Estimating Inflation Inequality

Jackson Mejia* Jonathan S. Hartley†

January 27, 2022

1 Introduction

This document constitutes the technical appendix for two distinct projects. First, it describes the data collection and analysis procedure for the Foundation for Research on Equal Opportunity's (FREOPP) Inflation Inequality Index (III). This index, the first of its kind published on a monthly basis, implements BLS methods to give updates of month-over-month and 12-month inflation for different income groups in the United States. In the future, the index will also show inflation inequality for different racial groups and different categories of income (e.g., pre-tax vs. post-tax income). Second, this appendix documents our methodology for collecting inflation inequality data to be used as an input into a work in progress, "Distributional Consequences of Monetary Policy."

A number of papers have already published estimates of inflation inequality. Our methodology largely follows the careful approach Cravino et al. (2020); hence our appendix will look quite similar to theirs. There are two reasons why we choose to write an appendix when a similar one already exists. First, we hope to provide a more granular level of detail, primarily because we hope that future researchers will choose to use our data and code rather than trudge through messy CEX and CPI data to recreate our result. Second, our methodology does differ in important respects. For example, our expenditure categories are somewhat more aggregated, which leads to critical differences in the concordance between the pre-1998 and post-1998 CPI data.

2 Source Data and Background

In order to construct measures of inflation for different income groups, we require expenditure and price data. Every year, the Bureau of Labor Statistics (BLS) conducts the Consumer Expenditure Survey (CEX) as a means of tracking changes in spending patterns among the general population and subgroups within the American populace, as well as providing data on income and other demographic characteristics at the individual level. Note that the CEX is the primary source data for expenditure weights used for the U.S. Consumer Price

*Massachusetts Institute of Technology, Department of Economics. Email: jpmejia@mit.edu.
†Stanford University, Department of Economics. Email: hartleyj@stanford.edu.

Index. The CEX collects data through two surveys: the Interview Survey and the Diary Survey. The former consists of a quarterly questionnaire conducted across five quarters and accounts for approximately 95% of total expenditures, while the latter records weekly spending and is intended to track smaller expenditures on, for example, groceries. In both surveys, expenditures are put into approximately 600 different expenditure categories called Universal Classification Codes (UCCs).

To accommodate technological and personal preference shifts, the CEX has been altered and improved many times since its inception (original source data are available starting in 1980). For example, it is possible to integrate both surveys for years going back to 1996 and roughly replicate expenditure shares published by BLS. The Interview Survey has about 350 UCCs, while the Diary Survey contains approximately 250. Additionally, starting in 2004, BLS began to impute pretax and after-tax income at the individual level rather than rely on complete reporting, which significantly distorts data prior to 2004.[1]. Hence it is possible to extract reliable annual expenditure and income estimates at the individual level from CEX starting in 2004. Unfortunately, the households in the Diary Survey are not the same as those in the Interview Survey and they are not the same over time either, so it is necessary to use a household matching procedure to put the surveys together.

Along with published estimates of aggregate price inflation, the BLS releases monthly, semi-annual, and annual estimates of price level changes at different levels of aggregation. Since 1998, BLS has formally embedded each series within a hierarchy, from lowest to highest, of Entry Level Items (ELIs), Item Strata, Expenditure Classes, and Major Groups. There are 303 ELIs, 211 item strata, 70 expenditure classes, and eight major groups. For example, the entry level item 450110 - New Cars falls under the item strata TA01 - New Vehicles, which in turn falls under the expenditure class TA - New and Used Motor Vehicles, which falls under the major group Transportation.[2] However, the BLS does not publish price level estimates for entry level items; the lowest level of aggregation is the item stratum. Prior to 1998, the CPI was not as hierarchical and the relationship between the CPI before and after 1998 is neither injective nor surjective. We detail how we deal with this below.

The BLS constructs the aggregate CPI by linking expenditure shares from the CEX with price level data. There are several concordances published by the BLS between UCCs in the CEX and ELIs in the CPI.[3] We document below how exactly we use these and the changes we made to account for idiosyncratic categories like healthcare, which rely on private data we do not have access to.

