# Generating household- and sub-state-level estimates of rental arrears in the COVID-19 pandemic

*Aaron Dibner-Dunlap, Staci Sutermaster, Peter Smittenaar, Sema Sgaier*
Surgo Ventures, July 2021

Explore the data: www.precisionforcovid.org/rental_arrears

Contact details:
- Technical inquiries: aarondibnerdunlap@surgoventures.org
- Media inquiries: bethanyhardy@surgoventures.org
- General inquiries: covid19@surgoventures.org

# Estimation Procedure: Overview

Our goal is to generate reliable local area estimates of rental arrears; however, our primary data source - the U.S. Census Bureau's [Household Pulse Survey] - is missing a fine geography indicator (like county), and indeed was not sampled to be used for sub-state estimates.

Our solution is to model a renting household's likelihood of being in arrears from the source data. Then, using this model, we predict likelihood of being in arrears for each of 344k renting households in a much larger and more representative dataset, the 2019 American Community Survey. We can then aggregate rental arrears up to several sub-state-level geographies, including county, using a combination of weighted means and geographic crosswalk files.

This approach has been used in other settings, notably to estimate local health outcomes using the Behavioral Risk Factor Surveillance System (Zhang et al., 2014), and is consistent with methodologies known collectively as unit-level small area estimation (Parker et al., 2020).

# Data Sources

## Household Pulse Survey Microdata

Our primary data source for predicting arrears is the Household Pulse Survey (hereafter, HPS), a biweekly household survey administered by the U.S. Census Bureau. HPS has had three distinct phases:

| Phase[1] | Dates Run | Wave Periodicity | Relevant changes from last phase |
|---|---|---|---|
| Phase 1 | April 23, 2020 - July 21, 2020 | Weekly | -- |
| Phase 2 | August 19, 2020 - October 26, 2020 | Every two weeks | Questions on Living Quarters and Receive Social Security Assistance added. |
| Phase 3 | October 28, 2020 - March 29, 2021 | Every two weeks | None |

[1] Source: U.S. Census Bureau, Household Pulse Survey. Available at https://www.census.gov/programs-surveys/household-pulse-survey/data.html (accessed 20 July 2021).

| | | | |
|---|---|---|---|
| Phase 3.1 | April 14, 2021 - July 5, 2021 | Every two weeks | Household unemployment reference period changed from “since the beginning of the pandemic” to “In the last four weeks”. |

The Census Bureau releases Public Use Microdata Files (PUF) very rapidly--two to three weeks after the conclusion of data collection--to enable rapid public consumption of the evolving results.

The survey has between 42,000 and 133,000 respondents per wave, sized to achieve reliable weighted estimates within 3 percentage points for most states and major metropolitan areas. Of these respondents, between 16 and 25 percent are renters, weighted to represent between 20 and 30 percent of American households (see Fig. 1, below). In practice, the household weights do not take rentership into account during calculation, and so may not reliably reflect the actual proportion of renters in a state. Therefore, while we use the HPS household weights and replicate weights to estimate survey means (per Census Bureau Bureau guidelines[2]), we do not rely on HPS weights for our modelling.

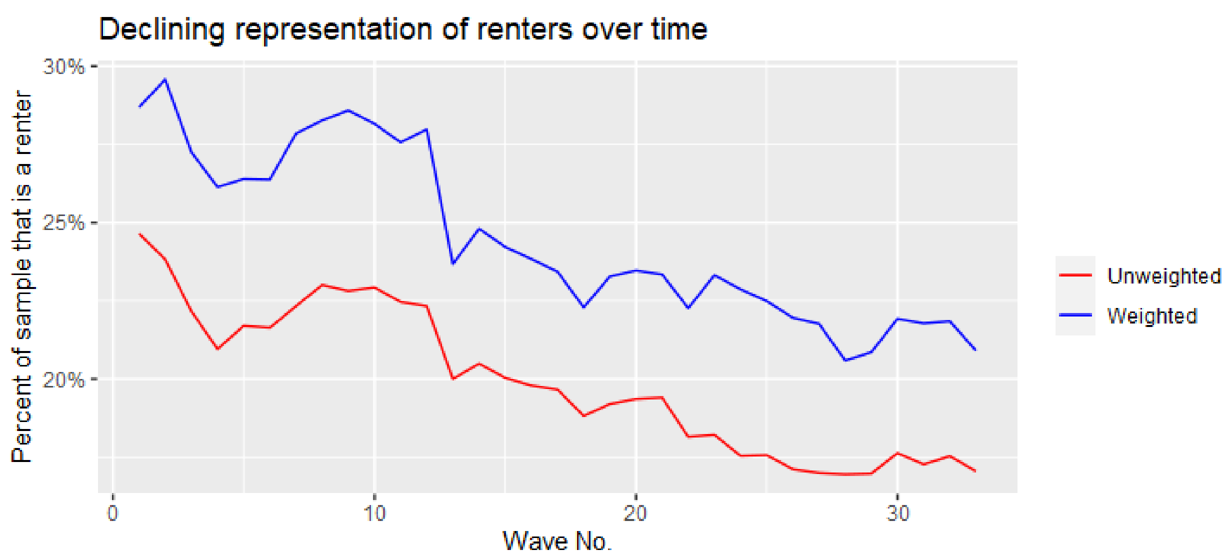


*Figure 1. Percent of sample that is a renter, by HPS wave. Red is unweighted, or observation-level percentage. Blue is weighted using household weights. “Renter” is defined as a household that self-reports as renting and typically pays positive dollar rent. Source: Authors’ calculations using Household Pulse Survey microdata.*

2 https://www2.census.gov/programs-surveys/demo/technical-documentation/hhp/Phase3-1_Source_and_Accuracy_Week_33.pdf, accessed 21 July 2021

## Definition of Rental Arrears

We are interested in modeling the likelihood of a household to be in arrears on rent; therefore our main outcome is the binary response to the question: "Is the household currently caught up on rent?" It is asked, unchanged, in every HPS wave, of any renting household who pays nonzero rent. We define a household as being in arrears if they answer "No" to this question.

HPS does not ask how many months behind on rent a household is, their monthly rent, or total arrearage, so we rely on our analysis and supporting assumptions to inform these metrics, described further below.

## American Community Survey Microdata

To estimate arrears for a dataset with high geographic granularity, we choose the 2019 American Community Survey (ACS) microdata 1% sample (Ruggles, Steven et al., 2021). The raw data have about 1.25 million respondents, of which nearly 344,000 are renters who pay nonzero rent.

Instead of counties, ACS microdata are assigned to a Public Use Microdata Area (PUMA) - a geographic area of around 100,000 population. At the end of our estimation process, we use a PUMA-to-county crosswalk file to obtain final county-level estimates.

Since ACS is also administered by the U.S. Census Bureau, many of the demographic variables are worded in a similar manner as the HPS; living quarters is one example, where the categories are easy to harmonize. Nonetheless, the different field conditions between the ACS and HPS mean that both response bias and nonresponse bias may diverge, which we acknowledge may have an impact on the quality of our estimates, but one that is difficult to measure.

## Other Data Sources

For state-level unemployment data, we rely on the U.S. Bureau of Labor Statistics' monthly Local Area Unemployment series.

For other state-level covariates such as race & ethnicity makeup, housing costs, poverty status and SNAP benefit participation rates, we rely on ACS 5-year estimates (2014-2019).

