DEBT COLLECTION LAB

January Advisors

Enriching Court Data to Identify Households in Consumer Debt and Eviction Lawsuits

August 2024

Jeff Reichman, January Advisors

Support for this project was provided by The Pew Charitable Trusts. The views expressed herein are those of the author(s) and do not necessarily reflect the views of Pew.

Summary

This study introduces a novel approach to enrich court docket data for debt and eviction lawsuits, revealing the broader household-level impact of these cases. Using this approach, researchers can associate cases between household members with more precision, leading to research that can analyze the effects of pooled income and shared debts among members of a household. We looked at a sample of 773 cases marked as Debt Claim or Eviction in Harris County, Texas, and found that 1-in-3 defendants had a household member with a Debt Claim or Eviction case filed against them. These household members accounted for an additional 443 cases attributable to the household, representing a 57% increase from the sample.

Introduction

In 2023, there were over 141,000 debt claim and eviction lawsuits filed in Harris County, Texas, representing an eviction rate of 10 evictions per 100 renter households and a litigation rate of 4.6 debt collection lawsuits per 100 people with a debt in collections. These staggering numbers underscore a growing crisis in consumer debt and housing instability, but they only tell part of the story.

Courts typically treat each case as a separate matter, rather than tracking a person or their household members across multiple cases. Subsequent research on eviction and debt cases also follows this structure and takes this case-centered approach, rather than a person-centered or household-centered approach.

For example, an eviction lawsuit might name a single defendant, but the actual eviction has an impact on the entire household. At the same time, a separate household member might be named in a debt lawsuit that involved family expenses. There is a potential relationship between these debts, but it is not possible to link people together into households using only court data.[1]

This experiment aimed to develop and test a method for identifying and associating debt and eviction cases at the household level. With this linkage, researchers can analyze household debts regionally (rather than nationally), identify upstream debts that occur prior to foreclosure or bankruptcy, and quantify the burden of consumer debt and eviction cases on the household, among other things.

We create these household linkages by enriching the data about the defendant with additional details that allow us to identify the same people across multiple addresses. We used the court data along

[1] The Center for Microeconomic Data at the Federal Reserve Bank of New York provides a nationwide overview of household debt and credit developments. However, this report is limited in its scope: it only provides data aggregated nationally, and for lawsuits, it is limited to bankruptcies and foreclosures. It does not include consumer debt or eviction lawsuits. ("Quarterly Report on Household Debt and Credit 2024: Q2," Federal Reserve Bank of New York, https://www.newyorkfed.org/medialibrary/interactives/householdcredit/data/pdf/HHDC_2024Q2).

with Trestle, a data broker, to pull current and historical addresses, unique personal IDs, and the names of up to five household members.

With the household member names, we were able to locate additional cases associated with the household. Our initial findings reveal a significant increase in cases associated with the household:

- 1-in-3 defendants have a household member with an Eviction or Debt Claim case filed against them.
- The cases against household members increases the original sample by 57%.

At the end of this paper, we propose additional research activities to refine these results and validate the trend. It is important to note, however, that data brokers are mostly unregulated, and their business is controversial. When used as tenant screening services, the information that data brokers provide have been used to discriminate against people.[2] Recently, there have been calls for regulation, with a specific focus on limiting the collection of location data and online activity.[3]

Although the experiment detailed in this paper was limited in its scope, new research ideas enabled by data brokers should be carefully considered in light of unintended consequences related to working with personal data. Researchers should treat public court data, enriched data provided by a data broker, and the household linkage as personally identifiable information (PII) and the work should be subject to IRB requirements.

The limitations of case data

In Harris County, consumer debt and eviction cases go through the Harris County Justice of the Peace Courts (JP Courts).[4] These courts make their case data publicly available, including defendant name and address.

Anecdotally, we know that some defendants have multiple debt and/or eviction cases. But the case information alone is not sufficient to reliably aggregate for several reasons:

- **Name matching is imperfect**. Using name alone, we don't know whether John Doe is the same person across multiple cases. There is no defendant ID number in Texas civil court like there is in Texas criminal court. When we do locate a John Doe in multiple cases, we can match based on both their name and address, but we run into problems in cases where John Doe has moved. This presents a particular problem for matching debt and eviction cases, as low-

[2] Boshart, Abby. "How Tenant Screening Services Disproportionately Exclude Renters of Color from Housing." Housing Matters, Urban Institute, 2022, housingmatters.urban.org/articles/how-tenant-screening-services-disproportionately-exclude-renters-color-housing.

[3] Emile Ayoub and Elizabeth Goitein. "Closing the Data Broker Loophole." *Brennan Center for Justice*, 13 Feb. 2024, https://www.brennancenter.org/our-work/research-reports/closing-data-broker-loophole.

