

WPS (DEPR): 05 / 2023

RBI WORKING PAPER SERIES

Reading Consumers' Minds: An Analysis of Inflation Expectations

Purnima Shaw



DEPARTMENT OF ECONOMIC AND POLICY RESEARCH
APRIL 2023

The Reserve Bank of India (RBI) introduced the RBI Working Papers series in March 2011. These papers present research in progress of the staff members of the RBI and at times also those of external co-authors when the research is jointly undertaken. They are disseminated to elicit comments and further debate. The views expressed in these papers are those of the authors and not necessarily those of the institution(s) to which they belong. Comments and observations may please be forwarded to the authors. Citation and use of such papers should take into account its provisional character.

Copyright: Reserve Bank of India 2023

Reading Consumers' Minds: An Analysis of Inflation Expectations

Purnima Shaw¹

Abstract

The heterogeneity in the consumption baskets of households is often deemed responsible for the deviation of households' inflation expectations from the headline inflation number. A novel approach is proposed in this paper to verify this by simulating heterogeneous population consumption baskets and estimating the mean inflation by sampling the baskets. The estimated mean inflation, treated as inflation sentiment, fails to display closeness with the survey numbers. Therefore, the paper proposes alternative logical methods for designing basket compositions and identifies the most suited method using which the estimated expectations are found to be close to and well-correlated with survey numbers. Such an attempt to find out the source(s) of disagreement in inflation expectations with respect to official inflation can be useful in understanding consumers' inflationary expectations better for inflation analysis. The findings suggest that a sudden rise in inflation in items of regular use can explain the deviation in households' inflation expectations from the official inflation. The deviation of survey expectations from the headline inflation can, thus, be explained effectively, over and above the other factors viz., demographic characteristics and exposure to media reports, which influence the formation of inflation expectations.

JEL Classification: D84, E20, E21, E30, E31

Keywords: Consumer price index, consumption basket, inflation, inflation expectations, sampling

¹ Purnima Shaw (pshaw@rbi.org.in) is Assistant Adviser in the Department of Statistics and Information Management (DSIM), Reserve Bank of India (RBI). Authors are thankful to Jayaraman Alur Raghavan, Sourajyoti Sardar, Pallavi Chavan, members of the DSIM Internal Advisory Group, participants of the Department of Economic and Policy Research (DEPR) Study Circle and the Editorial Board for providing useful comments which have enriched the content of this paper. The views and opinions expressed in this paper are those of the author and do not necessarily represent views of the RBI.

Reading Consumers' Minds: An Analysis of Inflation Expectations

Introduction

Inflation expectations are vital for monetary policy communication, as they affect the wage-setting behaviour, consumption, expenditure, investment, financial savings, *etc.* Hence, inflation targeting can be regarded as successful if the expectations of the economic agents are anchored effectively. Anchored inflation expectations reflect the trust of economic agents in central banks. The empirical measures of inflation expectations, therefore, form an important parameter in gauging the credibility of central banks. For this, the data on consumers sentiments need to be analysed carefully, including the shocks impacting households' inflation expectations.

Variations in households' consumption baskets are often deemed responsible for the deviation of their inflationary expectations from the official inflation numbers. However, this has not been verified empirically in the literature. It is anticipated that mean inflation in simulated heterogeneous consumer baskets should be well-correlated and close to the survey-based expectations. We design a novel approach to verify this empirically. A large number of heterogeneous consumption baskets are simulated. Treating this as the population, the mean inflation in the population baskets is estimated. The estimated mean inflation in the population baskets assumed as the simulated inflation sentiment fails to display a closeness with the survey numbers. Therefore, this paper explores alternative ways for simulating consumption baskets to find a logical method by which the simulated sentiments are closely related to the survey-based sentiments. The analysis helps to provide insights into the minds of consumers/ households as they form their expectations. This attempt to identify the source(s) of disagreement in inflation expectations with respect to the official inflation can help in better understanding and using the consumers' inflationary expectations.

Section II of this paper provides a study of the existing literature on the disagreement in survey-based expectations with respect to the official inflation. Section III starts with the presumption that heterogeneous consumption baskets is the main reason for such a deviation. It develops a logical method to simulate heterogeneous consumption baskets to form a population of baskets, and then, sample the baskets from this simulated population. We then proceed to test the initial presumption empirically. On finding limited empirical evidence to support this presumption, we develop other logical methods of simulating the consumption baskets. The motive is to find the best logical

method for simulating the consumption baskets such that the estimated inflation would be close in number bearing a high correlation with the survey-based sentiments. Section IV examines the proposed logics numerically using the Consumer Price Index (CPI) basket. Section V concludes.

II. Literature Review

Flexible inflation targeting critically hinges on measures to enhance agents' confidence in central banks. Forward-looking monetary policies aim to study the future inflation dynamics by feeding forward-looking inflation expectations into the New Keynesian Phillips Curve (NKPC). However, expectations may not be perfectly rational, and hence, may not fit the NKPC. Hence, the empirical studies for most inflation targeting countries use a hybrid version of NKPC taking both forward and backward-looking components (Taylor, 1982; Gali and Gertlar, 1999 and Pattanaik *et al.*, 2020). Another way of visualising the future price levels is using the well-known Dynamic Aggregate Demand – Surprise Aggregate Supply (DAD-SAS) model by which forward-looking (rational) inflation expectations assure an economy to reach the potential output level relatively faster than that of the backward-looking expectations through the wage channel (Alpanda *et al.*, 2013 and Man and Peterson, 2019).

Inflation expectations of agents can be influenced by various factors. For instance, Murphy and Rohde (2018) opine that food prices are weighted more heavily by consumers than other consumer goods in the US. Zhao's (2021) analysis brings out how inflation expectations of the US consumers is impacted by their sentiments on the prevailing economic conditions, their memory about news reports, and their social and demographic characteristics. Ciro and Zapata (2019) state that the disagreements in inflation expectations increase due to inflation volatility.

The Reserve Bank of India (RBI) has been tracking inflationary sentiments of households by conducting the Inflation Expectations Survey of Households (IESH) on a bimonthly basis in 19 cities in India. Generally, these expectations are in disagreement with the official CPI inflation. The correlations of households' expectations for different lags (0, 1 and 2) and leads (0 and 1) of the CPI-Urban (CPI-U) inflation are shown in Table 1. The gap between households' inflation expectations and the CPI-U inflation is also tabulated. A high and significant correlation is preferred. A significant gap between the CPI-U inflation and the expectations is not desirable. It is observed that although the correlations are significant, the gaps between actual and estimated sentiments are in the range of 3.5 to 5.1 percentage points (Table 1 and Chart 1).

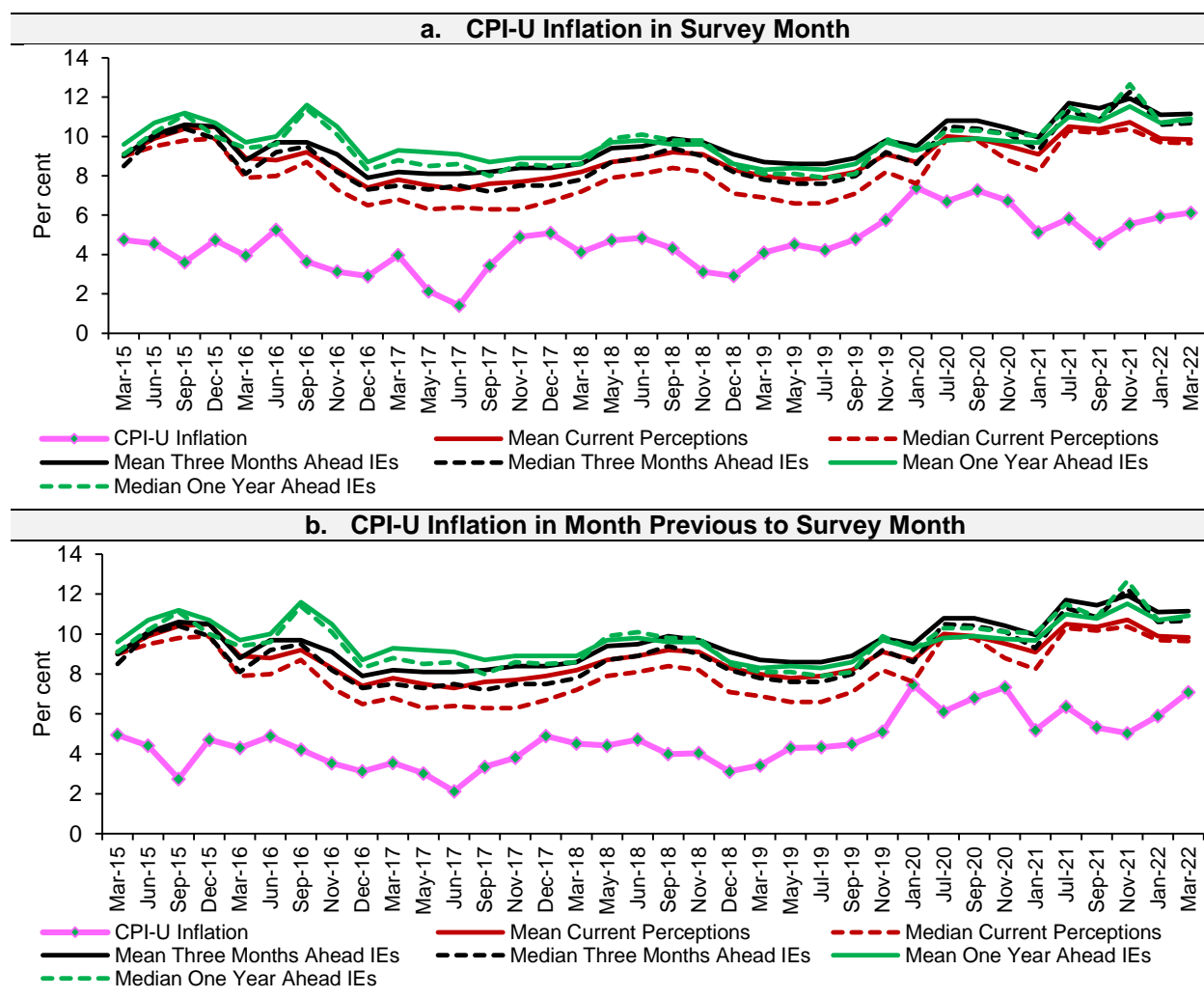
Table 1: Comparison of Households' Inflation Expectations with CPI-U Inflation

Reference to Survey Month	Measure	CPI-U Inflation	Current		Three Months Ahead		One Year Ahead	
			Mean	Median	Mean	Median	Mean	Median
Survey Month	Correlation	Lag 0	0.54	0.54	0.60	0.56	0.25	0.39
		Lag 1	0.61	0.62	0.63	0.62	0.39	0.52
		Lag 2	0.68	0.70	0.72	0.72	0.49	0.63
		Lead 1	0.39	0.38	0.44	0.40	0.05	0.23
		Lead 2	0.31	0.27	0.40	0.33	0.00	0.11
Previous Month	Gap	Lag 0	4.28	3.51	4.96	4.34	5.10	4.99
		Correlation Lag 0	0.56	0.57	0.63	0.59	0.32	0.44
		Gap Lag 0	4.26	3.49	4.93	4.32	5.08	4.97

Note: Figures in bold indicate significance at 5 per cent level.

