Federal Reserve Bank of Cleveland Working Paper Series

Improving the Median CPI: Maximal Disaggregation Isn't Necessarily Optimal

Christian Garciga, Randal Verbrugge, and
Saeed Zaman

Working Paper No. 24-02R

April 2025

Suggested citation: Garciga, Christian, Randal Verbrugge, and Saeed Zaman. 2025. "Improving the Median CPI: Maximal Disaggregation Isn't Necessarily Optimal." Working Paper No. 24-02R. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202402r.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

Improving the Median CPI: Maximal Disaggregation Isn't Necessarily Optimal*

Christian Garciga† Randal Verbrugge‡ Saeed Zaman§

First version: January 2024
This version: March 12, 2025

Abstract

For decades, the Federal Reserve Bank of Cleveland (FRBC) has produced the Median Consumer Price Index (CPI). It has proven useful in various contexts, such as forecasting and understanding post-COVID inflation dynamics. Historically, revisions/improvements to the FRBC methodology have involved increasing the level of disaggregation in the CPI components. Thus, it may be reasonable to assume that further disaggregation improves the properties of the median CPI. We theoretically demonstrate: not necessarily. We then empirically explore the impact of further disaggregation by examining fifteen candidate baskets of CPI items that vary by the level of disaggregation. In line with prior literature, we find that greater disaggregation in the shelter indexes improves the ability of the Median CPI to track the medium-term trend in CPI inflation and its predictive power over future CPI movements. In contrast, increasing disaggregation in the remaining components leads to a deterioration in performance. Our preferred Median CPI measure suggests lower trend inflation pre-pandemic, a faster acceleration in trend inflation in 2021, and a faster deceleration in trend inflation after 2022.

Keywords: inflation measurement, median CPI, trend inflation, disaggregates of inflation

JEL Codes: E31, E37, E52, C8

*This paper was previously circulated with the title "The Effect of Component Disaggregation on Measures of the Median and Trimmed-Mean CPI." We thank Larry Ball, Jorge Fornero, Edward Knotek, Michael McCracken, James Mitchell, Tony Richards, and participants at the American Economic Association-ASSA Annual Meeting 2025 and the Cleveland Fed Brown Bag for helpful conversations and comments. The views expressed herein are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

†Federal Reserve Bank of Cleveland, 1455 E. 6th St., Cleveland, OH 44114.
Email: christian.garciga@clev.frb.org

‡Federal Reserve Bank of Cleveland, 1455 E. 6th St., Cleveland, OH 44114.
Email: randal.verbrugge@clev.frb.org

§Federal Reserve Bank of Cleveland, 1455 E. 6th St., Cleveland, OH 44114.
Email: saeed.zaman@clev.frb.org

Funding: We received no outside funding for this project.
There are no potential sources of conflict of interest.

1 Introduction

Measuring the medium-term trend (MTT) in inflation is important for several reasons. First, it is a matter of central importance for monetary policy. Second, because the real interest rate matters for intertemporal decisions, estimating the MTT in inflation is useful in practically any real-world situation where intertemporal tradeoffs are involved. Third, it is useful in removing noise from the inflation series, noise that otherwise might obscure relationships and cloud empirical work involving inflation.[1]

While there are many approaches to estimating the MTT in inflation, a simple method that has stood the test of time is using a limited-influence statistic—such as a median—to estimate the center of the current cross-sectional distribution of CPI or PCE component price changes. Limited-influence statistics are useful for estimating the center of the inflation distribution because the sample distribution of CPI component price changes is highly leptokurtic, i.e., it frequently has big outliers. Official measures of inflation are computed as a sample mean across component price changes, but sample means are very sensitive to outliers. Thus, limited-influence statistics are more efficient estimators of the center of the distribution, and in practice, those statistics have been shown to be good estimates of the MTT in inflation.

Bryan and Pike (1991) and Bryan and Cecchetti (1994) introduced the Median CPI and Trimmed-Mean CPI, and shortly thereafter the Federal Reserve Bank of Cleveland (FRBC) began producing these series. These two series are generally reliable signals of the MTT in CPI inflation.[2] For this reason, these and other similar measures are increasingly used in empirical studies, not only in the forecasting context (e.g., Smith, 2004; Meyer, Venkatu and Zaman, 2013; Liu and Smith, 2014; Meyer and Zaman 2019; Verbrugge and Zaman, 2023,2024; Ocampo, Schoenle and Smith 2023; Richards,

[1]Mazumder (2017), Kishor and Koenig (2022), and de Veirman (2023) also make this point, and it is implicit in Andrle, Bruha and Solmaz (2017). The reduction of noise helps explain why Phillips curves estimated with, e.g., trimmed-mean inflation, are much more successful. To our knowledge, Alves (2014) is the only extant study using trimmed inflation in a DSGE model.

[2]In like manner, the trimmed-mean PCE has been shown to be a very good signal of the MTT in PCE inflation; see Mertens (2016). Limited-influence measures also greatly reduce noise at high frequencies (monthly or quarterly) and allow a more accurate assessment of the latest data, a point emphasized by Richards (2024). There are more sophisticated approaches to deducing the MTT. For instance, Goulet Coulombe et al. (2024) propose a machine learning approach to time-varying weighting and trimming, specifically calibrated for optimal forecasting performance. Carlomagno et al. (2023) propose an approach to selecting a particular trimmed mean (or fixed exclusion measure) based on a sophisticated loss function. However, simple approaches—such as the Median CPI—have other advantages, such as in communicating with the public.

2024) but in many other contexts as well, including: establishing and testing the robustness of stylized inflation facts (Bryan and Cecchetti, 1999; Verbrugge, 1999; Fang, Miller and Yeh, 2010); understanding inflation uncertainty (Metiu and Prieto, 2023); discriminating between models of price adjustment (Ashley and Ye, 2012); locating a stable Phillips curve (Ball and Mazumder, 2011; Ball and Mazumder, 2019a,b; Stock and Watson, 2020; Ashley and Verbrugge, 2025); studying the effects of oil supply shocks (Kilian, 2008); understanding inflation expectations and their relationship to inflation (Verbrugge and Zaman, 2021); and understanding post-Great Recession and post-COVID inflation dynamics (Ball and Mazumder, 2011; Mazumder, 2018; Ball et al., 2021; Ball, Leigh, and Mishra 2022; Verbrugge and Zaman 2023; Cotton et al., 2023). They have also been suggested as a superior measure for monetary policy targeting and communication (Cecchetti and Groshen, 2001; Dolmas and Koenig, 2019; Verbrugge, 2022).[3]

Central banks around the world make significant use of such measures in their monetary policy decision-making. The European Central Bank formally uses a median inflation measure as one of its estimates of "underlying inflation" (O'Brien and Nickel, 2018; Nickel et al., 2021), as does the Swedish Riksbank (Johansson et al., 2018). In Australia, "the most important indicators of underlying inflation are the trimmed mean and the weighted median" (Reserve Bank of Australia, 2025). In New Zealand, the trimmed mean and the weighted median are two of the five measures of core inflation (Reserve Bank of New Zealand, 2024); in Norway, the trimmed mean and median are two of four measures of core inflation (Norges Bank, 2025). Since 2017, the Bank of Canada has used a median and a trimmed-mean inflation measure as two of its three preferred core inflation measures and "main operational guides for monetary policy" (Lao and Steyn, 2019; see also the remarks of Deputy Governor Schembri (2017), who discussed the rationale for abandoning "core" CPI.). The Swiss National Bank calculates core inflation using a trimmed mean (Swiss National Bank, 2025). The Central Bank of Costa Rica uses a trimmed mean as its indicator of core inflation (Esquivel-Monge et al., 2011; Munoz-Salas and Rodriguez-Vargas, 2021).

Given the growing importance of the Median CPI, it is worth exploring whether it might be enhanced.[4] One potential approach lies in increasing the level of disaggregation, i.e., the number of

[3] Verbrugge (2022), in particular, highlights the severe theoretical and empirical deficiencies of ex-food-and-energy, or "core," inflation measures.

[4] While we provide a parallel analysis pertaining to the Trimmed-Mean CPI in the Appendix, we focus on the Median CPI in the main body. The degree of disaggregation that is best for the Median CPI need not coincide with

CPI components used to calculate it. Such changes have precedent. The earliest precursor to today's Median CPI had just seven components. When formally introduced, it had 36 components, which was later increased to 41. Subsequent research demonstrated that it would be desirable to increase the number of components to 45, where it stands today. But is this level of disaggregation *optimal*? Notably, trimmed-mean *PCE* inflation and median *PCE* inflation are calculated using significantly larger baskets of 178 and 201 PCE components, respectively (Dolmas 2009; Carroll and Verbrugge 2019). This paper investigates whether further disaggregation would enhance the Median CPI.

Our first result, which is a theoretical result, is surprising. We demonstrate that, when estimating the weighted median of the underlying distribution of CPI indexes, if some of those underlying indexes are unobserved because they are unpublished—as is the case with the CPI—the weighted median of the most highly disaggregated basket that is *feasible* (based upon indexes that are observed) is not necessarily the minimum mean squared error (MSE) estimator. Hence, the optimal level of disaggregation becomes an empirical question.

We therefore conduct such an empirical investigation, systematically investigating the impact of the level of disaggregation, taking the current FRBC Median CPI basket as our starting point. We disaggregate the CPI basket along two dimensions: shelter and non-shelter components. For shelter components, our starting point is the current FRBC methodology, which splits Owners' Equivalent Rent (OER) into four regional components. We then further divide each regional OER component into two components based on city size-classes. For Rent, we apply the same disaggregation, resulting in a total of eight OER indexes and eight Rent indexes at the highest level of shelter disaggregation. For non-shelter components, we use the current FRBC non-shelter component basket as our baseline and then construct four additional non-shelter component baskets using the disaggregation structure of the CPI. This results in a total of 15 candidate baskets of CPI components, ranging in size from the current FRBC basket of 45 components to a basket of 156 components. Next, we construct the Median CPI index corresponding to each basket. We then evaluate the performance of each new median measure using well-established criteria from the core inflation literature. Two in particular are worth highlighting. First, we assess how disaggregation affects the accuracy of median inflation in tracking an ex-post estimate of the MTT in CPI inflation. Second, we examine the extent to

the degree of disaggregation that is best for the Trimmed-Mean CPI.

which disaggregation impacts the predictive power of each measure over future movements in CPI inflation.

We find disaggregating both OER and Rent into eight components, in combination with only little-to-no-more non-shelter disaggregation relative to the FRBC basket, to be the optimal mix of disaggregation in terms of tracking the ex-post MTT in CPI inflation and maximizing in-sample predictability over future movements in CPI inflation. Notably, further disaggregation of non-shelter components generally results in clear performance losses. Thus, our findings are in keeping with those of our theoretical result, and are somewhat counterintuitive: increasing the level of disaggregation as far as possible is not optimal.

Our work extends a long thread of the literature that has derived or extended measures of trend CPI and PCE inflation using limited-influence estimators (e.g. Bryan and Pike 1991; Bryan and Cecchetti 1994; Bryan et al. 1997; Dolmas 2005; Brischetto and Richards 2007; Dolmas 2009; Higgins and Verbrugge 2015; Carroll and Verbrugge 2019; Rich et al. 2022). Our work also contributes to a large literature that has evaluated competing measures of the MTT in inflation in terms of metrics such as forecasting, explaining future headline inflation, and tracking ex-post measures of trend inflation (e.g., Clark 2001; Marques et al. 2003; Rich and Steindel 2007; Meyer and Pasaogullari 2010; Crone et al. 2013; Higgins and Verbrugge 2015; Gamber and Smith, 2019; Verbrugge and Zaman 2023, 2024). We also contribute to the literature examining how the properties of derived measures of inflation change as one changes the level of disaggregation of the components underlying the headline price indices (e.g., Mahedy and Shapiro 2017; Zaman 2019; Stock and Watson 2020). However, in contrast to prior work, we are the first to derive a theoretical result regarding the optimal level of disaggregation for a weighted-median estimator. Further, we are the first to propose disaggregating OER into 8 components, and disaggregating Rent at all, in the calculation of limited-influence estimators of MTT inflation. We are also the first to investigate the performance of MTT inflation measures by level of disaggregation in a systematic manner. Finally, we are also the first to demonstrate that there is no simple relationship between the frequency with which shelter components are selected as the median and the strength of the Phillips curve relationship in the resulting median CPI measure.

The rest of the paper is organized as follows. In Section 2, we provide a brief history of the FRBC

Median and 16% Trimmed-Mean CPI inflation measures, along with an overview of the methodology behind their calculation. In Section 3, we prove that the optimal level of disaggregation need not be the highest level of disaggregation possible. In Section 4, we propose a scheme for defining several collections of CPI components at increasingly finer levels of disaggregation of shelter and non-shelter components than the set of 45 components currently used to compute the FRBC measures. Section 5 examines how varying the level of CPI component disaggregation affects the ability of median CPI inflation measures to track an ex-post estimate of the MTT in CPI inflation and to explain future inflation. In Section 6, we discuss some practical implications of our preferred degree of disaggregation for the Median CPI. First, we look at how changing the degree of component disaggregation alters the distribution of components chosen as the median component over time. Next, we investigate the extent to which disaggregation influences the Phillips curve relationship. Finally, we examine the implications for trend CPI inflation: our preferred Median CPI measure suggests lower trend inflation pre-pandemic, a faster acceleration in trend inflation in 2021, and a faster deceleration in trend inflation after 2022. Section 7 concludes.

2 The FRBC Median and Trimmed-Mean CPI

2.1 A Brief History

The FRBC Median and 16% Trimmed-Mean CPI inflation originated from the seminal work of Bryan and Pike (1991) and Bryan and Cecchetti (1994), who were the first to propose a theoretical and statistical justification for the use of the median or trimmed mean as measures of "core" (or trend) inflation (Dolmas and Wynne 2008).[5] While the FRBC has published these limited-influence estimators of the MTT in CPI inflation for decades, the components of CPI inflation from which these measures are calculated have evolved over time.[6]

Prior to 1998, the FRBC calculated the Median and Trimmed-Mean CPI using 36 CPI components. In 1998, the Bureau of Labor Statistics (BLS) carried out its sixth comprehensive revision of

[5] The earliest precursor to today's Median CPI in Bryan and Pike (1991) was derived from just seven CPI components.

[6] The FRBC updates the Median and Trimmed-Mean CPI each month immediately following a new CPI data release by the Bureau of Labor Statistics (BLS) and makes these data available at https://www.clevelandfed.org/en/indicators-and-data/median-cpi.

the CPI, leading the FRBC to modify its component basket, for a revised total of 41 components.[7] Importantly, prior to 1998, Median and Trimmed-Mean CPI used the Shelter component, whereas after 1998, Shelter was split into: Rent of primary residence (Rent); Lodging away from home; Owners' Equivalent Rent of primary residence (OER); and Tenants' and household insurance.[8]

In July 2007, the FRBC again revised the Median and Trimmed-Mean CPI. Under this new "Revised Methodology," OER was split into four regional OER subindices, one each for the Northeast, Midwest, South, and West. This change was prompted by research by Brischetto and Richards (2007)–later confirmed by the FRBC (2007)–that found that breaking up OER improved the ability of trimmed-mean measures to track the trend in CPI inflation. Concurrently, the FRBC added the component Leased Cars and Trucks, bringing the total to 45 CPI components.[9]

Since its introduction in 2007, small methodological adjustments have since been made to the "Revised Methodology" Median and Trimmed-Mean CPI to ensure that it reflects the most recent statistical techniques and data availability. For example, the BLS does not seasonally adjust the four regional OER subindices despite the presence of seasonality in each (FRBC 2007). Since the FRBC Median and Trimmed-Mean CPI indices use seasonally adjusted (SA) data (see Higgins and Verbrugge 2015 for a discussion), the FRBC seasonally adjusts the regional OER series. Whereas the FRBC originally used the Census Bureau's X-12-ARIMA procedure to do this, it has since switched to the newer X-13-ARIMA-SEATS procedure.

In the next section, we review in detail the current methodology for constructing the FRBC Median and Trimmed-Mean CPI. We begin by explaining the procedure for calculating the monthly expenditure weights, and seasonally adjusting CPI indices, for the four regional components of OER. We then enumerate the steps we take to calculate the median and trimmed mean from any collection of CPI components.

[7] For more on this and other revisions, see https://www.bls.gov/cpi/additional-resources/historical-changes.htm

[8] We refer to measures calculated from either set of components as the "Old Methodology" Median and Trimmed-Mean CPI. The "Old Methodology" data begin in 1967 through July 2007. See the Online Appendix for both the pre-1998 and post-1997 sets of components used under the "Old Methodology."

[9] Data for the "Revised Methodology" measures begin in 1983, as this is when the BLS introduced the rental equivalence method of measuring the cost of owner-occupied shelter. See the online appendix for a list of the components used under the "Revised Methodology."

2.2 Methodology

2.2.1 OER

CPI expenditure weights (in CPI parlance, "relative importances") are estimated annually and released in December, based upon direct measures of consumer expenditures. Over the course of the following year, they are updated every month to reflect relative price movements that have occurred. Essentially, these describe or approximate how expenditure weights change, based upon changes in relative prices.[10]

In order to split OER into regional components and incorporate them into median and trimmed-mean CPI calculations, each component is weighted by its appropriate share in the overall CPI-U. Unlike for most components, the BLS does not publish the relative importances of each regional OER component relative to overall CPI-U. The FRBC must therefore compute these itself. Computing these weights is a two-step process.[11]

First, we calculate the **annual** relative importances for each region x by multiplying: (1) the relative importance of region x relative to the overall CPI-U; and (2) the relative importance of OER in region x relative to the overall CPI for that same region x. Why? Suppose we know that the weight of x in X is x/X and that the weight of X in Y is X/Y. What is the weight of x in Y, i.e. x/Y? It is simply x/Y = (x/X)*(X/Y). For example, taking the West region for concreteness, we know the weight of West OER (x) in the West CPI (X) is x/X, and we also know that the weight of the West (X) in the national CPI-U (Y) is X/Y. So to get the weight of West OER in the national CPI, we multiply the weight of West OER in the West CPI by the weight of the West region in the national CPI.

Given these annual relative importances, each month we pull the monthly price indices for each regional OER component and follow the BLS methodology to compute the current **monthly** relative importances for each regional OER component based upon price movements that occurred

[10]The updating procedure makes no attempt to approximate substitution behavior across components. Such substitution is picked up annually – with a very long lag – when the December relative importances are recomputed based upon Consumer Expenditure Survey data.

[11]When the Median CPI was introduced, the BLS did not publish monthly relative importances, nor did it provide documentation on the procedure it used for updating monthly weights. The FRBC developed a method for constructing monthly weights. The procedure used in this paper, which we outline below, is consistent with current BLS practice but differs slightly from the FRBC method. Our results demonstrate that the difference in the weighting formula has little impact.

that month. Let us consider a concrete example. Suppose we are in March and are given: (1) the values of the non-seasonally-adjusted (NSA) price index of component x for December (I^x_{Dec}) through March (I^x_{Mar}); (2) the values of the same for the headline CPI-U (I^{CPI}_{Dec} through I^{CPI}_{Mar}); and (3) the annual (December) relative importance of x, R^x_{Dec}. We wish to compute R^x_{Mar}. The current BLS method to construct the non-normalized weight R^x_{Mar} is given by:

$$R^x_{Mar} = R^x_{Dec} * \left(\frac{I^x_{Mar}}{I^x_{Dec}}\right)$$

To construct the normalized weight, one adjusts all the relative weights so as to ensure that they all add up to 100 by simply dividing the non-normalized weight by the analogous "updated relative importance" for the entire CPI – which has an initial "relative importance" of 100 in December – which is given by:

$$R^{CPI}_{Mar} = 100 * \left(\frac{I^{CPI}_{Mar}}{I^{CPI}_{Dec}}\right)$$

Hence the normalized weight for x, Φ^x_{Mar}, equals:

$$\Phi^x_{Mar} = \frac{R^x_{Mar}}{R^{CPI}_{Mar}}$$

One can also rewrite this as a recursive formula. Clearly:

$$R^x_{Mar} = R^x_{Dec} * \left(\frac{I^x_{Mar}}{I^x_{Dec}}\right) = R^x_{Dec} * \left(\frac{I^x_{Feb}}{I^x_{Dec}}\right)\left(\frac{I^x_{Mar}}{I^x_{Feb}}\right) = R^x_{Feb}\left(\frac{I^x_{Mar}}{I^x_{Feb}}\right)$$

Similarly:

$$R^{CPI}_{Mar} = R^{CPI}_{Feb}\left(\frac{I^{CPI}_{Mar}}{I^{CPI}_{Feb}}\right)$$

This implies:

$$\Phi^x_{Mar} = \frac{R^x_{Mar}}{R^{CPI}_{Mar}} = \frac{R^x_{Feb}}{R^{CPI}_{Feb}} * \frac{(I^x_{Mar}/I^x_{Feb})}{\left(I^{CPI}_{Mar}/I^{CPI}_{Feb}\right)} = \Phi^x_{Feb}\frac{(I^x_{Mar}/I^x_{Feb})}{\left(I^{CPI}_{Mar}/I^{CPI}_{Feb}\right)}$$

With this methodology, we compute for each month t the weights Φ^c_t for each component c in

Northeast OER, Midwest OER, South OER, and West OER.

