Quantifying uncertainty in HUD estimates of homelessness*

Chris Glynn†, Thomas H. Byrne,‡ and Dennis P. Culhane§

Abstract

Official government statistics are often reported with uncertainty intervals. Such margin of error estimates provide important information about the variability of reported data to end users and stakeholders. In this paper, we illustrate a method to construct uncertainty intervals around the estimated size of the homeless population in the United States. The intervals are constructed using predicted outcomes from synthetic homeless counts. Quantifying uncertainty in estimates of homelessness provides important contextual information on the possible total scope of and year-over-year changes in homeless populations – information that can be used to evaluate whether policy or programmatic interventions to address homelessness are having their desired impact. We find that variation in the 2017 national count of homeless is likely between 530,000 to 565,000 compared to the 553,000 people officially reported by HUD.

1 Introduction

Each year since 2007, the U.S. Department of Housing and Urban Development (HUD) has produced nationwide estimates of the number of persons experiencing homelessness in both sheltered and unsheltered locations on a single night. These estimates, known as point-in-time (PIT) counts, are based on local enumeration efforts conducted in roughly 400[1] Continuums of Care (CoCs) – geographic units at which efforts to address homelessness are coordinated and whose boundaries are typically coterminous with a single city, a single county, or a group of counties – throughout the United States (U.S. Department of Housing and Urban Development, 2017b). As part of their application for federal homeless assistance funds, CoCs are required to report PIT counts to HUD on an annual basis, and must conduct their own local counts on a single night during the last week of January. Both the national and local PIT counts are published each year as part of HUD's Annual Homeless Assessment Report to Congress (AHAR) [2] (U.S. Department of Housing and Urban Development, 2017b).

*This work was supported by Zillow. The authors thank Kaoru Irie, Surya Tokdar, Cory Hopkins, Alexander Casey, and Svenja Gudell for helpful discussions.

†Paul College of Business and Economics, University of New Hampshire, christopher.glynn@unh.edu

‡School of Social Work, Boston University, tbyrne@bu.edu

§School of Social Policy & Practice, University of Pennsylvania, culhane@upenn.edu

[1]The exact number of CoCs varies from year to year due to the creation or dissolution of CoCs or the merger of two or more existing CoCs. In 2007, there were 461 CoCs; in 2017 there were 399.

[2]As part of the AHAR, HUD also publishes nationwide estimates of the unduplicated number of persons who spend at least one night in an emergency shelter or transitional housing program over the course of an entire year. These estimates are based on data from local Homelessness Management Information Systems (HMIS) reported by CoCs to HUD. However, the CoC estimates are not made publicly available.

Prior to the publication of the first HUD PIT estimates in 2007, there were no systematic nationwide efforts to count the homeless population on a regular basis, with the best multi-jurisdiction data on the scope of homelessness in the United States available from several one-off enumeration attempts of varying methodological rigor and geographic coverage conducted throughout the 1980s and 1990s (Burt and Cohen, 1989; Burt et al., 1999; Metraux et al., 2001; Taeuber and Siegel, 1990; U.S. Department of Housing and Urban Development, 1984). The HUD PIT estimates are thus highly important because they provide a means by which to monitor changes in the size and characteristics of the homeless population over time. In turn, this information can be used to evaluate whether policy or programmatic interventions to address homelessness are having their desired impact at both the local and national level (Byrne et al., 2014; Corinth, 2017; Lucas, 2017). Likewise, PIT counts are useful for assessing the extent to which housing market conditions or other factors may account for observed trends in homelessness within and across communities over time (Byrne et al., 2013; Corinth and Lucas, 2018; Glynn and Fox, 2018).

Yet, a number of issues are likely to affect how many individuals are included in PIT counts and, consequently, the reliability of any observed changes in homelessness over time. First, even absent any actual change in the number of persons experiencing homelessness from one year to the next, PIT counts are likely to change due simply to sampling variability. Second, CoCs must contend with the well-documented methodological challenge of attempting to identify and count the number of persons experiencing homelessness in unsheltered locations (Martin, 1992; Martin et al., 1997; Rossi et al., 1987; United States General Accounting Office, 1991). Most CoCs use an approach that entails asking volunteer canvassers assigned to particular areas to count anyone who is visibly homeless. However, unsheltered individuals often seek out of the way spaces that are not easily visible to counters. As a result, it is believed that PIT counts differ from the total number of homeless on a given night, and a number of studies document the extent of the undercount of the unsheltered population. For example, Hopper et al. (2008) estimate that approximately 60% of homeless service users in Manhattan were definitely visible to counters in New York City's 2005 count. Finally, the accuracy of PIT counts may also depend on community-specific factors such as the proportion of the homeless population that is unsheltered and the count methodology. While CoCs must use a HUD approved methodology in conducting their count, exact methodologies vary across CoCs and can change from one year to the next within a given CoC.

