Measuring Housing Stability with Consumer Reference Data

David C. Phillips⋆

December 20, 2019

Abstract: Housing instability for low-income renters has drawn greater attention recently, but measurement has limited research on policies to stabilize housing. Address histories from consumer reference data can be used to increase the quantity and quality of research on low-income renters. Consumer data tracks housing moves throughout the entire United States for most of the adult population. This paper shows that such data can measure housing stability for groups with very low income and extreme instability. For example, the data can track housing moves during natural disasters, at demolition of public housing, for households at high risk of homelessness, and during gentrification. Consumer data can track housing instability outcomes that are more common than shelter entry and cheaper to collect than surveys. Relative to existing administrative address histories, consumer data allow researchers to track housing moves to exact addresses and across jurisdictions.

1 Introduction

Housing instability affects a large fraction of low-income renters in the United States. In 2018 landlords filed for the eviction of 2.3 million out of 38.4 million renter households, according to data recently aggregated by Desmond et al. (2018). According to Point-in-Time counts from the Department of Housing and Urban Development (HUD), more than half of a million people were homeless on a single night in January 2018. In-depth qualitative work such as Desmond (2016) has drawn new attention to the prevalence and high cost of such housing instability, and policymakers are responding to housing instability of low-income singles and families. For example, New York City has introduced a right to legal counsel in housing court, and homelessness interventions have rapidly transitioned from traditional models to 'housing first.' While the literature provides some rigorous evidence on the causal effects of major policy responses to housing instability, data constraints lead to large holes in the literature.[1] To measure housing stability, researchers typically rely on limited and difficult to access government administrative data or expensive surveys. Data constraints almost certainly limit the accumulation of additional evidence on how best to respond to housing stability for low-income renters.

In this paper, I show how researchers can use existing, publicly available consumer reference data to measure even extreme instances of housing instability. Consumer reference data companies accumulate records from various commercial transactions, e.g. subscription services, and create individual-level, national databases documenting the characteristics of an individual. I use an extract of data from one such company, Infutor Data Solutions, though competing companies exist (e.g. InfoUSA). The data attaches individual identities to address histories for essentially the full adult population of the United States, including exact street addresses and dates for which the address is valid. This data was introduced to the academic literature in a recent study of rent control (Diamond et al. (2019)). I use the same data to measure the number, timing, and location of housing moves for particularly unstably housed people. While measures of housing instability may in general include other characteristics (Frederick et al. (2014)), this paper focuses on how the data can measure housing instability as defined by frequent moves.

I validate the data, showing how housing moves recorded in consumer reference data can match known instances of extreme housing instability. First, I study public housing closures in Chicago. The existing literature studies how these closures affect children forced to move (Jacob (2004); Chyn (2018)) and neighborhood crime (Aliprantis and Hartley (2015); Sandler (2017)). I identify people living at one such housing complex, Robert Taylor Homes, and demonstrate that address moves in the consumer reference data match known building closure dates from the literature. Second, I examine the case of Hurricane Katrina in New Orleans. I use the consumer reference data to identify people living in New Orleans prior to the storm and track their moves afterward. The data correctly identifies the timing of the hurricane and effects on housing stability that vary with the extent of flooding. Third, I track housing instability for a widely reported case in which

---

[1]See Evans et al. (2019) for a more complete review of popular policy responses and the existing evidence base.

a change of management at a Washington, DC apartment complex induced significant housing instability. Finally, I validate the data by comparing it to more traditional census-type data. The consumer reference data omits most people under age 25 but otherwise matches statistics from the American Community Survey including annual move rates, state populations, and demographic characteristics. It is also more flexible and dynamic than traditional data, observing multiple moves within the same year.

Having validated the data, I demonstrate two applications of the data for future research. First, I construct a housing stability measure for use as an outcome for program evaluation. Evans et al. (2016) and Palmer et al. (2019) study the effect of emergency financial assistance provided through the Homelessness Prevention Call Center (HPCC) in Chicago for a sample of people at imminent risk of losing their housing. I match the consumer reference data to the same sample. I am able to match 61% of records of people eligible for emergency financial assistance to Infutor records. Matched people are similar to unmatched people on many characteristics including gender, income, and use of public assistance but are somewhat less likely to be young, Hispanic, or living in shared housing. Just as emergency financial assistance reduces emergency shelter entry in Evans et al. (2016), I find that address changes also fall for those referred to assistance relative to eligible applicants not referred to assistance. At the same time, I find that shelter entry and address changes are complementary measures in other ways. Address changes are more common, only rarely coincide with shelter entry, and correlate differently with control variables used in the prior papers.

Second, I show how consumer reference data can track location at fine detail and national scope. The data can track housing moves down to the exact building. For the HPCC study, I measure treatment effects of emergency financial assistance on neighborhood locations within Chicago. The data can also track moves across the country. I identify people moving from New Orleans after Katrina, tracking their moves to all counties throughout the United States. A final example requires both the interstate scope and the exact address detail of the data. I follow the moves of residents leaving gentrifying neighborhoods in Washington, DC as they exit to neighborhoods spread across not only DC but also neighboring Maryland and Virginia.

These results demonstrate the advantages of consumer reference data relative to existing measures of housing instability. The literature typically uses a few well-known methods: surveys, administrative data on shelter entry, or administrative data on addresses. Consumer reference data complements these existing sources in a few ways. Relative to consumer reference data, surveys are much more expensive and tend to experience high attrition rates when tracking unstably housed people. Administrative data often has a lower financial cost than either surveys or consumer reference data, but accessing it may be administratively difficult and labor intensive. Because shelter entry is an extreme outcome affecting only a very small fraction of households, it may be neither the obvious target of policy nor conducive to statistical power. Administrative address data from, for example, public assistance records is most similar to consumer reference data. However, such data cannot typically track long-distance moves, which may be confounded with not moving at all.

Of course, consumer reference data also has disadvantages. It typically must be purchased, making it more difficult to access than general purpose surveys. For-profit companies will be less likely to divulge their exact data construction methods than a public entity. Ethical considerations also differ from, for example, surveys with an informed consent process. Finally, the underlying consumer data will necessarily contain gaps for people with shorter paper trails, particularly younger people. Clearly, though, this data adds a new resource that complements more traditional sources of data on housing instability.

The remainder of the paper continues as follows. Section 2 discusses how the existing literature measures housing stability. Section 3 introduces the consumer reference data and validates it using examples from Chicago public housing demolitions, Hurricane Katrina, and a landlord change in Washington, DC. It also compares the consumer reference data to both Census-style data and the Homelessness Prevention Call Center data. Section 4 uses the Homelessness Prevention Call Center, Hurricane Katrina, and DC gentrification examples to demonstrate how the data can be used in program evaluation and for tracking moves within a city or across the country. Section 5 discusses some practical and ethical considerations that researchers using this data will face. Section 6 concludes.

2   Literature

Limited data availability almost certainly limits the literature on housing stability. The studies that do exist go to great lengths to measure housing stability. Four different approaches to measurement are most common: community-level measures, surveys, administrative shelter entry data, and administrative address histories. However, each of these methods has significant drawbacks.

Some studies examine homelessness at the community level. They leverage increasingly available Point-in-Time count and similar data from HUD to test if greater federal homelessness grant funding (Popov (2016); Lucas (2017)) or provision of permanent support housing (Corinth (2017), Evans et al. (2019)) affect community homelessness counts. While useful for some studies, community-level measures can, of course, only be used for community-level policy interventions, making them less relevant to the micro-level studies that are the focus of this paper. The Point-in-Time data are also rather new, inconsistently implemented in different locations (Schneider et al. (2016)), and can differ considerably from other measures such as school-based counts (Evans et al. (2019)).