This leaves the monthly CPI indices for each component. The BLS produces both SA and NSA versions of the headline CPI-U index and most components.[12] However, the BLS does not publish SA price indices for the four regional OER indices despite the presence of seasonality in each series. As a result, FRBC seasonally adjusts these series using the BLS methodology described above.

2.2.2 Calculating Median and Trimmed-Mean CPI Inflation

For a given collection of CPI components C, denote as

$$\pi_t^c = 100\big(\frac{I_t^c}{I_{t-1}^c} - 1\big)$$

the monthly inflation rate of component c in month t. For a regional OER component, I_t^c is the corresponding NSA CPI index that has been seasonally adjusted, as explained in the previous section. For other components, I_t^c is the SA index published by the BLS, if available. If a given index I_t^c is only available from the BLS in NSA form, then since that component does not display significant seasonality, we use the NSA index for that component. For each component c, in addition to π_t^c, we have available Φ_t^c, calculated as explained in the previous section for each regional OER component and taken from the BLS otherwise. To calculate the median and trimmed-mean in month t:

1. For each $c \in C$, if either π_t^c or Φ_t^c is missing, component c is dropped from any further calculations. Denote as $\tilde{C}$ the set of components C excluding components with missing data.[13]

2. Renormalize the weights such that for each $c \in \tilde{C}$: $\tilde{\Phi}_t^c = 100(\Phi_t^c / \sum_{c \in \tilde{C}} \Phi_t^c)$.

3. For all $c \in \tilde{C}$, sort π_t^c from smallest to largest. More formally, define a one-to-one mapping $c \leftrightarrow j$, $j = 1, ..., J$, where j denotes the relative position of π_t^c. For example, $c \leftrightarrow j = 1$ if π_t^c is the smallest monthly inflation rate, and $c \leftrightarrow j = J$ if π_t^c is the largest.

[12] As the BLS explains: "Seasonally adjusted data are computed using seasonal factors derived by the X-13ARIMA-SEATS Seasonal Adjustment Method. These factors are updated with the release of January data in February and reflect price movements from the previous calendar year. The new factors are used to revise the previous 5 years of seasonally adjusted data; older seasonally adjusted indexes are considered to be final." For more information on seasonal adjustment in the CPI, see https://www.bls.gov/cpi/seasonal-adjustment/.

[13] Missing data are rare, but can happen if BLS has insufficient source data to publish the component.

4. For each j, compute cumulative weight $w(j) = \sum_{k=1}^{j} \tilde{\Phi}_t^j$ where $\tilde{\Phi}_t^j \equiv \tilde{\Phi}_t^c$ if and only if $c \leftrightarrow j$.

5. Find the first j for which $50 \leq w(j)$. Denoting this index as j^{MED}, the median component is the component c^{MED} satisfying $c^{MED} \leftrightarrow j^{MED}$, and the median inflation rate is $\pi_t^{c^{MED}}$.

6. To calculate the 16% trimmed mean:

   (a) Find the first j for which $8 < w(j)$. Denote this index as j_S and set its normalized relative importance to $\tilde{\Phi}_t^{j_S} \equiv \tilde{\Phi}_t^j - 8$.

   (b) Find the first j for which $92 \leq w(j)$. Denote this index as j_E and set its normalized relative importance to $\tilde{\Phi}_t^{j_E} \equiv \tilde{\Phi}_t^j - \tilde{\Phi}_t^{j-1}$.

   (c) Calculate the trimmed mean:

$$\pi_t^{TM} = \frac{\sum_{j \in [j_S, j_E]} \pi_t^j \tilde{\Phi}_t^j}{\sum_{j \in [j_S, j_E]} \tilde{\Phi}_t^j} = \frac{\sum_{j \in [j_S, j_E]} \pi_t^j \tilde{\Phi}_t^j}{84}$$

3 More Disaggregation Need Not Be Better: A Theoretical Result

Given the historical record of improvements to the Median CPI, and given the much higher level of disaggreation in the Median PCE (for example) compared to the Median CPI, it may seem obvious that increasing the level of disaggregation in the Median CPI even further will lead to improvement. Moreover, even laying aside the empirical performance gains that have historically accompanied increased disaggregation, there is a more theoretical argument that seems to support the notion that "more disaggregation must be better." At the lowest level of aggregation —namely, no disaggregation at all —the median equals the mean., i.e., equals headline CPI. Next, consider a very low level of disaggregation, splitting the CPI into just core goods, core services, and energy. Most of the time, core services is likely to be chosen as the median. It is clear that the resulting median will be highly volatile, and also clear that very little has been gained over just using headline CPI. As the level of disaggregation increases, it seems obvious that the distinction between the mean and the median would sharpen, and that the median would become a closer and closer approximation to the "true median"—namely, the median of the full scope of underlying indexes that the statistical agency generates (even if it does not publish all of those indexes, owing to concerns such as inadequate sample sizes or confidentiality). However, it turns out that this intuition is incorrect. In this section, we prove that if researchers do not have access to the full scope of underlying indexes that the statistical agency generates in order to produce a weighted average statistic like the CPI, and wish to compute the weighted median of the sample distribution, it is not always better to use the most disaggregated data that are available. Thus, the optimal level of disaggregation in any particular case is an empirical issue: one must assess how the weighted average computed based upon various levels of disaggregation performs along a number of dimensions, such as accuracy against ex-post estimates of the MTT.

We begin with some definitions. Consider a discrete collection of N random variables: $A = \{X_j : j = 1, ...N\}$, each with an associated non-negative weight $w_j : j \in A$ with $\sum_{j=1}^{N} w_j = 1$. Denote a member of the set A by V. We define the weighted sample median as follows. After the random variables are realized, sort the random variables from smallest to largest, indexed by k, so that v_k is the k^{th} largest realization. The cumulative sum weight through index l is defined by $\sum_{k=1}^{l} w_k$. Then the weighted median of the sample of random variables A is defined as follows: $wmed\,(A) = v_l : \sum_{k=1}^{l} w_k \leq 0.5$ and $\sum_{k=1}^{l+1} w_k > 0.5$.

Proposition 1. *Suppose that there is a collection of N random variables: $B = \{X_j : j = 1, ...N\}$, each with an associated non-negative weight $w_j : j \in B$ with $\sum_{j=1}^{N} w_j = 1$. Suppose that there exists a set $C \subset B$, of cardinality r, whose elements are unobserved; instead, what is observed is their weighted mean $Y = \sum_{j \in C} w_j X_j$. Without loss of generality, assume that the indexes of the random variables in C are $\{M - r, M - r + 1, ..., M\}$. Moreover, there exists a second set $D \subset B$, of cardinality s, with $C \cap D = \emptyset$, with a weighted mean $Z = \sum_{j \in D} w_j X_j$. Without loss of generality, assume that the indexes of the random variables in C are $\{M - r - s, M - r - s + 1, ..., M - s - 1\}$.*

Let

$$G = \{Y, X_j : j = 1, ..., M - r - 1\}$$

and let

$$H = \{Y, Z, X_j : j = 1, ..., M - r - s\}$$

Then the following inequality need not hold:

$$E\left[wmed\left(G\right)-wmed\left(B\right)\right]^2\leq E\left[wmed\left(H\right)-wmed\left(B\right)\right]^2$$

Proof. We prove this via a counterexample. We consider a collection of 7 random variables $B = \{X_i, i = 1, ..., 7\}$, each equally weighted. For simplicitly, one of these, X_4, has the dirac delta function (centered on 0) as its distribution. The distributions of all the random variables are as follows:

$$X_1 \sim U\left[-8,-7\right]$$
$$X_2 \sim U\left[-4.5,-3.5\right)$$
$$X_3 \sim U\left[-2,-1\right]$$
$$X_4 \sim \{\delta\left(t\right)=0, t\neq 0\}$$
$$X_5 \sim U\left[1,2\right]$$
$$X_6 \sim U\left(3.5,4.5\right]$$
$$X_7 \sim U\left[7,8\right]$$

Given these distributions, the weighted sample median $wmed(B)$ is always x_4, whose realization is always 0, and this coincides with the population weighted median. But suppose that X_1 and X_5 are unobserved; instead, only their weighted average Y is observed. Note that realizations of Y lie within $[-3.5, -2.5]$. The most disaggregated *observed* collection of random variables is the set $G = \{X_2, Y, X_3, X_4, X_6, X_7\}$. For any realization, this ordering also corresponds to the sorted list, with 3/7 of the weight below X_3 and 3/7 of the weight above x_3; thus, for any realization, $wmed(G) = x_3$, which is at most -1. In this example, the median of the most disaggregated *observed* collection of random variables will never equal the median of the underlying distribution.

However, consider aggregating X_3 and X_7 into a variable Z. Realizations of Z lie within $[2.5, 3.5]$ and the weighted median of the set $H = \{X_2, Y, X_4, Z, X_6\}$ is always x_4, and equal to the population weighted median, 0.

In this example,

$$E\left[wmed\left(G\right)-wmed\left(B\right)\right]^2 > E\left[wmed\left(H\right)-wmed\left(B\right)\right]^2$$

Hence, the most disaggregated set available, G, need not yield the most accurate weighted median estimate. $\square$

The theoretical result in this paper is reminiscent of some results in the factor estimation literature. For instance, Boivin and Ng (2006) show that increasing the number of underlying indexes used to estimate factors is not necessarily better. However, the concern in that part of the literature relates to estimating a modest number of common factors, not a weighted median. Typically, in the factor literature, the researcher is considering additional series for inclusion that are of varied types, and not part of a unified group of series constructed and released by a statistical agency under a common sampling design. And even when the discussion does relate to using aggregates versus more disaggregated series, aggregation in that literature is often seen as desirable in that it can remove idiosyncratic noise (e.g., Alvarez, Camacho, and Perez-Quiros, 2016). "Beneficial aggregation" may also be driving the results of Gamber and Smith (2019), who find that—when constructing estimators of the MTT in PCE inflation using a principal components approach—computing principal components using 17 PCE components yielded superior MTT estimates compared with using 50 components. However, in our context, the true weighted median is a function of all of that idiosyncratic noise.

4 Finer Disaggregations of the CPI

Having established theoretically that more disaggregation is not necessarily better, we turn to an empirical investigation. We first describe our approach to disaggregation. To systematically explore the effect of disaggregation of CPI components on Median CPI inflation, we derive several novel splits of the CPI by disaggregating it along two distinct dimensions: shelter components (OER and Rent), and non-shelter components. (As will be evident below, this two-dimensional treatment is necessitated by the aggregation structure in the CPI.) We begin by outlining our methodology for achieving increasingly finer splits of the non-shelter components of the CPI, and then detail our splits for the shelter components. Finally, we combine each non-shelter split with every shelter split, thus forming several new complete sets of CPI subcomponents.

4.1 Splitting Non-Shelter Components of the CPI

To split non-shelter CPI components, we begin with the eight major groups of the CPI: "Food and beverages," "Housing," "Apparel," "Transportation," "Medical care," "Recreation," "Education and communication," and "Other goods and services." Next, we break up each of these components to the next lowest level possible by following the CPI item aggregation tree.[14] For example, "Food and beverages" is split into two components, "Food" and "Alcoholic beverages," while "Housing" is split into three components, namely "Shelter," "Fuels and utilities," and "Household furnishings and operations." We then repeat this process until we cannot achieve a finer level of disaggregation on any CPI component. In particular, for a given CPI component, we do not split it further if either of the following conditions hold:

1. A component only has a single subcomponent. For example, "Personal care services" has only one subcomponent: "Haircuts and other personal care services."
2. Splitting the component would introduce an "unsampled" item for which the BLS does not publish a price index. For example, splitting "Information technology, hardware, and services" would introduce the component "Unsampled information and information processing." However, for consistency with the current FRBC Median CPI, we make one exception and split "New and used motor vehicles" even though this introduces the component "Unsampled new and used motor vehicles."

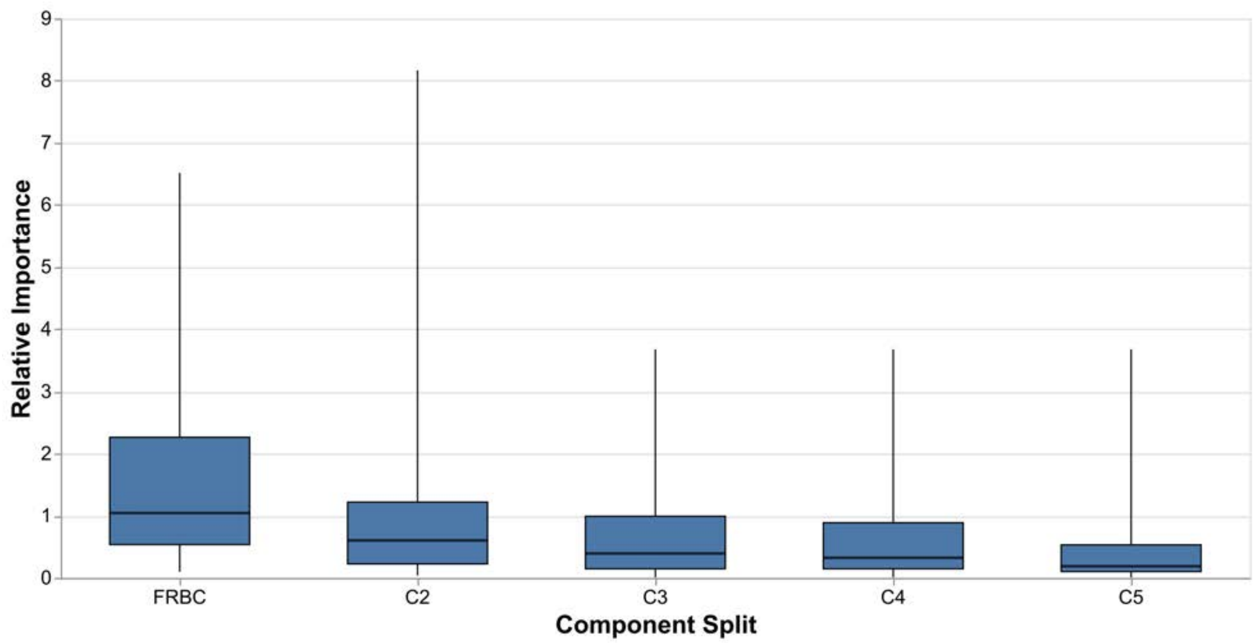
This yields six different collections of components, which we label C0, C1, C2, C3, C4, and C5, where ascending numbers indicate finer levels of disaggregation.[15] From the 8 components in C0, we have 25 in C1, 56 in C2, 90 in C3, 102 in C4, and 140 in C5. Importantly, the split of shelter into OER and Rent first appears in C2. Since the focus of this paper is on the effect of varying the degree of disaggregation in **both** shelter and non-shelter components, in the remainder of the paper, we drop collections C0 and C1 from further consideration.

[14]This tree is available from the BLS as a downloadable spreadsheet from https://www.bls.gov/cpi/additional-resources/cpi-item-aggregation.htm.

[15]In fact, this procedure yields eight different levels of disaggregation: C0,...,C7. However, C7 has just 16 more components than C5, all of which are relatively small components by weight within the Food at Home category. Therefore, we skip C5 and C6 and re-label C7 as C5.

As a baseline, we also add one more collection: FRBC, which consists of the components currently used to derive the official FRBC Median indicator. Overall, this is the least disaggregated of the the non-shelter component collections. This fact is evident in Figure 1, which displays boxplots of the set of relative importance levels associated with the (non-shelter) components in each collection. FRBC has the largest median, largest lower quartile, and largest upper quartile (though C2 has one component with a higher relative importance than FRBC).

Figure 1: Distribution of Component Weights in Each Collection



Notes: Component weights as published December 2023. For the purposes of this figure, each collection excludes the components OER and Rent.

Since the goods and services that are non-negligibly important in the consumer basket change over time, the CPI does not maintain a fully consistent item basket/disaggregation structure over time. Our approach to disaggregation allows us to collect these data with a non-interrupted disaggregation structure back to 2009M12. To go back further in time to 1997M12 requires adjusting for structural breaks in the components "Medical care commodities" and "Telephone services."[16],

[16]Prior to January 2010, "Medical care commodities" consisted of "Prescription drugs" and "Nonprescription drugs and medical supplies," the latter of which in turn consisted of "Internal and respiratory OTC drugs" and "Nonprescription medical equipment and supplies." From 2010 onward, "Medical care commodities" consists of "Medicinal Drugs" and "Medical equipment and supplies," the former of which in turn consists of "Prescription drugs" and "Nonprescription drugs."

[17] Due to this, the collection of subcomponents in Ci, $i = 2, 3, 4, 5$ differs slightly over 1997M12-2009M12 and from 2010M01 onward; additionally, the total number of subcomponents in C4 and C5 is higher by 1 over 1997M12-2009M12.[18]

Finally, for each component, we collect the monthly relative importances as published by the BLS, the SA price index (if published by the BLS), and the NSA price index.[19] Where the SA price index is available, we use it to calculate the month-over-month component price changes; otherwise, we use the NSA price index. If both are available, and if the SA index of a component begins after 1997M12 but the NSA index begins prior to the SA index, then prior calculating median CPI, we impute missing month-over-month price changes calculated from the SA index as far back as possible using the month-over-month price changes as calculated from the NSA index.[20]

4.2 Splitting Shelter Components of CPI

Unlike the other components in the CPI, OER and Rent cannot be split into further nationally representative subindices.[21] Further disaggregation can only be done by geography and city size-class.

We define three splits of OER and Rent:

- **OER4**: The first split, OER4, aligns with the current "Revised Methodology" of the Median and Trimmed-Mean CPI. OER4 breaks apart the national OER price indices and weights into

[17] Prior to January 2010, "Telephone Services" consisted of "Land-line telephone services, local charges," "Land-line telephone services, long distance," and "Wireless telephone services." From 2010 onwards, "Land-line telephone services, local charges," and "Land-line telephone services, long distance" are merged into a single "Land-line telephone services" component.

[18] Going back further in time before 1998 would require adjusting each component basket to align with the item structure of the CPI prior to the sixth comprehensive revision. We leave this extension to future research.

[19] Data are collected primarily using Haver Analytics. This is necessary because the BLS does not maintain a database of relative importances. For a list of the components in each collection and the Haver codes for each relative importance and price index series, see the Online Appendix. Price indices for "Nonprescription drugs and medical supplies," "Internal and respiratory OTC drugs," "Nonprescription medical equipment and supplies," "Land-line telephone services, local charges," and "Land-line telephone services, long distance," which were not in Haver and were obtained from the BLS website, are also documented in the Online Appendix.

[20] This is consistent with current BLS practices. As the BLS explains: "Each year the seasonal status of every series is reevaluated based upon certain statistical criteria. Using these criteria, BLS economists determine whether a series should change its status: from "not seasonally adjusted" to "seasonally adjusted", or vice versa." Therefore, if the SA price index begins later than the NSA price index for a component, then presumably the BLS did not detect seasonality in that component over the period where the NSA exists but the SA index does not, and so imputation in the manner in which we carry it out is valid. For more information on seasonal adjustment in the CPI, see: https://www.bls.gov/cpi/seasonal-adjustment/using-seasonally-adjusted-data.htm.

[21] Adams and Verbrugge (2021) criticize this practice, and propose a split along a structure-type dimension.

four regional OER price indices and weights, one for each of the four Census regions: Northeast, Midwest, South, and West. Rent is not split. Thus, OER4 consists of 5 components: OER for each of the four Census regions; and Rent.

- **OER8**: OER8 further splits each regional OER index into two parts, one for city size class A (corresponding to population size greater than 2.5 million) and size class B/C (population size 2.5 million or less), where size classes are defined by the BLS. Rent is not split. Thus, OER8 consists of 9 components: OER for City Size A and OER for City Size B/C for each of the four Census regions, respectively; and Rent.

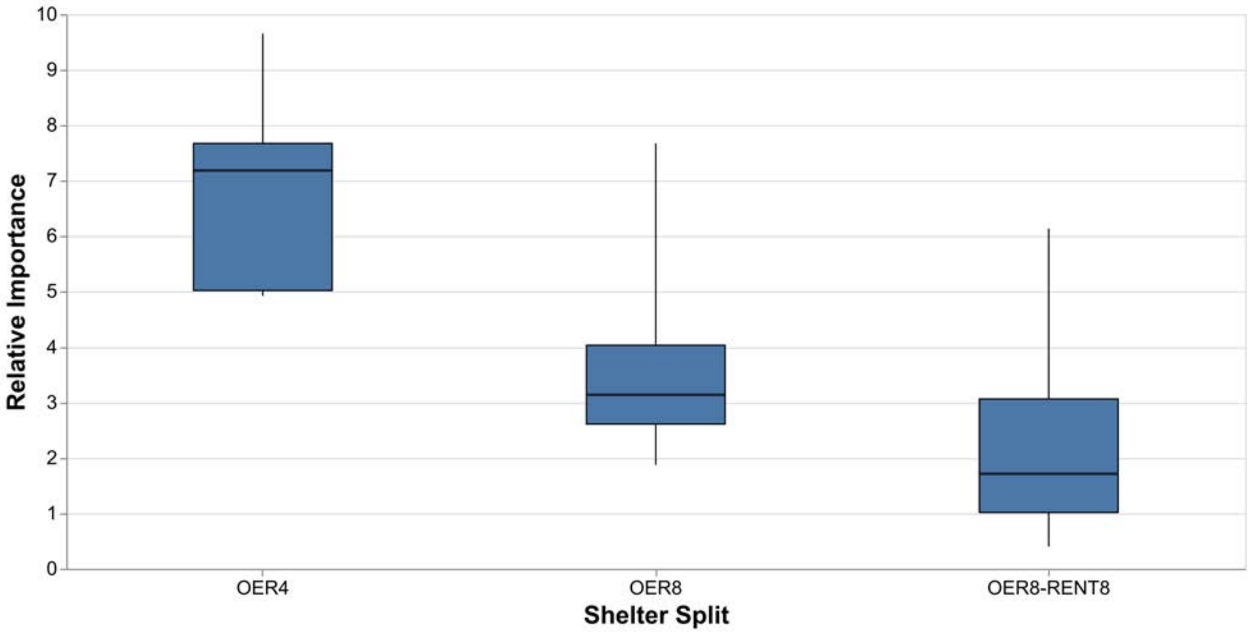
- **OER8-RENT8**: OER8-RENT8 retains the OER8 split and additionally splits Rent along the same eight regions as OER8. Thus, OER8-RENT8 consists of 16 components.[22]

As shown in Figure 2, on the basis of component weights, OER8 is more disaggregated than OER4, and OER8-RENT8 is more disaggregated than OER8, as expected. For all components of OER4, OER8, and RENT8, we (1) calculate monthly relative importances as outlined previously in the Methodology section[23]; and (2) seasonally adjust each price index using X-13-ARIMA-SEATS once using data starting in 1997M12.

