Distributional Consumer Price Indices*

Xavier Jaravel

LSE & CEPR

x.jaravel@lse.ac.uk

July 16, 2024

Abstract

This paper develops a new public database providing estimates of inflation heterogeneity across socio-demographic groups in the United States in real time. These distributional CPIs (D-CPIs) are fully consistent with the methodology of the official CPI and are available from 2002 to the present day, by household income, age, race and other characteristics. Using this data set, I establish three results showing that D-CPIs have important implications for the measurement of long-run trends in inequality and poverty, as well as of real wage dynamics during crises. First, I find that "real" income inequality across households between 2002 and 2019 has increased about 45 % faster with D-CPIs, compared to the official CPI. While the income gap between the top and bottom income quintiles increased by 15.6 % during this period according to the official CPI, it increased by 22.6 % with D-CPIs. Similarly large adjustments apply to consumption inequality and inequality in pre-tax and post-tax national incomes. Second, I find that today 2.3 million people are below the "real" poverty line using D-CPIs but above the poverty threshold using the official CPI. This population should become eligible for poverty alleviation programs tied to the poverty line, such as Medicaid. Third, focusing on the inflation burst in the years following the Covid-19 pandemic, I find that inflation was higher for the middle class, compared to low-income and high-income households. This pattern is driven by gas and vehicles and implies that the compression of "real" wages was about 25 % faster with D-CPIs than with the official CPI. Similar patterns of inflation heterogeneity hold in extensions allowing for geographic heterogeneity in inflation, non-homothetic price indices, and studying a longer period. Given that D-CPIs are available in real time (each month) and follow data construction steps that are identical to the official CPI, they can be readily adopted by statistical agencies for the production of statistics on inequality and poverty, for example in the context of distributional national accounts.

Keywords: Inflation; inequality; poverty; non-homotheticities.

*For stimulating conversations and useful feedback and suggestions, I am grateful to Joe Altonji, Orazio Attanasio, David Autor, Raj Chetty, Gabriel Chodorow-Reich, Larry Katz, Ricardo Reis, Emmanuel Saez, and Andrei Shleifer, as well as seminar participants at Berkeley, CREST, Norges Bank, University of Cologne, and Yale. I thank Antoine Ding, Ellen Munroe, Sylvia Tian, and Tyler Woodbury for outstanding research assistance, and UKRI for generous financial support. Finally, I am grateful to Joshua Klick and Anya Stockburger for their availability to discuss the data and methods of the Bureau of Labor Statistics.

1 Introduction

This paper develops a new public database providing estimates of inflation heterogeneity in the United States in real time. While a growing literature documents that there have been persistent gaps in inflation rates across income groups in the United States, two challenges remain unaddressed: (i) the available evidence is typically based on proprietary datasets or new linked datasets that are not necessarily consistent with the official aggregate Consumer Price Index; (ii) inflation inequality estimates are not available in real time.[1] The main contribution of this paper is to develop a new database addressing these two challenges, and to then use the new data to shed new light on the distributional effects of inflation over the past twenty years.

I develop a simple methodology to combine the information contained in high-frequency public data sources—including monthly price changes from the Consumer Price Index (CPI) and annual expenditure shares from the Consumer Expenditure Survey (CEX)— to obtain inflation statistics that can be distributed across socio-demographic groups while remaining consistent with the aggregate CPI. The methodology follows the exact same data construction steps as the CPI, which ensures that it is consistent with official inflation statistics — the only difference is that expenditure shares across product categories are computed by socio-demographic groups (e.g., income percentiles, age, race, urban vs. rural, etc.). Because they are fully consistent with the official CPI but can be disaggregated, I call this price series "Distributional Consumer Price Indices" (D-CPIs). I can thus track the distributional impacts of inflation from 2002 to the present day. All estimates can be updated with each monthly release of inflation data by the Bureau of Labor Statistics, within a few hours.

The new database constitutes a useful complement to "distributional national accounts" (see, e.g., Piketty et al. (2017) and Blanchet et al. (2022)), which have focused on changes in nominal inequality. Distributional national accounts provide inequality estimates that are consistent with macroeconomic aggregates and national accounts, but they use a single price index for all households. My approach extends the logic of distributional national accounts to allow for heterogeneity in inflation rates. D-CPIs are of direct relevance for the measurement of inequality and real wage dynamics, as well as the indexation of transfers, tax brackets and the poverty line. They can also be an important input for various economic applications, such as optimal redistribution with heterogeneous price changes (Jaravel and Olivi (2024)), optimal monetary policy with heterogeneous price changes (Olivi et al. (2024)), or the estimation of heterogeneous price responses to economic shocks.

Using the new database, I establish three main results. First, I analyze long-run trends in inequality before the Covid-19 pandemic. I find that "real" inequality increased about 45% faster with D-CPIs than with the official CPI. With the official CPI, the income gap between the top and bottom income quintiles increased by 15.6% between 2002 and 2019. In contrast, with D-CPIs, the income gap increases much more, by 22.6%. Using the Chained CPI formula to build group-specific price indices yields even larger differences across the income distribution. I also find large adjustments for trends in consumption inequality and inequality in pre-tax and post-tax national incomes. Together, these results show that

[1] Recent work on inflation heterogeneity includes several contributions by academics (e.g., Kaplan and Schulhofer-Wohl (2017), Jaravel (2019), Argente and Lee (2021), Jaravel and Lashkari (2023)), as well as by BLS researchers using confidential data from the Bureau of Labor Statistics (Klick and Stockburger (2021), Klick and Stockburger (2024)). I discuss the relationship between this paper and prior work at the end of the introduction.

D-CPIs can have important implications for the measurement of inequality.

Using decompositions of the price indices, I find that five product categories account for over 60 % of the inflation gap between the top and bottom income quintiles: rent, purchases of new or used vehicles, airline fares, cigarettes, and electricity. The differences across other socio-demographic groups – age, race, urban vs. rural – are less spectacular, although they can be meaningful. In particular, inflation for the 65+ age group has been consistently higher than average: in 2023, there is a 5 percentage point gap between the price index for households above 65 and the official CPI.

Second, I use D-CPIs to adjust the poverty line. The official CPI fails to account for the fact that inflation is higher for individuals in poverty, i.e. the poverty line should be indexed at a higher rate. Using D-CPIs, I find that, by the end of 2023, there are 2.3 million people who are below the "real" poverty line but above the standard threshold based on official CPI. This group should have access to poverty alleviation programs, for example Medicaid. Using D-CPIs is thus of direct policy relevance. Likewise, indexing Social Security retirement benefits on the price index for the population above 65 would have large budgetary implications, with an implied increase in annual retirement benefits of $50bn.

Third, I focus on the period of high inflation that started during the Covid-19 pandemic and the ensuing period of economic recovery, from May 2020 to May 2022. I document that there was meaningful inequality in inflation rates across socio-demographic groups during this period. Specifically, the cumulative inflation rates are inverse U-shaped, increasing from 13% at the bottom of the income distribution to 14.5% for the middle class, and falling back to 14% at the top of the income distribution. These estimates can be used to make adjustments to the compression of wages documented by Autor et al. (2023). Between May 2020 and May 2022, according to the official CPI, wages increased by 2% at the 10th percentile of the income distribution, compared to a fall of 4% at the median, i.e. there was a compression of the income distribution of 6pp. Using D-CPIs, the compression of the real wage distribution at the bottom is amplified by about 1.5pp, i.e. 25% of the baseline measure.

The difference in inflation rates across the income distribution is entirely driven by two product categories that experienced high inflation rates during the sample period: gas and new/used vehicles. Empirically, middle class households on average have higher expenditure shares on these categories, hence these households were more exposed to inflation during this period. Setting aside these categories, inflation rates are fairly homogeneous across the income distribution during the Covid period.

Finally, I present three extensions of the main analysis, which all confirm the patterns obtained in the main analysis. First, I allow for geographic heterogeneity in inflation. This analysis can be carried out with the publicly available data provided by the BLS for a subsample covering about 40 % of total national expenditures. I find that the inflation inequality series remain unchanged. Consistent with these results, using different data Molloy (2024) document that differences in housing inflation and location choices have not generated differences in inflation rates by income group.[2]

In the second extension, I depart from the BLS methodology and introduce non-homothetic price indices. Using the non-parametric algorithm of Jaravel and Lashkari (2023), I find that the non-homotheticity correction is relatively similar across income groups. This result implies that the increase in real income

[2]In contrast, Moretti (2013) documented that, between 1980 and 2000, college graduates concentrated in cities with rising cost of housing, suggesting that real income inequality was dampened relative to nominal income inequality. My analysis is the only one to use the same data as the official BLS series to document the role of geographic heterogeneity for inflation inequality.

inequality and the adjustment of the povery line remain similar to the baseline results.

In the third and final extension, I extend the analysis going back to 1983. Because of limitations with the publicly-available data, I must use a fixed expenditure shares methodology prior to 2002. Using the post 2002 data, I show that this approach should perform well: the level of inflation inequality measured with fixed end-of-period shares is similar to the baseline analysis with expenditure shares updated as in the baseline CPI.[3] I find that inflation inequality was also at play in the 1980s and 1990s and present the implications for the measurement of real households income growth across the income dsitribution, consumption inequality, and trends in pre-tax and post-tax national income inequality. These results illustrate how important inflation inequality can be for the measurement of inequality at longer horizons.

Before proceeding, it is worth highlighting the main limitations of the analysis. Since my goal is to stay as close as possible to the official CPI methodology, my analysis is naturally subject to any limitation affecting this index. The official CPI uses specific expenditure shares, described in Section 2 below, which may lead to substitution bias of potentially different magnitude for different socio-demographic groups. I can however directly address this potential issue by using the Chained CPI. A more fundamental limitation is that I cannot allow for inflation heteogeneity across socio-demographic groups *within* product category, because the BLS does not collect expenditures within categories. Heterogeneity in expenditure patterns and inflation rates can be large within product categories. Prior work has shown that, in the United States in recent decades, within-category inflation has been lower for the rich within consumer packaged goods (Kaplan and Schulhofer-Wohl (2017), Jaravel (2019), Argente and Lee (2021)) and within health care (Jaravel et al. (2024)). Instead, this paper focuses exclusively on estimating the "between category" component of inflation inequality – which can then be combined with "within category" inflation inequality estimates from other, complementary analyses. While prior work suggests that the "within" component may amplify inflation inequality, an important direction for future work is to collect granular data to measure inflation inequality in a larger number of product categories.[4]

This paper contributes to a growing literature on the measurement of inflation inequality. This literature documents that, over the past 20 years, on average annual inflation was lower for higher-income households. However, these papers build price indices that are not entirely consistent with the CPI methodology, making it difficult to use them to correct published statistics (e.g., about income inequality or poverty rates). First, several recent studies (Kaplan and Schulhofer-Wohl (2017), Jaravel (2019) and Argente and Lee (2021)) estimate inflation inequality within consumer packaged goods using Nielsen data, which account for a modest share of total expenditure (below 15%) and differ from the sample frame of the BLS. Besides being able to estimate "within category" inflation inequality, studies of scanner data have the advantage of accounting for inflation inequality arising from changes in product variety: Jaravel (2019) documents that changes in product variety strengthens inflation inequality within this sample. Second, Jaravel (2019) and Jaravel and Lashkari (2023) study the full consumption basket, combining BLS price series to consumption expenditures from the CEX; however, they use data construction steps and price index formulas that deviate from the official CPI.

Third, two papers compute inflation heterogeneity by income quintiles using confidential BLS data

[3]Intuitively, this happens because consumers' demand elasticities across products turn out to be relatively similar, on average, across income groups.

[4]This step would by definition depart from the data and methods of the BLS and is therefore outside of the scope of this paper.

and the same price index formulas as the official CPI: Klick and Stockburger (2021) build price indices by quintiles from 2003 to 2018; in contemporaneous work, Klick and Stockburger (2024) do so from 2006 to 2023. By using the same data and implementing the same index formulas as the official CPI, these papers avoid the approximations I am subject to with publicly-available data.[5] Using publicly-available data has the advantage of flexibility – providing price indices for any socio-demographic group observed in the consumer expenditure survey, from 1983 to the present day – and replicability. My results by income quintiles are close to those of Klick and Stockburger (2021) and Klick and Stockburger (2024) for the periods for which our analyses overlap, which validates the reliability of the approach using publicly-available data.[6]

The remainder of this paper is organized as follows. Section 2 presents the data and methodology to compute D-CPIs. Section 3 presents the main results, discussing in turn the implications of D-CPIs for the measurement of inequality and poverty over the past twenty years, and of real wage dynamics during the inflation burst in the wake of the Covid-19 pandemic. Finally, Section 4 presents the three extensions. Complementary results and methodological discussions are reported in the Appendix.

2 Data and D-CPI Methodology

This section present the main data set: Section 2.1 describes how to replicate CPI with the most disaggregated publicly-available statistics, and Section 2.2 presents my approach to build group-specific CPIs.

2.1 CPI Replication with Publicly-Available Statistics

The key goal is to use publicly-available statistics to build monthly price indices that are specific to particular socio-demographic groups and remain consistent with the official Consumer Price Index. The first step is therefore to replicate the official CPI with the most disaggregated publicly-available statistics.

I first briefly present the methodology of the BLS. I then describe the publicly available statistics used for replication, highlighting the main challenges I face due to data constraints.

A primer on the calculation of the Official CPI. The aggregate CPI is computed each month by the Bureau of Labor Statistics by combining two data sources: monthly price data and expenditure shares. The monthly price data are collected in the Commodities and Services Survey and the Housing Survey, while the expenditure shares use the Consumer Expenditure Survey (CEX).

The monthly price changes are measured by the BLS at the level of about 300 product categories called "entry-level items" (ELI).[7] These category-level price changes are themselves based on the aggregation of thousands of price quotes, as discussed in Appendix A. The expenditure data observed in the CEX use a more detailed product classification, using "universal categorization codes" (UCC), with approximately

[5] The main advantage of Klick and Stockburger (2021) and Klick and Stockburger (2024) is to have access to the full geographic granularity, while I only observe it for 40 % of expenditures. Other differences include the treatment of the CEX data: Klick and Stockburger (2021) and Klick and Stockburger (2024) follow the internal production methods of the BLS (e.g., using smoothed expenditure weights).

[6] Earlier work by BLS researchers on subgroup price indices includes Garner et al. (1996), Cage et al. (2002), and Martin (2022).

[7] Specifically, there are 294 ELIs after 2020, 296 ELIs between 2010 and 2020, and 303 ELIs prior to 2010.

600 UCCs corresponding to the ELIs used in the calculation of the CPI.[8]

To aggregate prices into the aggregate CPI, the BLS works with 243 "basic items", which are slightly more aggregated product categories than ELI, and keeps track of price changes in 32 areas. For simplicity, from here on I will call "items" the $7,776$ basic item-area cells (given by the product of 243 items times 32 areas) which constitute the level of observation at which the BLS applies the expenditure shares measured in the CEX.

To compute the item-level expenditure shares to be used for the aggregate CPI, the BLS takes two main steps. First, in December of every other year, the BLS uses a crosswalk from UCCs to ELIs to assign expenditure shares to each item, using CEX data from prior years. Specifically, prior to 2023, BLS assigned these expenditure shares biennially in December of odd-numbered years using CEX data from the most recent two years prior to the update year. For example, BLS computed a new set of expenditure shares in December 2017 using CEX data from 2015 and 2016.[9] Starting from 2023, in attempt to improve index accuracy and reduce the lag between the incidence and usage of spending data, BLS changed the update schedule of baseline expenditure weights to occur at an annual frequency, using expenditure data from a single calendar year to reflect the spending pattern from two years prior. For example, the CEX micro-data in 2021 are aggregated and used as baseline weights for January to December of 2023.[10] Let us denote these baseline December expenditure weights by $\omega_{i0(t)}$, where the notation "$0(t)$" refers to the most recent pivot December month prior to the month t when baseline expenditure weights are updated.

Second, for the period between the biennial or annual December updates, BLS computes the official CPI by taking the following weighted average with expenditure weights $\omega_{i0(t)}$:

$$\frac{P_t}{P_{0(t)}} \equiv \sum_k \left(\frac{p_{it}}{p_{i0(t)}} \cdot \omega_{i0(t)} \right), \tag{1}$$

where P_t denotes the overall price index in month t and i indexes the item. In words, the cumulative official CPI is a simple weighted average of cumulative price changes $\frac{p_{it}}{p_{i0(t)}}$ for each item i, using the baseline December expenditure shares $\omega_{i0(t)}$ as weights.

It is instructive to note that equation (1) is equivalent to the following monthly price index:

$$\frac{P_{t+1}}{P_t} = \sum_i \left(\frac{p_{i(t+1)}}{p_{it}} \cdot \omega_{it} \right), \tag{2}$$

with the monthly expenditure weights given by

$$\omega_{it} \equiv \frac{\frac{p_{it}}{p_{i0(t)}} \cdot \omega_{i0(t)}}{\sum_k \left(\frac{p_{kt}}{p_{k0(t)}} \cdot \omega_{k0(t)} \right)} = \frac{p_{it}/P_t}{p_{i0(t)}/P_{0(t)}} \cdot \omega_{i0(t)}.$$

[8]The exact number of UCCs varies across years. For example, in 2022 there were 656 UCCs, of which 608 were relevant for the CPI.

[9]For update year 2021, BLS decided to maintain the normal baseline weight updating practice and use CEX data from 2019 and 2020, after considering potential interventions to mitigate the impact on spending behaviour due to Covid-19.

[10]For more information regarding BLS's decision to change the update schedule of baseline spending weights, see `https://www.bls.gov/cpi/tables/relative-importance/weight-update-information-2022.htm`. Appendix B.1 provides more information about the calculation of the expenditure shares in December of every other year, addressing certain simplifications made here in the main text to facilitate reading.

Thus, the BLS uses monthly price changes to infer the way the expenditure shares should evolve across product categories every month. In words, for every item, to calculate the updated monthly expenditure weights ω_{it}, BLS takes the ratio between the item price index in the current month, p_{it}, and in the most recent baseline update month, $p_{i0(t)}$, and multiplies this ratio by the baseline expenditure weights $\omega_{i0(t)}$. The resulting monthly relative importance weights are then renormalized such that they sum up to 1 in the current month.

This formula can be rationalized with a CES price index with elasticity of substitution $\eta = 0$, i.e. the official CPI corresponds to a Leontief utility function: product categories and areas with rising relative prices will be assigned larger imputed expenditure shares over time. The purpose of this procedure is to provide updated expenditure shares in real time, obviating the need to use actual expenditures each month.

The BLS also computes a chained CPI, using actual monthly expenditure shares observed in the CEX for each product category, which may more accurately reflect consumers' substitution behaviors across items. But this index can only be produced with a lag, because the CEX data is released with a one-year lag. I further discuss the chained CPI below.

Publicly-available data and five associated challenges. The analysis requires being able to replicate the CPI calculation using public data only, which raises five challenges.

The first and main challenge pertains to the crosswalk between UCCs and ELIs. BLS only publishes the most recent UCC-to-ELI crosswalk. However, UCCs change frequently over time. While more stable, the set of ELIs used in the CPI calculation also changes from time to time. By contacting the BLS, I could obtain additional crosswalks for 2023, 2022, 2020 and 2010. To create the crosswalks for the remaining years, I used a concordance published by the BLS which tracks how UCCs change over time. These changes happen at the quarterly level, i.e. I created quarterly crosswalks, starting from the ELIs present in the 2023 and mapping the corresponding UCCs to every quarter in the past using the UCC changes concordance.[11] The crosswalk goes back to 1999, making it possible to compute inflation rates from year 2002. I make the final concordance public to facilitate future work on inflation with publicly-available data.

Second, the BLS does not publish any raw price information at the level of ELI. The most granular, complete, and mutually exclusive breakdown of CPI items for which price index data is publicly available at the national level consists of 211 categories called "item strata", out of which 209 are commodities and services, plus 2 strata for housing. Furthermore, the BLS only publishes price indices for item strata at the national level; the local-level price indices used in the construction of the official index are publicly avaiable only for a subsample of the data, covering about 40 % of national expenditures. In the main analysis, I work with national-level price indices for item strata, and I return to geographic heterogeneity in Section 4. There is a simple crosswalk between ELIs and item strata: the first four characters in the ELI code corresponds to the third to sixth characters in the CPI item code. Therefore, I convert the UCC-ELI concordance into a UCC-item strata concordance and conduct the analysis at the level of item strata in everything that follows. Specifically, I compute expenditure shares ω_{it} as described above, except

[11] In general there is a many-to-many mapping between UCCs and ELIs. Most of the time, if one UCC maps to many ELIs I distribute the UCC spending equally among the ELIs. However, in some cases I add specific weights obtained through correspondence with the BLS.

that i now indexes item strata rather than ELIs.

Third, only 181 item strata have published price series; the remaining 30 item strata require proxy price information. There are 26 "unsampled item strata", and 4 strata covered by "Health insurance" (item code *SEME*) for which no price series are available at the individual item level.[12] Unsampled item strata corresponds one-to-one with "unsampled ELIs", a situation occurring when the underlying product or service has reported expenditures in the Consumer Expenditure Survey and is in the scope of CPI, but it is infeasible or impractical to collect price information.[13] I use the price series of the most immediate overarching category for each of the 30 item strata without original price information. For example, I use the price series of *Men's apparel (SEAA)* as proxy for that of item stratum *Unsampled men's apparel (SEAA09)*. Four unsampled item strata do not have a published expenditure weight and are dropped from the analysis. Moreover, the four item strata covered by health insurance map to the same price series. The final dataset thus has 204 unique price series covering the full consumption basket.

Fourth, in some cases price changes are missing in a price series for specific months. Indeed, when the price index of any item fails to meet the publication quality threshold, it will be left out of any BLS publication. While these data points are not publicly available, they are still used by BLS internally in the official CPI calculation. This limitation can be directly addressed by imputing the missing price changes, which I do with a spline interpolation.

Fifth, the CEX expenditure data require various cleaning steps. First, to be consistent with the CEX published summary tables, prior to 2004 I restrict the dataset to households who reported all of their income. After 2004, the BLS started imputing missing income and this restriction is unnecessary. Second, while the analysis requires computing monthly expenditures, for many items the CEX survey respondents only report expenditures over the prior three months. For every respondent, I create a three-month panel, and distribute their expenditures evenly across each month. I use the survey weights so that, when aggregating these monthly expenditures to the yearly level, I match the published CEX summary tables. Note that aggregating expenditures at the monthly level in this way is very different from the methods described in the CEX documentation and sample code, which are only relevant for yearly aggregation. Finally, I must use the OPI dataset from the CEX microdata in order to obtain expenditures on "Owners' equivalent rent of primary residences" and "Owners' equivalent rent of secondary residences", which are not part of the CEX summary tables but which make up a large share of spending for the items relevant for the CPI. Instead of using Owners' equivalent rents, the CEX tracks the costs of home ownership through spending on categories such as mortgage interest payments, insurance and property taxes, all of which are absent from the CPI.

