Federal Reserve Bank of Cleveland Working Paper Series

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

Christian L. Garciga, Randal J. Verbrugge, and Saeed Zaman

Working Paper No. 24-02R2

February 2026

Suggested citation: Garciga, Christian L., Randal J. Verbrugge, and Saeed Zaman. 2026. "Improving the Median CPI: Maximal Disaggregation Isn't Necessarily Optimal." Working Paper No. 24-02R2. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202402r2.

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: December 11, 2025

Abstract

Median Consumer Price Indexes have proved useful in many contexts and are used worldwide as key policymaking inputs. Historically, US Median CPI improvements involved increasing the CPI component level of disaggregation; one might reasonably assume that further disaggregation would lead to further improvements. We theoretically demonstrate herein: not necessarily. We then empirically explore the impact of further disaggregation. Substantially more disaggregation in the shelter indexes and slightly more disaggregation in the remaining components improve the ability of Median CPI to track the medium-term trend in CPI inflation and its predictive power for future CPI movements. This new Median CPI suggests that trend inflation was lower pre-pandemic, accelerated faster in 2021, and decelerated faster after 2022.

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

JEL Codes: E31, E37, E52, C8

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

*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, Todd Clark, Jorge Fornero, Edward Knotek, Michael McCracken, James Mitchell, Tony Richards, and participants at the American Economic Association-ASSA Annual Meeting 2025, Midwest Macroeconomic Spring Meetings 2025, UEA 3rd Time Series Workshop 2025, European Central Bank and Cleveland Fed seminar meeting, 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

1 Introduction

Much attention is devoted to discerning the signal in the latest inflation reading. That signal is obscured by changing seasonality and transient fluctuations, but also by outright noise: the price component growth distribution is highly leptokurtic, so that monthly samples typically have big outliers. Official measures of inflation are sample means, but sample means are highly sensitive to outliers.

A simple and efficient approach to removing noise and estimating the center of a leptokurtic distribution is using a limited-influence statistic, such as the sample median or trimmed-mean. Introduced by Bryan and Pike (1991) and Bryan and Cecchetti (1994), the Median CPI has proven to be an efficient estimator of the center of the CPI component distribution, greatly reducing noise at higher frequencies.[1] This allows a more accurate assessment of the latest data (for more on this point, see the discussion in Richards 2024). Furthermore, median and trimmed-mean CPIs have been shown to provide accurate estimates of the medium-term trend (MTT) in inflation, an unobserved object thought to "reflect medium-term inflation developments linked to the business cycle" (Banbura et al., 2023).[2] Measuring the MTT in inflation is crucial in many contexts. For instance, it is of central importance for monetary policy, and it is necessary for accurate estimation of the real interest rate in the medium term, which is relevant for most intertemporal decisions. While much more sophisticated approaches to estimating the MTT exist (see, e.g., Goulet Coulombe et al. (2024) or Carlomagno et al. (2023)), simple approaches have key advantages, such as ease of communicating with the public and maintaining its trust.

For these reasons, median and trimmed-mean CPIs are increasingly used in empirical studies, not only for forecasting, but also for such topics as deducing and interpreting inflation uncertainty, locating a stable Phillips curve, distinguishing between price adjustment theories, and understanding post-COVID inflation dynamics (for citations and a more comprehensive list of uses, see Appendix H). 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). In fact,

[1] Noise may obscure relationships and cloud empirical work involving inflation; see also Mazumder (2017), Kishor and Koenig (2022), and de Veirman (2023), and this point is implicit in Andrle, Bruha and Solmaz (2017). Owing to the importance of reducing inflation noise, there is an entire literature devoted to the topic, and numerous approaches have been proposed.

[2] See also Mertens (2016).

central banks around the world make significant use of such measures in their monetary policy decision-making. Both the European Central Bank and the Sveriges Riksbank use a median inflation measure as one of their estimates of "underlying inflation" (O'Brien and Nickel, 2018; Nickel et al., 2021; Johansson et al., 2018). In New Zealand, Norway, and Canada, trimmed-mean and median measures comprise two of five or fewer measures of core inflation (Reserve Bank of New Zealand, 2024; Norges Bank, 2025; Lao and Steyn, 2019).[3] As "measures of underlying inflation," the Bank of Japan produces and uses a trimmed-mean and a weighted median (Shiratsuka, 2015). In Australia, "the most important indicators of underlying inflation are the trimmed-mean and the weighted median" (Reserve Bank of Australia, 2025). Both the Swiss National Bank and the Central Bank of Costa Rica use a trimmed-mean as their indicator of core inflation (Swiss National Bank, 2025; Esquivel-Monge et al., 2011; Munoz-Salas and Rodriguez-Vargas, 2021).

Given the growing importance of the median and trimmed-mean CPI, it is worth asking whether the methodology for constructing these measures could be further enhanced. One natural avenue worth exploring is increasing the level of disaggregation, i.e., the number of CPI components used to calculate the median CPI. It is a natural avenue to explore for at least two reasons. First, historically, all improvements of the Median CPI have involved increasing the level of disaggregation.[4] Second, the level of disaggregation in the current Median CPI measures is relatively low: just 45 components. Each of these components is an average of more disaggregated components, and as we have noted, sample averages are sensitive to outliers. In the limit, as less and less disaggregation is used, the sample median will converge to the sample mean. This suggests that more disaggregation is better. Furthermore, trimmed-mean *PCE* inflation and median *PCE* inflation are calculated using much larger baskets (178 and 201 PCE components, respectively: see Dolmas 2009 and Carroll and Verbrugge 2019). This paper investigates whether further disaggregation would enhance the Median CPI.[5]

We make theoretical and empirical contributions. On the theory front, we prove that in a real-world CPI context, more disaggregation does not necessarily yield the minimum mean squared error (MSE) estimator. The optimal level of disaggregation thus becomes an empirical question. On the

[3]In Canada, these are described as "main operational guides for monetary policy;" see also the remarks of Deputy Governor Schembri (2017), who discussed the rationale for abandoning traditional "core" CPI.

[4]For a brief history, see Appendix D.

[5]A parallel analysis pertaining to the Trimmed-Mean CPI is in Appendix A.

empirical front, we systematically investigate the impact of increasing the level of disaggregation, taking the current Federal Reserve Bank of Cleveland (FRBC) level of disaggregation as our starting point.[6] In particular, we further disaggregate the CPI basket along two dimensions: shelter and non-shelter components. For shelter components, starting from the current split of 4 regional OER components and 1 national Rent component, we explore further disaggregating each regional OER component into two city-size classes, and we explore disaggregating Rent along similar lines. For non-shelter components, we construct a total of four increasingly more disaggregated baskets, following the disaggregation structure of the CPI. All told, we compare 15 candidate baskets; the maximally disaggregated basket consists of 156 components.

We then evaluate the performance of each new median measure using well-established criteria from the core inflation literature, two of which are worth highlighting: accuracy in tracking an ex-post measure of the MTT in CPI inflation, and accuracy in predicting future CPI inflation. These criteria indicate that disaggregating both OER and Rent into eight components is optimal, but more extensive disaggregation of non-shelter components generally results in clear performance losses. Thus, our findings are consistent with our theoretical result: increasing the level of disaggregation as far as possible is not optimal.

Our work extends a long thread of the literature focused on limited-influence estimators (e.g., Bryan et al. 1997; Dolmas 2005; Brischetto and Richards 2007; Dolmas 2009; 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). 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 indexes (e.g., cyclical properties, similar to 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 on 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, in the calculation of

[6] For brevity, we omit results demonstrating that less disaggregation leads to worse estimators.

limited-influence estimators of inflation. We are also the first to investigate the performance of these inflation measures by level of disaggregation in a systematic manner. Finally, in another first, we 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.

Section 2 summarizes key details regarding the computation of the Median CPI. Section 3 sketches the intuition for our theoretical result: the optimal level of disaggregation need not be the highest level of disaggregation possible. Section 4 provides our approach to disaggregating the CPI. Results are presented in Section 5. Section 6 provides a discussion and two applications: how the Phillips curve relationship and how historical inferences about MTT dynamics change depending upon the version of the Median CPI used. Section 7 concludes.[7]

2 Computing the FRBC Median and Trimmed-Mean CPI

In this section, we first briefly outline the steps taken to calculate the median CPI from any given collection of CPI components. Then we explain how one may compute unpublished OER expenditure weights (and, by extension, unpublished Rent expenditure weights) for regional components of OER.

