Distributional Consumer Price Indices*

Xavier Jaravel

LSE & CEPR

x.jaravel@lse.ac.uk

March 27, 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. 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. 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. 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.

*I thank Ellen Munroe, Sylvia Tian, and Tyler Woodbury for outstanding research assistance. I am grateful to UKRI for generous financial support.

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.

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% with D-CPIs than with the official CPI. The cumulative D-CPIs from 2002 to 2019 range from about 84% at the bottom of the income distribution to about 69% at the top. 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%. These results show that D-CPIs can have important implications for the measurement of inequality. The differences across other socio-demographic groups – age, race, urban vs. rural – are less spectacular.

Second, I use D-CPIs to adjust the poverty line. The official CPI fails to account for the fact that

[1] Prior work on inflation heterogeneity includes several academic contributions (e.g., Kaplan and Schulhofer-Wohl (2017), Jaravel (2019), Argente and Lee (2021), Jaravel and Lashkari (2023)), as well as recent work from the Bureau of Labor Statistics which confirms the earlier findings using confidential BLS data (Klick and Stockburger (2021)). 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). Indeed, Kaplan and Schulhofer-Wohl (2017) and Argente and Lee (2021) only study consumer packaged goods, which account for a modest share of total expenditure (below 15%). Jaravel (2019), Klick and Stockburger (2021), and Jaravel and Lashkari (2023) study the full consumption basket but use data construction steps and price index formulas that deviate from the official CPI.

inflation is lower 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.

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, this compression is amplified by about 1.5pp. Thus, the compression of the real wage distribution at the bottom is amplified by 25% with D-CPIs.

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.

The remainder of this paper is organized as follows. Section 2 presents the data and methodology to compute D-CPIs. Section 3 presents the results, discussing in turn the implications of D-CPIs for the measurement of inequality and poverty in the long-run, and of real wage dynamics during the inflation burst in the wake of the Covid-19 pandemic. Finally, Section 4 presents several extensions, highlighting in particular that the results our robust to using non-homothetic D-CPIs. Complementary results and methodological discussions are reported in the Appendix.

2 Data and D-CPI Methodology

This section present our main data set. Section 2.1 describes how we replicate CPI with the most disaggregated publicly-available statistics; Section 2.2 presents our approach to build group-specific CPIs; finally, Section 2.3 discusses some limitations of our approach.

2.1 CPI Replication with Publicly-Available Statistics

Our 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 of our analysis is therefore to replicate the official CPI with the most disaggregated publicly-available statistics.

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

A primer on the calculation of the aggregate 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).[2] 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 600 UCCs corresponding to the ELIs used in the calculation of the CPI.[3]

To compute the ELI-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 ELI, 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.[4] 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.[5]

Second, in between the biennial or annual December updates, BLS obtains updated expenditure shares for every month by adjusting the shares using price data. Specifically, denoting price changes by p_{it}, where i indexes the product categories (ELIs) and t months, the expenditure shares ω_{it} are updated each month based on observed price changes according to the formula:

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

where $\omega_{i0(t)}$ is the baseline December weight and P_t the overall price index at t, with

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

The notation "$0(t)$" refers to the most recent pivot month prior to the month t, when baseline expenditure weights are updated as described in the first step above.[6]

[2] Specifically, we have 294 ELIs after 2020, 296 ELIs between 2010 and 2020, and 303 ELIs prior to 2010.

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

[4] For update year 2021, BLS decided to maintain the normal baseline weight updating practice and use CEX data from 2019 - 2020 after considering potential interventions to mitigate the impact on spending behaviour due to Covid-19, see https://www.bls.gov/cpi/notices/2021/2022-weight-update.htm.

[5] 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 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.

[6] Detailed description on how to estimate the relative importance for a component for a month other than months with new publication of baseline weights can be found at https://www.bls.gov/cpi/tables/relative-importance/home.htm.

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 ELI, BLS calculates the price index relative as the ratio between CPI in the current month and the most recent baseline update month, and multiplies this ratio by the baseline expenditure weights in use. The resulting relative importance weights are then renormalized such that they sum up to 1 in the current month.

The formula above effectively assumes that preferences are Leontief across ELI categories, i.e. product categories with rising relative prices will be assigned larger imputed expenditure shares. The purpose of this procedure is to provide updated expenditure shares in real time, obviating the need to use actual expenditures each month.

The official CPI is then computed as:

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

The BLS also computes a chained CPI, using actual monthly expenditure shares observed in the CEX for each product category. But this index can only be produced with a lag, because the CEX data is released with a one-year lag. We further discuss the chained CPI below.

Publicly-available data and five associated challenges. Our 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 directly we were able to obtain additional crosswalks for 2023, 2022, 2020 and 2010. To create the crosswalks for the remaining years, we used a concordance published by the BLS which tracks how UCCs change over time. These changes happen at the quarterly level which means we had to create quarterly crosswalks. This involved using the ELI present in the 2023, and mapping the corresponding UCCs to every quarter in the past using the UCC changes concordance.[7] Our crosswalk goes back to 1999, making it possible to compute inflation rates from year 2022. We 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. Therefore, we lose a little bit of granularity when working with the publicly-available item-strata price series, rather than with the 300 ELIs used by the BLS. However, 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, we extend the UCC-ELI concordance into a UCC-item strata concordance and conduct our analysis at the level of item strata in everything that

The monthly relative importance is also published in the CPI News Releases (Table 2), see https://www.bls.gov/bls/news-release/cpi.htm.

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

follows. Specifically, we compute expenditure shares ω_{it} as described above, except 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. Specifically 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.[8] 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.[9] We use the price series of the most immediate overarching category for each of the 30 item strata without original price information. For example, we 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, therefore we drop them from the analysis. Moreover, the four item strata covered by health insurance map to the same price series. We are thus left with 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. We address this limitation by imputing the missing price changes using 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 we restrict the dataset to household who reported all of their income. After 2004, the BLS started imputing missing income and this restriction became unnecessary. Second, while our analysis requires computing monthly expenditures, for many items the CEX survey respondents only report expenditures over the prior three months. For every respondent, we create a three-month panel, and distribute their expenditures evenly across each month. We carefully keep track of the survey weights so that, when aggregating these monthly expenditures to the yearly level, we match the published CEX summary tables. 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, we 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 Ownerss equivalent rent, 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.

[8] 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.

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

To ensure that we handle the CEX data in a manner consistent with the BLS' practice, we implement several checks. First, we leverage the fact that the BLS publishes the set of weights $\omega_{i0(t)}$ used in pivot months. We use these weights directly when calculating aggregate CPI to ensure that there is no source of error from the CEX. We will also check below that we obtain very similar results when we compute the shares directly from the CEX data. Finally, we will compare the patterns in our cleaned data set to the official CEX summary tables published by the BLS, as discussed in Section 2.2 below.