[22]Prior to December 2017, there were actually three city size classes: Size A (cities with a population size over 2.5 million), Size B/C (cities with a population size between 50 thousand and 2.5 million), and Size D (cities with a population less than 50,000). However, the BLS only published price indices and relative importances for Size D cities in the Midwest and South regions, and not for the Northeast and West regions. Therefore, between December 1997 and November 2017, we use **nine** regional indices for both the OER8 and RENT8 splits: the eight mentioned previously, as well as the index and weights for all Size D cities.

[23]All Haver codes for the input price indices and relative importances of shelter split components may be found in the Online Appendix. Additionally, to fill gaps in Haver annual relative importance data, we hand-collected annual relative importances for Size D cities from 1997 to 2016, as well as for every city size and Census region combination from 1997 to 2000, from electronic copies of CPI-U Regional Importance Reports. Finally, price index data for Size D cities were downloaded directly from FRED (Federal Reserve Economic Data). Collected non-Haver data may be found in the Online Appendix.

Figure 2: Distribution of Component Weights in Each Housing Split



Notes: Component weights as published December 2023 by the BLS.

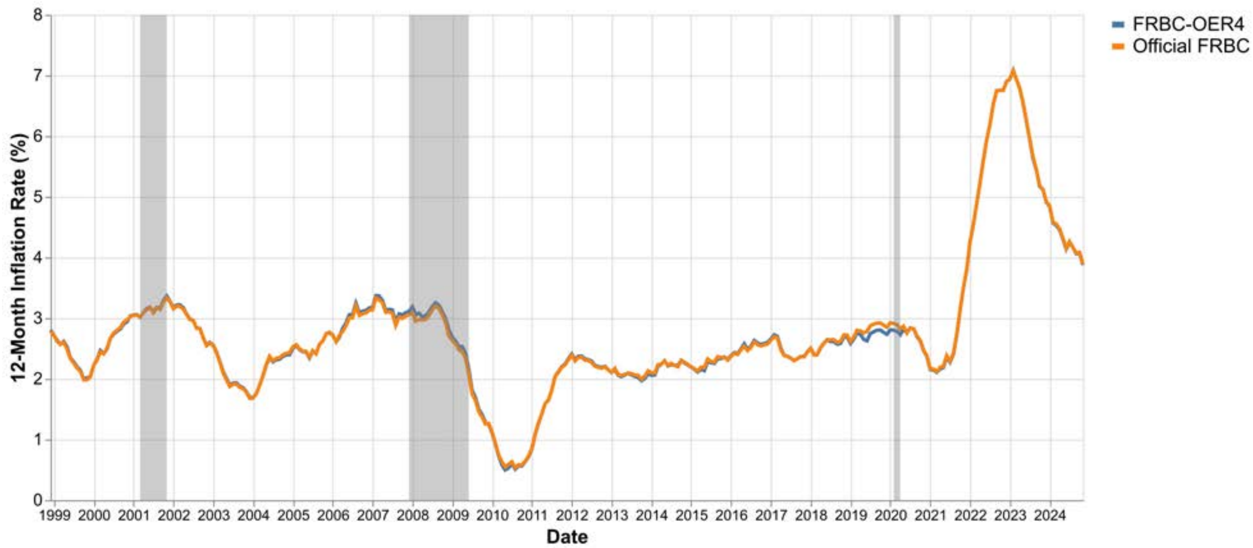
4.3 Combining Shelter and Non-Shelter Splits

Complete splits of the CPI are the cross-product of non-shelter component splits (FRBC, C2, C3, C4, C5) with shelter component splits (OER4, OER8, OER8-RENT8). This gives us 15 component collections in total. Our notation for a particular split is given by Ci-J where $i = 2, 3, 4, 5$ refers to the degree of disaggregation of the non-shelter components, and J $\in$(OER4, OER8, OER8-RENT8) refers to the degree of disaggregation of the shelter components. C5-OER8-RENT8, the most disaggregated split, has 156 components. FRBC-OER4, the least disaggregated CPI split, is our baseline.

Our reconstruction of the Median CPI is accurate. Figure 3 demonstrates that, despite minor methodological differences, the median measure we calculated from the FRBC-OER4 split is essentially identical to its official counterpart.[24]

[24]For the remainder of the paper, we continue to focus attention on the median measures. Appendix 1 provides results pertaining to the Trimmed-Mean CPI.

Figure 3: Comparing the FRBC-OER4 Median CPI Inflation to the Official Federal Reserve Bank of Cleveland Measure



From each of our 15 splits of the CPI, using the methodology outlined in Section 2.2.2, it is straightforward to derive measures of 12-month Median CPI inflation. Hereafter, when we refer to inflation, we are speaking of the 12-month rate of change, measured as a percent. In addition, we will refer to Median CPI inflation as derived from a particular split of the CPI by the name of the split; for example, taking split FRBC-OER8-RENT8 for concreteness, we refer to Median CPI inflation as derived from split FRBC-OER8-RENT8 as "Median FRBC-OER8-RENT8 inflation." Our sample starts in 1998M12 and ends in 2024M11.

5 Results

To empirically evaluate the effect of higher component disaggregation on the Median CPI, we primarily focus on two criteria that are standard in the core/MTT inflation evaluation literature.

First, we evaluate how accurately each measure of median inflation tracks medium-term movements in CPI inflation. We do so using two metrics. First, we assess how closely the mean of each candidate measure matches that of CPI inflation over our sample. Second, we examine how accurately each candidate measure tracks changes in a standard ex-post proxy of the "true" underlying MTT in CPI inflation.

Second, we assess the extent to which each candidate measure has predictive power over future movements in CPI inflation at six horizons: 1, 3, 6, 12, 24, and 36 months. We do so using a simple linear regression benchmark taken from the literature, and evaluate predictive power by first looking at measures of in-sample fit, and then by looking at out-of-sample predictive accuracy in a pseudo-real-time forecasting exercise.[25]

5.1 Accuracy in Estimating the MTT in CPI Inflation

5.1.1 Accuracy in mean

An MTT estimator should, over a long period of time, have an average as close as possible to the average rate of CPI inflation (Clark 2001; Rich and Steindel 2007; Higgins and Verbrugge 2015; Stock and Watson 2016). Therefore, we first examine how close the mean of each median measure was to the mean of CPI inflation over the pre-pandemic and full samples.

In Figure 4, for each median inflation candidate, we report the ratio of the average median inflation rate and the average CPI inflation rate. We find that median CPI measures, owing to skewness in the cross-sectional distribution of component growth rates, are upward-biased compared to the CPI.[26] However, there is variation in this bias across the median CPI measures. Most notably, we find that (1) as we move across a given row (i.e., as we increase the degree of non-shelter disaggregation), the mean of the median inflation series moves *further away* from the mean of headline CPI; and (2) as we move down any column (i.e. as we increase the degree of shelter disaggregation) the mean of the median inflation series moves *closer* to the mean of headline CPI. Hence, FRBC-OER8-RENT8 is the least biased, at 5% higher (6% higher) in the full (pre-pandemic) sample. This is a 3% reduction (5% reduction) relative to the benchmark FRBC-OER4 median inflation in the full (pre-pandemic) sample, the mean of which overstates that of headline CPI inflation by 8% (11%).

To investigate whether these differences are significant, we perform t-tests of the null hypothesis that the mean of the jth candidate median inflation measure, $j = 1, ..., 15$, is statistically indistin-

[25] High persistence and low variance are also often seen as desirable criteria for MTT indicators; see, e.g., Clark (2001), Jonanssen and Nordbo (2006), Silver (2007), Lao and Steyn (2019), or Richards (2024). We find that our alternative median measures have very similar persistence and variance properties in the full sample.

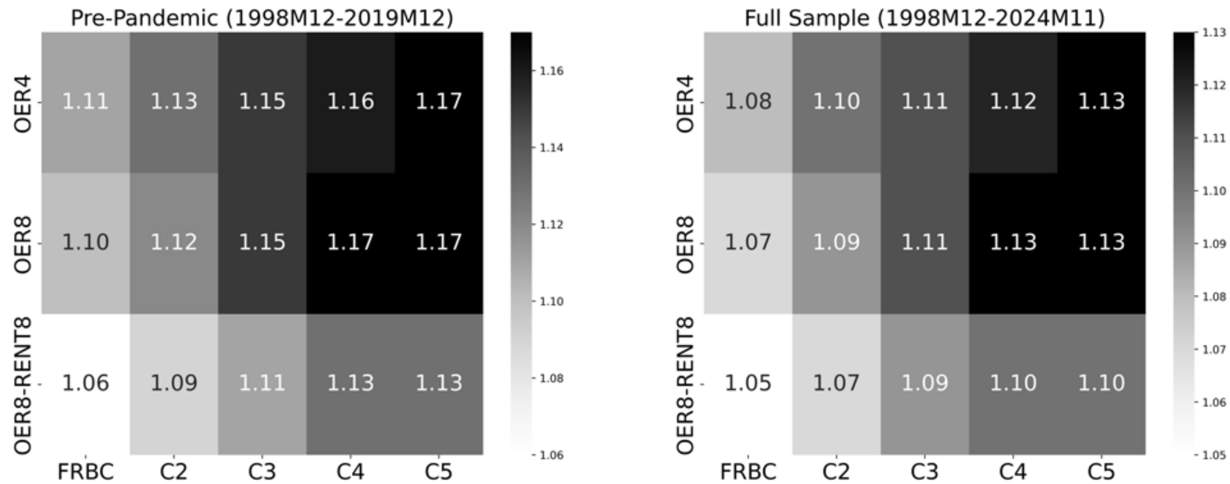
[26] Rich, Verbrugge, and Zaman (2022) discuss a similar phenomenon occurring in the components of PCE inflation.

guishable from the mean of CPI inflation. More formally, we test:

$$H_0 : \mathbb{E}[\pi_j^c] = \mathbb{E}[CPI]$$

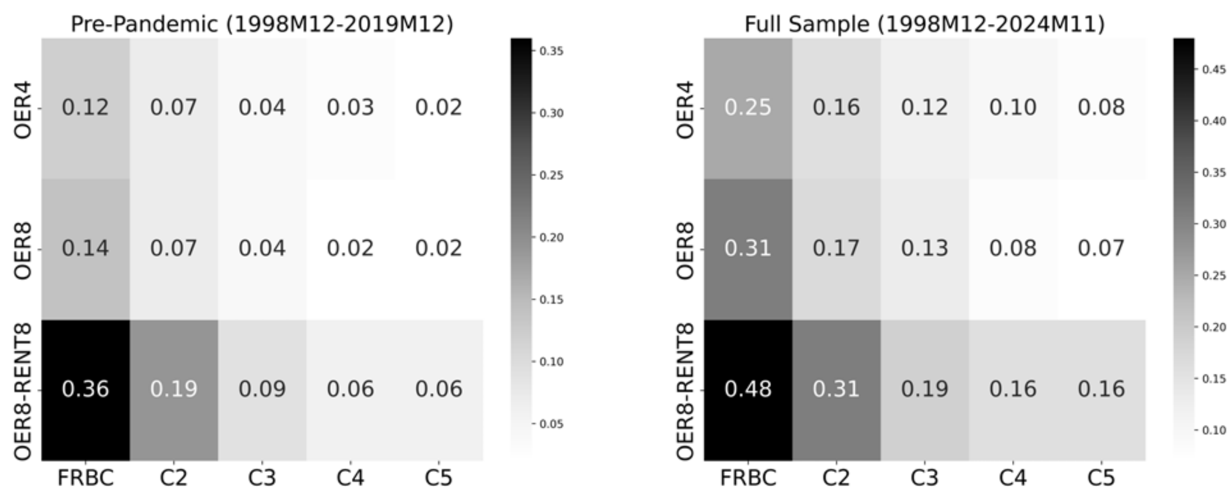
where π_j^c denotes the jth candidate median inflation measure, CPI denotes CPI inflation, and $\mathbb{E}[\cdot]$ is the expectation operator. Test statistics are calculated using heteroskedasticity-and-autocorrelation-consistent (HAC) standard errors. Test p-values are reported in Figure 5. These results reveal that the observed differences between average inflation in each candidate median measure and headline CPI inflation are not statistically significant at all common significance levels in both the pre-pandemic sample and the full sample for each FRBC split, as well as for the C2-OER8-RENT8 split.

Figure 4: Mean of Median Inflation Measures Relative to Mean of CPI Inflation



Notes: Reported figures are the ratio of the average of the median inflation rate and the average of CPI inflation. Both averages are computed as the mean of 12-month inflation rates, measured by percent changes, over the indicated period. Darker shading indicates higher values, while lighter shading indicates lower values.

Figure 5: p-Values of a Statistical Test of Equal Mean Relative to CPI Inflation



Notes: Reported figures are the p-values of a t-test of $H_0 : \mathbb{E}[\pi_j^c] = \mathbb{E}[CPI]$, where π_j^c denotes the jth candidate median inflation measure. The p-value is obtained by taking the difference of each median inflation measure from CPI inflation, and regressing this against a constant. The test statistic of the constant term is calculated using heteroskedasticity-and-autocorrelation-consistent (HAC) standard errors with a small sample correction and $\lfloor 4[T/100]^{2/9} \rfloor$ lags, where T refers to the size of the estimation sample. Darker shading indicates higher values, while lighter shading indicates lower values.

5.1.2 Accuracy versus a standard ex-post MTT estimate

An MTT estimator should closely track ex-post estimates of that MTT, thereby helping to distinguish persistent movements in underlying trend inflation from transitory price shocks (Clark 2001, Rich and Steindel 2007; Higgins and Verbrugge 2015). Following Bryan et al. (1997) and many subsequent studies, we use a 37-month centered moving average of 12-month CPI inflation as our ex-post estimate, or proxy, of the MTT. That is, in a given month, the proxy is equal to the average of inflation in the current month, the preceding 18 months, and the subsequent 18 months. We examine the RMSE of deviations between each median inflation measure and the MTT proxy. Hence, for each of our j candidates, $j = 1, ..., 15$, we compute:

$$RMSE(\bar{\pi}^{37MMA} - \pi_j^c) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\bar{\pi}_t^{37MMA} - \pi_{j,t}^c)^2} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(e_{j,t}^{37MMA})^2} \tag{1}$$

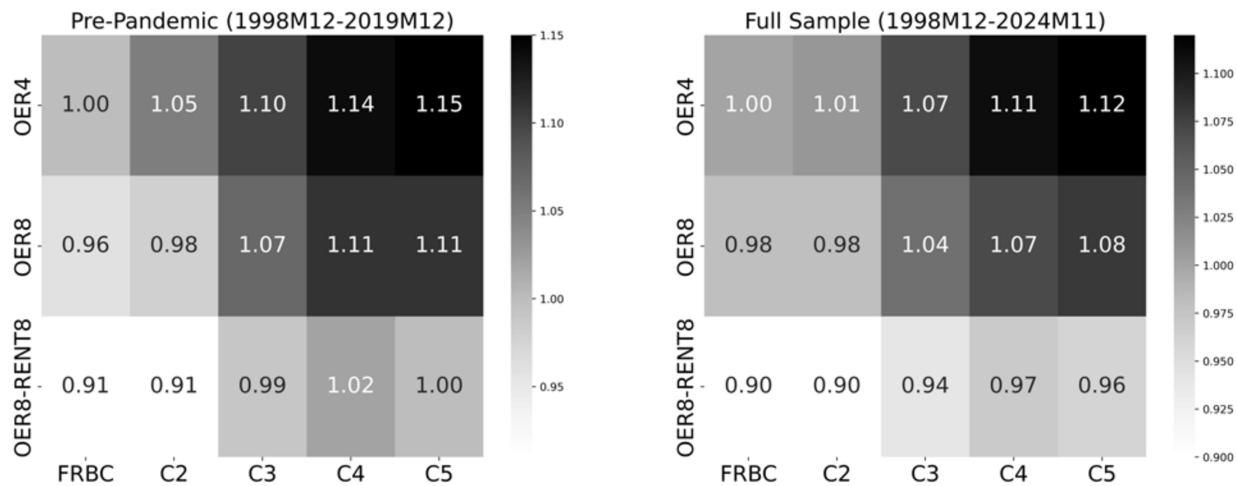
where $\bar{\pi}^{37MMA}$ denotes the 37-month centered moving average MTT proxy, π_j^c is a candidate median CPI inflation measure, and $e_{j,t}^{37MMA} = (\bar{\pi}_t^{37MMA} - \pi_{j,t}^c)$ measures the deviation between the two at month t.

In Figure 6, we report the RMSE of each measure of median inflation relative to the RMSE of our baseline median FRBC-OER4 inflation measure. By this metric, FRBC-OER8-RENT8 and C2-OER8-RENT8 equivalently outperform all other median candidates, in many cases by a wide margin. The RMSE of FRBC-OER8-RENT8 and C2-OER8-RENT8 is 9% lower than that of median FRBC-OER4 inflation pre-pandemic, and 10% lower in the full sample. As before, the results in Figure 6 show that greater shelter disaggregation *improves* the ability of the derived median inflation measure to track the trend inflation proxy, while greater non-shelter disaggregation *worsens* it.

To determine if differences in RMSEs between the baseline FRBC-OER4 measure and each of the other median inflation measures are statistically significant, we follow Rich and Steindel (2007) in constructing the Diebold-Mariano (1995) (DM) test statistic for each pairing. This allows us to consider the null hypothesis that FRBC-OER4 median inflation and another candidate median inflation measure j both track the 37-month centered moving average MTT proxy equally well against the alternative hypothesis of significantly different tracking ability. Test p-values are reported in Figure 7. The DM tests show that the observed reduction in the RMSE of FRBC-OER8-RENT8 and

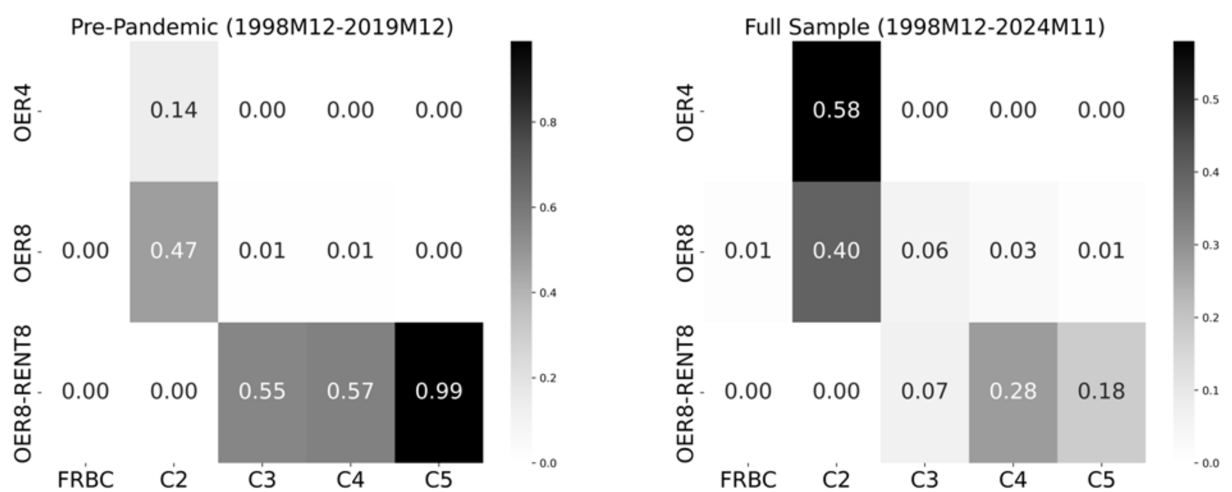
C2-OER8-RENT8 inflation relative to FRBC-OER4 inflation is statistically significant at all common significance levels in both the pre-pandemic sample and the full sample. Additionally, FRBC-OER8, while less accurate than FRBC-OER8-RENT8 and C2-OER8-RENT8, is also statistically significantly more accurate than FRBC-OER4 across samples. All other candidate median measures have either the same or higher RMSE than FRBC-OER4 in one or both sample periods, or their tracking ability is statistically indistinguishable from that of FRBC-OER4 at the 5% level and above.

