Population Subgroup Price Indexes:

Evidence of Heterogeneity or Measurement Error ?

by

Rob Cage, Joshua Klick, William Johnson

Rob Cage

Meeting of the Group of Experts on Consumer Price Indexes
United Nations Economic Commission for Europe
Geneva, Switzerland
May 7-9, 2018

Agenda

1. Understanding and meeting different user needs
2. Criticisms of population subgroup indexes
3. Evidence of price change heterogeneity
4. Improving population subgroup index estimates

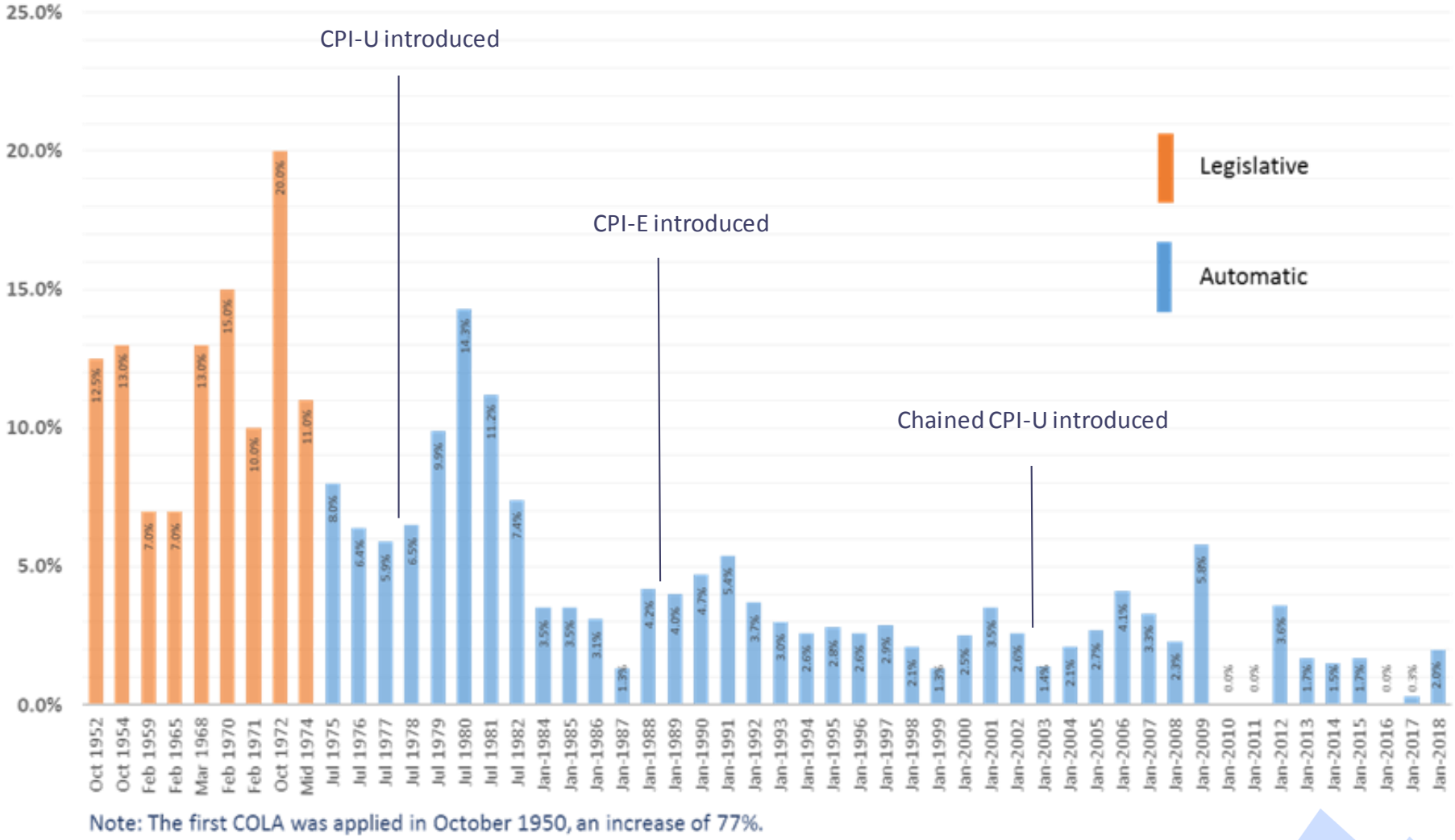
Any opinions expressed in this presentation are those of the author and do not constitute policy of the Bureau of Labor Statistics.

Understanding and Meeting Different User Needs:
BLS CPI products

CPI Users

- Policymakers
  - Congress, Federal Reserve Bank: inform fiscal and monetary policy
- Government agencies: execute laws stipulating use of CPI
  - Social Security Administration: cost-of-living-adjustment to benefits
  - Internal Revenue Service: adjust marginal tax bracket levels
- Contracting parties
  - Landlord\tenant: adjust lease payments
  - Employer\employee: wage escalation clauses
- Public at large: as barometer of overall inflation

History of Social Security COLAs



Note: The first COLA was applied in October 1950, an increase of 77%.

BLS Consumer Price Index products

INDEX	COHORT	YEAR INTRODUCED	NOTE	WEBSITE HITS[2]
CPI-W[1]	Wage-earners and clerical workers	1921	Index used for Social Security Cost of Living Adjustment (COLA)	6.0%
CPI-U	All urban consumers	1978	Headline Index	93.1%
CPI-E[2]	Elderly consumers	1988	Experimental	< 0.1%
Chained CPI-U	All urban consumers	2002	Published with ~ 1-year lag	0.9%

Notes:
1. Index is a population subgroup index.
2. Website hit percentages are for the 12-month period ending February 2018. CPI-E is not currently available on the website.

Differences in current index methods

Lower-level aggregation

$$I_{i,a;t-1\to t} = \frac{\sum_{k\epsilon i,a} w_{k,\theta}\left(\frac{P_{k,t}}{P_{k,\theta}}\right)}{\sum_{k\epsilon i,a} w_{k,\theta}\left(\frac{P_{k,t-1}}{P_{k,\theta}}\right)} \quad or \quad \prod_{k\epsilon i,a}\left(\frac{P_{k,t}}{P_{k,t-1}}\right)^{\hat{w}_{k,t-1},}$$

=

≠

Upper-level aggregation

$$I_{j,t-1\to t} = \frac{\sum_{i,a} w_{i,a,j,\theta}\left(\frac{I_{i,a,t}}{I_{i,a,\theta}}\right)}{\sum_{i,a} w_{i,a,j,\theta}\left(\frac{I_{i,a,t-1}}{I_{i,a,\theta}}\right)} \quad or \quad \prod_{i,a}\left(\frac{I_{i,a,t}}{I_{i,a,t-1}}\right)^{\frac{w_{i,at-1}+w_{i,a,t}}{2}}$$

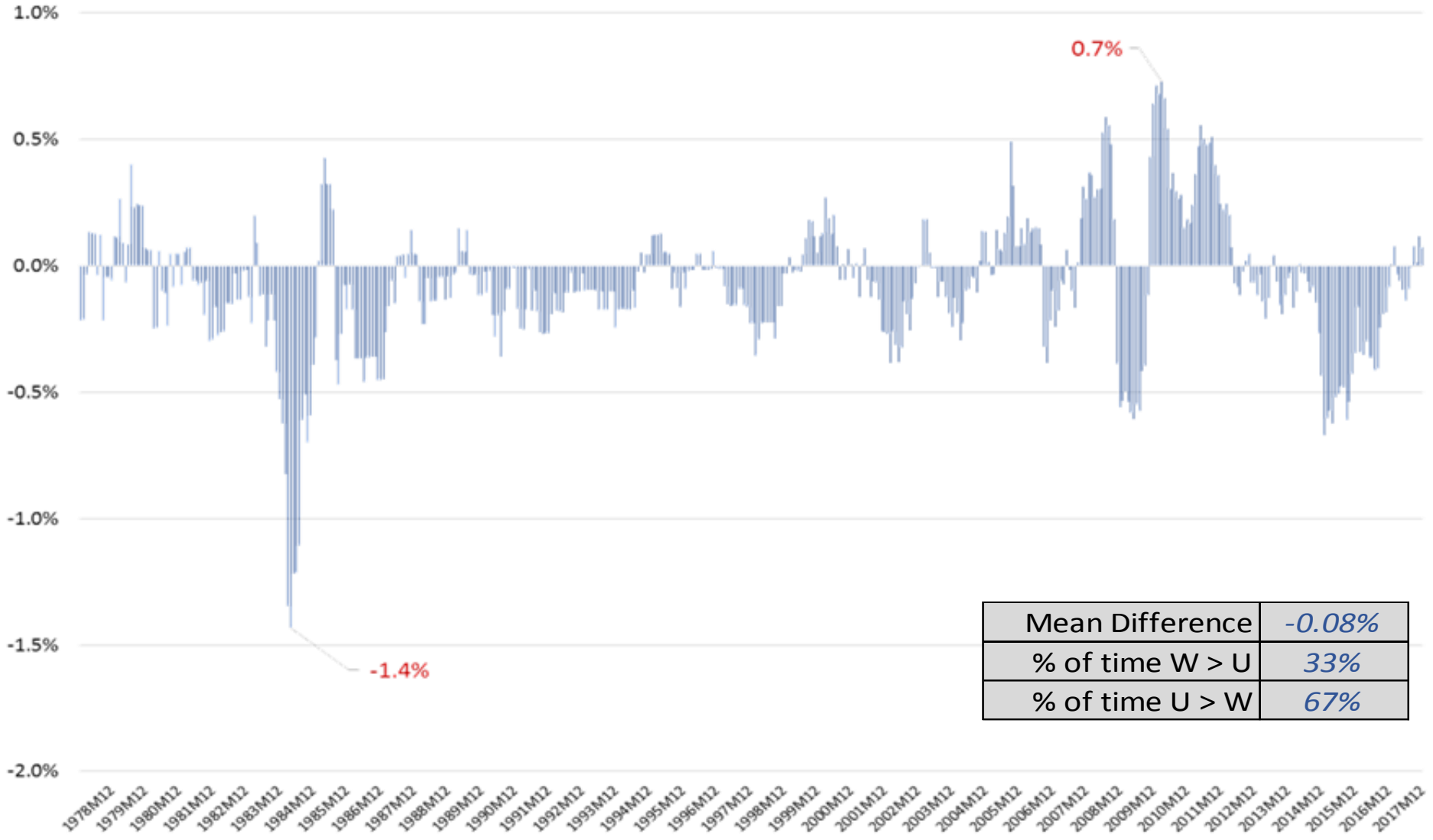
Note: Only calculated for all urban consumers.