The combination of these issues means that it is difficult to assess whether year-over-year changes in PIT counts are statistically meaningful, or are instead the product of sampling variability, changes in the accuracy of the unsheltered count or methodological changes that may also impact accuracy. This fact has not gone unnoticed, with some observers arguing that recent large decreases in the unsheltered homeless population are largely the product of miscounting rather than due to the success of federal policy initiatives, which are frequently credited for the observed reductions in street homelessness (Corinth, 2015).

A more transparent and useful approach to conducting and reporting PIT counts would account for the uncertainty in these estimates. Reporting an interval of uncertainty around observed PIT counts would provide valuable context to policymakers, program planners, researchers and other stakeholders and would enable them to make more robust community and temporal comparisons of the scope of homelessness. It may likewise be useful to develop estimates of the total homeless population by adjusting PIT counts to account for their relative accuracy. Thus, in the present study we develop and employ new methods to quantify uncertainty in PIT counts at the

CoC-level. Applying these methods allows us to construct uncertainty intervals around reported PIT counts at the CoC and national level, and we find that 2017 variation in the national count of homeless is between 530,000 and 565,000 compared to the 553,000 that were reported in the official HUD PIT statistics. We construct comparable uncertainty intervals around the PIT estimates for 386 of the 399 CoCs that were extant in 2017. Our methods also allow us to impute estimates of the total homeless population (and corresponding uncertainty intervals) that are based on the PIT counts and a set of plausible assumptions about the accuracy of these counts. Our imputed national estimate of the total homeless population exceeds the reported PIT count by approximately 100,000 people. We conclude with a discussion of the benefits and limitations of our approach and offer suggestions on how our methods might be applied and improved in the future.

2 Data

Data for the present study come from three sources. First, we use CoC-level PIT estimates, which are made available in downloadble files from HUD (U.S. Department of Housing and Urban Development, 2017a). We supplement these data with total population estimates from the U.S. Census Bureau's American Community Survey (ACS). As described below, these data are used to develop estimates of the total population of each CoC, which is an important input for the model we use to estimate uncertainty in the PIT counts. The third model input is an estimate of the cost of rental housing as a share of median income in each CoC. To build this housing affordability metric, Zillow custom computed a median rent price for each CoC based on the CoC's geographic boundaries and their established rent index methodology (Bun, 2012). Although PIT data are available for the period from 2007 to 2017, Zillow data were only available beginning in 2011. As such, we limit our analysis to the period from 2011-2017. The fourth and final model input is the CoC-level rate of extreme poverty, which is calculated from ACS data.

	PIT Count	CoC total population
2011	51,123	7,944,958
2012	56,672	8,009,322
2013	64,060	8,074,863
2014	67,810	8,159,782
2015	75,323	8,231,358
2016	73,523	8,268,601
2017	76,501	8,305,844

Table 1: PIT Counts and estimated total population for the New York City CoC (NY-600). Both the PIT count and estimated total population are utilized in a statistical model to impute the total number of homeless.

In 2017, PIT estimates were provided for 399 distinct CoCs across all 50 states, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, and Guam. In this analysis, we analyze PIT counts from 386 of 399 CoCs in each of the years 2011-2017. The 13 CoCs that are excluded are either (i) outside the 50 states and District of Columbia, or (ii) did not have data for at least one of the years from 2011-2017. The 386 CoCs included in our analysis account for more than 98.7% (546,000 of the roughly 553,000) of the counted homeless in 2017.

Comparing the scope of homelessness across communities requires accounting for the relative size of each community, measured by total population. It is the *rate* of homelessness that allows us to directly compare the magnitude of homelessness in communities like Los Angeles, California and Manchester, New Hampshire. Unfortunately, the total population of a CoC is not reported with the PIT count, and rates of homelessness require some care in estimating. Complicating matters is the discrepancy in the geographic boundaries of CoCs and the geographic units at which total population estimates are made available from the U.S. Census Bureau. CoCs represent the most granular unit at which PIT counts are available, whereas total population estimates from the U.S. Census Bureau are available for smaller areas, including for Census tracts. It is thus possible for us to aggregate Census tract total populations to the CoC-level, but not possible to disaggregate CoC-level PIT counts down to the Census tract. As a result, we adopt the CoC as our geographic unit of analysis.

