

Inflation at the Household Level[☆]

Greg Kaplan^{a,b}, Sam Schulhofer-Wohl^c

^a*Department of Economics, University of Chicago, Saieh Hall, 5757 S. University Ave., Chicago, IL 60637*

^b*National Bureau of Economic Research, 1050 Massachusetts Ave., Cambridge, MA 02138*

^c*Economic Research Department, Federal Reserve Bank of Chicago, 230 S. LaSalle St., Chicago IL 60604*

Abstract

We use scanner data to estimate inflation rates at the household level. Households' inflation rates are remarkably heterogeneous, with an interquartile range between 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. Lower-income households experience markedly higher inflation, with a difference in average inflation rates of nearly 1 percentage point per year between the highest- and lowest-income households. Nonetheless, most of the cross-sectional variation in inflation rates is uncorrelated with observable characteristics. 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. Together, the large cross-sectional dispersion and low serial correlation of household-level inflation rates 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.

[☆]We thank Cristian Alonso and Emily Moschini for research assistance, Joan Gieseke for editorial assistance, and our discussants Yuriy Gorodnichenko and Sarah Lein as well as numerous conference and seminar participants for helpful comments. An earlier version of this paper was completed while Schulhofer-Wohl was employed at the Federal Reserve Bank of Minneapolis. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Chicago, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.

Email addresses: gkaplan@uchicago.edu (Greg Kaplan), sam@frbchi.org (Sam Schulhofer-Wohl)

Keywords: inflation, heterogeneity

JEL classification: D12, D30, E31

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
5 (Federal Open Market Committee, 2016). 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 affect households' economic choices and well-being, we need to know how inflation behaves at the
10 household level.

We take 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 households buy different bundles of goods or because households pay different prices for the same goods. Previous research on inflation
15 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 (KNCP), a dataset that records the prices, quantities and specific goods purchased in 500 million transactions by about 50,000 U.S. households from 2004 through 2013, we can also consider variation in prices paid and
20 in the mix of goods within broad categories.¹ 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

¹Jaravel (2016) uses the KNCP data to estimate inflation rates as a function of income but assumes that all households in a given broad range of incomes (such as all households earning \$30,000 to \$100,000 per year) have the same consumption bundle and pay the same price for each good.

25 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 in-
flation shifts in parallel with aggregate inflation, and the central tendency of household-level
inflation closely tracks aggregate inflation.² Households with low incomes, more household
30 members, or older household heads experience higher inflation on average, whereas those
in the Midwest and West experience lower inflation. The cumulative differences in infla-
tion rates across income groups, in particular, are striking: Over the nine years from the
third quarter of 2004 through the third quarter of 2013, average inflation cumulates to 33
percent for households with incomes below \$20,000 but to just 25 percent for households
35 with incomes above \$100,000. The negative correlation of inflation with income implies that
inequality in real incomes is rising faster than inequality in nominal incomes. Nonetheless,
the differences in inflation rates across demographic groups are small relative to the total
variance of the inflation distribution, and observable household characteristics have little
power overall to predict household inflation rates.

40 The household-level inflation rates display reasonable substitution patterns in the aggre-
gate. Households substitute on average toward lower-priced goods, and the average substi-
tution bias that results from measuring inflation with initial-period consumption bundles
is comparable in magnitude to what has been estimated for aggregate indexes such as the
Consumer Price Index (CPI). However, as with inflation rates themselves, substitution pat-
45 terns display a great deal of heterogeneity. Many households are observed to substitute
toward higher-priced goods, and demographics have little power to explain differences in
substitution.

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
50 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

²Neither of these results is mechanical. If the rate of increase of the prices that a particular household
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.

price level is quite persistent. We use our estimates to provide a simple characterization of the stochastic process for deviations of household-level inflation from aggregate inflation. Together, the low serial correlation and high cross-sectional variation of household-level inflation 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 for both 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), 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 (Sims, 2003; Reis, 2006) 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.

Since the seminal work by Lucas (1972) and Lucas (1975), macroeconomics has had a long tradition of modeling monetary non-neutrality as arising because agents (either households or firms) have imperfect information about whether the changes they observe in nominal variables reflect economy-wide or agent-specific shocks. More sophisticated recent descendants of this literature include Angeletos and La'O (2009) and Nimark (2008). For more than forty years, this literature has remained essentially purely theoretical, partly because attempts to quantify the strength of the imperfect-information mechanism require some way of disciplining the amount of information about aggregate price changes contained in individual observations of prices. Our estimates of the stochastic process relating household-level inflation to aggregate inflation could, in principle, be used as a first step toward empirical quantification of these models. Although we do not take this step in this paper, our central finding — that almost all of the variation in household-level inflation is disconnected from movements in aggregate inflation — supports the idea that the informational frictions embedded in these models could be substantial.

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 infla-

tion distribution. Section 4 characterizes time-series properties of household-level inflation. Section 5 concludes.

2. Data and Estimation

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

2.1. Data

The KNCP tracks the shopping behavior of approximately 50,000 U.S. households over the period 2004 to 2013. Households in the panel provide information about each of their
120 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
125 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)
130 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
135 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).³

³The 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).

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 to resample from these households. We therefore treat the sample as if it is replenished at an annual frequency. We resample households within groups defined by geographic stratum and the year in which the household first appears in the data. 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 because of particular households' unique purchasing patterns. 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.

2.2. 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
 170 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
 175 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,
 180 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 percent of households that make any purchase in quarter t also make some purchase in quarter $t + 4$, and 72 percent buy at least five matched barcodes whose prices change 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

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

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

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

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

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

and

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

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

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

We compute three 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

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

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, $\pi_{it,t+4}^{P,BC}$ and $\pi_{it,t+4}^{F,BC}$, 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. 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 — and assigning to each purchase the average inflation rate for that item stratum, 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. We have two possible sources for average inflation rates at the stratum level: the prices observed in the KNCP and the stratum price indexes published by the BLS for the various strata that make up the CPI. 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

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

and the Paasche and Fisher indexes with CPI prices, $\pi_{i,t,t+4}^{P,CPI}$ and $\pi_{i,t,t+4}^{F,CPI}$, are defined analogously.⁴ To produce a similar index with stratum-average prices from the KNCP, we must first produce the analog to $p_{k,t+4}^{CPI}/p_{kt}^{CPI}$ with the KNCP data. The Laspeyres version of this stratum-level inflation rate is

$$\pi_{kt,t+4}^{L,S} = \frac{\sum_{\substack{i,j: j \in k \\ q_{ijt}, \\ q_{ij,t+4} > 0}} q_{ijt} \bar{p}_{j,t+4}}{\sum_{\substack{i,j: j \in k \\ q_{ijt}, \\ q_{ij,t+4} > 0}} q_{ijt} \bar{p}_{j,t}}. \quad (8)$$

⁴We calculate the quarterly value of the CPI as the average of the monthly values for the three months in the quarter.

Then the Laspeyres inflation rate at the household level with stratum-average prices is

$$\pi_{it,t+4}^{L,S} = \sum_k s_{ikt,t+4}^L \pi_{kt,t+4}^{L,S}, \quad (9)$$

where $s_{ikt,t+4}^L$ is the share of stratum k in household i 's total spending at date t . The Paasche and Fisher indexes with stratum-average prices are defined similarly.

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:

$$s_{jt,t+4}^L = \frac{\sum_{\substack{i: q_{ijt}, \\ q_{ij,t+4} > 0}} p_{ijt} q_{ijt}}{\sum_{\ell} \sum_{\substack{i: q_{i\ell t}, \\ q_{i\ell,t+4} > 0}} p_{i\ell t} q_{i\ell t}}, \quad (10)$$

and our aggregate inflation index is

$$\pi_{t,t+4}^{L,CPI} = \sum_j s_{jt,t+4}^L \frac{p_{k(j),t+4}^{CPI}}{p_{k(j),t}^{CPI}}. \quad (11)$$

¹⁹⁰ The aggregate index $\pi_{t,t+4}^{L,CPI}$ 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.⁵

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

$$\pi_{t,t+4}^{L,CPI} = \frac{\sum_i x_{it,t+4} \pi_{it,t+4}^{L,CPI}}{\sum_i x_{it,t+4}}, \quad (12)$$

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:

$$x_{it,t+4} = \sum_{\substack{j: q_{ijt}, \\ q_{ij,t+4} > 0}} p_{ijt} q_{ijt}. \quad (13)$$

Because $\pi_{t,t+4}^{L,CPI}$ 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.