Upper-level aggregation weights, 2013-2014

	Category	CPI-U	CPI-W	CPI-E
A1	Adult clothing	1.7%	1.6%	1.3%
A2	Children and infant clothing	0.4%	0.5%	0.1%
E0	Telephone and electronics	6.0%	6.9%	5.5%
E1	Tuition	2.7%	2.0%	1.2%
F1	Food at home	10.9%	12.6%	9.9%
F2	Food away from home	5.1%	5.1%	4.4%
F3	Alochol and tobacco	1.4%	1.6%	1.1%
H0	Shelter	32.1%	30.6%	35.5%
H1	Household utilities	5.5%	6.0%	5.7%
H2	Housefurnishings and operations	3.7%	2.6%	4.1%
M0	Medical Care	8.3%	6.5%	11.1%
R0	Entertainment and recreation	4.5%	3.6%	4.3%
T0	Vehicles	7.0%	7.9%	6.4%
T1	Gasoline and vehicle maintenance	9.5%	11.4%	8.0%
T2	Public transportation	1.3%	0.9%	1.3%

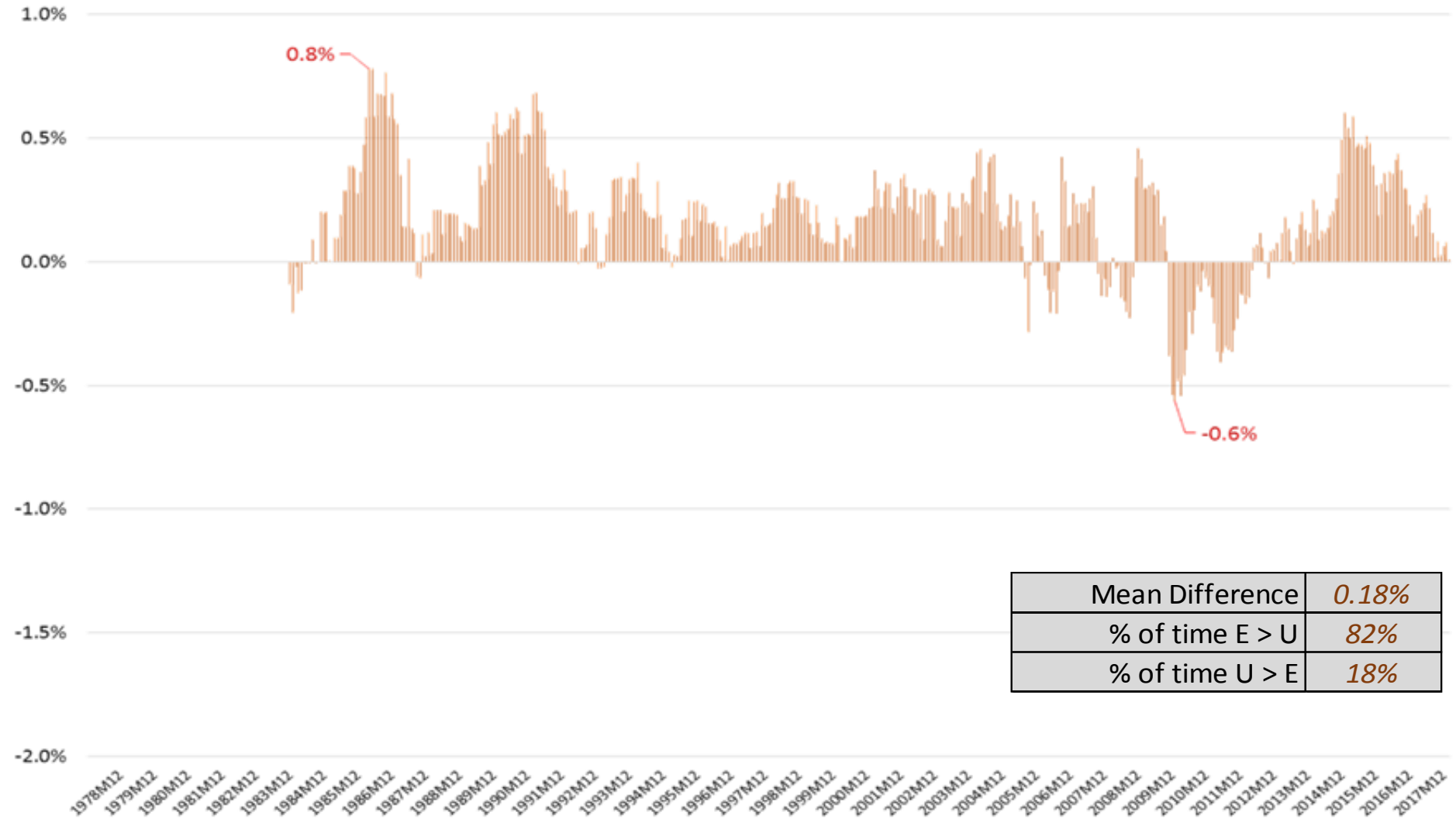
Note: Estimated from Consumer Expenditure Interview Survey, research sample.

Difference in annual inflation estimates, CPI-W minus CPI-U



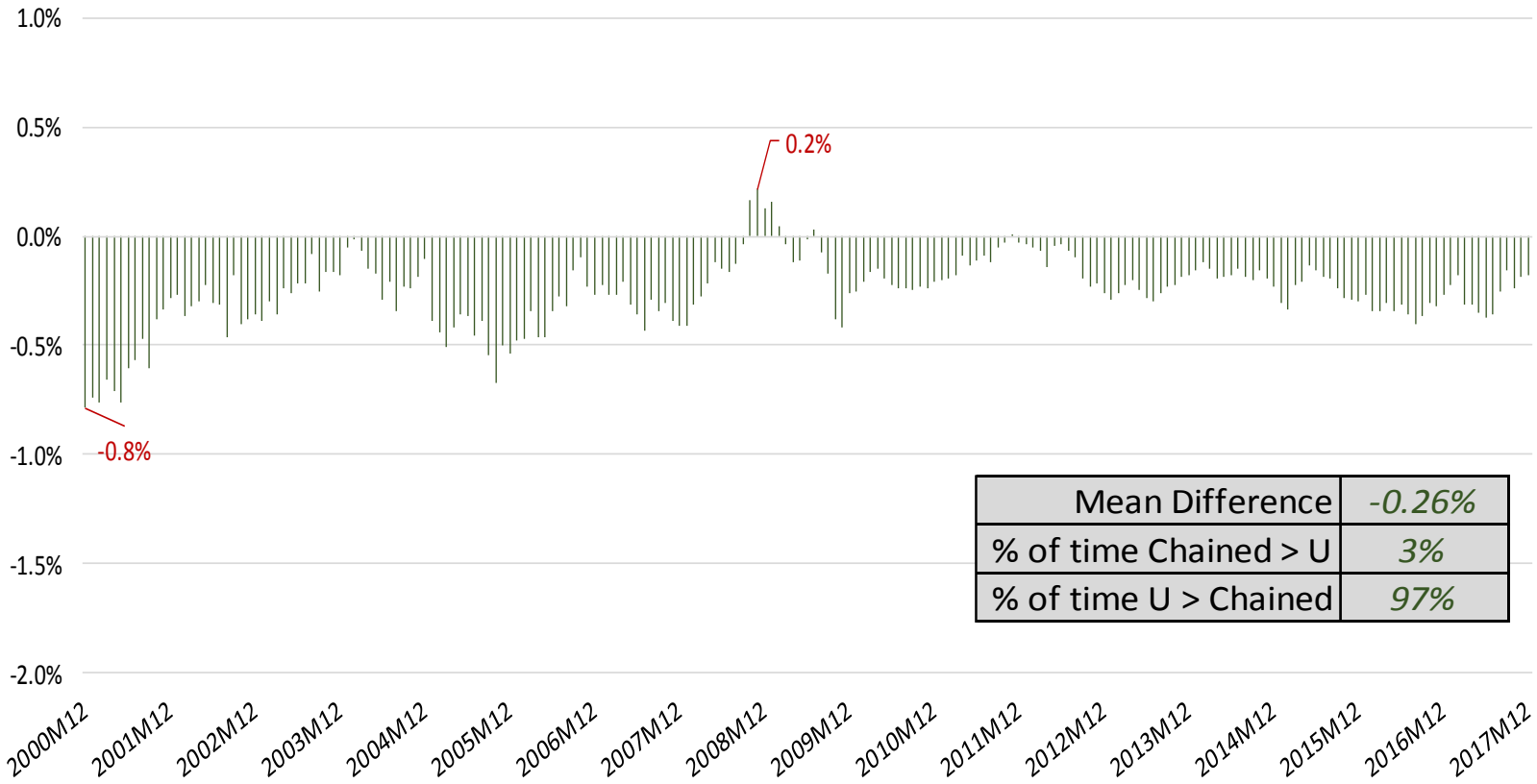
Mean Difference	*-0.08%*
% of time W > U	*33%*
% of time U > W	*67%*

Difference in annual inflation estimates, CPI-E minus CPI-U



Mean Difference	*0.18%*
% of time E > U	*82%*
% of time U > E	*18%*

Difference in annual inflation estimates, Chained CPI-U minus CPI-U



Mean Difference	*-0.26%*
% of time Chained > U	*3%*
% of time U > Chained	*97%*

COLA impact on Social Security outlays

2017

Number of beneficiaries	66,721,000
Average monthly benefit	$ 1,259
TOTAL	**$ 1,008,068,907,120**

SOURCE: U.S. Social Security Administration, Master Beneficiary Record and Supplemental Security Record, 100 percent data.

2018 OASDI Cost-of-Living Adjustment	CPI-W	CPI-U	CPI-E	Chained CPI-U
COLA	**2.0%**	**2.0%**	**2.1%**	**1.7%**
Average Monthly Benefit	$ 1,284	$ 1,284	$ 1,285	$ 1,280
Difference from CPI-W		$ -	$ 1	$ (4)
TOTAL 2018 OUTLAY	1.028 trillion	1.028 trillion	1.029 trillion	1.025 trillion
Aggregate Difference		0	+ 1.2 billion	- 3 billion

Criticisms of Population Subgroup Indexes:
The public debate over Social Security COLA

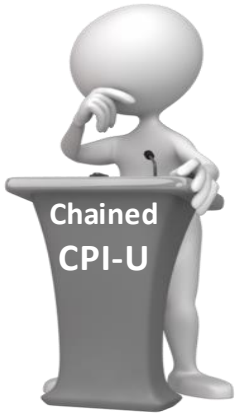
Inflation is a violent mugger,
as frightening as an armed robber,
and as deadly as a hit man.