To calculate estimates of the total population of each CoC, we use publicly available geospatial data from HUD on the boundaries of each CoC (U.S. Department of Housing and Urban Development, 2018) to match every census tract in the United States to a CoC. For each census tract, we compute the geographic centroid. If the centroid of a tract falls within the CoC's boundaries, we assign that census tract as a member of the CoC. Using this classification of census tracts and tract-level population estimates from the ACS 5-year estimates[3], we were able to construct measures for the CoC's total population for each year from 2011 to 2016. To achieve this, we use population estimates from the end-year of each version of the ACS 5-Year estimates that align with the actual year of the PIT count. For example, we use the 2007-2011 ACS 5-year estimates to calculate CoC-level population for 2011. We then use the PIT counts and the computed CoC-level total population counts to develop estimates of rates of homelessness. As the 2013-2017 5-year ACS estimates are not yet available, we extrapolated trends in population growth for each CoC. To estimate the 2017 total population, we assume the growth in each CoC from 2016-2017 is the same as growth observed between 2015-2016.

Table 1 presents the sequence of PIT counts and total population estimates for the New York City CoC. Because the total population of a CoC is not directly observed, we view our aggregation as a noisy estimate of each CoC's total population. As detailed below, to account for the uncertainty that this introduces in homeless rate calculations, we model the total population of a CoC as a random quantity centered around our estimate. As noted in the Introduction, it is believed that PIT counts of unsheltered individuals are not fully representative of the unsheltered homeless population. To address potential differences between the counted and actual total size of the homeless population at the CoC-level, we treat the total number of persons experiencing homelessness as missing data and impute the unobserved number of persons experiencing homelessness with a statistical model, which is discussed in Section 3.

3 Statistical Model for Quantifying Uncertainty

The statistical model we use for quantifying uncertainty in PIT counts requires accounting for uncertainty in our estimates of rates of homelessness across CoCs as well as some assumptions about the accuracy of counts in each CoC. In this section, we provide a conceptual overview of our modeling strategy and its connection to specific data challenges. Mathematical formalization of these model components is presented in the Appendix. For a full presentation of the model and computational strategy, see Glynn and Fox (2018) and Glynn et al. (2018).

One way of quantifying uncertainty in PIT counts is to predict the outcome of many additional

[3]The Census Bureau publishes 1-year, 3-year and 5-year ACS estimates. Only the 5-year ACS estimates are available at the Census tract level.

(hypothetical) counts. The resulting range of predictions from these hypothetical counts places the observed count in context. We adopt a Bayesian modeling and computational strategy that constructs uncertainty intervals on PIT counts via stochastic simulation methods. Our strategy has two stages. First, we compute posterior distributions for model parameters (including the rate of homelessness) as part of a Bayesian estimation procedure. These posterior distributions for parameters take into account observed PIT counts, total CoC population estimates, and assumptions about count accuracy, which we discuss in Section 3.2. In the second stage, samples from our posterior distributions are utilized to generate realizations of synthetic counts from posterior predictive distributions. These synthetic count values, while not actually observed, are informed by both (i) the actually observed PIT counts and (ii) the level of variation in the counts – both across and within communities. The model architecture allows uncertainty from estimates of total populations to flow through to inference on the total number of persons experiencing homelessness. Uncertainty in both a CoC's total population and its total homeless population, in addition to uncertainty in the accuracy of its PIT count, propagates to a sampling model for the counted number of homeless. Our hierarchical Bayesian model generates predictions of many synthetic PIT counts – conditioned on actually observed PIT counts and CoC total populations – enabling us to construct uncertainty intervals.

3.1 Framework for quantifying uncertainty in rates of homelessness

A critical step in quantifying uncertainty in PIT counts is modeling the homeless rate as a random quantity. Estimating the homeless rate requires identifying sources of uncertainty in both numerator and denominator. In the numerator, the reality that PIT counts do not reflect the size of the total homeless population necessitates modeling the total number of homeless as a latent variable. The homeless rate is also directly impacted by the CoC's total (overall) population in the denominator. While we construct an estimate of a CoC's total population from ACS 5-year estimates of total population at the census tract level, the CoC's total population is not directly observed. Modeling noise in both numerator and denominator allows for a more complete accounting of uncertainty in homeless rates, which is important when predicting the outcome of synthetic homeless counts. The logit transformation of the latent homeless rate is further modeled as a linear regression on the Zillow Rent Index as a share of median income and the rate of extreme poverty to account for CoC-level features. Glynn and Fox (2018) propose a frame-

