BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Prices and Living Conditions

Experimental CPI for lower and higher income households

Josh Klick
Anya Stockburger
U.S. Bureau of Labor Statistics

Working Paper 537
March 8, 2021

Experimental CPI for lower and higher income households[1]

Josh Klick, Anya Stockburger

Abstract

This paper examines CPI indexes for subsets of the target population defined by the bottom and top of the income distribution and compares results with the target population. We use data from the Consumer Expenditure Surveys (CE) to construct biennial and monthly market basket shares for groups of respondents based on their reported income, in order to calculate CPIs using modified Laspeyres and Tornqvist formulas respectively. From 2003 to 2018, we find the Laspeyres index for the lowest income quartile population rose faster than the index for all urban consumers. The Laspeyres index for the highest income quartile population rose slower than the index for all urban consumers. Chained CPI indexes for the income quartile populations rose slower than their Laspeyres counterparts. The measure of consumer substitution was lowest for the lowest income quartile population; the difference between the Laspeyres and Tornqvist index for the lowest income quartile population was less than half the difference for all urban consumers.

Introduction

The Consumer Price Index (CPI) measures the change in the cost of goods and services purchased by consumers between two time periods. The target population for the headline CPI is the urban population (CPI-U), however BLS also calculates estimates of price change for subsets of the target, including those aged 62 years and older (R-CPI-E) and those earning most of their income from a select list of wage-earning and clerical worker occupations (CPI-W). There is a lot of user interest in CPI indexes for lower income households. This paper examines CPI indexes for subsets of the target population defined by the bottom and top of the income distribution and compares results with the target population. We use data from the Consumer Expenditure Surveys (CE) to construct biennial and monthly market basket shares for groups of respondents based on their reported income, in order to calculate CPIs using modified Laspeyres and Tornqvist formulas respectively.

Almost 25 years ago, BLS researchers Thesia Garner, David Johnson, and Mary Kokowski published results for an experimental index for lower income households.[2] The authors used CE Interview Survey data from 1982-1984 and 1992-1994 to generate shares for lower income and lower expenditure households to calculate Laspeyres, Paasche, and Fisher indexes from 1984 to 1994. They found little difference in inflation between urban consumers and both lower income and lower expenditure households. In order to register differences between urban consumers and any subset of the target population, there must be significant differences in budget shares and price change differentials for

---

[1] Many thanks to David Popko for his contributions to earlier versions of this research and to Chris Miller and Greg Barbieri for their assistance compiling the data. We are also grateful to Robert Cage, Thesia Garner, and Sara Stanley for their insightful comments that improved the paper.

[2] https://www.bls.gov/opub/mlr/1996/09/art5full.pdf

those item categories. Without budget share differences, and relative price differences for those items, any measure of price change will be the same across different population definitions.

In a 2002 BLS working paper, Rob Cage, Thesia Garner, and Javier Ruiz-Castillo constructed household specific price indexes.[3] Compared to the earlier Garner et al. results, they found greater differences in inflation rates between urban consumers and the lower income population subset, perhaps because of the inclusion of budget shares for categories, particularly food, collected in the CE Diary Survey. Leslie McGranahan and Anna Paulson of the Federal Reserve Bank of Chicago conducted research over a longer time period to study inflation for lower income consumers and found little long-term differences.[4] Similar to prior BLS studies, this research focused on changing weights to reflect different consumption patterns across different population subsets.

More recently, several academic researchers have used scanner data linked with consumer information to account for differences in consumer behavior at a much finer level than possible with BLS data.[5] When accounting for the heterogeneity across consumers at the lowest levels, these studies generally find lower inflation rates for lower income consumers in the 1990s and early 2000s, and then higher inflation rates for more recent time periods.

In this paper we review the background and issues with calculating CPIs for population subsets, define two income-based populations (lowest and highest quartiles), and describe differences in their demographic characteristics. In the results section, we present (i) a comparison of expenditure share differences across the income-based populations, (ii) index results for both Laspeyres and Tornqvist formulas, and (iii) a comparison of upper level substitution bias. We conclude with final observations and remarks.

Background and Issues

This section begins with a brief explanation of the methods to construct the CPI for the target population, all-urban consumers. This foundation is helpful to explain how this methodology has been adapted to construct CPIs for subsets of the target population, and the various drawbacks due to those adaptations.

BLS selects cities to represent geographic strata (index areas) and sample units (goods or services) to represent consumption item strata. With market basket revisions, BLS may change the number of strata over time. As of January 2018, there were 32 index areas and 243 item strata. The product of these strata create 7,776 elementary index cells for which prices are collected and then aggregated in two stages.

At the first stage, changes in price are averaged across sampled units in each elementary index cell using either a geometric mean or modified Laspeyres formula.[6] The elementary index cells form the building

---

[3] https://www.bls.gov/osmr/research-papers/2002/pdf/ec020030.pdf

[4] https://www.chicagofed.org/publications/working-papers/2005/2005-20

[5] Many examples include Broda and Romalis (2009), Broda, Leibtag, and Weinstein (2009), Agente and Lee (2017), Jaravel (2017), and Kaplan and Schulhofer-Wohl (2017).

[6] The formula choice at the first stage of aggregation is based on the level of consumer substitution for that item category. Most goods and services use the geometric mean formula because consumers are generally able to substitute away from any particular item whose price is rising relative to others. Rent and Owner's Equivalent Rent

blocks for the second stage of estimation. The same calculated elementary index cells are used as building blocks to calculate the target (CPI-U) and subset (CPI-W and R-CPI-E) population indexes, as well as the chained CPI (C-CPI-U) which uses a different aggregation formula and weights. The building blocks for the target population are used as proxies for other populations. Building blocks are not produced independently for the consumption patterns of the population subset of interest.

At the second stage, BLS uses market basket shares to combine price changes across elementary index cells to calculate measures of aggregate price change. Market basket shares are calculated using data collected by the Census Bureau on behalf of BLS in the CE Diary and Interview surveys. Several adjustments are needed to modify CE data for CPI definitions of consumption, the most important of which is an adjustment for expenditures on owned homes to estimate a consumption value. Expenditures on the shelter component of the CPI include rent paid by renters and an estimate of the rent homeowners would pay to live in their home (Owners' Equivalent Rent).[7] BLS calculates market basket shares independently for each population, and these shares are the only index construction difference between populations. BLS uses the modified Laspeyres aggregation to calculate CPI-U indexes (as well as the CPI-W and R-CPI-E indexes described below). It computes an arithmetic average price change weighted by base period quantities. BLS uses the Tornqvist index formula to calculate the final version of the C-CPI-U. It is a geometric average of component price changes weighted using the average budget share for the previous and current month.