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:

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

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

These expenditure shares are only available with a lag, due to processing time with the CEX data, and so the Torqnvist 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. A CES index assumes a constant elasticity of substitution above zero. Using expenditure and price data from 2003 to 2014, the BLS they calculates an elasticity of substitution $\sigma \approx 0.6$ for most years. To update the weights from the most recent baseline period $0(t)$, we can use the CES update formula:

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

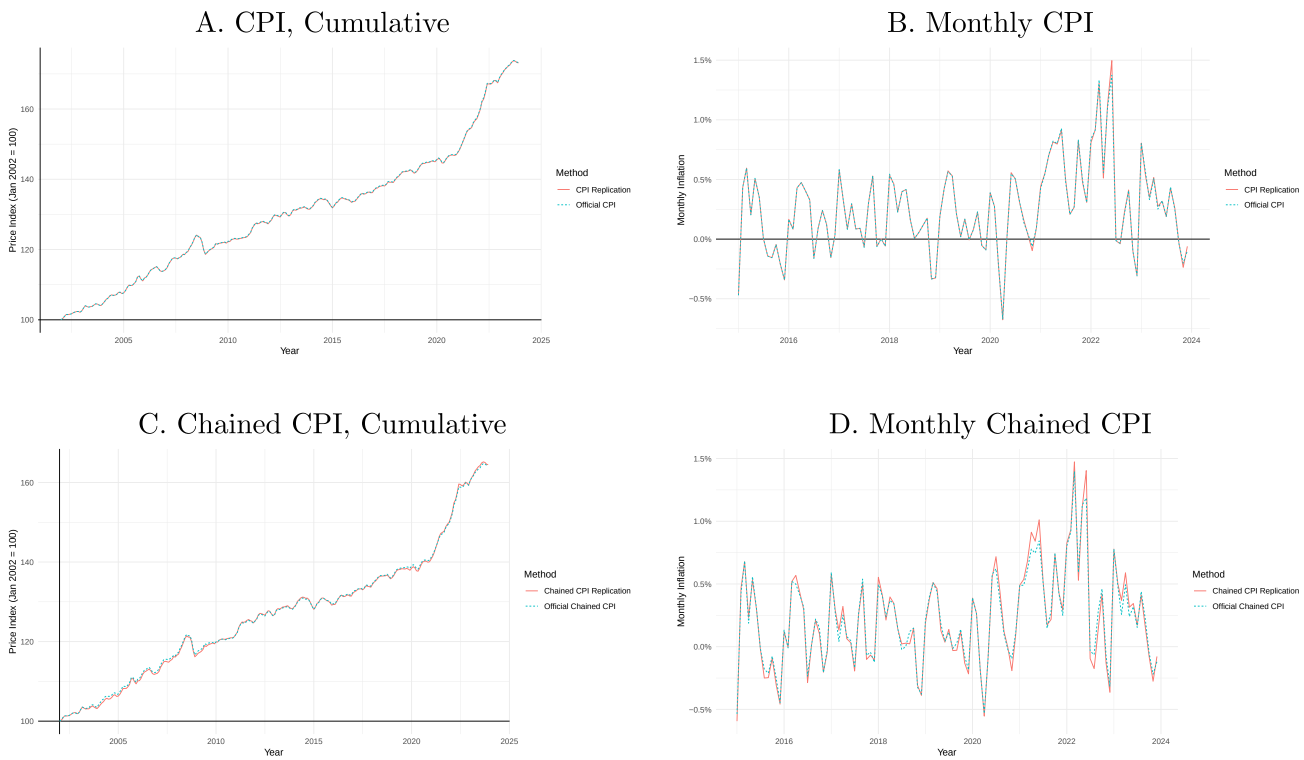
The final CES price index is then calculated as:

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

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. We follow the exact same procedure.

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 the data cleaning steps in the CEX survey – and the fact that prices series are only available at the item strata level of aggregation, rather than ELI, with missing data points in some cases requiring imputation. To assess how well we manage to replicate CPI, in Figure 1 we plot the results we obtain with the publicly-available data against the official statistics released by BLS, from 2002 to 2023.

Figure 1 Database Validation Tests



Notes: This figure compares the price indices published by the BLS to our 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.

We first present the comparison with the official CPI using the Laspeyres formula, in cumulative terms in panel A and by month in panel B. We 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 we use slightly different price series (at the item strata level rather than for ELIs, with imputation when needed). We find that our index is almost indistinguishable from the official index.

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

2.2 Computing D-CPIs

Having established in the previous section that we can replicate official indices 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 we obtain group-specific CPIs. We can apply this approach 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, we start from the official set of weights $\omega_{i0(t)}$ used in pivot months and published by BLS. We 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, we proceed in four steps:

First, we use the official CEX summary tables published by the CEX to check that we obtain the correct group-specific expenditure shares for each product.[10]

Second, using the UCC to item strata crosswalk, for each item strata we compute the share of sales to each socio-demographic group, indexed by g.

Third, using the shares from Step 2, we distribute the official expenditure weight of the item strata (used in the calculation of the official CPI) across group g. For example, say that we observe that 25% of expenditures on the item strata for car purchases come from households in the top 5% of the income distribution. We 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, we obtain expenditure shares $s_{i0(t)g}$, with $\sum_i s_{i0(t)g} = 1 \, \forall g$.

So far, we have obtained expenditure shares for each socio-demographic group in pivot months. Our last step is to use the same methodology as the BLS to obtain updated shares in other months, i.e. applying the formula $s_{itg} = \frac{p_{it}/P_{tg}}{p_{i0(t)}/P_{0(t)g}} \cdot s_{i0(t)g}$, with $\frac{P_{(t+1)g}}{P_{tg}} = \sum_i \frac{p_{i(t+1)}}{p_{it}} s_{itg}$. We thus obtain expenditure shares and price indices by socio-demographic groups over time, using a method fully consistent with the calculation of the official CPI.

Finally, in robustness analyses we will use a chained CPI index specific to each socio-demographic group. Indeed, we can compute the price index in equation (2) using monthly expenditure shares for each group, which we directly measure 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 54.72% of total spending in CPI.

Additional results are reported in the 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.

2.3 Limitations

Before proceeding to our results, it is worth highlighting a few limitations of our analysis.

First, since our goal is to stay as close as possible to the official CPI methodology, our analysis is naturally subject to any limitation affecting this index. These limitations include in particular the use of lagged (rather than current) expenditure shares in pivot months and the choice of a Leontief functional form to update the expenditure shares with price data starting from pivot months, which may lead to

[10]The BLS publishes a set of yearly expenditure summary tables by income quintile that we use to validate that we are processing the microdata correctly. These summary tables require substantial cleaning. For example, in some years these tables are in a text file where the names of some items span multiple lines. The format also changes substantially over time and the expenditures are only reported by the names of categories (rather than a code), which can also change over time. We take the expenditures from 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.

substitution bias (of potentially different magnitude for different socio-demographic groups). We can however directly address these potential issues by using the Chained CPI.