Surveys provide great detail but risk high attrition and force a trade-off between great expense and sample size. Researchers can ask respondents directly about their housing history. When studying long-term housing subsidies and support services for homeless individuals, i.e. permanent supportive housing, many randomized control trials use this approach (e.g. Stergiopoulos et al. (2015)). Surveys tend to be comprehensive, providing rich detail on housing instability. However, surveys tailored to studying housing stability can be prohibitively expensive, especially since tracking and interviewing unstably housing individuals typically requires in-person interviewing, significant respondent incentives, and intensive tracking activities between survey waves. Even in

the best case, attrition will be quite high. For example, even the Family Options Study experienced 22% attrition (Gubits et al. (2018)). If research on the housing instability must rely on special use surveys, these costs will likely keep the literature quite small. Of course, general purpose surveys such as the longitudinal PSID or the American Housing Survey, for which housing stability measures were recently updated, provide an alternative, and economies of scale in the use such surveys for many projects lowers their cost per project. However, many studies require detailed geographic information or large sample sizes for very specific sub-samples, which limits the use of general purpose surveys in many contexts.

Administrative data on shelter entry provides an inexpensive and widely available option but only measures an extreme outcome. Evans et al. (2016) estimate the effect of emergency financial assistance on sheltered homelessness in Chicago using shelter entry records from the Homelessness Management Information System. Rolston et al. (2013) and Goodman et al. (2016), use similar data in New York City to measure the impact of homelessness prevention services. Collinson and Reed (2018) use the same New York data to follow shelter entry for participants in housing court. When available, administrative data on shelter entry certainly provides a cheaper alternative to surveys. Many jurisdictions now participate in Homelessness Management Information Systems and collect close to comprehensive data on shelter entries. This means shelter entry data typically exists and is cheap to measure. However, shelter entry is an extreme and rare outcome. In Evans et al. (2016) only 3% of the group not receiving services enter a shelter. Other forms of housing instability are much more common. In Family Options Study usual care participants are 4 times more likely to double up with family and friends than to enter shelter (Gubits et al. (2018)). Similarly, school-based counts of homelessness that include doubling up imply dramatically higher levels of homelessness than the Point-in-Time counts, which do not include doubling up (Evans et al. (2019)).

Administrative data may also track individual address histories. For example, Chyn (2018) uses addresses in public assistance records to track tenants exiting demolished public housing in Chicago. Humphries et al. (2019) track tenants from Chicago housing court using credit agency records. Other options include children's school records. Such data can measure whether a household has a formal address, how frequently a household moves, and the characteristics of the neighborhood and unit. Administrative address data is similar to the consumer reference data I examine in this paper in many ways, but the two options differ in a couple important ways. First, the practical conditions for accessing the data are different. Local governments have not created administrative data for the purpose of disseminating it, so accessing such data poses significant administrative and legal barriers. Consumer reference data companies, on the other hand, pursue data sharing in exchange for a fee. Second, local government data typically covers a particular jurisdiction such as a city, county, or state, while consumer reference data is national in scope. Local administrative data will not be able to disentangle a household that does not move from one that moves out of the jurisdiction. Thus, consumer reference data may be a particularly useful option when data from local government is either inaccessible or too limited in scope.

The advantages and disadvantages of these various data sources are well-known in the literature. For this reason, the most ambitious studies track housing instability through all of the above sources. Large, federally-supported randomized control trials such as the Moving to Opportunity project (Ludwig et al. (2013)) and the Family Options Study (Gubits et al. (2018)) collect administrative data on address histories, shelter entry data, and conduct surveys. This revealed preference for multiple measures and the relative lack of such studies suggests a need for additional sources of data on housing stability.

3 Data

3.1 Infutor Data Solutions Address Histories

This paper uses national data on individual-level address histories from Infutor Data Solutions (Infutor), a consumer identity management company based in Chicago, Illinois. Infutor compiles a wide variety of data from sources such as public and private telephone data, deed and property information, subscription services and numerous other privacy and security- compliant sources. Infutor combines these records into a single identity graph that identifies individuals and links particular records to a single individual. The result is a list of unique individual residential histories. Individual names, dates of birth, and social security numbers can be attached to exact addresses and dates for which the address is valid.

Address changes can potentially measure housing stability that cannot be observed with other measures. In this paper, I typically define a move as either the current address ending or a new address starting. This measure recognizes the complexity of address histories for many unstably housed people, who may have multiple valid addresses that overlap in time. Traditional data sources measure housing stability by observing someone enter shelter or change their official address with one public agency. Consumer data that aggregates valid addresses from many sources can, at least in principle, provide a broader measure of housing moves, measuring if someone changes the address for their mobile phone bill but forgets to update their child's school.

Whether address histories developed for commercial purposes can be used to measure housing stability, though, is an open question. In the past, such data has been used for direct mail marketing and identity verification. More recently, academics have begun using it to measure residential status. Diamond et al. (2018) use Infutor data to measure the effect of rent control on residential moves. However, whether a dataset can accurately identify the right people to receive mail is somewhat different than whether that same data can correctly identify the timing and location of residential moves. This latter objective relies on the accuracy of the underlying data sources and matching algorithm. Since the data comes from a private company, both of these are necessarily proprietary, and such opacity necessitates validating the data for academic use.

3.2 Examples to Validate the Data

Consider three examples that can validate the ability of consumer reference address histories to identify the timing of housing moves.

Hurricane Katrina hit New Orleans and the Gulf Coast in August 2005, flooding large sections of the city. In particular, the Lower 9th Ward was totally inundated. Other areas on higher ground, such as the French Quarter, were significantly affected by the hurricane but with less dramatic flooding. In the Infutor data, I identify all people who lived in New Orleans between 1990 and 2005 either in a particular ZIP including much of the Lower 9th Ward (70117) or a ZIP including much of the French Quarter (70130). I then identify the first observed address at which the person appears after living in one of these ZIP codes. Figure 1 shows the number of people in the Infutor data moving away from these two New Orleans ZIP codes. I measure the number of people appearing at new addresses in a given month, relative to a typical month in 2004. Residents of the 9th Ward ZIP are 20 times more likely to appear at a new address just after the storm than in the years preceding it. As would be expected, residents of the French Quarter are also much more likely to move than before but are less responsive than people who experienced catastrophic flooding.

The closure and demolition of public housing in Chicago during the 1990s and 2000s has generated a large academic literature. Demolishing public housing complexes and giving the displaced tenants vouchers has little affect on schooling (Jacob (2004)) but increases future earnings (Chyn (2018)). Demolitions also reduced neighborhood crime near the demolition, with some debate about whether it reduced or simply redistributed overall crime (Aliprantis and Hartley (2015); Sandler (2017); Bruhn (2018)). One of the most well-known cases of public housing demolition are the Robert Taylor Homes. At its height, Robert Taylor Homes included 4,200 units in 21 buildings immediately next to the Dan Ryan Expressway on the South Side of Chicago. Popular press heavily reported on the housing complex, crime in the area, and the lives of those displaced (Garza (1999); Pollack (2000)). The complex was closed and demolished slowly over several years, and the academic literature above has defined closure dates for particular buildings.

Identifying housing stability of Robert Taylor Homes residents is straightforward using consumer reference data. I follow Jacob (2004) and Aliprantis and Hartley (2015) in attaching demolition dates to exact addresses of the buildings. I then identify any person in the Infutor data ever associated with a Robert Taylor Homes address. For such people, I measure moves to the first address outside Robert Taylor Homes with a start date after that of their Robert Taylor Homes address. Figure 2 shows the number of people per building per month appearing at new addresses in the Infutor data. Each sub-figure shows the number of people moving split out by the year the building closes. Vertical lines show the building closing dates as reported in the literature. The Infutor data correctly identifies closing dates of buildings as times of high residential instability. Particularly after 2000, large spikes in residential mobility occur at the time of closure.