BLS has a long history of calculating indexes for subsets of the target population. The CPI-W is the oldest measure of consumer inflation calculated by BLS.[8] In the 1978 revision of the CPI, the urban population was introduced as the target. The wage-earner and clerical worker population is a subset of the urban population, where only CE respondents who work full-time and earn most of their income from a select list of occupations are eligible for inclusion. Hence, when the CPI-U was introduced as the target population the wage earner population became a subset of the target population. BLS produces another index for a subset of the target population to measure price change for older consumers (R-CPI-E). This series began in 1988 at the request of Congress and is published on a research basis. The reasons why the R-CPI-E is published as a research index are listed in Table 1. These same caveats also apply to the CPI-W, which was not reclassified as an experimental index when the CPI-U was introduced, or any other population subset index calculated using the same methodology.

are calculated using a Laspeyres formula because consumers cannot easily move in response to changes in rent. There are a few other categories that use a Laspeyres formula due to the limited ability to substitute (such as prescription drugs).

[7] For more information on the calculation of price change for rent and owners' equivalent rent, see the factsheet https://www.bls.gov/cpi/factsheets/owners-equivalent-rent-and-rent.pdf

[8] The First 100 Years of the Consumer Price Index: a methodological and political history. Darren Rippy. Monthly Labor Review. April 2014. https://www.bls.gov/opub/mlr/2014/article/the-first-hundred-years-of-the-consumer-price-index.htm

Table 1: Primary caveats with BLS approach to calculating indexes for subsets of the target population

Experimental weights: the CE sample is designed to produce reliable weights for the population living in urban areas. The smaller sample of CE respondents used to calculate weights for subsets of the target population are expected to have higher sampling error compared to the full sample of respondents used to calculate urban population weights. The CE sample is also designed to produce expenditure weights for Laspeyres indexes that pool data over 24 months. Tornqvist indexes require spending estimates every month and data limitations constrain the ability to construct reliable monthly weights for demographic subsets of the CE sample.
Areas and outlets priced: the sample of cities is designed to represent the population living in urban areas. Within cities, the sample of retail establishments and rental units are designed to represent the total population. To the extent that subsets of the target population live in different cities (or in different parts of cities) and shop at different stores, the urban samples may not be representative.
Items priced: for goods and services sold in a retail establishment, the unique items selected for pricing are based on sales data within the store. If a subset of the target population purchases different items than the general population, then the items selected for pricing may not be representative.
Rental units priced: the realized sample of rental units may have rent-determining characteristics that are not representative of a subset of the target population.[9]
Prices collected: there is only one set of prices collected. Any discount given to particular groups (such as senior-citizens or veterans) are used in the CPI only in proportion to their use by the urban population as a whole. This could understate the prevalence of this type of discount in an index specifically designed for a population subset.

BLS is researching improvements to methodology to measure price change for a subset of the target population, drawing on the recent work in the international statistical community and academia.[10] In particular, BLS is investigating a different treatment of owner occupied housing. While the concept of owner's equivalent rent is an appropriate conceptual approach for aggregate economic measurement, it does not reflect price change experienced by individuals or households which is most useful for escalation purposes. For homeowners with a mortgage, the imputed rent is used in place of mortgage payments or other out-of-pocket expenses associated with owning a home. For populations with a large

[9] Research by BLS in 2019 and 2020 have shown there is a statistically significant different in rent changes by type of structure of the housing unit, for example whether it is a single family home or an apartment building. Beyond geographic differences, there could be rental unit characteristics that should be controlled to produce unbiased estimates of price change for subpopulations.

[10] The United Kingdom's Office on National Statistics in particular has made several improvements in the calculation of subpopulation indexes that BLS is investigating, including democratic aggregation and a payments approach to expenditures on owner occupied housing. Academic research on consumer heterogeneity, such as work by Greg Kaplan and Xavier Jaravel, also provide valuable insights into potential biases in subpopulation indexes. A summary of this work is outside the scope of this paper.

share of home owners with no mortgage (such as the E population), this process imputes a larger expense of owning a home than out-of-pocket spending.

BLS has been limited in its ability to assess the impact of the drawbacks listed in Table 1 on any particular population subset. In particular, BLS does not have the data needed to research issues related to the first stage of aggregation. Each of these caveats are unique and must be studied separately for each population of interest. For example, using a single set of prices collected might be the most important issue for the older subset, while the areas and outlets priced might be the most important issue for the lower income subset. BLS has produced population subset indexes with these caveats for many years, and there is growing interest in assessing the impact and potentially addressing these drawbacks.

Methodology

Price index number formula

We calculate Laspeyres and Tornqvist indexes following BLS methodology as described in Formulas 1 and 2, respectively.[11] The modified Laspeyres formula is a weighted arithmetic average of constituent elementary index cell price changes. The weights as described in Formula 1 as aggregation weights can be roughly interpreted as quantities corresponding to a 24 month reference period of consumer expenditures. For example, monthly indexes calculated from January 2018 to December 2019 use aggregation weights constructed from consumer spending in 2015 and 2016.

Formula 1: Modified Laspeyres Formula

$$IX_t^{[I,A]} = IX_{t-1}^{[I,A]} * \frac{\sum_{[i,a]\in[I,A]} AW_b^{[i,a]} IX_t^{[i,a]}}{\sum_{[i,a]\in[I,A]} AW_b^{[i,a]} IX_{t-1}^{[i,a]}}$$

Where:

$IX^{[I,A]}$ is the All-Items, All-US aggregate index

$IX^{[i,a]}$ are the elementary index cells

t and *t-1* are the current and previous months

$AW_b^{i,a}$ are the aggregation weights for elementary index cells, [*i,a]*, based on a biennial reference period, *b*

The Tornqvist index differs in both aggregation method and weights. The formula is a geometric average of price change weighted by average budget shares from the current and previous month.

[11] CPI Handbook of Methods, index calculation section. https://www.bls.gov/opub/hom/cpi/calculation.htm#index-calculation

Formula 2: Tornqvist formula

$$IX_t^{[I,A]} = IX_{t-1}^{[I,A]} * \prod_{[i,a]\in[I,A]} \left(\frac{IX_t^{[i,a]}}{IX_{t-1}^{[i,a]}}\right)^{\frac{s_t^{[i,a]}+s_{t-1}^{[i,a]}}{2}}$$

Where $IX^{[I,A]}$, $IX^{[i,a]}$, *t,* and *t-1* are defined as in Formula 1 and $s^{[i,a]}$ are the monthly expenditure shares for the elementary index cells.