# Estimation Pipeline

Estimates are calculated separately for each survey release, which we call "waves" but has been referred to as "weeks" elsewhere, including by the Census Bureau.

## 1. Download and preprocess data

The first step of the process is to download the HPS microdata and survey replicate weights via http. We transform the data to create our main explanatory variables for modeling (for example, we create 6 distinct age categories rather than use raw age).

## 2. Impute missing HPS values

Since HPS responses are voluntary, the data come with a minimal but nontrivial amount of missingness, especially for questions later in the survey form (see Fig. 2 below). Logistic regression, our main modeling tool, cannot handle missing data without special processing; therefore, we chose to impute our main covariates via Multiple Imputation by Chained Equations (White et al., 2011), using the R package MICE (Buuren and Groothuis-Oudshoorn, 2011). We then pooled the imputed values by selecting the modal response, assumed to be the most likely response. We do not impute our main outcome, rent delinquency, to prevent estimation error from modeling an unknown dependent variable. Respondents who failed to answer this question were excluded from analysis.

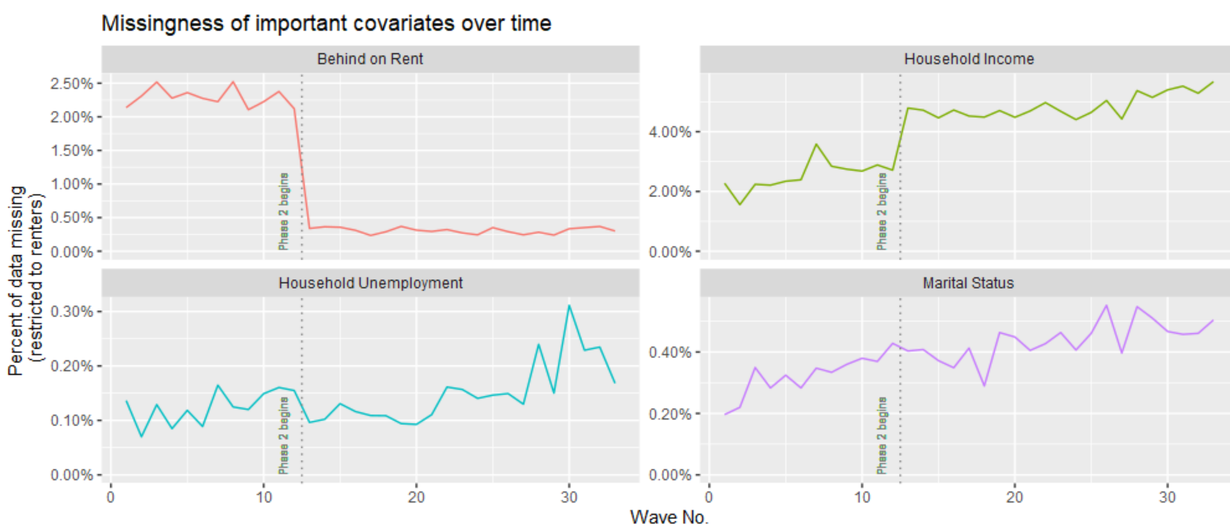


*Figure 2. Percent of the final HPS sample (renters only) that is missing some key variables, by survey wave. Missingness for our main outcome, Behind on Rent, drops dramatically after Phase 1. Household income typically has the highest missingness; it is one of the last questions in the questionnaire. Source: Authors' calculations using Household Pulse Survey microdata.*

## 3. Fit primary regression: rental arrears

As a reminder, our main outcome is the binary response to this question: "Is the household currently caught up on rent?" In this step we fit a logistic mixed effects model to estimate the probability of a household answering "No".

To build the model, we first restrict the data to renters who typically pay nonzero rent (i.e., they don't reside for free). This is typically about 20% of the sample (see Fig. 1, above). We further restrict the sample to respondents who have answered the primary question, which drops another 0.5%-2.5% respondents, depending on the wave (see Fig. 2, above). We are left with a final sample size of between 9,700 and 29,000 households each wave, with most samples from October 2020 onwards containing 12,000-14,000 households.

Using this dataset, we fit a mixed effects logistic regression model using the following categorical independent variables, with no interactions:

- Respondent-level:[3]
  - Age
  - Gender
  - Race & ethnicity
  - Education
  - Marital status
- Household-level:
  - Household income in 2019
  - Living quarters (added in Wave 13)
  - Whether Social Security assistance is being provided (added in Wave 13)
  - Number of children in the household
  - Census region
  - Whether or not the household has experienced unemployment since the beginning of the pandemic (Waves 1-27) or in the last four weeks (Waves 28-present).
- State-level random intercepts.

The final specification is identical for all waves (except for the inclusion of the 2 additional variables noted above as of wave 13). It was created by testing several specifications and combinations of input variables and selecting the model with the lowest AIC score.

In LME4 notation,[4] we estimate:

$$BehindOnRent \sim age_{factor} + race.eth_{factor} + education_{factor} + marit.status_{factor} + female + kids.in.hh_{factor} + hh.income_{factor} + census.region_{factor} + living.quarters_{factor} + receive.ssa + unemp.hh + (1 \mid state)$$

Fitted model results are available upon request.

[3] Since each household is sampled at most once per wave, we assume that the respondent assumes the role of household head, crucial for later projections with ACS data. Source: https://www2.census.gov/programs-surveys/demo/technical-documentation/hhp/Phase3-1_Source_and_Accuracy_Week_33.pdf, accessed 21 July 2021

[4] (Bates et al., 2015)

## 4. Fit supporting regression: household unemployment

The dataset on which we predict rent arrears--2019 ACS microdata--does not include timely measures of household unemployment during the pandemic. However, this binary measure is the single most important predictor of rental arrears according to our HPS-derived models (see Fig. 3 below for an indication of its importance). Therefore, it behooves us to impute or predict an additional feature on the ACS microdata: how likely each ACS microdata respondent is to have become unemployed during the pandemic.

