

Federal Reserve Bank of Minneapolis
Research Department

Inflation at the Household Level*

Greg Kaplan and Sam Schulhofer-Wohl

Working Paper #

February 2016

PRELIMINARY AND INCOMPLETE
DO NOT CITE OR CIRCULATE
WITHOUT PERMISSION

ABSTRACT

We use scanner data to estimate inflation rates at the household level. Households' inflation rates are remarkably heterogeneous, with an interquartile range of 6.2 to 9.0 percentage points on an annual basis. Most of the heterogeneity comes not from variation in broadly defined consumption bundles but from variation in prices paid for the same types of goods — a source of variation that previous research has not measured. The entire distribution of household inflation rates shifts in parallel with aggregate inflation. Deviations from aggregate inflation exhibit only slightly negative serial correlation within each household over time, implying that the difference between a household's price level and the aggregate price level is persistent. The large cross-sectional dispersion and low serial correlation of household-level inflation rates together mean that almost all of the variability in a household's inflation rate over time comes from variability in household-level prices relative to average prices for the same goods, not from variability in the aggregate inflation rate. We provide a characterization of the stochastic process for household inflation that can be used to calibrate models of household decisions.

Keywords: Inflation; Heterogeneity

JEL classification: D12, D30, E31

*Kaplan: Princeton University and National Bureau of Economic Research (gkaplan@princeton.edu). Schulhofer-Wohl: Federal Reserve Bank of Minneapolis (wohl@minneapolisfed.org). The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. Introduction

One major objective of monetary policy in most countries is to control the inflation rate. Policymakers typically measure inflation with an aggregate price index, such as, in the United States, the Bureau of Economic Analysis' price index for personal consumption expenditures (Federal Open Market Committee, 2015). But an aggregate price index is based on an aggregate consumption bundle and on average prices for various goods and services. It does not necessarily correspond to the consumption bundle, prices, or inflation rate experienced by any given household. To know how changes in monetary policy and inflation will impact households' economic choices and well-being, we need to know how inflation behaves at the household level.

This paper takes a first step toward an understanding of heterogeneity in inflation rates by using scanner data on households' purchases to characterize inflation rates at the household level. Inflation rates can vary across households because different households buy different bundles of goods or because different households pay different prices for the same goods. Previous research on inflation heterogeneity has focused exclusively on variation in consumption bundles and assumed that all households pay the average price for each broadly defined category of good. By employing data from the Kilts-Nielsen Consumer Panel Dataset (KNCP), which records the prices, quantities and specific goods purchased in 500 million transactions by around 50,000 U.S. households in 2004 through 2013, we can also consider variation in prices paid and in the mix of goods within broad categories. These new sources of variation are crucial to our results. We find that inflation at the household level is remarkably dispersed, with an interquartile range of 6.2 to 9.0 basis points annually, about five times larger than the amount of variation found in previous work. Almost two-thirds of the variation we measure comes from differences in prices paid for identical goods, and about one-third from differences in the mix of goods within broad categories; only 7 percent of the variation arises from differences in consumption bundles defined by broad categories.

Despite the massive degree of heterogeneity, the entire distribution of household-level inflation shifts in parallel with aggregate inflation, and the central tendency of household-level inflation closely tracks aggregate inflation.¹ Households with low incomes, more household

¹Neither of these results is mechanical. If the rate of increase of the prices that a particular household

members, or older household heads experience higher inflation on average, while those in the Midwest and West experience lower inflation, but these effects are small relative to the variance of the distribution, and observable household characteristics have little power overall to predict household inflation rates.

We also explore the evolution of households’ inflation rates over time. Deviations from mean inflation exhibit only slightly negative serial correlation within each household over time. As a result, inflation rates measured over time periods longer than a year are also very heterogeneous, and the difference between a household’s price level and the aggregate price level is quite persistent. The low serial correlation and high cross-sectional variation together imply that variation in aggregate inflation is almost irrelevant for variation in a household’s inflation rate. In a benchmark calculation, 91 percent of the variance of a household’s inflation rate comes from variation in the particular prices the household faces, and just 9 percent from variation in the aggregate inflation rate.

Previous research on household-level heterogeneity in inflation in the United States, such as Hobijn and Lagakos (2005) and Hobijn et al. (2009), has largely used microdata from the Consumer Expenditure Survey (CEX) to measure household-specific consumption bundles, then constructed household-level inflation rates by applying these household-specific consumption bundles to published indexes of average prices for relatively broad categories of goods, known as item strata. Such an analysis assumes both that all households pay the same price for a given good (for example, a 20-ounce can of Dole pineapple chunks) and that all households purchase the same mix of goods within each item stratum (for example, the same mix of 20-ounce cans of Dole pineapple, other size cans of Dole pineapple, other brands of pineapple, and other fruits within the category “canned fruits”). Relative to this literature, our key innovation is to use the detailed price and barcode data in the KNCP to also measure variation in the prices that households pay and the goods they choose within item strata. We observe the barcode of each good purchased and can thus account both for variation in the price of a specific good and variation across households in the mix of goods purchased within item strata. We thus build on the findings of Kaplan and Menzio (2015),

pays is correlated with the household’s consumption bundle, then the mean of household inflation rates need not match the aggregate inflation rate, and the distribution need not shift along with the aggregate.

who use the KNCP data to characterize variation over time and space in the prices at which particular goods are sold. However, because the KNCP focuses on goods sold in retail outlets — a universe that includes about 30 percent of household consumption (Kilts Center, 2013a) — we are unable, unlike the CEX-based literature, to measure the impact of heterogeneity in consumption of other goods and services. In particular, we cannot measure the impact of differences in spending on education, health care, and gasoline, which Hobijn and Lagakos (2005) found were important sources of inequality in inflation rates. Nonetheless, when we treat the data similarly to previous research by imposing common prices on all households, we measure a similar amount of inflation heterogeneity as in previous papers whose calculations encompassed a broader universe of goods and services.

Our findings are also related to the growing literature on households’ and small firms’ inflation expectations. Binder (2015) and Kumar et al. (forthcoming) find that households in the United States and small firms in New Zealand, respectively, do not have well-anchored inflation expectations and are poorly informed about central bank policies and aggregate inflation dynamics. One possible explanation is that if aggregate inflation is only a minor determinant of household- or firm-level inflation, then households and firms may rationally choose to be inattentive (Reis, 2006; Sims, 2003) to aggregate inflation and policies that determine it. More broadly, the weak link between aggregate inflation and household-level inflation may help explain the well-known long and variable lags in the impact of monetary policy on the economy: If household-level inflation is only loosely related to aggregate inflation, it may be difficult for households to detect changes in aggregate inflation, and hence households may react weakly or at least slowly to aggregate inflation.

The paper proceeds as follows. Section 2 describes the data and how we construct household-level inflation rates. Section 3 characterizes cross-sectional properties of the inflation distribution. Section 4 characterizes time-series properties of household-level inflation. Section 5 concludes.

2. Data and Estimation

This section describes the KNCP data and how we use the data to calculate inflation rates.

A. Data

Our data comes from the Kilts-Nielsen Consumer Panel (KNCP), a panel dataset that tracks the shopping behavior of approximately 50,000 households over the period 2004 to 2013. Households in the panel provide information about each of their shopping trips using a Universal Product Code (UPC), i.e. barcode, scanning device provided by Nielsen. Panelists use the device to enter details about each of their shopping trips, including the date and store where the purchases were made, and then scan the barcode of each purchased good and enter the number of units purchased. The price of the good is recorded in one of two ways, depending on the store where the purchase took place. If the good was purchased at a store that Nielsen covers, the price is set automatically to the average price of the good at the store during the week when the purchase was made. If the good was purchased at a store that Nielsen does not cover, the price is directly entered by the panelist. Panelists are also asked to record whether the good was purchased using one of four types of deals: (i) store feature, (ii) store coupon, (iii) manufacturer coupon, or (iv) other deal. If the deal involved a coupon, the panelist is prompted to input its value.