[4] In Harris County, Eviction cases (http://www.jp.hctx.net/evictions/filing.htm) and Debt Claim cases (http://www.jp.hctx.net/suits/filingDC.htm) in amounts up to $20,000 are filed in the JP Courts. Prior to September 1, 2020, the limit was $10,000.

income individuals often move frequently and evictions inherently involve changing addresses.[5]

- Data entry is unreliable. Sometimes names and addresses are misspelled or mistyped. This requires fuzzy matching to identify close matches, and a manual review of the data to ensure high quality matches.
- Case data lacks household information. The courts do not collect and publish supplementary information, such as household members, race, ethnicity, or age.

By itself, the case data is insufficient for household analysis. However, by enriching the data from a third-party data broker, we can use the results to gain additional insights about the defendant and their household.

Methodology for household matching

This is a proof-of concept study to identify cases related in a household. We started with data from the Harris County JP Courts, looking at cases filed over a 14-year period between January 4, 2010, and February 8, 2024. Our dataset consists of cases categorized as Debt Claim or Eviction.

We started by cleaning and standardizing the presentation of defendant names. For instance, we converted names formatted as 'last, first' to a uniform 'first last' format.

Next, we selected a random sample of cases and ran the defendant names and addresses through Trestle, a third-party data enrichment service.[6] Trestle matched defendants based on name variations and address history, and returned enriched data which included estimated age, gender, contact information, and the names of household members.

Trestle provides a unique person ID of the original query, as well as the names and person IDs of up to five household members. This allowed us to delve deeper into potential connections between the original defendant and other cases on the docket.

We used the names of household members and fuzzy matching[7] to identify a potential universe of additional cases in the case data. Each of these potential matches underwent manual review to

---

[5] Robin Phinney, "Exploring Residential Mobility among Low-Income Families," *Social Service Review*, vol. 87, no. 4, Dec. 2013, pp. 780-815. The University of Chicago Press.

[6] We selected Trestle as our data broker after evaluating the data requirements for this research, the accuracy of results, the programmatic methods for querying the system, and the overall costs. We appreciated that Trestle did not provide information that we did not want, limiting the personally identifiable information made available for this study. The Find Person API from Trestle contains over 1.79 billion name/address linkages with a 95%+ accuracy rate, and is billed based on volume at seven cents per match. https://trestleiq.com/find-person-api/

[7] Fuzzy matching looks for similarities between two names without seeking an exact match. For this experiment, we experimented with a range of match tolerances to narrow the list of possible matches. Ultimately, we settled on a tolerance of 0.1 which will permit string matches with small variations, such as a middle initial, suffix, or a minor misspelling.

confirm accuracy, where we removed non-matches and common names that yielded more than 10 results.

Finally, we ran these names and addresses of the refined list of household members back through Trestle. This step was crucial for verifying the connections, because we filtered the results based on matching the PersonID of the household member and the PersonID of the new query result. This method uncovered known cases filed against the household member of the defendant.

Results from Harris County

Using Harris County Justice of the Peace Court cases filed over a 14-year period between January 4, 2010, and February 8, 2024, we looked at three groups of people:

1. A random sample of 200 people with a Debt Claim case filed against them. In Harris County, the Debt Claim case category describes consumer debt cases in amounts up to $20,000.[8] These cases are often filed by debt buyers, credit card companies, and other personal loan providers.

2. A random sample of 250 people with an Eviction case filed against them. The Eviction case category describes real property possession cases in amounts up to $20,000, with almost all cases involving residential real estate rentals and the failure to pay rent.

3. A random sample of 700 people with both a Debt Claim and an Eviction case filed against them within a two-year period. This sample highlights people who are facing multiple financial difficulties over a short period of time.[9]

The results were consistent across all three groups. In all, we found 773 matches out of the initial sample of 1,150. Of these matches, we identified an additional 443 debt or eviction cases filed against household members.

	Debt Sample	Eviction Sample	Debt + Eviction Sample	Total
Original lookup *Random sample of cases from the original dataset*	200	250	700	1,150
Original case matches *Records where Trestle provided a match based on name and address*	**140**	**109**	**524**	**773**
Number of debt or eviction cases filed against household members *Some household members have more than one case*	**65**	**56**	**322**	**443**

[8] Prior to September 1, 2020, the Debt Claim and Eviction thresholds were $10,000. (http://www.jp.hctx.net/suits/filingDC.htm)

[9] This sample was derived by matching exact names and addresses across the Debt Claim and Eviction case datasets.

% additional cases for the household *The % of new cases added through household case matches*	46%	51%	61%	57%

Table 1: Cases filed against household members. In each sample, we identified a significant number of additional cases filed against household members.