To ensure that I handle the CEX data in a manner consistent with the BLS' practice, I implement several checks. First, I leverage the fact that the BLS publishes the set of weights $\omega_{i0(t)}$ used in pivot

[12] BLS tracks the price change of Health insurance using an indirect approach called the retained earnings method, instead of directly collecting information on premiums, because premium changes should not be compared across insurance plans of varying quality. This method utilizes industry data to determine the percentage of premiums that health insurance companies keep as retained earnings as opposed to payments for medical goods and services to providers. The four item strata covered by Health insurance correspond to different operating mechanisms of health plans (e.g., health maintenance (HMO) plans and Medicare) and the individual price indexes for these item strata are not available, likely due to the proprietary data used for construction. For more details, see the CPI Factsheet on medical care: https://www.bls.gov/cpi/factsheets/medical-care.htm#A2.

[13] For example, purchase of private jets might appear in the CEX expenditure data. BLS does not sample the price of private jets, but instead group it into "unsampled new and used motor vehicles."

months. I use these weights directly when calculating aggregate CPI to ensure that there is no source of error from the CEX; thus, in the main analysis I use the CEX data only to distribute aggregate spending acoss socio-demographic groups. I will also check below that I obtain very similar results when I compute the shares directly from the CEX data. Finally, I will compare the patterns in the cleaned data set to the official CEX summary tables published by the BLS, as discussed in Section 2.2.

Chained CPI. As mentioned above, the official CPI uses a Laspeyres formula with little room for substitution when prices change. To better account for potential substitution patterns, the BLS also publishes a different price index, the Chained CPI. This index uses a combination of a Tornqvist Index and a CES index. The Torqnvist Index uses actual spending shares from the current and previous periods and thus accounts for the observed changes in spending patterns due to price changes. The Tornqvist Index is calculated as follows:

$$P_t = P_{t-1} \prod_{i \in I} \left(\frac{p_{it}}{p_{i(t-1)}} \right)^{\frac{w_{i,t}+w_{i,t-1}}{2}}, \tag{3}$$

where $w_{i,t}$ are the actual monthly expenditure shares taken from the CEX data, with i indexing items.

These expenditure shares are only available with a lag, due to processing time with the CEX data, and so the Tornqvist Index cannot be calculated in real-time. To get around this limitation, the BLS calculates an interim version of the chained CPI using a CES price index, setting a constant elasticity of substitution above zero. Using expenditure and price data from 2003 to 2014, the BLS calculates an elasticity of substitution $\sigma \approx 0.6$ for most years in a regression of the Tornqvist index on the CES price index with free parameter σ. Thus, to update the weights from the most recent baseline period $0(t)$, one can use the CES update formula:

$$w_{i,t} = \left(\frac{p_{i,t}}{p_{i,o(t)}} \right)^{1-\sigma} w_{i,0(t)} \cdot \frac{1}{\sum_{j \in \mathcal{I}} \left(\frac{p_{i,t}}{p_{i,0(t)}} \right)^{1-\sigma} \cdot w_{j,0(t)}}.$$

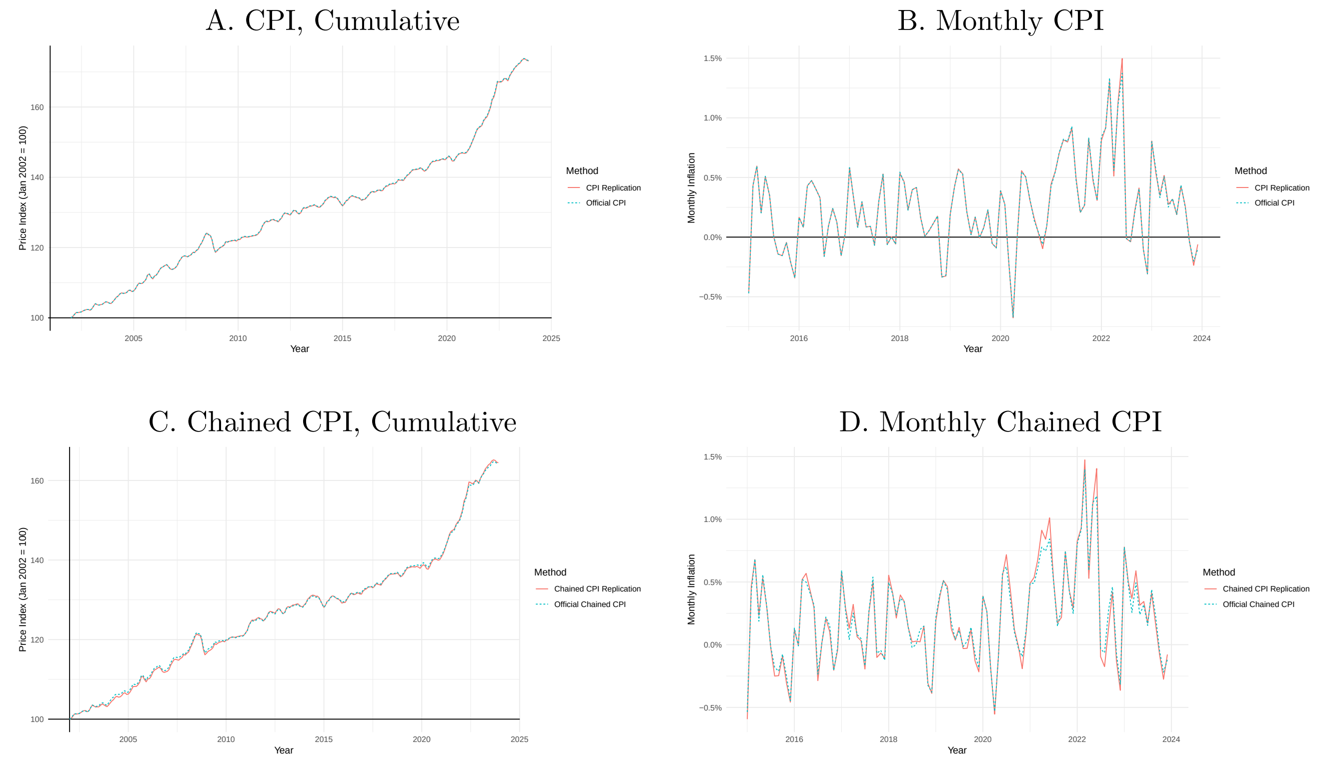
The final CES price index is then calculated as:

$$P_{i,t} = P_{i,t-1} \left(\sum_{i \in I} w_{i,t-t} \left(\frac{p_{i,t}}{p_{i,t-1}} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$$

The Chained CPI uses the Tornqvist Index for all months where actual monthly spending weights are available, creates an interim price index using the CES index afterwards, and then creates a revised final index after the new set of monthly weights become available. I follow the exact same procedure, except that in the baseline analysis I work with national-level price indices at the level of item strata (rather than at the item level as in the official Chained CPI, which features the geographic variation unavailable in public data).

Validation Tests. To sum up, replicating the official using publicly-available statistics can be challenging because of data construction steps – notably the ELI-UCC crosswalk and the data cleaning steps in the CEX survey – and the fact that prices series are only available at the item strata level of aggregation,

Figure 1 Database Validation Tests



Notes: This figure compares the price indices published by the BLS to price indices built using publicly-available data. Panel A and B use the official CPI, reporting a cumulative and monthly index respectively. Panels C and D report the results for the chained CPI.

rather than at the item level, with missing data points in some cases requiring imputation. To assess how well CPI can be replicated, Figure 1 plots the results obtained with the publicly-available data against the official statistics released by BLS, from 2002 to 2023.

I first present the comparison with the official CPI using the Laspeyres formula, in cumulative terms in panel A and by month in panel B. I use the official weights ω_{it} published by the BLS in the construction of our index, such that the potential discrepancies with the official index stem from the fact that I use slightly different price series (at the item strata level rather than for ELIs, with imputation when needed). I find that our index is almost indistinguishable from the official index.

Next, I report the comparison to chained CPI in panels C (cumulative) and D (monthly). I now use the CEX data directly to compute the expenditure shares every month. The indices are again almost indistinguishable, indicating that my treatment of the CEX data is consistent with BLS' practice.

2.2 Computing D-CPIs

Having established in the previous section that official indices can be replicated very well in real time (i.e., every month) using publicly-available data, the next step of our analysis is to distribute the aggregate expenditure shares across socio-demographic groups, so that I obtain group-specific CPIs. This approach can be applied to any socio-demographic group – by income, age, race, urban vs. rural, etc.

To obtain group-specific expenditure shares that remain consistent with macro aggregates, I start from the official set of weights $\omega_{i0(t)}$ used in pivot months and published by BLS. I then distribute these expenditures across socio-demographic groups using the CEX survey, and update the shares in the following month with price data, following the same methodology as the BLS. Specifically, I proceed in four steps:

First, I obtain the official set of weights ω_{i0} published by BLS.

Second, using the UCC to item strata crosswalk, for each item stratum i I compute the share of sales to each socio-demographic group g, denoted $\tilde{s}_{gi0}$.[14]

Third, using the shares $\tilde{s}_{gi0}$, I distribute the official expenditure weight of the item strata (used in the calculation of the official CPI) across groups indexed by g. For example, say that we observe that 25% of sales for the item strata for car purchases are accounted for by households in the top 5% of the income distribution. I then attribute 25% of the aggregate expenditure weight for cars to this household group. This step thus generates expenditure patterns for each household group g across all item strata. Normalizing by the sum of expenditures for each group, I obtain the group-specific expenditure shares $s_{i0(t)g} \equiv \frac{\tilde{s}_{gi0} \cdot \omega_{i0}}{\sum_k \tilde{s}_{gk0} \cdot \omega_{k0}}$, which are fully consistent with the calculation of the official CPI.

The last step is to use the same formula as the BLS to compute the price index for each group, using group-specific expenditure shares $s_{i0(t)g}$ in lieu of $\omega_{i0(t)}$ in formula (1) to obtain the cumulative CPI for each group. Likewise, the monthly CPI is obtained by applying formula (2), with group-specific expenditure shares updated each month, defined as $s_{itg} \equiv \frac{p_{it}/P_{tg}}{p_{i0(t)}/P_{0(t)g}} \cdot s_{i0(t)g}$. I thus obtain price indices by socio-demographic groups each month, using a method consistent with the calculation of the official CPI.

Finally, in robustness analyses I will use a chained CPI index specific to each socio-demographic group. Indeed, I can compute the price index in equation (3) using monthly expenditure shares for each group, which are directly measured in the CEX survey.

Expenditure shares by income groups. Table I summarizes the expenditure patterns by income groups, focusing on December 2013 for illustration. Panel A reports the patterns at the level of eight broad product categories covering the full consumption baskets. At this level of aggregation, there is relatively little heterogeneity in expenditure shares across income groups. Panel B presents the shares for the 10 largest most detailed categories (item strata), depicting much larger spending share heterogeneity across the income distribution. Together, these ten categories account for 54.7 % of total spending in CPI.

Additional results are reported in the appendix. Appendix Table A1 provides a full description of the expenditure patterns across all items strata, by income groups. Appendix Table A2 compares the expenditure weights in the CPI and CEX data for eight broad categories.

[14] I use the official CEX summary tables published by the CEX to check that I obtain the correct group-specific expenditures for each product. The BLS publishes a set of yearly expenditure summary tables by income quintile that I use to validate that the microdata is processed correctly. I implement this check with the expenditures for the twelve most aggregated categories, which only require minor standardization over time.

Table I Expenditure Shares by Income Quintile, December 2013

Panel A: Broad Item Categories

Item Name	CPI Weight	Bottom 5%	1	2	3	4	5	Top 5%
Housing	41.21	42.63	44.19	42.01	40.32	38.92	38.81	39.71
Transportation	16.67	13.53	13.53	16.94	18.61	19.50	18.47	17.84
Food and beverages	15.18	17.79	16.67	15.35	15.35	15.61	14.64	14.20
Medical care	7.21	5.98	7.17	8.47	7.95	7.63	6.35	6.02
Education and communication	6.78	8.13	6.82	5.25	5.49	6.00	8.25	8.50
Recreation	5.95	4.56	4,44	4.94	5.17	5.71	6.61	6.68
Apparel	3.62	3.40	3.30	3.38	3.57	3.34	3.82	3.95
Other goods and services	3.38	3.97	3.89	3.65	3.55	3.30	3.04	3.10

Panel B: Ten Largest Item Strata

Item Name	CPI Weight	Bottom 5%	1	2	3	4	5	Top 5%
Owners' equivalent rent of primary residence	22.78	16.19	18.41	19.94	21.42	22.71	24.22	24.97
Rent of primary residence	6.61	16.42	14.62	10.80	8.08	5.24	2.46	1.89
Gasoline (all types)	5.11	6.20	5.83	6.65	7.03	6.78	5.36	4.69
New vehicles	3.15	0.47	0.66	1.88	2.97	3.42	4.27	4.36
Electricity	2.89	3.53	3.70	3.46	3.06	2.69	2.08	2.02
Full service meals and snacks	2.72	1.90	1.90	2.18	2.50	2.90	3.12	3.26
Motor vehicle insurance	2.53	1.57	2.28	3.12	2.77	3.02	2.16	1.92
Limited service meals and snacks	2.30	2.78	2.33	2.20	2.43	2.52	2.09	1.86
Used cars and trucks	1.86	2.08	1.71	1.95	2.07	2.17	1.80	1.82
College tuition and fees	1.77	3.67	2.39	0.92	1.09	1.19	2.97	3.30

Notes: This table reports expenditure shares for various products across the income distribution. Panel A focuses on eight broad categories covering the full consumption basket. Panel B reports the patterns for the ten largest item strata, which account for 54.72% of total spending.

3 Main Results

This section present the main results. I first describe patterns of inflation inequality over the long run, going back to 2002. I then focus on the recent inflation burst, in the wake of the Covid-19 pandemic.

3.1 Long-Run Inflation Inequality

I examine in turn the extent of inflation heterogeneity by income group and for other socio-demographic group (age, race, urban vs. rural). The results show that the new price series have important implications for the measurement of inequality and the indexation of the poverty line.

3.1.1 Long-run inflation inequality by income percentile

Figure 2 reports inflation across income percentiles from January 2002 to December 2023. I find that inflation rates have been consistently higher for lower-income groups. Panel A show that the gap opens up gradually over time, plotting the full time series for selected income percentiles. Panel B reports the cumulative inflation rates across the income distribution, which ranges from about 84 % at the bottom of the income distribution to about 69 % at the top. Thus, the rate of increase in prices is about 25 % higher for the least affluent households, compared to the most affluent. At an annual frequency, during this period the annual inflation rate was 2.95 % at the bottom of the income distribution, compared to 2.53 % at the top, an annual inflation gap of 41 basis point.

How important are these trends for inequality? To address this question, it is useful to plots household income growth across the income distribution using the official CPI index and the indices accounting for inflation inequality. I use the official statistics of the U.S. Census Bureau to get household income growth by income quintile and for the top 5 %.[15] I focus on the period from 2002 to 2019, stopping the analysis before the onset of the Covid-19 pandemic, before turning to the pandemic period in the next subsection.

Figure 3 shows that, according to the official CPI, household real income growth between 2002 and 2019 was higher in higher income quintiles, ranging from 7.8 % from the bottom income quintile to 24.6 % in the top income quintile, and up to 26.5 % for the top 5 % of households. This gradient becomes considerably steeper with the income-group-specific price indices. After accounting for inflation inequality, household real income growth is only 2.4 % at the bottom of the distribution, i.e. earnings are almost stagnating, while income growth at the top is even faster, at 25.5 % for the top quintile and 27.8 % for the top 5 %.

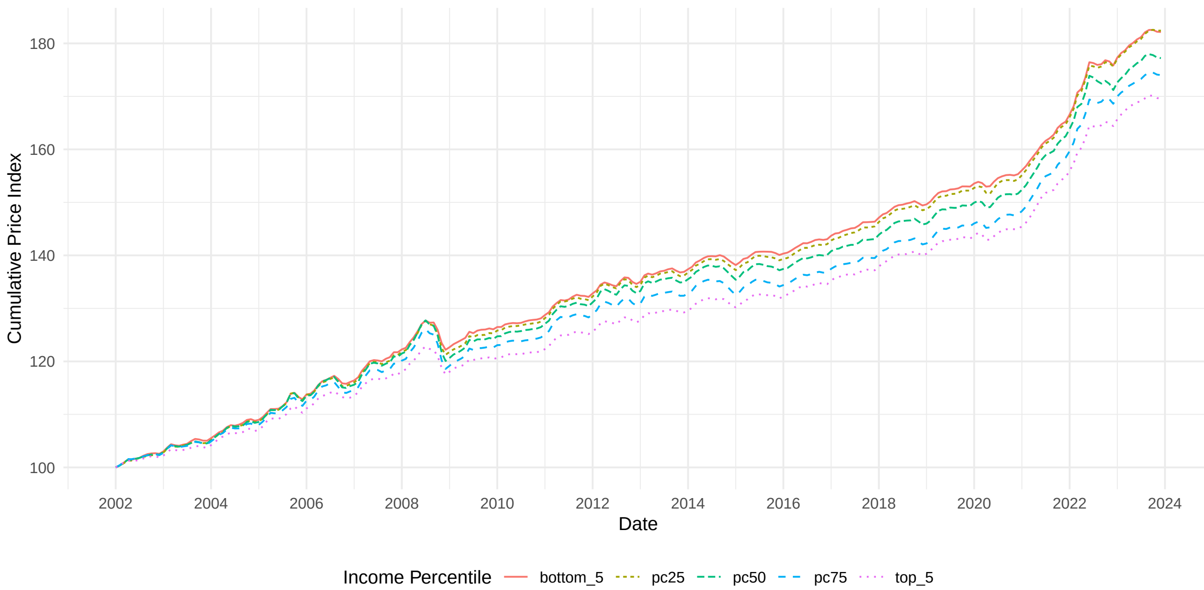
Thus, according to the official metric, the income gap between the top and bottom quintile increased by 15.6 % between 2002 and 2019 ($= 1.246/1.078 - 1$). When accounting for inflation inequality, the income gap increases much more, by 22.6 % ($= 1.255/1.024$): the rate of increase in real income inequality is about 45 % faster than with the official CPI.

In the appendix, I show that the results are qualitatively similar – and somewhat stronger quantitatively – with the Chained CPI formula, i.e. when using monthly expenditure shares. Specifically, Appendix Figure A2 documents that inflation heterogeneity across the income distribution is amplified with the chained CPI formula. The cumulative inflation rates across the income distribution in 2023 ranges from about 80 % at the bottom of the income distribution to about 60 % at the bottom (compared

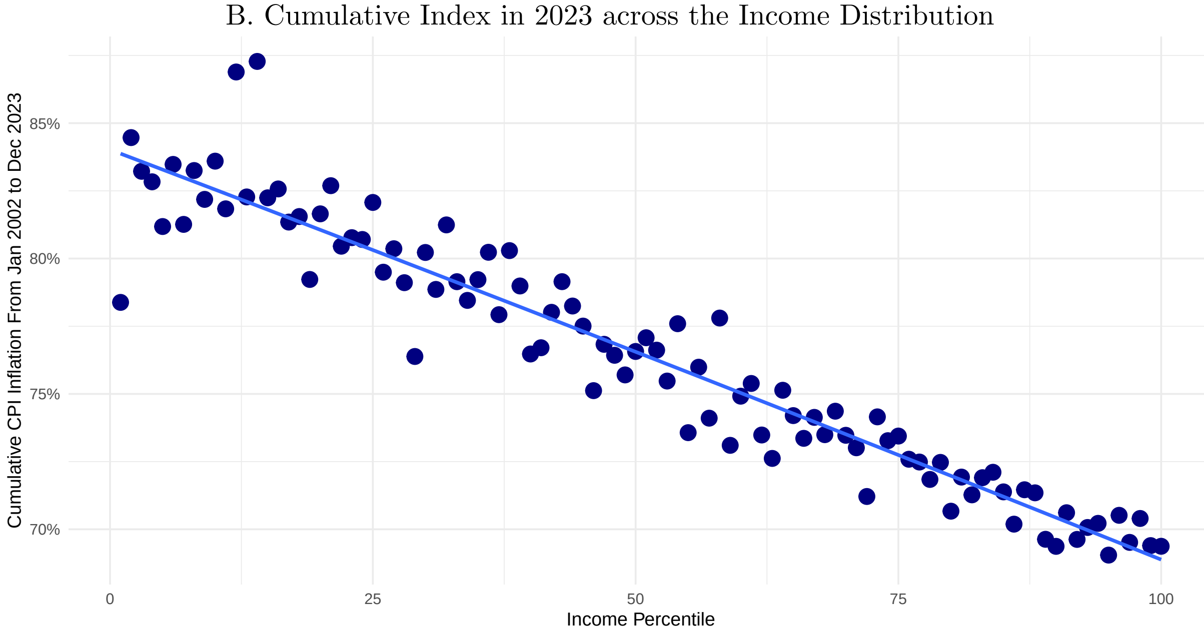
[15] The Census estimates are based on CPS data. I obtain virtually identical results when working with the CPS micro data directly.