- Ronald Reagan
1978

https://www.huffingtonpost.com/robert-reich/chained-cpi_b_3016471.html

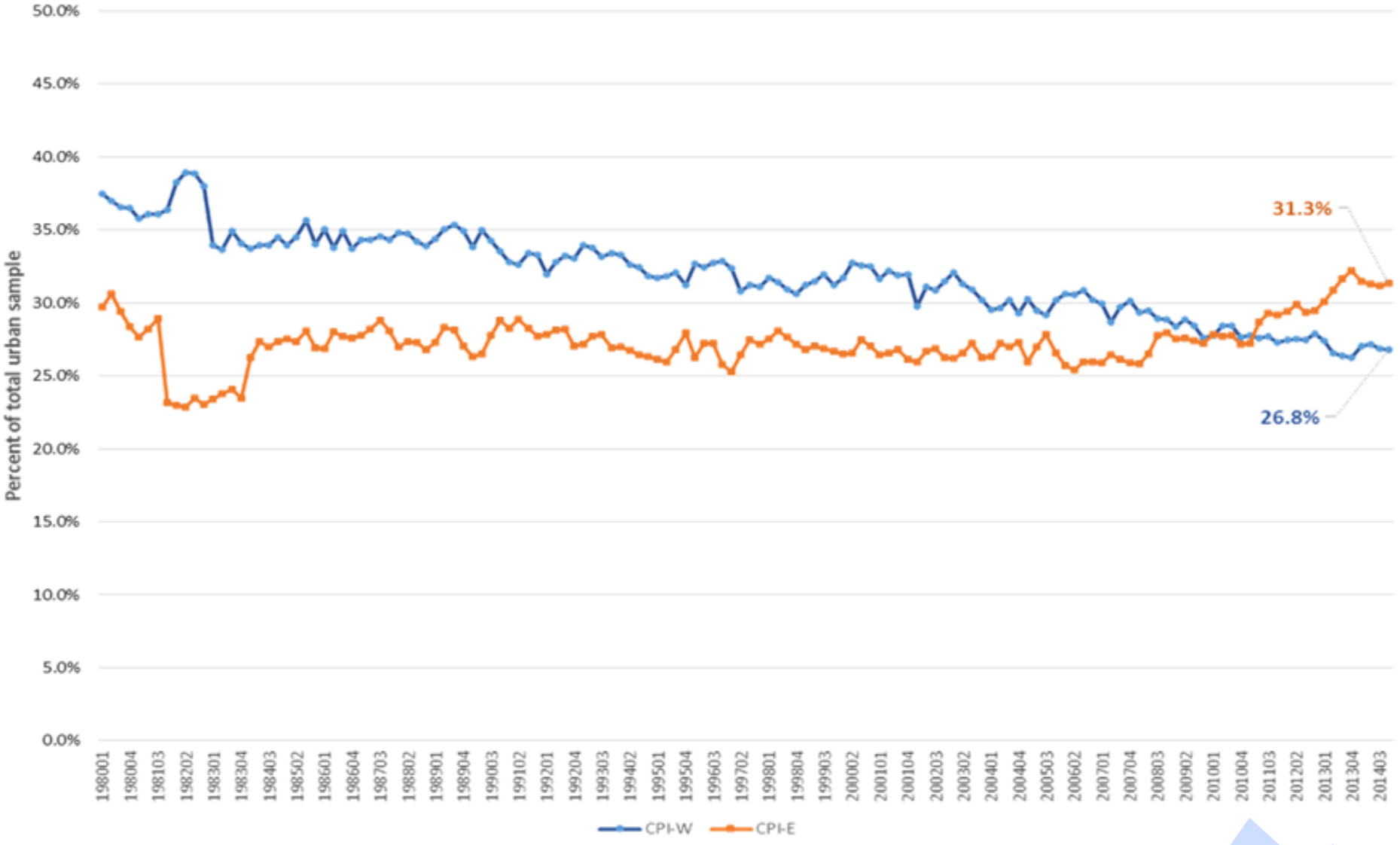
Indexation of Social Security benefits: which index?

- More appropriate cohort definition
- More representative of Social Security beneficiary spending behavior
- Hypothesized lower substitution elasticity
- Chained CPI weights reflect more than response to relative price change

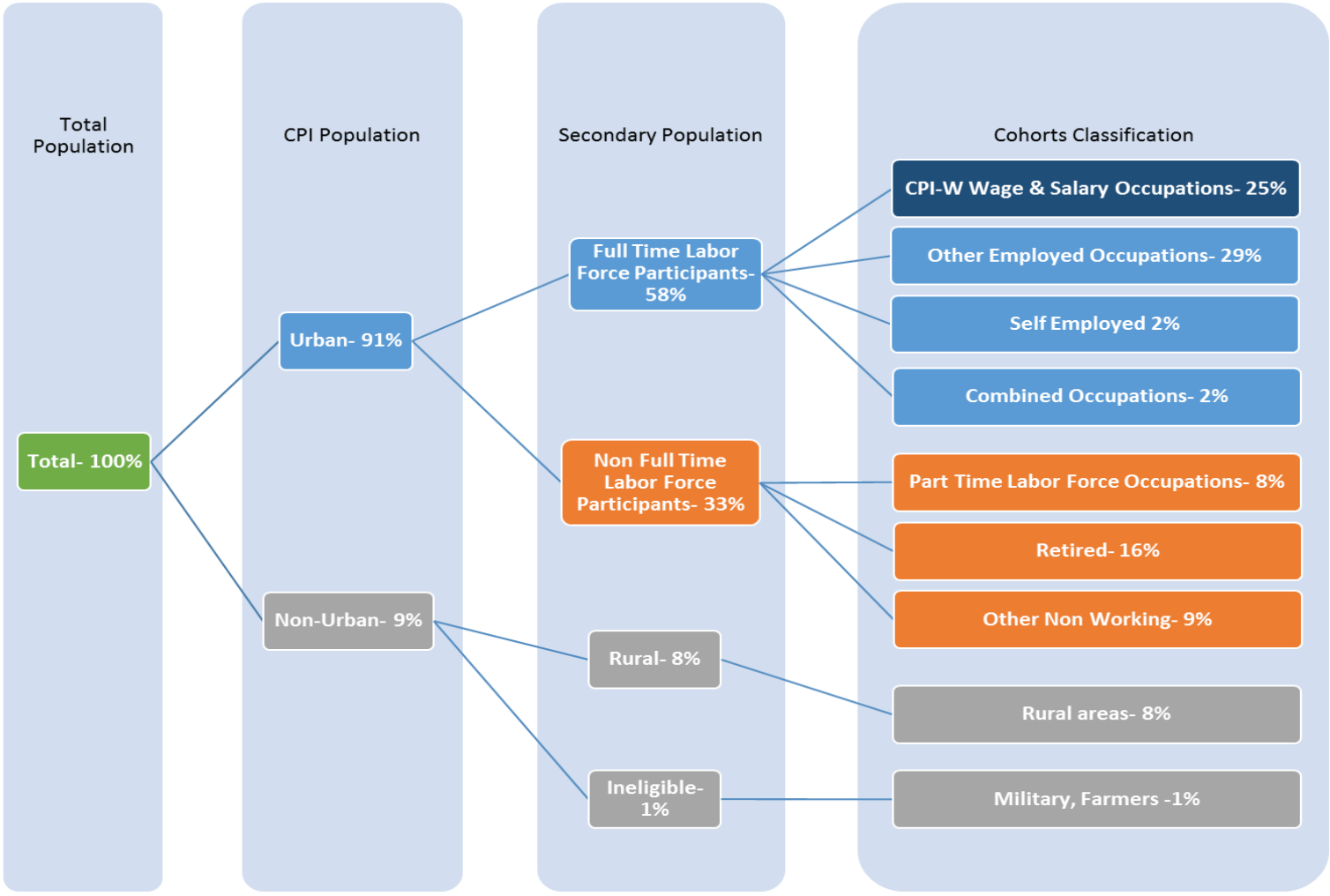


- CPI-E fails to control within stratum heterogeneity
- CPI-E aggregation weights associated with higher sampling error than the CPI-U
- CPI-E fails to address upper-level substitution bias
- Chained CPI-U more accurate measure of overall inflation

Declining CPI-W sample size

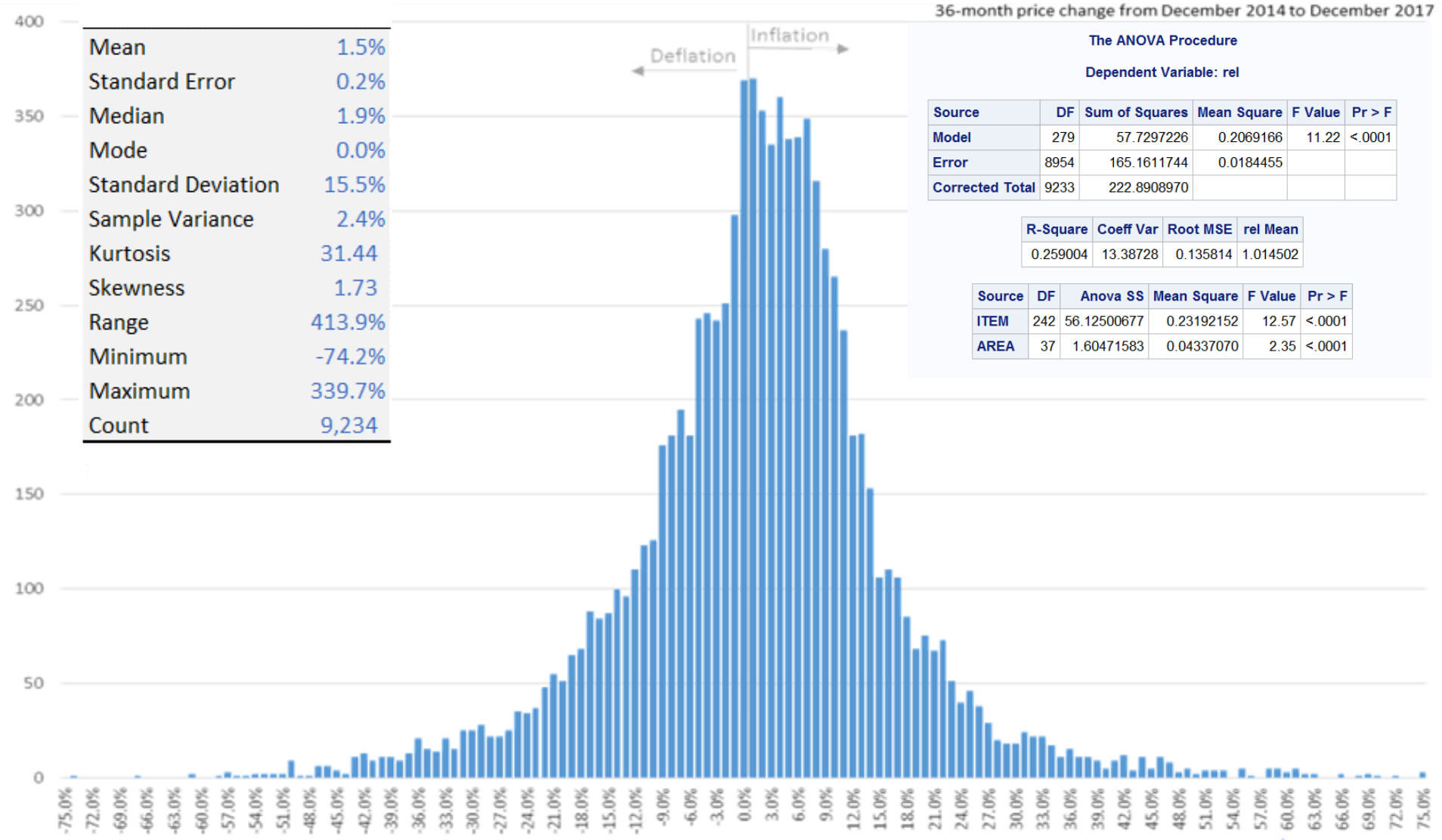


Population distribution by occupational status, 2013-2014



Evidence of Price Change Heterogeneity:
Within and across component items; across households

Evidence of component index variation



Mean	1.5%
Standard Error	0.2%
Median	1.9%
Mode	0.0%
Standard Deviation	15.5%
Sample Variance	2.4%
Kurtosis	31.44
Skewness	1.73
Range	413.9%
Minimum	-74.2%
Maximum	339.7%
Count	9,234