2.1 Selecting the Median CPI

In the interest of brevity, here we provide a high-level description of the method to compute median CPI. For a complete description of the method that the FRBC uses to compute the median and trimmed-mean CPIs, see Appendix E.

A brief description of the method used to compute the Median CPI is as follows. Starting with the J components in the basket, each month, sort these components' month-over-month inflation rates from smallest to largest, $j = 1, ..., J$. For each j, compute the cumulative expenditure weight $w(j) = \sum_{k=1}^{j} \omega_k$ where ω_k is the expenditure weight for component k. Identify the item category whose cumulative expenditure weight $w(j)$ falls at the 50th percentile; this item's inflation rate is the Median CPI inflation for that month.

[7] This paper has a supplementary online Appendix that has additional results and details about the history of the Median CPI and its computation.

Computing the trimmed-mean CPI is a little more involved, since this is computed as a weighted mean (so weights must be renormalized), and the specific items at the top and bottom trim percentages do not receive their full expenditure weights. For details, see Appendix E.2.

2.2 Computing the Unpublished OER Expenditure Weights

As is evident, expenditure weights are needed to compute the Median and Trimmed-Mean CPIs. 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, these relative importances are updated every month to reflect relative price movements that have occurred. Essentially, these approximate how expenditure weights change, based upon changes in relative prices.[8] Most of these relative importances are published.[9]

But not all. In order to disaggregate national OER into regional components, each component must be weighted by its appropriate share in the overall CPI-U. But the BLS does not publish the relative importances of each regional OER component relative to overall CPI-U, so these must be computed. This is a two-step process.

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. Then 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 indexes for

[8]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.

[9]When the Median CPI was introduced, the BLS did not publish any 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.

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 that month. With weights and indexes in hand, we apply seasonal adjustment as necessary. These procedures are explained in Appendix E.1.

3 More Disaggregation Need Not Be Better: A Counterexample

This paper takes as given the decades of research demonstrating the usefulness of the Median CPI as an estimator of the MTT in inflation, and this section asks: does constructing that median over a more disaggregated set of indexes necessarily lead to a more accurate estimate of the "true" median?

There is certainly a solid rationale for thinking so. Given the historical record of improvements to the Median CPI, and given the much higher level of disaggregation 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 theoretical argument that would appear 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 minimal level of disaggregation, splitting the CPI into just core goods, core services, and energy. Most of the time, the core services category is likely to be chosen as the median. It is clear that the resulting median will be highly volatile and still driven by outliers, only a modest improvement 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, that outliers would be more accurately identified, 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. Proposition 1, given in Appendix F, uses a simple counterexample to prove that using the median of the most disaggregated published index doesn't always provide a better (smaller mean-squared-error) estimate of the true median of the underlying distribution, compared to using a more aggregated set of indexes.

To give the intuition in the counterexample, consider a CPI comprised of 7 equally weighted indexes, whose growth-rate outcomes are driven by a coin flip, H or T. Outcomes are given in the table below:

Coin flip	Index						
	1	**2**	**3**	**4**	**5**	**6**	**7**
H	5	-5	0	1	-1	2	-2
T	-5	5	0	-1	1	-2	2

Besides index 3 (I_3), the other indexes come in pairs. For instance, if the coin flip is H, then index 1 (I_1) grows by 5, while I_2 grows by -5. The growth of the CPI would be the weighted average of the outcomes of I_1 through I_7.

For either H or T, the weighted median of the above basket is 0, matching the outcome of Index 3, i.e., I_3. Suppose that the statistical agency does not publish I_1 and I_5, but instead aggregates these two indexes into an index Z, whose weight is accordingly twice as much as the other indexes. This would not change the CPI outcome, but it would change the median outcome. In particular, when the coin flip is H, $Z = 2$, and the weighted median is $+1$. When the coin flip is T, $Z = -2$ and the weighted median is -1. If the statistical agency also offers an index Y that aggregates I_2 with I_4, then computing the weighted median of the set $\{Y, Z, I_3, I_6, I_7\}$ always yields the correct median.

This example is highly stylized, but the underlying idea is far more general. What is the upshot? Since there is no way to deduce the properties of the underlying indexes (as they are unobserved), there is no way to know which subset of aggregated and disaggregated indexes best estimate, on average, the true median of the distribution. Thus, the optimal level of disaggregation becomes an empirical issue. Testing must focus on performance along *measurable* dimensions, and clearly these dimensions should reflect end-use interests, such as accuracy against ex-post estimates of the MTT.

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 is about 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 is related 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 the context of this proof, the true weighted median is a function of all of that idiosyncratic noise.

4 Our Approach to Disaggregating 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 the 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. 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.[10] 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 holds:

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.[11] 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.

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 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. C2 has one com-

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

[11]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.

ponent with a higher relative importance than FRBC, but this is due to the fact that it follows the CPI item aggregation tree and so is derived by disaggregating C1, not FRBC.

[Figure 1 here]

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. But going back further in time, to 1997M12, required us to adjust for structural breaks in the components "Medical care commodities" and "Telephone services."[12,13] As a result, 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.

Going back further in time before 1998 is far more problematic. Not only was the item structure of the CPI even more dissimilar to the current one, extending our analysis prior to 1997:12 would eliminate an entire dimension of our analysis, namely the further disaggregation of shelter indexes (see Section 4.2). The BLS did not begin publishing monthly shelter indexes by city size class until December 1996 (prior to this, they were bimonthly, which are arguably not well-suited to even a quarterly analysis).[14]

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.[15] Where the SA price

[12]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."

[13]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 onward, "Land-line telephone services, local charges," and "Land-line telephone services, long distance" are merged into a single "Land-line telephone services" component.

[14]We acknowledge that our sample excludes some of previous major fluctuations in inflation, though it does include both the Global Financial Crisis and the COVID-19 inflation episode.

[15]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, please contact the authors. Price indexes for "Nonprescription drugs and medical supplies," "Internal and respiratory OTC drugs," "Nonprescription medical equipment and supplies," "Land-

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 to calculating the 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.[16]

4.2 Splitting Shelter Components of the CPI

Unlike the other components in the CPI, OER and Rent cannot be split into further nationally representative subindexes.[17] 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 FRBC "Revised Methodology" of the Median CPI. OER4 breaks apart the national OER price indices and weights into 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 shelter 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 the other for 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.[18]

line telephone services, local charges," and "Land-line telephone services, long distance," which were not in Haver and were obtained from the BLS website.

[16] 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.

[17] Adams and Verbrugge (2025) criticize this practice, and propose a split along a structure-type dimension.

[18] Prior to December 2017, there were actually three city size classes: Size A (cities with a population size over

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 Section 2.2;[19] and (2) seasonally adjust each price index using X-13-ARIMA-SEATS once using data starting in 1997M12.

[Figure 2 here]

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.[20]

[Figure 3 here]

From each of our 15 splits of the CPI, using the methodology outlined in Section 2, it is straightforward to derive measures of 12-month Median CPI inflation. Hereafter, when we refer to inflation,

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.

[19] All Haver codes for the input price indices and relative importances of shelter split components are available on request from the authors. 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 series are available on request from the authors.

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

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." After taking the 12-month percent change of the levels of Median CPI 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.

The first criterion entails evaluating 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.

The second criterion entails an assessment of the extent to which each candidate measure predicts 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 its 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.[21]

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

[21] 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.

was to the mean of CPI inflation over the pre-pandemic and full samples.

In Figure 4, panel (a), 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.[22] 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 indistinguishable 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 4, panel (b). 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 here]

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

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 5, panel (a), 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 5 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.[23]

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

[23] This is partly, but not entirely, driven by overall bias. For an illustration, see the Trimmed-Mean CPI results in Appendix A.

against the alternative hypothesis of significantly different tracking ability. Test p-values are reported in Figure 5, panel (b). 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 5 here]

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. Results are reported in Appendix I, in Figure I.3. Results are largely consistent with those for the 37-month moving average. 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. Only FRBC-OER8-RENT8 and C2-OER8-RENT8 inflation achieve a statistically significant improvement in tracking the 2SMA trend relative to FRBC-OER4 inflation in both the pre-pandemic sample and the full sample.[24]

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.