The online appendix gives details on the distribution of spending across types of goods in the KNCP. About 61 percent of spending in the KNCP is on food and beverages, a share that rises to 74 percent in the matched purchases that we use to measure household inflation. By contrast, food and beverages have only a 15 percent weight in the published CPI. But despite the heavy weight of food in the KNCP, many other types of purchases are represented, including housekeeping supplies, pet products, and personal care items. Housing and transportation, on the other hand, get much less weight in our data than in the CPI. Apparel is measured in the KNCP but gets zero weight in our household inflation rates because we observe no purchases of matched apparel barcodes in consecutive periods.

Figure 1 compares the aggregate index for the KNCP universe, $\pi^{L,CPI}$, with the published CPI and several CPI sub-indexes. The dates on the horizontal axis correspond to the initial

⁵Another, 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 , whereas the CPI uses a fixed base period.

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, as does the CPI sub-indexes for food at home. 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 sub-index. 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.

The foregoing analysis considers only differences in the categories of goods covered in the KNCP versus the CPI, applying the same CPI prices to both datasets. (Recall that we use CPI prices to compute the KNCP aggregate inflation index, $\pi^{L,CPI}$.) However, when we compute household-level inflation rates, we will use price data from the KNCP. In principle, the prices recorded in the KNCP could differ from those recorded in the CPI for similar goods. To assess this concern, we map each barcode in the KNCP to a CPI item stratum and compute inflation rates in the KNCP for each item stratum.⁶ Figure 2 shows two ways of comparing the resulting item stratum inflation rates in the KNCP with the corresponding published item stratum inflation rates for the CPI. In the upper panel, we show the distribution across item strata of the difference between the KNCP inflation rate

⁶We compute these inflation rates using the Laspeyres index and the volume-weighted average observed price for each barcode at each date.

and the published CPI inflation rate. The mean and median differences are small, and in
240 most quarters, the discrepancy is no more than 1 percentage point for roughly half of the
item strata. However, some item strata have substantially larger discrepancies. The lower
panel of the figure aggregates the item strata inflation rates from the KNCP and the CPI to
produce aggregate inflation rates, weighting each item stratum by total expenditure on that
stratum in the KNCP. The aggregate inflation rate using KNCP prices is almost identical to
245 that using CPI prices. Thus, on average, the prices recorded in the KNCP reflect the same
trends as the prices recorded in the CPI.

3. The Cross-Sectional Distribution of Inflation Rates

Figure 3 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
250 each panel, a kernel density estimate of the distribution of household-level inflation rates with
household-level prices is superimposed on estimates of the distributions of the household-
level rates with barcode-average prices, with stratum-average prices, and 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
255 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
260 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 variance of the index with CPI prices
is 2.5 percent of the variance of the index with household-level prices,⁷ and the interquartile

⁷We compute all 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.

265 range is just 0.95 percentage point. The inflation rates with stratum-average KNCP prices are slightly more dispersed, with an interquartile range of 1.25 percentage point and a variance that is 9.8 percent of the variance of the index with household-level prices. Relative to the CPI prices, stratum-average prices in the KNCP are likely to be a noisy measure of the true stratum-level price indexes. As a result, the household-level inflation rates with stratum-
270 average prices are likely to overstate the true amount of variation attributable to allocation of expenditure across item strata. Therefore, in the remainder of the analysis, we use CPI prices to assess the amount of variation due to the allocation of expenditure across item strata.

Differences in the barcodes that households buy within item strata and differences in the
275 prices that households pay for particular barcodes are both important sources of variation 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
280 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 middle panel of Figure 3, are almost as dispersed as those
285 for the baseline group of households with at least five matched barcodes, shown in the top panel. 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
290 the variance of the index with barcode-average prices is 31.0 percent of the variance of the index with household-level prices.

Spending on matched barcodes is only a minority of most households' spending. For the median household, across all quarters, 21 percent of spending is on matched barcodes, and for three-quarters of households, less than 30 percent of spending is on matched barcodes.

295 However, the heterogeneity in inflation rates is not driven by households for which especially little spending is on matched barcodes. The distributions of inflation rates for the one-fourth of households that devote at least 30 percent of their spending to matched barcodes, shown in the bottom panel of Figure 3, show similar dispersion to those for all households.

Figure 4 examines how the dispersion of household-level inflation rates evolves over time. 300 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 various 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 305 the Laspeyres index, and much more dispersed than household-level inflation rates with barcode-average, stratum-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 310 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 percentiles is nearly twice as large when we use barcode-average prices, allowing 315 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 Table 1 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. 320 However, in 2008, at the height of the Great Recession, the variance with CPI prices reached 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, made heterogeneity in broadly defined consumption bundles more important. Table 1 also shows that dispersion measured with the Fisher index is slightly lower than that measured with Laspeyres and Paasche indexes.

325 The bottom panel of Figure 4 shows how the distribution of household-level Laspeyres
inflation rates with household-level prices moves with the aggregate inflation rate $\pi^{L,CPI}$
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 index; thus, the democratic inflation rate in the sense of Prais (1959)
330 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 aggregate inflation
and on the median of household-level inflation to measure these shifts. Panel (1) of the
table shows that the lowest quantiles of household inflation move one for one with the
335 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. An
increase of 1 percentage point in the aggregate inflation index raises the gap between the
median and the 90th percentile of household inflation by 0.2 percentage point. This spreading
340 out comes from two sources. First, as panel (2) of the table shows, higher quantiles move
slightly more than one-for-one with median inflation. Second, median inflation itself moves
more than one-for-one with the aggregate index, as indicated by the slope coefficient for the
fifth decile in panel (1).

Table 3 examines the consequences of inflation inequality for income inequality. The
345 table shows the cumulative inflation rates for households at different income levels over the
nine-year period for which we have data, as well as some demographic characteristics of
households in the different income groups.⁸ Because there is substantial attrition in the
dataset, we use a synthetic cohort approach to calculate cumulative inflation rates: We
first compute the average one-year inflation rate in each income group for each year, then
350 cumulate the average one-year rates to find an average cumulative inflation rate. (The online
appendix illustrates the year-by-year inflation rates for each group.) Over the nine years of
data, the cumulative inflation rate is 8 to 9 percentage points higher for households with

⁸Household income is measured in nominal terms but is reported in bins, so we cannot group households
by real income in a consistent way over time.

incomes above \$100,000, compared with households with incomes below \$20,000. If these differences in inflation rates for goods in the KNCP extended to the universe of goods and services, they would imply that the difference in real incomes between the top and bottom groups was growing at a rate of nearly 1 percentage point per year faster than the difference in nominal incomes.

The lowest-income households are disproportionately likely to have elderly and less-educated heads and to live in the South, and disproportionately less likely to have children and to be Asian or Hispanic. But these demographic differences do not explain much of the difference in inflation rates across income groups. Table 4 examines the association of household-level inflation rates with a large set of household demographics in a regression framework. At each date, we compute the difference between each household's inflation rate and the aggregate inflation rate for the equivalent universe of goods, $\pi^{L,CPI}$. 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 two columns compute household-level inflation rates with household-level prices and the Laspeyres index, estimating the coefficients first with ordinary least squares regression and then with median regression, which is robust to outliers in the inflation distribution. The third and fourth columns compute household-level inflation rates with barcode-average and CPI prices, respectively, estimating the coefficients with median regression in both cases.

The results with household-level prices show several significant associations between demographics and household-inflation. Low-income households experience higher inflation, even after controlling for all other demographics. 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, holding other demographics fixed. The difference is even larger for mean inflation rates, measured with the OLS regression, and if cumulated over nine years would be nearly as large as the difference in Table 3 that does not control for other demographics. The inflation rate also is nearly

half a percentage point 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 with household-level prices are higher for households whose heads are older. Differences by race are not statistically significant.