[1] See the BLS document "Description of Income Imputation Beginning With 2004 Data" for details on the imputation procedure

[2] See the BLS Document "CPI Requirements for CE" for a complete listing of the hierarchy circa 2010. For a more complete description of the changes implemented in the 1998 revision of the CPI, see the BLS document "Changing the item structure of the Consumer Price Index ."

[3] See the CPI Methods Overview Page, in particular the documents "CPI Requirements for CE" and Appendix 5: UCC to ELI concordance.

Outline of Steps

Broadly, what we do is:

1. Get average expenditures on each UCC for each quantile.
2. Link UCC expenditures to ELIs and thence item strata and/or expenditure classes.
    - Because certain item strata and expenditure classes sometimes do not exist but the UCCs do, we manually decide whether to reallocate expenditures to a different item strata or expenditure class or omit entirely.
3. Sum expenditures by item strata and/or expenditure class per quantile per year.
4. Following the BLS expenditure share guide, we compute expenditure shares by combining two years of data and then computing shares for each expenditure class and/or item stratum per quantile.
5. Link expenditure shares to CPI item categories.
6. Compute inflation

3 Expenditure Shares

3.1 Procedure

We utilize three microdata file types per year of expenditures. For the Interview survey, we take the FMLI files, which contain detailed annual information about consumer unit characteristics, the MTBI files, which contain each expenditure reported by a consumer unit, and the ITBI file, which gives imputed income information for each consumer unit. Each consumer unit receives a unique identifying variable called NEWID. The corresponding files for the Diary survey are the FMLD, EXPD, and DTBD. Each year contains five files of each type for the Interview survey, each corresponding to a quarter. CEX gives five rather than four files because surveys are staggered such that a consumer may report for the previous three months going into a different year. Our procedure for handling this is detailed below. However, there are only four files per year for the Diary survey. Finally, we take advantage of the Integrated Hierarchy file published by CEX annually, which details which UCCs should be pulled from each survey to produce an aggregate measure (this matters because there is overlap between surveys).

The general procedure for aggregating expenditures for each year is as follows. First, we select the income variable we are interestd in grouping by. Typically, this is pretax income because it tends to be the most complete, though other types of income are also feasible to use. Then, using the Interview Survey income file, we construct the empirical CDF for the income distribution for each year. Following the CEX handbook of methods, income for a consumer unit is defined as the mean of the five income imputations that CEX conducts. Next, we assign each consumer unit to a particular quantile depending on where their income falls on the empirical CDF. Quantile choice is dependent on the granularity of income distribution

we are interested in; this ranges anywhere from 100 percentiles to a completely aggregated measure. Following this, we filter all UCCs in the interview expenditure file which meet the following conditions:

- Correspond to the interview survey in the hierarchical grouping file;
- Categorized as "FOOD" or "EXPEND" in the hierarchical grouping file;
- Are in the UCC-ELI-Item Strata-Expenditure Class hierarchy in either the document "CPI Requirements for CE" or Appendix 5: UCC to ELI concordance.

We are primarily interested in constructing weighted calendar year estimates. Consequently we construct a new variable called POPWT which, following the BLS Manual of Methods, is a variable that adjusts expenditures made by each consumer unit based on their population weight and when the expenditures took place. There are three critical values for making this adjustment: QINTRVMO, which gives the month the expenditure took place, QINTRVYR, which gives the year the expenditure occurred, and FINLWT21, the weight variable for each consumer unit. The following is the formula for the variable POPWT for each consumer unit i:

$$POPWT_i = MOSCOPE_{i,t} \times \frac{FINLWT21_i}{4}, \tag{1}$$

where MOSCOPE, according the BLS Handbook of Methods, is defined as

$$MOSCOPE_{i,t} = \begin{cases} \frac{QINTRVMO-1}{3} \text{ if } QINTRVMO < 4 \text{ and } Quarter \in \{1,2,3,4\} \\ 3 \text{ if } QINTRVMO \geq 4 \text{ and } QINTRVMO \leq 12 \\ \frac{4-QINTRVMO}{3} \text{ if } Quarter = 5 \end{cases}$$

If QINTVRVYR equals the survey year, then Quarter will take a value in $\{1,2,3,4\}$. If QINTVRYR is one greater than the survey year, then it means that some of the expenditures are taking place in the following year, so it must come from the fifth quarter of the expenditure file.