[24] A referee encouraged us to also investigate robustness to a longer window. Our results for a centered moving average of length 49 are qualitatively unchanged. Notably, both the FRBC-OER8-Rent8 and the C2-OER8-Rent8 measures offer economically- and statistically-significant gains over FRBC-OER4, and outperform all other measures (in some cases, by a large margin).

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.

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, $j = 1, ...15$. In other words, we use the current gap between headline inflation and the median inflation measure to predict the movement of headline inflation over the next h months. 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 6, panel (a), we report $\bar{R}^2$ from fitting Equation 2 on data in the pre-pandemic sample, and in Figure 6, panel (b), 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 to track medium-term movements in CPI inflation.

[Figure 6 here]

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.[25] This exercise provides an additional metric by which to measure how the median CPI's predictive accuracy over

[25] Our exercise is "pseudo" real-time because 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.

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 point 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 out-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 $2010M12 - h + 1$ to obtain a forecast for 2011M01, and so on, until we have made a forecast for the last month in our sample.

The density forecasts are computed by applying the parametric bootstrapping procedure to the estimated Equation 2; the details of the procedure, which accounts for both the shock and the parameter uncertainty, are relegated to Appendix G. We assess the accuracy of the density forecast using the widely used metric of log predictive score. The higher the log score, the more accurate the density forecast. This metric provides the broadest measure of the accuracy and calibration of the entire predictive density (see Geweke & Amisano, 2010).

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

In general, out-of-sample point forecast accuracy across most alternative splits is comparable to the benchmark; gains or losses are statistically distinguishable in a few cases, and the significance of the forecast gains varies across the pre-pandemic and full samples.[26,27,28] Specifically, 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.

Table 2 presents the density forecast evaluation results. Specifically, it reports the average log score of the predictive densities from the model with a given candidate median measure minus the average log score of the densities from the benchmark model with the FRBC-OER4 median measure. Thus, a positive number suggests that, on average, the density forecasts from the model with the listed candidate median measure are more accurate than those from the benchmark model. The top panel reports results for the pre-pandemic sample and the bottom panel for the full sample. Also reported is the statistical significance of the forecast differences based on the likelihood-ratio test of Amisano and Giacomini (2007).

Overall, density forecast results reinforce the point forecast results; if anything, they provide

[26]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 gains disappear in the extended sample, which incorporates data since the onset of the COVID pandemic. In the extended sample, there are seven cases with statistically significant forecast gains, but only one overlaps with the statistical gains observed in the shorter sample.

[27]A referee encouraged us to study robustness across other subsamples. Our forecasting results hold up if we end our estimation period in 2007, and study forecasts subsequent to then, and also hold up if we restrict attention to the past ten years (so our results are not driven by pre-2015 performance, either).

[28]We also assessed the statistical significance using the Clark and West test, and results are robust. Indeed if anything, the Clark and West (2007) test finds a greater number of accuracy gains to be statistically significant than the DM test.

stronger support for the FRBC-OER4-OER8 median measure. Similar to point forecast evaluation, the density forecast evaluation results indicate that accuracy gains over the benchmark model are concentrated for the OER8-RENT8 splits. For both sample periods, FRBC-OER4-OER8 consistently outperforms the benchmark model across all forecast horizons except the very near term, and several of the accuracy gains are statistically significant. But, in contrast to point forecast results, density results for horizons 24- and 30-months-ahead suggest that for the sample that includes the past 5 years of evaluation data, the density accuracy of most splits, with the exception of FRBC-OER8-RENT8, is statistically significantly inferior to the benchmark FRBC-OER4 model. A deeper inspection of the density estimates suggests that this statistically significant worsening in accuracy across several splits is due to narrower coverage intervals of the densities and inferior mean estimates of the forecast densities. The latter reasoning of inferior mean estimates is evident in relative RMSFE ratios greater than 1 in the point forecast evaluation for these splits at the relevant forecast horizons.

[Table 1 here]

[Table 2 here]

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 improves 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.[29]

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. Both the point and density evaluation results indicate that OER8-RENT8 splits offer the most improvements in forecast accuracy at multiple horizons. Overall, FRBC-OER8-RENT8 followed by C2-OER8-RENT8 appears 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

[29] In Appendix I, Figure I.1 (pre-COVID sample) and Figure I.2 (post-COVID sample) compare the main MTT proxy (37MMA) to the various median measures.

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.[30]

Thus, given the degree of interest in the distribution of the median component, before getting to our main results, we provide in Figure 7 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 7, panel (a) covers the pre-pandemic period (1998M12-2019M12) while panel (b) 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 in each figure). Splitting OER4 into 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

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

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 7 here]

Based on the findings presented in Section 5 and the research by Brischetto and Richards (2007), it would be reasonable to infer that there is a direct relationship between reducing the frequency of selecting shelter as the median component, and achieving performance gains in estimators of median CPI. However, if this relationship were strictly true, then the basket C3-OER8-RENT8 should have shown better performance than C2-OER8-RENT8 in our evaluation. This is because non-shelter components are chosen as the median more frequently in C3-OER8-RENT8, both before the pandemic and in the full sample. The fact that this was not the case suggests that the frequent selection of shelter as the median component may be a "feature" of the median CPI, and not a "bug" which is necessarily correlated with inferior performance. 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. In such scenarios, shelter would not be chosen as the median. Therefore, the frequent selection of shelter as the median might simply indicate that shelter costs generally align with 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, 2025) 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. Indeed, given the definition of an MTT, an MTT estimator that was *not* cyclically sensitive would be a poor estimator!

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.[31] 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, as the degree of shelter disaggregation increases, the estimated Phillips curve relationship weakens, but only marginally (and remains very robust). 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* simply 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 rows always results in a *weaker* Phillips curve relationship. The results are qualitatively similar for the Trimmed-Mean CPI and are reported in Appendix A.1.3.

[Table 3 here]

Recent work (e.g., Benigno and Eggertson, 2024; Ashley and Verbrugge, 2025) has located evidence for nonlinearities of various forms in Phillips curves. In Appendix B, we explore two different forms of nonlinearity in the Phillips curve, to investigate whether the above findings are robust. First, we approximate the curvature of an unknown nonlinear Phillips curve function using a piecewise-linear function, estimated using threshold regression with a continuity constraint, as in Doser et al. (2023). Second, we allow for frequency-dependence in the Phillips curve (i.e., we allow the Phillips curve relationship to differ by frequency), following Ashley and Verbrugge (2025) and Verbrugge and Zaman (2024).[32] We find convincing evidence for both types of nonlinearity. And broadly speaking,

[31] 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.

[32] We are grateful to a referee for encouraging this exploration.

we find the same general relationship with respect to level of shelter disaggregation: the strength of the Phillips curve tends to decline modestly as the level of shelter disaggregation increases, but remains quite robust.

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 8, 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 8 here]

In Figure 9, 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.

[Figure 9 here]

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. [33] 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. 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.[34]

[33] In Appendix C, 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.

[34] "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.

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.[35]

7 Conclusion

The Median CPI is a well-known estimator of the medium-term trend (MTT) in CPI inflation. Over time, various improvements to this measure have been implemented, and in this paper, we thoroughly investigate whether further disaggregation would result in further improvement. There are several reasons to think so. 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. One might think 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 that are well-established in the literature, focusing on two criteria in particular: accuracy vis-à-vis medium-term movements in CPI inflation, and predic-

[35] 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 authors' websites.

tive 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 the 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): perhaps *less* disaggregation would be helpful. Having said that, our work also suggests that in that context, there may be clear performance gains to disaggregating the OER and Rent components: at present, neither component is disaggregated at all.