Figure 2 Long-Run Inflation Inequality by Income Percentile

A. Cumulative Index from 2002 to 2023 for Selected Income Percentiles

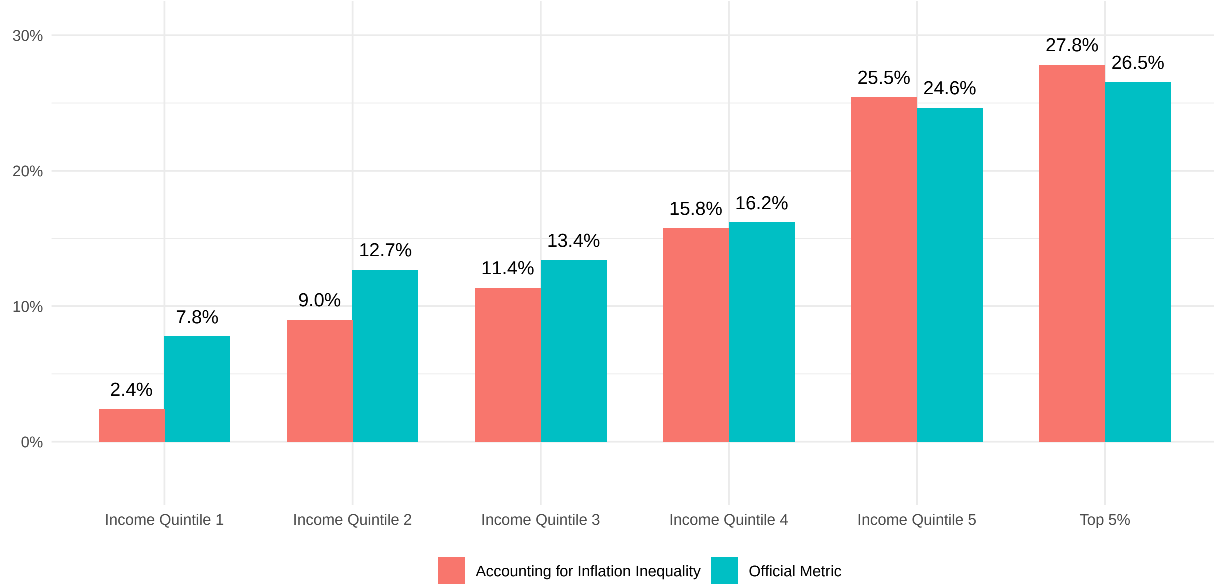


B. Cumulative Index in 2023 across the Income Distribution



Notes: This figure reports inflation rates by income percentile. Panel A show the monthly time series of the cumulative price index from January 2002 to December 2023 for selected income percentiles (bottom 5%, 25th, 50th, 75th, and top 5%). Panel B reports the cumulative CPI in December 2023 for all income percentiles, along with the OLS best-fit line.

Figure 3 Implications for Household Real Income Growth, 2002 to 2019



Notes: This figure reports cumulative real income growth from 2002 to 2019 by quintiles of the household income distribution, as well as for the top 5%. Two series are shown, with the official CPI and with our price indices specific to each income group.

to 84 %—69 % with the baseline CPI formula). Figure A2 shows that, with Chained CPI, the income gap between the top and bottom quintile increased by 15.6 % between 2002 and 2019, while with Chained D-CPIs it increases by 24.8 %. Thus, the rate of increase in real income inequality is about 60 % faster with Chained D-CPIs than with the official Chained CPI: the amplification of inflation inequality is even stronger with Chained D-CPIs than with the baseline D-CPIs.

Next, Table II examines the implications of D-CPIs for consumption inequality. I obtain consumption expenditures by household income quintile directly from the official CEX annual summary tables published by the Bureau of Labor Statistics, and then report the ratio of total consumption expenditures for households in the top and bottom income quintiles over time. With the official CPI, there is only a slight 2 % increase in this ratio between 2002 and 2019. With D-CPIs, there is a meaningful increase of about 8 %, four times as large as the baseline measure with the official CPI. While prior work has examined how measurement issues in surveys may bias estimates of consumption inequality (Aguiar and Bils (2015), Krueger and Perri (2006)), I find that heterogeneity in price indices is also an important factor for accurate measurement of consumption inequality.

Finally, Table III computes the pre-tax and after-tax national income ratios between the top and bottom income quintiles, using the data series of Auten and Splinter (2024). Here as well, the D-CPIs make a meaningful difference in the measurement of trends in inequality. Between 2002 and 2019, the pre-tax national income ratio between the top and bottom quintiles increased by 17 % with the conventional measure but by 24 % with D-CPIs. During this period, the post-tax income ratios fell by 2 % with the traditional measure but increased by 3.8 % with D-CPIs, which again illustrates the importance of inflation heterogeneity for the measurement of inequality.

Table II Trends in Consumption Inequality: Ratio of Top to Bottom Income Quintiles, 2002 to 2019

	Consumption ratios, top to bottom quintiles		% Change in consumption inequality
	2002	2019	2002 to 2019
	(1)	(2)	(3)
With common price index	4.16	4.24	+ 2.05 %
With D-CPIs	4.16	4.49	+ 8.09 %

Notes: Columns (1) and (2) of this table report the ratios of consumption expenditures of households in the top and bottom income quintiles, expressed in 2002 dollars. Consumption expenditures are obtained from the CEX annual summary tables. The first row uses the official CPI to deflate consumption expenditures in 2019, while the second row uses quintile-specific CPIs. Column (3) reports the percentages change in the consumption ratios from 2002 to 2019.

Table III Trends in Pre-tax and After-tax National Income: Ratio of Top to Bottom Income Quintiles, 2002 to 2019

	Pre-tax income ratios			Post-tax income ratios		
	2002	2019	Δ 2002–2019	2002	2019	Δ 2002–2019
	(1)	(2)	(3)	(4)	(5)	(6)
With common price index	15.11	17.70	17.120 %	5.21	5.10	- 2.01 %
With D-CPIs	15.11	18.75	24.14 %	5.21	5.41	+ 3.79 %

Notes: Columns (1) and (2) of this table report the ratios of pre-tax national income for households in the top and bottom quintiles, as defined by Auten and Splinter (2024). The first row is obtained from Auten and Splinter (2024) (Figure 5), while the second row uses quintile-specific CPIs to correct the ratios. Column (3) reports the percentages change in the ratios from 2002 to 2019. Columns (4) to (6) repeat the analysis for post-tax national income ratios.

Table IV Hierarchical Decomposition of the Cumulative Inflation Difference Between the Bottom and Top Income Quintiles, Jan 2002 to December 2023

Level of Aggregation	Between Percentage
Level 1, 8 product categories	33.6%
Level 2, 24 product categories	33.2%
Level 3, 79 product categories	80.3%
Level 4, 204 item strata	100%

Notes: This table reports the within-between decomposition of the log difference in cumulative inflation rates between the fifth and first household income quintiles, using equation (4). The "between" component is mechanically 100% for Level 4, which is the most detailed level of observation, with no within variation.

Table V Item Decomposition of the Cumulative Inflation Difference Between the Bottom and Top Income Quintiles, Jan 2002 to December 2023

Item Name	Share of Inflation Inequality	CPI Weight
Rent (Rent of primary residence + Owners' equivalent rent of primary residence)	24.2 %	29.39 %
Vehicles (New vehicles + Used cars and trucks + Sports vehicles including bicycles + Leased cars and trucks)	15.8 %	5.45 %
Airline fares	9.2 %	0.79 %
Cigarettes	6.5 %	0.75 %
Electricity	6.0 %	2.89 %

Notes: This table reports the within-between decomposition of the log difference in cumulative inflation rates between the fifth and first household income quintiles, using equation (5).

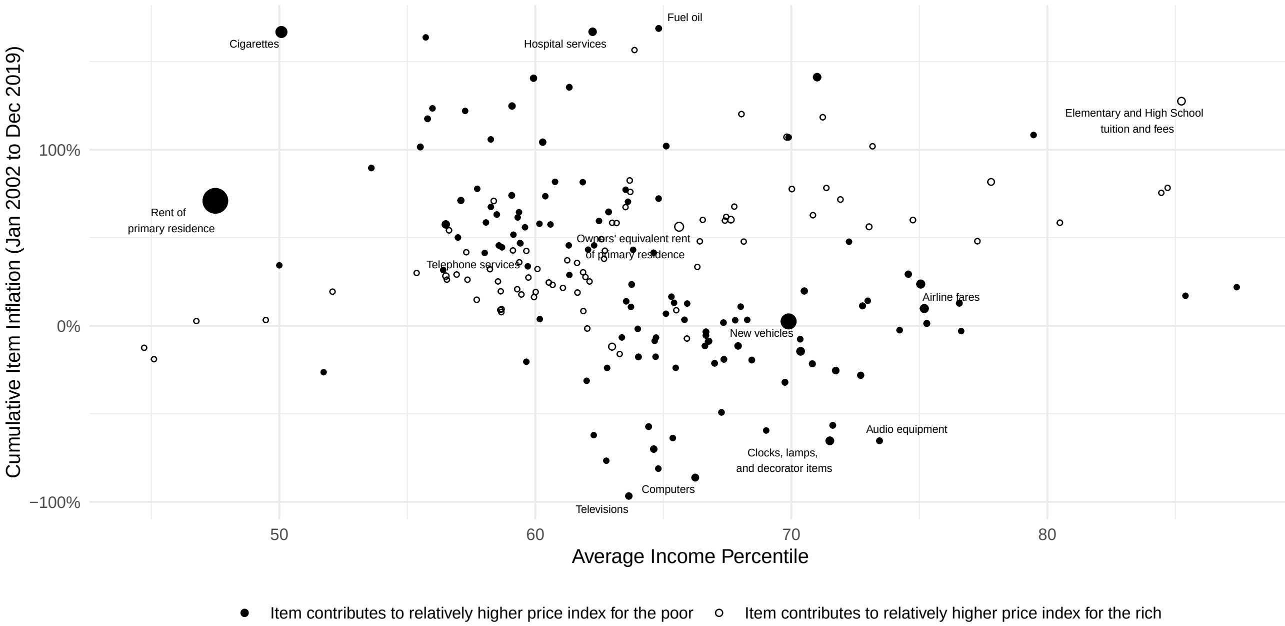
Decompositions Because it is a geometric average, the chained CPI lends itself to convenient, exact additive decompositions that can shed light on the drivers of inflation inequality over time.

First, it is instructive to assess whether broad or detailed categories drive inflation inequality. Using the BLS' official product hierarchy, the 204 items strata can be grouped into 79 product categories ("Level 3"), which can themselves be grouped in 24 categories ("Level 2"), themselves sorted across 8 broad product categories ("Level 1"). For any categorization of products into a set of categories indexed by I, I can decompose the difference in cumulative inflation rates that arises "between" and "within" categories (see Appendix C for the derivation). Specifically, denoting by $\log(P_T^Q/P_0^Q)$ the cumulative price index for household quintile Q, the difference between the top and bottom income quintiles is given by:

$$\Delta\pi_{0,T}^{Q1,Q5} \equiv \log(P_T^{Q5}/P_0^{Q5}) - \log(P_T^{Q1}/P_0^{Q1}) = \underbrace{\sum_{t=0}^{T}\sum_{I}\left(\bar{s}_{i,t}^{Q1,Q5}\Delta\pi_{i,t}^{Q1,Q5}\right)}_{\text{Within}} + \underbrace{\sum_{t=0}^{T}\sum_{I}\left(\bar{\pi}_{i,t}^{Q1,Q5}\Delta s_{i,t}^{Q1,Q5}\right)}_{\text{Between}}, \quad (4)$$

where $\bar{s}_{I,t}^{Q1,Q5} = .5(s_{I,t}^{Q1} + s_{I,t}^{Q5})$ is the expenditure shares of the two household income groups on products in category I; $\Delta s_{i,t}^{Q1,Q5} = s_{I,t}^{Q1} - s_{I,t}^{Q5}$ is the difference in expenditure shares of the two quintile on category I; $\Delta\pi_{i,t}^{Q1,Q5} = \pi_{I,t}^{Q5} - \pi_{I,t}^{Q1}$ is the difference in the inflation rates experienced by the two quintiles within category I; and $\bar{\pi}_{i,t}^{Q1,Q5} = .5(\pi_{I,t}^{Q1} + \pi_{I,t}^{Q5})$ is the average inflation rate experienced by both groups within category I.

Figure 4 Heterogeneous Inflation Rates by Item, 2002 to 2019



Notes: This figure plots the average income percentile of households buying an item (using sales weights to compute the average) against the cumulative inflation rate for this item from 2002 to 2019. The size of each dot is proportional to the contribution of the item to inflation inequality between the bottom and top income quintiles between 2002 and 2019.

Table IV reports the results of the decomposition, studying cumulative inflation between January 2002 and December 2023. The first level of aggregation, with only 8 broad product categories (listed in Panel A of Table I), already captures a third of overall inflation inequality. The "between" contribution does not increase when considering 24 categories rather than the 8 broadest categories. The third level of aggregation, with 79 categories, explains 80 % of overall inflation inequality, i.e. in the CPI sample inflation inequality can be measured quite accurately with relatively aggregate data.

To understand which products drive inflation inequality, it is useful to implement an item-level decomposition. The formula for Chained CPI in equation (3) directly implies an exact additive decomposition of the cumulative inflation gap between the top and bottom income quintiles:

$$\Delta\pi_{0,T}^{Q1,Q5} = \sum_i \left(\sum_{t=0}^{T} \Delta s_{i,t}^{Q1,Q5} \left(\log\left(\frac{p_{it}}{p_{i(t-1)}}\right) - \log\left(\frac{P_t}{P_{t-1}}\right) \right) \right). \tag{5}$$

Intuitively, item i contributes to inflation inequality if it has a higher (lower) inflation rate than average and a higher spending share from the poor (rich).

Table V reports the results for the top five item categories. Together, these products account for 62.5 % of the overall inflation gap between the top and bottom income quintiles from January 2002 to December 2023.

The first row pulls together rents of primary residence (for renters) and imputed rents (for homeowners),[16] which account for about a fourth of the overall difference. This stems from two reasons:

[16] It is useful to pool these two item categories together because they have strong, opposite relationship with expenditure

first, lower-income groups devote a higher share of spending to this aggregated rent categories, which has higher inflation than the rest of the basket; second, inflation is higher for actual rents (with higher spending shares from lower-income households) than for imputed rents (with higher spending shaers from higher-income groups).

Next, purchases of new or used vehicles, as well as leasing, contribute about 15 % of overall inflation inequality. Indeed, these categories have lower inflation than average and higher spending shares from the rich. The same patterns operate for airline fares, which account for 9.2 % of inflation inequality. Cigarettes are also an important source of inflation inequality, at 6.5%: expenditure shares from the poor are higher and the inflation rate for cigarettes is much higher due to rising taxes.[17] Electricity, which has higher expenditure shares from the low-income and higher inflation than average, contributes another 6 %.

While these five categories account for the bulk of inflation inequality, there is large heterogeneity across product categories. Figure 4 plots the heterogeneous contribution of all 204 item strata to inflation inequality. The size of each dot on the figure reflects the size of the contribution to inflation inequality. Black dots correspond to categories that increase inflation disproportionately for the poor, while hollow circles yields relatively higher inflation for the rich. The figure depicts the large heterogeneity across product categories. On average, categories that sell more to the rich have lower inflation, but there are many exceptions. For instance, tuition fees for elementary school or high school have particularly high inflation rates and have a larger weight in the consumption basket of the rich. Despite this wide heterogeneity at the product category level, the price indices for each income percentiles have a clear pattern in Figure 2.

3.1.2 The indexation of the poverty line

Besides the measurement of inequality, the higher rates of inflation for lower-income groups might matter for the indexation of the poverty line and the number of people in poverty. Figure 5 analyzes this question. I use CPS micro data to identify people considered to be in poverty according to the official CPI, comparing an individual's family income to the official poverty threshold.

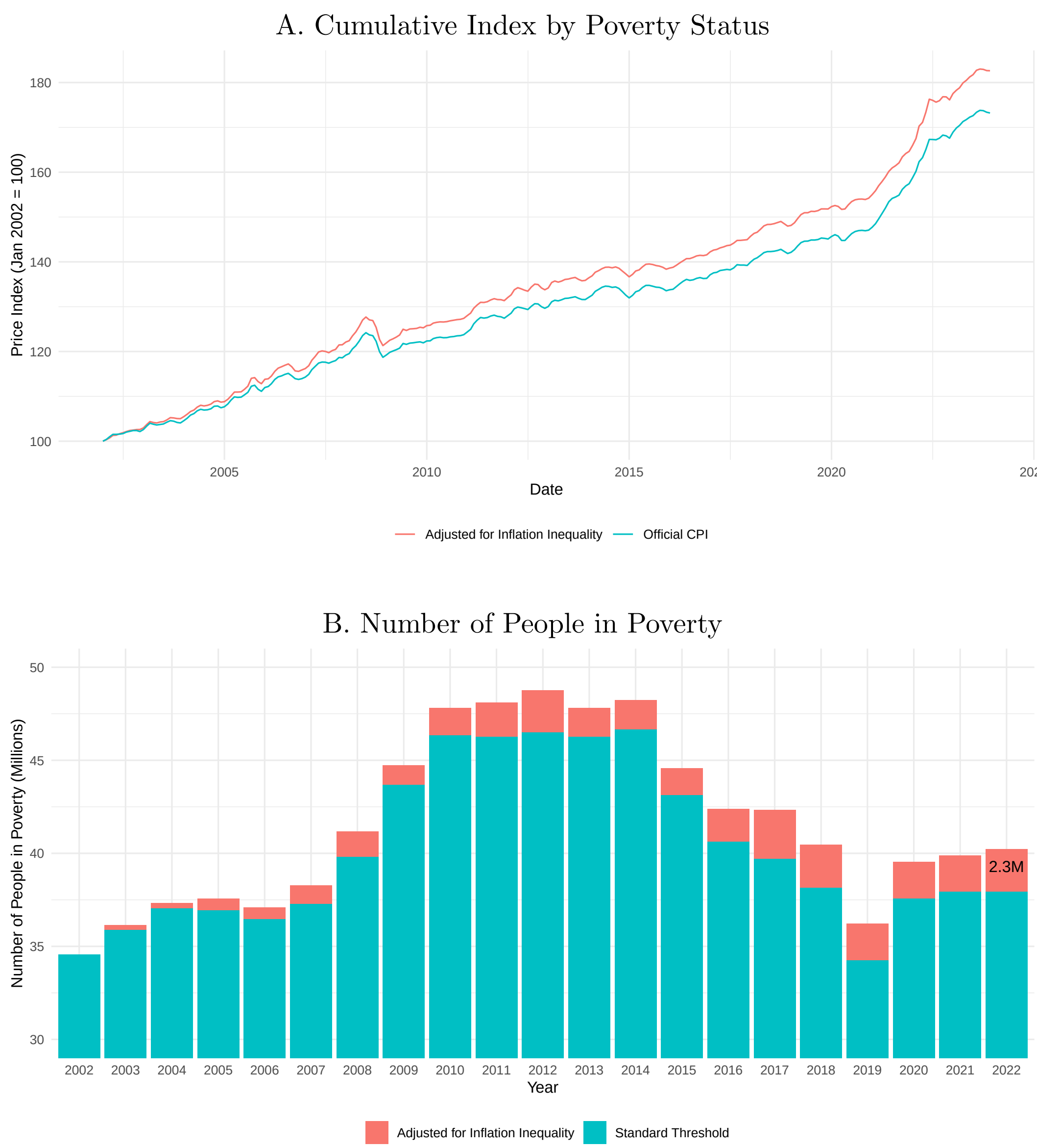
How to compute the price index relevant for the indexation of the poverty line? The official CPI fails to account for the fact that inflation is higher for individuals in poverty, i.e. the poverty line should be indexed at a higher rate. Instead, I use D-CPIs to keep track of the inflation rate experienced by individuals at the poverty line. Specifically, in each year I calculate the change in the price index for households at the 90th percentile of household income distribution conditional on being below the poverty threshold.[18] I implement this calculation iteratively: starting in 2002, I compute a poverty-specific price index which I then use to adjust the official poverty threshold for income inequality in the following year,

shares by income group (see Panel B of Table I).

[17] In a behavioral model with internalities, rising prices for cigarettes may improve the welfare of the poor (see, e.g., Allcott et al. (2019)). While this is an important caveat to keep in mind, I proceed with the methodology of the Bureau of Labor Statistics.

[18] I use the 90th percentile since there is a small tail of household with a large income even after the correction for household size. This occurs since I correct for household income by dividing by the square root of the household size, whereas the official poverty line is based on a more precise correction based on the number of adults and children in the household. This specific choice does not materially affect the results: the findings are similar when computing the average inflation rate for all individuals below the poverty line, rather than focusing on households at the 90th percentile.

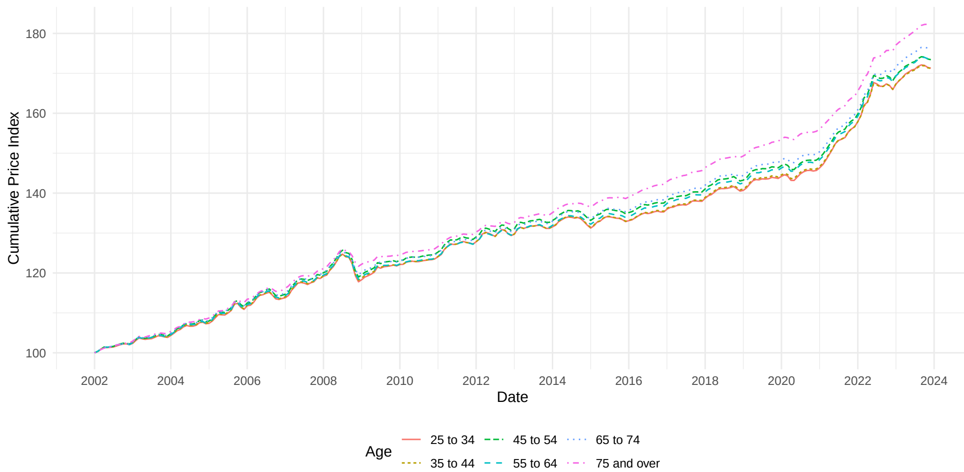
Figure 5 Implication for the Poverty Line



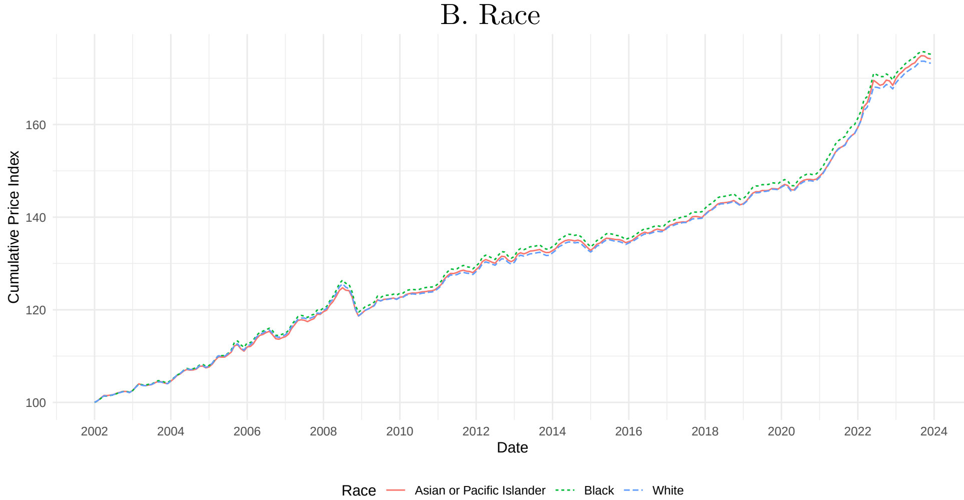
Notes: Panel A compares the cumulative price index from January 2002 to December 2023 for households in poverty to the official CPI. In panel B, we use the price index for households in poverty to index the poverty line over time and report the additional number of people who are under the poverty line (in red), compared to the official metric using CPI (in light green).

