Inflation at the Household Level: Online Appendix*

Greg Kaplan and Sam Schulhofer-Wohl

Revised August 2017

ABSTRACT

This appendix contains additional results on using scanner data to estimate inflation rates at the household level. There are seven sections. Section 1 provides details on the distribution of spending across types of goods in the KNCP. Section 2 shows cross-sectional distributions of Fisher and Paasche inflation rates and the full distribution of Laspyeres inflation rates. Section 3 shows the evolution over time of measures of dispersion of Fisher and Paasche inflation rates. Section 4 exhibits differences in mean inflation rates by income. Section 5 examines the relationship between household demographics and substitution patterns. Section 6 shows cross-sectional distributions of two-year inflation rates. Section 7 investigates how changes in a household's consumption bundle over time affect the estimated serial correlation of inflation rates.

Keywords: Inflation; Heterogeneity
JEL classification: D12, D30, E31

---

*Kaplan: University of Chicago and National Bureau of Economic Research (gkaplan@uchicago.edu). Schulhofer-Wohl: Federal Reserve Bank of Chicago (sam@frbchi.org). We thank Cristian Alonso and Emily Moschini for research assistance, Joan Gieseke for editorial assistance, and our discussants Yuriy Gorodnichenko, Bart Hobijn, Peter Klenow, and Sarah Lein as well as the editor, associate editor, referee, and 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.

1. Distribution of expenditure across categories in the KNCP

Table 1 compares the distribution of spending across types of goods in the KNCP with the weights used to construct the published CPI. All of the data in the table are for 2012.

The first column of the table shows the weights for the CPI for urban consumers, while the second column shows the distribution of spending in the Bureau of Labor Statistics' Consumer Expenditure Survey (CEX) for that year. The CEX distribution differs from the CPI weights because not all CEX households are urban.

The third column of the table shows the distribution of spending across all purchases in the KNCP data, and the fourth column considers only the purchases that we use to construct our household inflation rates — barcodes that a household purchases in both quarter t and quarter $t+4$, from households with at least five matched barcodes. 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 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, on the other hand, gets much less weight in our data than in the CPI, primarily because shelter, which has a 32 percent expenditure share in the CPI, is not measured in the KNCP. Similarly, the KNCP measures very little transportation spending. Apparel is measured in the KNCP, but we observe no purchases of matched apparel barcodes in consecutive periods, so apparel gets zero weight in our household inflation rates.

The fifth column shows the distribution of spending for the three-period fixed-basket sample considered in section 7; this sample consists of barcodes that a household buys in all of quarters t, $t+4$, and $t+8$, from households with at least five such matched barcodes. The distribution of spending is similar to that in the baseline sample in the fourth column, but the three-period fixed-basket sample includes less than half as much spending — measured in dollars, number of purchases, or unique UPCs — as the baseline sample.

Table 1: Percentage distribution of spending across categories in different datasets, 2012.

	(1)	(2)	(3)	(4) KNCP	(5)
	CPI-U	CEX	all spending	5+ matched UPCs	3 periods 5+ matched UPCs
Food and beverages	15.26	16.03	61.22	74.38	75.77
Food	14.31	15	58.08	67.61	67.16
Food at home	8.6	8.91	53.87	64.77	64.72
Cereals and bakery products	1.23	1.22	7.71	9.1	8.07
Cereals and cereal products	0.47	0.41	2.91	3.25	2.71
Bakery products	0.76	0.81	4.8	5.86	5.37
Meats, poultry, fish, and eggs	1.96	1.94	7.53	6.35	7.47
Meats, poultry, and fish	1.84	1.82	6.98	4.93	5.82
Eggs	0.11	0.12	0.55	1.42	1.65
Dairy and related products	0.91	0.95	7.92	13.15	14.81
Fruits and vegetables	1.29	1.66	7.36	6.9	6.38
Nonalcoholic beverages, beverage materials	0.94	0.84	6.85	13.52	15.18
Other food at home	2.28	2.3	14.84	15.75	12.81
Sugar and sweets	0.31	0.33	2.95	3.05	2.53
Fats and oils	0.26	0.26	1.58	2.4	2.52
Other foods	1.71	1.7	10.31	10.3	7.77
Food away from home	5.71	6.09	4.22	2.83	2.44
Alcoholic beverages	0.95	1.03	3.13	6.78	8.60
Housing	41.02	35.63	9.03	5.11	3.48
Shelter	31.68	22.52	-	-	-
Fuels and utilities	5.3	5.49	0.08	0.08	0.05
Household furnishings and operations	4.04	7.62	8.95	5.04	3.43
Window and floor coverings and other linens	0.27	0.32	-	-	-
Furniture and bedding	0.71	0.89	-	-	-
Appliances	0.29	0.67	1.17	0.09	0.03
Other household equipment and furnishings	0.48	-	1.07	0.14	0.11
Tools, hardware, outdoor equipment, supplies	0.68	-	1.09	0.21	0.09
Housekeeping supplies	0.89	1.39	5.62	4.59	3.19
Household operations	0.73	2.64	-	-	-
Apparel	3.56	3.95	8.4	-	-
Transportation	16.85	20.48	0.22	0.14	0.10
Private transportation	15.66	19.25	0.22	0.14	0.10
Public transportation	1.19	1.23	-	-	-
Medical care	7.16	8.1	6.92	4.85	3.56
Recreation	5.99	6.18	6.57	5.85	5.12
Video and audio	1.9	2.23	2.11	0.47	0.22
Pets, pet products and services	1.1	-	4.27	5.37	4.90
Sporting goods	0.46	-	-	-	-
Photography	0.11	-	0.16	0.01	0.01
Other recreational goods	0.45	-	0.02	-	0.00
Other recreation services	1.75	-	-	-	-
Recreational reading materials	0.23	-	-	-	-
Education and communication	6.78	5.57	-	-	-
Other goods and services	3.38	4.07	7.64	9.67	11.98
Tobacco and smoking products	0.81	0.76	1.87	6.46	9.46
Personal care	2.57	1.43	4.47	2.66	1.86
Total spending ($'000,000s)			291.7	19.3	8.8
Number of purchases ('000s)				4,507	2,052
Number of unique UPCs ('000s)				170	84

Subcategories are not exhaustive and do not necessarily add up to higher-level categories.

2. Cross-sectional distributions of Fisher and Paasche inflation rates

This section presents cross-sectional distributions of Fisher and Paasche inflation rates, similar to the distributions shown for Laspeyres indexes in Figure 3 of the main paper, as well as a version of Figure 3 with extended axes to show nearly the full distribution of Laspeyres indexes.

