UNITED STATES DEPARTMENT OF COMMERCE
U.S. Census Bureau
Washington, DC 20233-0001

December 3, 2024

2024 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES # ACS24-RER-03

MEMORANDUM FOR ACS Research and Evaluation Workgroup

From: Donna Daily
DONNA DAILY Digitally signed by DONNA DAILY Date: 2025.03.11 15:15:00 -04'00'
Chief, American Community Survey Office

Prepared by: Andrew Keller
Special Assistant to the Chief
Decennial Statistical Studies Division

Tom Mule
Senior Researcher for Administrative Records Research
Research and Methodology Directorate

Dorothy Barth
American Community Survey Office

Subject: Predicting Vacant Housing Units in the American Community Survey

Attached is the final American Community Survey (ACS) Research and Evaluation report for predicting vacant housing units to inform the ACS contact strategy in the nonresponse data collection phase. The Census Bureau's Disclosure Review Board (DRB) has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release. The DRB approval numbers are CBDRB-FY23-ACSO003-B0052 and CBDRB-FY24-ACSO003-0018. If you have any questions about this report, please contact Andrew Keller at 301-763-9308.

Attachment
cc:
acso.re.workgroup.list@census.gov
Katherine Mark ACSO
Jason Lizarraga
Judy Wang
Michelle Wiland
Paul Ehmann
Jennifer Williams Hill
James Cassese
Blake Krippendorf
Shevilla Hudson
Paige Miller

December 3, 2024

Predicting Vacant Housing Units in the American Community Survey

FINAL REPORT

Andrew Keller
Decennial Statistical Studies Division
Tom Mule
Research and Methodology Directorate
Dorothy Barth
American Community Survey Office

This page is intentionally blank.

TABLE OF CONTENTS

TABLE OF FIGURES

EXECUTIVE SUMMARY

In 2022, the U.S. Census Bureau was in danger of exceeding its budget for the American Community Survey (ACS) nonresponse follow-up operation, Computer-Assisted Personal Interviewing (CAPI). Many factors contributed to the unexpected rise in data collection costs, but the top two reasons were an increase in wages and a decline in survey response.

To offset the unplanned expenses, beginning in April 2022, the monthly CAPI workload was capped at 60,000 cases, and in some months, the cases were closed out before the end of the CAPI month. While these cost-saving measures got the program back on budget, they were a quick fix to an immediate issue. Given more time, the Census Bureau looked for a more data-driven approach to optimizing the CAPI workload to maintain the budget while minimizing the impact on data quality.

The first step involved adapting an approach used during the 2020 Census to minimize fieldwork. The Census Bureau implemented a vacancy prediction model to reduce contacts for likely vacant cases in the 2020 Census nonresponse follow-up phase of data collection. (*Likely vacant* housing units are those that receive a high probability score from the vacancy prediction model.) The ACS program adapted the model to identify likely vacant housing units in the CAPI sample.

This report provides a brief background on the modeling approach used for the 2020 Census and outlines the research analyses done to adapt the model for use in ACS CAPI data collection. After thorough analysis, we produced a model using a random forest approach to predict vacancy for ACS housing units. We analyzed the model output results for overall accuracy and accuracy by geography and by each monthly CAPI panel. The results of the analyses of the ACS vacancy prediction model are:

- The model shows consistent results when applied across multiple panels of ACS data. This is important because we want to be able to apply the same model to each CAPI survey panel in a consistent manner. Although all monthly panels use the same modeling inputs, the model will likely work differently in different months of the year.
- The model finds higher vacancy probabilities for states in the upper Midwest and lower vacancy probabilities in the Mountain West. This is expected because we know that areas with high concentrations of vacation rentals and seasonal homes will also have higher concentrations of vacancies in the off-season.
- The more cases we categorize as vacant from the model, the less accurate the model becomes at identifying true vacant housing units, with the measure of truth defined by a Field Representative identifying the unit as vacant. For this reason, we decided to conservatively use only the top 1% of likely vacant cases to be eligible for CAPI stop work orders.

The ACS vacancy prediction model has been used in ACS production since April 2023. For the top *three percent* of the cases identified by the model, Census Bureau interviewers must first personally visit the address to verify the vacancy (as opposed to making a phone call first). The interviewers are allowed two contact attempts for the top *one percent* of the cases identified by the model. If, after two attempts, we cannot obtain a survey interview, the case is removed from the workload. This approach allows the interviewers to verify the occupancy status of housing units and have more time to focus on getting survey responses from households more likely to be occupied, increasing the chances of obtaining a complete interview.

From April 2023 to August 2024, about 15.0% of the cases were removed from the CAPI workload for the predicted vacant cases eligible for stop work. Of those not removed from the workload, on average, 70.3% resulted in a vacant interview, 3.5% in an occupied interview, 3.3% in a late self-response, and 7.9% in a coded non-interview. We are closely monitoring production to see if we should modify the threshold to identify more (or fewer) cases for stop-work eligibility. We will also continue to investigate how to improve the predictive models to increase their accuracy.

1. INTRODUCTION

In 2022, the U.S. Census Bureau was in danger of exceeding its budget for the American Community Survey (ACS) nonresponse follow-up operation, Computer-Assisted Personal Interviewing (CAPI). Many factors contributed to the unexpected rise in data collection costs, but the top two reasons were an increase in wages and a decline in survey response.

To offset the unplanned expenses, beginning in April 2022, the monthly CAPI workload was capped at 60,000 cases, and in some months, the cases were closed out before the end of the CAPI month. While these cost-saving measures got the program back on budget, they were a quick fix to an immediate issue. Given more time, the Census Bureau looked for a more data-driven approach to optimizing the CAPI workload to maintain the budget while minimizing the impact on data quality.