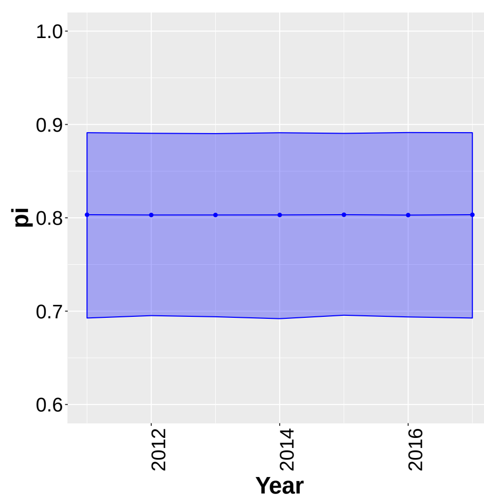


Figure 1: 95% uncertainty interval in the count accuracy for WA-500. The center of the prior distribution is chosen by assuming that 60% of unsheltered homeless are counted while 95% of sheltered homeless are counted. Though it is possible for the count accuracy to increase or decrease with time, we assume here that the center of the distribution is stable over time.

work where uncertainty in a CoC's total population, total number of homeless, and the rate of homelessness is integrated with a sampling model for PIT counts. We utilize the same Bayesian statistical framework for homeless counts in the 386 CoCs in order to (i) estimate the total number of homeless in each CoC; and (ii) aggregate the CoC-level estimates of total homeless to a single national estimate, and (iii) construct uncertainty intervals for the PIT counts in each CoC. Essential components of that modeling framework are summarized in the Appendix, and full details are available in Glynn and Fox (2018).

3.2 Modeling assumptions

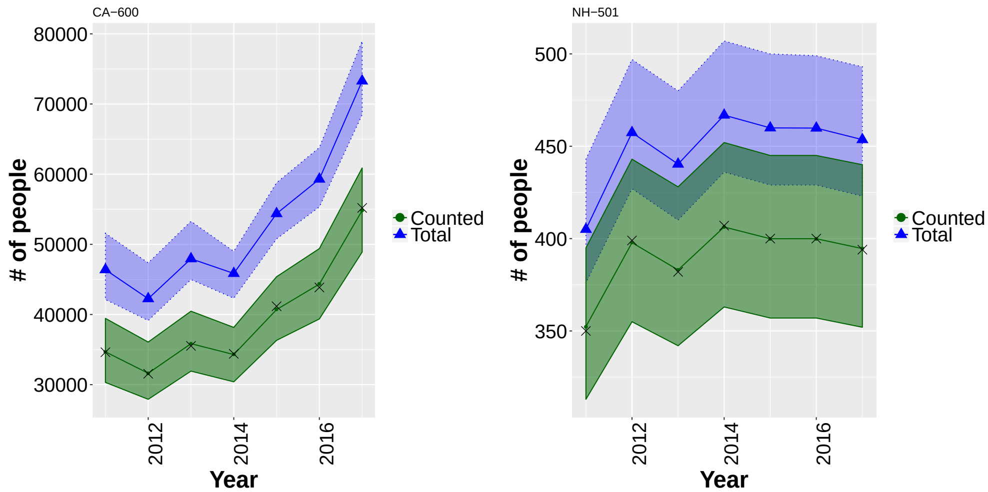
Our modeling approach makes several assumptions that we believe to be reasonable. First, we assume that the counted number of homeless reflected in PIT estimates is less than or equal to the total number of homeless in each community (see equation 3 in the Appendix). Second, we assume that the accuracy of the count varies from one CoC to another and depends on the size of the unsheltered population (see equation 4 in the Appendix). The count accuracy is the probability of a person who is actually homeless being included in the count. We follow Glynn and Fox (2018) in specifying a prior distribution for count accuracy rather than fixing a single number. By eliciting a prior distribution for count accuracy, we acknowledge our uncertainty in the true underlying quantity. The prior mean is chosen by assuming that 60% of unsheltered homeless are counted in each CoC – based on estimates by Hopper et al. (2008) – and that 95% of sheltered homeless are counted in each CoC, allowing for small discrepancies or administrative errors in counts of sheltered homeless. Note that these assumptions are used to compute the prior mean only, and that uncertainty in the count accuracy is of critical importance. Figure 1 illustrates our uncertainty in count accuracy in Seattle from 2011-2017.

Our prior distribution is consistent with a belief that if a person is experiencing homelessness in Seattle, there is between a 70-90% chance that person will be included in the Seattle / King County PIT count. While the prior distribution is flat over time, that does not mean that the count accuracy is the same from one year to the next. For example, the prior distribution in Figure 1 is also consistent with a count accuracy of 0.75 in 2016 and a count accuracy of 0.88 in 2017.