One concern with the two prior examples is that they focus on situations in which contact with federal emergency management or public assistance might make households easier to track in consumer data. Can consumer reference data track the timing of less dramatic, informal housing

instability, such as informal evictions after changes in property management? Consider the case of Terrace Manor Apartments in Washington, DC, which has been a prominent example involuntary moves induces by informal landlord action after a change in property management. Mills and Giambrone (2017) report in the local news:

> These days only about a dozen units at Terrace Manor remain occupied, the apartments shells of their former selves. When Sanford bought the property in 2012, more than 50 of the property's 61 units were occupied...Sanford made a series of changes after purchasing the property. Workers removed the benches that sat outside each buildings front door. They razed the little playground and replaced it with a Terrace Manor sign. Gradually, the laundry rooms closed. Heat and air conditioning broke down. Tenants began moving out in droves.

Similar to the Robert Taylor Homes example, I identify any person in the Infutor data ever associated with a Terrace Manor address and the start date of their first address after living at Terrace Manor. Figure 3 shows the number of moves by year as measured in the Infutor data. The data identifies an uptick in moves after the management change in 2012 that fades in later years as the apartment complex empties. This is but one small-sample example and does not prove that consumer reference data track all informal evictions. By definition, it is challenging to validate whether any dataset fully characterizes informal housing stability, and consumer reference data likely does worse tracking moves that have shorter paper trails. Still, this example shows that the data can at least record a reasonably large proportion of moves even in a case of informal, landlord-induced housing instability.

3.3   Comparison to Nationally Representative Data

Table 1 compares records from Infutor to nationally representative data. The first column shows summary statistics from IPUMS 2010 American Community Survey (ACS) micro data. The second column shows similar statistics for a 0.01% random sample of Infutor records of living individuals that have a valid address in April 2010, which was the time of the 2010 census. As is apparent immediately, the primary difference between census-style data and the Infutor database is age coverage. The entire US population in the ACS has a median age of 37 while people in Infutor records have a median age of 49. Most of this difference is due to children and young adults who do not show up in the Infutor records because they have not accumulated a large consumer paper trail. The third and fourth columns of Table 1 restrict both datasets to individuals 25 years and older. For those at least 25 years old, median age is quite similar between the two datasets: 50 in the ACS and 51 for Infutor. Similarly, the number of records for people age 25 and above in Infutor is similar to the ACS. The ACS estimates the total population of the US was 309 million in 2010, while Infutor has only 180 million records with valid addresses at that time. However, if we restrict the ACS to people aged 25 and older, the numbers are much more similar, 204 million versus 180 million. Figure 4 also shows that the total number of Infutor records varies reliably

across space. The Figure compares ACS estimates of age 25+ state populations with record counts from Infutor. ACS population estimates for nearly all states are slightly greater than the Infutor count, i.e. slightly above the 45-degree line in the graph.

Conditional on age, the ACS and Infutor measure similar gender balance, building types, and annual move rates. Returning to Table 1, the age 25+ population is 52% female in the ACS and 45% female in Infutor. In the ACS, 17% of the population lives in buildings with more than 1 family, while 15% of Infutor records are in multi-family dwellings. Most importantly, annual move rates are similar across the two datasets. To compare to the ACS, I measure moves within the past year for the sample of Infutor individuals observed near the 2010 census. I count a person as moving if they have any address that began or ended between April 2009 and April 2010. In the ACS, 13% of the sample has moved in the past year compared to 12% in Infutor.

The distribution of moves across locations differs somewhat between the ACS and Infutor data sources, though not in a simple way. Respondents to the ACS report a local move to a different residence within the same public use microdata area (PUMA)[2] 7.6% of the time. Moves within a state between PUMAs (2.1%) are less common. In the Infutor data, though, moves are somewhat more likely to be medium distance moves between PUMAs in the same state (5.1%) than local moves within a PUMA (4.4%). By definition, the Infutor data also misses international moves. On the other hand, inter-state moves within the US are tracked with roughly similar frequency in the ACS (1.8%) and Infutor (1.1%) datasets. While there are some differences in the moves tracked by census-style and consumer reference data, overall move rates are similar, and there is no sign that the Infutor data disproportionately misses long-distance moves within the United States.

In addition to overlapping with Census data, consumer reference data may do a better job capturing dynamics of moving. One concern with Census-type data is that they measure housing mobility and instability as time since the last move, but housing instability may take the form of multiple moves in quick succession. The consumer reference data appear to capture at least some of this repeated moving. As shown in the final row of Table 1, 3.6% of Infutor records observed at the time of the 2010 Census have moved more than once during the past year. This value is greater than what would be expected if multiple moves were statistically independent events,[3] which indicates the data is picking up the tendency for housing moves to be clustered together in time. Similarly, the data picks up many quick succession moves with gaps less than 1 year; the most common length of time between two moves in the consumer reference data is 1 month. See Appendix Figure 1 for a full histogram of the duration between the two most recent housing moves. These facts all suggest that consumer reference data can track interesting dynamics of housing moves that may be difficult to observe in other data.

---

[2]Public use microdata areas are areas of 100,000 to 200,000 residents within a state.

[3]In that case, a population with an overall move rate of 12% should have about 1.4% moving more than once because $0.014 = 0.12^2$.

3.4 Comparison within an Unstably Housed Population

As a final and perhaps more stringent test, I test whether the consumer reference records can match to a sample of individuals known to be at imminent risk of homelessness. I examine the Homelessness Prevention Call Center (HPCC) in Chicago as previously studied in Evans et al. (2016) and Palmer et al. (2019). The HPCC data include callers who request temporary financial assistance when at imminent risk of homelessness. As in the prior literature, I limit the sample to people who are eligible for assistance, i.e. people who are at imminent risk of homelessness, face a specific temporary crisis, can sustainably pay their rent after the assistance, and for whom the needed assistance is less than program funding caps. I match the HPCC call center data to Infutor address histories using name and date of birth. I use the sample of HPCC callers in the main sample of Palmer et al. (2019). I use an extract of Infutor address histories current as of June 2018 and limit the sample to current Illinois residents who have ever lived in Chicago. I then link any Infutor address record with the same month of birth, year of birth, first name soundex, and last name soundex to the HPCC caller record. Even focusing on exact matches for these relatively fuzzy variables, I match 61% of HPCC callers to an Infutor address history that existed prior to their call.

I test whether the Infutor data excludes the most vulnerable individuals by comparing the characteristics of callers who match to the full sample. Table 2 shows mean caller characteristics with columns corresponding to unmatched, matched, and all callers. The matched group is similar to the total sample on many characteristics. Gender, income, use of public assistance (SNAP), veteran status, household size, number of children, and the distribution of reasons for exiting current housing are quite similar across these two groups. A few characteristics differ somewhat between the two groups. As with the comparison to Census data, the consumer reference records are more likely to match to older HPCC callers. The average matched caller is 42 years old, compared to 40 for the HPCC callers overall. The matched sample also skews in favor of black rather than Hispanic residents. Hispanic residents compose 8% of all callers but only 6% of the matched sample. Finally, current housing situations differ somewhat, with 10% of matched callers living in shared housing compared to 12% of all callers. While the exact reason for any non-representative matching cannot be fully known, a few interpretations are consistent with the data. First, the consumer reference data misses some young adults. Second, it may also be more likely to miss other groups of people with limited paper trails, such as undocumented immigrants or people in shared housing. Third, we can still match to housing histories for a significant number of people with unstable and atypical housing situations.

4 Applications

4.1 A Housing Stability Outcome for Program Evaluation

Having validated the data, we can use consumer reference data to measure housing stability for unstably housed people. Within the HPCC sample, some people are referred to emergency financial assistance of roughly 1 month of rent paid to the landlord. Others are turned away if assistance is not available for them at that time. The prior papers argue that, conditional on eligibility and a few observable factors, referral to funds is effectively random. Thus, differences in outcomes between eligible callers referred to funds and those not referred to funds can be interpreted as the causal effect of emergency financial assistance. Evans et al. (2016) and Palmer et al. (2019) show that such funding reduces the likelihood of entering emergency shelter and being arrested for violent crime, respectively.