The results with barcode-average and CPI prices show similar correlations, except that the magnitude of the coefficients typically becomes smaller as prices are computed at higher levels of aggregation. For example, the difference in median inflation rates between the lowest and highest income groups is 0.1 percentage point with CPI prices and 0.4 percentage point with barcode-average prices, compared with 0.6 percentage point with household-level prices. The coefficients on age categories likewise shrink as prices are aggregated, as do the coefficients on region and household size. Thus, measures of household-level inflation that use aggregated prices may underestimate the differences in inflation rates across demographic groups. However, education is an exception to this pattern, with barcode-average prices producing larger coefficients than household-level prices.

Although Table 4 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 with household-level prices 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.

The remaining columns of Table 4 examine the association between demographics and the dispersion of inflation rates, as well as one possible explanation for differences in dispersion: differences in shopping behavior. Column (5) repeats the median regression of household inflation rates on demographics from column (2) but controls for the number of shopping

trips⁹ that the household made in quarters t and $t + 4$. All else equal, an increase in the number of shopping trips in quarter t is associated with a higher inflation rate from t to $t + 4$, while an increase in the number of shopping trips in quarter $t + 4$ is associated with a lower inflation rate from t to $t + 4$. This association suggests that households find lower prices when they make more shopping trips, as many search models would predict. However, the magnitudes of these effects are relatively small. For example, to reduce the inflation rate by one-fourth of one percentage point, one would need to double the number of shopping trips in the final quarter. The combined effect of equal percentage increases in shopping trips in both the initial and final quarters is close to zero, implying that, all else equal, a household's average level of search over time affects the household's price level but not its inflation rate. Controlling for the number of shopping trips does not much change the effects of demographics, compared with the equivalent coefficients in column (2). Thus, the demographic differences in inflation rates do not appear to result from differences in shopping behavior.

Columns (6) and (7) of Table 4 show how the dispersion of household inflation rates depends on demographics, without and then with controls for the number of shopping trips. The interquartile range of inflation rates is lower for higher-income, older and larger households, and those living outside the Northeast region, especially in the South. The magnitudes are substantial: The interquartile range is 0.87 percentage point smaller for the highest-income households relative to the lowest-income households, and at least 1 percentage point smaller for households with heads at least age 40 relative to households with heads under age 30. However, dispersion is higher for nonwhite households and has a U-shaped relationship with education. Controlling for the number of shopping trips does not much change these relationships, but households that make more shopping trips have less-dispersed inflation rates, again consistent with the predictions of many search models. All else equal, doubling the number of shopping trips in the initial quarter reduces the interquartile range by 0.13 percentage point, while double the number of trips in the final quarter reduces the interquartile range by 0.35 percentage point.

⁹The count of shopping trips excludes trips on which the household spent less than \$1.

440 When consumption shifts toward less-expensive goods as prices change, the Laspeyres index (which weights goods by initial-period consumption) shows higher inflation than the Paasche index (which weights by final-period consumption). Thus, differences between Laspeyres and Paasche inflation rates show the extent of substitution toward less-expensive goods. The top panel of figure 5 plots the mean differences between households' Laspeyres and Paasche inflation rates and their geometric mean, the Fisher index, for each quarter
445 in the sample. The Laspeyres inflation rate averages 0.33 percentage point higher than the Fisher index, which in turn averages 0.30 percentage point higher than the Paasche index. Hence, on average, the data show that households substitute toward less-expensive goods, as predicted by standard models of consumer demand. The discrepancy between Laspeyres and
450 Fisher indexes is known as substitution bias because the Fisher index is an approximately correct measure of the increase in the cost of living, whereas the Laspeyres index is biased upward as a measure of the cost of living because it ignores substitution toward lower-priced goods. The 0.33 percentage point average substitution bias in our household-level inflation rates is similar to typical estimates of substitution bias in aggregate data, such as the Boskin
455 Commission's estimate of a 0.4 percentage point substitution bias in the CPI (Boskin et al., 1996, Table 3).

The average substitution patterns mask a great deal of heterogeneity. The bottom panel of figure 5 shows the distribution of the difference between the Laspeyres and Paasche indexes for each household, $\pi_{it,t+4}^L - \pi_{it,t+4}^P$, for a typical period, the fourth quarter of 2004 to the
460 fourth quarter of 2005. Although the Paasche index measures a lower inflation rate than the Laspeyres index for the average household, this relationship is far from uniform. Slightly more than 40 percent of households fail to substitute in the expected direction and have a higher Paasche index than Laspeyres index.

The online appendix examines the relationship between household demographics and
465 substitution patterns. As with household-level inflation rates themselves, the difference between Laspeyres and Paasche indexes has an economically and statistically significant correlation with some demographics, but the R -squared is essentially zero: Demographics have almost no power to explain cross-sectional differences in substitution patterns.

We can also test whether household spending patterns are consistent with simple models

of intertemporal substitution. Models in which households can substitute spending over time generally predict that, all else equal, households will consume more at dates with low relative prices; the extent of this substitution will depend on the elasticity of intertemporal substitution and on any borrowing or savings constraints the household faces. We test whether intertemporal substitution goes in the expected direction by decomposing the cross-sectional variation in expenditure growth into variation in inflation and variation in quantity growth. For any household i , the growth rate of expenditure x between quarters t and $t + 4$ is

$$\ln x_{i,t+4} - \ln x_{it} = \ln \pi_{it,t+4} + \ln q_{i,t+4} - \ln q_{it} \quad (14)$$

where $\pi_{it,t+4}$ is a household-level inflation rate and q_t is an index of the real quantity consumed at date t . Hence the cross-sectional variation in spending growth can be decomposed as

$$\text{Var}(\ln x_{i,t+4} - \ln x_{it}) = \text{Var}(\ln \pi_{it}) + \text{Var}(\ln q_{i,t+4} - \ln q_{it}) + 2\text{Cov}(\pi_{it}, \ln q_{i,t+4} - \ln q_{it}). \quad (15)$$

Given expenditure growth, which we observe, and household-level inflation, which we calculate, we can recover the growth rate of the quantity index for each household directly from (14). We can then compute the variances in (15). We use the Laspeyres inflation rate with household-level prices to measure π_{it} , and expenditure on goods purchased at both t and $t+4$ to measure $x_{i,t+4}$ and x_{it} . (Thus, as throughout the analysis, this decomposition examines substitution on the intensive but not the extensive margin.) On average across quarters, the variance of expenditure growth is 0.116, the variance of log inflation is 0.007, the variance of quantity growth is 0.113, and the covariance between log inflation and quantity growth is -0.002. Thus, as theory predicts, households spend more when they face low relative prices. Given a specific model of intertemporal choice, one could use the variances and covariances we observe to recover the average elasticity of intertemporal substitution, though we do not pursue such an exercise here because the results would depend greatly on details of the model.

The combination of a positive variance of inflation and negative covariance between inflation and expenditure means that inequality in real consumption growth is almost the same as inequality in nominal consumption growth. However, while the amount of inequality is

485 the same in real and nominal terms, households' positions in the distribution of real expenditure growth are not the same as their positions in the distribution of nominal expenditure growth.

4. Time-Series Properties of Household-Level Inflation

490 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 6 shows how the persistence of household-level inflation evolves over time. For each household whose inflation rate is measured in two consecutive one-year periods, we compute an annualized inflation rate over the two years. The middle panel of the figure shows the cross-sectional standard deviation of these annualized two-year inflation rates, computed with 495 Laspeyres indexes. They remain highly dispersed, demonstrating that even over horizons longer than a year, households experience markedly different inflation rates. (The online appendix shows that results are similar for Fisher and Paasche indexes and exhibits the entire distribution of two-year inflation rates for an illustrative time period.) The standard deviation of two-year inflation rates with barcode-average prices is about halfway between 500 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 505 (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 510 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 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, we can use the serial correlation of inflation rates to characterize a simple stochastic process for household price levels. Assume that the deviation of a household's price level from the aggregate price level consists of a household fixed effect plus an AR(1) process at an annual frequency. That is,

$$\ln P_{it} - \ln P_t = \mu_i + \rho(\ln P_{i,t-4} - \ln P_{t-4} - \mu_i) + \epsilon_{it}, \quad (16)$$

where P_{it} is the price level of household i in quarter t , P_t is the aggregate price level in quarter t , μ_i is a household fixed effect, and ϵ_{it} is i.i.d. across households and over time with mean zero and variance σ^2 . Also assume that households' initial conditions are drawn from the ergodic distribution, so that the distribution of $\ln P_{it} - \ln P_t$ is stationary, which seems reasonable given the relative stability of the standard deviations and serial correlations shown in Figure 6. Then

$$\text{Var}(\pi_{it}) = \frac{2\sigma^2}{1 + \rho} \quad (17)$$

and

$$\text{Cov}(\pi_{it}, \pi_{i,t-1}) = \sigma^2 \frac{\rho - 1}{1 + \rho}, \quad (18)$$

from which it follows that

$$\rho = 1 + 2 \frac{\text{Cov}(\pi_{it}, \pi_{i,t-1})}{\text{Var}(\pi_{it})} = 1 + 2\text{Corr}(\pi_{it}, \pi_{i,t-1}), \quad (19)$$

515 where the second equality uses the stationarity of the distribution. Thus, a serial correlation of inflation rates of -0.1 implies a serial correlation of price levels of $\rho = 0.8$. As a result, shocks to households' price levels are persistent but not permanent, according to this stochastic process.