Households in the KNCP are drawn from 76 geographically dispersed markets, known as Scantrack markets, each of which roughly corresponds to a Metropolitan Statistical Area (MSA). Demographic data on household members are collected at the time of entry into the panel and are then updated annually through a written survey during the fourth quarter of each year. The collected information includes age, education, marital status, employment, type of residence, and race. For further details on the KNCP, see Kaplan and Menzio (2015).²

We use bootstrap standard errors to account approximately for the sampling design of the KNCP. The KNCP sample is stratified across 61 geographic areas, some of which include multiple Scantrack markets. Nielsen replenishes the sample weekly (Kilts Center, 2013b), but even at a quarterly frequency, there are quarters when too few new households join the sample in some geographic strata for us to be able resample from these households. We therefore treat the sample as if it is replenished at an annual frequency. We resample households within

²The KNCP has become an increasingly commonly used dataset for studies of prices and expenditure. Examples of recent studies that have used the KNCP include Einav, Leibtag, and Nevo (2010), Handbury (2013) and Bronnenberg et al. (2015).

groups defined by geographic stratum and the year in which the household first appears in the data the sample. As recommended by Rao, Wu, and Yue (1992), we ensure that the bootstrap is unbiased by resampling $N - 1$ households with replacement from a group containing N households. When we resample a household, we include all quarterly observations from that household in our bootstrap sample. Thus, our bootstrap accounts both for the geographic stratification of the original sample and for serial correlation over time within households, for example due to unique purchasing patterns of particular households. However, because we do not have access to all details of Nielsen’s sampling and weighting procedure, our bootstrap is only approximate. In particular, we do not know whether new households are chosen purely at random or based on observable characteristics. In addition, we cannot recompute the sampling weights in each bootstrap sample because we do not have Nielsen’s formula for computing the weights.

B. Calculating Household Inflation Rates

A price index measures the weighted average rate of change of some set of prices, weighted by some consumption bundle. Aggregate price indexes use the national average mix of consumption to define the consumption bundle. To construct household-level price indexes, we must define household-level consumption bundles and choose a time period over which to measure the change in prices.

We can measure the change between two dates in a household’s price for some good only if the household buys the good on both dates. On any given day, most households do not buy most goods — even goods that they buy relatively frequently. Therefore, although the KNCP data record the date of each purchase, we aggregate each household’s data to a quarterly frequency to increase the number of goods that a household is observed to buy in multiple time periods. If a household buys the same product (defined by barcode) more than once in a quarter, we set the household’s quarterly price for that product to the volume-weighted average of prices that the household paid.

Many prices exhibit marked seasonality. It would be virtually impossible to seasonally adjust the household-level price indexes because we do not observe a long time series for each household. Therefore, we remove seasonality in price changes by constructing price indexes

at an annual frequency, comparing prices paid in quarter t and quarter $t + 4$. Some residual seasonality may remain if consumption bundles change seasonally, but in practice we observe little seasonality in the annual price indexes we compute. To prevent mismeasured prices from distorting our estimates, we exclude a product from the calculation for a particular household at date t if the product's price for that household increases or decreases by a factor of more than three between t and $t + 4$.

Two commonly used price indexes are the Laspeyres index, which weights price changes between two dates by the consumption bundle at the initial date, and the Paasche index, which weights price changes by the consumption bundle at the final date. We compute both of these.

When we calculate a household's inflation rate between quarters t and $t + 4$, we consider only goods (defined by barcodes) that the household bought in both of those quarters. To reduce sampling error, we restrict the sample to households with at least five matched barcodes in the two quarters. (On average across all dates in the sample, 77% of households that make any purchase in quarter t also make some purchase in quarter $t + 4$, and 72% buy at least five matched barcodes whose prices changes by a factor no greater than three.) Thus, let q_{ijt} be the quantity of good j bought by household i in quarter t , and let p_{ijt} be the price paid. For each good j such that $q_{ijt} > 0$ and $q_{ij,t+4} > 0$, we define household i 's consumption share of good j at the initial date as

$$s_{ijt,t+4}^L = \frac{p_{ijt}q_{ijt}}{\sum_{\substack{\ell: q_{i\ell t}, \\ q_{i\ell,t+4} > 0}} p_{i\ell t}q_{i\ell t}}. \quad (1)$$

Similarly, we define household i 's consumption share of good j at the final date as

$$s_{ijt,t+4}^P = \frac{p_{ij,t+4}q_{ij,t+4}}{\sum_{\substack{\ell: q_{i\ell t}, \\ q_{i\ell,t+4} > 0}} p_{i\ell,t+4}q_{i\ell,t+4}}. \quad (2)$$

The Laspeyres and Paasche inflation rates for household i between t and $t + 4$ are then

$$P_{it,+4}^L = \sum_{\substack{j: q_{ijt}, \\ q_{ij,t+4} > 0}} s_{ijt,t+4}^L \frac{p_{ij,t+4}}{p_{ijt}} \quad (3)$$

and

$$P_{it,+4}^P = \sum_{\substack{j: q_{ijt}, \\ q_{ij,t+4} > 0}} s_{ijt,t+4}^P \frac{p_{ij,t+4}}{p_{ijt}}, \quad (4)$$

respectively. We also compute the Fisher index, which is the geometric mean of Laspeyres and Paasche:

$$P_{it,+4}^F = \sqrt{P_{it,+4}^L P_{it,+4}^P}. \quad (5)$$

We compute two sets of household-level inflation indexes with prices defined at a more aggregated level. First, we compute inflation indexes that assign to every household the average price for each barcode. Cross-sectional variation in this index comes only from variation in which barcodes each household buys, not from variation in the price changes for particular barcodes. The Laspeyres index at the household level with barcode-average prices is

$$P_{it,t+4}^{L,BC} = \sum_{\substack{j: q_{ijt}, \\ q_{ij,t+4} > 0}} s_{ijt,t+4}^L \frac{\bar{p}_{j,t+4}}{\bar{p}_{jt}}, \quad (6)$$

where $\bar{p}_{jt}$ is the volume-weighted average price for barcode j in quarter t . The Paasche and Fisher indexes with barcode-average prices, $P_{i,t,+4}^{P,BC}$ and $P_{i,t,+4}^{F,BC}$, are defined analogously. By comparing the indexes with household-level prices and the indexes with barcode-average prices, we can measure how much cross-sectional variation in household inflation rates comes from differences in which barcodes each household buys, and how much from differences in price changes for the same barcodes.

We next compute household-level inflation indexes that, similarly to the previous literature, account only for heterogeneity in broadly defined consumption bundles and not for heterogeneity in prices or in the selection of specific goods within each broad category of goods. The Bureau of Labor Statistics publishes price indexes for the various item strata that make up the CPI. By matching every purchase to the CPI price for the corresponding

item stratum — for example, by matching Dole canned pineapples to the canned fruit index — we can remove variation in prices for specific goods and variation in the mix of goods within item strata, leaving differences in how households spread their consumption across item strata as the only source of heterogeneity. Let $k(j)$ be the CPI category corresponding to barcode j , and let p_{kt}^{CPI} be the CPI sub-index for item stratum k at date t . The Laspeyres index at the household level with CPI prices is

$$P_{it,t+4}^{L,CPI} = \sum_{\substack{j: q_{ijt}, \\ q_{ij,t+4} > 0}} s_{ijt,t+4}^L \frac{p_{k(j),t+4}^{CPI}}{p_{k(j),t}^{CPI}}, \quad (7)$$

and the Paasche and Fisher indexes with CPI prices, $P_{i,t,t+4}^{P,CPI}$ and $P_{i,t,t+4}^{F,CPI}$, are defined analogously.³