4 Results

4.1 CoC-level estimates of total homelessness

For each of 386 CoCs, we construct uncertainty intervals for PIT counts as in Figure 2, which presents results for Los Angeles County and Manchester, NH – two CoCs of vastly different size. The observed PIT counts in LA and Manchester are presented with 'x' marks. The posterior predictive distribution for synthetic PIT counts is presented in green, with the center line representing the mean and the shaded green interval a 90% prediction interval. That is, if there were 1000 independent (hypothetical) PIT counts conducted, 900 of them would return values within the green shaded interval. The blue triangles present our estimate for each CoC's total homeless population, with the blue shaded region representing the 90% posterior predictive interval. While the distribution for the total number of homeless (blue) is sensitive to assumptions about accuracy of sheltered and unsheltered counts, the predictive distribution for synthetic PIT counts (green) is robust to these assumptions. Uncertainty intervals provide context for local counts. As



(a) Los Angeles County (CA-600)

(b) Manchester CoC (NH-501)

Figure 2: Uncertainty intervals for PIT counts in Los Angeles County and Manchester, NH. The observed PIT counts are presented with 'x' marks. The shaded green interval presents the 90% posterior predictive interval for synthetic PIT counts. The shaded blue represents our estimate of the CoC's total homeless population.

an example, the observed and reported PIT count in Los Angeles increased from 43,854 in 2016 to 55,188 in 2017. This increase of more than 11,000 is alarming; however, the large increase in the count may be a result of sampling variability. Our analysis suggests that it is possible that if the 2016 count had been repeated, as many as 49,300 could have been counted. In 2017, the observed count seems abnormally high. It is possible that if the PIT count were repeated, as few as 48,900 homeless could have been counted. In this alternate scenario for counts, the increase of more than 11,000 in the observed count becomes a reduction of 400 in the synthetic counts. We are not arguing that the homeless population in Los Angeles didn't increase from 2016 to 2017. We are arguing that the magnitude of the increase in the homeless count is less certain than that calculated from the raw PIT data.

An important feature of our method is that the uncertainty intervals naturally adjust based on the size of the community. In Los Angeles (Figure 2a), the shaded green interval spans more than 10,000 people. The larger the community and the larger the homeless population, the wider the uncertainty interval. By contrast, the uncertainty interval in Manchester, NH (Figure 2b) spans approximately 100 people. Uncertainty intervals that naturally adapt to account for the size of a CoC and its homeless population provide more helpful interpretations to stakeholders. PIT count uncertainty intervals and the predicted total homeless population in each CoC are available at https://github.com/G-Lynn/Inflection.

4.2 National estimates of total homelessness

By summing posterior samples for the synthetic PIT count in each CoC, we arrive at a predictive distribution for the 386 CoC's in total. Figure 3a illustrates the sampling variability inherent in national homeless counts in 2017. The vertical red line marks the sum of the observed PIT counts in each of the 386 CoCs in our analysis (546,566). The histogram demonstrates that, if repeated, the PIT count could plausibly be as low as 530,000 or as high as 565,000. A similar aggregation strategy for posterior samples of the total homeless population in each CoC provides a distribution for the total homeless population at the national level. Though sensitive to assumptions outlined in Section 3.2, Figure 3b illustrates that the total homeless population in 2017 is potentially 100,000 more than the 2017 PIT counts suggest, with estimates of the total homeless population ranging from 650,000 to 670,000 people and a mean of approximately 661,000 people.