520 Together, the high cross-sectional variance and low serial correlation of household-level inflation rates 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, whereas the cross-sectional standard deviation of household-level one-

year inflation rates averages 6.2 percentage points. If households' deviations from aggregate
525 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
— either the household fixed effect μ_i or the idiosyncratic shocks ϵ_{it} — and only 9 percent
from variability in aggregate inflation.

5. Conclusion

530 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
535 inflation rate; because extreme inflation rates cause larger welfare losses than small infla-
tion rates, optimal policy could be different if one accounted for heterogeneity in inflation
and for the 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
540 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 infla-
tion 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
545 informed about aggregate inflation and monetary policy (Kumar et al., forthcoming; Binder,
2015).

Our results also have implications for the measurement of income inequality. Because
inflation is higher on average for lower-income households for the goods in our data, inequality
in real incomes may be rising faster than inequality in nominal incomes. However, to fully
550 understand the effect of inflation heterogeneity on income inequality, it would be important
to extend the inflation measurements to a more comprehensive set of goods and services
than that measured in the KNCP. Future research could also investigate why inflation is

higher for lower-income households. Jaravel (2016) proposes a theory involving greater rates of innovation in product categories popular with high-income households, but other explanations are also possible, such as different shopping behavior among high-income people.

Heterogeneity in realized inflation could also help to explain heterogeneity in inflation expectations. Allowing heterogeneity only in the allocation of spending across CPI item strata, Johannsen (2014) shows that demographic groups with greater dispersion in realized inflation also have greater dispersion in inflation expectations and proposes an imperfect information model to explain this finding. Our results show that there is substantially more heterogeneity in realized inflation once we account for differences in the allocation of spending across goods within item strata and differences in prices paid for identical goods. The KNCP does not measure inflation expectations, but it would be valuable for future researchers to collect data on household-level inflation expectations and UPC-level purchasing patterns within the same dataset so that the relationship between expectations and realized inflation could be examined while allowing for all sources of heterogeneity.

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.

Angeletos, G. and Jennifer La'O, 2009, "Incomplete Information, Higher-Order Beliefs and Price Inertia," *Journal of Monetary Economics* 56, S19–S37.

Binder, Carola, 2015, “Fed Speak on Main Street,” manuscript, Haverford College.

Boskin, Michael J., Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, and Dale Jorgenson,
585 1996, “Toward a More Accurate Measure of the Cost of Living,” Final Report to the Senate
Finance Committee from the Advisory Commission to Study the Consumer Price Index,
accessed March 17, 2016, at <https://www.ssa.gov/history/reports/boskinrpt.html>.

Bronnenberg, Bart J., Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro, 2015,
“Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium,” *Quarterly*
590 *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*
595 *Economics* 8(2), 207–239.

Federal Open Market Committee, 2016, “Statement on Longer-Run Goals and Monetary
Policy Strategy,” adopted effective Jan. 24, 2012, amended effective Jan. 26, 2016. Ac-
cessed March 22, 2016, at [http://www.federalreserve.gov/monetarypolicy/files/
FOMC_LongerRunGoals_20160126.pdf](http://www.federalreserve.gov/monetarypolicy/files/FOMC_LongerRunGoals_20160126.pdf).

600 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
605 Experiences in the U.S.: A Comprehensive Approach,” Federal Reserve Bank of San
Francisco Working Paper 2009-19.

Jaravel, Xavier, 2016, “The Unequal Gains from Product Innovations,” manuscript, Stanford
University.

- Johannsen, Benjamin K., 2014, “Inflation Experience and Inflation Expectations: Dispersion and Disagreement Within Demographic Groups,” Finance and Economics Discussion Series 2014-89, Board of Governors of the Federal Reserve System.
- 610
- Kaplan, Greg, and Guido Menzio, 2015, “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.
- 615
- 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*.
- 620
- Lucas, Robert E., Jr., 1972, “Expectations and the Neutrality of Money,” *Journal of Economic Theory* 4(2), 103–124.
- Lucas, Robert E., Jr., 1975, “An Equilibrium Model of the Business Cycle,” *Journal of Political Economy* 83(6), 1113–1144.
- 625
- Nimark, Kristoffer, 2008, “Dynamic Pricing and Imperfect Common Knowledge,” *Journal of Monetary Economics* 55(2), 365–382.
- Prais, S.J., 1959, “Whose Cost of Living?” *Review of Economic Studies* 26(2), 126–134.
- 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.
- 630
- 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(3), 665–690.

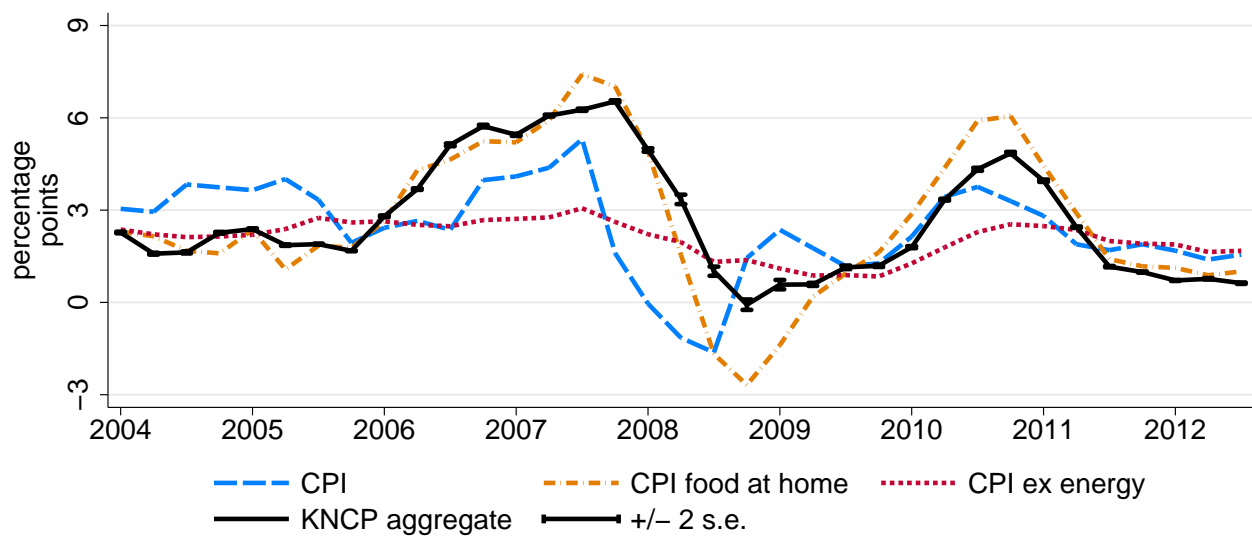


Figure 1: Aggregate inflation rates.

CPI inflation rates computed as annual percentage 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. Vertical bars show an interval of ± 2 bootstrap standard errors around each point estimate. Source for CPI data: Bureau of Labor Statistics.