Another limitation is that the official index is homothetic. Given our focus on indices by income groups, in an extension we repeat the analysis with a non-homotheticity correction (see Section 4).

Other limitations are inherent to the database. For instance, we cannot allow for heterogeneity in inflation rates *within* item strata. We can however supplement the publicly available data with more granular data from other sources, e.g. scanner data covering consumer packaged goods. Moreover, at the item strata level, inflation rates could be different in different locations. We return to these issues in Section 4 and proceed until then with the baseline price index.

3 Results

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

3.1 Long Run Inflation Inequality

We 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 our new price series have important implications for the measurement of inequality and the indexation of the poverty line.

Long-run inflation inequality by income percentile. Although prior work has examined inflation heterogeneity across income group, to date there is no evidence covering the full consumption baskets and using a methodology identical to the official CPI. Figure 2 report our results, analyzing inflation across income percentiles from January 2002 to December 2023.

We 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.

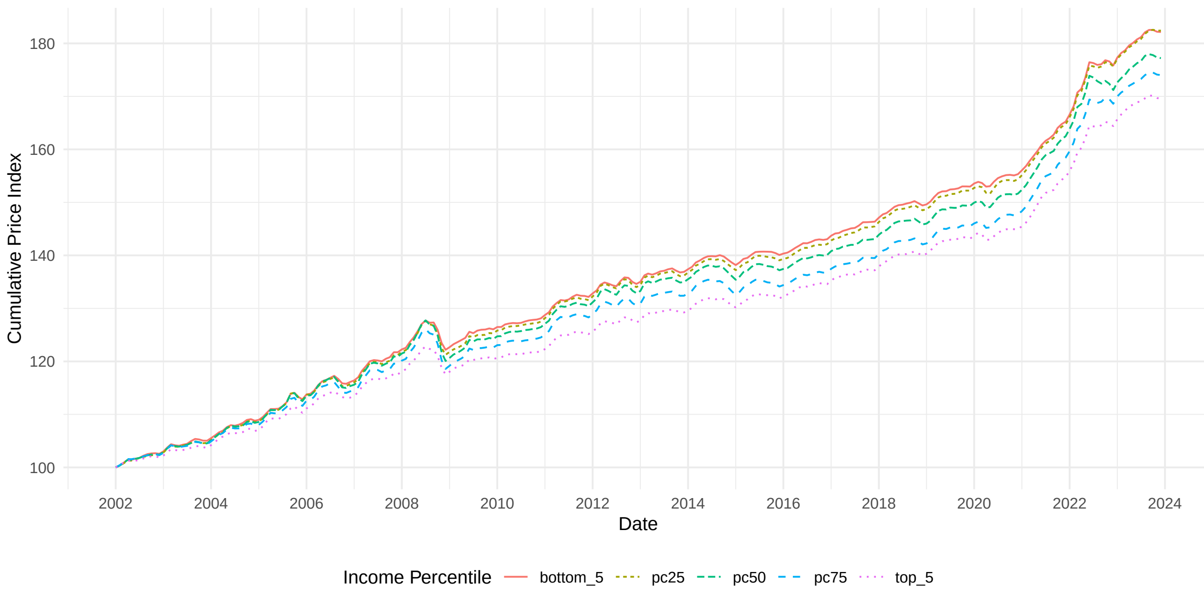
How important are these trends for inequality? To address this question, it is useful to plots household income growth across the income distribution using the official CPI index and our index adjusting for inflation inequality. We use the official statistics of the U.S. Census Bureau to get household income group by income quintile and for the top 5%.[11] We focus on the period from 2002 to 2019, stopping the analysis before the onset of the Covid-19 pandemic (before turning to this 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, up to 26.5% for the top 5% of households. This gradient becomes considerably steeper with our income-group-specific price indices. After accounting for inflation inequality, household real income income growth is only 2.4% at the bottom of the distribution, i.e. earnings are almost

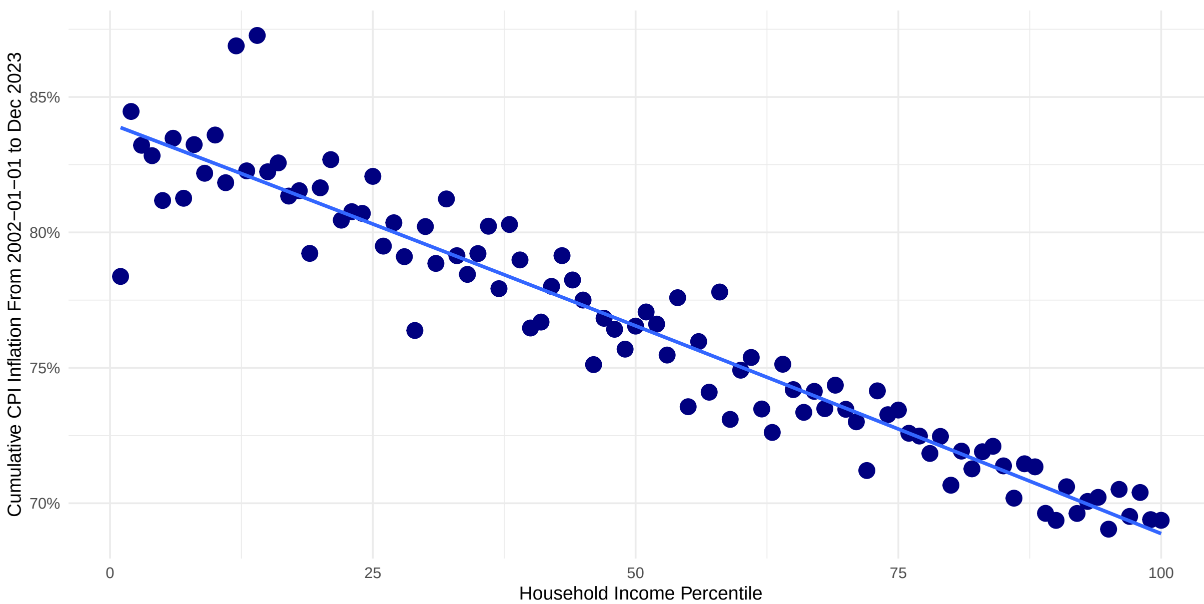
[11] The Census estimates are based on CPS data. We obtain virtually identical results when working with the CPS micro data directly.