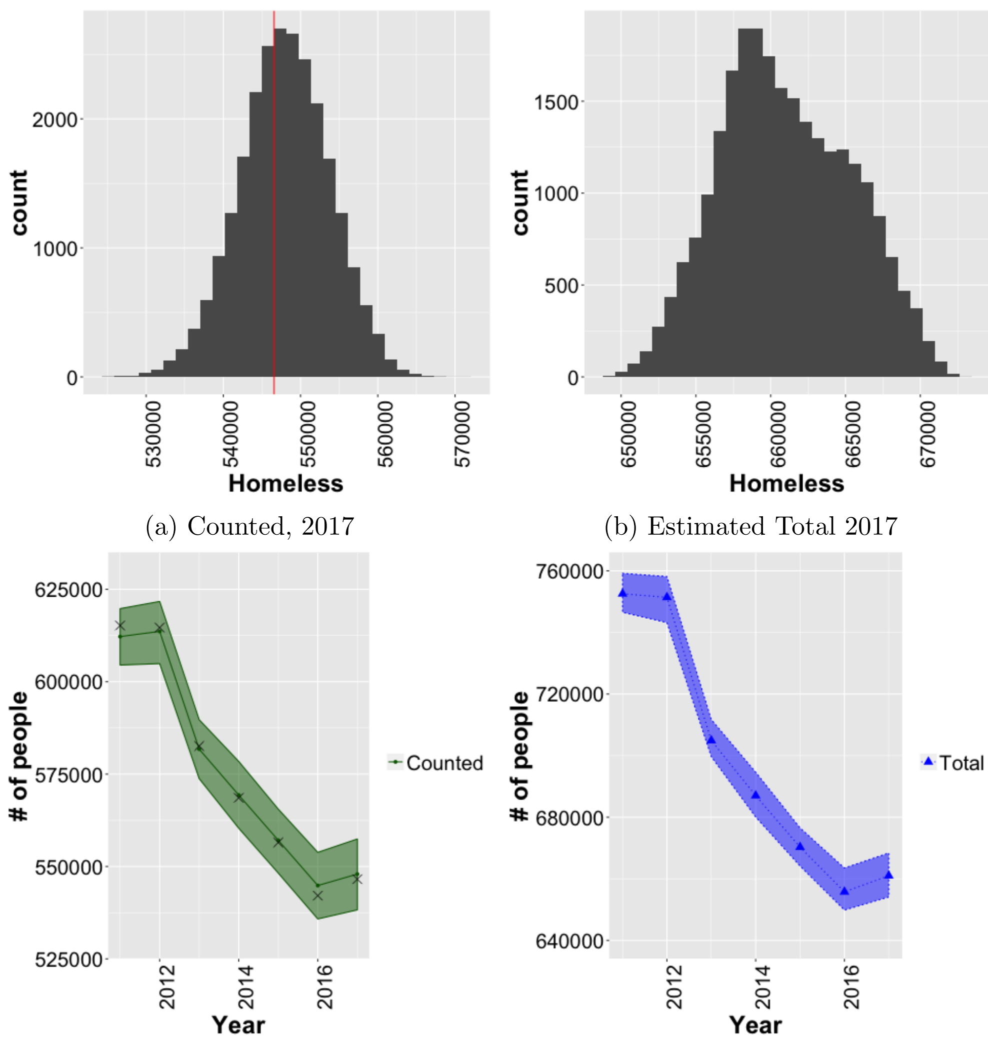
Our aim is not to identify a single number for the total homeless population. Our goal is to build methods capable of including outside sources of information – such as estimates of the undercount of unsheltered homeless – and advance the discussion on the scope of homelessness with estimates that include uncertainty intervals. To demonstrate why this latter point is important, consider the more than 30,000 person drop in the national homeless count between 2012 and 2013 shown in Figure 3c. Our uncertainty intervals suggest it is possible that if counts were repeated, the 2012 national count could be as low as 605,000 while in 2013, another count could have returned as many as 590,000 homeless nationwide. The 30,000 person reduction in homelessness according to the raw PIT data could have been as small as 15,000 if additional counts had occurred. Again, we are not arguing that there wasn't a significant reduction in homelessness nationwide between 2012 and 2013. We are arguing that there is less certainty in the size of the reduction than calculated from the raw PIT counts.

The feature of retrospective temporal smoothing is also present in our estimates of national homelessness. Observe in Figure 3c that the center green line is a smoothed version of the actual data. In 2011 and 2012, the method returns estimates that are likely lower than the PIT counts in those years. By contrast, in 2016 and 2017, the method returns estimates that are likely higher than the PIT counts. The same general trend of decreasing homelessness nationwide is present in Figure 3d. Although the estimates for the total homeless population shown in Figure 3d are sensitive to assumptions about count accuracy, it is possible for local count coordinators and individual CoCs to provide information as part of their reporting to HUD that would make this estimate significantly more informed.

5 Discussion

To our knowledge, the analysis reported here represents the first attempt to estimate uncertainty around HUD PIT counts. Our contributions are both methodological and applied. To advance statistical methods for analyzing homeless data, we present a framework for considering sampling variability in PIT counts with a Bayesian statistical model. The model naturally accounts for the size of each CoC, local costs of rental housing, and further incorporates outside sources of information like estimates of the extent to which unsheltered homeless populations are undercounted. We use prediction of hypothetical homeless counts in each year as a central tool to place observed PIT data in context. We use this model and baseline assumptions about count accuracy to (i) construct uncertainty intervals around reported PIT counts for 386 CoCs from 2011-2017; (ii) estimate the size of the total homeless population in each CoC; and (iii) aggregate these local estimates to the national level in a way that accounts for sampling variability in national homeless estimates.

