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

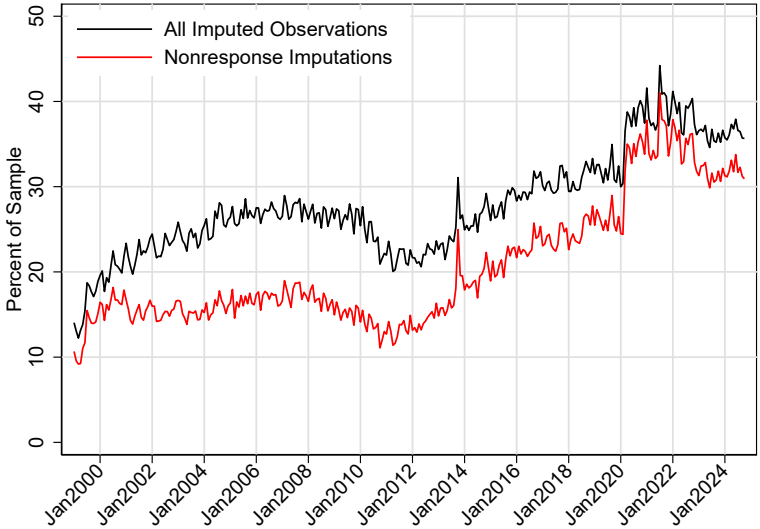
NBER-CRIW
16 July 2025

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

- Vacancies
- Non-responses

Significance

- Shelter represents about a third of CPI basket
- Recent imputation $>$ 30 % observations

Research Questions

- Does the imputation rate vary by observable rental unit characteristics?
  - Current approach accounts for difference by rent level
  - Does not account for difference by tenancy length
- Does this generate an imputation bias?
- How much uncertainty in the aggregate index generated by imputed observations?
- How significant are vacancies?

Preliminary Results

- Imputation rate and inflation rate varies by
  - Tenancy length
  - Structure type
  - Geography
  - Rent level
- Data not missing at random
- However, imputing by other covariates does not significantly affect aggregate index
- Non-response imputation slightly over-estimates rent changes
- Imputations don't add much noise

Still a work in progress.

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
- 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

Summary Statistics (1999-2024)

Unweighted

Imputation Type	Obs. (#)	Share (%)	Mean Rent ($)	Median Rent ($)	Mean Tenancy Length (months)	Median Tenancy Length (months)
Collected Obs	1,375,758	72	944	790	51	27
Vacant Obs	135,558	7	846	668	6	0
Non-response Obs	407,432	21	1132	936	70	49
All	1,918,748	100	977	803	52	29

Imputation: Non-response

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.

Procedure: Impute rent change by Rent Level-City-Date

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
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

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

Imputation Overview: Vacancy

Issue: No transacted rents for vacant units.

- Significant omission if largest rent changes for new tenants

Procedure: Impute rent change by Vacancy Length-City-Date

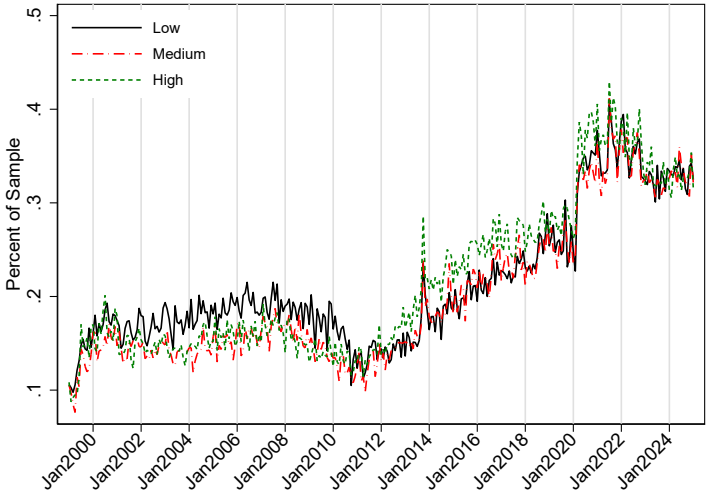
1. Separate housing units by city-date
2. For each city-date, sort vacant units into new and continuing vacancies
3. There are two source pools
   - New vacancies: use collected units with tenancy length $<= 6$ months
   - Continuing vacancies: use collected units with tenancy length > 6 months
   - Collapse by geography if insufficient units
4. Let the set $H_{c,v,t}$ denote the set of units in a source pool for cell (c,v,t). Calculate

