



JSPS Grants-in-Aid for Creative Scientific Research

Understanding Inflation Dynamics of the Japanese Economy

## Replicating Japanese CPI Using Scanner Data

Satoshi Imai  
Statistics Bureau of Japan

Tsutomu Watanabe  
University of Tokyo

May 20, 2015

## Scanner data based price index may be inconsistent with CPI

- Offer prices may have different time series properties from transaction prices.
  - Price tag information collected by price collectors may not be the same as information contained in scanner data, which is based on actual transaction prices.
  - As shown by the literature on asset pricing in financial markets, quoted prices, such as prices quoted by stock market dealers, have different time series properties than transaction prices in terms of volatility, serial correlation, and so on.
  - In Japan, price collectors are instructed by the statistics bureau to ask its regular price when the target product is on sale. On the other hand, one can apply various filtering techniques to scanner data to estimate regular prices. However, it is likely that regular prices estimated this way may differ from regular prices obtained by price collectors
- Differences in sampling procedures may yield different time series properties
  - Scanner data provides information on the number of customer visits for outlets and the quantities sold for products, which can be used when conducting outlet and product sampling. However, there is no guarantee that the set of outlets and the set of products chosen this way coincides with the one chosen based on the current procedure.

Table 1: Number of Outlets, Products and Observations in the Scanner Data

	No. of outlets	Entries	Exits	No. of products	No. of observations
2000	27	-	-	121,427	37,447,555
2001	27	0	0	129,848	40,632,653
2002	28	1	0	136,769	43,100,683
2003	28	0	0	136,663	39,347,502
2004	31	3	0	138,304	43,481,768
2005	28	1	4	135,222	44,197,393
2006	27	0	1	141,382	45,847,962
2007	32	5	0	146,165	44,291,942
2008	29	1	4	149,106	46,317,820
2009	28	0	1	142,518	45,808,810
2010	28	0	0	141,630	45,892,049
2011	28	0	0	143,821	45,559,906
2012	30	2	0	146,198	47,687,953
2013	32	3	1	151,387	50,038,122
2014*	30	0	2	124,933	29,430,411

\* January 2014 to July 2014

# Items and weights in the CPI data

Code	Item	Weight
1001	Non-glutinous rice (single ingredient, "Koshihikari")	23
1002	Non-glutinous rice (single ingredient, excluding "Koshihikari")	31
1011	Glutinous rice	3
1031	Boiled "Udon" (wheat noodles)	8
1041	Dried "Udon" (wheat noodles)	6
1042	Spaghetti	4
1051	Instant noodles	11
1052	Uncooked Chinese noodles	11
1071	Wheat flour	2
1081	Mochi, rice-cakes	9
1151	Satsumaage, fried fish-paste patties	6
1152	Chikuwa, baked fish-paste bars	4
1153	Kamaboko, steamed fish-paste cakes	7
1161	Dried bonito fillets	3
1163	Shiokara, salted fish guts	1
1166	Gyokai-tsukudani, fish boiled in soy sauce	3
1173	Canned tuna fish	6
1252	Ham	16
1271	Bacon	6
1303	Fresh milk (sold in stores, in cartons)	36
1311	Powdered milk	2
1321	Butter	3
1333	Yogurt	23
1341	Hen eggs	18
1451	Azuki, red beans	2
1453	Shiitake, Japanese mushrooms, dried	2
1463	Dried tangle	3
1471	Tofu, bean curd	14
1472	Fried bean curd	8
1473	Natto, fermented soybeans	8
1481	Konnyaku, devil's-tongue jelly	5
1482	Umeboshi, pickled Japanes apricot	9
1483	Pickled radishes	5
1485	Tangle prepared in soy sauce	4
1486	Pickled chinese cabbage	4
1601	Edible oil	7
1602	Margarine	2
1611	Salt	2
1621	Soy sauce	5
1631	Soybean paste	7
1632	Sugar	3
1633	Vinegar	4
1641	Worcester sauce	2
1642	Tomato ketchup	2
1643	Mayonnaise	4
1652	Instant curry mix	5
1654	Flavor seasonings	6

Code	Item	Weight
1655	Liquid seasonings	12
1656	Furikake, granular flavor seasonings	4
1701	Yokan, sweet bean jelly	14
1714	Pudding	7
1721	Biscuits	13
1732	Candies	7
1741	Senbei, Japanese rice crackers	19
1761	Chocolate	17
1772	Peanuts	2
1783	Potato chips	11
1784	Jelly	9
1791	Box lunch	47
1811	Salad	18
1851	Frozen croquettes	9
1871	Cooked curry	3
1881	Gyoza, Chinese meat dumpling	15
1902	Green tea ("Sencha")	16
1911	Black tea	4
1921	Instant coffee	6
1922	Coffee beans	6
1931	Fruit drinks (20-50% fruit juice)	6
1941	Vegetable juice	10
1951	Cola drinks	12
2003	Sake	14
2021	Beer	33
2033	Whisky (40% or more and less than 41% alcohol)	3
2041	Wine	4
4401	Food wrap	3
4412	Facial tissue	7
4413	Rolled toilet paper	10
4431	Liquid detergent, kitchen	9
4441	Detergent, laundry	11
4442	Fabric softener	4
4451	Insecticide	6
4461	Moth repellent for clothes	1
6101	Sanitary napkins	9
6141	Disposable diapers (baby)	5
9121	Notebooks	6
9124	Cellophane adhesive tape	2
9195	Dry batteries	5
9611	Toothbrushes	3
9621	Toilet soap	3
9622	Shampoo	9
9623	Toothpaste	6
9631	Hair dressing	8
9641	Hair tonic	4
9661	Face lotion	20

# Methodology

## ■ Sampling

- Outlet sampling
- Product sampling
- Price sampling
  - JSB price collectors collect prices on either Wednesday, Thursday, or Friday of the week which includes the 12th of the month.
  - JSB price collectors exclude “extra-low prices due to bargain, clearance, or discount sales, and quoted for less than eight days”

## ■ Quality adjustment

- Quantity ratio method
- Imputation method

## ■ Aggregation

- Unweighted arithmetic mean of individual prices (i.e., Dutot index) for the lower level aggregation
- Fixed base Laspeyres weighting for the upper level aggregation

# Number of Products that Meet the JSB Product Type Specifications

Item code	Description	No. of JAN codes (A)	No. of JAN codes that meet the product specifications (B)	(B/A)
1001	Rice-A (domestic)	11962	1649	0.138
1002	Rice-B (domestic)	11962	1905	0.159
1011	Glutinous rice	477	321	0.673
1031	Boiled noodles	4944	1213	0.245
1041	Dried noodles	2194	37	0.017
1042	Spaghetti	1410	237	0.168
1051	Instant noodles	6879	6	0.001
1052	Uncooked Chinese noodles	8042	2439	0.303
	⋮			
1321	Butter	369	30	0.081

## JSB Product Type Specifications for Butter

Jul 1996 - Jan 2001

“Snow Brand Hokkaido Butter”

Jan 2001 – present

200g. Packed in a paper container. Excluding unsalted butter.