Of the household members with any cases filed against them, between one quarter to one third have two or more cases. This indicates a deeper entanglement with debt and poverty than the case data alone would suggest.

	Debt	Eviction	Debt + Eviction	Total
Household member matches *The number of household members with any cases filed against them.*	42	38	192	272
Household members with 2+ cases	11	11	66	88
% of household members with 2+ cases	26%	29%	34%	32%

Table 2: Household members with multiple cases. Between one quarter to one third of household members have more than one case filed against them.

Refining and improving the methodology

This represents a first approach at assembling a household-level picture of Debt Claim and Eviction lawsuits. The results of this experiment are clear, but the methodology could be refined to achieve more precise results. Limitations of this approach include:

- The household count doesn't take into account names that appear 10 or more times in the case history. Some common names have hundreds of matches, and individual lookups for each name would be cost prohibitive.
- This approach does not take into account whether the household member was living with the original defendant at the time of the household member's lawsuit. In other words, we have not looked at whether both cases happened while the parties lived together.
- Unless a suffix is specified in the court record, we cannot differentiate between elder and younger generations that share the same name. We controlled for this in the case count by only allowing one match per case number. And while we know that match belongs to the household, we don't know exactly which member of the household was named in the lawsuit.

What does this approach enable?

This approach opens up new ideas for research that were previously difficult to explore. By leveraging enriched data and precise household matching, researchers can investigate the broader implications of household financial interdependence. This includes looking at how shared debts impact household financial and housing stability, and analyzing the role of pooled income in managing financial risks. Example research topics include:

Debt accumulation by household. Linking household members can shed light on debt accumulation patterns within families. Researchers could identify financial interdependencies and identify patterns or triggers that lead to multiple household members facing lawsuits. Researchers can also assess the types of debts that are shared or transferred among household members.

Delivering and evaluating holistic support services. Support services—such as financial counseling, legal aid, and other social services—affect all members of a household. Service providers can use this information to identify vulnerable households and offer holistic and coordinated support.

Furthering demographic disparity research. Analyzing demographic disparities in debt and eviction cases at the household level provides a more nuanced understanding of economic and social inequalities. Traditionally, research has focused on individual or neighborhood-level disparities, using case data, Census/ACS data, or a combination of both. However, this may overlook household dynamics. With this data enrichment approach, researchers can investigate how factors such as race, ethnicity, age, and gender interplay within households to influence the likelihood of facing multiple lawsuits.

Analyze the role of pooled income. Pooled income is a crucial factor in understanding household financial resilience. By leveraging enriched data to examine how households combine their financial resources, researchers can analyze how pooled income affects the house. This research can explore scenarios where pooled income provides a buffer against financial shocks, such as job loss or unexpected expenses, preventing households from falling into debt. Conversely, it can also identify situations where shared financial obligations and debts create additional stress.

Preliminary gender analysis at the household level. Through data enrichment, we obtain gender and age range[10] information. Using that data, along with other data sources, we can explore how the gender of the defendant in eviction and consumer debt cases aligns with other established patterns.

Potential next steps

Engage stakeholders. Through interviews and candid feedback, determine whether this information is useful to researchers, courts, legal aid, social service providers, and other consumer/tenant advocates.

[10] Trestle categorizes gender as Male, Female, or unspecified. It provides age in five-year ranges, such as 31-35 or 36-40.

Refine the methodology. Dedicate additional time and effort to refining the methodology to account for the above-described limitations. This may include:

- Using new and emerging methods to refine individual matches from common names and suffixes, as well as replace manual review in the fuzzy matching process. This must be supplemented with oversight to ensure any automation does not lead to false positives.
- Refining the address matching process by removing text matching and replacing it with point-in-polygon geospatial matches between the geocoded defendant addresses and other geographic information, such as building and parcel data.
- Implementing stricter match criteria by refining the dates of the study period, or the dates of the household member's tenure at the original case address.

Incorporate additional data sources. Currently, we rely on a single data broker for enrichment. In an effort to corroborate existing matches and identify new ones, we could add one or more new data providers to the lookup process.

Replicate the household matching process in different environments. The results discussed here are promising, but this method has not been tested extensively. There are "hard costs" associated with the matching process that requires modest funding to scale. With that funding secured, we recommend:

- Larger tests (1000+ sample) in Harris County.
- Smaller tests (100+ sample) in multiple jurisdictions across the country.
- Samples that include common names, which may increase household matches.

Determine populations of people who have both debt and eviction cases. There is still an outstanding question of how many matches exist between the debt and eviction datasets. Our approach to create samples for this research looked for exact name and address matches, but using this novel approach of data enrichment and PersonID matching, we can arrive at more precise estimates.