Figure 2 Long-Run Inflation Inequality by Income Percentile

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

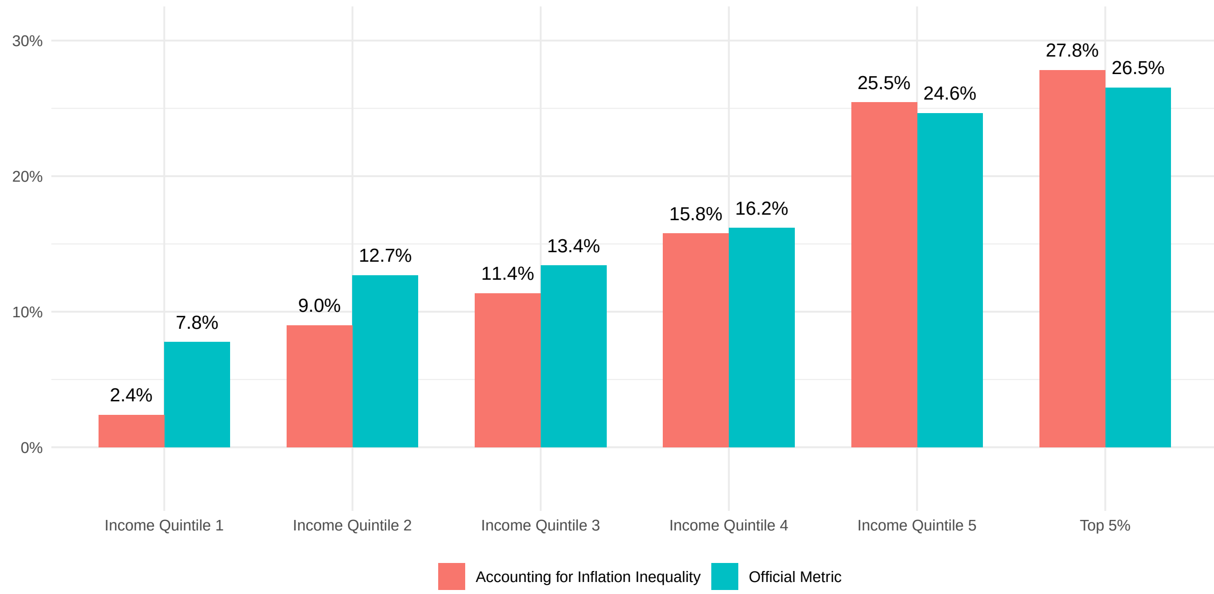


B. Cumulative Index in 2023 across the Income Distribution



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.

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, we show that the results are qualitatively similar – and somewhat stronger quantitatively – when using the Chained CPI formula (i.e., 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 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 (= 1.302/1.126 − 1). With Chained D-CPIs, the income gap increases by 24.8% (= 1.324/1.061), i.e. the rate of increase in real income inequality is about 60% faster than with the official Chained CPI. Thus, the amplification of inflation inequality is even stronger with Chained D-CPIs than with the baseline D-CPIs.

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 4 analyzes this question. We use CPS 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 lower for individuals in poverty, i.e. the poverty line should be indexed at a higher rate. Instead, we use our database to keep track of the inflation rate experienced by individuals at the poverty line. Specifically, we calculate the change in the price index for households at the 90th percentile of household income conditional on being below the poverty threshold.[12] We implement this calculation iteratively: starting in 2002, we compute a poverty-specific price index which we then use to adjust the official poverty threshold for income inequality, iterating this process year after year. For example, we update the 2003 official poverty threshold using the relative difference between the 2002 poverty-specific price change and the price change in the official CPI. We repeat this process in all future years, using the updated thresholds from the previous year.

Panel A of Figure 4 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 our revised price series is thus of direct policy relevance.

Long-run inflation inequality across other socio-demographic groups. Figure 5 reports inflation heterogeneity by age, race, and for urban vs. rural households.[13]

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.

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 we investigate further below –, the differences appear to be relatively short-lived.

3.2 Inflation Inequality after the Covid-19 Pandemic

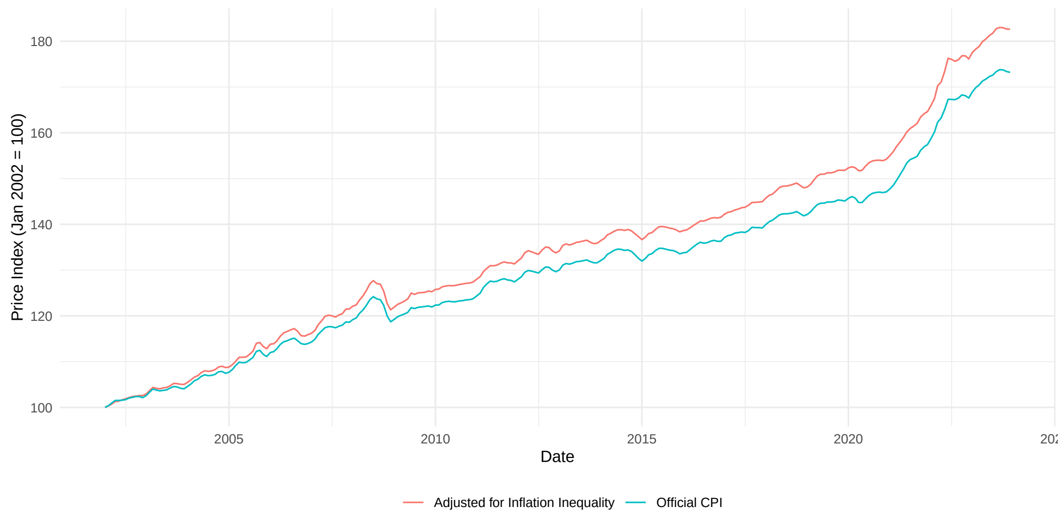
Having documented long-run trends in inflation inequality, we now focus more specifically on inflation heterogeneity across groups right after the Covid-19 pandemic, from May 2020 to May 2022. Inflation was high during this period, in particular because of two product categories, gas and new/used vehicles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

[12] We use the 90th percentile since there is a small tail of household with a large income even after our correction for household size. This occurs since we are correcting 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. These choices do not matter for our results: we obtain similar results when computing the average inflation rate for all individuals below the poverty line, rather than focusing on households at the 90th percentile.