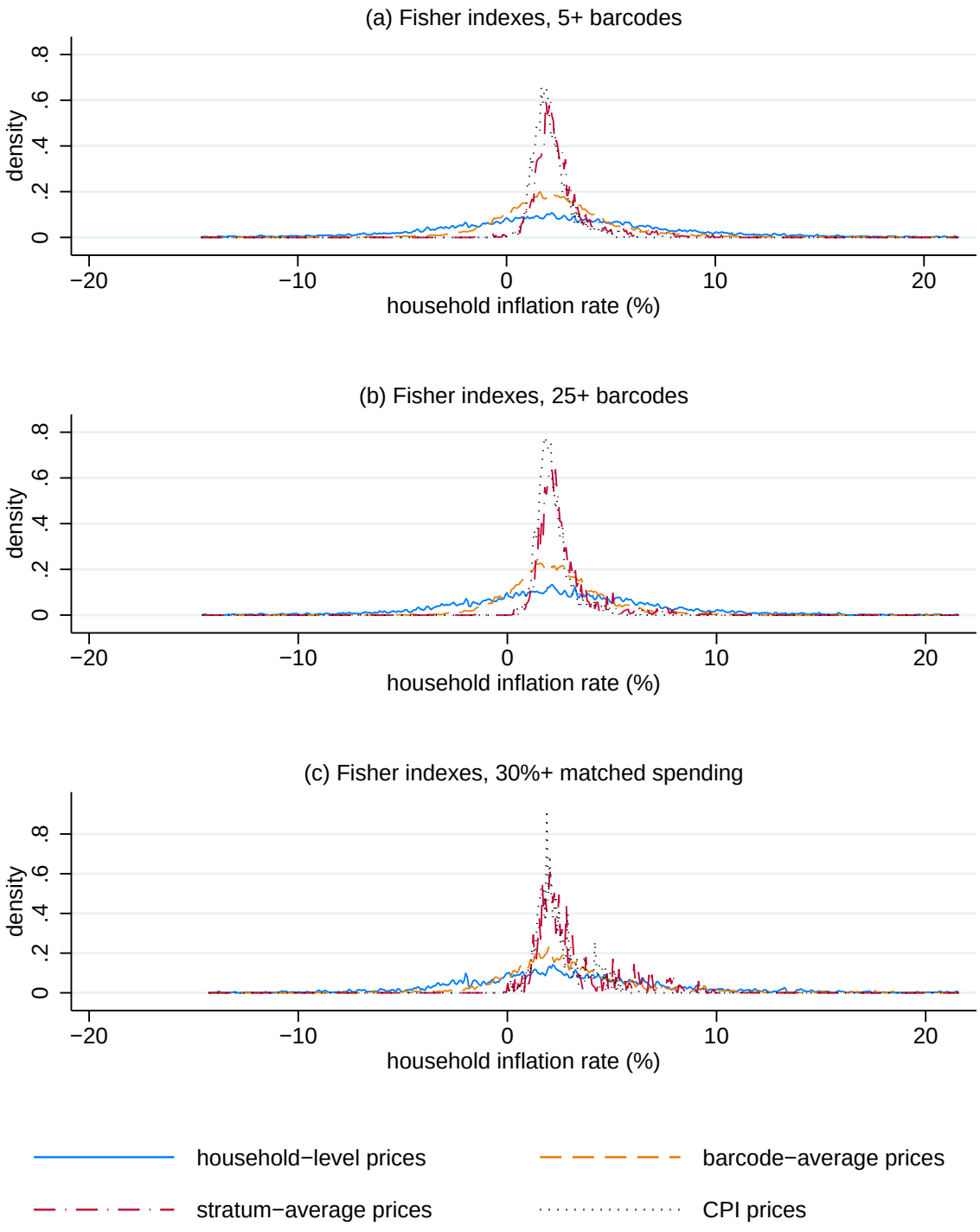

Figure 1: Distributions of household-level inflation rates (Fisher indexes) from fourth quarter of 2004 to fourth quarter of 2005.

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 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.

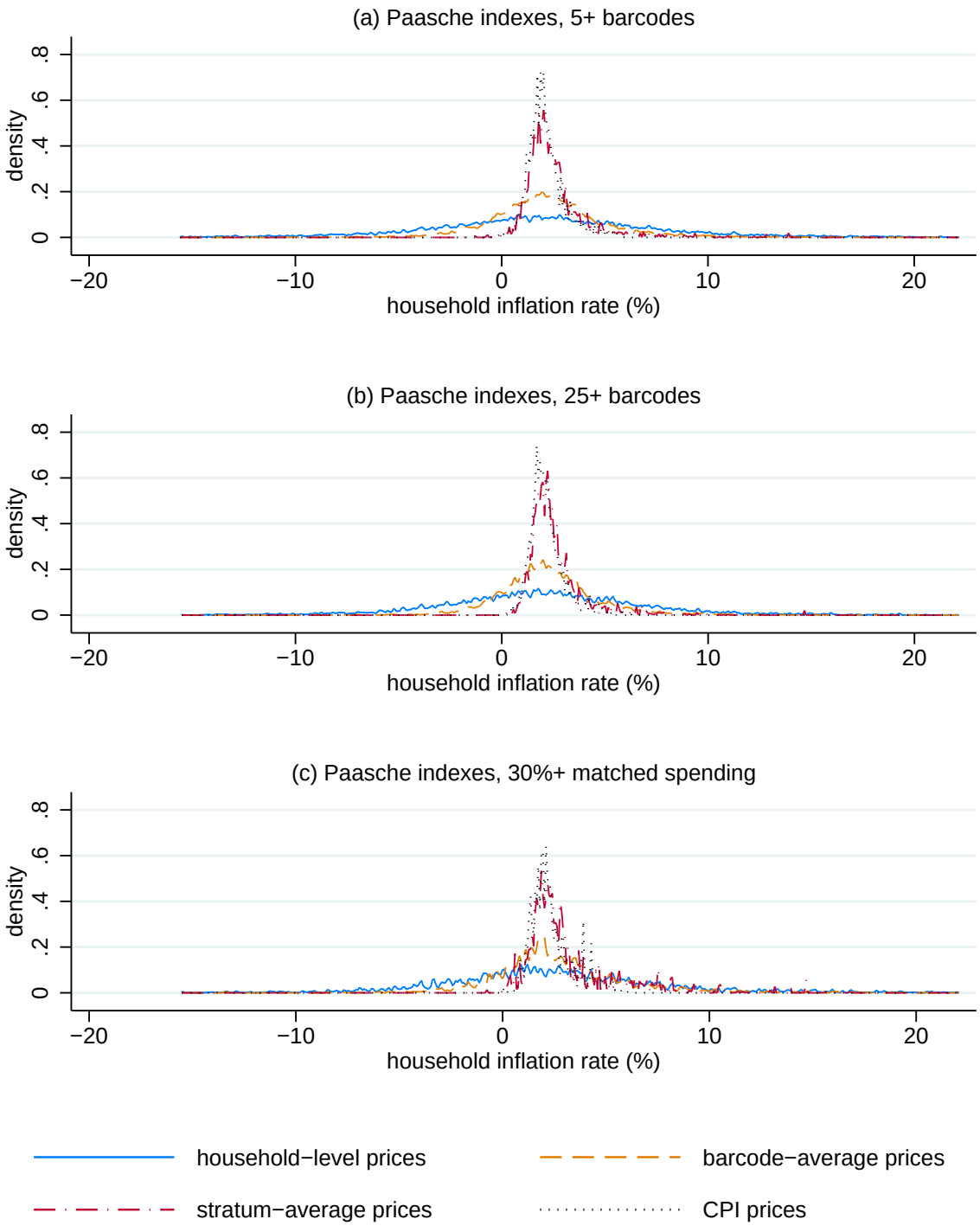

Figure 2: Distributions of household-level inflation rates (Paasche indexes) from fourth quarter of 2004 to fourth quarter of 2005.

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 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.