Figure 6: $RMSE(\hat{\pi}^{37MMA} - \hat{\pi}_j)$ of Median Inflation Measures Relative to FRBC-OER4



Notes: Reported figures are the RMSE of deviations of the median inflation measure from a 37-month centered moving average of CPI Inflation, divided by the same for median FRBC-OER4 inflation. In the pre-pandemic sample, the moving average is computed using CPI inflation through December 2019 only. Darker shading indicates higher values, while lighter shading indicates lower values.

Figure 7: p-Values of a Statistical Test of Equal Ability in Tracking $\hat{\pi}^{37MMA}$ for Median Inflation Measures, Relative to FRBC-OER4



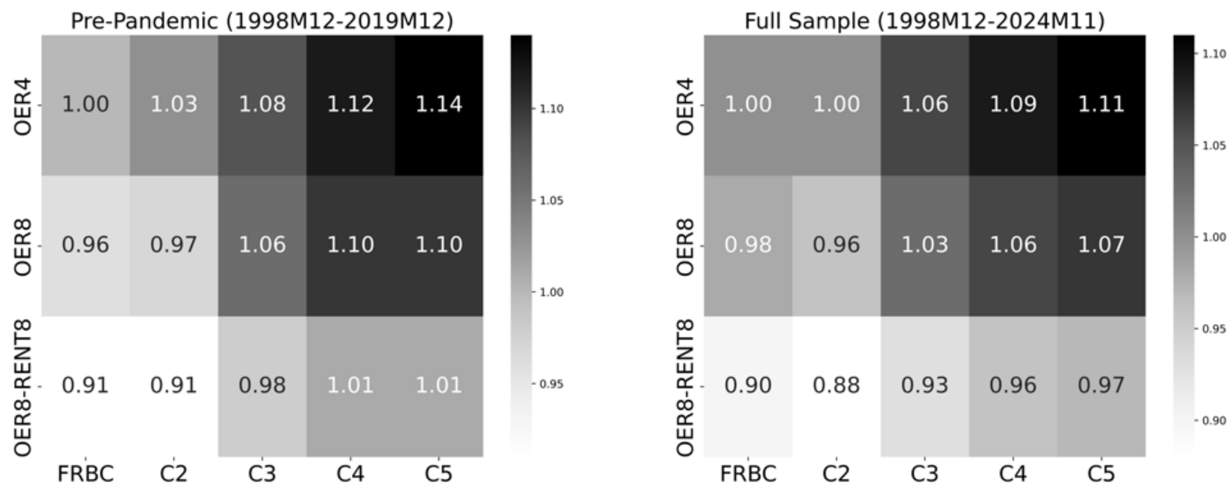
Notes: Reported figures are the p-values of a Diebold-Mariano (1995) test that $RMSE(\bar{\pi}^{37MMA} - \pi^c_{FRBC-OER4})$ and $RMSE(\bar{\pi}^{37MMA} - \pi^c_j)$ are equal, where j denotes the jth candidate median inflation measure. The p-value is obtained by taking the difference of the two squared errors series $e^{37MMA}_{j,t}$ and $e^{37MMA}_{FRBC-OER4,t}$, and regressing the resulting series against a constant. The test statistic of the constant term is then calculated using HAC standard errors with a small sample correction and $\lfloor 4[T/100]^{2/9} \rfloor$ lags, where T refers to the size of the estimation sample. Darker shading indicates higher values, while lighter shading indicates lower values.

As a robustness check, we repeat the previous exercise using a different ex-post estimate of the MTT. Following Higgins and Verbrugge (2015) and Carroll and Verbrugge (2019), we utilize the two-stage centered moving average, or 2SMA, trend. The 2SMA trend is constructed by first applying a 25-month centered moving average, followed by a 13-month centered moving average, to headline CPI inflation. This measure moves similarly to the centered 37-month moving average, but has the advantage that it eliminates all fluctuations lasting under 36 months. In Figure 8, we report the RMSE of each median inflation measure against the the 2SMA trend relative to the same for median FRBC-OER4 inflation. Results are largely consistent with those in Figure 6. FRBC-OER8-RENT8 and C2-OER8-RENT8 are once again the best performing measures in the pre-pandemic sample, achieving a 9% reduction in RMSE relative to the FRBC-OER4 baseline. However, C2-OER8-RENT8 slightly edges out FRBC-OER8-RENT8 in the full sample, achieving a 12% reduction in RMSE relative to a 10% reduction for FRBC-OER8-RENT8.

As before, in Figure 9 we report for each median inflation candidate the p-values of the DM test of equal ability in tracking the 2SMA trend estimate against the FRBC-OER4 baseline. Results show that only FRBC-OER8-RENT8 and C2-OER8-RENT8 inflation achieve a statistically significant improvement in tracking the 2SMA trend relative to FRBC-OER4 inflation at all common significance levels in both the pre-pandemic sample and the full sample.

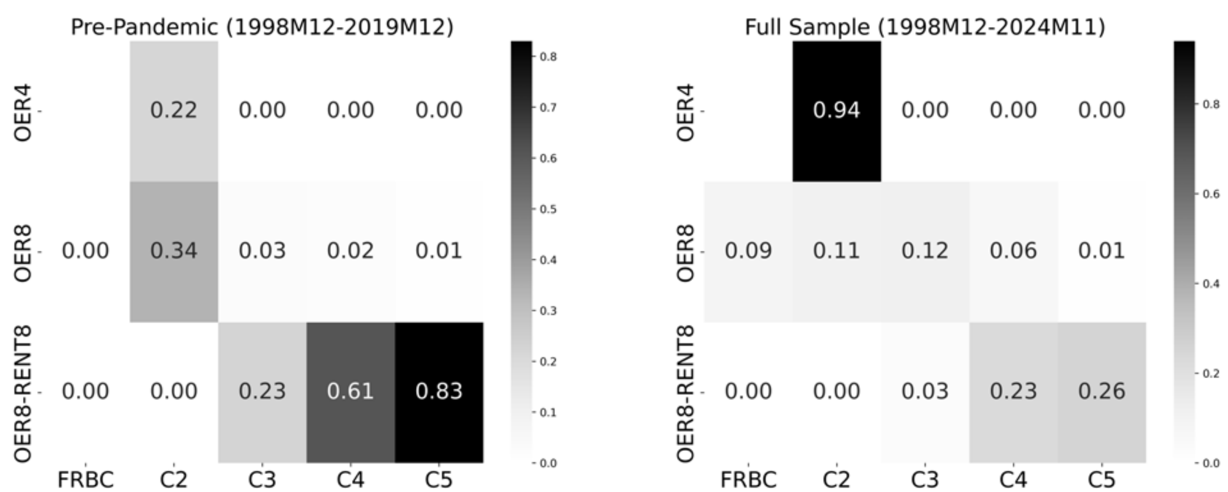
These results show that splitting OER improves the ability of median measures to track the underlying trend in CPI inflation, and splitting Rent yields still further benefits in the same direction. On the other hand, non-shelter disaggregation is generally associated with a *deterioration* in the median's trend-tracking capability as the degree of this disaggregation increases. These results largely reinforce the conclusions in the previous section, which together show that (1) in all cases, FRBC-OER8-RENT8 improves upon FRBC-OER4 in tracking the mean of CPI inflation over time as well as the underlying inflation trend; (2) in nearly each case, disaggregating the FRBC-OER8-RENT8 basket further, to C2-OER8-RENT8, yields the same or worse performance on these metrics; and (3) disaggregating further, to C3-OER8-RENT8 or beyond, hurts performance relative to both FRBC-OER8-RENT8 and C2-OER8-RENT8.

Figure 8: $RMSE(\hat{\pi}^{2SMA} - \hat{\pi}_j)$ of Median Inflation Measures Relative to FRBC-OER4



Notes: Reported figures are the RMSE of deviations of the median inflation measure from a two-stage centered moving average (2SMA) of CPI inflation, divided by the same for median FRBC-OER4 inflation. In the pre-pandemic sample, the moving average is computed using CPI inflation through December 2019 only. Darker shading indicates higher values, while lighter shading indicates lower values.

Figure 9: p-Values of a Statistical Test of Equal Ability in Tracking $\hat{\pi}^{2SMA}$ for Median Inflation Measures, Relative to FRBC-OER4



Notes: Reported figures are the p-values of a Diebold-Mariano (1995) test that $RMSE(\hat{\pi}^{2SMA} - \hat{\pi}_{FRBC-OER4})$ and $RMSE(\hat{\pi}^{2SMA} - \hat{\pi}_j)$ are equal, where j denotes the jth candidate median inflation measure. The p-value is obtained by taking the difference of the two squared errors series $\hat{e}_{j,t}^{2SMA}$ and $\hat{e}_{FRBC-OER4,t}^{2SMA}$, and regressing the resulting series against a constant. The test statistic of the constant term is then calculated using HAC standard errors with a small sample correction and $\lfloor 4[T/100]^{2/9} \rfloor$ lags, where T refers to the size of the estimation sample. Darker shading indicates higher values, while lighter shading indicates lower values.

5.2 Predictive Power over Future Inflation

5.2.1 In-sample explanatory power

It is desirable for an MTT estimator to have explanatory power over future inflation. To assess the in-sample explanatory power of each of our measures of median inflation, we follow previous research (e.g., Clark 2001; Rich and Steindel 2007) and estimate regressions of the form:

$$\pi_{t+h} - \pi_t = \alpha_{j,h} + \beta_{j,h}(\pi_t - \pi^c_{j,t}) + \epsilon_{j,t+h} \tag{2}$$

where h denotes the forecast horizon in months, π_t refers to the current reading of CPI inflation, and $\pi^c_{j,t}$ refers to the current reading of the jth indicator of median CPI inflation, $j = 1, ...15$. We consider six horizons h: 1, 3, 6, 12, 24, and 36 months.

For $h \in \{1, 3, 6\}$:

- $\pi_{t+h} = 100 \cdot [(P_{t+h}/P_t)^{12/h} - 1]$ is the h-month rate of inflation h months ahead, at an annualized rate.
- $\pi_t = 100 \cdot [(P_t/P_{t-h})^{12/h} - 1]$ is the current h-month rate of inflation at an annualized rate.

For $h \in \{12, 24, 36\}$:

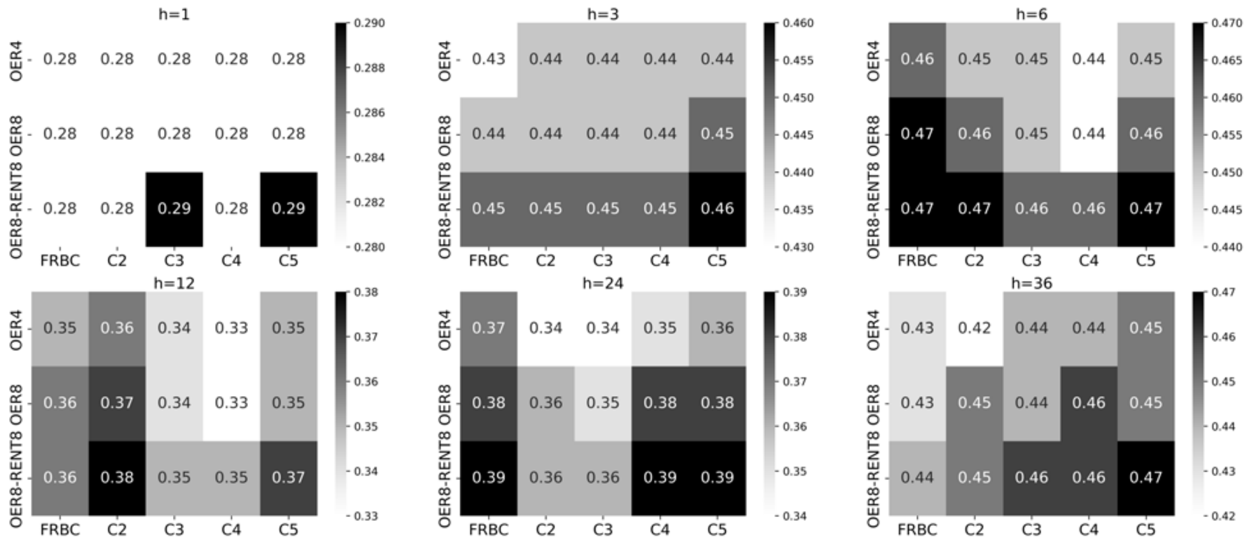
- $\pi_{t+h} = 100 \cdot [(P_{t+h}/P_{t+h-12}) - 1]$ is the 12-month rate of inflation h months ahead.
- $\pi_t = 100 \cdot [(P_t/P_{t-12}) - 1]$ is the current 12-month rate of inflation

Notice that the intercept, $\alpha_{j,h}$, allows for (fixed) bias adjustment for each candidate j - so that if a particular MTT estimator has a large bias versus headline inflation, the intercept will correct for that bias. Thus, candidates with higher bias are not automatically penalized.

We present the adjusted R^2, denoted $\bar{R}^2$, from these regressions for each measure of median inflation and for horizons $h \in \{1, 3, 6, 12, 24, 36\}$. In Figure 10 we report $\bar{R}^2$ from fitting Equation 2 on data in the pre-pandemic sample, and in Figure 11 we report the same for regressions estimated over the full sample.

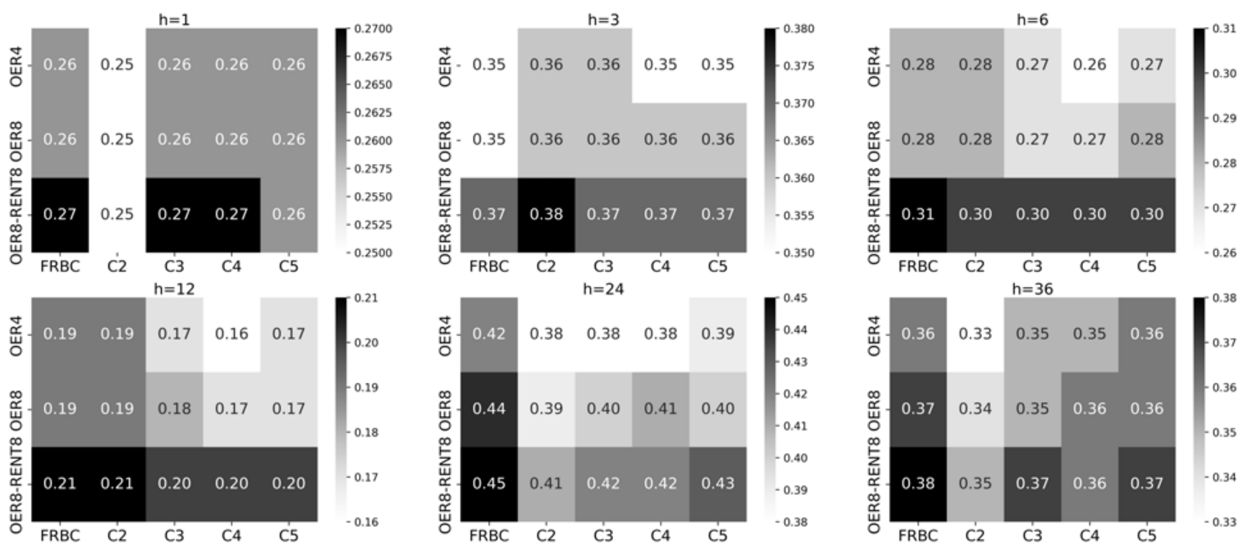
Overall, differences in in-sample fit between candidates are small. There are also no discernible trends as the degree of non-shelter disaggregation increases, although we note that in the full sample, the $\bar{R}^2$ of FRBC-OER8-RENT8 dominates in five of the six horizons. However, one finding stands out, and is consistent across both samples and all six horizons: for any given level of non-shelter disaggregation, as we increase shelter disaggregation, the $\bar{R}^2$ of the more shelter-disaggregated candidate weakly dominates that of the less shelter-disaggregated candidate. We therefore conclude that greater shelter disaggregation robustly improves the in-sample explanatory power of median CPI over future CPI inflation. This result is consistent with our earlier results showing that greater shelter disaggregation improves the ability of median CPI inflation in tracking medium-term movements in CPI inflation.

Figure 10: In-Sample Adjusted R^2 of Equation 2, Pre-Pandemic Sample (1998M12-2019M12)



Notes: Reported figures are the adjusted R^2 from fitting Equation 2 for each jth candidate median inflation measure. h denotes the horizon in months. Darker shading indicates higher values, while lighter shading indicates lower values.

Figure 11: In-Sample Adjusted R^2 of Equation 2, Full Sample (1998M12-2024M11)



Notes: Reported figures are the adjusted R^2 from fitting Equation 2 for each jth candidate median inflation measure. h denotes the horizon in months. Darker shading indicates higher values, while lighter shading indicates lower values.

5.2.2 Out-of-sample forecasting ability

In the previous section, we measured how increasing levels of component disaggregation impacted the explanatory power of median CPI over future headline CPI by looking at an in-sample measure of fit from a simple benchmark regression. In this section, we again turn to the same simple benchmark regression in order to conduct a pseudo-real-time-out-of-sample forecasting exercise.[27] This exercise provides an additional metric by which to measure how median CPI's predictive accuracy over headline CPI varies as we vary the levels of shelter and non-shelter disaggregation in the underlying components.

Consider again Equation 2, this time dropping the j subscript for ease of exposition:

$$\pi_{t+h} - \pi_t = \alpha_h + \beta_h(\pi_t - \pi_t^c) + \epsilon_{t+h}$$

Defining $y_{t+h} = \pi_{t+h} - \pi_t$ and $x_t = \pi_t - \pi_t^c$, we estimate regressions of the form:

$$y_t = \hat{\alpha}_h + \hat{\beta}_h x_{t-h}$$

through time t, where $(\hat{\alpha}_h, \hat{\beta}_h)$ denote the estimated parameters of the regression. We then form x_t and obtain our forecast $\hat{y}_{t+h}$ as:

$$\hat{y}_{t+h} = \hat{\alpha}_h + \hat{\beta}_h x_t$$

Finally, the implied forecast of $\hat{\pi}_{t+h}$ is then:

$$\hat{\pi}_{t+h} = \hat{y}_{t+h} + \pi_t$$

For each h, we estimate the regressions using an expanding window, such that the first our-of-sample forecast is obtained for 2010M12. Thus, our first estimation window runs through $2010M12-h$ to obtain a forecast for 2010M12, our second estimation window runs through $2010M12 - h + 1$

[27]Our exercise is "pseudo" real-time due to the fact that we are using seasonally adjusted headline CPI. We also construct our measures of median CPI using (1) seasonally adjusted CPI components, when available; and (2) seasonally adjusted OER and Rent components, which we seasonally adjust ourselves once, using all available data starting in 1997M12.

to obtain a forecast for 2011M01, and so on, until we have made a forecast for the last month in our sample.

Table 1 reports RMSFEs for the pre-pandemic sample relative to the RMSFE of forecasts made using the FRBC-OER4 benchmark. Table 1 also reports statistical significance results using the Diebold and Mariano (1995) (DM) test with the small-sample correction of Harvey, Leybourne, and Newbold (1997) for equal predictive accuracy between forecasts derived from a candidate median measure j and forecasts obtained using the FRBC-OER4 benchmark. Table 2 reports the same results for the full sample.

In general, out-of-sample forecast accuracy for alternative splits is comparable to the benchmark, gains or losses are only rarely statistically distinguishable (and usually only at the 10% level), and statistically significant forecast gains are not robust between the pre-pandemic and full samples.[28] Relative RMSFE improvements are concentrated on OER8-RENT8 splits. In addition, 4 out of 5 statistically significant forecast gains in the pre-pandemic sample and 6 out of 7 in the full sample are found among OER8-RENT8 splits. Aside from moving from the FRBC basket to C2, there is little tendency for point forecast accuracy to improve upon further non-shelter disaggregation. Overall, the two splits that appear to be the top performers are FRBC-OER8-RENT8 and C2-OER8-RENT8. The latter is superior pre-pandemic but the two have roughly equal performance when considering the full sample.

[28]Over the pre-pandemic sample, there are four cases in which statistically significant forecast gains are observed (mostly at the 36-month horizon) but these results are not robust once we extend the forecasting exercise to encompass the full sample. Similarly, over the full sample there are seven cases with statistically significant forecast gains, but for six of these cases, results are not robust when we consider the shorter sample.