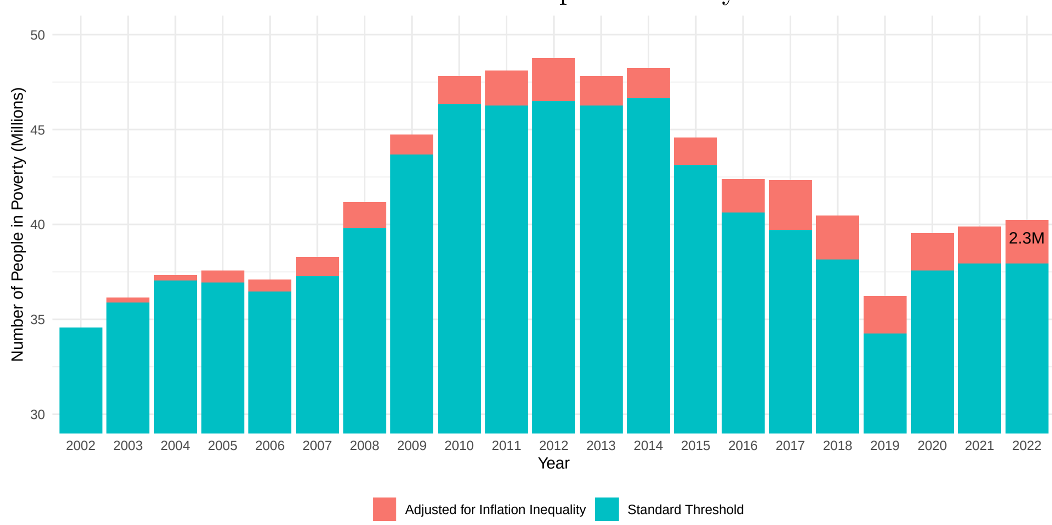
[13] We use the age and race of the reference person to build these price indices.

Figure 4 Implication for the Poverty Line

A. Cumulative Index by Poverty Status

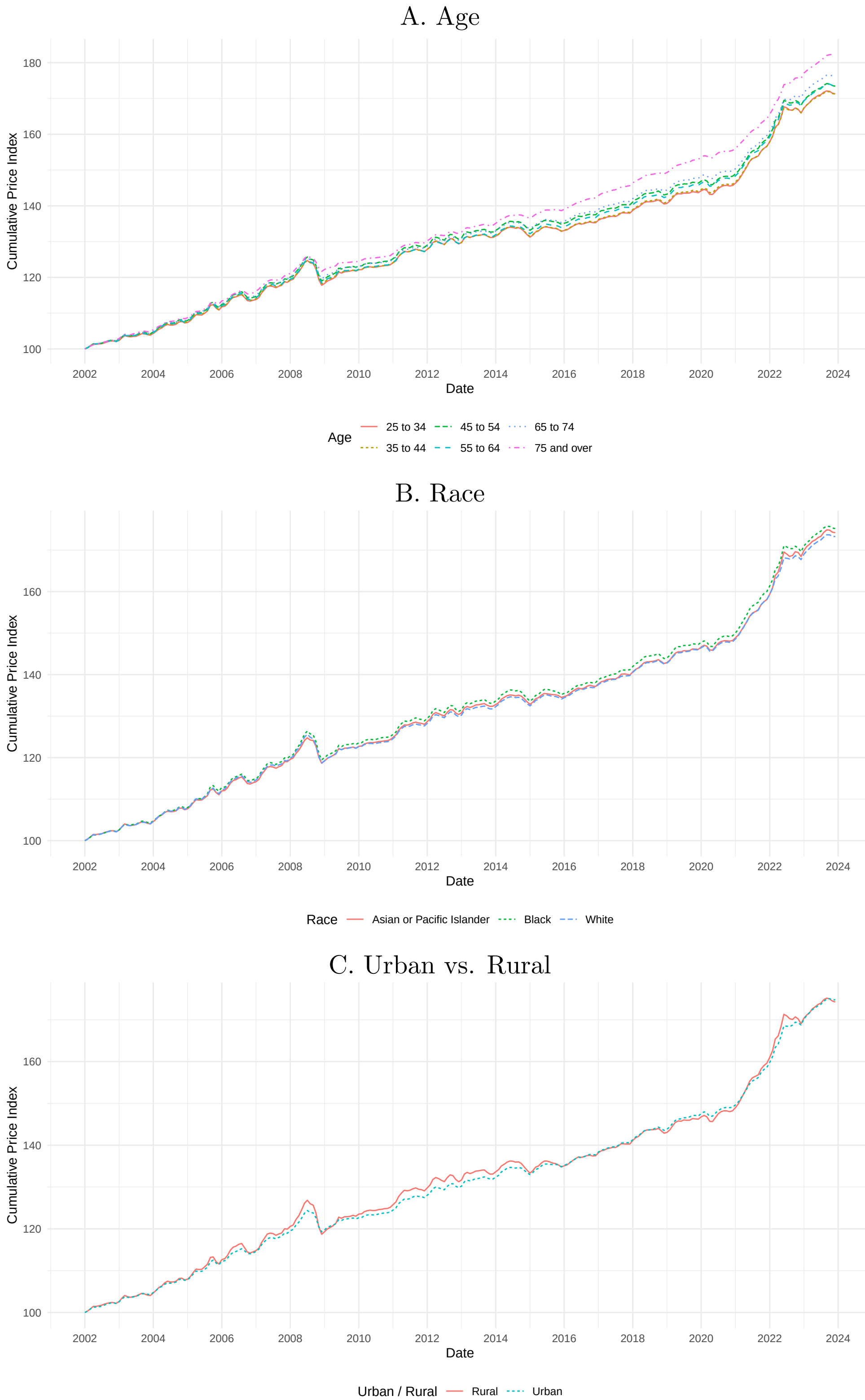


B. Number of People in Poverty



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).

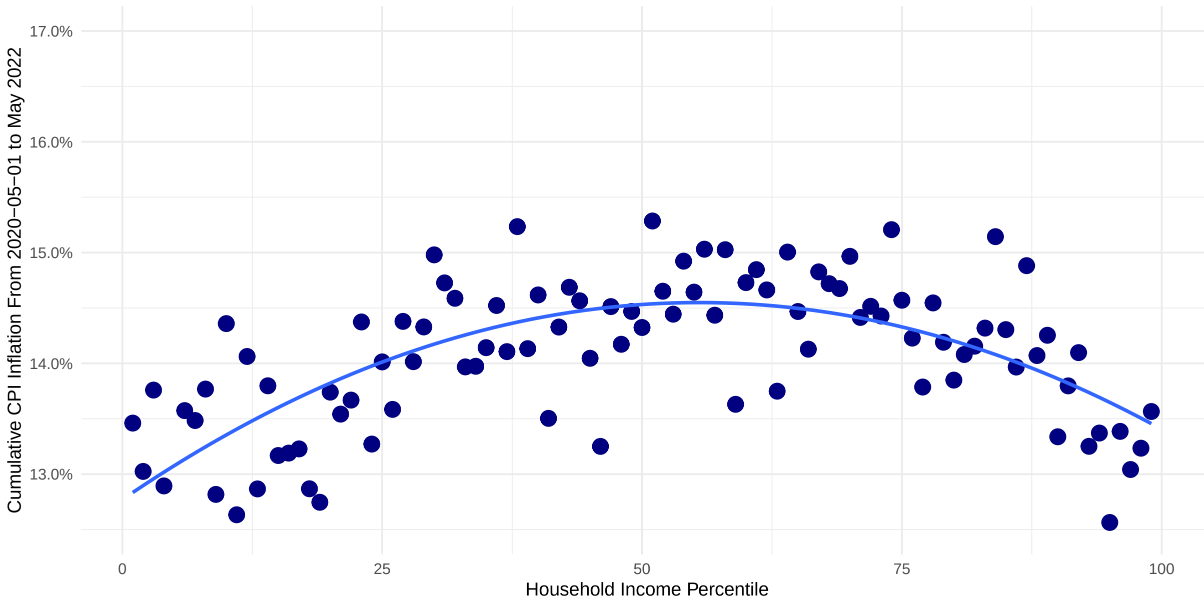
Figure 5 Long-Run Inflation Inequality across Other Socio-demographic Groups