Sources: MOSPI; RBI; and Author's calculation.

Chart 1. Inflation Perceptions and Expectations vis-à-vis CPI-U Inflation



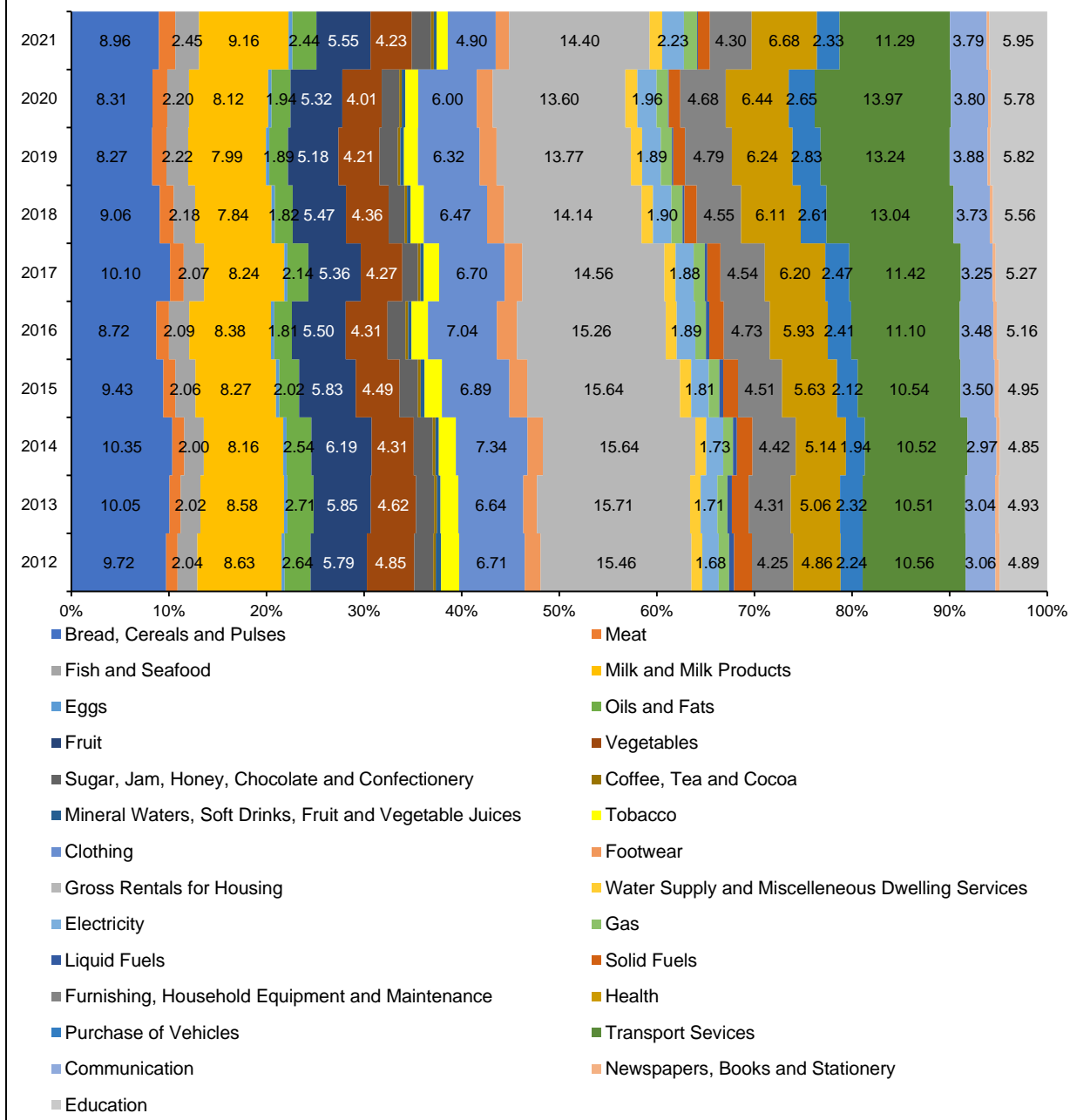
Sources: MOSPI; RBI; and Author's calculation.

India is not the only country showing a disagreement in survey-based inflation expectations with respect to the official inflation. It has been proven recently that such a disagreement exists in several other economies; India, in fact, is similar to Russia in this regard (Singh *et al.*, 2022). Shaw (2019) took an empirical Bayesian approach to handle this disagreement for predicting the future inflation trajectory in the Indian case.

Considering the pattern of private final consumption expenditure (2011-12), it is observed that per capita consumption on health, transport services and education increased (Chart 2). However, these changes are not reflected in the consumption weights in official inflation. Changes in consumption pattern may also be one of the factors responsible for disagreement in households' inflation expectations with respect to the official inflation.

Other factors, such as differences in consumption baskets, demography of consumers and exposure to media can also induce heterogeneity in households' inflationary sentiments. Some of these factors are documented in the literature from other economies. For example, Souleles (2004) shows the systemic correlation between consumers' inflation forecast errors and their demographic characteristics. About media exposure, Carroll (2003) is of the view that the stickiness in consumers' inflation expectations is due to their infrequent attention to media reports. Ehrmann *et al.* (2015) identified socio-economic factors *viz.*, income, age, gender and buying attitudes of consumers as crucial determinants of inflation expectations. This paper focuses on an important factor, as identified in the literature, namely, the differences in households' consumption baskets in explaining the disagreement in inflation expectations with respect to the official inflation.

Chart 2. Change in Consumption Pattern per capita (in Per cent)



Sources: CEIC Database and Author's calculation.

III. Methodology

In the following paragraphs, we propose different methods for simulating a population of consumption baskets; each one of these methods is based on different logic and set of assumptions.

III.1 Method 1

Using this method, a population of consumption baskets is simulated in such a way that a set of baskets contain a fixed number of items randomly drawn from the consumption groups and the rest of the baskets contain a varying number of items randomly drawn from the consumption groups. Let there be R items based on which official inflation figures are computed. To simulate a consumption basket, p_i (predefined), $0 < p_i < 1$, proportion of items are drawn randomly without replacement from each of the G subgroups and B_i such baskets are simulated. Next, based on this logic of drawing fixed proportion of items from each of the subgroups, I sets of population baskets are simulated by changing the values of p_i and B_i , $i = 1, 2, \dots, I$.

With an objective to add heterogeneity in the population consumption baskets, D_1 baskets are simulated, wherein the shares of items to be drawn without replacement from each of the subgroups are randomly drawn with replacement for each basket from a predefined set of proportions ($0 < p_1, p_2, \dots, p_I < 1$). Hence, the shares of items drawn from each of the subgroups vary across the baskets. Next, D_2 baskets are simulated, with the proportions of items to be selected without replacement from each of the subgroups are randomly selected with replacement for each basket from a predefined set of proportion numbers ($0 < p_{L1}, p_{L2}, \dots, p_{Lq} < 1$) which are low-valued proportions with an understanding to represent the low-income categories of households. Lastly, D_3 baskets are simulated, where, the proportions of items to be selected without replacement from each of the subgroups are randomly drawn with replacement for each basket from a predefined set of proportion numbers ($0 < p_{U1}, p_{U2}, \dots, p_{Uq} < 1$) which are comparatively high-valued proportions to represent the economically well-off sections of households.

Now, to derive the consolidated index for each of the $H = (\sum_{i=1}^I B_i + D_1 + D_2 + D_3)$ simulated baskets, let y_{thlk} be the price index for the k^{th} item in the l^{th} subgroup for the h^{th} basket in the t^{th} month. Then, the combined index pertaining to the h^{th} simulated basket for month t using Method 1 is,

$$CInF_t^{1h} = \frac{\sum_{l=1}^L W_l \frac{\sum_{k=1}^K w_{lk} y_{thlk}}{\sum_{k=1}^K w_{lk}}}{\sum_{l=1}^L W_l}, h = 1, 2, \dots, H, t = 1, 2, \dots, T \quad (1)$$

Where, w_{lk} is the weight of the k^{th} item in the l^{th} subgroup and W_l is the weight of the l^{th} subgroup. Let $CInF_t^{1h}$ denote the inflation at month t computed for basket h , $h = 1, 2, \dots, H$ using Method 1. It may be noted that in this method, the proportion of items selected from

each subgroup in any particular basket and the composition of each basket is kept unchanged across months.

III.1.i Sampling and Estimation

We now sample the baskets from the simulated population and estimate the average inflation in the baskets for each month. These estimates are considered as inflation estimates obtained from population baskets simulated using plausible logic of heterogeneous consumption baskets. Consider the finite population $U = (1, 2, \dots, h, \dots, H)$ on which the vector of values $CInf_t^h$ (obtained by using a particular method) for h in U of a real variable $CInf_t$ and x_h for h in U of a real variable x , which is considered as the size-measure, are defined. If the size-measure values are fractions or small numbers, then these are multiplied with a suitable integer power of 10 to convert into integers or manageable numbers. It is required to draw a sample, say, s , from U with a pre-assigned probability $p(s)$ admitting positive first and second-order inclusion-probabilities $\pi_h = \sum_{s \ni h} p(s)$ and $\pi_{hg} = \sum_{s \ni h, g} p(s)$, for $g \neq h$ and $h = 1, 2, \dots, H$. The population mean of the inflation in the H baskets at month t which is required to be estimated, can be denoted by \overline{CInf}_t , where,

$$\overline{CInf}_t = \frac{1}{H} \sum_{h=1}^H CInf_t^h \quad (2)$$

Taking, E_P and V_P as the design-based expectation and variance operators, respectively, consider the estimator $\widehat{\overline{CInf}}_t$, where,

$$\widehat{\overline{CInf}}_t = \frac{1}{H} \sum_{h \in s} \frac{CInf_t^h}{\pi_h} \quad (3)$$

Then, as $E_P(\widehat{\overline{CInf}}_t) = \overline{CInf}_t$, $\widehat{\overline{CInf}}_t$ is an unbiased estimator of \overline{CInf}_t . Using Chaudhuri and Pal (2002), variance of $\widehat{\overline{CInf}}_t$ is derived as,

$$V_P(\widehat{\overline{CInf}}_t) = \frac{1}{H^2} \left\{ \sum_h^H \sum_{<g}^H (\pi_h \pi_g - \pi_{hg}) \left(\frac{CInf_t^h}{\pi_h} - \frac{CInf_t^g}{\pi_g} \right)^2 + \sum_{h=1}^H \frac{\beta_h}{\pi_h} CInf_t^{h^2} \right\} \quad (4)$$

where,

$$\beta_h = 1 + \frac{1}{\pi_h} \sum_{g \neq h}^H \pi_{hg} - \sum_{h=1}^H \pi_h \quad (5)$$

If each sample s contains common number of distinct units, then, $\beta_h = 0 \forall h$. An unbiased estimator of $V_P(\widehat{CInf}_t)$ is given by,

$$\widehat{V}_P(\widehat{CInf}_t) = \frac{1}{H^2} \left\{ \sum_{h < g} \sum_{g \in s} \left(\frac{\pi_h \pi_g - \pi_{hg}}{\pi_{hg}} \right) \left(\frac{CInf_t^h}{\pi_h} - \frac{CInf_t^g}{\pi_g} \right)^2 + \sum_{h \in s} \frac{\beta_h}{\pi_h^2} CInf_t^{h^2} \right\} \quad (6)$$

with $\beta_h = 0$, whenever valid, such that, $E_P \{ \widehat{V}_P(\widehat{CInf}_t) \} = V_P(\widehat{CInf}_t)$.