# Outlet and Product Sampling

$$\Pr (\text{Outlet replacement}) = \left[ 1 + \left( k_S \frac{n_j^S}{n_i^S} \right)^{-1} \right]^{-1}$$

$n_i^S$  : Number of customer visits to outlet  $i$ , which is included in the sample

$n_j^S$  : Number of customer visits to outlet  $j$ , which is not included in the sample

$k_S$  : Parameter governing the frequency of outlet replacements

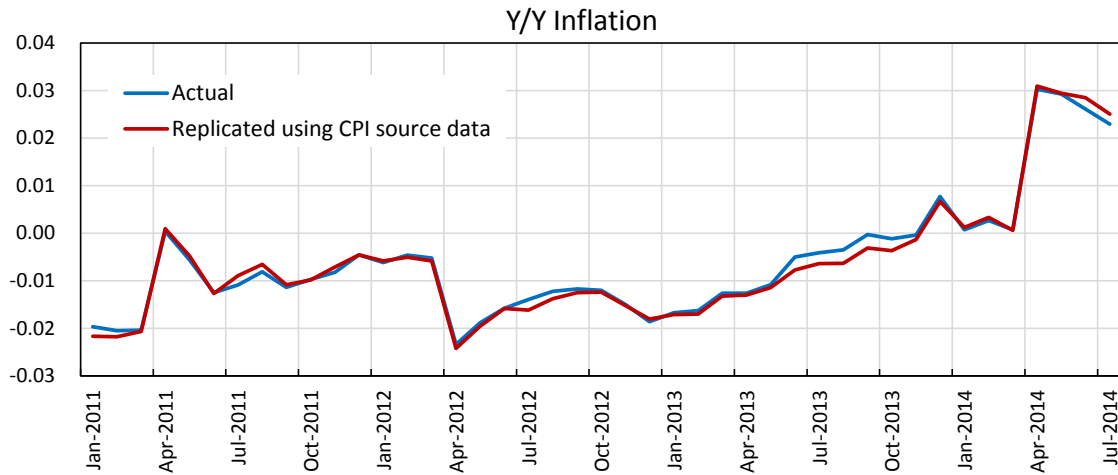
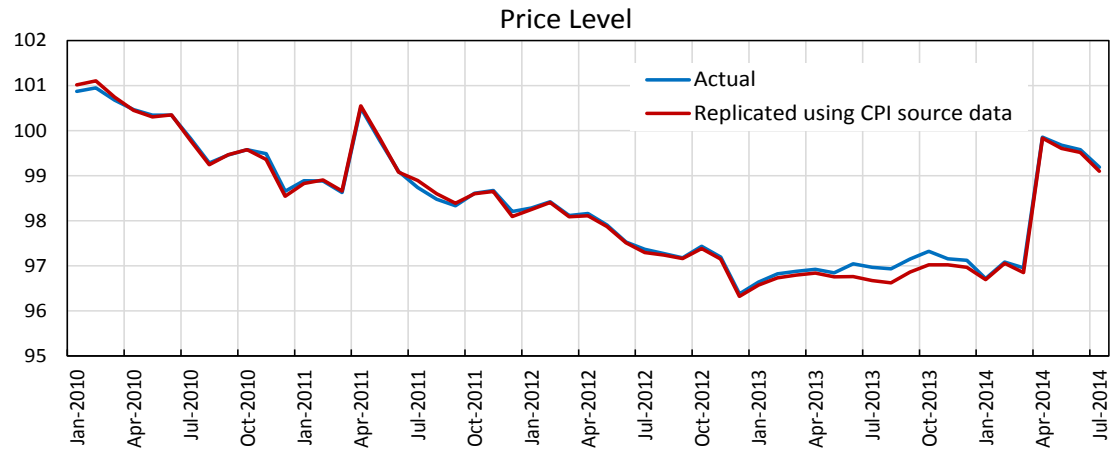
$$\Pr (\text{Product replacement}) = \left[ 1 + \left( k_P \frac{n_j^P}{n_i^P} \right)^{-1} \right]^{-1}$$

$n_i^P$  : Number of quantities sold for product  $i$ , which is included in the sample

$n_j^P$  : Number of quantities sold for product  $j$ , which is not included in the sample

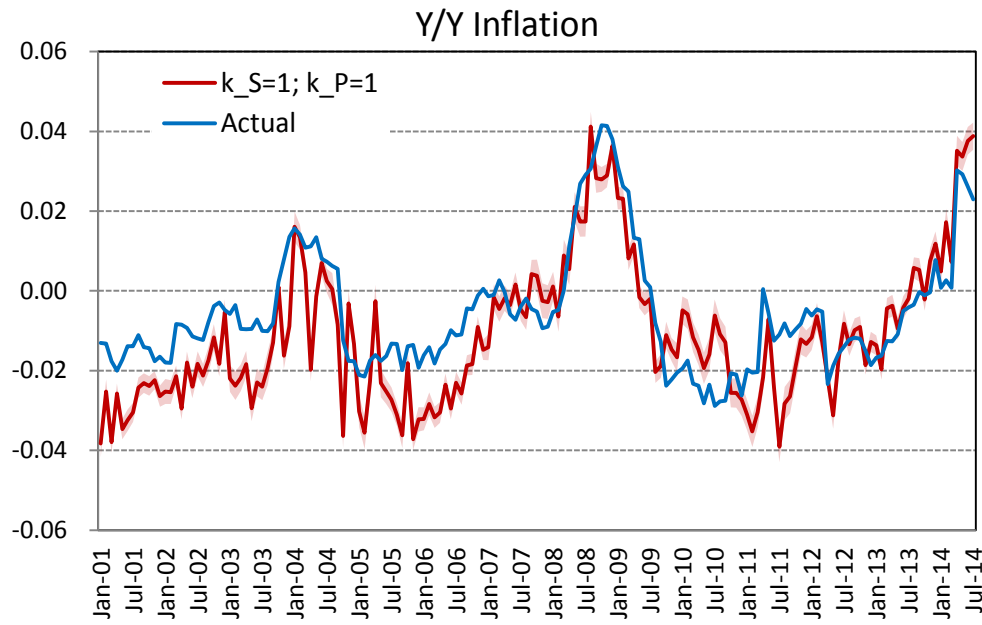
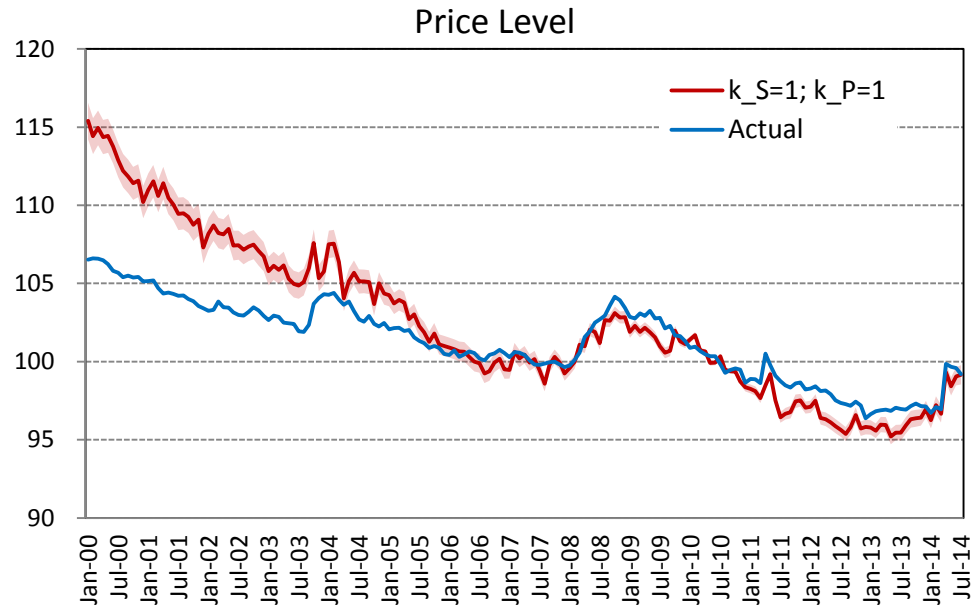
$k_P$  : Parameter governing the frequency of product replacements

