DISSERTATION

THREE ESSAYS ON HOMELESSNESS POLICY, FEDERAL HOUSING ASSISTANCE, AND RENTAL MARKET DYNAMICS IN THE UNITED STATES

Submitted by

Luke Maddock

Department of Economics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2026

Doctoral Committee:

Advisor: Anita Pena

Tim Komarek
Ray Miller
Jesse Burkhardt

Copyright by Luke Maddock 2026

All Rights Reserved

ABSTRACT

THREE ESSAYS ON HOMELESSNESS POLICY, FEDERAL HOUSING ASSISTANCE, AND RENTAL MARKET DYNAMICS IN THE UNITED STATES

This dissertation examines three dimensions of homelessness and housing policy in the United States: barriers to housing access for voucher holders, the effectiveness of federal homeless-assistance grants, and the measurement of homelessness during crisis-driven data disruptions. The first essay studies the effects of source-of-income protection laws on Housing Choice Voucher household mobility and rental market dynamics. Using staggered adoption across cities between 2013 and 2018 with a staggered difference-in-differences design, I find no detectable improvement in neighborhood quality or spatial dispersion for voucher households, but document a 4.9 percent increase in twenty-fifth percentile rents and a 1.8 percentage point decline in affordable rental stock, with effects concentrated where enforcement is stronger and vacancy is low. The second essay estimates the causal effect of federal Continuum of Care and Emergency Solutions Grant funding on homelessness and shelter capacity across 370 Continuums of Care in 2019. Instrumenting per-capita funding with the pre-1940 housing share from the Community Development Block Grant formula, I find that additional funding increases sheltered counts and expands emergency and transitional bed capacity but produces no detectable short-run reduction in unsheltered homelessness, with impacts varying sharply by geography, household type, and demographic subgroup. The third essay develops a two-phase machine learning framework to impute the 61.6 percent of unsheltered Point-in-Time counts missing in 2021 due to the COVID-19 pandemic. The first phase predicts baseline counts from pre-pandemic relationships; the second models COVID-specific deviations using the subset of communities that did conduct counts, with propensity score weighting to address selection bias. The resulting national estimate of 195,191 unsheltered people restores continuity to the national homelessness dataset and reveals that emergency relief fund-

ing dominated the pandemic adjustment. Together, these essays demonstrate that expanding legal protections and scaling federal funding face implementation frictions that limit their reach, and that the administrative data systems underlying policy evaluation remain vulnerable to disruption precisely when accurate measurement matters most.

ACKNOWLEDGEMENTS

I don't have a single person to dedicate this dissertation to, as I'm thoroughly indebted to many wonderful groups of people in my life. I hope I am able to fully express how thankful I am to each and every one of them in this short passage.

First, I want to thank my immediate family: my mom Susannah, my dad John, my brother Henry, and my sister Lucy. Both of my parents have been more supportive of me than I ever could have hoped for, both with guidance and motivation as I've furthered my career and with tangible help (often financial) when I've been struggling. My siblings have endured plenty of my likely insufferable ramblings about work and research over my time in grad school, but have often lended me significant stability and joy when I needed it the most. I wouldn't be in this position where I am today without the four of you, and I hope you know how much you all mean to me.

Similarly, I also want to thank my grandparents Don, Arlyce, Jennifer, John, Rachie, and Bapu for close presence in my life. I feel incredibly lucky that I get to share my experiences with all of you, and I'm so happy that I am able to do so.

Next, I would like to thank my incredible advisor, Dr. Anita Alves Pena. She's been offering me tremendous support since I came to her in October 2021 with an idea for a paper that is now my second chapter. I cannot do justice to how much I've benefited from her continued mentorship, and I am eternally grateful for her efforts to help me grow in this profession. I also want to thank Dr. Tim Komarek, Dr. Ray Miller, and Dr. Jesse Burkhardt for all of the time and energy spent helping me shape my research throughout this process, providing me with crucial feedback, suggestions, and ideas when I very much needed them. Furthermore, I want to thank the other faculty and staff members of the Colorado State Economics Department, as I've had the fortune of sharing positive experiences with every single person. I am so thankful that I was accepted into the CSU community, and I will always carry a sense of pride from being part of it.

Additionally, I also thank my incredible colleagues at the National Wildlife Research Center. Dr. Stephanie Shwiff gave me an opportunity to develop my research skills and have meaningful

experiences working for a great agency, and I'll forever be grateful for the time I spent working there. The incredibly talented pair of Dr. Levi Altringer and Dr. Sophie McKee helped mold me into a far higher-quality researcher, but more importantly, gave me invaluable advice for both my career and personal life. I've shared every minute of my time at the NWRC with Dr. Colin Jareb, Matt Harman, and Molly Selleck, and I deeply appreciate all of the ups and downs we've been through together.

I am especially thankful for the members of my PhD cohort: Milena Dehn, Paula-Leone Samuda, Nara Chung, Greg Miller, Sean Callahan, Matt Harman, and Thomas Gifford. My memories of this experience are in part shaped by the collective experiences we shared together, and I could not have asked for a better group of colleagues and friends. Additionally, I want to thank many other CSU students and now-alumni for the friendships made during my time here: Debora Nunes, Vinicius Cicero, Brendan Brundage, Kendall Stephenson, Austin Landini, Sal McCollum, Lucas Avelar, Brooke Fitzgerald, and countless more incredible people who made my time in the program immensely better.

I also want to acknowledge and thank some people outside of the immediate CSU ecosystem. I am so thankful to have William Cunningham, Matthew Eaton, Maxwell Bullard, and Jackson Vrieze as close friends to share all things related to work and life with. Additionally, I want to thank Bobbi Gifford as a good friend and for putting up with the relentless oddities in our close friend group related to the PhD program.

Lastly, I am forever grateful for the love and support of my fiance Tais Menezes. You are a truly amazing woman, and I will never be able to fully express what you mean to me. If I have achieved anything meaningful during this time, it's because of your endless kindness, patience, and stability you've given to me. I hope I can repay you the same feeling over the course of your journey as well. Eu te amo muito.

TABLE OF CONTENTS

ABSTRACT ii
ACKNOWLEDGEMENTS iv
LIST OF TABLES ix
LIST OF FIGURES xi

Chapter 1 Voucher Household Mobility and Rental Market Dynamics: Evidence from Source of Income Protection Laws 4
1.1 Introduction 4
1.2 Literature Review 9
1.2.1 Voucher search frictions and neighborhood access (direct channel) . . . 9
1.2.2 Evidence on source-of-income discrimination 9
1.2.3 Effects of SOI protections on voucher mobility and location 10
1.2.4 Research gap: market-wide rent effects 11
1.2.5 Contribution of this study 11
1.3 Institutional Background 12
1.3.1 The Housing Choice Voucher Program 12
1.3.2 Source of Income Protection Laws 13
1.3.3 Conceptual Framework 15
1.4 Data and Empirical Strategy 18
1.4.1 Policy Data 18
1.4.2 Data Construction and Sample 19
1.4.3 Outcome Variables 25
1.5 Empirical Strategy 27
1.5.1 Design and Estimator 27
1.5.2 Heterogeneity 29
1.5.3 Robustness 30
1.5.4 Identification and Validity 33
1.6 Results 34
1.6.1 Overall ATTs 34
1.6.2 HCV Mobility Results 36
1.6.3 Rental Market Results 36
1.6.4 Robustness Results 40
1.7 Discussion 44
1.7.1 Why rents rise after SOI protections even when HCV geography does not shift 44
1.7.2 Why the HCV mobility results differ from past studies 46
1.8 Conclusion 48

Chapter 2 Funding, Facilities, and the Face of Homelessness: Heterogeneous Impacts of Federal Grants on Sheltered and Unsheltered Counts 51
2.1 Introduction 51

2.2 Literature Review . . . 53
2.3 Conceptual Framework . . . 54
2.3.1 Funding Determination . . . 54
2.3.2 Heterogeneity in Utilization . . . 57
2.4 Data . . . 58
2.4.1 Point–in–Time Homelessness Counts . . . 58
2.4.2 Shelter and Housing Capacity . . . 60
2.4.3 Federal Homelessness–Assistance Funding . . . 61
2.4.4 Between-CoC Sample Construction . . . 62
2.4.5 Geographically Harmonised Covariates . . . 62
2.5 Methods . . . 65
2.5.1 Primary Specification . . . 65
2.5.2 Between-CoC IV Specification . . . 67
2.5.3 Robustness Checks . . . 68
2.6 Results . . . 68
2.6.1 First-Stage Estimates . . . 68
2.6.2 Sheltered and Unsheltered Homelessness . . . 69
2.6.3 Shelter Capacity: Total Beds and Permanent Supportive Housing . . . 70
2.6.4 Heterogeneity by CoC Category . . . 72
2.6.5 Heterogeneity by Individuals and Families . . . 73
2.6.6 Heterogeneity by Subpopulation . . . 74
2.6.7 Robustness . . . 76
2.7 Discussion . . . 77
2.7.1 Racial Disparities in Unsheltered Homelessness Response . . . 82
2.7.2 The Role of Permanent Supportive Housing . . . 83
2.7.3 Reconciling Results with Prior Studies . . . 84
2.7.4 Cost Per Bed and Per Person Sheltered . . . 86
2.7.5 Conclusion . . . 87
2.8 Tables and Figures . . . 88

Chapter 3 Imputing Missing 2021 Unsheltered Homelessness Counts: A Two-Phase Machine Learning Error Correction Approach . . . 101
3.1 Introduction . . . 101
3.2 Background . . . 105
3.2.1 Homelessness Data in the United States . . . 105
3.2.2 The 2021 Data Gap . . . 106
3.2.3 Research on Homelessness Count Data . . . 108
3.2.4 Machine Learning in Social Policy Research . . . 109
3.2.5 Imputation Methods and Error Correction Approaches . . . 110
3.3 Data and Method . . . 111
3.3.1 Methodological Overview . . . 111
3.3.2 Data Sources . . . 113
3.3.3 Data Construction and Sample Inclusion Criteria . . . 119
3.3.4 Phase 1: Baseline Prediction Model . . . 122
3.3.5 Phase 2: COVID Adjustment Model . . . 126

3.3.6 Uncertainty Quantification . . . 130
3.4 Results . . . 131
3.4.1 Phase 1: Baseline Model Performance . . . 131
3.4.2 Phase 2: COVID Adjustment Model . . . 138
3.4.3 Final Imputed Unsheltered Counts and National Estimates . . . 143
3.5 Discussion . . . 149
3.5.1 Key Findings . . . 149
3.5.2 Methodological Contributions . . . 150
3.5.3 Limitations . . . 152
3.5.4 Future Directions . . . 153
3.5.5 Conclusion . . . 155

Appendix A Supplementary Material for Chapter 1 . . . 168
A.1 Chapter 1 Appendix . . . 168
A.1.1 Example of "No Section 8" ad . . . 168
A.1.2 Callaway & Sant'Anna (2021) Estimation Details . . . 171
A.1.3 Anticipation Effects Event Study Graphs . . . 172
A.1.4 Event-time Support . . . 174

Appendix B Supplementary Material for Chapter 2 . . . 175
B.1 Chapter 2 Appendix . . . 175
B.1.1 Details of Tract-CoC Crosswalk Construction . . . 175
B.1.2 Robustness: Population in Poverty Results . . . 182
B.1.3 Robustness: Excluding New York City and Los Angeles CoCs . . . 183

Appendix C Supplementary Material for Chapter 3 . . . 184
C.1 Chapter 3 Appendix . . . 184

LIST OF TABLES

1.1 Places Adopting Source of Income Protection Policies, 2013–2018 20
1.2 Variable definitions and data sources 23
1.3 Baseline characteristics stratified by treatment status, 2012 Values 24
1.4 HCV Mobility Outcomes: Definitions and Sources 27
1.5 Rental Market Outcomes: Definitions and Sources 28
1.6 Summary Statistics for Outcome Variables - 2012 values 28
1.7 SOI Protections Policy Strength Scoring Framework 30
1.8 Places Adopting SOI Protection Policies by Law Strength and Vacancy Level, 2013–2018 31
1.9 Summary of Average Treatment Effects on the Treated (ATT) 35
1.10 Robustness Check: Synthetic Staggered Difference-in-Differences Results 43

2.1 Summary Statistics: Homelessness and Housing Inventory, 2015-2019 Averages 90
2.2 Summary Statistics: Federal Funding and Community Covariates, 2015-2019 Averages 91
2.3 First Stage: Effect of Pre-1940 Housing Share on Federal Funding 92
2.4 Main Results: Effect of Federal Funding on Sheltered and Unsheltered Homelessness . 93
2.5 Housing Inventory Results: Effect of Federal Funding on Bed Capacity 94
2.6 Effects on Seasonal and Overflow Bed Capacity 95
2.7 Heterogeneous Effects by CoC Geography: Urban vs. Rural/Suburban Areas - 2019 . . 96
2.8 Federal Funding Effects by Household Type and Shelter Status - 2019 97
2.9 Heterogeneous Effects on Sheltered Population Subgroups - 2019 98
2.10 Heterogeneous Effects on Unsheltered Population Subgroups - 2019 99
2.11 Comparison of Federal Funding Effects Across Studies 100

3.1 Phase 1 Predictor Variables: Descriptions and Data Sources (Part I) 115
3.2 Phase 1 Predictor Variables: Descriptions and Data Sources (Part II) 116
3.3 Phase 2 Predictor Variables: COVID-19 Adjustment Model 117
3.4 Phase 1 Modeling Sample (2015–2019) Summary Statistics 124
3.5 Phase 2: Covariate Balance by 2021 Counting Status 127
3.6 Phase 2 (2021) Summary Statistics: COVID Adjustment Predictors 128
3.7 CoC Urbanicity Composition by 2021 Count Completion Status 129
3.8 Phase 1: Model Comparison on 2020 Holdout Set 132
3.9 Phase 1: Prediction Accuracy by CoC Urbanicity (2020 Holdout) 133
3.10 Phase 1 Baseline Predictions: Summary Statistics (2021) 134
3.11 Phase 1: XGBoost Feature Importance (Top 20 Predictors) 135
3.12 Phase 1 Baseline Predictions: Random Sample by CoC Category 137
3.13 Phase 2: Overlap Weights Distribution (Treated CoCs) 138
3.14 Phase 2: COVID Adjustment Model Comparison 139
3.15 Phase 2 Predicted COVID Adjustments: Summary Statistics 140
3.16 Phase 2 COVID Adjustment Predictions: Random Sample by CoC Category 141
3.17 Phase 2: XGBoost Feature Importance (COVID Adjustment Model) 142

3.18 National Unsheltered Homelessness Estimates, 2015–2022 144
3.19 Top 30 CoCs by Imputed 2021 Unsheltered Count 145
3.20 Prediction Interval Widths by CoC Category 147
3.21 CoCs with Widest 90% Prediction Intervals 147
3.22 CoCs with Narrowest 90% Prediction Intervals 148

B.1 Overall Crosswalk Validation Statistics (2019) 181
B.2 Effects on Homelessness Rates Among People in Poverty 182
B.3 Robustness Check: Excluding New York City and Los Angeles CoCs 183

C.1 Imputed 2021 Unsheltered Counts: Part 1 of 5 (AK–CA-530) 190
C.2 Imputed 2021 Unsheltered Counts: Part 2 of 5 (CA-531–GA-508) 191
C.3 Imputed 2021 Unsheltered Counts: Part 3 of 5 (HI-500–MN-511) 192
C.4 Imputed 2021 Unsheltered Counts: Part 4 of 5 (MO-500–PA-603) 193
C.5 Imputed 2021 Unsheltered Counts: Part 5 of 5 (PA-605–WY-500) 194

LIST OF FIGURES

1.1 Adoption of SOI Protection Laws in US Cities 5
1.2 Geographic Distribution of SOI Protection Law Adoption, 2013–2018 6
1.3 Changes in Poverty Rate and HCV Household Share by Census Tract, 2012 $\rightarrow$ 2019 (Portland above; Dallas below). 26
1.4 Dynamic Effect Estimates for HCV Household Mobility Outcomes 37
1.5 Dynamic Effect Estimates for Weighted Tract Poverty Rate by Stratification 38
1.6 Dynamic Effect Estimates for Normalized HHI by Stratification 39
1.7 Dynamic Effect Estimates for Rental Affordability Outcomes 40
1.8 Dynamic Effect Estimates for 25th Percentile Rents by Stratification 41
1.9 Dynamic Effect Estimates for the Share of Rental Apartments Affordable at the $\leq$ 30% HAMFI Threshold by Stratification 42

2.1 Growth in Federally Allocated Funding for Homelessness Assistance vs. Counts of Sheltered and Unsheltered Homelessness in the U.S., 2015-2019, Authors' Calculations 89

3.1 Missing Counts in the HUD 2022 CoC Performance Profile 102
3.2 Map of CoCs by 2021 Unsheltered Count Status 107
3.3 Methodological Flowchart: Two-Phase Imputation Framework 114
3.4 Visual Diagram of Phase 1 Training Process 125
3.5 Phase 1 Actual-Predicted Scatterplot, Log Scale 133
3.6 Phase 1 SHAP Values 136
3.7 National Unsheltered Homelessness Time Series, 2015–2022 146

A.1 "No Section 8" Craigslist Housing Listings 169
A.2 Anticipation Check: Dynamic Effect Estimates for HCV Household Mobility Outcomes 172
A.3 Anticipation Check: Dynamic Effect Estimates for Rental Affordability Outcomes . . . 173
A.4 Contributing Treated Place-Year Units by Event Time 174

C.1 Phase 1 Feature Importance 184
C.2 Love Plot of Standardized Mean Difference Before vs. After Reweighting, Phase 2 . . 185
C.3 Phase 2 Feature Importance 186
C.4 Phase 2 SHAP Values 187
C.5 Cleveland Plot: Prediction Intervals of 30 CoCs by Category 188
C.6 Cleveland Plot: Prediction Intervals of Top 30 CoCs 189

Introduction

Homelessness in the United States persists despite decades of federal investment and policy innovation. Each January, the Department of Housing and Urban Development requires communities nationwide to count every person experiencing homelessness, and each year those counts reveal a population that has remained stubbornly above half a million people. The federal response operates through two principal channels: demand-side subsidies, primarily the Housing Choice Voucher program, that help households afford private-market housing; and supply-side grants, principally the Continuum of Care and Emergency Solutions Grant programs, that fund shelter beds, transitional housing, and permanent supportive housing. Whether these instruments effectively reduce homelessness and expand housing access depends on implementation details, market conditions, and measurement infrastructure that are often taken for granted. This dissertation examines all three dimensions through three independent essays unified by a common focus on the institutions, data systems, and market dynamics that shape homelessness policy in the United States.

Chapter 1 studies the demand-side access problem. The Housing Choice Voucher program serves roughly 2.3 million households annually, yet landlords routinely reject voucher applicants before any assessment of tenant quality. In response, a growing number of jurisdictions have adopted source-of-income protection laws that prohibit categorical rejection of voucher users. Using staggered adoption across cities with populations exceeding 65,000 between 2013 and 2018, I estimate the effects of these protections on two sets of outcomes: the neighborhoods where voucher households live and the rental market conditions in adopting places. Identification follows a staggered difference-in-differences design with doubly robust estimation to account for systematic differences between early adopters, late adopters, and never-adopters. I find no detectable improvement in voucher household neighborhood quality or spatial dispersion, but document significant increases in lower-tier rents and declines in affordable rental stock, with effects concentrated where enforcement is stronger, exemptions are fewer, and baseline vacancy is low. These results suggest that removing explicit refusal barriers may be insufficient to deliver neighborhood gains when supply-side constraints and early-stage screening frictions remain binding.

Chapter 2 turns to the supply-side funding channel. Despite a 16 percent real-term increase in HUD homeless-assistance appropriations between fiscal years 2015 and 2019, annual Point-in-Time counts remained above 550,000 and unsheltered homelessness persisted near 200,000 persons. This essay estimates the causal effect of combined Continuum of Care and Emergency Solutions Grant funding on both reported homelessness and shelter-bed capacity across 370 Continuums of Care in 2019. To address the mechanical endogeneity between measured homelessness and formula-based funding, I instrument each community's per-capita funding with its share of housing units built before 1940 in the Community Development Block Grant formula. The results show that additional federal dollars expand emergency and transitional shelter capacity and increase sheltered counts, but produce no detectable short-run reduction in unsheltered homelessness. Impacts vary sharply by geography, household type, and demographic subgroup, revealing that the same dollar amount operates through different channels depending on local conditions and population composition.

Chapter 3 addresses the measurement infrastructure that underlies both funding decisions and policy evaluation. In January 2021, the COVID-19 pandemic led 61.6 percent of Continuums of Care to forgo their unsheltered Point-in-Time count, creating the largest single-year data gap in the history of the national homelessness enumeration. This essay develops a two-phase machine learning framework to reconstruct the missing counts. The first phase trains predictive models on a 2015–2019 panel of community characteristics to establish what unsheltered counts would have been under pre-pandemic conditions. The second phase leverages the 149 communities that did conduct complete counts to model the COVID-specific deviation from baseline, applying propensity score weighting to address selection bias in the responding sample. The resulting national estimate of 195,191 unsheltered individuals, with a 90 percent prediction interval spanning 114,380 to 255,978, restores continuity to a dataset essential for longitudinal research. The framework also reveals that emergency relief funding, particularly ESG-CARES allocations, dominated the pandemic adjustment, consistent with direct shelter funding temporarily suppressing unsheltered homelessness during the crisis period.

Taken together, these essays show that the effectiveness of homelessness policy depends not only on the scale of resources but on market structure, implementation design, and the integrity of the data systems used to evaluate outcomes. Legal protections expand formal access without necessarily changing where voucher households live; federal grants expand shelter capacity without necessarily reducing unsheltered homelessness; and crisis events can destroy the measurement infrastructure needed to assess either channel. Each essay addresses a distinct empirical question, but the common thread is that policy levers interact with local conditions in ways that aggregate statistics and broad program evaluations can obscure.

Chapter 1

Voucher Household Mobility and Rental Market Dynamics: Evidence from Source of Income Protection Laws

1.1 Introduction

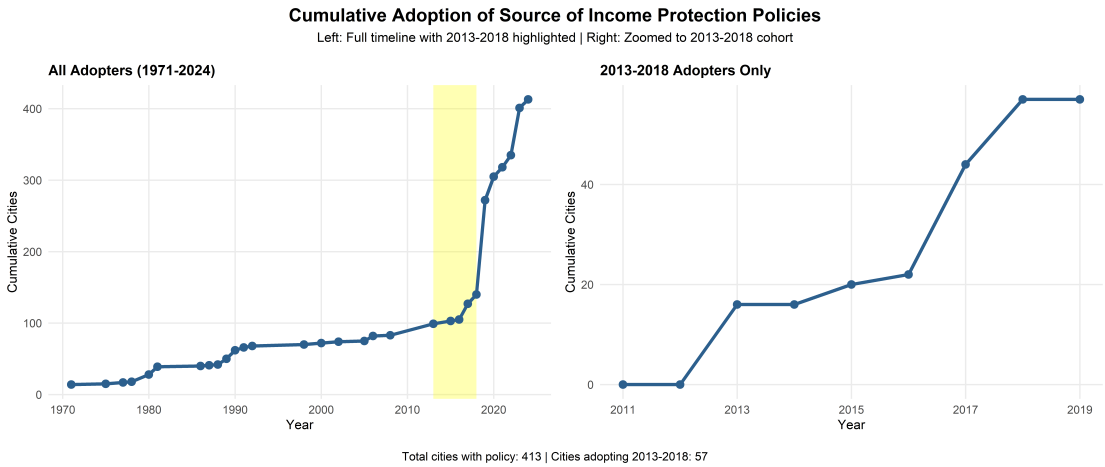
The Housing Choice Voucher (HCV) program serves about 2.3 million low-income households annually (U.S. Department of Housing and Urban Development, 2025), yet landlords routinely reject voucher applicants before any assessment of tenant quality.[1] In response, many jurisdictions adopted source-of-income (SOI) protections prohibiting categorical rejection of voucher users. Among census-designated places with populations exceeding 65,000, the cumulative number covered by SOI protections grew from 140 in 2018 to 413 by 2024. Adoption has proceeded unevenly, producing a patchwork of coverage and scope that varies widely in timing, enforcement strength, and exemptions. The cumulative adoption timing across all passed laws and those within the sample period of this paper can be seen in Figure 1.1.

Figure 1.2 illustrates the geographic distribution of this policy diffusion, showing that early adoption (2013-2016) occurred in scattered jurisdictions across the country, while the 2017 to 2018 period saw concentrated adoption along the east and west coasts.

The barrier is practical as well as legal. In jurisdictions without protections, voucher seekers encounter non-response, categorical refusals, and steering, often before any assessment of tenant quality. Field and administrative evidence point to high rejection rates for voucher applicants, slow lease-up, and concentration of successful placements in familiar, higher-poverty submarkets that openly solicit voucher tenants (Tighe et al., 2016; Phillips, 2017; Cunningham et al., 2018). These

[1] See Figure A.1 as an example.

Figure 1.1: Adoption of SOI Protection Laws in US Cities



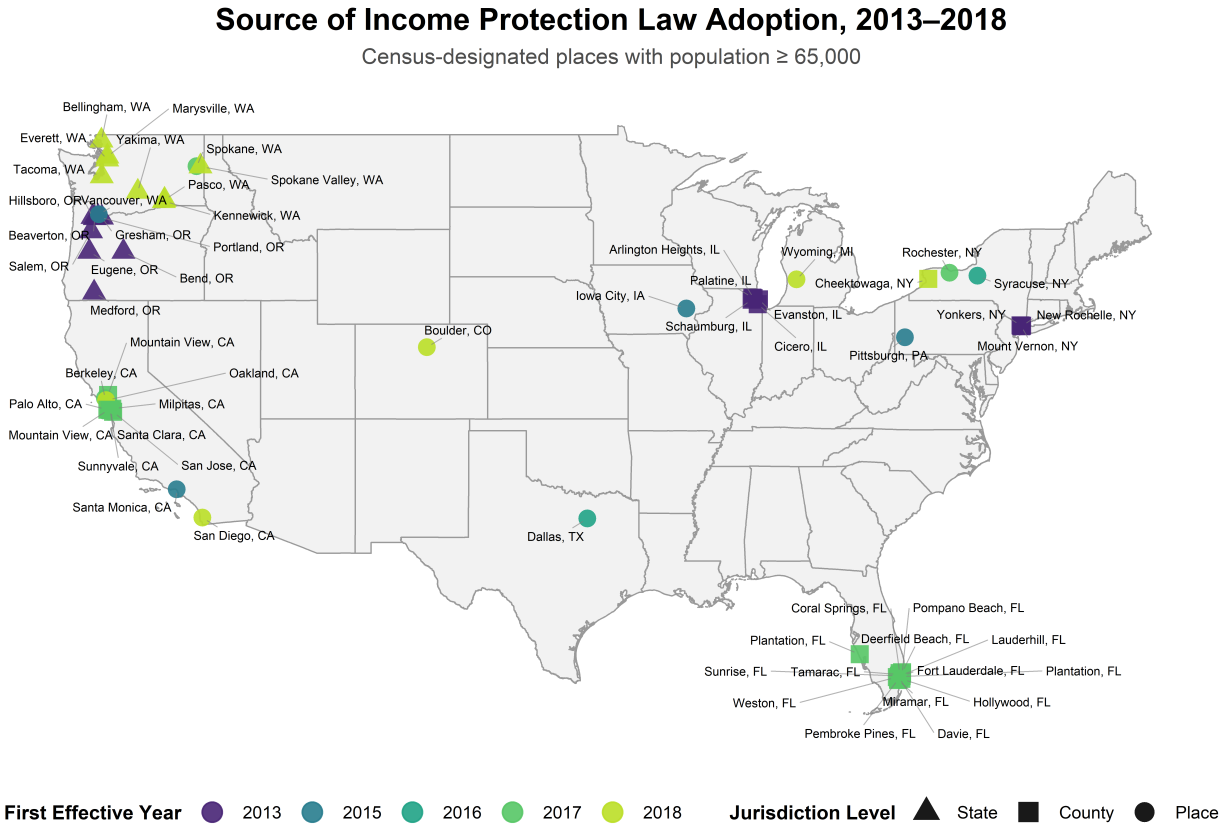
frictions interact with time limits, inspection scheduling, and rent-reasonableness requirements to compress search into segments where acceptance is predictable, undermining the program's deconcentration goals. Furthermore, many newly issued vouchers are never successfully used within standard search windows, underscoring the depth of frictions at the lease-up margin. Using HUD administrative records across 200 PHAs (2014–2022), Ellen et al. (2025) estimates that only about 60 percent of households issued a voucher successfully lease a unit within six months. This take-up shortfall coexists with continuing concentration of successful lease-ups in higher-poverty submarkets.

SOI protections target the explicit refusal margin by making "No HCV", "No Voucher", or "No Section 8"[2] policies unlawful. Whether those protections alter outcomes depends on details that vary across places. This variation is central to both the empirical strategy and the interpretation of effects.

This paper studies two questions that follow directly from these channels. First, do SOI protections expand neighborhood choice for voucher households, as reflected in changes to where voucher households live across tracts and how they are distributed within places? Second, do SOI

[2]The Housing Choice Voucher program is commonly referred to as "Section 8," referencing Section 8 of the Housing Act of 1937 that originally established the federal housing assistance framework. While the current voucher program was created by the Housing and Community Development Act of 1974 and substantially reformed in 1998, the "Section 8" terminology persists in both policy discussions and popular usage. Landlord advertisements stating "No Section 8" typically refer to rejection of Housing Choice Voucher recipients.

Figure 1.2: Geographic Distribution of SOI Protection Law Adoption, 2013–2018



Sources: Urban Institute; NLIHC; TIGRIS boundary files.
Note: Points represent place centroids. Color indicates first effective year; shape indicates jurisdiction level of enacting law.

Notes: Maps represent census-designated places above 65,000 population that passed a Source of Income protection law at any point between 2013 and 2018. Place information is gathered from ACS 1-Year surveys. Point geometry is gathered from TIGRIS boundary files.

protections generate broader general-equilibrium effects in rental markets, particularly at the lower end of the rent distribution where voucher-affordable units are concentrated?

The empirical setting uses staggered SOI adoption across incorporated places with populations above 65,000 during 2011–2019, focusing on first effective dates from 2013 to 2018 to avoid confounding variation from the Great Recession peak and the COVID-19 pandemic. The panel links detailed policy features to two sets of outcomes. The first set captures voucher mobility and spatial distribution at the place level: weighted tract poverty exposure, the average number of voucher households per occupied tract, the percentage of all census tracts with any HCV households, and a normalized Herfindahl index of voucher concentration. The second set captures rental market conditions: the twenty-fifth and seventy-fifth percentiles of contract rent, the rental vacancy rate, and the share of the rental stock affordable at or below 30 percent of Area Median Family Income.

Identification follows Callaway and Sant'Anna (2021). Treated units are places with first effective dates in 2013–2018; the counterfactual is the set of places that do not adopt within this window. Dynamic event-time estimates summarize the evolution of outcomes before and after adoption and allow inspection of pre-trends. Because eventual adopters and never-adopters differ systematically on observable characteristics that relate to both adoption and outcomes, dynamic treatment effects are estimated with a doubly robust approach that conditions on 2012 baseline covariates, mitigating bias from pre-policy differences while preserving the staggered-adoption design. Because policy design and market tightness plausibly shape the magnitudes of the dynamic effects, heterogeneity is examined along three dimensions emphasized by institutions and prior evidence: stronger versus weaker enforcement, fewer versus more exemptions, and lower versus higher baseline (2012) rental vacancy as a proxy for market tightness.

Across all mobility measures there is no detectable change in the neighborhoods where voucher households live or in the spatial dispersion of voucher households within places following SOI adoption. By contrast, lower-tier rents rise after adoption, with effects concentrated where enforcement is stronger, exemptions are fewer, and baseline vacancy is low. The twenty-fifth percentile rent rises by about 4.9 percent on average (5.5 percent in strong-law places and 5.2 percent in

low-vacancy places), while the share of units affordable at or below 30 percent of Area Median Family Income declines by 1.8 percentage points on average (2.1 percentage points in strong-law places and 1.9 percentage points in low-vacancy places). The pattern is consistent with adjustments in tight, voucher-relevant segments that shift prices without materially altering the geography of voucher households. These results remain robust under alternative empirical specifications: synthetic difference-in-differences estimation as an alternative to the Callaway & Sant'Anna approach, and incorporating one year of anticipation effects to account for forward-looking behavior by housing providers and HCV households.

This study contributes in three ways. First, it examines market-wide rental price effects of SOI protections using a staggered-adoption design, providing evidence on a dimension that remains underexplored. Second, it aligns outcome definitions with policy scope by distinguishing mover-focused findings in prior work from place-level stock measures, and by evaluating multiple mobility and spatial concentration metrics rather than a single neighborhood indicator. Third, it investigates heterogeneity along policy and market dimensions highlighted by institutional context and prior evidence (enforcement strength, exemptions, and baseline market tightness), offering insight into where and why effects are larger.

These findings speak to SOI protection and broader tenant protection policy design. Reducing explicit refusal may be insufficient to deliver neighborhood gains in the presence of early-stage screening frictions and limited short-run supply at the low end. Complementary tools, such as faster inspections and lease-up processes, targeted landlord engagement or bonuses, payment-standard reforms, and additional supply in voucher-affordable segments, may be necessary for translating legal protections into improved residential outcomes without amplifying rent pressure.

The remainder of the paper proceeds as follows. Section 2.2 reviews related literature on voucher discrimination and SOI protections. Section 1.3 describes institutional background and policy measurement. Section 3.3 documents data sources, outcome definitions, and baseline covariates. Section 2.5 outlines the estimation strategy and heterogeneity design. Section 3.4 presents

the main and subgroup results. Section 2.7 interprets mechanisms and policy implications, and Section 1.8 concludes.

1.2 Literature Review

1.2.1 Voucher search frictions and neighborhood access (direct channel)

A large literature documents persistent barriers that constrain Housing Choice Voucher recipients' residential choices. Early work established that voucher households disproportionately lease in higher-poverty neighborhoods, limiting progress toward deconcentration goals (Devine et al., 2003; Pendall et al., 2014). Subsequent studies confirm that these patterns have proven durable despite program refinements (McClure et al., 2015; Ellen, 2020). Qualitative research underscores that clustering reflects constraints rather than preferences: voucher households routinely encounter categorical refusals, informational gaps, tight timelines, and administrative frictions that compress searches into familiar submarkets (DeLuca, 2014; Graves, 2016; Galvez, 2010). These frictions interact with inspection scheduling and rent reasonableness to tilt successful lease-up toward segments where acceptance is predictable.

1.2.2 Evidence on source-of-income discrimination

Prevalence and geographic variation

Audit and correspondence studies consistently find substantial discrimination against voucher users. A HUD-sponsored multi-metro audit documented wide acceptance rate variation across markets and neighborhood types, with rejection especially pronounced in low-poverty areas (Cunningham et al., 2018). Correspondence designs reach similar conclusions in diverse settings: landlords frequently fail to respond to inquiries that disclose voucher use or impose conditions that effectively exclude voucher households (Turner et al., 1999; Phillips, 2017; Aliprantis et al., 2019; Hangen, 2022). These patterns suggest that, absent legal protections, explicit source-of-income screening is pervasive and systematically narrows the effective search set for voucher recipients.

Landlord responses under regulation (screening, pricing, timing)

Evidence also indicates that landlord behavior adapts as legal constraints evolve. Observational and practitioner reports describe early-funnel strategies—non-response, discouragement, and shifts to minimum income and credit thresholds—that are difficult to detect in complaint-driven systems (Unlock NYC et al., 2022). Correspondence experiments around application-screening reforms in Minneapolis show substitution toward earlier-stage discrimination against specific groups, highlighting how enforcement gaps at initial contact can blunt policy intent (Gorzig and Rho, 2025). Relatedly, field evidence that monetary incentives increase landlord willingness to consider vouchers points to acceptance margins that respond to expected returns and perceived costs (Aliprantis et al., 2019; Collinson and Ganong, 2018; Desmond and Perkins, 2016; Rosen, 2014). Recent work also documents landlord strategy adjustment under SOI protection regimes, such as adapting to screening on more restricive credit requirements or strategically reducing the chances of a successful housing authority inspection (Lucio and Cho, 2025; Cho and Lucio, 2025). Together, these studies map the margins (screening, pricing near payment standards, and timing) along which owners can comply formally while maintaining effective control over tenant selection.

1.2.3 Effects of SOI protections on voucher mobility and location

Research on the impacts of SOI protections has focused on utilization and locational outcomes, with a growing distinction between mover-specific effects and changes in the stock distribution. Early difference-in-differences analyses reported higher voucher utilization following adoption (Freeman, 2012) and modest improvements in neighborhood characteristics where voucher households reside (Freeman and Li, 2014). More recent work using administrative records and larger samples emphasizes temporal dynamics and mover focus: voucher movers experience greater reductions in tract poverty after adoption, with effects that tend to materialize several years post-enactment (Ellen et al., 2022). Event-study evidence on families with children similarly finds rising access to low-poverty areas with effects emerging over a longer horizon, and with larger gains for Black families (Teles and Su, 2022). Syntheses conclude that SOI protections yield meaningful

but gradual improvements, conditioned by local design features and enforcement capacity, and that discrimination persists through less visible channels (Galvez and Knudsen, 2024; Lens et al., 2011; Basolo and Nguyen, 2005). A notable gap in this literature concerns the effectiveness of enforcement practices and the extent to which exemptions limit coverage, issues that motivate attention to heterogeneity by enforcement strength and exemptions.

1.2.4 Research gap: market-wide rent effects

While several studies examine rents and affordability in voucher-accessible units, little evidence directly links SOI protections to market-wide price dynamics. Theoretically, landlord adjustments to higher expected administrative or legal costs, combined with short-run inelastic supply in low-tier segments, could generate upward pressure on rents; related policy contexts document such equilibrium responses (Abramson, 2024; Coulson et al., 2025). Whether similar general-equilibrium effects arise following SOI adoption remains largely untested, motivating explicit examination of lower-tier rents and the supply of affordable units alongside neighborhood outcomes.

1.2.5 Contribution of this study

This study extends the literature in three ways. First, it examines market-wide rental price effects of SOI protections using a staggered-adoption design, contributing evidence on a dimension that remains underexplored. Second, it aligns outcome definitions with policy scope by distinguishing mover-focused findings in prior work from place-level stock measures, and by evaluating multiple mobility and spatial concentration metrics rather than a single neighborhood indicator. Third, it investigates heterogeneity along policy and market dimensions highlighted by institutional context and prior evidence—enforcement strength, exemptions, and baseline market tightness—providing insight into where and why effects are likely to be larger.

1.3 Institutional Background

1.3.1 The Housing Choice Voucher Program

The HCV program represents the cornerstone of federal rental assistance policy in the United States. Administered by approximately 2,200 local Public Housing Authorities (PHAs) under federal oversight from the Department of Housing and Urban Development (HUD), the program currently serves 2.3 million low-income households with portable rental subsidies. Unlike the place-based public housing developments that dominated earlier federal housing policy, vouchers embody a market-oriented approach that allows recipients to seek housing throughout the private rental market.

The voucher program's payment standard structure mechanically concentrates voucher demand in the lower tier of the rental market. Payment standards are typically set between 90 and 110 percent of the Fair Market Rent (FMR), which itself is pegged at approximately the 40th percentile of area rents (Ellen et al., 2025). This creates a binding constraint: voucher households can only afford units priced at or below the payment standard without paying additional rent out-of-pocket. Once a household receives a voucher, it contributes approximately 30 percent of its adjusted income toward rent, while the PHA pays the landlord the difference up to the locally determined payment standard (Ellen, 2020).

This constraint has predictable geographic consequences. Collinson and Ganong (2018) document that under uniform metropolitan-area FMRs in Dallas, vouchers covered the cost of 68 percent of units in low-rent neighborhoods but only 15 percent in high-rent neighborhoods. More recent evidence confirms that voucher holders disproportionately lease units in ZIP codes where the Small Area FMR is below the metropolitan average (Ellen et al., 2025), concentrating transactions in the lower tier of each market's rent distribution.

For the empirical analysis, this institutional reality motivates the choice of the 25th percentile of contract rent, available through ACS 1-year data, as the policy-relevant outcome. This percentile captures the segment where voucher households actually compete for units and where landlords' responses to SOI protections—whether through pricing, screening, or participation decisions—are

most likely to materialize. By contrast, the 75th percentile contract rent serves as a falsification test: because voucher households are mechanically excluded from this segment by payment standard constraints, SOI laws should generate no effect there if the mechanism operates through voucher-specific landlord adjustments rather than city-wide rental market trends.

The voucher utilization process involves several critical steps that create potential friction points. After receiving a voucher, families typically have 60 to 90 days to locate suitable housing, though PHAs may grant extensions in difficult market conditions. The chosen unit must pass HUD's Housing Quality Standards inspection, and the landlord must agree to sign a Housing Assistance Payment (HAP) contract with the PHA. Only after these requirements are met can the family move in and begin receiving rental assistance. This timeline pressure is intensified by PHA performance incentives: housing authorities are expected to maintain utilization rates of 90-98 percent of their authorized vouchers, with funding consequences for those that fall significantly below these targets. The shortness of lease-up windows, inspections and rent-reasonableness determinations must be completed before move-in, and vouchers can expire absent extensions, which compresses search into submarkets with predictable acceptance (Tighe et al., 2016; Cunningham et al., 2018; Ellen, 2020).

Despite the program's market-based design, voucher holders face substantial barriers that limit their housing choices. Research consistently shows that many recipients struggle to use their vouchers within the allotted timeframe, with success rates varying dramatically across metropolitan areas and demographic groups. Those who do successfully lease units often find themselves concentrated in higher-poverty neighborhoods, undermining the program's poverty deconcentration objectives.

1.3.2 Source of Income Protection Laws

Legal Framework and Geographic Variation

Source-of-income protection laws target a central barrier faced by voucher households: categorical refusal by owners to consider tenants who pay with housing subsidies. SOI is not a

protected class under the federal Fair Housing Act; coverage is provided by state and local law (Schwemm, 2016). These SOI protection laws add "source of income" to state or local fair housing statutes and make it unlawful to reject an applicant solely because rent would be paid in part by a Housing Choice Voucher. Where covered, owners are expected to apply the same screening criteria used for other applicants—credit, rental history, and other non-voucher factors—rather than exclude voucher users per se.

However, statutory language differs in whether housing vouchers are explicitly included, implied, or excluded from the protected class described in each law. Recent statewide adoptions include Hawaii, which now prohibits discrimination based on participation in voucher programs, and Delaware, which enacted statewide protections in 2024 with renter coverage taking effect in 2026. By contrast, Wisconsin's state law has been interpreted not to treat federal vouchers as a protected "lawful source of income," leaving acceptance voluntary under state law unless a local ordinance applies. In several states, legislatures have preempted local governments from adopting their own SOI ordinances; Iowa and Texas are prominent examples.

Coverage also reflects political and market conditions. Jurisdictions adopting SOI protections tend to be higher-income, with an older and more college-educated population, and statutes differ in the breadth of units covered via exemptions and carve-outs (Cho and Lucio, 2025). In the period relevant for the empirical analysis, 2013–2019, adoption accelerated. Within the estimation sample of large incorporated places, 57 census-designated places adopted SOI protections with first effective dates between 2013 and 2018. This variation in timing, enforcement strength, and exemptions motivates the staggered research design and the heterogeneity analyses reported below.

Policy Mechanisms and Implementation Challenges

SOI protections are enforced through a mix of administrative and judicial channels. Most jurisdictions rely on complaint-driven processes administered by civil rights or human rights agencies; some provide for private rights of action with damages and fee-shifting. Enforcement capacity and remedies vary, and so does effective coverage: many laws include exemptions for specific property types or owners (for example, owner-occupied small buildings, certain small-landlord thresholds,

or units already governed by other programs), which can leave a substantial share of the stock outside the rule's reach. For measurement, the analysis distinguishes enforcement strength (agency authority, investigative tools or testing, penalty structure, private right of action) and the prevalence of exemptions; Section 3.3 details the coding rubric.

Implementation faces several structural challenges. First, early-funnel behavior is difficult to police in complaint-driven systems: non-response to inquiries, discouragement, or shifting to minimum-income and credit thresholds can sustain effective exclusion even where explicit refusal is unlawful (Unlock NYC et al., 2022; Cunningham et al., 2018; Phillips, 2017). Second, payment standards and rent-reasonableness tests cap what a voucher household can transact, so units priced even slightly above the standard may remain effectively out of reach despite formal protections (Ellen et al., 2025). Third, administrative requirements, such as inspections, Housing Assistance Payment contracts, and sequencing with PHA approval, impose time and process costs that can deter participation independent of legal obligations to consider voucher applicants. These features imply that the same statute may bind strongly in jurisdictions with robust enforcement and few exemptions, and bind weakly where coverage is narrow or markets are tight, motivating the focus on heterogeneity by enforcement, exemptions, and baseline vacancy (Lucio and Cho, 2025).

1.3.3 Conceptual Framework

SOI protection laws operate through two primary channels that can produce both intended benefits for voucher holders and unintended consequences for the broader rental market. This section outlines the theoretical mechanisms through which SOI policies affect housing market outcomes, focusing on the behavioral responses of both voucher recipients and landlords.

Voucher Holder Response Channel

The first channel through which SOI laws operate is by expanding the effective choice set available to voucher recipients. Prior to policy adoption, voucher holders face a constrained search process where a substantial portion of the rental market is explicitly off-limits due to "No Section 8" policies. Rational voucher holders, aware of this discrimination, may limit their search efforts

to landlords known to accept vouchers or neighborhoods with high concentrations of subsidized housing, even if they would prefer to live elsewhere.

When SOI protections are enacted, voucher holders gain legal recourse against discriminatory denials and can reasonably expect fairer treatment from a broader set of landlords. This expansion of the effective choice set should induce voucher holders to search more widely across neighborhoods and property types, including areas they might have previously avoided due to anticipated discrimination. The magnitude of this response depends on several factors: voucher holders' awareness of the new legal protections, their confidence in enforcement mechanisms, and their preferences for neighborhood characteristics versus the costs and uncertainty of expanded search.

For SOI laws to generate measurable effects on voucher holder outcomes, this behavioral response is necessary but not sufficient. Voucher holders must actually expand their search patterns and apply to landlords who would have previously rejected them outright. If recipients continue to limit their search to the same subset of accommodating landlords, the policy would have little impact regardless of its legal force.

Landlord Response Channel

The second channel operates through landlord behavioral responses to the new legal constraint on their tenant selection process. Prior to SOI adoption, landlords who preferred not to rent to voucher holders could simply refuse such applicants without legal consequence. This allowed them to avoid perceived costs or risks associated with voucher tenants—whether real administrative burdens, unfounded stereotypes, or preferences for particular tenant types[3].

SOI laws eliminate landlords' ability to categorically exclude voucher applicants, forcing them to evaluate such applicants using the same criteria applied to all prospective tenants. However, landlords retain considerable discretion in how they respond to this constraint, leading to several possible equilibrium adjustments.

[3]HUD issued a fact sheet about the HCV program to attempt to change landlord perception on voucher tenants - https://files.hudexchange.info/resources/documents/PIH-HCV-Landlord-Myth-Busting-and-Benefits-Fact-Sheet.pdfFact Sheet

First, some landlords may comply with the law's intent by genuinely evaluating voucher applicants on their merits. These landlords might discover that their previous aversion to voucher tenants was unfounded, potentially leading to increased acceptance rates and improved outcomes for voucher holders.

Second, landlords who continue to view voucher tenants as higher-cost or higher-risk may adjust their rental terms to reflect these perceived costs. This could manifest as higher security deposits, stricter screening criteria that disproportionately affect voucher holders, or most importantly for our analysis, higher rental prices. If landlords anticipate increased costs from a tenant pool that now includes more voucher holders, they may raise rents preemptively to maintain their expected returns.

Third, some landlords may attempt to circumvent the law through subterfuge, using pretextual reasons to reject voucher applicants or steering them away through discouraging behavior. The prevalence of such evasion depends critically on enforcement strength and the penalties for violations.

Finally, landlords operating in high-demand markets or luxury segments may find that SOI laws have little practical impact on their tenant selection. If these landlords receive numerous applications from qualified non-voucher tenants, they may rarely encounter voucher applicants in the first place, or may legitimately reject them based on income requirements that exceed voucher payment standards.

Market-Level Implications

The interaction of these two channels produces aggregate effects that may extend beyond the voucher population. If SOI laws successfully increase the effective demand for rental housing by enabling more voucher holders to compete for units throughout the market, and if landlords respond by raising rents to offset perceived costs, the result could be upward pressure on rental prices that affects all low-income renters.

However, rent effects may be limited or absent if SOI laws fail to meaningfully constrain landlord behavior in practice. Landlords may continue excluding voucher holders through several

mechanisms that render the legal prohibition ineffective. Weak enforcement capacity or complaint-driven systems may create insufficient deterrence, allowing continued discrimination with minimal consequences. The prevalence of legal exemptions—such as owner-occupied properties or small landlords—may preserve discriminatory exclusion for substantial portions of the rental market. Additionally, landlords in high-demand markets with numerous applications may rarely encounter situations where they must choose between renting to voucher holders or leaving units vacant, effectively nullifying the policy's binding constraint. In such contexts, SOI laws would produce limited behavioral change among landlords, resulting in minimal rent adjustments and smaller improvements in voucher holder access.

This rent effect could be particularly pronounced in the lower tier of the rental market, where voucher holders are most likely to compete with other low-income households. Higher rents in this segment could price out some non-voucher low-income households, potentially offsetting some of the gains achieved by voucher recipients. The net welfare effect would depend on the relative magnitudes of improved voucher utilization versus the displacement of other low-income renters.

The theoretical framework thus suggests that SOI laws should improve outcomes for voucher holders who successfully benefit from expanded access, while potentially generating negative spillovers for other market participants through rent increases. The empirical analysis that follows tests these predictions and quantifies the relative importance of these competing effects.

1.4 Data and Empirical Strategy

1.4.1 Policy Data

SOI policy information is assembled from the Urban Institute's State and Local Voucher Protection Laws database and the National Low Income Housing Coalition's Tenant Protection Database. Each statute or ordinance is matched to a census place via exact and fuzzy name matching verified against state and county identifiers, and time-stamped by the first effective date (passage dates are used only when coincident with effectiveness).

Within the estimation window, 57 census-designated places enter coverage between 2013 and 2018. Table 1.1 lists treated places, first effective year, and the adoption layer (state, county, municipal). Adoption-layer shares are summarized beneath the table. Definitions and sources for policy variables used below appear in Table 1.2.

1.4.2 Data Construction and Sample

Evidence from Cho and Lucio (2025) shows that SOI adoption may not be random: eventual adopters differ from never-adopters along observables tied to housing markets and local institutions, including higher incomes and rents, greater college attainment, and denser tenant-protection environments. These differences raise concerns about selection bias in a simple comparison of adopters versus never-adopters, since the same factors that drive SOI adoption may also influence housing outcomes independently of the policy itself.

To address this challenge, I assemble a comprehensive set of baseline place characteristics measured in 2012, before any place in the sample adopts SOI protections during the analysis window. The covariate set is designed to capture observable factors that theory and prior evidence suggest influence both the propensity to adopt SOI laws and the housing outcomes of interest. These include housing market fundamentals (median household income, median contract rent, rental vacancy rate), demographic composition (college-educated share, poverty rate, median age, Black population share, total population), economic conditions (unemployment rate, SNAP participation rate), the existing tenant protection environment (indicators for pre-2013 anti-retaliation protections, limits on fees, just-cause eviction standards, legal-defense funds, and right to counsel), voucher program presence (HCV units per 1,000 residents), and broader policy context (count of other state-level tenant policies active before 2013). Complete definitions and data sources for all variables appear in Table 1.2.