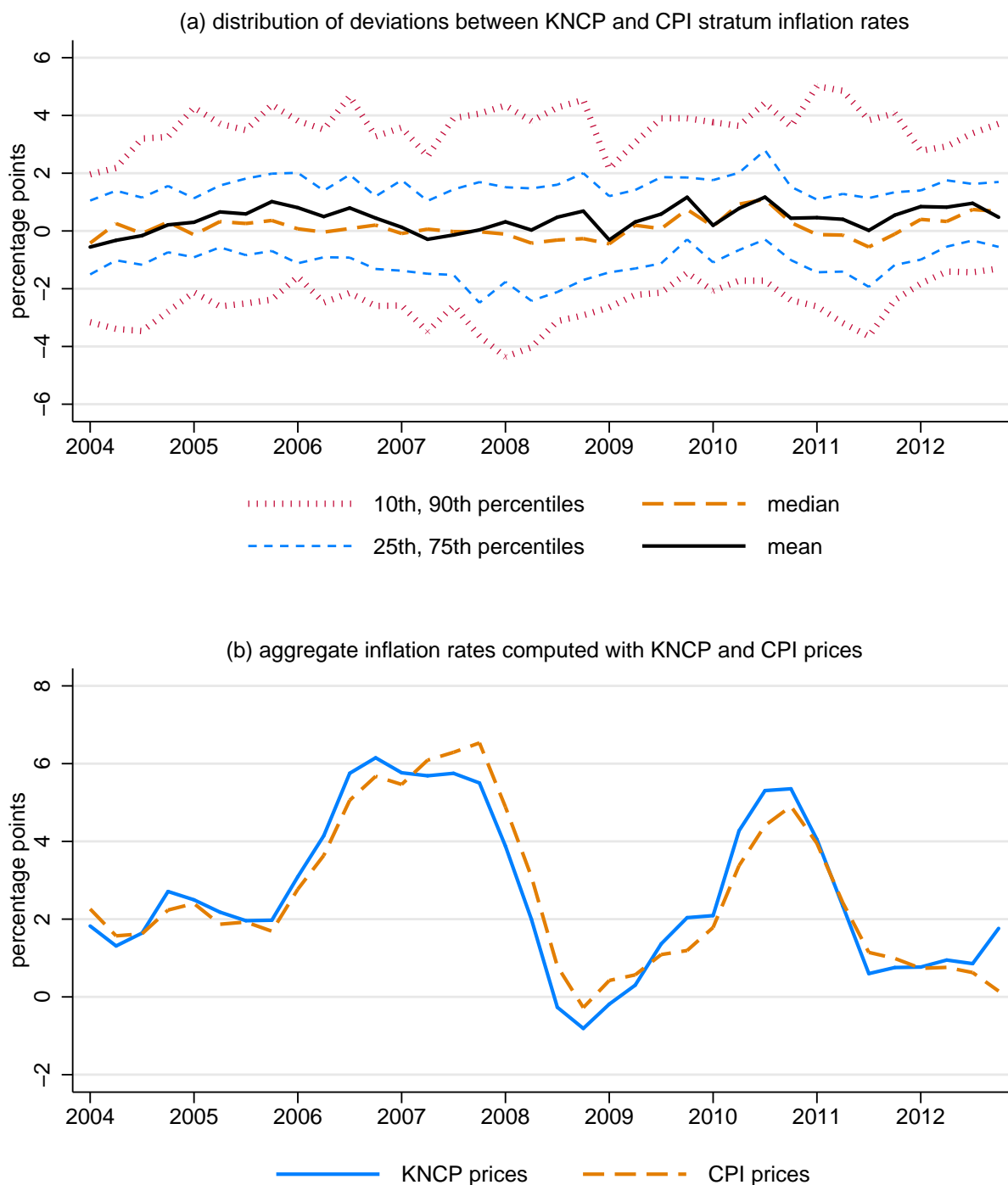


Figure 2: Comparison of KNCP and CPI inflation rates at the item stratum level. Panel (a) shows distribution across item strata of difference between inflation rate in KNCP data and that published for CPI. Panel (b) shows average inflation rates across all item strata using KNCP data and using published CPI indexes, weighted by distribution of spending in KNCP. CPI inflation rates are annual percent change in quarterly average of monthly indexes. Inflation rate plotted for each quarter is change in price index from that quarter to the quarter one year later.

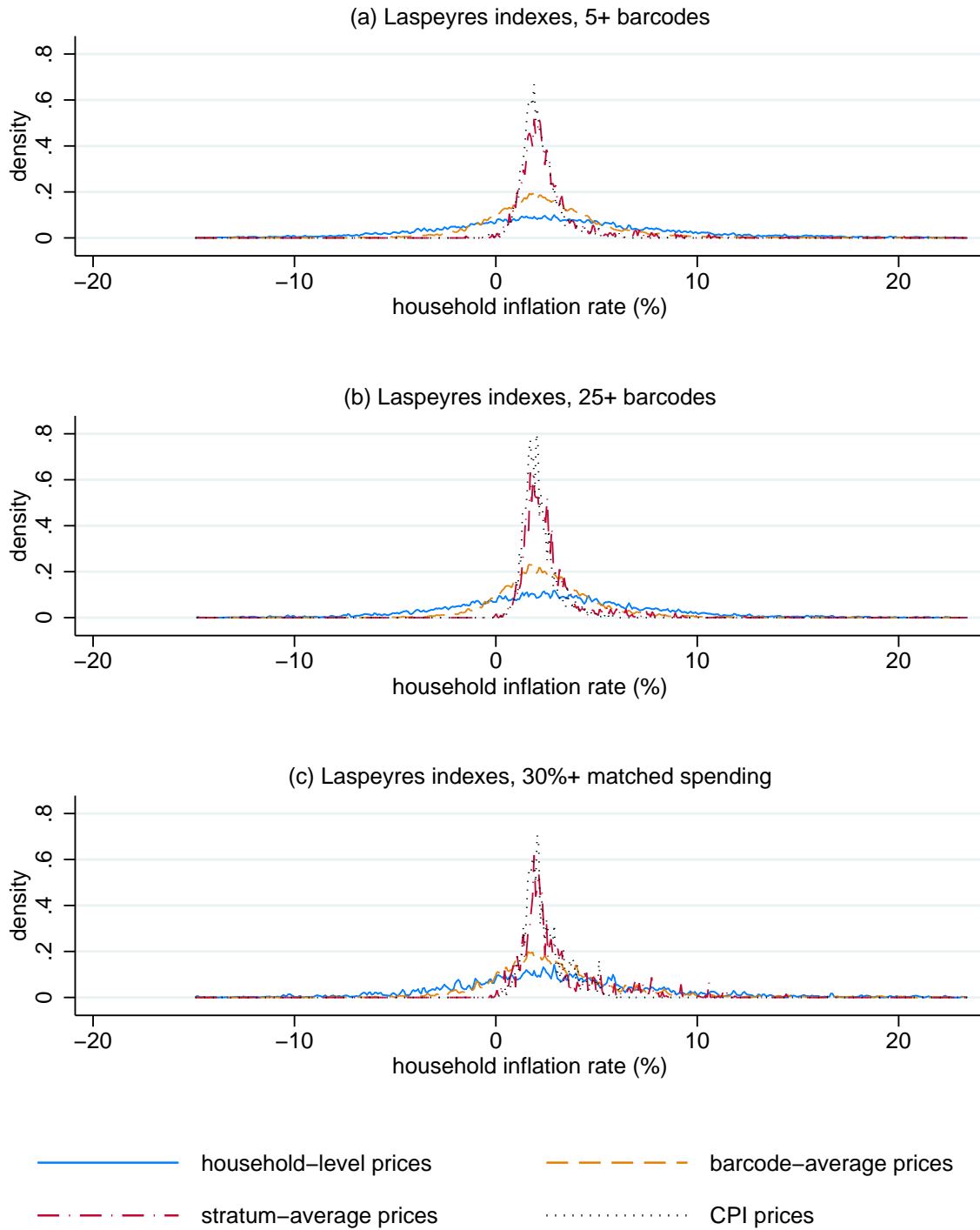


Figure 3: Distributions of household-level inflation rates, fourth quarter of 2004 to fourth quarter of 2005. Kernel density estimates using Epanechnikov kernel. Bandwidth is 0.05 percentage point for inflation rates with household-level and barcode-average prices and 0.005 percentage point for inflation rates with CPI prices. Data on 23,635 households with matched consumption in 2004q4 and 2005q4. Plots truncated at 1st and 99th percentiles of distribution of inflation rates with household-level prices.