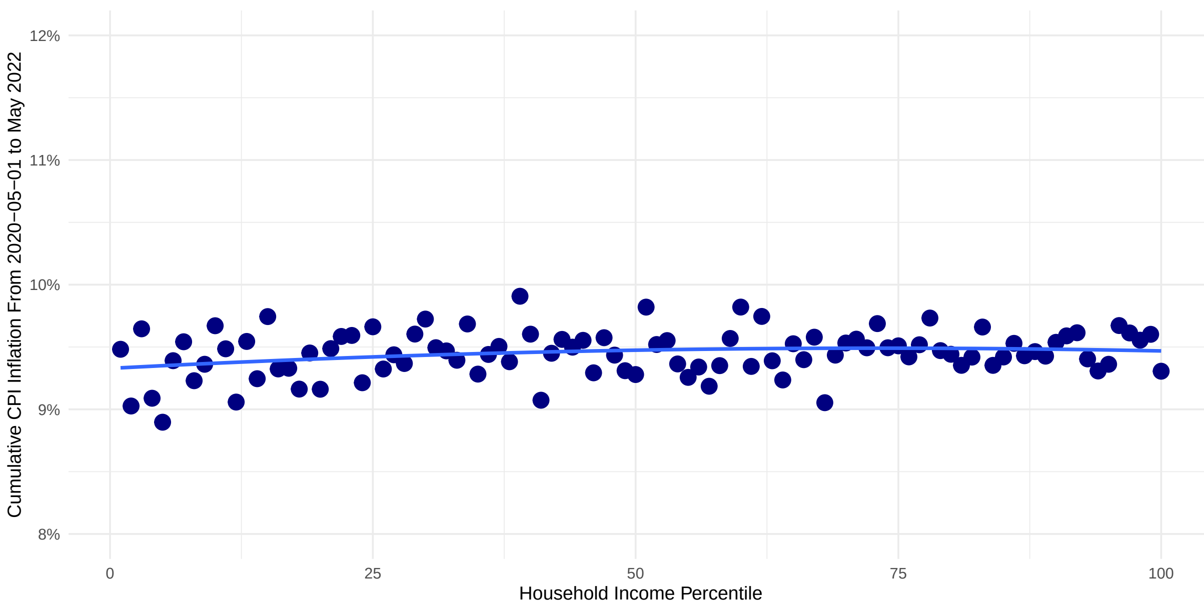
Notes: This figure reports cumulative price indices from January 2002 to December 2023 for various household groups, by age (panel A), race (panel B), and urban vs. rural households (panel C).

Figure 6 Short-Run Inflation Inequality by Income Percentile

A. All products, May 2020-May 2022

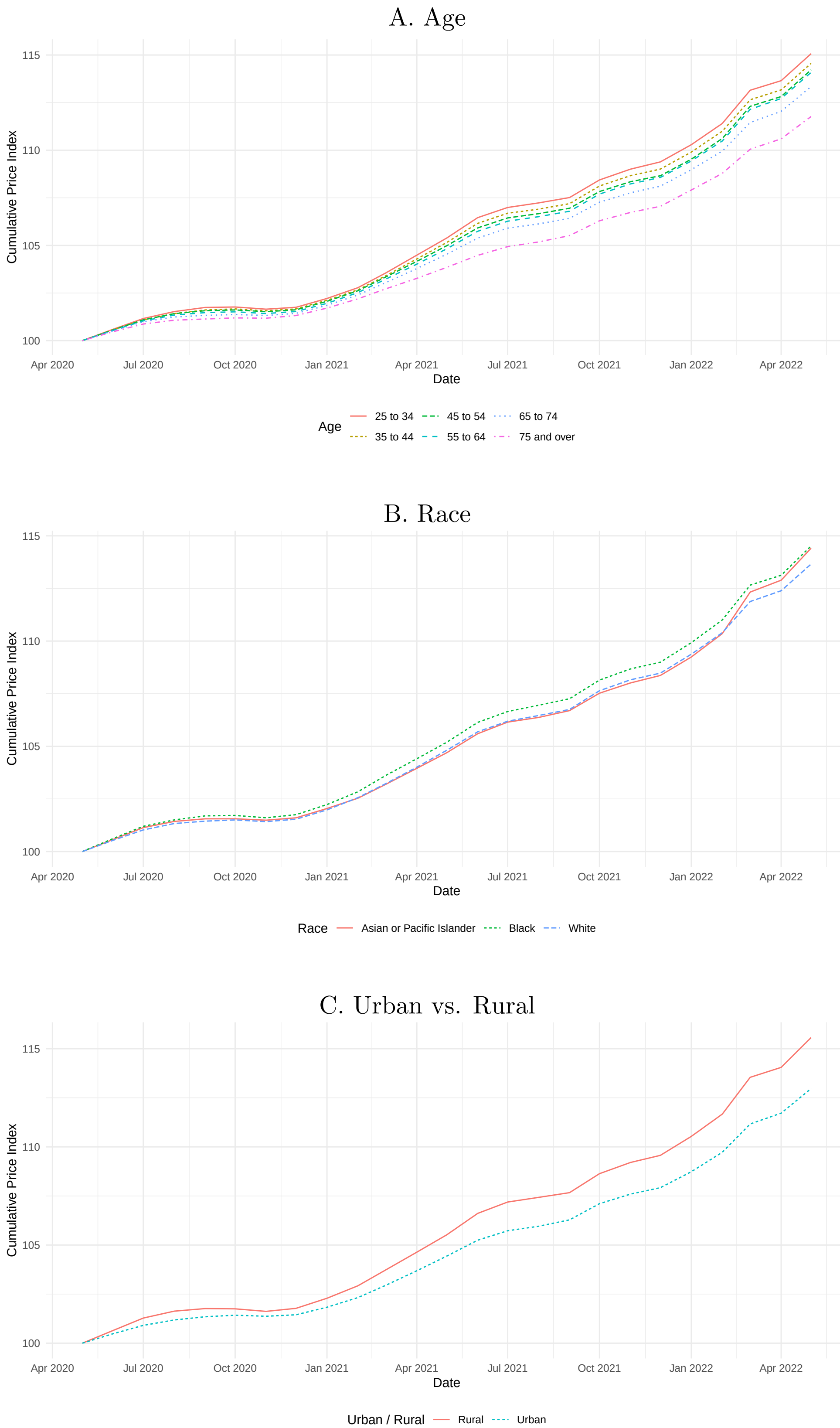


B. Excluding Gas & Vehicles, May 2020-May 2022



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 7 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).