I test whether the consumer reference data can measure similar treatment effects on address stability. To measure address stability, I allow multiple address histories to link to any given HPCC record. I then generate an address move outcome with the Infutor address histories. I define an indicator for whether a household moves x months after the call. It is an indicator for whether the household matches to an Infutor record that includes an address with a starting or ending date x months after the call. For comparison, I similarly match the sample to shelter entry records as in Evans et al. (2016). Changes in addresses in the data clearly measure housing instability. Figure 5 shows the probability of an address change in a particular month relative to the call month. The sample includes all callers, both those referred to funds those not referred. This graph shows the monthly flow probability of moving. As is apparent, address changes spike at the time of calls to the HPCC. This makes sense and matches the pattern for shelter entry. Callers select into and are screened as eligible precisely because they face imminent risk of homelessness. The address histories reflect this fact clearly.

While both spike at the time of the call, shelter entry and address changes appear to measure different aspects of housing instability. Two points that can be seen in Figure 5 are worth noting. First, address moves are roughly twice as common as shelter entry. Second, the two measures rarely overlap. If we observe someone entering shelter in the standard administrative data, we almost never observe an address change in the consumer reference data at the same time. The fact that the third, gray line rarely separates from the x-axis in Figure 5 demonstrates that we rarely observe a person entering shelter and changing recorded addresses at the same time. Out of 7168 people, we observe 501 people change addresses and 163 enter shelter within 1 year, but only 13 people do both. These two facts follow intuition and what is known in the literature. Shelter entry is an extreme event affecting a small minority of even those people requesting emergency assistance, and such individuals are likely to totally lack a formal address. A different but much larger group of people moves in with family and friends or manages to lease a different unit. We see evidence of such patterns in the two datasets.

Differences between treatment and control groups are similar for shelter entry and address

changes. I compute whether a household has a recorded address change within x months of the call month. I then calculate the difference in the proportion of households who move in the treatment group versus the control group. I make similar calculations for shelter entry. Panel (a) of Figure 6 shows this treatment-control difference in the cumulative probability of shelter entry (left) and housing moves (right). The graph on the left essentially replicates the analysis of Evans et al. (2016) in the larger sample of Palmer et al. (2019). The cumulative probability of entering shelter drops precipitously by 1.0 percentage points (0.06 σ) in the group referred to funds relative to other eligible callers not referred to funds. Following the interpretation of Evans et al. (2016), the results suggest this effect persists over time, indicating that emergency financial assistance does not simply delay homelessness. The graph on the right shows similar results for housing moves as measured by consumer reference data, with a difference of 2.3 percentage points (0.09 σ). Consumer reference data appears to track these different and much more common forms of housing instability.

Shelter entry and address changes show different treatment effects, though, when controlling for important observable characteristics. The natural experiment studied in Evans et al. (2016) and Palmer et al. (2019) is conditionally random with treatment assignment depending on a few observable factors. For example, people requesting payment of back rent are more likely to receive assistance than those requesting security deposits for a new apartment. Thus, the prior papers control for these factors, regressing a housing stability outcome on a treatment dummy while controlling for these observable factors. Figure 6.b replicates this analysis. Notice that for shelter entry in the left panels, there is little difference between the unconditional difference in panel (a) and the regression-adjusted difference in panel (b). However, for address changes, controlling for observable factors attenuates the effects somewhat. This occurs because address moves are correlated with the control variables; for example, people requesting back rent are less likely to actually move and are move likely to receive funding. These results suggest an additional way in which address moves provide a complementary measure to shelter entry: different types of people have different relative likelihood of moving versus entering a shelter.

4.2 Measuring Residential Location

The address history data describe not only when people move but where they move. Since the data has national coverage of exact addresses, it can provide detail on moves ranging from individual blocks to states. This feature distinguishes consumer reference data from local administrative data sources. Many administrative datasets are limited to a particular jurisdiction or limit the detail of geography observed. Consumer reference data can track moves from an exact address to another exact address on the opposite coast. Consider two examples.

First, for the HPCC study, I can measure differences in residential location between those referred to funds and those not. I match the data as described in the previous section and measure residential location for the most recent address of the person as of June 2018. I then compute the difference between the proportion of the treatment group versus the control group in each ZIP code in Chicago. Results are shown in Figure 7. For reference, the bottom panel shows poverty rates

by ZIP code in Chicago. While this analysis shows results at the ZIP code level, the data can in principle examine such effects at a very detailed level down to tracts, blocks, and even individual buildings.

The data can also track address moves nationally. Consider a similar exercise for New Orleans residents affected by hurricane Katrina. As before, I identify the first address observed after hurricane Katrina for residents of two New Orleans ZIP codes. Figure 8.a shows the densities of movers to counties in Louisiana, Texas, Arkansas, Mississippi, and Alabama. While a large density of movers remain near New Orleans, the data track interstate moves to known destinations of Katrina evacuees, e.g. Houston. The same exercise can be conducted for the entire continental United States, tracking long distance moves to Seattle, San Francisco, Chicago, Detroit, and Washington, DC.[4]

Some research applications require following individuals with both neighborhood-level detail and multi-jurisdiction scope. Much of the debate regarding the effects of gentrification centers on how it affects the original residents of a neighborhood, but even residential location of original residents of a particular neighborhood can be difficult to track longitudinally when people move across borders. Consider one example from Washington, DC. The zip code 20010 in Northwest Washington DC includes parts of the Mt. Pleasant, Columbia Heights, and Parkview neighborhoods. Much of this zip code, especially Columbia Heights, has experienced rapid neighborhood change over the past two decades. The proportion of black residents in 20010 decreased from 46% in the 2000 census to 32% in the 2010 census and 25% in the 2013-2017 ACS. Meanwhile, the proportion of white residents rose from 23% to 43% and then 49%. I use the Infutor data to identify residents who lived in 20010 between the years 2000 and 2005, i.e. prior to much of the gentrification. Figure 8.b shows the most recent residential location of these people as of June 2018. Zip code 20010 is shown in red. Most people in the Infutor data who leave 20010 remain in the District of Colombia (the diamond shape in the middle of the figure). However, many exit to neighboring Maryland, a phenomenon that would be difficult to measure in many administrative datasets. Zip codes in Prince George's County to the east and Montgomery County to the north are common destinations. On the other hand, few people move southwest to Virginia. The scope and detail of consumer reference data is particularly well-suited to this type of task, tracking individuals as they change neighborhoods across three separate jurisdictions.

5   Practical and Ethical Considerations

Actually completing the above exercises using consumer reference data requires consideration of both practical and ethical challenges. This section considers three main challenges faced by any user of consumer reference data. First, data access must be obtained through a contracting process. Second, the data raise some ethical questions. Third, computational challenges must be overcome.

The contracting process matters because consumer reference data are provided by for-profit

---

[4]See Figure 2 in the appendix.

companies that charge fees for data access. This arrangement provides both advantages and disadvantages relative to other data sources. The advantage is that these companies intentionally curate the data for outside users. They actively encourage use of the data and provide it quickly. This differs from some other administrative data sources for which data access processes may be opaque and difficult to navigate. Of course, the trade-off is that researchers must raise funds to pay for access to consumer reference data. While the exact terms of such contracts are typically not public knowledge due to non-disclosure agreements, such data are certainly less expensive than special purpose household surveys. The details of such contracts also matter. As with any limited access data source, researchers should take care in considering the extent to which the contract allows for data access throughout the entire life of a research project, including access for replication purposes.

