Nonresponse Imputations and Related Measurement Issues in the CPI for Shelter

Lara Loewenstein [1] Hugh Montag [2] Randal Verbrugge [1]

[1]Cleveland Federal Reserve

[2]Bureau of Labor Statistics

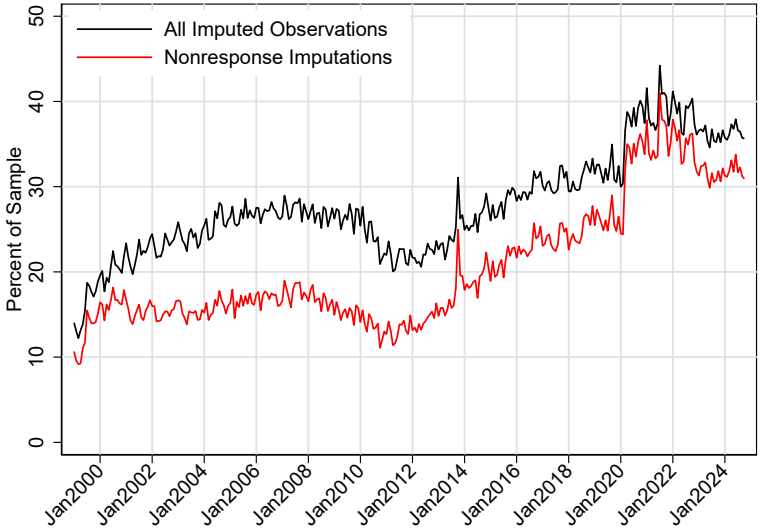
NBER/CRIW Conference on Measurement of Housing and the Housing Sector
12 March 2026

Notes

Disclaimer:
These notes reflect the views of the authors and don't necessarily reflect the official positions of the Federal Reserve Bank of Cleveland, the Federal Reserve Board, or the Federal Reserve System.

Disclaimer:
This presentation provides research results. The information is being released for statistical purposes, to inform interested parties, and to encourage discussion of work in progress. The presentation does not represent an official BLS statistical data product or production series.

Imputations are Significant



This Paper

This paper studies the imputation process for the BLS Housing Survey

- Non-responses

Significance

- Shelter represents about a third of Consumer Price Index (CPI) basket
- Imputation observations $\approx$ 30 %

Research Questions

- Does the imputation rate vary by rental unit characteristics?
- Which characteristics are significant for imputations?
- How accurate are current non-response imputations?
- What effect does this have on the rent index in the CPI?

BLS Housing Survey

- ≈40,000 rental units surveyed every 6 months
- Data from 1999–2024, with some changes in the survey design.
- Units are divided into 6-month panels (January-July, February-August, ...)
- Rental units selected within sampled Census Block Groups ("neighborhoods"), within each sampled metro areas
- Mostly continuing tenants, 18% are new tenants
- Respondents can be tenants, landlords or property management companies
- Missing data imputed with group means approach*
- Variables include:
  1. Tenant move-in date*
  2. Unit information, such as structure type (single family detached, single family attached, apartment, other)
  3. Indicator for whether unit is rent controlled.
  4. Weights used in CPI
  5. Imputation indicator

Tenancy Length Imputation

- Tenancy length is not observed for non-responding housing units
  - Ex. How many new tenants move-in if a unit non-responds for 3 years?
- Housing Survey microdata carries move-in date forward from last collected observation
- With full history of housing unit, can augment with move-in date of future responding tenant

Alternative non-response tenancy length imputation approaches

1. Use "raw" microdata move-in date
2. Assume no non-responding tenants move-out (lower bound)
3. Assume non-responding tenants move-out every 12 months (upper bound)
4. Supplement with American Community Survey (ACS) data to derive new tenant share of non-responses $x_{c,t}^{Nonresponse}$

$$x_{c,t}^{ACS} = w_{c,t}^{Collected} x_{c,t}^{Collected} + w_{c,t}^{Nonresponse} x_{c,t}^{Nonresponse}$$

Non-Response Imputation

Issue: Housing unit is initialized, but tenant and/or landlord cannot be contacted or is unwilling to continue participating in the survey.