Our household-level inflation rates are not directly comparable to published aggregate inflation rates because our data cover only a subset of goods. We construct an aggregate inflation rate that is comparable to our household-level indexes by measuring the aggregate consumption bundle in our data and using this bundle to aggregate the CPI item stratum price indexes. (We cannot make a similar comparison to the Bureau of Economic Analysis' personal consumption expenditure index because that program does not provide prices for detailed types of goods.) We use the Laspeyres index for this comparison because the aggregate CPI is a Laspeyres aggregate of item stratum prices (Bureau of Labor Statistics, 2015). The aggregate expenditure share of good j for this index is the good's share in total spending, counting only spending that is included in our index because it represents a household buying the same good at both dates:

$$s_{jt,t+4}^L = \frac{\sum_{\substack{i: q_{ijt}, \\ q_{ij,t+4} > 0}} p_{ijt} q_{ijt}}{\sum_{\ell} \sum_{\substack{i: q_{i\ell t}, \\ q_{i\ell,t+4} > 0}} p_{i\ell t} q_{i\ell t}}, \quad (8)$$

³We calculate the quarterly value of the CPI as the average of the monthly values for the three months in the quarter.

and our aggregate inflation index is

$$P_{t,t+4}^{L,CPI} = \sum_j s_{jt,t+4}^L \frac{p_{k(j),t+4}^{CPI}}{p_{k(j),t}^{CPI}}. \quad (9)$$

The aggregate index $P_{t,t+4}^{L,CPI}$ is a version of the CPI that is based on the same set of goods as our household-level price indexes. The price data in it are all from the CPI, and the consumption bundle is the aggregate of the bundles used to construct our household-level price indexes.⁴

The aggregate index $P_{t,t+4}^{L,CPI}$ can be rewritten as a weighted average of household-level indexes with CPI prices:

$$P_{t,t+4}^{L,CPI} = \frac{\sum_i x_{it,t+4} P_{it,t+4}^{L,CPI}}{\sum_i x_{it,t+4}}, \quad (10)$$

where the weights $x_{it,t+4}$ are each household i 's expenditure at date t on goods included in the household-level price index with household-level prices:

$$x_{it,t+4} = \sum_{\substack{j: q_{ijt}, \\ q_{ij,t+4} > 0}} p_{ijt} q_{ijt}. \quad (11)$$

Because $P_{t,t+4}^{L,CPI}$ weights households according to their spending, it is what Prais (1959) called a plutocratic price index. The published CPI is likewise a plutocratic index because it defines the consumption bundle based on each good's expenditure share in aggregate spending. One can alternatively construct democratic aggregate indexes that weight households equally, but because our goal here is to compare household-level indexes with an analog to the published CPI, we focus on the plutocratic index.

Figure 1 compares the aggregate index for the KNCP universe, $P^{L,CPI}$, with the published CPI and several CPI sub-indexes. The dates on the horizontal axis correspond to the initial quarter of the one-year period over which the inflation rate is calculated. The aggregate index for the KNCP data behaves similarly to the overall CPI but lags it somewhat,

⁴Another, minor difference between our index and the published CPI is that our index is chain-weighted, with new base-period weights defined at each date t , while the CPI uses a fixed base period.

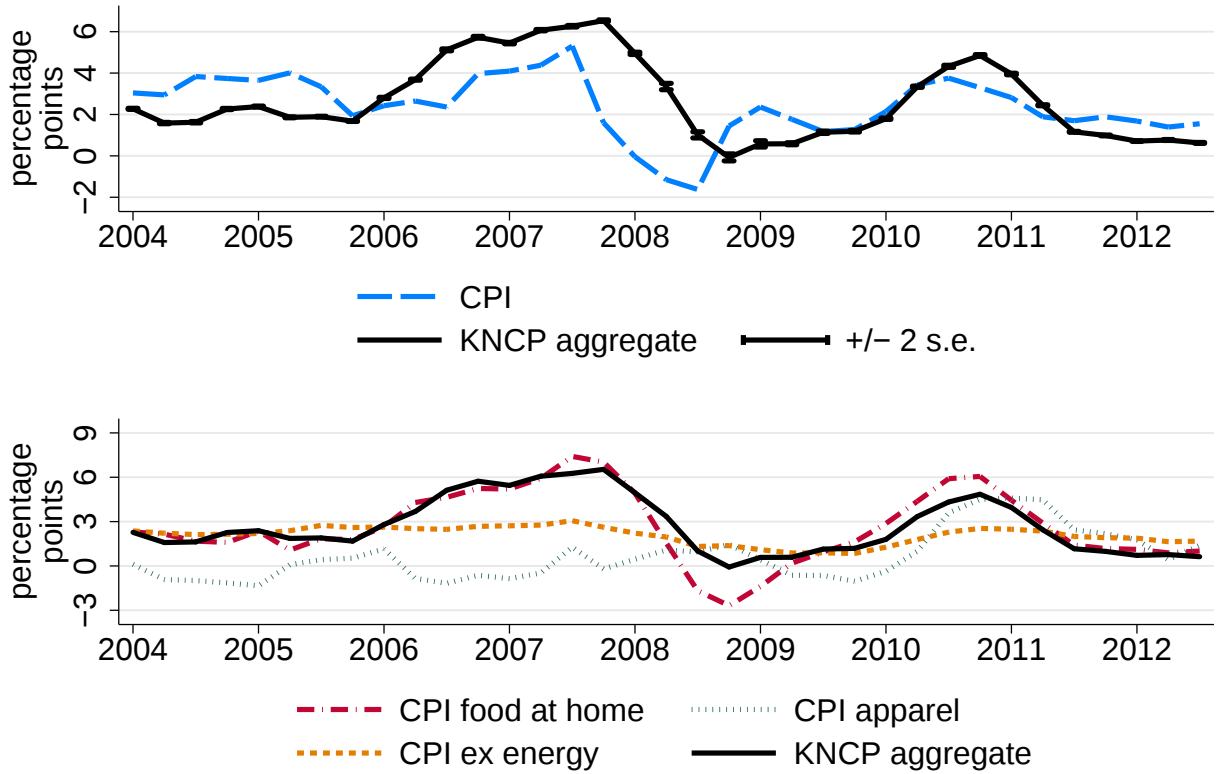


Figure 1: Aggregate inflation rates.

CPI inflation rates computed as annual percent change in quarterly average of monthly index values. Inflation rate plotted for each quarter is the change in price index from that quarter to the quarter one year later. Source for CPI data: Bureau of Labor Statistics.

as do the CPI sub-indexes for food at home and apparel, two categories that make up a disproportionate share of KNCP purchases. Unsurprisingly given the large share of food at home in the KNCP data, our index moves closely with the CPI for food at home, though our index is somewhat less volatile. The overweighting of food offsets the absence of energy in our data, so that our index's volatility is similar to that of the overall CPI and substantially greater than the volatility of the CPI excluding energy. Over the period we study, our aggregate inflation rate averages 2.6 percent with a standard deviation of 1.9 percentage points, compared with a mean of 2.4 percent and standard deviation of 1.5 percentage points for the published CPI, and a mean of 2.6 percent and standard deviation of 2.4 percentage points for the food-at-home subindex. Our index is precisely estimated; the bootstrap standard errors average 2 basis points.

Both when we compute the aggregate consumption bundle and when we compute household price indexes with CPI prices or barcode-level prices, we use only those specific goods — defined by barcodes — that each household purchases at both dates. Thus, in all of our indexes, we measure inflation for the subset of goods that households buy repeatedly. This inflation rate may differ from an inflation rate that includes goods bought less frequently, but we have no way to compute the latter rate at the household level.