Careful attention to details of research design can help alleviate possible ethical concerns. Some analyses can avoid the most difficult ethical questions by using only a subset of the data. For example, all of the analyses in this paper use only data that Infutor provides to external customers for non-research purposes: address histories, names, month of birth, and year of birth. Given their availability to customers outside research, these data may be viewed in discussions of ethics as public information, which reduces their ethical sensitivity in institutional review and other ethical discussions. Such discussions become more complicated when using additional information, such as SSNs. Of course, careful attention to ethics is still important even when limiting attention to address histories. The present paper was approved by a local Institutional Review Board, and ethics affected the choice of analysis in several ways. To give a particular example, Figure 7 is aggregated to ZIP code and does not show individual points. This presentation was chosen to avoid identifying individual HPCC callers. More generally, combining address history with other data that is not publicly available or publishing results with detailed geographic information, such as maps, raise ethical issues. It is imperative that researchers address these issues for any given study in partnership with a relevant Institutional Review Board.

Working with consumer reference data also requires computing resources. Since the data are curated for external use, considerable data cleaning has already been completed: standardizing addresses, attaching multiple aliases of an individual to one housing history, etc. However, considerable computational work remains. For example, the Hurricane Katrina example in Figure 8.a requires several steps. First, a relevant extract must be pulled from the full dataset. The full June 2018 dataset used in this paper is 355 GB, stored as 52 separate state (plus DC and Puerto Rico) files according to where the person most recently lived. So, the first step is to extract records for all individuals from each state in the Infutor data who have ever lived in ZIP codes 70117 or 70130. Reducing the dataset in this way using a simple numerical identifier, such as a ZIP code, dramatically reduces computing time in later steps. Second, the exact analysis sample must be identified using a more detailed screen. In the Katrina example, this entails using address start dates and ZIP codes in combination to identify which individuals lived in the two ZIP codes before the hurricane and which address of theirs was the next address. Third, the researcher analyzes the now-manageable dataset using standard tools. Of these steps, the initial data reduction step

is most computationally intensive and requires either facility with tools for managing very large datasets or significant computational resources.[5]

6 Conclusion

Consumer reference data can be used to measure housing stability. I show that measures of housing moves from such data can replicate known timing and relative magnitudes for cases of housing instability caused by public housing demolitions in Chicago, Hurricane Katrina in New Orleans, and changes in apartment building management in Washington, DC. At a national scale, this data captures moves within the United States at a similar rate as comparable census-style data. Having validated the data, I demonstrate two new applications. First, I show how this data can provide a housing stability outcome for an impact evaluation of emergency financial assistance in Chicago. Second, I show how the data can track residential location within the city of Chicago and across the country for people exiting New Orleans after Katrina. The data follow people across neighborhoods covering the multi-state DC metro area when one neighborhood gentrifies.

The data have some limitations. Most fundamentally, a housing move is not necessarily always negative. The data do not distinguish between voluntary and forced moves. While characteristics of the unit and neighborhood can surmount this issue to a large extent, the data will always have less detail than a full in-person survey. More practically, the data have some imperfections. Individuals must have a sufficiently long history of independent consumer transactions to be in the data. Hence, children and many young adults do not appear in the data. There are also some indications that the data undercount Hispanic people and people in shared housing. Matching in the data is not perfect. Some addresses belonging to the same individual may show up as different people in the data if sufficient information does not exist to match them, and while the data can be matched to outside datasets, merges based on name, geography, and date of birth will be necessarily imperfect. All of these factors suggest that consumer reference data is not a cure-all.

Even so, consumer reference data has the potential to facilitate research on housing stability that would otherwise be impossible or too costly to undertake. Consumer reference data is cheaper than a survey, has fewer non-pecuniary barriers than administrative data, measures an outcome more common than shelter entry, and has a broader scope than address histories based on local administrative records. Such data will lower the cost of existing descriptive and impact evaluation work. Perhaps most exciting, programs and policies which are thought too costly to evaluate may be evaluated with this data. Consumer reference data may be useful when no other administrative data exists, participants move too frequently across administrative boundaries, subjects are spread across many jurisdictions, or sample sizes are too small to detect effects on rare outcomes like shelter entry. Consumer reference data can substantially add to the toolkit for measuring housing

---

[5]The present paper was conducted entirely in STATA 15 on a server with 256 GB RAM, dual 12 core Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz Haswell processors, and a 1.4 TB Solid State Disk. Since STATA rather inefficiently stores entire large datasets in working memory, the large amount of available RAM was important for the completion of this paper using that tool.

stability, even in extreme situations.

References

Aliprantis, D. and D. Hartley (2015). Blowing it up and knocking it down: The local and city-wide effects of demolishing high concentration public housing on crime. *Journal of Urban Economics 88*, 67–81.

Bruhn, J. (2018). Crime and public housing: a general equilibrium analysis. Technical report, Unpublished working paper.

Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review 108*(10), 3028–56.

Collinson, R. and D. Reed (2018). The effects of evictions on low-income households. Technical report, Working paper, New York University, Wagner School.

Corinth, K. (2017). The impact of permanent supportive housing on homeless populations. *Journal of Housing Economics 35*, 69–84.

Desmond, M. (2016). *Evicted: Poverty and profit in the American city*. Broadway Books.

Desmond, M., A. Gromis, L. Edmonds, J. Hendrickson, K. Krywokulski, L. Leung, and A. Porton (2018). Eviction lab national database: Version 1.0.

Diamond, R., T. McQuade, and F. Qian (2018). The effects of rent control expansion on tenants, landlords, and inequality: Evidence from san francisco. Technical report, National Bureau of Economic Research.

Diamond, R., T. McQuade, and F. Qian (2019). The effects of rent control expansion on tenants, landlords, and inequality: Evidence from san francisco. *American Economic Review 109*(9), 3365–94.

Evans, W. N., S. Kroeger, C. Palmer, and E. Pohl (2019). Housing and urban development–veterans affairs supportive housing vouchers and veterans homelessness, 2007–2017. *American journal of public health 109*(10), 1440–1445.

Evans, W. N., D. C. Philips, and K. J. Ruffini (2019). Reducing and preventing homelessness: A review of the evidence and charting a research agenda. Technical report, National Bureau of Economic Research.

Evans, W. N., J. X. Sullivan, and M. Wallskog (2016). The impact of homelessness prevention programs on homelessness. *Science 353*(6300), 694–699.

Frederick, T. J., M. Chwalek, J. Hughes, J. Karabanow, and S. Kidd (2014). How stable is stable? defining and measuring housing stability. *Journal of Community Psychology 42*(8), 964–979.

Garza, M. (1999). 9 high-rises at taylor homes slated to close. *Chicago Tribune*.

Goodman, S., P. Messeri, and B. O'Flaherty (2016). Homelessness prevention in new york city: On average, it works. *Journal of housing economics 31*, 14–34.

Gubits, D., M. Shinn, M. Wood, S. R. Brown, S. R. Dastrup, and S. H. Bell (2018). What interventions work best for families who experience homelessness? impact estimates from the family options study. *Journal of Policy Analysis and Management 37*(4), 835–866.

Humphries, J. E., N. Mader, D. Tannenbaum, and W. van Dijk (2019). Does eviction cause poverty? quasi-experimental evidence from cook county, il. Technical report, Unpublished working paper.

Jacob, B. A. (2004). Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in chicago. *American Economic Review 94*(1), 233–258.

Lucas, D. S. (2017). The impact of federal homelessness funding on homelessness. *Southern Economic Journal 84*(2), 548–576.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu (2013). Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American Economic Review 103*(3), 226–31.

Mills, A. and A. Giambrone (2017). Life is hell for tenants of giant dc slumlord sanford capital. *Washington City Paper 2-2-2017*.

Palmer, C., D. C. Phillips, and J. X. Sullivan (2019). Does emergency financial assistance reduce crime? *Journal of Public Economics 169*, 34–51.

Pollack, N. (2000). Kicking and screaming. *Chicago Reader, Feb 17*.

Popov, I. (2016). Homelessness programs and social insurance. *Unpublished manuscript. Stanford University, Palo Alto, CA*.