As noted in the background section, the CE sample is designed to produce expenditure estimates pooled over a 24 month biennial reference period, *b*. To calculate monthly spending estimates used in the Tornqvist index calculation ($s^{[i,a]}$), BLS uses a ratio allocation approach to allocate national spending on an item category to index areas in order to minimize the number of elementary index cells with missing expenditure data. Where cells are still missing after this procedure, annual expenditures are set to $0.01 (or monthly expenditures of 1/12th of a penny) to synthesize with the CPI-U procedure.[12] The sparsity of data is the primary reason Tornqvist indexes for W and E populations are currently not produced.

The different aggregation method and weights in the Laspeyres and Tornqvist formulas result in different measures of price change. The resulting difference in inflation rates can be referred to as a measure of consumer substitution bias. This bias in the CPI-U index is one of many summarized in various reviews of CPI methodology.[13] In short, consumers tend to respond to price changes by substituting away from (or towards) items whose prices are rising (or falling) faster than average. Since the modified Laspeyres formula holds quantities fixed for two years (as captured by $AW_b^{i,a}$ in Formula 1), that index tends to overstate a true cost of living index when consumers exhibit substitution behavior. A Tornqvist index uses an average budget share from the current and previous time period, and reflects consumer substitution in response to relative price change. Tornqvist indexes (and other indexes that use both current and previous period weights) are closer approximations of a cost of living index than Laspeyres indexes (and other indexes with fixed weights). The generally upward bias in a Laspeyres index is called substitution bias, and at the upper level is measured by the difference in Tornqvist and Laspeyres indexes.[14]

Data and definitions

In this study we use CE data from both the Diary and Interview surveys. Although expenditures for some items are collected in both surveys, the CPI program selects one survey as the source for a particular reference year. We use the same survey source as was used in the production calculation of weights for the CPI-U index.

The time period of study is 2004 to 2018. BLS added an income imputation in 2004 that makes results prior to that time period not comparable. Also, as of this research, BLS had published Tornqvist indexes through July 2019, so we selected December 2018 as a terminus. The base period quantities used in the

[12] The CPI Handbook of Methods, Final C-CPI-U calculation section https://www.bls.gov/opub/hom/cpi/calculation.htm#final-c-cpi-u

[13] The Boskin Commission, CNSTAT *At What Price*, Moulton's NBER paper are a few references.

[14] Since there are two stages of index calculation, there are also two stages where consumer substitution bias can overstate inflation in a Laspeyres index. This measure of substitution bias is at the second stage, or upper level substitution bias.

modified Laspeyres formula are constructed using two years of CE data and updated in January of even years. For example, data from 2001 and 2002 are compiled to create aggregation weights used in index calculation from January 2004 through December 2005. We calculate Laspeyres indexes from December 2003 (2001/2002 weights) through December 2018 (2015/2016 weights). In order to preserve the same base period, we calculate Tornqvist indexes starting in December 2003, and ending in December 2018.[15] We made several adjustments to the data to account for minor CPI item structure changes over this longitudinal time period.

In this paper we define lowest and highest income populations by income quartiles. There are many other possible definitions of low and high income. We focused on a simple definition that ensures a quarter of CE respondents nationally are classified in the populations of interest. In order to define the income quartiles, we pooled respondents from each survey (Diary and Interview) by reference year, then ranked by income, and then divided into income quartiles. Our definition of income is total before-tax income, after imputation. We did not exclude households with incomes equal or below zero, but that is a definitional change that could be considered in future research. The populations we present in this report are the lowest income quartile and the highest income quartile.

Table 2 shows a comparison between these lowest and highest income quartile populations as well as the urban, wage earners, and elderly populations. The median annual income of urban CE respondents over this time period is $48,816. The wage earner and elderly subset of the urban population have slightly lower median annual incomes. By tautology, there are larger differences in annual income when grouping CE respondents by that variable. Looking at the quarter of CE respondents with the lowest and highest income reported, the median annual income was $13,500 and $122,800 respectively.

Table 2: Annual income for Population Cohorts: 2004-2018

Variable	Urban	Wage earner	Elderly	Lowest Income Quartile	Highest Income Quartile
Mean Annual Income	$67,109	$55,802	$51,156	$12,705	$155,045
Median Annual Income	$47,920	$46,099	$33,313	$13,570	$124,362

Source: CE integrated data from 2004-2018, population weighted to represent consumer units in the U.S.

We define the income bounds for the lowest and highest income quartile groupings in this paper based on CE data. Alternatively, one could define the income thresholds using an external source of information, for example to reflect a different benchmark income distribution of the population. According to an analysis conducted in 2019 to study nonresponse bias in the CE, the population earning less than $50,000 a year was over-represented by five to 20 percent when compared with the American

[15] Tornqvist indexes were also calculated from December 1999 through December 2001, but are not presented here for ease of explication. The results in these two early years are similar to the rest of the time period presented in this paper.

Community Survey (ACS).[16] The lowest and highest income quartiles in this paper might reflect lower incomes than corresponding income distribution levels as measured by the ACS, even after adjusting for different definitions of income. For example, CE respondents reporting an annual income in 2016 of less than $25,000 were included in the lowest income quartile in this paper. Using ACS data, the lowest quartile income cutoff is around $28,000. Similarly, based on the creation of the income quartile variable, CE respondents reporting an annual income greater than about $93,000 were include in the highest income quartile, compared to an ACS cutoff of around $110,000. This research could be repeated for other definitions of income.

A comparison of other demographic information across populations is also helpful context to explain differences in market basket shares. As shown in Table 3, relative to higher income respondents, lower income respondents have lower rates of home ownership and educational attainment and lower rates of labor force participation. Other demographic comparisons reveal the overlap in respondents included in the elderly and the lowest income populations. The lowest income population is by definition 25 percent of the urban population. The elderly population is around 30 percent of the urban population, 36 percent of the lowest income population, and 16 percent of the highest income population. This is likely the driving factor behind why, relative to higher income respondents, lower income respondents are older, more likely to be retired, and have higher rates of home ownership without a mortgage. Household size differences across populations are important to note and likely play an important role in the median income differences in Table 2. Income was not adjusted for household size and future research should control for household size to improve comparability across populations.[17]

---

[16] A Nonresponse Bias Study of the Consumer Expenditure Survey for the Ten-Year Period 2007-2016; Krieger et al. https://www.reginfo.gov/public/do/DownloadDocument?objectID=101978401

[17] There is a long literature using equivalence scales to adjust household income to account for different characteristics across households. Angela Daley, Thesia Garner, Shelley Phipps, Eva Sierminska, "Differences Across Place and Time in Household Expenditure Patterns: Implications for the Estimation of Equivalence Scales," BLS Working Paper, 2020 https://www.bls.gov/osmr/research-papers/2020/pdf/ec200010.pdf