Figure 6 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 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 6, 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 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 7 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.

4 Extensions

Several limitations of the current analysis stem from the fact that we closely follow the methodology of the BLS, which we work toward relaxing in current work.

First, the measurement exercise is based on publicly-available price series which may not be granular enough to capture the full extent of inflation inequality (see e.g. Jaravel (2019) for a discussion of aggregation bias). These data series could be supplemented with more granular data, e.g. scanner data available from the private sector. More generally, it would be valuable to collect more granular data for all sectors of the economy and develop appropriate procedures to disclose high-frequency public statistics constructed from private sector price data (e.g., in the spirit of Chetty et al. (2023)).

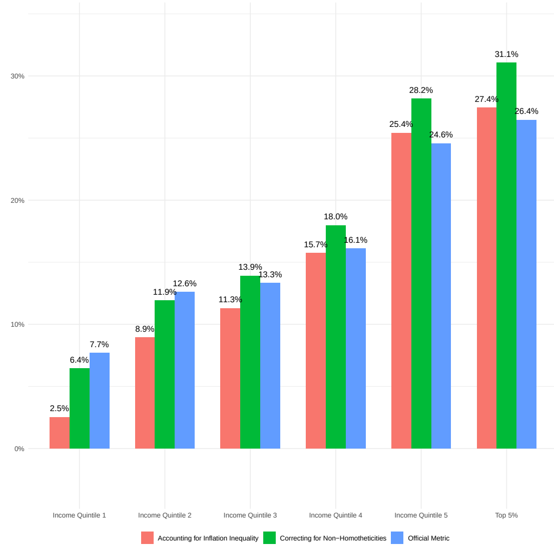
Second, our current analysis does not account for patterns of geographic heterogeneity. This limitation can be relaxed by using geographic-specific shares and price series from thirteen US regions. Out of the 204 item strata price series, 11 are published at the regional level, covering 44% to 49% of total spending depending on the year.[14]

Third, the analysis so far use homothetic price indices. Even when price indices where computed by income group, the maintained assumption was that each income group has homothetic preferences. This can be directly relaxed by implementing the algorithm of Jaravel and Lashkari (2023) to obtain the

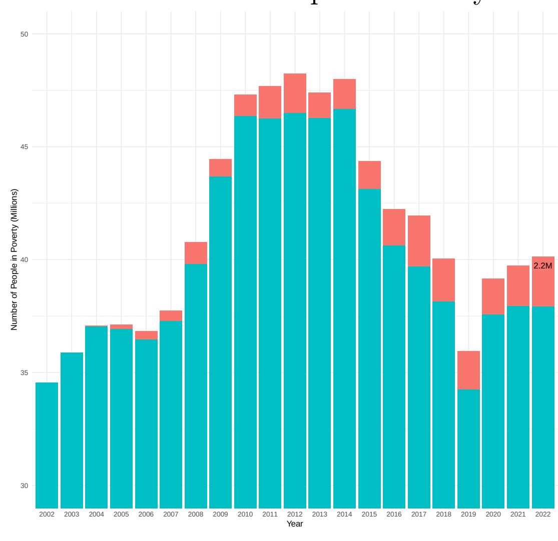
[14]Starting in 2015, the price series for these 11 items are available at the level of 23 different metropolitan areas.

Figure 8 Results with Non-homothetic 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).

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.

Figure 8 reports the results. Panel A focuses on real income growth across the income distribution, from 2002 to 2019. We 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), very 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: 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.

Fourth, another extension is to compute household-level (rather than group-level) price indices. This analysis is in the spirit of Kaplan and Schulhofer-Wohl (2017), who analyze inflation at the household level in a sample of consumer packaged goods. Using the same method as described above for groups, we can distribute aggregate expenditure shares across households and obtain a distribution of inflation rates across households for each month.

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. 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.

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.

References

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

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

Chetty, Raj, John N Friedman, Michael Stepner, and Opportunity Insights Team, "The Economic Impacts of Covid-19: Evidence from a New Public Database Built Using Private Sector Data," *The Quarterly Journal of Economics*, 10 2023, p. qjad048. 4

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, 4

_ **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, 8

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, 4

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

March 2024

A Data Appendix

In this appendix, we present 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 change 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 entry-level items. 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 strict 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 imputed 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, we obtian the count of unsampled ELIs by counting unsampled item strata and 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). We 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 we use the percentile across a given survey.

The BLS publish a series of yearly expenditure tables that contain total expenditures by income quintile at various levels of aggregation. We use these tables to validate that we are processing 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 Additional Information on the Calculation of Aggregate CPI

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

To illustrate the logic of this additional step, consider the computation of expenditure shares in December 2017. We explained in Section A that 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 price changes between 2015-2016 and December 2017.

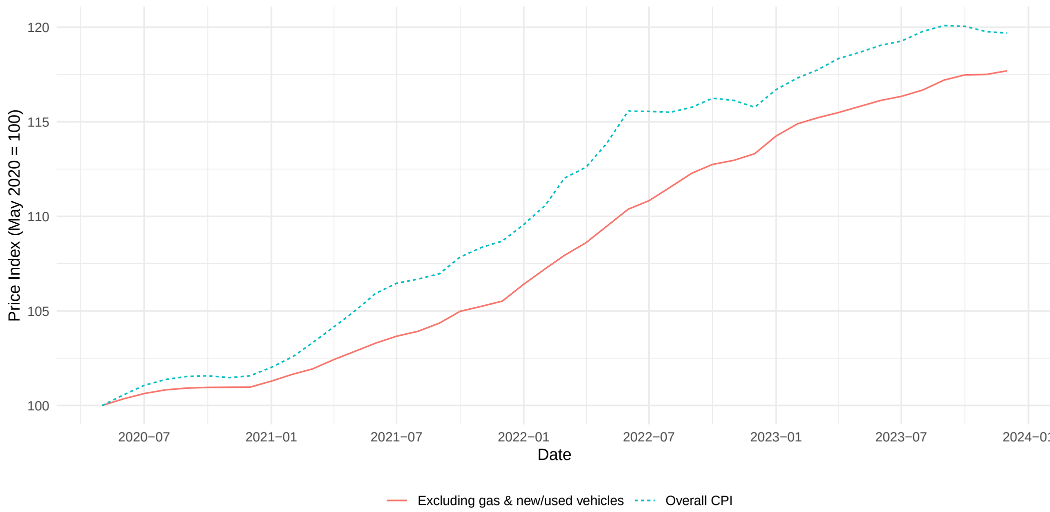
For consistency with the notation in the main text, let us denote the reference December period by $0(t)$, e.g. December 2017. $s_{ib(0(t))}$ denotes the expenditure share of item i in the base period $b(0(t))$ – e.g., with $0(t) =$ December 2017, $b(0(t))$ is 2015-2016. $s_{ib(0(t))}$ is computed directly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