The first step involved adapting an approach used during the 2020 Census to minimize fieldwork. The Census Bureau implemented a vacancy prediction model to reduce contacts for likely vacant cases in the 2020 Census nonresponse follow-up phase of data collection. (Throughout this report, *likely vacant* housing units are those that receive a high probability score from the vacancy prediction model.) The ACS program adapted the model to identify likely vacant housing units in the nonresponse follow-up CAPI data collection phase.

Identifying likely vacant housing units and stopping work on the cases after only two contact attempts gives interviewers a chance to verify vacancy status while allowing for more time to focus their efforts on getting survey responses from households that are more likely to be occupied, increasing the chances of obtaining a complete interview.

This report provides a brief background on the modeling approach used for the 2020 Census and outlines the research analyses done to adapt the model for use in ACS CAPI data collection.

2. BACKGROUND AND LITERATURE REVIEW

ACS data collection is multi-modal and occurs over three months for each monthly panel. In the first two months housing units may receive up to five mailings soliciting a survey response by internet, mail, or Telephone Questionnaire Assistance (TQA). Just before the beginning of the third month, a subsample of the remaining nonresponding housing units is selected for computer-assisted personal interviews (CAPI). CAPI data collection occurs during the third month. During the CAPI month, while telephone and personal-visit interviews are conducted, self-responses via mail or the internet continue to be accepted.

Because of rising costs of in-person interviewing and declining response rates, the Census Bureau has had to limit the number of housing units subsampled for CAPI interviews. However, more needed to be done to maintain the data collection budget while also obtaining the highest quality data possible. The 2020 Census successfully implemented reduced contact procedures in their Nonresponse Followup (NRFU) operation for likely vacant housing units

identified by their statistical predictive models, so the ACS program also decided to implement similar procedures. Throughout this report, we use the term *likely vacant* for housing units that receive a high probability score from the vacancy prediction model.

2.1 Using Administrative Records Data to Predict Vacant Housing Units for the Decennial Census

As part of the 2010 decennial census, about one-third of the US population was enumerated by a personal visit during the NRFU operation. The NRFU operation was a significant cost driver in the 2010 Census, with a total cost of about $1.6 billion. Consequently, to save money, the U.S. Census Bureau researched using administrative records to provide an occupancy status for some NRFU addresses in the 2020 Census.

Since the 1980s, using administrative records to supplement or replace traditional census-taking has been a topic of interest (Alvey and Scheuren 1982; Scheuren 1999).[1] However, the United States does not have a single administrative records system with a high coverage of the entire population (Mulry 2014). For the 2020 Census research, the Census Bureau was provided conditional access to data from organizations such as the Internal Revenue Service (IRS), Social Security Administration (SSA), Center for Medicare and Medicaid Services (CMS), and commercial data vendors. Even though each data source covers just a segment of the entire U.S. population, they provide information relevant to census enumeration, such as a person's tax filing address from the IRS and date of birth from the SSA.

In 2013 and 2014, the Census Bureau started developing methods to combine and use several administrative sources to identify occupied and vacant units before or after minimal NRFU fieldwork (Mule and Keller 2014). Mulry et al. (2021) document refinements to the vacancy prediction methodology made following the 2015, 2016, and 2018 Census Tests. Two notable advances were made to improve the ability to predict vacant units. First, Keller et al. (2018) improved the model by using a Euclidean distance metric to identify vacant units rather than linear programming methods. Second, the model was improved by refining the use of Undeliverable as Addressed (UAA) data provided by the United States Postal Service (USPS) when 2020 Census mailings could not be delivered to the census address.

The resulting AR predictive model was used to classify the occupancy status of addresses in the 2020 Census NRFU operation. Of the 151.8 million addresses in the 2020 Census, 3.20 percent were enumerated as AR Occupied, 1.15 percent as AR Vacant, and 0.24 percent as AR Delete.[2] This was 4.59 percent of the total address universe. Among the NRFU universe, 9.51 percent

[1] "Administrative records" refers to data collected by government agencies for the purpose of administering programs and providing services. They are also free, publicly available information or data bought from third-party vendors.

[2] Some examples of AR Deletes are uninhabitable units, units that are burned down, or former housing units that are converted to businesses.

were enumerated as AR Occupied, 3.43 percent were enumerated as AR Vacant, and 0.70 percent were enumerated as AR Delete. This was 13.64 percent of the NRFU address universe.[3]

2.2 Adapting the 2020 Census Vacancy Prediction Model for the ACS

Since the 2020 Census vacancy model was implemented successfully, we focused on adapting the model to create a similar model for the ACS program. Both the ACS model and the decennial census model use IRS, Medicare, and UAA data. They also both output a predicted vacant probability.

There is one major conceptual difference between the vacancy model implemented in the 2020 Census and the vacancy model used for ACS. The decennial census measures the population as of April 1 of the decennial year. Pursuant to that goal, the 2020 model intended to predict the vacancy status as of April 1, 2020, regardless of when the housing unit would be interviewed during the NRFU operation. The ACS measures the population on the day the survey is completed. As a result, the ACS model aims to predict vacancy status when the survey is in the field (i.e., during the CAPI operation). The ACS goes into the field about two months after the initial mailing is sent. This difference means that the same variable may have different predictive power when comparing its utility to the 2020 Census versus the ACS.

Additionally, the vacancy model will likely work differently during different months of the year. Addresses with seasonal occupancy, such as vacation homes, may have higher predicted vacant probabilities during periods of less temperate weather. Areas near large universities are another example of addresses that may have different vacancy probabilities depending on the time of year. Table 1 outlines the differences between the 2020 Census and the ACS vacancy models.

Table 1 Differences Between the 2020 Census and ACS Vacancy Models

Feature	2020 Census	ACS
Reference Date	April 1, 2020	The CAPI Interview Date
Variable Importance	All variables have the same importance since the model is only being fit once.	The variables may carry different levels of importance at different times of the year.
Model Fitting Universe	The vacancy model is fit over the entire 2010 Census Nonresponse Followup universe.	The vacancy model is fit over the CAPI universe of previous year's CAPI panel.