These covariates are then used to assess pre-treatment balance of observables. The analysis builds an annual panel for 2011–2019 at the census "place" level (incorporated cities and towns) using ACS one-year geographies. The sample frame includes all places with population at least

Table 1.1: Places Adopting Source of Income Protection Policies, 2013–2018

Place	First Year	Level	Place	First Year	Level
Berkeley, CA	2017	Place	Mount Vernon, NY	2013	County
Milpitas, CA	2017	County	New Rochelle, NY	2013	County
Mountain View, CA	2017	County	Rochester, NY	2017	Place
Oakland, CA	2018	Place	Syracuse, NY	2016	Place
Palo Alto, CA	2017	County	Yonkers, NY	2013	County
San Diego, CA	2018	Place	Beaverton, OR	2013	State
San Jose, CA	2017	County	Bend, OR	2013	State
Santa Clara, CA	2017	County	Eugene, OR	2013	State
Santa Monica, CA	2015	Place	Gresham, OR	2013	State
Sunnyvale, CA	2017	County	Hillsboro, OR	2013	State
Boulder, CO	2018	Place	Medford, OR	2013	State
Coral Springs, FL	2017	County	Portland, OR	2013	State
Davie, FL	2017	County	Salem, OR	2013	State
Deerfield Beach, FL	2017	County	Pittsburgh, PA	2015	Place
Fort Lauderdale, FL	2017	County	Dallas, TX	2016	Place
Hollywood, FL	2017	County	Bellingham, WA	2018	State
Lauderhill, FL	2017	County	Everett, WA	2018	State
Miramar, FL	2017	County	Kennewick, WA	2018	State
Pembroke Pines, FL	2017	County	Marysville, WA	2018	State
Plantation, FL	2017	County	Pasco, WA	2018	State
Pompano Beach, FL	2017	County	Spokane, WA	2017	Place
Sunrise, FL	2017	County	Spokane Valley, WA	2018	State
Tamarac, FL	2017	County	Tacoma, WA	2018	State
Weston, FL	2017	County	Vancouver, WA	2015	Place
Arlington Heights, IL	2013	County	Yakima, WA	2018	State
Cicero, IL	2013	County			
Evanston, IL	2013	County			
Palatine, IL	2013	County			
Schaumburg, IL	2013	County			
Iowa City, IA	2015	Place			
Wyoming, MI	2018	Place			
Cheektowaga, NY	2018	County			

Notes: This table lists all 57 places that adopted Source of Income Protection policies between 2013 and 2018. Sample includes all incorporated cities and towns with population $\geq$ 65,000 based on ACS 2019 estimates that enacted SOI legislation during the analysis window. First Year indicates the first year in which the policy was active at the place level. Level indicates the governmental level at which the policy was initially enacted (Place = local ordinance, County = county ordinance, State = state legislation).

65,000 on a fixed baseline (2012), yielding 521 places in 2012, of which 57 adopt a source-of-income (SOI) protection with first effective dates between 2013 and 2018, and 464 that never adopt

within the estimation window.[4] Places with SOI protections effective before 2011 are excluded to ensure a pre-policy baseline. Place boundaries follow the ACS one-year "place" geography (2010 vintage).

Table 1.3 documents the extent of pre-treatment imbalance by comparing means and standard deviations across eventual adopters and never-adopters in 2012. To quantify meaningful differences, I calculate standardized mean differences (SMD) defined as:

$$\text{SMD} = \frac{\bar{X}_{\text{treated}} - \bar{X}_{\text{control}}}{\sqrt{\frac{(n_T-1)s_T^2+(n_C-1)s_C^2}{n_T+n_C-2}}},$$

where $\bar{X}_{\cdot}$ and $s_{\cdot}$ are group means and standard deviations and n_T, n_C are the treated and control sample sizes. Balance is a property of the realized sample, so hypothesis tests and p-values are poor diagnostics because they fluctuate with sample size rather than substantive imbalance Ho et al. (2007). Standardized mean differences are scale-free, directly interpretable in SD units, and recommended as the primary check; a common rule of thumb treats $SMD \geq 0.15$ as large. Using SMDs focuses the diagnostic on imbalance itself, which is what drives bias and model dependence.

The results confirm substantial baseline differences between eventual adopters and never-adopters. Several variables exceed the 0.15 threshold, including state-level tenant-policy activity (SMD = 0.671), college-educated share (SMD = 0.429), median rent (SMD = 0.414), median age (SMD = 0.335), anti-retaliation protections (SMD = 0.268), and median household income (SMD = 0.188). These patterns indicate that places eventually adopting SOI protections were systematically different in 2012: they had higher rents and incomes, more college-educated populations, older median ages, greater prevalence of existing tenant protections, and operated in states with more active tenant-policy environments.

These baseline differences motivate the doubly robust estimation approach described in Section 2.5, which combines outcome regression with inverse probability weighting to condition on the full set of 2012 covariates. This adjustment helps ensure that the estimated treatment effects

[4] "Never-adopter within window" denotes places without an effective SOI date through 2019.

reflect the causal impact of SOI adoption rather than pre-existing differences between adopting and non-adopting places.

Table 1.2: Variable definitions and data sources

Variable	Definition	Source
Anti-retaliation ordinance (pre-2013)	Anti-retaliation protection in force before 2013 (1 = yes)	NLIHC
Limits on fees (pre-2013)	Caps or bans on application/late fees before 2013 (1 = yes)	NLIHC
Ability to expunge eviction records (pre-2013)	Tenants can seal or expunge eviction filings prior to 2013 (1 = yes)	NLIHC
Just-cause eviction standards (pre-2013)	Requirement that landlords cite "just cause" before 2013 (1 = yes)	NLIHC
Legal-defense fund (pre-2013)	Public fund covering tenant legal defense before 2013 (1 = yes)	NLIHC
Right to counsel (pre-2013)	Guaranteed right to counsel in eviction cases before 2013 (1 = yes)	NLIHC
HCV Units per 1,000 Total Population	Number of Housing Choice Voucher units per 1,000 residents	HUD PSH
Median contract rent (2019 $)	Median monthly contract rent	ACS 1-Year Surveys
Median household income (2019 $)	Median household income	ACS 1-Year Surveys
Rental vacancy rate (%)	Share of rental units that are vacant	ACS 1-Year Surveys
Unemployment rate (%)	Civilian unemployment rate	ACS 1-Year Surveys
College degree share (%)	Adults (25+) with bachelor's degree or higher	ACS 1-Year Surveys
SNAP participation rate (%)	Households receiving SNAP benefits	ACS 1-Year Surveys
Poverty rate (%)	Individuals below the federal poverty line	ACS 1-Year Surveys
Median age (years)	Median age of residents	ACS 1-Year Surveys
Black population share (%)	Residents identifying as Black or African American	ACS 1-Year Surveys
Total population	Total resident population	ACS 1-Year Surveys
Other state SOI laws	Count of SOI protections active in the same state before 2013	Policy databases

Note: SOI = Source of Income; HCV = Housing Choice Voucher; ACS = American Community Survey; HUD PSH = HUD Picture of Subsidized Households. Data is gathered to observe balance between treated and never-treated groups at baseline 2012 values. Policy variables denote the presence of a policy at any point before 2013, not just within the sample period. Other SOI Laws include SOI protections present at the city or county level in areas within the same state as the city of observation.

Table 1.3: Baseline characteristics stratified by treatment status, 2012 Values

	Stratified by treated			
Variable	**Overall**	**Never Treated**	**Treated**	**SMD**
Groups (unique census designated places)	521	464	57	–
Anti-retaliation protection, pre-2013 (mean, SD)	48.0 (50.0)	46.0 (50.0)	60.0 (49.0)	0.268(*)
Limits on fees, pre-2013 (mean, SD)	1.0 (10.0)	1.0 (9.0)	0.0 (0.0)	0.075
Just-cause eviction standards, pre-2013 (mean, SD)	3.0 (18.0)	3.0 (17.0)	5.0 (23.0)	0.112
Eviction legal-defense fund, pre-2013 (mean, SD)	2.0 (9.0)	2.0 (14.0)	2.0 (9.0)	< 0.001
Right to counsel, pre-2013 (mean, SD)	11.0 (67.0)	10.0 (56.0)	12.0 (45.0)	0.04
HCV Units per 1,000 Total Population (mean, SD)	9.11 (5.42)	8.97 (5.24)	9.21 (5.87)	0.077
Median rent, 2019 $ (mean, SD)	869.55 (287.94)	855.04 (279.10)	981.67 (330.71)	0.414(*)
Median household income, 2019 $ (mean, SD)	53,593.02 (18,495.00)	53,190.59 (18,430.24)	56,702.71 (18,869.29)	0.188(*)
Vacancy rate (%, mean, SD)	9.55 (5.44)	9.59 (5.43)	9.29 (5.59)	0.055
Unemployment rate (%, mean, SD)	10.25 (3.76)	10.26 (3.85)	10.18 (2.98)	0.021
College-educated population (%, mean, SD)	30.89 (13.84)	30.16 (13.29)	36.59 (16.57)	0.429(*)
SNAP participation rate (%, mean, SD)	13.68 (7.78)	13.56 (7.66)	14.62 (8.65)	0.129
Poverty rate (%, mean, SD)	17.62 (7.90)	17.74 (8.01)	16.75 (7.03)	0.130
Median age (mean, SD)	34.79 (4.35)	34.63 (4.36)	36.05 (4.13)	0.335(*)
Black population share (mean, SD)	0.14 (0.16)	0.15 (0.16)	0.13 (0.16)	0.109
Total population (mean, SD)	179,790.70 (261,944.95)	179,207.45 (262,280.87)	184,202.13 (261,744.19)	0.019
State-level tenant-policy count (mean, SD)	0.92 (1.47)	0.76 (1.10)	2.21 (2.85)	0.671(*)

Notes. This table compares means and standard deviations (in parentheses) of key covariates in 2012 for cities that never adopt an SOI law by 2018 and those that will adopt between 2013–2018. The column *SMD* reports the standardized mean difference. An SMD above 0.15 indicates a meaningful imbalance, marked by (*).

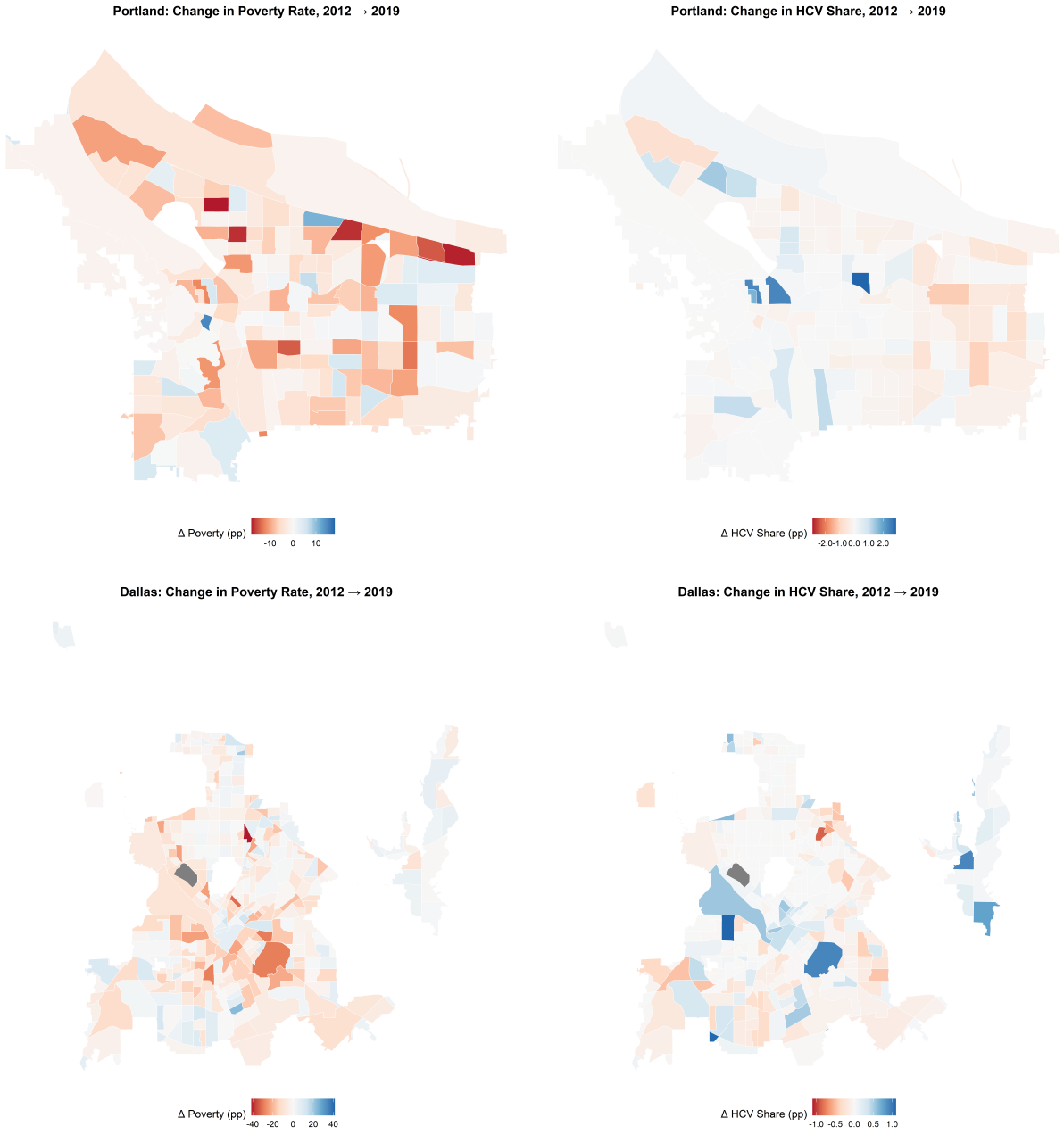
1.4.3 Outcome Variables

The outcome groups are chosen to align with the two channels through which SOI protections operate. For HCV mobility, seen in in Table 1.4, the goal is to track whether legal access translates into changes in where voucher households live within a place. Voucher household-weighted tract poverty exposure provides a standard measure of neighborhood economic conditions experienced by HCV households. The normalized Herfindahl index summarizes spatial concentration in a scale-free way that is robust to the number of tracts and the distribution of HCV presence across them, making it the preferred concentration metric when comparing places of different sizes. Two supplementary dispersion statistics—the share of tracts with any HCV presence and the average number of HCV households per occupied tract—help interpret movements in the primary indicators by distinguishing diffusion across new tracts from reallocation within already-active tracts.

Figure 1.3 visualizes tract-level changes between 2012 and 2019 in (i) poverty rate and (ii) the share of households using HCVs for two adopters (Portland, OR—SOI 2013; Dallas, TX—SOI 2016). Blues denote increases and reds denote decreases; color scales are symmetric across cities. These panels are descriptive only and serve to clarify the construction and interpretation of the HCV-weighted tract poverty outcome used below.

For market-wide rental conditions, seen in Table 1.5, the focus is on the price segment where vouchers transact and on an affordability quantity that should co-move with those prices. The twenty-fifth percentile of contract rent targets the lower tier of the market near payment-standard constraints, where small price adjustments are most likely to bind for voucher households. The seventy-fifth percentile provides a natural falsification outcome that should be largely insensitive if effects are concentrated in the voucher-relevant segment. The share of units affordable at or below 30 percent of HAMFI serves as an extensive-margin complement to the price measure, indicating whether shifts in low-tier rents translate into changes in the available stock of very low-rent units. Finally, the rental vacancy rate is reported to gauge short-run slack consistent with price movements, but it is treated as a mechanism rather than a primary target of policy effects.

Figure 1.3: Changes in Poverty Rate and HCV Household Share by Census Tract, 2012 → 2019 (Portland above; Dallas below).



Summary statistics for 2012 pre-treatment values of all outcome variables can be seen in Table 1.6.

Table 1.4: HCV Mobility Outcomes: Definitions and Sources

Variable	Definition	Source
Weighted HCV Household Tract Poverty Rate	Voucher-weighted tract poverty exposure: $\frac{\sum_i H_i \cdot \text{pov}_i}{\sum_i H_i}$, where H_i is the number of HCV households in tract i and pov_i is the tract poverty rate.	HUD Picture of Subsidized Households
Mean HCV Households per Occupied Census Tract	Average number of HCV households per occupied tract, H/n_{active}, where n_{active} is the count of tracts with ≥ 1 HCV household.	HUD Picture of Subsidized Households
Normalized Hirschman-Herfindahl Index	Normalized HHI of HCV spatial concentration: $\frac{\text{HHI} - 1/n}{1 - 1/n} \times 100$, with $\text{HHI} = \sum_i s_i^2$ and $s_i = H_i/H$. Ranges 0 (uniform) to 100 (complete concentration).	HUD Picture of Subsidized Households
Occupied Tract Share Percentage	Share of tracts with any HCV presence: $n_{\text{active}}/n_{\text{total}}$, where n_{total} is the total number of tracts in the place.	HUD Picture of Subsidized Households

Notes: Data comes from census tract level measures aggregated to the place level provided by the HUD Picture of Subsidized Households. Weighted HCV Household Tract Poverty is calculated and provided by HUD PSH. All other variables are the author's calculations. HCV = Housing Choice Vouchers. HHI - Herfindahl-Hirschman Index.

1.5 Empirical Strategy

1.5.1 Design and Estimator

The analysis follows the staggered difference-in-differences framework of Callaway and Sant'Anna (2021). Cohorts are defined by the first effective year of SOI coverage, $g \in \{2013, \ldots, 2018\}$, and in each calendar year t I estimate group-time average treatment effects on the treated relative to places that are never treated within 2011–2019. Not-yet-treated places are not used as controls after their effective dates. Standard errors are clustered at the place level.

Table 1.5: Rental Market Outcomes: Definitions and Sources

Variable	Definition	Source
Ln(25th Percentile Rent)	Log of the 25th percentile (P25) of the contract-rent distribution for renter-occupied and vacant-for-rent units at the place level; contract rent excludes utilities; CPI-U adjusted to 2019 dollars.	ACS 1-year
Ln(75th Percentile Rent)	Log of the 75th percentile (P75) of the contract-rent distribution; contract rent excludes utilities; CPI-U adjusted to 2019 dollars.	ACS 1-year
Rental Inventory Share Affordable at $\leq 30\%$ HAMFI	Share of total rental inventory (occupied + vacant-for-rent) with gross rent $\leq 30\%$ of HUD Area Median Family Income (HAMFI), using the 30% rent-to-income threshold.	HUD CHAS
Rental Vacancy Rate	Rental vacancy rate: vacant-for-rent units divided by total rental units.	ACS 1-year

Notes: HCV = Housing Choice Vouchers. HAMFI = HUD Area Median Family Income. Contract rent excludes utilities. CHAS data uses gross rent (contract rent + utilities) with HUD affordability thresholds based on 30% rent-to-income with HAMFI. Means and standard deviations calculated using 2012 values (pre-treatment for all units).

Table 1.6: Summary Statistics for Outcome Variables - 2012 values

Variable	N	Mean	SD	Min	Max
Tracts with any HCV (%)	521	80.26	20.42	4.00	100.00
HCV HH per occupied tract	521	41.28	25.81	1.00	218.44
HCV-weighted tract poverty (%)	521	20.50	8.96	4.00	60.00
Normalized HHI (0–100)	521	11.69	12.10	0.22	100.00
P25 rent ($2019)	521	698.11	288.46	295.00	2080.00
P75 rent ($2019)	521	1542.77	562.98	945.00	3350.00
Affordable share $\leq 30\%$ HAMFI (%)	521	9.28	5.09	0.51	46.66
Rental vacancy rate (%)	521	9.34	5.54	0.36	39.03

Notes: All measures are calculated as the means of both treated and untreated groups for 2012 values. HCV = Housing Choice Vouchers. HHI = Herfindahl Index. HAMFI = HUD Area Median Family Income. Contract rent excludes utilities. CHAS data uses gross rent (contract rent + utilities) with HUD affordability thresholds based on 30% rent-to-income with HAMFI. Means and standard deviations calculated using 2012 values (pre-treatment for all units).

Because eventual adopters and never-adopters differ on observables at baseline, I implement the doubly robust version of the estimator. Covariates are measured once, in 2012, as described in Section 3.3, and held fixed to avoid post-treatment conditioning. The estimator combines outcome regression with inverse-probability weighting so that consistency obtains if either the regression model or the propensity score is correctly specified. The weighting step improves overlap and covariate balance between treated cohorts and never-adopters, which is important given the baseline differences documented in the balance tests.

1.5.2 Heterogeneity

Mechanisms are examined in two prespecified breakouts that map policy design and market tightness into treatment intensity. First, a "strong-law" classification combines the enforcement and exemptions indicies provided by the Urban Institute, summarized in Table 1.7: places with an enforcement score of at least three and an exemptions score of at least 3 at adoption are labeled strong-law; all others are weak-law. I re-estimate the event study separately for strong- and weak-law adopters against the same pool of never-adopters, using the identical 2012 covariates.

Second, I stratify treated places by baseline tightness using the 2012 rental vacancy rate, defining low- and high-vacancy markets at the treated and non-treated sample median. Heterogeneity focuses on the outcomes that most directly test the two channels. For HCV mobility, I use the normalized HHI of voucher spatial concentration and the voucher household-weighted tract poverty exposure, which are the most robust measures of concentration and neighborhood quality relative to more mechanically sensitive dispersion statistics. For rental markets, I use the log of the twenty-fifth percentile of contract rent and the share of units affordable at or below 30 percent of HAMFI, where the former targets the low-rent segment relevant for voucher transactions and the latter provides a complementary extensive-margin check on affordability.

As shown in Table 1.8, the distribution across these categories is relatively balanced, with 18 places in the Strong-High quadrant, 19 in Strong-Low, 9 in Weak-High, and 11 in Weak-Low. This

variation provides sufficient power to test the hypothesis that SOI effects should be larger where laws are stronger and markets are tighter.

Table 1.7: SOI Protections Policy Strength Scoring Framework

Dimension	Component	Points	Description
Enforcement	Private Right of Action	+1	Allows private parties to file civil lawsuit directly in court
	Civil Damages	+1	Allows court to order monetary relief to winning plaintiff
	Attorney's Fees	+1	Allows court to order losing defendant to reimburse legal costs
	Criminal Penalties	+1	Allows court to impose fines or sentences in criminal cases
	Maximum	*4*	
Exemptions	Good Faith Business Decision	-1	Allows landlord to deny housing based on reasonable business judgment
	Minimum Income Requirement	-1	Allows landlord to require minimum income from non-voucher sources
	Owner Occupied/Property Size	-1	Exempts small properties or owner-occupied units
	Religious/Nonprofit Owner	-1	Exempts properties owned by religious or nonprofit organizations
	Maximum	*4*	

Notes: This table summarizes the scoring framework for SOI protection policy strength developed by the Urban Institute. The framework evaluates two key dimensions of policy design: enforcement mechanisms available and exemptions that weaken coverage. The scoring allows for systematic comparison of SOI protections policy strength across jurisdictions. Enforcement begins at a score of 0 and receives a point for each of the standardized components listed in the table. Exemptions begin at a score of 4 and lose a point for each of the standardized components that are present by place of adoption.

1.5.3 Robustness

Synthetic Staggered Difference in Differences

As a robustness check, I implement a synthetic staggered difference-in-differences (SDID) estimator. SDID blends the strengths of synthetic control and DiD by learning unit weights (across untreated donors) and time weights (across periods) that make the counterfactual path of treated units

Table 1.8: Places Adopting SOI Protection Policies by Law Strength and Vacancy Level, 2013–2018

Strong Laws, High Vacancy	First Year	Strong Laws, Low Vacancy	First Year
Berkeley, CA	2017	Milpitas, CA	2017
Oakland, CA	2018	Mountain View, CA	2017
Coral Springs, FL	2017	Palo Alto, CA	2017
Davie, FL	2017	Santa Monica, CA	2015
Deerfield Beach, FL	2017	Sunnyvale, CA	2017
Fort Lauderdale, FL	2017	Tamarac, FL	2017
Hollywood, FL	2017	Weston, FL	2017
Lauderhill, FL	2017	Arlington Heights, IL	2013
Pembroke Pines, FL	2017	Evanston, IL	2013
Plantation, FL	2017	Palatine, IL	2013
Pompano Beach, FL	2017	Schaumburg, IL	2013
Sunrise, FL	2017	Beaverton, OR	2013
Cicero, IL	2013	Bend, OR	2013
Syracuse, NY	2016	Eugene, OR	2013
Medford, OR	2013	Gresham, OR	2013
Pittsburgh, PA	2015	Hillsboro, OR	2013
Spokane, WA	2017	Portland, OR	2013
Tacoma, WA	2018	Salem, OR	2013
		Miramar, FL	2017
Weak Laws, High Vacancy	**First Year**	**Weak Laws, Low Vacancy**	**First Year**
San Diego, CA	2018	Boulder, CO	2018
Cheektowaga, NY	2018	Iowa City, IA	2015
Mount Vernon, NY	2013	Wyoming, MI	2018
New Rochelle, NY	2013	San Jose, CA	2017
Rochester, NY	2017	Santa Clara, CA	2017
Yonkers, NY	2013	Bellingham, WA	2018
Dallas, TX	2016	Everett, WA	2018
Kennewick, WA	2018	Marysville, WA	2018
Yakima, WA	2018	Pasco, WA	2018
		Spokane Valley, WA	2018
		Vancouver, WA	2015

Notes: This table cross-classifies the 57 places adopting SOI protections between 2013–2018 by law strength and baseline rental market tightness. Strong Laws have both enforcement and exemption scores ≥ 3 from the Urban Institute policy database; Weak Laws have at least one score ≤ 3. High Vacancy places had above-median rental vacancy rates in 2012 among treated and never-treated places; Low Vacancy places had below-median rates. The classification allows examination of heterogeneity by policy design features and market conditions that theory suggests should influence treatment intensity.

closely track their pre-treatment outcomes, then comparing appropriately weighted post-treatment means. This relaxes reliance on conventional parallel-trends and improves fit when unobserved

confounding evolves through latent unit×time factors, a setting where SDID has attractive robustness properties and good empirical performance. The method is first proposed from Arkhangelsky et al. (2021) and formalized by Porreca (2022).

In staggered-adoption panels, SDID is applied cohort-by-cohort (first treated in year g versus never-treated donors) and the cohort estimates are then aggregated, yielding a single ATT that preserves the staggered structure and avoids problematic comparisons to later-treated units. This cohort-wise synthetic construction addresses concerns about poor counterfactual matches or compositional issues in multi-period TWFE designs. When rich pre-policy covariates are available, a standard approach is to residualize outcomes on the vector of pre-treatment covariate values, then run SDID on the residual values. This absorbs systematic X-related differences while preserving SDID's outcome-based balancing (Porreca, 2022). In doing so, this approach mitigates the concerns of selection bias from observable characteristics as mentioned previously.

SDID supports several variance estimators. In practice, bootstrap, placebo, and jackknife are common, with coverage documented in simulations (Arkhangelsky et al., 2021; Porreca, 2022). I use placebo inference. For each adoption cohort g, fixing the pre/post split at g and the never-treated donor pool, I draw R placebo replications by (i) selecting a pseudo-treated subset of donors with the same cardinality as the actual cohort, (ii) assigning them placebo exposure $W_{it} = \mathbf{1}\{t \geq g\}$, (iii) re-estimating SDID on that placebo block (re-learning unit and time weights from pre-g outcomes), and (iv) recording $\tilde{\tau}_g^{(r)}$. The cohort-level standard error is the dispersion of the placebo distribution, $\widehat{\mathrm{se}}(\hat{\tau}_g) = \mathrm{sd}(\{\tilde{\tau}_g^{(r)}\}_{r=1}^{R})$, yielding Wald 95% CIs $\hat{\tau}_g \pm 1.96\,\widehat{\mathrm{se}}(\hat{\tau}_g)$; a randomization p-value is $\hat{p} = \frac{1+\sum_{r=1}^{R}\mathbf{1}\{|\tilde{\tau}_g^{(r)}|\geq|\hat{\tau}_g|\}}{R+1}$. This design respects cohort timing, matches the panel's dependence structure, and avoids contamination by restricting donors to never-treated units.[5]

Anticipation Effects

A potential threat to identification is that landlords and PHAs may learn about, prepare for, or partly comply with SOI protections before their first effective date that the policy is active for. In

[5]Implemented via `synthdid_se(method="placebo")`. R is set to $R = 400$ in the main results.

many jurisdictions, ordinances are passed months before they become enforceable, accompanied by public guidance, agency outreach, and media coverage. Owners may adjust screening or list prices in anticipation to avoid future frictions (inspection sequencing, rent-reasonableness negotiations, complaint risk). If such pre-treatment responses are material, dynamic effects estimated with the last pre-period year (e.g., $e = -1$) in the identifying set could understate true post-adoption changes or blur the timing of effects.

To assess this, I re-estimate the Callaway–Sant'Anna event studies, allowing an anticipation window of one year. Concretely, for each outcome I compute cohort- and time-specific doubly-robust ATT with the year immediately prior to first effectiveness dropped from the set of periods used to construct counterfactuals. To properly re-estimate the dynamic effects given baseline observable imbalances, I re-define all baseline covariates using 2011 mean values, which is the last period that is untreated and unanticipated for the earliest cohort (g=2013). I then aggregate to dynamic event time and cluster standard errors at the place level in the same manner as is done for the main results.

1.5.4 Identification and Validity

The causal interpretation of the results depends critically on the conditional parallel trends assumption: that treated and never-treated places would have followed similar outcome trajectories absent SOI adoption, conditional on the 2012 baseline covariates. This assumption is non-trivial given the substantial pre-treatment differences documented in Table 1.3, where several covariates exhibit standardized mean differences exceeding 0.4, indicating that eventual adopters were systematically different places with higher incomes, rents, education levels, and more active tenant-protection environments.

The doubly robust estimator addresses this challenge by combining outcome regression with inverse probability weighting to condition on the full vector of 2012 baseline characteristics. This approach provides protection against model misspecification: consistent estimates obtain if either the outcome regression or the propensity score model is correctly specified. The rich covariate set is

designed to capture observable factors that theory and prior evidence suggest influence both adoption propensity and housing market outcomes. While this adjustment cannot eliminate bias from unobserved time-varying confounders, the event-study profiles provide evidence on pre-treatment trends. Joint tests of lead coefficients assess whether differential trends existed before adoption; flat pre-period profiles support the conditional parallel trends assumption.

Several design features strengthen the identification strategy. The focus on incorporated places eliminates partial treatment concerns that would arise from using larger geographic units like metropolitan areas or counties, where SOI laws might cover only portions of the boundary. By matching policy coverage precisely to outcome measurement boundaries, the analysis captures the full intensity of treatment exposure. The staggered timing across 57 places and six adoption years (2013-2018) provides multiple quasi-natural experiment settings, reducing dependence on any single cohort or time period. The number of contributing treated place-year units across event times can be seen in Figure A.4.

The robustness check using SDID provides some reassurance by employing a different identification strategy that constructs counterfactuals through outcome-based matching rather than conditioning on pre-specified covariates. The consistency of results across estimators strengthens confidence in the findings, though both approaches share core identifying assumptions about the comparability of treated and control units. The current results should be interpreted as the causal effects of SOI adoption under the assumption that the rich baseline controls successfully account for systematic differences between adopting and non-adopting places.

1.6 Results

1.6.1 Overall ATTs

The pooled treatment effects reveal a clear pattern consistent with the event-study findings. Across all four HCV mobility measures, estimates are small and statistically indistinguishable from zero—weighted tract poverty (0.20 p.p.; 95% CI: 0.97, 0.57), normalized HHI (0.53 points; 95% CI: 1.12, 2.18), share of tracts with any HCV presence (0.23 p.p.; 95% CI: 4.92, 4.45), and

Table 1.9: Summary of Average Treatment Effects on the Treated (ATT)

Outcome	ATT	Std. Error	95% Conf. Int.
HCV Mobility Outcomes			
Occupied Tract Share (%)	-0.23	3.92	[-4.92, 4.45]
HCV HH per Occupied Tract	0.27	1.60	[-2.87, 3.41]
Normalized HHI (0-100)	0.53	0.84	[-1.12, 2.18]
HCV-Weighted Tract Poverty (%)	-0.20	0.39	[-0.97, 0.57]
Rental Market Outcomes			
Ln(25th Percentile Rent)	0.050**	0.014	[0.026, 0.078]
Affordable Share $\leq$ 30% HAMFI (%)	-1.70**	0.32	[-2.20, -0.87]
Ln(75th Percentile Rent)	0.010	0.018	[-0.026, 0.046]
Rental Vacancy Rate (%)	-0.60	0.42	[-1.43, 0.23]

Notes: This table reports overall average treatment effects on the treated (ATT) from the Callaway and Sant'Anna (2021) doubly robust estimator with event-study aggregation. Effects are estimated relative to never-adopters, conditioning on 2012 baseline covariates. Standard errors are clustered at the place level. ** indicates statistical significance at the 5% level (confidence interval does not include zero). HCV = Housing Choice Vouchers; HHI = Herfindahl-Hirschman Index; HAMFI = HUD Area Median Family Income.

HCV households per occupied tract (0.27; 95% CI: 2.87, 3.41)—indicating no detectable within-place change in voucher geography after SOI adoption. By contrast, rental market outcomes show effects concentrated at the low end: the 25th-percentile contract rent rises by about 5.0% (0.050 log points; 95% CI: 0.026, 0.078), and the share of units affordable at or below 30% of HAMFI falls by 1.7 p.p. (95% CI: 2.20, 0.87). Upper-tier rents remain unchanged (75th percentile: 0.010 log points; 95% CI: 0.026, 0.046), and the rental vacancy rate is negative but imprecise (0.60 p.p.; 95% CI: 1.43, 0.23). Taken together, SOI protections appear to generate price pressure at the low end of the rent distribution without materially altering the spatial distribution of voucher households within places.

The following figures denote the dynamic effects relative to time in years before and after treatment occurs. Across figures, two overarching features bear emphasis. First, for the majority of the dynamic effect estimates, the lead coefficients in all panels are centered near zero, providing graphical support for parallel pre-trends. Second, the post-adoption dynamics for rents and affordability are gradual rather than instantaneous, with effects building over one to several years and

showing little sign of reversal within the observed horizon, whereas the mobility outcomes remain flat throughout.

1.6.2 HCV Mobility Results

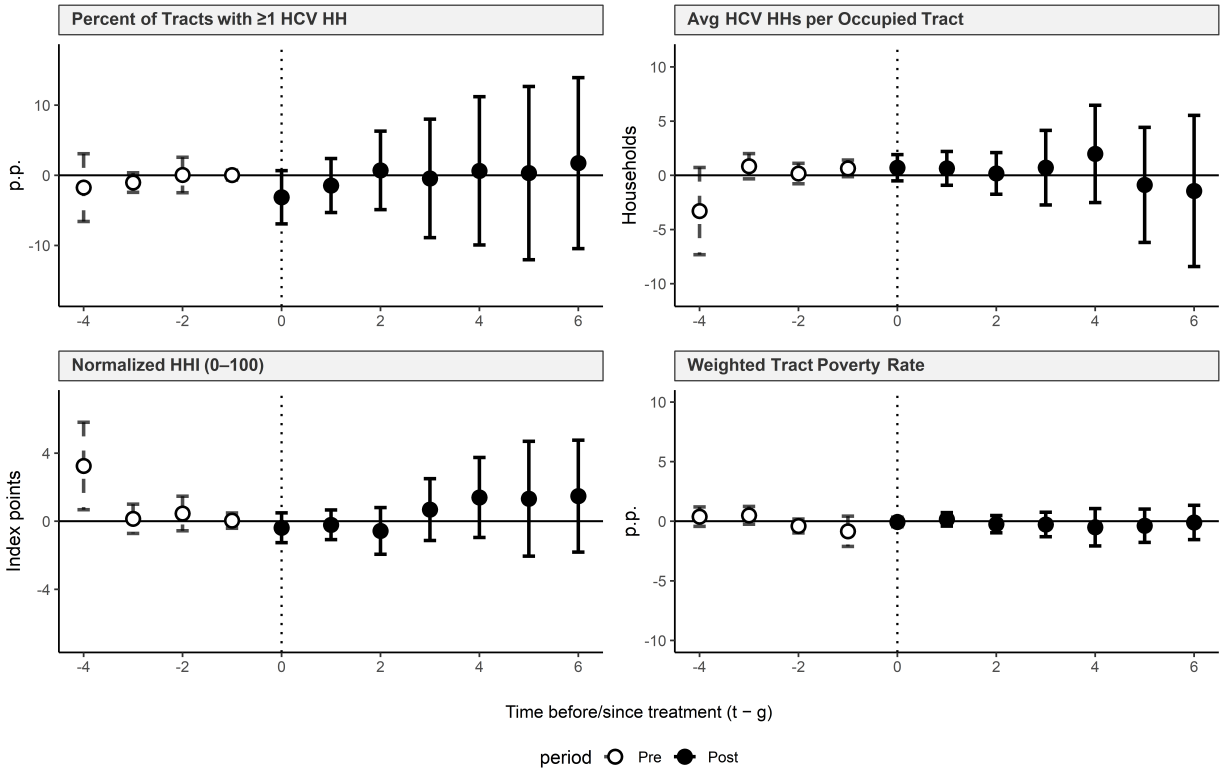
The HCV mobility event studies in Figure 1.4 show flat pre-trends and no discernible post-adoption movement across all four measures. Voucher-weighted tract poverty remains centered near zero (pooled ATT: 0.20 p.p.; 95% CI: 0.97, 0.57), indicating no detectable shift toward lower-poverty neighborhoods following SOI adoption. The normalized HHI exhibits the same pattern (0.53 index points; 95% CI: 1.12, 2.18), with no systematic rise or decline in concentration. The two dispersion statistics are noisier post-adoption but likewise show no persistent changes—share of tracts with any HCV presence: 0.23 p.p. (95% CI: 4.92, 4.45); HCV households per occupied tract: 0.27 (95% CI: 2.87, 3.41). Together, these profiles suggest that, within places, legal protections did not translate into measurable re-sorting of voucher households across tracts over the study horizon, and the null mobility findings are not an artifact of a single metric.

The stratified event studies for voucher neighborhood quality (Figure 1.5) and voucher concentration (Figure 1.6) mirror the pooled mobility results. By the strong- versus weak-law split, post-adoption effects for tract poverty and for the normalized HHI are statistically indistinguishable from zero in every event year (all 95% CIs span zero), with estimates hovering near zero before and after adoption. Splitting on baseline tightness yields visually opposite drifts in tract poverty—mildly downward in low-vacancy places and upward late in high-vacancy places—but the low-vacancy series shows non-flat leads and both series have wide post-period intervals (again, 95% CIs include zero throughout), so these patterns are not interpreted as causal. Taken together, the heterogeneous cuts provide no credible evidence of within-place re-sorting by voucher households; any vacancy-related divergence is suggestive at most and lacks robust pre-trend support.

1.6.3 Rental Market Results

Figure 1.7 turns to rental conditions and reveals a sharp split between the lower and upper tiers of the market. The log 25th-percentile contract rent drifts upward after adoption and remains

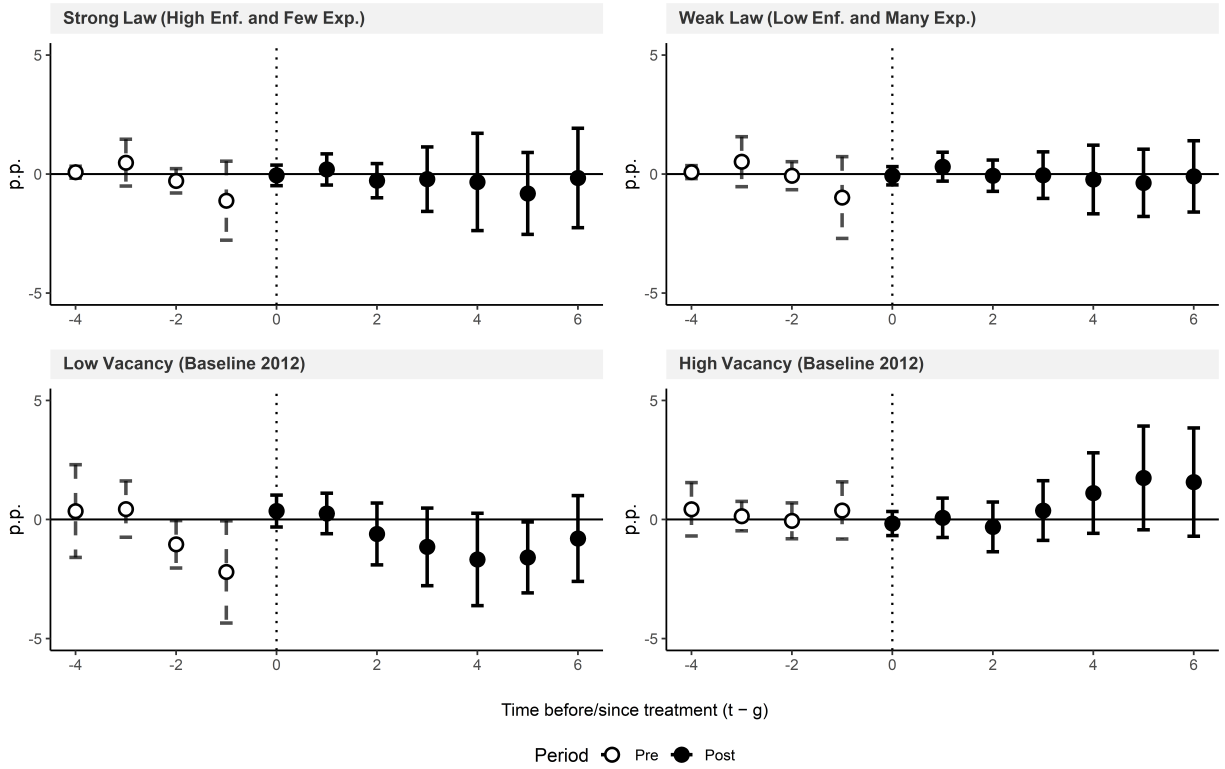
Figure 1.4: Dynamic Effect Estimates for HCV Household Mobility Outcomes



Notes: Figure displays 4 event study graphs of the dynamic effect of SOI Protection laws on HCV mobility related outcomes from Outcome Table A. Effects are estimated from Callaway & Sant'Anna (2021) doubly-robust approach. The counterfactual group consists of all census designated places that never passed an SOI protections law in the 2013 to 2018 time frame. The results account for conditional parallel trends pre-treatment 2012 baseline values of median rent, percent of population with a bachelor's degree, median age, median household income, the presence of anti-retaliation laws, the presence of limit fees laws, and the presence of SOI protection laws in other cities within the same state via the doubly-robust inverse probability weighting and outcome regression process. The x-axis represents time in years before or after treatment has occurred. "p.p" - Percentage points, "HHI" - Herschman-Herfindahl Index.

elevated (pooled ATT: 0.050 log points, about 5%; 95% CI: 0.026, 0.078). By contrast, the log 75th percentile is essentially flat (0.010 log points; 95% CI: 0.026, 0.046), a falsification consistent with effects concentrated in voucher-relevant segments. The affordability quantity moves as expected: the share of units with gross rent 30% of HAMFI declines (1.70 p.p.; 95% CI: 2.20, 0.87), mirroring the low-end price rise. The vacancy rate shows no clear contemporaneous shift (0.60 p.p.; 95% CI: 1.43, 0.23). Together, these patterns indicate price pressure concentrated where vouchers transact alongside a measurable contraction in the stock of very low-rent units.

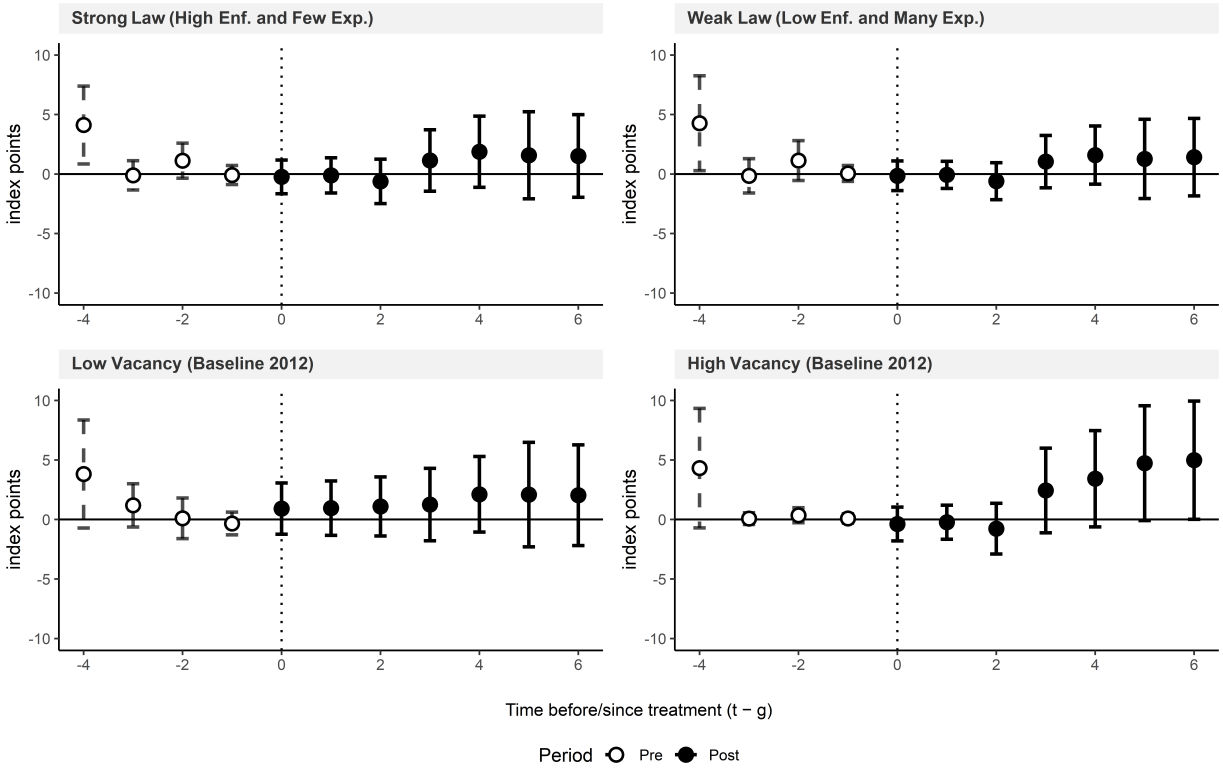
Figure 1.5: Dynamic Effect Estimates for Weighted Tract Poverty Rate by Stratification



Notes: Figure displays 4 event study graphs of the dynamic effect of SOI Protection laws on the weighted average poverty rate experienced by HCV Households by census tract across certain stratified groups. Effects are estimated from Callaway & Sant'Anna (2021) doubly-robust approach. The counterfactual group consists of all census designated places that never passed an SOI protections law in the 2013 to 2018 time frame. Strong Law refers to the policy's enforcement and exemption score both being greater than or equal to 3 for a given city. Low vacancy refers to the city having a lower than median rental vacancy rate in 2012 across the entire sample of treated and non-treated cities. The results account for conditional parallel trends pre-treatment 2012 baseline values of median rent, percent of population with a bachelor's degree, median age, median household income, the presence of anti-retaliation laws, the presence of limit fees laws, and the presence of SOI protection laws in other cities within the same state via the doubly-robust inverse probability weighting and outcome regression process. The x-axis represents time in years before or after treatment has occurred.

In contrast to the HCV mobility results, the stratified rental outcomes (Figures 1.8 and 1.9) show clear heterogeneity consistent with stronger binding where laws are tighter and markets are less slack. The 25th-percentile contract rent rises more in strong-law places than in weak-law places (0.067 log points; 95% CI: 0.036, 0.112; weak-law: 0.010 log points; 95% CI [-0.022, 0.048]), with the gap opening after adoption and persisting. A similar amplification appears by baseline tightness (low-vacancy: 0.075 log points; 95% CI: 0.038, 0.142; high-vacancy: -0.014 log points; 95% CI: -0.021, 0.008), with post-treatment estimates larger and more sustained in tighter

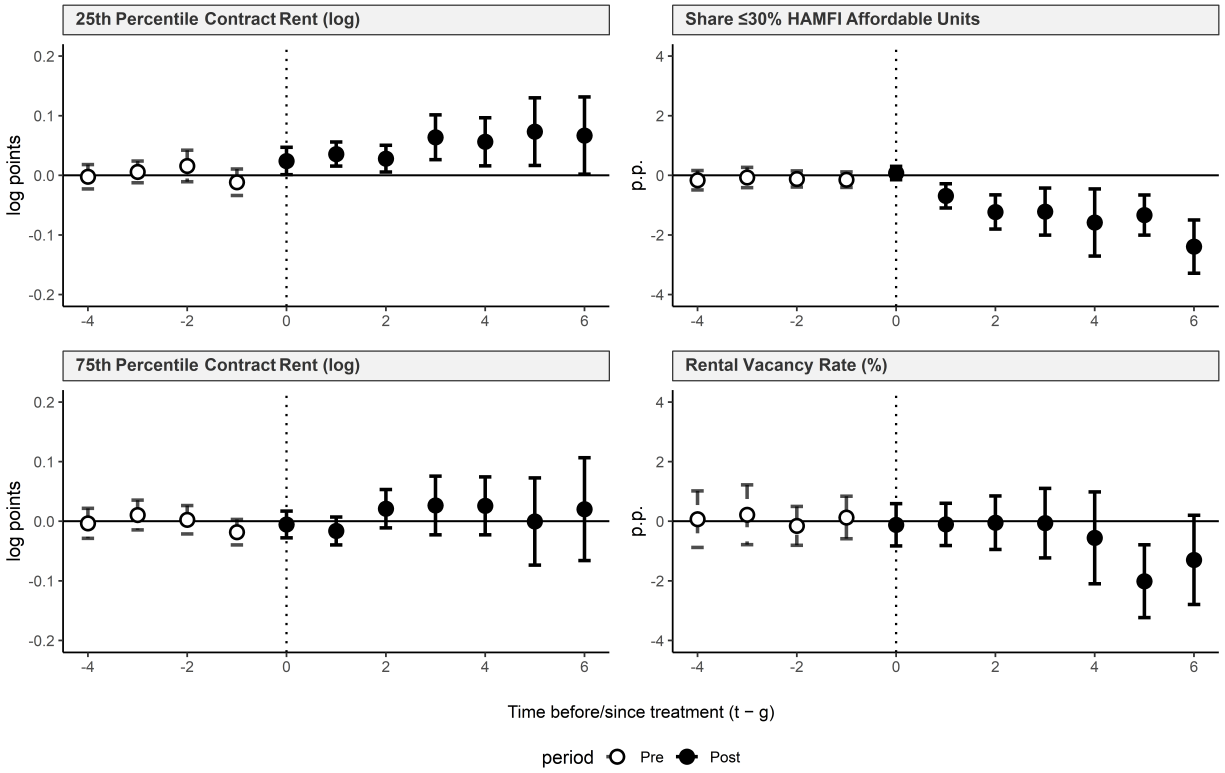
Figure 1.6: Dynamic Effect Estimates for Normalized HHI by Stratification



Notes: Figure displays 4 event study graphs of the dynamic effect of SOI Protection laws on the normalized HHI of HCV Households by occupied census tracts across certain stratified groups. Effects are estimated from Callaway & Sant'Anna (2021) doubly-robust approach. The counterfactual group consists of all census designated places that never passed an SOI protections law in the 2013 to 2018 time frame. Strong Law refers to the policy's enforcement and exemption score both being greater than or equal to 3 for a given city. Low vacancy refers to the city having a lower than median rental vacancy rate in 2012 across the entire sample of treated and non-treated cities. The results account for conditional parallel trends pre-treatment 2012 baseline values of median rent, percent of population with a bachelor's degree, median age, median household income, the presence of anti-retaliation laws, the presence of limit fees laws, and the presence of SOI protection laws in other cities within the same state via the doubly-robust inverse probability weighting and outcome regression process. The x-axis represents time in years before or after treatment has occurred. "HHI" - Herschman-Herfindahl Index.

markets. The affordability share moves in the opposite direction, falling more in strong-law and low-vacancy places (strong-law: -2.01 percentage points; 95% CI [-2.67, -0.97]; weak-law: 0.27 p.p.; 95% CI [-0.14, 0.31]; low-vacancy: -2.18 p.p.; 95% CI [-3.21, -1.16]; high-vacancy: 0.13 p.p.; 95% CI [-2.25, 2.45]). Together, these heterogeneous profiles align with the interpretation that stronger, broader-coverage statutes and tighter baseline conditions generate more pronounced price responses at the low end of the market while leaving voucher spatial distributions unchanged.