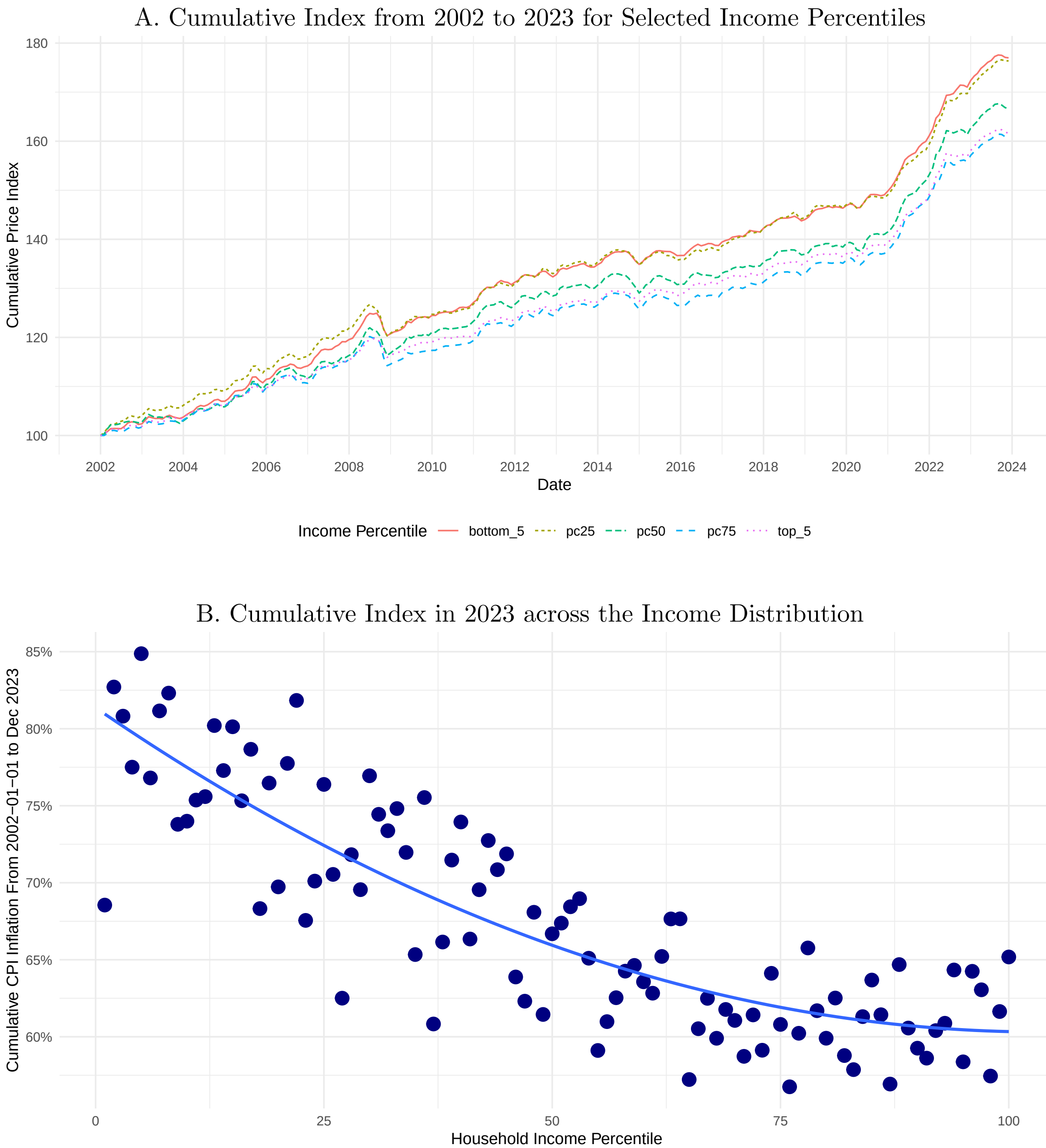
C Additional Figures and Tables

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



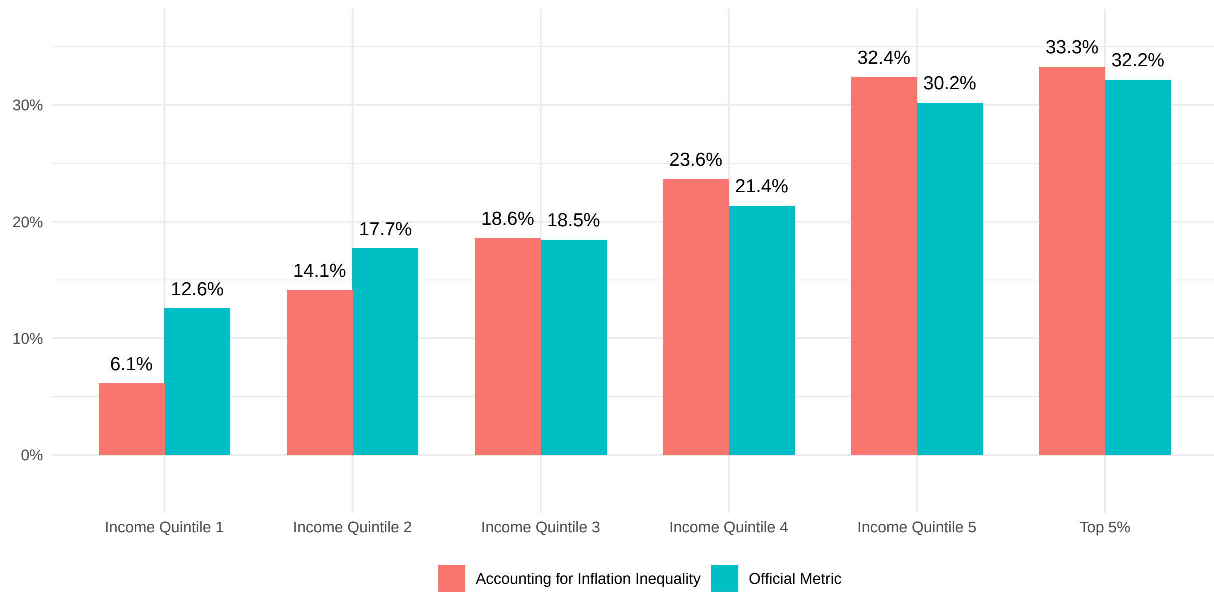
Notes: This figures plots the official CPI, as well as the CPI excluding gas and new/used vehicles, from May 2020 to December 2023.

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.

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".