References

[1] Adams, Brian, & Randal Verbrugge. 2025. Location, Location, Structure Type: Rent Divergence within Neighborhoods. *Journal of Housing Economics* 69: 102081. https://doi.org/10.1016/j.jhe.2025.102081

[2] Alvarez, Rocio, Maximo Camacho, & Gabriel Perez-Quiros (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] Amisano, Gianni & Raffaella Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics* Volume 25(2): 177-190. https://doi.org/10.1198/073500106000000332

[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] 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

[7] Ball, Laurence M., & Sandeep Mazumder (2019). 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

[8] Banbura, Marta., Elena Bobeica, Katalin Bodnár, Bruno Fagandini, Peter Healy, & Joan Paredes (2023). Underlying inflation measures: an analytical guide for the euro area. *ECB Economic Bulletin, Issue 5/2023*

[9] Benigno, Pierpaolo, and Gauti B. Eggertsson. (2024). The Slanted-L Phillips Curve. *AEA Papers and Proceedings* 114: 84-89.

[10] 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

[11] 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.

[12] 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

[13] 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.

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

[15] 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

[16] 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

[17] Cecchetti, Stephen G., & Erica L. Groshen (2001). Understanding Inflation: Implications for Monetary Policy. In: Drèze, Jacques. (ed) *Advances in Macroeconomic Theory*. International Economic Association. Palgrave Macmillan, London. https://doi.org/10.1057/9780333992753_7

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

[19] Clark, Todd E., & Kenneth D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* 138(1):291-311. https://doi.org/10.1016/j.jeconom.2006.05.023

[20] 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

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

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

[23] Dolmas, Jim (2005). trimmed-mean PCE Inflation. *Federal Reserve Bank of Dallas Working Paper* Number 05-06.

[24] Dolmas, Jim (2009). The 2009 Revision to the trimmed-mean PCE Inflation Series. Technical note. Federal Reserve Bank of Dallas.

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

[26] Doser, Alexander, Ricardo Nunes, Nikhil Rao, & Viacheslav Sheremirov (2023). Inflation expectations and nonlinearities in the Phillips curve. Journal of Applied Econometrics, 38(4), 453–471. https://doi.org/10.1002/jae.2963

[27] 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.

[28] 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

[29] 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.

[30] Geweke, John, & Gianni Amisano (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting* Volume 26(2): 216-230. https://doi.org/10.1016/j.ijforecast.2009.10.007

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

[32] 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

[33] 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

[34] 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

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

[36] 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

[37] 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

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

[39] 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.

[40] 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.

[41] 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_a_00549

[42] 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

[43] 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.

[44] 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?82b5ee80f9a840af58de447a1e25f86

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

[46] 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

[47] 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.

[48] 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.

[49] 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

[50] 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

[51] 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

[52] 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.

[53] Shiratsuka, Shigenori (2015). Performance of Core Indicators of Japan's Consumer Price Index. *Bank of Japan Review Series* 2015-E-7.

[54] 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

[55] 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.

[56] 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

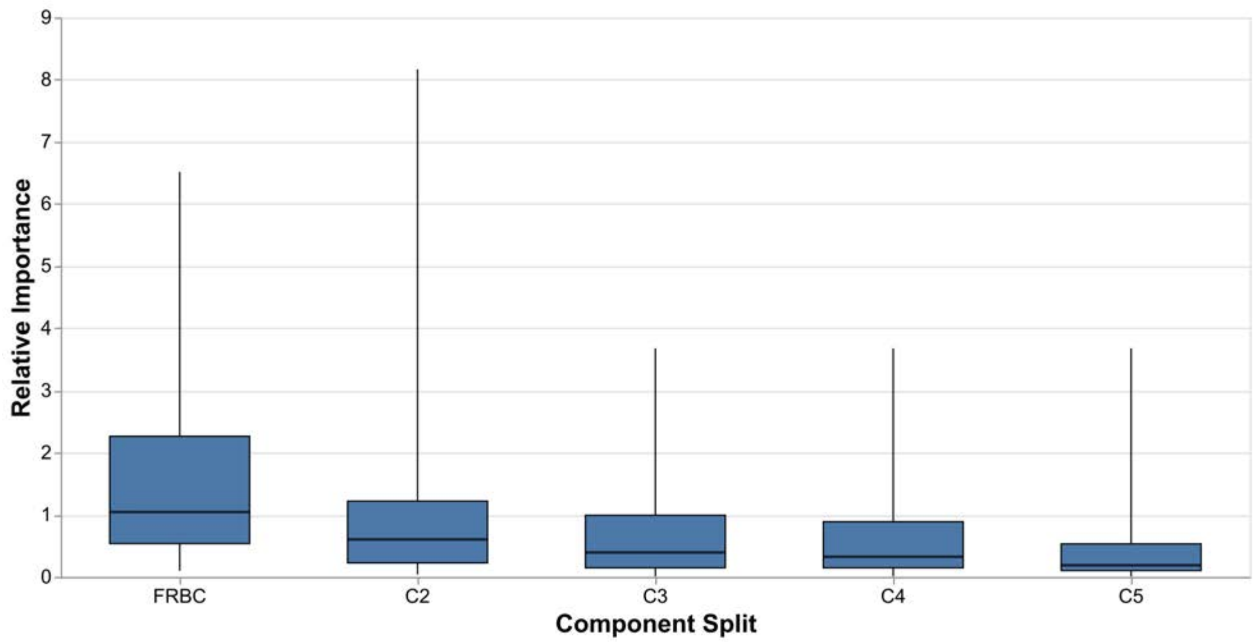
[57] Swiss National Bank (2025). Glossary. https://www.snb.ch/en/services-events/digital-services/glossary, accessed 2/19/2025.

[58] 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

[59] Verbrugge, Randal J., & Saeed Zaman. (2024) Post-COVID inflation dynamics: Higher for longer. *Journal of Forecasting* 43(4): 871-893.

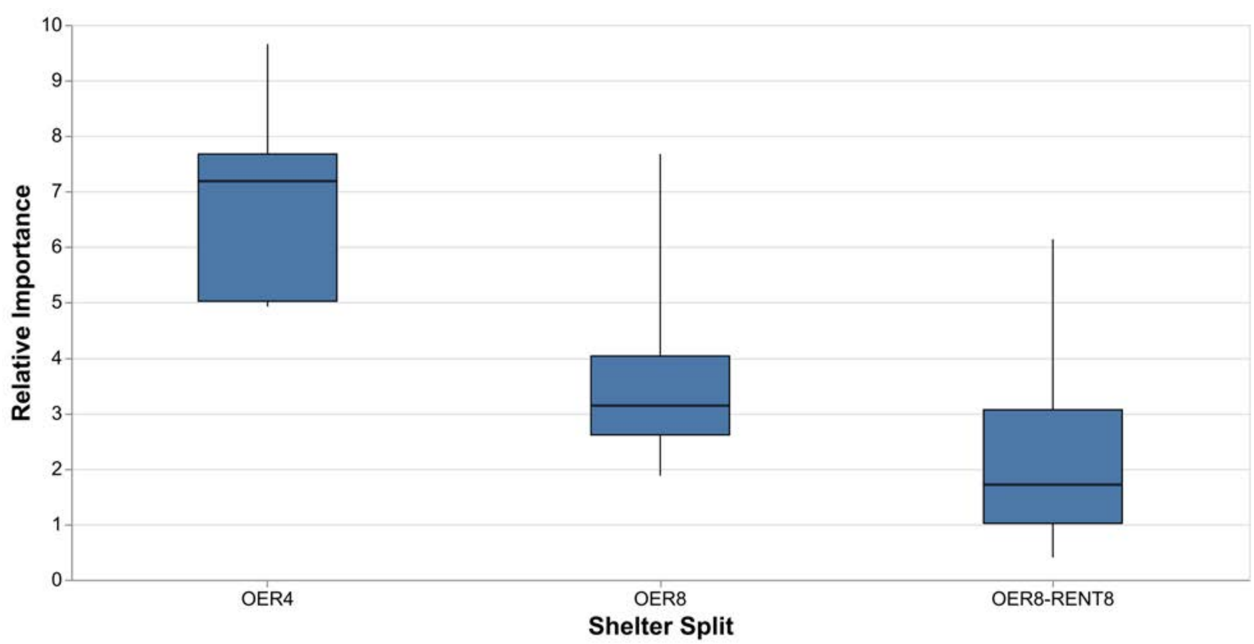
[60] Zaman, Saeed (2019). Cyclical versus Acyclical Inflation: A Deeper Dive. *Federal Reserve Bank of Cleveland Economic Commentary* Number 2019-13. doi:10.26509/frbc-ec-201913

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.