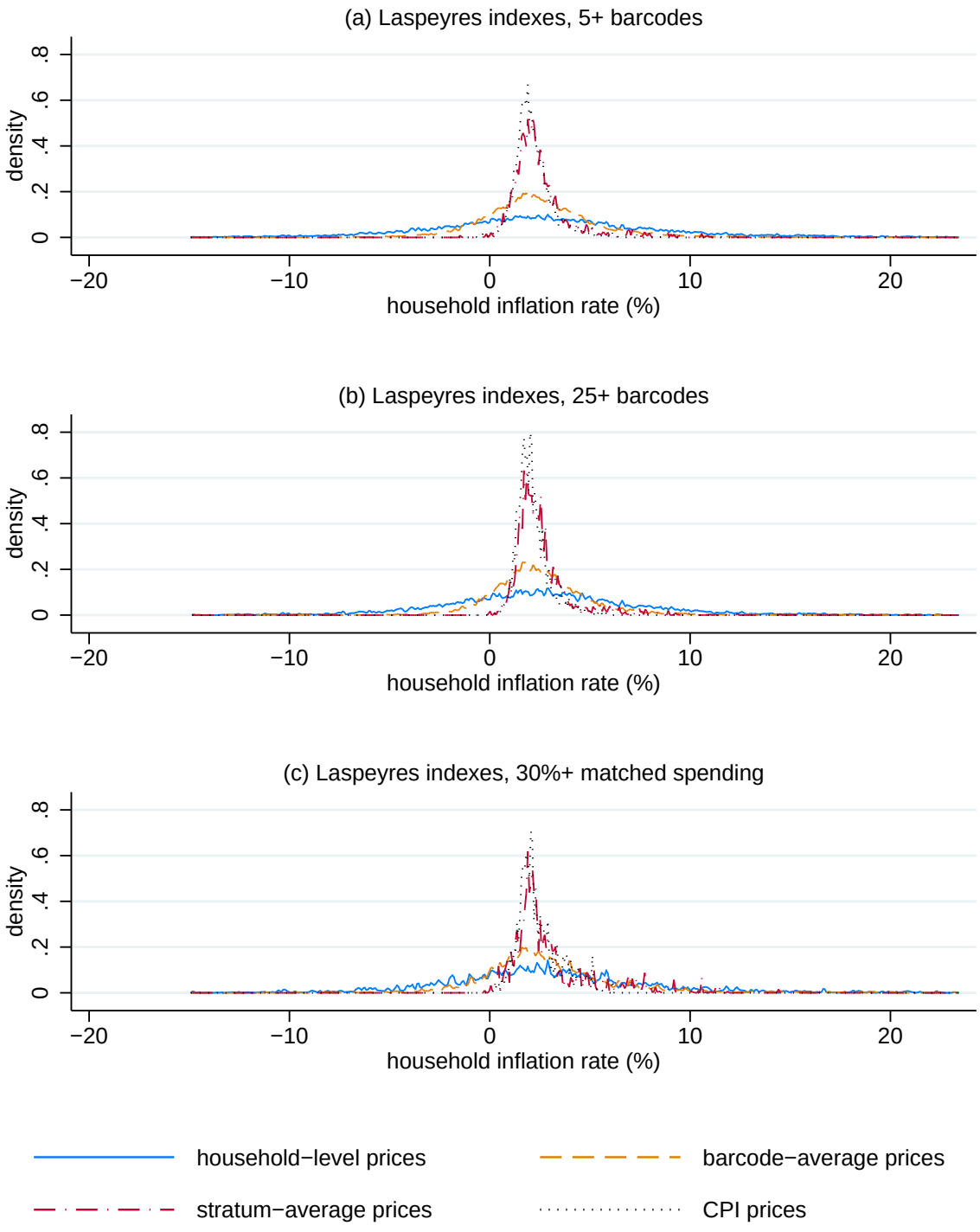

Figure 3: Distributions of household-level inflation rates (Laspeyres indexes), fourth quarter of 2004 to fourth quarter of 2005.

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.

3. Evolution over time of dispersion in inflation rates

This section presents time series of dispersion measures for Fisher and Paasche inflation rates, similar to the time series shown for Laspeyres indexes in Figure 4 of the main paper.

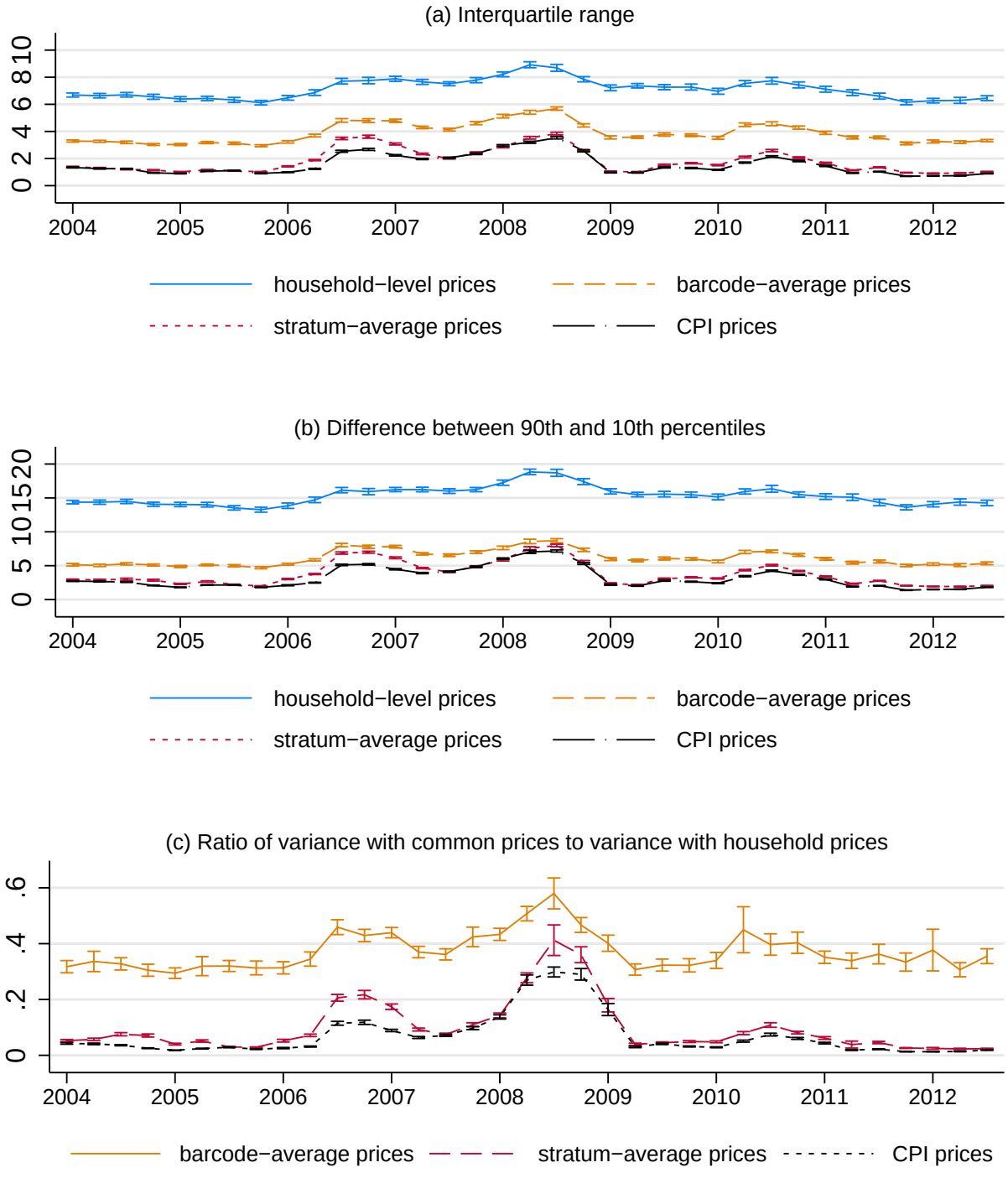

Figure 4: Measures of the dispersion of household-level inflation rates (Fisher indexes). Vertical bars show an interval of ± 2 bootstrap standard errors around each point estimate. In panel (c), variances are calculated on data from 1st to 99th percentiles of distribution of inflation rates with household-level prices at each date.