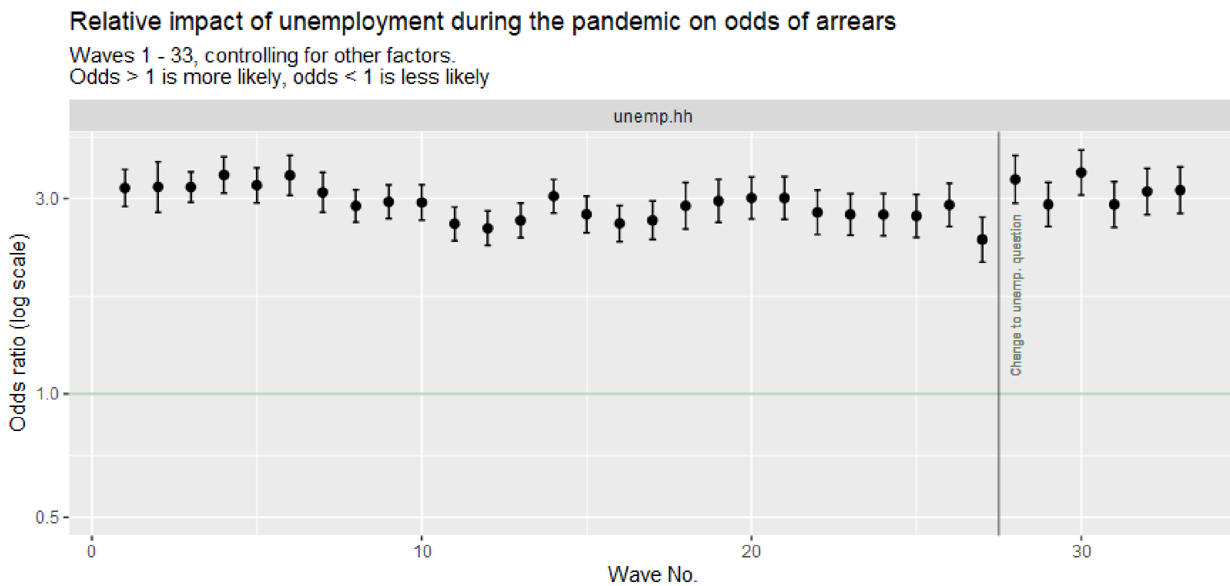


*Figure 3. We plot the odds ratios (exponentiated coefficients) for household pandemic unemployment (from our larger multivariate model) for each of HPS waves 1-30. Error bars represent 95% confidence intervals. All of the odds ratios are large and the confidence interval is far removed from one, indicating that it is a highly significant predictor for each wave. Source: Authors' calculations using Household Pulse Survey microdata.*

In order to incorporate some information from recent household unemployment, we build an auxiliary logistic model to estimate this parameter from our HPS microdata using static demographics (many of which are also used in our main arrears model). We use the single best linear unbiased predictor (BLUP) from this auxiliary model, transformed to express the predicted probability of household-level pandemic unemployment (between 0 and 1) and insert it into our ACS microdata in place of the missing binary "recent household unemployment" variable.

In essence, we treat "recent household unemployment" as an entirely missing variable suitable for multiple imputation. However, unlike a typical MICE strategy of using the modal predicted value (either 0 or 1) as the insertion value, we accommodate the uncertainty of estimating an unknown variable by using predicted probability (between 0 and 1) as the insertion value. This attenuates our estimates toward the population mean, but, as we will show, it improves the quality of our estimates over not using it.

To know the impact of this strategy on our arrears estimates, we ran a test using HPS responses whereby we predicted whether a respondent was in rental arrears. In Figs. 4a and 4b below, you see model metrics of household-level predicted probabilities of arrears using three techniques: full information (FI, blue, considered optimal), an identical model except it drops unemployment as a predictor (MU, green), and our estimation strategy using predicted unemployment (PU, red). You can see that, in all waves, estimation using PU matches or outperforms MU, meaning this additional feature is a worthwhile addition. The mean absolute errors for PU, a test of the calibration of our predictions, are extremely close to the FI model. Interestingly, the impact on MAE is larger than on AUC; we interpret this to indicate that the magnitude of misestimation declined, but the relative ranking of the respondents along the probability spectrum did not change.

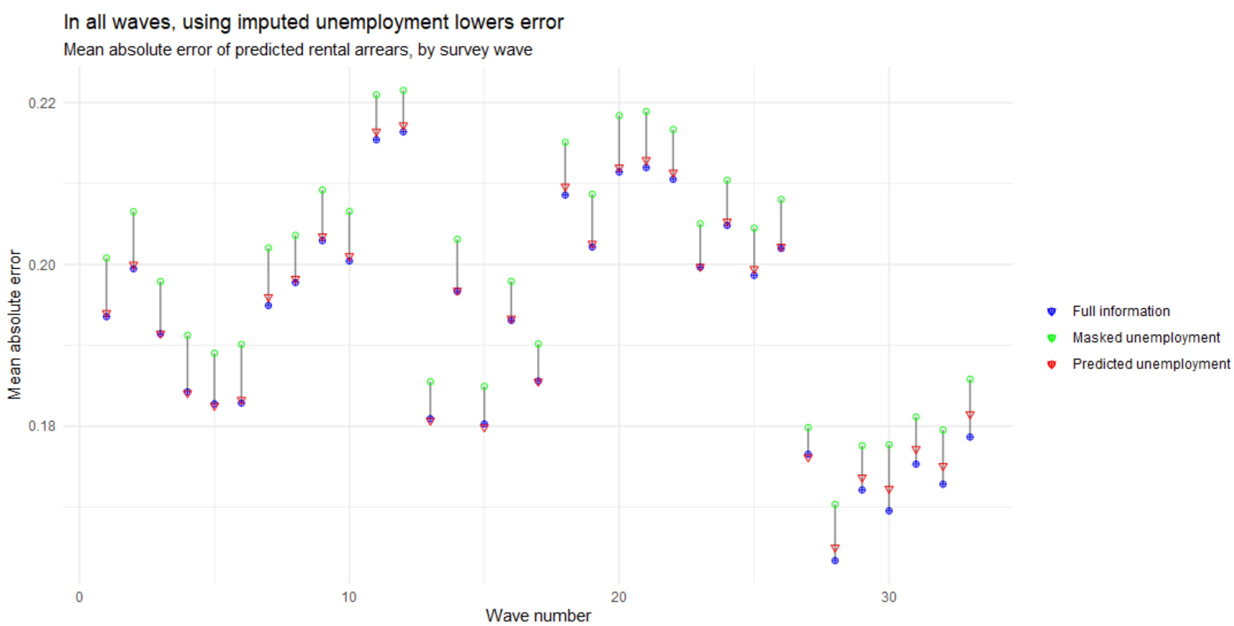


*Figure 4a. Mean absolute error (MAE), which is the absolute value of predicted arrears minus true arrears, for HPS waves 1-33 using three estimation strategies: full information (FI, blue, considered optimal), an identical model except for dropping unemployment as a predictor (MU, green), and our estimation strategy using predicted unemployment (PU, red). Lower MAE is better. Source: Authors' calculations using Household Pulse Survey microdata.*