Figure 1.7: Dynamic Effect Estimates for Rental Affordability Outcomes



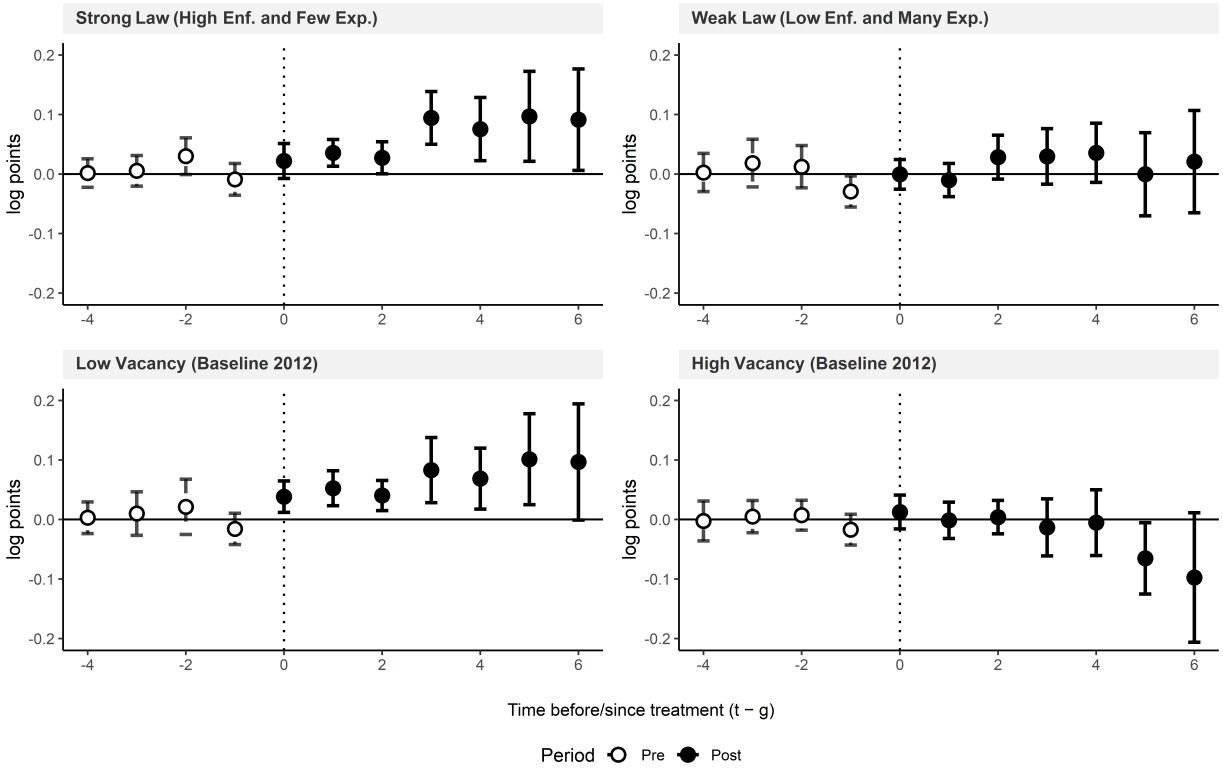
Notes: Figure displays 4 event study graphs of the dynamic effect of SOI Protection laws on Rental Affordability outcomes from Outcome Table B. Effects are estimated from Callaway & Sant'Anna (2021) doubly-robust approach. The counterfactual group consists of all census designated places that never passed an SOI protections law in the 2013 to 2018 time frame. The results account for conditional parallel trends pre-treatment 2012 baseline values of median rent, percent of population with a bachelor's degree, median age, median household income, the presence of anti-retaliation laws, the presence of limit fees laws, and the presence of SOI protection laws in other cities within the same state via the doubly-robust inverse probability weighting and outcome regression process. The x-axis represents time in years before or after treatment has occurred. "p.p" - Percentage points.

1.6.4 Robustness Results

Synthetic DiD Results

The synthetic staggered difference-in-differences results provide strong confirmation of the main findings. As with the Callaway and Sant'Anna (2021) estimator, all four HCV mobility outcomes remain statistically indistinguishable from zero, with point estimates of similar magnitude and direction. The rental market effects are likewise confirmed and, if anything, slightly amplified under the SDID approach. The 25th percentile rent increase is 6.0 percent (versus 5.0 percent in the main specification), while the decline in affordable units is 2.17 percentage points (versus 1.70 per-

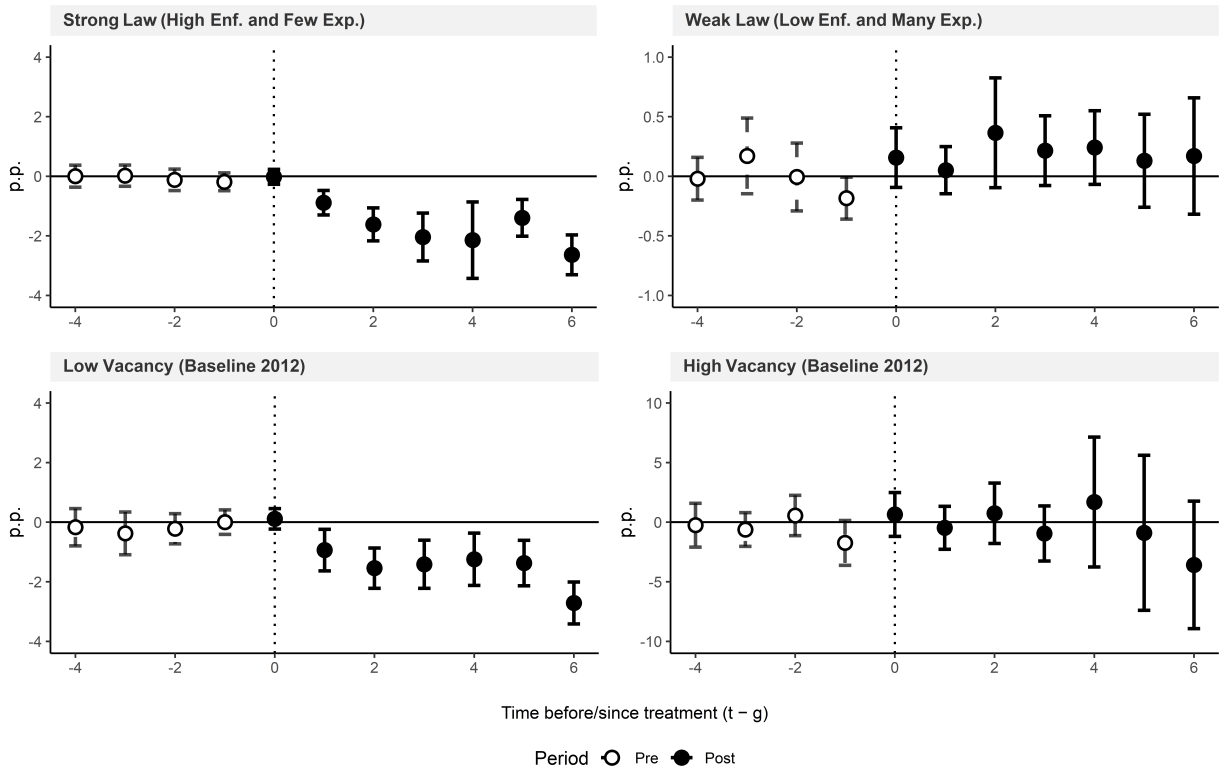
Figure 1.8: Dynamic Effect Estimates for 25th Percentile Rents by Stratification



Notes: Figure displays 4 event study graphs of the dynamic effect of SOI Protection laws on the 25th Percentile contract rent across certain stratified groups. Effects are estimated from Callaway & Sant'Anna (2021) doubly-robust approach. The counterfactual group consists of all census designated places that never passed an SOI protections law in the 2013 to 2018 time frame. Strong Law refers to the policy's enforcement and exemption score both being greater than or equal to 3 for a given city. Low vacancy refers to the city having a lower than median rental vacancy rate in 2012 across the entire sample of treated and non-treated cities. The results account for conditional parallel trends pre-treatment 2012 baseline values of median rent, percent of population with a bachelor's degree, median age, median household income, the presence of anti-retaliation laws, the presence of limit fees laws, and the presence of SOI protection laws in other cities within the same state via the doubly-robust inverse probability weighting and outcome regression process. The x-axis represents time in years before or after treatment has occurred.

centage points). The 75th percentile rent remains unchanged, preserving the falsification logic, and the vacancy rate continues to show no significant response. The consistency between SDID and the doubly robust Callaway and Sant'Anna (2021) results strengthens confidence that the findings are not artifacts of the particular identifying assumptions or estimation approach, but rather reflect robust treatment effects that emerge under alternative methods for constructing counterfactuals in staggered adoption settings.

Figure 1.9: Dynamic Effect Estimates for the Share of Rental Apartments Affordable at the $\leq 30\%$ HAMFI Threshold by Stratification



Notes: Figure displays 4 event study graphs of the dynamic effect of SOI Protection laws on the Share of Rental Apartments Affordable at the $\leq 30\%$ HAMFI Threshold across certain stratified groups. Effects are estimated from Callaway & Sant'Anna (2021) doubly-robust approach. The counterfactual group consists of all census designated places that never passed an SOI protections law in the 2013 to 2018 time frame. Strong Law refers to the policy's enforcement and exemption score both being greater than or equal to 3 for a given city. Low vacancy refers to the city having a lower than median rental vacancy rate in 2012 across the entire sample of treated and non-treated cities. The results account for conditional parallel trends pre-treatment 2012 baseline values of median rent, percent of population with a bachelor's degree, median age, median household income, the presence of anti-retaliation laws, the presence of limit fees laws, and the presence of SOI protection laws in other cities within the same state via the doubly-robust inverse probability weighting and outcome regression process. The x-axis represents time in years before or after treatment has occurred. "p.p." - Percentage points.

Anticipation Effect Results

Figures A.2 and A.3 show that the anticipation trim leaves the results materially unchanged. For the four HCV mobility outcomes, the pre-treatment leads at $e \leq -2$ remain centered near zero and the post-adoption paths stay flat, indicating no detectable re-sorting of voucher households within places even when the final pre-year is excluded. For rental market outcomes, the core pattern persists: the log 25th-percentile rent continues to drift upward over several post years, the

Table 1.10: Robustness Check: Synthetic Staggered Difference-in-Differences Results

Outcome	SDID ATT	Std. Error	95% Conf. Int.
HCV Mobility Outcomes			
Occupied Tract Share (%)	-1.04	3.56	[-5.94, 4.85]
HCV HH per Occupied Tract	0.93	2.27	[-3.51, 5.37]
Normalized HHI (0-100)	1.83	3.71	[-5.45, 9.10]
HCV-Weighted Tract Poverty (%)	-0.09	0.66	[-1.38, 1.19]
Rental Market Outcomes			
Ln(25th Percentile Rent)	0.060**	0.025	[0.026, 0.081]
Affordable Share $\leq$ 30% HAMFI (%)	-2.17**	0.44	[-2.99, -0.96]
Ln(75th Percentile Rent)	0.015	0.019	[-0.031, 0.045]
Rental Vacancy Rate (%)	-0.53	0.30	[-1.31, 0.38]

Notes: This table reports average treatment effects on the treated (ATT) from synthetic staggered difference-in-differences (SDID) estimation following Arkhangelsky et al. (2021) and Porreca (2022). SDID is applied cohort-by-cohort to avoid problematic comparisons with later-treated units, using never-adopters as donors. Outcomes are residualized on 2012 baseline covariates before SDID estimation. Standard errors are computed via placebo inference with 400 replications. ** indicates statistical significance at the 5% level (confidence interval does not include zero). HCV = Housing Choice Vouchers; HHI = Herfindahl-Hirschman Index; HAMFI = HUD Area Median Family Income.

share of units affordable at or below 30% of HAMFI declines, the log 75th-percentile rent remains near zero, and vacancy shows no clear contemporaneous break. The magnitudes and precision of the post-period estimates closely track the baseline specification, with any visual shift in onset reflecting the mechanical one-year relabeling of post-treatment time when the anticipation year-period is set to 1.

In sum, allowing for a one-year anticipation window neither reveals hidden pre-trends nor alters the substantive conclusions: SOI protections do not measurably change the within-place geography of voucher residence, while low-tier rents rise and the very-low-rent stock contracts. The robustness of these patterns to excluding the last pre-policy year suggests that pre-effectiveness adjustments, if present, are small in city-level aggregates or occur too close to the effective date to affect the identified dynamics.

1.7 Discussion

The results point to a consistent pattern. There is no detectable change in where voucher households live within places after SOI adoption, while prices at the low end of the rent distribution rise and the stock of very low–rent units shrinks. Effects are larger when laws are stricter and markets tighter. This section develops an interpretation of those dynamics, situates the findings in the existing evidence, and notes limitations and next steps.

1.7.1 Why rents rise after SOI protections even when HCV geography does not shift

SOI protections make categorical refusal costlier but leave ample scope for adjustment on other margins. Landlords that previously excluded vouchers can comply with the letter of the law while altering screening intensity, search and leasing practices, and asking rents. If voucher tenancy is perceived as higher cost (because of administrative steps, inspection sequencing, rent reasonableness determinations, delayed move-ins, or expectations about turnover and enforcement risk) owners can respond by raising prices where voucher demand is most likely to arrive. That response is most feasible in the part of the market close to payment standards and rent-reasonableness thresholds. In those segments, small price changes can both offset perceived costs and indirectly ration out some voucher applicants without violating the prohibition on categorical refusal.

The event studies line up with that logic. The twenty-fifth percentile of contract rent increases after adoption, while the seventy-fifth percentile is essentially flat. The affordability share moves inversely, consistent with a contraction in the stock of units priced at or below 30 percent of HAMFI. The absence of a clear vacancy response in the same window suggests adjustment is occurring primarily on the price margin rather than through short-run quantity changes.

Heterogeneity strengthens the inference in two complementary ways. Where laws are stronger, combining high enforcement and few exemptions, the policy more tightly constrains categorical refusal and raises the expected cost of noncompliance. If owners perceive a “compliance tax” (administrative time, inspection coordination, rent-reasonableness negotiations, legal risk), they can

restore target returns by marking up posted rents in the segment where voucher demand arrives and where payment standards and rent-reasonableness determinations provide headroom. Where baseline vacancy is low, short-run supply is inelastic and landlords face thick queues; the same compliance cost can be passed through more readily because the outside option of waiting for another non-voucher applicant is strong. In both dimensions, the event studies behave as the bindingness story predicts: stronger laws and tighter markets show larger increases at the twenty-fifth percentile and bigger declines in the very-low-rent stock, while upper-tier rents remain flat.

This pattern matches equilibrium responses in related protection regimes: models in which added frictions raise landlord costs predict price pass-through in tight segments (Abramson, 2024); reduced-form evidence links stronger tenant rights to higher rents and lower availability at the margin (Coulson et al., 2025); and experimental restrictions on screening have elicited substitution toward other exclusionary tactics rather than neutral compliance (Gorzig and Rho, 2025). The SOI setting fits that playbook: constrain one margin (explicit refusal) and adjustment shows up on margins still available (pricing and screening intensity) in the strata where those adjustments are most feasible.

This mechanism is also compatible with continued frictions early in the application funnel. Complaint-driven enforcement creates low, uncertain penalties ex ante, and discouragement or non-response is difficult to prove, so landlords can comply formally while shaping the pool of applicants who reach lease signing (Cunningham et al., 2018; Unlock NYC et al., 2022; Varady et al., 2017). Several levers remain legal and potent. Owners can quote asking rents just above payment standards or contest rent reasonableness, forcing vouchers to the sidelines without invoking "no vouchers." Minimum-income rules that ignore the subsidy, credit or eviction history screens, and document hurdles raise the cost of applying relative to non-voucher households and steer search back toward the same set of predictable, voucher-accepting tracts. On the supply side, inspection sequencing and Housing Assistance Payment contract timing impose real carry costs; when vacancy is scarce, waiting for a sure non-voucher tenant is a credible alternative. These frictions keep the composition of applicants who survive to the lease stage looking similar across tracts,

even if the legal choice set expands on paper. As a result, the within-place geography of voucher residence need not change, while the low-rent segment registers price adjustments where vouchers transact. In short, SOI protections appear to raise the effective cost of outright exclusion; landlords respond on margins they still control, and those adjustments show up where the program actually meets the market.

1.7.2 Why the HCV mobility results differ from past studies

Several features of the setting and design help explain the divergence from earlier work that finds modest improvements in voucher destinations following SOI adoption (Freeman and Li, 2014; Ellen et al., 2022; Teles and Su, 2022). First, the estimand differs. The analysis is at the place level and follows the stock of voucher households residing in a city each year. Studies that track movers' origin-destination pairs or analyze individual lease-ups are more sensitive to small composition shifts among recent movers. A change at the margin for movers can be washed out when averaged with the large stock of non-movers living in the same place, especially over relatively short horizons.

Second, the outcome set emphasizes robust, scale-free measures of within-place geography: voucher-weighted tract poverty and a normalized HHI of concentration. These measures are less sensitive to mechanical changes in the count of active tracts or to the distribution of small voucher flows across already-active tracts than simple dispersion statistics. The event studies and pooled estimates show those primary measures hovering near zero. Auxiliary dispersion measures tell the same story but are not the focus of inference.

Third, the period and sample differ. The study window covers first-effective dates from 2013 to 2018 in incorporated places with populations above 65,000. Many jurisdictions adopted SOI protections in already tight markets with binding payment standards. Where payment standards lag rents, a legal prohibition on categorical refusal does not expand the set of financially feasible units. Earlier studies detect improvements three to five years after adoption (Ellen et al., 2022; Teles and Su, 2022); effects of that magnitude may require longer horizons, complementary reforms

in inspections and payment standards, or targeted landlord engagement to translate formal access into realized moves. The null within-place changes here are therefore consistent with a world in which some movers experience improved opportunities at the margin but the aggregate residential distribution in a city remains broadly stable over the observed horizon.

Finally, enforcement and exemptions matter for any mobility response. Testing work shows substantial noncompliance even in protected jurisdictions and the substitution of new screens when old ones are curtailed (Cunningham et al., 2018; Phillips, 2017; Unlock NYC et al., 2022; Gorzig and Rho, 2025). The heterogeneity design partitions on law strength and baseline tightness. The absence of mobility effects in those splits, together with noisier vacancy-based profiles that lack clean pre-trend support, points to implementation frictions rather than a strong mobility channel in this period.

A useful contrast is payment-standard reform (e.g., Small Area Fair Market Rents (SAFMRs), which operates on the pricing margin rather than the legal-access margin I study. Ellen et al. (2025) finds that SAFMR-based reforms shift successful lease-ups toward higher-rent, lower-poverty neighborhoods without lowering overall lease-up rates and with roughly offsetting program costs across low- and high-rent areas. Because broad SAFMR rollout largely post-dates my window and does not cover the cities driving these estimates, it is not a confound here. Instead it points to a complementary policy bundle: align payment standards with spatial rent gradients while maintaining SOI protections as a backstop against categorical refusal. If the goal is neighborhood improvement for movers, pricing levers appear more directly effective than SOI protections alone in the short run.

Additionally, the null mobility finding may partly reflect revealed preferences among voucher households. Even where SOI protections expand the legal choice set, many households may prefer to remain in familiar neighborhoods due to proximity to family, friends, established social networks, or valued local amenities. If a substantial share of voucher holders would not move even absent discrimination, then removing the legal barrier addresses only one component of observed concentration patterns. Distinguishing preference-based stability from constraint-based concentra-

tion remains an empirical challenge, but the results are consistent with both mechanisms operating simultaneously.

1.8 Conclusion

The findings align with and extend three strands of evidence. First, they complement correspondence and audit studies that document continued discrimination under SOI laws by showing how general-equilibrium adjustments can surface, even when formal refusal is illegal (Cunningham et al., 2018; Phillips, 2017; Unlock NYC et al., 2022). Second, they connect to recent work on tenant protections more broadly, which emphasizes that stronger protections can deliver intended benefits for covered renters while inducing offsetting responses in prices or availability for others (Abramson, 2024; Coulson et al., 2025; Gorzig and Rho, 2025). The concentrated movement of the rent distribution at the lower quartile, together with a decline in the very-low-rent stock, is consistent with those models and with qualitative accounts of landlord strategy (Rosen, 2020; Lucio and Cho, 2025). Third, the results qualify the optimistic interpretation of earlier SOI studies that focus on mover destinations or utilization rates (Freeman, 2012; Freeman and Li, 2014; Ellen et al., 2022). The evidence here suggests SOI protections are not, on their own, a sufficient lever for changing the within-place geography of voucher residence in the short to medium run; complementary tools that reduce leasing frictions and realign payment standards with market rents may be necessary to realize the mobility rationale (Ellen, 2020; Aliprantis et al., 2019; Varady et al., 2017).

Certain limitations deserve emphasis. Policy measurement challenges are inherent in any analysis of heterogeneous local laws. While the Urban Institute's enforcement and exemption scores provide systematic categorization, they inevitably compress complex legal frameworks into broad indices that may miss important nuances. For instance, the enforcement score treats all "private rights of action" equally, but actual deterrent effects likely vary substantially depending on damage caps, fee-shifting provisions, and local legal culture. Similarly, exemptions are coded as binary indicators rather than capturing the share of rental stock actually affected. The policy coding reflects

characteristics at adoption or the nearest available year and assumes these features remain constant, but enforcement capacity, complaint processes, and exemption interpretations may evolve substantially post-adoption as agencies gain experience and face budget pressures.

Additionally, outcome measurement constraints limit the precision of effect detection across multiple dimensions. ACS rent quantiles are constructed from published distribution tables rather than unit-level microdata, introducing discretization error that may attenuate estimated price effects. The CHAS affordability measure relies on gross rent thresholds that ignore within-place variation in utility costs, tenant-paid fees, and informal side payments that could shift the effective price distribution. Most fundamentally, HUD administrative data capture residential location but not the search process itself—applications submitted, units toured, discriminatory encounters, or strategic self-sorting by voucher holders who anticipate rejection. This measurement gap means the analysis cannot distinguish between scenarios where legal protections expand search but do not change ultimate residential patterns versus scenarios where search behavior itself remains unchanged.

Furthermore, temporal and mechanism limitations constrain understanding of both short-run dynamics and long-run equilibrium effects. The maximum post-adoption observation window is six years for the earliest adopters, which may be insufficient to capture full adjustment in housing markets where lease terms, development cycles, and neighborhood change operate on longer timescales. Additionally, the study period predates several important program changes, including expanded Small Area FMR usage and Emergency Housing Voucher deployment during COVID-19, that may interact with SOI protections in ways not captured here.

Several extensions would sharpen inference on mechanisms and welfare. One avenue is to bring earlier stages of the leasing funnel into view by linking PHA leasing logs, inspection timing, and application records to measure delays, denials, and landlord participation dynamically. Another is to combine audit or platform-based inquiry data with enforcement records to quantify how complaint-driven systems miss discouragement and non-response. Unit-level or building-

level rent panels would allow decomposition of the P25 movement into within-unit price changes versus compositional shifts in the unit mix transacting at the low end.

Interactions between SOI protections and contemporaneous program changes (payment-standard increases, small-area FMR adoption, landlord signing bonuses) could identify complementary policy bundles that deliver mobility without amplifying price pressure. Border designs that compare treated places to contiguous never-adopters would probe local spillovers and strengthen counterfactual credibility. Finally, distributional consequences for non-voucher low-income renters merit direct study, given the decline in the affordable stock and the concentration of estimated price effects at the bottom of the distribution.

Taken together, the evidence indicates that SOI protections, as implemented in this period, changed landlords' incentives in ways that show up in prices rather than in the residential geography of voucher households. That pattern is not a verdict against SOI protections; rather, it underscores that legal access is a necessary but insufficient condition for mobility gains. Translating protections into moves likely requires reducing administrative frictions, aligning payment standards with market rents, and expanding supply where voucher demand concentrates, so that the cost of compliance does not pass through to the prices faced by the same low-income renters the policy aims to help.

Chapter 2

Funding, Facilities, and the Face of Homelessness: Heterogeneous Impacts of Federal Grants on Sheltered and Unsheltered Counts

2.1 Introduction

[6]

Despite a 16 percent real-term increase in U.S. Department of Housing and Urban Development (HUD) homeless-assistance appropriations, from $2.1 billion in fiscal year (FY) 2015 to $2.5 billion in FY 2019, the annual Point-in-Time (PIT) counts remained stubbornly high (above 550,000 people) and unsheltered homelessness has persisted near 200,000 persons. (Figure 2.1; U.S. Department of Housing and Urban Development (2019a)). This disconnect suggests that simply scaling up federal grants does not necessarily translate into fewer people sleeping rough or a commensurate expansion of shelter capacity.

Empirical evidence on the effectiveness of federal homeless-assistance grants—for both the Continuum of Care (CoC) grant and the Emergency Solutions Grant (ESG)—remains mixed. On one hand, more generous funding is theorized to relax local capacity constraints by underwriting new emergency shelter beds, improving service quality, and financing transitional as well as permanent supportive housing, thereby drawing individuals off the streets into stable accommodations (Moulton, 2013; Popov, 2016; Lucas, 2017). On the other hand, several studies argue that additional dollars disproportionately benefit those already marginally housed—households doubling up or in precarious rental situations—without materially reducing unsheltered counts captured in street-based PIT surveys (Culhane et al., 2011; Tsai and Alarcón, 2022). Compounding these diver-

[6]This paper has been accepted for publication at the *Southern Economic Journal.*

gent findings, HUD's formula-based targeting of resources to jurisdictions with higher measured homelessness creates a mechanical endogeneity: areas with rising PIT counts receive larger grants, so naïve correlations between funding and outcomes risk conflating cause and effect (Popov, 2016).

This paper provides new causal estimates of how combined CoC and ESG funding shapes both reported homelessness and Housing Inventory Count (HIC) shelter-bed capacity across 370 CoCs in 2019. Building on Popov (2016) and Lucas (2017), we instrument each CoC's per-capita funding with its share of housing units built before 1940 in the Community Development Block Grant (CDBG) formula (Congressional Research Service, 2014). Because the pre-1940 housing share reflects long-run stock vintage rather than contemporaneous homelessness shocks, it shifts allocations exogenously, satisfying the exclusion restriction and isolating the impact of marginal federal dollars. Focusing on a single, cross-sectional snapshot sidesteps concerns that year-to-year changes in the instrument may themselves correlate with local demolition or gentrification. We show that a $10,000 increase in annual funding per 10,000 residents leads to roughly 1 more sheltered person per 10,000, driven primarily by expansions in emergency and transitional beds, and yet produces no detectable short-run decline in unsheltered counts despite significant increases to permanent supportive housing (PSH) bed space capacity. We further document that impacts vary sharply by geography (urban vs. rural/suburban), household type (individuals vs. families), and demographic subgroup (age, gender, race, and chronicity).

Our contributions are threefold. First, we deliver the most recent national, causal estimate of federal homeless-assistance grants on both PIT homelessness—total, sheltered, unsheltered, and demographic subgroups—and on HIC bed categories (emergency, seasonal, overflow, and permanent supportive housing). Second, we document rich heterogeneity: funding differences disproportionately expand shelter use among people of different ages, gender, and race, while producing little-to-no meaningful short-run reduction in unsheltered counts. Third, we introduce a transparent, population-weighted tract-to-CoC crosswalk and share our geoprocessing code, enabling replication and extension.

The remainder of the paper is organized as follows. Section 2.2 reviews related economic and policy studies on homelessness and funding efficacy. Section 2.3 outlines the institutional mechanics of the federal grant allocations and the conceptual pathways through which grants may affect homelessness. Section 3.3 describes our 2019 outcome measures, funding data, the instrument, and the tract-to-CoC crosswalk for covariate construction. Section 2.5 presents the cross-sectional IV specification and discusses key identification assumptions. Section 3.4 reports our main and subgroup results, along with associational "between-CoC" contrasts. Finally, Section 2.7 interprets the findings, discusses policy implications, and outlines directions for future research.

2.2 Literature Review

Research on homelessness has evolved from early enumeration and descriptive studies to rigorous evaluations of policy interventions. Initial efforts focused on counting and profiling the homeless population. Elias (1952) developed one of the first systematic nighttime counts, foundational for the PIT, while Bahr and Caplow (1968) compared life trajectories of homeless men versus housed peers. Later on, sociologists such as Lamb (1984), Tsemberis and Eisenberg (2000), Yates et al. (1988), and Jencks (1994) documented nationwide increases in homelessness, exploring mental illness, substance use, and family breakdown among the unhoused. Appelbaum et al. (1991) examined how rent-control policies were blamed for homelessness despite broader structural factors.

Economists entered the field in the early 1990s, linking homelessness to housing and labor markets. Elliott and Krivo (1991) and Honig and Filer (1993) showed how local housing supply constraints, welfare generosity, and labor conditions influenced homelessness incidence. O'Flaherty (1995)'s bid-rent framework modeled housing choice under budget constraints, tying rent, wages, and welfare benefits to homelessness risk. However, these studies often assumed accurate PIT and HIC data—an assumption challenged by O'Flaherty (2004) and Early (2004), who highlighted undercounts and misclassifications in administrative counts. The Homeless Emergency Assistance

and Rapid Transition to Housing Act of 2009 (HEARTH) subsequently standardized reporting protocols to improve data quality (U.S. Department of Housing and Urban Development, 2012).

With better data, attention turned to funding efficacy. Culhane et al. (2011) advocated a prevention-centered approach, emphasizing rapid re-housing over emergency shelter. Byrne et al. (2013, 2014) showed that local investment in PSH correlated with lower chronic homelessness, even controlling for economic factors. Moulton (2013) found that PSH and rent subsidies reduced chronic homelessness more effectively than general CoC awards. In a landmark effort, Popov (2016) addressed endogeneity in funding allocations by instrumenting with annual fluctuations in the pre-1940 housing share, isolating causal impacts on shelter use and homelessness counts. Lucas (2017) extended this by disaggregating outcomes by demographic group, revealing that family groups responded more to funding increases than individuals.

Geographic and behavioral factors also matter. Cragg and O'Flaherty (1999) found that improved shelter conditions could attract individuals and families relative to the shelter options before modifications. Corinth and Lucas (2018) documented how winter climate affects PIT counts, echoed by Tsai and Alarcón (2022); Meyer et al. (2023), who proposed enumeration improvements for unsheltered populations.

Our study builds on these contributions by exploiting across-CoC variation in the pre-1940 housing share and employing a cross-sectional 2SLS design that examines both PIT and HIC outcomes with detailed subpopulation breakdowns. By creating a granular boundary crosswalk (Ferrara et al., 2024), we address time-heterogeneity and endogeneity concerns, providing robust causal estimates of how federal grants affect service utilization, capacity expansion, and the composition of homelessness across U.S. CoCs.

2.3 Conceptual Framework

2.3.1 Funding Determination

Under the CDBG program, HUD annually apportions block-grant funds to eligible entitlement jurisdictions (cities and counties) based on a two-part index that captures relative levels of

community development need. Each entitlement jurisdiction is evaluated on five core indicators expressed as shares of national totals: population growth lag (since 1960), overcrowded housing units, individuals in poverty, a composite "growth-poverty" index, and the stock of pre–1940 housing (Congressional Research Service, 2014; Collinson, 2014; Miller and Richardson, 2024). These five shares feed into two weighted formulas:

$$FundingShare = k \max \left\{ \underbrace{0.25 PopulationShare + 0.5 PovertyShare + 0.25 OvercrowdedShare}_{\text{Formula A}}, \right. \tag{2.1}$$

$$\left. \underbrace{0.2 GrowthLagShare + 0.3 PovertyShare + 0.5 Pre1940HousingShare}_{\text{Formula B}} \right\} \tag{2.2}$$

Each variable is expressed as that jurisdiction's share of the national total. The entitlement share is then

$$EntitlementShare_j = k \times max[FormulaA_j, FormulaB_j] \tag{2.3}$$

where k normalizes the maximum index so that shares sum to one across all entitlements. Prior to FY2013, all inputs were drawn from the decennial Census; since FY2013, HUD has relied on rolling American Community Survey (ACS) 5-year survey table estimates for these shares (Congressional Research Service, 2014; Miller and Richardson, 2024; Popov, 2016).

Each formula produces an index value for a jurisdiction by multiplying its local share of each indicator by the specified weight and summing the results. HUD then takes the maximum of Formula A and Formula B for each jurisdiction, yielding the preliminary entitlement index. A normalization constant ensures that the sum of all entitlement jurisdictions' indices equals one, so that each jurisdiction's entitlement share is the higher value of the output from Formula A or

Formula B, where k rescales national indices to budget shares (Congressional Research Service, 2014; Miller and Richardson, 2024). Once Congress enacts the annual CDBG appropriation, HUD applies these entitlement shares to allocate block-grant dollars to each jurisdiction.

Once CDBG shares are determined for each entitlement jurisdiction, HUD channels those allocations into the two principal homeless-assistance programs: the CoC grant and the ESG. To convert entitlement shares into CoC grant and ESG amounts, HUD applies each jurisdiction's CDBG-derived share to the total CoC appropriation, awarding funds directly to each CoC as its prorated allocation (U.S. Department of Housing and Urban Development, 2024). In this way, CDBG entitlement shares serve as the underlying distribution key for both ESG and CoC grants, ensuring that areas with higher composite need scores receive proportionally larger homeless-assistance awards each year.

Within each CoC, grant administrators allocate the awarded CoC and ESG funds across diverse program types to expand and manage local homeless-assistance capacity. The ESG is oriented toward the immediate crisis end of the continuum: current regulations permit spending on street outreach, emergency shelter construction and operations, shelter resident services, rapid re-housing assistance, homelessness prevention, and the maintenance of the local Homeless Management Information System, with administrative expenses capped at seven and one-half per cent of the award (U.S. Department of Housing and Urban Development, 2019c). CoC grants, by contrast, finance longer-term or system-level solutions such as permanent supportive housing, rapid re-housing projects, transitional housing, joint transitional–rapid re-housing initiatives, safe-haven beds, supportive-services-only projects, CoC planning activities, enhancements to local Homeless Management Information System (HMIS) capabilities, and unified funding agency costs (U.S. Department of Housing and Urban Development, 2020).

Some funding allocations convert swiftly into additional bed inventory, such as leasing new shelter units, expanding emergency or seasonal beds, and opening overflow facilities, thereby increasing year-round and emergency-supply capacity in the immediate term (Moulton, 2013; Stergiopoulos et al., 2015). Other allocations underwrite capital projects for PSH or TH, which involve

land acquisition, construction or rehabilitation, and service coordination; these activities unfold over longer horizons, yielding gradual increases in housing stock and supportive-services infrastructure (Popov, 2016). Furthermore, planning and HMIS investments strengthen a CoC's operational capacity—improving bed utilization tracking, placement protocols, and data quality—but do not directly add new beds. By combining rapid-deployment resources (e.g., ESG-funded rental assistance and emergency-shelter contracts) with capital and service investments (e.g., CoC-funded PSH development and planning), CoC administrators dynamically adjust both short-run shelter availability and long-run housing infrastructure in response to annual funding variations.

2.3.2 Heterogeneity in Utilization

As CoCs expand bed capacity, people experiencing homelessness respond to the availability of new resources in heterogeneous ways. Overnight shelters frequently target certain population groups: individual adult men, women and families with children, or people suffering from severe mental illness or substance use. Emergency and transitional beds often attempt to accommodate families with children, people fleeing domestic violence, and other vulnerable subpopulations who prefer shelter over doubling up or couch-surfing when space opens (Cragg and O'Flaherty, 1999; Wong et al., 1997; Culhane et al., 2011; O'Flaherty, 2019). Conversely, individuals with chronic patterns of unsheltered homelessness may face substantial barriers to entering new shelters, such as eligibility restrictions, perceived stigma, or behavioral health challenges, and therefore may remain unsheltered even as beds increase (Stefancic and Tsemberis, 2007; Byrne et al., 2014). Rapid rehousing subsidies and overflow/seasonal options can draw in marginally housed persons who were not previously counted in street-based PIT surveys, causing a rise in reported shelter-use without a corresponding decline in unsheltered PIT counts (Schneider et al., 2016; Tsai and Alarcón, 2022). By contrast, PSH units take longer to develop but eventually provide stable housing for those with long-term needs, potentially reducing chronic unsheltered homelessness only over extended horizons (Stergiopoulos et al., 2015).

Behavioral and spillover pathways further complicate local utilization patterns. First, as some CoCs gain new beds, individuals may migrate from neighboring jurisdictions to access more resources (Tsemberis and Eisenberg, 2000; Popov, 2016). Such movements may depress unsheltered counts in lower-funded CoCs while boosting sheltered utilization in higher-funded ones, confounding direct comparisons of funding and outcomes. Second, because HUD's formula indices embed prior homelessness indicators—such as overcrowding and poverty—there exists endogenous targeting: areas with historically high PIT counts receive larger entitlement shares in subsequent years, creating reverse-causality concerns (Popov, 2016; Bunce, 1979; Collinson, 2014). By instrumenting contemporaneous funding changes with the exogenous annual fluctuations in the pre–1940 housing share, our design isolates the causal effect of new bed capacity on utilization and prevents bias from these spillover and targeting channels.

To connect the conceptual framework to the empirical implementation, we now describe the data sources and measurement procedures used to capture both the variation in CoC-level funding allocations and the resulting changes in homeless counts and bed capacity. Specifically, we begin by detailing how annual PIT counts of sheltered and unsheltered persons are collected at the Continuum of Care level, which form our primary outcome measures.

2.4 Data

2.4.1 Point–in–Time Homelessness Counts

Every January, during the final ten days of the month, each CoC must conduct a PIT enumeration under the McKinney–Vento Homeless Assistance Act, as amended by the HEARTH Act of 2009. HUD is authorized by the McKinney–Vento Act and implements this authority through the CoC Program Interim Rule, which obligates CoCs to conduct a one-night count of sheltered and unsheltered persons at least biennially (U.S. Department of Housing and Urban Development, 2012). Although the statutory minimum for unsheltered counts remains once every two years, HUD has required—and CoCs now effectively conduct—annual PIT counts to satisfy both CoC

Program Competition and Emergency Solutions Grants reporting requirements (U.S. Department of Housing and Urban Development, 2024).

Sheltered counts, which cover people in emergency shelters, transitional housing, and safe havens, are compiled primarily from Homeless Management Information System (HMIS) bed-night records, supplemented by facility surveys for programs not participating in HMIS (U.S. Department of Housing and Urban Development, 2014). Unsheltered counts enumerate persons residing in places not meant for human habitation—such as streets, parks, vehicles, and encampments—using route-based canvassing, short intercept interviews, and partnerships with outreach workers, law enforcement, and other service providers to identify and survey individuals (U.S. Department of Housing and Urban Development, 2014). Data collection protocols—including sampling frames, volunteer deployment, use of mobile survey tools, and quality-assurance checks—are prescribed in HUD's annual guidance but implemented locally, resulting in variation in precision and coverage across CoCs (Government Accountability Office, 2020, 2021).

PIT count data and accompanying methodology documentation are submitted each spring via the Homelessness Data Exchange as a non-waivable condition for HUD CoC and ESG funding eligibility. HUD uses these submissions both to produce the Annual Homelessness Assessment Report to Congress and as scoring factors in the CoC Program Competition, meaning that failure to submit compliant PIT data can directly affect a CoC's funding allocation (U.S. Department of Housing and Urban Development, 2012). Our '2019' cross-section uses Point-in-Time counts conducted in January 2020 and federal funding allocations for fiscal year 2019 (October 2018–September 2019). Thus, the PIT count occurs approximately four months after the end of the fiscal year, allowing time for most allocated funds to be obligated and spent.

This study draws raw CoC-level tabulations from HUD's Homelessness Data Exchange and the publicly released Annual Homeless Assessment Report (AHAR). For each CoC reported within the AHAR in 2019, the dataset captures total homelessness, the sheltered and unsheltered subtotals, mutually exclusive demographic splits by age, race, gender, and chronicity, and four non-exclusive

subpopulations—severe mental illness, chronic substance abuse, domestic-violence survivorship, and veterans. [7]

All PIT counts are converted to rates per ten thousand total residents within a CoC to achieve comparability across CoCs and to align with the funding denominators. Year-specific total population comes from the tract-level American Community Survey five-year files aggregated to the CoC level, which is described later in this section.

2.4.2 Shelter and Housing Capacity

To observe how federal dollars translate into physical capacity, the analysis incorporates the annual HIC that HUD collects in tandem with the PIT enumeration. Each CoC submits a census of beds and housing units that are dedicated to people experiencing homelessness as of the same night as the PIT count. The HIC distinguishes year-round beds in emergency shelters (ES), transitional housing (TH) and safe-haven projects (SH); seasonal and overflow ES beds; and PSH beds, among other components (U.S. Department of Housing and Urban Development, 2023).

Each program describes a different form of sheltered experience for a person utilizing its space. First, *total year-round ES–TH–SH beds* measure the stock of beds that operate every night in emergency shelters, transitional housing, and safe-haven programs. Emergency shelters provide short-term crisis accommodation, transitional-housing projects offer time-limited stays (usually up to twenty-four months) coupled with supportive services, and safe havens are low-barrier facilities intended for persons with severe mental-health conditions who may not initially engage with standard shelter settings (U.S. Department of Housing and Urban Development, 2023). Second, *year-round PSH beds* represent permanent supportive-housing capacity; these beds come with indefinite tenure and wrap-around services and are therefore the principal vehicle for the "Housing First" model emphasised in Continuum of Care funding. Third, *seasonal ES beds* are activated only during defined periods—typically winter months or extreme-weather declarations—and thus expand crisis capacity when exposure risk is highest. Fourth, *overflow ES beds* are temporary

[7] Because many subpopulation cells appear only in image-based tables, they are digitised with optical-character-recognition using the `tesseract` package in R (U.S. Department of Housing and Urban Development, 2019b).

placements, often in congregate spaces such as cots in cafeteria areas, that CoCs open whenever demand exceeds regular and seasonal supply; they provide the most flexible but least stable form of shelter. We measure the response of the bed-capacity of the four programs to better understand how CoC coordinators allocate funding resources across varying forms of shelter programs.

The raw data are published through HUD's Homelessness Data Exchange in a series of yearly spreadsheets, where each row represents a Continuum of Care and columns record bed counts by project type. We retain four aggregate measures: total year-round ES, TH and SH beds; total year-round PSH beds; total seasonal ES beds; and total overflow ES beds. We transform the bed counts similarly to allow each capacity measure to be expressed per ten thousand residents of the corresponding CoC.

2.4.3 Federal Homelessness–Assistance Funding

Funding award information is obtained from the HUD Exchange *Awards and Allocations* workbooks for 2019 (U.S. Department of Housing and Urban Development, 2020). CoC grant totals are reported at the CoC level and therefore require no further geographic manipulation. ESG allocations, issued to states, counties, and entitlement cities, are matched to CoCs using the tract-based spatial crosswalk described in Section 2.4.5. When a state possesses a designated Balance-of-State CoC, the statewide ESG portion is assigned in full to that CoC; otherwise the statewide amount is prorated across constituent CoCs according to the same population weights employed elsewhere in the study. All dollar figures are expressed in 2011 values using the Consumer Price Index.

For each CoC, we measure the total amount of funding through the summation of the ESG and CoC grant award as our primary funding variable of interest. The amount is expressed in thousands of dollars and divided by the corresponding CoC population and multiplied by ten thousand residents, yielding per-10,000 funding rates that are directly comparable with the population-scaled homelessness outcomes introduced in Section 2.4.1.

Despite their differences, both ESG and CoC allocations ultimately derive from an underlying Community Development Block Grant (CDBG) allocation mechanism, which assigns entitlement

shares to cities and counties based on a weighted combination of need-based and demographic variables. The CDBG formula computes two indices for each entitlement jurisdiction in a given fiscal year, as seen in equations (1) and (2) above.

Popov (2016) and Lucas (2017) observe that the pre-1940 housing share generates idiosyncratic variation in entitlement shares that is plausibly orthogonal to differences in homelessness across the recipient communities. In our study, we exploit exactly this feature to instrument CoC and ESG funding with the ACS-derived pre-1940 share, thereby isolating exogenous differences in grant allocations. We discuss the plausibility of the instrumental variable strategy further in Section 2.5.

2.4.4 Between-CoC Sample Construction

Our primary causal analysis focuses on a cross-section of CoCs in 2019. To provide auxiliary "between-CoC" estimates leveraging longer-run variation in funding and homelessness, we also construct CoC-level averages across 2015–2019. We apply the data assembly steps described in Sections 2.4.1, 2.4.2, and 2.4.3 for each year from 2015 through 2019, yielding observations for 373 unique CoCs across five years (total observations $N = 1{,}840$), with annual coverage from 365 CoCs in 2015 to 370 in 2019. Missingness reflects intermittent non-submission of PIT or HIC data and the creation/merging of new CoCs. Chi-square tests confirm that neither geographic nor temporal patterns of missing observations correlate with funding or outcomes ($\chi^2_{\text{state, completeness}} = 34.89$, $p = 0.92$; $\chi^2_{\text{year, completeness}} = 4.4$, $p = 0.35$), indicating randomness in incomplete data. While our main specification exploits only 2019 data for a purely cross-sectional IV analysis, we also compute five-year averages (2015–2019) for each CoC to support robustness checks and descriptive between-CoC comparisons, illustrating how sustained differences in funding levels relate to average homelessness and capacity outcomes across CoCs.

2.4.5 Geographically Harmonised Covariates

The empirical strategy requires that every explanatory variable share a common CoC geography. Because socioeconomic data most relevant to this problem are published only at more orthodox scales, a Census tract–to–CoC spatial crosswalk is constructed and applied uniformly to

all tract–based inputs. The procedure follows best practice in Ferrara et al. (2024) and improves on the county merges used by Popov (2016) and Lucas (2017).

The crosswalk is rebuilt for each study year because HUD occasionally redraws CoC boundaries. Annual CoC shapefiles are projected into the USA Contiguous Albers equal–area coordinate system (EPSG 5070) and intersected with the same-year Census tracts, obtained from TIGRIS-line files. The intersection yields a complete partition: each tract is split into one or more polygons whose union exactly reconstructs the parent tract. We overlay the Global Human Settlement Population raster (100m, R2015-R2019) and compute, for each polygon, its share of tract population; where raster data are unavailable the polygon's equal–area share is used. These weights sum to one by construction and ensure that allocations reflect where people actually live rather than where land area is largest. Diagnostic statistics show that the median tract retains 96% of its population in a single CoC, but 17% (about 15,000 total) of tracts straddle two or more CoCs, warranting the population-weighting scheme (Ferrara et al., 2024).

For our main (cross-sectional) IV analysis, we focus on 2019 covariates, but we construct the same covariates for 2015–2019 to support our between-CoC robustness checks for each CoC via the population-weighted crosswalk. These include total population, vacant housing units, units built before 1940, persons below poverty, unemployed individuals, overcrowded households, rent-burdened households, households receiving SNAP and SSI benefits, uninsured households, non-Hispanic Black and Hispanic population counts, persons aged sixty-five or older, veterans, and persons with disabilities. After summing each count across tracts within a CoC, we compute shares or rates as appropriate: for example, the pre-1940 housing share is the ratio of pre-1940 units to total housing units, while demographic and program-receipt shares divide subgroup counts by total population. This aggregation-then-ratio approach preserves internal consistency and avoids the ratio-of-ratios bias detailed in Ferrara et al. (2024). This effort results in a cross-section for the year 2019 of 370 CoCs and a cross-section of 5-year average values from 2015 to 2019 of

373 Continuums of Care, each endowed with a rich set of harmonized housing, demographic, and socioeconomic covariates[8].

Weather controls are derived from the National Oceanic and Atmospheric Administration's (NOAA) daily county–level data (Durre et al., 2022). For every county and year we compute the mean January temperature and the total January precipitation. Counties are then intersected with the same CoC polygons used in the tract procedure, except that simple polygon area shares replace population weights because climate variables are physical rather than demographic quantities. Area-weighted means and totals are aggregated to the CoC level. Tying the climate window to January matches the conditions encountered by PIT enumerators and the visibility of unsheltered homelessness (Elliott and Krivo, 1991; Corinth and Lucas, 2018).

Spatial spillovers pose a significant threat to identification if individuals experiencing homelessness relocate across CoC boundaries in response to differences in service availability. To account for the relative strength of nearby funding, we compute for each CoC–year a "nearby funding ratio," defined as the total contemporaneous per-capita funding summed over all proximate CoCs within a fifty-mile radius, divided by the focal CoC's own per-capita funding. Formally, letting F_{jt} denote the per-10,000-resident funding level in CoC j and $N(c)$ the set of nearby CoCs c, we first calculate:

$$NearbySum_{ct} = \sum_{j \in \mathcal{N}(c)} F_{jt}, \tag{2.4}$$

and then construct

$$\text{Nearby\_ratio}_{ct} = \frac{NearbySum_{ct}}{F_{ct}}. \tag{2.5}$$

By expressing proximate resources in units of the focal CoC's own funding intensity, this ratio absorbs both demand-side migration incentives and supply-side strategic complementarity in program provision across adjacent jurisdictions. Omitting this control would risk conflating the effect of a CoC's own awards with the influence of better- or worse-resourced neighbours—an is-

[8]For more technical information about the construction of the crosswalk, see Appendix B.1.1

sue Popov (2016) demonstrates is empirically important, as people experiencing homelessness are able to move toward higher-resource areas.

Table 2.1 displays summary statistics of our outcome variables of interest, which includes homelessness rates per 10,000 CoC residents by sheltered status and demographic group as well as HIC bed space counts per 10,000 CoC residents. Table 2.2 displays the funding values of the ESG and CoC grants in thousands of dollars per 10,000 CoC residents, as well as the full set of harmonised covariates at the CoC level.

2.5 Methods

2.5.1 Primary Specification

Our core identification strategy exploits the cross-section of 370 CoCs in 2019, using each CoC's share of pre-1940 housing stock—measured in the 2015–2019 ACS five-year estimates—as an instrument for its per-10,000 resident federal grants (CoC + ESG). By anchoring the analysis in a single year, we avoid concerns that year-to-year changes in old-housing shares might reflect contemporaneous demolition or renovation activity correlated with local homelessness trends. Instead, we draw on long-run, historically determined variation in housing-stock vintage while leveraging modern ACS data.

Concretely, let i index the 370 CoCs observed in 2019. Our funding variable of interest, $Fund_i$, is the sum of CoC and ESG grant dollars (in 2011-adjusted thousands of dollars) per 10,000 residents in CoC i. The instrument, $OldHousing_i$, is the fraction of all housing units in CoC i constructed prior to 1940, drawn from the 2015–2019 ACS five-year estimates; as shown in Section 2.4.5, this share varies meaningfully across CoCs but is predetermined with respect to 2019 homelessness. A rich set of control variables, X_i, captures contemporaneous CoC characteristics: demographic shares (non-Hispanic Black, Hispanic, age groups, SNAP and SSI receipt, veteran status, disability status), housing-market indicators (vacancy rate, overcrowding rate, rent-burdened households), the ratio of nearby CoC funding to own CoC funding (nearby ratio), and climatic factors (January mean temperature and total precipitation).

Our two-stage least squares (2SLS) procedure proceeds as follows, with the first stage as:

$$Fund_i = \pi OldHousing_i + X_i'\gamma + \epsilon_i \tag{2.6}$$

Here, π captures how long-run housing-age differences predict cross-CoC funding levels, after adjusting for contemporaneous covariates X_i.

The second stage is:

$$Outcome_i = \beta Fund_i + X_i'\delta + u_i \tag{2.7}$$

where $Outcome_i$ is one of our per-10,000 CoC measures in 2019 (total, sheltered or unsheltered PIT counts; subgroup counts; or HIC bed-capacity metrics). The coefficient β captures the local average treatment effect of an exogenous increase in federal funding.

All standard errors are heteroskedasticity-robust and clustered by CoC. First-stage F-statistics on $OldHousing_i$ exceed 40 in our preferred specifications, comfortably above the conventional threshold, confirming strong instrument relevance.

Although HUD's CDBG mechanism classifies individual entitlement jurisdictions as Formula A or Formula B, our unit of analysis—the CoC—typically aggregates multiple entitlement areas. Even CoCs whose largest entitlement share is driven by the "Formula A" index often contain at least one county or city whose allocation is determined by the pre-1940 housing share (the Formula B inputs) (Popov, 2016). As shown in Popov (2016), the choice between A and B formulas at the entitlement level is a "max" operation that does not fully mute the B-formula's influence once grants are pooled, so even "Formula A" CoCs receive an identifiable dose of pre-1940 driven variation, warranting the full sample inclusion of all CoCs across the time period. Within the 2019 sample, 85% of CoCs were "Formula B" communities[9].

[9]Author's calculations using ACS 5-year survey values of CDBG formula variables.

2.5.2 Between-CoC IV Specification

While our primary 2SLS estimates report a causal interpretation, it is also informative to examine how sustained differences in average funding levels relate to average homelessness and capacity outcomes across CoCs. To that end, we complement the cross-sectional 2SLS analysis with a "between" IV specification that operates on five-year averages of each CoC's outcomes, funding, and covariates. Although this approach does not eliminate time-invariant CoC heterogeneity and thus cannot be interpreted as strictly causal, it descriptively links long-run, average funding intensity to long-run, average outcome levels, offering a perspective on the cumulative scale of federal grants and their association with homelessness and bed capacity.

Formally, let $\overline{Fund_i}$ be equal to the average of total CoC grant and ESG funding in CoC i over the years 2015 to 2019. Additionally, let $\overline{OldHousing_i}$ represent the average level of pre-1940's housing share and $\overline{X_i}$ to represent average levels of the covariate vector across the same time span. The between-CoC first stage is then:

$$\overline{Fund_i} = \pi \overline{OldHousing_i} + \overline{X_i}'\gamma + \epsilon_i \tag{2.8}$$

The corresponding second stage is then:

$$\overline{Outcome_i} = \beta \overline{Fund_i} + \overline{X_i}'\delta + u_i \tag{2.9}$$

Where $\overline{Outcome_i}$ is the average of the sheltered and unsheltered PIT counts and the HIC counts across 2015 to 2019 for each CoC.

These between-CoC regressions offer a useful complement to our primary cross-sectional IV results, highlighting the scale of longer-run associations between federal grant levels and both homelessness and shelter capacity.

