Leo Kitchell

Professor Magilke

ECON180

September 16, 2021

Senior Thesis Prospectus

Over the past 5 years, the size and complexity of Natural Language Processing (NLP) algorithms has increased exponentially, creating new opportunities for economists to effectively leverage novel text data. In early 2018, state of the art NLP models used up to 100 million parameters; today’s state of the art models use upwards of 175 billion parameters, which translates to greater accuracy in a variety of tasks.[[1]](#footnote-1) This computational revolution has been noticed by economists, as evidenced by the nearly 100 working papers which include the keyword “Natural Language Processing” on the National Bureau of Economic Research website.[[2]](#footnote-2) However, given the speed of computational advances, the corpus of economic research using modern NLP algorithms may be limited outside of these working papers. In addition, the spread of COVID-19 has created and continues to create unusual heterogeneity across many markets, which researchers have had fewer than two years to investigate.

Among the markets disrupted by COVID-19, the U.S. housing market is a ripe choice for research given its importance in previous economic contractions and its well documented recomposition.[[3]](#footnote-3)[[4]](#footnote-4) One particularly robust source of text based housing data is the Airbnb open datasets, which provide forward looking prices and descriptions for each Airbnb listing in a given geography.[[5]](#footnote-5) Utilizing a panel of these panel data, the effect of changes in listing descriptions can be isolated to identify which words or phrases correspond to abnormal rental prices. Notably, variation in quality of housing and aggregate demand can be absorbed by investigating on a rental-by-rental basis and creating daily expected mean price variables, respectively.

If the remaining variation is predicted by the textual descriptions, there would be a host of significant implications. Using new releases of the Airbnb data, real estate investors could evaluate all listings for phrases predicted to increase future rent and preemptively purchase the listing or rental cash streams at a discount. With future research on the relationship between short term rental prices and community development, insurance companies and banks may be able to better predict the risk or default rates of loans proximate to groupings of listings predicted to increase in price. Conversely, these textual data may describe listings or neighborhoods where predicted rents are lower than posted and home values may be inflated.[[6]](#footnote-6) In light of housing bubbles’ impact in previous recessions, developing tools to better discover abnormal valuations may help reduce the size and severity of future housing crashes.[[7]](#footnote-7)

1. Brown, Tom B., et al. "Language models are few-shot learners." arXiv preprint arXiv:2005.14165 (2020). [↑](#footnote-ref-1)
2. “Working Papers.” NBER, National Bureau of Economic Research, www.nber.org/papers?page=1&per Page=50&sortBy=public\_date. [↑](#footnote-ref-2)
3. Christiano, Lawrence J., Martin S. Eichenbaum, and Mathias Trabandt. "Understanding the great recession." American Economic Journal: Macroeconomics 7.1 (2015): 110-67. [↑](#footnote-ref-3)
4. Toura, Varvara. "Emerging socio-economic inequities during the Covid-19 pandemic." URBAN CORPORIS: 224. [↑](#footnote-ref-4)
5. “Inside Airbnb: Adding Data to the Debate.” Inside Airbnb, Airbnb, insideairbnb.com/about.html. [↑](#footnote-ref-5)
6. Garcia, Brett & Miller, Keaton & Morehouse, John. (2020). In Search of Peace and Quiet: The Heterogeneous Impacts of Short-Term Rentals on Housing Prices. 10.13140/RG.2.2.25842.32965. [↑](#footnote-ref-6)
7. Christiano, et al. “Understanding the great recession” [↑](#footnote-ref-7)