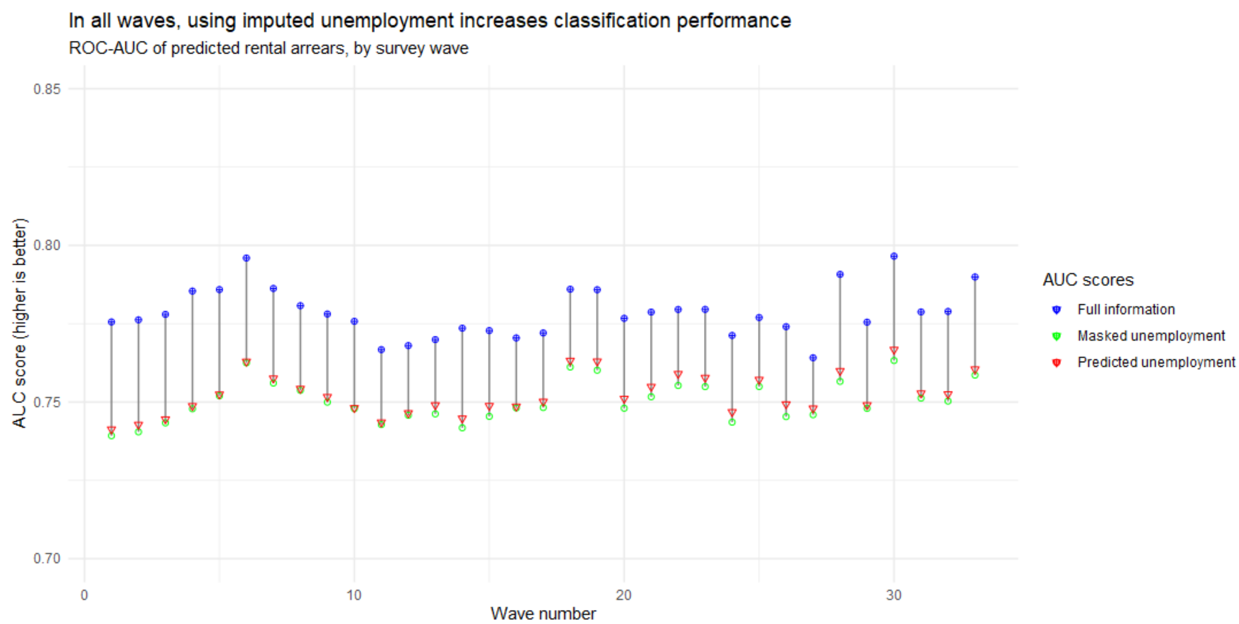


*Figure 4b. Area under the receiver operating curve (ROC-AUC, or simply AUC), a measure of classification performance, for HPS waves 1-33 using three estimation strategies: full information (FI, blue, considered optimal), an identical model except for dropping unemployment as a predictor (MU, green), and our estimation strategy using predicted unemployment (PU, red). Higher AUC is better. Source: Authors' calculations using Household Pulse Survey microdata.*

Since imputing unemployment in this manner utilizes many of the same covariates already present in the main model, another way to think of this method is as building nonlinearity into otherwise linear predictors, since each variable may have different relative weights in our main and auxiliary models.

## 5. Build ACS dataset for estimation

We predict arrears using the 2019 ACS microdata 1% sample, restricted to renters who pay nonzero gross rent. This yields 343,915 respondents who collectively represent nearly 42 million renter households in the US when using ACS household weights.

To be able to feed these data into our fitted models, we featurize the ACS data identically to HPS. For individual-level covariates (education, age, sex, race/ethnicity) we take the household head's information.

We generate predicted probabilities for recent household unemployment from our auxiliary model.

## 6. Generate estimated arrears via parametric bootstrap

To estimate likelihood of household arrears with standard errors, we use a parametric bootstrap to simulate the likelihood of generating our estimates from multiple survey populations (Luke, 2017). To accomplish this, we:

1) Resample the original HPS data, with replacement.
2) Fit a new model for rental arrears using the resampled HPS data
3) Generate predicted arrears for our ACS microdata

We do this 200 times for each wave, resulting in 200 estimates for each ACS respondent per wave.

Note that we do not refit our auxiliary model (unemployment) as part of this process. We chose to do this because we consider unemployment to be imputed using the best available data (all original HPS observations), and therefore our "best guess" at likelihood of unemployment. Furthermore, we already incorporate the uncertainty of our unemployment estimates by using a probability rather than the likeliest response.

We take the mean and standard deviation of the 200 estimates to produce a single estimate for the household (with error). We also retain the raw bootstrap results for aggregation to larger geographies.

## 7. Estimate auxiliary outcomes

Although we predict the probability of the household being behind on rent, we are also interested in the amount that they owe, in dollars. To estimate this, we need to know two things: their monthly rent obligations, and the number of months they are behind.

For monthly rent, fortunately ACS microdata contains the variable RENTGRS which is the monthly gross rent, inclusive of utilities. According to IPUMS USA, this measure is considered more comparable than pure contract rent, as the inclusion of some utilities in lease agreements varies widely.[5] Therefore, it is better to include all utilities.

For the number of months in arrears, we exploited the sampling strategy of Phase 1 of the Household Pulse Survey, which allowed for some resampling of households across waves. (Phases 2 and beyond have no shared samples.) Our objective was to estimate the likelihood a household was in arrears for multiple consecutive months.

We examined households with identical IDs that appeared in two non-sequential waves that are 1 month apart (e.g., waves 4 and 6, which ran from May 21-26 2020 and June 4-9 2020, respectively) to calculate the probability of a renter being behind in a prior month if they were also behind in the current month. Over 3 months (April-May, May-June, and June-July 2020), we found the pooled probability of prior arrears to be about 69%. That is, if a household said they were behind in rent in Month T, it was also 69% likely to have been behind in Month (T-1).

[5] See IPUMS USA, https://usa.ipums.org/usa-action/variables/RENTGRS#description_section , accessed 19 July 2021.

We used this prior probability of arrears to stochastically estimate the number of months of arrears for each household going back to the beginning of the pandemic (March 2020 in the US). In practice, for later waves, the average number of months in arrears is about 3.4.

We can then calculate dollars in arrears as RENTGRS * (Months in arrears). Note that for simplicity and to limit the number of potential sources for error, we assume that, on average, households fall behind in whole-month increments with the above-mentioned 69% probability.

While this may seem like an unreasonable assumption, we show at the end of this document that our estimates align with other national, state, and county estimates that make different assumptions about debt accumulation; therefore, we are satisfied that it meets appropriate sanity checks. In reality, this may be an underestimate, since the average awards under actual ERA conditions exceed our estimates. As more data accumulates in ERA programs, we will update our estimates accordingly.

## 8. Aggregate to geographic units

The ACS microdata contain region, state, and Public Use Microdata Area (PUMA) indicators, as well as household weights, allowing us to generate weighted area means for our outcomes of interest (percent in arrears, dollars in arrears). PUMAs are non-overlapping geographical areas of at least 100,000 population. Some PUMAs contain multiple small counties; some large counties contain multiple PUMAs.

In order to maintain appropriate error structures, we use the raw bootstrapped estimates of arrears probability (200 per wave) as input for our area means. For each bootstrap run, we calculate the weighted mean for all households in the area. The mean across these 200 area means becomes our final area mean, and the standard deviation becomes our standard error of the mean.

For dollars in arrears, which does not use such a bootstrapping technique, we take the area-weighted mean.

We perform this method of aggregation at the national, state, and PUMA level.

## 9. Obtain final county-level estimates

To obtain our final county estimates, it is necessary to disaggregate PUMAs according to the counties they comprise. We use the PUMA12-to-county14 crosswalk calculated by the

Missouri Census Data Center.[6] For large counties that contain multiple PUMAs, we take population-weighted means of our outcomes of interest, including standard errors. For small counties contained by a single PUMA, we allow the counties to jointly share PUMA-level estimates and standard errors. For small counties, this presents a more confident estimate than is warranted by the data, and we are working on methods to better reflect this uncertainty for small counties.

# Estimating Eligibility for Rental Assistance

The federal government, under both the Trump Administration and the Biden Administration, has allocated nearly $50 billion in Emergency Rental Assistance (ERA), to be distributed to states and large localities (cities, counties) with over 200,000 inhabitants. We are interested in estimating who, within our estimates of arrears, would be eligible for such assistance. Assistance is available to households that meet all of the following criteria:

1. Someone in the household has been economically impacted by the COVID-19 pandemic;
2. The household is at demonstrable risk of housing instability if evicted for nonpayment of rent; and
3. The household meets income requirements--primarily, annual income below 80% of Area Median Income (AMI), as calculated by the Secretary of the U.S. Department of Housing and Urban Development (HUD).[7]

We chose the third criterion, income, to estimate eligibility. We lost the ability to soundly estimate (1) due to a change in wording of the unemployment question as of April 2021. We consider (2) extremely challenging to estimate and outside the scope of this project. As we use only 1 out of 3 criteria, our estimates will be overestimates of the number of people eligible for ERA.