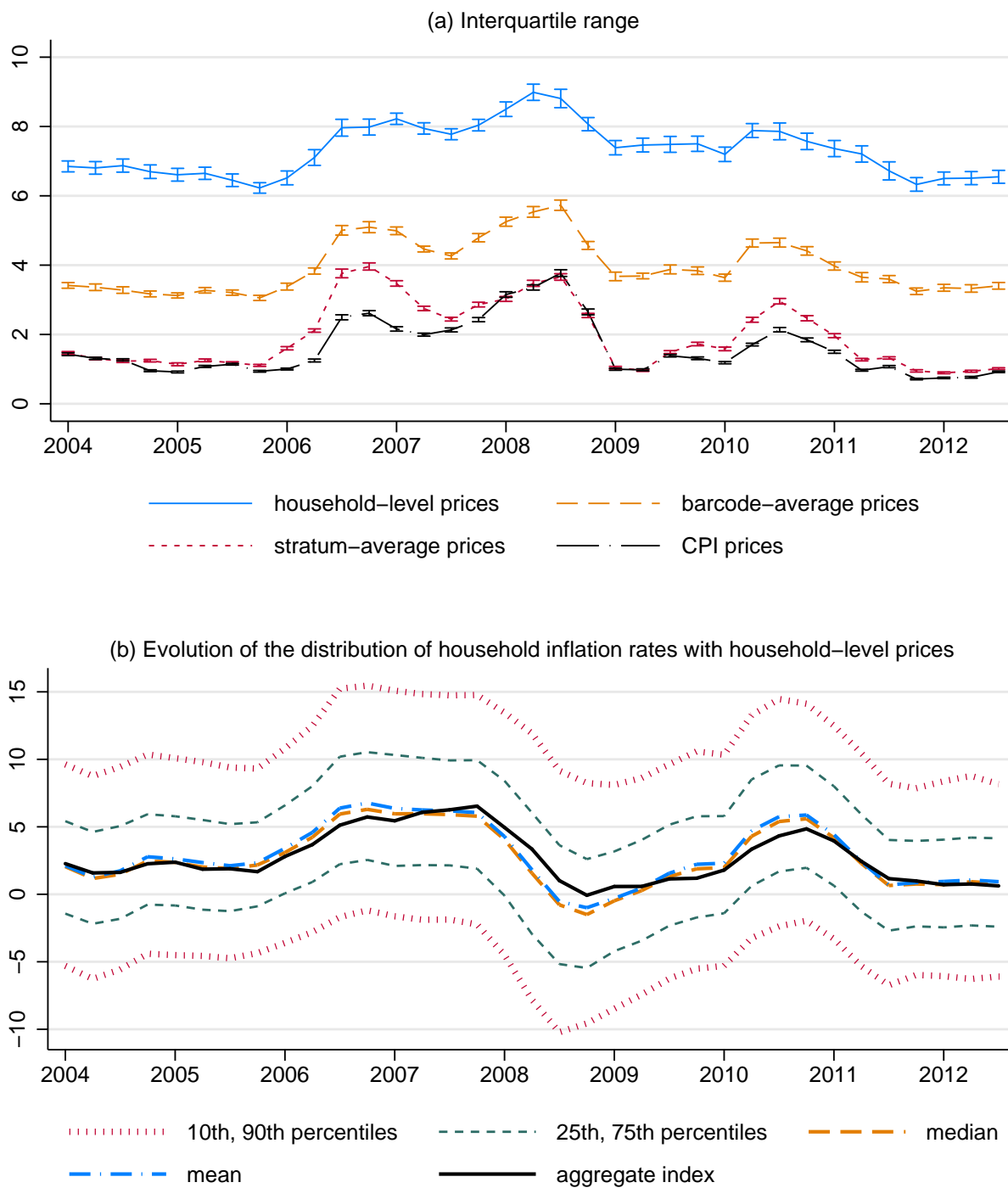


Figure 4: Measures of the dispersion of household-level inflation rates. Calculated with Laspeyres indexes. Vertical bars in panel (a) show an interval of ± 2 bootstrap standard errors around each point estimate. Mean in panel (b) is calculated on data from 1st to 99th percentiles of distribution of inflation rates with household-level prices at each date.

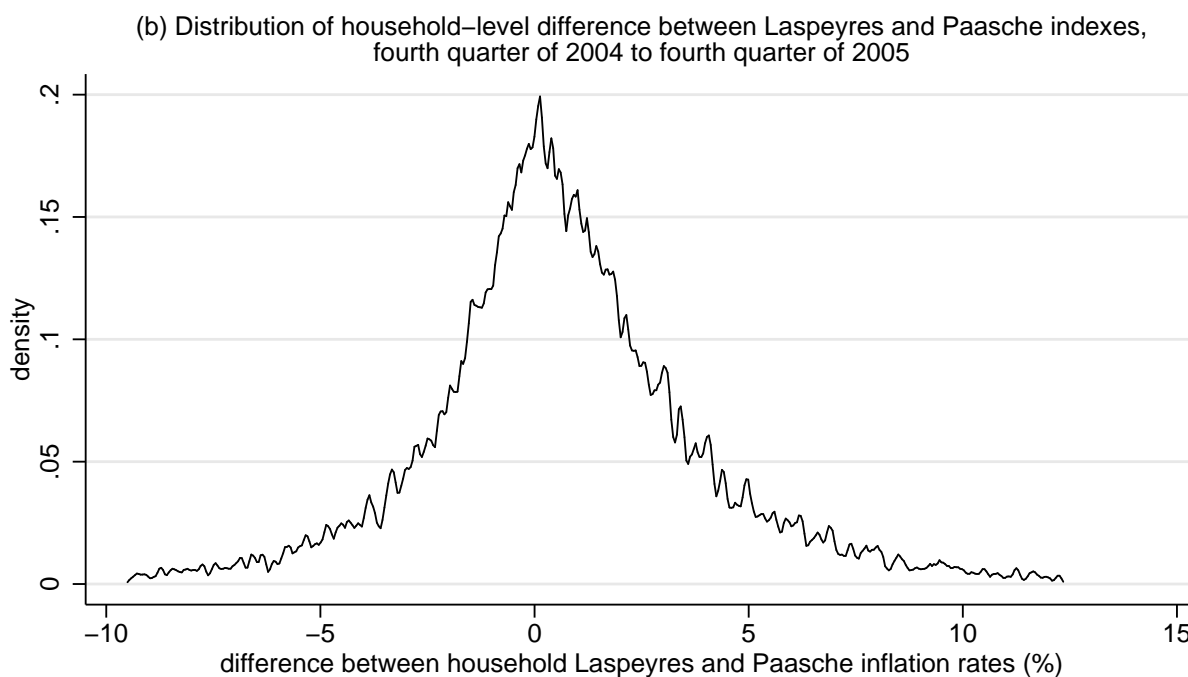
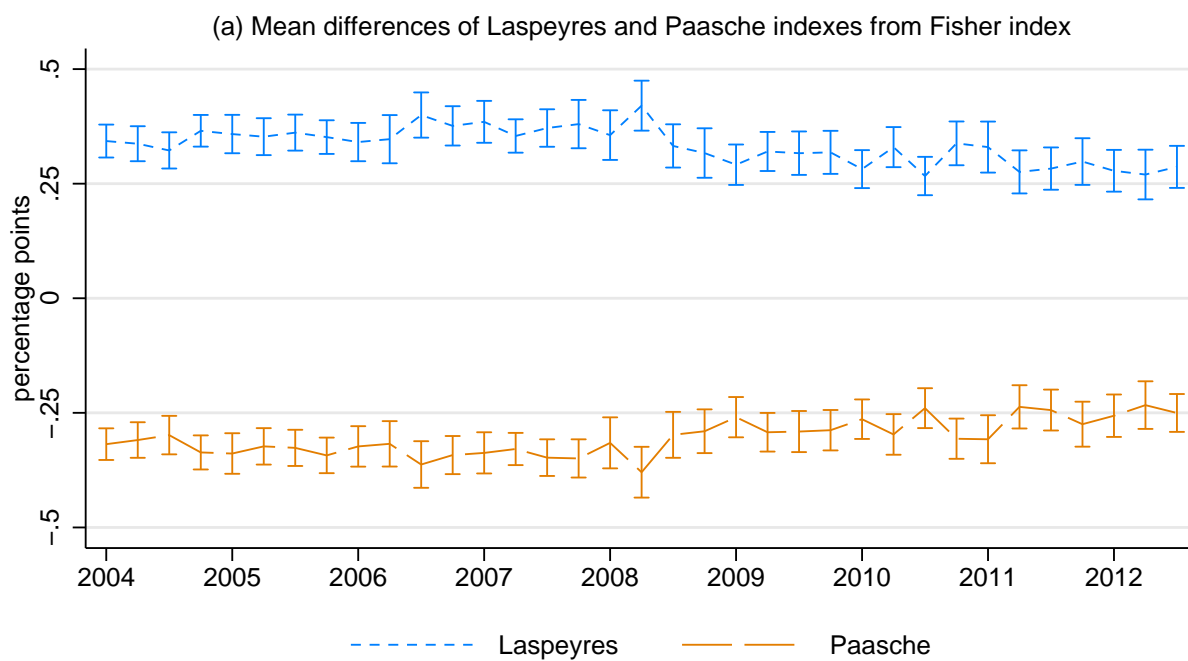