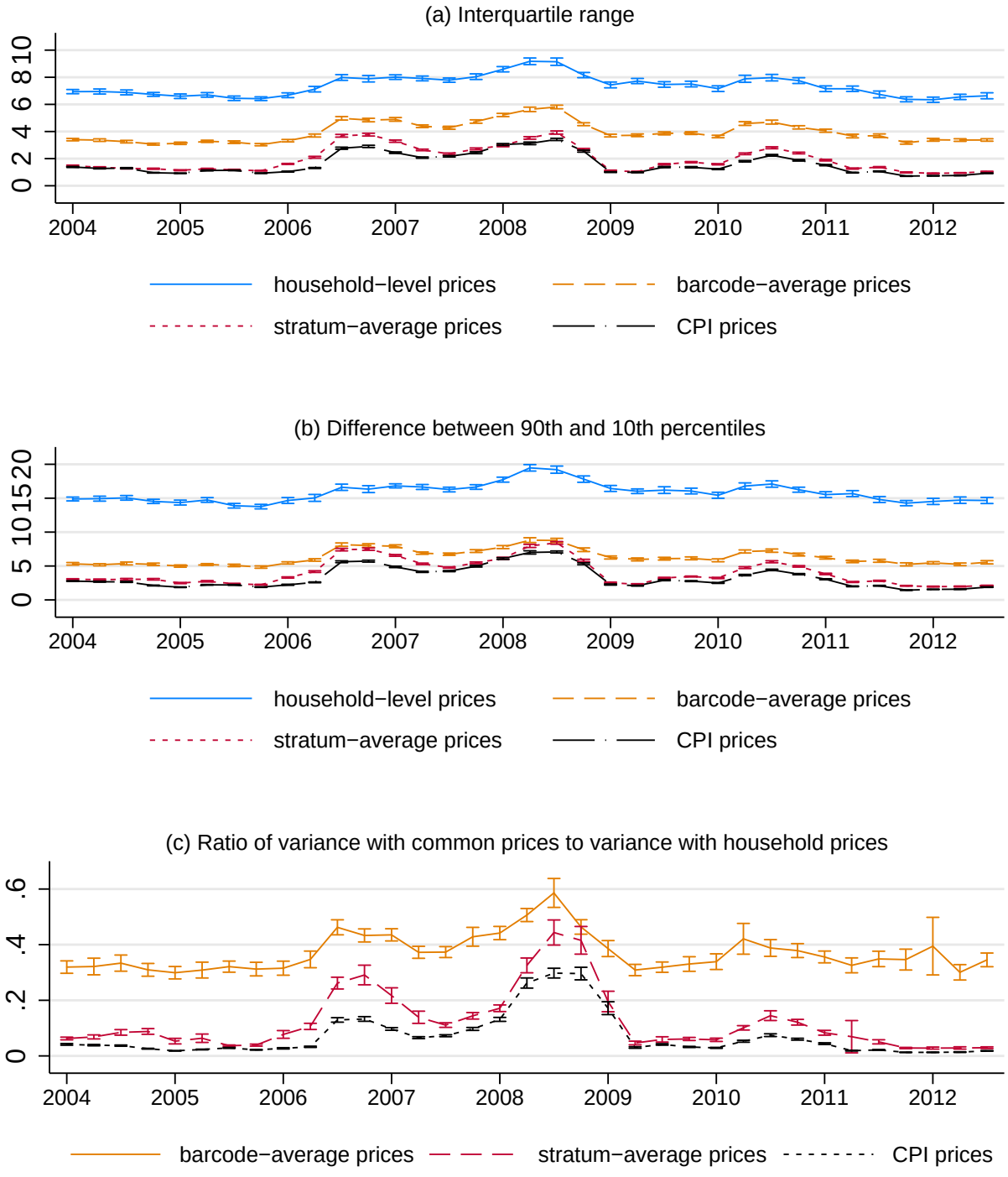

Figure 5: Measures of the dispersion of household-level inflation rates (Paasche indexes). Vertical bars show an interval of ± 2 bootstrap standard errors around each point estimate. In panel (c), variances are calculated on data from 1st to 99th percentiles of distribution of inflation rates with household-level prices at each date.

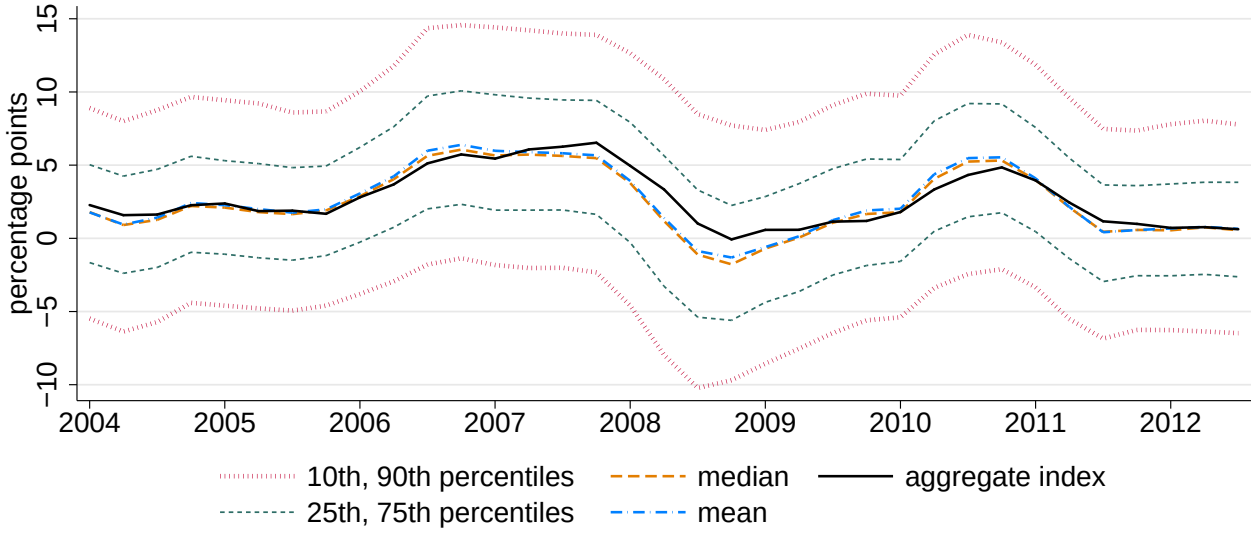

Figure 6: Evolution of the distribution of household inflation rates (Fisher indexes) with household-level prices.
Mean is calculated on data from 1st to 99th percentiles of distribution of inflation rates at each date.