Table 3: Demographic Comparisons between Population Cohorts

Variable	Urban	Wage earner	Elderly	Lowest income quartile	Highest income quartile
Home Ownership					
Percent Owner (incl. unknown mortgage status)	64.3%	56.0%	79.0%	41.3%	87.4%
Percent Owner with Mortgage	39.8%	39.8%	26.7%	13.0%	69.5%
Percent Owner no Mortgage	23.1%	14.4%	50.4%	26.6%	17.3%
Age and Household Size					
Median Age of Householder	49	43	70	55	48
Mean Household Size	2.5	2.9	1.8	1.8	3.1
Education Level					
Percent High School Diploma or Above	87.4%	83.5%	82.9%	75.6%	97.1%
Percent Associate's Degree or Above	42.0%	26.8%	35.5%	21.5%	67.8%
Employment Status					
Percent Not Working/Any Reason	31.4%	12.8%	69.7%	58.0%	12.1%
Person Not Working Disabled or Taking Care of Family	10.7%	8.5%	7.6%	20.2%	5.8%
Percent Not Working/Retired	19.0%	3.4%	61.9%	33.5%	5.8%

Source: CE integrated data from 2004-2018, population weighted to represent consumer units in the U.S.

Results

Using these examples, we constructed expenditure weights for lower and higher income populations and used them as input to calculate Laspeyres and Tornqvist indexes. First we present a comparison of expenditure weights for the population definitions, and then index results.

Expenditure Weights

Recall a caveat to the method BLS uses to calculate indexes for subsets of the target is the potential for increased sampling error of expenditure weights. This caution is particularly relevant for populations defined by income. As we show in Table 4, biennial expenditure weights calculated for the 7,776 elementary index cells are rarely missing for the urban population (3 percent of the time during the study period). The rate of missing cells is higher for subsets of the target population, and the highest for the lowest income quartile population. The item structure is defined for the urban population, and some item categories might be less relevant for a subset of the target population. For example there are missing expenditures for the item category Sports vehicles (which includes bicycles, boats, and snowmobiles) in 39 percent of the areas for the urban population and 86 percent of the areas for the lowest income population. Very low (or no) expenditures might be an appropriate proxy for spending by some populations on certain item categories. Nonetheless, the high rate of missing cells is a concerning quality metric that should be further studied.

Table 4: Rate at which expenditure data are missing for elementary index cells (average 2004-2018)

Type of weights	Urban	Wage earner	Elderly	Lowest Income (Q1)	Highest Income (Q4)
Biennial expenditure weight	3%	9%	11%	17%	6%
Monthly expenditure weight	19%	44%	45%	55%	36%

Source: CPI expenditure weights from 2004 to 2018

As we stretch CE data further to calculate monthly expenditure weights for the Tornqvist index, the number of elementary index cells with missing expenditure data increases substantially. BLS publishes a Tornqvist index for the urban population, with a missing rate of 19%. In the past, the high rate of missing data for subsets of the target population monthly expenditures has been a primary reason Tornqvist indexes for the W and E populations have not been explored. This same caveat applies to populations defined by income. Indeed, on average over half of the elementary index cells for monthly expenditure weights are missing expenditure data for the lowest income quartile. Here, the imputation techniques used for the urban population are applied to population subsets to enable calculation of Tornqvist indexes. Imputation of missing expenditure data is another area that could be improved upon in future research.

After imputing missing expenditures to fully populate the elementary index cells, there are several notable differences in market basket shares between populations as displayed in Table 5. The eight major group categories are presented, along with some notable subcategories. Although the market basket shares vary over the time period of study, the comparison of 2015-2016 differences is illustrative of general differences. Note these shares are not identical to CPI relative importances published on the BLS website, which are inflation adjusted to reflect snapshots of weights used in CPI index calculation.

> *Food*: Although spending shares on food in total by the lowest income quartile population is similar to all households, more of their budget is spent on food at home rather than food away from home (such as restaurants) and alcoholic beverages.
>
> *Housing*: The share of spending (or consumption in the case of owners) on shelter (rent and owner's equivalent rent) is highest for the lowest income quartile population. Although there are roughly twice as many renters in the lowest quartile compared to the highest quartile, their budget share allocated to rent is more than four times as high[18]. Recall from the background

[18] Recent research by BLS and Census Bureau linking CE data with rent subsidy information collected by the Department of Housing and Urban Development could have interesting implications for the budget share of rent

section that owner's equivalent rent is an imputed cost of the shelter services provided by owned homes. The imputed budget share for owner's equivalent rent is the lowest for the lowest income quartile population, but more than proportional to their smaller share of homeowners. The lowest income quartile population also spends more on household utilities than the other populations in Table 5 and less on household furnishings and lodging away from home (including hotels and motels).

Recreation: The lowest income quartile population spends more of their budget share on televisions than any other population presented in Table 5. With that one exception, spending shares on all other recreation categories were lowest for the lowest income quartile than any other population.

Education and communication: Spending shares on these item categories are very similar between the urban population and the lowest income quartile population. The highest income quartile population spends more of their budget share on education than the other populations listed and the spending share of the older population is the lowest.

Apparel: Spending shares on jewelry and watches had the largest dispersion across the income distribution. Spending shares on other apparel categories were fairly similar across the income distribution and lowest for the older population.

Medical care: The older population spends the highest budget shares on all medical care categories. The lowest income quartile population spends the least share on physician's services and health insurance.[19] The impact of programs such as Medicaid and Medicare on the budget shares for different populations is an interesting area for future research.

Transportation: Spending shares on transportation goods and services are the lowest for the lowest income quartile population mostly due to differences in expenditures on vehicles and vehicle maintenance and public transportation which includes all forms of non-private transportation (such as fares for air, bus, train, ship, taxis, and ride sharing).

Other goods and services: Overall spending shares on other goods and services are highest for the lowest income quartile population. This is due to larger spending shares on cigarettes and

---

for the lower income population. Future research should explore this particular impact of CE data quality on the calculation of weights specifically for a lower income population.
Garret Christensen, Laura Erhard, Thesia Garner, Brett McBride, Nikolas Pharris-Ciurej, John Voorheis, "The promises and challenges of linked rent data from the Consumer Expenditure Survey and Housing and Urban Development," paper presented at the Joint Statistical Meetings Annual Conference 2019, Denver, Colorado, July 27–August 1, 2019 (U.S. Census Bureau, 2019). See https://www.census.gov/newsroom/press-kits/2019/jsm.html for conference proceedings, including links to all of the papers presented at the conference.

[19] BLS uses an indirect method to measure the price change for health insurance. CE respondents report out-of-pocket spending on health insurance which is mostly allocated to the health care services that are covered by health insurance. The remainder is included in a health insurance retained earnings category which also includes the costs incurred by insurance companies to process claims. Since the factors used to allocate health insurance spending are fixed across populations, the lower overall budget shares of the lowest income quartile population on health insurance retained earnings can be accurately described as lower shares on out-of-pocket health insurance.

miscellaneous personal services (a category that includes legal, funeral, laundry, and banking services).