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

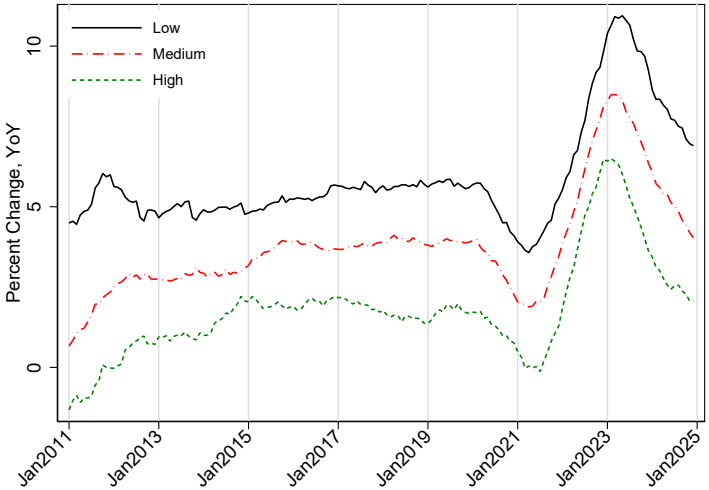
5. For each vacant unit, calculate imputed rent

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

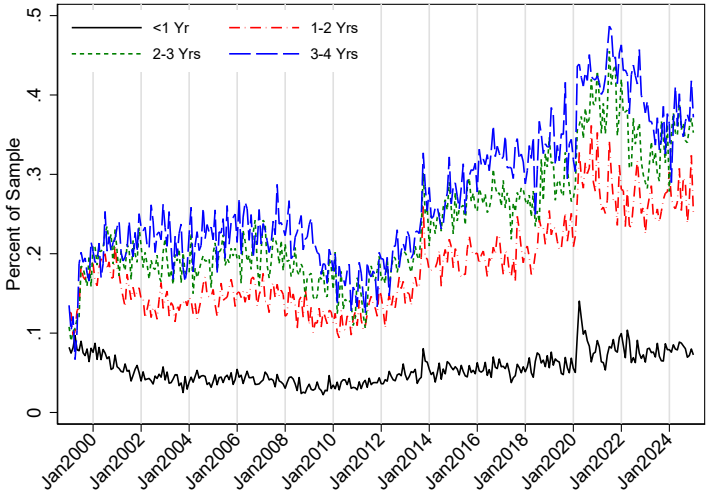
Non-Response Imputation Rate by Rent Level