In reality, while conducting the survey for estimating inflation perceptions and expectations of the households, it may be useful to sample the households using varying probability sampling schemes so as to make the sample more representative. However, the motive here is not to estimate representative inflation expectations, but to prove that the estimated inflation figures using simulated population baskets (following the logic of heterogeneity in the consumption baskets), deviate from the official inflation figures, *i.e.*, they are close and well-correlated to the currently published households' inflation expectations. Hence, the estimates that would be derived using the sample of consumption baskets drawn from the simulated population baskets would not be sampling-scheme specific. In fact, the motive here is to show that the disagreement in households' inflation expectations with respect to the official inflation can be explained by estimating inflation from consumption baskets sampled by employing a general sampling scheme from a heterogenous population of consumption baskets simulated using a particular logic of select items.

III.1.ii Testing the Performance of Method 1

In order to demonstrate that huge variation in households' consumption baskets is one of the crucial factors responsible for the deviation of their inflationary expectations from the official inflation numbers, the above simulation method is followed for 299 consumer items in the CPI basket.² However, due to the lack of continuous time series inflation data of 13 consumer items *viz.*, "jackfruit", "singara", "mango", "kharbooza", "pears/ nashpati", "berries", "leechi", "chips (gm)", "sewing machine", "electric iron, heater, toaster, oven and other electric heating appliances", "VCD/ DVD hire (incl. instrument)", "cinema: new release (normal day)" and "library charges", 286 items are considered. The following steps³ are followed:

² State-wise analysis is not performed as CPI-inflation figures for items are unavailable for the states.

³ Here, the numbers of baskets to be simulated in the population and proportions of items to be drawn from each of the consumption subgroups of items at each step are indicative. Other values may be considered.

- i. Simulate $B_1 = 20,000$ consumption baskets by sampling $p_1 = 0.05$ proportions of items randomly without replacement from each of the 23 subgroups in the CPI basket.
- ii. Repeat i with $B_2 = 18,000$ and $p_2 = 0.10$.
- iii. Repeat i with $B_3 = 14,000$ and $p_3 = 0.15$.
- iv. Repeat i with $B_4 = 10,000$ and $p_4 = 0.20$.
- v. Repeat i with $B_5 = 10,000$ and $p_5 = 0.25$.
- vi. Repeat i with $B_6 = 2,000$ and $p_6 = 0.30$.
- vii. Thus, steps i to vi result in 74,000 simulated population baskets with fixed proportion of items from each of the subgroups.
- viii. Next, to simulate heterogeneous population consumption baskets, $D_1 = 2,000$ baskets are simulated, wherein the proportions of items to be drawn without replacement from each of the 23 subgroups for each of the baskets, are randomly drawn with replacement from $(p_1 = 0.05, p_2 = 0.10, p_3 = 0.15, p_4 = 0.20, p_5 = 0.25, p_6 = 0.30)$.
- ix. $D_2 = 20,000$ baskets are simulated, with the proportions of items to be selected without replacement from each of the subgroups for each of the baskets are randomly selected with replacement from $(p_{L1} = 0.02, p_{L2} = 0.07, p_{L3} = 0.10, p_{L4} = 0.17, p_{L5} = 0.20, p_{L6} = 0.25)$, which are relatively low proportions to represent the low income category of households who can afford to consume only a limited number of items.
- x. Lastly, $D_3 = 4,000$ baskets are simulated, where, the proportions of items to be selected without replacement from each of the subgroups for each of the baskets are randomly drawn with replacement from $(p_{U1} = 0.40, p_{U2} = 0.45, p_{U3} = 0.50, p_{U4} = 0.55, p_{U5} = 0.60, p_{U6} = 0.65)$ which are relatively high proportions to represent the economically sound class of households who can afford to consume large number of items.
- xi. Thus, 26,000 population baskets are simulated by following the steps from viii to x, resulting in $H = 1,00,000$ population baskets.
- xii. This is followed by the computation of the consolidated index for each of the H simulated baskets for each time point from January 2014 to March 2022, using equation (1).

xiii. For each time point from January 2015 to March 2022, inflation figures⁴ for each of the above simulated population baskets are calculated.

Next, a sample, say, s of baskets, of size $n(\geq 2) = 100^5$ is drawn from the population baskets using random sampling schemes *viz.*, Simple Random Sampling with Replacement (SRSWR), Simple Random Sampling without Replacement (SRSWOR) and Midzuno (1952). In the first two schemes, equal probabilities are assigned for selection to each of the population baskets. In Midzuno (1952) sampling technique, varying probabilities are assigned to the baskets using auxiliary or size-measure variable, for sampling the baskets from the population. For this sampling schemes, the number of family members in each household (corresponding to each population basket) are simulated and considered as the size-measure for selection of baskets. The sampling procedures and their estimation methodologies are performed following Chaudhuri and Pal (2022). Once the samples are selected, the inflation figures in each of the baskets are transformed into ranges⁶ of '< 1%', '1 – 2%', ..., '15 – 16%' and '≥ 16%' following the response options provided in the IESH questionnaire. Mean inflation estimates for the sampled baskets are calculated.

Next, $R_S = 1,000$ re-samples are drawn from the population using each of the four sampling schemes. Using re-samples, the average estimate for month t is:

$$AE_t = \frac{1}{R_S} \sum_{rs=1}^{R_S} \widehat{CInf}_{t_{rs}} \quad (7)$$

where, $\widehat{CInf}_{t_{rs}}$ is the unbiased estimator of \overline{CInf}_t , obtained from the rs^{th} re-sample. The Average Variance Estimate at t , (AVE_t) is,

$$AVE_t = \frac{1}{R_S} \sum_{rs=1}^{R_S} \hat{V}_p \left(\widehat{CInf}_{t_{rs}} \right) \quad (8)$$

⁴ Item-wise CPI indices are unavailable for the months March to May 2020 due to the Covid-19 pandemic and the resulting restrictions.

⁵ The bimonthly IESH conducted by the RBI forms one of the important inputs to the bimonthly Monetary Policy. Given the limited time frame, it covers about 6,000 households (survey centre Jammu, which was added recently, is not considered for the analysis) out of the total 1,77,28,937 households (as per Census 2011) in the 18 survey centres considered for the survey. Keeping this in consideration, a small sampling fraction is considered here for the numerical illustration.

⁶ For the samples drawn from the population baskets simulated using Method 3 (to be described later in the paper), as all figures are above 16 per cent, these are not transformed into ranges.

where, $\hat{V}_P(\widehat{CInf}_{t_{rs}})$ is the unbiased estimator of $V_P(\widehat{CInf}_{t_{rs}})$, obtained from the rs^{th} re-sample. Next, the Average Relative Bias at t , (ARB_t) is,

$$ARB_t = \left| \frac{AE_t - \widehat{CInf}_t}{\widehat{CInf}_t} \right| \quad (9)$$

Low values of AVE and ARB indicate the efficiency of the estimates. The Average Coefficient of Variation at t , (ACV_t) is given by,

$$ACV_t = \frac{1}{R_S} \sum_{rs=1}^{R_S} \left\{ \frac{\sqrt{\hat{V}_P(\widehat{CInf}_{t_{rs}})}}{\widehat{CInf}_{t_{rs}}} 100 \% \right\} \quad (10)$$

An estimate is considered outstanding, if $ACV < 10$ per cent, it is satisfactory, if ACV is in the range of (10 per cent, 20 per cent), poor but tolerable if (20 per cent $< ACV < 30$ per cent) and it is to be discarded, if $ACV > 30$ per cent (Chaudhuri, 2018). The estimated confidence interval for \widehat{CInf}_t , obtained from the rs^{th} re-sample is denoted by $[L_{t_{rs}}, U_{t_{rs}}]$, where,

$$L_{t_{rs}} = \widehat{CInf}_{t_{rs}} - \left\{ \tau_{\frac{\alpha}{2}} \sqrt{\hat{V}_P(\widehat{CInf}_{t_{rs}})} \right\}, U_{t_{rs}} = \widehat{CInf}_{t_{rs}} + \left\{ \tau_{\frac{\alpha}{2}} \sqrt{\hat{V}_P(\widehat{CInf}_{t_{rs}})} \right\} \quad (11)$$

with $\tau_{\frac{\alpha}{2}}$ being the upper $\frac{\alpha}{2}$ point of Standard Normal distribution. The Actual Coverage Percentage at t , (ACP_t) viz., the percentage of replicates out of R_S , for which the estimated confidence interval covers \widehat{CInf}_t is also computed. High values of ACP indicate efficiency of the estimates.

Next, it is intended to compare the households' inflation perceptions and expectations in terms of both gap and correlation with the simulated baskets' inflation of the survey month as well as of the month previous to the survey month (presuming households remember events of the immediate past). Table 2 shows the correlations of households' expectations for different lags (0, 1 and 2) and leads (0 and 1) of the estimated inflation in the simulated baskets. The gap between households' inflation expectations and estimated inflation is also tabulated. It is preferable to get high and significant correlation figures. In addition, there should not be enough evidence to support significant gap between the expectations and the estimated inflation in the simulated baskets.

Table 2: Performance of Method 1

Sampling Scheme	Reference to Survey Month	Measure	Estimated Inflation from Simulated Baskets	Current		Three Months Ahead		One Year Ahead				
				Mean	Median	Mean	Median	Mean	Median			
SRSWR	Survey Month	Correlation	Lag 0	0.73	0.75	0.75	0.74	0.55	0.63			
			Lag 1	0.70	0.72	0.69	0.69	0.58	0.65			
			Lag 2	0.68	0.69	0.69	0.71	0.63	0.70			
			Lead 1	0.62	0.63	0.65	0.63	0.39	0.51			
			Lead 2	0.55	0.52	0.61	0.58	0.33	0.43			
	Previous Month	Gap	Lag 0	4.23	3.45	4.90	4.29	5.04	4.93			
			Correlation	Lag 0	0.71	0.75	0.73	0.71	0.58	0.62		
				Gap	Lag 0	4.27	3.49	4.94	4.33	5.09	4.97	
			SRSWOR	Survey Month	Correlation	Lag 0	0.73	0.75	0.75	0.74	0.56	0.63
						Lag 1	0.70	0.72	0.69	0.70	0.58	0.65
Lag 2	0.68	0.69				0.69	0.71	0.63	0.70			
Lead 1	0.63	0.64				0.65	0.63	0.39	0.51			
Lead 2	0.55	0.52				0.61	0.58	0.33	0.43			
Previous Month	Gap	Lag 0	4.23	3.45	4.90	4.29	5.04	4.93				
		Correlation	Lag 0	0.71	0.75	0.73	0.72	0.58	0.62			
			Gap	Lag 0	4.27	3.49	4.94	4.33	5.09	4.97		
		Midzuno (1952)	Survey Month	Correlation	Lag 0	0.73	0.75	0.75	0.74	0.55	0.63	
					Lag 1	0.70	0.72	0.69	0.69	0.58	0.65	
Lag 2	0.68				0.69	0.69	0.71	0.63	0.70			
Lead 1	0.62				0.63	0.65	0.63	0.39	0.51			
Lead 2	0.55				0.52	0.61	0.58	0.33	0.43			
Previous Month	Gap	Lag 0	4.23	3.45	4.90	4.29	5.04	4.93				
		Correlation	Lag 0	0.71	0.75	0.73	0.71	0.58	0.62			
			Gap	Lag 0	4.27	3.49	4.94	4.33	5.09	4.97		

Note: Figures in bold indicate significance at 5 per cent level.