Figure 6 Long-Run Inflation Inequality across Other Socio-demographic Groups

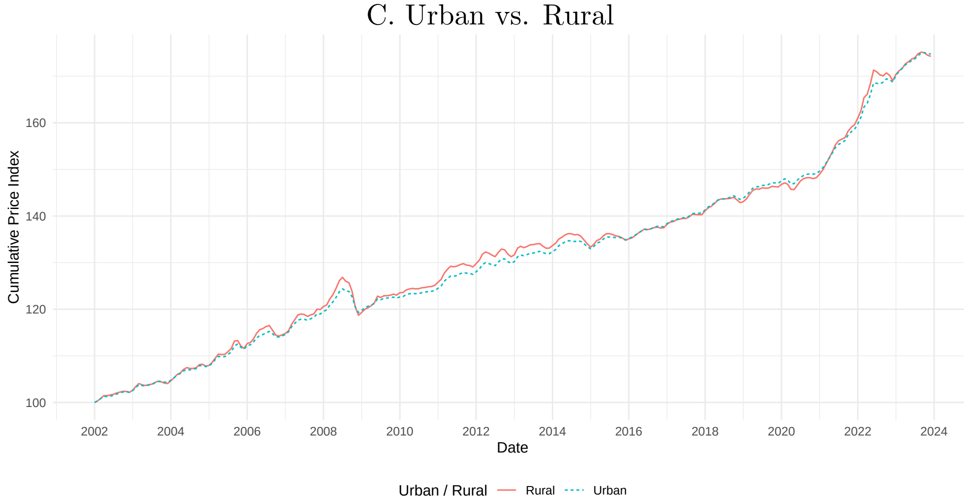
A. Age



B. Race



C. Urban vs. Rural



Notes: This figure reports cumulative price indices from January 2002 to December 2023 for various household groups, by age (panel A), race (panel B), and urban vs. rural households (panel C).

iterating this process year after year.

Panel A of Figure 5 plots the price index for the poverty line, compared to the official CPI. A gap emerges gradually and becomes substantial by the end of the period. Panel B computes the number of people in poverty with the standard threshold and a revised threshold using the price index relevant for the population in poverty. The figure shows that the number of people who are misclassified – i.e., who are considered to be above the poverty line while they are really below – becomes substantial over time. By the end of 2023, there are 2.3 million people who are below the "real" poverty line but above the standard threshold based on official CPI. This group should have access to poverty alleviation programs, for example Medicaid. Using D-CPIs is thus of direct policy relevance.

3.1.3 Long-run inflation inequality across other socio-demographic groups

Figure 6 reports inflation heterogeneity by age, race, and for urban vs. rural households.[19]

Panel A shows that, from 2002 to 2023, inflation rates were higher for older households. The cumulative price index in 2023 is about ten percentage point higher for households above the age of 75, compared to those between 25 and 34.

Panel B reports the patterns by race. A gap gradually emerges over time, with higher inflation for African-American households. Whites have the lowest inflation rate, while Asian households experienced a slightly larger inflation rate compared to Whites. These differences are however relatively modest compared to those observed across income groups.

Finally, in Panel C we document the difference in inflation rates between urban and rural households. The figure shows little difference. While gaps open in specific periods – for example right after the Covid-19 pandemic, which I investigate further below –, the differences appear to be relatively short-lived.

For the indexation of government transfers, it is particularly instructive to compute a price index for households above the age of 65, who are eligible to receive Social Security Retirement benefits. Figure 7 shows the results, documenting that the population above 65 experiences higher inflation rates between 2002 and 2023. In 2023, there is a 5 percentage point gap between the price index for household above 65 and the official CPI. Indexing Social Security Retirement benefits on this alternative price series would have large budgetary implications. The total cost of annual Social Security Retirement benefits was about $1 trillion in 2023, i.e. using an age-specific index would increase the total cost of pension benefits by $50bn. When using the Chained CPI instead of the official CPI formula, the inflation gap increases further, to 6.7 percentage points.

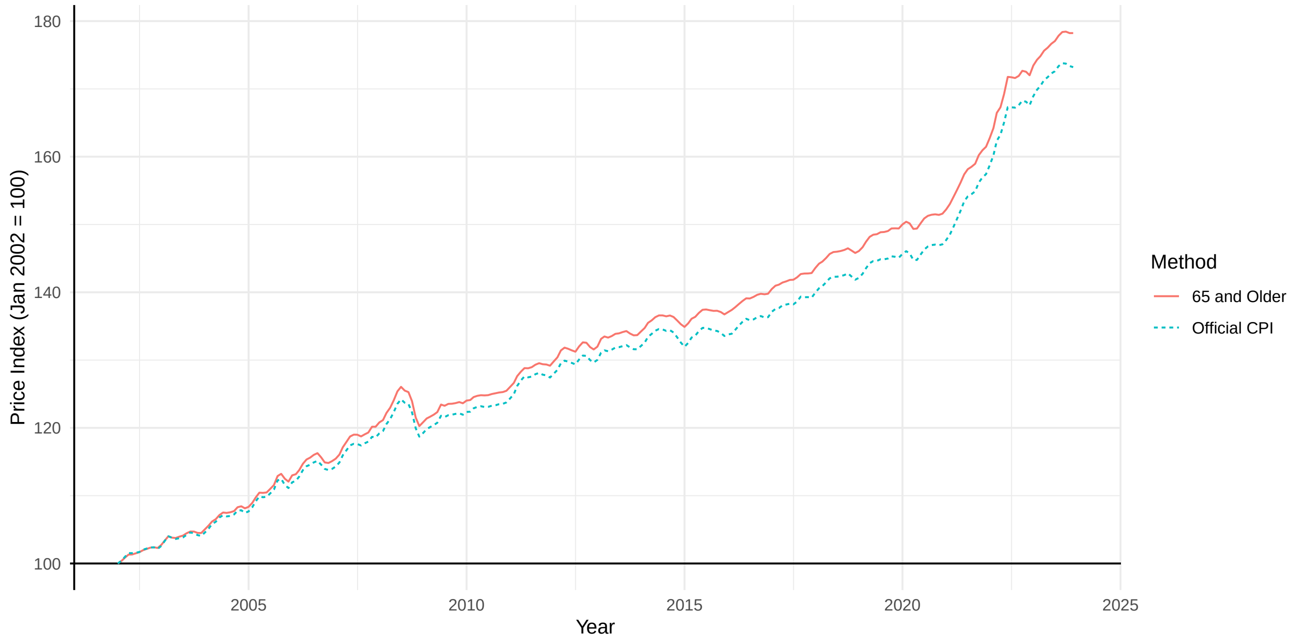
3.2 Inflation Inequality after the Covid-19 Pandemic

Having documented long-run trends in inflation inequality, I now focus on inflation heterogeneity across groups right after the Covid-19 pandemic, from May 2020 to May 2022. Inflation was high during this period, in particular because of two product categories, gas and new/used vehicles (see Appendix Figure A1).

Figure 8 plots the results by income percentile. Panel A, considering all products, shows an invert U-shaped pattern: inflation was a bit higher for the middle class during the inflation burst. While

[19] I use the age and race of the reference person to build these price indices.

Figure 7 65+ D-CPI and Official CPI



Notes: This figure reports cumulative price indices from January 2002 to December 2023 for various households with a household head above the age of 65, compared to the official CPI.

cumulative inflation between May 2020 and May 2022 was 13% at the bottom of the income distribution, it was about 14.5% at the 50th percentile, and 14% at the top.

These estimates can be used to compare the compression of "real" wages during this period to the compression of nominal wages documented by Autor et al. (2023). Between May 2020 and May 2022, according to the official CPI, wages increased by 2% at the 10th percentile of the income distribution, compared to a fall of 4% at the median, i.e. there was a compression of the income distribution of 6pp. From Panel A of Figure 8, this compression is amplified by about 1.5pp when we account for inflation heterogeneity. Thus, the compression of the real wage distribution at the bottom is amplified by 25%.

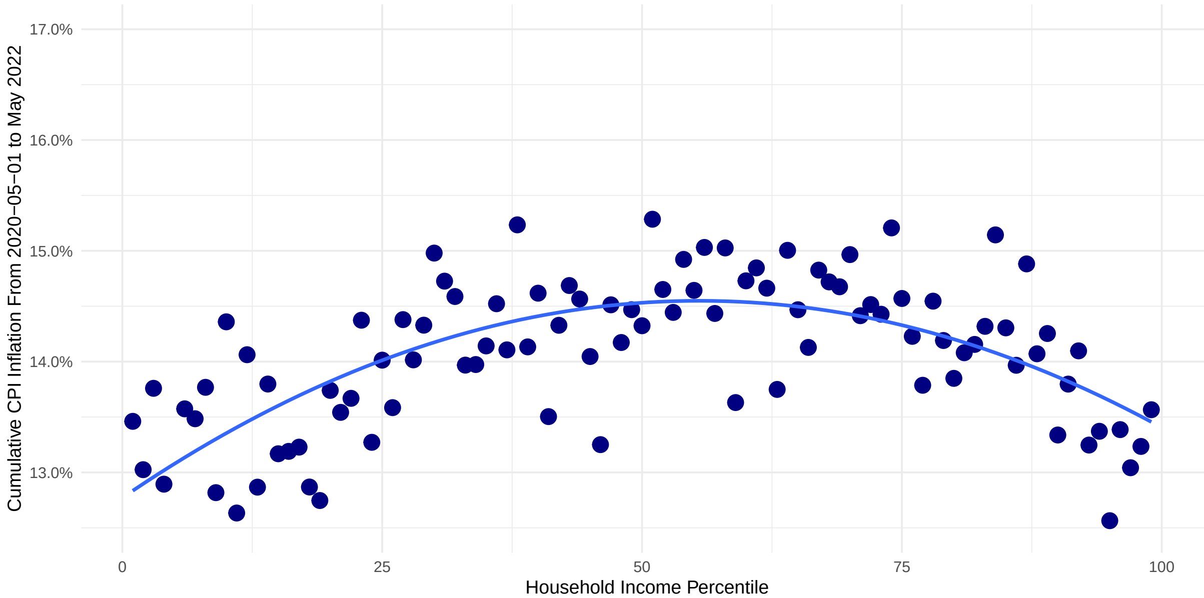
Panel B of Figure 8 shows that this pattern of inflation heterogeneity in the wake of the Covid-19 pandemic is entirely driven by two product categories. When excluding gas and new/used vehicles, there is no difference in inflation rates across the income distribution during this period. People who drive were particularly hit by the increase in gas prices and by high inflation rates for cars – caused by the semi-conductor crisis in 2021-2022.

Next, Figure 9 repeats the analysis by age, race, and for urban vs. rural households. Panel A shows that younger households experienced significantly higher inflation between May 2020 and May 2022. Panel B documents that White households faced somewhat lower inflation during this period – the cumulative inflation rate was about 1pp higher for African-Americans or Asian as of May 2022. Finally, panel C document substantial differences between urban and rural households. As expected, rural households – who more frequently need to drive – experience higher inflation rates. Their cumulative inflation rate was 2.5pp higher than that of urban households as of May 2022.

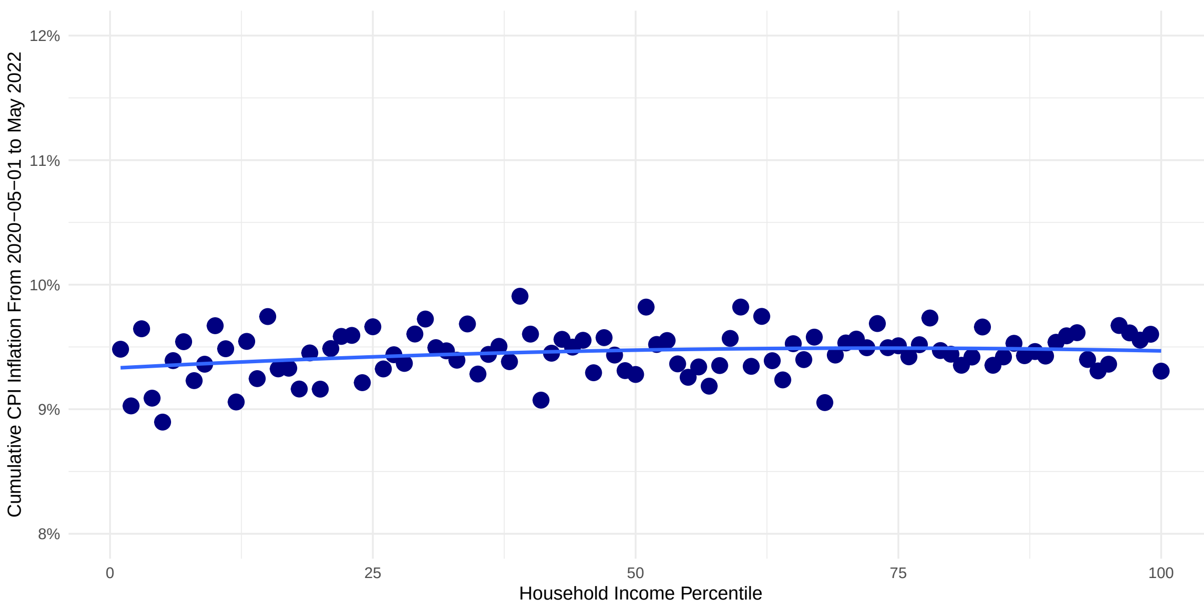
Overall, these results show that the patterns of inflation heterogeneity can vary across periods and are not always in line with the long-run trends documented in Section 3.1. While the poor experience higher inflation in the long run, during the pandemic the middle class was hit more strongly. This illustrates the usefulness of benig able to compute D-CPIs at a monthly frequency.

Figure 8 Short-Run Inflation Inequality by Income Percentile

A. All products, May 2020-May 2022

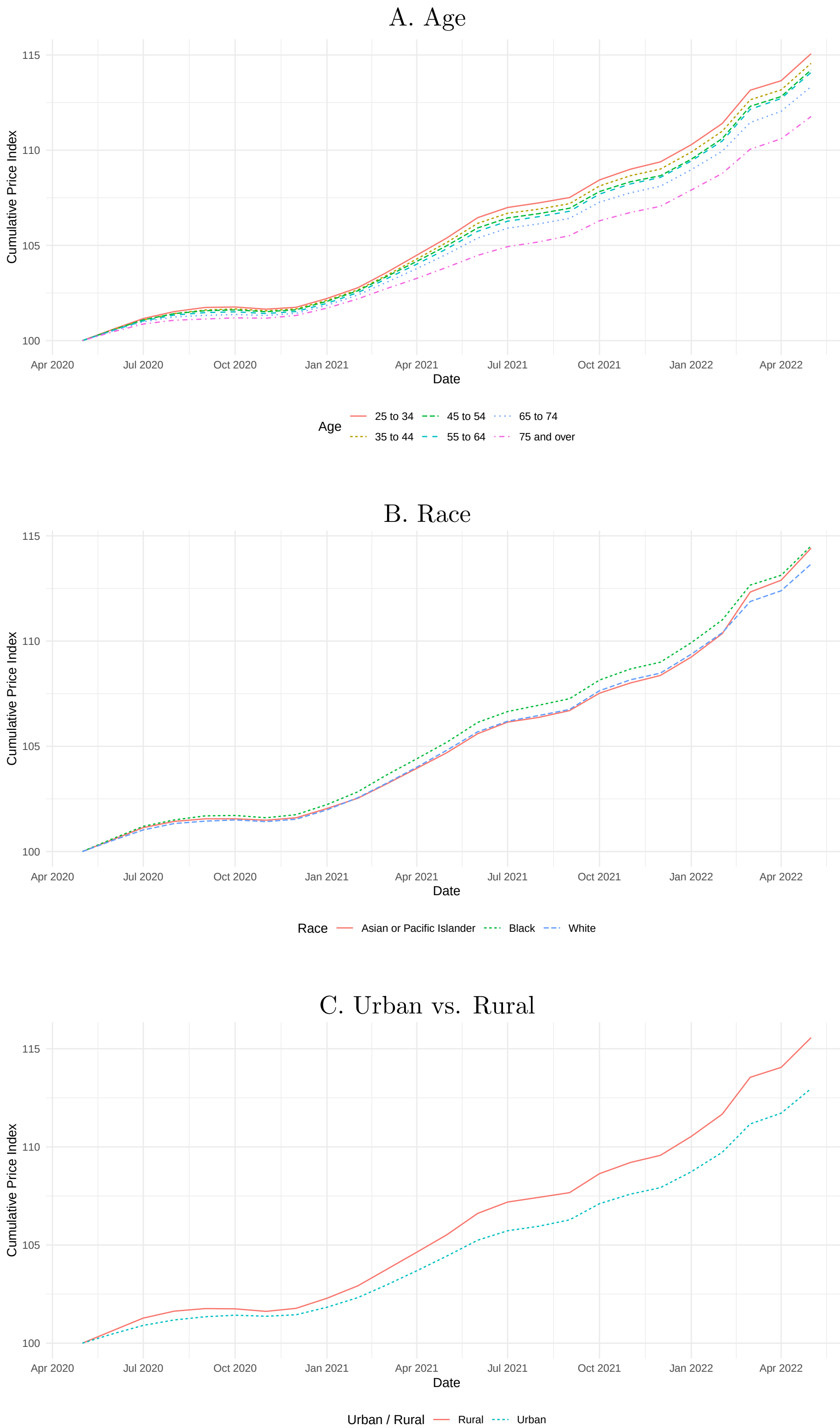


B. Excluding Gas & Vehicles, May 2020-May 2022



Notes: This figure plots the cumulative price index by income percentile from May 2020 to May 2022. While panel A includes all products, panel B excludes gas and vehicles.

Figure 9 Short-Run Inflation Inequality by Socio-demographic Groups, May 2020 – May 2022



Notes: This figure reports cumulative price indices from May 2020 to May 2022 for various household groups, by age (panel A), race (panel B), and urban vs. rural households (panel C).

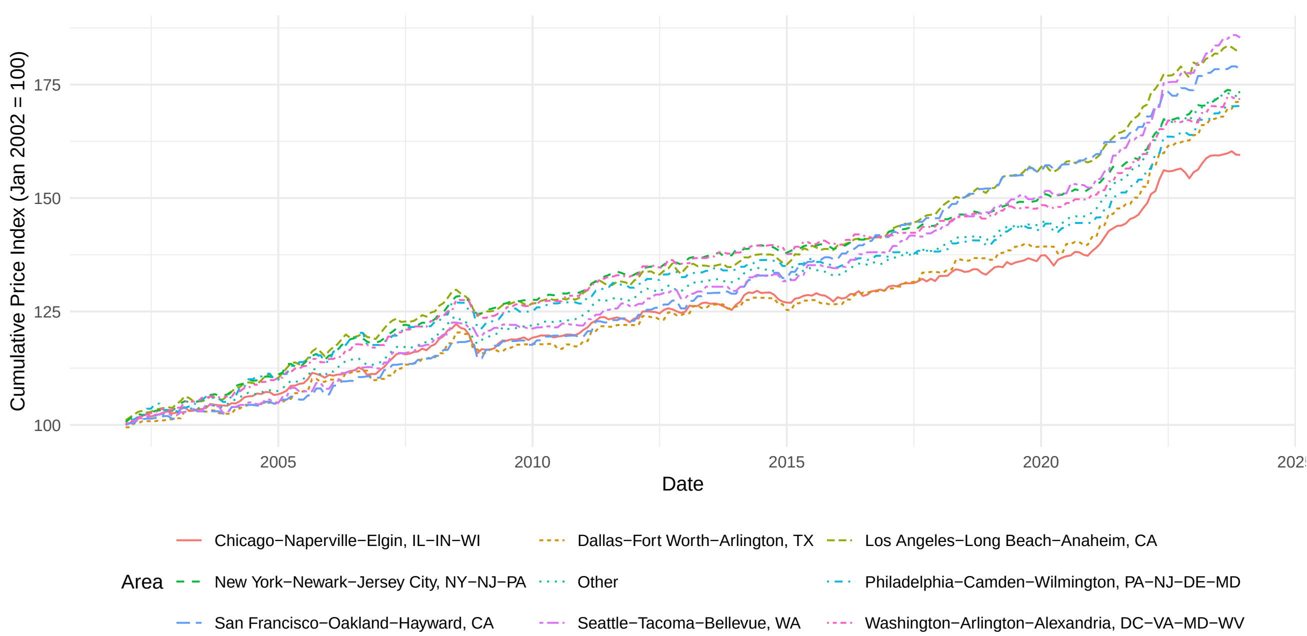
4 Extensions

I now present a series of extensions, first allowing for geographic heterogeneity in inflation, then presenting the results with a non-homothetic price index, and finally extending the analysis further back in time. All extensions confirm the patterns presented in the main analysis, with systematically lower inflation for higher income groups.

4.1 Allowing for Geographic Heterogeneity

I now examine whether the results change when price dynamics are allowed to vary across cities. This analysis can be conducted for a subsample of the data covering 23 cities accounting for 40% of total national expenditures, for which the BLS makes the local price series publicly available. Figure 10 plots the price series for major cities between 2002 and 2023. Seattle has the highest inflation rate, while Chicago has the lowest during this period.