[3] Results are from 2020 Census Data Quality Metrics: Release 1 (DRB Clearance CBDRB-FY21-DSSD007-0012) 2020 Census Data Quality (Accessed August 2021).

3. METHODOLOGY

3.1 Creating the ACS Vacancy Prediction Model

The underlying approach for the ACS vacancy model approximates the tack taken in developing the vacancy model used from the 2020 Census. The ACS vacancy prediction model incorporates data from four general sources: administrative records, operational data, address data, and the Census Bureau's planning database.

The administrative records data consists of federal and commercial sources. The federal sources include data from the IRS, USPS, and CMS. The commercial datasets are aggregated public information purchased by the Census Bureau, which consists of local tax, deed, and mortgage information. The commercial data also includes address-level data indicating land use, whether the owner lives at the address, and the ownership rights on the unit. We use person-level information providing information about persons at the address.

For operational data, we use UAA data from the USPS. This data indicates why an ACS mailing could not be delivered to the address. For example, the address could be vacant or the address number may not exist when the postal worker attempted to deliver the ACS mailing at the beginning of survey operations. We also used indications of vacancy as indicated by the ACS internet survey response instrument.[4] These cases are followed up during the CAPI month as part of ACS operations.

For address-level data, we were interested in data that would help indicate the nature of the addresses' occupancy status. We looked at the biyearly status of the address as indicated by the Delivery Sequence File (DSF) provided to the Census Bureau by the USPS. This shows whether the address is residential or commercial, excluded from delivery statistics, or not on the DSF. We also used information indicating whether the housing unit was a single unit, part of a multi-unit complex, trailer, or some other type of building. We also incorporated information indicating the type of mailbox that serves the address.

Finally, we incorporated ACS survey results that characterize the tracts. The Census Bureau's Planning Database contains housing, demographic, and socio-economic statistics developed from past five-year ACS estimates.[5] We applied those estimates to each address within the tract to provide contextual information about the geography.

To begin, we acquired the 2021 and 2022 ACS datasets to serve as separate training and external validation universes, respectively, in model development. We also collected the four data sources that reflected the proper timing. For example, we collected IRS 1040 records from

[4] Respondents can self-report that a housing unit is vacant on the internet response instrument. Vacant households are asked some questions about the housing unit, but no questions about the people residing in the unit.

[5] Census Bureau's Planning Database (census.gov)

the 2020 and 2021 tax years to reflect the 2021 and 2022 calendar years since taxes are filed for the previous year's income.

In addition to analyzing the value of each covariate, we also looked at multiple modeling functions. Ultimately, we found that a random forest approach performed better than a logistic model. Logistic regression assumes a linear relationship between the independent variables and the outcome variable and tries to fit a function to the data. It is best used for predicting binary outcomes. On the other hand, random forest is a method that uses multiple decision trees to make a prediction. It creates many decision trees, each based on a random subset of the input variables, and then combines the information to make a final prediction. It is more accurate than logistic regression when dealing with complex and non-linear relationships between the input and output variables.

3.2 Analyzing the ACS Vacancy Prediction Model

As previously stated, the intention of creating a model to predict likely vacant housing units is to issue a stop work order in the field after two attempts are made to obtain an interview from the housing unit. Identifying likely vacant housing units and stopping work on the cases after only two contact attempts gives the interviewers more time to focus their efforts on getting survey responses from households that are more likely to be occupied, increasing the chances of obtaining a complete interview. Allowing two contact attempts also allows the interviewer to verify the correct occupancy status, limiting the effect of model error.

After creating the model, we analyzed several aspects to determine its efficacy in ACS production. We analyzed the model to assess the quality of the vacancy predictions and determine what cutoff we would use to flag cases as likely vacant and, therefore, eligible for stop work. We also analyzed the top percentages of cases predicted to be likely vacant to see the geographic distribution of the cases.

We ran simulations using the 2021 and 2022 ACS mailable CAPI universes to determine the quality of the model using the following steps:

1) Start with the mailable CAPI universe cases from the 2021 and 2022 ACS universe.
2) Fit the model on 2021 data.
3) Score the model on 2022 data.
4) Sort the predicted vacant probabilities from greatest to least.
5) Iterate over-the-top percentages by picking a threshold (for example, the top 10% or 5% of predicted vacant probabilities).

3.2.1 Quality of Vacancy Predictions

RQ1. How accurate is the vacancy prediction model?

We analyzed the model using the top ten percentage points of vacancy predictions. For example, if the mailable CAPI universe had fifty thousand cases, we looked at the quality over

the top 500 cases, 1000 cases, and so on, up to 5,000 cases. For each case, we measured quality by observing how often a Field Representative identified cases falling within the vacancy prediction threshold as vacant. We looked at the agreement rates at different thresholds, allowing program managers to decide the cutoff percentage errors they would tolerate and the maximum number of contact attempts to allow interviewers when applying the vacancy model predictions during ACS field operations. We also observed the match rates for each month to ensure that the model can be used consistently for each ACS CAPI monthly panel.

3.2.2 Geographic Distributions of Vacancy Predictions

RQ2. What is the geographic distribution of the likely vacant housing units?

We analyzed the models to detect geographic differences using the same vacancy prediction thresholds. One possible reason for vacancy is that a housing unit is a vacation home, so we naturally assumed that specific geographies (or vacation spots) would have more occurrences of units identified by the model. We performed a geographic analysis to verify the assumption and see the geographic outcomes of the vacancy predictions at various thresholds.