Figure 5: Difference between Laspeyres, Paasche, and Fisher indexes.

Vertical bars in panel (a) show an interval of ± 2 bootstrap standard errors around each point estimate. Panel (b) shows a kernel density estimate using Epanechnikov kernel and bandwidth of 0.05 percentage point; data are on 23,635 households with matched consumption in 2004q4 and 2005q4, and plot is truncated at 1st and 99th percentiles of distribution of inflation rates with household-level prices.

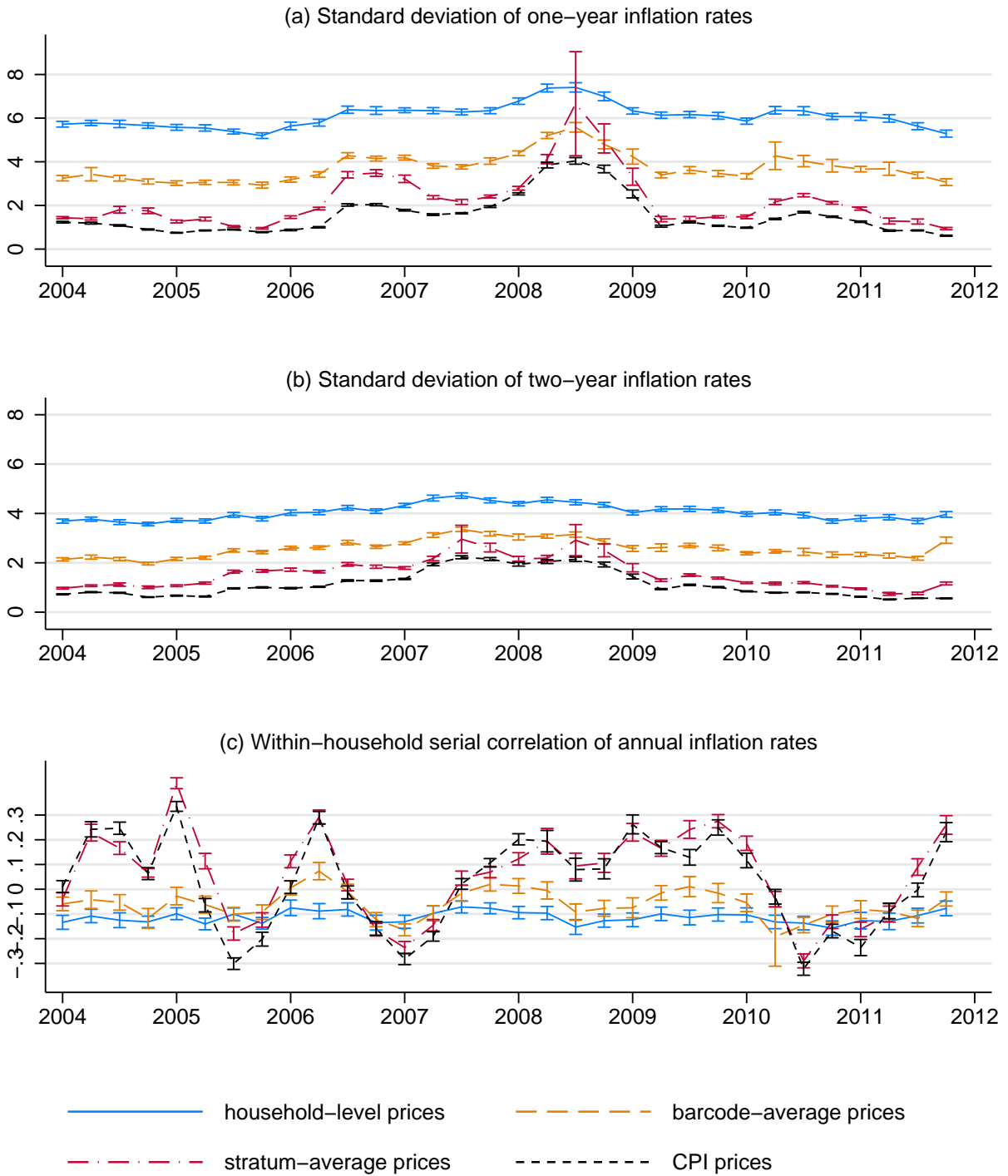


Figure 6: Evolution of the persistence of household-level inflation rates. Calculated with Laspeyres indexes. Calculations for each quarter use the subset of households for which inflation with household-level prices is observed and falls between the 1st and 99th percentiles of the distribution in both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

Table 1: 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
Stratum-average prices:				
Laspeyres	1.96	0.95	0.89	3.96
Fisher	1.83	0.88	0.91	3.84
Paasche	1.95	0.92	0.92	3.92
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
<hr/>				
<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
Stratum-average prices:				
Laspeyres	4.12	1.93	1.93	7.92
Fisher	3.78	1.75	1.92	7.99
Paasche	4.07	1.89	1.97	8.41
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
<hr/>				
<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
Stratum-average prices:				
Laspeyres	0.14	0.15	0.03	0.71
Fisher	0.10	0.09	0.02	0.41
Paasche	0.12	0.11	0.03	0.44
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.

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

Decile	(1) Aggregate index		(2) Median inflation	
	Slope	Intercept	Slope	Intercept
1	1.011 (0.015)	-7.602 (0.058)	0.997 (0.011)	-7.445 (0.044)
2	1.013 (0.009)	-4.609 (0.039)	0.973 (0.006)	-4.421 (0.022)
3	1.026 (0.008)	-2.810 (0.031)	0.966 (0.005)	-2.606 (0.019)
4	1.052 (0.008)	-1.448 (0.027)	0.978 (0.003)	-1.224 (0.013)
5	1.093 (0.007)	-0.264 (0.026)	1.000 (0.000)	0.000 (0.000)
6	1.137 (0.009)	0.944 (0.030)	1.030 (0.004)	1.242 (0.014)
7	1.198 (0.010)	2.286 (0.034)	1.073 (0.006)	2.641 (0.022)
8	1.243 (0.012)	4.189 (0.046)	1.100 (0.009)	4.595 (0.035)
9	1.305 (0.019)	7.491 (0.066)	1.112 (0.014)	8.036 (0.053)

The table shows the slope and intercept from quantile regressions of household-level inflation rates, computed with Laspeyres indexes and household-level prices, on measures of the overall inflation rate. In panel (1), the overall inflation rate is the aggregate CPI for the KNCP universe of goods. In panel (2), the overall inflation rate is the median of household inflation rates. Sample contains 835,386 household-quarter observations. Bootstrap standard errors are in parentheses.

Table 3: Cumulative inflation rates at different levels of household income.