Figure 2: Distribution of Component Weights in Each Housing Split



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

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

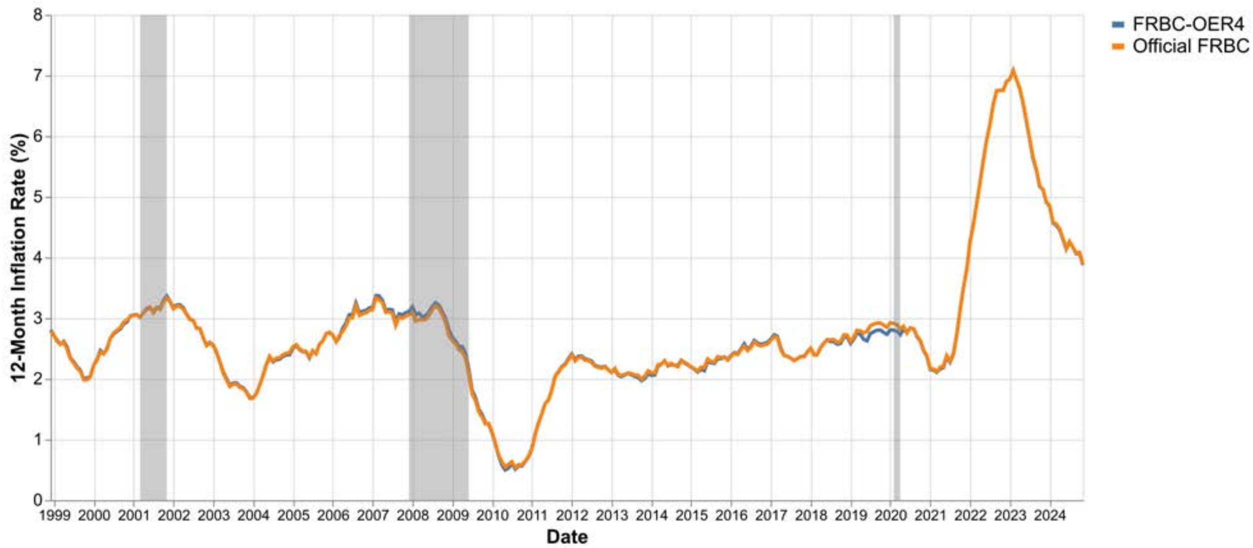
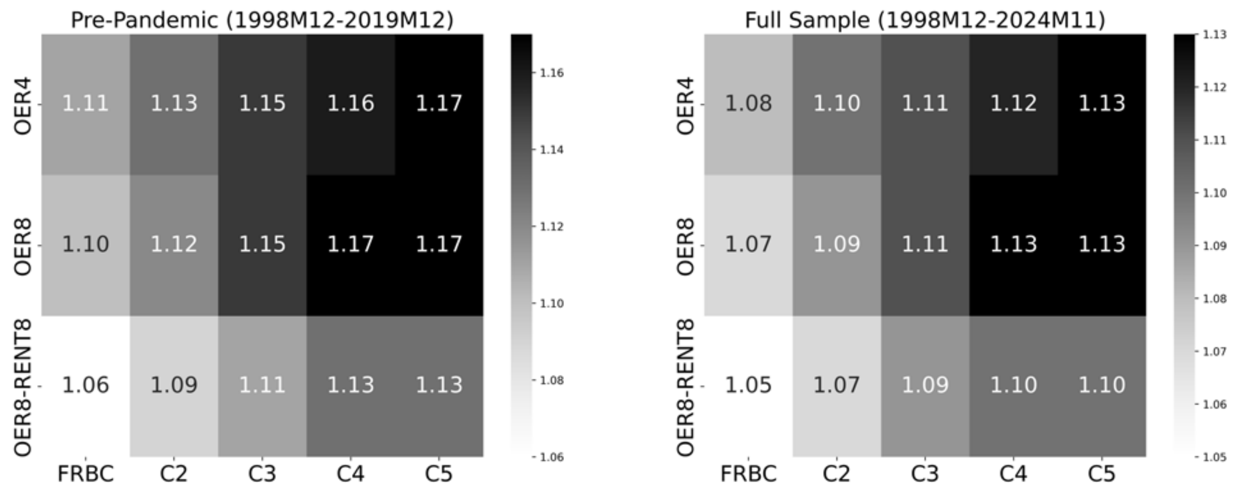


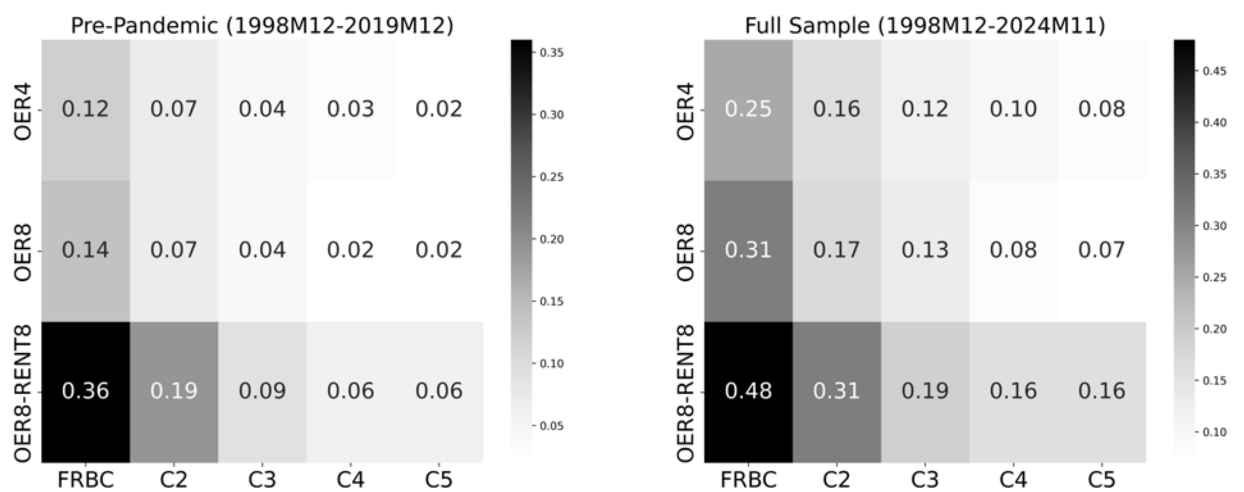
Figure 4: Median Inflation Measures Relative to Mean of CPI Inflation

Panel A: 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.

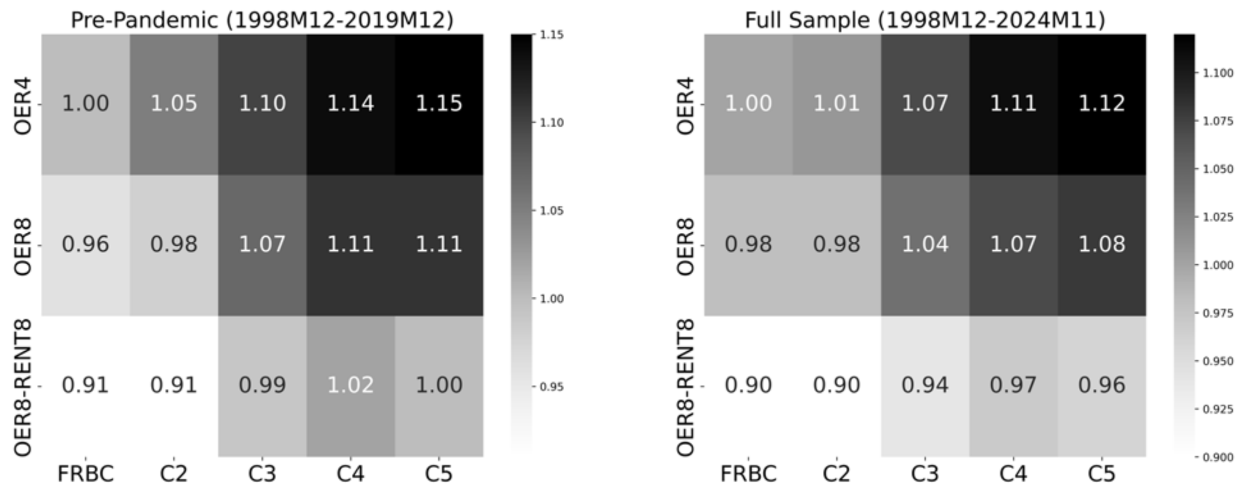
Panel B: p-Values of a Statistical Test of Equal Mean Relative to CPI Inflation



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.