4. ASSUMPTIONS AND LIMITATIONS

One of the concerns about developing a model for continuous monthly identification of vacant housing units in the ACS was whether we could ingest and process the various data sources in real time. The experience with the 2020 Census demonstrated that various data sources could be integrated to pursue a tailored contact strategy in real time. However, like the implementation of the vacancy model in 2020, we somewhat limited the extent of data sources that were researched in this project. For example, we did not consider state-level data sources like the Supplemental Nutritional Assistance Program (SNAP) records. Although SNAP data might aid vacancy prediction, investigating its utility was not pursued due to the compressed time limit needed to get the vacancy model into production.

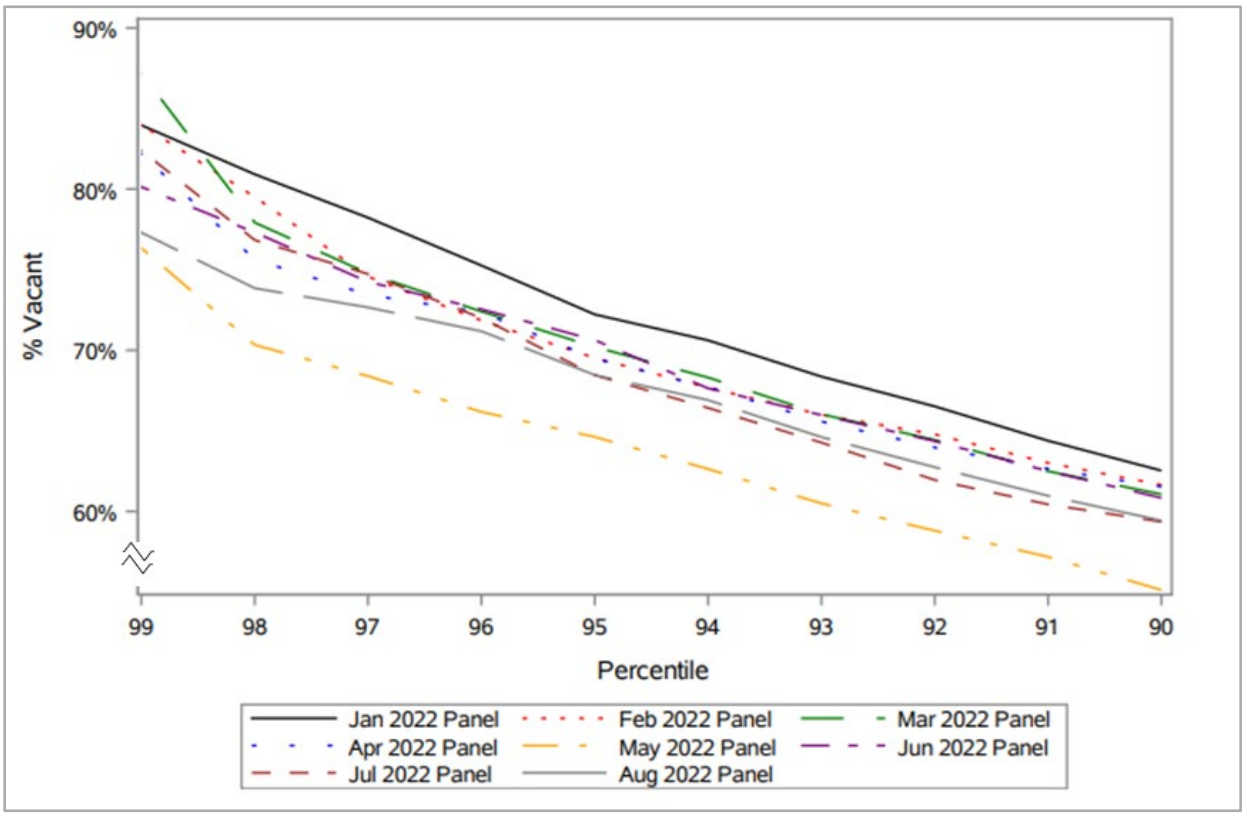
5. RESULTS

5.1 Quality of Vacancy Predictions

RQ1. How accurate is the vacancy prediction model?

To determine the model's accuracy, we calculated a true positive rate by looking at when predictions ended up being a vacant housing unit, as reported by a Field Representative. We calculated the true positive rates for the January 2022 through the August 2022 sample panels. Figure 1 (below) shows a line graph where each line represents a panel. The *x*-axis shows the percentile. For example, the 99^{th} percentile indicates the top one percent of vacant probabilities associated with the panel. Each percent is about five hundred cases. The *y*-axis represents the true positive rate. For example, for the 97^{th} percentile for the January 2022 panel, the true positive rate is 78 percent.

Figure 1 True Positive Vacancy Rates from the ACS Vacancy Prediction Model by Percentile and Panel

