CID Technical Notes Series 2024-03:

Measuring Longitudinal HUD Benefit Receipt with Administrative Data[1]

Angela Wyse
University of Chicago

Gillian Meyer
University of Pennsylvania

Derek Wu
University of Virginia

Bruce D. Meyer
University of Chicago,
NBER, and AEI

May 6, 2024

I. Overview

In this note, we describe our methodology for measuring longitudinal housing benefit receipt using administrative data from the U.S. Department of Housing and Urban Development (HUD). Each month, HUD aids over 4.5 million households through a variety of programs, including public housing, tenant-based housing, and project-based housing.[2]

We draw on administrative datasets from HUD to create household- and individual-level datasets indicating monthly receipt and benefit amounts for 1995-2016. This memo describes our approach to constructing these datasets, including the steps we take to impute housing benefit amounts where missing and to construct housing assistance spells from available variables. We compare year-by-year estimates of households receiving HUD assistance in our datasets to public aggregates and find that, between 2004 and 2016, our estimates are within 4.5 percent of those sources in each year, with the exception of 2009, where our estimates are within 5.7 percent of public aggregates. Our estimates benchmark less well against public aggregates in earlier years, falling within 9.6 percent of those sources between 1995 and 2000, a fact that likely reflects the incompleteness of the PIC/TRACS datasets in these years. Comparable public aggregates are not available for 1999 and 2001-2003.

II. Description of Administrative Data

We begin with raw data from HUD's Public and Indian Housing Information Center (PIC) and Tenant Rental Assistance Certification System (TRACS) databases. PIC and TRACS are HUD's

---

two main computer systems for managing administrative housing assistance data. The PIC database tracks participation in public housing, Indian housing, and Section 8 rental subsidy programs, while TRACS tracks participation in other types of multifamily project-based housing projects. The datasets available to Census researchers include benefits received from 1995 to 2016.

In the raw data, each observation represents an action from a single household. Actions fall into three categories: move-ins, move-outs, and annual recertifications. The raw data also indicate the start date of the housing assistance spell ("admission date") and the date on which the action occurred. The data also indicate tenant payments and, except for public housing units, the gross rent of the unit. These data can be linked to other data at the person level using Protected Identification Keys (PIKs).

III. Constructing the Longitudinal Dataset

This section describes our approach to constructing the longitudinal dataset. We begin with a handful of basic cleaning steps, including collapsing actions into the three above categories and standardizing geographic variables. Next, we impute gross rent and total tenant payment in cases where these values are missing as a precursor to calculating housing benefit amounts. Finally, we group actions into housing assistance spells and identify or impute a start and end date for each spell. We then transform these data into datasets indicating monthly household- and person-level benefit receipt and amounts for 1995-2016.

Imputing missing housing benefit amounts

We define housing benefit amount as the difference between gross rent (total housing costs, including rent and utilities) and total tenant payment (the minimum amount households must contribute toward rent and utilities). Because public housing units do not have a monthly rent amount, we impute the value of gross rent in these cases. We also impute gross rent for private housing actions where gross rent is either missing or implausibly low ($5/month or less, a value that appears to have been used as a placeholder for missing gross rent values).

Where necessary, we first attempt to impute gross rent by drawing from surrounding actions for the same household. If gross rent is missing on all actions for a given household, we impute gross rent as the mean of gross rent in a cell defined by the number of bedrooms and zip code. We expand each cell as necessary using progressively larger geographic units. We follow a similar procedure to impute total tenant payment with one important modification. Specifically, before proceeding to impute based on nearby households, we first attempt to impute a value based using HUD's formula for determining tenant payments, a formula that draws on total and adjusted monthly income values which may be reported in our data.

We selected this imputation procedure after considering several alternatives, including models that incorporate HUD's fair market rent values, fixed effect regression models, and traditional hot decking methods. We evaluated these alternatives by removing gross rent from a randomly selected subset of observations and imputing a value using each method. Our chosen method produced the smallest mean squared error of the alternatives considered.

Constructing housing assistance spells

We aim to identify the start and end dates for all housing assistance spells. However, not all spells have move-in and move-out actions, nor do they all have recertifications at least once a year, as required in principle. This is because the Public Housing Authorities (PHAs) that report these actions to HUD do so incompletely. According to published HUD reports, PHAs only reported about 75% of all actions in 1996 but have improved their reporting rates over time to about 95-100% in 2008-2016.[3]

Our procedure for constructing housing assistance spells consists of four steps:

1. **Group actions into housing spells.** We group actions into spells using admission date, if available, or otherwise unique address identifiers called MAFIDs.
2. **Identify or impute start dates for each spell.** We identify spell start dates using admission date or move-in date if available. If these are missing, for example if the first action in the spell is a recertification or move-out, we impute a spell start date 18 or 30 months prior to the first observed action (18 months for actions in 2000 and later and 30 months for actions before then).[4]
3. **Identify or impute end dates for each spell.** If the final action in a spell is a move-out, we take this to be the spell end date. Otherwise, we impute a spell end date 18 or 30 months after the last observed action (18 months for actions in July 1998 or later, 30 months for actions before then).
4. **Eliminate overlap between spells if present.** Our final step eliminates a small number of overlapping spells by setting the end date of any overlapping spell equal to the day before the next spell's start date.