Figure 10 Inflation Heterogeneity Across Selected Cities

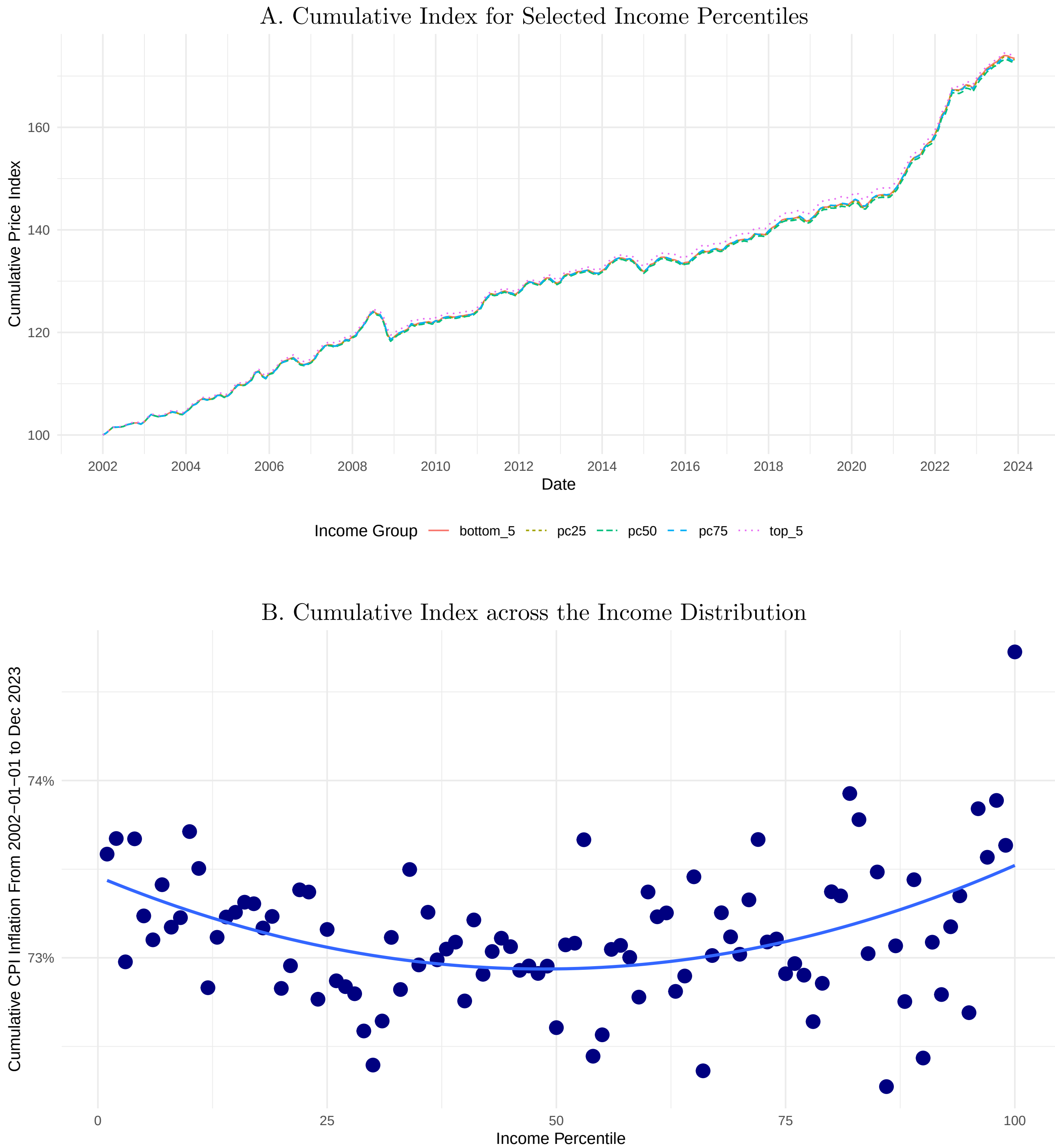


Notes: This figure reports cumulative inflation rates in a selected set of cities.

Using this data, I implement the same methodology as for the main analysis, except that i now indexes items in a specific city. Appendix B.2 describes the data construction steps to allow for geographic heterogeneity. The analysis thus takes into account that inflation rates may be different across space, generating potential differences in inflation across income groups, who are unequally sorted across cities.

To assess the importance of geographic heterogeneity, I start by computing the level of inflation inequality that arises only *between* cities, i.e. I compute price indices where all income groups are assumed to experience the same inflation rates within cities but have unequal expenditure shares across cities, as measured in the CEX data. Figure 11 reports the results. Panel A reports the cumulative index over time for selected income percentiles, which all experience very similar inflation rates throughout the period.

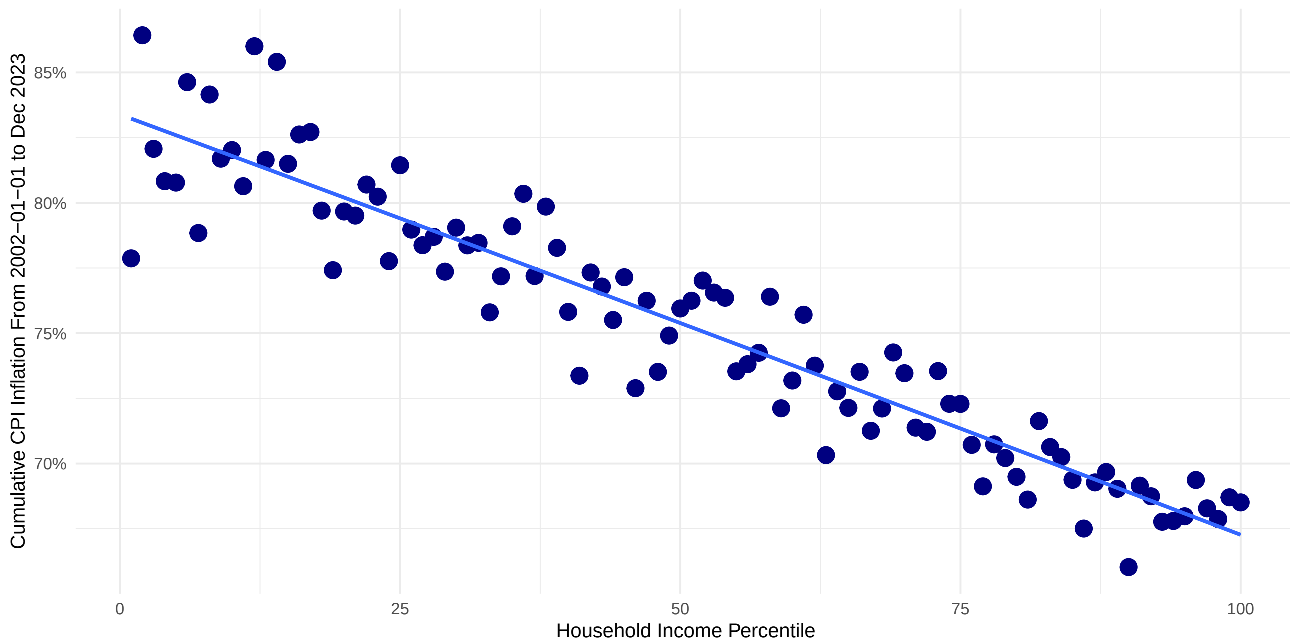
Figure 11 Inflation Inequality from Geographic Heterogeneity Alone



Notes: This figure reports inflation rates by income percentile. Panel A show the monthly time series of the cumulative price index from January 2002 to December 2023 for selected income percentiles (bottom 5%, 25th, 50th, 75th, and top 5%). Panel B reports the cumulative CPI in December 2023 for all income percentiles, along with the OLS best-fit line. In both panels, only the heterogeneity in inflation rates arising *across* 23 cities is taken into account.

Figure 12 Inflation Inequality including Geographic Heterogeneity

A. Cumulative Index for Selected Income Percentiles



Notes: This figure reports inflation rates by income percentile, inclusive of geographic heterogeneity.

Panel B reports the cumulative inflation rate from 2002 to 2023 by household income percentile, showing a flat pattern. Thus, geographic heterogeneity does *not* affect inflation inequality. I confirm this result in Figure 12, computing overall inflation inequality inclusive of geographic heterogeneity. The results are essentially unchanged compared to the baseline results without inflation heterogeneity (Figure 2).

In prior work, Moretti (2013) built city-specific CPIs and documented that, between 1980 and 2000, college graduates concentrated in cities with high cost of housing, suggesting that inequality in purchasing power is lower than commonly thought based on nominal wage differences. I instead study a different period, focusing on income groups. Consistent with my results, Molloy (2024) different housing and location choices have not generated materially different shelter components of inflation across the income distribution.[20]

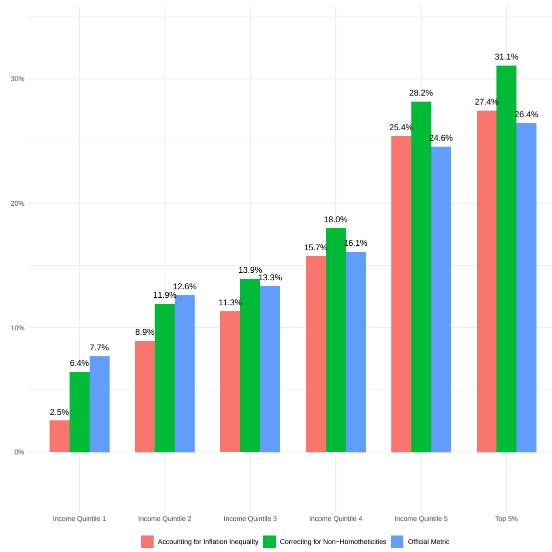
4.2 Nonhomothetic Price Index

The analysis so far uses homothetic price indices, in line with BLS methods. Even though price indices were computed by income group, the maintained assumption was that each income group had homothetic preferences. This can be directly relaxed by implementing the algorithm of Jaravel and Lashkari (2023), which delivers a non-parametric non-homotheticity correction. Since luxuries have lower inflation rates during the study period, as people get richer their preferences shift toward goods who relative prices are falling. Using 2002 price as base, the non-homotheticity correction implies that real income growth is higher than with the conventional homothetic index. If this correction is similar for all income groups,

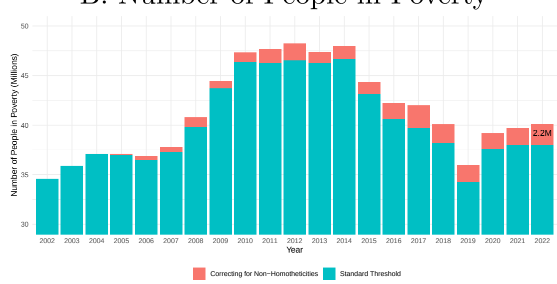
[20]While Molloy (2024) calculates changes in rents in the American Housing Survey, I use official BLS data, including imputed rents.

Figure 13 Results with Nonhomothetic Price Indices

A. Household Real Income Growth, 2002 to 2019



B. Number of People in Poverty



Notes: This figure accounts for non-homotheticities in the computation of inflation, using the algorithm of Jaravel and Lashkari (2023). Panel A reports real income growth across the income distribution under the official CPI (blue), with D-CPIs without the non-homotheticity correction (red), and with D-CPIs including the non-homotheticity correction (green). In panel B, we use the price index for households in poverty to index the poverty line over time, including the non-homotheticity correction, and report the additional number of people who are under the poverty line (in red), compared to the official metric using CPI (in light green).

then the baseline estimates of inflation inequality may not change substantially.

Figure 13 reports the results. Panel A focuses on real income growth across the income distribution, from 2002 to 2019. I find that the non-homotheticity correction is relatively similar across income groups, implying that the increase in real income inequality remain similar to the results documented in Section 3.1. Indeed, with the non-homotheticity correction the income gap between the bottom and top income quintiles increases by 23.2% (= 1.311/1.064), close to the rate of 22.6% obtained with the D-CPIs without the non-homotheticity correction. Panel B repeats the analysis of the poverty line, now including the non-homotheticity correction. The poverty line is indexed at a slightly lower rate than without the non-homotheticity correction, but the effect is modest relative to baseline: we find that 2.2 million people should be eligible for poverty alleviation programs today, rather than 2.3 million without the non-homotheticity correction.

Overall, these analyses show that it is straightforward to account for non-homotheticities, with limited effects on the measurement of inequality and poverty.

4.3 Going Further Back in Time

I now extend the analysis going back to 1983. The main analysis stops in 2002 because of additional challenges in building crosswalks between the consumption and price datasets before that date. However, it is simple to keep expenditure shares fixed in 2002 and build the price index in prior years using these shares, going back to 1983.[21]

As a validation test of this approach, I first compare my reconstructed CPI, with fixed shares prior to 2002, to the official CPI over time. Panel A of Appendix Figure A5 shows that the two price series are closely aligned, despite the fixed shares. In addition, I compute inflation inequality from 2002 to 2023 with fixed shares in 2023, comparing the results to the full series using updated shares. Panel B of Appendix Figure A5 shows that the level of inflation inequality measured with fixed end-of-period shares is similar to the baseline analysis with updated shares in Figure 2. These results suggest that using fixed shares in 2002 may offer a good approximation to measure the patterns of inflation inequality prior to that date.

Figure 14 present the results, documenting inflation inequality from 1983 onward. In Panel A, the price series after 2002 are identical to the results shown in Figure 2. The figure shows that the trend of inflation inequality started before 2002. Panel B show the cumulative inflation patterns by income percentile from 1983 to 2023, while Panel C focuses on the period from 1983 to 2001. While the relationship between household income percentiles and inflation was close to linear after 2002, as shown on Panel B of Figure 2, Panel C of Figure 14 shows that prior to 2002 the relationship was non-linear: inflation inequality before 2002 primarily affected households below the median of the income distribution.

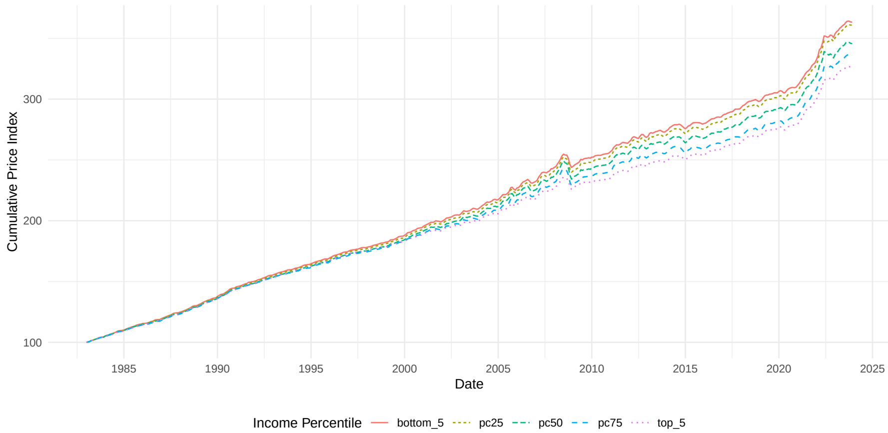
These results have several implications for the measurement of inequality. First, Figure 15 plots household real income from 1983 to 2019. Real income for the bottom quintile increased by 12.5% according to the official CPI, but by only 2.1% with D-CPI. With D-CPIs, real income inequality between the top and bottom income quintiles increased 30 % faster than with the nominal CPI.[22] Second, Table

[21] The methodology of the BLS changed substantially before that date, in particular regarding the treatment of housing, therefore I do not extend the analysis before that date.

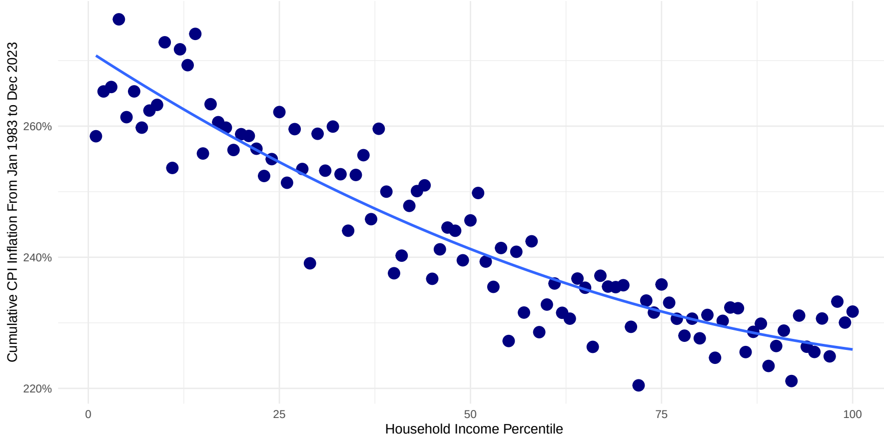
[22] Appendix Figure 15 reports the results from 1983 to 2002.

Figure 14 Inflation Inequality by Income Percentile from 1983 to 2023

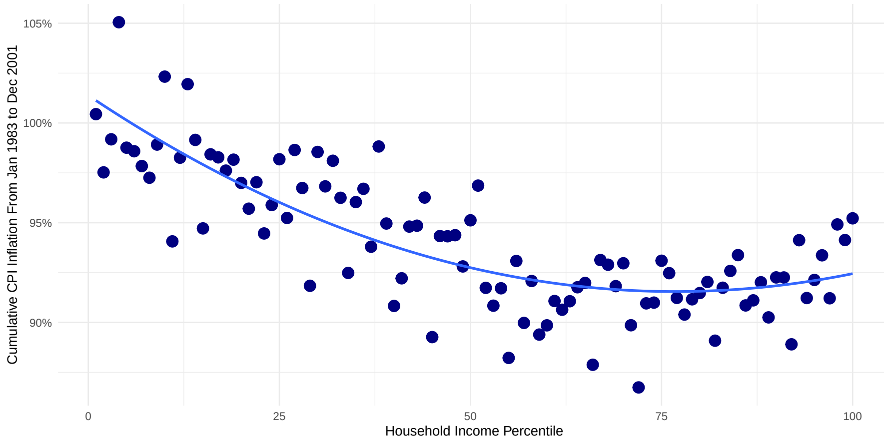
A. Cumulative Index from 1983 to 2023 for Selected Income Percentiles



B. Cumulative Index from 1983 to 2023 across the Income Distribution

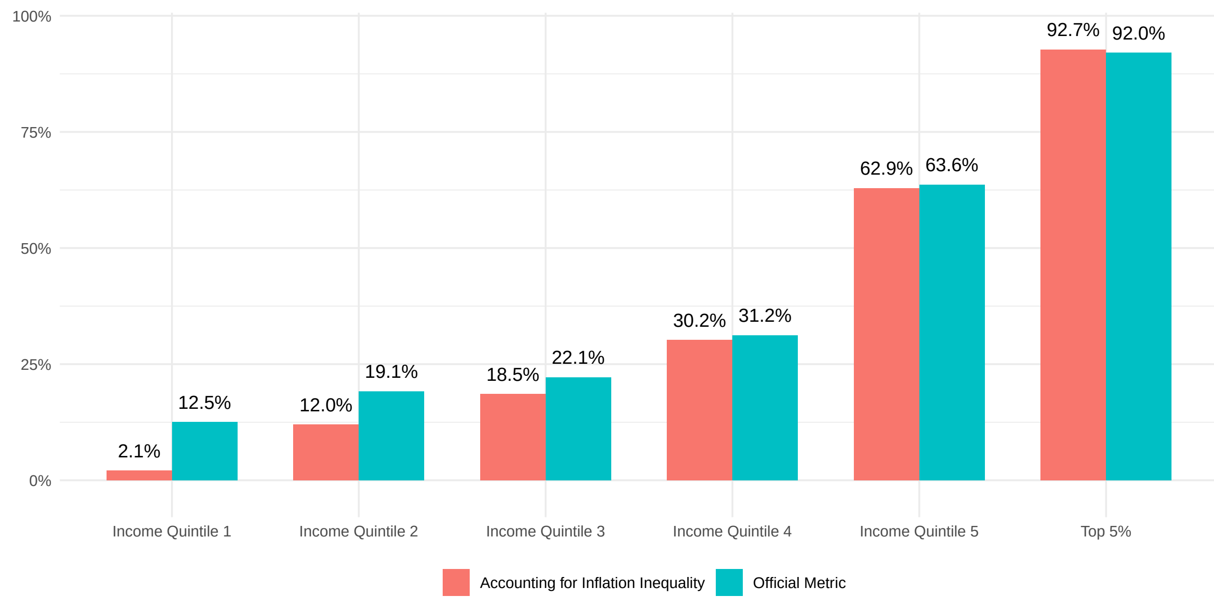


C. Cumulative Index from 1983 to 2001 across the Income Distribution



Notes: This figure reports inflation rates by income percentile. Panel A show the monthly time series of the cumulative price index from January 2002 to December 2023 for selected income percentiles (bottom 5%, 25th, 50th, 75th, and top 5%). Panel B reports the cumulative CPI in December 2023 for all income percentiles, along with the OLS best-fit line. Panel C repeats the analysis from 1983 to 2001.

Figure 15 Implications for Household Real Income Growth, 1983 to 2019



Notes: This figure reports cumulative real income growth from 1983 to 2019 by quintiles of the household income distribution, as well as for the top 5%. Two series are shown, with the official CPI and with our price indices specific to each income group.

Table VI Trends in Consumption Inequality: Ratio of Top to Bottom Income Quintiles, 1984 to 2019

	Consumption ratios, top to bottom quintiles		% Change in consumption inequality
	1984 (1)	2019 (2)	1984 to 2019 (3)
With official CPI	3.84	4.24	+ 10.44 %
Accounting for inflation inequality with D-CPIs	3.84	4.63	+ 20.71 %

Notes: Columns (1) and (2) of this table report the ratios of consumption expenditures of households in the top and bottom income quintiles, expressed in 1984 dollars. Consumption expenditures are obtained from the CEX annual summary tables. The first row uses the official CPI to deflate consumption expenditures in 2019, while the second row uses quintile-specific CPIs. Column (3) reports the percentage changes in the consumption ratios from 1984 to 2019.

Table VII Trends in Pre-tax and After-tax National Income: Ratio of Top to Bottom Income Quintiles, 1983 to 2019

	Pre-tax income ratios			Post-tax income ratios		
	1983 (1)	2019 (2)	Δ 1983–2019 (3)	1983 (4)	2019 (5)	Δ 1983–2019 (6)
With common price index	14.02	17.70	+ 26.32 %	5.21	5.10	+ 16 %
With D-CPIs	14.02	19.35	+ 38.07 %	5.21	5.41	+ 27 %

Notes: Columns (1) and (2) of this table report the ratios of pre-tax national income for households in the top and bottom quintiles, as defined by Auten and Splinter (2024). The first row is obtained from Auten and Splinter (2024), while the second row uses quintile-specific CPIs to correct the ratios. Column (3) reports the percentages change in the ratios from 1983 to 2019. Columns (4) to (6) repeat the analysis for post-tax national income ratios.

VI reports the implications for consumption inequality. While consumption inequality between the top and bottom income quintiles increased by 10 % with the official CPI, it more than doubles, at 21 %, with D-CPIs. Third and finally, Table VII reports the results for pre-tax and post-tax national income ratios, applying the D-CPIs to the data of Auten and Splinter (2024). Here as well, the adjustments relative to the standard metric are substantial. Together, these results illustrate how important inflation inequality can be for the measurement of inequality at long horizons.

5 Conclusion

This paper has shown how to build a public database to measure inflation rates in real time (monthly) across socio-demographic groups in the United States.