3. The Cross-Sectional Distribution of Inflation Rates

Figure 2 shows the distributions of household-level inflation rates for a typical period, the fourth quarter of 2004 to the fourth quarter of 2005, calculated with Laspeyres indexes. In each panel, a kernel density estimate of the distribution of household-level inflation rates with household-level prices is superimposed on an estimate of the distribution of the household-level rates with CPI prices. Household-level inflation rates with household-level prices are remarkably dispersed. The interquartile range of annual inflation rates is 6.7 percentage points, and the difference between the 10th and 90th percentiles is 14.8 percentage points. (The distributions of the Fisher and Paasche indexes are similar and are shown in the web appendix.) For legibility, the plots run from the 1st to 99th percentiles of the distribution of household-level inflation with household-level prices, but a few households have much more extreme inflation rates; the smallest observed inflation rate at the household level with the Laspeyres index in the fourth quarter of 2004 was -43 percent, and the largest was 102 percent.

The vast majority of the cross-sectional variation in household-level price indexes comes from cross-sectional dispersion in prices paid for goods within an item stratum, not variation in the allocation of expenditure to item strata. The interquartile range of the index with CPI prices is just 0.95 percentage point, while the variance of the index with CPI prices is 2.5 percent of the variance of the index with household-level prices.⁵

Differences in the barcodes that households buy within item strata and differences in the prices that households pay for particular barcodes are both important sources of variation

⁵We compute both variances on the subset of observations that fall between the 1st and 99th percentiles of the distribution of the index with household-level prices, to remove some extreme cases that would inflate the variance with household-level prices.

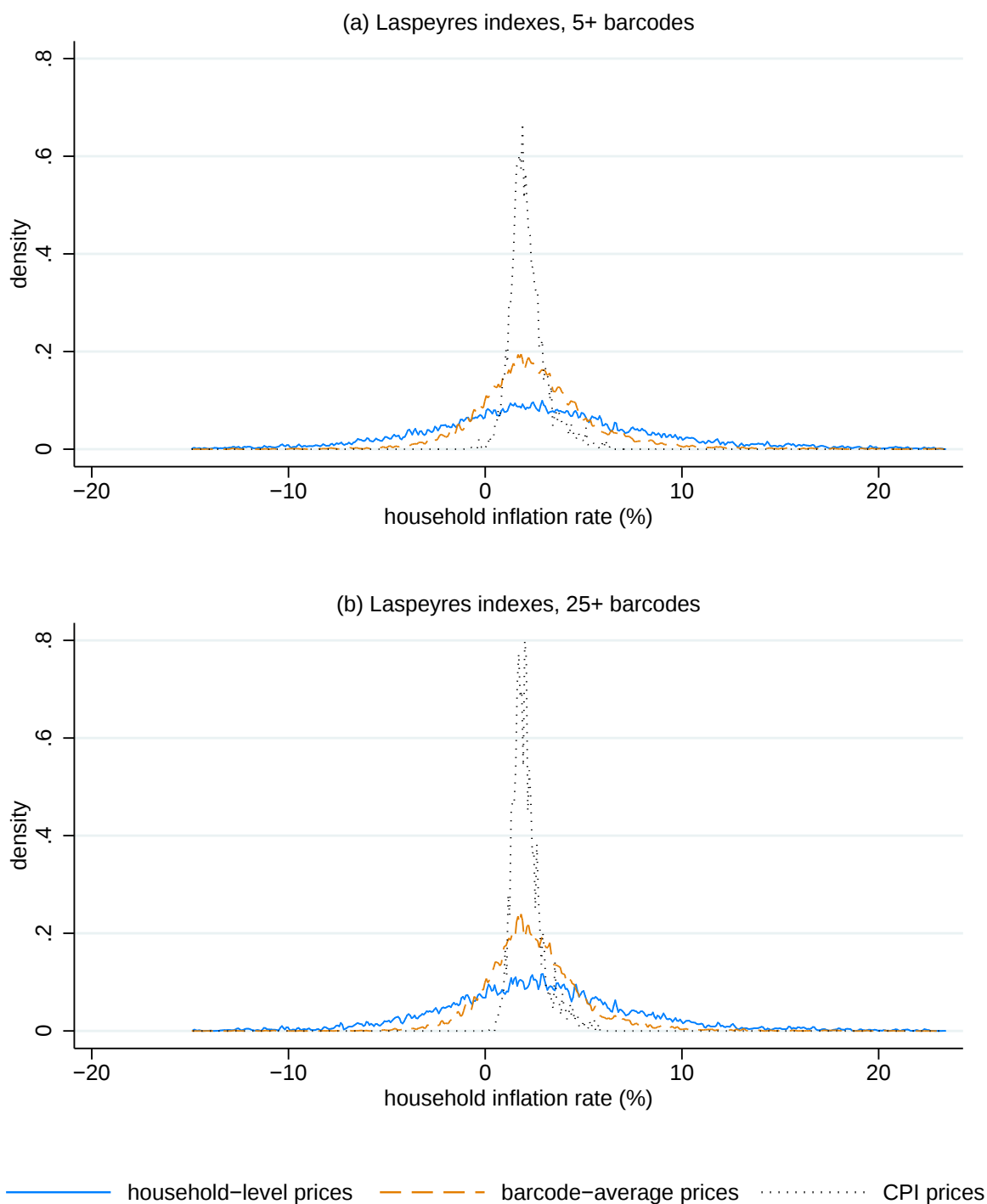


Figure 2: Distributions of household-level inflation rates from fourth quarter of 2004 to fourth quarter of 2005.

Kernel density estimates using Epanechnikov kernel. Bandwidth is 0.05 percentage point for inflation rates with household-level and barcode-average prices and 0.005 percentage point for inflation rates with CPI prices. Data on 24,412 households with matched consumption in 2004q4 and 2005q4. Plots truncated at 1st and 99th percentiles of distribution of inflation rates with household-level prices.

in household-level price indexes. The variance of the Laspeyres index with barcode-average prices is 30.6 percent of the variance of the index with household-level prices, implying that 66.9 percent of the variation in the index with household-level prices comes from variation across households in the prices paid for given barcodes, 30.6 percent from variation in the choice of barcodes within item strata, and 2.5 percent from variation in the mix of consumption across item strata.

The heterogeneity in inflation rates is not driven by households for which we can match only a few barcodes across quarters. The distributions for households with at least 25 matched barcodes, shown in the bottom panel of Figure 2, are almost as dispersed as those for all households, shown in the left column of the figure. Among households with at least 25 matched barcodes, the interquartile range of Laspeyres inflation rates with household prices is 5.8 percentage points, the difference between the 10th and 90th percentiles is 12.1 percentage points, the variance of the index with CPI prices is 2.7 percent of the variance of the index with household-level prices, and the variance of the index with barcode-average prices is 31.0 percent of the variance of the index with household-level prices.

Figure 3 examines how the dispersion of household-level inflation rates evolves over time. The graphs show results calculated from Laspeyres indexes, but graphs based on Paasche and Fisher indexes, shown in the web appendix, are almost identical; table 1 summarizes the dispersion measures from all three indexes. The patterns observed in the fourth quarter of 2004 are quite typical. Household-level inflation rates with household-level prices are enormously dispersed, with interquartile ranges of 6.2 to 9.0 percentage points using the Laspeyres index, and much more dispersed than household-level inflation rates with barcode-average or CPI prices. The bootstrap standard errors show that the amount of dispersion is precisely estimated at each date.