After we calculate the POPWT variable, we are then able to calculate aggregate expenditures by UCC and household income percentile. We do the following:

- First, we add up all expenditures for each consumer unit on particular expenditure categories (UCCs). Call this variable COST.
- Recall that households have been sorted into quantiles. Consequently, we are able to summarize average expenditures on each UCC by each quantile. Then for all households (NEWID) i in quantile q, we find average expenditures on each UCC j by obtaining weighted expenditures on a particular UCC j for each household, summing over all households in the quantile, then dividing by the sum of the weight of all households in the quantile:

$$X_{q,j} = \frac{\sum_{NEWID \in q} COST_{i,j} \times FINLWT21_i}{\sum_{NEWID \in q} POPWT_i}$$

This process is repeated for the Diary survey with three key exceptions. First, because observations are weekly rather than quarterly, we multiply each expenditure by 13. Second, each Diary survey year has only four files, so there is no need to be concerned about collecting expenditures out of scope. Finally, we take the empirical CDF defined for Interview Survey income and use those quantiles to categorize household income in the Diary Survey. That is, if the cutoff is $10,000 in annual income for the first decile in the Interview Survey, we also use that cutoff for the first decile for the Diary survey. After that, we append the Diary survey quantile expenditure summaries to the Interview results.

We use imputed income measures (in most cases, pretax income), which is only available since 2004. Prior to that, BLS relied on complete income reporters. We are confident that, following Cravino et al. (2020), computing disaggregated inflation prior to 2004 would result in unusably biased results, so we only compute expenditure shares going back to 2004.

3.1.1 Adjustments

Our procedure for most types of expenditures is quite simple, but it is necessary to make adjustments for certain categories. All adjustments proceed prior to computing average expenditures by UCC and quantile. All adjustments come from "CPI Requirements for CE" and Appendix 5: UCC to ELI concordance. Broadly, there are two types of adjustments. The first occurs when there are multiple UCCs allocated to different ELIs. In these cases, we follow the instructions of the BLS in the document "CPI Requirements for CE". If instructions do not exist, then we allocate equally across different ELIs. For example, the UCC 360350 - "Mens Swimsuits/Warm-Up/Ski Suits" is allocated to four different ELIs (AA013, AA022, AA033, and AA041). In this case, we give each ELI 25% of the expenditure to avoid double-counting. The second type of adjustment involves scaling down an expenditure's cost, usually because it involves an investment component.

3.1.2 Housing

We make two adjustments to housing. The first, following the BLS document "How the CPI measures price change of Owners' equivalent rent of primary residence (OER) and Rent of primary residence (Rent)", is an estimate of owner-equivalent rent (OER), which typically makes up approximately 23% of total expenditures. The UCC categories under housing tend to lump together a number of ongoing costs, such as property taxes and mortgage interest payments, which the CPI does not consider to be consumer goods. According to BLS, a house is properly considered a capital good but some portion of housing (rent and its equivalent) is considered a consumer good. Hence, following "CPI Requirements for CE," we instead consult the FMLI file, which asks the following question: "If someone were to rent your home today, how much do you think it would rent for monthly, unfurnished and without utilities?" We take the answer from the FMLI file using the variable RENTEQVX, multiply it by three to get a quarterly value (rather than monthly), and create an artificial UCC that we put into the UCC-ELI hierarchy under HC011.

Second, the BLS makes an adjustment for expenditures on household maintenance, insurance, and major appliances to separate out the consumption component from the investment

component. BLS assumes that, on average, 43% of these expenditures are consumption. According to Cravino et al. (2020), this is based on the likelihood that a renter in the same situation would make those expenditures. The relevant UCCs are listed in the document "CPI Requirements for CE". Each expenditure corresponding to one of those UCCs is multiplied by 0.43.

3.1.3 Medical Care

In the UCC-ELI concordance, expenditures on insurance are counted multiple times under different categories because insurance is fungible across categories (for example, Medicare expenditures can be used on many different types of medical goods). It is therefore required to allocate these expenditures to different categories such that the total cost is only counted once. BLS accomplishes this by redistributing weights from private health insurance and Medicare premia using the National Health Expenditure tables, which are calculated according to a formula which we do not have access to. Following Cravino et al. (2020), we approximate the BLS factors using *Table 20 Private Health Insurance Benefits and Net Cost; Levels, Annual Percent Change and Percent Distribution, Selected Calendar Years 1960-2015*.