Distributional CPIs (D-CPIs) have important implication for the measurement of long–run trends in inequality and poverty. While the income gap between the top and bottom income quintiles increased by 15.6% between 2002 and 2019 according to the official CPI, the income gap increased by 22.6% with D-CPIs. The amplification of inequality is even stronger with Chained D-CPIs and also affects the measurement of consumption inequality and trends in pre- and post-tax national income inequality. Moreover, 2.3 million people are below the "real" poverty line using D-CPIs but above the poverty threshold using the official CPI. These people should become eligible for poverty alleviation programs tied to the poverty line, e.g. Medicaid. Finally, during the inflation burst following the Covid-19 pandemic, inflation was higher for the middle class and the compression of "real" wages was 25% faster with D-CPIs than with the official CPI. The results are similar when D-CPIs are adjusted with a non-parametric non-homotheticity correction, when allowing for inflation heterogeneity across space, or when studying a longer period going back to 1983.

Given that D-CPIs are available in real time (each month) and follow data construction steps that are identical to the official CPI, they can be readily adopted by statistical agencies and researchers for the production of statistics on inequality and poverty, for example in the context of distributional national accounts.

References

Aguiar, Mark and Mark Bils, "Has consumption inequality mirrored income inequality?," *American Economic Review*, 2015, *105* (9), 2725–2756. 3.1.1

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky, "Regressive sin taxes, with an application to the optimal soda tax," *The Quarterly Journal of Economics*, 2019, *134* (3), 1557–1626. 17

Argente, David and Munseob Lee, "Cost of living inequality during the great recession," *Journal of the European Economic Association*, 2021, *19* (2), 913–952. 1

Auten, Gerald and David Splinter, "Income inequality in the United States: Using tax data to measure long-term trends," *Journal of Political Economy*, 2024, *132* (7), 000–000. 3.1.1, III, VII, 4.3

Autor, David, Arindrajit Dube, and Annie McGrew, "The Unexpected Compression: Competition at Work in the Low Wage Labor Market," Working Paper 31010, National Bureau of Economic Research March 2023. 1, 3.2

Blanchet, Thomas, Emmanuel Saez, and Gabriel Zucman, "Real-Time Inequality," Working Paper 30229, National Bureau of Economic Research July 2022. 1

Cage, Robert A, Thesia Isedora Garner, and Javier Ruiz Castillo, "Constructing household specific consumer price indexes: An analysis of different techniques and methods," *BLS Working Paper*, 2002. 6

Garner, Thesia I, David S Johnson, and Mary F Kokoski, "An experimental consumer price index for the poor," *Monthly Lab. Rev.*, 1996, *119*, 32. 6

Jaravel, Xavier, "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector," *The Quarterly Journal of Economics*, 12 2019, *134* (2), 715–783. 1

_ **and Alan Olivi**, "Prices, non-homotheticities, and optimal taxation," *Working Paper*, 2024. 1

_ **and Danial Lashkari**, "Measuring Growth in Consumer Welfare with Income-Dependent Preferences: Nonparametric Methods and Estimates for the United States," *The Quarterly Journal of Economics*, 09 2023, *139* (1), 477–532. 1, 4.2, 13

_ **, Jonathan Kolstad, Kurt Lavetti, and Michael Nguyen-Mason**, "Health Inflation Inequality: Evidence from Utah," *Working Paper*, 2024. 1

Kaplan, Greg and Sam Schulhofer-Wohl, "Inflation at the household level," *Journal of Monetary Economics*, 2017, *91*, 19–38. The Swiss National Bank/Study Center Gerzensee Special Issue: âModern Macroeconomics: Study Center Gerzensee Conference in Honor of Robert G. King" Sponsored by the Swiss National Bank and the Study Center Gerzensee. 1

Klick, Josh and Anya Stockburger, "Experimental CPI for lower and higher income households," *Working Paper*, 2021. 1, 5

— **and** — , “Examining U.S. inflation across households grouped by equivalized income,” *Monthly Labor Review*, 2024. 1, 5

Krueger, Dirk and Fabrizio Perri, “Does income inequality lead to consumption inequality? Evidence and theory,” *The Review of Economic Studies*, 2006, *73* (1), 163–193. 3.1.1

Martin, Robert S, “Democratic Aggregation: Issues and Implications for Consumer Price Indexes,” *Review of Income and Wealth*, 2022. 6

Molloy, Raven, “JUE insight: Differences in rent growth by income from 1985 to 2021 and implications for inflation,” *Journal of Urban Economics*, 2024, *142*, 103669. 1, 4.1, 20

Moretti, Enrico, “Real wage inequality,” *American Economic Journal: Applied Economics*, 2013, *5* (1), 65–103. 2, 4.1

Olivi, Alan, Vincent Sterk, and Dajana Xhani, “Optimal Monetary Policy during a Cost of Living Crisis,” *Working Paper*, 2024. 1

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman, “Distributional National Accounts: Methods and Estimates for the United States,” *The Quarterly Journal of Economics*, 10 2017, *133* (2), 553–609. 1

For Online Publication

Appendix to "Distributional Consumer Price Indices"

Xavier Jaravel, *LSE & CEPR*
July 2024

A Data Appendix

This appendix presents background information on the CPI and CEX databases.

Consumer Price Index. The Consumer Price Index is a set of official price indexes that capture price changes experienced by urban consumer in the US. BLS employs a multi-stage sampling design for the pricing surveys to select rotating samples of geographic areas, retail outlets, specific goods and services, and residential housing units. Each month, the surveys collect approximately 94,000 prices for commodities and services, and 8,000 rental housing unit quotes to compute rental price and owners' equivalent rent of residences. The CPI target population is all urban consumers, which covers 93% of the U.S. population.

BLS defines a specific scope of goods and services for CPI calculation that differs from other published consumption statistics such as the annual expenditure summary tables based on the Consumer Expenditure Survey. At the most granular level of item classification, BLS defines 273 mutually exclusive and exhaustive entry-level items (hence denoted as "ELI") for which price information is sampled, plus 26 unsampled ELIs.[1]

After the collection of initial price data, BLS constructs basic price indexes for each unique combination of 32 basic areas and 243 basic items (hence denoted as "basic-price-index items"), which follow a one-to-many mapping to the ELIs. These basic indexes serve as the building blocks for any published CPI series, but aren't available to the public. The most granular, complete, and mutually exclusive breakdown of CPI items for which price index data is publicly available at the national level consists of 211 "item strata".

BLS publishes multiple versions of the price series for each item stratum and each aggregate item group. Any series can be uniquely identified by its item code, geographic location, targeted population, seasonality adjustment, and base period. Not seasonally adjusted data are typically used for official purposes including monthly update of relative importance weights, and therefore chosen for our calculations.

BLS adheres to a regular publication schedule for the price series, which tends to be around the end of the second week of every month.[2] Other than the published price series, one can also find monthly summary information and relative importance weights in CPI News Release, which is made available concurrently with the newest CPI series. BLS also publishes many useful appendices to accompany the

[1] Note that the total number of ELIs may change if BLS decides to update the item classification convention in future years. For a full list of sampled entry-level items and the content of each item under the current definition, see Appendix 2 of the CPI Handbook of Methods at https://www.bls.gov/cpi/additional-resources/entry-level-item-descriptions.htm. Since the appendix only includes sampled ELIs, I obtain the count of unsampled ELIs by counting unsampled item strata, assuming a one-to-one mapping between the two classifications.

[2] The schedule of release can be found at: https://www.bls.gov/schedule/news_release/cpi.htm.

index data, one of which is the concordance table between UCC and ELI that allows users to identify the CPI-relevant UCCs and their associated expenditures in the CEX micro-data, and link the expenditure data to price series in a manner that is consistent with BLS' practice.

Consumer Expenditure Survey. While BLS sources the raw price information from the Commodities & Services survey and Housing survey, it uses data from the Consumer Expenditure survey to compute the weights that are used in index construction and aggregation. The expenditure shares obtained from the CEX are called "relative importance weights" by the BLS. These expenditure shares are published for pivot months (see discussion in the main text). I also use the CEX expenditure micro-data to obtain the spending patterns of consumers in each socio-demographic group.

The CEX conducts two different surveys to analyze consumption patterns. The interview survey asks respondents to report spending over large consumption categories over the previous three months. The diary survey in contrast asks respondents to keep a detailed log of all purchases made over a week. Consumption is aggregated to a set of Universal Categorization Codes (UCC). Each survey also tracks a set of demographic and household information about the respondents. However, each respondent is only in one survey. When calculating income percentiles, I use the percentile across a given survey.

The BLS publishes a series of yearly expenditure tables that contain total expenditures by income quintile at various levels of aggregation. I use these tables to validate that I process the CEX micro-data correctly. However, the set of UCCs that are part of these CEX tables are not the same as the ones relevant for the CPI. The product scopes differ primarily for the category *owned dwelling*: for CPI this is captured in *owners' equivalent rent of residences* (OER), which is defined as the implicit rent that owner occupants would have to pay if they were renting their homes unfurnished; for the CEX expenditure summary this includes mortgage interests, property taxes and insurance, and expenses for repairs and maintenance.

B Price Index Calculations

B.1 Additional Information on the Calculation of Aggregate CPI

In Section A, the discussion of the calculation of item-level expenditure shares by the BLS omitted a step for simplicity, which I describe here.

To illustrate the logic of this additional step, consider the computation of expenditure shares in December 2017. As discussed in Section A, BLS computes a new set of expenditure shares in December 2017 using CEX data from 2015 and 2016. In fact, the BLS also makes an adjustment for price changes between the years 2015-2016 and December 2017, inferring how expenditure shares should have changed given relative price changes between 2015-2016 and December 2017.

For consistency with the notation in the main text, let us denote the reference December period by $0(t)$, e.g. December 2017. $s_{ib(0(t))}$ denotes the expenditure share of item i in the base period $b(0(t))$ – e.g., with $0(t) =$ December 2017, $b(0(t))$ is 2015-2016. $s_{ib(0(t))}$ is computed directly in CEX data. Then the expenditure share assigned at $0(t)$ for category i is:

$$\omega_{i0(t)} \equiv \frac{\frac{p_{i0(t)}}{p_{ib(0(t))}} \cdot s_{ib(0(t))}}{\sum_k \left(\frac{p_{k0(t)}}{p_{kb(0(t))}} \cdot s_{kb(0(t))}\right)},$$

where $p_{ib(0(t))}$ denotes the average price index of the item over period $b(0(t))$, while $p_{i0(t)}$ is the price index in the focal December month.

B.2 Price Indices with Geography Heterogeneity

The BLS only publishes local price indices for a subset of items and a subset of locations. The consumption within these local areas represents roughly 40 % of total consumption and within each local area, consumption on items with a published prices series represents around 40-50 % of total consumption within that area.

Let S be the set of all items and $I \subset S$ be the set of all items with a published local price series. Let C be the set of cities with published local prices. $p^c_{i,t}$ is the price of good i, in period t, in city c, P^c_t is the overall price index for city c, and $r_{i,t}$ is the published Relative Importance weight for item i in period t.

The first goal is to calculate the city-specific importance weights. As in the main text, $0(t)$ denotes the reference period and $b(t)$ is the base period for month t. For instance, for January 2018 to December 2019, $0(t)$ is December 2017 and $b(t)$ is January 2015 to December 2016. The BLS publishes the set of importance weights only for the reference period. These weights represent the consumption shares from the base period that have been updated to reflect any price changes that have occured. We can calculate the implied importance weights from the base period by inverting the update formula:

$$r_{i,b(t)} = r_{i,0(t)} \frac{p_{i,b(t)}}{p_{i,0(t)}} / \frac{P_{b(t)}}{P_{0(t)}}$$

Using the microdata, one can caculate the fraction of spending on each item i that is from area c, which is denoted by $S^c_{i,t}$. Using the weight update formula applied to the city-specific prices yields:

$$r^c_{i,b(t)} = r_{i,b(t)} S^c_{i,b(t)}$$

$$r^c_{i,0(t)} = r^c_{i,b(t)} \frac{p^c_{i,0(t)}}{p^c_{i,b(t)}} / \frac{P_{0(t)}}{P_{b(t)}} \quad = r_{i,0(t)} \frac{p^c_{i,0(t)}}{p^c_{i,b(t)}} / \frac{p_{i,0(t)}}{p_{i,b(t)}} S^c_{i,b(t)}$$

Local price indices available. When only a subset of the local goods are priced, one can calculate an implied local price index for these goods which is consitent with the overall price index for that area. Let $S^c_t = \sum_{i \in S} r^c_{i,t}$ be the share of total consumption going to area c. To measure all shares in the base period $b(t)$, one has to calculate $S^c_{0(t)}$using the share update formula:

$$S^c_{0(t)} = S^c_{b(t)} \frac{P^c_{0(t)}/P^c_{b(t)}}{P_{0(t)}/P_{b(t)}}$$

One can then calculate changes in the local price index as

$$\begin{aligned}\frac{P^c_t}{P^c_{0(t)}} &= \sum_{i\in I} \frac{r^c_{i,0(t)}}{S^c_{0(t)}} \frac{p^c_{i,t}}{p^c_{i,0(t)}} + \sum_{i\in I^c} \frac{r^c_{i,0(t)}}{S^c_{0(t)}} \frac{p^c_{i,t}}{p^c_{i,0(t)}} \\ &\equiv w^c_{I,0(t)} \frac{P^c_{I,t}}{P^c_{I,0(t)}} + (1-w^c_{I,0(t)}) \frac{P^c_{I^c,t}}{P^c_{I^c,0(t)}}\end{aligned}$$

where $P^c_{I,t}$ is the local price index for area c restricted to the set of goods I, and $w^c_{I,t} = \sum_{i\in I} \frac{r^c_{i,t}}{S^c_t}$ is the total share of consumption in area c that goes to the goods in set I. From this, one can calculate the change in the price index for the set of unpriced goods I^c:

$$\frac{P^c_{I^c,t}}{P^c_{I^c,0(t)}} = \left(\frac{P^c_t}{P^c_{0(t)}} - w^c_{I,0(t)} \frac{P^c_{I,t}}{P^c_{I,0(t)}}\right)/(1-w^c_{I,0(t)}).$$

For $i \in I^c$ the $r^c_{i,0(t)}$ are unknown. However, one can make the simplifying assumption that

$$\frac{p^c_{i,0(t)}}{p^c_{i,b(t)}} / \frac{p_{i,0(t)}}{p_{i,b(t)}} = \alpha^c_{0(t)},$$

i.e. the ratio of the change in an item's local price to the change in the item's national price is constant across items. One can then solve for $\alpha^c_{0(t)}$:

$$\begin{aligned}(1-w^c_{I,0(t)}) &= \sum_{i\in I^c} \frac{r^c_{i,0(t)}}{S^c_{0(t)}} \\ &= \sum_{i\in I^c} \frac{r_{i,0(t)} \frac{p^c_{i,0(t)}}{p^c_{i,b(t)}} / \frac{p_{i,0(t)}}{p_{i,b(t)}} S^c_{i,b(t)}}{S^c_{0(t)}} \\ &= \sum_{i\in I^c} \frac{r_{i,0(t)} a^c_{0(t)} S^c_{i,b(t)}}{S^c_{0(t)}} \\ a^c_{0(t)} &= \frac{(1-w^c_{I,0(t)}) S^c_{0(t)}}{\sum_{i\in I^c} r_{i,0(t)} S^c_{i,b(t)}}.\end{aligned}$$

Plugging back into the formula for the relative importance weights yields:

$$\begin{aligned}r^c_{i,0(t)} &= r_{i,0(t)} \frac{p^c_{i,0(t)}}{p^c_{i,b(t)}} / \frac{p_{i,0(t)}}{p_{i,b(t)}} S^c_{i,b(t)} \\ &= (1-w^c_{I,0(t)}) S^c_{0(t)} \frac{r_{i,0(t)} S^c_{i,b(t)}}{\sum_{i\in I^c} r_{i,0(t)} S^c_{i,b(t)}}.\end{aligned}$$

One can then use these implied relative importance weights to calculate local item prices that are consistent with the overall local price index of unpriced goods calculated above. One needs a similar simplifying assumption to the one above:

$$\frac{p^c_{i,t}}{p^c_{i,0(t)}} / \frac{p_{i,t}}{p_{i,0(t)}} = \beta^c_t.$$

Plugging into the defintion of the local price index yields:

$$\begin{aligned}
\frac{P^c_{I^c,t}}{P^c_{I^c,0(t)}} &= \sum_{i\in I^c} \frac{r^c_{i,0(t)}}{S^c_{0(t)}} \frac{p^c_{i,t}}{p^c_{i,0(t)}} / (1 - w^c_{I,0(t)}) \\
&= \sum_{i\in I^c} \frac{r^c_{i,0(t)}}{S^c_{0(t)}} \frac{p_{i,t}}{p_{i,0(t)}} \beta_t / (1 - w^c_{I,0(t)}). \\
\beta_t &= \frac{\frac{P^c_{I^c,t}}{P^c_{I^c,0(t)}} (1 - w^c_{I,0(t)}) S^c_{0(t)}}{\sum_{i\in I^c} r^c_{i,0(t)} \frac{p_{i,t}}{p_{i,0(t)}}}.
\end{aligned}$$

Implied price index for location without local price series.

Around 60 % of spending occurs in areas with no local published prices. Denote by u the set of all areas that do not have local prices. One can calculate local prices within u so that the implied aggregate index for the whole US is consistent with the published index. Doing so follows almost the same steps as above. Let E be the set of all items. Then

$$\begin{aligned}
\frac{P_t}{P_{0(t)}} &= \textstyle\sum_{c\in C} \sum_{i\in E} r^c_{i,0(t)} \frac{p^c_{i,t}}{p^c_{i,0(t)}} + \sum_{i\in E} r^u_{i,0(t)} \frac{p^u_{i,t}}{p^u_{i,0(t)}} \\
&\equiv w_{0(t)} \tfrac{P^C_t}{P^C_{0(t)}} + (1 - w_{0(t)}) \tfrac{P^u_t}{P^c_{0(t)}},
\end{aligned}$$

where $\frac{P^C_t}{P^C_{0(t)}}$ is the change in the price index for all goods across all areas where local prices are availible, and $w_{0(t)} = \sum_{c\in C} \sum_{i\in E} r^c_{i,0(t)}$ is the share of total consumption that occurs in these areas. One can calculate the local price index as

$$\frac{P^C_t}{P^C_{0(t)}} = \sum_{c\in C} \sum_{i\in E} \frac{r^c_{i,0(t)}}{w_{0(t)}} \frac{p^c_{i,t}}{p^c_{i,0(t)}}.$$

From this forumla, one can calculate the implied price index in areas with no local prices:

$$\frac{P^u_t}{P^u_{0(t)}} = (\frac{P_t}{P_{0(t)}} - w_{0(t)} \frac{P^C_t}{P^C_{0(t)}}) / (1 - w_{0(t)}).$$

The $r^u_{i,0(t)}$ are unknown. However, one can make the simplifying assumption that

$$\frac{p^u_{i,0(t)}}{p^u_{i,b(t)}} / \frac{p_{i,0(t)}}{p_{i,b(t)}} = \gamma^u_{0(t)},$$

so that the ratio of the change in an item's local price to the change in the item's national price is constant across items. One can then solve for $\gamma^u_{0(t)}$

$$\begin{aligned}(1-w_{0(t)}) &= \sum_{i\in E} r^u_{i,0(t)} \\ &= \sum_{i\in E} r_{i,0(t)} \frac{p^u_{i,0(t)}}{p^u_{i,b(t)}} / \frac{p_{i,0(t)}}{p_{i,b(t)}} S^u_{i,b(t)} \\ &= \sum_{i\in E} r_{i,0(t)} \gamma^u_{0(t)} S^c_{i,b(t)} \\ \gamma^u_{0(t)} &= \frac{(1-w_{0(t)})}{\sum_{i\in E} r_{i,0(t)} S^u_{i,b(t)}}.\end{aligned}$$

Plugging back into the formula for the relative importance weights yields:

$$\begin{aligned}r^u_{i,0(t)} &= r_{i,0(t)} \frac{p^u_{i,0(t)}}{p^u_{i,b(t)}} / \frac{p_{i,0(t)}}{p_{i,b(t)}} S^u_{i,b(t)} \\ &= (1-w_{0(t)}) \frac{r_{i,0(t)} S^u_{i,b(t)}}{\sum_{i\in E} r_{i,0(t)} S^u_{i,b(t)}}.\end{aligned}$$

One can then use these implied relative importance weights to calculate local item prices that are consistent with the overall local price index of unpriced goods we calculated above. One again needs a similar simplifying assumption to the one above:

$$\frac{p^u_{i,t}}{p^u_{i,0(t)}} / \frac{p_{i,t}}{p_{i,0(t)}} = \delta^u_t.$$

Plugging into the defintion of the local price index yields:

$$\begin{aligned}\frac{P^u_t}{P^u_{0(t)}} &= \sum_{i\in E} r^u_{i,0(t)} \frac{p^u_{i,t}}{p^u_{i,0(t)}} / (1-w_{0(t)}) \\ &= \sum_{i\in E} r^u_{i,0(t)} \delta^u_t \frac{p_{i,t}}{p_{i,0(t)}} / (1-w_{0(t)}) \\ \delta^u_t &= \frac{\frac{P^u_t}{P^u_{0(t)}} (1-w_{0(t)})}{\sum_{i\in E} r^u_{i,0(t)} \frac{p_{i,t}}{p_{i,0(t)}}}\end{aligned}$$

C Price Index Decomposition

Consider a set of item categories $E = \cup I_1, \ldots, I_n$,where $I_i \cap I_j = \emptyset$ when $i \neq j$. We are interested in decomposing the inflation difference experienced between two household groups as the sum of the difference experienced *within* each of the sets I_i and the difference experienced *between* them.

This decomposition is straightforward using the Tornqvist Index, i.e. the Chained CPI, where the price index for group g can be written as $P^g_T/P^g_0 = \prod_{t=0}^T \prod_{i\in E} (p_{i,t}/p_{i,t-1})^{s^g_{i,t}}$, where $s^g_{i,t} = .5(r^g_{i,t} + r^g_{i,t-1})$

is the average expenditure share for group g, on item i between period t and $t-1$.[3] We define $p_{i,-1} = p_{i,0}$ to normalize the price index to 1 at $t = 0$.