On the whole, the inflation inequality we measure with CPI prices is similar to what has been reported in previous literature that uses a wider universe of goods and services and imposes CPI prices on all households; for example, Hobijn et al. (2009) found a gap of 1 to 3 percentage points between the 10th and 90th percentiles. But our results show that assuming all households face the same prices and buy the same mix of goods within CPI item strata misses most of the heterogeneity in inflation rates. The gap between the 10th and 90th

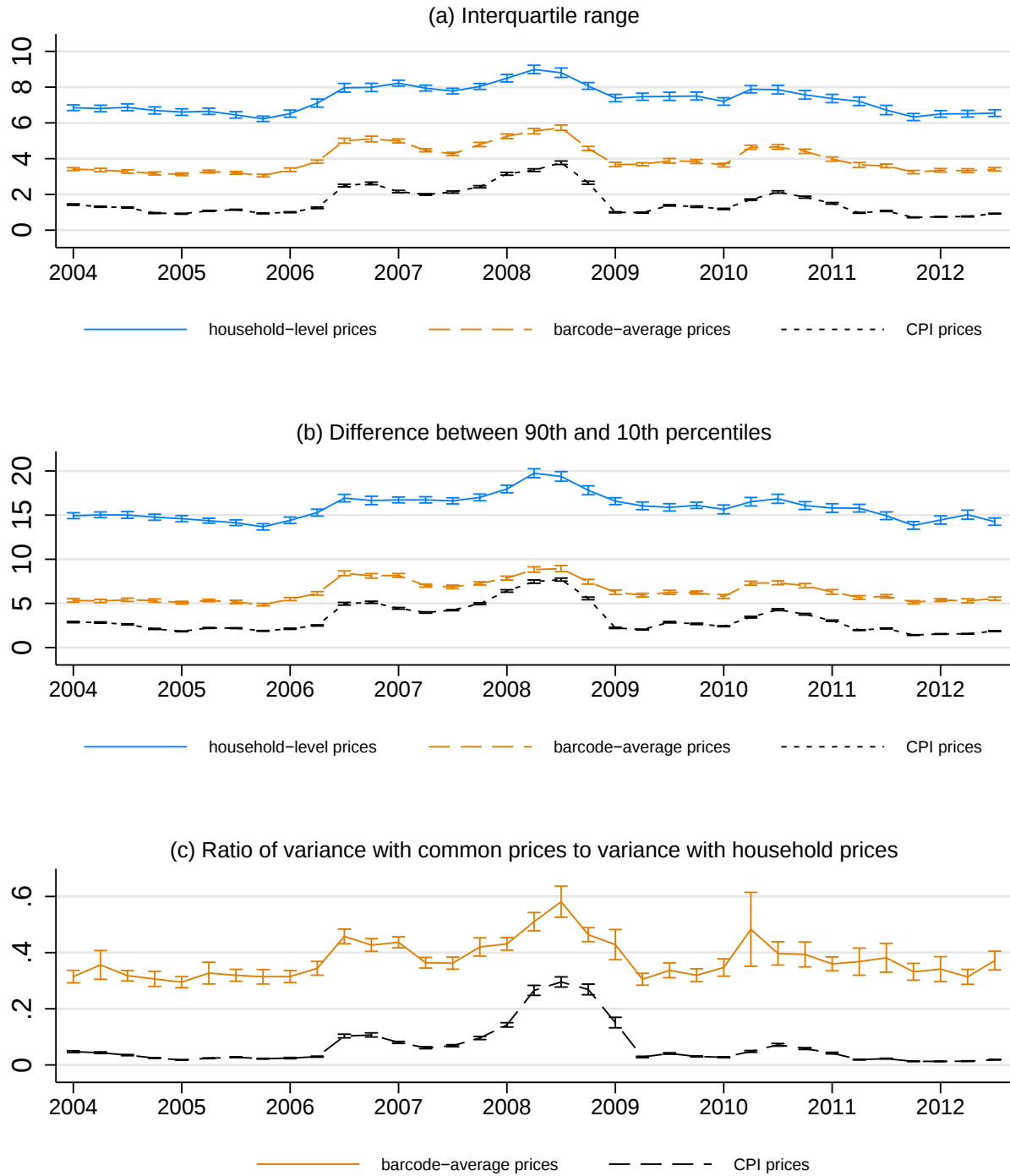


Figure 3: Measures of the dispersion of household-level inflation rates. Calculated with Laspeyres indexes. Vertical bars show an interval of ± 2 bootstrap standard errors around each point estimate. In panel (c), variances are calculated on data from 1st to 99th percentiles of distribution of inflation rates with household-level prices at each date.

Table 1: Measures of the dispersion of household-level inflation rates.

	mean	s.d.	min	max
<i>A. Interquartile range</i>				
Household-level prices:				
Laspeyres	7.33	0.74	6.23	8.99
Fisher	7.13	0.72	6.12	8.92
Paasche	7.37	0.76	6.34	9.18
Barcode-average prices:				
Laspeyres	3.99	0.77	3.06	5.73
Fisher	3.87	0.75	2.95	5.68
Paasche	3.98	0.76	3.03	5.81
CPI prices:				
Laspeyres	1.61	0.80	0.71	3.77
Fisher	1.57	0.77	0.70	3.53
Paasche	1.62	0.78	0.71	3.42
<i>B. Difference between 90th and 10th percentiles</i>				
Household-level prices:				
Laspeyres	15.87	1.44	13.67	19.74
Fisher	15.32	1.36	13.27	18.84
Paasche	15.83	1.38	13.76	19.48
Barcode-average prices:				
Laspeyres	6.39	1.18	4.85	8.94
Fisher	6.15	1.14	4.70	8.69
Paasche	6.31	1.14	4.84	8.82
CPI prices:				
Laspeyres	3.30	1.66	1.41	7.69
Fisher	3.21	1.58	1.38	7.17
Paasche	3.33	1.61	1.44	7.07
<i>C. Ratio of variance with common prices to variance with household prices</i>				
Barcode-average prices:				
Laspeyres	0.38	0.07	0.29	0.58
Fisher	0.37	0.07	0.29	0.58
Paasche	0.37	0.07	0.30	0.59
CPI prices:				
Laspeyres	0.07	0.07	0.01	0.30
Fisher	0.07	0.08	0.01	0.30
Paasche	0.07	0.08	0.01	0.30

Averages from 2004q1 through 2012q3 of dispersion measures for each date.

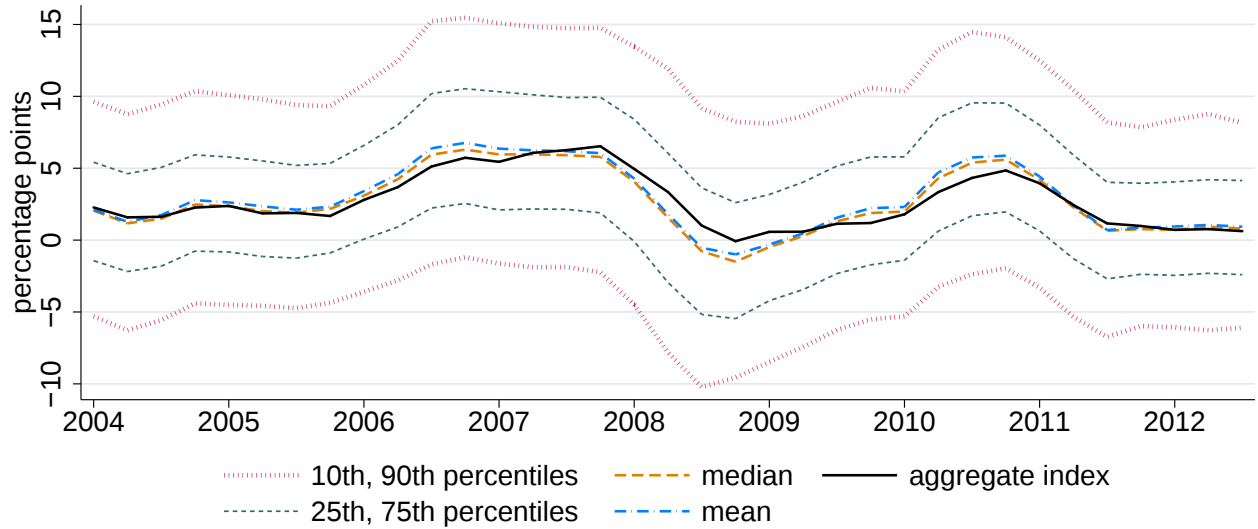


Figure 4: Evolution of the distribution of household inflation rates with household-level prices. Calculated with Laspeyres indexes. Mean is calculated on data from 1st to 99th percentiles of distribution of inflation rates at each date.

percentiles is nearly twice as large when we use barcode-average prices, allowing the mix of goods within item strata to vary across households, as when we use CPI prices. And if we also allow different households to pay different prices for the same barcode, the gap between the 10th and 90th percentiles is five times larger than when we use CPI prices.