Since our income data are from 2019, we utilize the 2019 HUD methodology for calculating AMI thresholds from the HUD website, which use Median Family Income as the relevant base metric.[8] We obtain Median Family Income from the 2019 ACS table B19113 (1-year estimates) for each PUMA.[9] We scale households' reference AMI based on the number of

[6] Geocorr 2018: Geographical Correspondence Engine. Available at https://mcdc.missouri.edu/applications/geocorr2018.html (accessed 19 July 2021).
[7] Source: U.S. Department of the Treasury. "Emergency Rental Assistance Frequently Asked Questions." 24 June 2021. Available at https://home.treasury.gov/system/files/136/ERA_FAQs_6-24-21.pdf. Accessed 19 July 2021. Footnote 2 describes the AMI methodology as following HUD practices.
[8] https://www.huduser.gov/portal/datasets/il/il19/Medians2019r.pdf. Accessed 19 July 2021.
[9] The reference area encompassing median income estimates developed by HUD are Fair Market Rent Areas, which are not directly attributable to census areas in our data (source: https://www.huduser.gov/portal/datasets/il/il2020/area_definitions.odn?systype=IL&year=2020&wherefrom=geo, accessed 19 July 2021). Therefore, we use PUMA as the reference area for AMI calculations, the smallest geography directly attributable to ACS microdata. There are approximately

people in the household, per HUD practice. At this point, it is trivial to mark households below the required 80% threshold, and we deem these households “potentially ERA-eligible” and aggregate them to form geographical weighted estimates. The national summary for Wave 33 is as follows:

| Estimate | All renter households | Potentially ERA-eligible renter households |
|---|---|---|
| *Total households paying non-zero rent* | 41.9 million | 23.7 million |
| *Percent of renter households in arrears* | 14.7% | 15.4% |
| *Number of renter households in arrears* | 6.18 million | 3.65 million |
| *Debt owed per household in arrears* | $3,786 | $3,330 |
| *Total rent debt owed* | $23.4 billion | $12.2 billion |

# Assumptions & Known Limitations

There are several known limitations to our strategy, as well as necessary assumptions. We enumerate them here, along with our strategies for minimizing their impact.

First, we assume that respondents to HPS are heads of households, to align their responses with ACS head-of-household indicators. This is a crucial assumption, as it allows us to disaggregate estimates along axes of race, ethnicity, gender, and the like. Other works make the same assumption, such as the National Equity Atlas estimates.

Second, we acknowledge that the estimation strategy for months in arrears is undifferentiated among households. We did initially find that the probability of prior arrears was affected by whether or not the household had experienced unemployment, but a change in measurement to this question in Wave 28 compelled us to remove this responsive estimation. A more nuanced strategy would model the probability of prior arrears; however, there were not enough cross-wave respondents in Phase 1 to build an adequate model.

Third, we assume data from 2019, used in ACS, is essentially unchanged. In reality, many situations may have changed with the pandemic, including living conditions (NW et al., 2020). We also assume household income did not change significantly between the

2,500 FMRAs and 2,350 PUMAs.This means that in some circumstances our AMI thresholds will not match those published by HUD. However, we assume this error is minimal and averages to 0.

recording of ACS microdata (12 months prior to interview date)[10], and calendar year 2019, which is recorded in HPS microdata..

Fourth, we assume that covariates exhibit fixed effects across the country, with no interactions at the state level. This is unlikely to be the case in reality; however, we rely on estimate pooling to obtain reliable estimates across states. The effect is to attenuate state-level estimates toward the national mean (see Fig. 8, below), which is a common feature of this strategy.

# Results

We have included some of the results from our estimates below. For more discussion of results, please see the white paper accompanying this methodological report.

[10] See https://usa.ipums.org/usa-action/variables/HHINCOME#description_section for a discussion of the time frames for this ACS variable.

# Recent trends

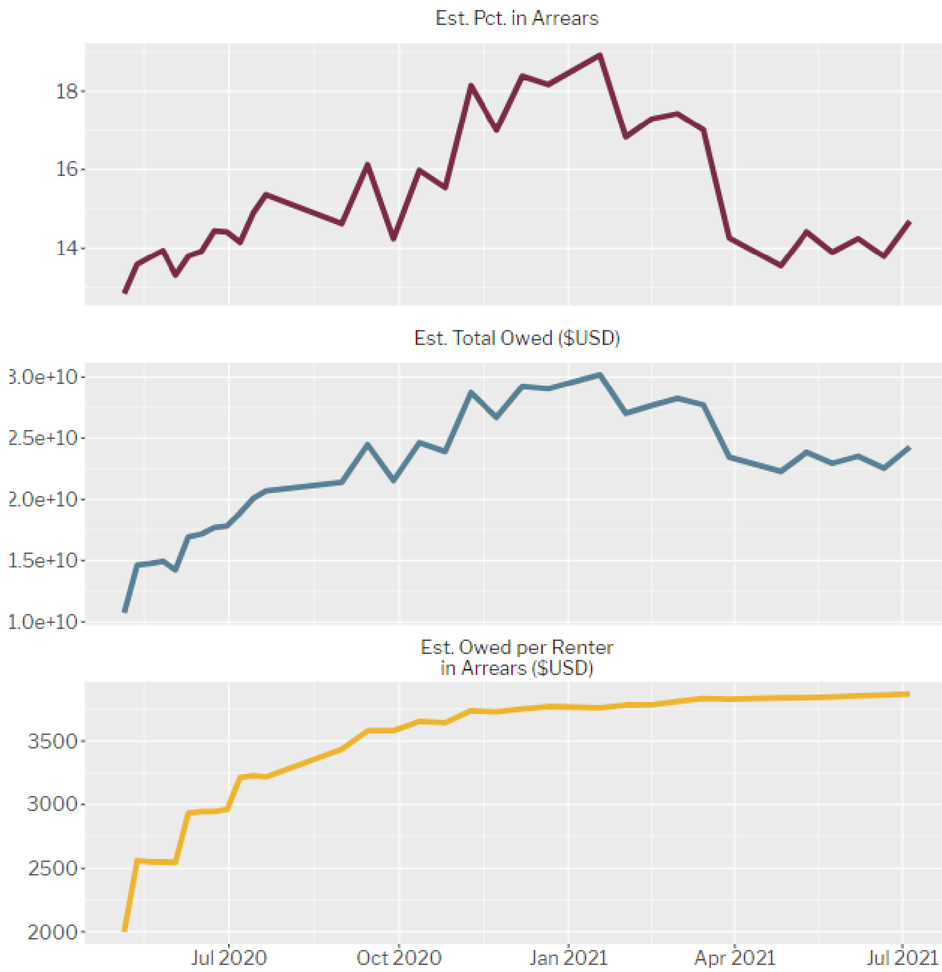


Figure 5. National trends in the estimated percent of renters in arrears with rental debt, estimated total dollar amounts owed, and estimated dollars owed per renter in arrears overtime. Data as of July 2021, Household Pulse Survey Wave 33.