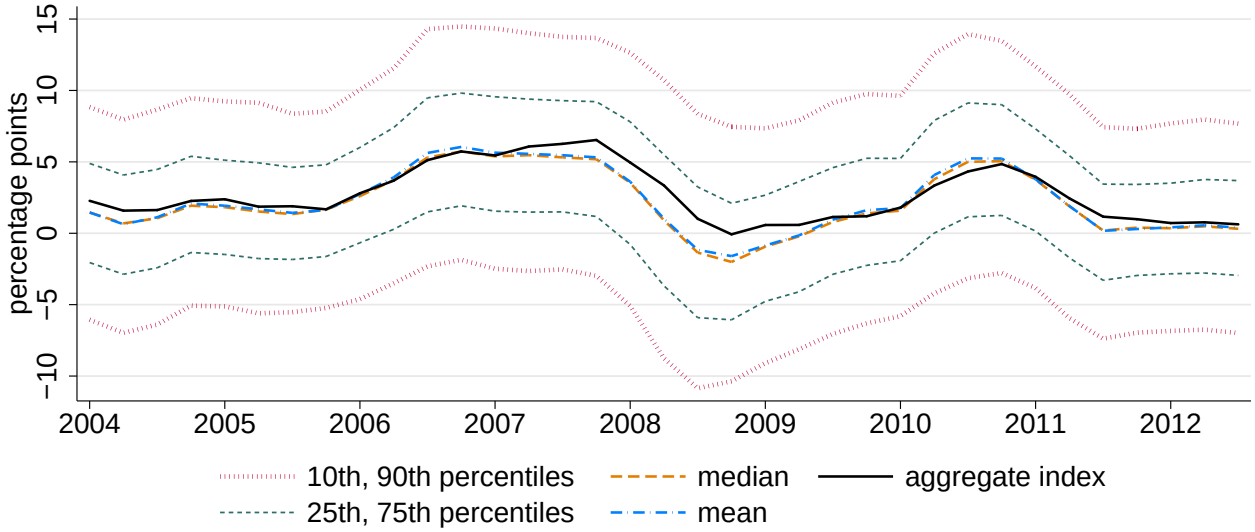

Figure 7: Evolution of the distribution of household inflation rates (Paasche indexes) with household-level prices.
Mean is calculated on data from 1st to 99th percentiles of distribution of inflation rates at each date.

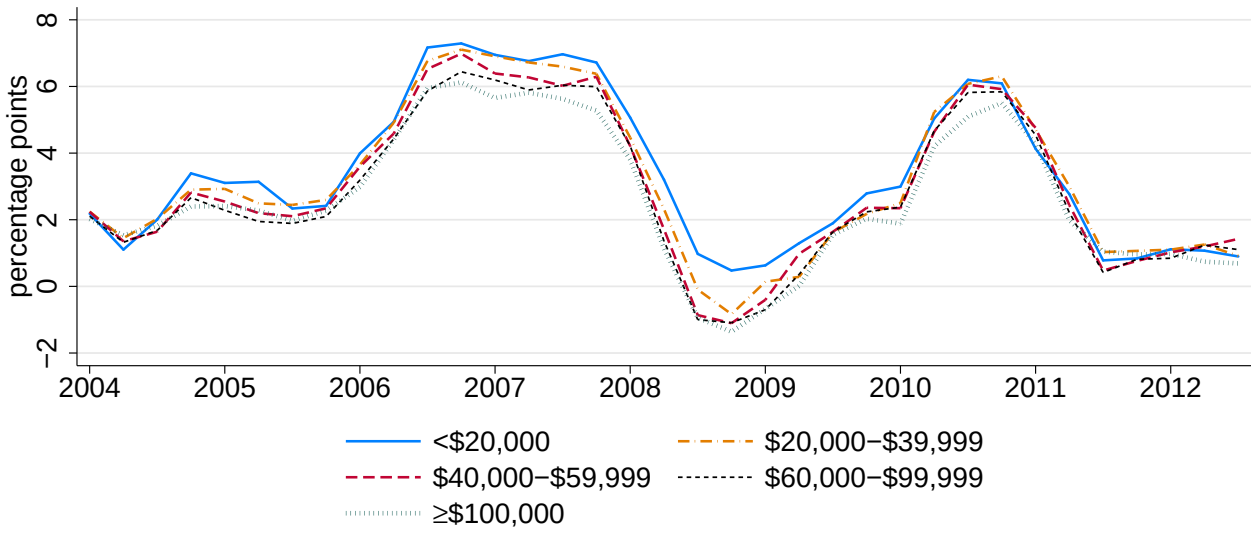

Figure 8: Mean inflation rates at different levels of household income.
Calculated with Laspeyres indexes and household-level prices.

4. Inflation rates and household income

Figure 8 shows the mean inflation rates among households at different income levels. Inflation is consistently higher for lower-income households — so much so that, even as the depth of the Great Recession produced widespread deflation, households with incomes below $20,000 still had a positive inflation rate.

5. Household demographics and substitution patterns

Table 2 measures the relationship between household demographics and substitution patterns. We use ordinary least squares and median regressions to examine the association of the Laspeyres-Paasche difference with household demographics, and a linear probability model to examine how demographics relate to the probability that a household's Laspeyres inflation rate is greater than its Paasche inflation rate. These regressions use the data for all quarters and control for time effects. The largest effects are found for age, income, and household size. Households with heads between ages 40 and 70 have an average Laspeyres-Paasche difference about 0.2 percentage point larger than households with heads between ages 20 and 29; this is substantial relative to the mean difference of 0.6 percentage point. Households with children also show stronger substitution, as do those with relatively low, but not the lowest, incomes. Nonetheless, as with household-level inflation rates themselves, the low R-squared in the regressions shows that demographics have almost no power to explain differences between households' Laspeyres and Paasche inflation rates.

Table 2: Regressions of household-level difference between Laspeyres and Paasche inflation rates on household demographics.

	$\pi^L_{it,t+4} - \pi^P_{it,t+4}$				$\pi^L_{it,t+4} > \pi^P_{it,t+4}$	
	(1) OLS		(2) Median		(3) OLS	
	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.
household income						
\$20,000–\$39,999	0.095	(0.029)	0.074	(0.022)	0.013	(0.003)
\$40,000–\$59,999	0.031	(0.031)	0.036	(0.022)	0.008	(0.004)
\$60,000–\$99,999	0.001	(0.035)	0.013	(0.024)	0.006	(0.004)
≥\$100,000	-0.004	(0.040)	0.007	(0.031)	0.007	(0.004)
average age of household head(s)						
30–39	0.119	(0.097)	0.103	(0.055)	0.017	(0.009)
40–49	0.196	(0.096)	0.166	(0.051)	0.030	(0.008)
50–59	0.228	(0.097)	0.184	(0.053)	0.030	(0.008)
60–69	0.195	(0.098)	0.175	(0.054)	0.030	(0.009)
≥70	0.100	(0.098)	0.116	(0.054)	0.020	(0.009)
highest education of household head(s)						
high school diploma	0.056	(0.062)	0.048	(0.048)	0.008	(0.007)
some college	0.051	(0.062)	0.036	(0.047)	0.003	(0.007)
bachelor's degree	0.008	(0.065)	0.025	(0.051)	0.001	(0.008)
graduate degree	0.042	(0.066)	0.023	(0.054)	0.001	(0.008)
Census region						
Midwest	0.054	(0.029)	0.049	(0.020)	0.010	(0.003)
South	-0.104	(0.025)	-0.086	(0.017)	-0.003	(0.003)
West	0.026	(0.032)	-0.001	(0.024)	0.004	(0.004)
# household members	0.047	(0.013)	0.047	(0.009)	0.007	(0.001)
has children	0.132	(0.068)	0.177	(0.053)	0.018	(0.007)
has children × # household members	-0.036	(0.021)	-0.038	(0.015)	-0.004	(0.002)
black	0.046	(0.039)	0.017	(0.025)	-0.005	(0.004)
Asian	-0.137	(0.064)	-0.069	(0.053)	-0.022	(0.007)
other nonwhite	0.050	(0.055)	0.013	(0.032)	0.001	(0.005)
Hispanic	-0.046	(0.036)	-0.019	(0.023)	-0.002	(0.004)
R^2	0.0010				0.0011	
R^2 (time dummies only)	0.0004				0.0003	
N	835,386		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 are 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 \$20,000; white; non-Hispanic; heads' average age less than 30; heads' highest education less than high school diploma; Northeast region.