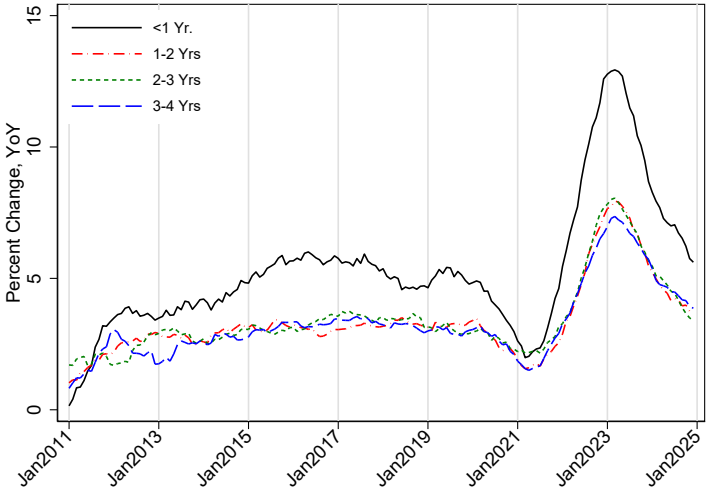
Inflation Rate by Rent Level



Non-Response Rate by Tenancy Length



Inflation Rate by Tenancy Length

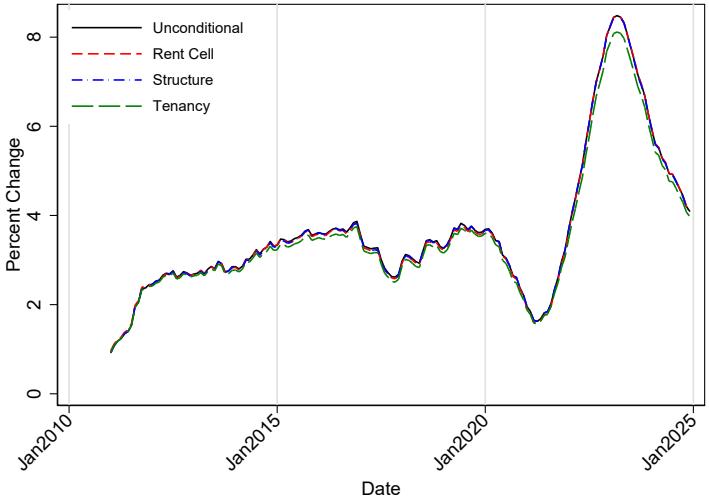


Alternative Non-response Imputations

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

- Rent Level (CPI Approach)
  - Low
  - Medium
  - High
- Structure
  - Single Family Detached
  - Single Family Attached
  - Multi-family
- Tenancy Length
  - Short ($<=$ 6 months)
  - Long ($>$ 6 months)
- Unconditional

Inflation Rate by Imputation Type, YoY



Conditional Imputation Investigation

Q: Why does choice of variable for class means imputation have a small effect on the national index?

- Similar results if conditioning on rent cell, structure type, or tenancy length.
- Even unconditional imputations similar.
- However, non-response rates and inflation rates differ by covariate levels.
- Data is not missing at random.

Conditional Imputation Investigation

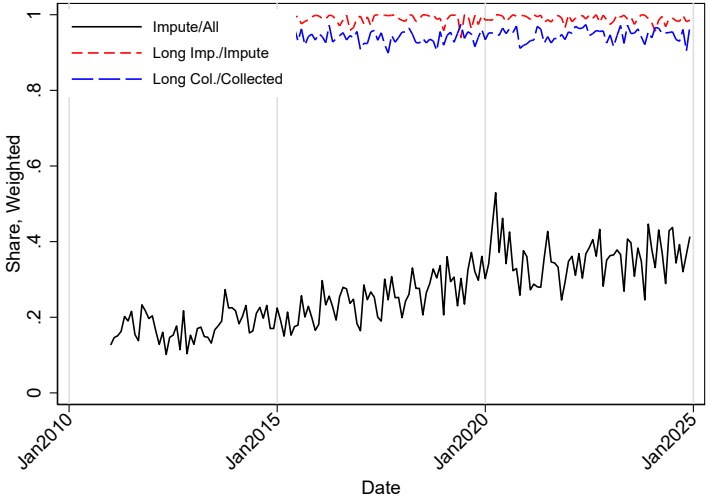
Q: Why does choice of variable for class means imputation have a small effect on the national index?

- Similar results if conditioning on rent cell, structure type, or tenancy length.
- Even unconditional imputations similar.
- However, non-response rates and inflation rates differ by covariate levels.
- Data is not missing at random.

Example: New York City (NYC) and imputation by tenancy length

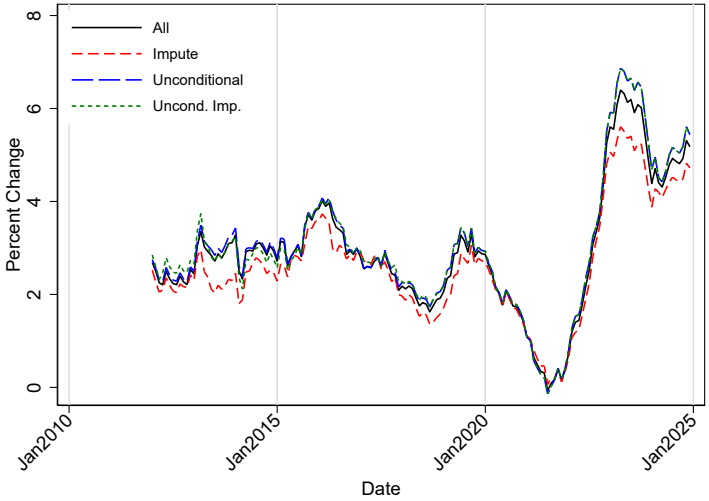
- Long tenancy ($>$ 6 months) rent changes small
- New tenant ($<=$ 6 months) rents change large