The bottom panel of Figure 3 shows that in most years, the variance of inflation rates with CPI prices is only a few percent of the variance of inflation rates with household-level prices. However, in 2008, at the height of the Great Recession, the variance with CPI prices reaches 30 percent of the variance with household-level prices, as the sharp shift in the relative price of food at home, shown in Figure 1, makes heterogeneity in consumption bundles more important. The variance of inflation rates with barcode-average prices is typically about one-third of the variance with household-level prices, but rises in 2008. Table 1 shows that dispersion measured with the Fisher index is slightly lower than that measured with Laspeyres and Paasche indexes.

Figure 4 shows how the distribution of household-level Laspeyres inflation rates with household-level prices moves with the aggregate inflation rate $P^{L,CPI}$ computed for the corresponding universe of goods. The distribution does not exhibit any noticeable seasonality. The mean and median of the household-level Laspeyres indexes closely match the aggregate

Table 2: Quantile regressions of household-level inflation rates on aggregate price index.

Decile	Coefficient on aggregate index	Intercept
1	1.011 (0.015)	−7.602 (0.058)
2	1.013 (0.009)	−4.609 (0.039)
3	1.026 (0.008)	−2.810 (0.031)
4	1.052 (0.008)	−1.448 (0.027)
5	1.093 (0.007)	−0.264 (0.026)
6	1.137 (0.009)	0.944 (0.030)
7	1.198 (0.010)	2.286 (0.034)
8	1.243 (0.012)	4.189 (0.046)
9	1.305 (0.019)	7.491 (0.066)

The table shows the slope and intercept from quantile regressions of household-level inflation rates, computed with Laspeyres indexes and household-level prices, on the aggregate inflation rate for the equivalent universe of goods. Sample contains 835,386 household-quarter observations. Bootstrap standard errors in parentheses.

index; thus, the democratic inflation rate in the sense of Prais (1959) differs little from the plutocratic rate. Moreover, the entire distribution shifts roughly in parallel with the aggregate index.

Table 2 uses quantile regressions of household-level inflation rates on the aggregate index to measure these shifts. The lowest quantiles move one for one with the aggregate index. However, higher quantiles move somewhat more than one-for-one with the aggregate index, with the relationship strongest for the highest quantiles. Thus, the distribution spreads out, especially in the upper half, when aggregate inflation is higher. A one-percentage-point increase in the aggregate inflation index raises the gap between the median and the 90th percentile of household inflation by 0.2 percentage point.

Table 3 examines the association of household-level inflation rates with household demographics. At each date, we compute the difference between each household’s inflation rate, computed with household prices and the Laspeyres index, and the aggregate inflation rate for the equivalent universe of goods, $P^{L,CPI}$. Then we regress this difference on household characteristics, controlling for time dummies to absorb any aggregate effects that might be correlated with the sample distribution of household characteristics. (Although the regression

contains multiple observations on most households, we do not control for household fixed effects because they would absorb the variation in household characteristics, which are mostly time invariant.) The first column of the table shows results from ordinary least squares regression, while the second column shows results from median regression, which is robust to outliers in the inflation distribution.

Low-income households experience higher inflation. According to the median regression, the median annual inflation rate is 0.6 percentage point higher for a household with income below \$20,000, compared with a household with income of at least \$100,000. The difference is even larger for mean inflation rates, measured with the OLS regression. The inflation rate also is nearly half a percent lower for households in the West than in the East, and almost 0.2 percentage point lower in the Midwest than in the East. Larger households have higher inflation rates; the median inflation rate is one-fourth of a percentage point higher for a family of two adults and two children than for a single adult. In both the OLS and median regressions, inflation rates are higher for households whose heads are older. Differences by race are not statistically significant.

Although Table 3 reveals several statistically and economically significant associations of household inflation with observables, these effects are small relative to the total amount of heterogeneity. For example, moving from the bottom to the top of the age distribution raises median inflation by less than one-tenth as much as moving from the 25th to the 75th percentile of the overall distribution of inflation. As a result, observables have little power to predict household inflation rates in the cross section. The R -squared in the OLS regression is just 1.2 percent, of which three-fourths is explained by the time dummies; household characteristics explain just 0.3 percent of the cross-sectional variation in inflation rates.

4. Time-Series Properties of Household-Level Inflation

The long-run impact of heterogeneous inflation rates on households' welfare depends on whether the heterogeneity is persistent or whether a household that experiences high inflation in one year tends to experience an offsetting low inflation rate in the next year.

Figure 5 examines this persistence for a particular pair of years, 2004–2005 and 2005–2006. The figure shows the distributions of household inflation rates in 2004–2005 and in

Table 3: Regressions of household-level inflation rates on household demographics.

	Deviation of household inflation from aggregate inflation			
	(1) OLS		(2) Median	
	coefficient	std. err.	coefficient	std. err.
household income				
\$20,000–\$39,999	-0.206	(0.055)	-0.126	(0.039)
\$40,000–\$59,999	-0.420	(0.052)	-0.257	(0.041)
\$60,000–\$99,999	-0.587	(0.059)	-0.468	(0.045)
≥\$100,000	-0.731	(0.065)	-0.597	(0.050)
average age of household head(s)				
30–39	0.384	(0.205)	0.244	(0.145)
40–49	0.451	(0.203)	0.297	(0.137)
50–59	0.511	(0.206)	0.362	(0.137)
60–69	0.552	(0.199)	0.430	(0.139)
≥70	0.552	(0.203)	0.457	(0.140)
highest education of household head(s)				
high school diploma	-0.064	(0.127)	-0.029	(0.108)
some college	-0.138	(0.127)	-0.102	(0.107)
bachelor’s degree	-0.251	(0.128)	-0.163	(0.110)
graduate degree	-0.285	(0.139)	-0.137	(0.118)
Census region				
Midwest	-0.179	(0.049)	-0.175	(0.040)
South	-0.140	(0.046)	-0.063	(0.033)
West	-0.517	(0.054)	-0.429	(0.046)
# household members	0.089	(0.022)	0.110	(0.018)
has children	-0.115	(0.118)	0.074	(0.094)
has children × # household members	0.002	(0.034)	-0.033	(0.029)
black	0.058	(0.064)	0.063	(0.049)
Asian	-0.074	(0.156)	-0.089	(0.125)
other non-white	-0.033	(0.087)	-0.025	(0.067)
Hispanic	-0.068	(0.068)	-0.085	(0.056)
R^2	0.012		-	
R^2 (time dummies only)	0.009		-	
N	835,386		835,386	

The dependent variable is the difference between the household inflation rate, computed with household prices and the Laspeyres index, and the aggregate inflation rate for the equivalent universe of goods. Bootstrap standard errors in parentheses. Column (1) shows results from ordinary least squares regression, and column (2) from median regression. Regressions include time dummy variables. Omitted categories of categorical variables are: income less than \$25,000; white; non-Hispanic; heads’ average age less than 30; heads’ highest education less than high school diploma; Northeast region.