Sources: MOSPI; RBI; and Author's calculation.

The initial presumption was that if the survey expectations are close to and well-correlated with the inflation estimates obtained from the simulated baskets, then it would provide an empirical evidence to the general understanding that heterogeneity in the composition of consumption baskets is a crucial reason for the deviation of households' inflation expectations from the official CPI-U inflation figures. From the findings of Method 1, it is understood that although the correlation figures are high and significant, the gaps with respect to the estimated inflation figures obtained from the simulated baskets, are significantly different from zero. Given this limitation of Method 1, we go on to design alternative logical methods as illustrated below.

III.2 Method 2

This method of basket simulation is based on the logic that the households while making expectations consider inflation of those items whose inflation figures are relatively higher than the other items in a particular month. Here, they do not consider their own

consumption baskets at all; rather they look independently at the items that rank top with respect to the respective inflation figures. Thus, the items considered by the households are dynamic, *i.e.*, they change over the months according to their inflation. Inflation in the CPI groups may display abrupt changes whose impact on households' minds may be asymmetric and irrespective of the consumption weight.

At first, all the R items are ordered in decreasing order of their inflation in each month. Then, a random fraction is selected from a set of predefined proportions ($0 < p_1, p_2, \dots, p_N < 1$), say, p_K . Next, the first $n_K = Rp_K$ number of items are filtered in the basket for a household for each month (n_K remaining same in each month) and the rest of the items are not considered for further analysis. Following this procedure, H baskets are prepared such that n_K varies across the baskets but remains same for a basket across the months.

Now, to find the consolidated index for each of the H baskets, let y_{thk} be the price index for the k^{th} , $k \in (n_1, n_2, \dots, n_N)$ item for the h^{th} basket in the t^{th} month. Then, the combined index pertaining to the h^{th} simulated basket for month t is calculated as,

$$CI_t^{2h} = \frac{\sum_{k \in (n_1, n_2, \dots, n_N)} w_{tk} y_{thk}}{\sum_{k \in (n_1, n_2, \dots, n_N)} w_{tk}}, h = 1, 2, \dots, H, t = 1, 2, \dots, T \quad (12)$$

where, w_{tk} is the weight considered as the relative inflation of the k^{th} item with respect to inflation in the rest of the items in a particular month, *i.e.*,

$$w_{tk} = \frac{\frac{y_{tk} - y_{(t-12)k}}{y_{(t-12)k}}}{\sum_{r=1}^R \frac{y_{tr} - y_{(t-12)r}}{y_{(t-12)r}}}, t = 1, 2, \dots, T \quad (13)$$

From these aggregate indices, year-on-year inflation figures are calculated for each basket in the population. Let $CI_n F_t^{2h}$ denote the inflation in basket h , $h = 1, 2, \dots, H$. It is clear that in this method, the number of items selected in any particular basket is constant across months. However, the composition of any basket varies across the months based on the items appearing at the top in the list of their respective inflation numbers.

A major disadvantage of Method 2 is that the consolidated indices corresponding to each of the baskets are not comparable across months or year, as the items may change. As a result, the year-on-year inflation figures do not actually represent the percentage change in a fixed basket of items.

III.3 Method 3

To overcome the above-noted limitation of Method 2, instead of computing the weighted index of the customised baskets, the inflation numbers of all the items in a basket are aggregated during each month under Method 3. This is done by calculating the following:

$$CInF_t^{3h} = \frac{\sum_{k \in (n_1, n_2, \dots, n_N)} w_{tk} InF_{thk}}{\sum_{k \in (n_1, n_2, \dots, n_N)} w_{tk}}, h = 1, 2, \dots, H, t = 1, 2, \dots, T \quad (14)$$

where, InF_{thk} is inflation of k^{th} item in h^{th} basket in t^{th} month.

III.4 Method 4

This method goes a step ahead of Method 3. Here, it is presumed that households while making their inflation sentiments, not only consider those items whose inflation is among the highest in a month but also those items which they consume more frequently or in higher quantity than other items – identified by consumption weights. It is based on the understanding that if inflation in the items which are not in frequent use by most of the households rises, households will not be bothered much. Households may be bothered more about the rise in inflation of the necessary items of basic consumption. Given the skewed income distribution in India resulting in a skewed consumption pattern, this presumption seems logical.

At first, all the items are ordered in decreasing order of their consumption weights and inflation figures in each month. Then, a random fraction is selected from a set of predefined proportions ($0 < p_1, p_2, \dots, p_N < 1$), say, p_K . From the first $n'_K < R$ items in the ordered list, $n_K (< n'_K) = Rp_K$ number of items are filtered in the basket for a household for each month and the rest of the items are not considered for further analysis. Here, p_K is sampled afresh for each month and hence, n_K varies accordingly over time. Following this procedure, H baskets are prepared such that n_K varies across the baskets as well as across the months.

Now, to find the aggregate inflation for each of the H baskets, let Inf_{thk} be the inflation for the k^{th} , $k \in (n_1, n_2, \dots, n_N)$ item for the h^{th} basket in the t^{th} month. Then, the combined inflation pertaining to the h^{th} simulated basket for month t is calculated as,

$$CInf_t^{4h} = \frac{\sum_{k \in (n_1, n_2, \dots, n_N)} w_{tk} Inf_{thk}}{\sum_{k \in (n_1, n_2, \dots, n_N)} w_{tk}}, h = 1, 2, \dots, H, t = 1, 2, \dots, T \quad (15)$$

where, w_{tk} is consumption weight of the k^{th} item. It is to be noted that in this method, the number of items selected in any particular basket is different over months. In addition, the composition of any basket also varies across the months based on the items appearing at the top in the list of their consumption weights as well as respective inflation numbers. Hence, this method assumes a dynamic framework.

III.5 Method 5

In this method, the consumption baskets are simulated following a procedure almost similar to the previous method. The main difference here is that, items considered for the baskets are filtered based on a threshold on the CPI inflation and a threshold on the consumption weight in each month. The thresholds are introduced here with an understanding that households remember items whose inflation outpaced more than the tolerance level and are of frequent use. These thresholds are not derived theoretically by optimisation method, but different permutations of various values are considered for performing this simulation.

At first, all the R items whose CPI inflation figures are above a predefined number, say, $CPIUI$, are filtered in for each month, separately. Next, among these items, all the items whose CPI consumption weights are above a predefined proportion, say, CW (kept constant in each month), are filtered in for each month, separately. Suppose, the above two filtering criteria result in a list of R' items. Then, a random fraction is selected from a set of predefined proportions ($0 < p_1, p_2, \dots, p_N < 1$), say, p_K . From the list of items filtered, $n_K = R'p_K$ items are filtered in the basket for a household for each month and the rest of the items are not considered for further analysis. Here, p_K is sampled afresh for each month and hence, n_K varies accordingly over the months. Following this procedure, H baskets are prepared such that n_K varies across the baskets as well as across the months.

Now, to find the aggregate inflation for each of the H baskets, let Inf_{thk} be the inflation for the k^{th} , $k \in (n_1, n_2, \dots, n_N)$ item for the h^{th} basket in the t^{th} month. Then, the combined inflation pertaining to the h^{th} simulated basket for month t is calculated similar to Method 4 and denoted by $CInf_t^{5h}$. It is to be noted that similar to that in the previous method, the proportion of items selected using this method in any particular basket is different over time. In addition, the composition of any basket also varies across the months based on the items appearing at the top in the list of their consumption weights as well as respective inflation numbers, with both being above certain predefined thresholds. Hence, this method also relies on a dynamic framework.

IV. Numerical Illustrations to Test the Performance of the Alternative Methods

IV.1 Methods 2 and 3

The following are the steps involved in these two methods:

- i. At first, all the CPI consumer items are ordered in decreasing order of their inflation in each month from January 2015 to March 2022.
- ii. Then, a random fraction, say p_K , is selected randomly from $(p_1 = 0.01, p_2 = 0.02, \dots, p_{50} = 0.50)$ ⁷.
- iii. Next, the first $n_K = Rp_K = 286p_K$ items are filtered in the basket for a household for each month (n_K remaining same in each month) and the rest of the items are not considered in the basket.
- iv. Following this procedure, $H = 1,00,000$ population baskets are simulated such that n_K varies across the baskets but remains same for a basket across the months.

For Method 2: This is followed by the computation of the consolidated index for each of the H simulated population baskets for each time point from January 2015 to March 2022, using equations (12) and (13).

For Method 3: The consolidated inflation for each of the H simulated population baskets for each time point from January 2015 to March 2022, using equation (14) are calculated.

- v. **For Method 2:** For each time point from January 2016 to March 2022, inflation figures of the above simulated population baskets are calculated. Item-wise CPI indices are unavailable for the months March to May 2020 due to the Covid-19 pandemic and the resulting restrictions.

IV.2 Method 4

The following are the steps involved in this method:

- i. At first, all the CPI consumer items are ordered in decreasing order of their consumption weights and inflation figures in each month from January 2015 to March 2022.

⁷ Low fractions are considered here with the presumption that households can remember a limited number of items while making expectations.

- ii. Then, a random fraction is selected randomly from $(p_1 = 0.01, p_2 = 0.02, \dots, p_5 = 0.05)$, say, p_K .
- iii. From the first $n'_K = 50$ items in the ordered list, $n_K = 286p_K$ number of items are filtered in the basket for a household for each month and the rest of the items are not considered in the population basket. Here, p_K is sampled afresh for each month and hence, n_K varies accordingly over time as well as across the baskets.
- iv. Following this procedure, $H = 1,00,000$ baskets are simulated.
- v. Next, the inflation figures for the sampled items in each of the H baskets are aggregated for each month using equation (15).

IV.3 Method 5

The following steps are followed:

- i. At first, all the items whose CPI inflation figures are above $CPIUI = 4$ (constant in each month), are filtered in for each month separately.
- ii. Next, among these items, all the CPI items whose consumption weights for urban areas are above $CW = 0.4$ (constant in each month), are filtered in for each month, separately.
- iii. Suppose, the above two filtering criteria result in a list of R' items. Then, a random fraction is selected from $(p_1 = 0.01, p_2 = 0.02, \dots, p_5 = 0.05)^8$, say, p_K .
- iv. From the list of items filtered above, $n_K = R'p_K$ items are filtered in the basket for a household for each month and the rest of the items are not considered in the basket.
- v. The fraction p_K is sampled afresh for each month.
- vi. Following this procedure, $H = 1,00,000$ baskets are simulated such that n_K varies across the baskets as well as across the months.
- vii. Next, the inflation figures for the sampled items in each of the H baskets are aggregated for each month using equation (15).
- viii. The above procedure is repeated to simulate more population baskets of size $H = 1,00,000$ using the combinations as given in Table 3.