NYC Observation Shares



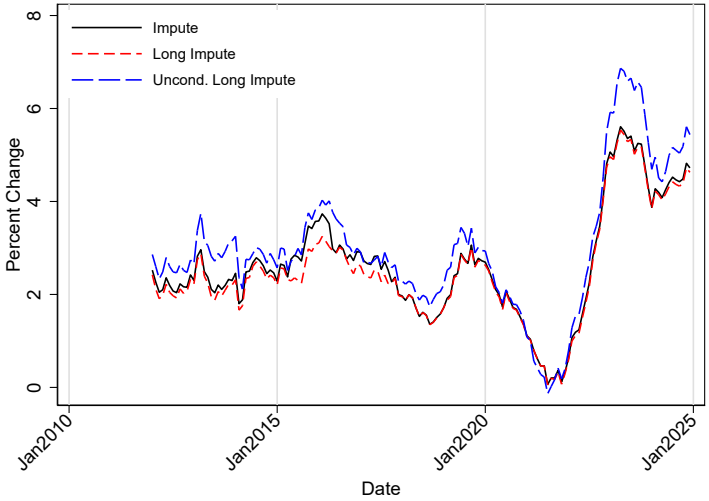
NYC Rent Inflation Rate, All vs. Imputes

Using Imputation By Tenancy Length



NYC Rent Imputes Inflation Rate, All vs. Long

Using Imputation By Tenancy Length



Bootstrap Setup

We estimate a confidence interval for the year-on-year change using a bootstrap approach.

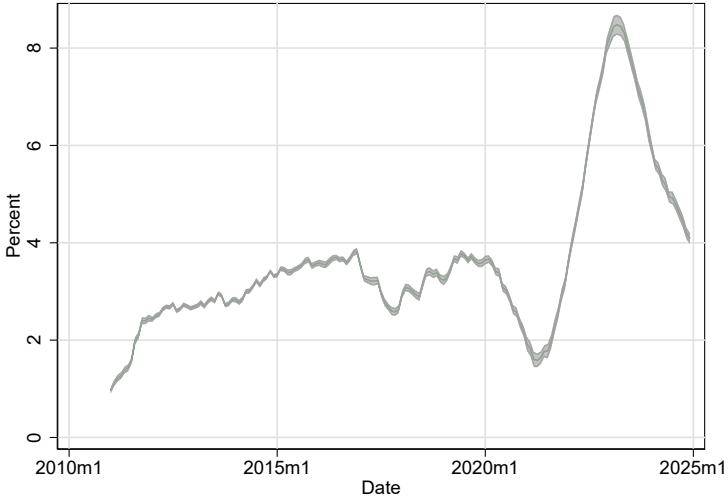
- Goal: Evaluate precision of rent level non-response imputation method
- $n = 30$ resamples

For each resample

1. Resample the collected observations with replacement for each city-date
2. Sort housing units in resample by lagged rent, divide into low, medium and high cells
3. Calculate average rent change for each rent level
4. Impute for non-response observations

Can apply same approach to structure and tenancy imputations

Rent Level Imputation Bootstrap



Jackknife Setup

Q: How accurate are the non-response imputations?

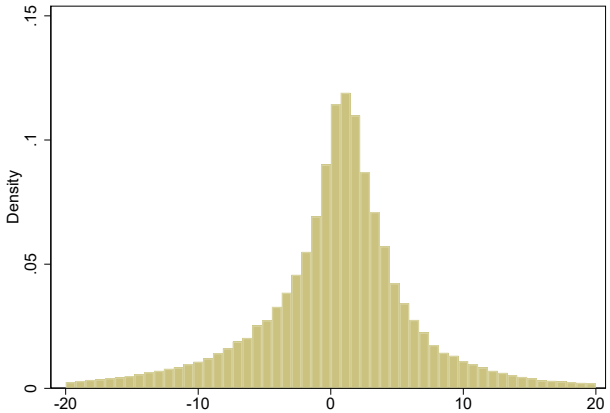
- Difficult to answer since we don't observe them
- Solution: jackknife approach