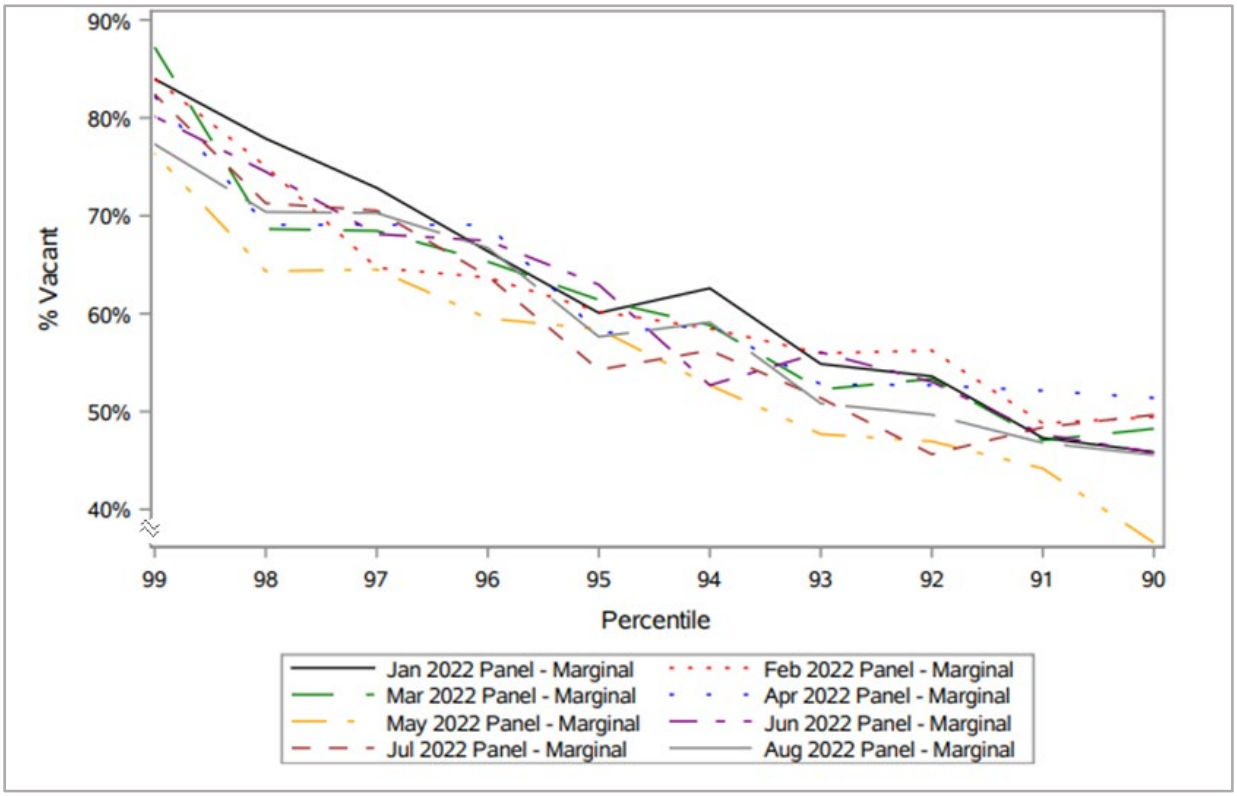


Source: U.S. Census Bureau, 2022 American Community Survey paradata, DRB Approval Number: CBDRB-FY23-ACSO003-B0052
NOTE: A "true positive vacant" is determined by a Census Bureau field representative identifying a housing unit as "vacant." The May 2022 panel (July CAPI) had its CAPI data collection cut shorter than the other panels and thus is less representative of the general trend shown by the other panels.

As all panels move from left to right, the vacancy rate decreases. This is expected—more cases are being treated as vacant, but fewer are observed as vacant. For a fixed set of nonresponding addresses, as you move from left to right on the graph, you have already identified the best cases and are now identifying the cases that are relatively the NEXT best. So, if your model works well and the top 1% have been identified very well, the NEXT 1% (giving you the best 2%) will be slightly worse, with a somewhat lower percentage of vacants. The further right you move on the graph, the less accurate the model becomes.

Figure 2 looks at each additional one percent in isolation. That is, the marginal gain for each one percent in predicted probabilities. Like Figure 1, Figure 2 shows line graphs where each line represents a panel. The x-axis shows the percentile. For example, the 99th percentile indicates the top one percent of vacant probabilities associated with the panel. However, the 97th percentile indicates the one percent of cases with between the best 2% and 3% of predicted probabilities. The y-axis represents the true positive rate of the one percent marginal. For example, for the cases between the 98th and 97th percentile for the January 2022 panel, the true positive rate is 73 percent.

Figure 2 True Positive Vacancy Rates from the ACS Vacancy Prediction Model by Percentile and Panel and 1% Marginals



Source: U.S. Census Bureau, 2022 American Community Survey paradata, DRB Approval Number: CBDRB-FY23-ACSO003-B0052
NOTE: A "true positive vacant" is determined by a Census Bureau field representative identifying a housing unit as "vacant." The May 2022 panel (July CAPI) had its CAPI data collection cut shorter than the other panels and thus is less representative of the general trend shown by the other panels.

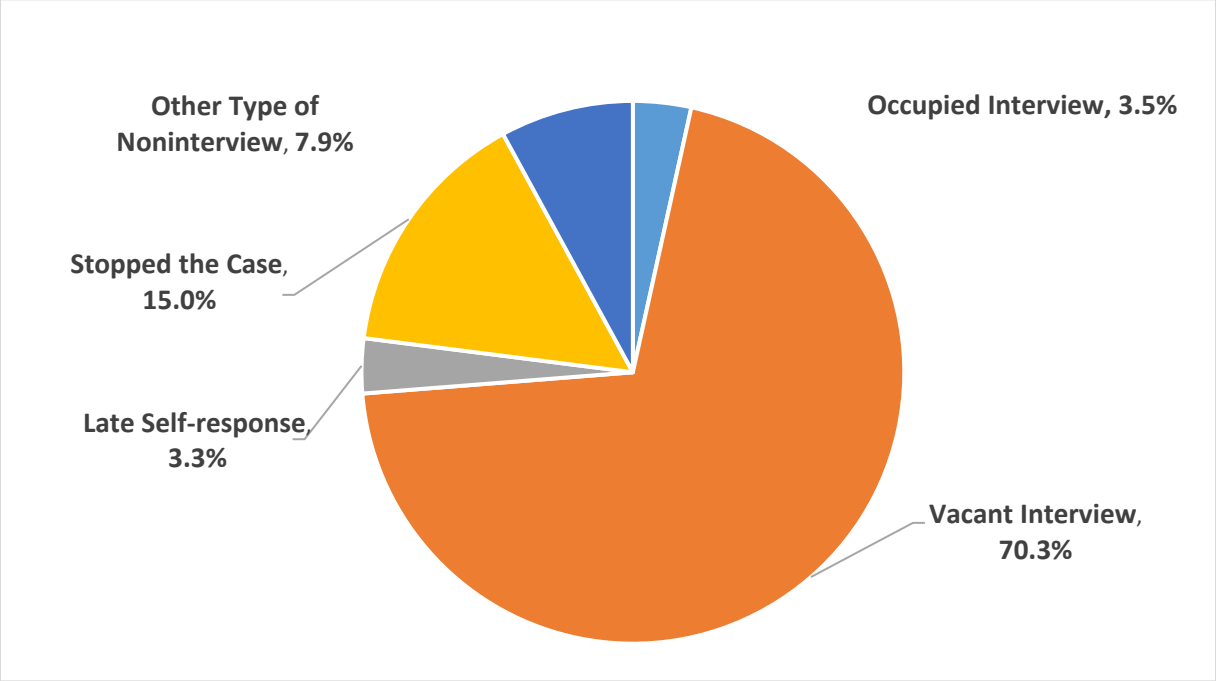
Compared to Figure 1, Figure 2 generally decreases in nature—although in a more jagged fashion. The decreasing probabilities show that the next one percent is generally a little worse, i.e., identifies a lower percentage of vacants. However, there are some percentages where the percentile to the right has a higher vacancy rate. This indicates that the next percentile performs a little better in terms of vacancy prediction.