Rolston, H., J. Geyer, G. Locke, S. Metraux, and D. Treglia (2013). Evaluation of homebase community prevention program. *Final Report, Abt Associates Inc, June 6*, 2013.

Sandler, D. H. (2017). Externalities of public housing: The effect of public housing demolitions on local crime. *Regional Science and Urban Economics 62*, 24–35.

Schneider, M., D. Brisson, and D. Burnes (2016). Do we really know how many are homeless?: An analysis of the point-in-time homelessness count. *Families in Society 97*(4), 321–329.

Stergiopoulos, V., S. W. Hwang, A. Gozdzik, R. Nisenbaum, E. Latimer, D. Rabouin, C. E. Adair, J. Bourque, J. Connelly, J. Frankish, et al. (2015). Effect of scattered-site housing using rent supplements and intensive case management on housing stability among homeless adults with mental illness: a randomized trial. *Jama 313*(9), 905–915.

Table 1: Comparing average characteristics of Census and Infutor data, entire US

| | All | | Age 25+ | |
	ACS	Infutor	ACS	Infutor
Median Age	37	49	50	51
Proportion Female	0.51	0.45	0.52	0.45
Proportion High Rise	0.18	0.15	0.17	0.15
Proportion Moved in Past Year	0.15	0.12	0.13	0.12
–in PUMA	0.091	0.043	0.076	0.044
–out of PUMA, in State	0.026	0.050	0.021	0.051
–out of State, in US	0.021	0.011	0.018	0.011
–out of US	0.005	–	0.004	–
–location missing	0.011	0.028	0.010	0.029
Moved More than Once in Past Year	–	0.036	–	0.036

ACS estimates come from 2010 ACS micro data from IPUMS. Infutor values are for a 0.01% sample of all living individuals in the Infutor data whose history includes an address with valid date range spanning April 2010.

Table 2: Comparison of matched vs. unmatched call center records

	No Matched Record	Matched Record	Total
Age	38.2	41.7	40.3
Female	0.79	0.80	0.80
Black	0.83	0.90	0.88
Hispanic	0.12	0.056	0.080
Monthly Income, Thousands	1.11	1.13	1.12
SNAP Recipient	0.65	0.66	0.65
Veteran	0.026	0.029	0.028
Minors in HH	1.37	1.42	1.40
Adults in HH	1.39	1.42	1.41
Living Situation - Owner	0.026	0.031	0.029
Living Situation - Rent	0.83	0.86	0.85
Living Situation - Shared	0.14	0.10	0.12
Reason - Benefit Loss	0.11	0.14	0.13
Reason - Can't Afford Bills	0.035	0.030	0.032
Reason - Exit Shared Housing	0.065	0.048	0.055
Reason - Fleeing Abuse	0.013	0.0096	0.011
Reason - Job Loss	0.31	0.28	0.29
Reason - Exiting Shelter	0.029	0.023	0.026

Each value is a mean of a characteristic recorded by the Homelessness Prevention Call Center (HPCC) during the process of determining eligibility of the caller for financial assistance. Column 1 shows HPCC records not matched to any Infutor records. Column 2 shows HPCC records that match to at least 1 Infutor record. Column 3 combines the samples.
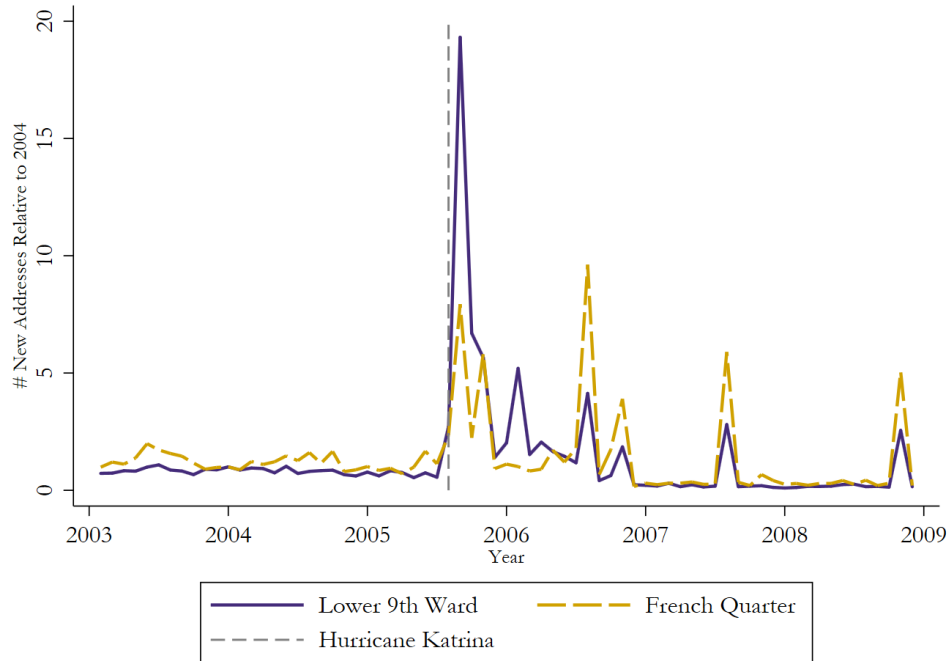
Figure 1: Frequency of housing moves from two New Orleans ZIP codes, relative to pre-Katrina levels
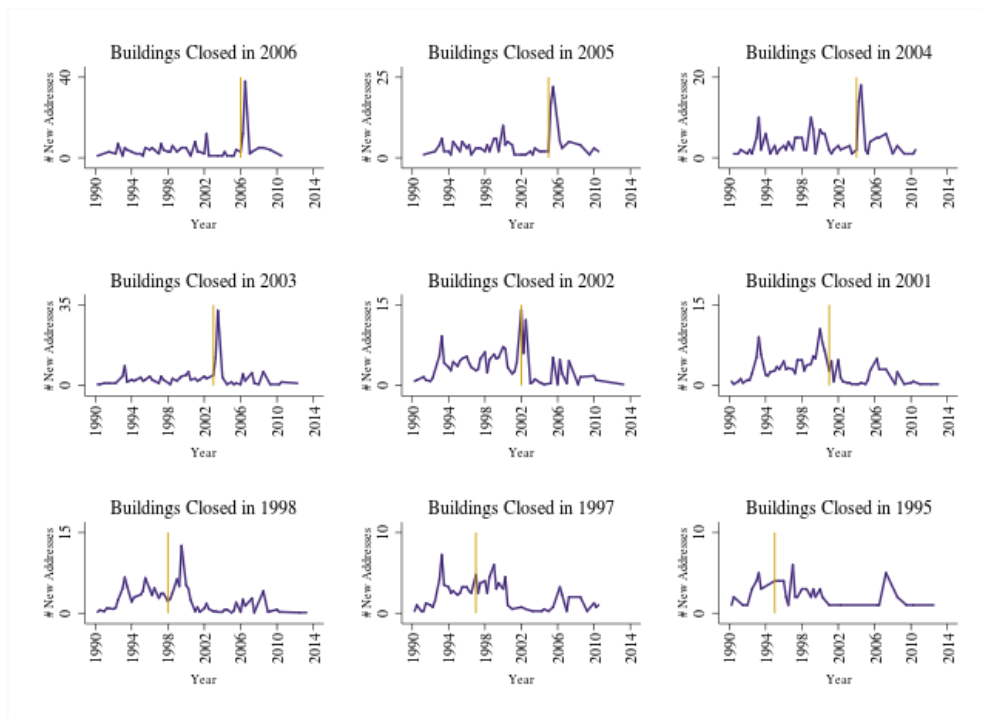

Figure 2: Number of housing moves out of Robert Taylor Homes, by year of building closure