⁸ Small percentages are considered here with a presumption that households may remember only a few items in mind while making their inflationary sentiments.

Table 3: Combinations of Consumption Weights and Inflation

Number	CW	CPIUI(in per cent)
1	0.50	4
2	0.60	4
3	0.40	5
4	0.50	5
5	0.60	5
6	0.40	6
7	0.50	6
8	0.60	6
9	0.40	7
10	0.50	7
11	0.60	7
12	'Combined Population', wherein 8,334 baskets are selected randomly from the first 4 populations and 8,333 baskets are selected randomly from the remaining populations, resulting in 1,00,000 baskets in this population. This is done to inject heterogeneity in the population baskets.	
13	<p>'Logic Population', wherein 'Method 5' is followed considering dynamic $CPIUI$ (varies over the months), <i>i.e.</i>, $CPIUI_t$, where,</p> $CPIUI_t = \frac{1}{K} \sum_{k=1}^K InF_{tk} \quad (16)$ <p>and dynamic CW, <i>i.e.</i>, CW_t, where,</p> $CW_t = \frac{1}{K_t} \sum_{k=1}^{K_t} w_k \quad (17)$ <p>where, InF_{tk} is the inflation of the k^{th} item in the t^{th} month, $t = 1, 2, \dots, T$ and w_k is the consumption weight of the k^{th} item with inflation greater than $CPIUI_t$.</p>	

ix. Thus, we simulate 14 populations, each of size $H = 1,00,000$.

Tables 4 to 6 display correlation and gap figures for the estimates of inflation derived from baskets sampled from the population baskets simulated using Method 5, 'Combined Population and 'Logic Population' as described in Table 3 above.⁹

⁹ The Tables on performances of Methods 2 to 4 and of eleven combinations of $CPIUI$ and CW for Method 5 are refrained from reporting here to conserve space, but available on request.

Table 4: Performance of Method 5 ($CPIUI = 4$ per cent and $CW = 0.60$)

Sampling Scheme	Reference Month	Measure	Estimated Inflation from Simulated Baskets	Current		Three Months Ahead		One Year Ahead			
				Mean	Median	Mean	Median	Mean	Median		
SRSWR	Survey Month	Correlation	Lag 0	0.58	0.57	0.67	0.63	0.26	0.45		
			Lag 1	0.63	0.62	0.71	0.68	0.37	0.56		
			Lag 2	0.68	0.70	0.77	0.74	0.45	0.61		
			Lead 1	0.45	0.46	0.54	0.50	0.11	0.31		
			Lead 2	0.32	0.31	0.44	0.38	-0.05	0.15		
	Previous Month	Gap	Lag 0	0.96	0.19	1.64	1.02	1.78	1.67		
			Correlation	Lag 0	0.59	0.58	0.68	0.65	0.31	0.49	
				Lag 0	0.97	0.19	1.64	1.03	1.78	1.67	
			Survey Month	Correlation	Lag 0	0.58	0.57	0.67	0.63	0.26	0.46
					Lag 1	0.63	0.62	0.71	0.68	0.37	0.56
SRSWOR	Survey Month	Correlation	Lag 2	0.68	0.70	0.77	0.74	0.45	0.61		
			Lead 1	0.45	0.46	0.54	0.50	0.11	0.32		
			Lead 2	0.32	0.31	0.44	0.37	-0.05	0.15		
			Gap	Lag 0	0.97	0.19	1.64	1.03	1.78	1.67	
				Previous Month	Correlation	Lag 0	0.59	0.58	0.68	0.65	0.31
	Lag 0	0.97	0.19			1.64	1.03	1.79	1.67		
	Midzuno (1952)	Survey Month	Correlation	Lag 0	0.58	0.57	0.67	0.63	0.26	0.45	
				Lag 1	0.63	0.62	0.71	0.68	0.37	0.56	
				Lag 2	0.68	0.70	0.77	0.74	0.45	0.61	
				Lead 1	0.45	0.46	0.54	0.50	0.11	0.32	
Lead 2				0.32	0.31	0.44	0.37	-0.05	0.15		
Previous Month		Gap	Lag 0	0.96	0.19	1.63	1.02	1.78	1.67		
			Correlation	Lag 0	0.58	0.58	0.68	0.65	0.31	0.48	
				Lag 0	0.96	0.19	1.64	1.02	1.78	1.67	
			Survey Month	Correlation	Lag 0	0.58	0.57	0.67	0.63	0.26	0.45
					Lag 1	0.63	0.62	0.71	0.68	0.37	0.56

Note: Figures in bold indicate significance at 5 per cent level.

Sources: MOSPI; RBI; and Author's calculation.

Table 5: Performance of Method 5 Using 'Combined Population'

Sampling Scheme	Reference Month	Estimated Inflation from Simulated Baskets	Simulated IE	Current		Three Months Ahead		One Year Ahead			
				Mean	Median	Mean	Median	Mean	Median		
SRSWR	Survey Month	Correlation	Lag 0	0.35	0.33	0.44	0.39	0.15	0.29		
			Lag 1	0.42	0.42	0.49	0.47	0.27	0.44		
			Lag 2	0.49	0.54	0.59	0.57	0.36	0.50		
			Lead 1	0.26	0.26	0.35	0.33	0.01	0.18		
			Lead 2	0.18	0.17	0.28	0.23	-0.08	0.04		
	Previous Month	Gap	Lag 0	-0.65	-1.43	0.02	-0.59	0.16	0.05		
			Correlation	Lag 0	0.36	0.34	0.46	0.42	0.21	0.35	
				Lag 0	-0.73	-1.50	-0.06	-0.67	0.09	-0.02	
			Survey Month	Correlation	Lag 0	0.35	0.33	0.44	0.39	0.15	0.29
					Lag 1	0.42	0.42	0.49	0.47	0.27	0.44
SRSWOR	Survey Month	Correlation	Lag 2	0.49	0.54	0.59	0.57	0.36	0.50		
			Lead 1	0.26	0.26	0.35	0.33	0.01	0.18		
			Lead 2	0.18	0.17	0.28	0.22	-0.08	0.04		
			Gap	Lag 0	-0.66	-1.44	0.01	-0.60	0.16	0.04	
				Correlation	Lag 0	0.35	0.34	0.46	0.42	0.20	0.35

	Previous Month	Gap	Lag 0	-0.73	-1.51	-0.06	-0.67	0.09	-0.03	
Midzuno (1952)	Survey Month	Correlation	Lag 0	0.35	0.33	0.44	0.39	0.15	0.29	
			Lag 1	0.41	0.42	0.49	0.47	0.26	0.43	
			Lag 2	0.49	0.54	0.59	0.57	0.35	0.50	
			Lead 1	0.26	0.27	0.35	0.33	0.01	0.18	
			Lead 2	0.18	0.17	0.28	0.22	-0.08	0.03	
	Previous Month	Gap	Lag 0	-0.66	-1.44	0.01	-0.60	0.16	0.04	
			Correlation	Lag 0	0.35	0.33	0.46	0.41	0.20	0.34
			Gap	Lag 0	-0.73	-1.51	-0.06	-0.67	0.08	-0.03

Note: Figures in bold indicate significance at 5 per cent level.

Sources: MOSPI; RBI; and Author's calculation.

Table 6: Performance of Method 5 Using 'Logic Population'

Sampling Scheme	Reference Month	Estimated Inflation from Simulated Baskets	Simulated IE	Current		Three Months Ahead		One Year Ahead		
				Mean	Median	Mean	Median	Mean	Median	
SRSWR	Survey Month	Correlation	Lag 0	0.66	0.66	0.69	0.67	0.51	0.56	
			Lag 1	0.64	0.66	0.65	0.65	0.57	0.64	
			Lag 2	0.63	0.67	0.67	0.68	0.61	0.67	
			Lead 1	0.59	0.59	0.61	0.60	0.35	0.46	
			Lead 2	0.50	0.49	0.53	0.50	0.22	0.28	
	Previous Month	Gap	Lag 0	0.06	-0.71	0.74	0.13	0.88	0.77	
			Correlation	Lag 0	0.65	0.67	0.69	0.67	0.55	0.59
			Gap	Lag 0	0.14	-0.63	0.81	0.20	0.96	0.84
SRSWOR	Survey Month	Correlation	Lag 0	0.66	0.66	0.69	0.67	0.51	0.56	
			Lag 1	0.64	0.66	0.65	0.65	0.57	0.64	
			Lag 2	0.63	0.67	0.67	0.68	0.61	0.67	
			Lead 1	0.59	0.59	0.61	0.60	0.35	0.46	
			Lead 2	0.50	0.49	0.53	0.50	0.22	0.28	
	Previous Month	Gap	Lag 0	0.06	-0.71	0.73	0.12	0.88	0.76	
			Correlation	Lag 0	0.65	0.67	0.69	0.67	0.55	0.58
			Gap	Lag 0	0.14	-0.64	0.81	0.20	0.96	0.84
Midzuno (1952)	Survey Month	Correlation	Lag 0	0.66	0.66	0.69	0.67	0.51	0.56	
			Lag 1	0.64	0.66	0.64	0.64	0.57	0.63	
			Lag 2	0.63	0.67	0.67	0.68	0.61	0.67	
			Lead 1	0.58	0.58	0.61	0.60	0.35	0.45	
			Lead 2	0.50	0.48	0.53	0.49	0.22	0.28	
	Previous Month	Gap	Lag 0	0.06	-0.71	0.73	0.12	0.88	0.77	
			Correlation	Lag 0	0.65	0.67	0.69	0.67	0.55	0.58
			Gap	Lag 0	0.14	-0.63	0.81	0.20	0.96	0.85

Note: Figures in bold indicate significance at 5 per cent level.

Sources: MOSPI; RBI; and Author's calculation.