2.5.3 Robustness Checks

The primary sheltered and unsheltered specifications are repeated for the rates of people experiencing homelessness per 10,000 total CoC population living below the federal poverty line (Lucas, 2017). This alternative scale helps control for secular shifts in local poverty levels, and tests the concern that a CoC with rising poverty rates could mechanically exhibit more "homeless per 10,000" just because its denominator decreased.

Additionally, the sheltered and unsheltered specifications are conducted after dropping observations for the New York City CoC (NY-600) and the Los Angeles City/County CoC (CA-600), as those areas experienced growth in homelessness at incomparable rates to all other parts of the U.S. over this period[10] (Lucas, 2017; O'Flaherty, 2019).

2.6 Results

2.6.1 First-Stage Estimates

Table 2.3 presents our four first-stage specifications, in which the dependent variable is total federal funding (CoC + ESG) per 10,000 residents. Columns (1) and (2) use the 2019 cross-section; Columns (3) and (4) use CoC-level five-year averages (2015–2019).

In Column (1), we regress 2019 per-capita funding on the pre-1940 housing share without any additional covariates. A one-percentage-point higher share of pre-1940 housing is associated with an extra $4,332 of funding per 10,000 residents (SE = 0.291, $p < 0.01$), and the first-stage F-statistic on the instrument is 220.9, confirming very strong relevance. Column (2) adds the full set of time-varying controls—demographic shares (e.g. Black population share, share over age 65), housing-market indicators (vacancy, overcrowding, rent burden), special-population shares (SNAP, SSI, veteran, disability), the nearby-funding ratio, and January temperature and precipitation. Even after adjusting for these covariates, the coefficient on the pre-1940 share remains

[10] In 2015, New York City and Los Angeles together held nearly 15% of the nation's sheltered beds and 20% of unsheltered counts, even though they represent just 2 of 370 nationwide CoCs.

virtually unchanged at 4.340 (SE = 0.364, $p < 0.01$), and the instrument's F-statistic remains high at 48.6.

Panels (3) and (4) repeat the exercise using five-year CoC averages. In the unadjusted between-CoC regression (Column 3), a one-point increase in average pre-1940 housing share is associated with \$3,852 more funding per 10,000 residents (SE = 0.264, $p < 0.01$, F = 213.7). Adding the same covariate averages in Column (4) leaves the point estimate at \$3,961 (SE = 0.366, $p < 0.01$, F = 39.3). Although these between-CoC estimates do not purge fixed, unobserved CoC characteristics, they confirm that long-run variation in pre-1940 housing share captures the bulk of cross-sectional funding differences driven by HUD's CDBG formula.

Together, these results demonstrate that the ACS-derived pre-1940 housing share is a highly relevant instrument: it generates substantial within-CoC variation in annual funding (Columns 1–2) and is associated with large cross-CoC differences in average funding levels (Columns 3–4), even after controlling for a rich set of contemporaneous community characteristics.

2.6.2 Sheltered and Unsheltered Homelessness

Table 2.4 reports our 2SLS estimates of how per-capita federal funding affects sheltered and unsheltered homelessness, using both the 2019 cross-section (Columns 1–2 and 4–5) and the 2015–2019 between-CoC averages (Columns 3 and 6). All specifications instrument total funding per 10,000 residents with the pre-1940 housing share and cluster standard errors at the CoC level.

In the simple cross-section without additional controls (Column 1), a \$1,000 increase in annual funding per 10,000 residents raises the sheltered count by 0.123 persons (SE = 0.011, $p < 0.01$). When we add the full suite of demographic, housing-market, special-population, nearby-funding, and January climate controls (Column 2), the point estimate falls slightly to 0.100 (SE = 0.017, $p < 0.01$), but remains highly significant.

Turning to the between-CoC specification (Column 3), which regresses five-year average sheltered rates on five-year average funding and covariates, the coefficient is 0.092 (SE = 0.019,

$p < 0.01$). Although smaller than the within-CoC estimates, this long-run association confirms that CoCs with persistently higher funding support modestly higher average sheltered counts.

In contrast, higher funding does not translate into fewer unsheltered individuals in the same year. In the unadjusted cross-section (Column 4), the point estimate is -0.034 (SE = 0.014, $p < 0.05$), suggesting a small negative effect; however, once we add the full control set (Column 5), the coefficient attenuates to -0.001 (SE = 0.015, $p > 0.10$) and loses statistical significance. The between-CoC estimate (Column 6) is -0.004 (SE = 0.016, $p > 0.10$), again indistinguishable from zero.

These results indicate that exogenous increases in federal homeless-assistance funding robustly expand sheltered placements within the same year—on the order of about 1 additional sheltered persons per $10,000 per 10,000 residents—while having no meaningful short-run impact on street-sleeping counts. The between-CoC findings reinforce this pattern over longer horizons: sustained funding differentials are correlated with modestly higher average shelter use but do not alter unsheltered rates. This divergence suggests that additional beds primarily draw in people who were not previously counted as unsheltered, rather than directly converting street homelessness into shelter stays.

2.6.3 Shelter Capacity: Total Beds and Permanent Supportive Housing

Table 2.5 presents 2SLS estimates of how per-capita federal funding affects two measures of bed capacity: total year-round shelter beds (emergency, transitional, and safe-haven) and permanent supportive housing (PSH). Columns (1) and (3) use the 2019 cross-sectional specification; Columns (2) and (4) report the 2015–2019 between-CoC averages.

In the cross-section (Column 1), a $1,000 increase in annual funding per 10,000 residents is associated with 0.087 additional shelter beds per 10,000 (SE = 0.018, $p < 0.01$). This result confirms that same-year grants rapidly translate into expanded emergency and transitional capacity—consistent with the notion that ESG-funded shelter contracts and CoC-funded rapid re-housing can be deployed quickly when funds arrive (Moulton, 2013). In the between-CoC

specification (Column 2), CoCs with $1,000 higher average funding per 10,000 residents have, on average, 0.085 more beds per 10,000 (SE = 0.020, $p < 0.01$). Although correlational, this association illustrates how sustained resource differences map into baseline shelter infrastructure.

PSH capacity also responds significantly to funding differences. In the 2019 within-CoC model (Column 3), the point estimate is 0.131 PSH beds per 10,000 per $1,000 funding (SE = 0.013, $p < 0.01$), suggesting that a portion of funding does flow into PSH after allocation. In the between-CoC regression (Column 4), CoCs with $1,000 higher mean funding per 10,000 residents are associated with 0.111 more PSH beds per 10,000 (SE = 0.013, $p < 0.01$). This demonstrates that areas with persistently higher funding levels accumulate greater PSH capacity, consistent with the capital-intensive nature of permanent supportive housing development (Rog et al., 2014; Stergiopoulos et al., 2015).

Table 2.6 examines two additional bed types that may surge during the January PIT count. In the 2019 within-CoC model (Column 1), a $1,000 per 10,000 funding increase yields 0.009 overflow beds per 10,000 (SE = 0.003, $p < 0.01$), indicating that some new funding goes to surge-capacity arrangements (e.g. overflow cots) when demand spikes (Byrne et al., 2013). The effect on seasonal beds (Column 3) is small (0.004) and not statistically significant. Between-CoC associations for both bed types (Columns 2 and 4) are modest and imprecise, underscoring that these temporary categories fluctuate year-to-year with policy and weather conditions rather than long-run funding levels.

These combined patterns raise a critical question: if federal grants are swiftly constructing new temporary beds and sheltered counts rise accordingly, why do we not observe a corresponding decline in unsheltered homelessness? One explanation is that the additional shelter capacity is drawing in individuals who were never counted as unsheltered in the first place. For example, families and individuals doubling up in overcrowded or precarious housing (e.g. those coping with domestic violence, severe mental-health conditions, or substance-use challenges) may opt into stable shelter programs without ever appearing in the street-based Point-in-Time enumeration (Byrne et al., 2013; O'Flaherty, 2019; Tsai and Alarcón, 2022; Wong et al., 1997). A second,

complementary mechanism is geographic re-sorting: as Popov (2016) argues, people experiencing homelessness will migrate toward jurisdictions with more robust service networks. Our finding that neighboring CoC funding is negatively associated with local unsheltered counts supports this view, suggesting that some portion of the sheltered increase reflects in-migration of individuals seeking shelter rather than direct transitions out of street homelessness.

2.6.4 Heterogeneity by CoC Category

Disaggregating our analysis by HUD's four CoC geographic categories—"Major City CoC," "Other Largely Urban CoC," "Largely Suburban CoC," and "Largely Rural CoC" (U.S. Department of Housing and Urban Development, 2012)—allows us to probe whether funding effects vary with population density and service-delivery infrastructure. HUD assigns each CoC to one of these categories based on census definitions of urbanized areas, metropolitan adjacency, and rural composition, under the logic that densely populated centers operate very different shelter and outreach systems than more dispersed jurisdictions.

Table 2.7 reports our 2SLS estimates separately for (i) Urban CoCs (Major City + Other Largely Urban) and (ii) Rural/Suburban CoCs (Largely Suburban + Largely Rural). In both settings, exogenous increases in per-capita federal funding predict nearly identical boosts in sheltered homelessness: roughly 0.09 additional sheltered persons per 10,000 residents for each extra $1,000 of funding, significant at the 1 percent level in both strata. This uniformity suggests that, regardless of density, communities are able to deploy new grants rapidly into emergency, transitional, or safe-haven beds.

However, the impact on unsheltered homelessness diverges sharply. In Urban CoCs, higher funding yields no statistically detectable change in street-sleeping counts (coefficient = 0.023, $p > 0.10$), whereas in Rural/Suburban CoCs a $1,000 funding increase is associated with a significant decline of 0.096 unsheltered individuals per 10,000 residents ($p < 0.05$). This pattern implies that in less densely settled areas, where baseline shelter capacity and outreach networks are thinner, marginal resources can directly draw more people off the streets. In contrast, in urbanized

CoCs—often already operating near capacity or facing logistical constraints—additional dollars primarily expand formal shelter placements without immediately reducing the unsheltered population. Understanding this geographic heterogeneity is critical for tailoring federal grant strategies: urban centers may need complementary investments in outreach, transportation, or data-driven targeting to convert shelter-capacity gains into declines in street homelessness, whereas rural and suburban areas appear to benefit from more straightforward capacity additions.

CoC boundaries reflect both HUD administrative decisions and local service-provider coordination, resulting in substantial variation in geographic scale—from single-county metropolitan areas to multi-county rural regions or entire states (Balance-of-State CoCs). While this heterogeneity complicates direct comparisons, our rich covariate set and separate urban/rural analyses help account for these structural differences.

2.6.5 Heterogeneity by Individuals and Families

Disaggregating our analysis by household type—single individuals versus families—sheds light on which groups capture new shelter capacity and who remains on the street when funding rises. Under HUD's PIT framework, "individuals" are single adults or unrelated persons counted separately, while "families" include at least one adult and one or more children sharing living quarters, often housed in dedicated family units with wrap-around services (U.S. Department of Housing and Urban Development, 2014; O'Flaherty, 2019).

Table 2.8 presents our 2SLS estimates, instrumenting per-capita federal grants with the pre-1940 housing share. In the cross-sectional 2019 model, a $1,000 increase in funding per 10,000 residents leads to an additional 0.047 sheltered individuals per 10,000 (SE = 0.009, $p < 0.01$) and an additional 0.052 sheltered family members per 10,000 (SE = 0.012, $p < 0.01$). Both effects are highly significant, but the slightly larger coefficient for families suggests that when new beds become available—often larger or more specialized units—households with children marginally outcompete single adults for these slots.

In contrast, the unsheltered margins barely budge. We estimate 0.006 additional unsheltered individuals per 10,000 (SE = 0.014, $n.s.$) and a 0.007 change for unsheltered family members (SE = 0.002, $p < 0.01$). Although the decline in unsheltered families is statistically significant, its magnitude is small: roughly one fewer person in a family moves out of unsheltered homelessness for an additional $143,000 in funding per 10,000 CoC residents. This negligible shift implies that most of the rise in sheltered counts reflects the intake of people who were previously doubled-up or in precarious housing—particularly families—rather than a direct conversion of street-sleepers into shelter clients.

These findings underscore two important points. First, federal grants rapidly expand formal shelter capacity in proportion to need and design: family units respond slightly more strongly in CoCs with higher funding levels compared to individuals. Second, the minimal movement among unsheltered individuals suggests that simple bed-count increases are mostly ineffective in moving people experiencing unsheltered homelessness into immediate shelter. Policymakers seeking to reduce unsheltered homelessness may therefore need to pair capacity expansions with targeted outreach, transportation supports, or flexible shelter entry policies, especially for single adults who face unique barriers to accessing congregate or gender-segregated facilities.

2.6.6 Heterogeneity by Subpopulation

Table 2.9 presents the 2SLS estimates for demographic and subpopulation groups in the sheltered counts. In nearly every sheltered category, a $1,000 increase in per-10,000 funding yields a statistically significant rise in shelter use. Among age groups, adults over 24 show the largest response with 0.061 additional sheltered individuals per 10,000 (SE = 0.010, $p < 0.01$), followed by children under 18 at 0.031 per 10,000 (SE = 0.007, $p < 0.01$), and young adults aged 18–24 with a smaller but significant effect of 0.008 per 10,000 (SE = 0.001, $p < 0.01$). Both women (0.044, SE = 0.009, $p < 0.01$) and men (0.054, SE = 0.009, $p < 0.01$) experience significant increases in shelter access, with men showing a slightly larger response. Among racial groups, Black individuals demonstrate a stronger shelter response (0.055, SE = 0.010, $p < 0.01$) compared to white

individuals (0.038, SE = 0.009, $p < 0.01$). Subpopulations facing chronic challenges also respond positively: those with mental illness (0.021, SE = 0.004, $p < 0.01$), substance abuse issues (0.022, SE = 0.004, $p < 0.01$), and domestic violence survivors (0.006, SE = 0.002, $p < 0.01$) all show significant increases. People experiencing chronic homelessness increase by 0.018 per 10,000 in response to additional funding (SE = 0.003, $p < 0.01$).

An advantage of our heterogeneity analysis is that the subgroup coefficients can be interpreted as shares of the overall sheltered response. Recall that the total within-CoC effect of funding on sheltered homelessness is 0.099 per 10,000. Because the age groups are mutually exclusive, we can decompose the total effect mechanically: adults over 24 account for approximately 61% of the overall sheltered increase (0.061/0.099), children under 18 contribute 31% (0.031/0.099), and young adults aged 18–24 represent 8% (0.008/0.099). Similar calculations for gender yield that men comprise 54% of the shelter response (0.054/0.099) and women 44% (0.044/0.099), while for race, Black individuals represent 55% of the total response (0.055/0.099) compared to 38% for white individuals (0.038/0.099). By contrast, the subpopulations defined by severe conditions overlap—an individual may belong to multiple categories—so their coefficients (0.021, 0.022, 0.006) imply a combined share ranging from 22% (0.022/0.099) to 49% ((0.021+0.022+0.006)/0.099) of the total sheltered effect, depending on overlap assumptions. People experiencing chronic homelessness represent 18% of the shelter response (0.018/0.099). This decomposition reveals that working-age adults and men drive the majority of the shelter response, while highlighting substantial representation across demographic groups and vulnerable subpopulations.

While our main analysis finds no overall effect of increased federal grants on street homelessness, a more granular look across demographic subgroups uncovers important patterns (Table 2.10). Across age and gender cohorts, the changes are uniformly small: a modest 0.004-person uptick per 10,000 young adults and a 0.004 decline among children under 18 both reach only borderline significance, while adults over 24, women, men, and chronically homeless individuals exhibit effects indistinguishable from zero.

The most pronounced—and sensitive—finding emerges by race. For white individuals, a $1,000 bump in per-capita funding is associated with a statistically significant drop of 0.020 unsheltered persons per 10,000 (SE = 0.010, $p < 0.05$). By contrast, black individuals see a significant increase of 0.015 per 10,000 (SE = 0.005, $p < 0.01$) in their unsheltered count. Both groups nonetheless experienced positive shelter-entry responses (Table 2.9), implying that net unsheltered changes reflect more than simple movements into shelter.

Interpreting this racial divergence demands caution. One possibility is improved outreach and enumeration: better-funded CoCs may simply count more black individuals who were previously missed, narrowing longstanding undercounts in the Point-in-Time survey. Alternatively, systemic barriers—such as location of new shelters, transportation gaps, or discriminatory practices—may limit black individuals' ability to access bed space even as it expands for others . Finally, underlying housing-market shifts could displace black households while drawing more visible white homelessness into the count.

All other special-condition groups—those with mental illness, substance-use disorders, or fleeing domestic violence—show no significant unsheltered response, reinforcing that federal grants primarily operate by increasing formal shelter capacity rather than directly diminishing street-sleeping counts among these vulnerable populations. Taken together, the nearly offsetting racial effects help explain our aggregate null result, underscoring the imperative to assess distributional impacts when evaluating homelessness interventions.

2.6.7 Robustness

In Table B.2, we re-estimate the 2019 2SLS specification using homelessness rates per 10,000 people in poverty (rather than per 10,000 total residents). The point estimate on funding for sheltered homelessness remains positive and highly significant—at 0.687 ($SE = 0.115, p < 0.01$)—indicating that a $10,000 increase in per-10,000 funding corresponds to roughly 7 additional sheltered individuals per 10,000 poor residents. The unsheltered coefficient (0.074, SE = 0.117) remains statistically indistinguishable from zero.

Similarly, Table B.3 presents the within-CoC estimates after dropping New York City (NY-600) and Los Angeles City/County (CA-600). The sheltered-homelessness coefficient is unchanged at 0.099 ($SE = 0.016, p < 0.01$) and the unsheltered coefficient remains small (-0.001, SE = 0.015) and insignificant. Hence, neither the poverty-scaled outcome nor the exclusion of these two outliers overturns our main result: federal grants causally raise sheltered PIT counts without a concomitant decrease in unsheltered homelessness.

2.7 Discussion

This paper provides new, causal evidence on how federal homeless-assistance grants affect reported homelessness and shelter capacity across U.S. CoCs in 2019. By exploiting cross-sectional variation in the ACS-derived pre-1940 housing share as an instrument for combined CoC and ESG funding, our 2SLS design yields several key insights that advance understanding of federal homelessness policy effectiveness.

First, contemporaneous grant allocations induce sizable increases in sheltered homelessness—with roughly 1 additional sheltered placement per $10,000 per 10,000 residents—yet have no detectable short-run impact on unsheltered counts. In other words, ESG and CoC dollars principally expand formal shelter access rather than immediately reducing street sleeping. This pattern could arise because new beds attract individuals who were previously "hidden" in doubled-up or precarious living arrangements and thus never counted in the unsheltered PIT enumeration (Byrne et al., 2013; O'Flaherty, 2019; Culhane et al., 2011).

Second, federal dollars flow rapidly into all shelter modalities, but PSH displays the largest per-dollar expansion. While emergency and transitional beds respond within the same year, PSH capacity increases even more for each $1,000 of funding—consistent with HUD's "Housing First" emphasis and a strong evidence base demonstrating PSH's impacts on stability for chronically homeless populations (Tsemberis and Eisenberg, 2000; Rog et al., 2014; Byrne et al., 2014). This finding indicates that CoCs are effectively channeling federal assistance into long-term housing

infrastructure, even as they retain the ability to scale up emergency shelter in response to immediate crises.

Third, heterogeneity analysis reveals that the shelter response to increased funding is concentrated among specific demographic groups, with working-age adults, men, and black individuals accounting for the majority of new shelter placements. This demographic composition represents a notable shift from earlier research by Popov (2016) and Lucas (2017), who found that families and children were most responsive to similar funding increases. This shift could reflect a reorientation of funding and program design toward adult-focused services in recent years, or differing demographic changes as a result of the more spatially fine crosswalk process to obtain relevant covariates for the specifications.

Fourth, we uncover stark racial disparities in unsheltered outcomes. Although both white and black individuals gain shelter access as funding rises, the net effect on street homelessness diverges: white individuals experience significant declines in unsheltered rates, whereas black individuals see significant increases. This finding is particularly concerning given overall declines in black homelessness reported between 2015 and 2019, and suggests that structural barriers—such as discriminatory intake practices, mismatches in shelter location, or uneven outreach—may prevent equitable utilization of new resources.

These empirical findings align with a few interlocking theoretical mechanisms. Under a standard capacity-constraint framework, CoC and ESG grants immediately loosen local bottlenecks in emergency and transitional shelter supply, creating "fast-stock" beds that can be stood up with minimal lead time (e.g. leasing contracts, temporary cots) (O'Flaherty, 1995). However, our finding that each dollar of funding yields even larger gains in PSH capacity suggests that, beyond sheer speed of deployment, federal funding formulas and local CoC priorities tilt investments toward long-term housing solutions. PSH development does indeed carry longer lead times—site acquisition, capital financing, and administrative approvals all take months or years—but federal regulations (e.g. CoC grant spending restrictions) channel a larger share of marginal dollars into PSH relative to emergency shelter space (Rog et al., 2014; Byrne et al., 2014). The result is both a

rapid surge in temporary shelter and an outsized expansion of the PSH stock from larger funding allocations across CoCs.

Geographic spillovers further shape these findings. While the coefficient on the nearby funding ratio is not significantly different from zero in most specifications, our strong negative result for unsheltered homelessness in urban settings suggests that individuals experiencing street homelessness in more densely populated areas relocate toward jurisdictions with greater available resources, influencing counts in less advantaged communities while inflating sheltered populations in well-funded CoCs (Popov, 2016). Even households with marginal housing stability (e.g. doubling up with family or friends) may choose to migrate into better-funded CoCs to secure more reliable shelter placements, again boosting local sheltered counts without a one-for-one conversion of the truly unsheltered. Nonetheless, we show that, in the aggregate, cross-CoC funding differences lead to strong compositional changes in sheltered homeless populations even after accounting for possible spillover incentives.

Finally, by expanding the capacity of formal programs, marginal federal dollars can "reveal" a hidden population—households in precarious but roofed situations who opt into shelter when it becomes available (Byrne et al., 2013; O'Flaherty, 2019). These individuals were never captured in the unsheltered Point-in-Time enumeration, so their entry into shelter programs raises reported sheltered homelessness without generating an offsetting drop in recorded street homelessness. In effect, the outside option of doubling up, overcrowding, or couch-surfing becomes dominated by the newly created alternative option of program-funded shelter, muting any immediate decline in unsheltered counts.

Recent evidence from Christopher et al. (2025) provides a useful contrast to our findings. Studying Los Angeles County's winter shelter program (2014-2019), they find that temporary expansions of 1,000-1,500 beds directly reduced unsheltered homelessness, with approximately 90 people sheltered per 100 beds added. Their high-frequency, within-city design captures immediate shelter uptake during winter surges, whereas our national cross-section reflects steady-state funding differences across diverse CoC types. The difference in results suggests that targeted,

short-term shelter expansions can directly convert street homelessness into shelter occupancy, while sustained federal funding increases primarily expand access for marginally housed populations without commensurate unsheltered declines. This reinforces our interpretation that CoC and ESG grants draw in individuals from precarious housing rather than directly reducing visible street homelessness.

Despite these contributions, certain limitations should be noted. First, PIT and HIC data are subject to measurement error. PIT counts notoriously undercount street-sleeping individuals—particularly those in concealed or rural encampments—while HIC counts may overstate usable capacity if providers lack sufficient staffing or operational resources (National Law Center on Homelessness and Poverty, 2017; Schneider et al., 2016; Government Accountability Office, 2020; Meyer et al., 2023). Hard-to-locate groups such as people sleeping in vehicles, abandoned buildings, rural encampments, or concealed urban niches are often missed, and the same is true for women, youth, and LGBTQ+ individuals who avoid public spaces (National Law Center on Homelessness and Poverty, 2017; Schneider et al., 2016). Furthermore, because the PIT is a snapshot, administrative micro-panels suggest annual incidence can exceed the single-night count by a factor of three or more (Ward et al., 2024). We mitigate these concerns with extensive covariate adjustment and robustness checks, but readers should interpret our coefficients as effects on reported homelessness rather than the unobserved, underlying population.

Second, by focusing on contemporaneous and average (between) effects, we do not fully capture longer-run lags in PSH development or delayed reductions in unsheltered homelessness. PSH projects often require many years to materialize from award to occupancy; since our time-frame ends in 2019, we cannot observe these extended timelines (Rog et al., 2014; Byrne et al., 2014). Future research should extend the time horizon through at least 2024 to estimate the long-run evolution of PSH capacity and its impacts on chronic homelessness.

Third, related to timing, complications in the PIT counts conducted over the COVID-19 pandemic[11] in the U.S. do not allow for us to examine the relationship over that period. Our results are valid for the 2015–2019 timeframe, during which both homelessness measurement and federal funding allocation followed consistent methodologies. The pandemic period introduced substantial disruptions to both outcomes and treatment variables: emergency COVID-19 relief dramatically increased ESG funding by over 15 times the previous yearly average[12] (U.S. Department of Housing and Urban Development, 2020), while PIT count reliability deteriorated due to modified or dropped enumeration procedures and health restrictions. Consequently, we cannot determine whether the estimated relationships between federal funding and homelessness outcomes persist in the post-pandemic environment, where both the scale of resources and the nature of service delivery have evolved considerably. Future research extending this analysis to the post-2020 period will require careful attention to these structural breaks in both measurement and policy regimes.

Lastly, an important limitation is that we observe only federal CoC and ESG allocations, not the full portfolio of resources deployed against homelessness in each community. If federal grants crowd out local contributions or private philanthropy, our estimates overstate the net impact of federal policy. Conversely, if federal dollars leverage additional state and local matching funds—as ESG explicitly requires—our estimates may understate the total resource effect.

Nonetheless, our findings carry three key implications for policy and theory. First, the rapid surge in emergency and overflow beds following increases in CoC and ESG funding demonstrates that annual grants are effective tools for scaling up immediate, need-based responses. To translate these short-run gains into durable reductions in chronic homelessness, policymakers should embed multi-year commitments—such as earmarking a fixed share of CoC funds for PSH capital projects over a rolling three- to five-year horizon—to align incentives toward long-term infrastructure growth (Rog et al., 2014). Second, our subgroup results underscore that the most responsive beneficiaries-black individuals, men, adults over 24, and those suffering from serious behavioral-

[11] Due to health and safety concerns, many CoCs were unable to conduct their annual PIT counts in 2021, with HUD allowing jurisdictions not submit unsheltered counts if the count would conflict with local lockdown restrictions.

[12] Author calculations from ESG award data.

health conditions-face acute vulnerability and may derive outsized welfare gains from targeted expansions of specialized beds. Incorporating welfare weights or cost-benefit comparisons across subpopulations could sharpen the case for prioritizing these groups. Third, the racial divergence in unsheltered outcomes highlights the need for equity-focused policy design: funding formulas and program rules must be assessed for potential barriers that may increase the number of black individuals who experience unsheltered homelessness, even as white unsheltered counts decline.

Additionally, our subgroup analysis highlights the importance of heterogeneity within the homeless population—women, children, people with chronic spells, and people suffering from serious conditions—who are most responsive to newly available beds. Targeting additional capacity toward these groups may yield greater welfare gains than a uniform expansion of emergency beds.

2.7.1 Racial Disparities in Unsheltered Homelessness Response

A particularly concerning finding from Table 2.10 is the divergent racial response to increased federal funding in unsheltered populations. While white individuals experience a statistically significant decrease of 0.020 persons per 10,000 for each $1,000 increase in funding ($p < 0.05$), Black individuals experience a significant increase of 0.015 persons per 10,000 ($p < 0.01$). This pattern emerges despite both groups showing positive shelter-entry responses in Table 2.9, suggesting that the racial disparity operates through channels beyond simple differences in shelter access. We propose two potential explanations, though our cross-sectional design prevents definitive causal identification among these mechanisms. First, increased funding may improve Point-in-Time enumeration quality, and if historical PIT counts systematically undercounted Black individuals at higher rates (Schneider et al., 2016; National Law Center on Homelessness and Poverty, 2017; Government Accountability Office, 2020), then funding increases would mechanically raise measured Black unsheltered counts without reflecting true increases in the underlying population.

Second, geographic mobility combined with shelter capacity constraints may generate this pattern. Black individuals may migrate toward better-funded CoCs in response to expanded service

availability, successfully entering shelter programs at higher rates than white individuals. However, if these same CoCs face binding capacity constraints or if new shelter beds are geographically concentrated in areas less accessible to Black migrants, some portion of the Black inflow may appear in rising unsheltered counts even as Black shelter utilization increases. This mechanism would reconcile the simultaneous increases in both sheltered and unsheltered Black homelessness with differential mobility patterns across racial groups (Byrne et al., 2013; Popov, 2016).

Distinguishing among these mechanisms requires individual-level panel data linking residential histories, program participation, and enumeration protocols, which are unavailable at the national scale. Nonetheless, the stark racial divergence underscores the critical importance of equity considerations in federal homeless-assistance policy. If improved funding primarily reduces street homelessness for white individuals while leaving Black individuals behind—or if measurement improvements reveal previously invisible disparities in homelessness prevalence—the distributional consequences of current grant formulas warrant serious policy attention (Tsai and Alarcón, 2022; Hopper et al., 2008).

2.7.2 The Role of Permanent Supportive Housing

An important consideration in interpreting our results is the substantial PSH expansion documented in Table 2.5: each $1,000 funding increase generates 0.131 PSH beds per 10,000 residents—the largest capacity response across all bed types. Because PSH residents are not counted as homeless in PIT enumerations (Corinth, 2017; Rog et al., 2014), our estimated shelter effects represent net impacts after PSH exits.

This has two implications. First, absent PSH investments, emergency shelter capacity might have expanded more than the 0.087 beds per $1,000 we observe. Second, PSH absorbs substantial resources ($13,000-$20,000 per person annually (Culhane et al., 2002; Perlman and Parvensky, 2006; Culhane et al., 2025)), meaning our cost-per-bed estimates ($4,587 per bed, $10,000 per person sheltered) reflect portfolio allocation across program types rather than pure emergency shelter productivity. Using Corinth (2017)'s bound of 0.72 persons housed per PSH bed, our 13.1 beds

created per $100,000 could plausibly house 9-10 individuals who would otherwise appear in PIT counts.

These considerations reinforce our core finding: federal grants rapidly expand formal service capacity, with PSH absorbing the largest marginal share despite longer development timelines. The puzzle of why unsheltered counts do not decline persists even accounting for PSH, as individuals entering PSH must come from somewhere—either shelter, streets, or the "hidden homeless" population. Observing shelter increases, PSH increases, and no unsheltered decreases suggests the primary response margin remains marginally housed populations opting into services when capacity expands.

2.7.3 Reconciling Results with Prior Studies

Table 2.11 systematically compares our main results with those of Popov (2016) and Lucas (2017). Direct numerical comparison is complicated by the fact that Popov reports effects on total homeless counts per CoC while Lucas and our study report effects per 10,000 residents. Because Popov does not report mean CoC population, precise unit conversion is infeasible. However, the qualitative patterns remain clear and interpretable across all three studies.

First, regarding the direction of effects, all three studies find significant positive effects on sheltered homelessness. For unsheltered homelessness, however, the results diverge sharply: Popov (2016) finds a significant reduction (-0.46, $p < 0.05$), Lucas (2017) finds small, statistically insignificant effects (0.09–0.127), and we find a null effect (-0.00). Second, the relative magnitudes within studies are consistent: in all three analyses, the sheltered effect is larger in absolute value than the unsheltered effect, indicating that funding primarily expands formal shelter utilization rather than reducing street homelessness. Third, a clear temporal pattern emerges: effect sizes appear to diminish over time. Popov's 2011 estimates show the largest impacts (both positive for sheltered and negative for unsheltered), Lucas's 2013–2015 estimates show intermediate effects with declining unsheltered impacts, and our 2019 estimates show the smallest sheltered effects and null unsheltered effects.

We argue that this temporal pattern is most plausibly explained by the phase-out of the Homelessness Prevention and Rapid Re-Housing Program (HPRP). This $1.5 billion American Recovery and Reinvestment Act–funded initiative, created in response to the Great Recession, distributed emergency assistance in 2009 with the requirement that all funds be spent by 2012. Like CoC and ESG grants, HPRP allocations were determined by the CDBG formula and distributed to cities, counties, and states, though neither Popov (2016) nor Lucas (2017) explicitly discuss this program in their analyses.

The magnitude of HPRP was substantial relative to the programs we study. In 2011, the year of Popov (2016)'s analysis, HPRP spending was at or near peak levels while CoC grants totaled $1.67 billion and ESG grants $250 million—making HPRP roughly 78 percent the size of CoC grants alone and more than six times the size of ESG. By 2013, when Lucas (2017)'s primary cross-section was collected, HPRP funds were nearly exhausted. By contrast, at the time of our main 2019 study period, HPRP had been completely absent for at least seven years. Critically, unlike CoC and ESG programs that primarily fund shelter operations and permanent supportive housing, HPRP explicitly targeted homelessness prevention and rapid re-housing—interventions potentially more effective at reducing unsheltered counts by keeping at-risk individuals from losing housing in the first place.

This interpretation has important policy implications. The temporal pattern suggests that large-scale homelessness prevention efforts—such as HPRP's focus on rental assistance and rapid re-housing for at-risk households—may be more effective at reducing unsheltered homelessness than shelter expansion alone. Culhane et al. (2011) advocate precisely this prevention-centered paradigm, arguing that intervening before housing loss occurs is both more humane and more cost-effective than managing chronic homelessness through emergency shelter systems. Our results thus represent the steady-state effectiveness of ongoing federal homeless-assistance programs, which is arguably more policy-relevant for understanding long-run program efficacy than estimates that conflate sustained funding with one-time emergency interventions. While CoC and ESG grants successfully expand shelter capacity and increase formal shelter utilization, additional

policy mechanisms, such as targeted prevention subsidies or expanded rapid re-housing vouchers, may be necessary to achieve meaningful reductions in street homelessness.

2.7.4 Cost Per Bed and Per Person Sheltered

To assess the plausibility of our estimates relative to prior work and industry benchmarks, we conduct a back-of-the-envelope calculation comparing our implied costs per bed created and per person sheltered with documented program expenditures. From Table 2.5, a $1,000 increase in per-capita federal funding produces 0.087 additional emergency, transitional, and safe-haven beds per 10,000 residents and 0.131 additional permanent supportive housing (PSH) beds per 10,000 residents. Scaling to $100,000 for direct comparison with Popov (2016), we obtain 21.8 total beds across all programs, implying a cost per bed created of $4,587 per year. From our main results table, the same $100,000 funding increase moves 10.0 individuals into sheltered status, implying a cost per person sheltered of $10,000 per year. By contrast, Popov (2016) reports that $100,000 supports 152 additional beds and shelters 46 people, implying a cost per bed of only $658—implausibly low relative to documented shelter operating costs. Industry benchmarks reveal that Culhane and An (2021) estimate average costs of $21,250 per bed per year for nonprofit temporary housing, Khadduri et al. (2010) report emergency-shelter costs ranging from $1,634 to over $20,000 per person annually depending on service intensity, and Culhane et al. (2002), Perlman and Parvensky (2006), Mondello et al. (2007), and Culhane et al. (2025) report PSH costs between $13,000 and $20,115 per person per year.

Our estimate of $4,587 per bed created appears plausible when federal CoC and ESG grants are leveraged with state, local, and philanthropic funds at typical cost-sharing ratios of 40 to 50 percent federal contribution, implying total operating costs of $9,000 to $11,500 per bed—well within the documented range. Additionally, not all grant funds flow directly to bed creation; CoC and ESG awards support HMIS infrastructure, coordinated-entry systems, street outreach, homelessness prevention, and administrative overhead, so the marginal federal dollar finances a portfolio of activities beyond physical bed capacity. The estimate from Popov (2016) of $658

per bed cannot be reconciled with industry data unless one assumes implausibly high leveraging ratios (15:1 or greater). As discussed in Section 5.8, the 2011 analysis from Popov (2016) coincides with peak HPRP spending, suggesting the instrument captures the joint effect of CoC, ESG, and prevention-focused HPRP funding, artificially inflating per-dollar productivity. Lucas (2017) does not estimate bed-creation effects directly but notes that Corinth (2017) bounds PSH impact at 0.72 homeless persons per bed and combines this with $13,000 to $17,000 PSH costs to derive a minimum $18,000 per one-person reduction in homelessness. This plausibility check strengthens confidence in our main finding: federal CoC and ESG grants expand shelter capacity and increase reported sheltered homelessness in the short run, but the magnitude of these effects is smaller (and more consistent with administration program costs) than earlier estimates.

2.7.5 Conclusion

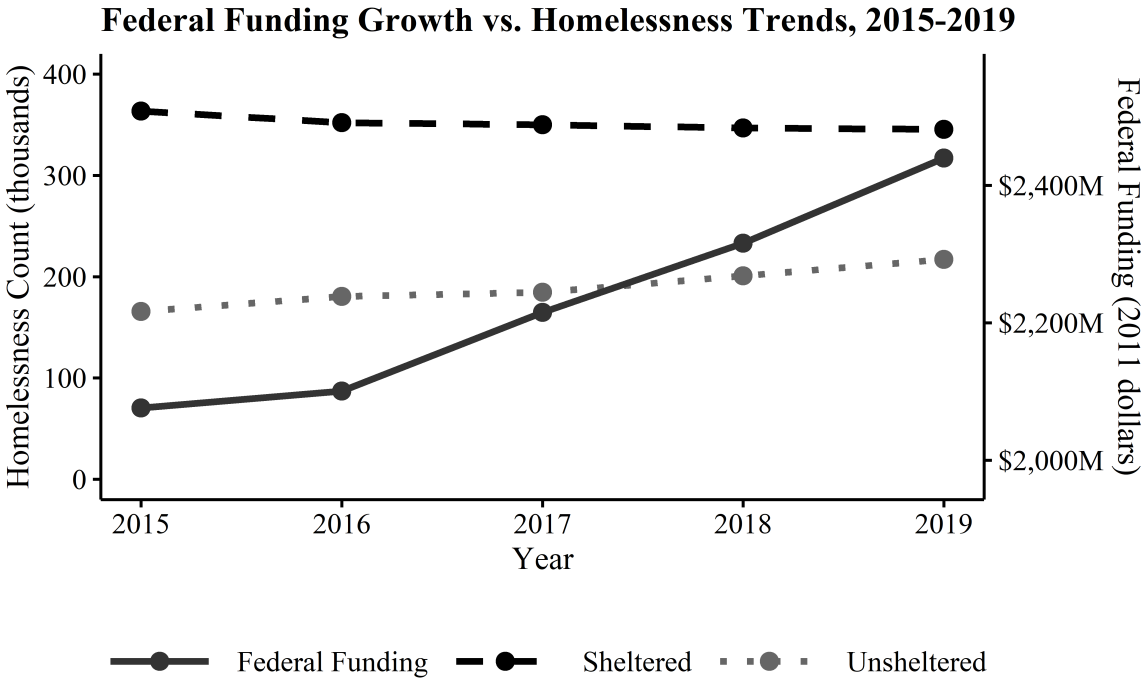
In summary, this paper advances understanding of how federal homeless-assistance grants translate into reported homelessness, shelter capacity, and the composition of beneficiaries across U.S. CoCs. Our 2SLS snapshot shows that CoC and ESG grants expand immediate shelter access by drawing into the system households who were previously doubled-up or otherwise hidden, without producing any detectable short-run decline in street homelessness.

Simultaneously, CoCs with higher per-capita funding in 2019 exhibit larger stocks of permanent supportive housing, underscoring the role of sustained resource levels in building infrastructure suitable for long-term reductions in homelessness. Future work should refine measurement of unsheltered counts, explore the equity of access across demographic groups, and evaluate how one-off funding shocks compare to ongoing allocations in shaping durable reductions in various heterogeneous groups of people experiencing homelessness. Only by coupling rapid-response shelter funding with durable PSH commitments—and by ensuring those resources reach the most vulnerable—can federal grants fulfill their full promise of both immediate relief and lasting stability.

While the federal government distributes sizeable funding toward homeless-assistance services, the nature of the relationship between this funding and the composition of the homeless population over time remains only partly understood. Our estimates note that for $1,000,000 allocated to a given CoC, about 100 more people are counted as staying in an emergency overnight shelter, transitional housing unit, or safe haven bed space. Heterogeneity analysis reveals that these new placements are more often family members rather than single individuals. In fact, of those 100 people, about 31 of them are under the age of 18 while 61 are older than 24, 44 are women and 54 are men, 38 are white and 55 are black, 18 have been counted as homeless for at least one year already, and between 22 and 49 of those people suffer from a severe condition. However, the origin of these 100 people remains mostly unclear, since there is no robustly detectable relationship between federal funding and unsheltered homelessness counts. Overall, our findings suggest that larger federal grant allocations across CoCs predominantly expand shelter capacity without immediately reducing street homelessness, underscoring the need for complementary strategies (such as targeted outreach and long-term supportive housing) to reach unsheltered individuals and address chronic homelessness.

2.8 Tables and Figures

Figure 2.1: Growth in Federally Allocated Funding for Homelessness Assistance vs. Counts of Sheltered and Unsheltered Homelessness in the U.S., 2015-2019, Authors' Calculations



Source: HUD Point-in-Time Counts and Federal CoC/ESG Grant Allocations

Table 2.1: Summary Statistics: Homelessness and Housing Inventory, 2015-2019 Averages

Variable	N	Mean	Std. Dev.	Min	Max
Overall Homelessness Counts (per 10,000 residents)					
Total Homeless	1,840	17.61	18.61	1.23	205.71
Sheltered	1,840	11.37	12.54	0.13	160.89
Unsheltered	1,840	6.24	11.72	0.00	123.69
Sheltered Population by Demographics (per 10,000 residents)					
Under 18	1,840	2.97	4.44	0.00	65.11
Age 18–24	1,840	0.90	1.10	0.00	13.69
Over 24	1,840	7.49	7.99	0.02	135.52
Women	1,840	4.89	5.76	0.00	65.65
Men	1,840	6.44	7.10	0.07	99.15
White	1,840	6.06	7.18	0.05	144.61
Black	1,840	4.23	7.55	0.00	115.83
Chronically Homeless	1,840	1.31	1.93	0.00	22.00
Individuals	1,840	6.64	6.86	0.00	115.64
People in Families	1,840	4.73	7.70	0.00	108.66
Mental Illness[a]	1,792	2.06	2.55	0.00	28.08
Substance Abuse	1,792	1.83	2.75	0.00	28.47
Domestic Violence	1,790	1.29	1.42	0.00	14.46
Unsheltered Population by Demographics (per 10,000 residents)					
Under 18	1,840	0.38	1.11	0.00	11.52
Age 18–24	1,840	0.60	1.83	0.00	25.20
Over 24	1,840	5.26	9.81	0.00	120.25
Women	1,840	1.86	3.49	0.00	39.82
Men	1,840	4.34	8.37	0.00	95.45
White	1,840	4.17	8.79	0.00	102.92
Black	1,840	1.21	2.44	0.00	34.47
Chronically Homeless	1,840	1.97	3.72	0.00	40.42
Individuals	1,840	5.65	10.84	0.00	122.43
People in Families	1,840	0.60	1.96	0.00	34.81
Mental Illness	1,792	1.58	3.44	0.00	37.83
Substance Abuse	1,792	1.33	2.79	0.00	33.26
Domestic Violence	1,790	0.73	2.30	0.00	38.22
Housing Inventory (beds per 10,000 residents)					
Emergency Shelter, Transitional & Safe Haven	1,840	12.33	12.96	0.00	181.90
Permanent Supportive Housing	1,840	11.17	12.63	0.00	125.28
Seasonal Beds	1,840	0.92	1.58	0.00	13.74
Overflow Beds	1,840	0.86	1.91	0.00	22.37

[a]Counts for Mental Illness, Substance Abuse, and Domestic Violence were extracted via OCR from image-based tables in HUD's CoC reports. Some CoCs did not report these categories in all years, resulting in a small percentage of missing observations.

Table 2.2: Summary Statistics: Federal Funding and Community Covariates, 2015-2019 Averages

Variable	N	Mean	Std. Dev.	Min	Max
Federal Funding (thousands of dollars per 10,000 residents)					
CoC Grant ($1000s)	1,840	64.72	69.99	0.15	555.20
Emergency Solutions Grant ($1000s)	1,840	5.49	7.51	0.00	62.30
Total Federal Funding ($1000s)	1,840	71.17	75.28	0.15	573.06
Housing and Demographic Characteristics					
CoC Total Population (divided by 10,000)	1,840	84.09	119.69	3.33	1,090.29
CoC Total Population in Poverty (divided by 10,000)	1,840	11.98	19.28	0.34	188.32
Pre-1940's Housing Unit Share (% of total housing units)	1,840	13.64	11.60	0.20	64.75
Black Population Share (%)	1,840	12.31	12.66	0.42	80.17
Housing Vacancy Rate (%)	1,840	12.49	6.83	3.33	48.66
Overcrowding Rate (%)	1,840	0.82	0.74	0.06	5.48
Rent Burden Rate (%)	1,840	13.57	4.49	4.61	33.91
Socioeconomic Characteristics					
Poverty Rate (%)	1,840	14.07	4.49	3.38	39.77
SNAP Recipients Share (%)	1,840	12.48	4.86	2.52	42.79
SSI Recipients Share (%)	1,840	5.44	1.93	1.41	15.79
College Education Rate (%)	1,840	66.11	22.62	8.15	88.11
Unemployment Rate (%)	1,840	3.35	0.97	1.41	10.30
Health and Special Populations					
Veteran Population Rate (%)	1,840	79.73	37.00	1.58	99.74
Disabled Population Share (%)	1,840	20.47	4.70	5.76	34.18
Uninsured Population Share (%)	1,840	25.31	11.26	5.02	80.86
Population Over 65 Share (%)	1,840	25.45	12.05	7.59	45.65
Geographic and Climate Controls					
Nearby Funding Ratio (Nearby CoC Funding / Own, $1000s)	1,840	21.36	102.74	0.00	2,116.89
January Average Temperature (°C)	1,840	2.09	6.91	−16.73	21.01
January Precipitation (inches)	1,840	2.76	2.19	0.00	21.02

Table 2.3: First Stage: Effect of Pre-1940 Housing Share on Federal Funding

	Dependent Variable: Total Federal Funding (per 10,000 residents)			
	Cross-Section, 2019		Between CoC, 2015-2019	
	(1)	(2)	(3)	(4)
Instrument				
Old Housing Share (pre-1940, %)	4.332***	4.340***	3.852***	3.961***
	(0.291)	(0.364)	(0.264)	(0.366)
Black Population Share (%)		0.672**		0.981***
		(0.273)		(0.256)
Population Over 65 Share (%)		−2.273**		−2.960*
		(1.098)		(1.559)
SNAP Recipients Share (%)		−0.111		0.325
		(1.227)		(1.166)
SSI Recipients Share (%)		4.852*		−0.427
		(2.839)		(2.416)
Poverty Rate (%)		0.429		−0.355
		(1.432)		(1.231)
College Education Rate (%)		5.346***		10.685***
		(0.572)		(2.448)
Unemployment Rate (%)		15.436***		7.426
		(5.796)		(4.865)
Vacancy Rate (%)		0.225		−0.359
		(0.496)		(0.456)
Overcrowding Rate (%)		12.894***		2.220
		(4.945)		(5.053)
Rent Burden Rate (%)		0.576		2.924**
		(1.163)		(1.151)
Veteran Population Rate (%)		−0.725		−6.775***
		(1.990)		(1.839)
Disabled Population Share (%)		5.352**		9.145**
		(2.330)		(3.893)
Nearby Funding Ratio		−0.067**		−0.031
		(0.026)		(0.025)
January Average Temperature (°C)		2.205***		1.916***
		(0.589)		(0.621)
January Precipitation (inches)		−4.752***		0.995
		(1.744)		(1.841)
Constant	16.415***	−217.486***	16.666***	−338.998***
	(5.112)	(30.186)	(4.727)	(113.872)
Model Specifications				
Observations	370	370	373	373
R^2	0.375	0.688	0.366	0.639
F-Statistic	220.898***	48.630***	213.743***	39.318***

Notes: Robust standard errors clustered at CoC level in parentheses.
Cross-section specifications use 2019 data. Between specifications
use 2015–2019 averages. *p<0.1; **p<0.05; ***p<0.01. All regressions are unweighted.

Table 2.4: Main Results: Effect of Federal Funding on Sheltered and Unsheltered Homelessness

	Dependent Variables (per 10,000 residents)					
	Sheltered			**Unsheltered**		
	2019 (No Covs)	2019 (Full)	2015-2019 Between	2019 (No Covs)	2019 (Full)	2015-2019 Between
	(1)	(2)	(3)	(4)	(5)	(6)
Total Federal Funding ($1000s)	0.123***	0.099***	0.092***	−0.034**	−0.001	−0.004
	(0.011)	(0.017)	(0.019)	(0.014)	(0.015)	(0.016)
Black Pop. Share (%)		−0.103*	−0.097*		−0.342***	−0.266***
		(0.056)	(0.056)		(0.051)	(0.047)
Pop. Over 65 (%)		0.003	−0.098		−0.233	−0.464*
		(0.218)	(0.325)		(0.199)	(0.272)
SNAP Recipients (%)		0.585**	0.719***		−0.814***	−0.606***
		(0.247)	(0.242)		(0.225)	(0.203)
SSI Recipients (%)		−0.239	0.063		0.452	0.315
		(0.591)	(0.495)		(0.540)	(0.414)
Poverty Rate (%)		−0.484*	−0.601**		0.473*	0.432**
		(0.287)	(0.256)		(0.262)	(0.214)
College Educ. (%)		−0.189	−0.149		0.081	−0.258
		(0.158)	(0.555)		(0.144)	(0.464)
Unemployment (%)		−2.905**	−2.663***		6.164***	3.128***
		(1.173)	(1.002)		(1.070)	(0.839)
Vacancy Rate (%)		0.103	0.228**		0.238***	0.246***
		(0.101)	(0.094)		(0.092)	(0.079)
Overcrowding (%)		0.145	−0.404		5.580***	3.287***
		(1.026)	(1.047)		(0.936)	(0.876)
Rent Burden (%)		0.919***	1.031***		0.144	0.075
		(0.238)	(0.256)		(0.217)	(0.215)
Veteran Pop. (%)		0.375	0.134		0.982***	0.190
		(0.404)	(0.406)		(0.369)	(0.340)
Disabled Pop. (%)		−0.101	−0.482		0.063	−0.799
		(0.487)	(0.801)		(0.445)	(0.671)
Nearby Ratio		0.009*	0.008		0.001	−0.002
		(0.005)	(0.005)		(0.005)	(0.004)
Jan. Temp. (°C)		−0.138	−0.045		0.207**	0.260***
		(0.100)	(0.110)		(0.091)	(0.092)
Jan. Precip. (in.)		0.377	−0.043		1.929***	2.425***
		(0.359)	(0.378)		(0.327)	(0.316)
Constant	1.680*	1.753	9.191	9.499***	−25.437***	15.108
	(0.952)	(7.713)	(24.048)	(1.218)	(7.038)	(20.135)
Observations	370	370	373	370	370	373
R^2	0.256	0.400	0.433	−0.105	0.547	0.545

Notes: 2SLS with pre-1940 housing share as instrument. Robust SEs clustered at CoC level.
2019 = Cross-section 2019 data. Between = 2015–2019 averages. *p<0.1; **p<0.05; ***p<0.01.

Table 2.5: Housing Inventory Results: Effect of Federal Funding on Bed Capacity

	Dependent Variables (beds per 10,000 residents)			
	Total Shelter Beds		Permanent Supportive Housing	
	2019	2015-2019	2019	2015-2019
	(1)	(2)	(3)	(4)
Total Federal Funding ($1000s)	0.087***	0.085***	0.131***	0.111***
	(0.018)	(0.020)	(0.013)	(0.013)
Black Population Share (%)	−0.083	−0.053	−0.016	0.020
	(0.061)	(0.060)	(0.043)	(0.039)
Population Over 65 Share (%)	−0.036	−0.028	−0.079	−0.175
	(0.238)	(0.325)	(0.165)	(0.210)
SNAP Recipients Share (%)	0.583**	0.715***	−0.104	0.066
	(0.269)	(0.260)	(0.187)	(0.168)
SSI Recipients Share (%)	−0.137	−0.096	−0.515	−0.460
	(0.644)	(0.530)	(0.448)	(0.341)
Poverty Rate (%)	−0.366	−0.351	0.292	0.288
	(0.313)	(0.274)	(0.217)	(0.177)
College Education Rate (%)	−0.043	−0.052	0.268**	0.336
	(0.172)	(0.595)	(0.119)	(0.383)
Unemployment Rate (%)	−2.683**	−3.247***	1.275	0.354
	(1.277)	(1.078)	(0.888)	(0.695)
Vacancy Rate (%)	0.110	0.210**	0.017	0.019
	(0.110)	(0.101)	(0.076)	(0.065)
Overcrowding Rate (%)	1.056	−0.157	1.155	0.849
	(1.117)	(1.126)	(0.777)	(0.726)
Rent Burden Rate (%)	0.868***	0.973***	0.210	0.177
	(0.259)	(0.276)	(0.180)	(0.178)
Veteran Population Rate (%)	0.564	0.071	0.542*	−0.161
	(0.440)	(0.434)	(0.306)	(0.280)
Disabled Population Share (%)	0.174	−0.357	0.600	−0.504
	(0.530)	(0.844)	(0.369)	(0.544)
Nearby Funding Ratio	0.008	0.009	0.0005	0.0001
	(0.006)	(0.005)	(0.004)	(0.004)
January Average Temperature	−0.064	0.030	−0.084	−0.121
	(0.109)	(0.118)	(0.075)	(0.076)
January Precipitation	0.012	−0.231	−0.403	−0.101
	(0.390)	(0.406)	(0.271)	(0.262)
Instrumental Variable	Pre-1940 Housing Share (%)			
Summary Statistics				
Observations	370	373	370	373
R^2	0.353	0.381	0.740	0.737

Notes: All estimates from 2SLS regressions with pre-1940 housing share as instrument.
Total Shelter Beds include Emergency Shelter, Transitional Housing, and Safe Haven beds.
Robust standard errors clustered at CoC level in parentheses. Cross-section = 2019 data.
2015–2019 = Between CoC averages. *p<0.1; **p<0.05; ***p<0.01. All regressions are unweighted.