We use factors from the table referenced above to redistribute insurance expenditures to different medical categories. Expenditures are mapped to existing UCCs or to a new artificial UCC that underlies an existing item stratum. The exact mapping for doing this is available in the file CEX_CPI_CONCORDANCE.XLSX and in the functions code file cexis_functions.R under the function average_ann_expenditures.

3.1.4 Transportation

Two adjustments are made to UCCs in the transportation category. First, although CPI has different ELIs for regular (*TB011*), midgrade (*TB012*), and premium gas (*TB013*), we are unable to distinguish these at the UCC level. BLS overcomes this issue by distributing expenditures according to factors which we do not have access to. Hence, following Cravino et al. (2020), we simply assume that expenditures on each category are equivalent. This is likely a poor assumption but it ultimately matters very little because the item strata for all three is simply *TB01* anyway. Second, following Cravino et al. (2020), we adjust expenditures on used cars and trucks (UCCs 450116, 450216, 460110, 460901, 860100, and 860200) such that they only reflect dealer value-added. This amounts to multiplying expenditures on each of these UCCs by 0.5.

3.2 Other Adjustments

Several UCCs are adjusted because they appear multiple times in other categories. For these expenditures, we do not have a good rationale for distributing factors in any way other than equally. The UCCs 340906 and 550310 are repeated twice, so we divide each by 2. Similarly, the UCCs 360350, 370904, 380340, and 390230 are repeated four times, so we divide expenditures on each UCC by 4.

3.3 Household Aggregation

The next two steps are interrelated: aggregate UCCs into item strata or expenditure classes and then calculate expenditure shares. The task is not straightforward for three reasons. First, the CPI underwent a major revision in 1998, introducing the ELI-item strata-expenditure class hierarchy. Consequently, it is necessary to construct a concordance between pre-1998 CPI series and the post-1998 revision such that there is some continuity. We do this in two ways. First, we rely on the BLS document "1998 Item Concordance". However, this concordance is incomplete, so when we are unable to match a series, we rely on the concordance created by Nakamura et al. (2018). The latter concordance matches at the ELI level rather than item strata or expenditure classes, which the former relies on. It is frequently the case that either multiple ELIs get matched to the same series or multiple series get matched to the same ELI. The first case is not a problem. To address the second case, we allocate expenditures to each series according to expenditure weights from Nakamura et al. (2018).

Second, certain series are not introduced until the middle of the period. We deal with this in three different ways. If a related item strata or expenditure class has an available CPI item series, then we reallocate expenditures to that or multiple item strata according to what the particulars of the situation call for. Second, if no related series is available, then we exclude the expenditures entirely. An important example of this is expenditure class *EE* (information technology hardware, and services), which does not have a CPI series until December 1988. But because in our judgment, expenditures from 2004 on something like information technology would overstate their importance prior to the introduction of this series, we omit those expenditures entirely. A related issue comes up for health insurance in the post-1998 revision. Finally, some ELIs are matched to multiple series, some of which do not begin until the middle of the period so that, for example, an ELI might be matched to a series that is present for the duration of the period and to another that does not begin until 1988. In such cases, we use coefficients based on the Nakamura et al. (2018) concordance expenditure weights. These coefficients are stored in the sheet "pre98_concordance" under the names "factor", "factor_split", and "factor_split_2". The relevance of each is determined by how often new series are introduced for particular ELIs. See the sheets "Pre-98 Plan" and "pre98_concordance" in the document "CEX_CPI_CONCORDANCE.xlsx" for details.