Table 5: Distribution of total CPI market basket expenditures, snapshot of 2015-2016[20]

Item Category	All urban households (U)	62 years or older (E)	Lowest income quartile	Highest income quartile
Food, total	14.6%	12.4%	15.6%	14.2%
Food at home	7.7%	7.1%	9.5%	6.7%
Food away from home	5.9%	4.5%	5.4%	6.3%
Alcoholic beverages	1.0%	0.8%	0.7%	1.2%
Housing, total	41.0%	45.8%	45.2%	39.5%
Shelter	30.5%	34.4%	34.6%	28.8%
Rent	7.5%	4.7%	15.6%	3.4%
Owner's equivalent rent	23.0%	29.7%	19.1%	25.5%
Household utilities	4.7%	5.0%	5.9%	3.8%
House furnishings and other household services	4.5%	4.8%	3.7%	5.1%
Lodging away from home	1.0%	1.0%	0.6%	1.4%
Recreation	5.9%	5.7%	4.6%	6.7%
Education and communication, total	7.1%	4.4%	6.8%	8.0%
Education	3.0%	0.8%	2.7%	4.4%
Communication	4.1%	3.7%	4.1%	3.6%
Apparel	3.2%	2.1%	2.9%	3.5%
Medical care	8.5%	12.0%	8.2%	7.9%
Health insurance	1.0%	1.2%	0.8%	1.1%
Professional services	3.3%	4.6%	3.0%	3.2%
Transportation, total	16.6%	14.7%	13.0%	17.2%
Motor Fuel	3.7%	2.9%	3.4%	3.3%
Vehicles and vehicle maintenance	9.0%	8.0%	6.0%	9.7%
Motor vehicle insurance	2.1%	2.1%	2.4%	1.9%
Public transportation	1.3%	1.2%	0.9%	1.7%
Other goods and services	3.2%	3.0%	3.6%	3.0%

Source: CE integrated data with CPI division adjustments based on CE data from 2015 – 2016.

Price Indexes

We calculated Laspeyres indexes using biennial budget shares (expenditure weights), like the 2015-2016 shares described in the previous section. Indexes for the lowest and highest income quartile populations are shown in graph 1, along with the indexes for the CPI-W and R-CPI-E for comparison purposes. We show index results at the all items and major group levels in Table 6. The annualized percent change over the time period of study (December 2001 to December 2018) is defined in Formula 3.

[20] Due to rounding, the figures presented may not add to exactly 100%. The component items displayed are not exhaustive so the sum of their market basket shares may not equal the major group.

At the all items level, the annualized change in the lowest income quartile index is larger than that for the urban population (and R-CPI-E index) and the annualized percent change for the highest income quartile index is lower than that for the urban population. The annualized percent change in the lowest income quartile index is greater than the urban population index for the education and communication, other goods and services, housing, recreation, and transportation major groups. The annualized percent change in the highest income quartile index is less than the urban population index for the other goods and services, housing, recreation, and transportation major groups. As a reminder, these indexes differ only in the market basket shares at the elementary index level.

Between 2002 and 2018, the 12-month change in the lowest income quartile index is consistently greater than the urban population index, in 152 out of 169 months. The remaining 17 months occur in 2006, 2009, 2010, and 2011. Further study is needed to understand the cause of these months that are different than the rest. Similarly, the 12-month change in the highest income quartile index is consistently less than the urban population index. Future research should include variance estimation so confidence intervals can be calculated to statistically compare these index results.

Formula 3: Annualized percent change

$$Annualized\ percent\ change = \left(\frac{Terminal\ Index\ Value}{100}\right)^{12/number\ of\ months}$$

Graph 1: Monthly Laspeyres indexes for lowest and highest income quartiles: December 2003 to December 2018