Another observation is that the line graphs, while not completely overlapping, largely follow the same path. This indicates a degree of consistency of results across panels—indicating that a general model can be applied across the different months of ACS. That being noted, the prediction models will likely work differently in various months of the year. This might affect the entire U.S., or perhaps only states whose housing unit vacancy depends on the season, such as vacation homes and university housing in the housing unit universe. For this work, we developed a national-level model that was fit on the same panel month as the previous year. Allowing the modeling coefficients to vary across the panel allowed us to account for temporal variation in the reasons for vacancy. In addition, developing sub-national models could also

affect the empirical results. However, this was thought to be difficult to implement, given the compressed time limit for implementation.

Based on this analysis, the ACS program decided to conservatively use the top 1% of likely vacant cases to stop interviewing attempts after two visits beginning in April 2023. Looking at production data from April 2023 to August 2024, we see that 70% of all cases in the top one percent ended up with a vacant interview. This percentage would be higher if we did not stop work on the likely vacant cases because some cases would likely have been resolved as vacant interviews.

Figure 3 Average Percentage of Outcomes for All Cases Identified in the Top 1% by the ACS Vacancy Prediction Model, ACS Production from April 2023 to August 2024

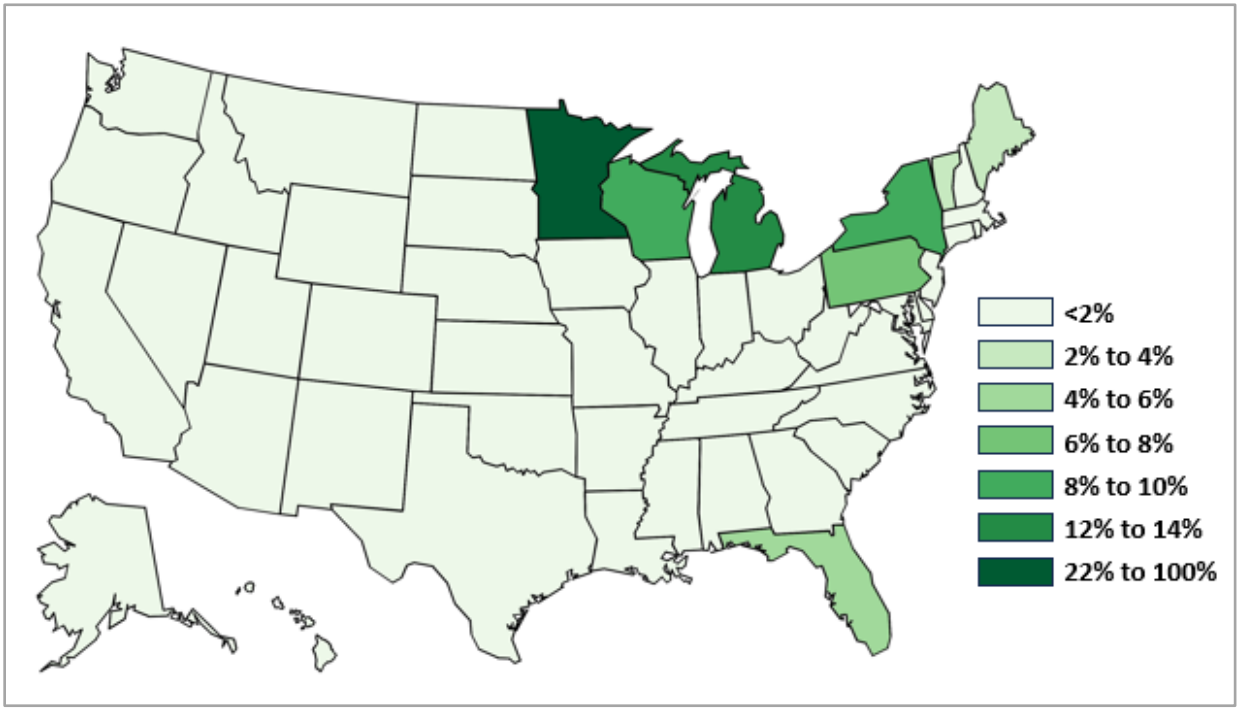


Source: U.S. Census Bureau, 2023 and 2024 American Community Survey paradata, DRB Approval Number: CBDRB-FY24-ACSO003-0018