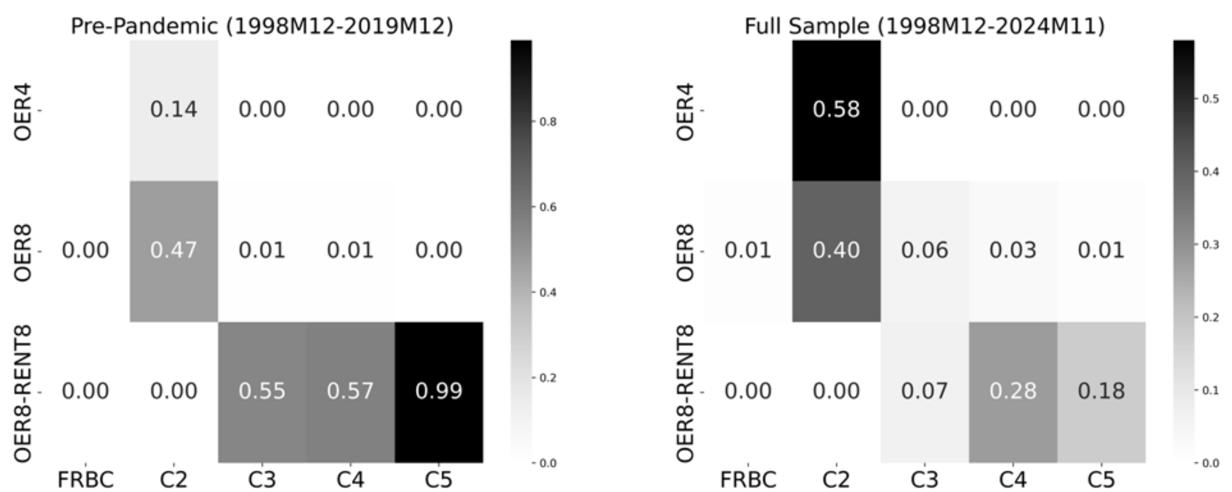
Figure 5: Accuracy of Median Inflation Measures Relative to FRBC-OER4: 37 MMA

Panel A: $RMSE(\hat{\pi}^{37MMA} - \hat{\pi}_j)$ 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.

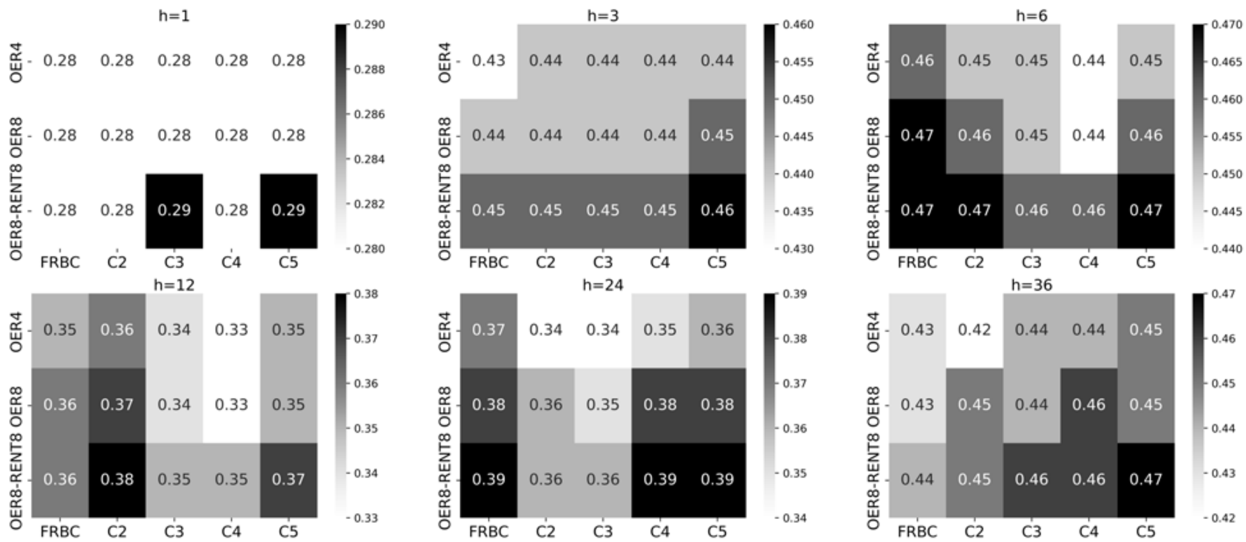
Panel B: p-Values of a Statistical Test of Equal Ability in Tracking $\hat{\pi}^{37MMA}$ for Median Inflation Measures, Relative to FRBC-OER4



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.

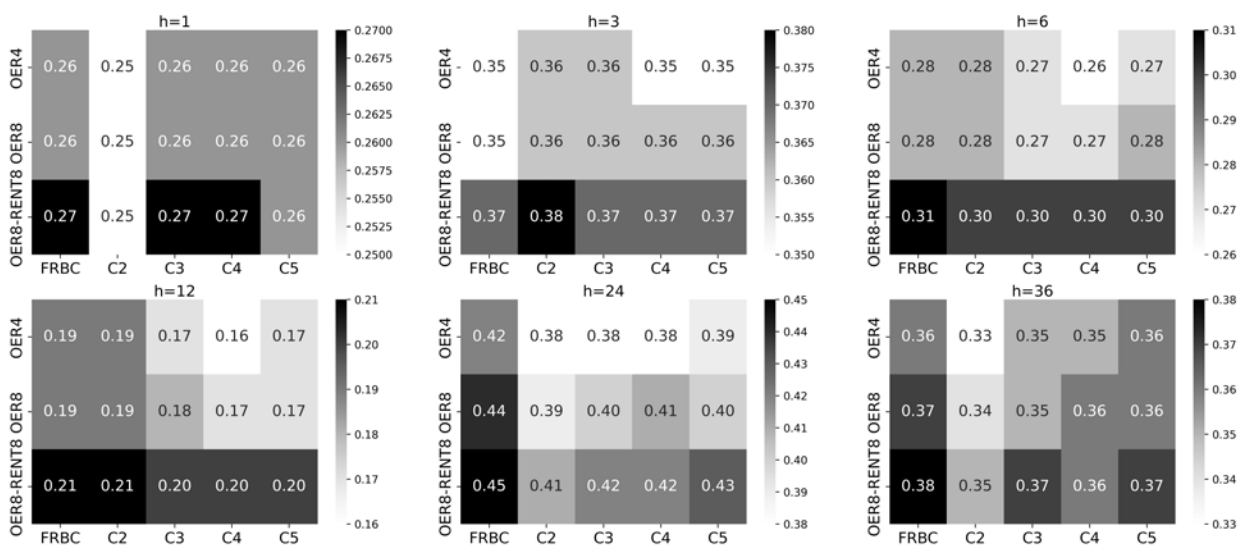
Figure 6: In-Sample Adjusted R^2 of Equation 2

Panel A: 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.