Letting $\pi_{i,t} = \log(p_{i,t}/p_{i,t-1})$ and $s^g_{I,t} = \sum_{i \in I} s^g_{i,t}$, one can define the one period inflation experienced by group g at time t within set I as $\pi^g_{I,t} = \sum_{i\in I} \frac{s^g_{i,t}}{\sum_{j\in I} s^g_{j,t}} \pi_{i,t} = \sum_{i\in I} \frac{s^g_{i,t}}{S^g_{I,t}} \pi_{i,t}$. With this notation, one can decompose $\log(P^g_T/P^g_0) - \log(P^q_T/P^g_0)$ for any two groups g and q:

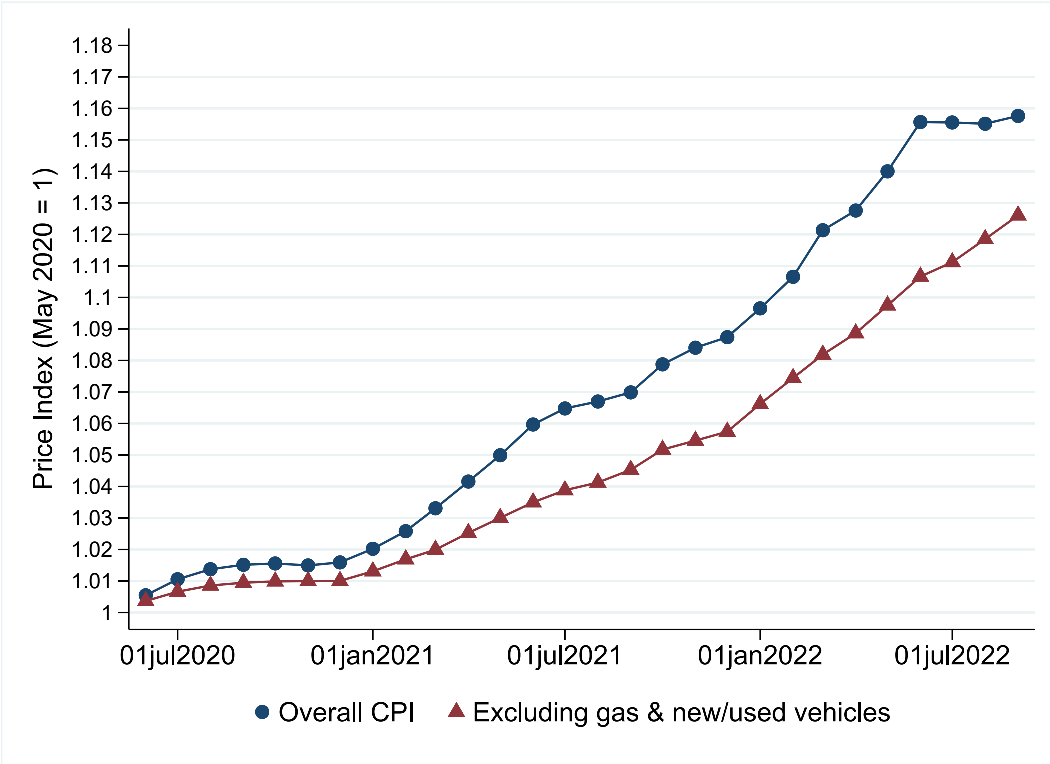
$$
\begin{aligned}
&\log(P^g_T/P^g_0) - \log(P^q_T/P^q_0) \\
&= \sum_{t=0}^{T} \sum_{i\in E} s^g_{i,t}\pi_{i,t} - \sum_{t=0}^{T} \sum_{i\in E} s^q_{i,t}\pi_{i,t} \\
&= \sum_{t=0}^{T} \sum_{I} \sum_{i\in I} \left(s^g_{i,t}\pi_{i,t} - s^q_{i,t}\pi_{i,t} \right) \\
&= \sum_{t=0}^{T} \sum_{I} \left(s^g_{I,t}\pi^g_{I,t} - s^q_{I,t}\pi^q_{I,t} \right) \\
&= \sum_{t=0}^{T} \sum_{I} \left(.5 s^g_{I,t}\pi^g_{I,t} + .5 s^q_{I,t}\pi^g_{I,t} - .5 s^g_{I,t}\pi^q_{I,t} - .5 s^q_{I,t}\pi^q_{I,t} + .5 s^g_{I,t}\pi^g_{I,t} - .5 s^q_{I,t}\pi^g_{I,t} + .5 s^g_{I,t}\pi^q_{I,t} - .5 s^q_{I,t}\pi^q_{I,t} \right) \\
&= \sum_{t=0}^{T} \sum_{I} \left(.5(s^g_{I,t} + s^q_{I,t})(\pi^g_{I,t} - \pi^q_{I,t}) + .5(\pi^q_{I,t} + \pi^g_{I,t})(s^g_{I,t} - s^q_{I,t}\right) \\
&= \underbrace{\sum_{t=0}^{T} \sum_{I} \left(\bar{s}^{g,q}_{i,t} \Delta\pi^{g,q}_{i,t} \right)}_{\text{Within}} + \underbrace{\sum_{t=0}^{T} \sum_{I} \left(\bar{\pi}^{g,q}_{i,t} \Delta s^{g,q}_{i,t} \right)}_{\text{Between}},
\end{aligned}
$$

where $\bar{s}^{g,q}_{I,t} = .5(s^g_{I,t} + s^q_{I,t})$ is the average value of $s_{I,t}$ for groups g and q, and $\Delta s^{g,q}_{i,t} = s^g_{I,t} - s^q_{I,t}$ is the difference.

[3] $r^g_{i,t}$ denotes the expenditure share for each household group on item i in month t, as measured in the CEX data crosswalked to item strata.

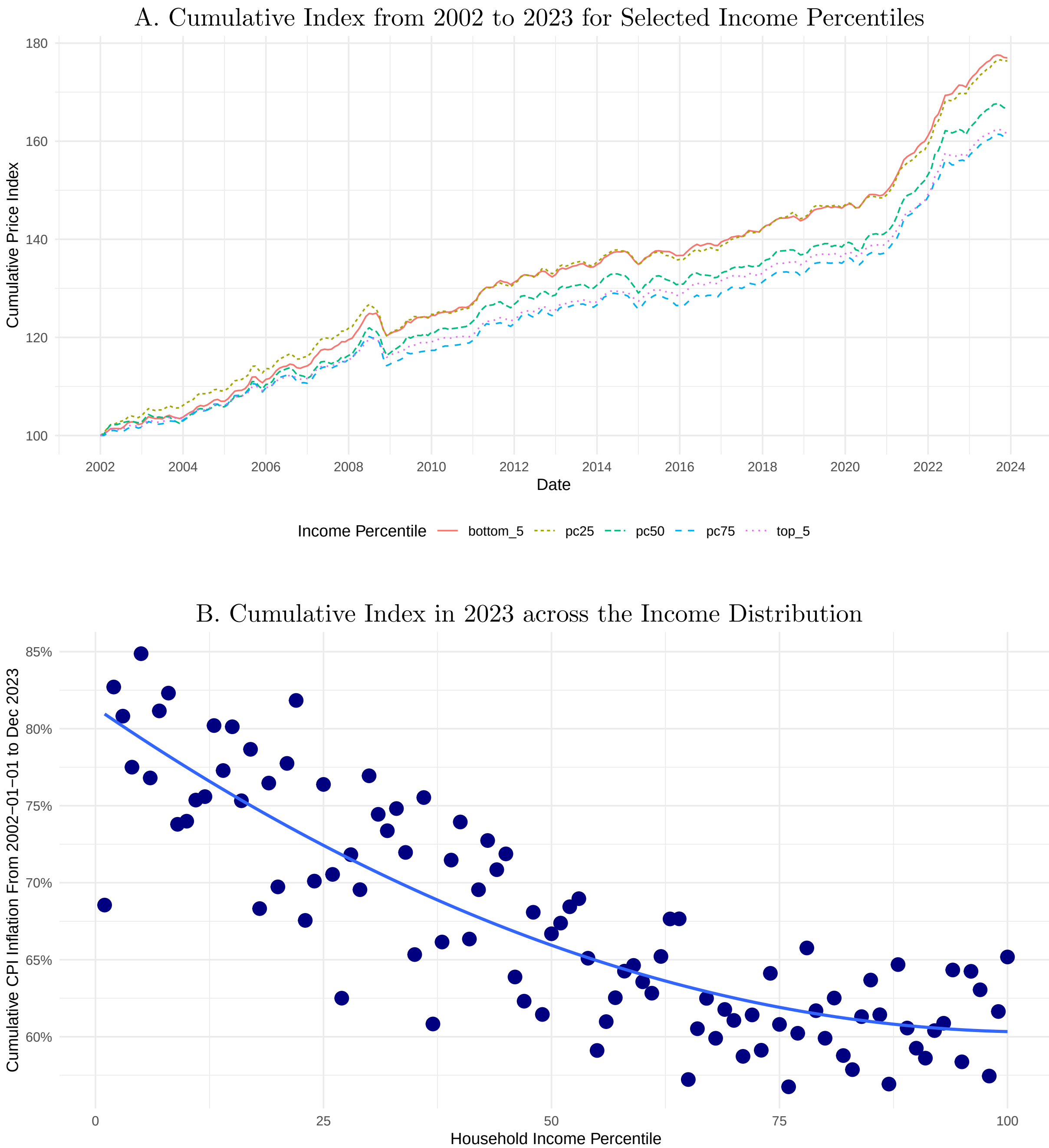
D Additional Figures and Tables

Figure A1 Inflation in the Wake of the Covid-19 Pandemic



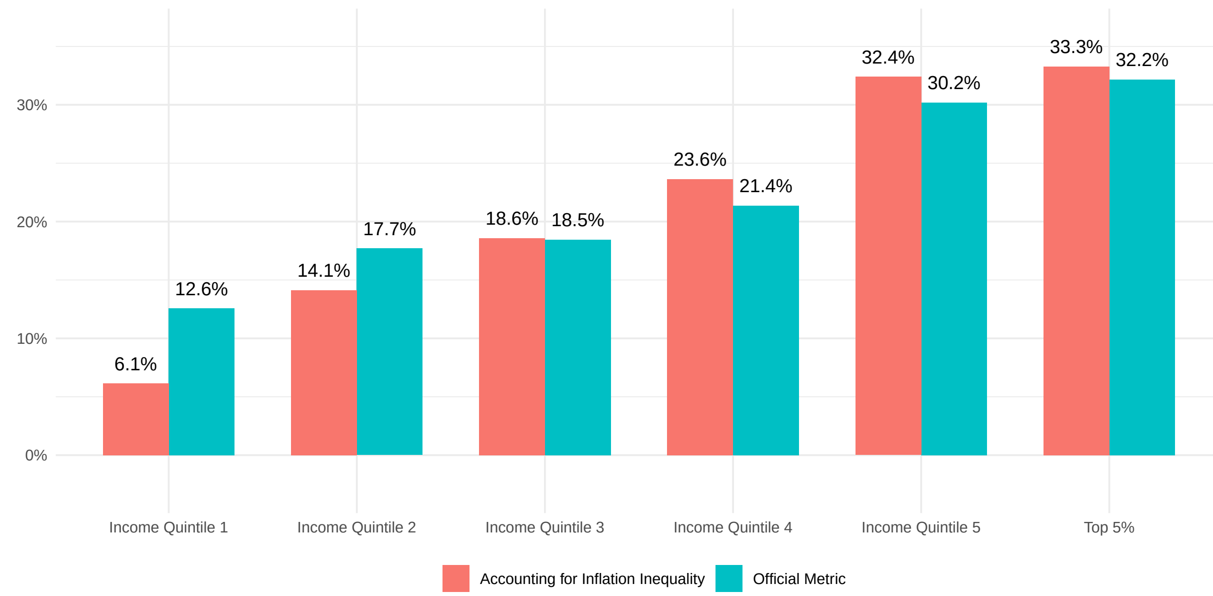
Notes: This figures plots the official CPI, as well as the CPI excluding gas and new/used vehicles, from May 2020 to December 2023.

Figure A2 Long-Run Inflation Inequality by Income Percentile



Notes: This figure reports inflation rates by income percentile. Panel A show the monthly time series of the cumulative price index from January 2002 to December 2023 for selected income percentiles (bottom 5%, 25th, 50th, 75th, and top 5%). Panel B reports the cumulative CPI in December 2023 for all income percentiles, along with the OLS best-fit line.

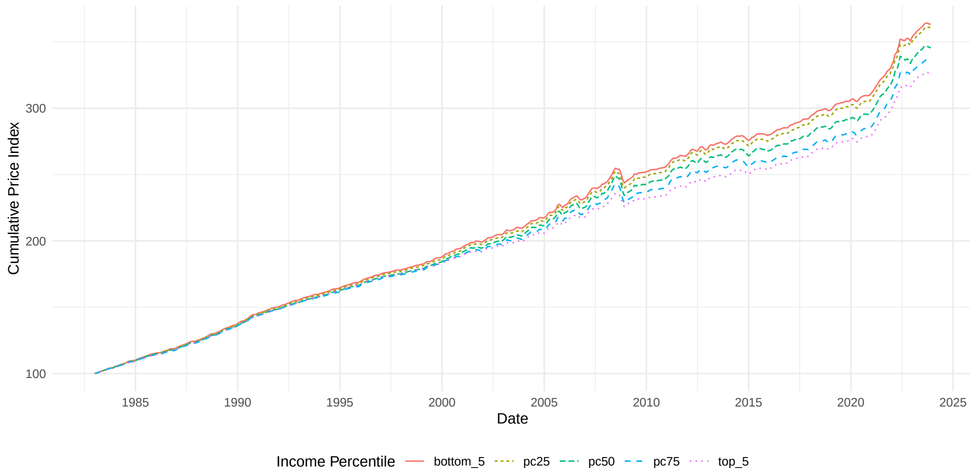
Figure A3 Implications for Household Real Income Growth, Chained CPI, 2002 to 2019



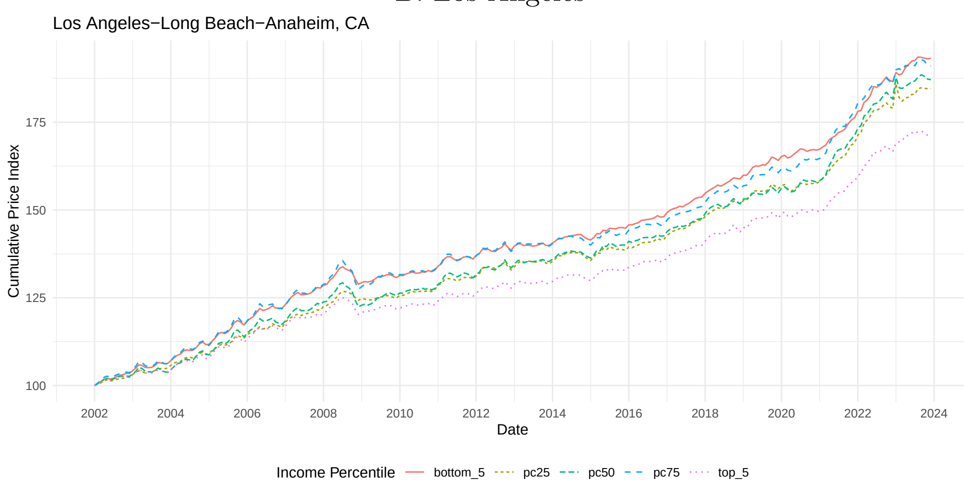
Notes: This figure reports cumulative real income growth from 2002 to 2019 by quintiles of the household income distribution, as well as for the top 5%. Two series are shown, with the official chained CPI and with our price indices specific to each income group.

Figure A4 Inflation Inequality within Selected Cities

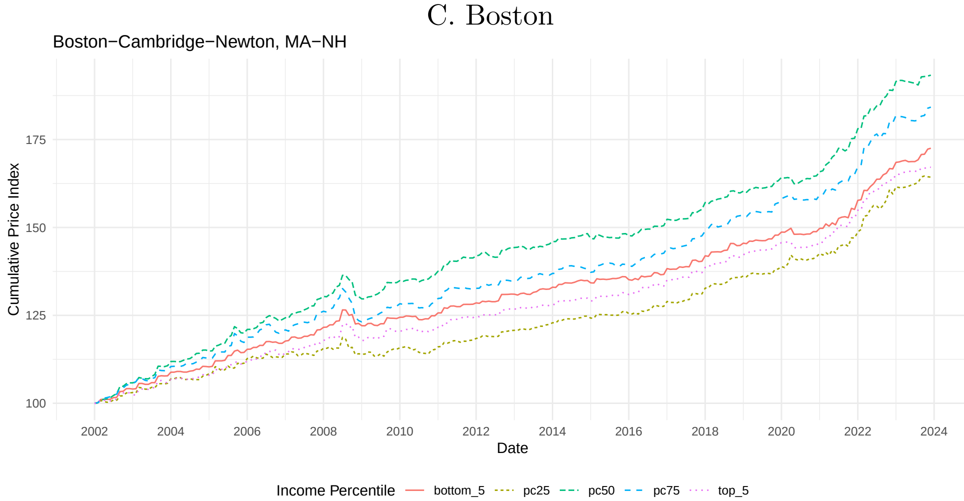
A. Chicago



B. Los Angeles



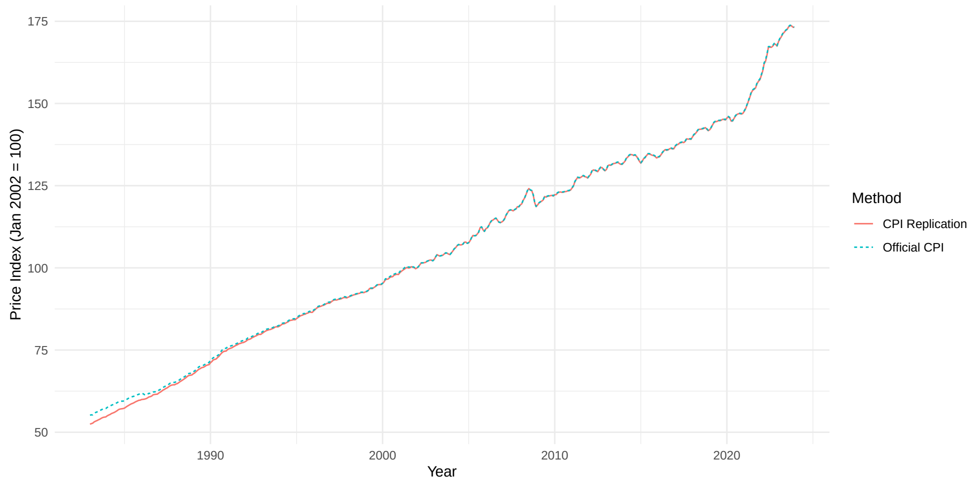
C. Boston



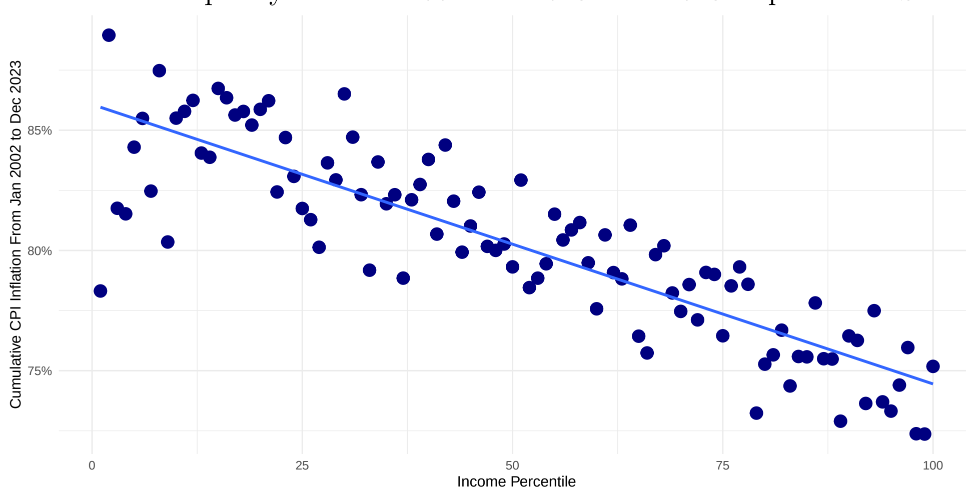
Notes: This figure reports inflation rates by income percentile for three cities.

Figure A5 Validation Tests for Analysis with Fixed Expenditure Shares

A. Comparison to Official CPI, using Fixed Expenditure Shares prior to 2002

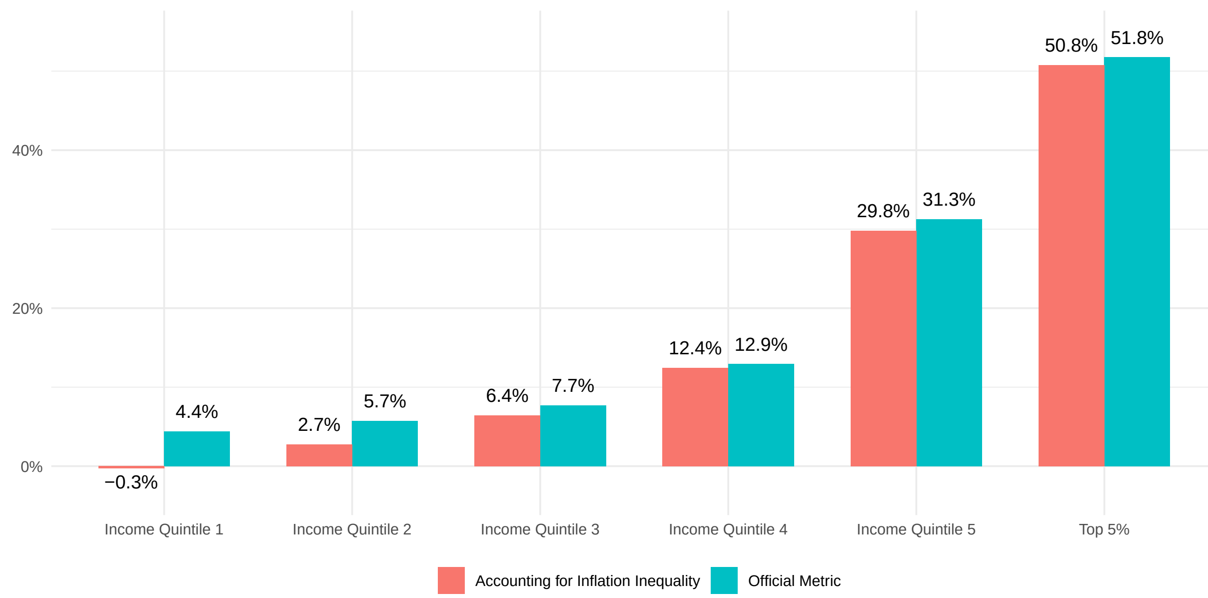


B. Inflation Inequality between 2002 and 2023 with 2023 Expenditure Shares



Notes: This figure reports two tests of the reliability of price indices built with fixed shares. Panel A compares the official CPI to my reconstructed CPI, using fixed 2002 expenditure shares for all years prior to 2002 (and updating the shares thereafter as in Figure 1). The two price series behave very similarly, including prior to 2002. Panel B reports the cumulative inflation rates from 2002 to 2023, using 2023 expenditure shares. The gap between the top and bottom of the income distribution is around 12 percentage points, which is similar to the difference observed in our baseline analysis updating shares every year (see Figure 2, where the inflation difference is about 15 percentage points).