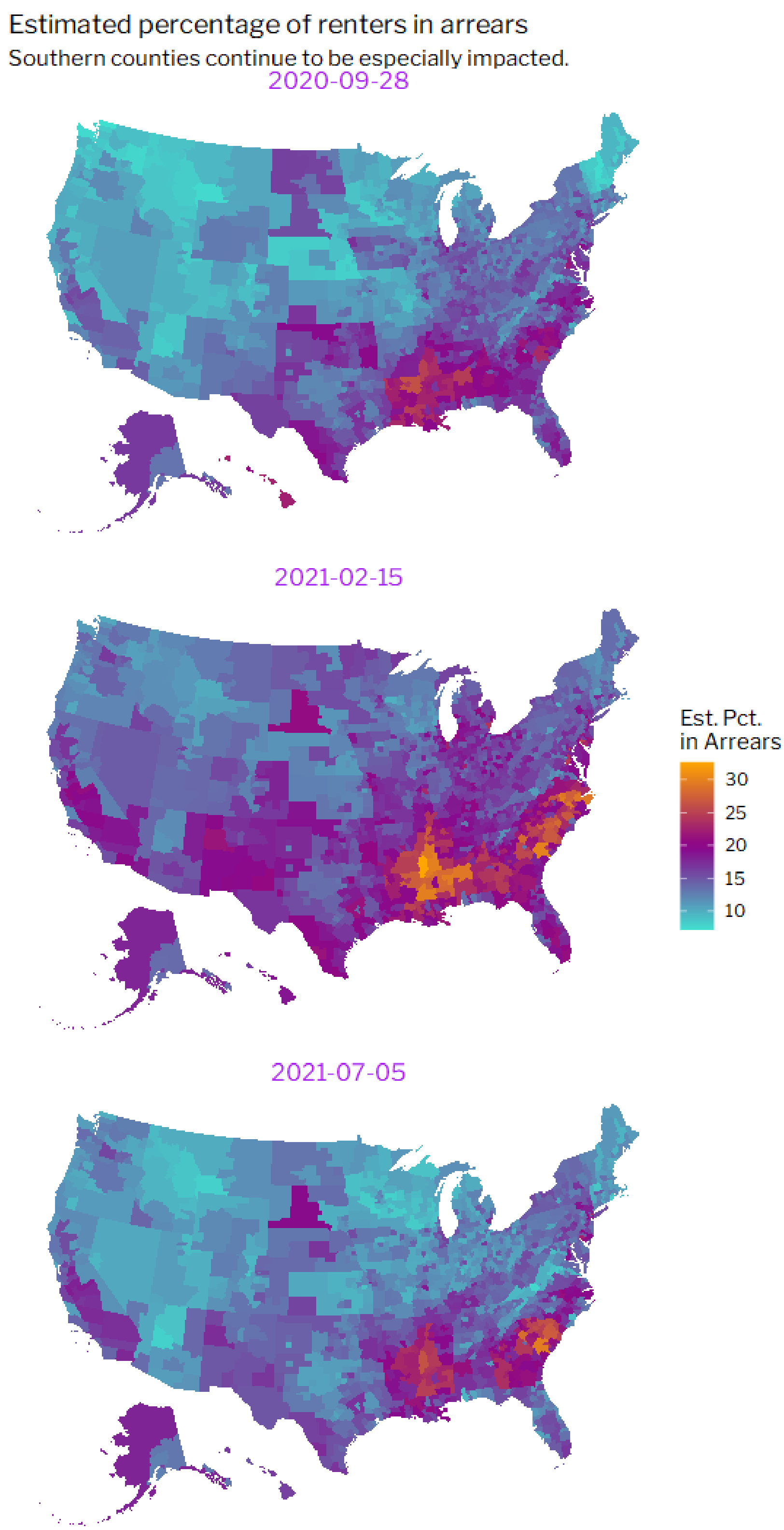


Figure 6. Estimated percent of county-level renters in arrears with rental debt during a) early pandemic (September 2020), b) peak arrears (February 2021), and c) current stages (July 2021). Data as of Household Pulse Survey Waves 15, 24, and 33.

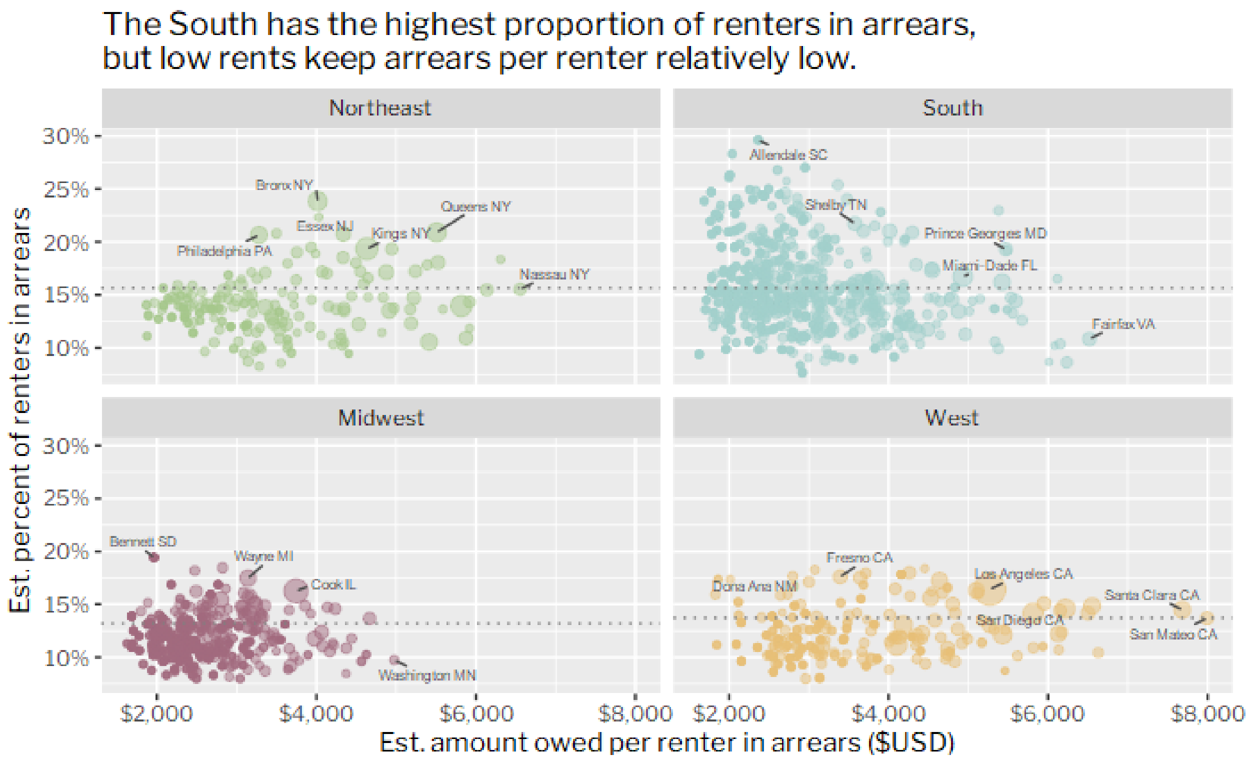


Figure 7. County-level percent of renters in arrears vs. arrear debt owed per renter by US Census Region. Data as of July 2021 , Wave 33. Counties are sized according to their renting population; regional means weighted by county renter population size are indicated by the dotted line.