6. Cross-sectional distributions of two-year inflation rates

This section presents cross-sectional distributions of two-year inflation rates for an illustrative time period, the fourth quarter of 2004 through the fourth quarter of 2006. The distributions are presented with Laspeyres, Fisher, and Paasche indexes. Each figure shows the distributions of household inflation rates in 2004–2005 and in 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 distributions of inflation rates for the two one-year periods are similar, whereas the annualized two-year inflation rates are somewhat less dispersed but still very heterogeneous.

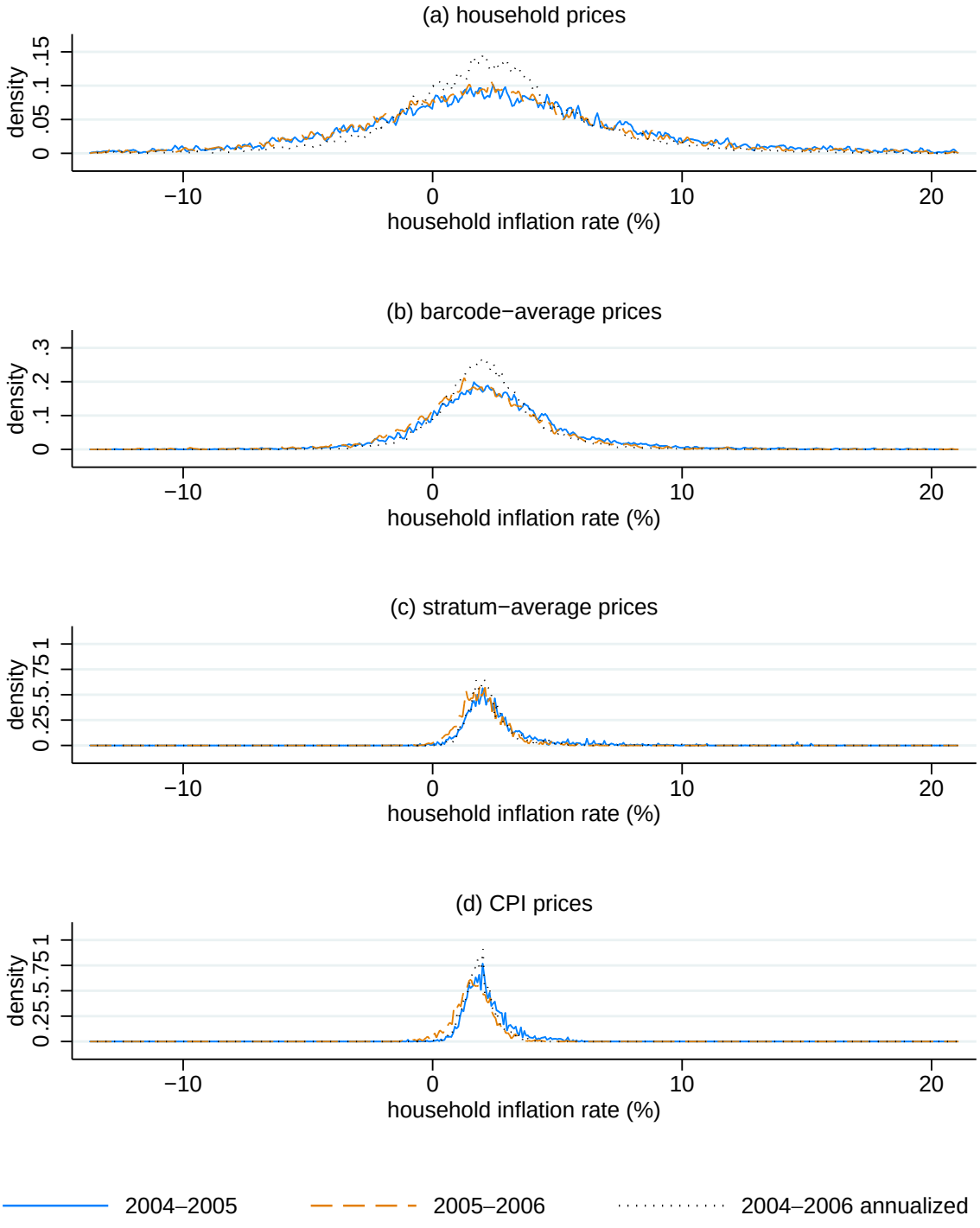

Figure 9: Distributions of one-year and two-year household-level inflation rates, 2004q4–2005q4 and 2005q4–2006q4.

Calculated with Laspeyres indexes. 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. Sample limited to 19,252 households with inflation rates calculated for both 2004q4–2005q4 and 2005q4–2006q4.