The ANOVA Procedure

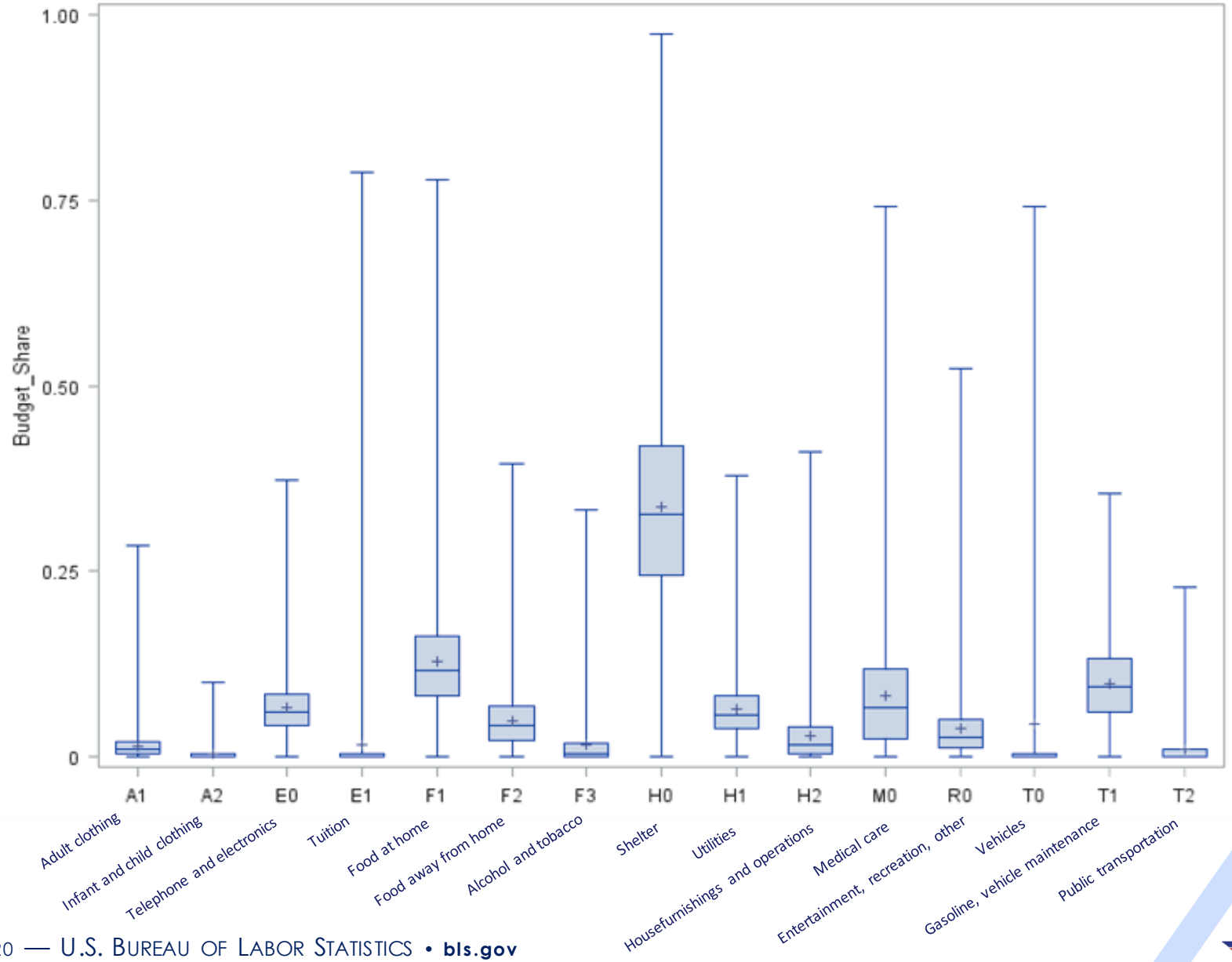
Dependent Variable: rel

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	279	57.7297226	0.2069166	11.22	<.0001
Error	8954	165.1611744	0.0184455		
Corrected Total	9233	222.8908970			

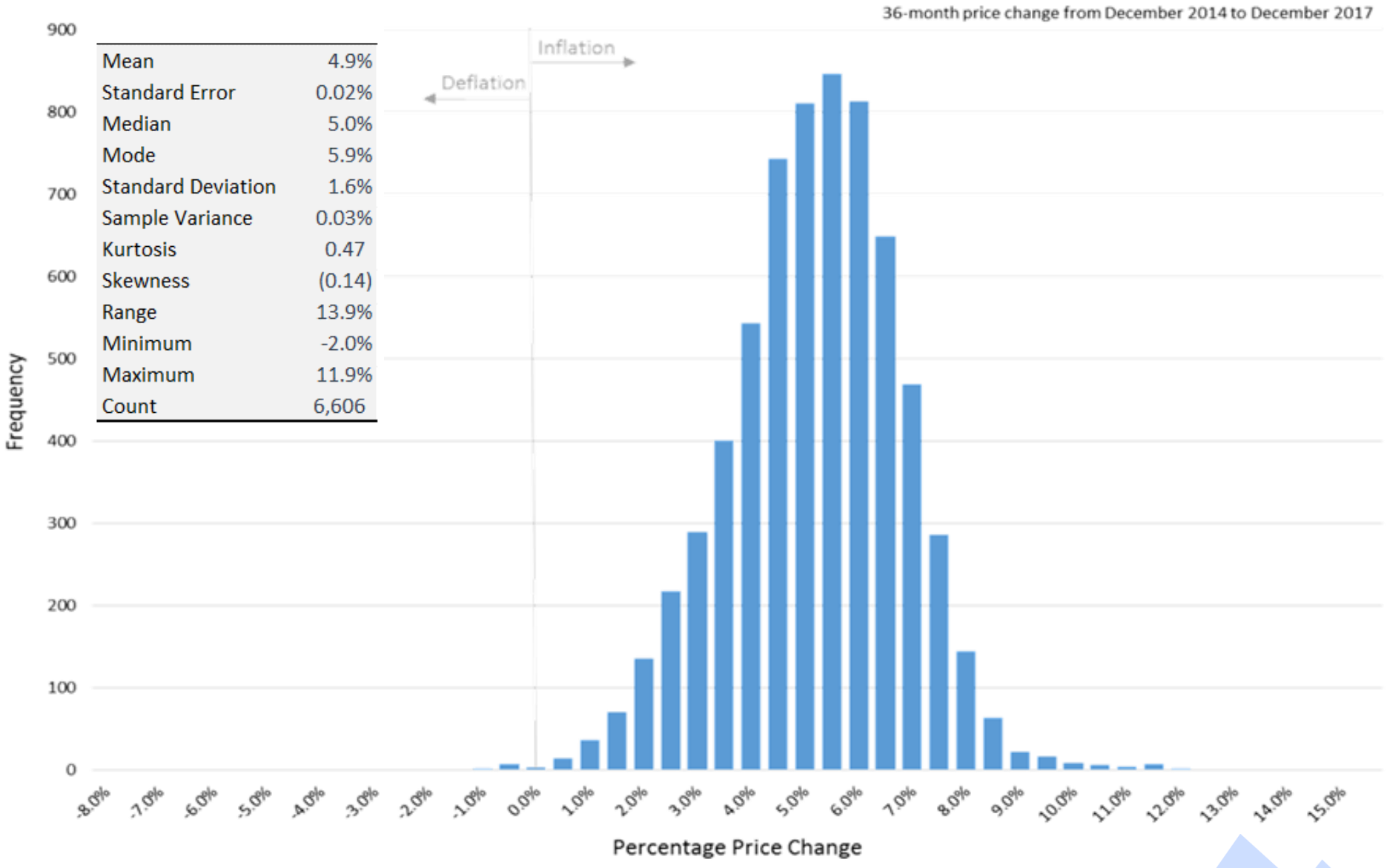
R-Square	Coeff Var	Root MSE	rel Mean
0.259004	13.38728	0.135814	1.014502

Source	DF	Anova SS	Mean Square	F Value	Pr > F
ITEM	242	56.12500677	0.23192152	12.57	<.0001
AREA	37	1.60471583	0.04337070	2.35	<.0001

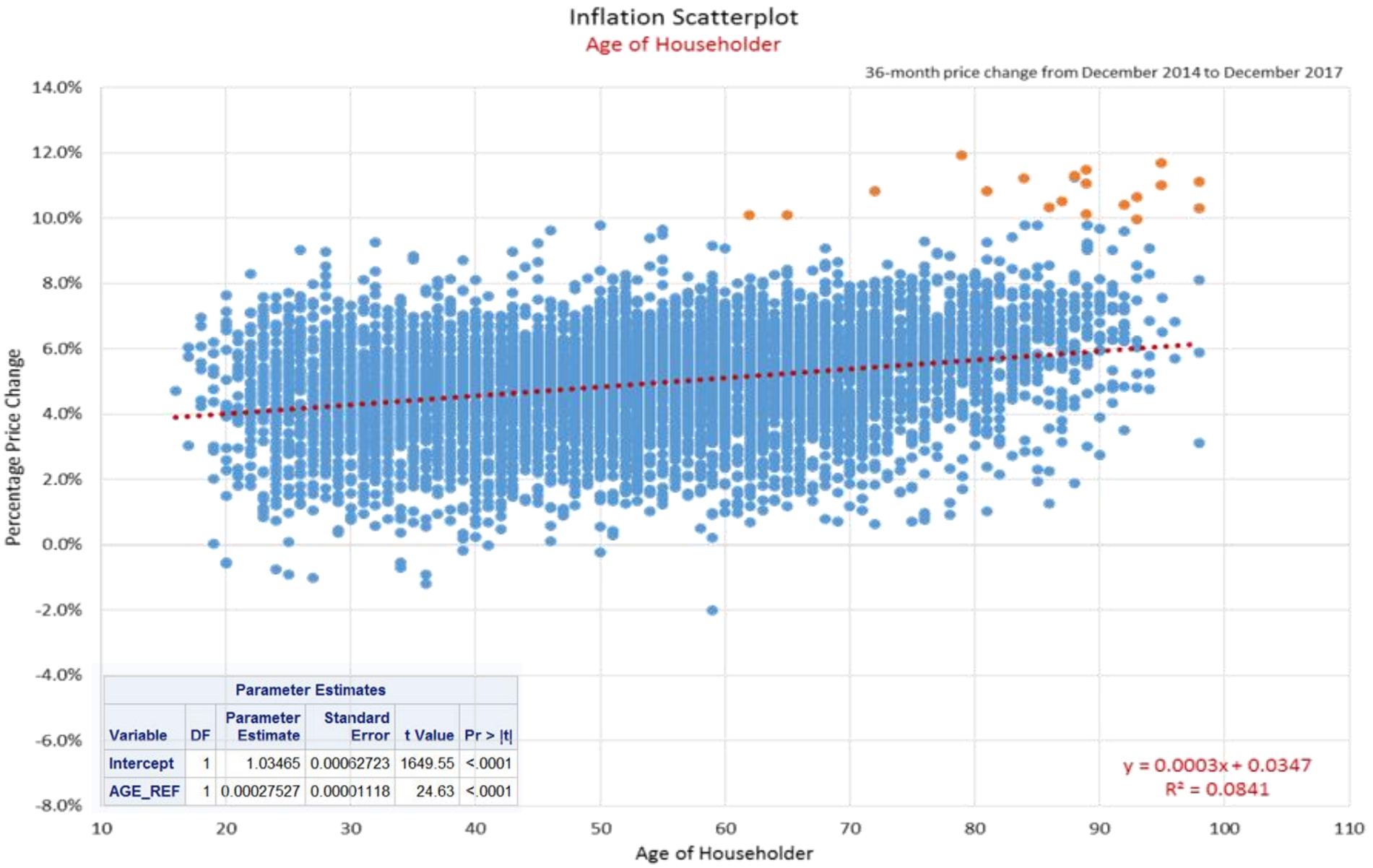
Evidence of household budget share variation,
All urban consumers, 2013-2014