## Concordance with HPS State Estimates

Below are charts showing how our point estimates fall along the estimates generated directly from HPS data, using household-level replicate weights to calculate confidence intervals. Our models sometimes struggle to match early estimates in Phase 1; however, they generally perform quite well as of August 2020 onward.

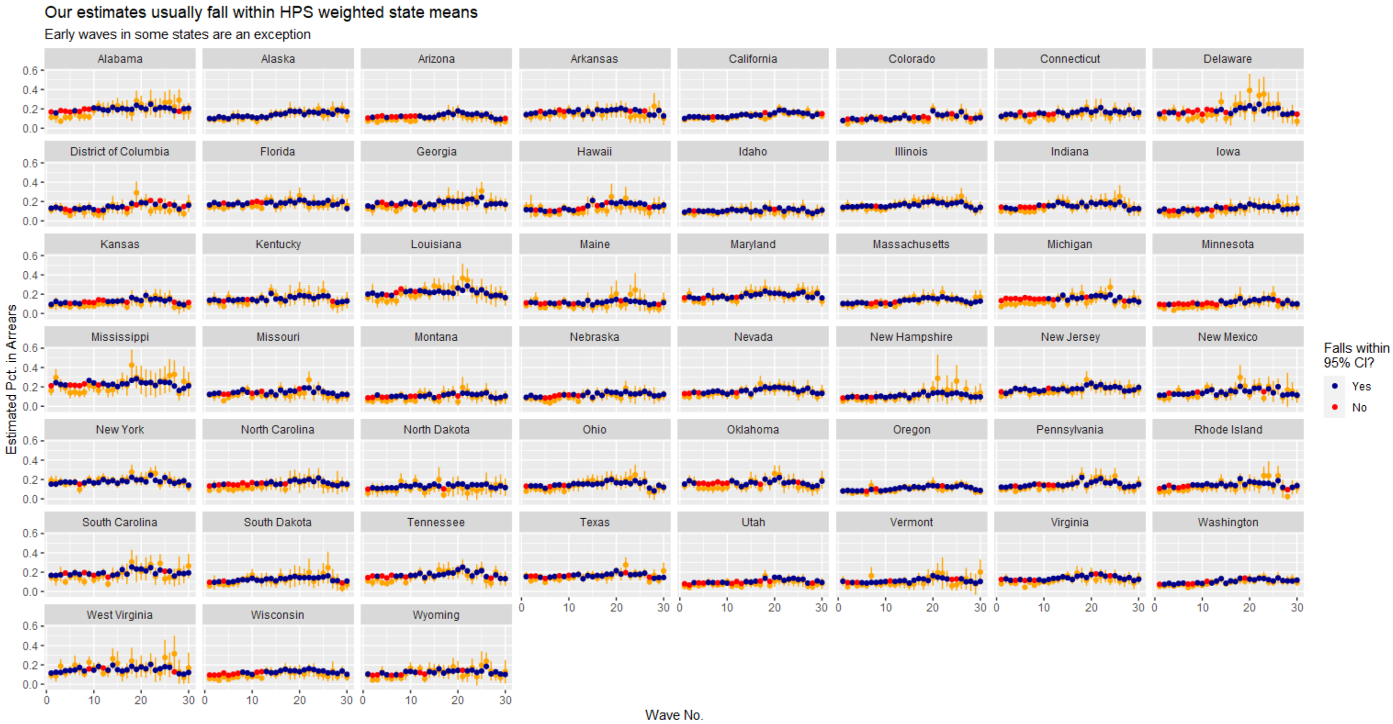


Figure 8. Our state-level estimates of rental arrears, expressed as blue or red dots, compared with HPS estimates, expressed as yellow bars indicating a 95% confidence interval. A blue dot means the point estimate falls within the 95% confidence interval of HPS; a red dot indicates it falls outside the confidence interval.

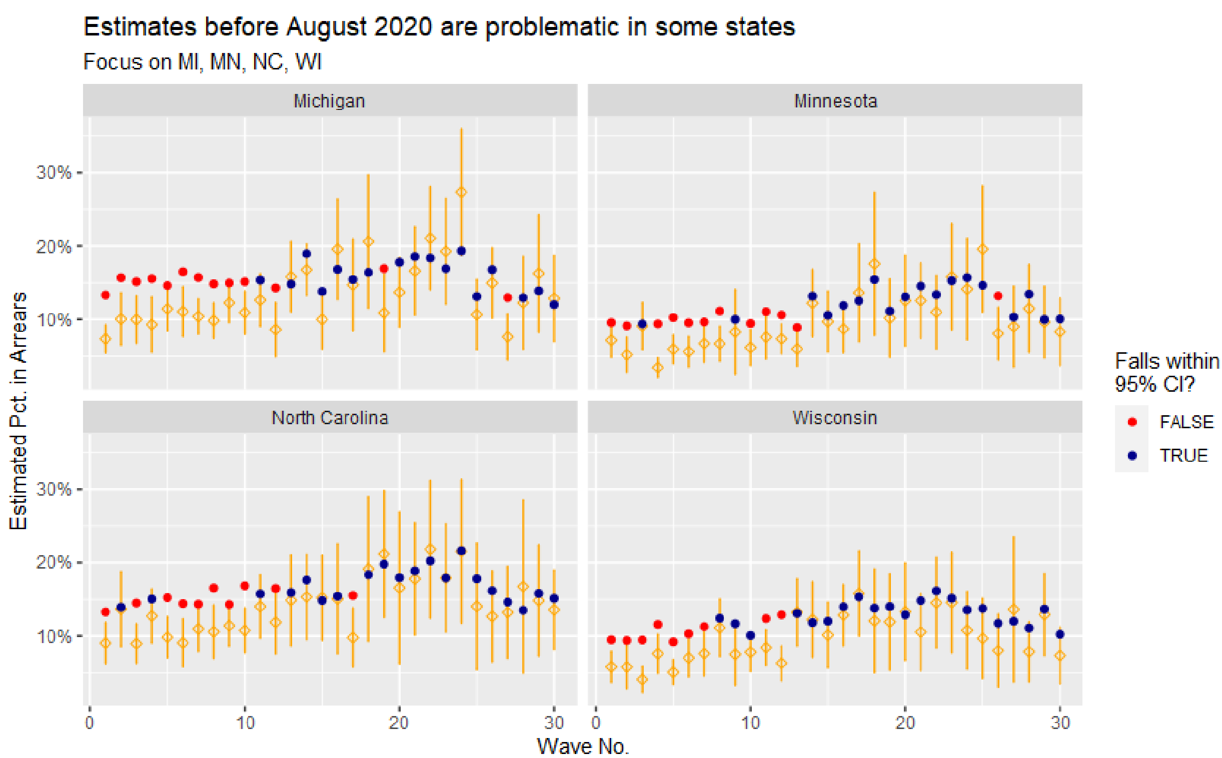


Figure 9. Our state-level estimates of rental arrears, expressed as blue or red dots, compared with HPS estimates, expressed as yellow bars indicating a 95% confidence interval. Here we focus on four states where our model tended to overestimate

arrears during Phase 1: Michigan, Minnesota, North Carolina, and Wisconsin. A blue dot means the point estimate falls within the 95% confidence interval of HPS; a red dot indicates it falls outside the confidence interval.

# Concordance with Other State & County Estimates

## National Equity Atlas

National Equity Atlas (NEA) has also estimated county-level arrears for 40-42 states and 15 metro areas, although they employ a different methodology, relying on statewide survey-derived estimates from the most recent wave and clustering by broad income category. To our knowledge, NEA is the only other organization to release county-level estimates of rent arrears.