## IV. Benchmarking the Cleaned Dataset

We benchmark the cleaned dataset by comparing the annual number of recipient households in the cleaned data to aggregates from HUD's publicly available Picture of Subsidized Households (PSH). Figure 1 displays the number of households with assistance in a single month according to published aggregates ("HUD public aggregate – single month"), along with a series indicating the number of households with assistance in that same month according to our clean data ("Cleaned PIC/TRACS – single month"). We find that the single month estimates from our clean data are within 9.6 percent of public aggregates in 1996-2000 and within 5.7 percent in 2004-2016. Under-coverage in earlier years likely reflects the incompleteness of administrative data in those years due to relatively low reporting rates from public housing authorities.

---

[3] HUD's Picture of Subsidized Housing (PSH) 1996-2016 reports
[4] We are less conservative in early years when PHAs reported a smaller share of actions to the PIC/TRACS system.

IV. Figure

Figure 1: Total Households Receiving HUD Assistance, 1995-2016

Total Households Receiving HUD Assistance, 1995-2016
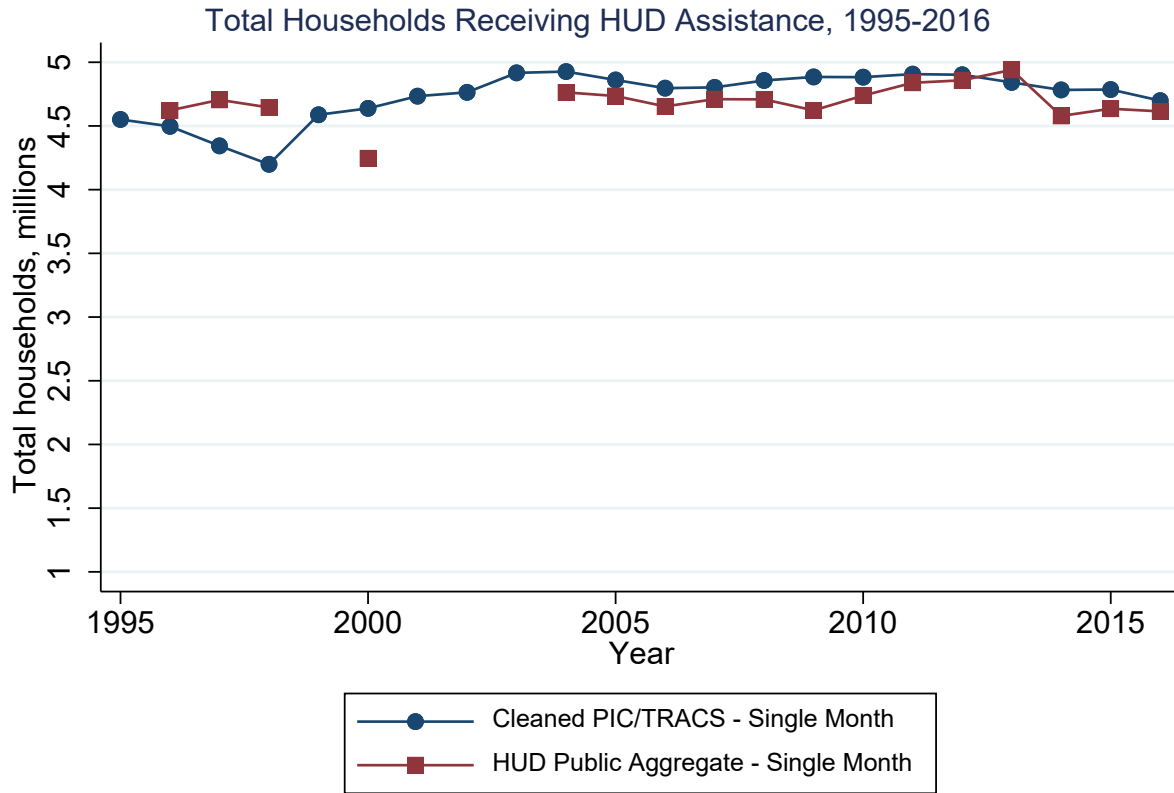
Source: HUD PIC/TRACS longitudinal data (1995-2016), HUD's Picture of Subsidized Housing (PSH) 1996-2016 reports