Evidence of household index variation

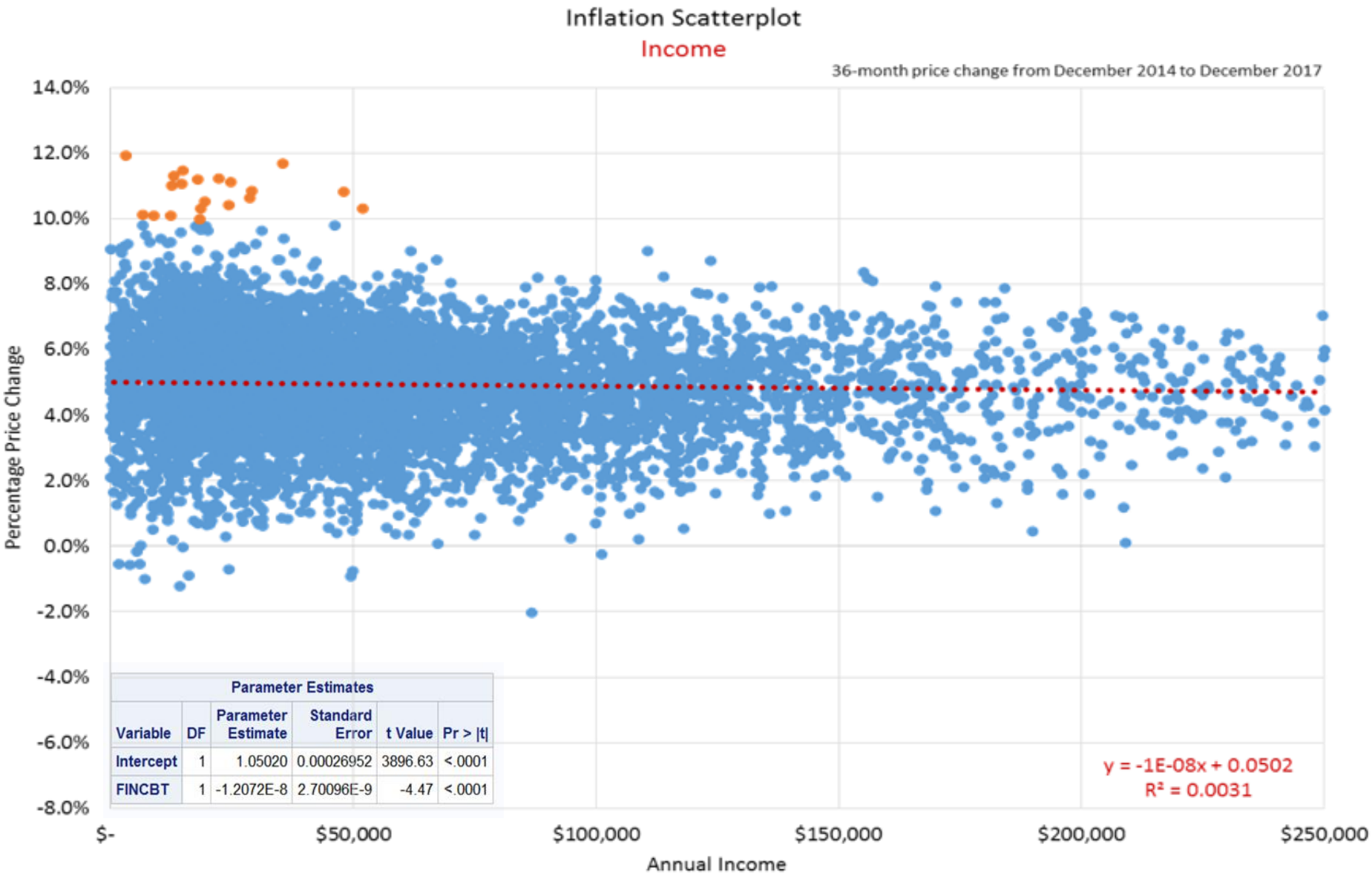


Mean	4.9%
Standard Error	0.02%
Median	5.0%
Mode	5.9%
Standard Deviation	1.6%
Sample Variance	0.03%
Kurtosis	0.47
Skewness	(0.14)
Range	13.9%
Minimum	-2.0%
Maximum	11.9%
Count	6,606

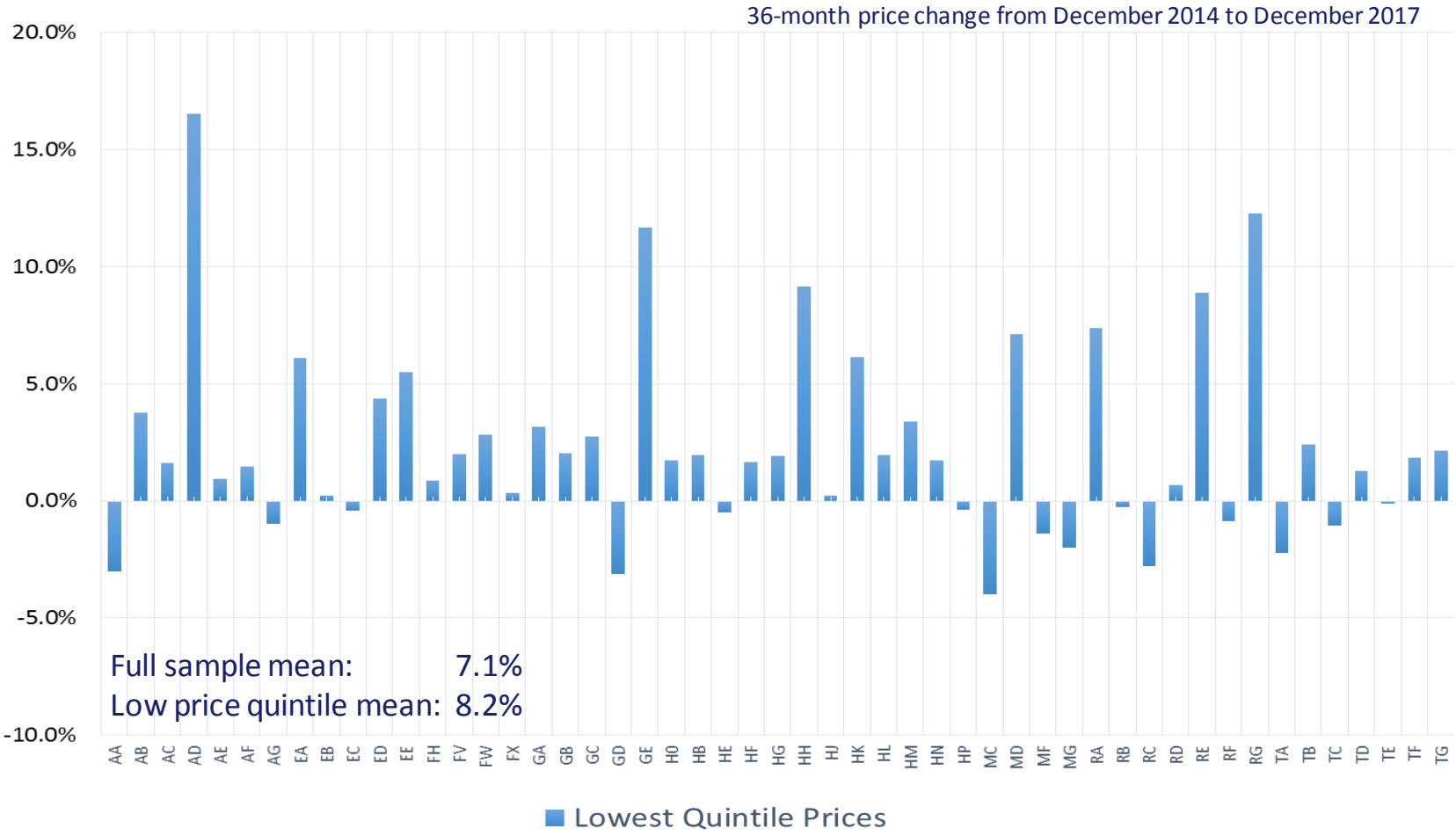


Parameter Estimates

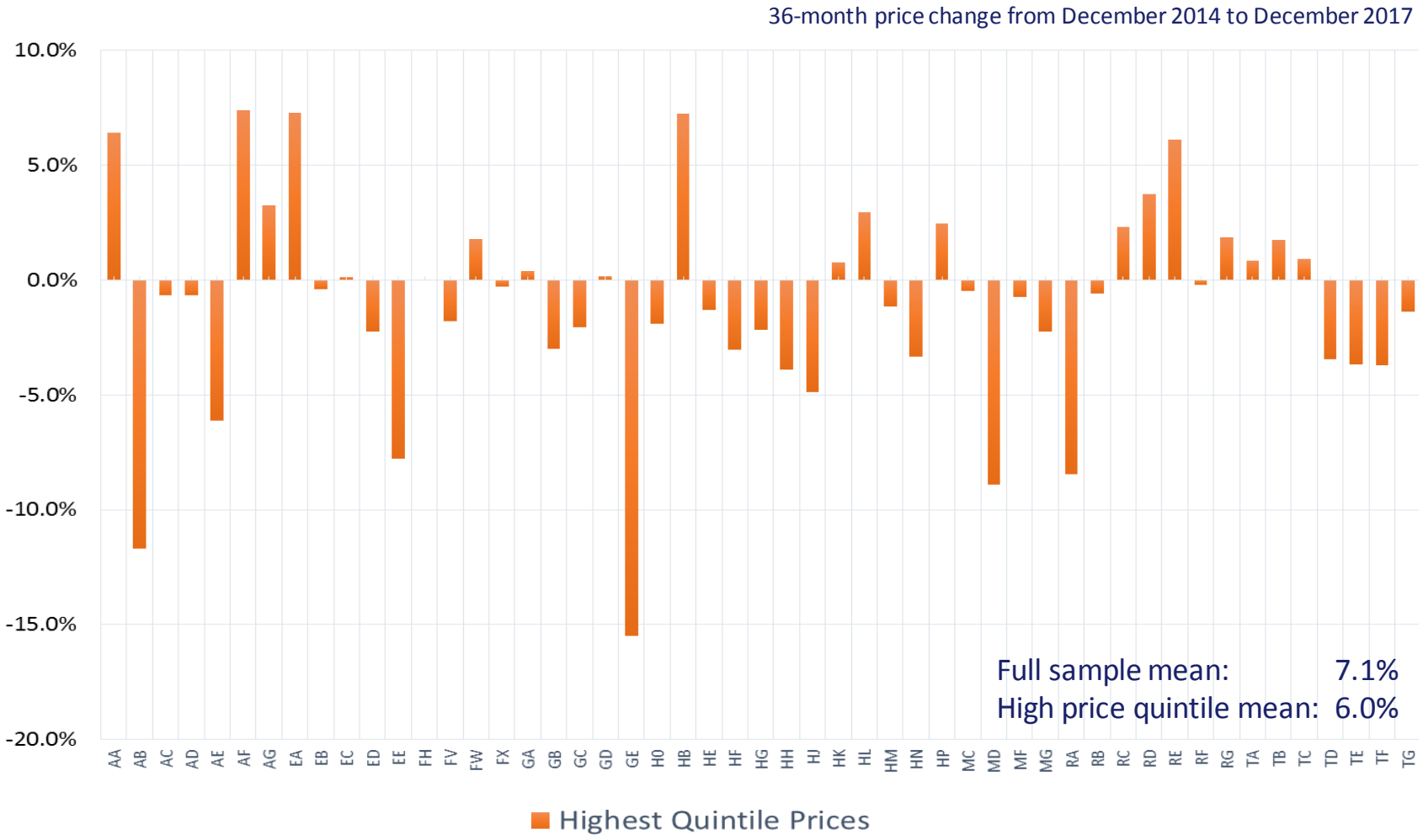
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.03465	0.00062723	1649.55	<.0001
AGE_REF	1	0.00027527	0.00001118	24.63	<.0001