21

Figure 3: Number of housing moves out of Terrace Manor, management change in 2012
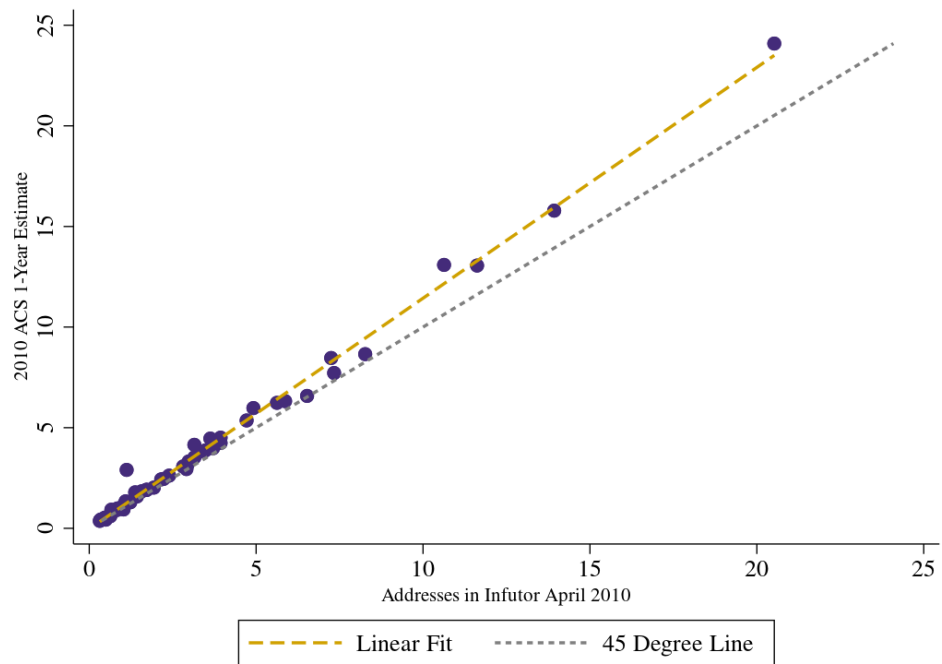

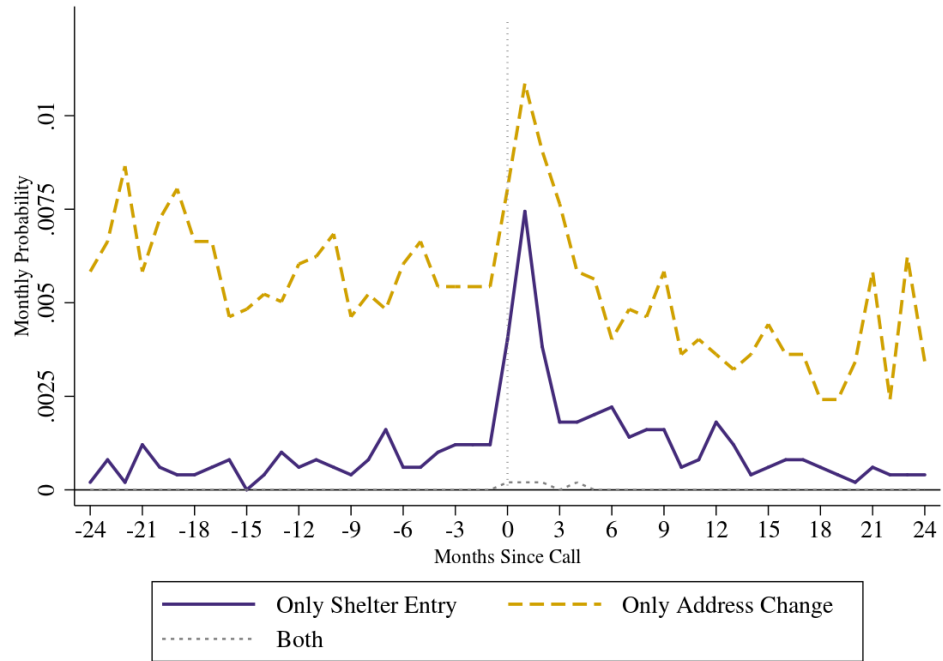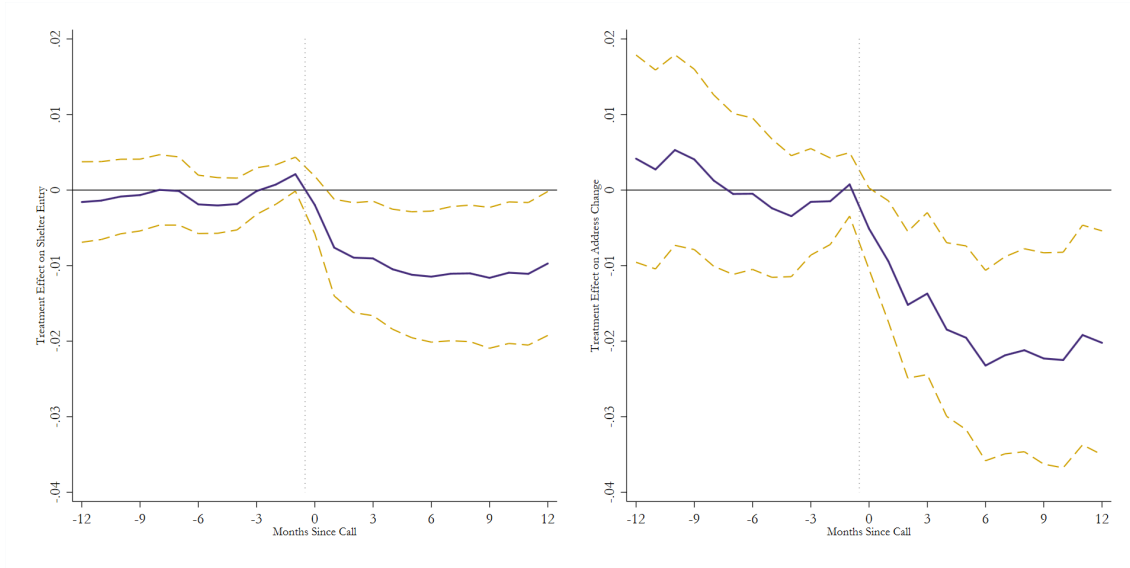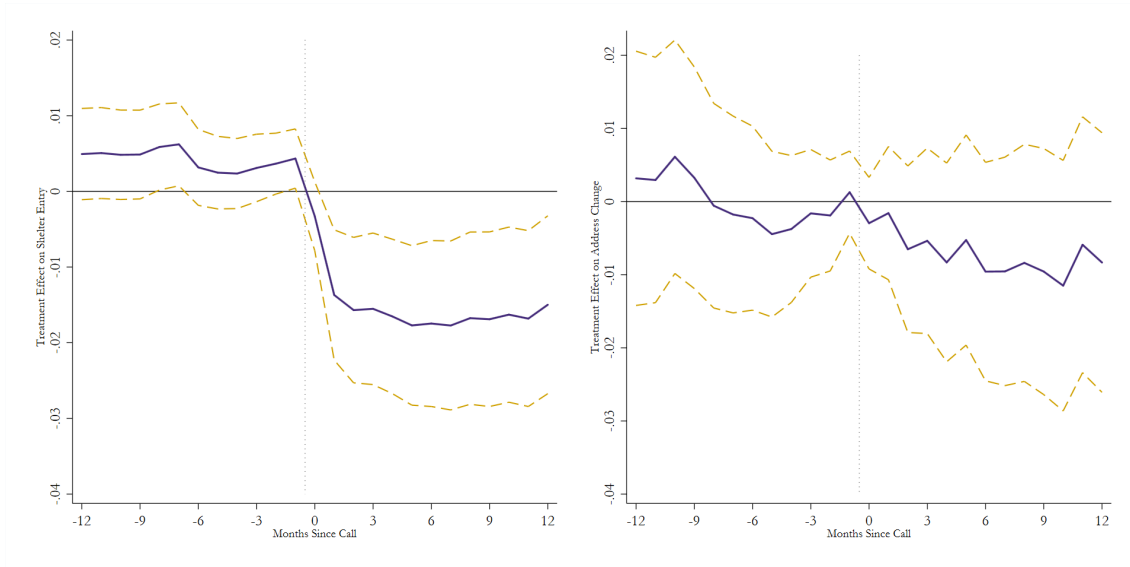Figure 4: State population 25+ (millions), ACS vs. Infutor