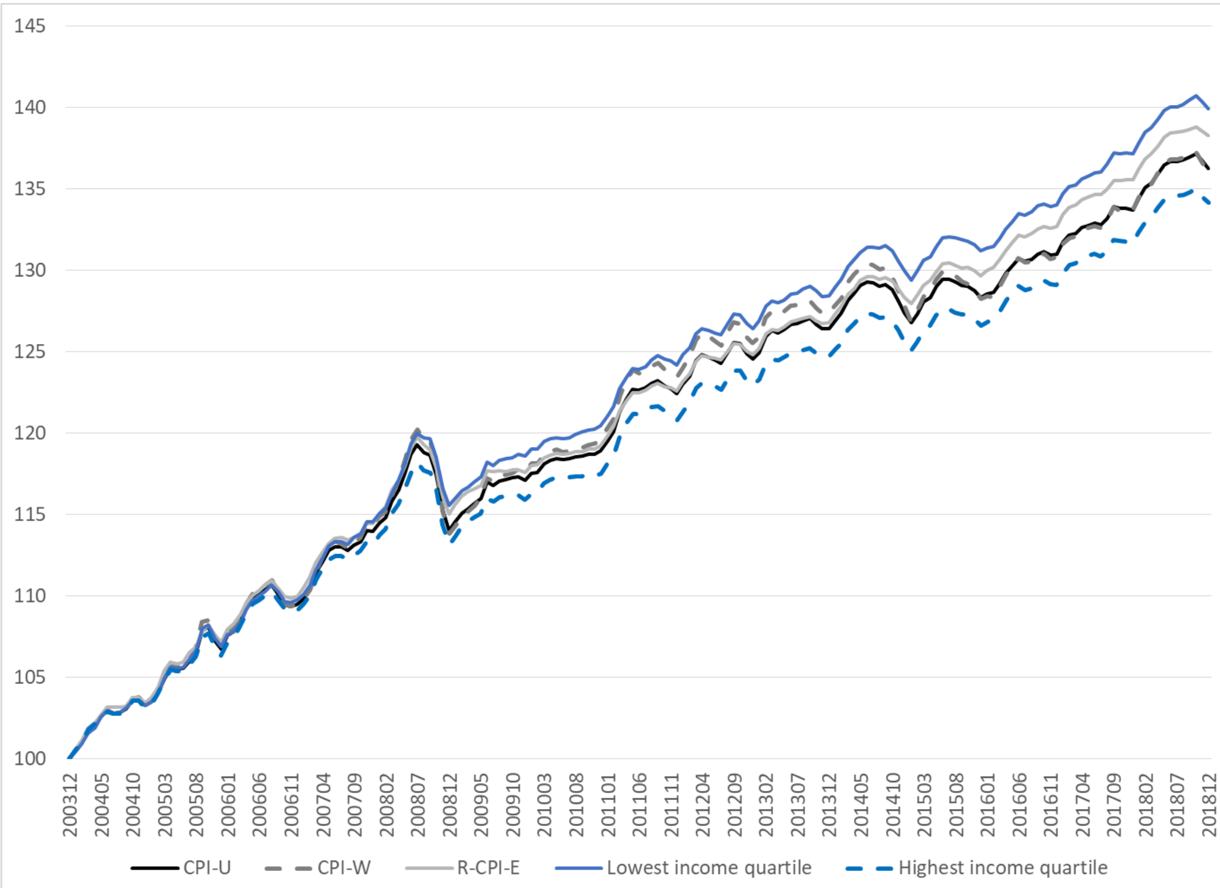
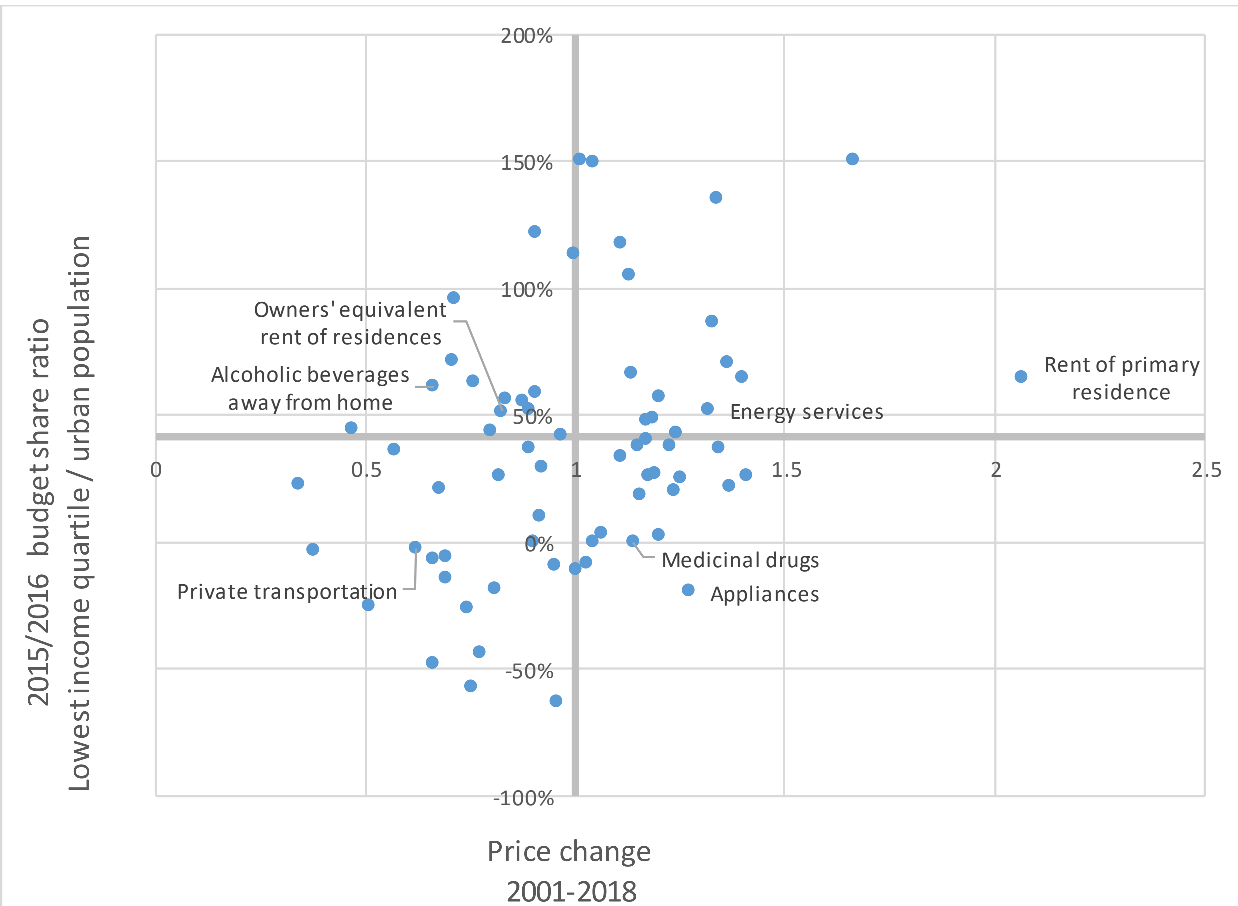


Table 6: Laspeyres index annualized percent changes from December 2003 to December 2018

Item Category	**All urban households (U)**	**62 years or older (E)**	**Wage earner (W)**	**Lowest income quartile**	**Highest income quartile**
All items	**2.07**	**2.17**	**2.06**	**2.25**	**1.97**
Apparel	0.14	0.05	0.10	-0.09	0.23
Education and communication	1.39	0.69	0.86	1.84	1.77
Food and beverages	2.19	2.14	2.18	2.13	2.23
Other goods and services	2.65	2.52	3.07	3.03	2.25
Housing	2.31	2.32	2.36	2.45	2.17
Medical care	3.21	3.08	3.29	3.11	3.29
Recreation	0.70	1.17	0.54	0.92	0.63
Transportation	1.85	1.92	1.93	2.11	1.68

Indexes for subset populations differ from the urban population when there are meaningful differences in budget shares and price change. The scatterplot in graph 2 displays the relationship between long term price change (the percent change in indexes from December 2003 to December 2018) and the difference in market basket shares (the ratio of 2015-2016 biennial shares for the lower income and urban populations) for expenditures classes at the national level.[21] The bolded x-axis shows the percent change of the All Items, US City Average index over this time period (41.6%). They bolded y-axis shows budget share ratios equal to one, where observations greater than one reflect greater spending shares for the lowest income quartile compared to the urban population.

Graph 2: Relationship between price change and budget share differences for the lowest income quartile and urban population



Source: Consumer expenditure survey data from 2015-2016, Consumer Price Index data from December 2001 and December 2018

The population in the lowest income quartile spent more than the urban population on rent and energy services (which includes electricity), as a share of total spending, and the indexes for both of those items rose faster than average over this time period (upper right quadrant of the graph). Conversely, the lower income population had lower budget shares for items whose indexes rose slower than average (or fell) such as private transportation, which includes new and used vehicles (lower left quadrant of the graph).

[21] The 243 item strata that form the building blocks of CPI estimation are grouped into 70 expenditure classes.

These indexes likely contributed to a larger measure of price change for the lowest income quartile for the housing and transportation major groups as well as the all items level.

Item categories in the upper left or lower right quadrants of the graph likely contributed to a smaller measure of price change for the lowest income quartile. For example, the lowest income quartile population spent less on owner's equivalent rent and alcohol away from home whose prices rose faster than average over the time period (upper left quadrant) and spent more on medicinal drugs and appliances whose prices fell faster than the average over the time period (lower right quadrant). The distribution of item categories across the four quadrants does not reveal a clear pattern between price change and budget share differences.