# Replication using the CPI source data

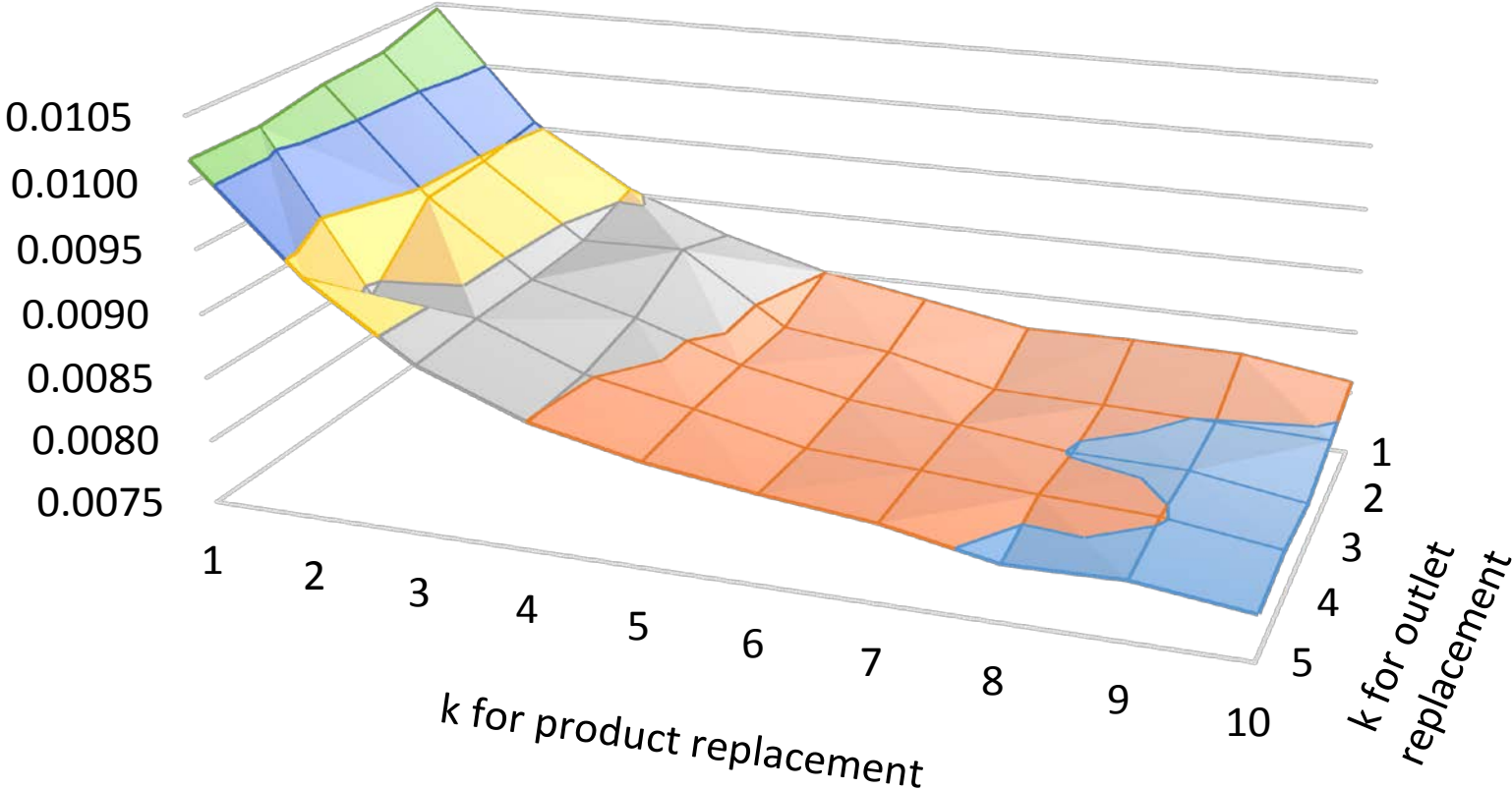




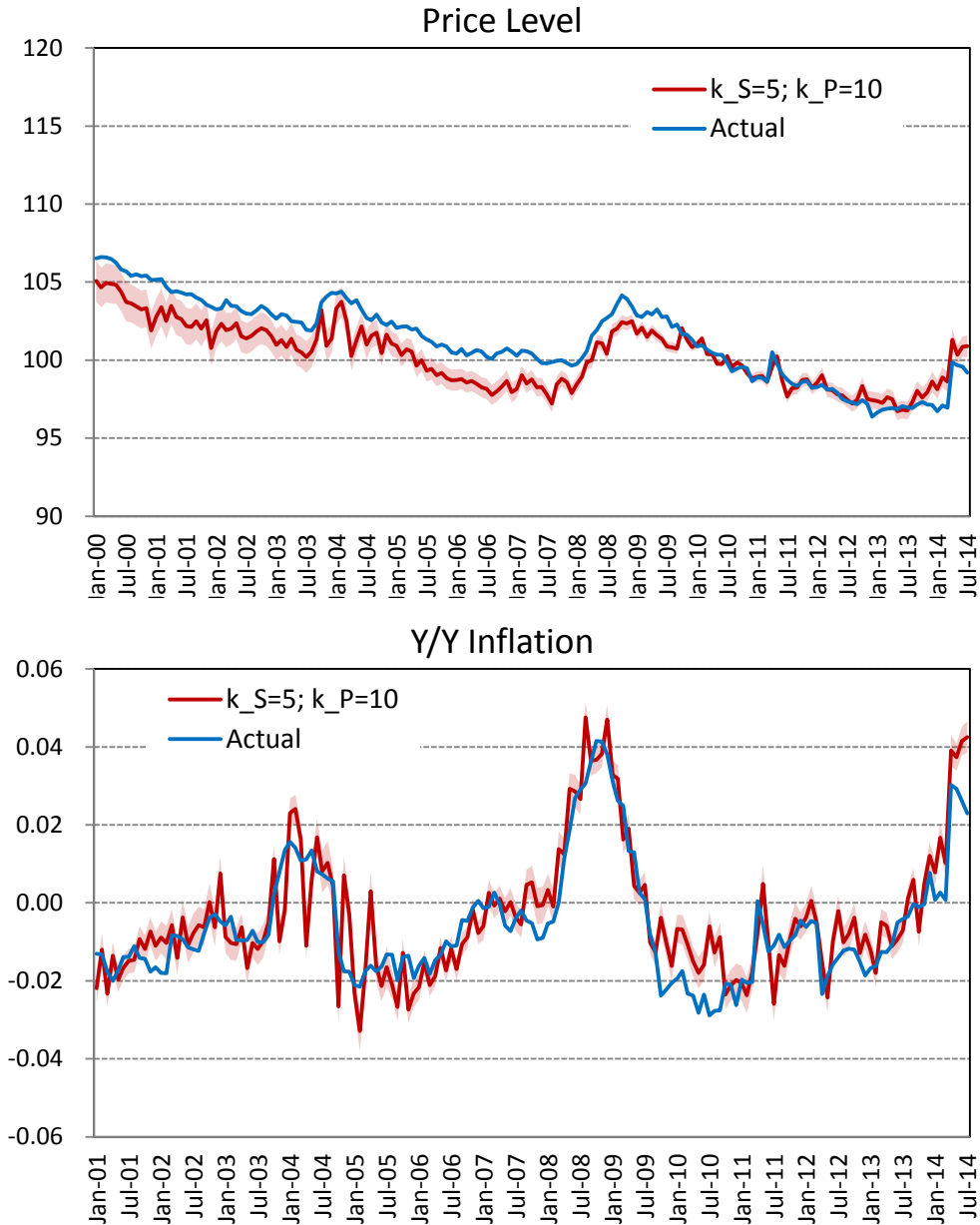
# Replication using the