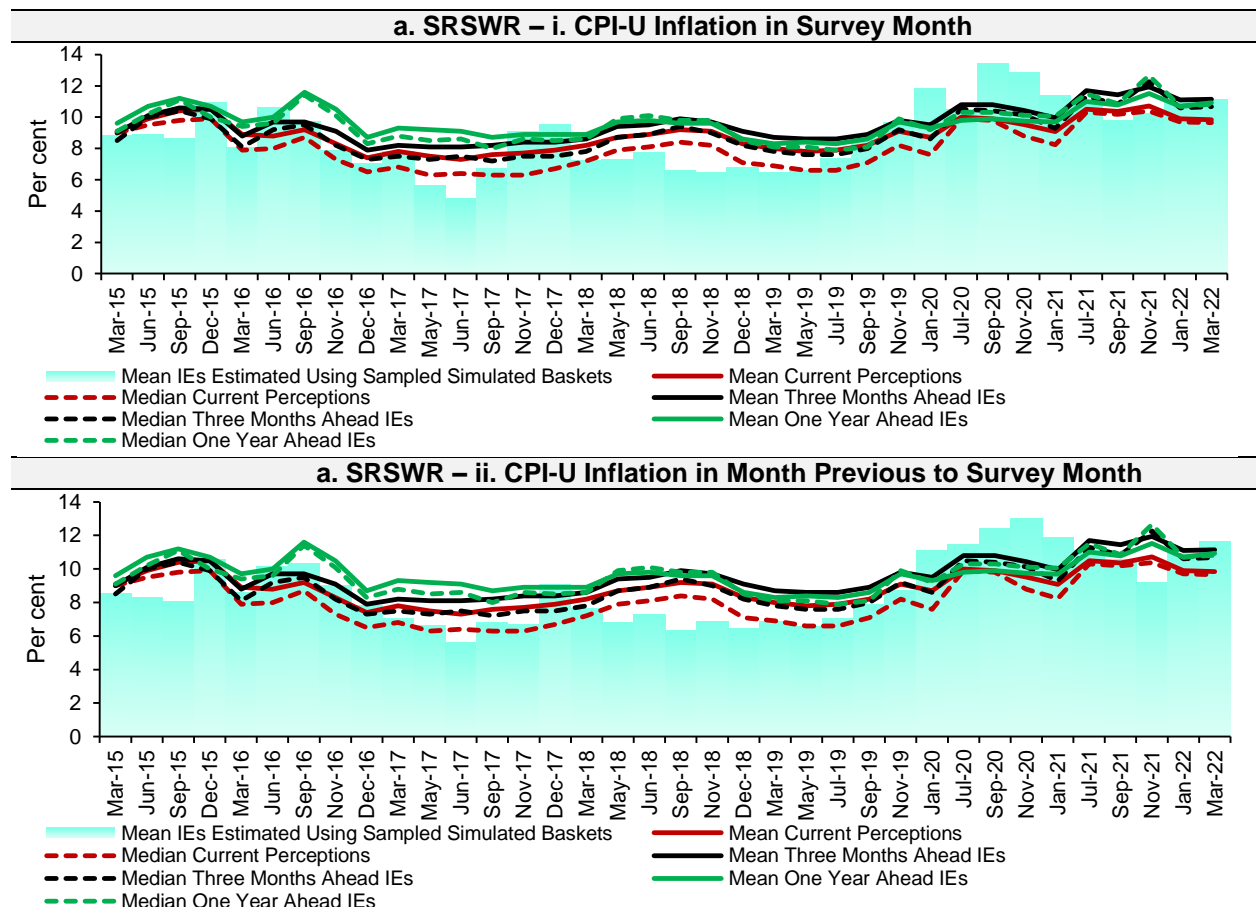
Methods 2 and 3 display low correlation values and significant gap figures. Under Method 4, although correlations are significant, gaps are about 5 percentage points. In the combinations of $CPIUI = 5$ per cent and $CW = 0.40$, $CPIUI = 5$ per cent and $CW = 0.50$ and $CPIUI = 5$ per cent and $CW = 0.60$, correlation figures are significant but gaps are significantly different from zero. For most of the combinations of $CPIUI$ and CW in

Method 5, the correlation and gap figures are not satisfactory. However, $CPIUI = 4$ per cent and $CW = 0.60$, the correlations are significant and the gaps are also significantly higher than zero.

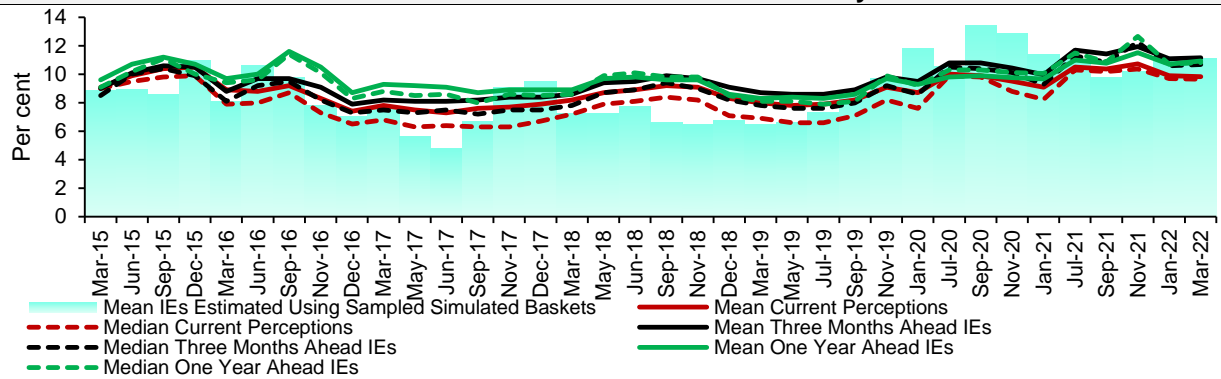
The performance of the estimates derived from the 'Combined Population' is satisfactory. However, in the estimates obtained from samples drawn from the 'Logic Population'; not only are the correlations high and significant, there is not much evidence to reject the null hypothesis of gaps equal to zero.

Thus, the results from the 'Logic Population' are the best in the lot of combinations studied here. These are also evident from Chart 3 below. Further, when these results are superimposed on the efficiency of the estimates measured by Average Variance Estimate (AVE), Average Relative Bias (ARB), Average Coefficient of Variation (ACV) and Actual Coverage Percentage (ACP) (Appendix), results reflect that irrespective of the sampling scheme, the 'Logic Population' produces inflation estimates which are found to be close, and significantly correlated with the survey-based inflation expectations.

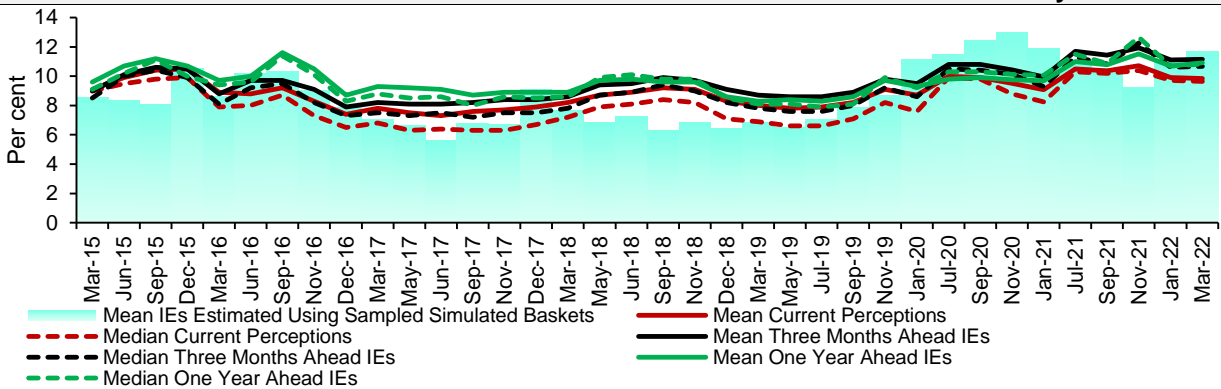
Chart 3. Inflation Perceptions and Expectations vis-à-vis Estimated Expectations Computed Using Samples Drawn From the 'Logic Population'



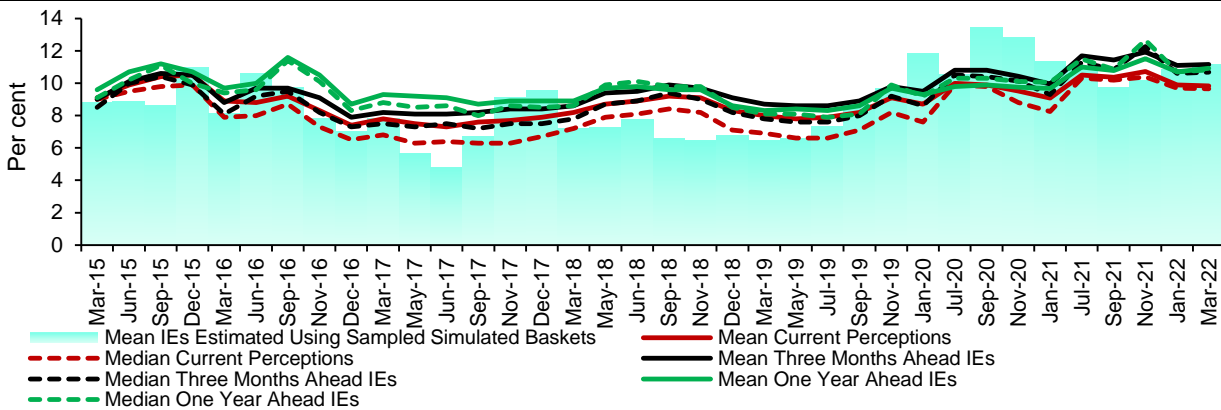
b. SRSWOR – i. CPI-U Inflation in Survey Month



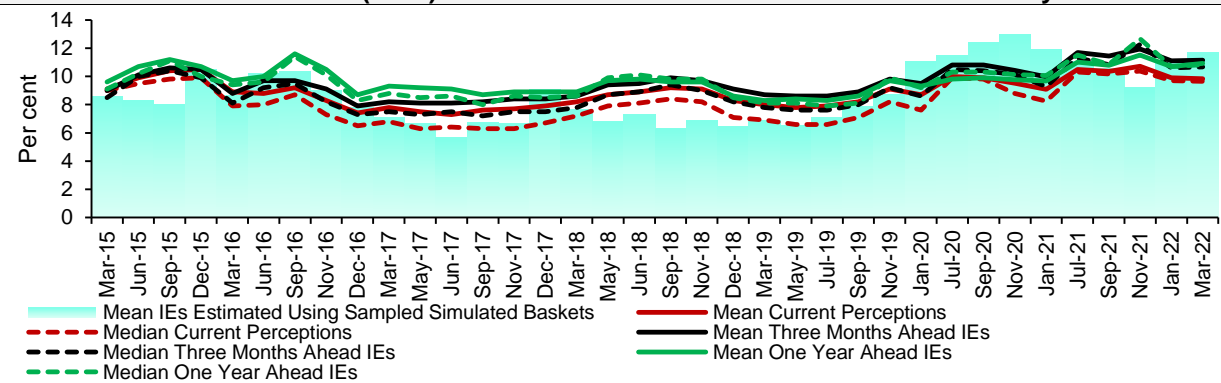
b. SRSWOR – ii. CPI-U Inflation in Month Previous to Survey Month