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 $x^{I}_{i,c,t}$

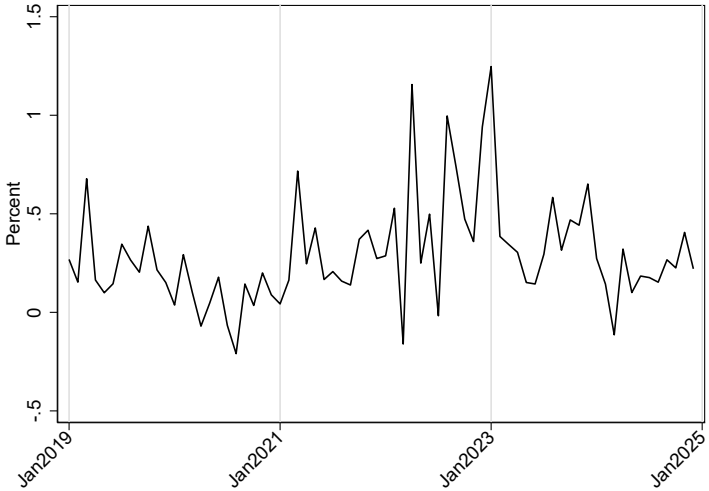
Then evaluate accuracy of $\{x^{I}_{i,c,t}\}$

Rent Level Imputation Jackknife Histogram



Mean = 0.40%, Median = 0.82%, Std. = 5.89%.

Rent Level Imputation Mean Jackknife Error



Imputation Inflation

Q: Is the inflation rate for all (vacancy and non-response) imputations similar to the inflation rate for collected observations?

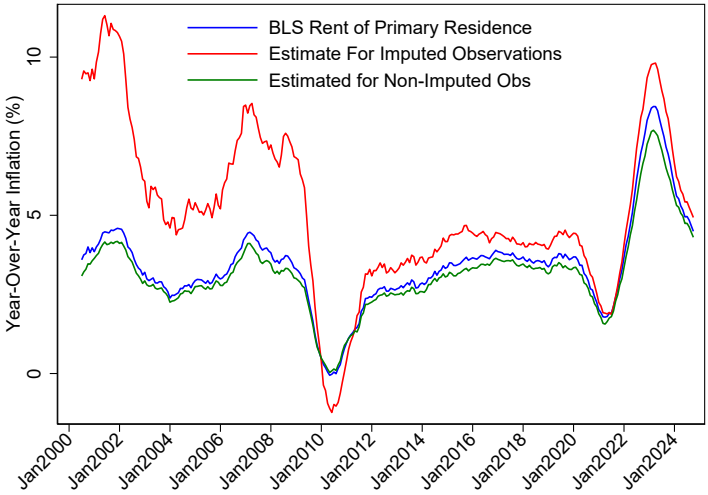
Calculate three inflation rates

- All observations
- Observations with imputed rent at time t
- Observations with collected rent at time t

Let $T_{c,t}$ denote set of imputed observations for city c and time t.

$$\pi_{c,t}^{\text{imputed}} = \frac{\sum_{i \in I_{c,t}} \omega_{i,c,t} r_{i,c,t}}{\sum_{i \in I_{c,t}} \omega_{i,c,t-6} r_{i,c,t-6}}$$

Inflation Rates, by Imputation



Conclusion

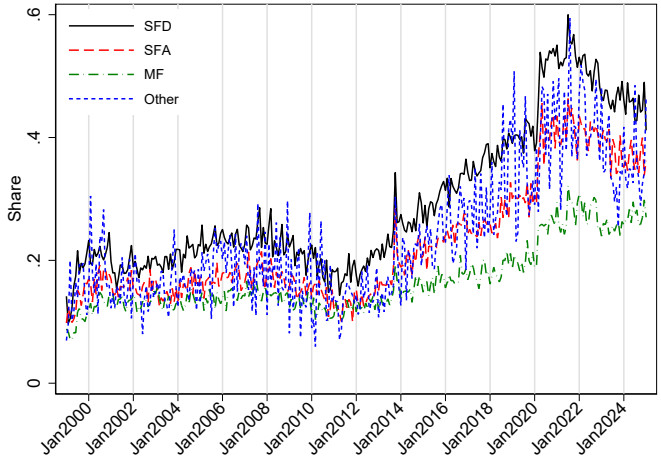
- Imputations rates have been increasing over time
- Imputation rates and inflation rates vary by numerous observables, such as tenancy length
  - Aggregate rent index not significantly affected by which observables are used in non-response imputation
- Nonresponse imputations are biased upwards to some extent, driven by using too high a proportion of short tenancy observations