	Household income					difference (<\$20,000 vs. ≥\$100,000)
	<\$20,000	\$20,000– \$39,999	\$40,000– \$59,999	\$60,000– \$99,999	≥\$100,000	
cumulative inflation (%) over 9 years ending in:						
2013q1	34.35 (0.90)	32.37 (0.58)	29.90 (0.60)	27.84 (0.55)	25.74 (0.65)	8.61 (1.10)
2013q2	33.25 (0.66)	31.11 (0.57)	28.26 (0.63)	25.86 (0.56)	24.23 (0.71)	9.02 (0.99)
2013q3	32.96 (0.86)	30.61 (0.64)	27.64 (0.60)	25.72 (0.60)	24.98 (0.63)	7.98 (1.08)
fraction of population	0.17	0.25	0.19	0.22	0.16	
average age of household head(s)						
<30	0.01	0.02	0.02	0.01	0.01	
30–39	0.08	0.1	0.14	0.16	0.13	
40–49	0.13	0.16	0.22	0.27	0.31	
50–59	0.22	0.21	0.26	0.29	0.34	
60–69	0.22	0.22	0.2	0.17	0.15	
≥70	0.34	0.29	0.16	0.09	0.07	
highest education of household head(s)						
less than high school	0.1	0.05	0.02	0.01	0	
high school diploma	0.47	0.44	0.33	0.2	0.09	
some college	0.31	0.33	0.35	0.32	0.23	
bachelor’s degree	0.1	0.14	0.22	0.31	0.37	
graduate degree	0.02	0.04	0.08	0.15	0.31	
Census region						
Northeast	0.19	0.17	0.17	0.19	0.21	
Midwest	0.22	0.23	0.24	0.23	0.21	
South	0.42	0.39	0.39	0.35	0.32	
West	0.18	0.2	0.21	0.23	0.26	
mean # household members	1.74	2.15	2.51	2.72	2.79	
has children	0.16	0.21	0.29	0.33	0.31	
white	0.81	0.8	0.79	0.78	0.79	
black	0.13	0.12	0.11	0.1	0.08	
Asian	0.01	0.01	0.02	0.03	0.06	
other nonwhite	0.05	0.07	0.08	0.08	0.07	
Hispanic	0.07	0.1	0.11	0.12	0.12	

Calculated with Laspeyres indexes and household-level prices. Bootstrap standard errors are in parentheses. Demographics are averages over the sample period.

Table 4: Regressions of household-level inflation rates on household demographics and shopping behavior.

Deviation of household inflation from aggregate inflation														
	(1) OLS, household prices		(2) Median, household prices		(3) Median, barcode prices		(4) Median, CPI prices		(5) Median household prices		(6) Interquartile range household prices		(7) Interquartile range household prices	
	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.
household income														
\$20,000-\$39,999	-0.206	(0.055)	-0.126	(0.039)	-0.093	(0.025)	-0.061	(0.010)	-0.136	(0.039)	-0.399	(0.079)	-0.338	(0.076)
\$40,000-\$59,999	-0.420	(0.052)	-0.258	(0.041)	-0.185	(0.027)	-0.077	(0.011)	-0.271	(0.039)	-0.597	(0.085)	-0.553	(0.085)
\$60,000-\$99,999	-0.587	(0.059)	-0.468	(0.045)	-0.267	(0.028)	-0.099	(0.013)	-0.479	(0.043)	-0.706	(0.086)	-0.655	(0.084)
≥\$100,000	-0.731	(0.065)	-0.597	(0.050)	-0.404	(0.035)	-0.124	(0.013)	-0.599	(0.047)	-0.873	(0.096)	-0.830	(0.097)
average age of household head(s)														
30-39	0.384	(0.205)	0.244	(0.145)	0.160	(0.089)	0.066	(0.028)	0.288	(0.162)	-0.775	(0.270)	-0.621	(0.268)
40-49	0.451	(0.203)	0.297	(0.137)	0.218	(0.090)	0.085	(0.025)	0.360	(0.152)	-1.031	(0.257)	-0.774	(0.260)
50-59	0.511	(0.206)	0.362	(0.137)	0.262	(0.090)	0.111	(0.027)	0.437	(0.156)	-1.187	(0.251)	-0.879	(0.255)
60-69	0.552	(0.199)	0.430	(0.139)	0.287	(0.088)	0.104	(0.027)	0.515	(0.156)	-1.293	(0.257)	-0.892	(0.263)
≥70	0.552	(0.203)	0.453	(0.140)	0.276	(0.090)	0.065	(0.027)	0.527	(0.158)	-1.214	(0.256)	-0.872	(0.258)
highest education of household head(s)														
high school diploma	-0.064	(0.127)	-0.029	(0.108)	-0.120	(0.054)	-0.062	(0.023)	-0.040	(0.103)	-0.280	(0.167)	-0.222	(0.172)
some college	-0.138	(0.127)	-0.102	(0.107)	-0.162	(0.054)	-0.092	(0.024)	-0.115	(0.102)	-0.118	(0.165)	-0.072	(0.169)
bachelor's degree	-0.251	(0.128)	-0.163	(0.110)	-0.238	(0.058)	-0.125	(0.025)	-0.169	(0.104)	-0.099	(0.180)	-0.028	(0.184)
graduate degree	-0.285	(0.139)	-0.137	(0.118)	-0.267	(0.063)	-0.141	(0.025)	-0.146	(0.114)	0.024	(0.185)	0.054	(0.188)
Census region														
Midwest	-0.179	(0.049)	-0.175	(0.040)	-0.029	(0.025)	-0.025	(0.009)	-0.169	(0.041)	-0.350	(0.084)	-0.380	(0.081)
South	-0.140	(0.046)	-0.063	(0.033)	0.041	(0.023)	0.006	(0.008)	-0.061	(0.034)	-1.018	(0.078)	-1.064	(0.078)
West	-0.517	(0.054)	-0.429	(0.046)	-0.254	(0.027)	-0.039	(0.011)	-0.430	(0.045)	-0.285	(0.091)	-0.310	(0.093)
# household members	0.089	(0.022)	0.110	(0.018)	0.058	(0.012)	0.026	(0.005)	0.109	(0.018)	-0.363	(0.030)	-0.313	(0.032)
has children	-0.115	(0.118)	0.074	(0.094)	-0.153	(0.066)	0.048	(0.024)	0.066	(0.100)	-0.932	(0.199)	-0.838	(0.190)
has children × # household members	0.002	(0.034)	-0.033	(0.029)	0.012	(0.020)	-0.029	(0.007)	-0.032	(0.030)	0.281	(0.054)	0.234	(0.053)
black	0.058	(0.064)	0.063	(0.049)	-0.019	(0.027)	-0.061	(0.011)	0.074	(0.048)	0.825	(0.099)	0.907	(0.095)
Asian	-0.074	(0.156)	-0.089	(0.125)	-0.013	(0.070)	-0.010	(0.019)	-0.072	(0.136)	1.442	(0.213)	1.535	(0.222)
other nonwhite	-0.033	(0.087)	-0.025	(0.067)	0.053	(0.045)	0.007	(0.017)	-0.035	(0.070)	0.317	(0.141)	0.337	(0.137)
Hispanic	-0.068	(0.068)	-0.085	(0.056)	-0.092	(0.036)	-0.010	(0.013)	-0.065	(0.060)	-0.110	(0.094)	-0.114	(0.097)
log(# of shopping trips) initial quarter														
final quarter														
R^2	0.012		-		-		-		-		-		-	
R^2 (time dummies only)	0.009		-		-		-		-		-		-	

Dependent variable is difference between household inflation rate and aggregate inflation rate for the equivalent universe of goods. Sample contains 835,386 household-
quarter observations. Bootstrap standard errors in parentheses. In columns (1), (2), (5), (6), and (7), household inflation rate is computed with household-level prices;
column (3) uses barcode-average prices, and column (4) uses CPI prices. All columns use Laspeyres indexes. Column (1) shows results from ordinary least squares
regression; columns (2), (3), (4), and (5) from median regression; and columns (6) and (7) from interquartile range regression. Regressions include time dummy variables.
Omitted categories are: income less than \$20,000; white; non-Hispanic; heads' average age less than 30; heads' highest education less than high school diploma; Northeast
region. Count of shopping trips excludes trips on which the household spent less than \$1.