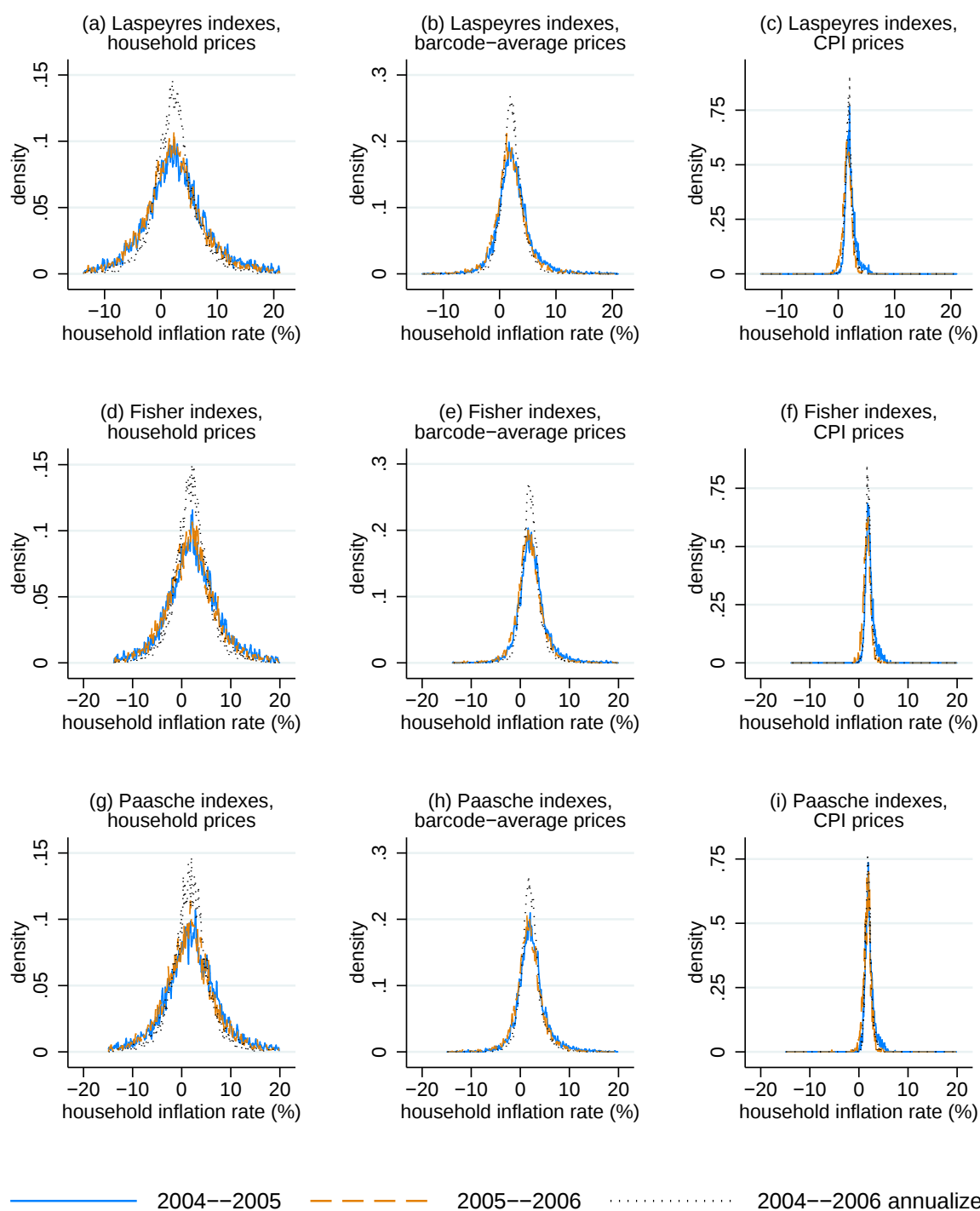


Figure 5: Distributions of one-year and two-year household-level inflation rates, 2004q4–2005q4 and 2005q4–2006q4.

Kernel density estimates using Epanechnikov kernel. Bandwidth is 0.05 percentage point for inflation rates with household-level prices and barcode-average and 0.005 percentage point for inflation rates with CPI prices. Plots truncated at 1st and 99th percentiles of distribution of inflation rates with household-level prices. Sample limited to 19,252 households with inflation rates calculated for both 2004q4–2005q4 and 2005q4–2006q4.

2005–2006, as well as the distribution of the annualized inflation rate that each household experienced over the two-year period from 2004 to 2006. The left column of the figure shows inflation rates with household-level prices, while the middle column shows inflation rates with barcode-average prices and the right column uses CPI prices. The distributions of inflation rates for the two one-year periods are similar, while the annualized two-year inflation rates are somewhat less dispersed but still very heterogeneous. Among the households with inflation rates measured in both one-year periods, the interquartile range of the Laspeyres inflation rates with household prices is 6.57 percentage points in 2004–2005, 6.10 percentage points in 2005–2006, and 4.29 percentage points for annualized inflation rates over 2004–2006. Results are similar for Fisher and Paasche indexes, and inflation rates with barcode-average prices are likewise quite persistent. Thus, even over longer horizons than a year, households experience markedly different inflation rates.

Figure 6 shows how the persistence of household-level inflation evolves over time. The middle panel of the figure shows the cross-sectional standard deviation of annualized inflation rates computed over two-year periods. The two-year inflation rates with household-level prices remain highly dispersed, much more so than two-year inflation rates with CPI prices, throughout the sample period. The standard deviation of two-year inflation rates with barcode-average prices is about halfway between the results for CPI prices and household-level prices, demonstrating that, as with one-year rates, variation in prices for a given barcode and variation in the mix of barcodes within an item stratum are both important sources of variation in household-level inflation rates. For comparison, the top panel of the figure computes the cross-sectional standard deviation of one-year inflation rates among the sample of households used to compute two-year rates (i.e., those with inflation rates observed in consecutive years). The one-year rates are more dispersed than the two-year rates, showing that heterogeneity in inflation rates does diminish when the inflation rate is computed over a longer time horizon.

The bottom panel of Figure 6 computes the cross-sectional correlation between a household’s inflation rate in quarter t and its inflation rate in quarter $t + 4$. The one-year serial correlation of inflation rates using household-level prices is approximately -0.1 , and precisely estimated, throughout the sample period. This correlation is much less negative