- Rent changes extremely heterogeneous, especially based on observables.
- If unit responds in future survey, then any imputation error is corrected then.
- Sample rotation began in 2016, non-response unit has 1/12th chance of exiting.

Official Procedure ("Rent Tier Approach")

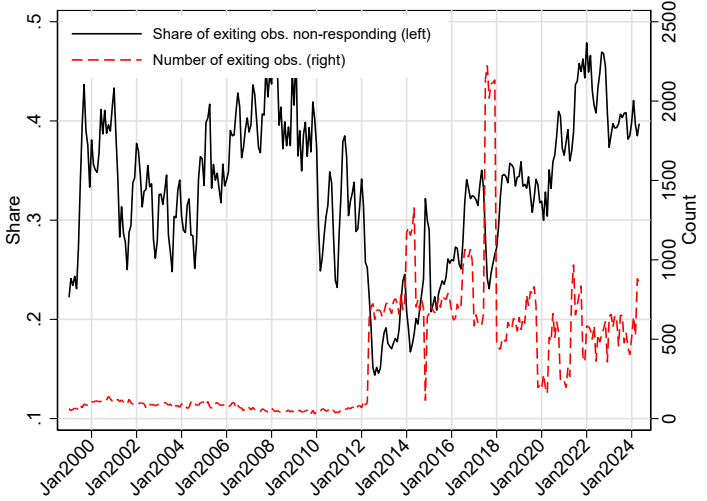
1. Separate housing units by city-date
2. For each city-date, sort housing units by rent at last collection period (*t-6*)
3. Partition units by rent level into tertiles (low, medium, and high rent)
   - Collapse cells by rent level if statistically insufficient units (<units)
4. Let the set $H_{c,r,t}$ denote the set of collected and imputed vacant units in cell (c,r,t). Calculate

$$NII_{c,r,t} = \frac{\sum_{i \in H_{c,r,t}} \omega_{i,t} rent_{i,t}}{\sum_{i \in H_{c,r,t}} \omega_{i,t} rent_{i,t-6}}$$

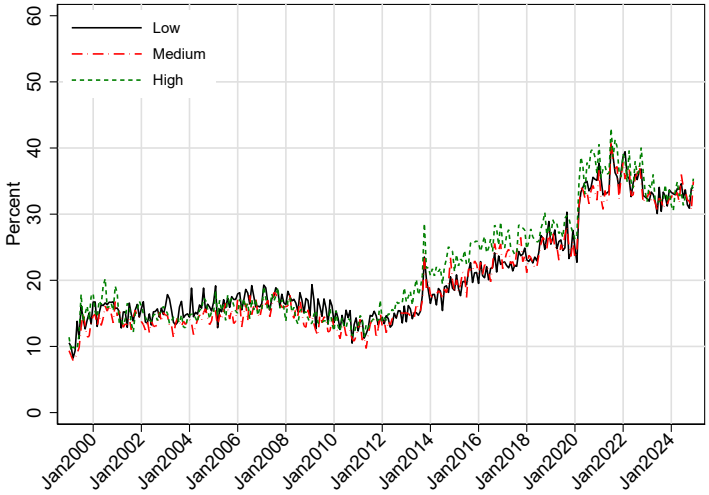
5. For each non-response unit, calculate imputed rent