scanner data



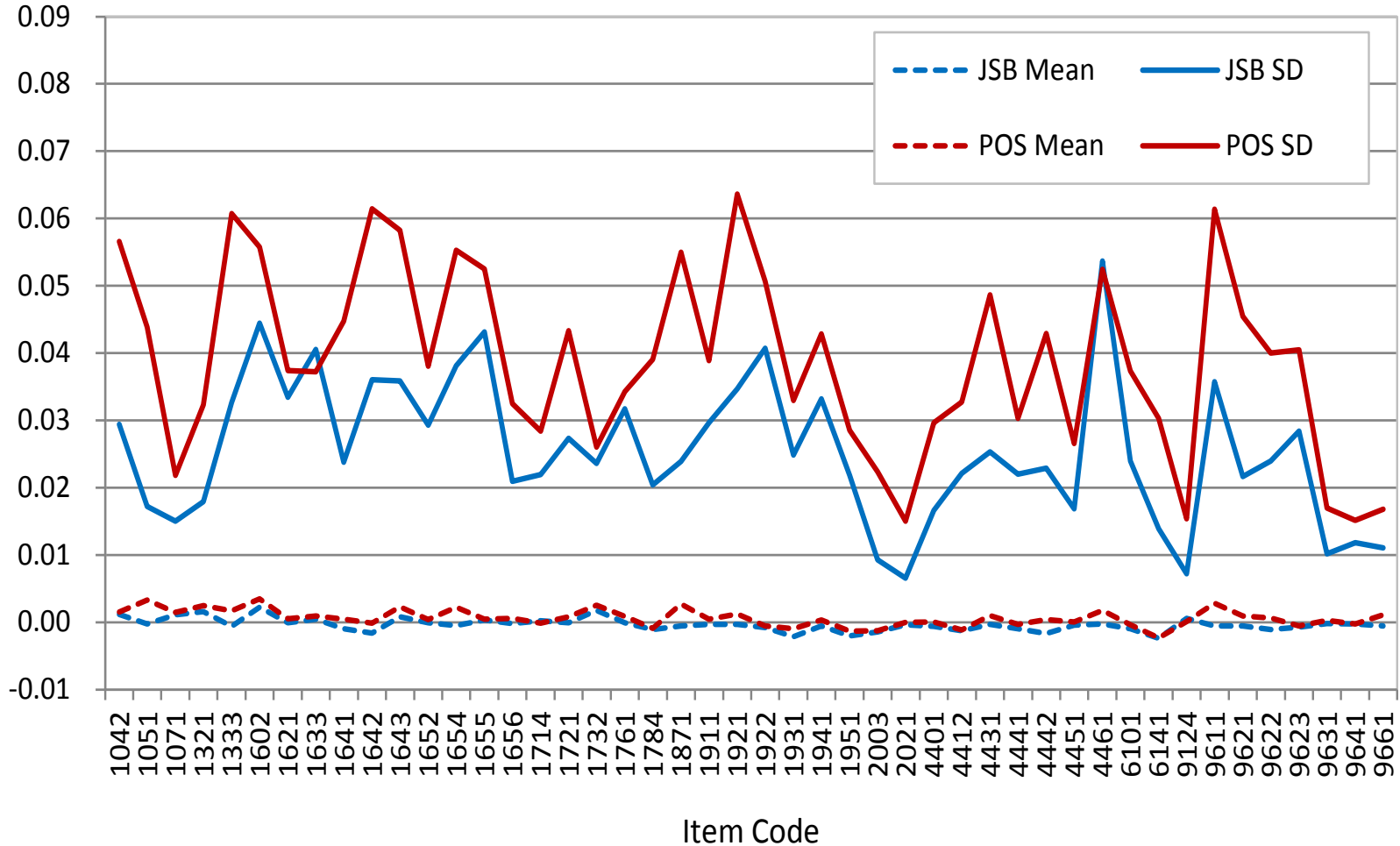
# Root Mean Square Error for the Difference between Scanner Data Based Inflation and CPI Inflation