Table 1: Relative RMSFEs of Out-of-Sample Forecasts Using Equation 2, Pre-Pandemic Sample (1998M12-2019M12)

Forecast Horizon (Months)	1	3	6	12	24	36
FRBC-OER4 RMSFE	2.23	1.86	1.6	1.39	1.28	0.92
FRBC-OER8	1.01*	1.01	1.01	1.01	0.99	1.02
FRBC-OER8-RENT8	1.01	1.01	1.01	0.99	0.98	1.02
C2-OER4	1.0	0.99	1.02	1.0	1.02	0.96
C2-OER8	1.01	0.98	1.0	0.98	0.99	**0.96***
C2-OER8-RENT8	1.0	**0.97***	0.98	0.93	0.99	0.97
C3-OER4	1.0	1.0	1.0	1.0	1.03	0.97
C3-OER8	1.01	1.01	1.01	1.0	1.02	0.97
C3-OER8-RENT8	1.0	1.0	0.99	0.97	1.01	**0.97***
C4-OER4	1.0	1.0	1.01	1.02	1.02	0.96
C4-OER8	1.01	1.0	1.01	1.0	0.97	0.94
C4-OER8-RENT8	1.0	0.99	0.99	0.98	0.98	**0.96*****
C5-OER4	1.0	1.01	1.02	1.03	1.02	0.94
C5-OER8	1.01	0.99	1.01	1.01	0.99	0.95
C5-OER8-RENT8	1.0	1.0	0.99	0.97	0.98	**0.94*****

Notes: RMSFE is the root mean squared forecast error. The row "FRBC-OER4 RMSFE" reports the raw RMSFE of the FRBC-OER4 benchmark for each forecast horizon. All other rows report relative RMSFEs, with the RMSFE of FRBC-OER4 for forecast horizon h taken as the denominator for relative RMSFEs in column h. Relative RMSFEs less than 1 are highlighted in green. For each relative RMSFE, we calculate the Diebold and Mariano (1995) (DM) test with the small-sample correction of Harvey, Leybourne, and Newbold (1997) for equal predictive accuracy between a given forecast and the forecast from the FRBC-OER4 benchmark. Relative RMSFEs in bold and with *, **, or *** denote rejections of the null hypothesis of equal predictive accuracy of the alternative median CPI measure and the FRBC-OER4 benchmark at the 10%, 5%, or 1% level, respectively, based on the DM test.

Table 2: Relative RMSFEs of Out-of-Sample Forecasts Using Equation 2, Full Sample (1998M12-2024M11)

Forecast Horizon (Months)	1	3	6	12	24	36
FRBC-OER4 RMSFE	2.76	2.54	2.39	2.33	2.25	1.98
FRBC-OER8	1.01*	1.01	1.01	1.0	**0.98***	0.99
FRBC-OER8-RENT8	1.0	**0.98***	0.98	0.98	0.98	0.99
C2-OER4	1.01	1.0	0.99	1.0	1.03	1.02
C2-OER8	1.02	0.99	0.99	1.0	1.03	1.03
C2-OER8-RENT8	1.02	**0.98***	**0.97***	**0.97***	1.01	1.02
C3-OER4	1.0	1.01	1.0	1.01	1.03	1.01
C3-OER8	1.01	1.01	1.0	1.0	1.02	1.01
C3-OER8-RENT8	1.0	0.99	**0.97***	0.97	1.01	1.0
C4-OER4	1.0	1.01	1.0	1.02	1.04	1.01
C4-OER8	1.0	1.0	0.99	1.0	1.02	1.01
C4-OER8-RENT8	0.99	0.98	**0.97***	0.98	1.01	1.01
C5-OER4	1.01	1.01	1.01	1.02	1.03	1.01
C5-OER8	1.01	1.01	1.0	1.01	1.02	1.01
C5-OER8-RENT8	1.0	1.0	0.98	0.98	1.0	1.0

Notes: RMSFE is the root mean squared forecast error. The row "FRBC-OER4 RMSFE" reports the raw RMSFE of the FRBC-OER4 benchmark for each forecast horizon. All other rows report relative RMSFEs, with the RMSFE of FRBC-OER4 for forecast horizon h taken as the denominator for relative RMSFEs in column h. Relative RMSFEs less than 1 are highlighted in green. For each relative RMSFE, we calculate the Diebold and Mariano (1995) (DM) test with the small-sample correction of Harvey, Leybourne, and Newbold (1997) for equal predictive accuracy between a given forecast and the forecast from the FRBC-OER4 benchmark. Relative RMSFEs in bold and with *, **, or *** denote rejections of the null hypothesis of equal predictive accuracy of the alternative median CPI measure and the FRBC-OER4 benchmark at the 10%, 5%, or 1% level, respectively, based on the DM test.

5.3 Summary of Results

Our analysis evaluates the impact of higher shelter and non-shelter component disaggregation on median CPI measures, focusing on their accuracy in tracking the medium-term trend in CPI inflation and predictive power over future CPI inflation.

To measure how median inflation tracks medium-term movements in CPI inflation, we first compare the mean of each median inflation measure to the mean of CPI inflation. We find that median CPI measures are generally upward-biased compared to CPI, but this bias decreases as shelter disaggregation increases, particularly with the FRBC-OER8-RENT8 measure, which is the least biased. We then use a 37-month centered moving average (37MMA) and a two-stage centered moving average (2SMA) as proxies for the MTT, and calculate the RMSE of deviations between each median measure and the MTT proxies. FRBC-OER8-RENT8 and C2-OER8-RENT8 consistently outperform other measures, with statistically significant improvements in tracking the MTT. We also find more generally that splitting OER and Rent improve the ability of median measures to track the underlying trend in CPI inflation, while increasing non-shelter disaggregation is associated with a deterioration in the median's trend-tracking capability.

To measure the predictive power of median inflation over future CPI inflation, we use a simple linear regression from the literature to assess the in-sample explanatory power of each median inflation measure. In-sample fit as measured by $\bar{R}^2$ shows that greater shelter disaggregation improves in-sample fit, with FRBC-OER8-RENT8 performing the best across most horizons in the full sample. Finally, we conduct a pseudo-real-time out-of-sample forecasting exercise using the same regression framework. The results indicate that OER8-RENT8 splits offer the most improvements in forecast accuracy at multiple horizons. Overall, FRBC-OER8-RENT8 and C2-OER8-RENT8 appear to have the best out-of-sample forecasting performance.

Taken together, our findings strongly suggest that splitting shelter to the OER8-RENT8 level significantly enhances the ability of median measures to track the MTT in CPI inflation and explain future inflation, while further non-shelter disaggregation beyond the level of FRBC and C2 is detrimental. These findings support the use of FRBC-OER8-RENT8 or C2-OER8-RENT8 as the preferred median measure for tracking the MTT in CPI inflation and explaining future CPI inflation. Since FRBC-OER8-RENT8 represents a smaller change from the current official FRBC Median CPI, we would recommend this split.

6 Discussion

In this section, we explore some practical implications of finer disaggregations of CPI components for median CPI inflation.

First, we investigate how varying the level of disaggregation affects the frequency with which the OER, Rent, and Non-Shelter (i.e., non-OER, non-Rent) components are chosen as the median component in the median CPI inflation measure.

Next, we examine empirically how increasing disaggregation can alter the relationship between median CPI and other key economic variables. It has been suggested (e.g., Dolmas and Koenig 2019 or Ball et al. 2021) that a good MTT estimator should co-move inversely with economic slack. Accordingly, for each of our median CPI candidate measures, we estimate a parsimonious empirical Phillips curve between the median CPI measure and the unemployment gap. This estimation allows us to observe how changing disaggregation affects the strength of that relationship.

Finally, given our focus on tracking the medium-run trend in CPI inflation, we compare the historical time-paths of the baseline FRBC-OER4 median to those from our preferred measure, FRBC-OER8-RENT8, over our sample period. By doing so, we aim to determine how inferences about the medium-run trend in inflation would change when using one measure versus the other.

6.1 Variation in the Median Component of Median CPI

In any given month, the median CPI is entirely determined by the rate of change in the particular component chosen as the median component. In the past, there has been considerable interest in the frequency with which shelter components are chosen as the median. Indeed, even at the inception of median CPI as an indicator of interest, Bryan and Cecchetti (1994) noted the disproportionate frequency with which shelter was chosen as the median component. Brischetto and Richards (2007) found that splitting OER into four subcomponents decreased the frequency with which OER was selected as the median component.[29]

Thus, given the degree of interest in the distribution of of the median component, before getting to our main results, we provide in Figure 12 and Figure 13 the frequency with which the following types of components are chosen as the median component: OER, Rent, and Non-shelter (i.e., non-OER, non-Rent) components. Figure 12 covers the pre-pandemic period (1998M12-2019M12) while Figure 13 covers the full sample.

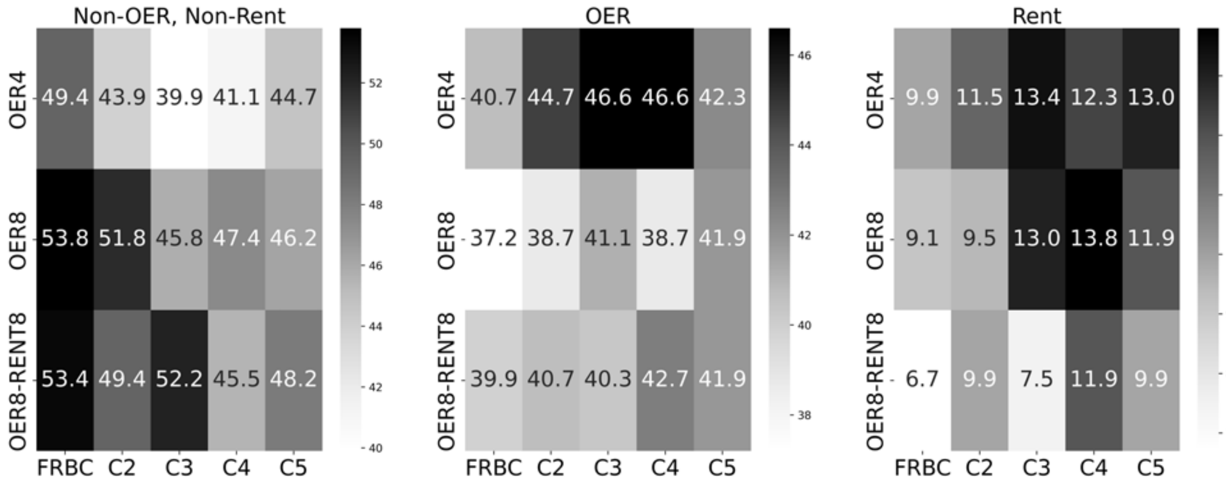
We find that over both samples, non-shelter components are chosen as the median component less than 50% of the time in the baseline FRBC-OER4 split (upper left entry in left-most matrix

[29]Stock and Watson (2020) also draw attention to the frequency with which shelter components are chosen; see Section 6.3

in each figure). Splitting OER4 to OER8, as one would expect, decreases the frequency with which an OER component is chosen as the median component and increases the fraction of non-shelter median components across all non-shelter CPI component splits (FRBC, C2, C3, C4, C5). Splitting shelter further, from OER8 to OER8-RENT8, decreases the frequency with which a Rent component is chosen as the median component across nearly all non-shelter CPI component splits (with C2 as the sole exception), and has a mixed effect on the frequency with which non-shelter is selected as the median component, as it generally causes the frequency of OER selection to increase slightly or remain nearly unchanged. Finally, at each shelter split, increasingly disaggregating non-shelter components generally tends to reduce the frequency with which non-shelter components are chosen as the median.

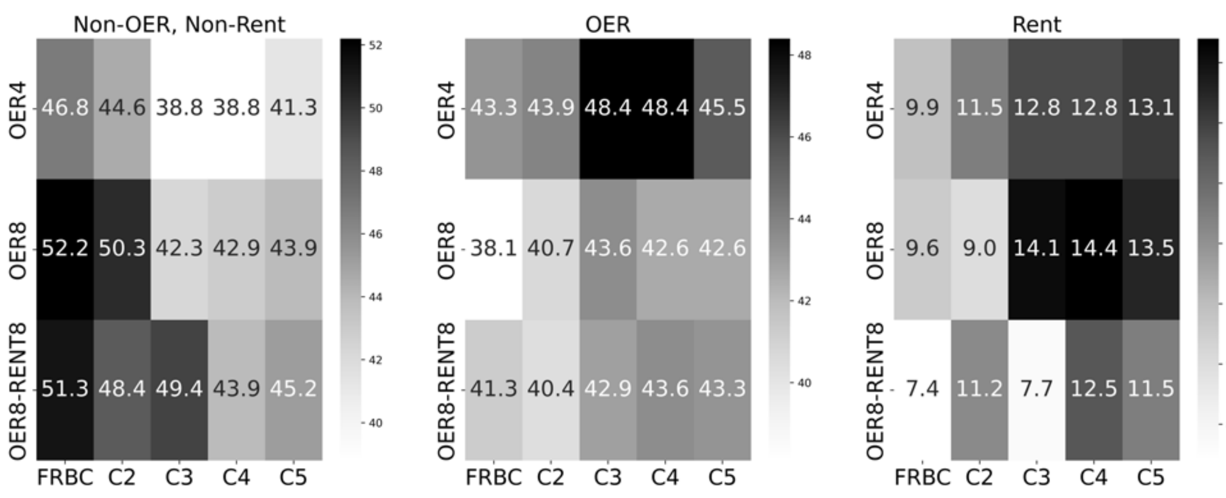
On net, we find that a combination of more disaggregated shelter components and less disaggregated non-shelter components leads to a higher probability that non-shelter components are selected as the median component. In fact, non-shelter components are chosen as the median component with over 50% probability in the full sample in just three CPI splits: FRBC-OER8, FRBC-OER8-RENT8, and C2-OER8. What these three splits have in common is the relatively low level of disaggregation in the non-shelter components.

Figure 12: Percentage Frequency of Selection as the Median Component, Pre-Pandemic (1998M12-2019M12)



Notes: Reported figures are the percentage of months over the indicated sample period in which OER, Rent, and Non-shelter (i.e., non-OER, non-Rent) components are chosen as the median component. Darker shading indicates higher values, while lighter shading indicates lower values.

Figure 13: Percentage Frequency of Selection as the Median Component, Full Sample (1998M12-2024M11)



Notes: Reported figures are the percentage of months over the indicated sample period in which OER, Rent, and Non-shelter (i.e., non-OER, non-Rent) components are chosen as the median component. Darker shading indicates higher values, while lighter shading indicates lower values.

A reasonable inference, given our results in Section 5 and those of Brischetto and Richards (2007), would be that there is a direct link between reducing the frequency with which shelter is chosen as the median component and performance gains in estimators of median CPI. On the other hand, if that were strictly the case, then C3-OER8-RENT8 should have outperformed C2-OER8-RENT8 in our evaluation, given that non-shelter is chosen as the median more frequently in the former basket both pre-pandemic and in the full sample. This suggests that, on its own, choosing shelter as the median component more frequently may simply be a feature of the median CPI, and is not necessarily correlated with inferior performance by median CPI inflation as a measure of the MTT in inflation. For instance, if shelter is highly cyclical, but the remaining components that are typically near the median of the distribution are not, then whenever cyclical forces are strong, they would pull shelter away from the median of the distribution, such that shelter would not be chosen as the median. In that case, the fact that shelter is often chosen as the median would simply mean that shelter generally reflects what is happening to components near the center of the distribution.

6.2 Empirical Application: Phillips Curve Relationship

Recent research has shown that empirical Phillips curve-type relationships (i.e., estimated reduced-form regressions relating inflation to labor market slack) are strong and stable over time when the inflation variable used in regressions is the Median PCE (Ball and Mazumber, 2019), trimmed mean PCE (Ashley and Verbrugge, 2023) or Median CPI (Stock and Watson, 2020). Stock and Watson (2020) argue that this is because, month after month, the component selected as the median CPI is often an inflation component with a strong sensitivity to economic conditions, such as a shelter component. They demonstrate that the dynamics of median CPI inflation are "quite similar" to their cyclically sensitive inflation indicator. Above, we argued that the cyclical sensitivity of the median CPI results from the fact that components near the median of the distribution are cyclically sensitive, not because a shelter component per se is often selected. Nonetheless, the conjecture does raise the question: are candidate MTT estimators that select shelter components less frequently in turn relatively less cyclically sensitive? Once again, the answer we find is both interesting and counterintuitive: no!

To assess the cyclical sensitivity, for each of our median indices, we estimate a parsimonious

Phillips curve formulation similar to the one used in Zaman (2019):

$$\pi_{j,t} = \alpha_j + \beta_j x_t + e_{j,t} \tag{3}$$

where $\pi_{j,t}$ is the 12-month inflation rate of the jth median or trimmed-mean CPI index, x_t is defined as the average of the unemployment gap over the preceding 12 months,

$$x_t = \frac{1}{12}\Sigma_{i=1}^{12}(U_{t-i} - U_{t-i}^N) \tag{4}$$

where U_t is the overall unemployment rate and U_t^N is the Congressional Budget Office's (CBO) estimate of the long-run unemployment rate. β_j, which can be thought as the slope of the Phillips curve, determines the strength of the cyclical relationship between the median or trimmed-mean inflation measure and the labor market slack.

Table 3 reports the estimated β_j for each median candidate. Also reported are the p-values to provide an assessment as to whether each estimated β_j is statistically different from zero.[30] To abstract from the extreme volatility in the unemployment rate data at the onset of and during the COVID-19 pandemic, we also report estimates based on estimating the Phillips curve model over the pre-COVID sample for each inflation measure. As shown in the table, each median measure exhibits a statistically significant Phillips curve relationship. The β_j estimates also clearly indicate that for a given non-shelter split, the greater the shelter disaggregation, the weaker the estimated Phillips curve relationship, although the differences in the estimates are marginal. For example, the estimated β_j for median FRBC-OER4 inflation is -0.408, for FRBC-OER8 it is -0.392, and for FRBC-OER8-RENT8 it is -0.370. While our results show that the level of shelter disaggregation is inversely related to the strength of the Phillips curve relationship among the Median CPI candidates, this inverse relationship is *not* driven by a reduction in the percentage of time a shelter component is chosen as the median. Recall from Section 6.2 that upon increasing the level of shelter disaggregation by disaggregating Rent into 8 components, a shelter component was (typically) chosen *more* frequently as the median - and yet, moving from OER8 rows to the OER8-RENT8 row

[30]Similar to Zaman (2019), to account for the possibility of serial correlation in the regression residuals, we compute Newey-West standard errors. The lag length is set equal to $(4*(T/100)^{2/9})$, where T refers to the size of the estimation sample.

always results in a less strong Phillips curve relationship. The results are qualitatively similar for the Trimmed-Mean CPI and are reported in Appendix A.1.3.

Table 3: Estimated Phillips Curve Slope

Series	Median CPI			
	Sample: 2000-2019		Sample: 2000-2024	
	Beta	p-value	Beta	p-value
FRBC-OER4	-0.271	0.00	-0.408	0.00
FRBC-OER8	-0.263	0.00	-0.392	0.00
FRBC-OER8-RENT8	-0.241	0.00	-0.370	0.00
C2-OER4	-0.272	0.00	-0.408	0.00
C2-OER8	-0.251	0.00	-0.390	0.00
C2-OER8-RENT8	-0.227	0.00	-0.359	0.00
C3-OER4	-0.290	0.00	-0.432	0.00
C3-OER8	-0.282	0.00	-0.416	0.00
C3-OER8-RENT8	-0.251	0.00	-0.380	0.00
C4-OER4	-0.305	0.00	-0.450	0.00
C4-OER8	-0.278	0.00	-0.415	0.00
C4-OER8-RENT8	-0.250	0.00	-0.382	0.00
C7-OER4	-0.308	0.00	-0.451	0.00
C7-OER8	-0.283	0.00	-0.422	0.00
C7-OER8-RENT8	-0.250	0.00	-0.384	0.00

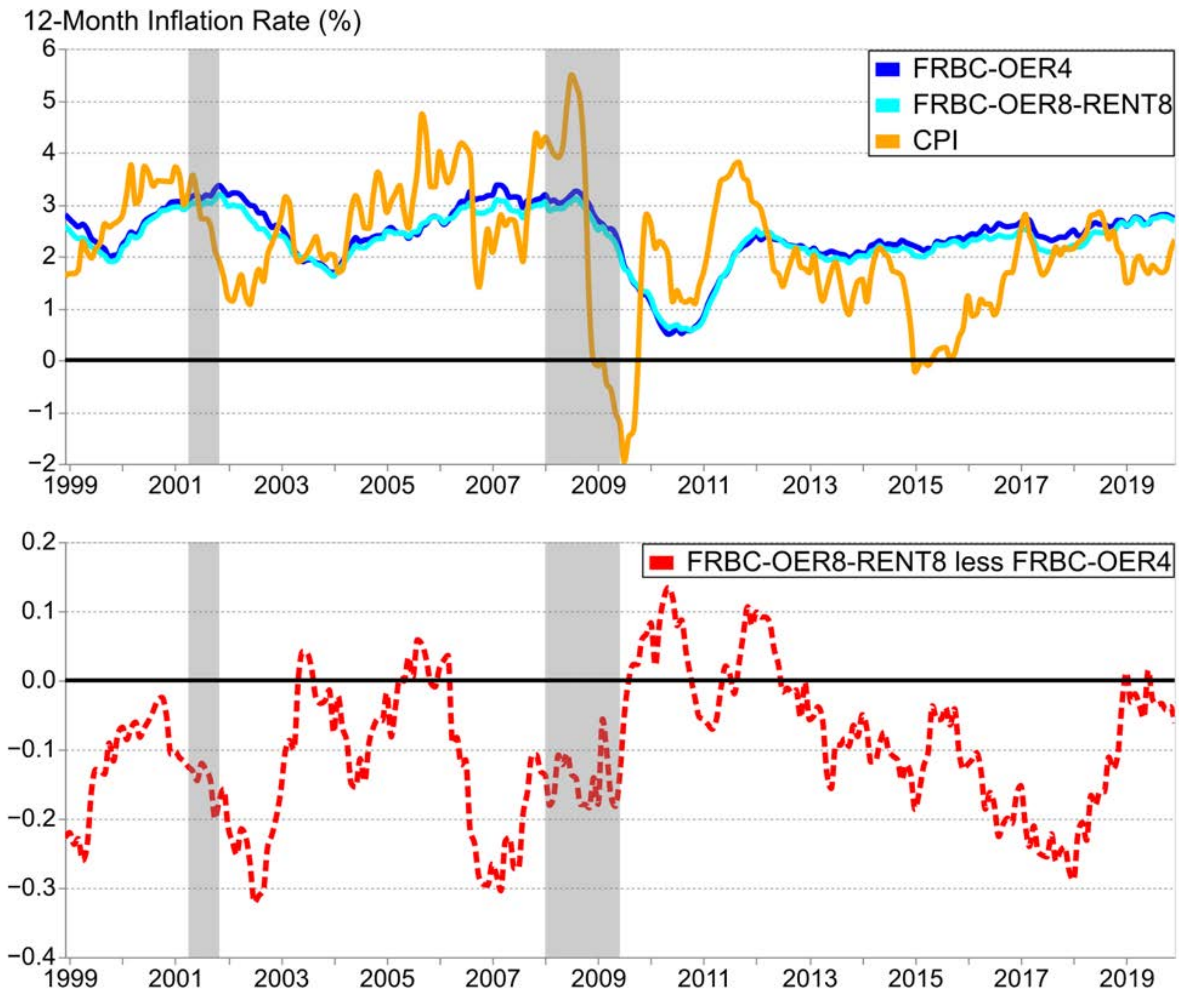
Note: The estimates shown are for two different estimation samples: 2000M1 through 2019M12 (denoted Sample: 2000-2019) and 2000M1 through 20204M11 (denoted Sample: 2000-2024). The data from 1999M1 through 1999M12 are used to compute the lagged value of the unemployment rate gap.

