), 181-208.

Botterill, J. (2013) 'Property Porn: An Analysis Of Online Real Estate Advertising', in M. McAllister and E. West, *The Routledge Companion to Advertising and Promotional Culture*, Routledge, New York.

Bourassa, S. C., Hoesli, M., & Peng, V. S. (2003). Do housing submarkets really matter?. *Journal of Housing Economics*, 12(1), 12-28.

Bramley, G., & Pawson, H. (2002). Low demand for housing: incidence, causes and UK national policy implications. *Urban Studies*, *39*(3), 393-422.

Brown, P. J., & Hincks, S. (2008). A framework for housing market area delineation: principles and application. *Urban Studies*, *45*(11), 2225-2247.

Choi, H., & Varian, H. (2012). Predicting the present with google trends. *Economic Record*, *88*(s1), 2-9.

Clark, W.A.V. (1982). *Modelling housing market search*. London: Croom Helm.

Clark, W.A.V., & Flowerdew, R. (1982). A review of search models and their application to search in the housing market. In W. A. V. Clark (Ed.), *Modelling Housing Market Search* (pp. 134-159). London: Croom Helm.

Dunning, R. & Watkins, C. (2012) *Information, Search and Housing Choice*, RICS, London.

Experian Hitwise (2013) *Hitwise Market Intelligence*, available at http://www.experian.com/hitwise, data from March 2013.

Ferrari, E., & Rae, A. (2011). Local housing market volatility. *York: Joseph Rowntree Foundation*.

Goodman, A. C., & Thibodeau, T. G. (1998). Housing market segmentation. *Journal of housing economics*, 7(2), 121-143.

Goodman. A. and Thibodeau, T. (2007) 'The Spatial Proximity of Metropolitan Area Housing Submarkets', *Real Estate Economics*, 35(2), 209-232.

Google Trends (2013) Search data available at: http://www.google.co.uk/trends

Huff, J.O. (1982) Spatial aspects of residential search, In W.A.V. Clark (Ed.), *Modelling Housing Market Search* (pp. 134-159). London: Croom Helm.

Hwang, S. & Thill, J-C. 2009. Delineating urban housing submarkets with fuzzy clustering. *Environment and Planning B* 36:865-882.

Jones, C. (2002). The definition of housing market areas and strategic planning. *Urban Studies*, *39*(3), 549-564.

Jones, C., & Watkins, C. (2009). *Housing markets and planning policy* (Vol. 41). Wiley-Blackwell.

Jones, C., Coombes, M., Dunse, N., Watkins, D., & Wymer, C. (2012). Tiered Housing Markets and their Relationship to Labour Market Areas. *Urban Studies*, *49*(12), 2633-2650.

Levin, E.J., & Pryce, G. (2009) *A regional house price model of excess demand for housing*. Royal Institution of Chartered Surveyors, London.

Maclennan, D. (1982). *Housing economics: an applied approach*. Harlow: Longman.

Maclennan, D., & O'Sullivan, A. (2012). Housing markets, signals and search. *Journal of Property Research*, 29(4), 324-340.

Maclennan, D., & Tu, Y. (1996). Economic perspectives on the structure of local housing systems. *Housing Studies*, 11(3), 387-406.

MacLennan, D., & Wood, G. (1982). 'Information acquisition: Patterns and strategies'. In W. A. V. Clark (Ed.), *Modelling Housing Market Search* (pp. 134-159). London: Croom Helm.

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.

Meen, G. (2001) *Modelling Spatial Housing Markets: Theory, Analysis, and Policy*, Kluwer, Norwell, MA.

Palm, R. (1976) 'Real estate agents and geographical information', *Geographical Review*, Vol. 66, 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.

Royuela, V., & Vargas, M. A. (2009). Defining Housing Market Areas Using Commuting and Migration Algorithms: Catalonia (Spain) as a Case Study. *Urban Studies*, *46*(11), 2381-2398.

Sun, S. & Manson, S. (2010) 'An Agent-based Model of Housing Search and Intraurban Migration in the Twin Cities of Minnesota', International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada.

Varma, A. (2004) *Working Paper 7: Defining and Analysing London's Housing Submarkets*, GLA Economics, London.

Watkins, C. A. (2001). The definition and identification of housing submarkets. *Environment and Planning A*, *33*(12), 2235-2254.

Wheaton, W. C. (1990). Vacancy, search, and prices in a housing market matching model. *Journal of Political Economy*, 1270-1292.

Wu, L., & Brynjolfsson, E. (2013). 'The future of prediction: How Google searches foreshadow housing prices and sales', *National Bureau for Economic Research*, available at: http://pages.stern.nyu.edu/~bakos/wise/papers/wise2009-3b3_paper.pdf