# Replication with Inertia in Outlet and Product Replacement



# Mean and SD of Monthly Inflation By Item



# Decomposition of Monthly Inflation into Extensive and Intensive Margins

$$\pi_{ct} \equiv \text{Fr}_{ct} \times dP_{ct}$$

Month on month  
Inflation for item c  
in month t

Fraction of products in item c  
that experience price changes  
in month t

Average size of price changes  
for those products in item c  
that experience price changes  
in month t

➔  $Var(\pi_{ct})$

$$= \underbrace{Var(dP_{ct}) [E(\text{Fr}_{ct})]^2}_{\text{Intensive margin}} + \underbrace{Var(\text{Fr}_{ct}) [E(dP_{ct})]^2}_{\text{Extensive margin}} + \text{Other terms}$$

Intensive margin

Extensive margin

$$Var(\pi_{ct})$$

$$= \underbrace{Var(dP_{ct}) [E(Fr_{ct})]^2}_{\text{Intensive margin}} + \underbrace{Var(Fr_{ct}) [E(dP_{ct})]^2}_{\text{Extensive margin}} + \text{Other terms}$$

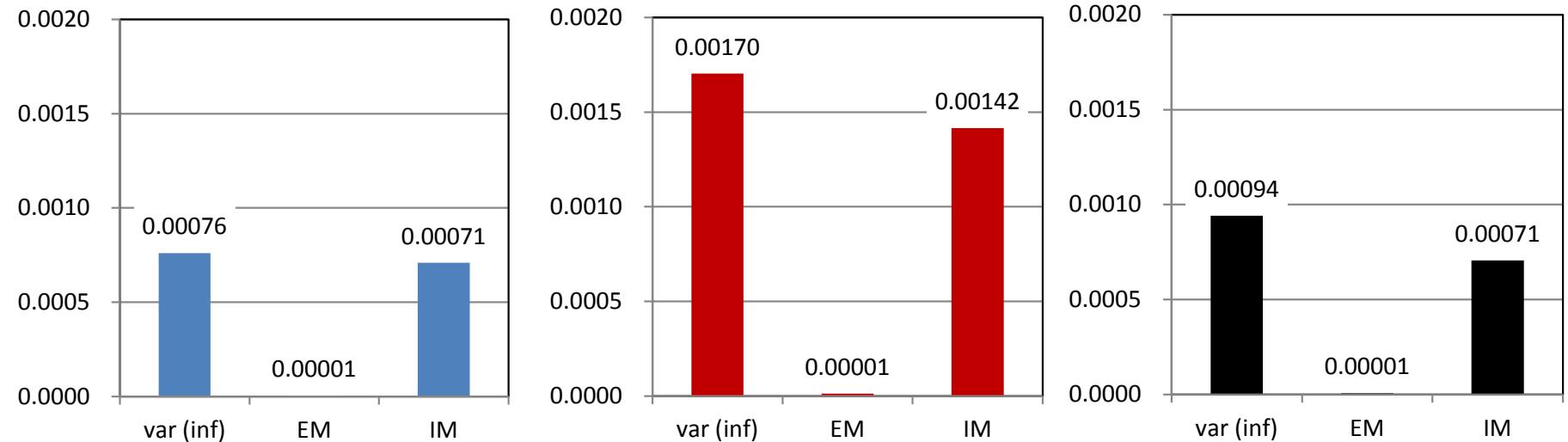
Intensive margin

Extensive margin

JSB

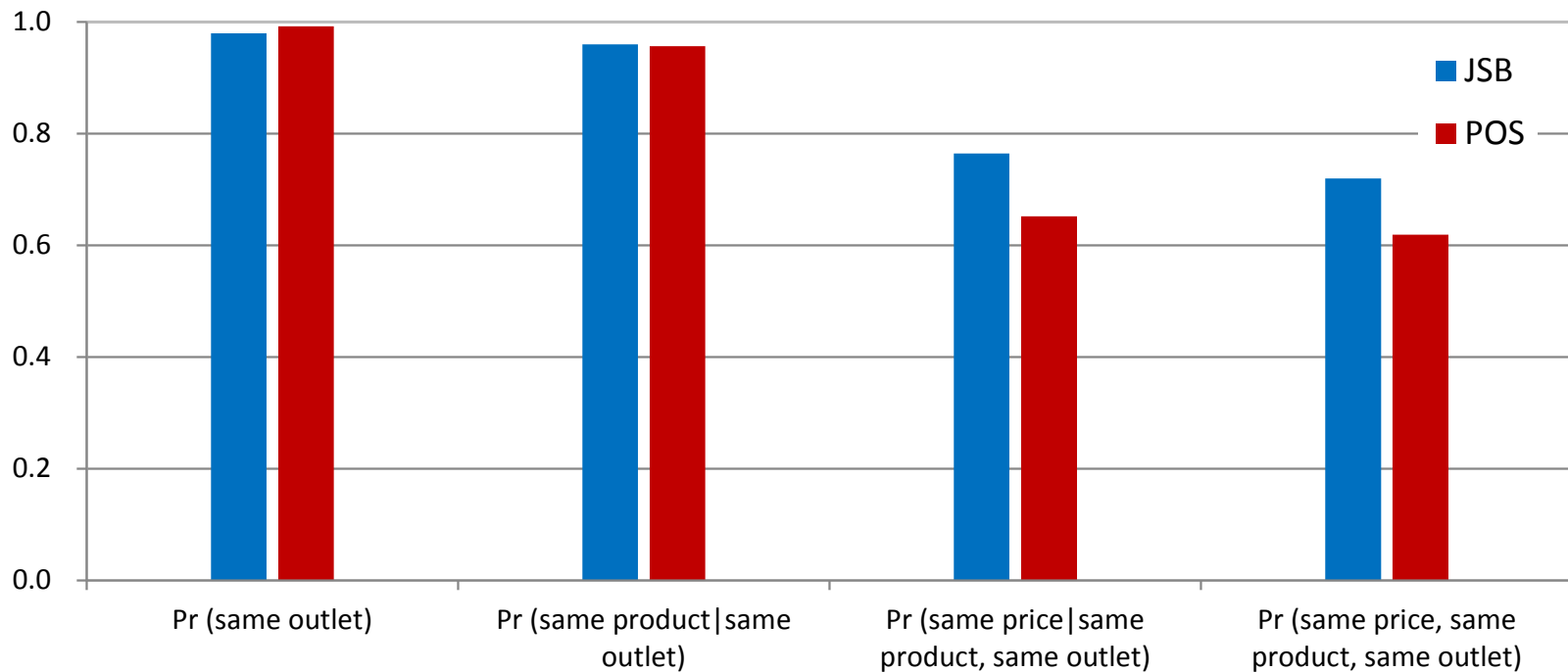
POS

Difference

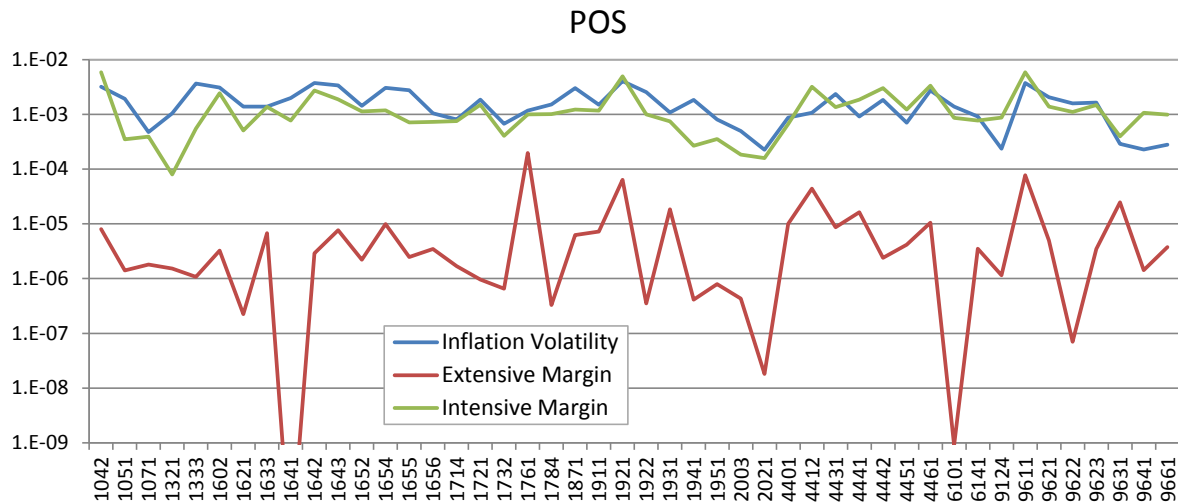
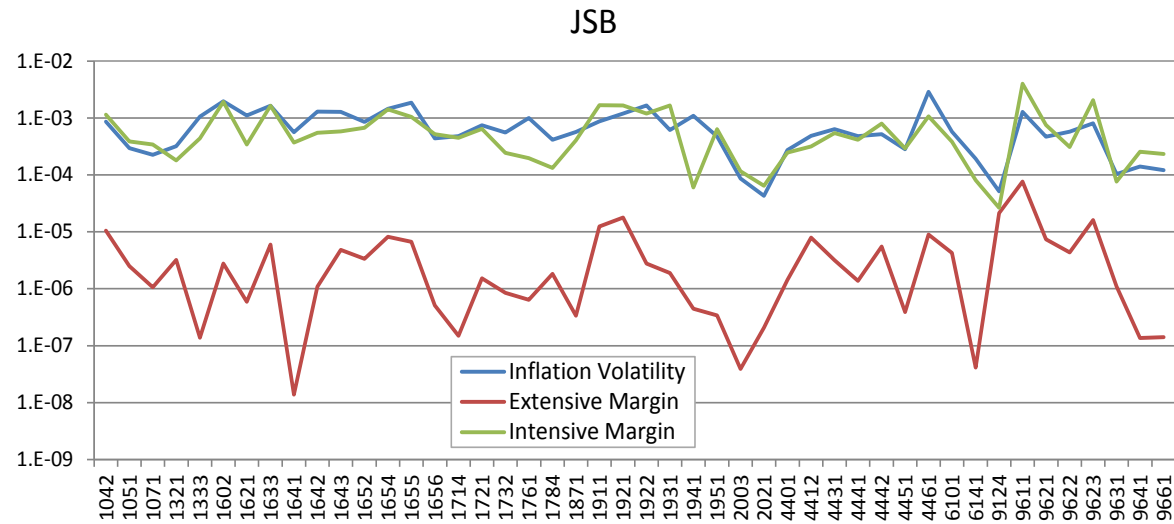


Diff in IM	0.00071
Due to Var (dP)	0.00002
Due to E (Fr)^2	<b>0.00068</b>

$$\begin{aligned}
 \text{Freq of price changes} &= 1 - \Pr(\text{No price change, No product repl., No outlet repl.}) \\
 &= 1 - \Pr(\text{No price change} \mid \text{No product repl., No outlet repl.}) \\
 &\quad \times \Pr(\text{No product repl.} \mid \text{No outlet repl.}) \\
 &\quad \times \Pr(\text{No outlet repl.})
 \end{aligned}$$