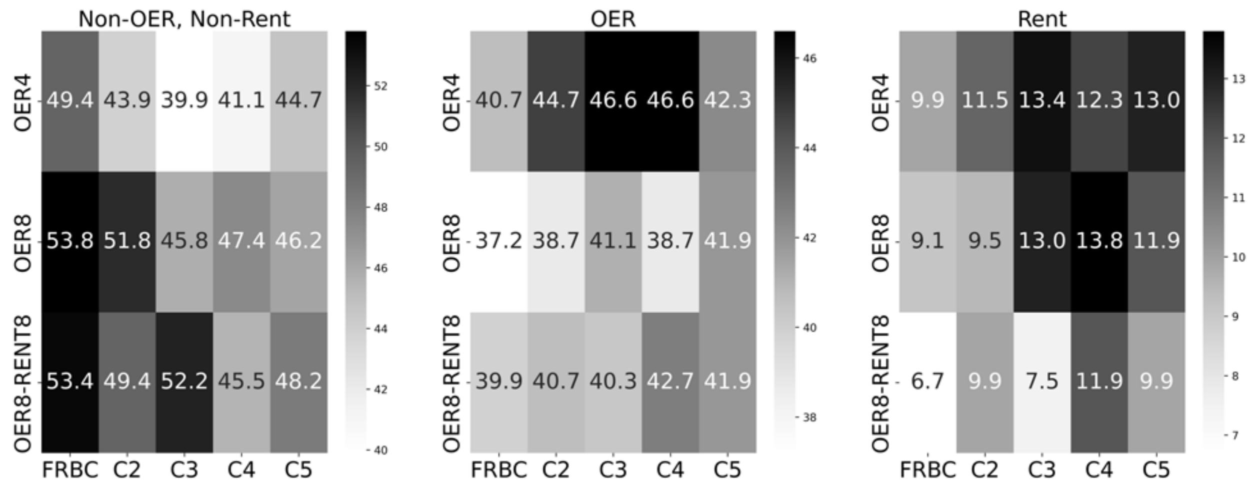
Panel B: Full Sample (1998M12-2024M11)



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.

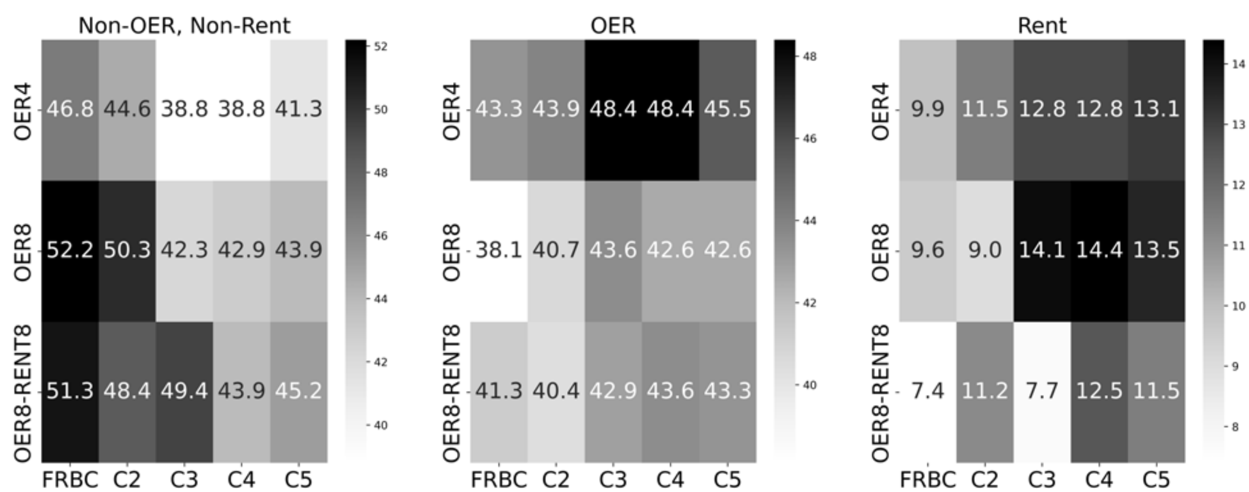
Figure 7: Percentage Frequency of Selection as the Median Component

Panel A: 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.

Panel B: 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.

Figure 8: Comparing FRBC-OER4 and FRBC-OER8-RENT8 Measures of Median CPI Inflation, Pre-Pandemic Sample

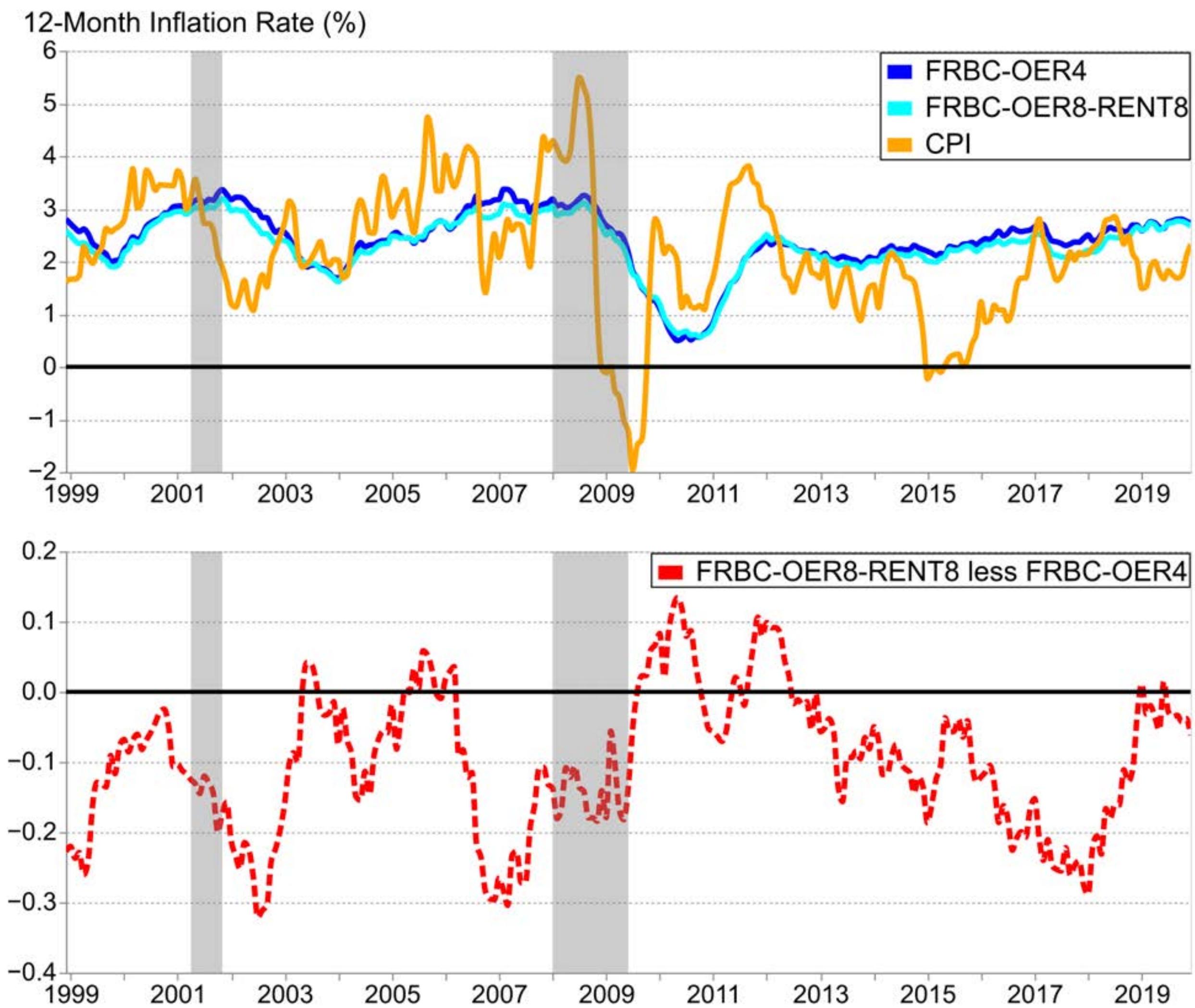


Figure 9: Comparing FRBC-OER4 and FRBC-OER8-RENT8 Measures of Median CPI Inflation, Post-Pandemic Sample