$$\text{rent}^{*}_{i,t} = NII_{c,r,t} \text{rent}_{i,t-6}$$

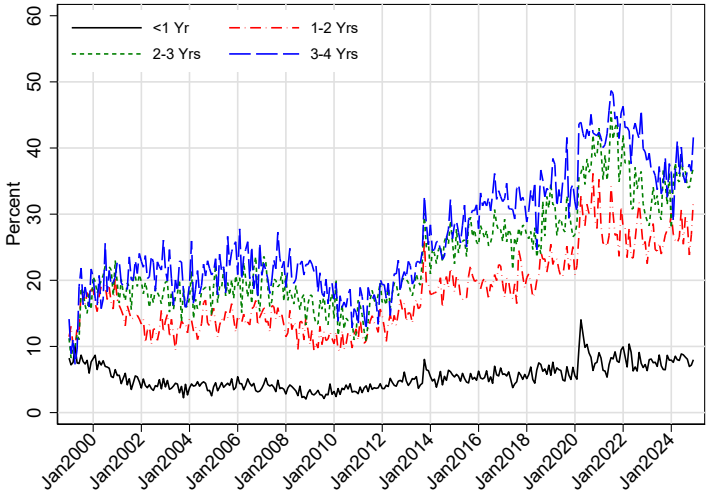
Non-Response Exits



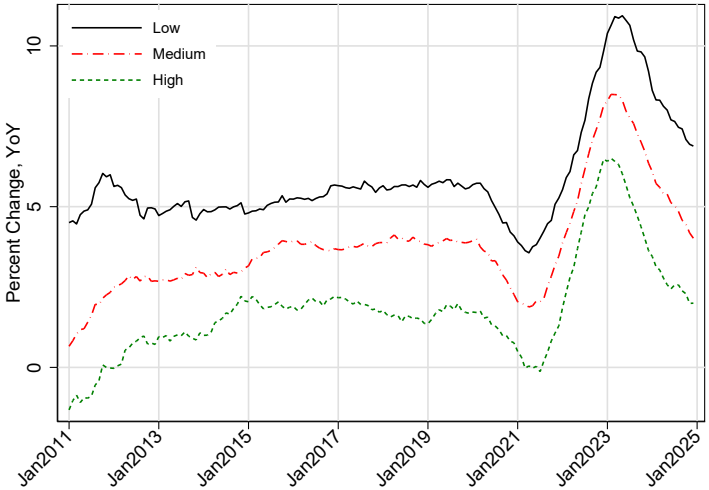
Non-Response Imputation Rate by Rent Tier



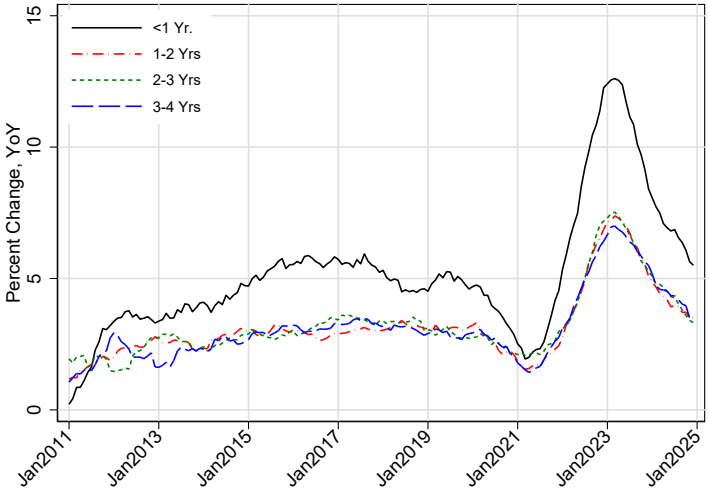
Non-Response Rate by Raw Tenancy Length



Inflation Rate by Rent Tier



Inflation Rate by Raw Tenancy Length



Alternative Non-response Imputations

Q: Non-response rates and inflation rates vary along multiple characteristics. How do alternate imputations fare?