Figure A6 Implications for Household Real Income Growth, 1983 to 2002



Notes: This figure reports cumulative real income growth from 1983 to 2002 by quintiles of the household income distribution, as well as for the top 5%. Two series are shown, with the official CPI and with our price indices specific to each income group.

Table A1 Expenditure Shares by Income Quintile

Item Name	CPI Weight	Bottom 5%	1	2	3	4	5	Top 5%
Owners' equivalent rent of primary residence	22.78	16.19	18.41	19.94	21.42	22.71	24.22	24.97
Rent of primary residence	6.61	16.42	14.62	10.80	8.08	5.24	2.46	1.89
Gasoline (all types)	5.11	6.20	5.83	6.65	7.03	6.78	5.36	4.69
New vehicles	3.15	0.47	0.66	1.88	2.97	3.42	4.27	4.36
Electricity	2.89	3.53	3.70	3.46	3.06	2.69	2.08	2.02
Full service meals and snacks	2.72	1.90	1.90	2.18	2.50	2.90	3.12	3.26
Motor vehicle insurance	2.53	1.57	2.28	3.12	2.77	3.02	2.16	1.92
Limited service meals and snacks	2.30	2.78	2.33	2.20	2.43	2.52	2.09	1.86
Used cars and trucks	1.86	2.08	1.71	1.95	2.07	2.17	1.80	1.82
College tuition and fees	1.77	3.67	2.39	0.92	1.09	1.19	2.97	3.30
Physicians' services	1.62	1.24	1.54	1.85	1.77	1.72	1.42	1.31
Hospital services	1.60	1.44	1.65	1.95	1.89	1.82	1.47	1.34
Unsampled owners' equivalent rent of secondary residences	1.43	0.45	0.83	1.16	1.06	1.21	1.93	2.28
Cable and satellite television service	1.42	1.58	1.73	1.68	1.57	1.43	1.09	0.98
Wireless telephone services	1.40	1.20	1.18	1.37	1.43	1.40	1.06	0.95
Prescription drugs	1.33	1.20	1.64	1.84	1.55	1.34	1.00	0.94
Residential telephone services	0.96	1.17	1.36	1.21	1.04	0.89	0.71	0.68
Water and sewerage maintenance	0.93	1.03	1.10	1.14	1.06	1.03	0.84	0.77
Utility (piped) gas service	0.90	0.77	0.88	0.91	0.85	0.80	0.71	0.70
Airline fares	0.79	0.32	0.39	0.47	0.62	0.77	1.28	1.36
Day care and preschool	0.79	0.24	0.31	0.44	0.48	0.81	1.22	1.22
Dental services	0.78	0.69	0.59	0.84	0.85	0.82	0.74	0.79
Cigarettes	0.75	1.47	1.35	1.22	1.01	0.79	0.35	0.22
Pets and pet products	0.68	0.72	0.66	0.74	0.71	0.72	0.57	0.52
Health insurance	0.66	0.40	0.55	0.72	0.71	0.68	0.56	0.53
Admissions	0.64	0.44	0.32	0.36	0.47	0.59	0.89	0.97
Haircuts and other personal care services	0.63	0.45	0.48	0.53	0.57	0.61	0.71	0.72
Other miscellaneous foods	0.63	0.76	0.86	0.64	0.61	0.60	0.56	0.53
Motor vehicle repair	0.60	0.44	0.50	0.60	0.63	0.64	0.58	0.59
Other lodging away from home including hotels and motels	0.60	0.20	0.22	0.27	0.38	0.51	0.82	0.88
Women's suits and separates	0.59	0.54	0.54	0.50	0.49	0.52	0.62	0.61

Internet services and electronic information providers	0.57	0.50	0.46	0.54	0.61	0.61	0.48	0.41
Club membership for shopping clubs, fraternal, or other organizations, or participant sports fees	0.57	0.19	0.21	0.29	0.39	0.47	0.88	0.96
Motor vehicle maintenance and servicing	0.46	0.37	0.41	0.43	0.46	0.46	0.46	0.45
Pet services including veterinary	0.42	0.15	0.20	0.40	0.31	0.43	0.57	0.49
Women's underwear, nightwear, swimwear, and accessories	0.40	0.38	0.40	0.38	0.39	0.37	0.47	0.53
Nonfrozen noncarbonated juices and drinks	0.40	0.55	0.49	0.43	0.41	0.37	0.33	0.31
Elementary and high school tuition and fees	0.40	0.04	0.12	0.10	0.14	0.30	0.83	0.94
Alcoholic beverages away from home	0.38	0.35	0.24	0.30	0.37	0.40	0.46	0.49
Services by other medical professionals	0.38	0.23	0.29	0.36	0.36	0.42	0.38	0.36
Leased cars and trucks	0.37	0.21	0.16	0.18	0.25	0.32	0.48	0.57
Other food away from home	0.37	0.28	0.26	0.17	0.21	0.30	0.59	0.74
Tenants' and household insurance	0.36	0.35	0.35	0.35	0.35	0.36	0.35	0.36
Outdoor equipment and supplies	0.35	0.15	0.18	0.29	0.23	0.33	0.42	0.49
Household cleaning products	0.35	0.49	0.46	0.38	0.35	0.35	0.28	0.25
Hair, dental, shaving, and miscellaneous personal care products	0.34	0.29	0.33	0.32	0.30	0.31	0.32	0.31
Living room, kitchen, and dining room furniture	0.33	0.14	0.21	0.30	0.29	0.27	0.39	0.49
Women's footwear	0.33	0.28	0.34	0.33	0.32	0.29	0.29	0.28
State motor vehicle registration and license fees	0.32	0.64	0.45	0.40	0.36	0.32	0.25	0.23
Snacks	0.32	0.38	0.37	0.34	0.34	0.34	0.31	0.28
Unsampled recreation services	0.31	0.13	0.15	0.16	0.25	0.30	0.41	0.40
Nonprescription drugs	0.31	0.25	0.32	0.30	0.31	0.29	0.27	0.24
Toys	0.31	0.26	0.24	0.23	0.24	0.26	0.29	0.27
Legal services	0.30	0.14	0.20	0.35	0.25	0.28	0.36	0.45

Garbage and trash collection	0.30	0.29	0.32	0.31	0.32	0.30	0.27	0.25
Cosmetics, perfume, bath, nail preparations and implements	0.30	0.26	0.28	0.28	0.28	0.28	0.30	0.30
Milk	0.29	0.44	0.46	0.38	0.33	0.31	0.25	0.22
Frozen and freeze dried prepared foods	0.29	0.36	0.37	0.34	0.30	0.29	0.21	0.19
Breakfast cereal	0.29	0.39	0.39	0.36	0.28	0.28	0.24	0.21
Cheese and related products	0.28	0.37	0.34	0.29	0.30	0.30	0.27	0.26
Spices, seasonings, condiments, sauces	0.28	0.35	0.32	0.28	0.28	0.28	0.24	0.24
Miscellaneous household products	0.28	0.25	0.25	0.26	0.25	0.28	0.29	0.29
Chicken	0.28	0.37	0.41	0.32	0.30	0.26	0.22	0.21
Carbonated drinks	0.28	0.45	0.41	0.33	0.31	0.25	0.20	0.17
Tires	0.28	0.21	0.21	0.26	0.31	0.30	0.31	0.28
Beer, ale, and other malt beverages at home	0.27	0.42	0.29	0.26	0.30	0.29	0.24	0.19
Intracity transportation	0.27	0.31	0.36	0.29	0.22	0.22	0.30	0.34
Food at employee sites and schools	0.26	0.48	0.23	0.21	0.25	0.31	0.29	0.28
Other meats	0.26	0.33	0.36	0.30	0.29	0.28	0.23	0.22
Domestic services	0.25	0.11	0.17	0.13	0.13	0.13	0.44	0.52
Eyeglasses and eye care	0.25	0.18	0.21	0.25	0.24	0.25	0.24	0.22
Girls' apparel	0.24	0.26	0.22	0.27	0.25	0.25	0.26	0.24
Household paper products	0.24	0.31	0.33	0.27	0.27	0.23	0.20	0.18
Other fresh vegetables	0.24	0.28	0.28	0.26	0.23	0.24	0.23	0.22
Sports vehicles including bicycles	0.24	0.07	0.08	0.08	0.16	0.26	0.38	0.44
Laundry and dry cleaning services	0.24	0.29	0.35	0.26	0.20	0.18	0.25	0.31
Gardening and lawncare services	0.24	0.12	0.21	0.19	0.16	0.17	0.32	0.40
Other bakery products	0.23	0.32	0.28	0.25	0.24	0.23	0.20	0.19
Fees for lessons or instructions	0.23	0.06	0.07	0.06	0.12	0.19	0.41	0.46
Clocks, lamps, and decorator items	0.23	0.13	0.11	0.14	0.13	0.17	0.28	0.31
Other fresh fruits	0.23	0.24	0.25	0.24	0.21	0.22	0.24	0.23
Bedroom furniture	0.23	0.17	0.16	0.20	0.18	0.24	0.25	0.31
Jewelry	0.22	0.06	0.08	0.14	0.17	0.25	0.35	0.38
Bread	0.22	0.33	0.31	0.27	0.24	0.22	0.18	0.17

Wine at home	0.22	0.10	0.12	0.13	0.14	0.21	0.29	0.33
Computers, peripherals, and smart home assistants	0.22	0.20	0.15	0.13	0.14	0.18	0.21	0.21
Fuel oil	0.22	0.14	0.28	0.32	0.31	0.30	0.28	0.27
Uncooked ground beef	0.22	0.37	0.36	0.29	0.28	0.25	0.17	0.15
Men's shirts and sweaters	0.22	0.11	0.16	0.20	0.20	0.20	0.25	0.26
Educational books and supplies	0.22	0.76	0.43	0.17	0.17	0.19	0.25	0.26
Parking and other fees	0.22	0.16	0.13	0.14	0.16	0.21	0.31	0.32
Financial services	0.22	0.13	0.18	0.19	0.20	0.22	0.25	0.25
Men's footwear	0.21	0.35	0.28	0.24	0.25	0.16	0.19	0.22
Uncooked beef steaks	0.21	0.23	0.24	0.23	0.25	0.22	0.20	0.18
Sports equipment	0.20	0.12	0.09	0.18	0.13	0.19	0.20	0.21
Miscellaneous personal goods	0.20	0.14	0.17	0.16	0.18	0.18	0.21	0.19
Infants' and toddlers' apparel	0.19	0.23	0.23	0.23	0.21	0.19	0.16	0.17
Men's underwear, nightwear, swimwear and accessories	0.19	0.15	0.16	0.17	0.18	0.19	0.21	0.23
Boys' apparel	0.19	0.25	0.19	0.18	0.19	0.18	0.19	0.21
Cakes, cupcakes, and cookies	0.19	0.22	0.23	0.21	0.19	0.19	0.17	0.17
Other motor fuels	0.18	0.24	0.12	0.17	0.28	0.28	0.26	0.26
Candy and chewing gum	0.18	0.22	0.19	0.19	0.18	0.19	0.17	0.16
Other dairy and related products	0.18	0.22	0.21	0.19	0.17	0.18	0.17	0.16
Women's dresses	0.18	0.06	0.07	0.15	0.14	0.22	0.17	0.18
Tools, hardware and supplies	0.17	0.16	0.13	0.14	0.15	0.23	0.15	0.15
Fresh fish and seafood	0.16	0.18	0.21	0.18	0.16	0.15	0.18	0.20
Housing at school, excluding board	0.16	0.24	0.17	0.06	0.08	0.09	0.31	0.34
Funeral expenses	0.16	0.53	0.33	0.18	0.18	0.15	0.10	0.11
Boys' and girls' footwear	0.15	0.23	0.16	0.21	0.18	0.12	0.13	0.11
Major appliances	0.15	0.06	0.08	0.10	0.14	0.17	0.16	0.18
Men's pants and shorts	0.15	0.21	0.17	0.13	0.13	0.14	0.15	0.13
Processed fish and seafood	0.15	0.20	0.21	0.17	0.16	0.15	0.14	0.12
Canned fruits and vegetables	0.15	0.21	0.19	0.17	0.15	0.15	0.12	0.11
Other intercity transportation	0.15	0.04	0.06	0.07	0.07	0.12	0.25	0.25
Bacon, breakfast sausage, and related products	0.14	0.22	0.23	0.20	0.18	0.15	0.11	0.09
Postage	0.14	0.14	0.20	0.14	0.15	0.15	0.13	0.14
Other linens	0.14	0.12	0.12	0.09	0.13	0.13	0.15	0.15
Unsampled items	0.14	0.10	0.10	0.07	0.28	0.19	0.09	0.15

Unsampled tools, hardware, outdoor equipment and supplies	0.14	0.09	0.11	0.11	0.13	0.13	0.15	0.12
Vehicle accessories other than tires	0.14	0.15	0.15	0.19	0.19	0.14	0.10	0.09
Nursing homes and adult day services	0.14	0.05	0.13	0.16	0.15	0.15	0.13	0.16
Ice cream and related products	0.13	0.21	0.17	0.15	0.14	0.14	0.13	0.11
Newspapers and magazines	0.13	0.11	0.13	0.14	0.13	0.12	0.13	0.14
Coffee	0.13	0.16	0.18	0.16	0.14	0.15	0.13	0.11
Rice, pasta, cornmeal	0.13	0.20	0.17	0.15	0.13	0.13	0.11	0.11
Men's suits, sport coats, and outerwear	0.12	0.10	0.11	0.08	0.10	0.11	0.15	0.19
Televisions	0.12	0.07	0.06	0.07	0.08	0.08	0.08	0.10
Other furniture	0.12	0.07	0.08	0.07	0.08	0.10	0.15	0.16
Other fats and oils including peanut butter	0.12	0.18	0.18	0.15	0.13	0.11	0.10	0.09
Other appliances	0.11	0.09	0.09	0.09	0.13	0.11	0.09	0.11
Citrus fruits	0.11	0.12	0.13	0.12	0.11	0.10	0.09	0.09
Eggs	0.11	0.18	0.19	0.16	0.14	0.12	0.09	0.09
Fresh biscuits, rolls, muffins	0.11	0.14	0.12	0.13	0.11	0.12	0.11	0.10
Unsampled video and audio	0.11	0.08	0.09	0.15	0.12	0.10	0.08	0.10
Unsampled tuition, other school fees, and childcare	0.11	0.10	0.07	0.07	0.07	0.10	0.16	0.15
Women's outerwear	0.11	0.14	0.12	0.11	0.10	0.09	0.13	0.14
Propane, kerosene, and firewood	0.11	0.17	0.17	0.11	0.13	0.12	0.10	0.10
Video discs and other media, including rental of video	0.10	0.11	0.09	0.10	0.12	0.11	0.09	0.09
Indoor plants and flowers	0.10	0.05	0.08	0.06	0.07	0.09	0.13	0.13
Recreational books	0.10	0.06	0.06	0.07	0.07	0.09	0.11	0.12
Soups	0.09	0.14	0.12	0.10	0.09	0.09	0.07	0.07
Other beverage materials including tea	0.09	0.12	0.11	0.10	0.10	0.09	0.08	0.07
Watches	0.09	0.04	0.05	0.05	0.24	0.04	0.06	0.05
Moving, storage, freight expense	0.09	0.09	0.09	0.07	0.07	0.10	0.09	0.12
Frozen fruits and vegetables	0.09	0.12	0.11	0.10	0.10	0.09	0.08	0.07
Other pork including roasts, steaks, and ribs	0.09	0.12	0.11	0.11	0.11	0.09	0.08	0.07

Apples	0.09	0.10	0.10	0.09	0.08	0.09	0.08	0.08
Care of invalids and elderly at home	0.08	0.23	0.15	0.11	0.06	0.07	0.07	0.09
Uncooked beef roasts	0.08	0.09	0.10	0.10	0.10	0.09	0.09	0.08
Tomatoes	0.08	0.10	0.10	0.10	0.09	0.07	0.07	0.07
Food from vending machines and mobile vendors	0.08	0.16	0.10	0.11	0.10	0.09	0.06	0.04
Nonelectric cookware and tableware	0.08	0.09	0.10	0.06	0.06	0.07	0.09	0.08
Repair of household items	0.08	0.04	0.04	0.06	0.07	0.09	0.10	0.13
Ham	0.08	0.11	0.10	0.11	0.09	0.07	0.06	0.05
Bananas	0.08	0.10	0.10	0.09	0.07	0.07	0.06	0.06
Telephone hardware, calculators, and other consumer information items	0.08	0.06	0.06	0.07	0.08	0.08	0.06	0.05
Potatoes	0.08	0.10	0.10	0.09	0.08	0.08	0.06	0.06
Medical equipment and supplies	0.08	0.07	0.09	0.10	0.07	0.07	0.06	0.06
Other uncooked poultry including turkey	0.07	0.09	0.09	0.09	0.08	0.08	0.07	0.07
Unsampled household operations	0.07	0.05	0.06	0.06	0.06	0.07	0.09	0.10
Window coverings	0.07	0.01	0.03	0.03	0.06	0.05	0.11	0.12
Butter and margarine	0.07	0.14	0.11	0.09	0.09	0.08	0.07	0.06
Baby food	0.07	0.09	0.10	0.09	0.09	0.07	0.05	0.08
Distilled spirits at home	0.07	0.07	0.06	0.05	0.07	0.07	0.07	0.07
Unsampled new and used motor vehicles	0.07	0.03	0.01	0.03	0.05	0.15	0.06	0.07
Audio equipment	0.07	0.07	0.04	0.03	0.06	0.05	0.07	0.09
Car and truck rental	0.07	0.03	0.03	0.03	0.05	0.06	0.09	0.11
Lettuce	0.07	0.08	0.08	0.08	0.07	0.07	0.06	0.06
Salad dressing	0.06	0.08	0.08	0.07	0.07	0.07	0.06	0.06
Other sweets	0.06	0.08	0.08	0.07	0.06	0.06	0.05	0.05
Pork chops	0.06	0.10	0.11	0.08	0.07	0.06	0.04	0.03
Technical and business school tuition and fees	0.06	0.00	0.03	0.04	0.04	0.04	0.11	0.13
Sewing machines, fabric and supplies	0.06	0.04	0.04	0.06	0.06	0.07	0.06	0.06
Motor vehicle body work	0.06	0.05	0.04	0.04	0.07	0.05	0.07	0.07
Photographers and photo processing	0.06	0.03	0.03	0.02	0.05	0.08	0.07	0.10

Other processed fruits and vegetables including dried	0.05	0.08	0.08	0.06	0.06	0.05	0.04	0.04
Tobacco products other than cigarettes	0.05	0.09	0.08	0.06	0.06	0.06	0.05	0.03
Sugar and sugar substitutes	0.05	0.10	0.09	0.07	0.07	0.05	0.04	0.03
Uncooked other beef and veal	0.05	0.07	0.08	0.07	0.05	0.06	0.05	0.05
Flour and prepared flour mixes	0.05	0.07	0.06	0.05	0.06	0.05	0.04	0.03
Photographic equipment and supplies	0.05	0.08	0.03	0.03	0.03	0.04	0.06	0.06
Dishes and flatware	0.04	0.02	0.03	0.03	0.03	0.03	0.05	0.05
Recorded music and music subscriptions	0.04	0.04	0.03	0.03	0.04	0.04	0.04	0.04
Computer software and accessories	0.04	0.04	0.03	0.03	0.03	0.04	0.04	0.04
Music instruments and accessories	0.04	0.09	0.03	0.04	0.03	0.03	0.05	0.03
Floor coverings	0.04	0.01	0.01	0.01	0.03	0.03	0.06	0.06
Unsampled service policies	0.04	0.01	0.01	0.03	0.04	0.04	0.04	0.02
Apparel services other than laundry and dry cleaning	0.03	0.02	0.02	0.02	0.02	0.03	0.05	0.05
Other video equipment	0.02	0.02	0.01	0.01	0.02	0.02	0.02	0.02
Unsampled motor vehicle fees	0.02	0.01	0.00	0.01	0.00	0.03	0.04	0.04
Unsampled recreation commodities	0.02	0.01	0.01	0.01	0.01	0.02	0.02	0.01
Unsampled women's apparel	0.02	0.02	0.01	0.01	0.01	0.02	0.02	0.02
Frozen noncarbonated juices and drinks	0.01	0.02	0.02	0.02	0.01	0.02	0.01	0.01
Unsampled information and information processing	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01
Delivery services	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Unsampled sporting goods	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01
Unsampled men's apparel	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01
Unsampled personal care products	0.01	0.02	0.01	0.01	0.01	0.00	0.00	0.00
Unsampled furniture	0.01	0.01	0.02	0.01	0.01	0.00	0.00	0.00
Unsampled tobacco and smoking products	0.01	0.03	0.01	0.01	0.01	0.01	0.00	0.00
Unsampled recreational reading materials	0.00	0.03	0.01	0.00	0.01	0.00	0.00	0.00

Unsampled public transportation	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
Unsampled appliances	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00
Unsampled photography	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table A2 Comparison of Expenditure Shares in CEX and CPI, December 2013

Item Name	CEX Weight*	CPI Weight
Housing	38.95	41.21
Transportation	19.99	16.67
Food and beverages	16.14	15.18
Medical care	7.94	7.21
Education and communication	6.24	6.78
Recreation	4.10	5.95
Apparel	4.02	3.62
Other goods and services	2.61	3.38

Notes: This table compares expenditure shares in the CEX micro data to the CPI expenditure weights, at the level of eight broad categories. Even at this level of aggregation, there is not a 1:1 mapping between CEX categories and CPI categories. For instance "Computer information services" is classified as "Housing" in the CEX data but gets mapped to "Education and communication" in the CPI categories. We map the following CEX categories to "Other Goods and Services": Miscellaneous, Personal care products and services, Tobacco products and services. We also map "Entertainment" and "Reading" to "Recreation".