Table 2.6: Effects on Seasonal and Overflow Bed Capacity

	Dependent Variables (beds per 10,000 residents)			
	Overflow Beds		Seasonal Beds	
	2019	2015-2019	2019	2015-2019
	(1)	(2)	(3)	(4)
Total Federal Funding ($1000s)	0.009***	0.005	0.004	−0.0005
	(0.003)	(0.003)	(0.004)	(0.003)
Model Specifications				
Full Control Set	Yes	Yes	Yes	Yes
Instrumental Variable		Pre-1940 Housing Share		
Summary Statistics				
Observations	370	373	370	373
R^2	0.139	0.151	0.192	0.206

Notes: All estimates from 2SLS regressions with pre-1940 housing share as instrument.
Full control set includes all demographic, socioeconomic, housing market, special population, and geographic/climate controls as shown in previous tables.
2015-2019 = Between CoC Averages. *p<0.1; **p<0.05; ***p<0.01. All regressions are unweighted.

Table 2.7: Heterogeneous Effects by CoC Geography: Urban vs. Rural/Suburban Areas - 2019

	Dependent Variables (per 10,000 residents)			
	Urban CoCs		**Rural/Suburban CoCs**	
	Sheltered	Unsheltered	Sheltered	Unsheltered
	(1)	(2)	(3)	(4)
Total Federal Funding	0.087***	0.023	0.091***	−0.096**
($1000s)	(0.026)	(0.014)	(0.034)	(0.045)
Black Population Share (%)	0.108	−0.087	−0.143**	−0.390***
	(0.138)	(0.077)	(0.057)	(0.076)
Population Over 65 Share (%)	0.440	0.126	0.187	0.024
	(0.773)	(0.428)	(0.198)	(0.264)
SNAP Recipients Share (%)	0.661	−0.508*	0.296	−1.121***
	(0.533)	(0.295)	(0.255)	(0.340)
SSI Recipients Share (%)	−2.927**	−0.254	2.919***	2.227**
	(1.233)	(0.682)	(0.721)	(0.963)
Poverty Rate (%)	−0.689	−0.114	−0.875***	0.318
	(0.732)	(0.405)	(0.308)	(0.411)
College Education Rate (%)	−0.068	0.073	−0.200	0.207
	(0.334)	(0.185)	(0.165)	(0.221)
Unemployment Rate (%)	−2.164	2.786*	−4.276***	7.130***
	(2.870)	(1.588)	(1.089)	(1.454)
Vacancy Rate (%)	−0.563	−0.449	0.160*	0.204*
	(0.562)	(0.311)	(0.083)	(0.110)
Overcrowding Rate (%)	1.158	6.276***	−0.123	3.838**
	(2.007)	(1.110)	(1.120)	(1.495)
Rent Burden Rate (%)	1.165**	−0.273	1.015***	1.220***
	(0.522)	(0.289)	(0.336)	(0.449)
Veteran Population Rate (%)	−0.453	−0.113	0.288	1.373***
	(1.001)	(0.554)	(0.393)	(0.524)
Disabled Population Share (%)	1.755	1.426*	−0.867**	−0.318
	(1.433)	(0.793)	(0.420)	(0.561)
Nearby Funding Ratio	−0.105	−0.176***	0.009*	−0.006
	(0.112)	(0.062)	(0.005)	(0.006)
January Average Temperature (°C)	−0.586**	0.341**	0.101	−0.047
	(0.251)	(0.139)	(0.118)	(0.157)
January Precipitation (inches)	1.081	1.911***	−0.376	1.889***
	(0.947)	(0.524)	(0.338)	(0.451)
Constant	−10.873	−15.852	5.612	−40.978***
	(18.138)	(10.034)	(8.423)	(11.240)
Instrumental Variable	Pre-1940 Housing Share (%)			
Summary Statistics				
Observations	106	106	264	264
R^2	0.463	0.705	0.486	0.483

Notes: All estimates from 2SLS regressions with pre-1940 housing share as instrument. Urban CoCs include "Major City CoC" and "Other Largely Urban CoC" categories. Rural/Suburban CoCs include "Largely Rural CoC" and "Largely Suburban CoC" categories. Robust standard errors clustered at CoC level in parentheses. Cross-section = 2019 data. *p<0.1; **p<0.05; ***p<0.01. All regressions are unweighted.

Table 2.8: Federal Funding Effects by Household Type and Shelter Status - 2019

	Dependent Variables (per 10,000 residents)			
	Sheltered		**Unsheltered**	
	Individuals (1)	Families (2)	Individuals (3)	Families (4)
Total Federal Funding ($1000s)	0.047***	0.052***	0.006	−0.007***
	(0.009)	(0.012)	(0.014)	(0.002)
Black Population Share (%)	−0.053*	−0.049	−0.325***	−0.017**
	(0.029)	(0.041)	(0.047)	(0.008)
Population Over 65 Share (%)	−0.022	0.025	−0.230	−0.003
	(0.114)	(0.159)	(0.184)	(0.031)
SNAP Recipients Share (%)	−0.068	0.654***	−0.892***	0.078**
	(0.129)	(0.180)	(0.208)	(0.035)
SSI Recipients Share (%)	−0.852***	0.613	0.433	0.019
	(0.308)	(0.431)	(0.499)	(0.084)
Poverty Rate (%)	0.247	−0.731***	0.526**	−0.053
	(0.150)	(0.209)	(0.242)	(0.041)
College Education Rate (%)	0.051	−0.240**	0.046	0.034
	(0.082)	(0.115)	(0.133)	(0.022)
Unemployment Rate (%)	−0.091	−2.813***	5.898***	0.266
	(0.611)	(0.855)	(0.989)	(0.166)
Vacancy Rate (%)	0.099*	0.004	0.213**	0.025*
	(0.052)	(0.073)	(0.085)	(0.014)
Overcrowding Rate (%)	0.259	−0.113	4.988***	0.592***
	(0.534)	(0.747)	(0.865)	(0.146)
Rent Burden Rate (%)	0.408***	0.511***	0.142	0.002
	(0.124)	(0.174)	(0.201)	(0.034)
Veteran Population Rate (%)	0.318	0.057	0.937***	0.045
	(0.211)	(0.295)	(0.341)	(0.057)
Disabled Population Share (%)	0.251	−0.353	0.002	0.061
	(0.254)	(0.355)	(0.411)	(0.069)
Nearby Funding Ratio	0.004	0.006	0.001	−0.0002
	(0.003)	(0.004)	(0.005)	(0.001)
January Average Temperature	−0.070	−0.068	0.199**	0.008
	(0.052)	(0.073)	(0.084)	(0.014)
January Precipitation	0.075	0.303	1.820***	0.108**
	(0.187)	(0.261)	(0.302)	(0.051)
Constant	−6.976*	8.729	−22.672***	−2.765**
	(4.018)	(5.620)	(6.506)	(1.094)
Instrumental Variable	Pre-1940 Housing Share (%)			
Summary Statistics				
Observations	370	370	370	370
R^2	0.391	0.305	0.561	0.136

Notes: All estimates from 2SLS regressions with pre-1940 housing share as instrument. Dependent variables disaggregate homelessness by household type (individuals vs. families) and shelter status. Robust standard errors clustered at CoC level in parentheses. Cross-section = 2019 data. *p<0.1; **p<0.05; ***p<0.01. All regressions are unweighted.

Table 2.9: Heterogeneous Effects on Sheltered Population Subgroups - 2019

Panel A: Age and Gender Groups						
Dependent Variables: Sheltered Homeless by Subgroup (per 10,000 residents)						
	Under 18	Age 18–24	Over 24	Women	Men	Chronically Homeless
	(1)	(2)	(3)	(4)	(5)	(6)
Total Federal Funding ($1000s)	0.031***	0.008***	0.061***	0.044***	0.054***	0.018***
	(0.007)	(0.001)	(0.010)	(0.009)	(0.009)	(0.003)
Observations	370	370	370	370	370	370
Full Control Set Used?	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.311	0.367	0.419	0.357	0.417	0.460
Panel B: Race and Special Conditions						
Dependent Variables: Sheltered Homeless by Subgroup (per 10,000 residents)						
	White	Black	Mental Illness	Substance Abuse	Domestic Violence	
	(7)	(8)	(9)	(10)	(11)	
Total Federal Funding ($1000s)	0.038***	0.055***	0.021***	0.022***	0.006***	
	(0.009)	(0.010)	(0.004)	(0.004)	(0.002)	
Observations	370	370	359	359	359	
Full Control Set Used?	Yes	Yes	Yes	Yes	Yes	
R^2	0.396	0.425	0.400	0.323	0.201	
Model Specifications						
Instrumental Variable: Pre-1940 Housing Share						

Notes: All estimates from 2SLS regressions with pre-1940 housing share as instrument.
Robust standard errors clustered at CoC level in parentheses. Cross-section = 2019 data.
*p<0.1; **p<0.05; ***p<0.01. All regressions are unweighted.

Table 2.10: Heterogeneous Effects on Unsheltered Population Subgroups - 2019

Panel A: Age and Gender Groups
Dependent Variables: Unsheltered Homeless by Subgroup (per 10,000 residents)

	Under 18	Age 18–24	Over 24	Women	Men	Chronically Homeless
	(1)	(2)	(3)	(4)	(5)	(6)
Total Federal Funding	−0.004**	0.004**	−0.001	−0.003	0.0003	0.002
($1000s)	(0.002)	(0.002)	(0.013)	(0.005)	(0.011)	(0.006)
Observations	370	370	370	370	370	370
Full Control Set Used?	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.130	0.342	0.555	0.513	0.524	0.486

Panel B: Race and Special Conditions
Dependent Variables: Unsheltered Homeless by Subgroup (per 10,000 residents)

	White	Black	Mental Illness	Substance Abuse	Domestic Violence
	(7)	(8)	(9)	(10)	(11)
Total Federal Funding	−0.020**	0.015***	0.004	0.007	0.001
($1000s)	(0.010)	(0.005)	(0.004)	(0.004)	(0.002)
Observations	370	370	359	359	359
Full Control Set Used?	Yes	Yes	Yes	Yes	Yes
R^2	0.497	0.392	0.463	0.429	0.337

Model Specifications

Instrumental Variable: Pre-1940 Housing Share

Notes: All estimates from 2SLS regressions with pre-1940 housing share as instrument.
Robust standard errors clustered at CoC level in parentheses. Cross-section = 2019 data.
*p<0.1; **p<0.05; ***p<0.01. All regressions are unweighted.

Table 2.11: Comparison of Federal Funding Effects Across Studies

Outcome	**Popov (2016)** 2011 (Total Counts)	**Lucas (2017)** 2013 | 2015 (per 10,000)	**This Study** 2019 (per 10,000)
Panel A: Main Effects			
Total Homeless	0.73^{**}	0.32^{***} | 0.27^{***}	0.10^{***}
	(0.31)	(0.09) | (0.06)	(0.02)
Sheltered	1.19^{**}	0.18^{***} | 0.18^{***}	0.10^{***}
	(0.30)	(0.04) | (0.06)	(0.02)
Unsheltered	-0.46^{**}	0.127 | 0.09	-0.00
	(0.19)	(0.09) | (0.05)	(0.02)
Individuals	0.74^{**}	0.06^{**} | 0.13^{**}	0.05^{***}
	(0.09)	(0.02) | (0.05)	(0.01)
Families	1.38^{**}	0.12^{***} | 0.14^{**}	0.05^{***}
	(0.30)	(0.03) | (0.04)	(0.01)
Panel B: Study Characteristics			
Years	2011	2013, 2015	2015–2019
Homelessness Unit	Total Counts	Per 10,000 CoC Population	Per 10,000 CoC Population
Observations	367	374, 371	370
First-stage F	51.7	11.1	48.6
Time Since HPRP Allocation	2 Years	4-6 Years	10 Years
Panel C: Control Variables			
Crosswalk method	Undefined	County	Pop-weighted tract
Climate	No	Yes	Yes
Housing market	Yes	Yes	Yes
Demographics	Limited	Extensive	Extensive
Nearby CoC funding	No	No	Yes

Notes: Standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Popov (2016) uses total counts; Lucas (2017) and this study use rates per 10,000. Lucas reports results for multiple years; where single estimates are shown, these represent the most comparable specification to other studies.

Chapter 3

Imputing Missing 2021 Unsheltered Homelessness Counts: A Two-Phase Machine Learning Error Correction Approach

3.1 Introduction

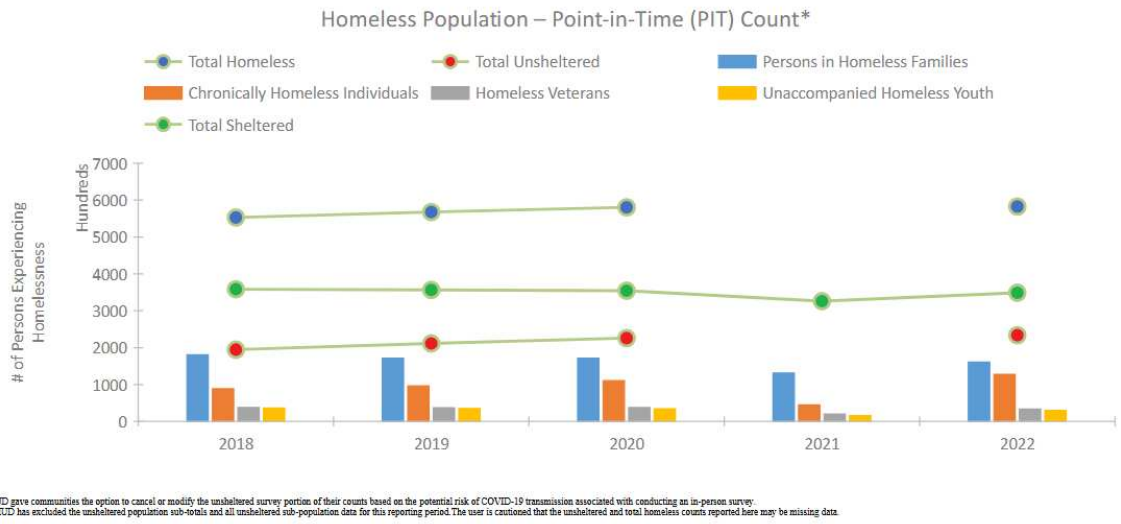
Every January, the U.S. Department of Housing and Urban Development (HUD) requires each of the nation's Continuums of Care (CoCs) to conduct a Point-in-Time (PIT) count of all people experiencing homelessness: both those residing in emergency shelters and transitional housing, as well as those sleeping in unsheltered locations such as streets, parks, and encampments. These counts form the backbone of federal homelessness policy: they produce the primary national estimate of homelessness in the United States, help inform billions of dollars in annual funding allocations through HUD's Notice of Funding Opportunity (NOFO) competition, and provide the primary data source for longitudinal research on homelessness trends (O'Flaherty, 2019; Henry et al., 2022; de Sousa et al., 2023).

In January 2021, this system broke down. As the COVID-19 pandemic entered its second year, HUD encouraged CoCs to assess whether conducting unsheltered street counts posed unacceptable public health risks (Henry et al., 2022). The response was overwhelming: 61.6 percent of CoCs (236 out of 385) either conducted only a sheltered count or performed merely a partial unsheltered enumeration. The remaining 149 CoCs that did carry out complete counts were not representative of the national landscape, skewing toward smaller, less complex communities with substantially lower baseline unsheltered populations (Henry et al., 2022). The result was the largest single-year data gap in the history of the PIT count.

The consequences of this gap are far-reaching. First, no reliable national estimate of unsheltered, and therefore total, homelessness exists for 2021. The Annual Homeless Assessment Report

(AHAR) to Congress, typically the definitive accounting of homelessness in the United States, was forced to restrict its 2021 findings entirely to sheltered populations (Henry et al., 2022). This means that the federal government's own assessment of homelessness during the most significant public health and economic crisis in a generation is fundamentally incomplete, as seen visually in the trend graph present in the HUD 2022 CoC Performance Profile Report[13] (Figure 3.1).

Figure 3.1: Missing Counts in the HUD 2022 CoC Performance Profile



Second, the missing data severely constrains researchers' ability to study how the pandemic affected homelessness. The COVID-19 period produced an unprecedented combination of economic shocks, such as mass unemployment, business closures, and housing instability. Alongside those shocks were equally unprecedented policy responses, including eviction moratoria, emergency rental assistance, CARES Act relief, and expanded shelter capacity. Whether these countervailing forces produced a net increase or decrease in unsheltered homelessness, and how those effects varied across communities, remains an open empirical question that cannot be answered with the existing data.

[13] de Sousa et al. (2023)

Third, the data gap had direct implications for federal funding. HUD's annual NOFO competition allocates approximately $2.7 billion to CoCs, with scoring criteria that historically incorporated PIT count data to assess community need and system performance (U.S. Department of Housing and Urban Development, 2021; de Sousa et al., 2023). In the absence of reliable unsheltered counts, HUD was forced to modify its evaluation framework for the FY 2021 competition, explicitly noting that "most communities could not conduct an unsheltered count in 2021 that is comparable to previous counts" and that it would evaluate only sheltered data for that year's funding decisions (U.S. Department of Housing and Urban Development, 2021). Whether this methodological shift altered the relative priority rankings of CoCs and the resulting distribution of federal resources is an empirical question with significant policy implications.

This paper addresses the 2021 data gap by developing a two-phase imputation framework that reconstructs missing unsheltered counts while explicitly accounting for the pandemic's disruption to homelessness dynamics. The core methodological challenge is straightforward: one cannot simply predict missing 2021 counts using historical relationships between community characteristics and unsheltered homelessness. The COVID-19 pandemic fundamentally altered those relationships through simultaneous, countervailing channels: economic dislocation pushed individuals toward homelessness while emergency policy responses pulled them away from it. A model trained solely on pre-pandemic data would produce estimates reflecting a world in which COVID-19 never occurred, systematically mischaracterizing what actually happened in 2021.

The framework proceeds in two phases. In the first phase, I train multiple predictive models on a panel of CoC-level data spanning 2015 through 2019, drawing on established determinants of unsheltered homelessness, including housing market conditions, economic indicators, demographic composition, climate, and homeless services infrastructure, to generate baseline predictions of what 2021 unsheltered counts would have been under pre-pandemic conditions. Out-of-sample accuracy is assessed on a locked 2020 temporal holdout before the selected model is retrained on the full panel and used to produce 2021 baseline predictions. In the second phase, I leverage the 149 CoCs that did conduct complete unsheltered counts in 2021 to model the discrepancy between

these baseline predictions and observed reality. This residual captures the net effect of COVID-19 on unsheltered homelessness, and I model it as a function of pandemic-specific factors: COVID-19 severity, emergency relief funding, and local policy responses such as eviction moratoria. Because the CoCs that counted differ systematically from those that did not, I apply overlap-based propensity score weighting to address selection bias before extrapolating the estimated COVID adjustment to non-counting CoCs. The final imputed values reflect both the long-run structural determinants of unsheltered homelessness and the unique conditions of the pandemic period, with uncertainty quantified through end-to-end clustered bootstrap prediction intervals.

This paper makes two contributions. First, it produces a complete set of 2021 unsheltered homelessness estimates for all CoCs in the United States with sufficient feature variable data available, accompanied by prediction intervals that quantify the uncertainty inherent in imputed values. These estimates restore the continuity of a national dataset essential for longitudinal policy analysis and enable, for the first time, a comprehensive accounting of unsheltered homelessness during the COVID-19 pandemic. Second, the paper demonstrates a generalizable methodological framework for imputing missing administrative data when standard assumptions of random missingness are violated by exogenous shocks. The two-phase residual correction approach of establishing a counterfactual baseline and then modeling the shock-specific deviation offers a template applicable beyond homelessness to any domain where crisis events simultaneously disrupt data collection and alter the quantity being measured.

The remainder of the paper proceeds as follows. Section 3.2 reviews the institutional background of PIT counts, the 2021 data gap, and the relevant literature on homelessness determinants, machine learning in social policy, and missing data imputation. Section 3.3 describes the data sources, sample construction, and the two-phase modeling and uncertainty quantification methodology. Section 3.4 presents results for both phases and the final national estimates. Section 3.5 discusses findings, methodological contributions, limitations, and directions for future research.

3.2 Background

3.2.1 Homelessness Data in the United States

The systematic measurement of homelessness in the United States is a relatively recent undertaking. Congress first directed the Department of Housing and Urban Development to develop an unduplicated count of the homeless population in the fiscal year 2001 appropriations act, though the first Annual Homeless Assessment Report was not produced until 2007, and the quality of its early data was admittedly limited (O'Flaherty, 2019). The AHAR has been produced annually since then under contracts with Abt Associates and the University of Pennsylvania, and it remains the most widely cited source of national homelessness statistics (Meyer et al., 2021).

The AHAR draws on three data components: a Point-in-Time count of sheltered and unsheltered homeless individuals conducted on a single night in late January, information on the characteristics of sheltered individuals over the course of the year, and a Housing Inventory Count cataloging the beds and units available to the homeless population (O'Flaherty, 2019; Meyer et al., 2021). Responsibility for collecting and reporting these data falls to the nation's approximately 385 Continuums of Care—administrative units that receive HUD funding for homeless programs and that vary widely in geographic scope, encompassing individual cities, counties, parts of states, or entire states (O'Flaherty, 2019). CoCs are required to report annually but are only mandated to count unsheltered individuals in odd-numbered years, though many conduct unsheltered counts every year.

The PIT count, while indispensable as the only national enumeration of both sheltered and unsheltered homelessness, carries well-documented limitations. The unsheltered count in particular relies on loosely supervised volunteers who canvas streets, parks, and other public spaces on a single January night (O'Flaherty, 2019). The diligence, training, and consistency of these volunteers vary considerably across CoCs, and the methodology systematically excludes individuals in locations that volunteers cannot access—restaurants, parking garages, stairwells, and other private property (O'Flaherty, 2019). People experiencing homelessness are also exceptionally difficult to enumerate for reasons that extend beyond methodology: poor mental health, substance use, lack

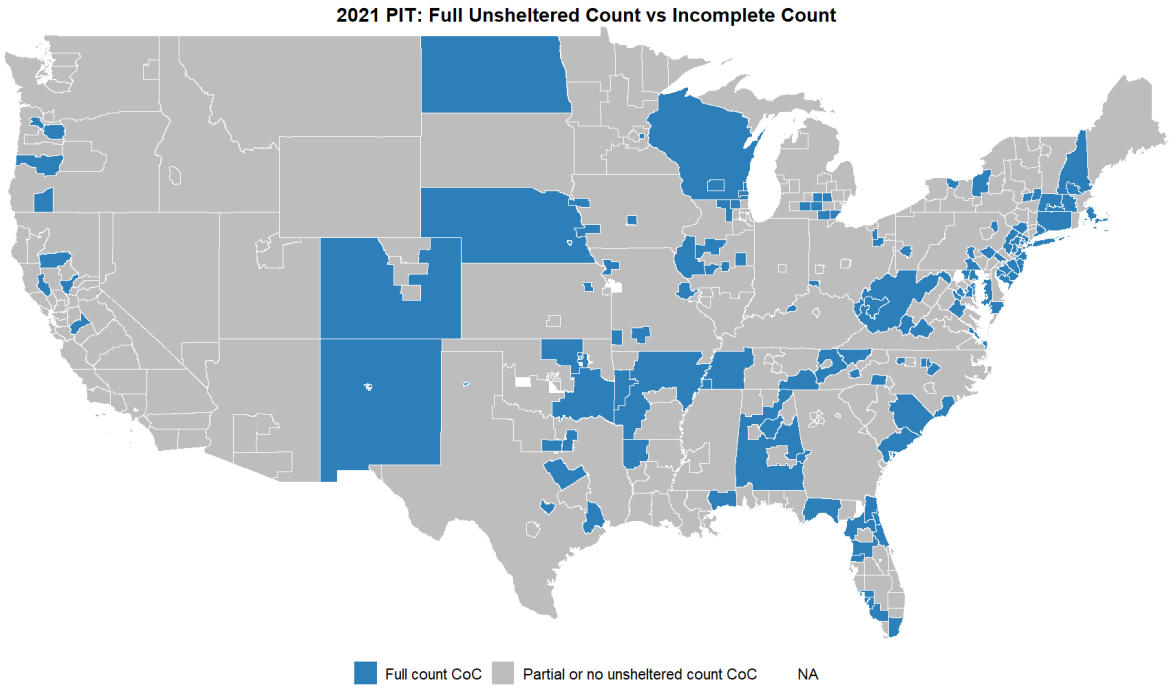
of a fixed location, and active avoidance of being found all contribute to undercounting (Meyer et al., 2021; Glasser et al., 2014). These difficulties are particularly acute for the unsheltered population, where no institutional point of contact—such as a shelter check-in—exists to facilitate enumeration.

The limitations of PIT data have important implications for research. As O'Flaherty (2019) notes, when PIT counts serve as dependent variables in regression analyses, two concerns arise: heteroskedasticity, if measurement error is systematically larger in some CoCs than others, and correlation between errors and policies of interest, if CoCs with particular characteristics tend to systematically overcount or undercount. Despite these shortcomings, the PIT count remains the only data source that captures unsheltered homelessness at a national scale, and it is the primary input into HUD's funding allocation decisions. Other national data sources—such as the Census Bureau's Special Report on the Emergency and Transitional Shelter Population (Smith et al., 2012) or the 1996 National Survey of Homeless Assistance Providers and Clients (Burt et al., 1999)—either exclude the unsheltered population entirely or are decades out of date. Localized studies drawing on shelter administrative records or CoC-level surveys offer richer detail but cannot be generalized nationally, given the substantial geographic heterogeneity in homelessness trends, housing markets, and shelter capacity across the country (Meyer et al., 2021).

3.2.2 The 2021 Data Gap

In January 2021, the COVID-19 pandemic was entering a severe winter wave. Vaccines had received emergency use authorization only weeks earlier and were not yet widely available to the general public, and communities across the country remained under varying degrees of public health restrictions. Under these conditions, conducting unsheltered PIT counts, which require mobilizing large teams of volunteers to canvas streets, parks, and encampments through the night, posed serious logistical and public health challenges. HUD recognized this reality and encouraged CoCs to evaluate whether conducting an unsheltered count would pose unacceptable risks of exac-

Figure 3.2: Map of CoCs by 2021 Unsheltered Count Status



erbating COVID-19 transmission among people experiencing homelessness, homeless assistance staff, and volunteers (Henry et al., 2022).

The majority of CoCs concluded that it would. Of the nation's 385 CoCs present in 2021, only 149 (38.4 percent) conducted a complete sheltered and unsheltered count. Another 74 CoCs performed partial unsheltered counts (enumerating some but not all unsheltered subpopulations or geographic areas) while the remaining 162 CoCs conducted sheltered-only counts, forgoing the unsheltered component entirely. Following standard practice in the literature, I classify CoCs with partial unsheltered counts alongside those with no unsheltered count, as incomplete enumerations cannot be treated as comparable to full counts. This yields 236 CoCs (61.6 percent of the national total) with missing unsheltered count data. The geographic distribution of those CoCs can be seen in Figure 3.2.

The consequences were immediate and far-reaching. The 2021 AHAR to Congress was forced to exclude unsheltered populations entirely from its national estimates, noting that the communities that did conduct counts "are not representative of all communities across the United States" (Henry

et al., 2022). This exclusion means that the only national assessment of homelessness during the most significant economic and public health crisis in decades captures, at best, half the picture. The 2022 AHAR subsequently reported a three percent increase in unsheltered homelessness between 2020 and 2022, alongside a seven percent increase in sheltered homelessness between 2021 and 2022 that likely reflected the easing of pandemic-era shelter capacity restrictions (de Sousa et al., 2023). But without reliable 2021 unsheltered data, whether the increase observed in 2022 represents a continuation, acceleration, or reversal of trends that began during the pandemic cannot be determined.

3.2.3 Research on Homelessness Count Data

The empirical study of homelessness determinants has developed in tandem with improvements in national data collection. Early work relied on limited cross-sectional data: Honig and Filer (1993) and Elliott and Krivo (1991) established that local housing supply constraints, labor market conditions, and welfare generosity predicted homelessness incidence across metropolitan areas, while O'Flaherty (1995) provided the foundational theoretical framework, modeling homelessness as arising from a mismatch between the lower tail of the income distribution and the lower tail of the housing price distribution. These studies identified the broad structural forces—housing costs, poverty, and labor market weakness—that subsequent work has consistently confirmed, but were constrained by the absence of nationally standardized count data.

The introduction of HUD's Point-in-Time counts in the mid-2000s enabled a new generation of CoC-level analyses. Byrne et al. (2013) used this data to examine community-level determinants, finding housing market variables—particularly median rent and rent burden—among the strongest predictors. Corinth and Lucas (2018) documented the role of January temperatures in shaping the geographic distribution of unsheltered homelessness. Hanratty (2017) provided panel fixed-effects estimates showing that a ten percent increase in median rents was associated with a nine percent increase in homelessness rates. On the funding side, Moulton (2013) found that permanent supportive housing investments reduced chronic homelessness more effectively than general CoC

awards, while Popov (2016) and Lucas (2017) addressed the endogeneity of funding allocations using instrumental variables strategies.

Recent methodological advances have improved homeless population measurement and prediction. Glynn and Fox (2019) develop a dynamic Bayesian model distinguishing counted from true homeless populations across 25 major metros, finding rental cost effects strongest in New York, Los Angeles, and Seattle. Nisar et al. (2019) demonstrate that housing market factors—rental costs, crowding, and evictions—most consistently predict community-level homelessness rates. Meyer et al. (2023) use linked Census-HMIS microdata to estimate approximately 400,000 sheltered and 200,000 unsheltered homeless individuals nationally, with over 90 percent of those sheltered counted in the Census though often misclassified. Chien et al. (2024) show that administrative and citizen-generated data can predict unsheltered homelessness magnitude and spatial distribution in Los Angeles between annual Point-in-Time counts. These studies demonstrate both the feasibility of improving measurement through novel data linkages and the persistent challenges of enumerating a mobile population.

While this body of work has produced robust evidence on what drives cross-sectional and temporal variation in homelessness counts, it has focused almost exclusively on estimation and inference rather than prediction and imputation. No study, to my knowledge, has addressed the problem of reconstructing missing PIT count data at a national scale, particularly in a setting where missingness is driven by an exogenous shock rather than random nonresponse. This paper adapts the predictor relationships established in the determinants literature to an imputation objective, using the same variables that prior work has shown to explain variation in homelessness counts as inputs to a predictive modeling framework designed to fill the 2021 data gap.

3.2.4 Machine Learning in Social Policy Research

Machine learning methods have seen growing adoption in social science and policy research, particularly for prediction tasks where the goal is accurate out-of-sample forecasting rather than unbiased estimation of causal parameters. In settings with large numbers of candidate predictors

and potentially nonlinear relationships, ensemble methods such as random forests and gradient boosting machines (Breiman, 2001; Friedman, 2001; Chen and Guestrin, 2016) have consistently demonstrated superior predictive performance relative to traditional linear regression (Embaye et al., 2021; Ruhnke et al., 2022; Downing, 2025). This advantage has been documented across a range of applied contexts: Embaye et al. (2021) show that boosting, bagging, and random forest methods outperform OLS in predicting rental housing values in household surveys across multiple countries and years, while Ruhnke et al. (2022) find that random forest imputation of immigration legal status in nationally representative survey data yields greater accuracy and less bias than regression-based approaches.

The present application shares key features with this literature. The prediction task involves a moderately high-dimensional set of community-level predictors with plausible nonlinearities and interactions—precisely the conditions under which tree-based ensemble methods tend to outperform linear models. At the same time, the imputation context demands more than point prediction accuracy: downstream users of the imputed data need reliable uncertainty quantification, and the predictions must generalize to observations (non-counting CoCs) that may differ systematically from the training sample. These requirements motivate the multi-model comparison strategy adopted in this paper, in which linear and machine learning approaches are evaluated side by side, and the selection bias inherent in the Phase 2 training sample is addressed explicitly.

3.2.5 Imputation Methods and Error Correction Approaches

The standard statistical framework for handling missing data is multiple imputation, formalized by Rubin (1987) and operationalized through methods such as multivariate imputation by chained equations. These approaches assume that data are missing at random—that is, conditional on observed covariates, the probability of missingness is unrelated to the missing values themselves. Recent work has extended this framework by incorporating machine learning into the imputation step: Chen and Xu (2025) provide a unified treatment of ML-based imputation, inverse propensity score weighting, and doubly robust methods for high-dimensional settings, demonstrating that

methods such as XGBoost and deep neural networks can handle the nonlinear relationships that challenge traditional imputation in large-scale datasets. Dang et al. (2026) show that even basic imputation models with modest predictor sets can produce reliable poverty estimates when survey data are missing, though accuracy improves with richer covariates.

The 2021 PIT count data gap, however, violates the missing-at-random assumption in a fundamental way. Missingness was driven by an exogenous shock (the COVID-19 pandemic) that simultaneously altered the outcome of interest. Communities did not simply fail to report their unsheltered counts; the pandemic changed what those counts would have been through economic dislocation, emergency policy responses, and shifts in shelter capacity. A standard imputation model trained on pre-pandemic data would recover the counterfactual (what counts would have been without COVID) rather than the actual 2021 values, while a model incorporating 2021 covariates would conflate the stable structural determinants of homelessness with transient pandemic effects. This paper's two-phase approach addresses the problem by separating these components: the first phase estimates the counterfactual using pre-pandemic relationships, and the second phase models the COVID-specific deviation as a function of pandemic-related variables observed for the subset of CoCs that did conduct counts. The logic parallels the selection correction tradition in econometrics (Heckman, 1976), where modeling the selection process explicitly allows for unbiased estimation in the presence of non-random sample composition, though the objective here is prediction rather than causal identification.

3.3 Data and Method

3.3.1 Methodological Overview

The imputation framework developed in this paper proceeds in two phases, each addressing a distinct component of the prediction problem. The central insight motivating this structure is that predicting missing 2021 unsheltered counts requires solving two problems simultaneously: estimating what counts would have been under normal conditions, and estimating how the COVID-19 pandemic altered those conditions. Collapsing these into a single model would conflate the stable,

long-run determinants of unsheltered homelessness with the transient shock of the pandemic, making it difficult to disentangle structural relationships from crisis-specific effects. Separating them allows each phase to be estimated on the data best suited to identify it.

A critical feature of the temporal structure warrants emphasis: Point-in-Time counts conducted in January of year t reflect conditions from the preceding calendar year $t-1$. People experiencing unsheltered homelessness in January 2020, for example, arrived at that state through processes unfolding in 2019: job losses, evictions, relationship dissolutions, or reductions in shelter capacity that occurred months earlier. Throughout this paper, predictors are therefore temporally aligned: the January 2020 PIT count is predicted using 2019 community characteristics, the January 2021 count from 2020 characteristics, and so forth. This one-year lag ensures that predictions capture the causal priority of community conditions over homelessness outcomes and avoids the mechanical relationship that would arise from using contemporaneous measures.

In the first phase, I estimate a baseline prediction model using a panel of CoC-level observations from 2015 through 2019. This model captures the historical relationship between community characteristics, such as housing markets, economic conditions, demographics, climate, shelter infrastructure, and unsheltered homelessness counts. The outcome is modeled on the log scale to address heteroskedasticity in count data, with predictions back-transformed to the count scale for interpretability. Multiple predictive algorithms are trained and compared, ranging from linear regression to ensemble machine learning methods, with out-of-sample predictive accuracy assessed on a locked 2020 holdout set. The selected model is then retrained on the full 2015–2019 panel and used to generate 2021 predictions for every CoC. These predictions represent a counterfactual: the unsheltered counts one would expect if pre-pandemic relationships had continued unchanged into 2021.

The second phase models the deviation between this counterfactual and observed reality. For the 149 CoCs that conducted complete unsheltered counts in 2021, I compute the residual—the difference between the actual count and the Phase 1 prediction. This residual isolates the COVID-specific component: the portion of the 2021 count that cannot be explained by historical patterns

alone. I then model these residuals as a function of pandemic-related variables, including COVID-19 severity, emergency relief programs, changes in shelter capacity, and local policy responses. Because the CoCs that conducted complete counts in 2021 differ systematically from those that did not, this phase incorporates overlap-based propensity score weighting to address the resulting selection bias and ensure that the estimated COVID adjustment generalizes to the non-counting CoCs.

The final imputed value for each non-counting CoC is the sum of its Phase 1 baseline prediction and its Phase 2 predicted COVID adjustment. All imputed values are accompanied by prediction intervals constructed via end-to-end clustered bootstrap, propagating uncertainty from both phases and providing downstream users with a transparent measure of estimate precision. Importantly, this framework does not claim to recover the true unsheltered count for any CoC; rather, it produces principled, uncertainty-quantified estimates that are preferable to the available alternatives of dropping 2021 from longitudinal analyses entirely, analyzing only the non-representative subset of CoCs that did count, or naively extrapolating from pre-pandemic trends.

3.3.2 Data Sources

The analysis draws on ten data sources spanning community characteristics, health conditions, climate, shelter infrastructure, federal funding, and pandemic-specific conditions. Tables 3.1 and 3.2 enumerate the full set of Phase 1 predictors along with their units and sources; Table 3.3 lists the Phase 2 predictors.

Figure 3.3: Methodological Flowchart: Two-Phase Imputation Framework

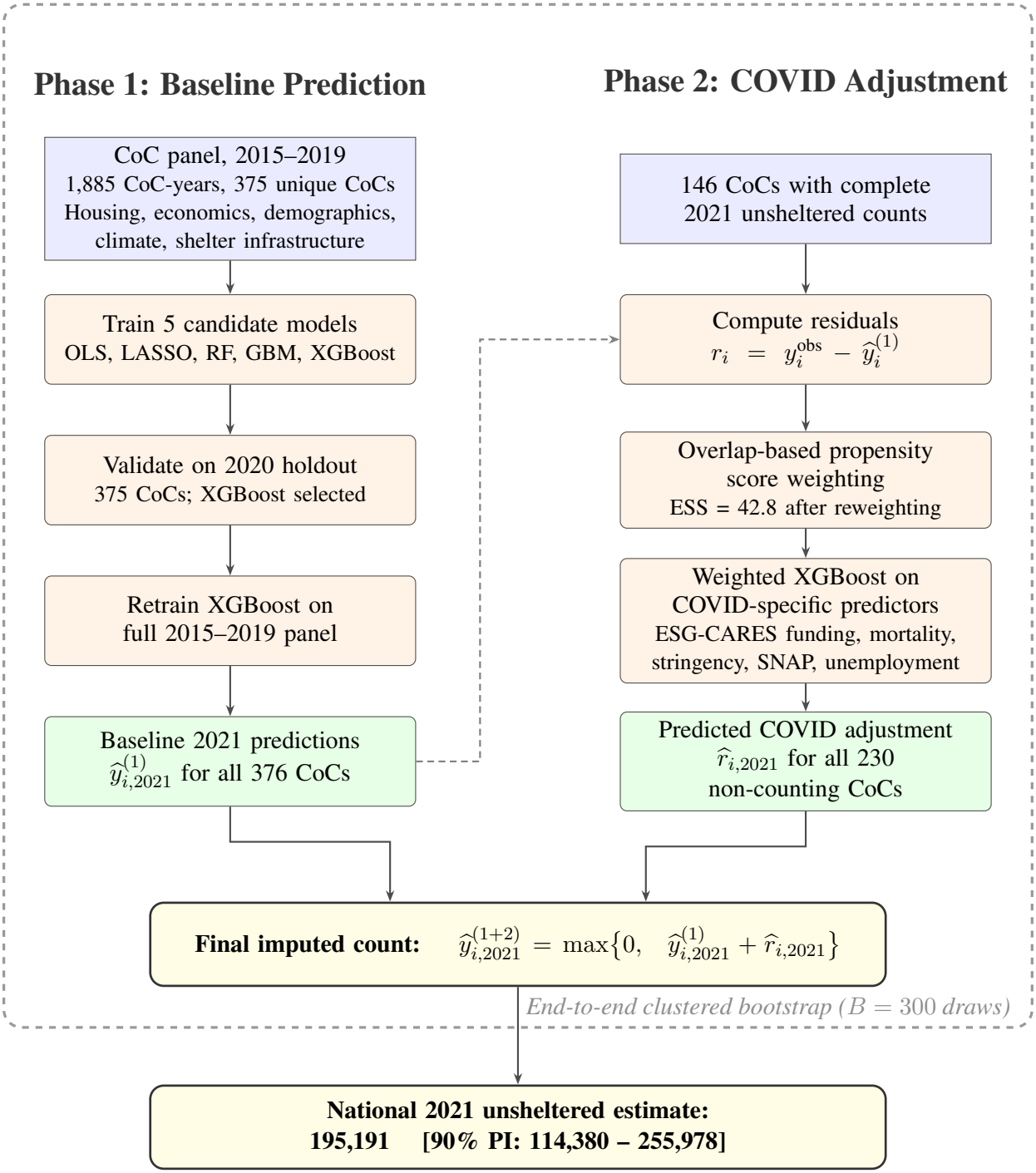


Table 3.1: Phase 1 Predictor Variables: Descriptions and Data Sources (Part I)

Variable	Unit	Data Source
Homeless Services Infrastructure		
Emergency Shelter Beds	Count	Housing and Urban Development Housing Inventory Count
Transitional Housing Beds	Count	Housing and Urban Development Housing Inventory Count
Permanent Supportive Housing Beds	Count	Housing and Urban Development Housing Inventory Count
Seasonal Shelter Beds	Count	Housing and Urban Development Housing Inventory Count
Overflow Shelter Beds	Count	Housing and Urban Development Housing Inventory Count
Housing Market		
Median Gross Rent	USD	American Community Survey 5-Year Estimates
Rent-Burdened Households	% households	American Community Survey 5-Year Estimates
Rental Vacancy Rate	%	American Community Survey 5-Year Estimates
Economic Conditions		
Unemployment Rate	%	American Community Survey 5-Year Estimates
Poverty Rate	%	American Community Survey 5-Year Estimates
Median Household Income	USD	American Community Survey 5-Year Estimates
Demographics and Mobility		
% Black Population	% population	American Community Survey 5-Year Estimates
% Hispanic Population	% population	American Community Survey 5-Year Estimates
Population Age 65+	% population	American Community Survey 5-Year Estimates
Population Age 18–24	% population	American Community Survey 5-Year Estimates
Bachelor's Degree or Higher	% pop. 25+	American Community Survey 5-Year Estimates
One-Person Households	% households	American Community Survey 5-Year Estimates
Divorced Population	% adults	American Community Survey 5-Year Estimates
Moved from Different State	% population	American Community Survey 5-Year Estimates
Moved from Abroad	% population	American Community Survey 5-Year Estimates

Note: All variables are measured at the Continuum of Care (CoC) level via spatial aggregation to CoC boundaries using annual HUD CoC shapefiles. American Community Survey variables reflect 5-year estimates aggregated from county-level data using a population-weighted crosswalk. All variables in the Phase 1 estimation panel span the years 2015–2019. All predictors from year t are aligned with the January year $t + 1$. t+1 PIT count outcome.

Table 3.2: Phase 1 Predictor Variables: Descriptions and Data Sources (Part II)

Variable	Unit	Data Source
Health and Social Conditions		
Mental Health Providers	Count	Centers for Medicare & Medicaid Services
Adult Obesity Rate	% adults	Behavioral Risk Factor Surveillance System
Teen Birth Rate	Per 1,000 females 15–19	National Center for Health Statistics Natality Files
Uninsured Adults	% adults	Small Area Health Insurance Estimates
Diabetes Prevalence	% adults	Behavioral Risk Factor Surveillance System
Households Receiving SNAP	% households	American Community Survey 5-Year Estimates
Population with Disability	% population	American Community Survey 5-Year Estimates
Veteran Population	% adults	American Community Survey 5-Year Estimates
Single-Parent Households	% households	American Community Survey 5-Year Estimates
Households with No Vehicle	% households	American Community Survey 5-Year Estimates
Very-Low-Income Renter Households	% renter hh.	American Community Survey 5-Year Estimates
Climate (January)		
Average Temperature	Degrees Celsius	National Oceanic and Atmospheric Administration
Total Precipitation	Millimeters	National Oceanic and Atmospheric Administration
CoC Characteristics and Federal Funding		
CoC Urbanicity Classification	Categorical	Housing and Urban Development Point-in-Time Count
Tenant Protection Policies	Count	National Low Income Housing Coalition Tenant Protection Database

Note: All Health and Social Condition variables are aggregated to the CoC level using an area-weighted spatial join to CoC boundaries. Climate variables are derived from county weather station data spatially interpolated to CoC centroids. All predictors from year t are aligned with the January year $t + 1$.

Table 3.3: Phase 2 Predictor Variables: COVID-19 Adjustment Model

Variable	Unit	Data Source
COVID-19 Severity and Policy		
COVID-19 Deaths	Count	Centers for Disease Control and Prevention
Any Mask Mandate	Binary (0/1)	CDC State, Territory, and County Public Mask Mandates
Lockdown Policy Stringency	Score (0-100)	Oxford Coronavirus Government Response Tracker
Economic Conditions		
Unemployment Rate	%	American Community Survey 5-Year Estimates
Households Receiving SNAP	% households	American Community Survey 5-Year Estimates
Emergency Relief and Housing Policy		
ESG-CARES Funding	USD (total)	Housing and Urban Development Allocations and Awards
Emergency Rental Assistance	Binary (0/1)	National Low Income Housing Coalition Tenant Protection Database
Any Eviction Moratorium	Binary (0/1)	National Low Income Housing Coalition Tenant Protection Database

Note: Phase 2 is estimated on the cross-section of CoCs that conducted complete unsheltered counts in 2021. The dependent variable is the residual between the Phase 1 naive prediction and the observed 2021 unsheltered count, which was conducted in January 2021. All variables reflect 2020 annual values. The mask mandate variable is coded as 1 if any county within the CoC had an active mask mandate policy; the underlying CDC dataset records mandate status at the county-day level. Emergency rental assistance and eviction moratorium indicators are coded as 1 if any jurisdiction within the CoC had an active policy as of January 2021.

The primary outcome variable, the annual unsheltered Point-in-Time count, comes from HUD's Point-in-Time Count dataset, which reports CoC-level counts of sheltered and unsheltered people on a single night each January. I use counts from 2015 through 2022. All predictors are temporally aligned to reflect the one-year lag between community conditions and homelessness outcomes: the January 2020 PIT count is predicted using 2019 community characteristics, and so forth. Phase 1 models the outcome on the log scale to address heteroskedasticity, with predictions back-transformed to counts for interpretation. The Housing Inventory Count, also administered by HUD, supplies annual CoC-level data on shelter bed capacity, including emergency shelter, transitional housing, permanent supportive housing, seasonal, and overflow beds. Both datasets are available from HUD's public data portal and are reported at the CoC level, requiring no spatial aggregation.

The bulk of the Phase 1 predictor set comes from the American Community Survey (ACS) five-year estimates, accessed via the Census Bureau's API. Health indicators, including adult obesity rates, teen birth rates, uninsured rates, and the count of mental health providers, come from sources including the Behavioral Risk Factor Surveillance System, the National Center for Health Statistics, and the Centers for Medicare and Medicaid Services National Provider Identification file. These data are also available at the county level.

Average January temperature and precipitation, two predictors with well-documented associations with unsheltered homelessness, come from the NOAA Global Historical Climatology Network Daily dataset. CoC Program funding allocations, used to construct the federal grant funding per capita predictor, come from HUD's annual awards data. Tenant protection policies are sourced from the National Low Income Housing Coalition's Tenant Protection Database, which documents the universe of active renter protections at the state and local level, including just-cause eviction requirements, rent stabilization ordinances, and right-to-counsel provisions among others.

The Phase 2 predictors introduce pandemic-specific data sources. COVID-19 death counts at the county level come from the Centers for Disease Control and Prevention. County-level mask mandate records come from the CDC's State and Territorial Public Mask Mandates dataset, which

records mandate status at the county-day level from April 2020 through January 2021. I aggregate this to a CoC-level binary indicator equal to one if any county within the CoC had an active mandate. Emergency Solutions Grants-CARES Act funding allocations come from HUD's awards data. Eviction moratorium and Emergency Rental Assistance policy indicators are constructed from the National Low Income Housing Coalition's Tenant Protection Database, restricted to policies enacted between March 2020 and January 2021.

3.3.3 Data Construction and Sample Inclusion Criteria

A central challenge in constructing the analysis dataset is that CoC boundaries do not correspond to any standard Census geography. CoC boundaries are administrative units defined by HUD that may encompass individual cities, single counties, multi-county regions, or entire states, and they frequently cross county and metropolitan area lines. Because demographic, economic, and housing market variables are collected at the Census tract or county level, and health indicators and policy data are reported at the county or jurisdiction level, a systematic procedure is required to aggregate sub-CoC data up to the CoC level.

I adopt a centroid-based spatial assignment approach. For each year in the panel, I obtain the corresponding CoC boundary shapefiles from HUD and county or jurisdiction geometries from the Census Bureau. I compute the geographic centroid of each sub-CoC unit—county, Census place, or other jurisdiction—and assign it to the CoC whose boundaries contain that centroid. Variables are then aggregated to the CoC level using summation for count variables (e.g., total COVID-19 cases, total ESG-CARES funding) and population-weighted averaging for rate and proportion variables (e.g., poverty rate, unemployment rate, percent rent-burdened). Matching jurisdiction-level policy data, such as tenant protection laws and COVID-era eviction moratoria, to CoC boundaries requires an additional step since policy databases report jurisdictions by name rather than by standardized geographic identifier. I match jurisdiction names to Census place and county geographies using exact string matching where possible and Levenshtein-distance fuzzy matching for the remainder,

with manual corrections applied to consolidated city-counties and other non-standard jurisdictions that resist automated matching.

The final consolidated dataset spans 2015 through 2019 and merges data from HUD's Point-in-Time Counts and Housing Inventory Counts, the American Community Survey, the Behavioral Risk Factor Surveillance System, the Small Area Health Insurance Estimates, the Center for Disease Control (CDC), NOAA climate records, CoC Program funding allocations, tenant protection policy databases, CDC COVID-19 case and mortality data, county-level mask mandate records, state and local lockdown stringency measures, ESG-CARES emergency funding allocations, and Emergency Rental Assistance and eviction moratorium policy records. Both modeling phases impose complete-case requirements—observations with missing outcome or predictor data are excluded—but differ in their treatment of panel structure and temporal coverage.

Phase 1 estimates baseline relationships using a panel of CoC-year observations from 2015 through 2019. The sample construction proceeds in three steps. First, the temporally shifted panel is restricted to non-territory CoCs, as U.S. territories exhibit fundamentally different housing markets, federal funding structures, and data reporting patterns that make pooled estimation inappropriate. Second, CoC-years with missing outcome data (unsheltered PIT count) are dropped, removing 107 of 2,214 observations. Third, CoC-years with missing values on any Phase 1 predictor are excluded via complete-case analysis, removing an additional 622 observations and yielding a final training sample of 1,485 CoC-years spanning 375 unique CoCs. This is done as the unsheltered count outcomes are transformed to the log scale (specifically, log(count + 1)) to address heteroskedasticity in count data.