The general patterns of price change between populations using a Laspeyres index also hold true for the Tornqvist indexes. The lowest income quartile index displays the highest rate of inflation, and the highest income quartile displays the lowest. Tornqvist indexes for the E and W populations were also calculated, however note these are research indexes as opposed to the Laspeyres indexes for the E and W populations which are produced by the BLS production systems. Graph 3 shows the Tornqvist indexes and Table 7 shows the annualized percent changes at the all items level.[22]

[22] Results at the major group level were presented for Laspeyres indexes to explain the differences at the all items level. This is possible because the Laspeyres index formula is consistent in aggregation, meaning the weighted sum of the major group level is equal to a direct calculation of the all items level. The Tornqvist index formula is not consistent in aggregation, therefore a presentation of major group level indexes would not necessarily explain differences at the all items level.

Graph 3: Monthly Tornqvist indexes for lowest and highest income quartiles: December 2003 to December 2018

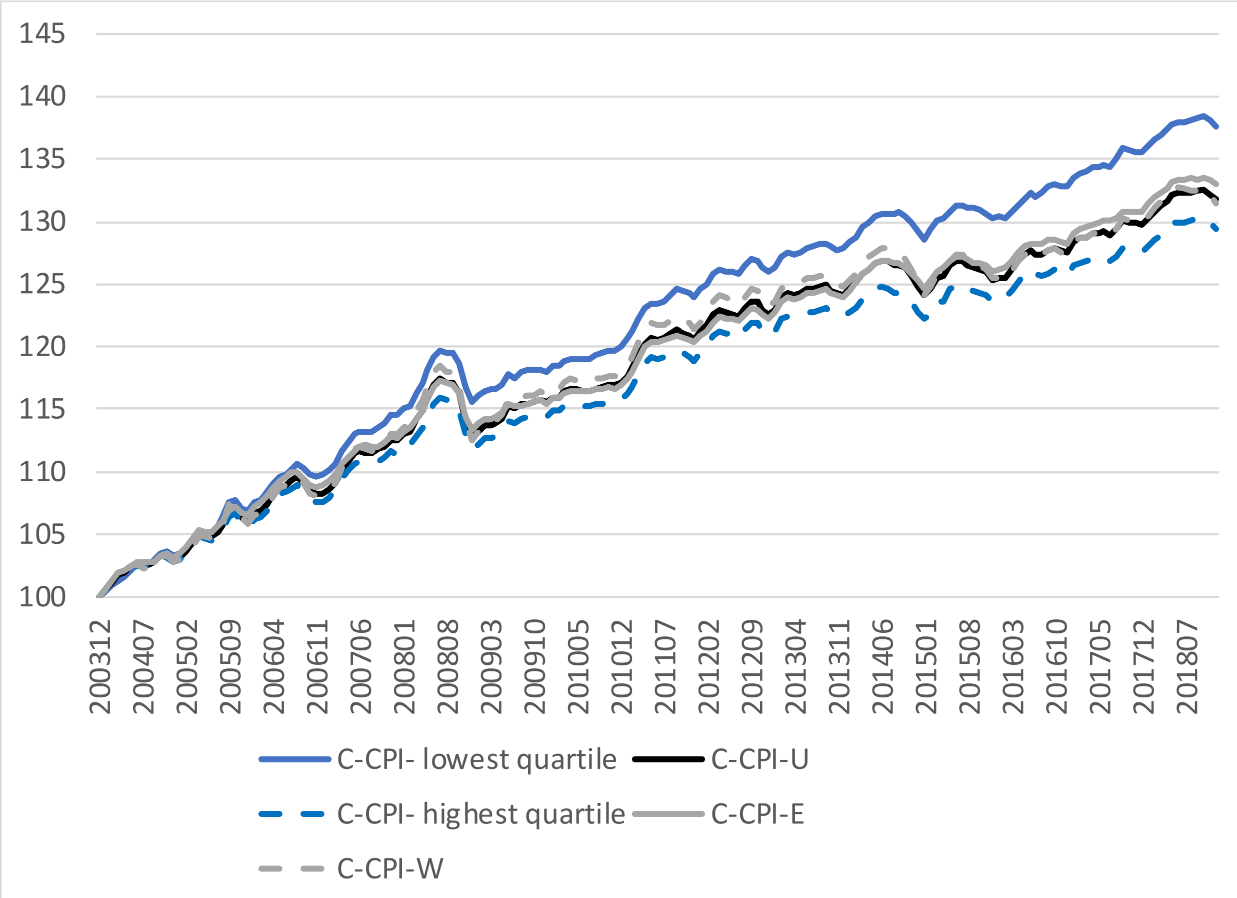


Table 7: Tornqvist index annualized percent changes: December 2003 – December 2018

	All urban households (U)	62 years or older (E)	Wage earner (W)	Lowest income quartile (I1)	Highest income quartile (I4)
Index value December 2018 (December 2003 = 100)	131.7	133.0	131.5	137.6	130.0
Annualized percent change	1.84%	1.91%	1.83%	2.14%	1.76%

For each population, the Tornqvist formula generally displays a lower measure of price change than the Laspeyres index. The graphs in Appendix 1 and Table 8 show the difference in substitution bias between the populations, defined as the difference in annual inflation rates measured by the Tornqvist and Laspeyres indexes. The graphs in Appendix 1 shows the difference in the annual rate of change each month for each population, and Table 8 shows the difference in the annualized percent changes over the 16 year period.

Table 8: Difference in Laspeyres and Tornqvist annualized percent change: December 2003 – December 2019

Annualized percent change	All urban households (U)	62 years or older (E)	Wage earner (W)	Lowest income quartile	Highest income quartile
Tornqvist	1.84%	1.91%	1.83%	2.14%	1.76%
Laspeyres	2.07%	2.17%	2.06%	2.25%	1.97%
Substitution bias	0.23%	0.26%	0.23%	0.11%	0.21%