6.3 Historical Trend in CPI Inflation

To close this section, we explore a practical implication of our proposed finer disaggregation of CPI components for median CPI inflation: historically, how would inferences about medium-run trend inflation have changed, when using the FRBC-OER8-RENT8 measure versus the baseline FRBC-OER4?

In Figure 14, we compare the evolution of both median measures over our pre-pandemic sample. During most of the period, median FRBC-OER8-RENT8 inflation was consistently lower than median FRBC-OER4 inflation, at times running about 30 basis points (bps) below FRBC-OER4; a gap of this size is considered meaningful in monetary policy discussions. We also find that FRBC-OER8-RENT8 showed less variability over this time, with a standard deviation 8% lower than that of FRBC-OER4. Our impression is that when the gap between the two measures is notable, FRBC-OER8-RENT8 is giving a modestly superior reading of the trend in inflation. For instance, in the years prior to the COVID-19 pandemic, FRBC-OER8-RENT8 was more in line with the views of economists and monetary policymakers at the time, namely, that medium-run trend inflation was more subdued than FRBC-OER4 suggested.

Figure 14: Comparing FRBC-OER4 and FRBC-OER8-RENT8 Measures of Median CPI Inflation, Pre-Pandemic Sample



In Figure 15, we compare the evolution of both median measures over our post-pandemic sample.

Our impression about the superiority of the alternative measure is the same. Notably, the FRBC-OER8-RENT8 median provided more of an early warning, relative to FRBC-OER4, in early 2021 about the surge in CPI inflation that was about to arrive in Q2 of 2021. By March 2021, median FRBC-OER8-RENT8 exceeded median FRBC-OER4 by just over 20 bps, the largest gap since the aftermath of the Great Recession. Thereafter, this gap increased to a high of 40 bps in September 2021, just as CPI inflation was exceeding 5%. Conversely, as CPI inflation fell back to more normal levels in 2023 and 2024 after hitting a post-pandemic peak of 9% in June 2022, median FRBC-OER8-RENT8 indicates a more rapid and substantial easing of underlying inflationary pressures than FRBC-OER4, with the gap between the two hitting a record low of nearly -40 bps when CPI inflation was hovering at just over 3% in early 2024.

Still, one might wonder: why did the gap between the FRBC-OER8-RENT8 measure and headline inflation remain so large for so long over, for example, the mid-2021 to late 2022 period? Does this gap represent a failure of the Median CPI? Our emphatic answer is: No. Median CPI should not, and does not, track the CPI when outliers are driving the CPI. To illustrate this, consider just two months, June 2021 and June 2022. [31] In June 2021, one item, public transportation, experienced an inflation rate of 59%, a clear outlier (the next highest category came in at 29%, itself a long way away from the bulk of the observations, which were clustered between -5% and +9%). But public transportation was not the biggest outlier. Used cars and trucks came in at 133%, while car and truck rental came in at nearly 300%! The latter two categories have a combined aggregation weight of just 3.1% in the CPI basket, yet those two categories *alone* lifted the CPI in that month by 4.4 percentage points (at an annual rate). The monthly official Median CPI reading that month was 4.2%; headline CPI came in at 10.8%. In June 2022, overall inflation had risen notably. Monthly inflation peaked that month, coming in at 16.7%; few components read less than +3%. In that month, even excluding outliers, there was a pronounced rightward skew: a notable number of categories experienced inflation rates exceeding 11%. The Median CPI came in at 7.2%, indicating that half of the items in the CPI basket (by aggregation weight) experienced inflation rates at or below that level. Still, just three categories–motor fuel, public transportation, and fuel oils and other fuels, coming in at 61%, 169% and 339%–lifted headline CPI by 4.6 percentage points.

[31] In Appendix B, we provide a discussion of the sensitivity of the median versus the mean to outliers, along with providing some illustrative examples taken from CPI data in the post-2020 period.

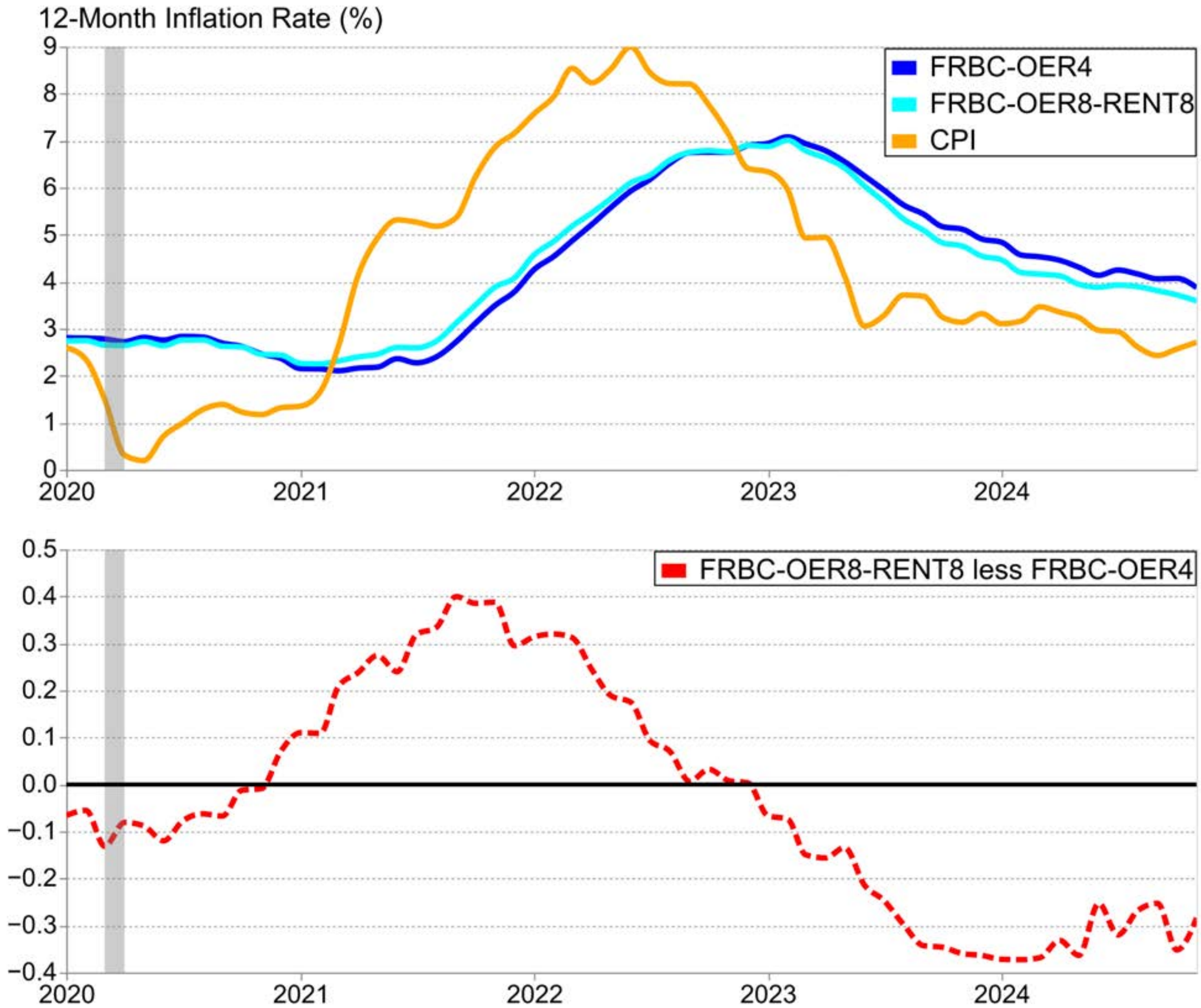
Dropping those three categories (and reweighting), the CPI reading would have been 9.8%. In both of these episodes, the median CPI clearly indicated that outliers were exerting an undue influence on headline CPI.[32] Inflation rose during the COVID-19 pandemic, but outliers have played a dominant role over the entire period of mid-2021 to the present.

That said, our analysis demonstrates that FRBC-OER8-RENT8 median CPI inflation can systematically and persistently diverge across time from the FRBC-OER4 baseline. Thus, the FRBC-OER8-RENT8 inflation measure may offer better insights into trend inflation dynamics that diverge from those provided by the existing baseline.[33]

[32]"Core" CPI does not reliably signal when outliers are driving the CPI: in June 2022, core CPI was 8.4%, fairly close to Median CPI, correctly indicating that outliers (energy) were huge drivers of headline CPI that month; but in June 2021, core CPI came in at 9.8%, providing no indication that outliers were driving headline CPI.

[33]Qualitative results for C2-OER8-RENT8 are similar. FRBC-OER4, FRBC-OER8-RENT8, C2-OER8-RENT8, and the rest of our median inflation rates may be downloaded from the Online Appendix.

Figure 15: Comparing FRBC-OER4 and FRBC-OER8-RENT8 Measures of Median CPI Inflation, Post-Pandemic Sample



7 Conclusion

The Median CPI measure developed by Federal Reserve Bank of Cleveland (FRBC) researchers is a well-known estimator of the medium-term trend (MTT) in CPI inflation. Over time, various improvements to this measure have been implemented. Historically, revisions to the underlying methodology have usually involved increasing the level of disaggregation; and whenever further disaggregation has been investigated, it has always improved the performance of the Median CPI. Moreover, *less* disaggregation, in the limit, leads to an index identical to headline CPI, suggesting that more disaggregation should be better. Finally, the current level of disaggregation in the Median CPI is far lower than that used in, e.g., the Median PCE. It may seem obvious that increasing the level of disaggregation would improve the performance of the Median CPI—and indeed, that increasing the level of disaggregation by as much as possible would result in the highest performance gains. But this paper provides evidence to the contrary.

We first demonstrate theoretically, in Proposition 1, that the minimum mean squared error estimator of the median of the underlying distribution need not be the one associated with the most disaggregated basket that is feasible. Hence, the optimal level of disaggregation must be established empirically.

We conduct this analysis using criteria well-established in the literature, focusing on two criteria in particular: accuracy vis-a-vis medium-term movements in CPI inflation, and predictive power for future movements in headline CPI inflation at various horizons. We systematically investigate the impact of further disaggregation by constructing 15 distinct median CPI inflation measures with varying levels of disaggregation.

For reducing bias and tracking the ex-post MTT in inflation, we find improvements in accuracy from disaggregating the shelter components (OER and Rent) by as much as possible. However, in accordance with Proposition 1, we find that for non-shelter components, increasing the level of disaggregation–at least once one goes beyond a small amount–generally worsens accuracy. This demonstrates empirically that increasing disaggregation does *not* always further enhance these measures. Our predictive accuracy findings are congruent: increasing housing disaggregation by as much as possible, but increasing non-shelter component disaggregation only a small amount at most, generally yields marginal predictive accuracy gains.

We further show that increasing the level of shelter disaggregation does not necessarily increase the frequency of choosing a non-shelter component as the median. And we show that similar to the official FRBC median measure, all median CPI measures we considered exhibit a statistically significant relationship with labor market slack. Contrary to popular belief, there is not a simple relationship between the frequency with which a shelter component is selected as a median, and the strength of the resulting Phillips curve.

We end the paper by exploring a very practical question: over the historical period, would inferences about the medium-term trend in inflation have been different, had the FRBC-OER8-Rent8 Median CPI been in use? Our answer is that our new measure is modestly superior. First, we show that prior to the COVID-19 pandemic, our preferred median measure, derived from our FRBC-OER8-RENT8 split of CPI, was somewhat more in line with the views of economists and monetary policymakers at the time, namely, that medium-run trend inflation was more subdued than the headline inflation rate suggested. During the COVID-19 pandemic and beyond, our preferred measure more closely tracked the medium-run trend in CPI inflation, rising more quickly in 2021 than the official measure, and falling more rapidly starting in early to mid-2023.

The optimal level of disaggregation for a limited-influence estimator has received almost no attention in the literature (though prior work has, in a specific case, demonstrated the usefulness of increasing the level of disaggregation starting from a modestly sized baseline level). As noted above, our study demonstrates, for the first time, that maximal disaggregation is not necessarily optimal; instead, the optimal level of disaggregation must be investigated empirically. This is an impetus for revisiting other measures, such as the Trimmed-Mean and Median PCE measures, which are currently built on a very high level of disaggregation (on the order of 200 components).

References

[1] Adams, Brian, & Randal Verbrugge. 2021. Location, Location, Structure Type: Rent Divergence within Neighborhoods. *Federal Reserve Bank of Cleveland Working Paper* 21-03. https://doi.org/10.26509/frbc-wp-202103

[2] Alvarez, Rocio, Maximo Camacho, & Gabriel Perez-Quiro (2016). Aggregate versus disaggregate information in dynamic factor models. *International Journal of Forecasting* Volume 32(3): 680-694. https://doi.org/10.1016/j.ijforecast.2015.10.006

[3] Alves, Felipe Alduino (2014). Underlying inflation in a DSGE Model. *PhD Dissertation, Departamento de Economia, PUC-Rio. 37298.PDF (puc-rio.br)*

[4] Andrle, Michal & Jan Brůha, & Serhat Solmaz (2017). On the sources of business cycles: implications for DSGE models. *Working Paper Series 2058, European Central Bank*

[5] Ashley, Richard & Randal Verbrugge (2025). The Intermittent Phillips Curve: Finding a Stable (But Persistence-Dependent) Phillips Curve Model Specification. Forthcoming, *Economic Inquiry.*

[6] Ashley, Richard & Haichun Ye (2012). On the Granger causality between median inflation and price dispersion. *Applied Economics* 44:32, 4221-4238. DOI: 10.1080/00036846.2011.587788

[7] Ball, Laurence M., Carlos Carvalho, Christopher Evans, & Luca Antonio Ricci (2023). Weighted Median Inflation Around the World: A Measure of Core Inflation. *National Bureau of Economic Research* Working Paper No. w31032. https://doi.org/10.3386/w31032

[8] Ball, Laurence M., Daniel Leigh, & Prachi Mishra (2022). Understanding U.S. Inflation During the COVID Era. *IMF working papers* No. 2022/208.

[9] Ball, Laurence M., Daniel Leigh, Prachi Mishra, and Antonio Spilimbergo (2021). Measuring US core inflation: the stress test of COVID-19. *National Bureau of Economic Research* Working Paper No. 29609. https://doi.org/10.3386/w29609

[10] Ball, Laurence M., & Sandeep Mazumder (2011). Inflation dynamics and the great recession. *Brookings Papers on Economic Activity* SPRING 2011, pp. 337-405. https://doi.org/10.1353/eca.2011.0005

[11] Ball, Laurence M., & Sandeep Mazumder (2019a). The Nonpuzzling Behavior of Median CPI Inflation. *NBER Working Paper* 25512. https://doi.org/10.3386/w25512

[12] Ball, Laurence M., & Sandeep Mazumder (2019b). A Phillips curve with anchored expectations and short-term unemployment. *Journal of Money, Credit and Banking* 51(1): 111-137. https://doi.org/10.1111/jmcb.12502

[13] Boivin, Jean, & Serena Ng. (2006). Are more data always better for factor analysis? *Journal of Econometrics* 132, pp. 169-194. https://doi.org/10.1016/j.jeconom.2005.01.027

[14] Brischetto, Andrea, & Anthony Richards (2007). The Performance of Trimmed Mean Measures of Underlying Inflation. Paper presented at the Conference on Price Measurement for Monetary Policy, May 24-25, 2007, jointly sponsored by the Federal Reserve Banks of Cleveland and Dallas.

[15] Bryan, Michael F., & Christopher J. Pike (1991). Median Price Changes: An Alternative Approach to Measuring Current Monetary Inflation. *Federal Reserve Bank of Cleveland Economic Commentary* December 1. https://doi.org/10.26509/frbc-ec-19911201

[16] Bryan, Michael F., & Stephen G. Cecchetti (1994). Measuring Core Inflation. *National Bureau of Economic Research Studies in Business Cycles, vol. 29, Monetary Policy, ed. N. Gregory Mankiw (Chicago: University of Chicago Press)* pages 195–215.

[17] Bryan, Michael F., & Stephen G. Cecchetti (1999). Inflation and the Distribution of Price Changes. *The Review of Economics and Statistics* 81 (2): 188–196. https://doi.org/10.1162/003465399558148

[18] Bryan, Michael F., Stephen Cecchetti, & Rodney Wiggins (1997). Efficient Inflation Estimation. *NBER Working Paper* Number 6183. doi:10.3386/w6183

[19] Carlomagno, Guillermo, Jorge Fornero, & Andres Sansone (2023). A Proposal for Constructing and Evaluating Core Inflation. *Latin American Journal of Central Banking* 4:100094. https://doi.org/10.1016/j.latcb.2023.100094

[20] Carroll, Daniel, & Randal Verbrugge (2019). Behavior of a New Median PCE Measure: A Tale of Tails. *Federal Reserve Bank of Cleveland Economic Commentary* Number 2019-10. https://doi.org/10.26509/frbc-ec-201910

[21] Cecchetti, Stephen G., & Erica L. Groshen (2001). Understanding inflation: implications for monetary policy. Palgrave Macmillan UK. https://doi.org/10.1057/9780333992753$_7$

[22] Clark, Todd E. (2001). Comparing Measures of Core Inflation. *Federal Reserve Bank of Kansas City Economic Review* Vol. 86, no. 2, pp. 5-31.

[23] Cotton, Christopher, Vaishali Garga, Giovanni Olivei, & Viacheslav Sheremirov (2023). The Distribution of Sectoral Price Changes and Recent Inflation Developments. *Federal Reserve Bank of Boston Research Paper Series Current Policy Perspectives Paper* 96665

[24] Crone, Theodore, N. Neil K. Khettry, Loretta J. Mester, & Jason A. Novak (2013). Core Measures of Inflation as Predictors of Total Inflation. *Journal of Money, Credit and Banking* 45(2-3):505-519. doi:10.1111/jmcb.12013

[25] De Veirman, Emmanuel (2023). How does the Phillips curve slope vary with repricing rates? *ECB Working Paper Series* No 2804.

[26] Diebold, F.X., & R.S. Mariano. 1995. Comparing predictive accuracy *Journal of Business Economic Statistics* 13:253-263. https://doi.org/10.1198/073500102753410444

[27] Dolmas, Jim (2005). Trimmed Mean PCE Inflation. *Federal Reserve Bank of Dallas Working Paper* Number 05-06.

[28] Dolmas, Jim (2009). The 2009 Revision to the Trimmed Mean PCE Inflation Series. Technical note. Federal Reserve Bank of Dallas.

[29] Dolmas, Jim, & Evan F. Koenig (2019). Two Measures of Core Inflation: A Comparison. Working Paper 1903. https://doi.org/10.24149/wp1903

[30] Dolmas, Jim, & Mark A. Wynne (2008). Measuring Core Inflation: Notes from a 2007 Dallas Fed Conference. Federal Reserve Bank of Dallas Staff Papers, 4.

[31] Esquivel-Monge, Manfred, Adolfo Rodríguez-Vargas, &osé Pablo Vásquez-Carvajal. 2011. Trimmed Means of the CPI as an Indicator of Core Inflation for Costa Rica." Banco Central de Costa Rica *Documento de Trabajo* No 001/2011.

[32] Fang, W., S.M. Miller, & C-C. Yeh (2010). Does a threshold inflation rate exist? Quantile inferences for inflation and its variability. *Empirical Economics* 39, 619–641. https://doi.org/10.1007/s00181-009-0327-1

[33] Federal Reserve Bank of Cleveland (2007). Methodological Adjustments to the Median and 16 Percent Trimmed-Mean CPI Estimators. https://www.clevelandfed.org/en/indicators-and-data/median-cpi#background

[34] Gamber, Edward N., & Julie K. Smith (2019). Constructing and evaluating core inflation measures from component-level inflation data. *Journal of Forecasting*- 38 (8), 833-852. https://doi.org/10.1002/for.2595.