Critically, Phase 1 allows CoCs to enter and exit the panel across years. Of the 375 unique CoCs in the training sample, 367 (98 percent) appear in all five years (2015–2019), six appear in four years, one appears in three years, and one appears in fewer than three years. This unbalanced panel structure reflects real-world administrative dynamics: CoC boundaries are periodically redrawn, new CoCs are created when regions split or reorganize their homeless assistance systems, and data quality improves over time as reporting infrastructure matures. Restricting the sample to a balanced

panel would exclude CoCs experiencing boundary changes and bias the training data toward stable, administratively mature jurisdictions. The machine learning methods employed in Phase 1 handle unbalanced panels naturally by treating each CoC-year as an independent observation, avoiding the fixed-effects logic that requires within-CoC variation (Efron and Hastie, 2016). The 2020 validation set contains 371 CoCs, nearly matching the 375 unique CoCs in training, and the 2021 baseline prediction set contains 386 CoCs, representing effectively complete national coverage of non-territory jurisdictions.

Phase 2 training imposes stricter requirements due to the cross-sectional structure and the need for comparability with Phase 1 predictions. While there are 385 CoCs that reported a PIT count of any kind in 2021, 5 come from US territories (Guam, Puerto Rico, the Virgin Islands, and the Mariana Islands) that lack sufficient feature variable data to be included. Furthermore, 4 CoCs (AR-505, MO-604, MD-514, and OK-504) lacked crucial COVID-related feature variable data that necessitated inclusion as well. Because of this, 376 CoCs in total are considered for the Phase 2 model process, with 230 CoCs listed as having partial or no unsheltered counts (the CoCs to impute for) and 146 CoCs as having completed the full unsheltered count in 2021.

The complete-case approach adopted here is conservative but appropriate for the prediction objective. Multiple imputation of missing predictors would introduce additional uncertainty into already-uncertain imputations and complicate the interpretation of prediction intervals. Moreover, the variables with the most missingness (health indicators) contribute modestly to Phase 1 predictive accuracy, as evidenced by feature importance rankings that place them outside the top ten predictors. The near-complete coverage of the 2021 baseline prediction set (376 of approximately 380 non-territory CoCs) confirms that predictor missingness is concentrated in the training period rather than the application year, mitigating concerns about coverage bias in the final imputed estimates.

3.3.4 Phase 1: Baseline Prediction Model

Phase 1 constructs a pre-pandemic baseline mapping from CoC characteristics to subsequent unsheltered PIT counts. The goal is not causal identification, but a counterfactual prediction: what each CoC's January 2021 unsheltered count would have been under the pre-2020 relationship between local conditions and unsheltered homelessness. These baseline predictions serve two purposes. First, they provide a "no-pandemic" benchmark against which observed 2021 counts can be compared. Second, for CoCs that conducted complete PIT counts in 2021, the baseline prediction errors (residuals) isolate the component of the 2021 deviation that is not explained by pre-pandemic covariates and therefore becomes the target for the Phase 2 COVID-era adjustment model.

I compare five prediction algorithms spanning a spectrum from interpretable linear models to flexible nonparametric ensembles. The linear benchmark is ordinary least squares (OLS), which provides a transparent baseline and a reference point for incremental gains from more flexible methods. The second model is LASSO regression, which augments the OLS objective with an L_1 penalty and therefore performs shrinkage and implicit variable selection when predictors are numerous and potentially collinear; the penalty parameter is selected by cross-validation. The remaining models are regression tree-based ensembles that accommodate nonlinearities and interactions without requiring them to be specified ex ante. Random Forest averages many trees grown on bootstrap samples to reduce variance and typically performs well in settings with complex predictor interactions (Breiman, 2001). Gradient Boosting Machines (GBM) build trees sequentially to reduce bias, fitting each new tree to the residual structure of the existing ensemble (Friedman, 2001). Finally, Extreme Gradient Boosting (XGBoost) is an efficient boosting implementation that combines tree boosting with regularization and subsampling features designed to improve generalization in high-dimensional settings (Chen and Guestrin, 2016). Collectively, these models provide a disciplined way to evaluate whether the data support a strong nonlinear or interaction structure beyond what linear specifications capture.

The predictor set comprises community characteristics from the prior year, but notably excludes the lagged unsheltered count as a predictor. This design choice departs from much of the homelessness forecasting literature, which routinely includes lagged outcomes to capture the strong empirical persistence of homelessness levels (Byrne et al., 2013; O'Flaherty, 2019; Hanratty, 2017; Corinth and Lucas, 2018). Three considerations motivate the exclusion.

First, methodological portability: a framework reliant on lagged counts cannot be applied when outcome data are missing for multiple consecutive years, precisely the scenario that arises during prolonged crises. Second, structural break concerns: flexible machine learning methods may overweight lagged outcomes and approximate autoregressive rules, which becomes problematic when pandemic disruptions decouple current homelessness from prior-year levels. Third, the cross-sectional imputation context creates a recursive dependency: including a 2020 lagged count would require first imputing that 2020 count, compounding uncertainty. Omitting the lag forces the model to extract signal from community characteristics alone and maintains a clear separation between the baseline counterfactual (Phase 1) and the pandemic-specific adjustment (Phase 2). All models are estimated with the unsheltered count outcome transformed to log(count + 1) to address heteroskedasticity in count data, with predictions back-transformed to the count scale for evaluation. Summary statistics for the full Phase 1 feature variable set is seen in Table 3.4.

For each algorithm, I adopt a common training and validation workflow. To evaluate which candidate model performs the best, I train all models on 2015 to 2017 data with hyperparameters being selected by grid search using grouped cross-validation on the 2018 set, with the objective of minimizing out-of-sample prediction error. For Random Forest, the grid spans the number of trees, the number of predictors sampled at each split, and minimum node size. For GBM and XGBoost, the grid spans tree depth, learning rate, subsampling fractions, and regularization settings, with the number of boosting rounds chosen using early stopping. Using a unified holdout and consistent tuning protocol ensures that comparisons reflect predictive performance rather than differences in evaluation design. Once the optimal hyperparameter values are determined, each candidate model is evaluated on its performance in predicting the held-out test sample of 2019 data. Once the most

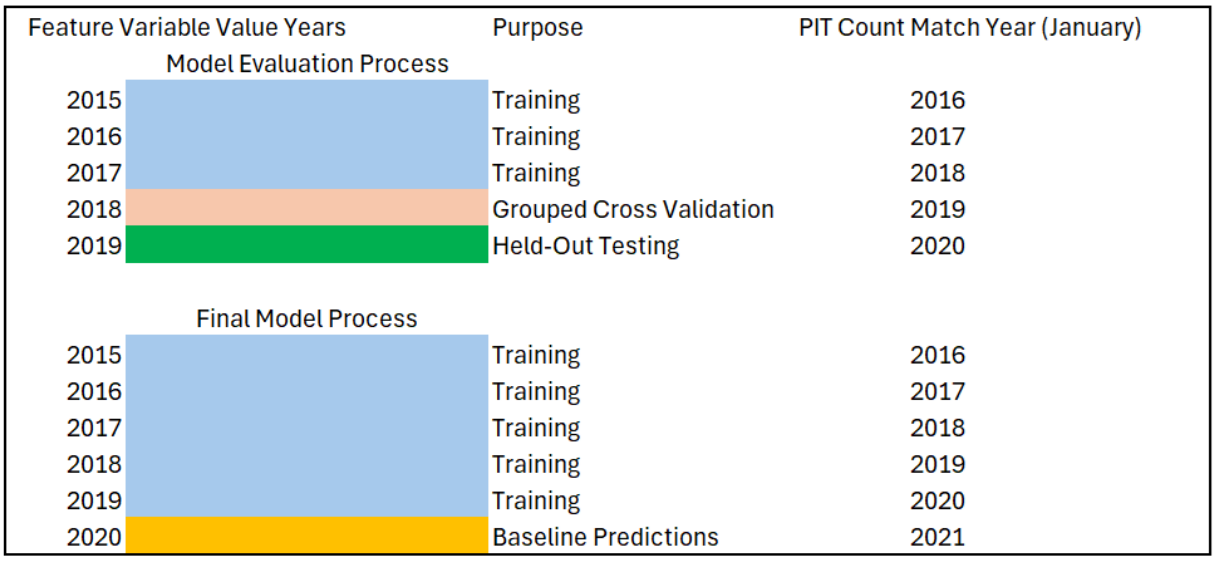
Table 3.4: Phase 1 Modeling Sample (2015–2019) Summary Statistics

Variable	N	Mean	SD	Min	Median	Max
Homelessness outcome						
Total unsheltered homeless	1,885	521.8	2,124.4	0	113	42,471
Homeless services infrastructure						
Emergency shelter beds	1,885	739	3,853	0	266	75,245
Transitional housing beds	1,885	303	507	0	147	6,760
Permanent supportive housing beds	1,885	942	2,167	0	380	32,150
Seasonal shelter beds	1,885	56	119	0	15	1,791
Overflow shelter beds	1,885	45	89	0	12	1,101
Housing market						
Median rent (USD)	1,885	1,064	321	597	986	2,439
Rent-burdened households (%)	1,885	50.1	4.9	36.4	49.5	66.5
Rental vacancy rate (%)	1,885	6.4	2.8	1.7	6.0	24.7
Tenant protection policies (count)	1,885	1.0	1.8	0	0	10
Economic conditions						
Unemployment rate (%)	1,885	6.4	2.0	1.9	6.0	22.2
Poverty rate (%)	1,885	14.3	4.6	3.4	14.2	39.5
Median household income (USD)	1,885	64,489	17,989	27,901	60,003	147,536
Demographics and mobility						
% Black population	1,885	12.2	12.6	0.4	8.2	79.6
% Hispanic population	1,885	13.4	13.3	1.0	8.7	84.2
Population age 65+ (%)	1,885	15.7	3.7	7.4	15.5	39.3
Population age 18–24 (%)	1,885	10.0	2.7	5.3	9.3	28.3
Bachelor's degree or higher (%)	1,885	31.2	10.6	12.6	29.7	79.0
One-person households (%)	1,885	28.3	4.4	12.7	28.5	48.2
Divorced (%)	1,885	11.2	1.9	5.9	11.2	17.4
Moved from different state (%)	1,885	2.5	1.3	0.5	2.2	8.4
Moved from abroad (%)	1,885	0.6	0.5	0.02	0.5	4.9
Health and social conditions						
Mental health providers (count)	1,885	2,071	3,577	60	1,067	31,986
Obesity rate (%)	1,885	28.5	4.7	14.8	28.8	39.8
Teen birth rate (per 1,000)	1,885	29.1	12.9	3.6	27.8	90.0
Uninsured rate (%)	1,885	11.9	5.2	2.1	11.3	33.1
Diabetes rate (%)	1,885	10.2	2.0	4.2	10.1	16.6
Households with SNAP (%)	1,885	12.3	4.8	2.5	12.1	42.4
Disability rate (%)	1,885	13.1	3.0	5.4	13.1	22.5
Veteran population (%)	1,885	8.2	2.7	1.8	8.2	20.5
Single-parent households (%)	1,885	13.9	3.4	6.5	13.6	32.3
Households without vehicle (%)	1,885	7.9	5.3	2.3	6.7	54.8
Very low income renters (%)	1,885	28.3	7.9	7.3	28.6	54.1
Climate (January)						
Average temperature (°C)	1,885	2.3	7.0	−16.7	1.2	23.2
Total precipitation (mm)	1,885	90.5	71.1	0.0	78.5	651.7

Note: Summary statistics for the Phase 1 complete-case modeling sample (N = 1,885 CoC-years from 375 unique CoCs, 2015–2019). This sample excludes CoC-years with missing outcomes or predictors. All variables are measured at the Continuum of Care level via spatial aggregation. The outcome variable (total unsheltered homeless) is modeled on the log scale; summary statistics shown here are in levels (raw counts).

accurate model is determined, the final model used for the Phase 1 predictions is re-trained on all 2015 to 2019 data to make the 2021 baseline predictions. Figure 3.4 displays this process.

Figure 3.4: Visual Diagram of Phase 1 Training Process



Model performance is summarized using mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2) on the held-out test data. Because CoCs vary substantially in size and urbanicity, I additionally examine holdout performance by CoC category and by size strata to assess whether prediction error is systematically concentrated in particular parts of the CoC distribution (e.g., large urban CoCs versus smaller Balance-of-State CoCs).

I select the Phase 1 model as the specification (algorithm $\times$ predictor set) that yields the best holdout performance under the primary accuracy criterion (RMSE, with MAE used to assess robustness to outliers). After this selection, I then generate baseline predictions for each CoC using its predictor-year 2020 covariates (aligned to the January 2021 PIT outcome) to maximize information used to estimate the baseline mapping, producing $\widehat{U}_{c,2021}^{\text{baseline}}$, the predicted unsheltered count absent pandemic-era disruptions.

3.3.5 Phase 2: COVID Adjustment Model

Phase 1 produces a baseline prediction for each CoC's 2021 unsheltered count using pre-pandemic relationships learned from the Phase 1 panel. However, 2021 is not a typical missing-data year: the pandemic plausibly induced systematic deviations from historic patterns in ways that differ across CoCs. Phase 2 is designed to explicitly model this COVID-era deviation and use it to adjust the Phase 1 baseline for CoCs that did not conduct a complete unsheltered count in 2021.

This is the primary concern in Phase 2: that CoCs that completed full unsheltered counts in 2021 may differ systematically from those that did not, which can be seen in Table 3.5. If so, a residual model trained on full-count CoCs could extrapolate poorly to the non-counting group. To diagnose and mitigate this covariate shift, I estimate a propensity score model for the probability of completing a full 2021 count using a broader set of contextual covariates (population size, income, rents and vacancy rates, poverty, age composition, insurance and health measures, and shelter capacity measures), along with COVID-era variables and CoC category. Specifically, I fit a logistic regression

$$\Pr(T_i = 1 \mid X_i) = \frac{1}{1 + \exp(-X_i'\beta)},$$

where $T_i = 1$ indicates a full 2021 count and X_i' is the covariate set shown in Table 3.5.

From this model I compute a propensity score $\widehat{p}_i$ and inspect overlap between full-count and non-full-count CoCs by comparing the distributions of $\widehat{p}_i$ across groups. To improve transportability of the residual model to non-counting CoCs, I then construct an inverse-propensity reweighting scheme that reweights full-count CoCs to resemble the covariate distribution of non-full-count CoCs. Concretely, I weight each full-count CoC by

$$w_i \;=\; \frac{1 - \widehat{p}_i}{\widehat{p}_i} \quad \text{for } T_i = 1,$$

which reweights the treated (counting) sample to approximate the covariate distribution of the non-counting CoCs, ensuring that the residual model learns patterns representative of the population it will be applied to. To stabilize the procedure, propensity scores are truncated away from 0 and

1, and extreme treated weights are capped at the 99th percentile. Finally, I compute standardized mean differences (SMDs) before and after weighting as a balance diagnostic.

Table 3.5: Phase 2: Covariate Balance by 2021 Counting Status

	Mean		
Variable	Full Count	No Count	SMD
ESG-CARES funding (USD thousands)	58,337	99,149	−0.437
Average policy stringency score	4.7	5.2	−0.328
COVID-19 deaths (2021)	1,451	2,203	−0.247
Total population	706,147	959,486	−0.214
Has eviction moratorium (indicator)	0.02	0.09	−0.195
Has mask mandate (indicator)	0.90	0.96	−0.186
Overflow shelter beds	54	87	−0.177
Unemployment rate (%)	5.4	5.6	−0.122
Obesity rate (%)	30.8	30.2	0.112
Population age 65+ (%)	17.0	16.7	0.088
Poverty rate (%)	12.6	12.9	−0.086
Has Emergency Rental Assistance (indicator)	0.11	0.17	−0.067
Seasonal shelter beds	48	60	−0.065
Emergency shelter beds	964	707	0.062
Median rent (USD)	1,165	1,188	−0.061
Median household income (USD)	75,834	74,662	0.057
Rental vacancy rate (%)	5.97	5.91	0.047
Uninsured rate (%)	9.41	9.62	−0.041
Diabetes rate (%)	10.9	10.8	0.041

Note: Comparison of observable characteristics between CoCs that conducted complete unsheltered counts in 2021 (Full Count, N = 146) and those that did not (No Count, N = 230). SMD = standardized mean difference, calculated as $(m_1 - m_0)/\sqrt{(s_1^2 + s_0^2)/2}$, where m and s denote means and standard deviations. Values sorted by absolute SMD. CoCs that did not count are systematically larger, received more emergency funding, experienced higher COVID mortality, and faced stricter policy environments. CoC category distribution: Full Count (48% largely suburban, 20% other urban, 20% largely rural, 12% major city); No Count (40% largely suburban, 34% largely rural, 13% other urban, 13% major city).

Phase 2 is trained only on CoCs that conducted a full 2021 count (sheltered and unsheltered), because these CoCs provide the only observations where the true 2021 unsheltered count is known.

For each of these CoCs, I define the dependent variable as the level residual

$$r_{i,2021} \;=\; y^{\text{obs}}_{i,2021} \;-\; \widehat{y}^{(1)}_{i,2021},$$

where $y^{\text{obs}}_{i,2021}$ is the observed 2021 unsheltered PIT count and $\widehat{y}^{(1)}_{i,2021}$ is the Phase 1 baseline prediction. A key modeling choice is that residuals are kept entirely in levels (no transformations). This is necessary because true deviations may be negative, and the residual model should be allowed to predict negative adjustments when supported by the data.

The Phase 2 feature set is intentionally COVID-specific, capturing channels through which 2021 unsheltered counts may have departed from pre-pandemic expectations. The final predictor set includes: COVID mortality burden (total deaths in 2021), policy environment (average stringency index and indicator variables for eviction-related policies such as moratoria and mandates), emergency resources (total ESG-CARES funding), and household economic stress proxies (e.g., SNAP participation and unemployment), along with a CoC category factor to allow systematic differences by urbanicity. These predictors are assembled into a one-row-per-CoC modeling frame for 2021, with consistent indicator coding and factor handling prior to model fitting. Table 3.6 displays summary statistics for the final predictor set of all CoCs in the 2021 data.

Table 3.6: Phase 2 (2021) Summary Statistics: COVID Adjustment Predictors

Variable	N	Mean	SD	Min	Median	Max
COVID-19 deaths (2021)	376	1,918	3,295	0	1,042	37,139
Average policy stringency score	376	35.01	31.53	1.00	25.36	76.77
ESG-CARES funding (USD thousands)	376	83,803	101,153	4,390	48,352	616,388
Has Emergency Rental Assistance	376	0.13	0.34	0	0	1
Has eviction moratorium	376	0.02	0.15	0	0	1
Has mask mandate	376	0.90	0.29	0	1	1
Households with SNAP (%)	376	11.4	4.6	2.3	11.0	36.9
Unemployment rate (%)	376	5.5	1.5	2.4	5.3	15.3

Note: Summary statistics for Phase 2 COVID-specific predictors measured at the CoC level in 2021. Sample includes all non-territory CoCs with complete Phase 2 predictor data (N = 376). ESG-CARES funding scaled to thousands for readability. Binary indicators (ERA, moratorium, mandate) take values 0 or 1.

Table 3.7: CoC Urbanicity Composition by 2021 Count Completion Status

CoC category	Share (Full count)	Share (Not full)	Diff.
Largely Rural CoC	19.7	33.6	-13.9
Largely Suburban CoC	48.3	40.1	8.2
Other Largely Urban CoC	20.4	13.4	7.0
Major City CoC	11.6	12.9	-1.4

Notes: Shares are computed within each group (Full count vs. Not full) for the 2021 cross-section of non-territory CoCs. "Diff." reports the difference in shares (Full count − Not full). The urbanicity classification is the HUD CoC category used throughout the analysis to stratify model performance and to assess selection into conducting a full unsheltered count in 2021.

Using the full-count CoCs, the Phase 2 residual model is estimated as a supervised learning problem with outcome $r_{i,2021}$ and predictors described above. I fit the same five candidate model families used in Phase 1: weighted OLS, weighted LASSO, weighted random forest, weighted gradient boosting (GBM), and weighted XGBoost. All models are trained with the overlap-based weights w_i applied to the treated (full-count) sample.

To ensure consistent feature construction across methods, I implement a single preprocessing pipeline using a recipe that removes zero-variance predictors and one-hot encodes factor variables. This produces a stable design matrix for the linear models and for the tree-based learners. For the ensemble methods (RF, GBM, and XGB), I evaluate a small hyperparameter grid and select the best-performing specification using cross-validated error on the treated sample, subject to the constraint that the sample size (146 full-count CoCs) limits how aggressive tuning can be without overfitting.

For every in-scope CoC, the chosen Phase 2 model generates a predicted residual $\widehat{r}_{i,2021}$. The final imputed 2021 unsheltered count is then formed as

$$\widehat{y}_{i,2021}^{(1+2)} \;=\; \widehat{y}_{i,2021}^{(1)} + \widehat{r}_{i,2021}.$$

Because population counts cannot be negative, I enforce a hard floor at zero:

$$\widehat{y}_{i,2021}^{(1+2,\text{clamp})} \;=\; \max\{0,\; \widehat{y}_{i,2021}^{(1)} + \widehat{r}_{i,2021}\}.$$

Importantly, I also retain an indicator for whether the unclamped prediction was negative, which serves as a diagnostic flag for CoCs where the model implies an unusually large negative adjustment relative to the Phase 1 baseline.

To contextualize model performance against a widely used ad hoc alternative, I compare Phase 1 and Phase 1+2 predictions to midpoint interpolation based on adjacent PIT counts. For CoCs with observed 2021 unsheltered counts, and where both 2020 and 2022 counts are available, the midpoint benchmark is defined as

$$y_{i,2021}^{\text{mid}} \;=\; \frac{y_{i,2020}^{\text{obs}} + y_{i,2022}^{\text{obs}}}{2}.$$

I evaluate prediction accuracy using RMSE and MAE for: (i) midpoint interpolation, (ii) Phase 1 baseline, (iii) Phase 1+2 adjusted predictions, and (iv) the clamped Phase 1+2 predictions. This comparison is reported overall and by CoC category, providing a transparent check on whether the two-phase procedure improves upon a simple, commonly used interpolation rule.

3.3.6 Uncertainty Quantification

To construct uncertainty intervals that correctly propagate error from both stages of the imputation pipeline, I implement an end-to-end clustered bootstrap at the CoC level. The bootstrap resamples CoCs with replacement from the Phase 1 training panel (clustered by CoC to preserve within-CoC serial dependence), refits both the Phase 1 preprocessing recipe and the log-scale baseline model on each resample, back-transforms predictions to the count scale, and generates a bootstrap draw of $\widehat{y}_{i,2021}^{(1)}$ for all CoCs. Conditional on that draw, I then rebuild the Phase 2 training residuals $r_{i,2021}$ for full-count CoCs, refit the Phase 2 residual model using the overlap-based weights computed on the full sample, and generate bootstrap residual predictions $\widehat{r}_{i,2021}$ for all CoCs. Each bootstrap draw therefore yields an end-to-end imputed prediction

$$\widehat{y}_{i,2021,b}^{(1+2)} \;=\; \widehat{y}_{i,2021,b}^{(1)} + \widehat{r}_{i,2021,b},$$

along with the clamped counterpart $\max\{0, \widehat{y}^{(1+2)}_{i,2021,b}\}$. Repeating this procedure across $B = 300$ bootstrap draws produces an empirical distribution of predictions for each CoC, from which I report percentile-based intervals (e.g., 5th and 95th percentiles for 90% intervals) and the bootstrap frequency of negative *unclamped* predictions as an additional diagnostic.

This end-to-end bootstrap is intentionally conservative: it treats the two-stage prediction pipeline as a single composite estimator and captures uncertainty from both the baseline model and the COVID adjustment model, rather than conditioning on Phase 1 predictions as fixed.

3.4 Results

3.4.1 Phase 1: Baseline Model Performance

The Phase 1 baseline model predicts 2021 unsheltered counts using only pre-pandemic structural determinants—housing markets, economic conditions, demographics, climate, shelter infrastructure, and federal funding—without relying on lagged outcome data. This design choice prioritizes methodological portability: the framework can be applied to settings where outcome data are missing for multiple consecutive years, a common feature of administrative data gaps during prolonged crises.

I compare five candidate algorithms on a locked 2020 holdout set: ordinary least squares, LASSO, random forest, gradient boosting (GBM), and XGBoost. Table 3.8 reports test-set performance. XGBoost substantially outperforms all alternatives, achieving an RMSE of 905 and an R^2 of 0.972, compared to the next-best model (random forest: RMSE = 1,457, R^2 = 0.934). The linear models perform poorly by comparison—OLS and LASSO produce RMSEs exceeding 9,000—indicating that the relationship between community characteristics and unsheltered homelessness exhibits strong nonlinearities that tree-based ensembles capture but linear specifications miss. XGBoost is selected as the final Phase 1 model and retrained on the full 2015–2019 panel before generating 2021 baseline predictions.

Prediction accuracy varies systematically by CoC type. Table 3.9 stratifies holdout performance by urbanicity. The model performs best in largely suburban CoCs (RMSE = 18, MAE = 8)

Table 3.8: Phase 1: Model Comparison on 2020 Holdout Set

Model	N	RMSE	MAE	R^2
XGBoost	375	905	164	0.972
Random Forest	375	1,457	250	0.934
GBM	375	1,834	302	0.771
LASSO	375	9,233	796	0.475
OLS	375	9,886	831	0.474

Note: Out-of-sample performance on the 2020 validation set (year 2019 predictors $\rightarrow$ January 2020 PIT outcome). All models trained on 2016–2018 data (using 2015-2017 predictors) for cross-validation and hyperparameter tuning, with validation on 2020 holdout (using 2019 predictors). Selected model retrained on full 2015-2019 panel for final predictions. RMSE = root mean squared error; MAE = mean absolute error; R^2 = coefficient of determination. XGBoost selected as Phase 1 model based on superior RMSE.

and other largely urban CoCs (RMSE = 12, MAE = 7), where unsheltered counts are moderate and relatively stable. Performance degrades slightly in largely rural CoCs (RMSE = 27, MAE = 11), likely reflecting greater heterogeneity in local conditions and smaller sample sizes for model training. The largest errors occur in major city CoCs (RMSE = 409, MAE = 112), where unsheltered populations are both large and volatile. Importantly, however, even in major cities the model captures the broad magnitude of unsheltered homelessness—mean predicted and actual counts differ by less than 3 percent (2,338 versus 2,410)—suggesting that the imputation procedure will produce defensible estimates across the full CoC distribution, albeit with wider uncertainty intervals for large urban jurisdictions.

Figure 3.5 plots predicted versus actual 2020 unsheltered counts on a log scale. The tight clustering around the 45-degree line confirms that the model predicts well across the full range of CoC sizes, with no systematic over- or underprediction at either tail of the distribution. Table 3.10 reports summary statistics of the baseline predictions.

Table 3.9: Phase 1: Prediction Accuracy by CoC Urbanicity (2020 Holdout)

CoC Category	N	RMSE	MAE	Mean Actual	Mean Predicted
Largely Suburban CoC	162	18	8	302	299
Largely Rural CoC	102	27	11	394	388
Other Largely Urban CoC	60	12	7	215	214
Major City CoC	47	409	112	2,410	2,338

Note: XGBoost model performance stratified by CoC urbanicity category on the 2020 validation set. RMSE and MAE reported in levels (unsheltered count). Mean actual and predicted counts show the model captures central tendency well across all categories, with proportionally larger absolute errors in major cities due to scale.

Figure 3.5: Phase 1 Actual-Predicted Scatterplot, Log Scale

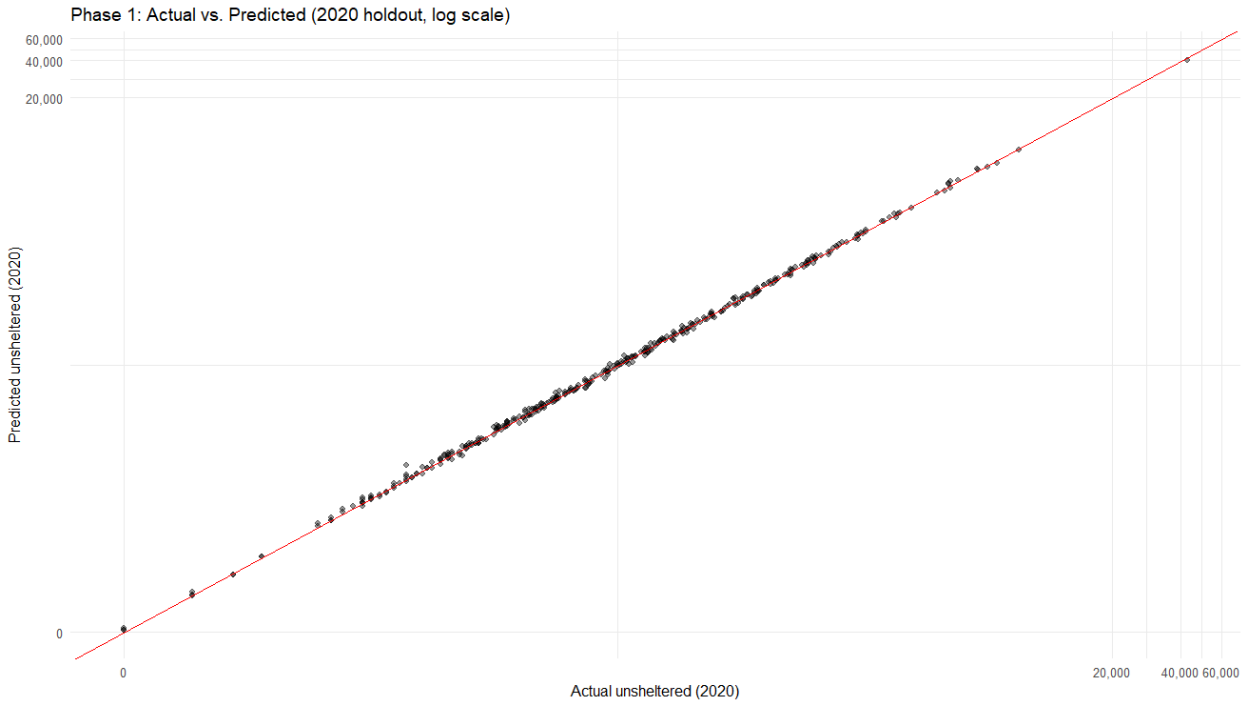


Table 3.10: Phase 1 Baseline Predictions: Summary Statistics (2021)

	N	Mean	SD	Min	Median	Max
Overall	376	468	1,545	0.6	136	26,830
By CoC Category						
Major City CoC	47	1,916	3,977	106	902	26,830
Largely Rural CoC	105	307	579	0.6	141	4,706
Largely Suburban CoC	163	277	487	6	105	3,412
Other Largely Urban CoC	61	196	296	10	94	1,418

Note: Summary statistics for Phase 1 baseline predictions of 2021 unsheltered counts, generated using 2020 community characteristics and the log-scale XGBoost model retrained on 2015–2019 data. Predictions generated for all 376 CoCs in the 2020 predictor set. Major City CoCs have substantially higher predicted counts on average (mean = 1,916) compared to other categories, reflecting both larger populations and structural conditions conducive to unsheltered homelessness. The overall distribution is right-skewed (median = 136, mean = 468), consistent with the observed distribution of unsheltered counts.

What drives the baseline predictions? Feature importance rankings (Table 3.11) reveal that January average temperature dominates the model, accounting for 23.3 percent of total predictive gain. This aligns with prior research documenting that mild winter climates enable year-round outdoor habitation and attract unsheltered populations through migration (Lucas, 2017; Corinth and Lucas, 2018). The next most important predictors are mental health provider availability (12.6 percent), emergency shelter bed capacity (10.4 percent), transitional housing beds (7.5 percent), and permanent supportive housing beds (6.8 percent). Together, these five variables account for over 60 percent of the model's predictive power, underscoring the centrality of climate and service infrastructure in shaping unsheltered homelessness patterns.

Figure C.1 displays the top 20 features by gain, and Figure 3.6 presents SHAP values for the full feature set. SHAP values decompose each prediction into feature-specific contributions, revealing both the direction and magnitude of each variable's effect on the model's output for individual observations. The SHAP beeswarm plot reveals directional effects consistent with the homelessness literature: higher January temperatures, more emergency shelter beds, and larger shares of Hispanic populations are associated with higher predicted unsheltered counts, while higher mental health provider density, more permanent supportive housing, and higher poverty rates are associ-

Table 3.11: Phase 1: XGBoost Feature Importance (Top 20 Predictors)

Rank	Feature	Gain
1	January average temperature (°C)	0.233
2	Mental health providers (count)	0.126
3	Emergency shelter beds	0.104
4	Transitional housing beds	0.075
5	Permanent supportive housing beds	0.068
6	% Hispanic population	0.051
7	% Black population	0.040
8	% Divorced	0.025
9	Uninsured rate (%)	0.021
10	Median rent (USD)	0.019
11	January precipitation (mm)	0.017
12	% Veteran population	0.016
13	% Households without vehicle	0.015
14	% Moved from different state	0.014
15	Rental vacancy rate (%)	0.012
16	Population age 18–24 (%)	0.012
17	% Rent-burdened households	0.011
18	% One-person households	0.011
19	Population age 65+ (%)	0.011
20	Disability rate (%)	0.011

Note: Feature importance ranked by total gain contribution in the Phase 1 XGBoost model. Gain measures the improvement in prediction accuracy from splits on each variable, summed across all trees. Top 20 of 38 total features shown. January temperature dominates with 23.3% of total gain.

ated with lower counts. The negative association with poverty is initially counterintuitive but likely reflects the model learning that CoCs with extreme poverty concentrate in rural areas with small unsheltered populations, while large unsheltered populations concentrate in high-cost urban areas where poverty rates are moderate. Similarly, the positive association between emergency shelter beds and unsheltered counts likely reflects reverse causality in the training data: jurisdictions build more shelter capacity in response to large unsheltered populations, not the other way around.

Figure 3.6: Phase 1 SHAP Values

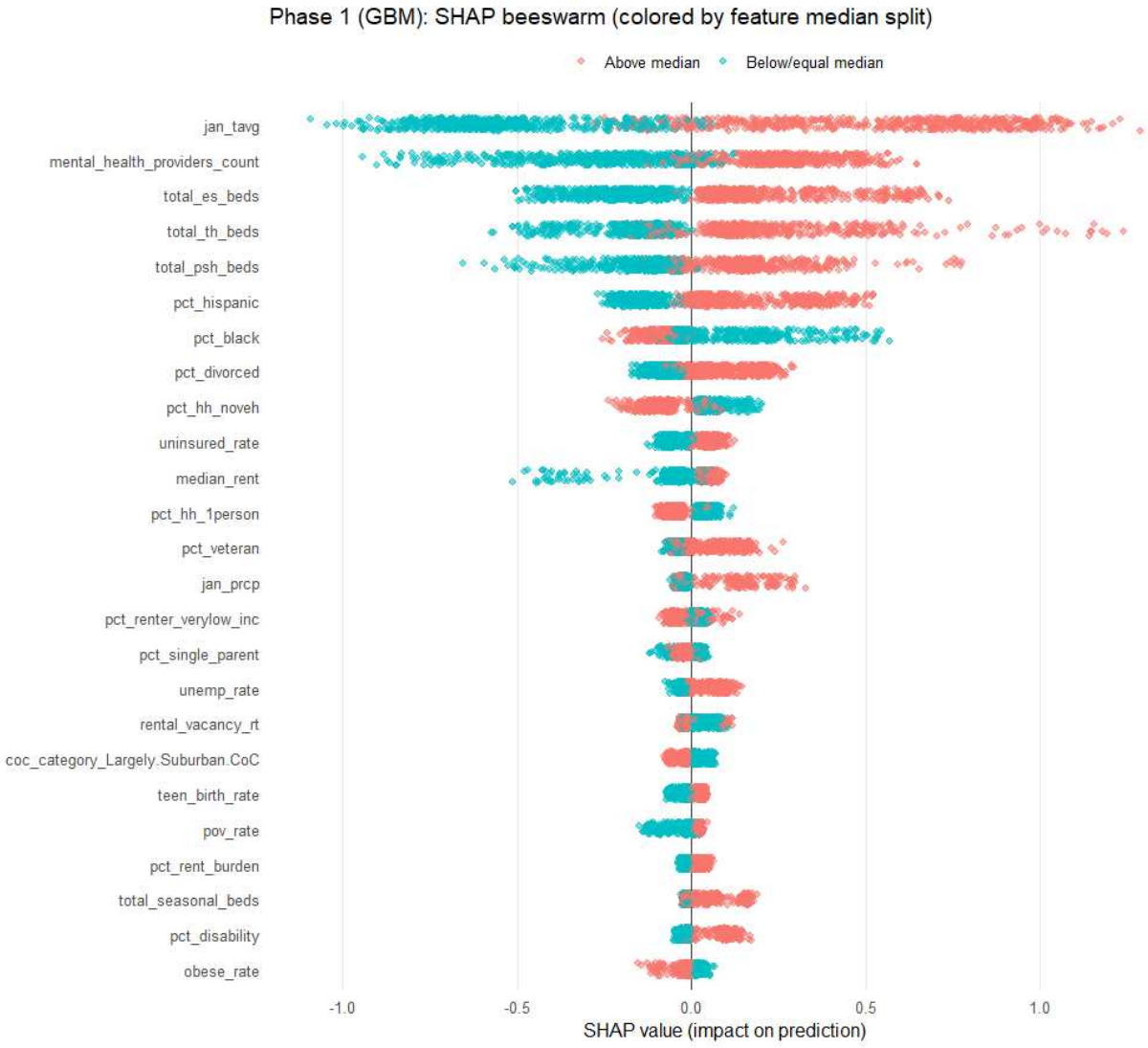


Table 3.12 illustrates Phase 1 prediction accuracy across the CoC distribution with a random sample of five CoCs from each urbanicity category. The model captures variation reasonably well, with errors ranging from near-perfect predictions (Miami-Dade: 892 actual vs. 981 predicted) to larger misses in smaller jurisdictions (New Mexico Balance of State: 362 actual vs. 753 predicted). These examples confirm the pattern in Table 3.9: proportional errors are larger in smaller CoCs, while absolute errors are larger in major cities.

Table 3.12: Phase 1 Baseline Predictions: Random Sample by CoC Category

CoC ID	CoC Name	Actual 2021	Predicted 2021
Largely Rural CoC			
AL-507	Alabama Balance of State	98	168
CO-500	Colorado Balance of State	528	359
MI-511	Lenawee County	0	16
NM-501	New Mexico Balance of State	362	753
TX-604	Waco/McLennan County	41	119
Largely Suburban CoC			
MA-506	Worcester City & County	247	148
MD-506	Carroll County	18	21
NJ-504	Newark/Essex County	91	452
NY-500	Rochester, Irondequoit, Greece/Monroe County	33	55
NY-505	Syracuse, Auburn/Onondaga, Oswego, Cayuga Counties	12	25
Major City CoC			
FL-510	Jacksonville-Duval, Clay Counties	230	476
FL-600	Miami-Dade County	892	981
KY-501	Louisville-Jefferson County	257	229
OK-501	Tulsa City & County	287	375
TX-601	Fort Worth/Arlington/Tarrant County	479	902
Other Largely Urban CoC			
GA-505	Columbus-Muscogee	27	66
IA-500	Sioux City/Dakota, Woodbury Counties	16	10
IL-516	Decatur/Macon County	31	28
VA-505	Newport News/Hampton/Virginia Peninsula	30	90
VA-600	Arlington County	27	33

Note: Random sample of 5 CoCs per urbanicity category from the 146 CoCs that conducted complete unsheltered counts in 2021. Actual 2021 = observed January 2021 PIT unsheltered count. Predicted 2021 = Phase 1 baseline prediction using 2020 community characteristics and the log-scale XGBoost model. Sample selected via random sampling within category.

The final Phase 1 model, retrained on 2015–2019 data, generates baseline 2021 predictions for all 376 CoCs. These predictions range from 1 to 26,830 (median = 136, mean = 468), closely matching the distribution of observed unsheltered counts in pre-pandemic years. For the 146 CoCs that conducted complete 2021 counts, the residual (the difference between the observed count and the Phase 1 baseline prediction) isolates the COVID-specific component of the 2021 outcome. This residual becomes the dependent variable in Phase 2, to which I now turn.

3.4.2 Phase 2: COVID Adjustment Model

The Phase 2 residual model addresses the central challenge of the 2021 data gap: the 146 CoCs that conducted complete unsheltered counts are not representative of the national landscape. CoCs that counted differ systematically from those that did not, creating a covariate shift problem that would bias naive extrapolation of the residual pattern. I address this through overlap-based propensity score weighting, reweighting the 146 full-count CoCs to resemble the covariate distribution of the 230 non-counting CoCs. After weighting, standardized mean differences on included covariates fall below 0.1 for all but 6 predictors, indicating drastically improved balance (see Appendix Figure C.2). The effective sample size after weighting is 42.8 (Table 3.13), reflecting moderate efficiency loss from reweighting but preserving sufficient information for residual modeling.

Table 3.13: Phase 2: Overlap Weights Distribution (Treated CoCs)

N	Min	5th %	Median	95th %	Max	ESS
146	0.034	0.333	1.02	4.32	17.8	42.8

Note: Distribution of overlap-based propensity score weights $w_i = (1 - \widehat{p}_i)/\widehat{p}_i$ for the 146 treated CoCs (complete 2021 unsheltered counts) used in Phase 2 training. ESS = effective sample size, calculated as $(\sum w_i)^2 / \sum w_i^2$. The ESS of 42.8 reflects moderate efficiency loss from reweighting but preserves sufficient information for residual modeling.

I fit five weighted candidate models to predict the residual as a function of COVID-specific variables: mortality burden, policy stringency, emergency relief funding, eviction policy, and eco-

nomic stress. Each model is trained on 110 randomly sampled CoCs and their performance is evaluated on the remaining 36 held-out CoCs. Table 3.14 reports residual prediction performance on the held-out sample. XGBoost performs well again, achieving an RMSE of 181 and explaining 47.8 percent of residual variation. The next-best model (gradient boosting: RMSE = 252, R^2 = 0.369) performs substantially worse, and the linear models fail entirely: weighted LASSO and OLS produce negative R^2 values, indicating they predict worse than a horizontal line through the residual mean. The baseline-only benchmark (predicting zero adjustment for all CoCs) yields an RMSE of 318, confirming that the Phase 2 model meaningfully improves over naive baseline predictions. The sample size constraint (N = 146, ESS = 42.8 after weighting) limits the complexity of models that can be reliably estimated, but XGBoost's strong performance suggests the COVID adjustment pattern is sufficiently systematic to be learned even with a small effective sample. The summary statistics of the residual predictions can be seen in Table 3.15.

Table 3.14: Phase 2: COVID Adjustment Model Comparison

Model	RMSE	MAE	R^2
Phase 2 XGBoost	181	54	0.478
Phase 2 GBM	252	82	0.369
Phase 2 Random Forest	276	74	0.245
Phase 2 LASSO	339	192	-0.136
Phase 2 OLS	374	182	-0.386

Note: Each model is trained on 75% (110) randomly sampled CoCs, and the performance metrics above are computed from evaluation on the remaining 25% (36) CoCs. All models use overlap-based propensity score weights from the model fitted on Table 3.5. Dependent variable is $r_{i,2021} = y^{\text{obs}}_{i,2021} - \widehat{y}^{(1)}_{i,2021}$ (actual minus Phase 1 baseline). Negative R^2 indicates the model performs worse than predicting the mean. XGBoost selected as Phase 2 model.

Table 3.15: Phase 2 Predicted COVID Adjustments: Summary Statistics

	N	Mean	SD	Min	Median	Max
All Imputed CoCs	230	88	332	$-1,068$	-8	1,288
By CoC Category						
Major City CoC	30	146	621	$-1,068$	-66	1,288
Other Largely Urban CoC	31	123	227	-279	30	762
Largely Suburban CoC	92	109	322	-393	-16	1,191
Largely Rural CoC	77	26	185	-766	-1	682

Note: Summary statistics for Phase 2 predicted residuals (COVID adjustments) for the 230 CoCs with missing 2021 unsheltered counts. Residuals represent the predicted deviation from Phase 1 baseline due to COVID-19 effects. Negative values indicate predicted reductions in unsheltered homelessness relative to pre-pandemic baselines. Of the 230 imputed residuals, 122 (53%) are negative. Major City CoCs show the largest average positive adjustments (mean = 146), while Largely Rural CoCs show the smallest (mean = 26). The near-zero overall median (-8) indicates COVID-era forces pushed unsheltered counts both above and below baseline expectations with roughly equal frequency.

Table 3.16: Phase 2 COVID Adjustment Predictions: Random Sample by CoC Category

CoC ID	CoC Name	Baseline	Actual 2021	Actual Resid.	Pred. Resid.
Largely Rural CoC					
AL-507	Alabama Balance of State	168	98	−70	−98
CO-500	Colorado Balance of State	359	528	169	68
MI-511	Lenawee County	16	0	−16	−37
NM-501	New Mexico Balance of State	753	362	−391	−364
TX-604	Waco/McLennan County	119	41	−78	−68
Largely Suburban CoC					
MA-506	Worcester City & County	148	247	99	77
MD-506	Carroll County	21	18	−3	8
NJ-504	Newark/Essex County	452	91	−361	−456
NY-500	Rochester/Monroe County	56	33	−22	2
NY-505	Syracuse/Onondaga Counties	25	12	−13	−2
Major City CoC					
FL-510	Jacksonville-Duval Counties	476	230	−246	−207
FL-600	Miami-Dade County	981	892	−89	−122
KY-501	Louisville-Jefferson County	229	257	28	20
OK-501	Tulsa City & County	375	287	−88	−103
TX-601	Fort Worth/Tarrant County	902	479	−423	−474
Other Largely Urban CoC					
GA-505	Columbus-Muscogee	66	27	−39	−51
IA-500	Sioux City/Woodbury Counties	10	16	6	19
IL-516	Decatur/Macon County	28	31	3	−6
VA-505	Newport News/Virginia Peninsula	90	30	−60	−44
VA-600	Arlington County	33	27	−6	−30

Note: Random sample of 5 CoCs per urbanicity category from the 146 CoCs that conducted complete 2021 unsheltered counts. Baseline = Phase 1 prediction using pre-pandemic relationships. Actual 2021 = observed January 2021 PIT unsheltered count. Actual Resid. = Actual 2021 − Baseline (true COVID adjustment). Pred. Resid. = Phase 2 XGBoost predicted COVID adjustment. The relationship Actual 2021 = Baseline + Actual Resid. holds by construction. The Phase 2 model achieves solid residual predictions in this sample (overall RMSE = 181, R^2 = 0.478). Sample selected via random sampling within category.

What explains the COVID-specific deviations from baseline? Table 3.17 reports feature importance rankings for the Phase 2 XGBoost model. Emergency relief funding dominates: total ESG-CARES allocations account for 57.6 percent of predictive gain, more than four times the contribution of the next most important variable. COVID mortality burden ranks second (12.4 percent of gain), followed by SNAP participation (9.5 percent), policy stringency (8.3 percent), and unemployment (7.6 percent). Local eviction moratoria and Emergency Rental Assistance programs contribute minimally to the model (under 1 percent each), likely because these policies were widespread and exhibited limited cross-sectional variation in 2021. Figure C.3 displays the full feature importance ranking.

Table 3.17: Phase 2: XGBoost Feature Importance (COVID Adjustment Model)

Rank	Feature	Gain
1	ESG-CARES funding (USD)	0.576
2	COVID-19 deaths (2021)	0.124
3	Households with SNAP (%)	0.095
4	Average policy stringency score	0.083
5	Unemployment rate (%)	0.076
6	Has Emergency Rental Assistance	0.009
7	Has eviction moratorium	0.009
8	CoC category: Major City	0.009
9	CoC category: Largely Suburban	0.009
10	CoC category: Largely Rural	0.006
11	CoC category: Other Largely Urban	0.002
12	Has mask mandate	0.001

Note: Feature importance ranked by total gain contribution in the Phase 2 XGBoost residual model. ESG-CARES emergency shelter funding dominates with 57.6% of total gain, more than four times the contribution of the next most important variable (COVID-19 deaths). Local eviction moratoria and Emergency Rental Assistance contribute minimally, likely due to limited cross-sectional variation in 2021.

The Phase 2 model generates residual predictions for all 230 CoCs that did not conduct complete unsheltered counts. The predicted adjustments range from $-1,068$ to 1,288 (median = -8, mean = 88), closely matching the residual distribution in the training sample. However, 122 of the

230 imputed residuals are negative, and for 26 CoCs the predicted negative adjustment exceeds the Phase 1 baseline prediction, yielding a negative final count. These predictions are clamped at zero, as unsheltered populations cannot be negative, but the prevalence of large negative adjustments is a limitation worth noting. The model appears to learn that CoCs with very low baseline unsheltered populations and high emergency relief funding experienced sharp reductions in 2021, but in cases where the baseline was already near zero, the model extrapolates implausibly. This affects primarily small rural CoCs (e.g., NY-514, IL-512) and a handful of suburban jurisdictions (e.g., NJ-514, CA-612).

3.4.3 Final Imputed Unsheltered Counts and National Estimates

The two-phase imputation framework produces a national 2021 unsheltered estimate of 195,191 individuals (90% prediction interval: [114,380, 255,978]), combining 146 observed complete counts with imputed estimates for 230 CoCs that did not conduct complete unsheltered counts. This represents the first comprehensive accounting of unsheltered homelessness during the COVID-19 pandemic, filling a critical gap in the administrative data record and enabling longitudinal analysis of homelessness trends through the crisis period.

Table 3.18 presents the 2015–2022 unsheltered time series. To ensure comparability across years, the observed totals for all non-2020 years are restricted to the consistent sample of 376 non-territory CoCs included in the 2021 analysis. The 2021 imputed estimate falls 13.7 percent below the 2020 observed count of 226,080, suggesting that emergency relief measures continued to suppress unsheltered homelessness through early 2021 even as their coverage began to erode, such as the expanded shelter capacity funded by ESG-CARES, eviction moratoria, and Emergency Rental Assistance. This decline reverses sharply in 2022, when the consistent-sample count surges 25.2 percent to 244,320 as emergency protections fully expired. The 2022 count lies within the upper portion of the 2021 prediction interval, confirming that the two-phase procedure captures the broad trajectory: emergency relief temporarily suppressed unsheltered homelessness below pre-pandemic baselines, but the effect proved transient once protections lapsed.

Table 3.18: National Unsheltered Homelessness Estimates, 2015–2022

Year	Total Unsheltered	Interval	YoY Change	YoY % Change
2015	166,322		—	—
2016	176,900		10,578	6.4
2017	187,835		10,935	6.2
2018	197,215		9,380	5.0
2019	215,682		18,467	9.4
2020	226,080		10,398	4.8
2021	195,191	[114,380, 255,978]	$-30,889$	-13.7
2022	244,320		49,129	25.2

Note: National unsheltered PIT counts restricted to a consistent sample of 376 non-territory CoCs present in the 2021 analysis to ensure comparability across years, with the exception of 2020, which uses official HUD national totals. 2021 estimate combines observed counts from 146 CoCs with imputed estimates for 230 CoCs; 90% prediction interval shown in brackets. YoY = year-over-year.

Figure 3.7 displays the time series visually. The 2021 imputed estimate (shown with 90% prediction interval) falls substantially below 2020, consistent with emergency relief measures suppressing unsheltered homelessness through early 2021. The subsequent 2022 surge places the observed count within the upper range of the 2021 prediction interval, validating the imputed estimate as a plausible midpoint in the post-pandemic trajectory and reinforcing the interpretation that pandemic-era protections produced a temporary, not permanent, reduction in unsheltered homelessness.

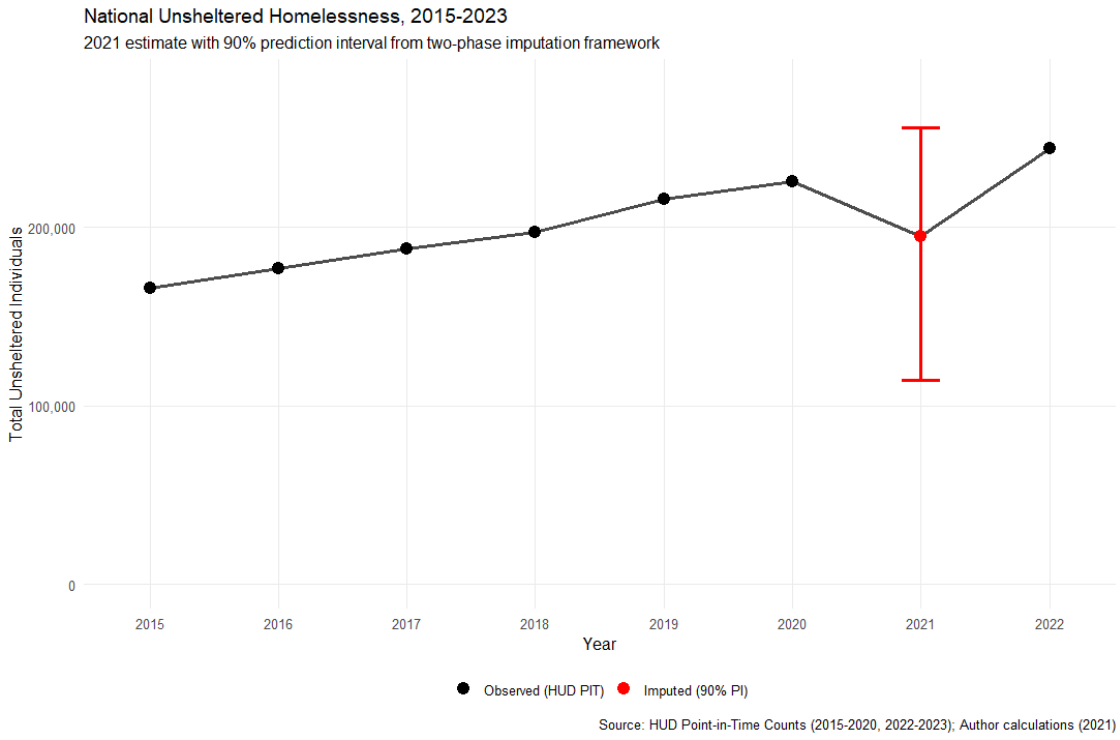
The imputed unsheltered population concentrates heavily in California and major cities. Los Angeles (CA-600) alone accounts for an estimated 28,038 unsheltered individuals—more than the combined total of the bottom 150 CoCs—with a 90% prediction interval spanning [4,466, 35,180]. The interval width of 30,714 reflects extreme uncertainty in both the baseline prediction and the COVID adjustment for the nation's largest unsheltered population, where volatile housing markets and heterogeneous policy responses resist precise prediction. Beyond Los Angeles, California dominates the top of the distribution: nine of the ten largest imputed unsheltered populations are in California, and fourteen of the thirty largest overall (Table 3.19). This concentration underscores the geographic inequality of unsheltered homelessness in the United States, where a handful of jurisdictions bear a disproportionate share of the crisis.