Notably, though NEA and Surgo use different sources of data to estimate the number of months in arrears, our conclusions are similar: 3.8 months in the case of NEA, and about 3.2 months in ours. Like our estimates, NEA does not tailor its estimate of the number of months in arrears to demographic or geographic conditions.

We have compared our estimates for Massachusetts counties for Wave 33 (see Table 1, below). In sum, our estimates are over 25% higher for both total arrears and number of households. However, some of our estimates are quite close, such as Bristol and Barnstable, while others are quite different, such as Suffolk and Essex. This could be an artefact of NEA's decision to rely on MSA estimates where available (Suffolk and Essex both lie in the Boston MSA), whereas we pool both MSA and state observations at the state level and allow fine distinctions between household characteristics (e.g., the type of rental unit) to accommodate local conditions.

Our estimated debt per household is similar, owing to the fact that we both source this information from ACS data (microdata in our case, summary table estimates in theirs).

We believe both sets of estimates play a valuable role in informing public policy. Ours tend to be higher in urban areas and as a result show more intrastate variability; policymakers wishing to highlight intrastate distributional efficiencies should use ours. As ERA applications are fulfilled, time will tell which estimates are more accurate.

| | Estimated total arrears | | Estimated Debt per Behind Household | | Estimated number of households behind on rent | |
|---|---|---|---|---|---|---|
| **County Name** | ***NEA*** | ***SV*** | ***NEA*** | ***SV*** | ***NEA*** | ***SV*** |
| ***Barnstable*** | **$9,279,753** | $8,661,903 | $3,871 | **$4,348** | **2,397** | 2,167 |
| ***Berkshire*** | **$7,077,341** | $4,478,204 | $2,648 | **$2,676** | **2,673** | 1,556 |
| ***Bristol*** | $31,352,578 | **$33,109,760** | $2,739 | **$3,434** | **11,445** | 11,141 |
| ***Dukes*** | **$824,766** | $763,629 | $4,201 | **$4,313** | **196** | 151 |
| ***Essex*** | $38,507,358 | **$56,122,549** | $3,853 | **$4,500** | 9,994 | **15,831** |
| ***Franklin*** | **$4,284,723** | $2,363,795 | **$3,064** | $2,876 | **1,398** | 811 |
| ***Hampden*** | **$30,282,220** | $28,386,487 | $2,756 | **$2,761** | 10,987 | **11,842** |
| ***Hampshire*** | **$9,303,501** | $5,442,648 | **$3,384** | $3,194 | **2,749** | 2,724 |
| ***Middlesex*** | $88,202,128 | **$129,775,196** | $4,919 | **$5,475** | 17,930 | **28,600** |
| ***Nantucket*** | **$468,611** | $446,266 | **$5,258** | $4,313 | **89** | 97 |
| ***Norfolk*** | $33,101,792 | **$42,102,807** | **$4,886** | $4,772 | 6,774 | **13,700** |
| ***Plymouth*** | $15,446,595 | **$20,210,568** | $3,851 | **$4,205** | 4,011 | **6,843** |
| ***Suffolk*** | $76,629,876 | **$123,003,333** | $4,390 | **$4,672** | 17,454 | **26,310** |
| ***Worcester*** | **$46,431,669** | $45,785,666 | $3,133 | **$3,606** | **14,820** | 14,381 |
| | | | | | | |
| **Total:** | $391,192,911 | **$500,652,811** | -- | -- | 102,917 | **136,154** |

Table 1: Comparison of Wave 33 county-level estimates in Massachusetts from National Equity Atlas (NEA) and Surgo Ventures (SV). For each metric, larger figures are in bold. Data as of July 2021. Source: National Equity Atlas & Surgo Ventures.

## National Council of State Housing Agencies / Stout

In September 2020, Stout prepared a report of current and projected rent shortfalls in the US, with HPS again as the primary data source.[11] Although they did not furnish county-level estimates, the report represents another major attempt to estimate total rent debt. Here we will compare national estimates; a future version of this document may compare state-level estimates as well.

Stout made an robustly flexible estimation framework, with key assumptions about number of months behind on rent that rely on a different HPS question, how confident the

[11] Stout, "Analysis of Current and Expected Rental Shortfall and Potential Evictions in the US," September 25, 2020, https://www.ncsha.org/wp-content/uploads/Analysis-of-Current-and-Expected-Rental-Shortfall-and Potential-Evictions-in-the-US_Stout_FINAL.pdf. Accessed 21 July 2021

household is able to make next month's payment. They estimated a more rapid pace of rent debt growth, even as the number of households remained the same. They also estimated less in rent debt per household in arrears due to assumptions about partial payments, which we have not incorporated into our estimates.

| Estimation period | Stout | Surgo Ventures |
|---|---|---|
| September 14, 2020 | • Between 9.7 million and 14.2 million households in arrears<br>• Between $12.2 billion and $16.7 billion in rent debt | • Approx. 6.8 million households in arrears<br>• Approx. $23.6 billion in rent debt |
| January 2021 | • Between 9.7 million and 14.2 million households in arrears (same as September 2020)*<br>• Between $25.1 billion and $34.3 billion in rent debt* | • Approx. 7.9 million households in arrears<br>• Approx. $29.1 billion in rent debt |

Table 2: Comparison of September and January estimates produced by Stout and Surgo Ventures. *Stout's January estimate is a projection, which is inherently more challenging.

# Changelog

- July 29, 2021
    - Corrected description of NEA methodology; like ours, their county estimates use the latest wave of HPS microdata.

# References


Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Buuren, S. van, Groothuis-Oudshoorn, K., 2011. mice: Multivariate Imputation by Chained Equations in R. J. Stat. Softw. 45, 1–67.

Luke, S.G., 2017. Evaluating significance in linear mixed-effects models in R. Behav. Res. Methods 49, 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

NW, 1615 L. St, Washington, S. 800, Inquiries, D. 20036 U.-419-4300 | M.-857-8562 | F.-419-4372 | M., 2020. About a fifth of U.S. adults moved due to COVID-19 or know someone who did. Pew Res. Cent. URL https://www.pewresearch.org/fact-tank/2020/07/06/about-a-fifth-of-u-s-adults-moved-due-to-covid-19-or-know-someone-who-did/ (accessed 7.23.21).

Parker, P.A., Janicki, R., Holan, S.H., 2020. Unit Level Modeling of Survey Data for Small Area Estimation Under Informative Sampling: A Comprehensive Overview with Extensions. ArXiv190810488 Stat.

Ruggles, Steven, Flood, Sarah, Foster, Sophia, Goeken, Ronald, Pacas, Jose, Schouweiler, Megan, Sobek, Matthew, 2021. IPUMS USA: Version 11.0. https://doi.org/10.18128/D010.V11.0

White, I.R., Royston, P., Wood, A.M., 2011. Multiple imputation using chained equations: Issues and guidance for practice. Stat. Med. 30, 377–399. https://doi.org/10.1002/sim.4067

Zhang, X., Holt, J.B., Lu, H., Wheaton, A.G., Ford, E.S., Greenlund, K.J., Croft, J.B., 2014. Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. Am. J. Epidemiol. 179, 1025–1033. https://doi.org/10.1093/aje/kwu018