Finally, the task is made difficult by the fact that CEX does not report imputed income measures until 2004 and income estimates prior to then are hopelessly biased. Note that to construct a Laspeyres inflation index (which is what the CPI is; see equation 2), it is necessary to have fixed expenditure weights coming from some *reference period*. That is, Laspeyres indices measure changes in the aggregate price level by keeping expenditures fixed and evaluating the change in price for the same basket. The BLS measures price changes for different categories of goods, then constructs an aggregate price level by multiplying changes in the prices of those goods by their relative importance, or expenditure share. The expenditure shares in the past, i.e., in the reference period, form the consumption basket for the consumer in question in the future when price changes are measured. Consequently, following Cravino et al. (2020), we use expenditures from 2004-05 for all years until 2008,

then follow the BLS schedule after that.[4] See Table 1 for details. Note that this implies that we use a Paasche index for years prior to 2006 and a Laspeyres index thereafter.

$$CPI = \frac{\sum_i P_{i,t} Q_{i,0}}{\sum_i P_{i,0} Q_{i,0}} \tag{2}$$

Years	Expenditure Reference Period
1977-2008	2004-05
2009-2010	2006-07
2011-2012	2008-09
2013-2014	2010-11
2015-2016	2012-13
2017-2018	2014-15
2019-2020	2016-17
2021-2022	2018-19

Table 1: Matching expenditure shares with years 1977-2021. Note that the left column means that prices from those years use expenditures from the years in right column to form the aggregate CPI.

After compiling average expenditures per UCC by quantile per year, we proceed to construct expenditure shares. The procedure is substantially different for different years. This is because even though UCCs are basically constant over time, item strata and even expenditure classes are not. Hence there is a constant need to manually reallocate expenditures to different item strata and expenditure classes. Suppose we have already constructed a concordance between UCCs and item strata (this is not straightforward and will be discussed in some detail below). We will proceed as if expenditure shares are developed at the slightly less aggregated item strata level rather than expenditure class. Then the basic procedure is

1. match UCCs with item strata using the BLS concordance (see the spreadsheet "CEX_CPI_CONCORDANCE.xlsx" for details).
2. Sum all expenditures made by a quantile within the reference period. Call this object $COST_q$.
3. Sum all expenditures by quantile and item strata within the reference period. Call this object $COST_q, i$ for quantile q and item strata i.

[4]See "Chapter 17. The BLS Handbook of Methods". According to Cravino et al. (2020), the aggregate series constructed using BLS aggregate expenditure shares and the aggregate series constructed using this methodology are quite similar.

4. Obtain expenditure shares by item strata and quantile by taking $COST_{q,i}/COST_q$ for every quantile q and item strata i. Within each quantile, these should sum to one.

1977-1982

We use expenditures from 2004-05, with the following adjustments. Expenditures falling under item strata *HL02*, *RD02*, *HB02*, *RF03*, and *RF01* are redistributed to all remaining item strata under the same expenditure class according to the proportion of expenditures the other item strata take up within the expenditure class. For example, suppose expenditure class *HL* contains four item strata: *HL01*, *HL02*, *HL03*, and *HL04*. The CPI item-level series that best matches with *HL02* does not exist prior to 1983, so we would take all expenditures from *HL02* and redistribute them to each remaining item strata with the following weights: $HL01/(HL01+HL03+HL04)$, and $HL03/(HL01+HL03+HL04)$, $HL04/(HL01+HL03+HL04)$. In this appendix, the word *redistribute* always refers to this procedure. Next, we remove all expenditures made in the expenditure classes, *EE* (information technology hardware, and services) and *AG* (jewelry and watches). It is not so much that we do not believe people made expenditures on either of these categories; rather, there is not a feasible match for either in this period. Finally, BLS did not collect data on owner-equivalent rent prior to 1983. However, because it is such a large proportion of total expenditures (20-25%) and is closely related to the rent category, we think we are justified in redistributing those expenditures to the item stratum *HA01* (rent of primary residence). All remaining item strata are matched with a CPI series and multiplied by a variable called "factor" as defined in the sheet pre98_concordance in the spreadsheet "CEX_CPI_CONCORDANCE.xlsx". Health insurance expenditures are completely reallocated to all other medical care categories and the expenditure class *MG* expenditures are reallocated to the item strata *MF01* and *MF02* expenditures are redistributed.

1982-1983

Adjustments are the same as 1977-1982, except owner-equivalent rent is no longer allocated to *HA01*, *HB02* is not redistributed among *HB* item strata. Health insurance expenditures are completely reallocated to all other medical care categories and the expenditure class *MG* expenditures are reallocated to the item strata *MF01* and *MF02* expenditures are redistributed.