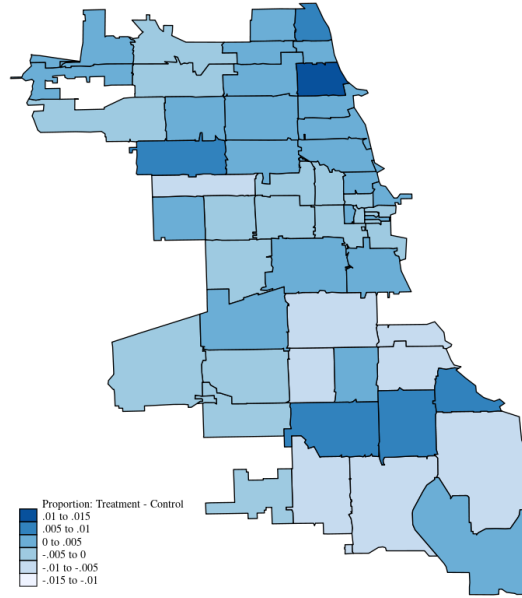Figure 5: Monthly flow probability of address moves and shelter entry, callers to HPCC

(a) Raw difference
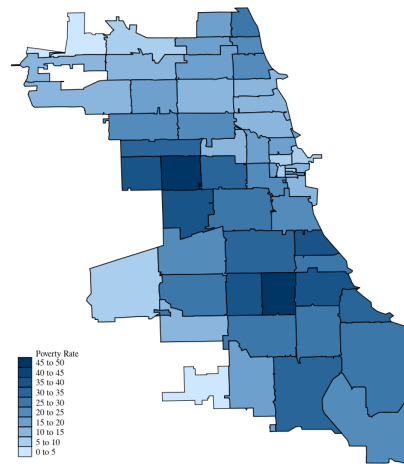

(b) Regression-adjusted difference

Figure 6: Difference in cumulative probability of housing instability, referred to funds vs. not referred
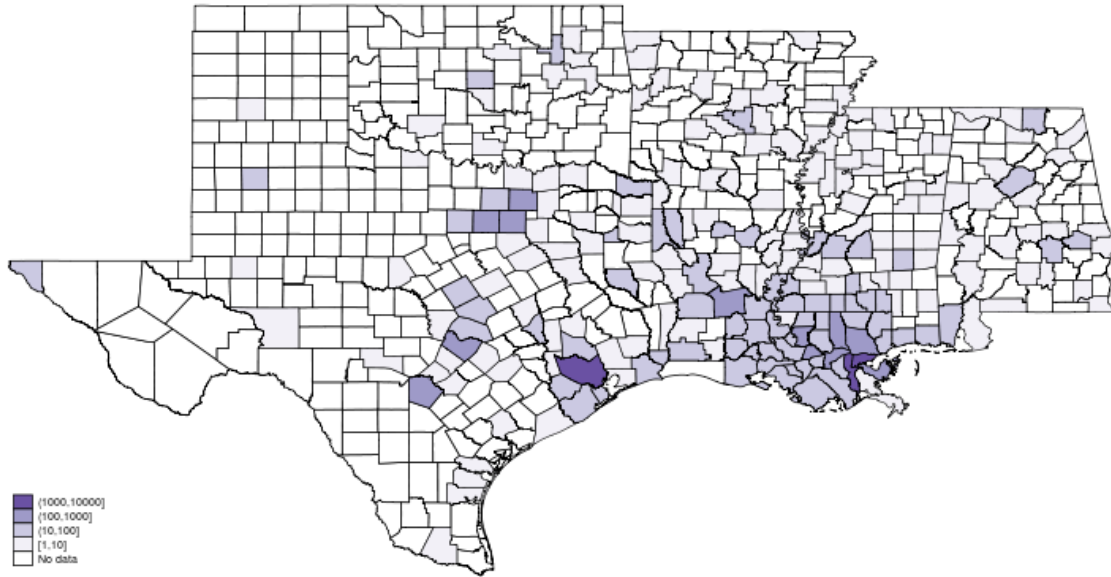
Proportion: Treatment - Control
- .01 to .015
- .005 to .01
- 0 to .005
- -.005 to 0
- -.01 to -.005
- -.015 to -.01

(a) Difference in most recent residential location, referred to funds vs. not referred

Poverty Rate
- 45 to 50
- 40 to 45
- 35 to 40
- 30 to 35
- 25 to 30
- 20 to 25
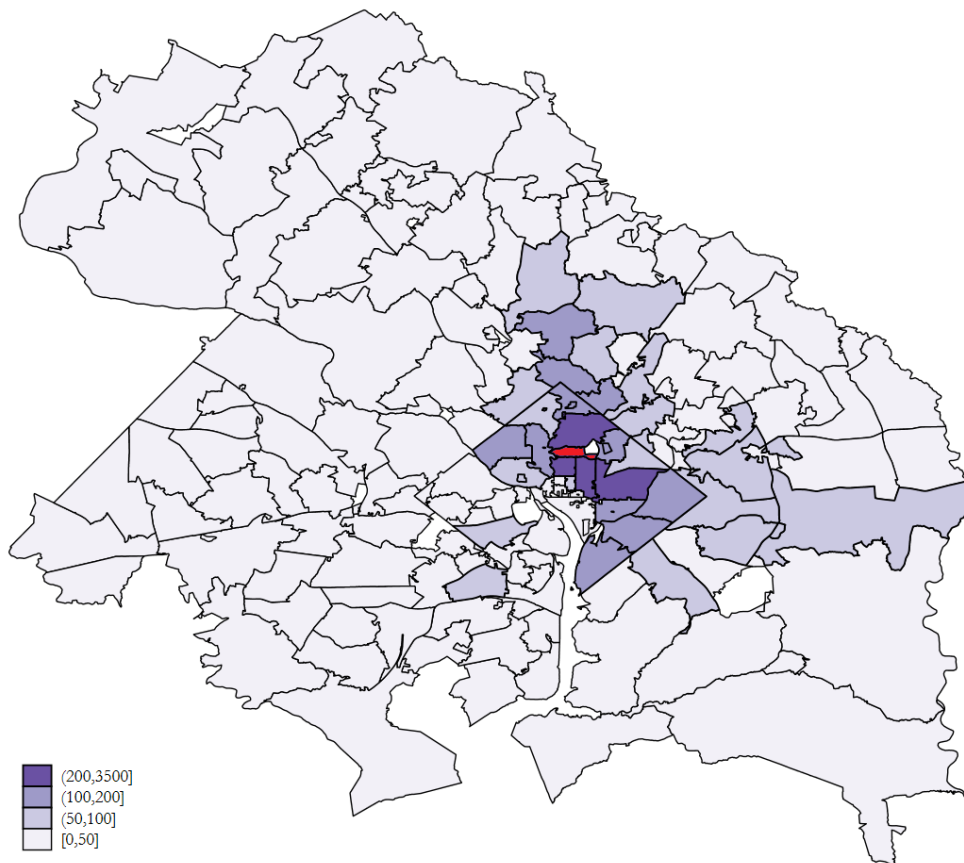- 15 to 20
- 10 to 15
- 5 to 10
- 0 to 5

(b) Poverty rate

Figure 7: Chicago ZIP codes

(a) First residential location after Hurricane Katrina, number of people by county

(b) Most recent residential location across DC metro area, people residing in zip 20010 between years 2000 and 2005

Figure 8: Applications to measuring residential location