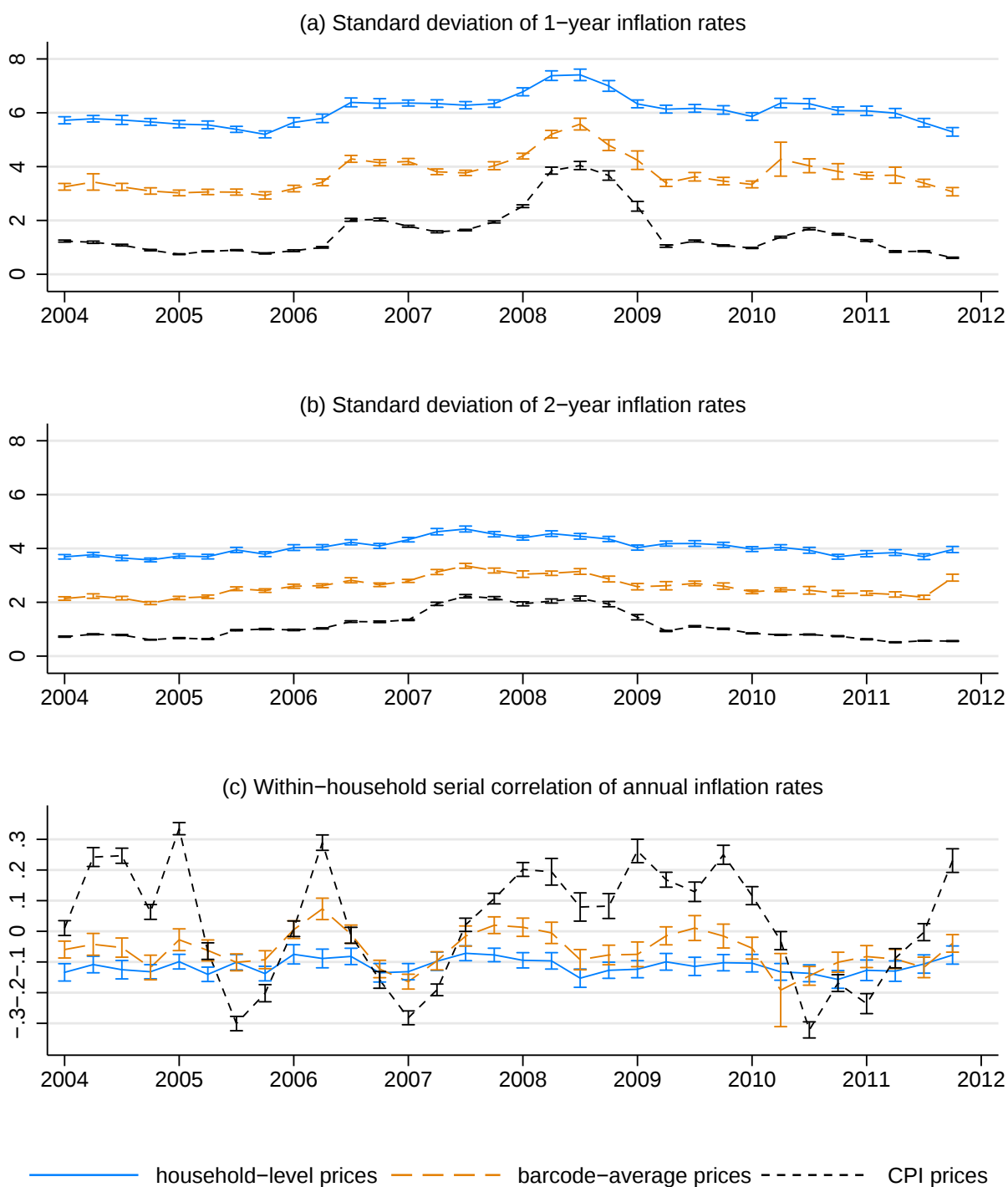


Figure 6: Evolution of the persistence of household-level inflation rates. Calculations for each quarter use the subset of households for which inflation with household-level prices is observed and falls between the 1st and 99th percentiles of the distribution in both that quarter and the quarter one year ahead. Vertical bars show an interval of ± 2 bootstrap standard errors around each point estimate.

than we would expect if households drew their *price levels* at random each period from the cross-sectional distribution of price levels; in that case, the serial correlation would be -0.5 . Indeed, if the deviation of a household's price level from the aggregate price level is an AR(1) process, a serial correlation of inflation rates of -0.1 implies that the household's price level relative to the aggregate has a serial correlation of 0.8 . Unfortunately, we lack sufficiently long data on each household to determine whether households' price levels relative to the aggregate are an AR(1) process or something more complicated.

The high cross-sectional variance and low serial correlation of household-level inflation rates together suggest that for individual households, the aggregate inflation rate is almost irrelevant as a source of variation in the household-level inflation rate. Over the sample period, our aggregate inflation rate averages 2.7 percent with a standard deviation of 1.9 percentage points, while the cross-sectional standard deviation of household-level one-year inflation rates averages 6.2 percentage points. If households' deviations from aggregate inflation are independent of the aggregate inflation rate, these figures mean that, over time, 91 percent of the variance of a household's annual inflation rate comes from heterogeneity, and only 9 percent from variability in aggregate inflation.

5. Conclusion

This paper documents massive heterogeneity in inflation rates at the household level, an order of magnitude larger than that found in previous work, owing to differences across households in prices paid within the same categories of goods. Such heterogeneity poses a range of challenges for monetary economics. Optimal policy in most monetary models is calculated to maximize the welfare of a representative household that faces the aggregate inflation rate; because extreme inflation rates cause larger welfare losses than small inflation rates, optimal policy could be different if one accounted for heterogeneity in inflation and for policy's effect on heterogeneity. In addition, even in models that relax the representative-agent assumption by allowing uninsured shocks to generate heterogeneity in assets and consumption, all households typically face the same inflation rate and hence the same real interest rate (see, e.g., Kaplan, Moll, and Violante, 2015). Optimal policy might differ if models allowed households to face identical nominal interest rates but, because inflation rates vary,

different real rates. Furthermore, the heterogeneity we observe suggests that movements of the aggregate price level may not be an important determinant of individual agents' inflation rates, potentially explaining why households and small firms fail to be well informed about aggregate inflation and monetary policy (Binder, 2015; Kumar et al., forthcoming).

However, the implications of household-level inflation heterogeneity depend importantly on whether households can forecast where they will fall in the cross-sectional distribution. If a household has no idea where in the inflation distribution its particular inflation rate will fall each year, the household's best way to forecast its own inflation rate is still to forecast the aggregate inflation rate. In such a case, while heterogeneity in realized inflation rates may have distributional consequences, it should have little impact on inflation expectations or forward-looking decisions. By contrast, if households can predict whether their own inflation rates will be above or below average, heterogeneity in inflation rates will affect expectations and dynamic choices. Our data show that inflation rates at the household level are only weakly correlated with observables and nearly serially uncorrelated. Thus, we have little ability to forecast household-level deviations from aggregate inflation, either in the cross-section or over time, with the limited information available to us as econometricians. Whether households can use their much larger information sets to make better forecasts of their idiosyncratic inflation rates is an important question that we leave for further research.

References

- Binder, Carola, 2015, "Fed Speak on Main Street," manuscript, Haverford College.
- Bronnenberg, Bart J., Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro, 2015, "Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium," *Quarterly Journal of Economics* 130(4), 1669–1726.
- Bureau of Labor Statistics, 2015, "The Consumer Price Index," in *Handbook of Methods*, Washington: Bureau of Labor Statistics.
- Einav, Liran, Ephraim Leibtag, and Aviv Nevo, 2010, "Recording discrepancies in Nielsen Homescan data: Are they present and do they matter?", *Quantitative Marketing and Economics* 8(2), 207–239.

- Federal Open Market Committee, 2015, “Statement on Longer-Run Goals and Monetary Policy Strategy,” adopted effective Jan. 24, 2012, amended effective Jan. 27, 2015. Accessed Dec. 24, 2015, at http://www.federalreserve.gov/monetarypolicy/files/FOMC_LongerRunGoals.pdf.
- Handbury, Jessie, 2013, “Are Poor Cities Cheap for Everyone? Non-homotheticity and the Cost of Living Across U.S. Cities,” manuscript, University of Pennsylvania.
- Hobijn, Bart, and David Lagakos, 2005, “Inflation Inequality in the United States,” *Review of Income and Wealth* 51(4), 581–606.
- Hobijn, Bart, Kristin Mayer, Carter Stennis, and Giorgio Topa, 2009, “Household Inflation Experiences in the U.S.: A Comprehensive Approach,” Federal Reserve Bank of San Francisco Working Paper 2009-19.
- Kaplan, Greg, and Guido Menzio, forthcoming, “The Morphology of Price Dispersion,” *International Economic Review* 56(4), 1165–1206.
- Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante, 2015, “Monetary Policy According to HANK,” manuscript, Princeton University.
- Kilts Center Archive of the Nielsen Company, 2013a, “Consumer Panel Dataset FAQ’s.”
- Kilts Center Archive of the Nielsen Company, 2013b, “Consumer Panel Dataset Manual.”
- Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko, forthcoming, “Inflation targeting does not anchor inflation expectations: Evidence from firms in New Zealand,” *Brookings Papers on Economic Activity*.
- Prais, S.J., 1959, “Whose Cost of Living?”, *Review of Economic Studies* 26(2), 126–34.
- Rao, J.N.K., C.F.J. Wu, and K. Yue, 1992, “Some recent work on resampling methods for complex surveys,” *Survey Methodology* 18, 209–217.
- Reis, Ricardo, 2006, “Inattentive consumers,” *Journal of Monetary Economics* 53(8), 1761–1800.

Sims, Christopher A., 2003, “Implications of rational inattention,” *Journal of Monetary Economics* 50, 665–690.