5.2 Geographic Distributions of Vacancy Predictions

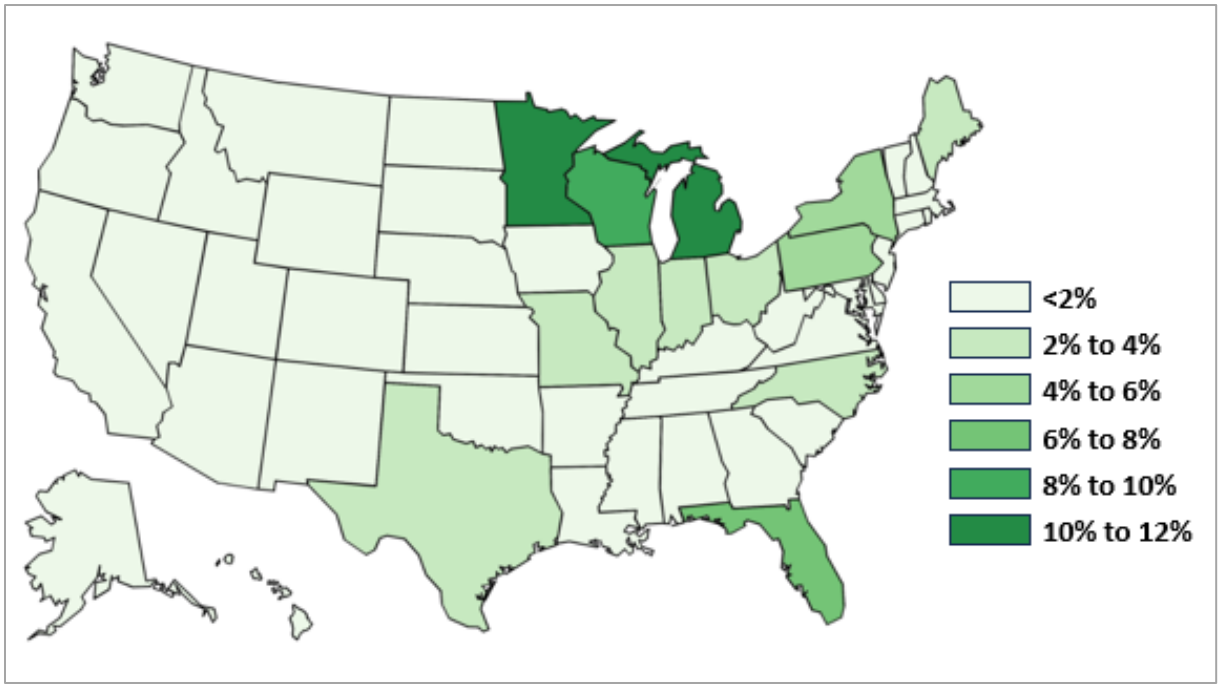
To see how the predicted vacant housing units were distributed among the different states, we plotted the distribution of vacancies among the ACS CAPI universe from January 2022 through August 2022 panels. The top one percent, three percent, and five percent, respectively, are shown in the figures below.

Figure 4 Distribution of the Top 1% of Vacancy Predictions by State: January to August 2022 ACS CAPI Panels



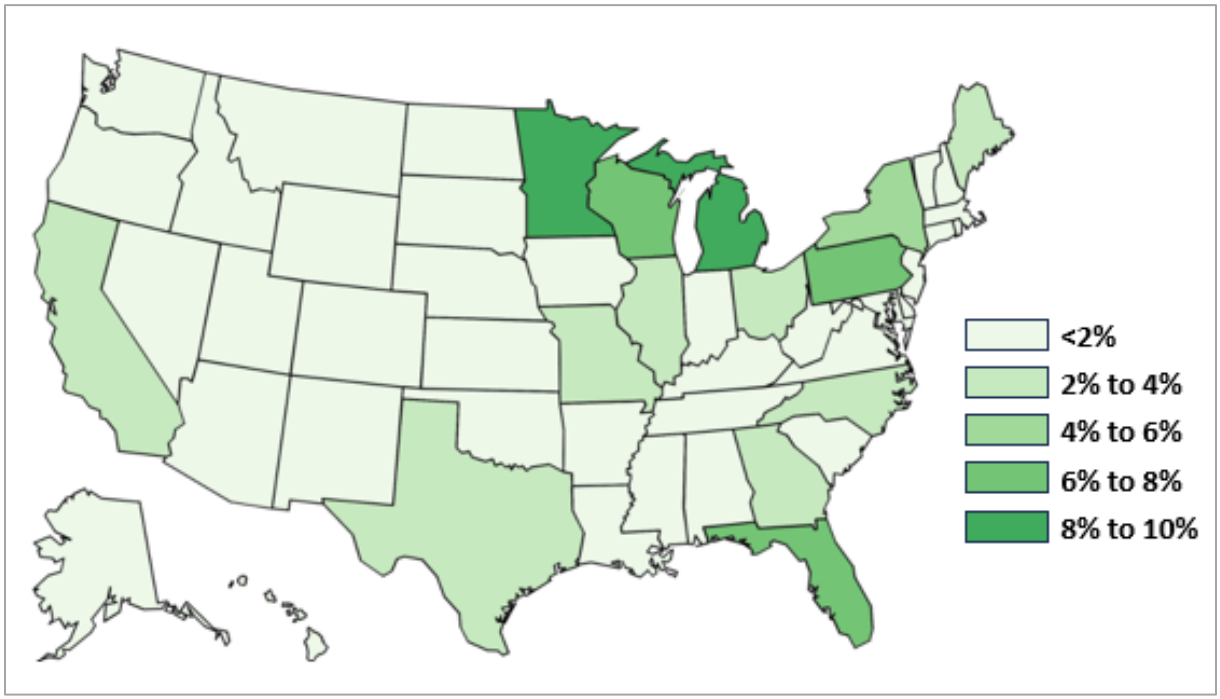
Source: U.S. Census Bureau, 2022 American Community Survey paradata, DRB Approval Number: CBDRB-FY24-ACSO003-0018