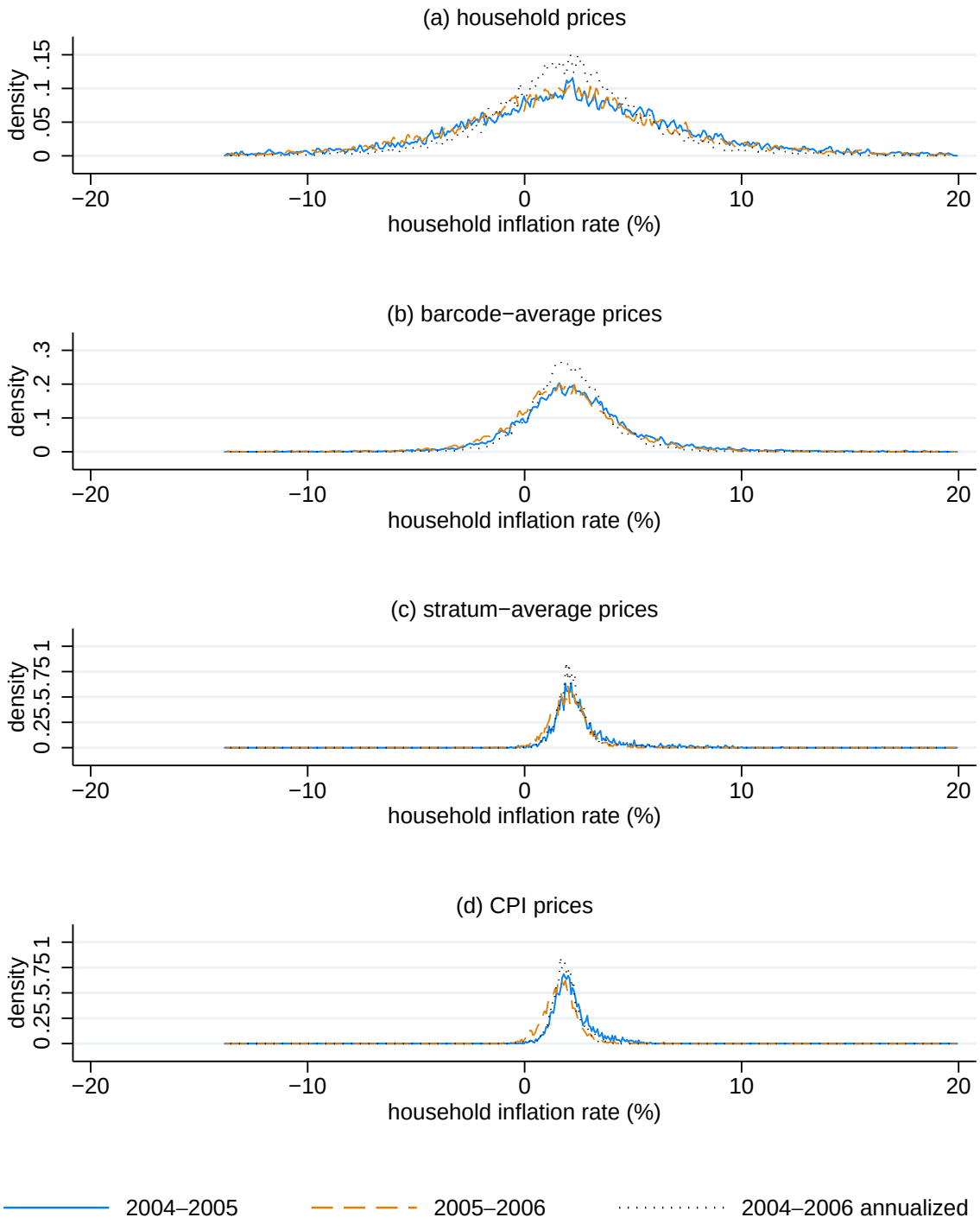

Figure 10: Distributions of one-year and two-year household-level inflation rates, 2004q4–2005q4 and 2005q4–2006q4.

Calculated with Fisher indexes. 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. Sample limited to 19,252 households with inflation rates calculated for both 2004q4–2005q4 and 2005q4–2006q4.

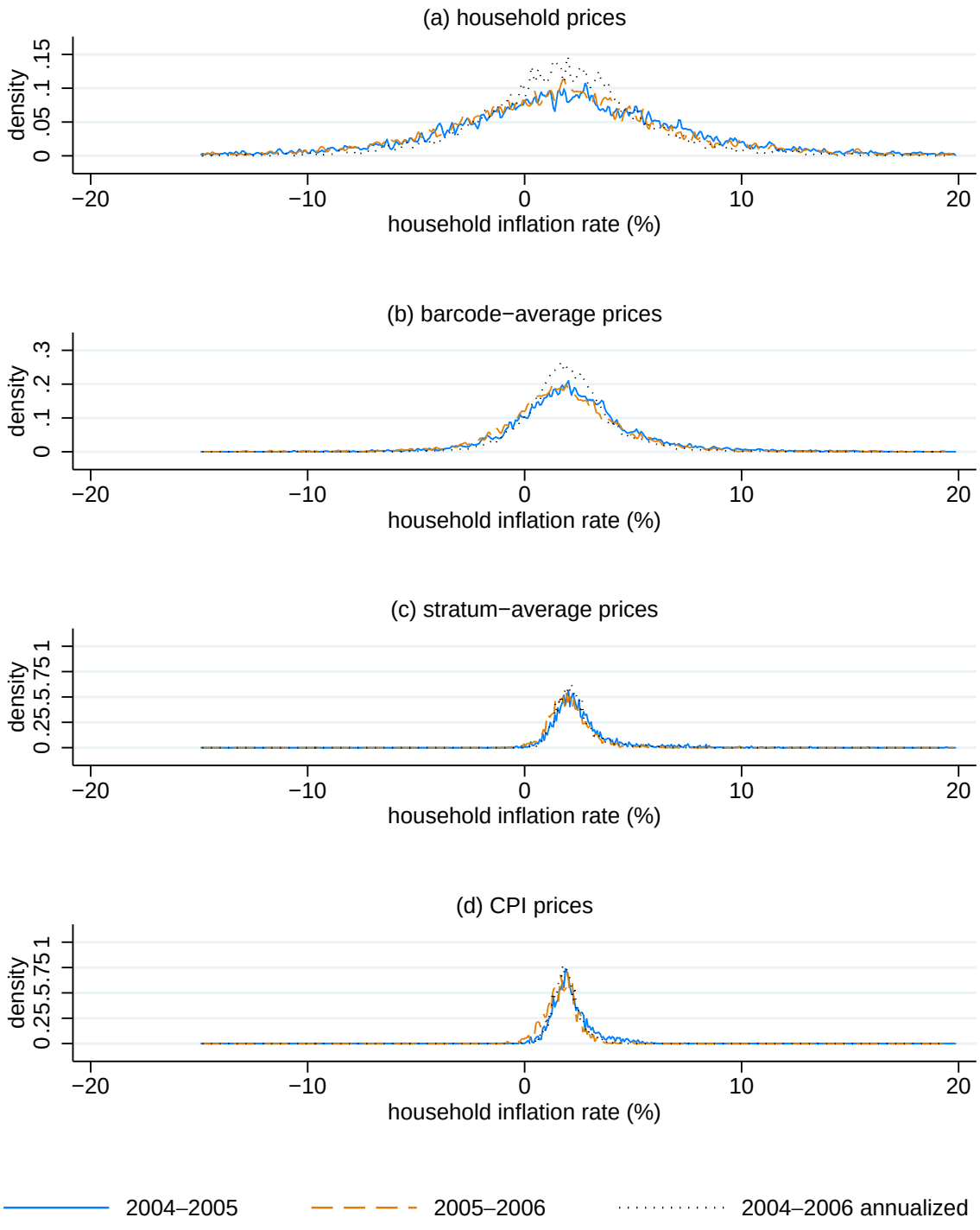

Figure 11: Distributions of one-year and two-year household-level inflation rates, 2004q4–2005q4 and 2005q4–2006q4.

Calculated with Paasche indexes. 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. Sample limited to 19,252 households with inflation rates calculated for both 2004q4–2005q4 and 2005q4–2006q4.

7. Constant vs. time varying consumption baskets and the serial correlation of inflation rates

In the main analysis of serial correlation and persistence, the basket of goods used to compute a household's inflation rate between $t+4$ and $t+8$ consists of all goods bought at both $t+4$ and $t+8$, and thus differs from the basket used to compute the household's inflation rate between t and $t+4$. In this section, we investigate how changes in the basket over time affect the estimated serial correlation.

At any given date t, we begin by creating a "fixed-basket sample" of households that bought at least five matched UPCs at all three relevant dates — t, $t+4$, and $t+8$. The fixed-basket sample is a subset of the sample used in the main text.