Future Work

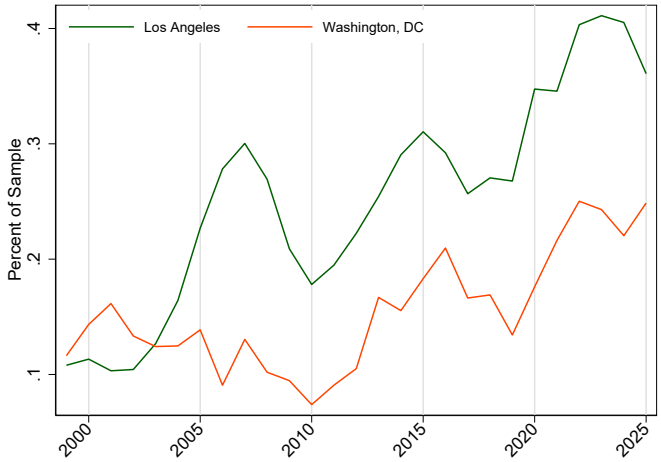
- Increase number of bootstraps
- Confirm results for sub-national geographies
- Investigate vacancy imputation process in similar detail
- Owners' equivalent rent imputation by structure type

Thank you

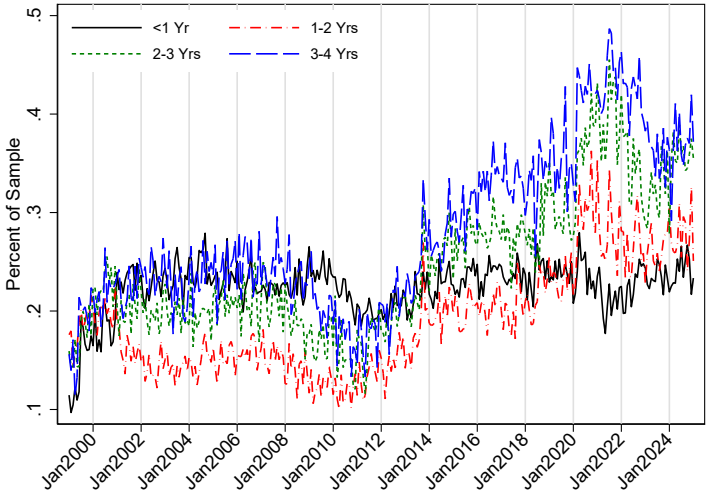
Non-Response Imputation Rate by Structure



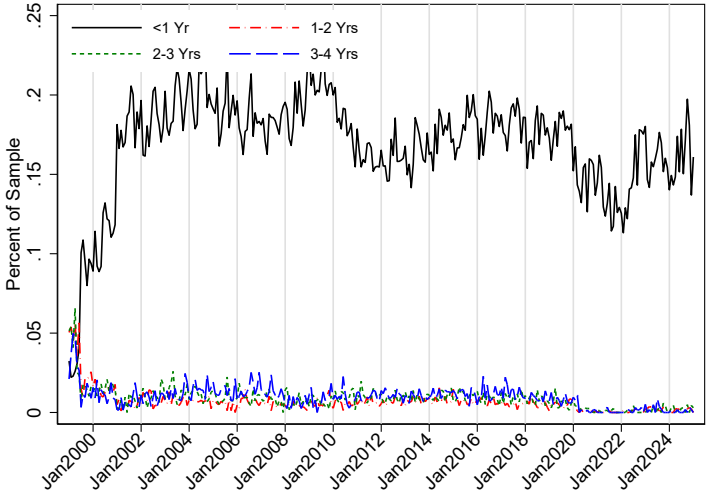
Non-Response Imputation Rate by Geography