1984-1986

Adjustments are the same as 1982-1983, except for series multiplication by factor. For this period the item strata *RE01*, *RA03*, *RA04*, *HM01*, *GE*, *HP02*, *HP03*, *HK01*, *RE02*, *TA01*, and *TD02* are multiplied by the variable "factor" while all other item strata are multiplied by "factor_split". Health insurance expenditures are completely reallocated to all other medical care categories and the expenditure class *MG* expenditures are reallocated to the item strata *MF01* and *MF02* expenditures are redistributed.

1987-1988

Adjustments are the same as 1984-1986 except only *HL02* is redistributed. Additionally, the item strata *RE01*, *RA03*, and *RA04* are multiplied by the variable "factor", while all other item strata are multiplied by the variable "factor_split_2". Finally, item strata in the expenditure class *AA* can be matched with a CPI series, so those expenditures are counted in total expenditure and hence begin to have an expenditure weight. Health insurance expenditures are completely reallocated to all other medical care categories and the expenditure class *MG* expenditures are reallocated to the item strata *MF01* and *MF02* expenditures are redistributed.

1989-1990

The only adjustments made in this period are that *HL02* is redistributed. All series are multiplied by the variable "factor_split_2" and *EE* expenditures begin to count in total expenditures. Health insurance expenditures are completely reallocated to all other medical care categories and the expenditure class *MG* expenditures are reallocated to the item strata *MF01* and *MF02* expenditures are redistributed.

1991-1997

No adjustments are made except that health insurance expenditures are completely reallocated to all other medical care categories and the expenditure class *MG* expenditures are reallocated to the item strata *MF01* and *MF02* expenditures are redistributed.

1998-2001

Since it is the post-1998 period, UCCs are concorded according to the ELI-Item Strata-Expenditure Class hierarchy using a different concordance (see above for details). Health insurance expenditures are completely reallocated to all other medical care categories and the expenditure class *MG* expenditures are reallocated to the item strata *MF01* and *MF02* expenditures are redistributed. *ED04* expenditures are reallocated to *ED03*, *MD03* expenditures are reallocated to *MD02*, and *TA03* expenditures are redistributed.

2002-2005

Same as 1998-2001 except *TA03* is not redistributed.

2006-2008

Same as 2002-2005 except *ME* expenditures are not completely reallocated; a CPI series exists for health insurance starting December 2005.

2009

Exactly the same adjustments as 2006-2008 except this is now done using 2006-07 expenditures instead of 2004-05.

2010-January 2021

No adjustments.

January 2021-April 2021

The series corresponding to *HP04* is no longer available so we redistribute these expenditures.

April 2021-Present

The series corresponding to *TA03* is no longer available so we redistribute *TA03.* Additionally, the series corresponding to *HP04* is no longer available so we redistribute these expenditures.

4 Household-specific CPIs

In this section, we describe our procedure for selecting specific item-level CPI series. BLS releases consumer price indices each month for series starting at the item strata level all the way up to the aggregate CPI-U. Price index identifiers follow a common naming convention. The first two letters determine whether the series is from the current survey or a prior revision (CU for the former, MU for the later). The next two letter indicates whether the series is seasonally adjusted (U for unadjusted, S for adjusted). The fourth letter gives the periodicity (R indicates monthly, S semiannual). The next four numbers determine which area is being sampled. We exclusively use national price indices. All characters after those numbers indicate the item code. Thus, for example, CUSR0000SEAA is the series ID for a seasonally adjusted, national price index from the current survey for men's apparel.

Across all periods, we only use national price indices at a monthly frequency (sampling all urban consumers). The post-1998 revision is the current survey, while all pre-1998 series are the prior revision, so they begin with "MU". When available, we use seasonally adjusted series. Frequently, these are not available for the duration of the sample, so we create "hybrid" series that typically begin with an unadjusted series and become an adjusted series when available. These series are indicated in our code and data with the letter "H" in place of "S" or "U" as the third letter in the series ID.