[35] Goulet Coulombe, Philippe, Karin Klieber, Christophe Barrette & Maximilian Gobel. 2024. Maximally Forward-Looking Core Inflation. *arXiv preprint* arXiv:2404.05209 (2024). https://doi.org/10.48550/arXiv.2404.05209

[36] Harvey D., S. Leybourne, & P. Newbold. 1997. Testing the equality of prediction mean squared errors *International Journal of Forecasting* 13:281-291. https://doi.org/10.1016/S0169-2070(96)00719-4

[37] Higgins, Amy, & Randal Verbrugge. 2015. Is a Nonseasonally Adjusted Median CPI a Useful Signal of Trend Inflation? *Federal Reserve Bank of Cleveland Economic Commentary* Number 2015-13. doi:10.26509/frbc-ec-201513

[38] Jonansen, Morten & Einar Woien Nordbo. 2006. Indicators of underlying inflation in Norway. Norges Bank *Economic Bulletin* 4/06 (Vol. 77), 195-204. http://hdl.handle.net/11250/2504291

[39] Johansson, Jesper, Mårten Löf, Oliver Sigrist & Oskar Tysklind. 2018. Measures of core inflation in Sweden. *Sveriges Riksbank Economic Commentaries* No. 11 2018.

[40] Kilian, Lutz (2008). A comparison of the effects of exogenous oil supply shocks on output and inflation in the G7 countries. *Journal of the European Economic Association* 6, (1): 78-121. https://doi.org/10.1162/JEEA.2008.6.1.78

[41] Kishor, N. Kundan, & Evan F. Koenig (2022). Finding a Role for Slack in Real-Time Inflation Forecasting. *International Journal of Central Banking* Volume 18, (2): 245-282

[42] Lao, Helen, & Ceciline Steyn. (2019). A comprehensive evaluation of core inflation for Canada: an update. *Bank of Canada Staff Discussion Paper* No. 2019-9. https://doi.org/10.34989/sdp-2019-9

[43] Liu, Dandan, & Julie K. Smith (2014). Inflation forecasts and core inflation measures: Where is the information on future inflation? *The Quarterly Review of Economics and Finance* 54(1): 133-137. https://doi.org/10.1016/j.qref.2013.07.006

[44] Mahedy, Tim, & Adam Shapiro (2017). What's Down with Inflation? Federal Reserve Bank of San Francisco, *Economic Letter* 2017-35.

[45] Marques, Carlos Robalo, Pedro Duarte Neves, & Luis Morais Sarmento (2003). Evaluating core inflation indicators. *Economic Modelling* 20(4): 765-775. https://doi.org/10.1016/S0264-9993(02)00008-1.

[46] Mazumder, Sandeep (2017). Output gains from accelerating core inflation. *Journal of Macroeconomics* 51:63-74. https://doi.org/10.1016/j.jmacro.2016.12.001.

[47] Mazumder, Sandeep (2018). Inflation in Europe after the Great Recession. *Economic Modelling* 71:202-213. https://doi.org/10.1016/j.econmod.2017.12.014.

[48] Mertens, Elmar (2016). Measuring the level and uncertainty of trend inflation. *Review of Economics and Statistics* 98(5): 950-967. https://doi.org/10.1162/REST$_{a0}$0549

[49] Metiu, Norbert, & Esteban Prieto (2023). The Macroeconomic Effects of Inflation Uncertainty. http://dx.doi.org/10.2139/ssrn.4381311

[50] Meyer, Brent H., & Mehmet Pasaogullari (2010). Simple Ways to Forecast Inflation: What Works Best? *Federal Reserve Bank of Cleveland Economic Commentary* 2010-17. https://doi.org/10.26509/frbc-ec-201017

[51] Meyer, Brent, & Saeed Zaman (2019). The Usefulness of the Median CPI in BVARs used for Macroeconomic Forecasting and Policy. *Empirical Economics* 57(2):603-630. https://doi.org/10.1007/s00181-018-1472-1

[52] Meyer, Brent, Guhan Venkatu, & Saeed Zaman (2013). Forecasting Inflation? Target the Middle. *Federal Reserve Bank of Cleveland Economic Commentary* Number 2013-05. https://doi.org/10.26509/frbc-ec-201305

[53] Munoz-Salas, Evelyn, & Adolfo Rodríguez-Vargas. 2021. Monetary policy in Costa Rica: an assessment based on the neutral real interest rate. *Centro de Estudios Monetarios LantinoAmericanos Joint Research Paper* 2021-01.

[54] Nickel, Christiane, Annette Fröhling, Caroline Willeke, Giordano Zevi, Chiara Osbat, Ioannis Ganoulis, Gerrit Koester, et al. (2021). Inflation Measurement and its Assessment in the ECB's Monetary Policy Strategy Review. *ECB Strategy Review 2020-21 Occasional Papers No. 265* https://www.ecb.europa.eu/pub/pdf/scpops/ecb.op265 a3fb0b611d.en.pdf?82b5ee80f9a840af58de447a1e25f86e

[55] Norges Bank (2025). Inflation Indicators. https://www.norges-bank.no/en/topics/Statistics/Inflation-indicators/, accessed 2/19/2025.

[56] O'Brien, Derry, and Christiane Nickel (2018). The ECB's measures of underlying inflation for the euro area. VoxEU. https://cepr.org/voxeu/columns/ecbs-measures-underlying-inflation-euro-area

[57] Ocampo, Sergio, Raphael Schoenle, & Dominic A. Smith (2023). Extending the Range of Robust PCE Inflation Measures. Working Paper arXiv:2207.12494. https://doi.org/10.48550/arXiv.2207.12494

[58] Reserve Bank of Australia (2025). Inflation and its measurement. https://www.rba.gov.au/education/resources/explainers/inflation-and-its-measurement.html, accessed 2/14/2025.

[59] Reserve Bank of New Zealand (2024). Monetary Policy Statement, February 2024. https://www.rbnz.govt.nz/hub/publications/monetary-policy-statement/2024/monetary-policy-statement-february-2024, accessed 2/14/2025.

[60] Rich, Robert, and Charles Steindel (2007). A Comparison of Measures of Core Inflation. *Federal Reserve Bank of New York Economic Policy Review December*. doi:10.2139/ssrn.1072923

[61] Rich, Robert W., Randal J. Verbrugge, & Saeed Zaman (2022). Adjusting Median and Trimmed-Mean Inflation Rates for Bias Based on Skewness. *Federal Reserve Bank of Cleveland Economic Commentary* Number 2022-05. https://doi.org/10.26509/frbc-ec-202205

[62] Richards, Anthony, J. (2024). The Use of Trimmed-Mean Measures of Inflation in Monetary Policy Making: Some Differences between Old and New World Central Banks. *SSRN*. doi:10.2139/ssrn.4971521

[63] Schembri, Lawrence. (2017). Getting to the core of inflation. Remarks at the Department of Economics, Western University, London Ontario. https://www.bankofcanada.ca/2017/02/getting-core-inflation/, accessed Feb. 18, 2025.

[64] Silver, Mick. (2007) Core inflation: measurement and statistical issues in choosing among alternative measures. *IMF Staff Papers* 54(1), 163-190. https://doi.org/10.1057/palgrave.imfsp.9450006

[65] Smith, Julie K. (2004). Weighted median inflation: is this core inflation? *Journal of Money, Credit and Banking* 36(2):253-263. https://doi.org/10.1353/mcb.2004.0014

[66] Stock, James H., & Mark W. Watson (2016). Core Inflation and Trend Inflation. *Review of Economics and Statistics* 98(4):770–784. doi:10.1162/REST_a_00608.

[67] Stock, James H., & Mark W. Watson (2020). Slack and cyclically sensitive inflation. *Journal of Money, Credit and Banking* 52, no. S2:393-428. https://doi.org/10.1111/jmcb.12757

[68] Swiss National Bank (2025). Glossary. https://www.snb.ch/en/services-events/digital-services/glossary, accessed 2/19/2025.

[69] Verbrugge, Randal (1999). Cross-sectional inflation asymmetries and core inflation: A comment on Bryan and Cecchetti *The Review of Economics and Statistics* 81, no. 2:199-202. https://doi.org/10.1162/003465399558166

[70] Verbrugge, Randal (2022). Is it time to reassess the focal role of core PCE inflation in assessing the trend in PCE inflation? *Economia* 45, no. 89:73-101. https://doi.org/10.18800/economia.202201.004

[71] Verbrugge, Randal J., & Saeed Zaman (2021). Whose Inflation Expectations Best Predict Inflation? *Federal Reserve Bank of Cleveland Economic Commentary*, Number 2021-19. https://doi.org/10.26509/frbc-ec-202119

[72] Verbrugge, Randal J., & Saeed Zaman (2023). The hard road to a soft landing: Evidence from a (modestly) nonlinear structural model. *Energy Economics* 123: 106733. https://doi.org/10.1016/j.eneco.2023.106733

[73] Verbrugge, Randal J., & Saeed Zaman (2024). Improving Inflation Forecasts Using Robust Measures. *International Journal of Forecasting*, Vol 40(2):735-745. https://doi.org/10.1016/j.ijforecast.2023.05.003

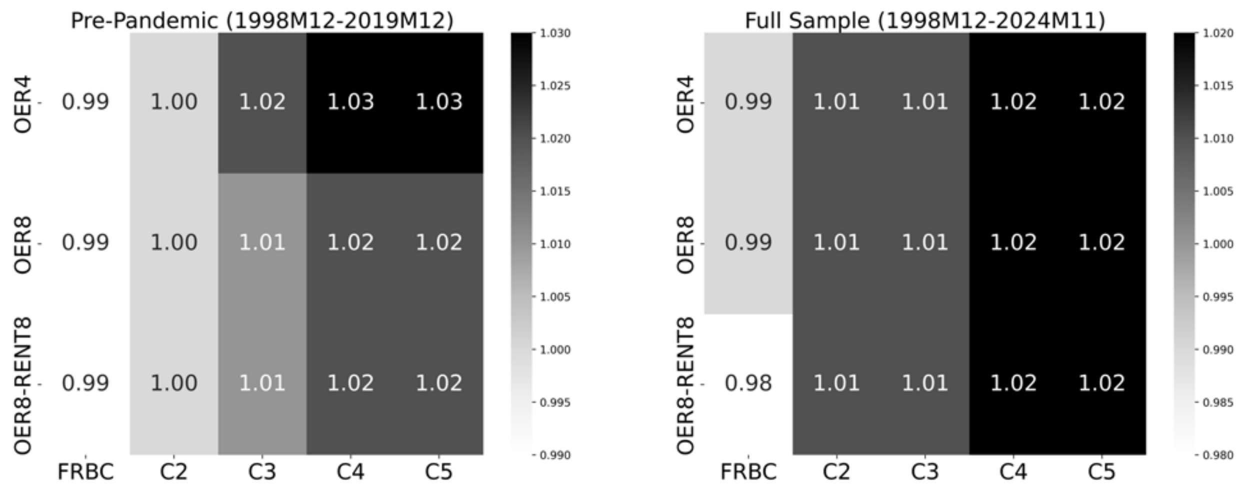
[74] Zaman, Saeed (2019). Cyclical versus Acyclical Inflation: A Deeper Dive. *Federal Reserve Bank of Cleveland Economic Commentary* doi:10.26509/frbc-ec-201913

A Appendix 1: Trimmed-Mean CPI Results

A.0.1 Accuracy in mean

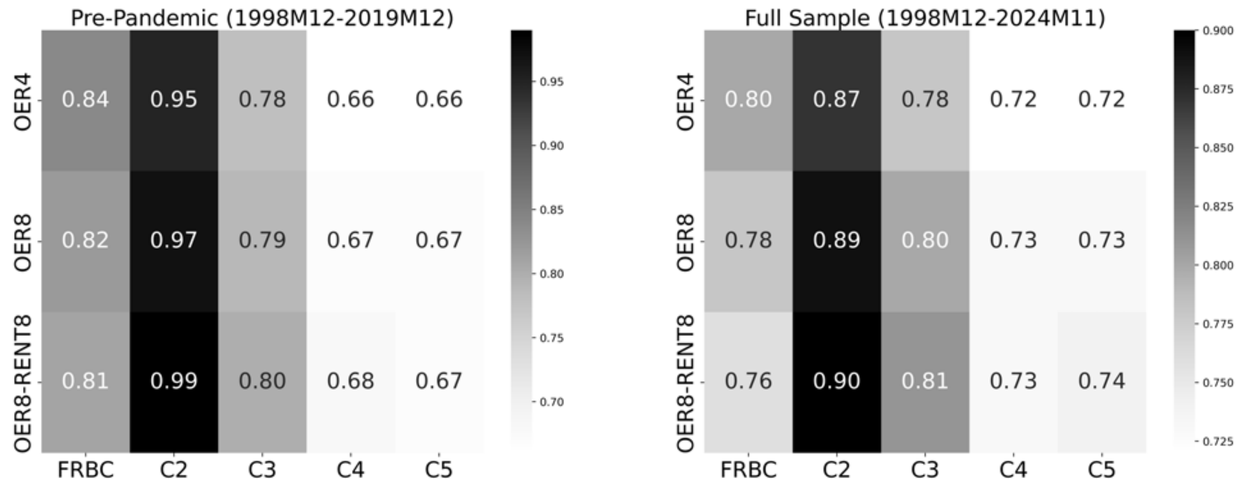
In Figure A.1, for each trimmed-mean inflation candidate, we report the ratio of the average trimmed-mean inflation rate and the average CPI inflation rate. All ratios are approximately equal to 1. In Figure A.2 we report p-values for t-tests of equality of means. Results show that the observed differences between average inflation in each candidate trimmed-mean measure and headline CPI inflation are not statistically significant

Figure A.1: Mean of Trimmed-Mean Inflation Measures Relative to Mean of CPI Inflation



Notes: Reported figures are the ratio of the average of the trimmed-mean inflation rate and the average of CPI inflation. Both averages are computed as the mean of 12-month inflation rates, measured by percent changes, over the indicated period. Darker shading indicates higher values, while lighter shading indicates lower values.

Figure A.2: p-Values of a Statistical Test of Equal Mean Relative to CPI Inflation

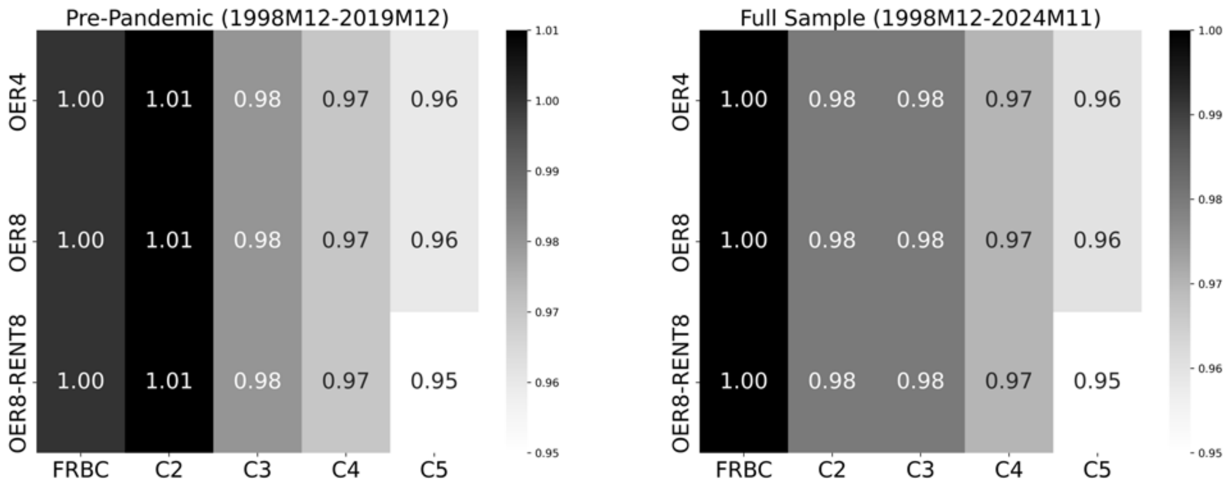


Notes: Reported figures are the p-values of a t-test of $H_0 : \mathbb{E}[\pi_j^c] = \mathbb{E}[CPI]$, where π_j^c denotes the jth candidate trimmed-mean inflation measure. The p-value is obtained by taking the difference of each trimmed-mean inflation measure from CPI inflation, and regressing this against a constant. The test statistic of the constant term is calculated using heteroskedasticity-and-autocorrelation-consistent (HAC) standard errors with a small sample correction and $\lfloor 4[T/100]^{2/9} \rfloor$ lags, where T refers to the size of the estimation sample. Darker shading indicates higher values, while lighter shading indicates lower values.

A.0.2 Accuracy versus a standard ex-post MTT estimate

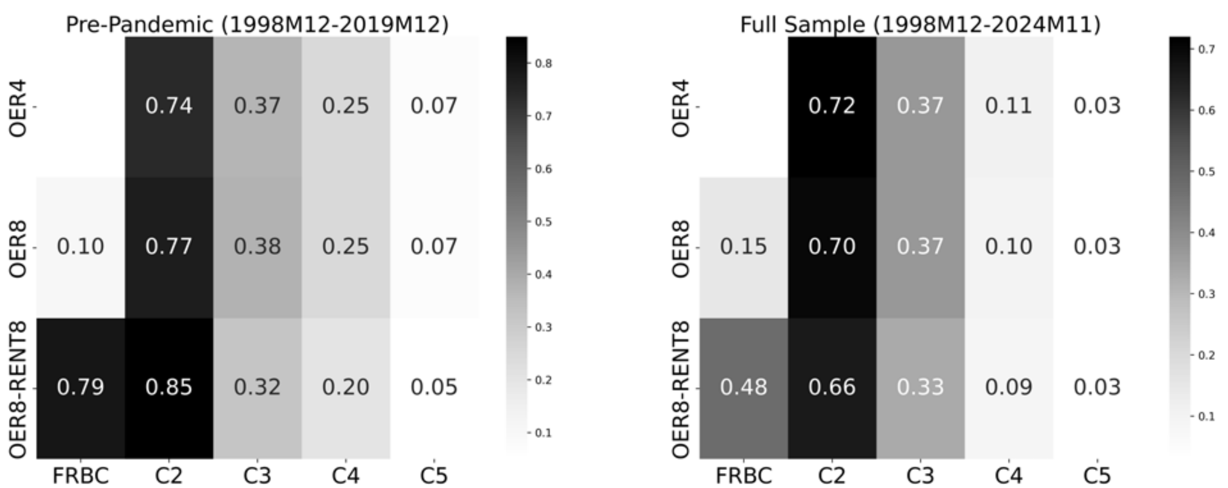
In Figure A.3, we report the RMSE of each measure of trimmed-mean inflation against a 37-month centered moving average (37MMA) of 12-month CPI inflation relative to the RMSE of the baseline trimmed-mean FRBC-OER4 inflation measure. C5-OER8-RENT8 outperforms, with an RMSE that is 5% lower than that of trimmed-mean FRBC-OER4 inflation pre-pandemic and in the full sample. DM p-values reported in Figure A.4 show that the observed reduction in the RMSE of C5-OER8-RENT8 relative to FRBC-OER4 is statistically significant at the 10% level in the pre-pandemic sample and at the 5% level in the full sample.

Figure A.3: $RMSE(\hat{\pi}^{37MMA} - \hat{\pi}_j)$ of Trimmed-Mean Inflation Measures Relative to FRBC-OER4



Notes: Reported figures are the RMSE of deviations of the trimmed-mean inflation measure from a 37-month centered moving average of CPI Inflation, divided by the same for trimmed-mean FRBC-OER4 inflation. In the pre-pandemic sample, the moving average is computed using CPI inflation through December 2019 only. Darker shading indicates higher values, while lighter shading indicates lower values.

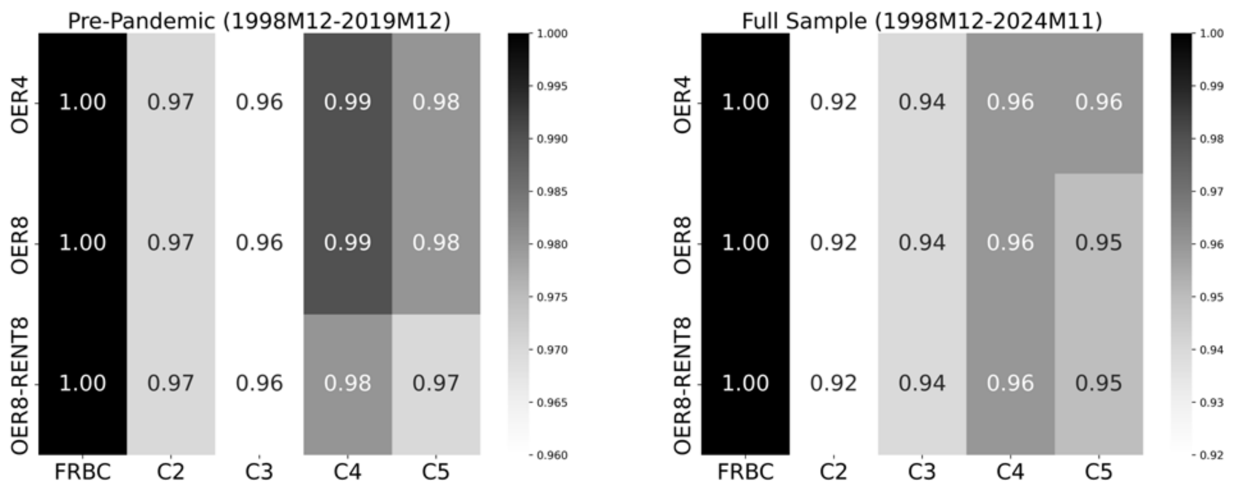
Figure A.4: p-Values of a Statistical Test of Equal Ability in Tracking $\hat{\pi}^{37MMA}$ for Trimmed-Mean Inflation Measures, Relative to FRBC-OER4



Notes: Reported figures are the p-values of a Diebold-Mariano (1995) test that $RMSE(\bar{\pi}^{37MMA} - \pi^c_{FRBC-OER4})$ and $RMSE(\bar{\pi}^{37MMA} - \pi^c_j)$ are equal, where j denotes the jth candidate trimmed-mean inflation measure. The p-value is obtained by taking the difference of the two squared errors series $e^{37MMA}_{j,t}$ and $e^{37MMA}_{FRBC-OER4,t}$, and regressing the resulting series against a constant. The test statistic of the constant term is then calculated using HAC standard errors with a small sample correction and $\lfloor 4[T/100]^{2/9} \rfloor$ lags, where T refers to the size of the estimation sample. Darker shading indicates higher values, while lighter shading indicates lower values.