Inflation Scatterplot
Income
36-month price change from December 2014 to December 2017
Percentage Price Change
14.0%
12.0%
10.0%
8.0%
6.0%
4.0%
2.0%
0.0%
-2.0%
-4.0%
-6.0%
-8.0%
$-
$50,000
$100,000
$150,000
$200,000
$250,000
Annual Income
Parameter Estimates
Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t|
Intercept | 1 | 1.05020 | 0.00026952 | 3896.63 | <.0001
FINCBT | 1 | -1.2072E-8 | 2.70096E-9 | -4.47 | <.0001
y = -1E-08x + 0.0502
R² = 0.0031

Evidence of within stratum variation:
Restricted sample index minus full sample index



Evidence of within stratum variation:
Restricted sample index minus full sample index



Improving Population Subgroup Index Estimates:
A stratified cluster aggregation framework

Experimental subgroup index methodology

	CURRENT APPROACH	EXPERIMENTAL FRAMEWORK
COHORT DEFINITION	Wage-earners and clerical workers	Social Security recicpients
OWNER-OCCUPIED HOUSING	Rental equivalence	Payment approach
COMPONENT INDEXES	All urban consumers	Variable (low price vector for low income households)
AGGREGATION WEIGHTS	Plutocratic	Democratic
AGGREGATION FORMULA	Lowe	Tornqvist
CLASSIFICATION OF HOUSEHOLDS	None	Stratified by income Clustered by budget shares

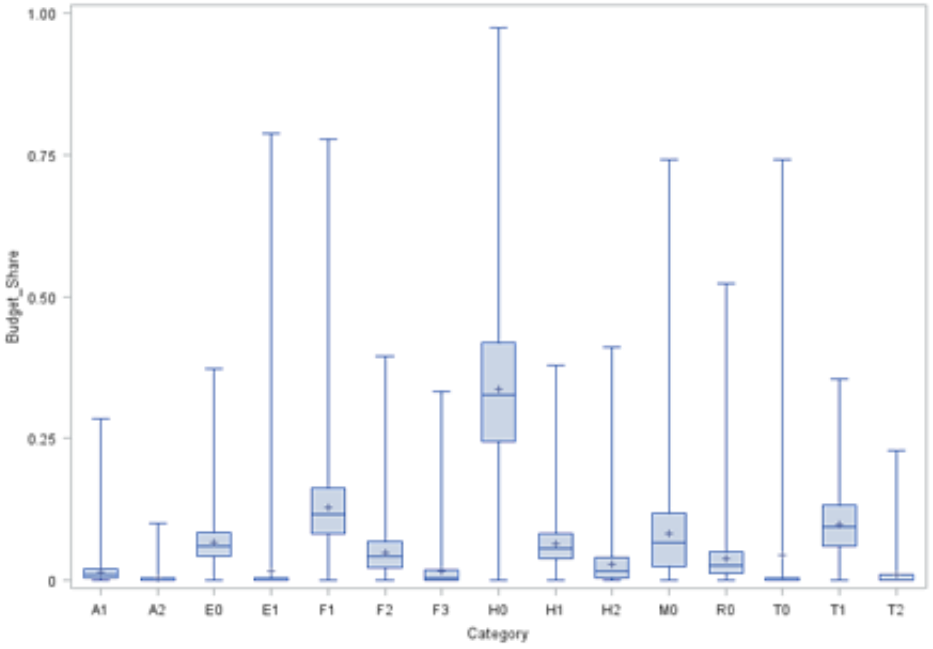
Process: Stratified cluster aggregation

Step 1. Stratify households into like-kind groups

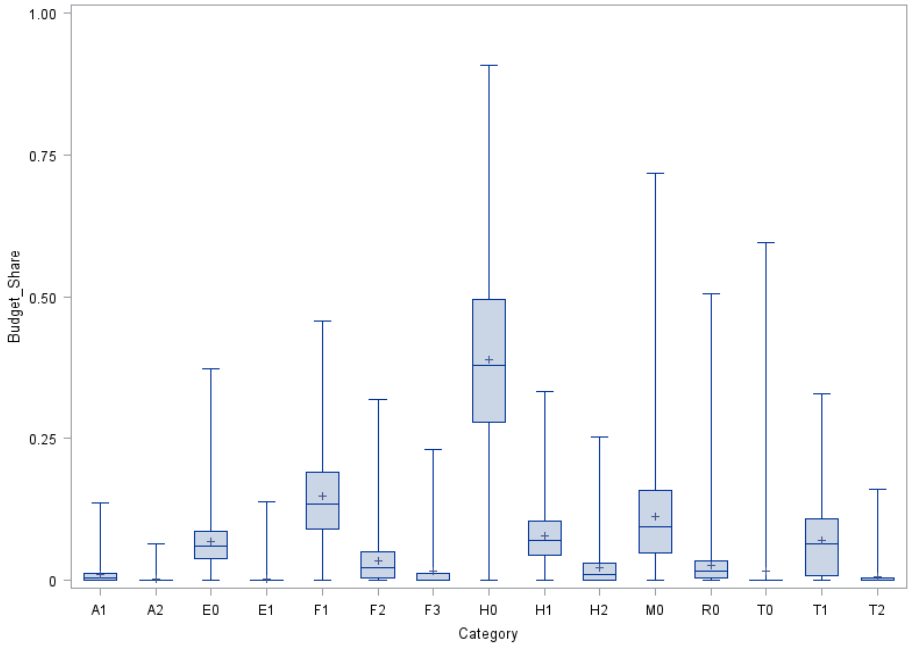
- (3) Income x (2) Social Security status

Code		n	%
SL	Social Security households:	1424	*22%*
SM	Low income	648	*10%*
	Middle Income	776	*12%*
NL	Non Social Security households:	5171	*78%*
NM	Low income	671	*10%*
NH	Middle income	3181	*48%*
	High income	1319	*20%*

All urban consumers



Social Security, Low Income Quintile



Process: Stratified cluster aggregation cont'd

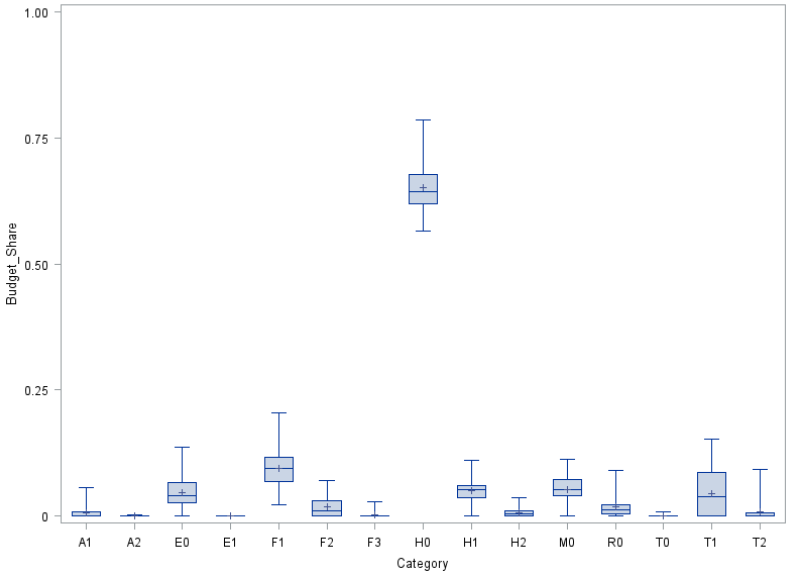
Step

2. Execute LASSO to identify budget shares for clustering
   - $I^G{}_{j;\theta\to t} = \sum_{i=1}^{15} \beta_i w_{k_i}$ where the error is $\sum_{observations}\left(I^G{}_{j;\theta\to t} - \sum_{i=1}^{15} \beta_i w_{k_i}\right)^2 + \lambda \sum_{i=1}^{15} |\beta_i|$
   - Selections:
     - A2-Childrens apparel, F1-Food at home, F2-Food away, H0-Shelter, H1-Utilities, M0-Medical care, T1-Gasoline\vehicle maintenance, E1-Telephone, electronics, internet

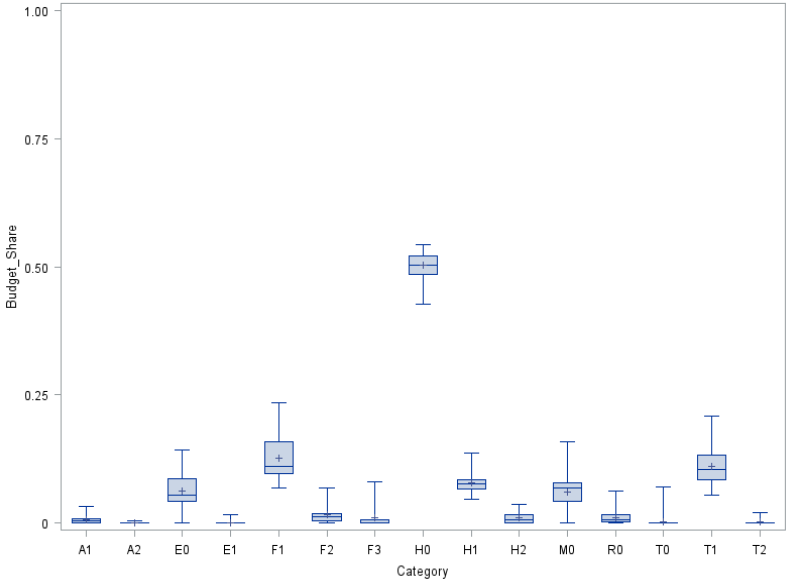
3. Execute Hierarchical Clustering within cohort strata on budget shares
   - Ward's minimum variance method
   - Resulted in 40 clusters per strata (75 for NM stratum)
   - Average sample size of 30 households per cluster

$$D(K,L) = \frac{\|\overline{X_K} - \overline{X_L}\|^2}{\frac{1}{N_K} + \frac{1}{N_L}}$$

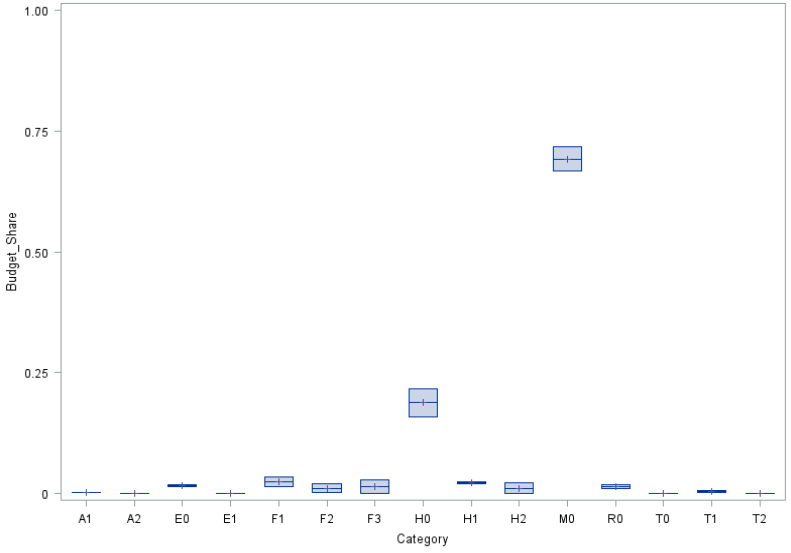
STRATUM: Social Security, Low Income Quintile
CLUSTER = 1