c. Midzuno (1952) – i. CPI-U Inflation in Survey Month



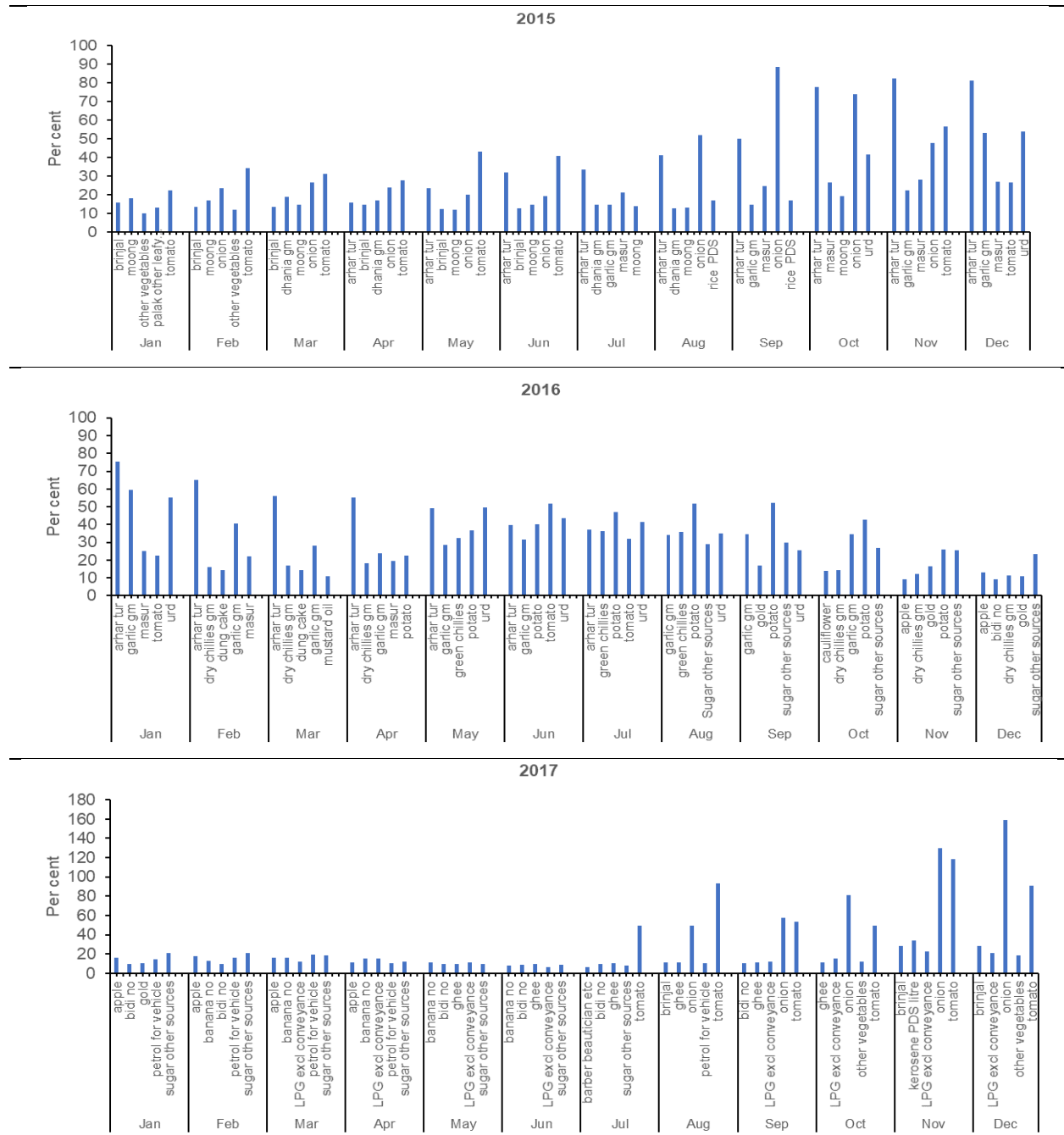
c. Midzuno (1952) – ii. CPI-U Inflation in Month Previous to Survey Month



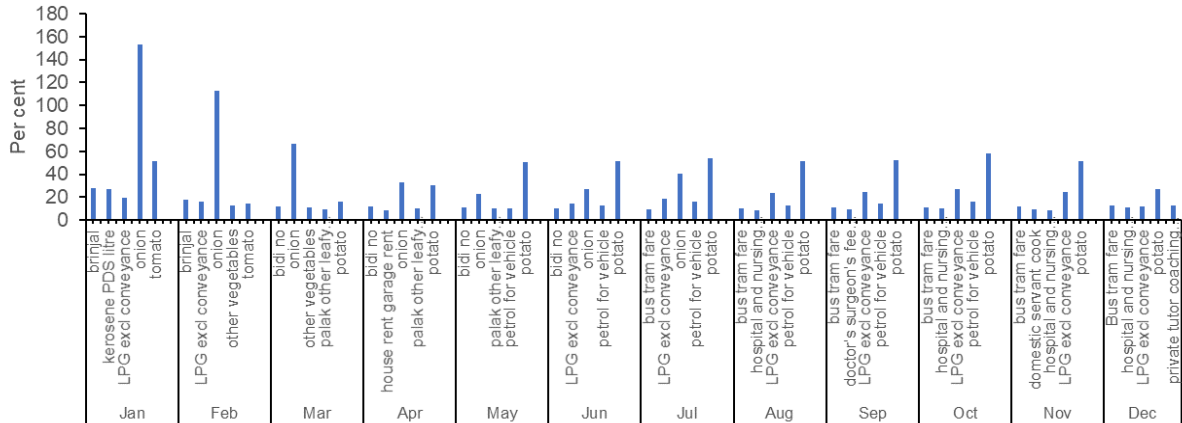
Sources: MOSPI; RBI; and Author's calculation.

Chart 4 provides an overview of select five items that rank top in inflation in the 'Logic Population' (list of items with consumption above the mean thresholds among items with inflation above the monthly thresholds). The major items of concern were pulses and vegetables from 2015 to 2016, vegetables, fuel and conveyance charges from 2017 to 2019 and vegetables (especially potato), refined oil and fuel during the COVID-19 pandemic.

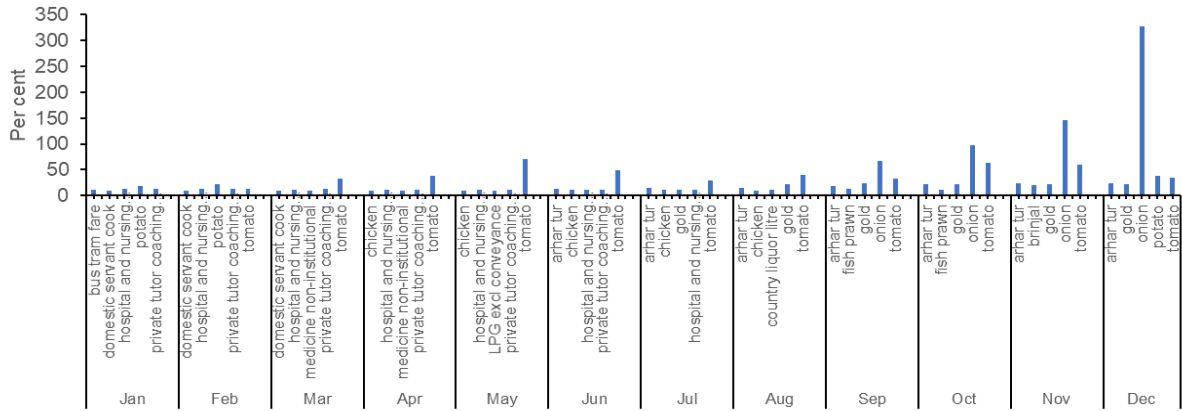
Chart 4. Five Items Ranking Top in Inflation from 'Logic Population' Baskets



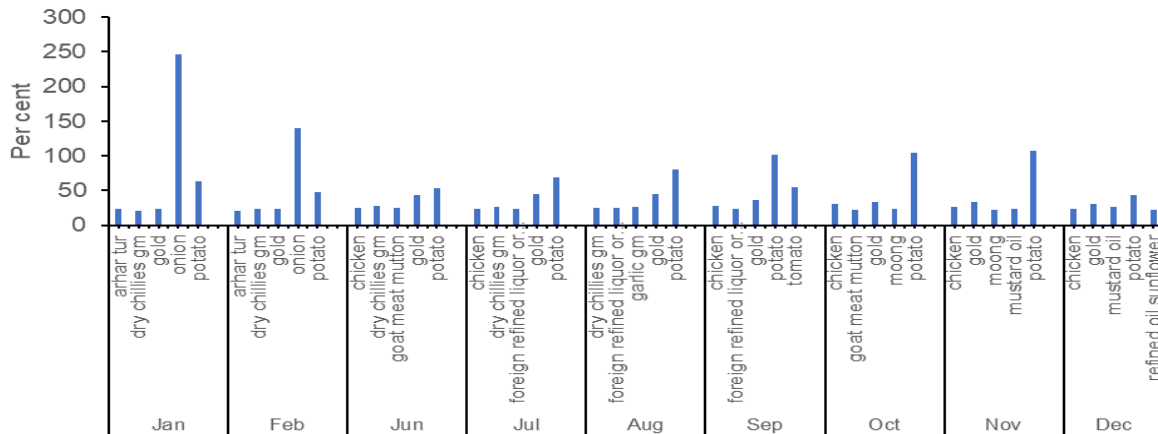
2018

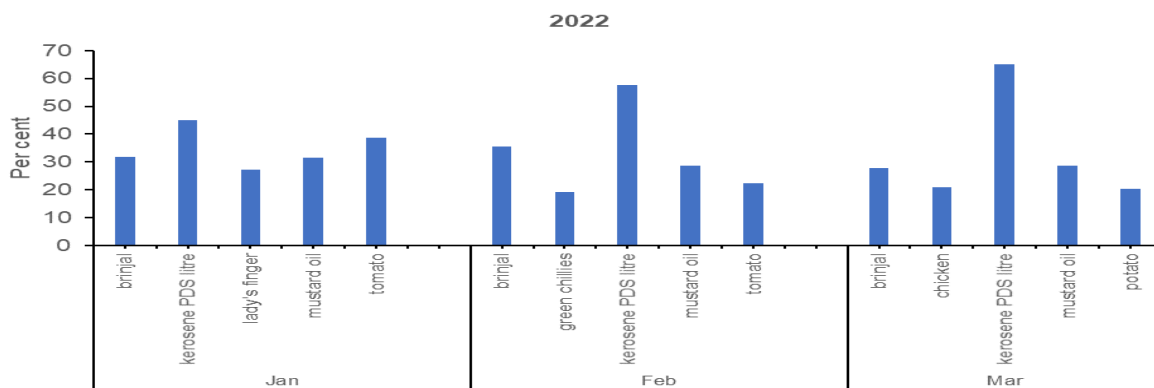
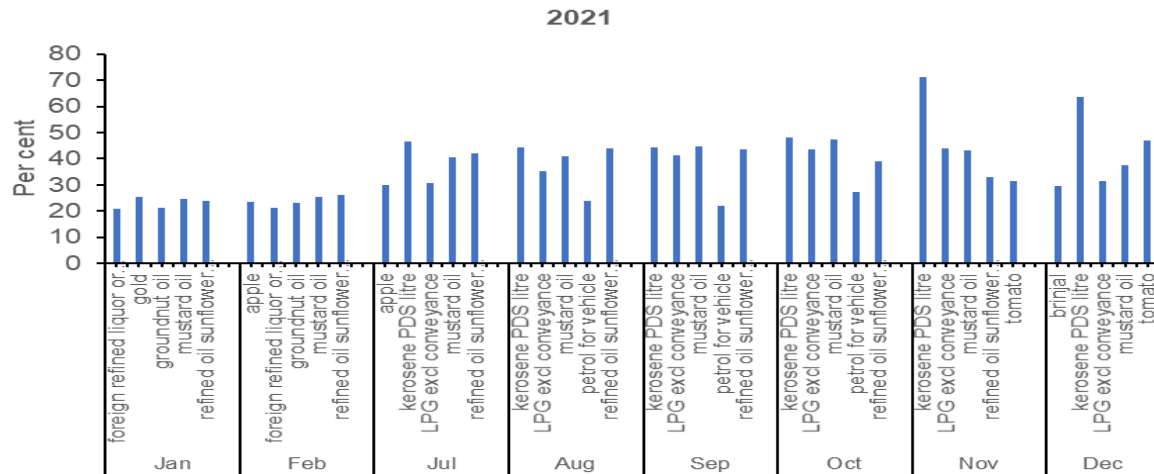


2019



2020





Sources: MOSPI and Author's calculation.