- Rent Tier (CPI Approach)
  - Low
  - Medium
  - High
- Structure
  - Single Family Detached
  - Single Family Attached
  - Multi-family
- Unconditional
- Raw Tenancy Length
  - Short ($<=$ 6 months)
  - Long ($>$ 6 months)
- Extreme Tenancy Assumptions
  - Upper bound (move-out every 12 months)
  - Lower bound (never move-out)
  - Use lagged data instead of cell collapsing
- ACS-based Tenancy Imputation
  - Weighted average of short and long tenancy using $x_{c,t}^{Nonresponse}$
- Hedonic Approach (Erickson-Pakes)

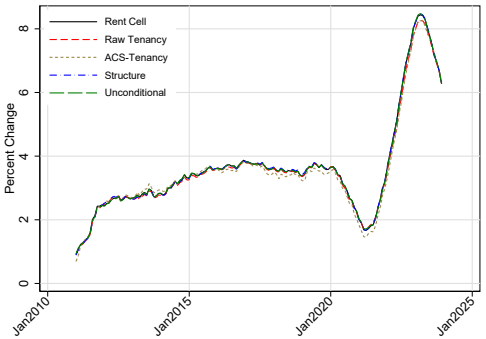
Rent Inflation Rate by Imputation Type, YoY



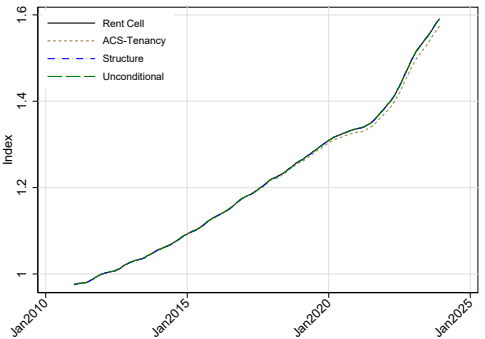
Index Type	$E[\lvert\pi - \pi_{RentCell}\rvert]$
Raw Tenancy	0.042%
Structure	0.024%
Uncond.	0.024%

Rent Inflation Rate by Tenancy Imputation

Tenancy Inflation Rate



Index (Base = 2012)



Index Type	$Ind_{Dec2023}$
Rent Cell	1.590%
ACS-Tenancy	1.574%
Structure	1.592%
Unconditional	1.591%

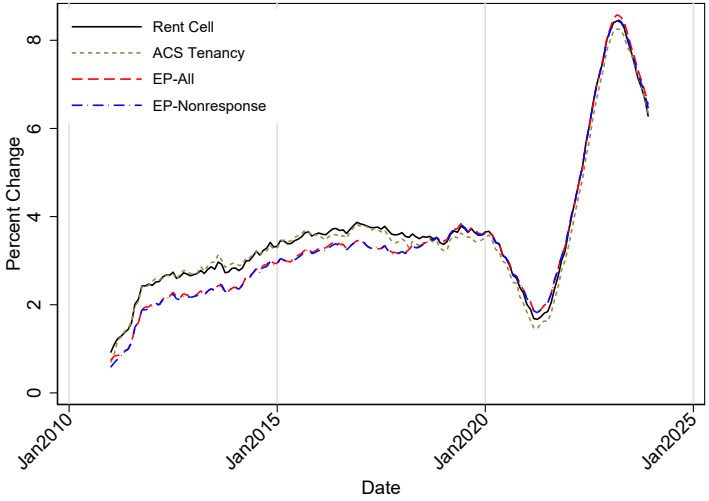
Erickson-Pakes Approach

- Group Means approach falls to curse of dimensionality
  - Even rent tier approach requires cell collapsing due to cell insufficiency
- Erickson-Pakes (2011) hedonic regression permits more characteristics and less cell collapsing
- For each city c and time period t, run the following regression on collected units

$$\frac{ln(rent_{i,t})}{ln(rent_{i,t-6})} = \alpha_{c,t} + \beta_{c,t} T_{i,t} \sum_{j=0}^{2} \gamma_{j,c,t} S_{i,t} + \varepsilon_{i,t} \quad (1)$$

  - $T_{i,t} = 1[tenure_{i,t} < 6 months]$
  - $S_{1,i,t} = 1[Structure_{i,t} = SingleFamilyDetached]$
  - $S_{2,i,t} = 1[Structure_{i,t} = SingleFamilyAttached]$
  - $S_{3,i,t} = 1[Structure_{i,t} = Multifamily]$
- Use ACS-based tenancy measure $x_{c,t}^{Nonresponse}$ to impute $T_{i,t}$ for non-responses
  - $T_{i,t} \in [0, 1]$
- Two approaches
  1. Use fitted equation (1) to impute for non-response observations
  2. Use fitted equation (1) to impute for all observations, including collected observations

Erickson-Pakes: Lower Inflation in the 2010s



Jackknife Setup

Q: How accurate are the non-response imputations?