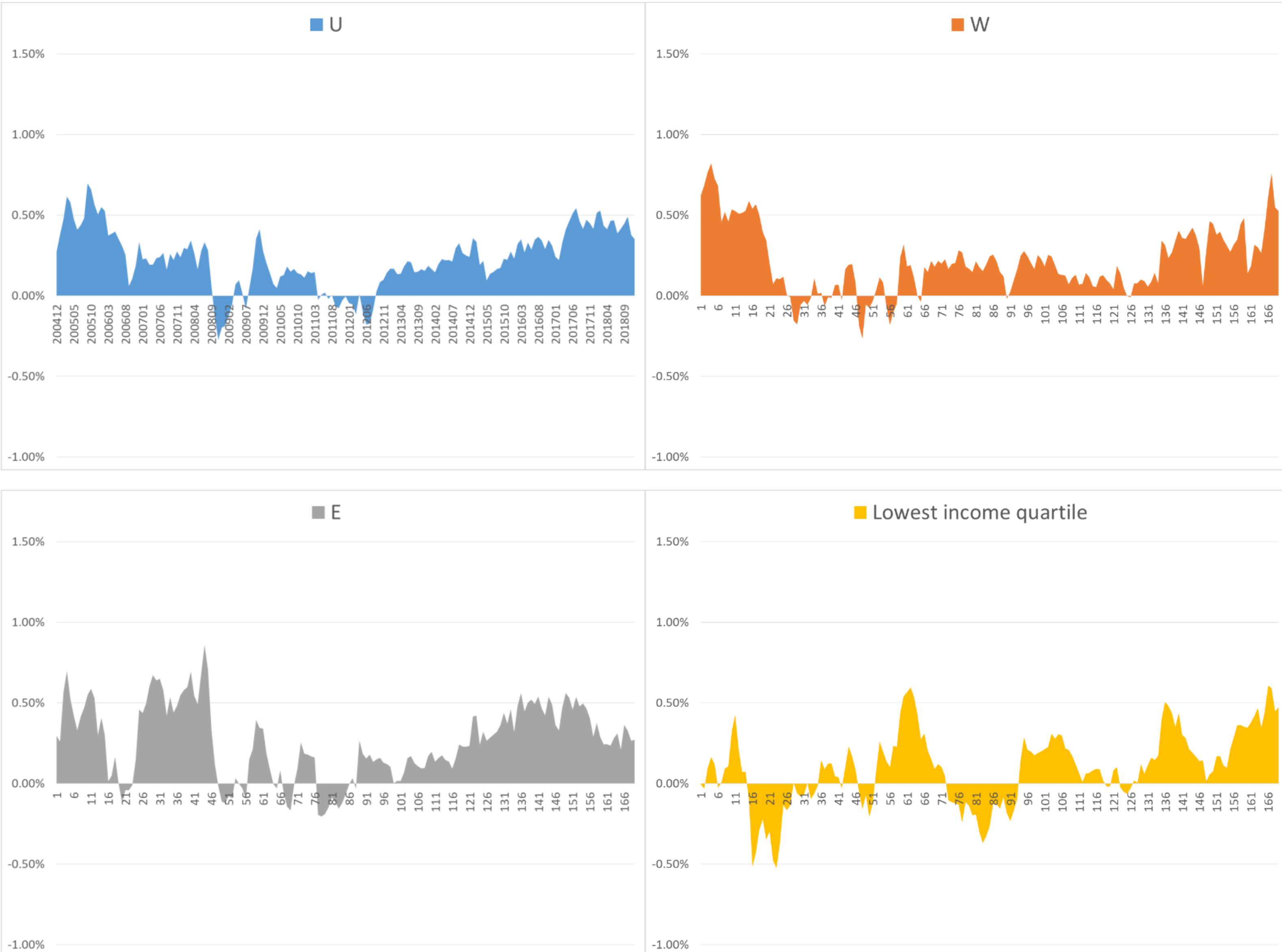
While the Tornqvist index for each population rises more slowly than its Laspeyres counterpart, the difference in the rate of change is smallest for the lowest income quartile population. Indeed, the measure of consumer substitution bias over the 2003 to 2018 time period for the lowest income quartile population is less than half that of all urban consumers. The highest income quartile population had a similar consumer substitution effect as all urban consumers. Future research should explore the extent of consumer substitution (and the elasticity of substitution) across populations.

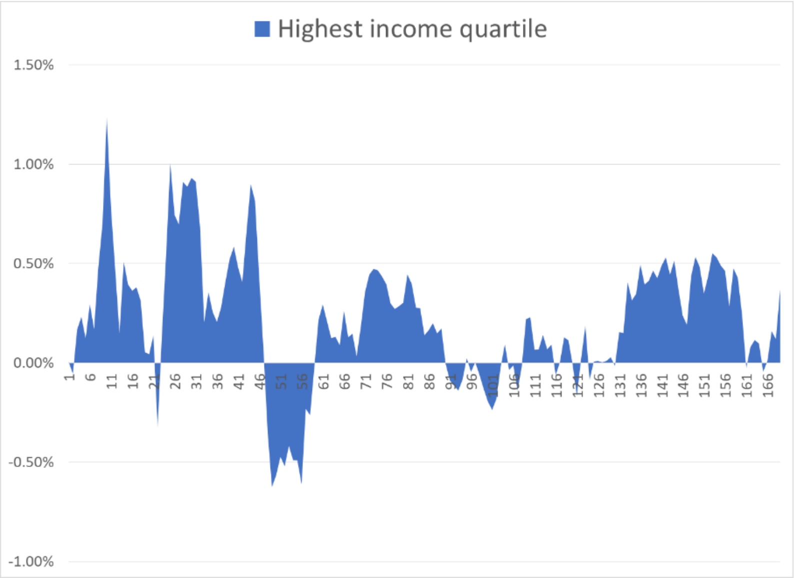
Summary and conclusion

In this paper we present results of estimating CPI indexes for the lowest and highest income quartiles of CE respondents. From 2003 to 2018, the Laspeyres index for the lowest income quartile population rose faster than the index for all urban consumers. The Laspeyres index for the highest income quartile population rose slower than the index for all urban consumers. Chained CPI indexes for the income quartile populations rose slower than their Laspeyres counterparts. The measure of consumer substitution was lowest for the lowest income quartile population; the difference between the Laspeyres and Tornqvist index for the lowest income quartile population was less than half the difference for all urban consumers.

We present these results with many caveats. Future research can improve upon the work of this paper by redefining the income groups, either by using an equivalence scale to adjust for varying household sizes, using externally defined income bands that are more representative of the population, or defining income quartiles at the index area level (as opposed to nationally). Other improvements include using a more sophisticated imputation methodology for missing expenditure weights (ideally sensitive to population spending patterns) and calculation of variances to enable a statistical comparison of index results. Additionally, the lowest income quartile population exhibits the largest number of missing expenditure weights and the lowest measure of consumer substitution bias. Further research is needed to understand the spending patterns of the lower income quartile subpopulation, which appear to be unique from the W and E subpopulations previously defined.

Appendix 1: Difference in Laspeyres and Tornqvist annual percent change: December 2003 – December 2019




Highest income quartile
1.50%
1.00%
0.50%
0.00%
-0.50%
-1.00%
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
96
101
106
111
116
121
126
131
136
141
146
151
156
161
166