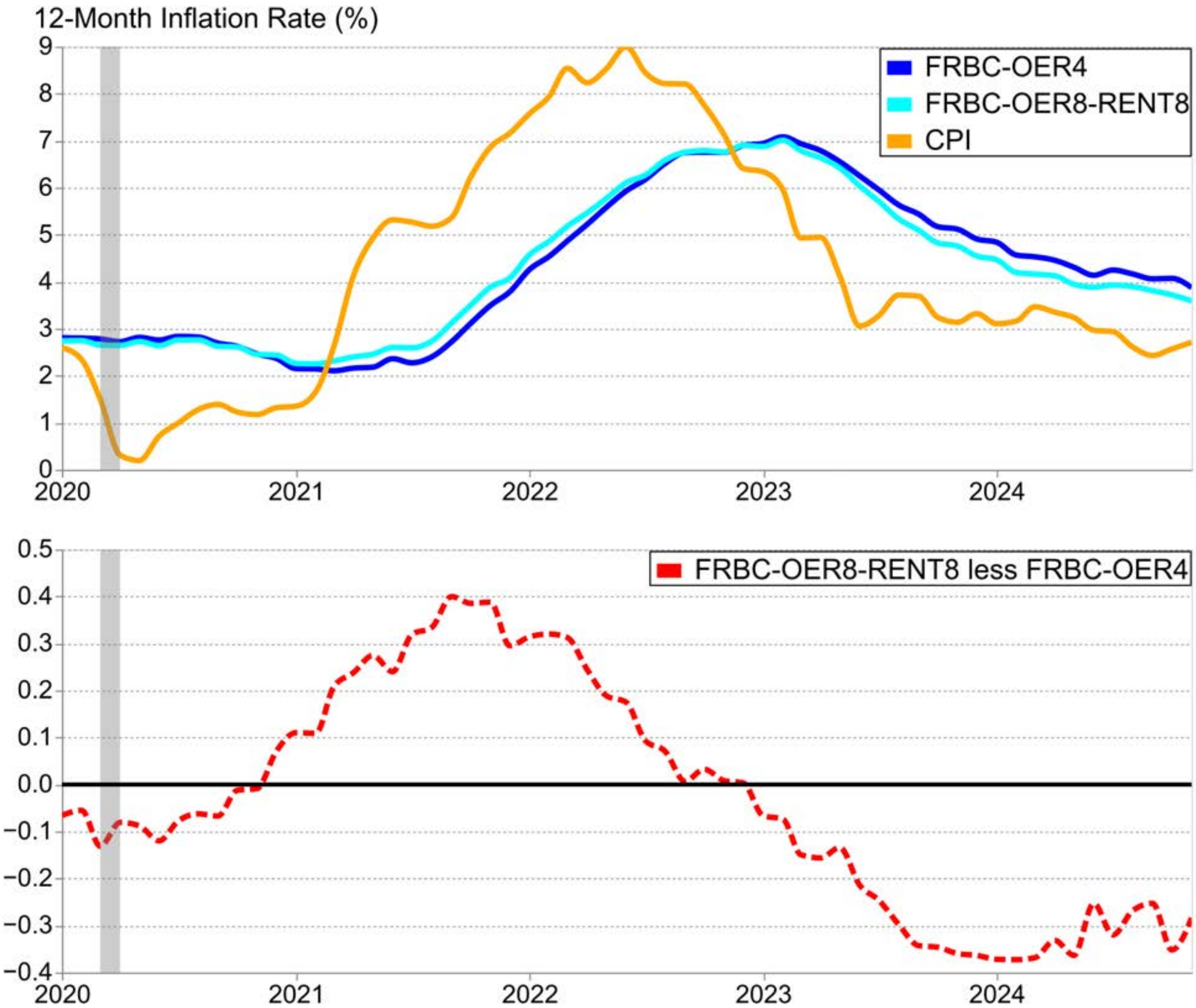


Table 1: Relative RMSFEs of Out-of-Sample Point Forecasts Using Equation 2

Panel A: Pre-Pandemic Sample (1998M12-2019M12)

Forecast Horizon (Months)	h=1	h=3	h=6	h=12	h=24	h=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*****

Panel B: Full Sample (1998M12-2024M11)

Forecast Horizon (Months)	h=1	h=3	h=6	h=12	h=24	h=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. In both panels, 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 at the 10%, 5%, or 1% level, respectively.

Table 2: Relative Log-scores of Out-of-Sample Density Forecasts Using Equation 2

Panel A: Pre-Pandemic Sample (1998M12-2019M12)

Forecast Horizon (Months)	h=1	h=3	h=6	h=12	h=24	h=36
FRBC-OER4 Log score	-2.63	-2.50	-2.13	-1.77	-1.75	-1.61
FRBC-OER8	-0.007*	-0.008	0.011	-0.009	0.018*	-0.002
FRBC-OER8-RENT8	-0.007*	0.003	0.014	0.012	0.036*	0.004
C2-OER4	-0.002	-0.004	-0.010	0.002	-0.033	-0.016
C2-OER8	-0.006	0.001	0.012	0.023*	0.001	0.000
C2-OER8-RENT8	-0.007	0.006	0.037**	0.064**	0.001	-0.003
C3-OER4	0.002	-0.001	-0.001	0.000	-0.041**	0.007
C3-OER8	-0.008*	-0.003	-0.002	-0.001	-0.026*	0.010
C3-OER8-RENT8	-0.001*	0.000	0.021	0.030	-0.011	0.013
C4-OER4	0.001	-0.002	-0.013	-0.023**	-0.028*	0.008
C4-OER8	-0.008	-0.002	-0.011	0.002	0.030**	0.018*
C4-OER8-RENT8	-0.005	0.003	0.014	0.020	0.025**	0.018**
C5-OER4	0.001	-0.003	-0.024*	-0.029**	-0.015	0.025**
C5-OER8	-0.007	0.001	0.001	-0.008	0.013*	0.025***
C5-OER8-RENT8	-0.006	0.006	0.017	0.031	0.022***	0.033***

Panel B: Full Sample (1998M12-2024M11)

Forecast Horizon (Months)	h=1	h=3	h=6	h=12	h=24	h=36
FRBC-OER4 Log score	-2.72	-2.63	-2.57	-2.53	-2.53	-2.33
FRBC-OER8	-0.003	-0.003	-0.006	-0.010	0.037***	0.021*
FRBC-OER8-RENT8	0.002	0.025**	0.061**	0.053	0.036*	0.037*
C2-OER4	-0.008*	-0.005	0.023	0.006	-0.043**	-0.036***
C2-OER8	-0.014*	-0.005	0.033	-0.012	-0.049*	-0.063**
C2-OER8-RENT8	-0.012	0.014	0.076***	0.059**	-0.022	-0.071**
C3-OER4	-0.001	-0.005	0.005	-0.009	-0.054***	-0.035*
C3-OER8	-0.001	-0.001	0.012	-0.006	-0.036*	-0.036*
C3-OER8-RENT8	-0.005	0.010	0.074*	0.081*	-0.012	-0.035
C4-OER4	0.000	-0.008	-0.003	-0.020	-0.072**	-0.034*
C4-OER8	-0.005	-0.004	0.019	0.019	-0.043	-0.054*
C4-OER8-RENT8	-0.004	0.009	0.064**	0.053	-0.031	-0.054*
C5-OER4	-0.003	-0.014	-0.010	-0.042	-0.071**	-0.033
C5-OER8	-0.007*	-0.009	0.006	-0.013	-0.071	-0.027
C5-OER8-RENT8	-0.005	0.004	0.045*	0.034	-0.021	-0.034

Notes: The numbers reported in the first row are the logarithmic predictive score (Log score) from the model with the FRBC-OER4 median measure, while the rows below it are relative Log scores (relative to FRBC-OER4). Thus, a relative Log score that is negative indicates that the model with the FRBC-OER4 measure is more accurate on average than the model being compared. Similarly, the positive value of relative Log score indicates the model being compared is more accurate on average; these cells are highlighted in green. The forecast performance is based on an expanding window of estimation. *, **, or *** denote rejections of the null hypothesis of equal predictive accuracy at the 10%, 5%, or 1% level, respectively, based on the LR test of Amisano and Giacomini (2007).

Table 3: Estimated Phillips Curve Slope

Series	Median CPI			
	Sample: 2000-2019		Sample: 2000-2024	
	$\hat{\beta}$	p-value	$\hat{\beta}$	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
C5-OER4	-0.308	0.00	-0.451	0.00
C5-OER8	-0.283	0.00	-0.422	0.00
C5-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 2024M11 (denoted Sample: 2000-2024). The data from 1999M1 through 1999M12 are used to compute the lagged value of the unemployment rate gap.