We view our method and analysis not as the final word on this subject. Rather, it is one step



(a) Counted, 2017 (b) Estimated Total 2017

(c) Estimated counts, 2011-2017 (d) Estimated total, 2011-2017

Figure 3: National estimates of homelessness from 2011-2017 using data from 386 CoCs. Top left: Histogram of the synthetic PIT counts from 2017. The vertical red line marks 546,566 – the sum of actual 2017 PIT counts from 386 CoCs. Top right: Histogram of predictive distribution for total homeless population in 2017 based on assumptions outlined in Section 3.2. Bottom left: Posterior predictive distribution for synthetic PIT counts from 2011-2017. The black 'x' marks present the aggregated raw PIT counts from 386 CoCs. The shaded interval marks the 90% posterior predictive interval. Bottom right: Estimate of total homeless population. The triangles mark the posterior mean in each year, and the shaded interval marks the 90% credible interval.

in an iterative process of refining local and national estimates of homelessness. PIT counts are an absolutely critical source of information on the scope of homelessness; however, external sources of information like count accuracy, unsheltered rates, and total housing costs can be used to augment PIT counts. Integrating PIT counts with additional data and domain expertise provides policymakers, program evaluators, and stakeholders with a more comprehensive view of progress and remaining challenges.

Our analysis underscores the importance of quantifying the degree of uncertainty around estimates of the size of the homeless population from several perspectives. From a resource allocation and program planning perspective, this information is crucial for policymakers, service providers and other stakeholders to have a complete understanding about the actual scope of the problem and how it is (or is not) changing over time. From a research perspective, accounting for uncertainty in estimates of homelessness could help improve the quality of the growing body of research that relies on PIT count data to examine factors that explain variation in homelessness across communities and over time. Finally, from the perspective of public discourse on homelessness, providing information about uncertainty around PIT counts acknowledges longstanding concerns about the perceived inaccuracy of official counts of homelessness. Such an acknowledgment may help shift the public conversation away from debates about the reliability of homeless counts and towards solutions to prevent and end homelessness.

Our study has a number of limitations that are important to acknowledge. First, our model uses the same prior estimate for the degree to which the unsheltered and sheltered populations are undercounted for all CoCs. The different prior distributions across CoCs come from the different relative mix of the unsheltered and sheltered populations in each CoC. Because local information about the undercount of the unsheltered populations was not available from all CoCs, we based the selection of this prior on research conducted in one jurisdiction (New York City). However, the extent to which the unsheltered population is undercounted (or potentially overcounted) is likely to vary across communities. Second, our analysis only considers the total population of persons experiencing homelessness. It is likely that the degree of uncertainty varies among the numerous sub-populations (e.g. persons in families with children, military Veterans, those experiencing chronic homelessness) who are included in the PIT count. Thus, the exact composition of the homeless population in a community may affect the degree of uncertainty around the overall PIT count. Third, HUD has retroactivley made adjustments to PIT counts for some years. We use data as they were reported in 2017, and our method doesn't account for ad hoc adjustments made to PIT counts over time. Future work could further model these post-count adjustments by HUD.

Future work should address these limitations and would result in improvements to our model. In the ideal case, more community-specific information could be integrated into the modeling approach. For example, local information about the extent to which individuals in unsheltered locations are undercounted could improve the accuracy of the assumptions underpinning our model. Obtaining this information is not without its challenges, but HUD could incentivize communities to conduct plant-capture studies or post-count surveys as part of their PIT enumeration efforts.

We wish to conclude by emphasizing that the goal of our analysis is not to call into question the validity of the current PIT counts or the efforts of those that participate in their execution. Conducting PIT counts is difficult and essential work. Rather, our intent is to provide a methodology that can be used to provide credible estimates of the otherwise unknown and unreported degree of uncertainty around these counts. Reporting this type of information is done as a matter of standard practice with a wide array of other official government statistics, public opinion surveys and political polls. We argue that the same should be true of official estimates of the number of persons experiencing homelessness, and hope that our model might be used and refined for developing and reporting uncertainty estimates for future PIT counts.

References

Bun, Y. (2012). Zillow rent index: Methodology. https://www.zillow.com/research/zillow-rent-index-methodology-2393/. [Online; accessed 04/2/2017].

Burt, M. R., Aron, L. Y., Douglas, T., Valente, J., Lee, E., and Iwen, B. (1999). Homelessness: Programs and the People they Serve. Urban Institute, Washington, D.C.

Burt, M. R. and Cohen, B. (1989). America's Homeless: Numbers, Characteristics and the Programs that Serve Them. Technical report, Urban Institute, Washington, D.C.

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., Fargo, J., Montgomery, A., and Culhane, D. (2013). New Perspectives on community-level determinants of homelessness. Journal of Urban Affairs, 35(5):607–625.

Corinth, K. (2015). Street homelessness: A disappearing act? Technical report, American Enterprise Institute, Washington, D.C.

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, 41:45–56.

Glynn, C., Byrne, T. H., and Culhane, D. P. (2018+). Inflection points in community-level homeless rates.