Occasionally, a more complicated issue arises when expenditures are reallocated from one series to another, a phenomenon described in detail in the previous section. Take owner-equivalent rent as an example. From 1977-1982, it was allocated to the price index associated with rent. But when the owner-equivalent rent price index begins to be published in 1983, it starts at a price level of 100, while the rental series is at a number well above that (around 146). The practical consequence of this is that, because OER is such a large share of total expenditures, it looks like a massive deflation has taken place when really it is simply expenditures going to a new series. To deal with this, we take two steps. First, we compute the percentage monthly percentage change in the new series, $\pi_{i,t}$. Then we take the last monthly price index value of the series to which expenditures were previously allocated and

set that as the base for the new series so that instead of OER starting at 100, it would start at 146. We then iteratively compute what the price level is in the following months using the monthly inflation rates for the new series. For example, suppose the OER series begins in December 1982 at 100 and the rent series in November 1982 is 146. Using the OER series, we compute a new inflation variable and set the December 1982 price index for OER at 146. To determine what the price level is for the OER series in January 1983, we multiply 146 by one plus the inflation rate for the original OER series between December 1982 and January 1983, and so on. We follow this procedure every time expenditures are reallocated.

We then match expenditure shares with these series as described in the previous section and are therefore ready to compute aggregate inflation.

4.1 Aggregation Formula

"Chapter 17. The BLS Handbook of Methods" provides an aggregation formula for the CPI. Following this and the lead of Cravino et al. (2020), we construct the income quantile CPI as follows:

$$PIX_t^h = PIX_v^h \cdot \sum_{j \in J} \left(\omega_{j,\beta}^h \times \frac{P_{j,t}}{P_{j,v}} \right)$$

where

- PIX_t^h is the consumer price index for household at quantile h at time t
- v is the pivot year and month prior to the month when expenditure wights from reference period β are first used in the CPI. Note that the way we construct the weights means that these will not always start at 100, in contrast to some methodologies.
- β is the predetermined expenditure reference period
- $P_{j,t}$ price of item j at time t
- $\omega_{j,\beta}^h$ is the expenditure weight of household at quantile h on time j during the predetermined expenditure reference period β

As a robustness check, we plot the aggregate annual inflation rates from 1979-2020 using our measure versus the official statistic published by the BLS in figure 1:

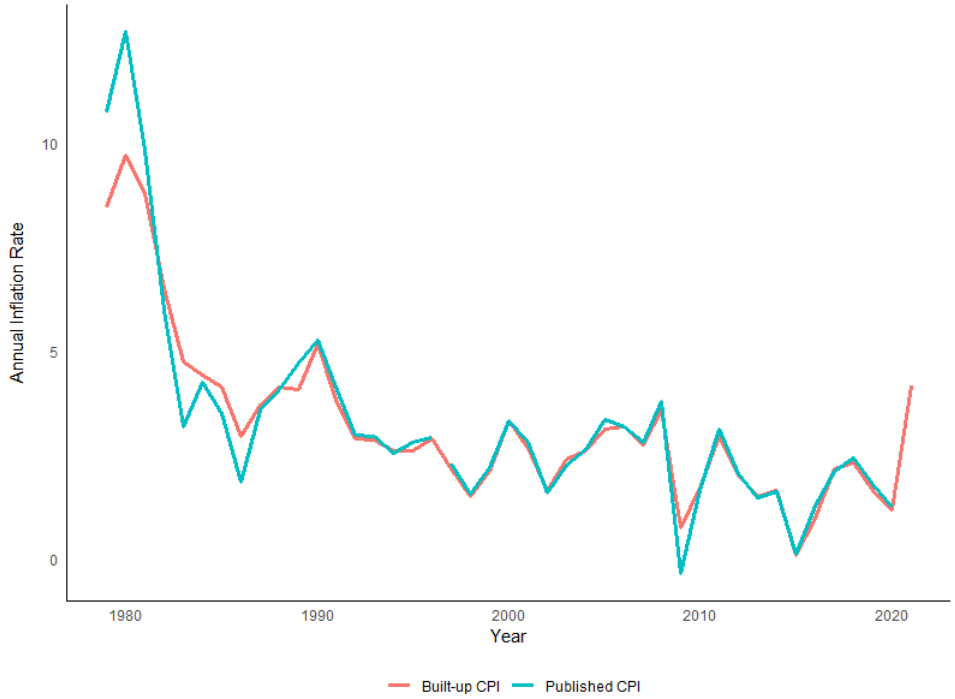

Figure 1: Annual Inflation 1979-2020. The orange line is our measure while the blue line is the measure published by BLS.