Table 3.19: Top 30 CoCs by Imputed 2021 Unsheltered Count

Rank	CoC ID	Category	Median Estimate
1	CA-600	Major City	28,038
2	CA-502	Major City	5,316
3	CA-601	Major City	5,046
4	CA-503	Major City	4,392
5	CA-500	Major City	4,417
6	CA-602	Largely Suburban	4,386
7	CA-501	Major City	4,279
8	TX-607	Largely Rural	3,859
9	CA-608	Largely Suburban	3,166
10	WA-500	Major City	3,601
11	CA-514	Major City	2,877
12	NY-600	Major City	2,383
13	OR-501	Major City	2,376
14	WA-501	Largely Rural	2,202
15	NV-500	Major City	2,142
16	HI-501	Largely Suburban	1,842
17	TX-503	Major City	2,238
18	CA-510	Largely Suburban	2,201
19	CA-504	Other Urban	2,180
20	CA-506	Largely Rural	2,167
21	CA-604	Major City	1,995
22	OR-500	Other Urban	1,990
23	GA-501	Largely Suburban	1,897
24	CA-611	Other Urban	1,830
25	CA-508	Largely Suburban	1,694
26	CA-603	Largely Suburban	1,628
27	AZ-502	Major City	2,823
28	CA-511	Largely Suburban	2,384
29	CA-505	Largely Suburban	2,350
30	CA-609	Largely Suburban	3,391

Note: Thirty largest imputed 2021 unsheltered populations (bootstrap median). Los Angeles accounts for 28,038 individuals. Twenty-three of thirty are in California or the West. Category labels abbreviated. Mix of observed and imputed counts; table focuses on imputed CoCs to illustrate geographic concentration.

Figure 3.7: National Unsheltered Homelessness Time Series, 2015–2022



Prediction intervals quantify the uncertainty inherent in the imputed estimates and vary dramatically across the CoC distribution. The median 90% prediction interval width is 259, indicating moderate uncertainty for typical CoCs, but the distribution is severely right-skewed: the mean width is 671, and the maximum reaches 30,714 (Table 3.20). Uncertainty varies systematically by CoC type. Major City CoCs exhibit the widest intervals (median width = 813), reflecting both larger population scales and greater volatility in unsheltered counts, where housing markets are tighter and policy responses more heterogeneous. Largely Rural and Largely Suburban CoCs have narrower intervals (median widths of 221 and 222, respectively), consistent with smaller baseline populations and more stable COVID effects.

The widest prediction intervals concentrate overwhelmingly in California and the West (Table 3.21). Los Angeles dominates with an interval width of 30,714, an order of magnitude larger than any other CoC. The remaining wide intervals cluster in California major cities (San Jose, Oakland, San Francisco, Sacramento, San Diego) and western urban centers (Seattle, Las Vegas), where large baseline unsheltered populations combine with substantial COVID adjustments to

Table 3.20: Prediction Interval Widths by CoC Category

CoC Category	N	Median Width	Mean Width	Max Width
Major City CoC	30	813	2,335	30,714
Other Largely Urban CoC	31	245	326	725
Largely Suburban CoC	92	222	367	1,913
Largely Rural CoC	77	221	359	4,046
Overall	230	259	671	30,714

Note: Distribution of 90% prediction interval widths for the 230 imputed CoCs, stratified by urbanicity category. Width = $p_{95} - p_{05}$ from end-to-end clustered bootstrap (B = 300). Major City CoCs exhibit substantially wider intervals (median = 813) than other categories, reflecting both larger population scales and greater volatility in COVID effects.

produce extreme uncertainty. The narrowest intervals, conversely, concentrate in small suburban and rural CoCs in the Northeast and Midwest (Table 3.22), where baseline predictions are low and COVID adjustments near zero. One CoC (NY-523: Glens Falls, Saratoga Springs) has a degenerate interval—width of zero—because the bootstrap consistently predicts zero unsheltered individuals in every draw, reflecting a small baseline and large predicted negative COVID adjustment that the zero-clamping enforces uniformly.

Table 3.21: CoCs with Widest 90% Prediction Intervals

CoC ID	CoC Name	Category	p_{50}	p_{95}	Width
CA-600	Los Angeles City & County	Major City	28,038	35,180	30,714
CA-500	San Jose/Santa Clara	Major City	4,417	6,906	4,707
CA-502	Oakland, Berkeley/Alameda	Major City	5,316	6,734	4,110
TX-607	Texas Balance of State	Largely Rural	3,859	5,107	4,046
WA-500	Seattle/King County	Major City	3,601	4,973	4,013
CA-501	San Francisco	Major City	4,279	5,717	3,517
NV-500	Las Vegas/Clark County	Major City	2,142	3,027	2,364
CA-503	Sacramento City & County	Major City	4,392	5,175	2,247
CA-601	San Diego City and County	Major City	5,046	6,008	1,949
CA-602	Orange County	Largely Suburban	4,386	5,438	1,913

Note: Ten CoCs with widest 90% prediction intervals. Los Angeles dominates with width of 30,714. Nine of ten are in California or the West. Width = $p_{95} - p_{05}$. CoC names abbreviated for space.

Table 3.22: CoCs with Narrowest 90% Prediction Intervals

CoC ID	CoC Name	Category	p_{50}	p_{95}	Width
NY-523	Glens Falls, Saratoga Springs	Largely Suburban	0	0	0
MN-508	Moorhead/West Central MN	Largely Rural	0	6	6
MI-523	Eaton County	Largely Suburban	0	17	17
PA-603	Beaver County	Largely Suburban	0	23	23
MN-511	Southwest Minnesota	Largely Rural	0	28	28
IL-508	East St. Louis, Belleville	Largely Suburban	0	32	32
IL-502	Waukegan, North Chicago	Largely Suburban	0	33	33
NY-520	Franklin, Essex Counties	Largely Rural	0	35	35
NY-525	New York Balance of State	Largely Rural	0	36	36
MD-504	Howard County	Largely Suburban	11	37	37

Note: Ten CoCs with narrowest 90% prediction intervals. NY-523 has degenerate interval (width = 0) because bootstrap consistently predicts zero. Narrow intervals concentrate in small suburban and rural CoCs in Northeast and Midwest. CoC names abbreviated.

An important caveat warrants emphasis: these are prediction intervals, not confidence intervals. Prediction intervals quantify uncertainty in individual CoC-level imputations and include both parameter estimation error and irreducible forecast error. They cannot be aggregated across CoCs without additional assumptions. The national-level interval reported above (90% PI: [114,380, 255,978]) is constructed by summing the 5th and 95th percentiles of the bootstrap distribution across all imputed CoCs, which treats CoC-level prediction errors as independent. This is conservative but likely overstates national uncertainty, since common shocks—federal relief policy, nationwide policy stringency—induce positive correlation in prediction errors across CoCs, partially offsetting in aggregate. Nonetheless, even under conservative assumptions, the 2021 imputed national unsheltered count is precise enough to support policy analysis: the interval spans 141,598 individuals, or 73 percent of the point estimate, comparable to the uncertainty in large-scale household surveys and substantially narrower than the alternative of dropping 2021 from longitudinal analyses entirely.

3.5 Discussion

The 2021 Point-in-Time count, disrupted by COVID-19, left 61.6 percent of Continuums of Care without complete unsheltered enumeration data. This paper develops a two-phase machine learning framework to fill that gap, producing the first comprehensive national estimate of unsheltered homelessness during the pandemic. The imputed national unsheltered count of 195,191 individuals (90% prediction interval: [114,380, 255,978]), combined with the observed sheltered count of 326,126, yields a reconstructed total of 521,317 people experiencing homelessness in January 2021. Beyond filling a critical data gap, the methodology demonstrates how to handle missing administrative data when both the missingness mechanism and the outcome process are altered by crisis events.

3.5.1 Key Findings

Three substantive findings emerge from the analysis. First, emergency relief funding overwhelmingly explains how the pandemic reshaped local unsheltered homelessness. ESG-CARES allocations account for 57.6 percent of the Phase 2 model's predictive gain, more than four times the next most important variable. Meanwhile, COVID mortality, policy stringency, and economic stress play comparatively minor roles. The implication is that direct shelter funding, not the severity of the health crisis itself, determined whether a community's unsheltered population grew or shrank relative to pre-pandemic levels. This extends evidence from the Great Recession (Popov, 2016) into a fundamentally different crisis setting: whereas conventional recessions displace people through job loss and foreclosure, the pandemic simultaneously destroyed livelihoods and deployed unprecedented countervailing resources. That targeted relief appears to have buffered unsheltered homelessness even under these conditions suggests emergency shelter investment operates through channels robust to the specific mechanism of economic disruption.

Second, the pandemic amplified the geographic concentration of unsheltered homelessness rather than dispersing it. California jurisdictions account for nine of the ten largest imputed unsheltered populations, and Los Angeles alone exceeds the combined total of the bottom 150 CoCs.

This pattern is not simply an artifact of California's pre-existing crisis. ESG-CARES allocations followed existing CoC grant formulas that weight population and poverty but do not scale with baseline unsheltered severity, meaning that jurisdictions already bearing outsized burdens received relief proportional to their general need rather than their crisis magnitude. The result is a federal funding structure that, during an emergency, mechanically widens the gap between high-burden and low-burden communities.

Third, the imputed 2021 estimate sits plausibly within the trajectory defined by adjacent observed counts. The national unsheltered total declines 13.7 percent from 2020, consistent with emergency protections still operating in early 2021, then rebounds sharply by 2022 as those protections expired. This V-shaped pattern is difficult to reconcile with alternative explanations such as measurement artifact or model overfitting, since it aligns with the known timeline of federal relief expiration and independently corroborates the Phase 2 finding that emergency funding drove short-run reductions. The 2022 observed count falls within the upper portion of the 2021 prediction interval, providing external validation that the two-phase framework captures the correct order of magnitude and directional trend even if point estimates carry substantial uncertainty.

3.5.2 Methodological Contributions

The central methodological contribution is a framework for imputing missing administrative data when the standard missing-at-random assumption fails in two simultaneous ways: the shock that causes missingness also changes the quantity being measured, and the units that do report are not representative of those that do not. Neither problem alone is unusual: selection models address non-random missingness, and structural break methods handle regime changes. However, their combination creates a setting where no single-stage approach suffices. A model trained on pre-crisis data recovers the counterfactual, not the actual outcome. A model trained on crisis-year responders extrapolates from a biased sample. The two-phase decomposition addresses both problems by assigning each to the estimation stage where it can be most credibly handled.

This decomposition carries a broader lesson for applied work with disrupted administrative data. The logic does not depend on the specific context of homelessness counts or COVID-19. Any setting where a crisis event simultaneously interrupts routine data collection and alters the data-generating process shares the same structure: vital statistics registration after natural disasters, census enumeration during civil conflict, and labor force surveys suspended by pandemics. In each case, pre-disruption panel data can anchor the counterfactual, and partial observations from the disruption period can identify the deviation. The two-phase template is agnostic to the domain; what matters is that the researcher can credibly separate stable structural relationships from transient shock effects and has at least some crisis-year observations to learn from.

Three technical choices strengthen the framework's credibility in practice. The overlap-based propensity score weighting in Phase 2 directly confronts the selection problem: CoCs that conducted full counts in 2021 skew smaller, less urban, and less affected by COVID restrictions than those that did not, and naive residual extrapolation would systematically mischaracterize the adjustment for non-counting CoCs. Reweighting cannot eliminate bias from unobservables, but balance diagnostics confirm that it substantially reduces observable covariate shift. The end-to-end clustered bootstrap treats the entire two-stage pipeline as a single composite estimator, propagating uncertainty from Phase 1 into Phase 2 rather than conditioning on baseline predictions as though they were known. This is conservative by design, as it produces wider intervals than stage-wise approaches, but it avoids the well-documented problem of understating imputation uncertainty when estimation error in early stages is ignored (Rubin, 1987). Finally, the systematic model comparison across linear and tree-based methods disciplines the functional form choice. XGBoost dominates in both phases, but the margin of improvement over linear models differs starkly: modest in Phase 2 (where the small effective sample constrains complexity) and dramatic in Phase 1 (where nonlinearities in climate, shelter capacity, and demographics are strong). Reporting the full model comparison makes this choice transparent rather than stipulated.

3.5.3 Limitations

Five limitations warrant emphasis. First, selection bias in Phase 2 is mitigated but not eliminated. Propensity score reweighting reduces observable covariate imbalance between counting and non-counting CoCs, but the effective sample size of 42.8 after weighting signals that only a fraction of the 146 full-count CoCs carry meaningful weight in the residual model. More fundamentally, the decision to conduct an unsheltered count during a pandemic likely reflects unobserved institutional characteristics that plausibly correlate with how effectively a community deployed emergency resources, such as administrative capacity, political will, volunteer infrastructure, organizational culture around data collection. If CoCs that counted also happened to manage pandemic response more competently, the residual model would learn an adjustment pattern systematically different from what non-counting CoCs actually experienced, and no amount of reweighting on observables can correct this.

Second, the imputed estimates are modeled quantities, not observed data, and should be treated accordingly. The national prediction interval spans roughly 73 percent of the point estimate, reflecting genuine uncertainty that cannot be wished away through methodological sophistication. Researchers incorporating these estimates into downstream analyses—panel regressions, trend decompositions, funding simulations—should propagate this uncertainty rather than treating imputed values as equivalent to observed counts. Practical approaches include conducting sensitivity analysis at the prediction interval bounds, using multiple draws from the bootstrap distribution as parallel datasets, or restricting analyses to CoCs where interval widths are narrow enough to support the inferential claims being made.

Third, the zero-clamping of negative predictions introduces asymmetric bias at the bottom of the distribution. Nineteen CoCs receive imputed counts of zero because the Phase 2 model predicts a negative adjustment larger than the Phase 1 baseline. While zero unsheltered counts do occur in PIT data (15 CoCs reported exactly zero in 2019), the clamping mechanically prevents the model from distinguishing genuine near-zero populations from cases where the additive structure of the two-phase framework overshoots. This affects small rural and suburban CoCs where baselines

are already low and ESG-CARES allocations are high relative to population, and it biases the national total modestly downward. The alternative of allowing negative counts would be incoherent, making the floor at zero a defensible, if imperfect, compromise.

Fourth, the imputation inherits every limitation of the underlying PIT count methodology. Volunteer-driven street enumeration on a single January night systematically undercounts unsheltered individuals, exhibits high measurement error, and varies in quality across jurisdictions (O'Flaherty, 2019; Meyer et al., 2023). The framework predicts what a PIT count would have found, not the true size of the unsheltered population: a distinction that matters for interpreting both the point estimates and the prediction intervals. If a CoC's pre-pandemic counts systematically undercounted by 40 percent, the imputed 2021 value will reproduce that undercount, not correct it.

Fifth, the analysis is predictive, not causal. The dominance of ESG-CARES funding in the Phase 2 model is an associational finding: communities that received more emergency shelter funding saw smaller deviations from baseline, but this correlation reflects both the direct effect of funding and the confounded selection process through which funding was allocated. CoCs receiving large ESG-CARES awards tend to have stronger existing shelter infrastructure, more experienced grant administrators, and greater political engagement with HUD: all factors that could independently suppress unsheltered homelessness during a crisis. Credible causal identification would require exploiting exogenous variation in funding, such as discontinuities in formula-based allocation rules or quasi-random timing of grant disbursement, which lies beyond the present scope.

3.5.4 Future Directions

The most immediate application of these estimates is substantive rather than methodological. The 2021 data gap has forced researchers studying pandemic-era homelessness to either drop the year entirely, restrict analysis to the non-representative subset of CoCs that counted, or treat sheltered counts as a proxy for total homelessness. The imputed estimates remove this constraint. With a complete 2021 cross-section in hand, researchers can trace unsheltered homelessness tra-

jectories through the full pandemic arc, decompose the 2022 surge into a rebound component and a new-inflow component, and test whether communities that received more emergency funding experienced durably lower homelessness or merely delayed an inevitable increase. The prediction intervals make it possible to distinguish findings that are robust to imputation uncertainty from those that depend on taking the point estimates at face value.

The framework itself invites extension in at least two directions. First, applying the same two-phase procedure to other partially disrupted PIT count years, such as sporadic missingness in non-pandemic years, would produce a more complete national panel and allow researchers to study whether the structural relationships estimated in Phase 1 remain stable over time or drift in ways that affect imputation quality. Second, the current framework treats each CoC independently in both phases. Spatial models that allow prediction errors to correlate across neighboring CoCs could tighten national uncertainty bounds, since common regional shocks induce dependence that the current bootstrap treats as independent noise. Hierarchical Bayesian approaches offer a related avenue: partial pooling across CoCs would borrow strength from data-rich jurisdictions to improve estimates for data-poor ones, a particularly attractive property when the effective sample size after propensity score weighting is small.

Beyond homelessness, the methodological template applies wherever administrative data collection is interrupted by the same event that changes the quantity being measured. The specific technical choices are modular and can be adapted to different data structures and sample sizes. What is less clear, and worth formalizing in future work, is when the two-phase decomposition actually improves over simpler alternatives. If the crisis-year deviation is small relative to baseline prediction error, a single-stage model may suffice. If the responding sample is so selected that no reweighting scheme achieves adequate balance, the Phase 2 adjustment may introduce more bias than it corrects. Developing diagnostic criteria for these boundary conditions (when to decompose and when not to) would make the framework more practically useful for applied researchers confronting the next disrupted dataset.

3.5.5 Conclusion

The 2021 PIT count disruption left a hole in the national homelessness record at precisely the moment that record mattered most. This paper fills it, imperfectly but transparently. The two-phase framework produces a national unsheltered estimate of 195,191 individuals (90% prediction interval: [114,380, 255,978]) and, more importantly, makes the uncertainty in that estimate explicit rather than leaving it as an unquantified absence in the data. The substantive picture that emerges is consistent with the known timeline of federal intervention and independently corroborated by adjacent observed counts: emergency relief temporarily suppressing unsheltered homelessness before a sharp post-pandemic rebound.

The deeper contribution is methodological. Missing administrative data during crises is not a one-time inconvenience; it is a recurring structural problem. Climate disasters disrupt vital statistics systems. Armed conflicts prevent census enumeration. Future pandemics will again force trade-offs between data collection and public safety. In each case, the missingness is entangled with the very phenomenon being measured, violating the assumptions on which standard imputation methods rest. The two-phase decomposition, anchoring a counterfactual in pre-crisis data and modeling the crisis-specific deviation from partial observations, offers one principled response. It will not be the only response needed, but it demonstrates that the choice between accepting a permanent data gap and pretending the gap does not exist is a false one. With appropriate uncertainty quantification and honest acknowledgment of limitations, defensible reconstruction is possible.

Bibliography

Abramson, B. (2024). The equilibrium effects of eviction policies. Working paper.

Aliprantis, D., Martin, H., and Phillips, D. (2019). Can landlords be paid to stop avoiding voucher tenants? Working Paper 19-02, Federal Reserve Bank of Cleveland, Cleveland, OH.

Appelbaum, R. P., Dolny, M., Dreier, P., and Gilderbloom, J. I. (1991). Scapegoating rent control: Masking the causes of homelessness. *Journal of the American Planning Association*, 57(2):153–164.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.

Bahr, H. M. and Caplow, T. (1968). Homelessness, affiliation, and occupational mobility. *Social Forces*, 47(1):28–33.

Basolo, V. and Nguyen, M. T. (2005). Does mobility matter? the neighborhood conditions of housing voucher holders by race and ethnicity. *Housing Policy Debate*, 16(3–4):297–324.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Bunce, H. L. (1979). An evaluation of the community development block grant formula. *Urban Affairs Quarterly*, 14(4):491–510.

Burt, M. R., Aron, L. Y., Douglas, T., Valente, J., Lee, E., and Iwen, B. (1999). Homelessness: Programs and the people they serve. Findings of the National Survey of Homeless Assistance Providers and Clients.

Byrne, T., Fargo, J. D., Montgomery, A. E., Munley, E., and Culhane, D. P. (2014). The relationship between community investment in permanent supportive housing and chronic homelessness. *Social Service Review*, 88(2):234–263.

Byrne, T., Munley, E. A., Fargo, J. D., Montgomery, A. E., and Culhane, D. P. (2013). New perspectives on community-level determinants of homelessness. *Journal of Urban Affairs*, 35(5):607–625.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Chen, S. and Xu, C. (2025). On the use of machine learning methods for missing data problems. In *Handbook of Statistics*, volume 53, pages 161–174. Elsevier.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM.

Chien, J., Henwood, B. F., St. Clair, P., Kwack, S., and Kuhn, R. (2024). Predicting hotspots of unsheltered homelessness using geospatial administrative data and volunteered geographic information. *Health & Place*, 88:103267.

Cho, S. and Lucio, J. (2025). Vouchers welcome here: Source of income antidiscrimination laws in american cities. *Housing Policy Debate*, pages 1–21.

Christopher, D. A., Duggan, M., and Martin, O. H. (2025). The social and individual effects of homeless shelter: Evidence from temporary shelter provision. Working Paper 34376, National Bureau of Economic Research.

Collinson, R. and Ganong, P. (2018). How do changes in housing voucher design affect rent and neighborhood quality? *American Economic Journal: Economic Policy*, 10(2):62–89.

Collinson, R. A. (2014). Assessing the allocation of cdbg to community development need. *Housing Policy Debate*, 24(1):91–118.

Congressional Research Service (2014). Community development block grants: Recent funding history. Technical report, Congressional Research Service. Analyst in Federalism and Economic Development Policy.

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

Corinth, K. and Lucas, D. S. (2018). When warm and cold don't mix: The implications of climate for the determinants of homelessness. *Journal of Housing Economics*, 41(C):45–56.

Coulson, N. E., Le, T., Ortego-Marti, V., and Shen, L. (2025). Tenant rights, eviction, and rent affordability. *Journal of Urban Economics*, 147:103762.

Cragg, M. and O'Flaherty, B. (1999). Do homeless shelter conditions determine shelter population? the case of the dinkins deluge. *Journal of Urban Economics*, 46(3):377–415.

Culhane, D., Fowler, M., and Moses, J. (2025). How much would it cost to provide housing first to all households staying in homeless shelters? Technical report, National Alliance to End Homelessness.

Culhane, D. P. and An, S. (2021). Estimated revenue of the nonprofit homeless shelter industry in the united states: Implications for a more comprehensive approach to unmet shelter demand. *Housing Policy Debate*, 32(6):823–836.

Culhane, D. P., Metraux, S., and Byrnes, M. (2011). A prevention-centered approach to homelessness assistance: A paradigm shift? *Housing Policy Debate*, 21(2):295–315.

Culhane, D. P., Metraux, S., and Hadley, T. (2002). Public service reductions associated with placement of homeless persons with severe mental illness in supportive housing. *Housing Policy Debate*, 13(1):107–163.

Cunningham, M., Galvez, M., Aranda, C. L., Santos, R., Wissoker, D., Oneto, A., Pitingolo, R., and Crawford, J. (2018). A pilot study of landlord acceptance of housing choice vouchers.

Technical report, U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Washington, DC.

Dang, H.-A., Kilic, T., Hlasny, V., Abanokova, K., and Carletto, C. (2026). Using survey-to-survey imputation to fill poverty data gaps at a low cost: Evidence from a randomized survey experiment. *The World Bank Economic Review*. lhaf037.

de Sousa, T., Andrichik, A., Cuellar, M., Marson, J., Prestera, E., and Rush, K. (2023). The 2022 annual homeless assessment report (AHAR) to congress. Technical report, U.S. Department of Housing and Urban Development, Office of Community Planning and Development. Prepared by Abt Associates.

DeLuca, S. (2014). Why don't more voucher holders escape poor neighborhoods? Technical report, NYU Furman Center, New York.

Desmond, M. and Perkins, K. L. (2016). Are landlords overcharging housing voucher holders? *City & Community*, 15(2):137–162.

Devine, D. J., Gray, R. W., Rubin, L., and Tahiti, L. B. (2003). Housing choice voucher location patterns: Implications for participant and neighborhood welfare. Technical report, U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Washington, DC.

Downing, N. J. (2025). Missing value imputation in environmental, social, and governance data: An impact on emissions scores. *Finance Research Letters*, 85:107818.

Durre, I., Squires, M. F., Vose, R. S., Arguez, A., Gross, W. S., Rennie, J. R., and Schreck, C. J. (2022). NOAA's nClimGrid-Daily Version 1 – Daily gridded temperature and precipitation for the Contiguous United States since 1951. Technical report / data set, NOAA National Centers for Environmental Information.

Early, D. W. (2004). The determinants of homelessness and the targeting of housing assistance. *Journal of Urban Economics*, 55(1):195–214.

Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press, Cambridge.

Elias, G. (1952). A measure of "homelessness". *The Journal of Abnormal and Social Psychology*, 47(1):62–66.

Ellen, I. G. (2020). What do we know about housing choice vouchers? *Regional Science and Urban Economics*, 80:103380.

Ellen, I. G., O'Regan, K. M., and Harwood, K. W. (2022). Advancing choice in the housing choice voucher program: Source of income protections and locational outcomes. *Housing Policy Debate*, 33(4):941–962.

Ellen, I. G., O'Regan, K. M., and Strochak, S. (2025). Pricing for opportunity: The impact of spatially varying rent subsidies on housing voucher neighborhoods and take-up. *Journal of Public Economics*, 250(105465).

Elliott, M. and Krivo, L. J. (1991). Structural determinants of homelessness in the united states. *Social Problems*, 38(1):113–131.

Embaye, W. T., Zereyesus, Y. A., and Chen, B. (2021). Predicting the rental value of houses in household surveys in Tanzania, Uganda and Malawi: Evaluations of hedonic pricing and machine learning approaches. *PLOS ONE*, 16(2):e0244953.

Ferrara, A., Testa, P. A., and Zhou, L. (2024). New area- and population-based geographic crosswalks for u.s. counties and congressional districts, 1790–2020. Working Paper 32206, National Bureau of Economic Research.

Freeman, L. (2012). The impact of source of income laws on voucher utilization. *Housing Policy Debate*, 22(2):297–319.

Freeman, L. and Li, Y. (2014). Do source of income anti-discrimination laws facilitate access to less disadvantaged neighborhoods? *Housing Studies*, 29(1):88–107.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.

Galvez, M. (2010). What do we know about housing choice voucher program location outcomes? Technical report, Urban Institute, Washington, DC.

Galvez, M. and Knudsen, B. (2024). Discrimination against voucher holders and the laws to prevent it: Reviewing the evidence on source of income discrimination. *Cityscape: A Journal of Policy Development and Research*, 26(2):145.

Glasser, I., Hirsch, E., and Chan, A. (2014). Reaching and enumerating homeless populations: Geospatial methods and promising practices. *Cityscape*, 16(3):167–188.

Glynn, C. and Fox, E. B. (2019). Dynamics of homelessness in urban America. *The Annals of Applied Statistics*, 13(1):573–605.

Gorzig, M. M. and Rho, D. (2025). The impact of renter protection policies on rental housing discrimination. *Contemporary Economic Policy*.

Government Accountability Office (2020). Homelessness: Better HUD oversight of data collection could improve estimates of homeless population. *Report GAO-20-433*.

Government Accountability Office (2021). Homelessness: HUD should help communities better leverage data to estimate homelessness. *Report GAO-22-104445*.

Graves, E. (2016). Rooms for improvement: A qualitative metasynthesis of the housing choice voucher program. *Housing Policy Debate*, 26(2):346–361.

Hangen, F. (2022). The choice to discriminate: How source of income discrimination constrains opportunity for housing choice voucher holders. *Urban Affairs Review*, 59(5).

Hanratty, M. (2017). Do local economic conditions affect homelessness? impact of area housing market factors, unemployment, and poverty on community homeless rates. *Housing Policy Debate*, 27(4):1–16.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5(4):475–492.

Henry, M., de Sousa, T., Tano, C., Dick, N., Hull, R., Shea, M., Morris, T., and Morris, S. (2022). The 2021 annual homeless assessment report (AHAR) to congress. Technical report, U.S. Department of Housing and Urban Development, Office of Community Planning and Development. Prepared by Abt Associates.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3):199–236.

Honig, M. and Filer, R. K. (1993). Causes of intercity variation in homelessness. *American Economic Review*, 83(1):248–255.

Hopper, K., Shinn, M., Laska, E., Meisner, M., and Wanderling, J. (2008). Estimating numbers of unsheltered homeless people through plant-capture and postcount survey methods. *American Journal of Public Health*, 98(8):1438–1442.

Jencks, C. (1994). *The Homeless*. Harvard University Press.

Khadduri, J., Leopold, J., Spellman, B. E., and Sokol, B. (2010). Costs associated with first-time homelessness for families and individuals. Report 233, U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Prepared by Abt Associates.

Lamb, H. R. (1984). Deinstitutionalization and the homeless mentally ill. *Psychiatric Services*, 35(9).

Lens, M. C., Ellen, I. G., and O'Regan, K. (2011). Do vouchers help low-income households live in safer neighborhoods? evidence on the housing choice voucher program. *Cityscape*, 13(3):135–159.

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

Lucio, J. and Cho, S. (2025). Can i (still) refuse housing vouchers? source of income protections and landlord strategies. *Journal of Urban Affairs*, pages 1–16.

McClure, K., Schwartz, A. F., and Taghavi, L. B. (2015). Housing choice voucher location patterns a decade later. *Housing Policy Debate*, 25(2):215–233.

Meyer, B. D., Wyse, A., and Corinth, K. (2023). The size and census coverage of the u.s. homeless population. *Journal of Urban Economics*, 136:103559.

Meyer, B. D., Wyse, A., Grunwaldt, A., Medalia, C., and Wu, D. (2021). Learning about homelessness using linked survey and administrative data. Working Paper 28861, National Bureau of Economic Research.

Miller, G. and Richardson, T. (2024). An evaluation of the cdbg formula's targeting to community development need. Technical report, U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Mondello, M., Gass, A., McLaughlin, T., and Shore, N. (2007). Cost of homelessness: Cost analysis of permanent supportive housing, state of maine, greater portland. Technical report, Corporation for Supportive Housing.

Moulton, S. (2013). Does increased funding for homeless programs reduce chronic homelessness? *Southern Economic Journal*, 79(3):600–620.

National Law Center on Homelessness and Poverty (2017). Don't count on it: How the HUD point-in-time count underestimates the homelessness crisis in america.

Nisar, H., Vachon, M., Horseman, C., and Murdoch, J. (2019). Market predictors of homelessness: How housing and community factors shape homelessness rates within continuums of care. Technical report, U.S. Department of Housing and Urban Development.

O'Flaherty, B. (1995). An economic theory of homelessness and housing. *Journal of Housing Economics*, 4:13–49.

O'Flaherty, B. (2004). Wrong person and wrong place: For homelessness, the conjunction is what matters. *Journal of Housing Economics*, 13(1):1–15.

O'Flaherty, B. (2019). Homelessness research: A guide for economists (and friends). *Journal of Housing Economics*, 44:1–25.

Pendall, R., Dawkins, C., Blumenberg, E., Hayes, C., George, A., McDade, Z., Jeon, J. S., Knaap, E., Pierce, G., and Smart, M. (2014). Driving to opportunity: Understanding the links among transportation access, residential outcomes, and economic opportunity for housing voucher recipients. Technical report, Urban Institute, Washington, DC.

Perlman, J. and Parvensky, J. (2006). Denver housing first collaborative cost benefit analysis and program outcomes report. Technical report, Colorado Coalition for the Homeless, Denver, CO.

Phillips, D. C. (2017). Landlords avoid tenants who pay with vouchers. *Economics Letters*, 151:48–52.

Popov, I. (2016). Homeless programs and social insurance. Working paper, Stanford Institute for Economic Policy Research.

Porreca, Z. (2022). Synthetic difference-in-differences estimation with staggered treatment timing. *Economics Letters*, 220:110874.

Rog, D. J., Marshall, T., Dougherty, R. H., George, P., Daniels, A. S., Ghose, S. S., and Delphin-Rittmon, M. E. (2014). Permanent supportive housing: Assessing the evidence. *Psychiatric Services*, 65(3):287–294.

Rosen, E. (2014). Rigging the rules of the game: How landlords geographically sort low-income renters. *City & Community*, 13(4):310–340.

Rosen, E. (2020). *The Voucher Promise: "Section 8" and the Fate of an American Neighborhood*. Princeton University Press, Princeton, NJ.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Ruhnke, S. A., Wilson, F. A., and Stimpson, J. P. (2022). Using machine learning to impute legal status of immigrants in the National Health Interview Survey. *MethodsX*, 9:101848.

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

Schwemm, R. (2016). Source-of-income discrimination and the fair housing act. *Case Western Reserve Law Review*, 70(1):573–600.

Smith, A. C., Holmberg, C., and Jones-Puthoff, M. (2012). Emergency and transitional shelter population: 2010.

Stefancic, A. and Tsemberis, S. (2007). Housing first for long-term shelter dwellers with psychiatric disabilities in a suburban county: A four-year study of housing access and retention. *The Journal of Primary Prevention*, 28(3):265–279.

Stergiopoulos, V., Hwang, S. W., Gozdzik, A., Nisenbaum, R., Latimer, E., Rabouin, D., Adair, C. E., Bourque, J., Connelly, J., Frankish, J., Katz, L. Y., Mason, K., Misir, V., O'Brien, K., Sareen, J., Schütz, C. G., Singer, A., Streiner, D. L., Vasiliadis, H.-M., and Goering, P. N. (2015). Effect of scattered-site housing using rent supplements and intensive case management on housing stability among homeless adults with mental illness. *Journal of the American Medical Association*, 313(9):905–915.

Teles, D. and Su, Y. (2022). Source of income protections and access to low poverty neighborhoods. Brief, Urban Institute, Washington, DC.

Tighe, J. R., Hatch, M. E., and Mead, J. (2016). Source of income discrimination and fair housing policy. *Journal of Planning Literature*, 32(1):3–15.

Tsai, J. and Alarcón, J. (2022). The annual homeless point-in-time count: Limitations and two different solutions. *American Journal of Public Health*, 112(4):633–637.

Tsemberis, S. and Eisenberg, R. F. (2000). Pathways to housing: Supported housing for street-dwelling homeless individuals with psychiatric disabilities. *Psychiatric Services*, 51(4):487–493.

Turner, M. A., Popkin, S. J., and Cunningham, M. (1999). Section 8 mobility and neighborhood health: Emerging issues and policy challenges. Technical report, The Urban Institute, Washington, DC.

Unlock NYC, Neighbors Together, Anti-Eviction Mapping Project, and Housing Data Coalition (2022). An illusion of choice: How source of income discrimination and voucher policies perpetuate housing inequality.

U.S. Department of Housing and Urban Development (2012). Community planning and development: Continuum of care program interim rule. *Federal Register*, 77(213):65806–65879. 24 CFR 578.

U.S. Department of Housing and Urban Development (2014). *Point-in-Time Count Methodology Guide*. Office of Community Planning and Development.

U.S. Department of Housing and Urban Development (2019a). Annual homeless assessment report (AHAR) to congress. Accessed 2025.

U.S. Department of Housing and Urban Development (2019b). Coc homeless populations and subpopulations reports. HUD Exchange. Accessed 2025.

U.S. Department of Housing and Urban Development (2019c). Emergency solutions grant program requirements. HUD Exchange. Accessed 2025.

U.S. Department of Housing and Urban Development (2020). Coc program awards and allocations. HUD Exchange. Accessed 2025.

U.S. Department of Housing and Urban Development (2021). Notice of funding opportunity (NOFO) for fiscal year (FY) 2021 continuum of care competition and noncompetitive award of youth homeless demonstration program renewal and replacement grants. Technical Report FR-6500-N-25, U.S. Department of Housing and Urban Development, Community Planning and Development.

U.S. Department of Housing and Urban Development (2023). *PIT and HIC Data Collection Notice*.

U.S. Department of Housing and Urban Development (2024). *2024 CoC Program Competition: HIC and PIT Data Collection Guidance*. HUD Exchange.

U.S. Department of Housing and Urban Development (2025). Housing Choice Vouchers. https://www.hud.gov/helping-americans/housing-choice-vouchers. Accessed: 2025-3-01.

Varady, D. P., Jaroscak, J., and Kleinhans, R. (2017). How to attract more landlords to the housing choice voucher program: A case study of landlord outreach efforts. *Urban Research & Practice*, 10(2):143–155.

Ward, J. M., Garvey, R., and Hunter, S. B. (2024). Annual trends among the unsheltered in three los angeles neighborhoods: The los angeles longitudinal enumeration and demographic survey (LA LEADS) 2023 annual report. Research report rr-a1890-4, RAND Corporation.

Wong, Y.-L. I., Culhane, D. P., and Kuhn, R. (1997). Predictors of exit and reentry among family shelter users in new york city. *Social Service Review*, 71(3):441–462.

Yates, G. L., MacKenzie, R., Pennbridge, J., and Cohen, E. (1988). A risk profile comparison of runaway and non-runaway youth. *American Journal of Public Health*, 78(7):820–821.

Appendix A

Supplementary Material for Chapter 1

A.1 Chapter 1 Appendix

A.1.1 Example of "No Section 8" ad

Figure A.1: "No Section 8" Craigslist Housing Listings

$3,000 / 4br - 1600ft^2 - Pet Friendly [redacted] Home [redacted]

[redacted]

4 bd 2 ba pet friendly home with HUGE backyard just a short walk t[redacted] Located in the very lively and active University neighborhood, this home is perfect for grad students and families alike. Fresh paint & new carpet.

Pets ok! Possible rental time frame can be less than 1 year.

No Section 8.

$1,095 / 2br - *Move-In Special* 2 Bedroom 1.5 Bath Apartment

[redacted]

Charming 2-bedroom, 1.5-bath apartment conveniently located near[redacted] The eat-in kitchen is equipped with a stove and refrigerator, perfect for everyday meals. This unit includes washer and dryer connections, central heat and air, and comes with blinds installed. Water, trash, pest control, and lawn maintenance are all covered by the owner, making this a hassle-free living experience!

No Smoking, No Section 8, No Cats

A.1.2 Callaway & Sant'Anna (2021) Estimation Details

The estimation process begins by defining the cohort-and-time average treatment effect for places first treated in year g at calendar time t:

$$\mathrm{ATT}_{g,t} \;=\; \mathbb{E}\big[Y_{it}(1) \;-\; Y_{it}(0) \;\big|\; G_i = g,\; T_i = t\big], \tag{A.1}$$

where $Y_{it}(d)$ denotes the potential outcome for place i in year t under treatment status $d \in \{0, 1\}$, G_i is the year of first SOI policy adoption for place i, and T_i is the calendar year of the observation.

Under the usual identifying assumptions of conditional parallel trends and overlap, construct a doubly-robust estimator by first fitting an outcome regression $\hat{m}_d(X_i) = \widehat{\mathbb{E}}[Y_{it} \mid D_i = d, X_i]$ and cohort-specific propensity scores $\hat{e}_g(X_i) = \widehat{\Pr}(G_i = g \mid X_i)$, where X_i is a vector of baseline covariates and $D_i = \mathbf{1}\{T_i \geq G_i\}$. The DR estimator for each (g, t) is then

$$\hat{\tau}^{\mathrm{DR}}_{g,t} = \frac{1}{n_{g,t}} \sum_{i=1}^{n} \Big\{ \big[\hat{m}_1(X_i) - \hat{m}_0(X_i)\big] + \frac{\mathbf{1}\{G_i = g,\, T_i = t\}}{\hat{e}_g(X_i)} \big(Y_{it} - \hat{m}_1(X_i)\big) - \frac{\mathbf{1}\{G_i > t\}}{1 - \sum_{g' \leq t} \hat{e}_{g'}(X_i)} \big(Y_{it} - \hat{m}_0(X_i)\big) \Big\}, \tag{A.2}$$

where $n_{g,t} = \sum_i \mathbf{1}\{G_i = g, T_i = t\}$. This estimator is consistent so long as either the outcome model $\hat{m}_d$ or the propensity model $\hat{e}_g$ is correctly specified.

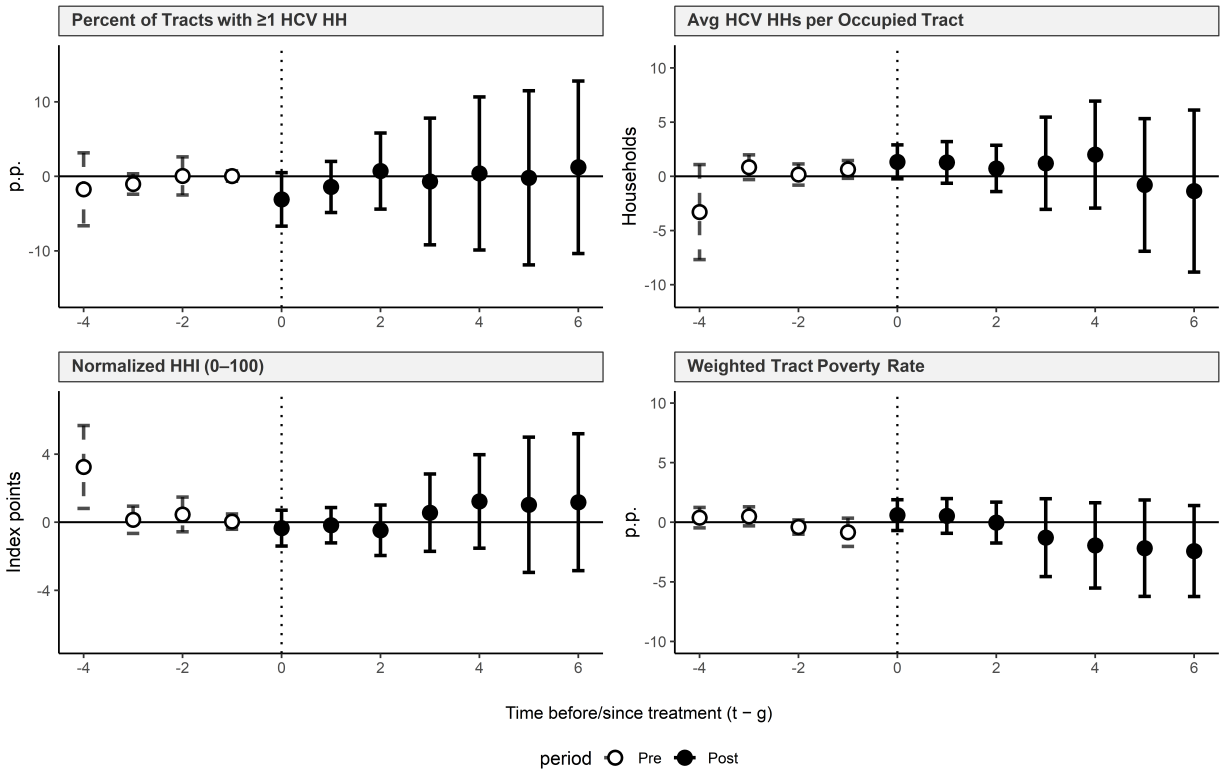
To visualize the dynamic policy response, aggregate these group-time estimates into an event-study curve in relative time $e = t - g$:

$$\mathrm{ATT}(e) \;=\; \sum_{g} w_g \, \hat{\tau}^{\mathrm{DR}}_{g,\, g+e}, \tag{A.3}$$

where weights w_g reflect the relative size of each cohort.

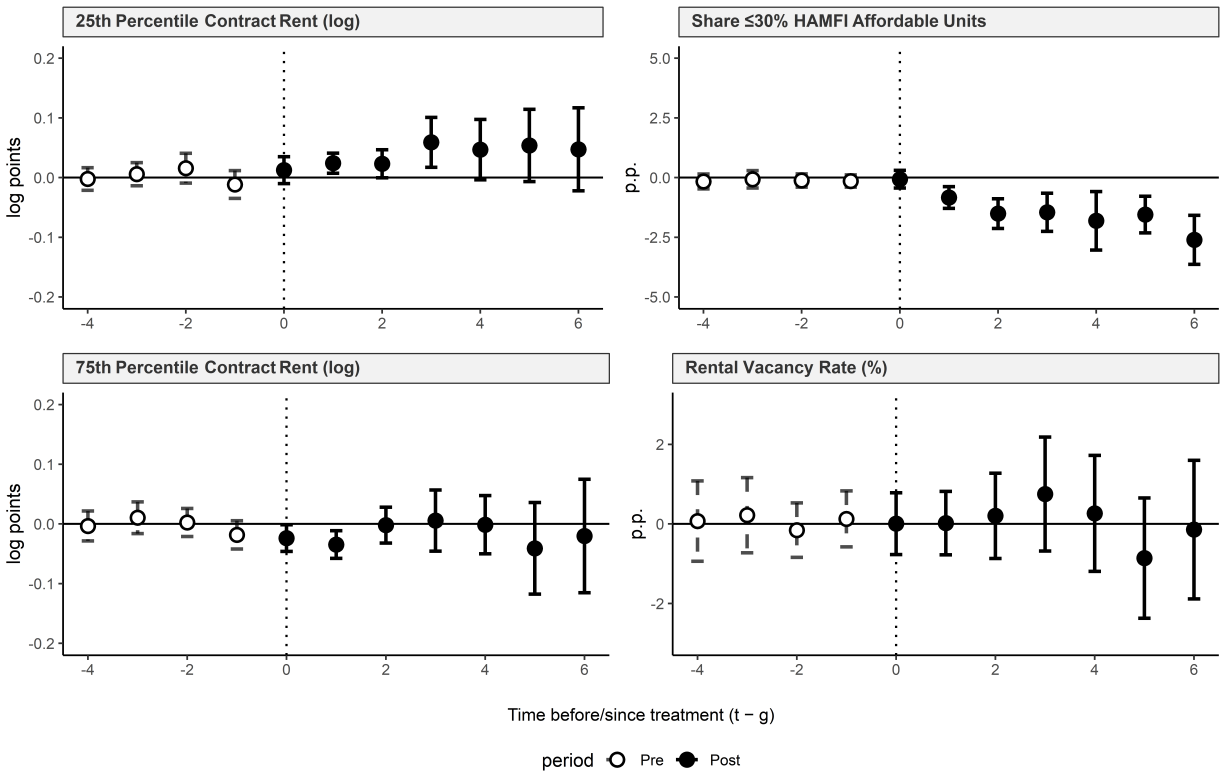
A.1.3 Anticipation Effects Event Study Graphs

Figure A.2: Anticipation Check:
Dynamic Effect Estimates for HCV Household Mobility Outcomes



Notes: Figure displays 4 event study graphs of the dynamic effect of SOI Protection laws on HCV mobility related outcomes from Outcome Table A with 1 year of anticipation in the Callaway & Sant'Anna (2021) doubly-robust approach. The counterfactual group consists of all census designated places that never passed an SOI protections law in the 2013 to 2018 time frame. The results account for conditional parallel trends pre-treatment 2011 baseline values of median rent, percent of population with a bachelor's degree, median age, median household income, the presence of anti-retaliation laws, the presence of limit fees laws, and the presence of SOI protection laws in other cities within the same state via the doubly-robust inverse probability weighting and outcome regression process. The x-axis represents time in years before or after treatment has occurred. "p.p" - Percentage points, "HHI" - Herschman-Herfindahl Index.

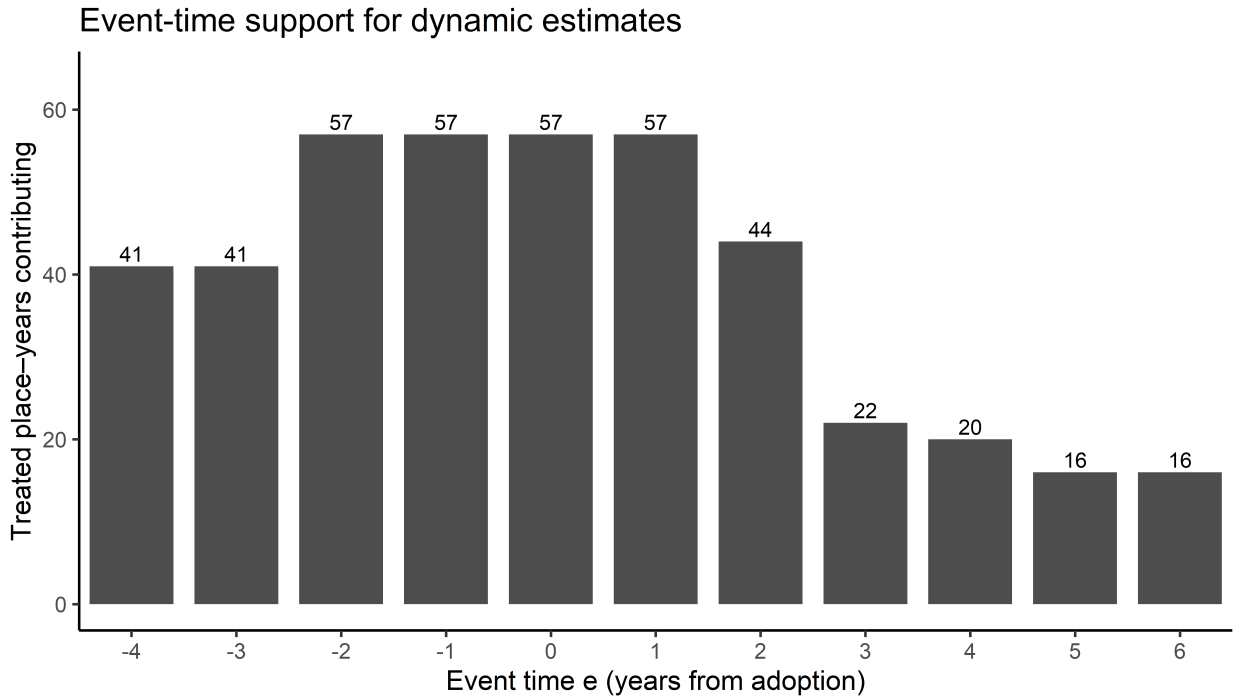
Figure A.3: Anticipation Check:
Dynamic Effect Estimates for Rental Affordability Outcomes



Notes: Figure displays 4 event study graphs of the dynamic effect of SOI Protection laws on Rental Affordability outcomes from Outcome Table B with 1 year of anticipation in the Callaway & Sant'Anna (2021) doubly-robust approach. The counterfactual group consists of all census designated places that never passed an SOI protections law in the 2013 to 2018 time frame. The results account for conditional parallel trends pre-treatment 2011 baseline values of median rent, percent of population with a bachelor's degree, median age, median household income, the presence of anti-retaliation laws, the presence of limit fees laws, and the presence of SOI protection laws in other cities within the same state via the doubly-robust inverse probability weighting and outcome regression process. The x-axis represents time in years before or after treatment has occurred. "p.p" - Percentage points.

A.1.4 Event-time Support

Figure A.4: Contributing Treated Place-Year Units by Event Time



Notes: Figure displays event-time support for dynamic estimates. Bars show the number of treated place–years contributing to each event time e = t - g; 0 marks the first effective year.