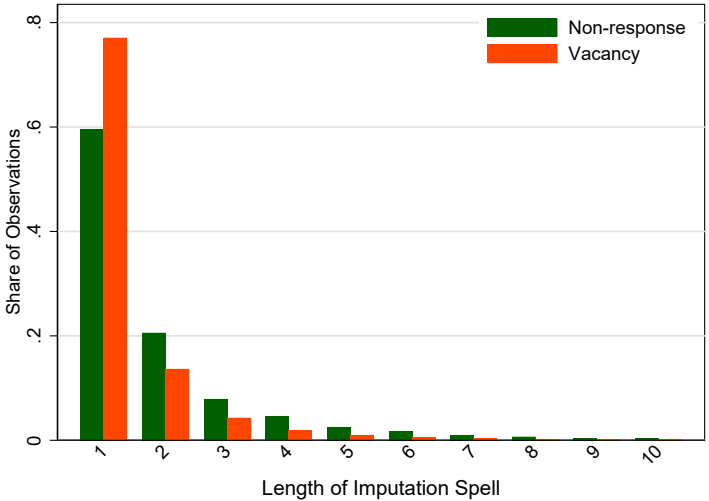
Imputation Rate by Tenancy Length



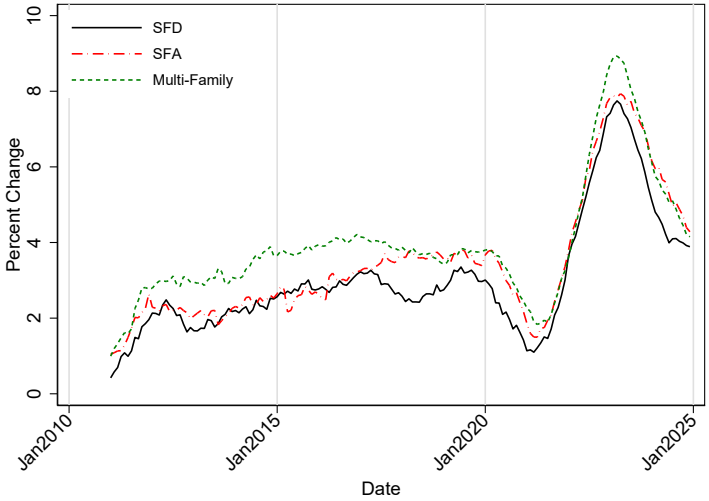
Vacancy Rate by Tenancy Length Length



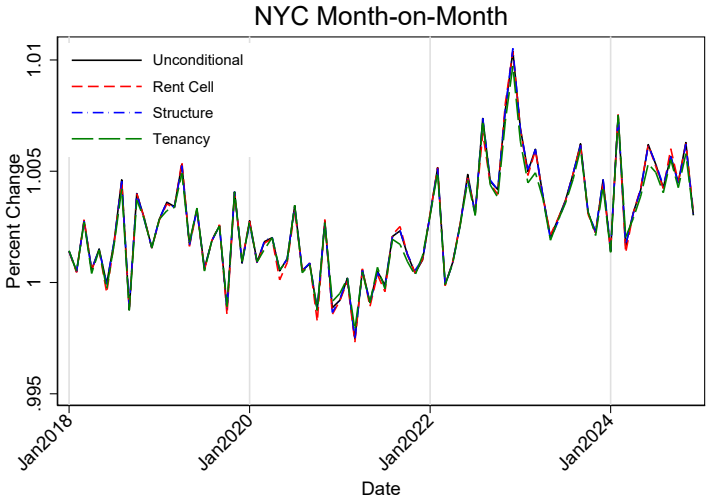
Length of Imputation Spell



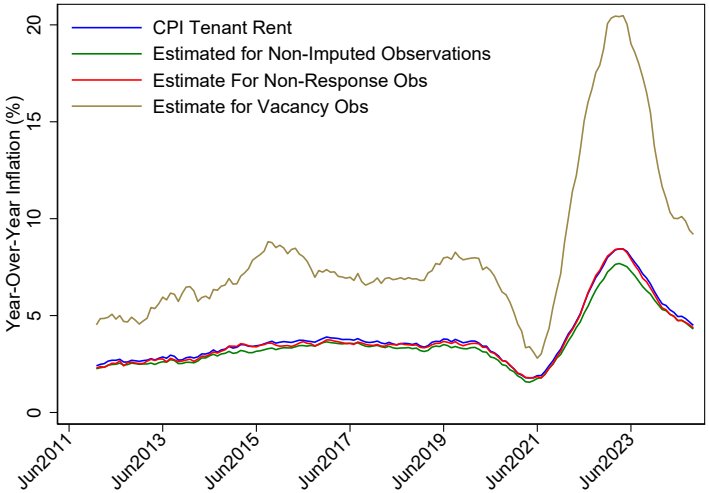
Inflation Rate by Structure Type



Inflation Rate by Imputation Type, NYC



Inflation Rates, by Imputation



Reentered Imputation Evaluation

Q: How accurate are the non-response imputations?

- Difficult to answer since we don't observe them
- Solution: Examine non-responses that start responding again

Example:

- Housing unit responds in February 2021
- Does not respond August 2021-August 2023. Rents are imputed
- Responds again in February 2024
- Does the imputed February 2024 rent match what we observe?

Reentered Imputation Histogram