Glynn, C. and Fox, E. B. (2018). Dynamics of homelessness in urban America. Annals of Applied Statistics, forthcoming.

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.

Lucas, D. S. (2017). The Impact of Federal Homelessness Funding on Homelessness. Southern Economic Journal, 84(2):548–576.

Martin, E. (1992). Assessment of S-Night Street Enumeration in the 1990 Census. Evaluation Review, 16(4):418–438.

Martin, E., Laska, E., Hopper, K., and Wanderling, J. (1997). Issues in the use of a plant-capture method for estimating the size of the street dwelling population. Journal of Official Statistics, 13(1):59.

Metraux, S., Culhane, D., Raphael, S., White, M., Pearson, C., Hirsch, E., Ferrell, P., Rice, S., Ritter, B., and Cleghorn, J. S. (2001). Assessing Homeless Population Size Through the Use of Emergency and Transitional Shelter Services in 1998: Results from the Analysis of Administrative Data from Nine US Jurisdictions. Public Health Reports, 116(4):344–352.

Rossi, P., Wright, J., Fisher, G., and Willis, G. (1987). The urban homeless: estimating composition and size. Science, 235(4794):1336–1341.

Taeuber, C. M. and Siegel, P. M. (1990). Counting the Nation's Homeless Population in the 1990 Census. Technical report, U.S. Census Bureau, Washington, D.C.

United States General Accounting Office (1991). Limitations in Methods and Procedures to Include the Homeless. Technical report, Washington, DC.

U.S. Department of Housing and Urban Development (1984). A Report to the Secretary on the Homeless and Emergency Shelters. Technical report, U.S. Government Printing Office, Washington, D.C.

U.S. Department of Housing and Urban Development (2017a). PIT and HIC Data Since 2017.

U.S. Department of Housing and Urban Development (2017b). The 2017 Annual Homeless Assessment Report to Congress, Part 1: Point-in-Time Estimates of Homelessness. Technical report, U.S. Department of Housing and Urban Development, Washington, D.C.

U.S. Department of Housing and Urban Development (2018). Continuum of Care GIS Tools.

A Appendix

We present essential components of the modeling framework used to quantify uncertainty in rates of homelessness.

Model for a CoC's total population. The overall population for CoC i in year t is modeled with a Poisson random variable in equation 1.

$$N_{i,t} \sim Poisson(\lambda_{i,t}) \tag{1}$$

The expected total population in year t, $\lambda_{i,t}$, is further modeled over time in a way that admits an efficient computational algorithm for estimation.

Model for a CoC's total homeless population. A binomial thinning step in equation 2 is employed so that $H_{i,t}$ represents some small fraction of the CoC's total population.

$$H_{i,t}|N_{i,t}, p_{i,t} \sim Binomial(N_{i,t}, p_{i,t}) \tag{2}$$

The homeless rate, $p_{i,t}$, is further modeled over time. See Section 3.2 of Glynn and Fox (2018) for further detail. Note that neither $H_{i,t}$ nor $p_{i,t}$ are actually observed, and they are treated as latent variables that will be estimated as part of the Bayesian model fitting procedure.

Model for a CoC's PIT count. The counted number of homeless in a CoC is denoted $C_{i,t}$. The $C_{i,t}$ count is a quantity less than or equal to the total number of homeless, $H_{i,t}$. We again use the binomial thinning strategy based on a count accuracy parameter for each CoC, $\pi_{i,t}$ – the probability that a homeless person is included in the PIT count.

$$C_{i,t}|H_{i,t}, \pi_{i,t} \sim Binomial(H_{i,t}, \pi_{i,t}) \tag{3}$$

$$\pi_{i,t} \sim Beta(r_{i,t}, s_{i,t}) \tag{4}$$

The important part is that we do not ever observe $\pi_{i,t}$ or have any way of estimating it, since we do not observe the true homeless count, $H_{i,t}$. Instead of learning $\pi_{i,t}$, we integrate it out of the model for $C_{i,t}$ so that our marginal distribution for $C_{i,t}|H_{i,t}$ is Beta-Binomial.

A posterior predictive distributions for new PIT counts. The uncertainty interval for $C_{i,t}$, the observed PIT count for CoC i in year t is constructed from the posterior predictive distribution

$$C^*_{i,t}|C_{1:386,1:T}, N_{1:386,1:T}. \tag{5}$$

Conceptually, equation 5 is the predictive distribution where unobserved latent variables $p_{i,t}$, the predicted CoC total population, and the total number of homeless are integrated over – resulting in a prediction for a new homeless count that incorporates these various sources of uncertainty and is informed by the observed PIT counts and CoC total population estimates.