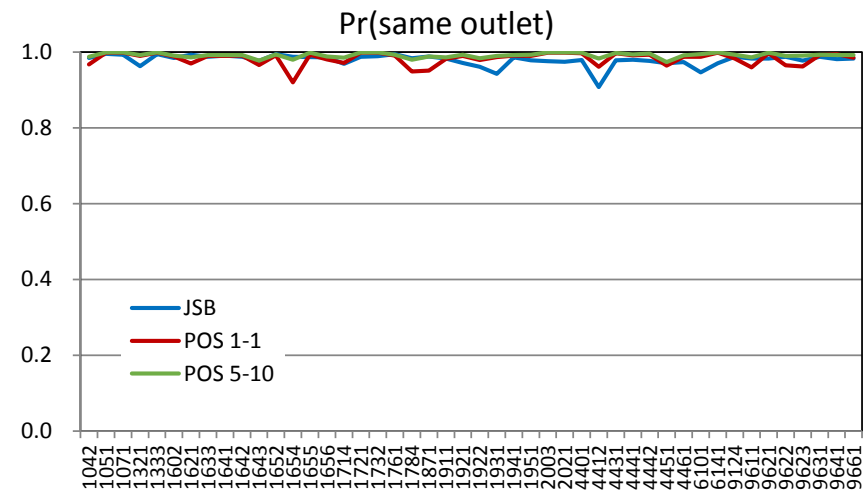
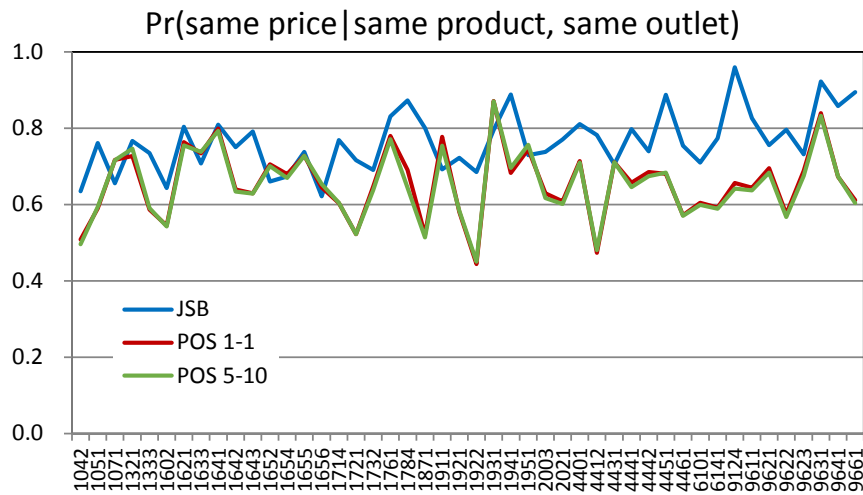
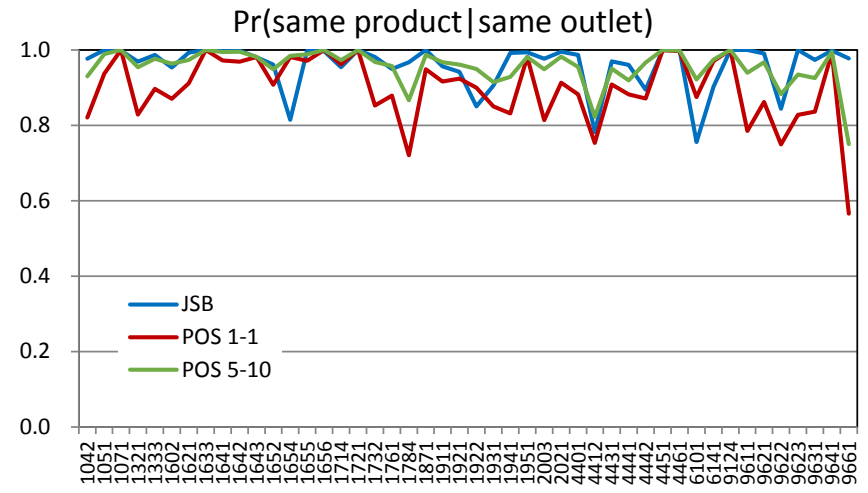
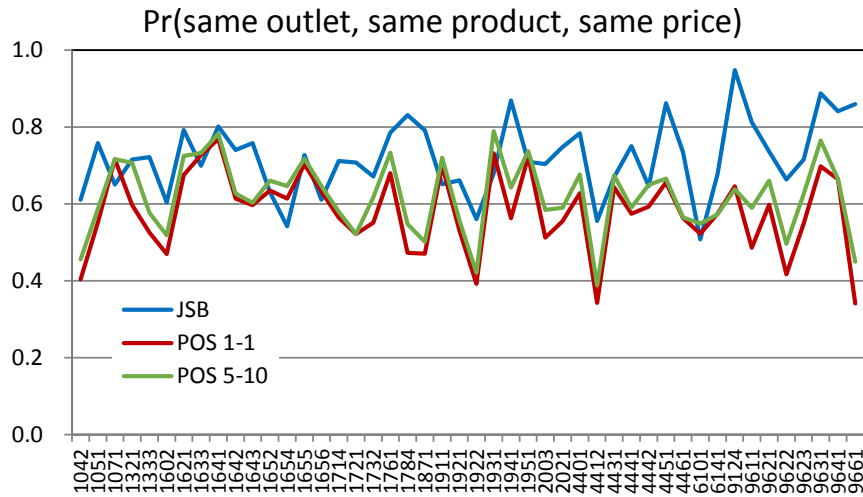


# Decomposition of Inflation Volatility into Extensive and Intensive Margins

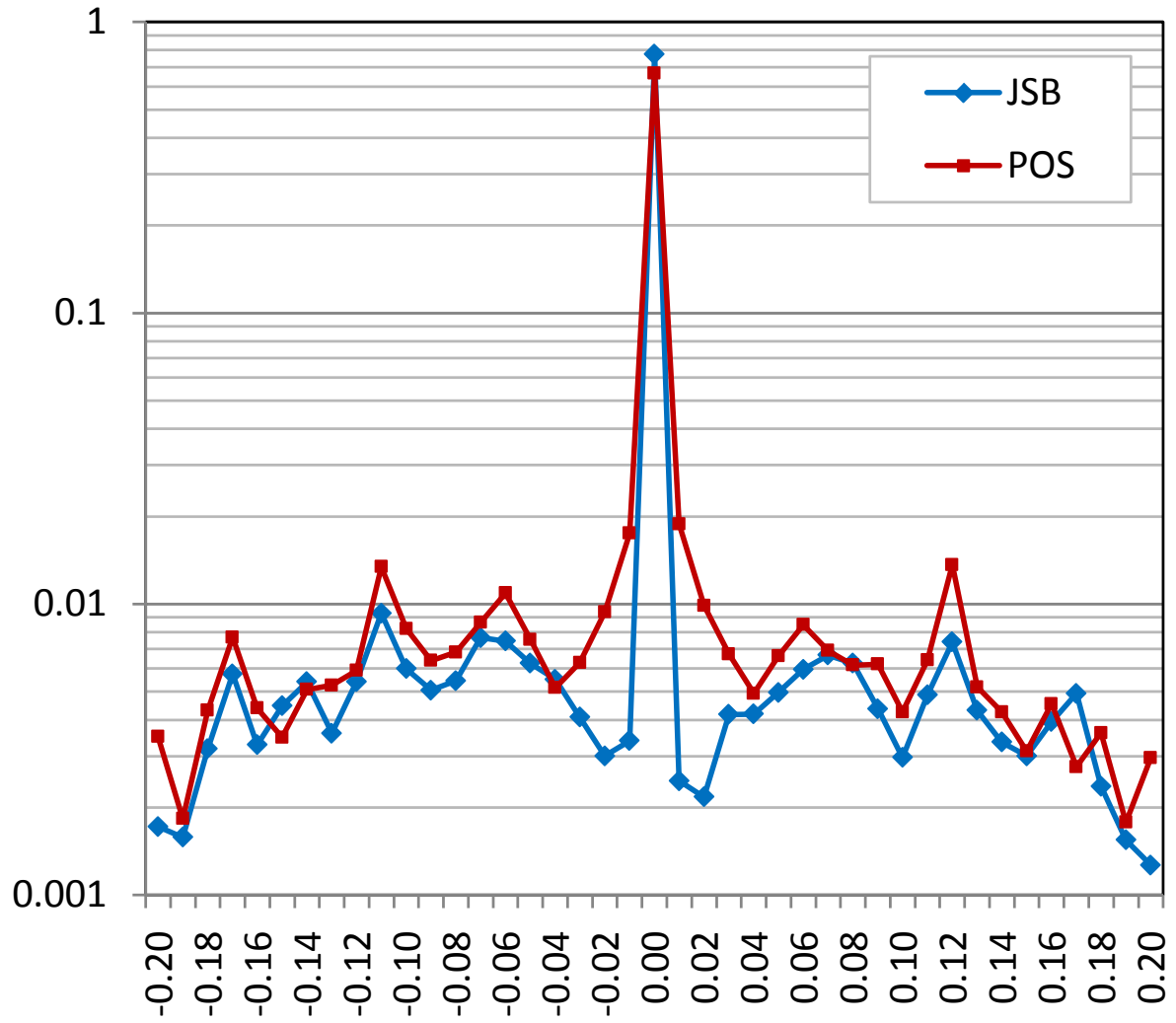
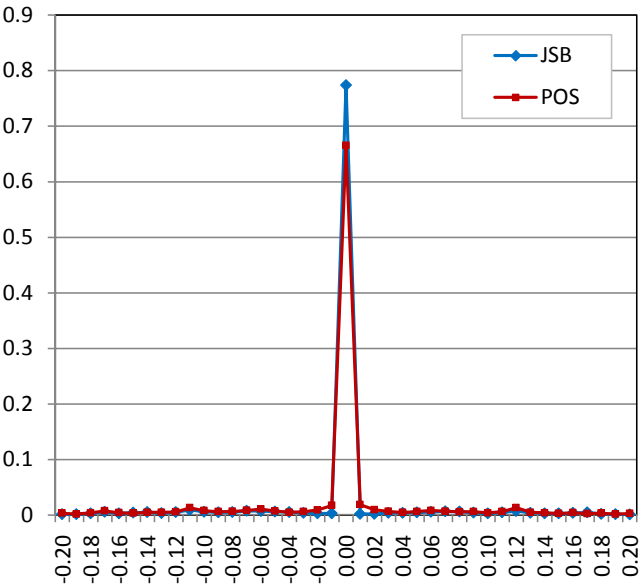