For each household in the fixed-basket sample, we call the set of UPCs bought at all three relevant dates the three-period basket. We can construct an inflation rate both from t to $t+4$ and from $t+4$ to $t+8$ using the three-period basket. We define the Laspeyres inflation rate between t and $t+4$ using the three-period basket as:

$$\pi^L_{it,t+4|t,t+4,t+8} = \frac{\sum_{\substack{j:\, q_{ijt},\ q_{ij,t+4},\\ q_{ij,t+8}>0}} p_{ij,t+4} q_{ijt}}{\sum_{\substack{j:\, q_{ijt},\ q_{ij,t+4},\\ q_{ij,t+8}>0}} p_{ijt} q_{ijt}}, \tag{1}$$

and the Laspeyres inflation rate between $t+4$ and $t+8$ using the three-period basket as:

$$\pi^L_{i,t+4,t+8|t,t+4,t+8} = \frac{\sum_{\substack{j:\, q_{ijt},\ q_{ij,t+4},\\ q_{ij,t+8}>0}} p_{ij,t+8} q_{ij,t+4}}{\sum_{\substack{j:\, q_{ijt},\ q_{ij,t+4},\\ q_{ij,t+8}>0}} p_{ij,t+4} q_{ij,t+4}}. \tag{2}$$

These formulas are identical to those used to calculate Laspeyres inflation rates with household-level prices in the main text, except that we now restrict the calculation to goods in the three-period basket.

For households in the fixed-basket sample, we can compare the cross-sectional distribution of $\pi^L_{it,t+4|t,t+4,t+8}$ (a 1-year inflation rate calculated using the goods in the three-period basket) with that of $\pi^L_{it,t+4}$ (a 1-year inflation rate calculated using goods bought at both t and $t+4$). We can also compare the serial correlation of inflation using a fixed basket

($\text{Cov}[\pi^L_{it,t+4|t,t+4,t+8}, \pi^L_{i,t+4,t+8|t,t+4,t+8}]$) with the serial correlation of inflation using a time-varying basket ($\text{Cov}[\pi^L_{it,t+4}, \pi^L_{i,t+4,t+8}]$). These comparisons show the effect of using a fixed basket instead of one that changes each period. We can also see how the properties of $\pi^L_{it,t+4}$ in the fixed-basket sample compare with the properties of $\pi^L_{it,t+4}$ in the full sample; these comparisons show the effect of restricting attention to the fixed-basket sample instead of the full sample, using the same time-varying basket to calculate inflation in both cases. Table 1 describes the distribution of spending in this sample.

Figure 12 shows the results. In the top panel, the standard deviation of 1-year inflation rates with time-varying baskets ($\pi^L_{it,t+4}$) is lower in the fixed-basket sample than the full sample, indicating that households that remain in the fixed-basket sample are less heterogeneous than households that do not. (Households fail to be in the fixed-basket sample if they attrit from the survey or have unstable purchasing patterns. To the extent the relationship between stable purchasing, attrition, and heterogeneity over two years also holds over a span of one year, this finding may indicate that our restriction in the main text to households that remain in the survey over at least a year and buy at least five matched barcodes over that year biases our measurement of heterogeneity downward.) The same pattern holds for two-year inflation rates, shown in the middle panel. However, within the fixed-basket sample, inflation rates with fixed baskets are more heterogeneous than inflation rates with time-varying baskets, over both one and two years.

The bottom panel shows the results on serial correlation. Over the 2004–2010 period — we cannot calculate inflation in 2011 in the fixed-basket sample — the serial correlation of inflation averages -0.11 with time-varying baskets in the full sample, -0.13 with time-varying baskets in the fixed-basket sample, and -0.23 with time-varying baskets in the fixed-basket sample. Each of these averages has a bootstrap standard error of 0.003. Thus, although the serial correlation of inflation is somewhat more negative when we used fixed baskets, some of this change is due to studying a narrower set of households, and in any event the serial correlation remains clearly well above -0.5; the evidence remains strongly against the hypothesis that households draw price levels at random each period.

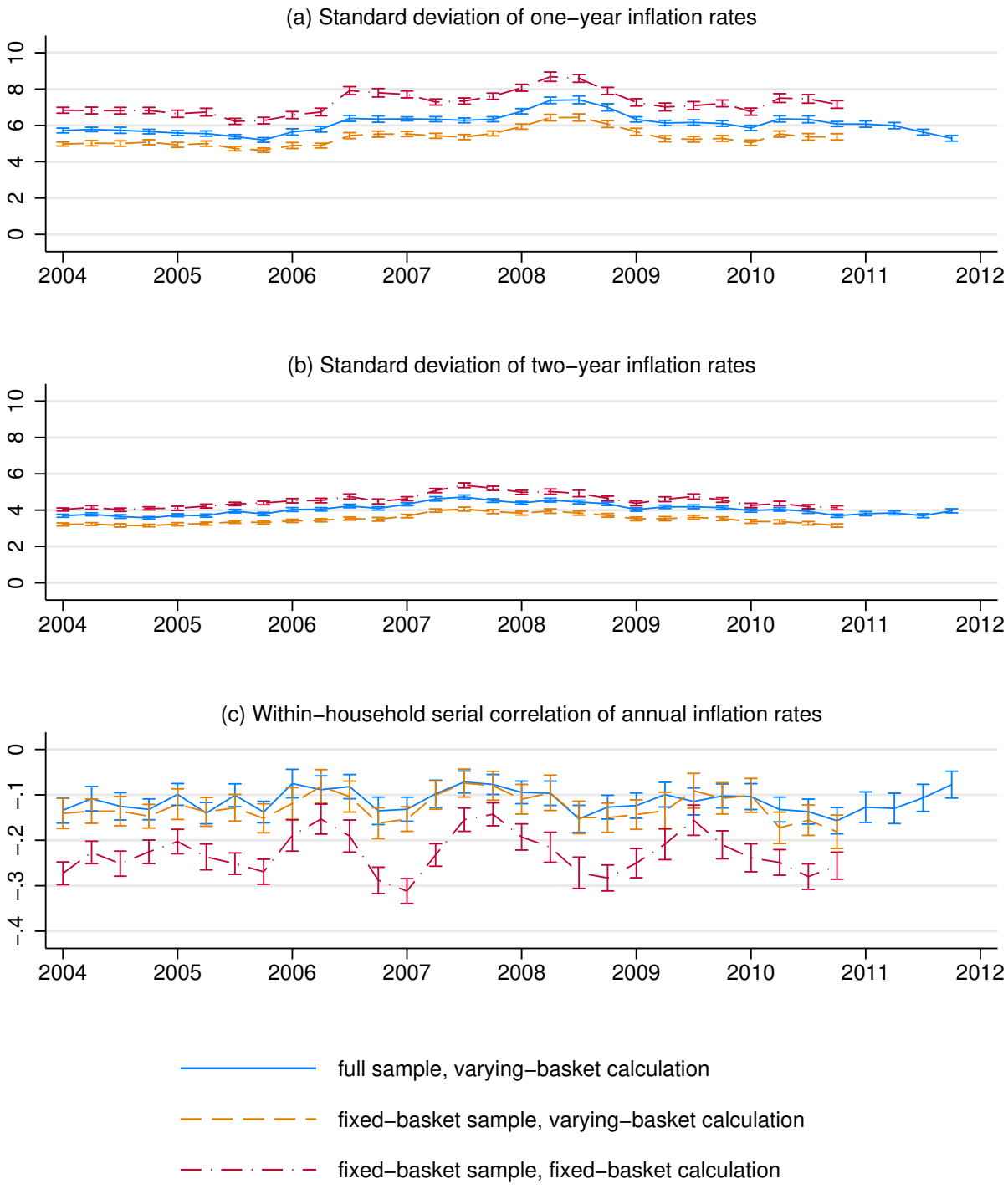

Figure 12: Evolution of the persistence of household-level inflation rates, fixed vs. time-varying bundles.

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.