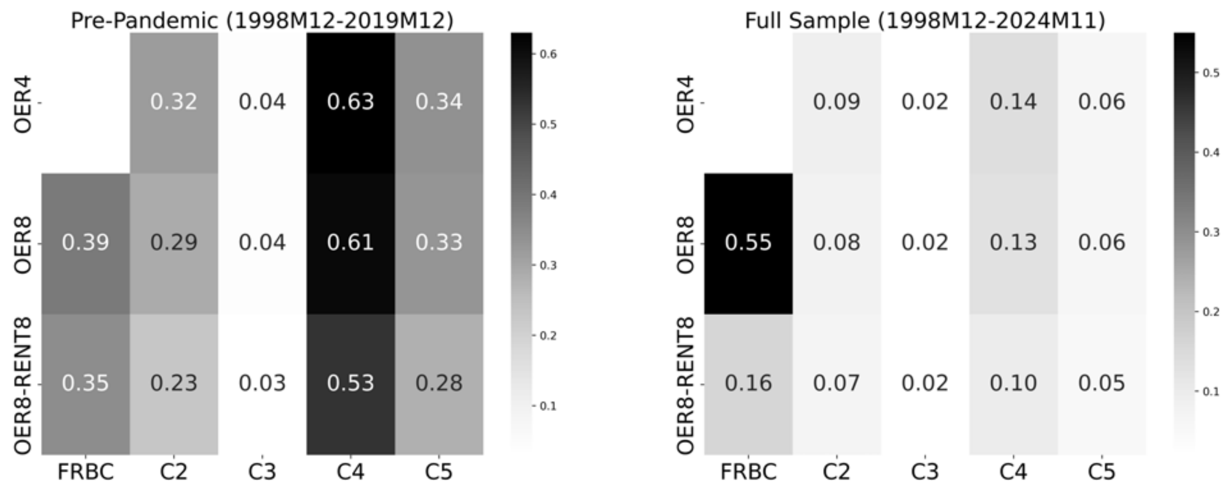
In Figure A.5, we report the RMSE of each trimmed-mean inflation measure against the 2SMA trend relative to the same for trimmed-mean FRBC-OER4 inflation, and in Figure A.6 we report for each trimmed-mean candidate the p-values of the DM test of equal ability in tracking the 2SMA trend estimate against the FRBC-OER4 baseline. In contrast to results using the 37MMA trend, C3 splits perform best in the pre-pandemic sample, achieving a 4% reduction in RMSE relative to the FRBC-OER4 baseline, while C2 splits perform best in the full sample, achieving an 8% reduction in RMSE, both of which are statistically significant at the 5% level.

Figure A.5: $RMSE(\hat{\pi}^{2SMA} - \hat{\pi}_j)$ of Trimmed-Mean Inflation Measures Relative to FRBC-OER4



Notes: Reported figures are the RMSE of deviations of the trimmed-mean inflation measure from a two-stage centered moving average (2SMA) of CPI inflation, divided by the same for trimmed-mean FRBC-OER4 inflation. In the pre-pandemic sample, the moving average is computed using CPI inflation through December 2019 only. Darker shading indicates higher values, while lighter shading indicates lower values.

Figure A.6: p-Values of a Statistical Test of Equal Ability in Tracking $\hat{\pi}^{2SMA}$ for Trimmed-Mean Inflation Measures, Relative to FRBC-OER4



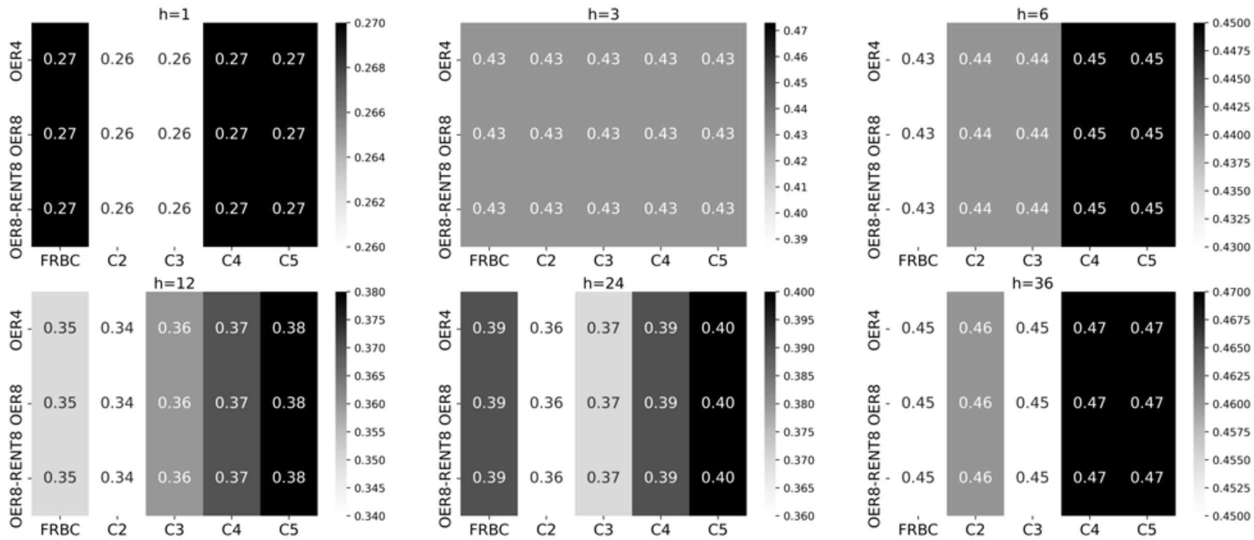
Notes: Reported figures are the p-values of a Diebold-Mariano (1995) test that $RMSE(\hat{\pi}^{2SMA} - \hat{\pi}_{FRBC-OER4})$ and $RMSE(\hat{\pi}^{2SMA} - \hat{\pi}_j)$ are equal, where j denotes the jth candidate trimmed-mean inflation measure. The p-value is obtained by taking the difference of the two squared errors series $\hat{e}^{2SMA}_{j,t}$ and $\hat{e}^{2SMA}_{FRBC-OER4,t}$, and regressing the resulting series against a constant. The test statistic of the constant term is then calculated using HAC standard errors with a small sample correction and $\lfloor 4[T/100]^{2/9} \rfloor$ lags, where T refers to the size of the estimation sample. Darker shading indicates higher values, while lighter shading indicates lower values.

A.1 Predictive power over future inflation

A.1.1 In-sample explanatory power

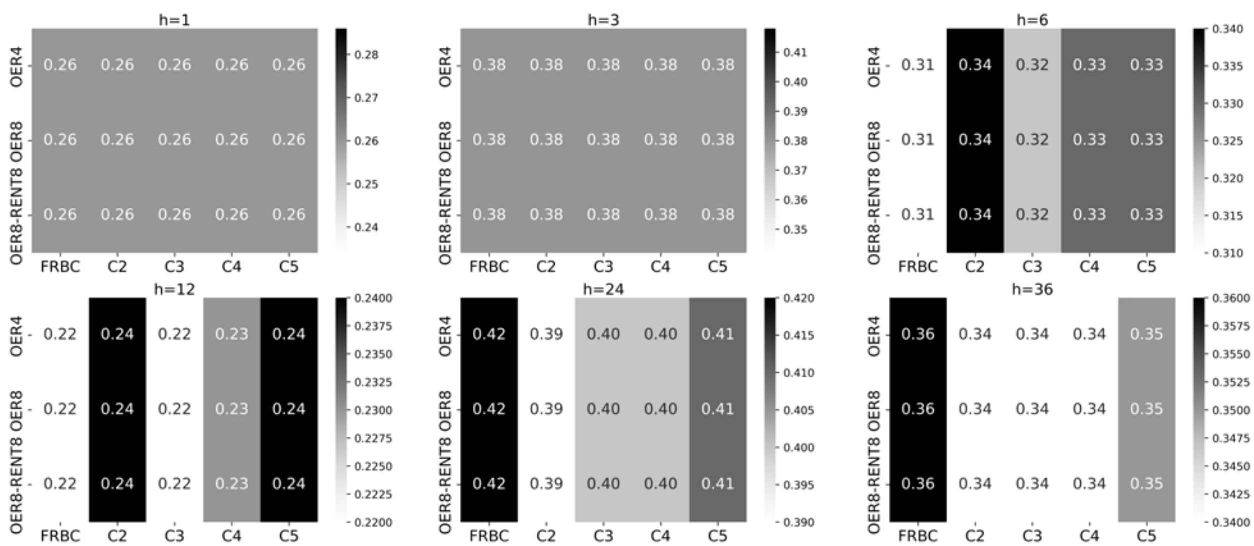
In Figure A.7 and Figure A.8 we report the adjusted R^2 from fitting Equation 2 for each measure of trimmed-mean inflation and for horizons $h \in \{1, 3, 6, 12, 24, 36\}$ in the pre-pandemic and full samples, respectively. Pre-pandemic, C5 measures match or exceed the rest, although differences between measures are small. Results are more mixed in the full sample, with small differences between measures once again.

Figure A.7: In-Sample Adjusted R^2 of Equation 2, Pre-Pandemic Sample (1998M12-2019M12)



Notes: Reported figures are the adjusted R^2 from fitting Equation 2 for each jth candidate trimmed-mean inflation measure. h denotes the horizon in months. Darker shading indicates higher values, while lighter shading indicates lower values.

Figure A.8: In-Sample Adjusted R^2 of Equation 2, Full Sample (1998M12-2024M11)



Notes: Reported figures are the adjusted R^2 from fitting Equation 2 for each jth candidate trimmed-mean inflation measure. h denotes the horizon in months. Darker shading indicates higher values, while lighter shading indicates lower values.

A.1.2 Out-of-sample forecasting ability

To measure trimmed-mean CPI's predictive accuracy over headline CPI, we report in Table A.1 and Table A.2 out-of-sample-forecast RMSFEs relative to the RMSFE of forecasts made using the FRBC-OER4 benchmark for the pre-pandemic and full samples, respectively.

In the pre-pandemic sample, the largest statistically significant gains in forecast accuracy occur at the 6-month (C4-OER8-RENT8 at the 5% level and C5-OER8-RENT8 at the 10% level) and 12-month horizons (C5 measures, at the 10% level). In the full sample, the largest statistically significant gains in forecast accuracy occur at the 6-month horizon (C2 measures, at the 10% level).

Table A.1: Relative RMSFEs of Out-of-Sample Forecasts Using Equation 2, Pre-Pandemic Sample (1998M12-2019M12)

Forecast Horizon (Months)	1	3	6	12	24	36
FRBC-OER4 RMSFE	2.25	1.8	1.53	1.24	1.21	0.98
FRBC-OER8	1.0	1.0	1.0	1.0	1.0	1.01
FRBC-OER8-RENT8	1.0	1.0	1.0*	1.0*	1.0	1.0
C2-OER4	1.0	0.99	0.99	0.97	1.04	0.98
C2-OER8	1.0	0.99	0.99	0.97	1.04	0.99
C2-OER8-RENT8	1.0	0.99	0.99	0.97	1.03	0.98
C3-OER4	1.0	0.99	0.99	0.99	1.02	0.99
C3-OER8	1.0	0.99	0.99	0.99	1.02	0.99
C3-OER8-RENT8	1.0	0.99	**0.99***	0.99	1.01	0.99
C4-OER4	1.0	1.0	**0.98****	0.98	1.01	1.0
C4-OER8	1.0	1.0	**0.98****	0.98	1.01	1.0
C4-OER8-RENT8	1.0	1.0	**0.97****	0.97	1.0	1.0
C5-OER4	1.0	0.99	0.97	**0.96***	0.99	1.02
C5-OER8	1.0	0.99	0.97	**0.96***	0.99	1.02
C5-OER8-RENT8	1.0	0.99	**0.97***	**0.96***	0.99	1.02

Notes: RMSFE is the root mean squared forecast error. The row "FRBC-OER4 RMSFE" reports the raw RMSFE of the FRBC-OER4 benchmark for each forecast horizon. All other rows report relative RMSFEs, with the RMSFE of FRBC-OER4 for forecast horizon h taken as the denominator for relative RMSFEs in column h. Relative RMSFEs less than 1 are highlighted in green. For each relative RMSFE, we calculate the Diebold and Mariano (1995) (DM) test with the small-sample correction of Harvey, Leybourne, and Newbold (1997) for equal predictive accuracy between a given forecast and the forecast from the FRBC-OER4 benchmark. Relative RMSFEs in bold and with *, **, or *** denote rejections of the null hypothesis of equal predictive accuracy of the alternative median CPI measure and the FRBC-OER4 benchmark at the 10%, 5%, or 1% level, respectively, based on the DM test.

Table A.2: Relative RMSFEs of Out-of-Sample Forecasts Using Equation 2, Full Sample (1998M12-2024M11)

Forecast Horizon (Months)	1	3	6	12	24	36
FRBC-OER4 RMSFE	2.71	2.35	2.17	2.18	2.26	2.03
FRBC-OER8	1.0*	1.0	1.0	1.0	1.0	1.0*
FRBC-OER8-RENT8	1.0	1.0	1.0**	1.0**	1.0	1.0
C2-OER4	1.0	0.99	**0.96***	0.97	1.03*	1.02
C2-OER8	1.0	0.99	**0.96***	0.97	1.03*	1.02
C2-OER8-RENT8	1.0	0.99	**0.96***	0.97	1.03	1.02
C3-OER4	1.0	1.0	0.99	1.0	1.02	1.02
C3-OER8	1.0	1.0	0.99	1.0	1.02	1.02
C3-OER8-RENT8	1.0	0.99	0.99	1.0	1.02	1.02
C4-OER4	1.01	0.99	**0.98***	1.0	1.02	1.02
C4-OER8	1.01	0.99	**0.98***	1.0	1.02	1.03
C4-OER8-RENT8	1.01	0.99	**0.98***	0.99	1.02	1.02
C5-OER4	1.01	1.0	0.98	0.99	1.02	1.02
C5-OER8	1.01	1.0	0.98	0.99	1.02	1.02
C5-OER8-RENT8	1.01	1.0	0.98	0.99	1.02	1.02

Notes: RMSFE is the root mean squared forecast error. The row "FRBC-OER4 RMSFE" reports the raw RMSFE of the FRBC-OER4 benchmark for each forecast horizon. All other rows report relative RMSFEs, with the RMSFE of FRBC-OER4 for forecast horizon h taken as the denominator for relative RMSFEs in column h. Relative RMSFEs less than 1 are highlighted in green. For each relative RMSFE, we calculate the Diebold and Mariano (1995) (DM) test with the small-sample correction of Harvey, Leybourne, and Newbold (1997) for equal predictive accuracy between a given forecast and the forecast from the FRBC-OER4 benchmark. Relative RMSFEs in bold and with *, **, or *** denote rejections of the null hypothesis of equal predictive accuracy of the alternative median CPI measure and the FRBC-OER4 benchmark at the 10%, 5%, or 1% level, respectively, based on the DM test.

A.2 Summary of results

We find that trimmed-mean CPI is insensitive to increasing shelter disaggregation, but can benefit from increasing the level of non-shelter disaggregation beyond FRBC. However, based on our evaluation, the optimal level of non-shelter disaggregation is unclear, as it appears to shift depending on the criteria and the sample period over which performance on the criteria is evaluated.

A.3 Empirical application: Phillips curve relationship

Table A.3: Estimated Phillips Curve Slope

Series	Trimmed-Mean CPI			
	Sample: 2000-2019		Sample: 2000-2024	
	Beta	P-value	Beta	P-value
FRBC-OER4	-0.195	0.00	-0.310	0.00
FRBC-OER8	-0.193	0.00	-0.308	0.00
FRBC-OER8-RENT8	-0.193	0.00	-0.308	0.00
C2-OER4	-0.182	0.00	-0.292	0.00
C2-OER8	-0.181	0.00	-0.290	0.00
C2-OER8-RENT8	-0.181	0.00	-0.290	0.00
C3-OER4	-0.190	0.00	-0.308	0.00
C3-OER8	-0.190	0.00	-0.307	0.00
C3-OER8-RENT8	-0.190	0.00	-0.307	0.00
C4-OER4	-0.178	0.00	-0.299	0.00
C4-OER8	-0.178	0.00	-0.298	0.00
C4-OER8-RENT8	-0.178	0.00	-0.298	0.00
C7-OER4	-0.180	0.00	-0.300	0.00
C7-OER8	-0.179	0.00	-0.299	0.00
C7-OER8-RENT8	-0.179	0.00	-0.299	0.00

Note: The estimates shown are for two different estimation samples: 2000M1 through 2019M12 (denoted Sample: 2000-2019) and 2000M1 through 20204M11 (denoted Sample: 2000-2024). The data from 1999M1 through 1999M12 are used to compute the lagged value of the unemployment rate gap.

B Appendix 2: Median Versus Mean, for Finding the Center of the Cross-sectional Distribution of Monthly Inflation Rates

The median of a sample is its middle value, i.e., half of the points in the sample are higher than the median and half are lower.

But some samples have weights, i.e., each point has an associated weight. This is the case in the CPI, where each point represents the percentage change in the index of a given category of goods or services, and the weights correspond to the aggregation weights in the CPI. The CPI in a given month is the weighted mean of this sample.

Corresponding to a weighted mean, one can also construct a weighted median. A weighted median of a sample is constructed using the same logic as an unweighted median: the weighted median is that value such that half of the weight of the data lies below it. In that sense, like the median, it identifies the middle value (i.e., the center of the distribution).

A weighted median is very insensitive to outliers, as the following thought experiment demonstrates. Consider taking just a single point in the sample that lies above the median, and replacing it with a point that is arbitrarily large. The weighted median will not change its value (at all): this outlier has not contaminated the information in the weighted median.

Meanwhile, as noted above, the CPI is a weighted *average*. Under ordinary circumstances, a weighted average provides a very good signal of the center of the sample. But a weighted mean has a serious drawback. Very much unlike the weighted median, a weighted mean is extremely sensitive to outliers. For instance, one may increase the weighted mean by an arbitrary amount by taking just one single point that lies above the weighted mean–no matter how small its weight–and increasing it by a sufficiently large amount.

This means that if a given sample has unusually big, or unusually small, points – and particularly if those points are associated with non-negligible weights – the sample mean will be pulled very strongly in the direction of the unusual points, and fail to provide an efficient signal of the center of the distribution. Indeed, this provides the rationale for limited-influence estimators, such as the median or trimmed mean.

This behavior is well-illustrated for the four months of the CPI's underlying component growth rates depicted in Figure B.1: June 2021, June 2022, June 2023 and June 2024. Depicted in the four panels of the figure are month-over-month percentage changes in the CPI components, along with the official CPI reading of that month, and the corresponding Median CPI reading (corresponding to our FRBC-OER8-Rent8 variant). In June 2021, for instance, there are three extreme positive outliers (as discussed in the main body). They are massively larger than the bulk of the inflation observations that month. The median CPI (in M/M terms), depicted with an orange dotted line, reads 0.3 for that month. It is easy to see that it is providing an accurate read of the center of the bulk of the inflation component readings. Conversely, the CPI (in red) – pulled as it was by those outliers – came in at more than double that amount, namely, at 0.86. In this month, the CPI was clearly not a reasonable signal of the center of the distribution. In June 2022, we see something qualitatively similar: there is a mass of observations lying between -0.4 and +1.3, and it is unsurprising to see that the median is 0.61. The CPI, conversely, read +1.30 that month, more than double the value of the median, and well above the bulk of the observations, pulled as it was by a few outliers (including one extreme outlier). Conversely, we see that in June 2023, the sample distribution of CPI component growth rates was only modestly negatively skewed with only two outliers that were not nearly as extreme. The CPI ended up close to the Median value (0.26 vs. 0.25); but in this month, unlike the previous two cases we just looked at, the CPI seems a reasonable reading of the center of the distribution. Finally, in June 2024, the sample distribution is clearly displaying a more pronounced negative skew, with a fairly large number of observations in the deeply negative range. These negative outliers pulled the CPI down to a reading of 0.0, while the Median appears to be providing a more reasonable signal of the center of the distribution, at +0.19.

Figure B.1: Cross-sectional distribution of inflation in CPI price index components