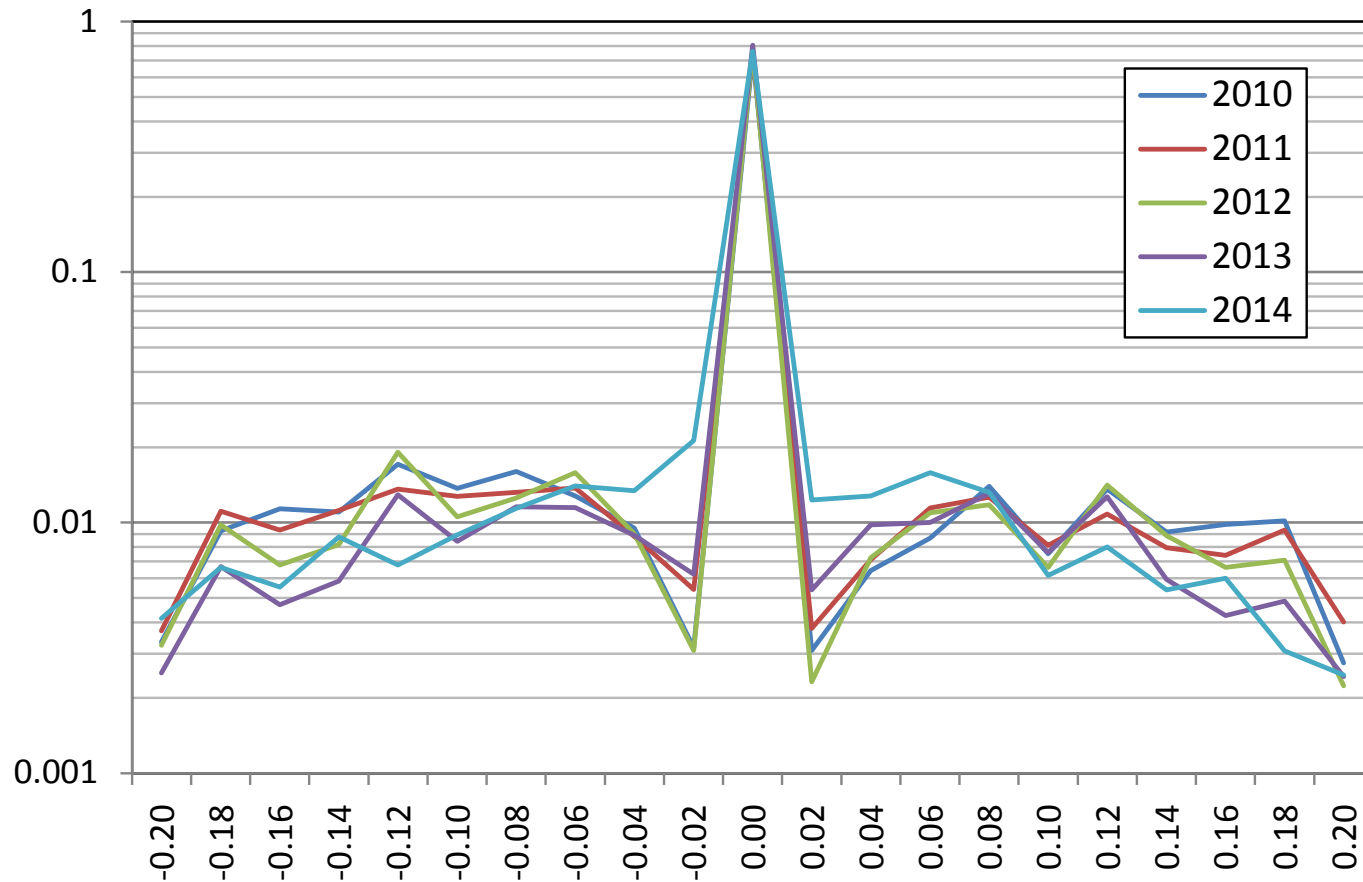
# Probability of No Price Adjustments



# Distributions of price changes for individual products, $dP$



# JSB Price Change Distributions by Year



# Main findings of the paper

- Scanner data based price index depends crucially on how often product substitution occurs. The deviation of scanner data based price index from actual CPI is not negligible when we choose parameter values for product substitution such that substitutions occur very frequently, while it becomes smaller as we pick parameter values such that substitutions occur only infrequently.
- **Scanner data based inflation differs significantly from actual CPI inflation in terms of volatility.** We decompose the difference in the variance of monthly inflation into various factors to find that the difference in inflation volatility mainly stems from the difference in the frequency of price adjustments for individual products. Actual CPI inflation is less volatile since individual prices in the CPI data are stickier.
- **Small-sized price changes are less likely to occur in the CPI data than in the scanner data. Together with the fact that prices are stickier in the CPI data, this suggests that menu costs play a more important role in the CPI data.**