Figure 5 Distribution of the Top 3% of Vacancy Predictions by State: January to August 2022 ACS CAPI Panels



Source: U.S. Census Bureau, 2022 American Community Survey paradata, DRB Approval Number: CBDRB-FY24-ACSO003-0018

Figure 6 Distribution of the Top 5% of Vacancy Predictions by State: January to August 2022 ACS CAPI Panels



Source: U.S. Census Bureau, 2022 American Community Survey paradata, DRB Approval Number: CBDRB-FY24-ACSO003-0018

Across the three charts, the results show that the model finds higher vacancy probabilities for states in the upper Midwest (Minnesota, Michigan, and Wisconsin) and lower probabilities of vacancy in the Mountain west (Colorado and Wyoming). It is also evident that the top vacant probabilities get spread across more states as a higher vacancy threshold is used. To observe this, we can focus on two states—Minnesota and Texas. Figure 4 shows that Minnesota comprises over 22 percent of the distribution while Texas comprises less than 2 percent among the top 1 percent of vacant probabilities. Figure 5 shows that Minnesota's share decreases—now comprising between 10 and 12 percent of the distribution while Texas' share increases—now comprising between 2 and 4 percent among the top 3 percent of vacant probabilities. Figure 6 shows that Minnesota's share again decreases—now comprising between 8 and 10 percent of the distribution while Texas' share still comprises between 2 and 4 percent among the top 5 percent of vacant probabilities.

In short, the charts show that the top vacant probabilities get distributed more equally among the states as a greater pool of top vacant probabilities is considered. Therefore, geographies may be differentially affected if a smaller share of vacant probabilities is considered for an operational treatment.

6. CONCLUSION

This project aimed to predict vacant units to inform the ACS CAPI contact strategy. To do this, refinements were made to the vacancy model used in the 2020 Census to account for the design of ACS field operations. Both models incorporate information from multiple sources. This includes independent variables from four general sources: administrative records, operational data, address data, and the Census Bureau's planning database.

The results show that the model finds higher vacancy probabilities for various geographies. It also shows that if more cases are treated as vacant, the model decreases in its predictive quality. Lastly, the model performs similarly over multiple months of ACS data.

The ACS vacancy prediction model has been used in ACS production since April 2023. For the top *three percent* of the cases identified by the model, Census Bureau interviewers must first personally visit the address to verify the vacancy (as opposed to making a phone call first). The interviewers are allowed two contact attempts for the top *one percent* of the cases identified by the model. If, after two attempts, we cannot obtain a survey interview, the case is removed from the workload. This approach allows the interviewers to verify the occupancy status of housing units to have more time to focus on getting survey responses from households more likely to be occupied, increasing the chances of obtaining a complete interview.

From April 2023 to August 2024, about 15.0% of the cases were removed from the CAPI workload for the predicted vacant cases eligible for stop work. Of those not removed from the workload, on average, 70.3% resulted in a vacant interview, 3.5% in an occupied interview, 3.3% in a late self-response, and 7.9% in a coded non-interview. We are closely monitoring production to see if we should modify the threshold to identify more (or fewer) cases for stop-work eligibility. We will also continue to investigate how to improve the predictive models to increase their accuracy.

7. REFERENCES

Alvey, W. and F. Scheuren. 1982. "Background for an Administrative Record Census." in JSM Proceedings, Social Statistics Section, American Statistical Association, Cincinnati, OH, August 16–19, 1982. Washington, DC: American Statistical Association. 137–152.

Keller, A., T. Mule, D.S. Morris, and S. Konicki. 2018. "A Distance Metric for Modeling the Quality of Administrative Records for Use in the 2020 U.S. Census," Journal of Official Statistics, vol. 34(3), No. 3, 2018, pp. 599–624, September. http://dx.doi.org/10.2478/JOS-2018-0029.

Mule, V.T. and A. Keller. 2014. "Using Administrative Records to Reduce Nonresponse Followup Operations." in JSM Proceedings, Survey Research Methods Section, American Statistical Association, Boston, MA, August 2–7, 2014. Alexandria, VA: American Statistical Association. 3601–3608.

Mulry, M.H. 2014. “Measuring Undercounts for Hard-to-Survey Groups.” In *Hard-to-Survey Populations*, edited by R. Tourangeau, N. Bates, B. Edwards, T. Johnson, and K. Wolter, chapter 3, pp. 37-57. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9781139381635.005

Mulry, M.H., Mule, T., Keller, A., and S. Konicki. 2021. “Administrative Record Modeling in the 2020 Census.” Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/administrative-record-modeling-in-the-2020-census.pdf (accessed October 16, 2024).

Scheuren, F. 1999. “Administrative Records and Census Taking.” *Survey Methodology* 25(2): 151–160.

Appendix A. Summary of Input Data for the ACS Vacancy Prediction Model

Administrative Records Data

- Federal Data
  - Internal Revenue Service (Individual Taxpayer Form 1040 information)
  - Center for Medicare and Medicaid Services (Medicare Enrollment)
  - United States Postal Service (National Change of Address Information)
- Aggregated Third-party Data
  - Local tax, deed, and mortgage information
  - Land use, absence of owner at address, and ownership rights on the unit

Operational Data

- Mailing operations (U*ndeliverable as Addressed* and reasons why from USPS)
- Indication of vacancy from internet responses
- Vacancies coded by interviewers in CAPI data collection

Address-level Data

- Delivery Sequence File status (Residential, Commercial, Excluded from Delivery Statistics).
- Housing Unit Type (Multi, Single, Trailer, Other)
- Delivery Point Type (Type of mailbox)

Block Group-level Data (from the Census Bureau's Planning Database)

- Poverty and Rental Rates
- Type of Language Spoken Rates
- Hispanic Origin Rates