- Difficult to answer since we don't observe them
- Solution: jackknife approach

Test: For each collected observation i at date t in city c

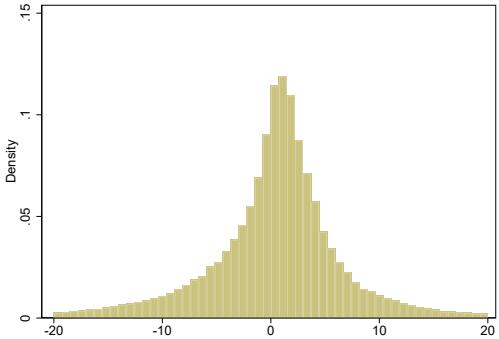
- Treat $x_{i,c,t}$ as a missing observation
- Perform the imputation process with the remaining collected observations
- Impute rent $x^I_{i,c,t}$

$$x^I_{i,c,t} = NII_{-i,c,t} x_{i,c,t-6}$$

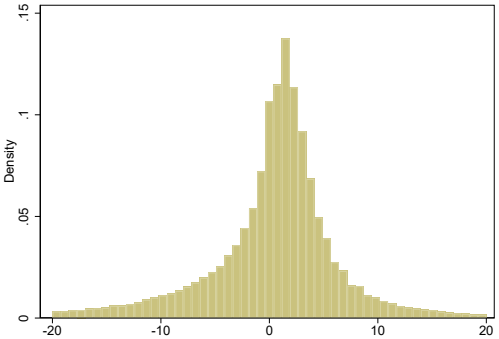
- Then evaluate accuracy of $\{x^I_{i,c,t}\}$

Imputation Jackknife Histogram

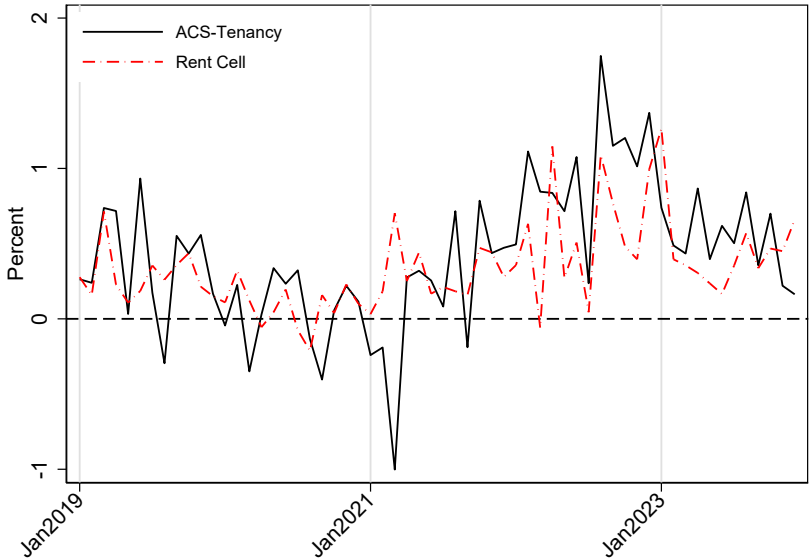
(a) Rent Level



(b) ACS-Tenancy



Imputation Mean Jackknife Error



Conclusion

- Non-response imputations rates have been increasing over time
- Exit rates jumped starting in 2012, increasing significance of imputation accuracy
- Rent cell inflation rates do not differ significantly from unconditional imputations
- Imputing with tenancy length yields different inflation rate, but challenges to measuring non-response tenancy lengths
- Rent cell and ACS-tenancy length imputations appear similarly accurate
- Hedonic imputations result in notably lower inflation estimates

Thank you