5 Cross-Survey Matching

Because our ultimate goal is to build a dataset linking microdata across surveys connecting quantile-level unemployment with inflation expectations and inflation, we match our inflation data described above from CEX with unemployment from the Current Population Survey (CPS) and inflation expectations from the University of Michigan's Survey of Consumers. We describe the matching procedure below.

5.1 Current Population Survey (CPS)

The Current Population Survey (CPS) consists of a set of surveys at different frequencies designed to provide basic information about individual and household finances, with about 60,000 households participating. We use two of these surveys: the Basic Monthly Survey (BMS) and the Annual Social and Economic Supplement (ASEC). The Basic Monthly Survey, conducted on a monthly basis, surveys households across multiple periods to track changes in labor force statistics and financial outcomes. ASEC is conducted once per year in March, supplementing the BMS with more extensive demographic data about each household, including household income. Relative to BMS, ASEC oversamples the population but can replace the BMS for March of the relevant survey year.

Hence, to get household income and unemployment statistics, we took the following procedure. After downloading CPS BMS and ASEC surveys from 1976-2019 from IPUMS, we

filtered out all oversampled households in ASEC, i.e., those households for which monthly unemployment data do not exist and merged ASEC with BMS. This is possible to do using the variable CPSIDP; all CPSIDP entries with value zero are not in the BMS. In our merged sample, we could then observe household income using the variable HHINCOME and compute quantile-level monthly unemployment rates using the variables EMPSTAT and WTFINL, where the former gives the employment status of each person in the survey and the latter gives survey weight. In particular, depending on the survey period, we computed the total number of unemployed by observing whether EMPSTAT was coded as 20, 21, or 22 and the employed by whether EMPSTAT took value 10 or 12, then weighting by WTFINL.

The difficulty in this procedure is determining how to set income thresholds and consequently which income quantile to put households income (and thereby loosely match across surveys). Ultimately, we decided that two factors determined the order of importance for which survey would determine income thresholds: importance of the survey and quality of the survey over time. Since it is quite feasible to run different specifications of the New Keynesian Phillips curve without reference to the unemployment rate, but it is not possible to exclude inflation expectations or inflation. Therefore, the CEX and Survey of Consumers data are more important. However, the Survey of Consumers has the lowest quality data early on. Hence, when possible, we use CEX income thresholds as described above. Note that because good data on pretax income is only available after 2004 in the CEX, we end up determining quantile-level thresholds for CPS and the Survey of Consumers for all years prior to 2004.

5.2 Survey of Consumers (SC)

We obtained data on household-level inflation expectations from 1978-2019 using microdata from the University of Michigan's Survey of Consumers (SC). After removing households that do not provide a quantitative estimate of inflation expectations (PX1Q) and do not report household income (INCOME), we weighted each household's year-ahead inflation expectations using the variable WT and sorted into household income quantiles using thresholds set by CEX from 2004-2019 and CPS from 1978-2003. After that, we grouped by income quantile and computed median expected inflation rates for each quantile.[5] The precision of SC's household income data varies substantially over time. Until the mid-1980s, households were grouped in fewer than ten discrete bins, which makes it infeasible to do quintile-level analysis during that time period.

References

Cravino, J., T. Lan, and A. A. Levchenko (2020, 4). Price stickiness along the income distribution and the effects of monetary policy. *Journal of Monetary Economics 110*(1), 19–32.

[5]Following the procedure adopted by Curtin (1996) in publishing SC estimates of inflation expectations, we take the median rather than the mean to avoid skewing the responses.

Curtin, R. T. (1996, 1). Procedure to Estimate Price Expectations. Technical report, University of Michigan.

Nakamura, E., J. Steinsson, P. Sun, and D. Villar (2018, 11). The Elusive Costs of Inflation: Price Dispersion during the U.S. Great Inflation. *The Quarterly Journal of Economics 133*(4).