Search in Real Estate Markets
Research Proposal

Arpit Gupta* Theresa Kuchler† Venky Venkateswaran‡

April 5, 2019

Geolocation Data: We make novel use of a data resource based on unstructured geolocation data. Our data are drawn from applications on individual smartphones which track the locations of around 60 million Americans over the course of Summer 2016–Summer 2017. Our data contain no personally identifying information, but allow us to connect the precise geolocations of a large sample of individuals combined with exact timestamps for those pings.

Computing Resources: The size of our data (60 terabytes) pose substantial computing challenges. To address these, we have worked extensively to set up an appropriate computing environment for our data. We conduct our analysis primarily using a Hadoop Cluster. Computing resources available on this platform include 1280 CPU cores and 280 terabytes of raw HDFS. We access the data using Spark, which allows for rapid data access and quick processing.

We take a number of steps to ensure privacy of individuals in our data. Our server access provides secure connections, and restricts files download. We are careful in our analysis to avoid the personal identification of any individual on our analysis.

Motivation of Core Analysis: Despite the importance of search behavior on economic decisions, direct information on the search patterns of individuals is difficult to obtain in practice. For real estate in particular, information acquisition through directly visiting listed properties is an important component of the home purchasing process. Yet, we know relatively little about how individuals search for homes to buy. Our project seeks to understand the process of physical real estate search by examining individual mobility data which allow us to track the physical locations of a sample of 60 million Americans across their process of real estate search and purchase.

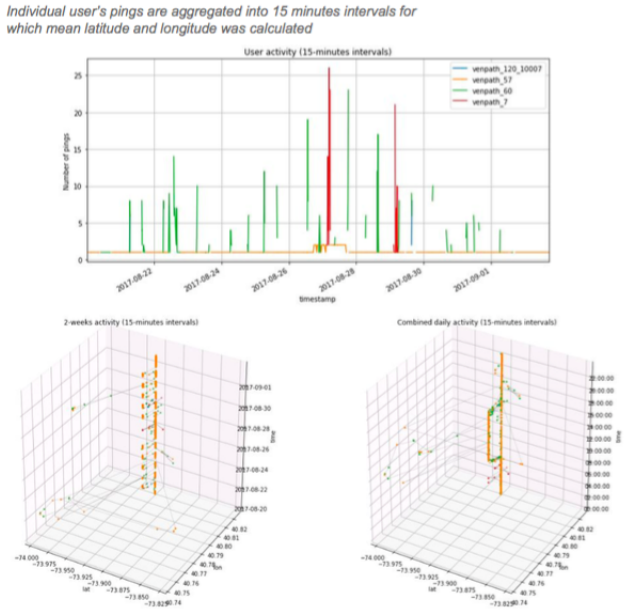
Sample Used: We combine several other administrative datasets with our geolocation records to provide a novel perspective at understanding housing search behavior. We include housing Deeds data, which contain detailed information on housing attributes, as well as changes in ownership. Open home listing data include a comprehensive set of all homes listed. Finally, data from Infutor contain information on the residential move history of 170 million Americans, over their last 10 addresses over the past 30 years. We host this data on a common computing platform.

The combination of these datasets allows us to 1) Identify home movers, using the Infutor dataset; 2) Isolate the set of homes available on the market, in the time that they were searching; 3) Identify which of these properties individuals were searching for on the basis of visits to local open homes; and 4) Identify which homes people did in fact ultimately purchase, on the basis of transactions cross-verified in the Deeds and Infutor datasets.

Additionally, our data allow us to identify other attributes of individuals. Aside from the identification of home locations using the mobile phone dataset (on the basis of the residence in

*NYU Stern, arpit.gupta@stern.nyu.edu
†NYU Stern, tkuchler@stern.nyu.edu
‡NYU Stern, vvenkate@stern.nyu.edu

Figure 1: Estimation of home and work locations



which individuals can be most commonly located in night hours; our analysis allows us to identify as well work locations. Figure 1 illustrates this process for one sample individual who lives in the Bronx and works in Manhattan.

Research Questions:

1. Understand individuals' real estate search process.
    - What influences in which areas homeowners visit homes?
    - How many homes do people typically visit before buying? How long do they spend in each house? How long have people searched and visited properties before buying? How many potential buyers end up purchasing a home?
    - How does real estate search impact ultimate purchase price? Does prior search behavior affect prices? Do the characteristics of other homes on the market at the same time affect buyers' decision and ultimately home prices?
    - Are there differences between local buyers and out-of-town buyers? Is the search behavior of out-of-town buyers influenced by the real estate market in their previous location?
2. How do individual's mobility patterns interact with their real estate search?
    - What role does commuting distance to place of work play?
    - Does distance to old residence and prior experience with the location affect location decisions?
    - How do distance and location of places typically visited influence location and home purchase decision?

Current Status and Next Steps: Our analysis so far has identified home locations; definite movers, and has identified open home visits on the basis of both 1) physical proximity to specific open homes, as well as 2) time spent on location. We have also identified work locations on the basis of typical daytime commuting patterns. This pattern of connected data is sufficient to establish a dataset of individuals engaged in real estate search.