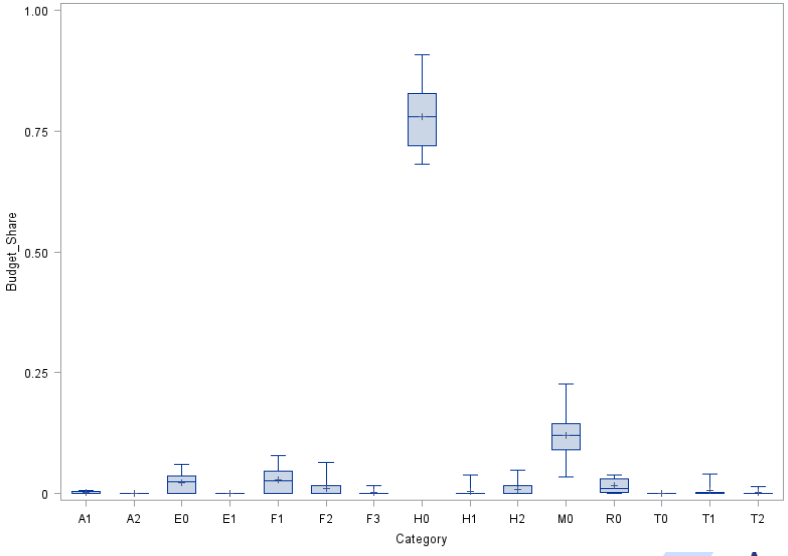
STRATUM: Social Security, Low Income Quintile
CLUSTER = 13



STRATUM: Social Security, Low Income Quintile
CLUSTER = 30

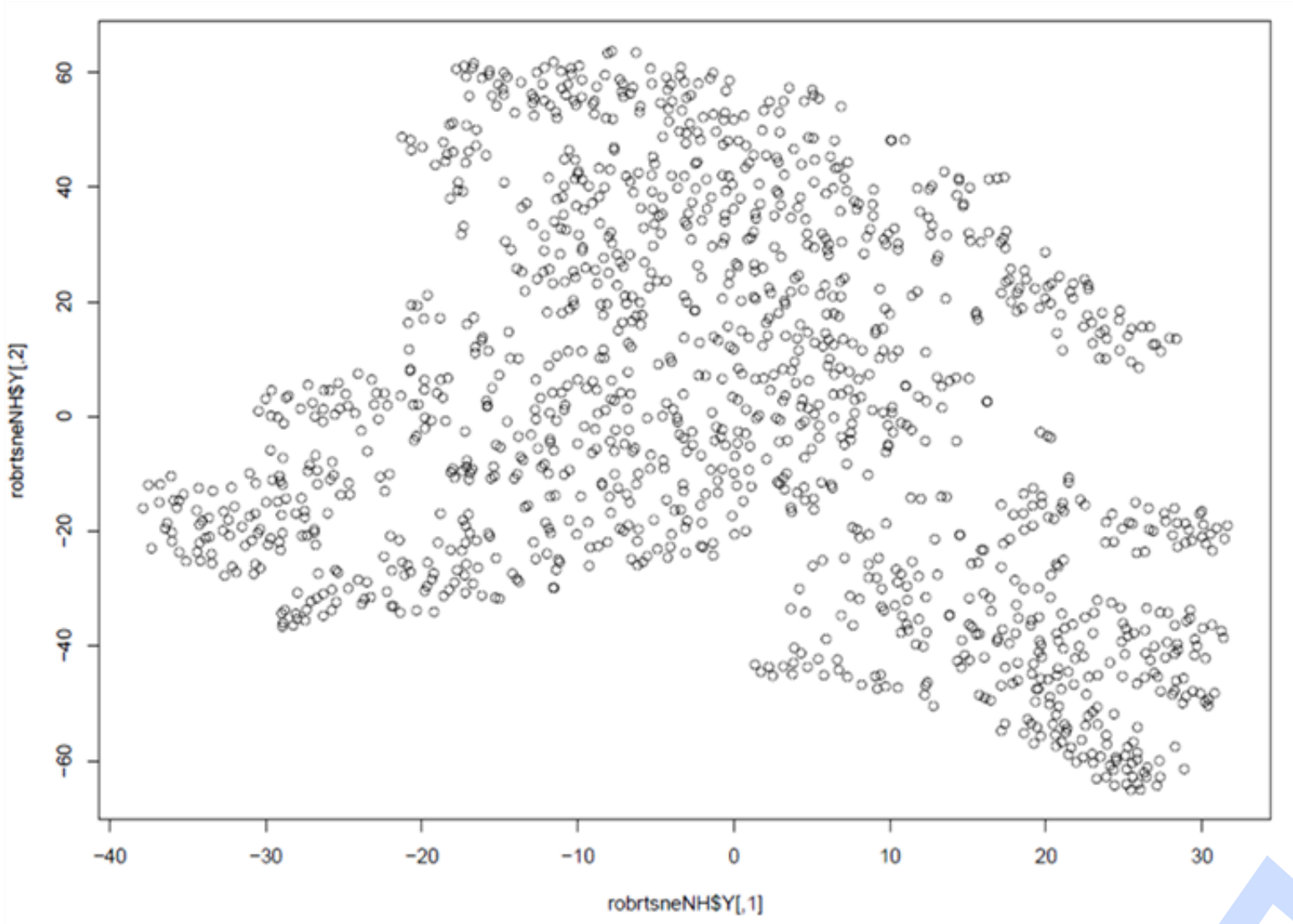


STRATUM: Social Security, Low Income Quintile
CLUSTER = 11



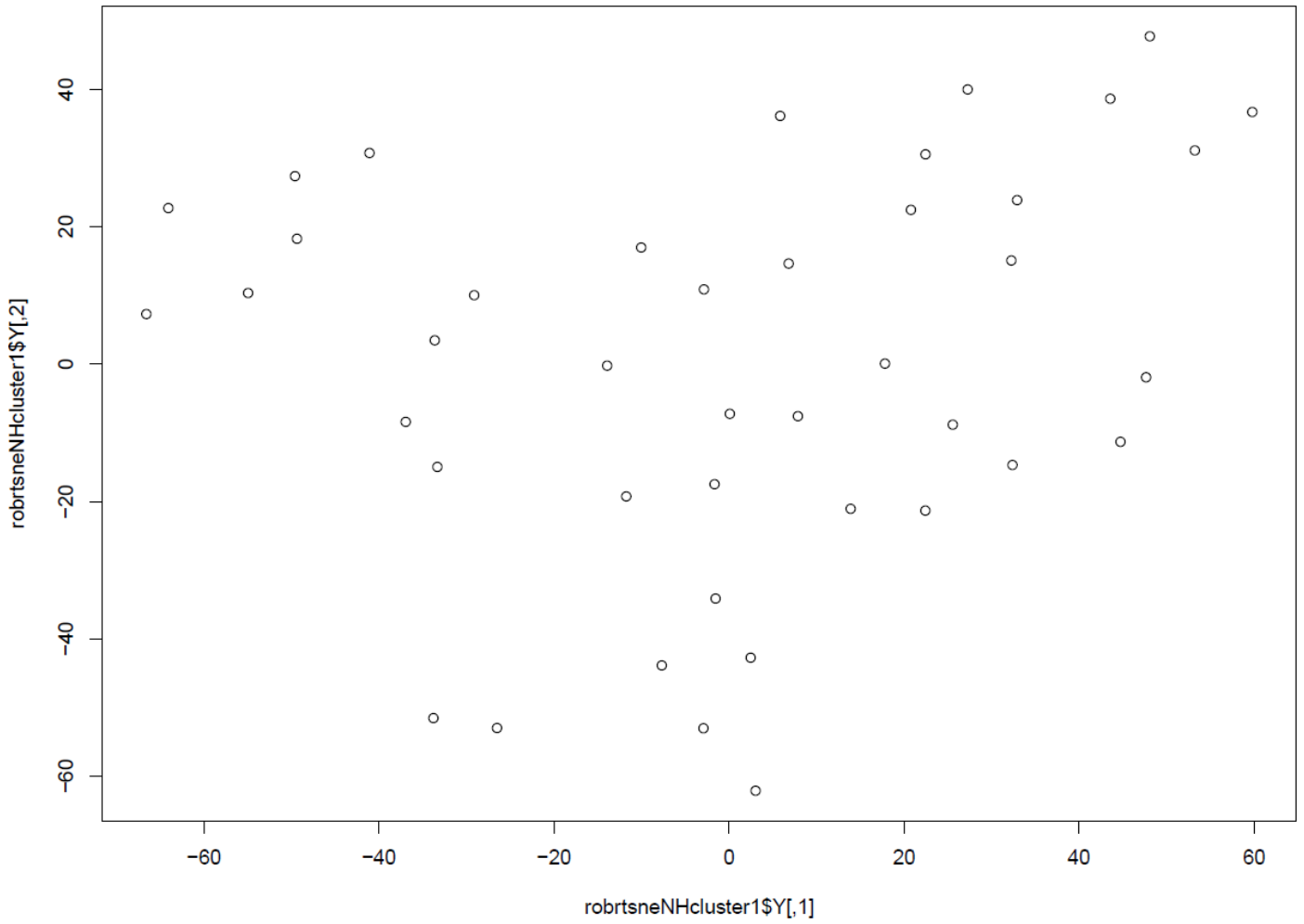
t-distributed Stochastic Neighbor Embedding visualization

STRATUM: Non-Social Security, High Income Quintile



t-distributed Stochastic Neighbor Embedding visualization

STRATUM: Non-Social Security, High Income Quintile
CLUSTER = 1



Process: Stratified cluster aggregation cont'd

Step

4. Calculate cluster (c) indexes within each stratum (s)

$$E_{k,s,c,0} = \sum_{j \in s,c} (PQ)_{k,j,0}\,, \quad w_{k,s,c,0} = \frac{E_{k,sc,0}}{\sum_{k \in s,c} E_{k,sc,0}}$$

$$I^{L}_{s,c;} = \sum_{k} w_{k,s,c} \left(\frac{P_{k,t}}{P_{k,0}}\right) \qquad I^{LM}_{s,c;} = \left[\sum_{k} w_{k,s,c,0} \left(\frac{P_{k,t}}{P_{k,0}}\right)\left(\frac{P_{k,t}}{P_{k,0}}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$$

- Tornqvist: execute K-nearest neighbor classification machine learning algorithm on discriminate function of demographic variables, to assign current-period household sample into base-period stratum-clusters (future research)

5. Aggregate into stratum indexes
   - Subgroups: Democratic cluster weights
   - National index: Plutocratic cluster weights

6. Aggregate into final index products
   - Subgroups: Democratic stratum weights
   - National index: Plutocratic cluster weights

Results

36-month price change from December 2014 to December 2017

Formula	Owner Occupied Housing	Component Indexes	Low Income Quintile	High Income Quintile	Social Security Cohort	All Urban Consumers
Laspeyres	Payment	Full sample, urban	4.24%	4.50%	4.33%	4.19%
Laspeyres	Payment	Restricted price vector	6.12%*	4.44%	5.47%	4.71%
Laspeyres	OER	Full sample, urban	5.06%	4.86%	5.56%	4.80%
Laspeyres	OER	Restricted price vector	7.05%*	4.75%	6.81%*	5.34%
CES σ=0.6	Payment	Restricted price vector	5.98%*	4.32%	5.31%	
CES σ=0.6	OER	Restricted price vector				5.21%

* Outside 95% confidence interval of official CPI-U (3.85% to 5.91%)

Summary of findings

- Experimental indexes not significantly different from CPI-U
  - Exception: Low income quintile and Social Security cohort estimates using low price vector and rental equivalence are larger
  - Social Security CES index higher than national CES index by 0.03% per annum
- Treatment of owner occupied-housing and within-item inflation variation have large impact
  - Payment approach yields lower inflation estimates during study period, notably for Social Security cohort
  - Use of restricted samples for component indexes (low price quintile) for low income cohorts yields higher overall inflation estimates
- Stratified clustering of households may improve accuracy