Appendix B

Supplementary Material for Chapter 2

B.1 Chapter 2 Appendix

B.1.1 Details of Tract-CoC Crosswalk Construction

This appendix describes in detail how we construct the tract–to–Continuum of Care (CoC) crosswalk used to harmonize all tract-level covariates to the CoC geography, following the general approach of Ferrara et al. (2024). Although Ferrara et al. (2024) develop population-weighted crosswalks for county- and congressional-district-level data, our implementation adapts their logic to the tract and CoC level. In brief, for each year from 2015 to 2019, we (i) load CoC boundaries and Census tracts (TIGER/Line), (ii) intersect tracts with CoCs to create *tract-pieces*, (iii) assign a population weight to each piece using a high-resolution raster (Global Human Settlement), (iv) download ACS five-year estimates at the tract level, (v) multiply each tract's ACS variables by its tract-piece population share, and (vi) sum up all weighted tract-piece values within each CoC. The result is an annual panel of CoC-level covariates that exactly preserves the original tract-level data in aggregate. Below we describe each step with sufficient detail to reproduce our results.

Data Sources

- **CoC shapefiles (2015–2019):** We obtain annual CoC boundary shapefiles from HUD's Homelessness Data Exchange. Each shapefile contains one polygon per CoC with a unique six-digit `CoC_Code`.

- **Census tract shapefiles (2010 TIGER, projected to five-year Census years):** For each year 2015–2019, we download the TIGER/Line tract shapefiles corresponding to that year's ACS 5-year release (e.g. 2015 tracts come from 2015 TIGER, etc.). These are obtained via the `tigris` R package (argument `year = YEAR`, `cb = TRUE`), returning one polygon per 2010 Census tract whose boundaries are valid for the ACS reference period.

- **Population raster (Global Human Settlement, 2020 edition):** We use the GHSL population-distribution raster (`GHS_POP_E2020_GLOBE_R2020A_100m.tif`) to assign high-resolution estimates of population to each tract-piece. This raster provides estimated population counts at approximately 100 m × 100 m resolution for the year 2020, which we project into the Albers equal-area CRS (EPSG 5070).

- **ACS 5-year tract data (2015–2019):** We download tract-level five-year ACS estimates of all covariates of interest (population, poverty, housing units pre-1940, vacant units, overcrowding, rent burden, race, age, eligibility for SNAP/SSI, disability, veteran status, etc.) via the `tidycensus` R package, specifying `geography = "tract"`, `year = YEAR`, and `survey = "acs5"`. We retain only the census variables listed in Section 2.4.5.

Overall Workflow

For each year $t = 2015, 2016, 2017, 2018, 2019$, we perform the following steps:

1. **Load and preprocess CoC boundaries:**

   - Read `All_CoCs_<t>.shp` into R as an `sf` object (column `CoC_Code`), then transform to the USA Contiguous Albers equal-area CRS (EPSG 5070), and apply `st_make_valid()` to ensure valid polygons.
   - Extract the six-digit CoC identifier by taking the last six characters of `CoC_Code`, and store in a new column `id`. Call this object `coc_shp`.
   - Compute $\bigcup_{c \in \mathcal{C}} \mathrm{CoC}_c$ (the union of all CoC polygons) and store as `coc_union` for later use in constructing "outside" pieces.

2. **Load and preprocess Census tracts:**

   - For each state in the continental U.S. $\{\mathrm{AL}, \ldots, \mathrm{WY}, \mathrm{DC}\}$, download the TIGER/Line tract shapefile (`tracts(state = st, year = t, cb = TRUE)`) and bind all states together into one `sf` (`tr_yr`). Then transform `tr_yr` into EPSG 5070 and apply `st_make_valid()`.

- At this point, `tr_yr` contains exactly the 2010 Census tracts $(GEOID)$ as used by ACS for year t. We keep only the `GEOID` and the geometry in `tr_yr`.

3. **Split tracts into "inside" and "outside" CoC pieces:**

   - *Inside pieces*: Compute

   $$\text{pcs\_in} \;=\; \text{st\_intersection}\big(\texttt{tr\_yr},\, \texttt{coc\_shp}\big),$$

   keeping the original `GEOID` and storing the intersected CoC's `id` as `coc_id`. These are all the "tract-pieces" that lie *inside* some CoC.

   - *Outside pieces*: Compute

   $$\text{pcs\_out} \;=\; \text{st\_difference}\big(\texttt{tr\_yr},\, \texttt{coc\_union}\big),$$

   again extracting polygons. Label all of these with `coc_id = "NONCOC"`. This ensures that for each tract τ, $\big(\text{pcs\_in}(\tau)\big) \cup \big(\text{pcs\_out}(\tau)\big)$ exactly covers the entire tract.

   - Combine `pcs_in` and `pcs_out` (with identical columns) into `pieces`. Each row of `pieces` is a *tract-piece* with `GEOID`, `coc_id` $\in$ {six-digit CoC, "NONCOC"}, and a valid polygon geometry.

4. **Compute population weights via GHSL raster:**

   - Load the GHSL 100 m population raster (projected to EPSG 5070). Call this raster `gpw_proj`.

   - For each tract-piece in `pieces`, compute the total raster population via

   $$\texttt{pop\_gpw} \;=\; \text{exact\_extract}(\texttt{gpw\_proj}, \texttt{pieces}, \text{"sum"}).$$

   This yields one `pop_gpw` ≥ 0 per tract-piece.

- Within each original tract (group by GEOID), define

$$\texttt{pop\_share} = \begin{cases} \dfrac{\texttt{pop\_gpw}}{\sum_{\text{pieces of } \texttt{GEOID}} \texttt{pop\_gpw}}, & \text{if } \sum \texttt{pop\_gpw} > 0, \\ \dfrac{\text{area(piece)}}{\sum \text{area(all pieces of } \texttt{GEOID})}, & \text{if } \sum \texttt{pop\_gpw} = 0. \end{cases}$$

  In other words, if the GHSL raster assigns nonzero population to any piece in the tract, we use those raster-based population shares. Otherwise (e.g. water-only tracts), we revert to an area-share fallback area(piece)/area(total tract).

- Store:

- $\texttt{pop\_share} \in [0, 1]$ for each tract-piece in pieces.

- By construction $\sum_{\text{pieces of } \tau} \texttt{pop\_share}(\tau, \text{piece}) = 1$ for every tract τ.

5. **Download and attach ACS tract variables:**

   - Via get_acs(geography="tract", year=t, variables=acs_vars, survey="acs5", geometry=FALSE), retrieve the vector of ACS five-year estimates $V_{\tau,g}$ for each tract τ and each variable $g \in \{\text{pop\_total}, \text{pop\_poverty}, \dots\}$.

   - Keep only the integer estimate columns (e.g. pop_totalE, pop_povertyE, ...) and rename them to pop_total, pop_poverty, etc. Call this data frame acs_df, keyed by GEOID.

   - Left-join pieces $\leftarrow$ pieces acs_df by GEOID, dropping geometry for the join. After this merge, each tract-piece has all ACS variables at the tract level, plus pop_share.

6. **Aggregate to CoC-level:**

- For each $g \in \{\text{ACS variables}\}$, define a CoC-level weighted sum

$$\text{CoC}_{c,g}(t) \;=\; \sum_{\substack{\text{tract-pieces } i \\ \text{with } i.\texttt{coc\_id}=c}} \big(\texttt{acs\_df}_{\tau(i),\,g}\big) \;\times\; \texttt{pop\_share}_i,$$

  where $\tau(i)$ is the original tract ID for piece i.

- Record the sum for each CoC c as $\text{sum}_{c,g}(t)$, and also store $\text{total\_pop}_c(t)$. We then compute any ratios (e.g. $\text{pct\_poverty} = 100 \times$ pop_poverty/pop_total) as needed.

- Append $\text{CoC}_{c,\cdot}(t)$ to a list `coc_results`$[[t]]$.

7. **Diagnostics:**

   - For each year t, we compute summary diagnostics on the tract-piece population shares:

$$\texttt{diag\_stats}(t) \;=\; \Big\{\text{for each tract } \tau,\; \big|\{\text{pieces of } \tau\}\big|,\; \max_i \texttt{pop\_share}(i),\; \ldots\Big\},$$

     confirming that no tract's population share is degenerate (unless entire-raster zeros, then by area share) and that the median tract has a largest piece share near 0.96 (i.e. only 4% of a typical tract's population is split across multiple CoCs).

   - Save these diagnostics to `diag_results`$[[t]]$ for completeness.

After looping over $t = 2015$–2019, we bind all `coc_results`$[[t]]$ into one balanced (or nearly balanced) panel `coc_panel` with columns $\{\texttt{coc\_id}, t, \text{sum}_{c,g}(t)\}$. This final data frame contains, for each CoC–year, every ACS-derived covariate aggregated via population shares, plus any derived ratios (e.g. poverty rate, pre-1940 share, housing vacancy, rent burden, race shares, age shares, veteran share, disability share, etc.). We then merge this `coc_panel` with our outcome variables (PIT counts, HIC counts, funding, etc.) by $\{\texttt{coc\_id}, t\}$ to form the final estimation data.

At this point, `coc_panel` contains one row per $\{\texttt{coc\_id}, \texttt{year}\}$ with all ACS-derived covariates scaled by population shares. We then merge `coc_panel` with our outcome data (TOT

PIT counts, funding, HIC counts, etc.) by $\{\texttt{coc\_id}, \texttt{year}\}$. This yields the final data set used in all FE–2SLS regressions.

Key Validation and Diagnostics

- **Population-share diagnostics:** For each year t, we summarized $\{\max_i \texttt{pop\_share}(i)\}$ across all tract IDs. The mean largest share per tract is at least 0.96, and the minimum nonzero share is around 0.00017, confirming that (i) every tract splits no more than 25 pieces on average, and (ii) every tract has some nonzero weight in at least one CoC.

- **State-level consistency:** After constructing `coc_panel`, we aggregated total population by $\{\text{state}, t\}$, including $\texttt{coc\_id} = "NONCOC"$ pieces. These state-level sums closely match the ACS state-level population counts (within 0.2%) for each year, indicating no substantial leakage or double-counting.

- **Comparison with area-only crosswalk:** In ancillary tables (not shown), we compared key covariates—poverty rate, Black population share, rent burden—aggregated via population weights versus naive area-weights (i.e. area(piece)/area(tract)). For most metropolitan CoCs, the population-weighted poverty rate is 1–2 percentage points higher than the area-weighted one, reflecting the fact that poverty clusters in dense neighborhoods. This confirms that population weights do correct meaningful within-tract heterogeneity.

Crosswalk Comparison to Byrne (2017)

To validate our population-weighted crosswalk against the established Byrne (2017) centroid-based method, we conducted a comprehensive analysis of tract-CoC boundary intersections using 2019 data. Of 72,684 census tracts, 312 (0.43%) exhibit meaningful splits (10-90% of area in multiple CoCs), affecting 1.67 million people (0.51% of U.S. population). While small in aggregate, these splits disproportionately impact certain CoCs: 23% of CoCs contain at least one meaningfully split tract, and in affected CoCs, centroid-based assignment would misattribute 14-48% of population. For example, GA-502 would experience 47.8% population misattribution, OK-503 would experience 36.0%, and TX-607 would misattribute 13.9% (affecting 1.77 million

people). The distribution of split severity is nearly uniform across the 10-90% range, indicating that plurality-assignment rules would introduce systematic rather than random measurement error.

We implemented comprehensive diagnostic checks across all 1,850 CoC-year observations. All computed percentage variables lie within [0,100] with zero violations across nine measures. All count variables satisfy logical bounds (unemployed $\leq$ labor force, poverty count $\leq$ poverty universe, etc.) with 100% compliance. Our crosswalk recovers 99.2-99.6% of national population across 2015-2019, with the small shortfall (0.4-0.8%) reflecting tracts entirely outside CoC service areas rather than methodological error. This near-perfect recovery, combined with zero logical inconsistencies, demonstrates that our intersection-based method successfully allocates tract-level ACS data to CoCs without systematic double-counting or population loss.

Table B.1: Overall Crosswalk Validation Statistics (2019)

Metric	**Value**
Total CoCs	392
CoCs with 10%+ splits	91
% CoCs with splits	23.2%
Total census tracts	72,684
Tracts touching multiple CoCs	12,682
Tracts with meaningful splits (10%+)	312
% tracts split	0.43%
Total population in split tracts	1,667,367
% US population in split tracts	0.51%

Notes: "Meaningful splits" defined as tracts with 10–90% of area in at least two CoCs. Population figures based on ACS 2015–2019 5-year estimates.

A detailed comparison with Byrne's methodology, including tract-level split analyses and ground-truth validation using county-equivalent CoCs, is available from the authors.

B.1.2 Robustness: Population in Poverty Results

Table B.2: Effects on Homelessness Rates Among People in Poverty

	Dependent Variables (per 10,000 people in poverty)	
	Sheltered Homeless	Unsheltered Homeless
	(1)	(2)
Total Federal Funding	0.687***	0.074
($1000s)	(0.115)	(0.117)
Black Population Share (%)	−0.848**	−2.122***
	(0.381)	(0.389)
Population Over 65 Share (%)	−0.211	−1.650
	(1.478)	(1.510)
SNAP Recipients Share (%)	3.162*	−5.097***
	(1.672)	(1.708)
SSI Recipients Share (%)	−0.666	4.820
	(4.001)	(4.088)
Poverty Rate (%)	−7.478***	−0.845
	(1.943)	(1.985)
College Education Rate (%)	−0.481	0.781
	(1.066)	(1.090)
Unemployment Rate (%)	−17.603**	28.064***
	(7.937)	(8.109)
Vacancy Rate (%)	1.099	2.112***
	(0.681)	(0.696)
Overcrowding Rate (%)	5.530	47.925***
	(6.940)	(7.091)
Rent Burden Rate (%)	4.449***	0.382
	(1.612)	(1.647)
Veteran Population Rate (%)	3.039	6.926**
	(2.736)	(2.796)
Disabled Population Share (%)	−1.537	0.412
	(3.296)	(3.367)
Nearby Funding Ratio	0.070*	0.019
	(0.037)	(0.037)
January Average Temperature	−0.730	2.337***
	(0.675)	(0.689)
January Precipitation	3.759	14.219***
	(2.426)	(2.478)
Summary Statistics		
Observations	370	370
R^2	0.401	0.549
Residual Std. Error	62.981	64.347 (df = 353)

Notes: All estimates from 2SLS regressions with pre-1940 housing share as instrument. Dependent variables are homelessness rates per 10,000 people in poverty rather than per 10,000 total residents. Robust standard errors clustered at CoC level in parentheses. Cross-section = 2019 data. *p<0.1; **p<0.05; ***p<0.01

B.1.3 Robustness: Excluding New York City and Los Angeles CoCs

Table B.3: Robustness Check: Excluding New York City and Los Angeles CoCs

	Dependent Variables (per 10,000 residents)	
	Sheltered Homeless	Unsheltered Homeless
	(1)	(2)
Total Federal Funding	0.099^{***}	-0.001
($1000s)	(0.016)	(0.015)
Model Specifications		
Full Control Set	Yes	Yes
Instrumental Variable	Pre-1940 Housing Share	
Summary Statistics		
Observations	368	368
R^2	0.404	0.536
Residual Std. Error	8.705	8.474 (df = 351)

Notes: All estimates from 2SLS regressions with pre-1940 housing share as instrument. Sample excludes New York City CoC and Los Angeles City/County CoCs to test robustness to potential outliers. Robust standard errors clustered at CoC level in parentheses. Cross-section = 2019 data. Full control set includes all demographic, socioeconomic, housing market, special population, and geographic/climate controls as shown in previous tables. *p<0.1; **p<0.05; ***p<0.01

Appendix C

Supplementary Material for Chapter 3

C.1 Chapter 3 Appendix

Figure C.1: Phase 1 Feature Importance

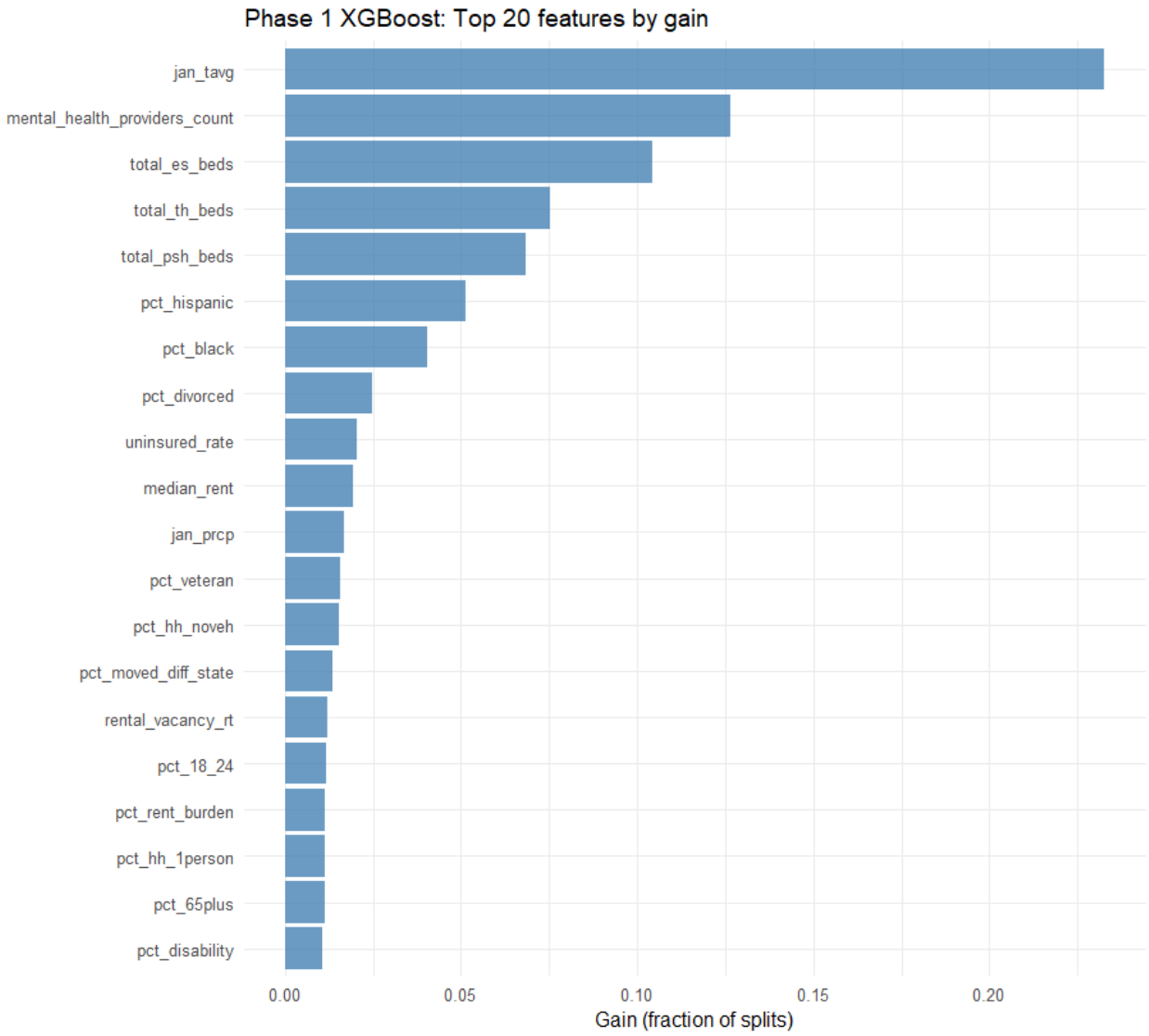


Figure C.2: Love Plot of Standardized Mean Difference Before vs. After Reweighting, Phase 2

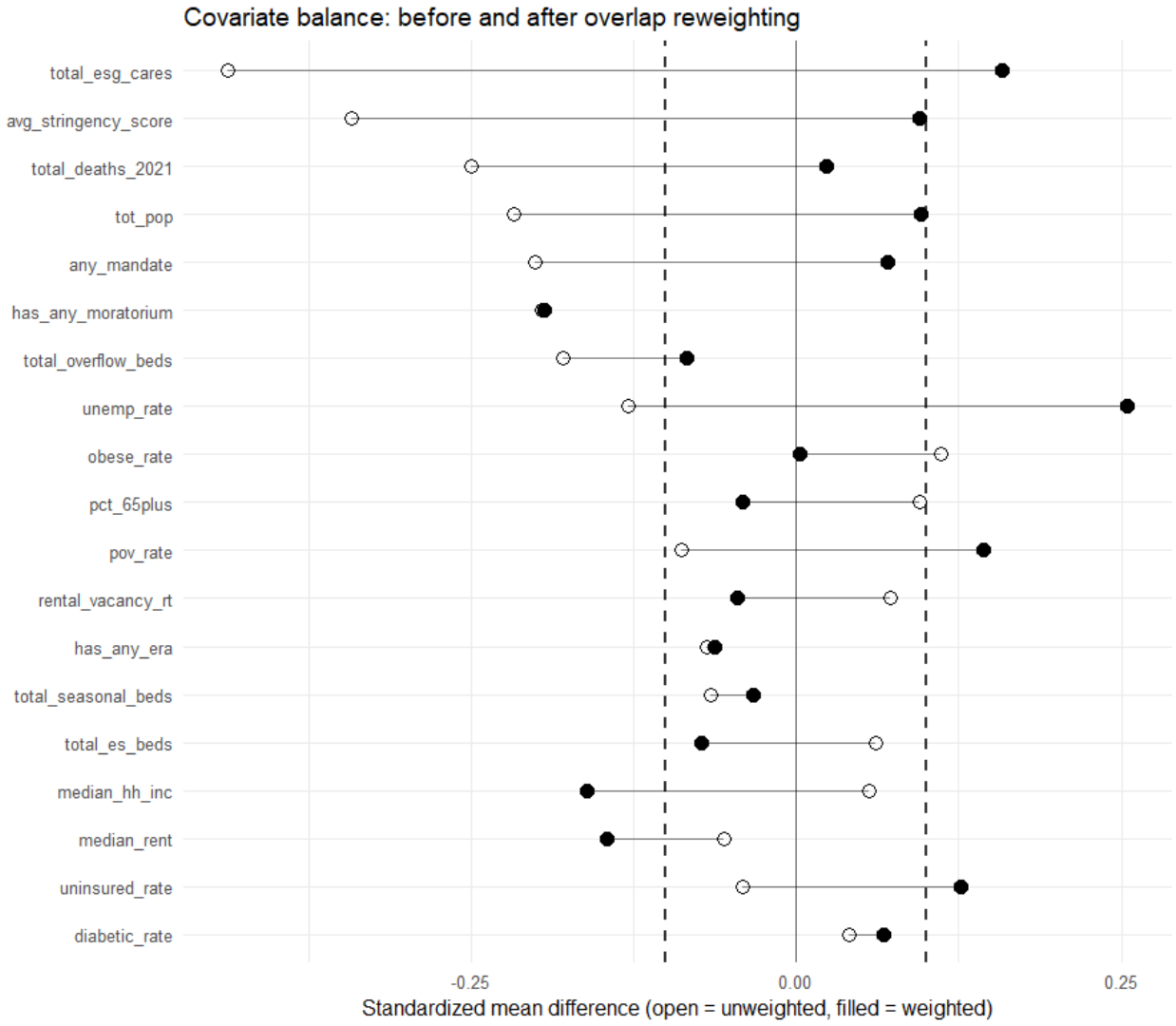


Figure C.3: Phase 2 Feature Importance

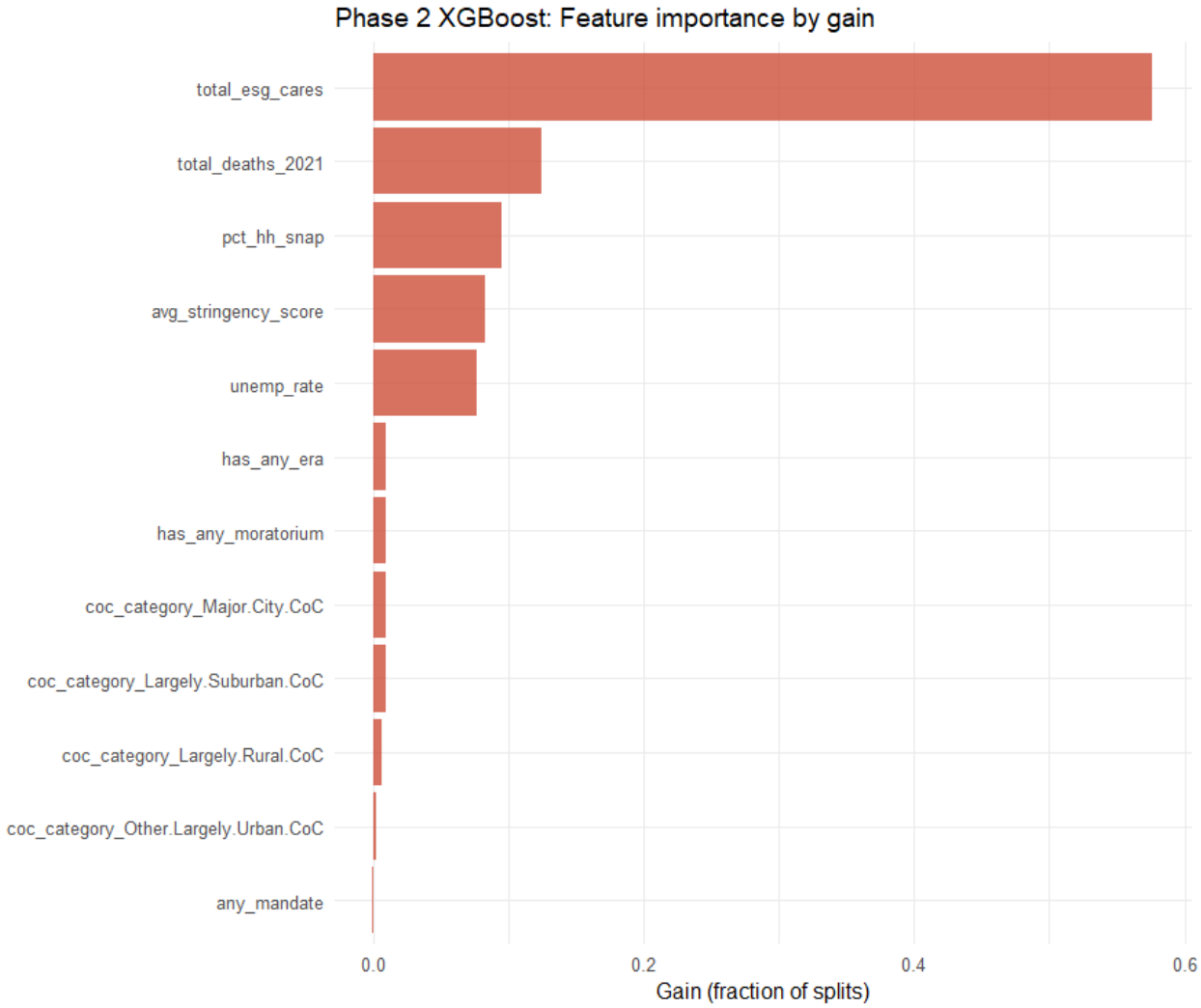


Figure C.4: Phase 2 SHAP Values

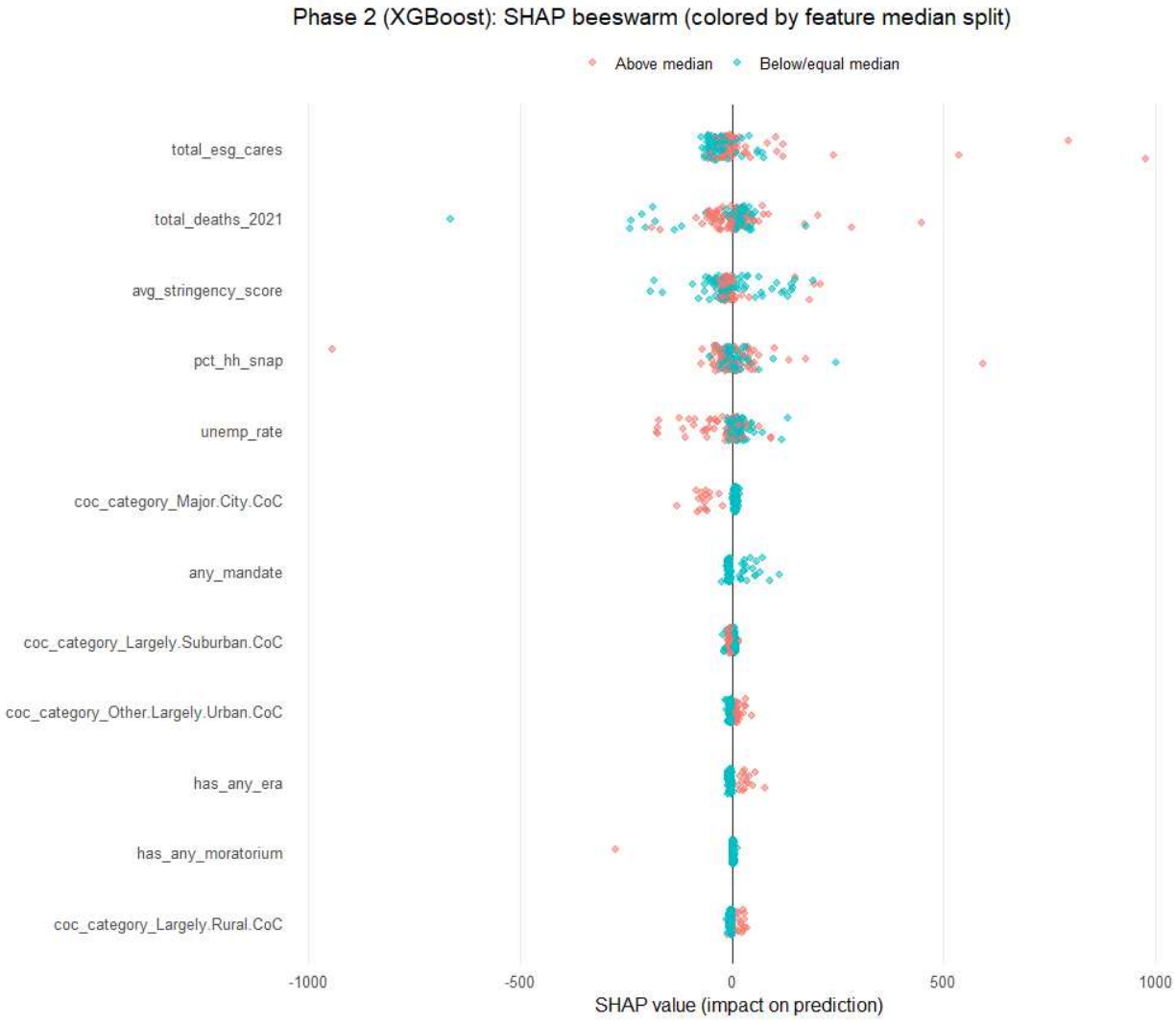


Figure C.5: Cleveland Plot: Prediction Intervals of 30 CoCs by Category

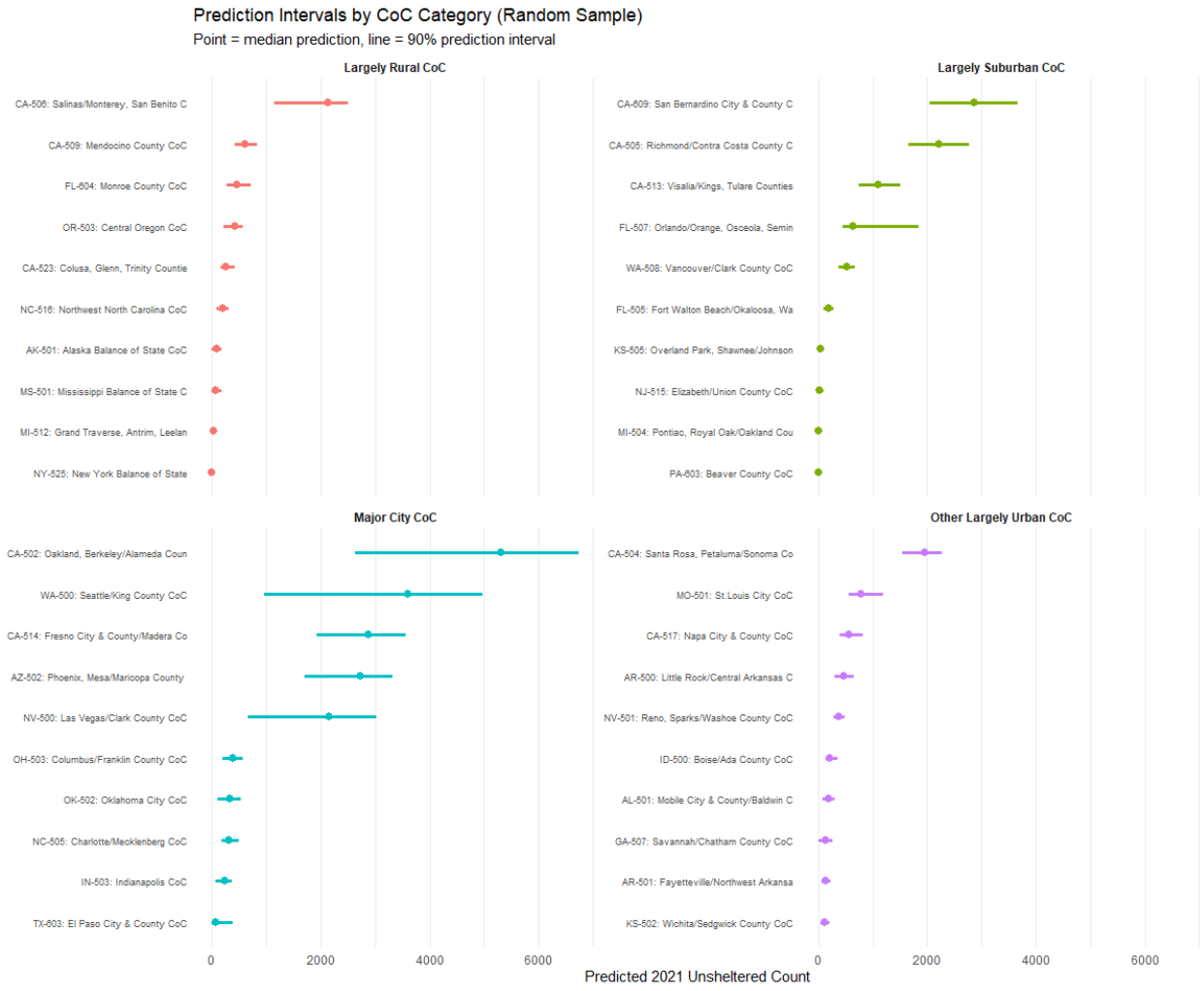


Figure C.6: Cleveland Plot: Prediction Intervals of Top 30 CoCs

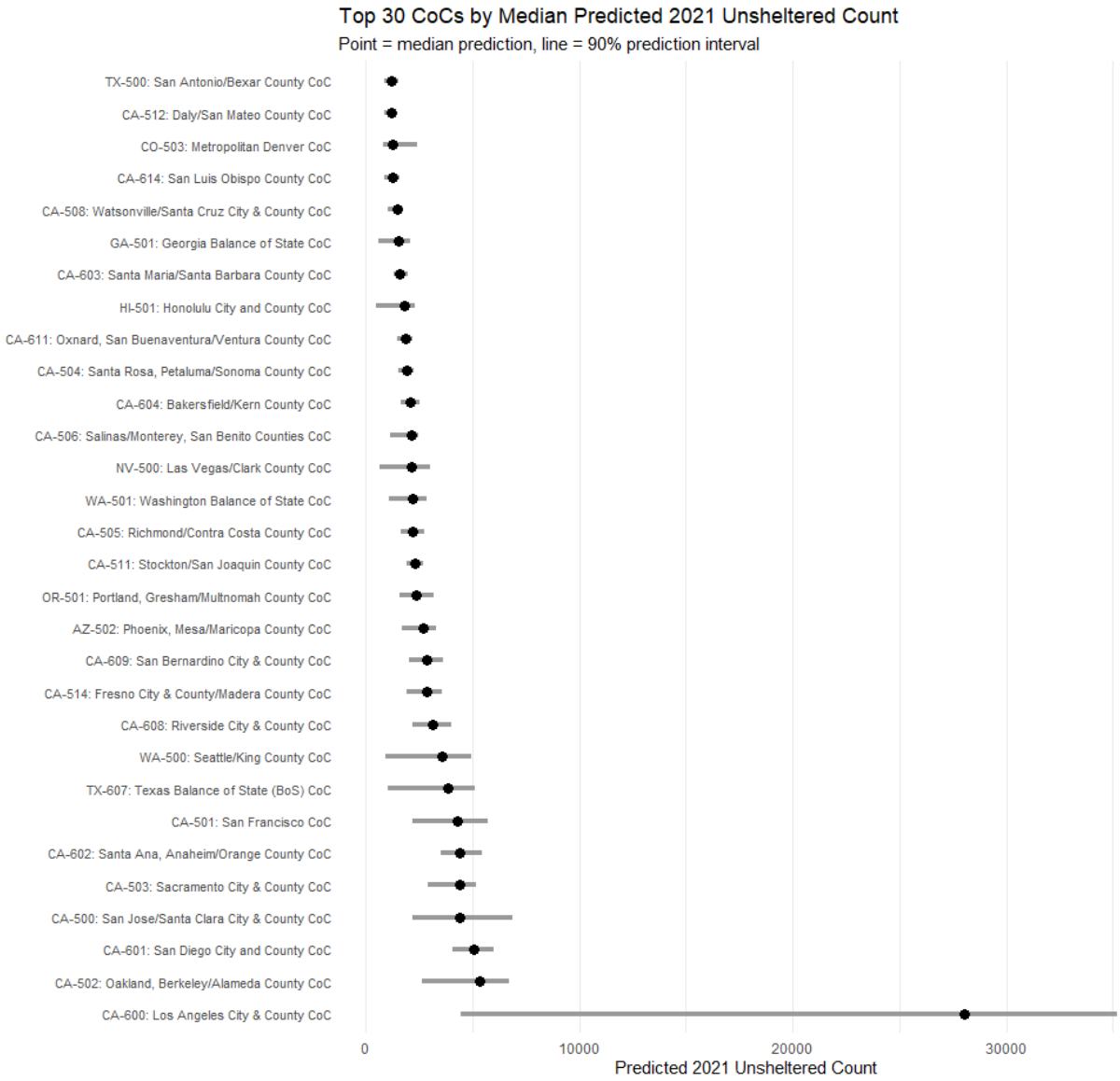


Table C.1: Imputed 2021 Unsheltered Counts: Part 1 of 5 (AK–CA-530)

CoC ID	CoC Name	Median	p_{05}	p_{95}
AK-501	Alaska Balance of State	86	0	181
AL-501	Mobile City & County/Baldwin County	197	86	298
AL-502	Florence/Northwest Alabama	87	0	188
AL-503	Huntsville/North Alabama	194	129	286
AL-504	Montgomery City & County	158	56	274
AL-505	Gadsden/Northeast Alabama	248	149	373
AL-506	Tuscaloosa City & County	118	48	187
AR-500	Little Rock/Central Arkansas	474	303	646
AR-501	Fayetteville/Northwest Arkansas	129	59	230
AZ-500	Arizona Balance of State	746	361	957
AZ-501	Tucson/Pima County	587	326	953
AZ-502	Phoenix, Mesa/Maricopa County	2,720	1,704	3,321
CA-500	San Jose/Santa Clara City & County	4,417	2,198	6,906
CA-501	San Francisco	4,279	2,200	5,717
CA-502	Oakland, Berkeley/Alameda County	5,316	2,625	6,734
CA-503	Sacramento City & County	4,392	2,928	5,175
CA-504	Santa Rosa, Petaluma/Sonoma County	1,954	1,553	2,272
CA-505	Richmond/Contra Costa County	2,217	1,648	2,774
CA-506	Salinas/Monterey, San Benito Counties	2,135	1,152	2,503
CA-507	Marin County	980	701	1,170
CA-508	Watsonville/Santa Cruz City & County	1,490	1,043	1,762
CA-509	Mendocino County	602	415	836
CA-511	Stockton/San Joaquin County	2,292	1,954	2,674
CA-512	Daly/San Mateo County	1,199	908	1,443
CA-513	Visalia/Kings, Tulare Counties	1,096	744	1,500
CA-514	Fresno City & County/Madera County	2,877	1,926	3,569
CA-515	Rocklin/Roseville/Placer County	943	745	1,156
CA-516	Redding/Shasta, Siskiyou Counties	896	639	1,166
CA-517	Napa City & County	566	396	814
CA-518	Vallejo/Solano County	1,003	712	1,252
CA-519	Chico, Paradise/Butte County	1,110	912	1,357
CA-520	Merced City & County	806	538	1,154
CA-521	Davis, Woodland/Yolo County	662	501	832
CA-522	Humboldt County	700	213	959
CA-523	Colusa, Glenn, Trinity Counties	249	157	410
CA-525	El Dorado County	628	442	790
CA-526	Amador, Calaveras, Mariposa Counties	646	456	881
CA-530	Alpine, Inyo, Mono Counties	279	147	509

Note: Part 1 of 5. Complete list of 230 imputed 2021 unsheltered counts. Median = bootstrap median prediction; p_{05} and p_{95} = 5th and 95th percentiles (90% prediction interval). All estimates from end-to-end clustered bootstrap (B = 300 draws). CoC names abbreviated for space.

Table C.2: Imputed 2021 Unsheltered Counts: Part 2 of 5 (CA-531–GA-508)

CoC ID	CoC Name	Median	p_{05}	p_{95}
CA-531	Nevada County	527	326	760
CA-600	Los Angeles City & County	28,038	4,466	35,180
CA-601	San Diego City and County	5,046	4,059	6,008
CA-602	Santa Ana, Anaheim/Orange County	4,386	3,525	5,438
CA-603	Santa Maria/Santa Barbara County	1,623	1,302	1,978
CA-604	Bakersfield/Kern County	2,086	1,673	2,549
CA-606	Long Beach	1,157	597	1,580
CA-607	Pasadena	650	413	981
CA-608	Riverside City & County	3,166	2,220	4,008
CA-609	San Bernardino City & County	2,874	2,052	3,660
CA-611	Oxnard/Ventura County	1,859	1,491	2,212
CA-612	Glendale	139	0	725
CA-613	Imperial County	1,095	529	1,578
CA-614	San Luis Obispo County	1,282	861	1,595
CO-503	Metropolitan Denver	1,268	826	2,417
CO-504	Colorado Springs/El Paso County	208	2	365
CO-505	Fort Collins/Larimer, Weld Counties	245	143	388
DE-500	Delaware Statewide	133	28	284
FL-500	Sarasota/Manatee, Sarasota Counties	324	205	462
FL-501	Tampa/Hillsborough County	572	379	780
FL-502	St. Petersburg/Pinellas County	822	522	1,000
FL-503	Lakeland, Winterhaven/Polk County	198	60	357
FL-505	Fort Walton Beach/Okaloosa Counties	196	100	274
FL-507	Orlando/Orange, Osceola Counties	643	440	1,848
FL-509	Fort Pierce/St. Lucie Counties	827	306	1,035
FL-511	Pensacola/Escambia, Santa Rosa	214	82	345
FL-513	Palm Bay, Melbourne/Brevard County	348	234	455
FL-514	Ocala/Marion County	197	33	374
FL-515	Panama City/Bay, Jackson Counties	229	86	328
FL-517	Hendry, Hardee, Highlands Counties	376	182	519
FL-518	Columbia, Hamilton Counties	637	332	997
FL-601	Ft Lauderdale/Broward County	865	517	1,127
FL-604	Monroe County	455	274	721
FL-605	West Palm Beach/Palm Beach County	867	391	1,050
GA-500	Atlanta	538	141	775
GA-501	Georgia Balance of State	1,565	580	2,069
GA-502	Fulton County	261	177	375
GA-503	Athens-Clarke County	88	33	153
GA-504	Augusta-Richmond County	600	387	1,009
GA-506	Marietta/Cobb County	409	250	590
GA-507	Savannah/Chatham County	140	0	259
GA-508	DeKalb County	248	153	374

Note: Part 2 of 5. See Table C.1 notes for details.

Table C.3: Imputed 2021 Unsheltered Counts: Part 3 of 5 (HI-500–MN-511)

CoC ID	CoC Name	Median	p_{05}	p_{95}
HI-500	Hawaii Balance of State	1,048	362	1,267
HI-501	Honolulu City and County	1,842	489	2,304
IA-501	Iowa Balance of State	109	0	212
ID-500	Boise/Ada County	204	137	354
ID-501	Idaho Balance of State	497	110	663
IL-502	Waukegan/Lake County	0	0	32
IL-504	Madison County	7	0	50
IL-506	Joliet/Will County	377	247	510
IL-508	East St. Louis/St. Clair County	0	0	32
IL-510	Chicago	1,125	350	1,627
IL-511	Cook County	0	0	121
IL-512	Bloomington/Central Illinois	85	3	167
IL-514	DuPage County	0	0	61
IL-515	South Central Illinois	13	0	98
IL-517	Aurora, Elgin/Kane County	108	64	182
IL-518	Rock Island/Northwestern Illinois	36	0	111
IL-520	Southern Illinois	67	0	139
IN-502	Indiana Balance of State	637	282	912
IN-503	Indianapolis	235	62	367
KS-502	Wichita/Sedgwick County	107	38	206
KS-505	Overland Park/Johnson County	47	19	88
KS-507	Kansas Balance of State	172	78	258
KY-500	Kentucky Balance of State	842	554	1,270
KY-502	Lexington-Fayette County	59	0	126
LA-500	Lafayette/Acadiana	165	69	284
LA-503	New Orleans/Jefferson Parish	528	264	716
LA-505	Monroe/Northeast Louisiana	231	0	393
LA-506	Slidell/Southeast Louisiana	80	0	222
LA-507	Alexandria/Central Louisiana	241	72	436
LA-509	Louisiana Balance of State	153	45	266
MA-502	Lynn	147	0	312
MA-516	Massachusetts Balance of State	150	0	353
MD-501	Baltimore	317	0	556
MD-504	Howard County	11	0	37
MD-601	Montgomery County	118	54	202
ME-500	Maine Statewide	202	73	617
MI-500	Michigan Balance of State	217	102	335
MI-501	Detroit	23	0	265
MI-502	Dearborn/Wayne County	0	0	51
MI-503	Warren/Macomb County	1	0	62
MI-504	Pontiac/Oakland County	0	0	57
MI-505	Flint/Genesee County	193	73	410
MI-506	Grand Rapids/Kent County	76	18	149
MI-507	Kalamazoo City & County	62	24	101
MI-509	Washtenaw County	42	1	98
MI-510	Saginaw City & County	565	400	876
MI-512	Grand Traverse Counties	31	0	84
MI-516	Muskegon City & County	19	0	89
MI-519	Holland/Ottawa County	44	2	107
MI-523	Eaton County	0	0	17
MN-500	Minneapolis/Hennepin County	526	275	730
MN-502	Rochester/Southeast Minnesota	58	15	94
MN-503	Dakota, Anoka Counties	55	0	123
MN-504	Northeast Minnesota	2	0	43
MN-505	St. Cloud/Central Minnesota	83	0	188
MN-506	Northwest Minnesota	15	0	60
MN-508	Moorhead/West Central Minnesota	0	0	6
MN-509	Duluth/St.Louis County	113	0	208
MN-511	Southwest Minnesota	0	0	28

Note: Part 3 of 5. See Table C.1 notes for details.

Table C.4: Imputed 2021 Unsheltered Counts: Part 4 of 5 (MO-500–PA-603)

CoC ID	CoC Name	Median	p_{05}	p_{95}
MO-500	St. Louis County	20	0	97
MO-501	St.Louis City	788	555	1,199
MO-606	Missouri Balance of State	411	222	561
MS-500	Jackson/Rankin, Madison Counties	26	0	146
MS-501	Mississippi Balance of State	64	0	179
MT-500	Montana Statewide	273	129	407
NC-503	North Carolina Balance of State	1,168	698	1,626
NC-504	Greensboro, High Point	154	111	195
NC-505	Charlotte/Mecklenberg	312	183	497
NC-511	Fayetteville/Cumberland County	280	77	415
NC-516	Northwest North Carolina	186	81	302
NJ-514	Trenton/Mercer County	61	0	121
NJ-515	Elizabeth/Union County	20	0	95
NV-500	Las Vegas/Clark County	2,142	663	3,027
NV-501	Reno, Sparks/Washoe County	379	289	491
NV-502	Nevada Balance of State	148	51	231
NY-501	Elmira/Steuben Counties	0	0	50
NY-507	Schenectady City & County	23	0	111
NY-508	Buffalo/Erie, Niagara Counties	63	0	167
NY-510	Ithaca/Tompkins County	0	0	45
NY-511	Binghamton/Broome Counties	53	0	157
NY-513	Wayne, Ontario Counties	0	0	53
NY-514	Jamestown/Chautauqua County	110	0	226
NY-518	Utica, Rome/Oneida Counties	0	0	55
NY-519	Columbia, Greene Counties	0	0	58
NY-520	Franklin, Essex Counties	0	0	35
NY-522	Jefferson, Lewis Counties	80	23	146
NY-523	Glens Falls/Saratoga Counties	0	0	0
NY-525	New York Balance of State	0	0	36
NY-601	Poughkeepsie/Dutchess County	2	0	53
NY-604	Yonkers/Westchester County	202	59	704
NY-606	Rockland County	8	0	48
NY-608	Kingston/Ulster County	93	40	143
OH-501	Toledo/Lucas County	305	158	471
OH-502	Cleveland/Cuyahoga County	83	0	227
OH-503	Columbus/Franklin County	380	199	574
OH-504	Youngstown/Mahoning County	0	0	48
OH-505	Dayton/Montgomery County	183	120	281
OH-507	Ohio Balance of State	568	184	859
OK-502	Oklahoma City	332	101	531
OK-503	Oklahoma Balance of State	53	0	142
OK-504	Norman/Cleveland County	109	25	179
OK-505	Northeast Oklahoma	324	200	504
OK-506	Southwest Oklahoma Regional	32	0	156
OR-501	Portland/Multnomah County	2,376	1,601	3,221
OR-503	Central Oregon	412	220	567
OR-504	Salem/Marion, Polk Counties	935	721	1,239
OR-505	Oregon Balance of State	1,129	743	1,623
PA-500	Philadelphia	850	429	1,177
PA-502	Delaware County	11	0	113
PA-503	Wilkes-Barre/Luzerne County	118	9	267
PA-505	Chester County	25	0	83
PA-509	Eastern Pennsylvania	362	212	596
PA-510	Lancaster City & County	69	1	140
PA-511	Bristol/Bucks County	461	321	586
PA-601	Western Pennsylvania	138	28	276
PA-603	Beaver County	0	0	23

Note: Part 4 of 5. See Table C.1 notes for details.

Table C.5: Imputed 2021 Unsheltered Counts: Part 5 of 5 (PA-605–WY-500)

CoC ID	CoC Name	Median	p_{05}	p_{95}
PA-605	Erie City & County	225	132	440
RI-500	Rhode Island Statewide	141	68	268
SC-501	Greenville/Anderson Upstate	295	66	466
SC-502	Columbia/Midlands	375	260	516
SD-500	South Dakota Statewide	171	83	274
TN-502	Knoxville/Knox County	211	136	301
TN-503	Central Tennessee	472	335	630
TN-504	Nashville/Davidson County	388	217	563
TN-506	Oak Ridge/Upper Cumberland	501	410	610
TN-510	Murfreesboro/Rutherford County	271	170	400
TX-500	San Antonio/Bexar County	1,194	904	1,515
TX-603	El Paso City & County	68	0	378
TX-607	Texas Balance of State	3,859	1,061	5,107
TX-624	Wichita Falls/Wichita Counties	125	0	265
TX-701	Bryan/Brazos Valley	108	0	218
UT-500	Salt Lake City & County	237	118	540
UT-503	Utah Balance of State	53	0	173
UT-504	Provo/Mountainland	27	0	123
VA-500	Richmond/Henrico Counties	101	37	178
VA-501	Norfolk/Chesapeake Counties	132	70	224
VA-504	Charlottesville	11	0	48
VA-513	Harrisonburg/Western Virginia	24	0	72
VA-521	Virginia Balance of State	344	214	447
VA-602	Loudoun County	39	5	79
VT-500	Vermont Balance of State	135	51	272
VT-501	Burlington/Chittenden County	101	27	227
WA-500	Seattle/King County	3,601	960	4,973
WA-501	Washington Balance of State	2,202	1,096	2,866
WA-502	Spokane City & County	493	352	626
WA-503	Tacoma/Pierce County	673	421	921
WA-504	Everett/Snohomish County	269	161	421
WA-508	Vancouver/Clark County	519	382	671
WV-500	Wheeling/Weirton Area	2	0	48
WY-500	Wyoming Statewide	214	134	307

Note: Part 5 of 5. Complete list of 230 imputed 2021 unsheltered counts, sorted alphabetically by CoC ID. Median = bootstrap median prediction; p_{05} and p_{95} = 5th and 95th percentiles (90% prediction interval). All estimates from end-to-end clustered bootstrap (B = 300 draws). CoC names abbreviated for space. National total (imputed only): Median = 50,832; 90% PI: [24,021, 81,619]. Combined with 146 observed complete counts yields national 2021 unsheltered estimate of 195,191 [114,380, 255,978].