Next steps

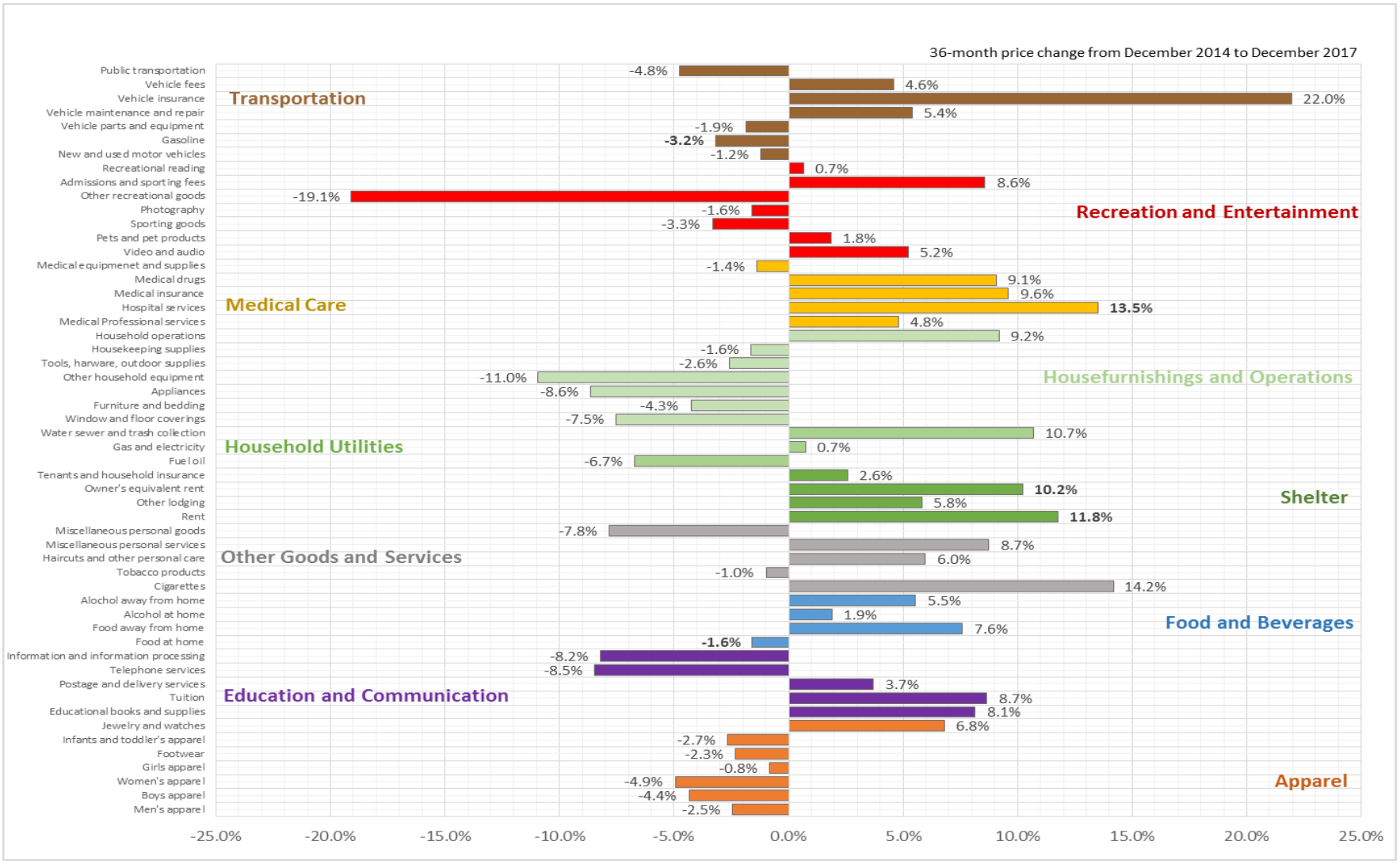
- Stratified Cluster Aggregation Research:
  - Apply technique at item-area level (control for geographic variation)
  - Apply technique over extended time period (to estimate Tornqvist index)
- Consumer Expenditure Survey Redesign:
  - Need larger sample size to stratify CPI samples by item, area, *and* population
  - Addition of outlet questions planned for 2020
    - Should enable stratified outlet sample selection by population cohort
    - Could link selected outlets to specific item reported in CE > first stage of unique item selection
    - Allows for some control of within stratum heterogeneity across subgroups
- Subgroup indexes
  - Reevaluate CPI-W definition and options for improvement
  - Reexamine subgroup index calculation methodology

Contact Information

Rob Cage
Economist
U.S. Department of Labor | Bureau of Labor Statistics
2 Massachusetts Ave, N.E. | Washington, DC
www.bls.gov/cpi
202-691-6959
cage.rob@bls.gov

Reference Slides

Price change for CPI Expenditure Class categories



Descriptive statistics by cohort

				Income x Social Security strata				
2013-2014 Sample Statistics	**CPI-U**	**CPI-W**	**CPI-E**	**NL**	**NM**	**NH**	**SL**	**SM**
Sample size	24,049	6,140	6,416	671	3,181	1,319	648	776
Demographic variable:								
One or more members age 16 or under	38.9%	50.5%	15.6%	38.2%	43.7%	59.0%	8.6%	11.6%
One or more members age 64 or older	26.0%	10.5%	83.3%	14.3%	19.5%	17.1%	77.2%	89.2%
Single consumer household	32.0%	23.6%	41.7%	46.7%	24.9%	5.0%	75.6%	30.0%
Living in center city of metro area	32.0%	35.8%	29.3%	40.5%	30.0%	23.7%	30.7%	21.5%
One or more owned vehicles	84.0%	86.0%	83.6%	65.4%	91.8%	96.1%	65.7%	90.7%
Owner	58.6%	48.6%	77.6%	37.3%	67.0%	91.4%	62.5%	87.2%
Percent of owners without mortgage	41.7%	31.7%	65.3%	58.0%	37.1%	24.4%	74.6%	71.5%
One or more members earning income	76.9%	100.0%	42.7%	59.3%	92.1%	97.8%	6.6%	22.2%
50% or more of total income sourced from Social Security	16.9%	0.0%	50.0%	0.0%	0.0%	0.0%	100.0%	100.0%
50% or more of total income sourced from earnings	59.0%	100.0%	27.0%	30.0%	74.3%	83.9%	0.0%	0.0%
Income before taxes:								
q1	$21,607	$27,223	$18,220	$ 3,881	$38,000	$117,400	$ 11,022	$39,724
q3	$84,000	$72,258	$62,928	$14,406	$74,343	$190,000	$ 16,467	$25,093
Median	$45,000	$44,912	$33,205	$ 9,856	$54,704	$140,908	$ 13,851	$30,398
Mean	$64,300	$54,779	$51,298	$ 9,528	$56,610	$172,366	$134,801	$33,868

Budget shares for income, Social Security strata

RENTAL EQUIVALENCE

	Category	NL	NM	NH	SL	SM
A1	Adult clothing	1.2%	1.6%	2.1%	1.1%	1.0%
A2	Children and infant clothing	0.6%	0.4%	0.4%	0.1%	0.1%
E0	Telephone and electronics	6.8%	6.6%	5.3%	6.2%	5.5%
E1	Tuition	2.6%	1.9%	4.9%	0.3%	0.5%
F1	Food at home	15.7%	11.9%	8.8%	12.3%	10.2%
F2	Food away from home	4.2%	5.1%	5.8%	3.6%	3.9%
F3	Alochol and tobacco	1.8%	1.5%	1.3%	1.3%	1.0%
H0	Shelter	35.2%	31.7%	30.2%	39.4%	34.5%
H1	Household utilities	7.2%	5.8%	4.3%	7.1%	6.3%
H2	Housefurnishings and operations	2.2%	3.0%	4.9%	2.6%	3.4%
M0	Medical Care	5.7%	7.6%	7.3%	12.0%	13.9%
R0	Entertainment and recreation	3.3%	3.9%	5.9%	2.8%	3.7%
T0	Vehicles	3.8%	7.0%	8.0%	3.2%	6.9%
T1	Gasoline and vehicle maintenance	8.9%	10.7%	8.8%	7.2%	8.3%
T2	Public transportation	0.9%	1.0%	1.8%	0.7%	0.9%

PAYMENTS APPROACH

	Category	NL	NM	NH	SL	SM
A1	Adult clothing	1.4%	1.7%	2.3%	1.4%	1.3%
A2	Children and infant clothing	0.7%	0.5%	0.5%	0.2%	0.2%
E0	Telephone and electronics	7.5%	7.2%	5.7%	7.8%	7.1%
E1	Tuition	2.8%	2.0%	5.2%	0.4%	0.6%
F1	Food at home	17.2%	13.0%	9.5%	15.6%	13.1%
F2	Food away from home	4.6%	5.6%	6.3%	4.6%	5.0%
F3	Alochol and tobacco	2.0%	1.7%	1.4%	1.7%	1.2%
H0	Shelter	28.8%	25.6%	25.1%	23.1%	16.1%
H1	Household utilities	7.9%	6.4%	4.7%	9.0%	8.1%
H2	Housefurnishings and operations	2.4%	3.3%	5.3%	3.3%	4.3%
M0	Medical Care	6.3%	8.3%	7.8%	15.2%	17.8%
R0	Entertainment and recreation	3.6%	4.3%	6.3%	3.5%	4.7%
T0	Vehicles	4.2%	7.6%	8.6%	4.1%	8.9%
T1	Gasoline and vehicle maintenance	9.8%	11.7%	9.5%	9.1%	10.6%
T2	Public transportation	1.0%	1.1%	2.0%	0.9%	1.1%

PAYMENT APPROACH minus RENTAL EQUIVALENCE

	Category	NL	NM	NH	SL	SM
A1	Adult clothing	0.1%	0.1%	0.2%	0.3%	0.3%
A2	Children and infant clothing	0.1%	0.0%	0.0%	0.0%	0.0%
E0	Telephone and electronics	0.7%	0.6%	0.4%	1.7%	1.6%
E1	Tuition	0.3%	0.2%	0.4%	0.1%	0.1%
F1	Food at home	1.5%	1.1%	0.6%	3.3%	2.9%
F2	Food away from home	0.4%	0.5%	0.4%	1.0%	1.1%
F3	Alochol and tobacco	0.2%	0.1%	0.1%	0.3%	0.3%
H0	Shelter	-6.3%	-6.1%	-5.0%	-16.2%	-18.4%
H1	Household utilities	0.7%	0.5%	0.3%	1.9%	1.8%
H2	Housefurnishings and operations	0.2%	0.3%	0.4%	0.7%	1.0%
M0	Medical Care	0.6%	0.7%	0.5%	3.2%	3.9%
R0	Entertainment and recreation	0.3%	0.4%	0.4%	0.7%	1.0%
T0	Vehicles	0.4%	0.6%	0.6%	0.9%	1.9%
T1	Gasoline and vehicle maintenance	0.9%	1.0%	0.6%	1.9%	2.3%
T2	Public transportation	0.1%	0.1%	0.1%	0.2%	0.2%

NL Non Social Security, Low Income Quintile
NM Non Social Security, Middle Income Quintiles
NH Non Social Security, High Income Quintiles
SL Social Security, Low Income Quintile
SM Social Security, Middle Income Quintiles

Substitution bias evidence, by cohort