Following are the key findings from the analysis in this section:

1. Households' inflation expectations are largely dependent on the items that carry a high rank in inflation (more than the average inflation of items in a month) and among them, those which are consumed the most (more than the average consumption weight of items). The list of such items is dynamic and varies across months. This results in disagreement in households' inflation expectations with respect to the official inflation figures. The results stand valid irrespective of the sampling scheme used for sampling the baskets from the simulated population baskets.
2. The inflation figures estimated using sampled baskets drawn from population consumption baskets simulated using a logic of considering items with high inflation and of regular consumption, are not only close to the survey-based expectations but also significantly correlated. Thus, a sudden rise in inflation in items of regular use can explain the deviation in households' inflation expectations from the official inflation. These findings indicate that using a

logical method for the selection of items in consumption baskets, the disagreement in survey-based expectations with respect to the official inflation can be explained.

V. Conclusions

It is generally argued in the literature that the deviation of households' inflation expectations from the official inflation number can be the result of the diversity in households' consumption baskets. Using data from the IESH conducted by the RBI, this paper attempts to provide empirical evidence to support this observation in the case of India. The paper takes a novel approach to verify this observation empirically. It simulates a large number of heterogeneous consumption baskets. Treating this as the population, it estimates the mean inflation in the population baskets by using sampled baskets from the simulated population. It is anticipated that the mean inflation in numerous simulated heterogeneous consumer baskets should be well-correlated and close to the survey-based expectations. Such a correlation would provide empirical support to the observation that the heterogeneity in the composition of consumption baskets is a crucial reason for the deviation of households' inflation expectations from the official CPI-U inflation figures. This effort of finding the source(s) of disagreement in households' inflation expectations with respect to the official inflation can assist in better understanding and using the inflationary expectations.

The paper, however, finds that although the estimated inflation figures obtained from the simulated baskets are significantly correlated with the survey-based expectations, the survey-based numbers bear significant gap with respect to the estimated inflation figures obtained from the simulated baskets.

As a result, we develop other alternative logical methods for simulating the baskets of consumption items which would provide a reason for the deviation of expectations from the CPI-U inflation. The motive is to find the best logical method for simulating the consumption baskets that would ensure that (a) the estimated inflation would deviate from the headline inflation figures; and (b) the estimated inflation would be close in number and display a high correlation with the survey-based expectations.

While most of the proposed methods do not result in satisfactory results, we obtain the desired results from the samples drawn from 'Logic Population' method (under Method 5 in the paper) wherein the items are considered if their respective inflation figures

are more than the average inflation of items in a month and among them, those whose consumption weight is more than the average consumption weight.

The expectation estimates obtained from the 'Logic Population' are not only significantly correlated with the survey-based expectations but the gap between the two series is not found to be statistically significant. Irrespective of the sampling scheme used for sampling the baskets from the simulated population baskets, the 'Logic Population' method produces inflation estimates that are found to be close to and significantly correlated with the survey-based inflation expectation estimates.

We, therefore, conclude that households' inflation expectations are largely dependent on the items carrying a high ranking in inflation (more than the average inflation of items in a month) and among these, the ones which carry a high weight in consumption (more than the average consumption weight of items). Evidently, the list of such items is dynamic and can vary across months.

These findings indicate that using a logical method for the selection of items in consumption baskets, the disagreement in survey-based expectations with respect to the official inflation can be explained.

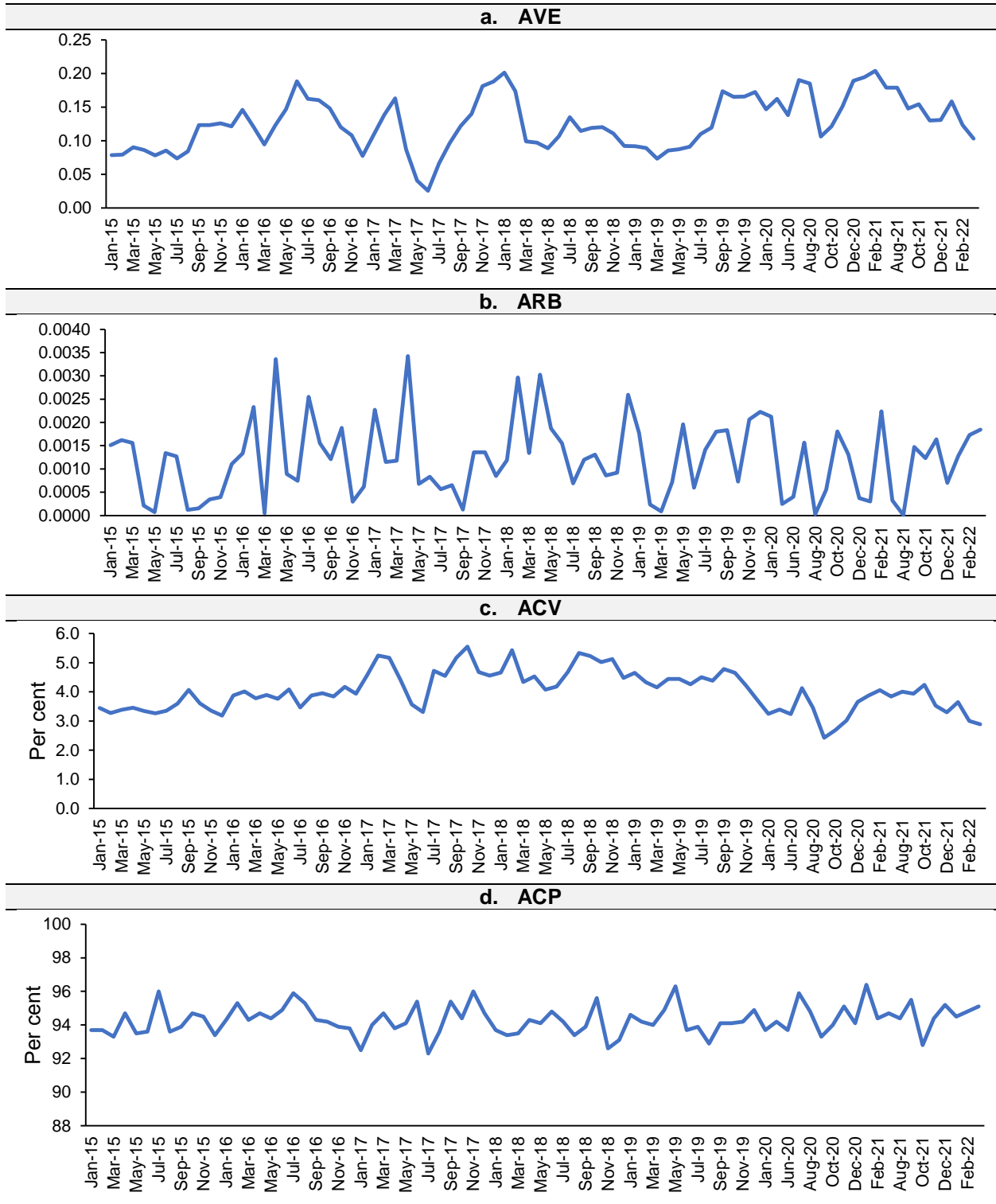
References

- Alpanda, S., Honig, A., & Woglom, G. (2013). Extending the textbook dynamic AD-AS framework with flexible inflation expectations, optimal policy response to demand changes, and the zero-bound on the nominal interest rate. *Modern Economy*, 4, 145-160.
- Carroll, C. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics*, 118(1), 269-298.
- Chaudhuri, A. (2018). *Survey sampling*. Chapman and Hall: CRC Press, Taylor and Francis Group.
- Chaudhuri, A., & Pal, S. (2002). On certain alternative mean square estimators in complex survey sampling. *Journal of Statistical Planning and Inference*, 104(2), 363-375.
- Chaudhuri, A., & Pal, S. (2022). *A comprehensive textbook on sample surveys*. Springer.
- Ciro, J. C. G., & Zapata, J. C. A. (2019). Disagreement in inflation expectations: empirical evidence for Colombia. *Applied Economics*, 51(40), 4411-4424.
- Ehrmann, M., Pfajfar, D., & Santoro, E. (2015). Consumers' attitudes and their inflation expectations. Finance and Economics Discussion Series 2015-015. Washington: Board of Governors of the Federal Reserve System, <http://dx.doi.org/10.17016/FEDS.2015.015>.
- Galí, J., & Gertler, M. (1999). Inflation dynamics: a structural econometric analysis. *Journal of Monetary Economics*, 44, 195–222.
- Man, C.H., & Peterson, M. (2019). How does inflation expectation explain the undershooting of inflation target in Japan? Time series analysis within the frame of hybrid Phillips Curve model. Jönköping University, Jönköping International Business School, Economics.
- Midzuno, H. (1952). On the sampling system with probabilities proportionate to sum of sizes. *Annals of the Institute of Statistical Mathematics*, 3, 99–107.
- Murphy, R., & Rohde, A. (2018). Rational bias in inflation expectations. *Eastern Economic Journal*, 44, 153–171.
- Pattanaik, S., Muduli, S., & Ray, S. (2020). Inflation expectations of households: do they influence wage-price dynamics in India? *Macroeconomics and Finance in Emerging Market Economies*, 13(3), 244-263.

- Shaw, P. (2019), Using rational expectations to predict inflation”, *Reserve Bank of India Occasional Papers*, 40(1), 85-104.
- Singh, D. P., Mishra, A., & Shaw, P. (2022). Taking cognisance of households’ inflation expectations in India. *Reserve Bank of India Working Paper Series No. 02*.
- Souleles, N. S. (2004). Expectations, heterogeneous forecast errors, and consumption: micro evidence from the Michigan Consumer Sentiment surveys. *Journal of Money, Credit, and Banking*, 36(1).
- Taylor, J. B. (1982). The role of expectations in the choice of monetary policy. Monetary Policy Issues in the 1980s, Economic Symposium Conference Proceedings August 9-10, 1982, pages 47-76.
- Zhao, Y. (2021). Uncertainty and disagreement of inflation expectations: evidence from household-level qualitative survey responses. *Journal of Forecasting*, 41(4), 810-828.

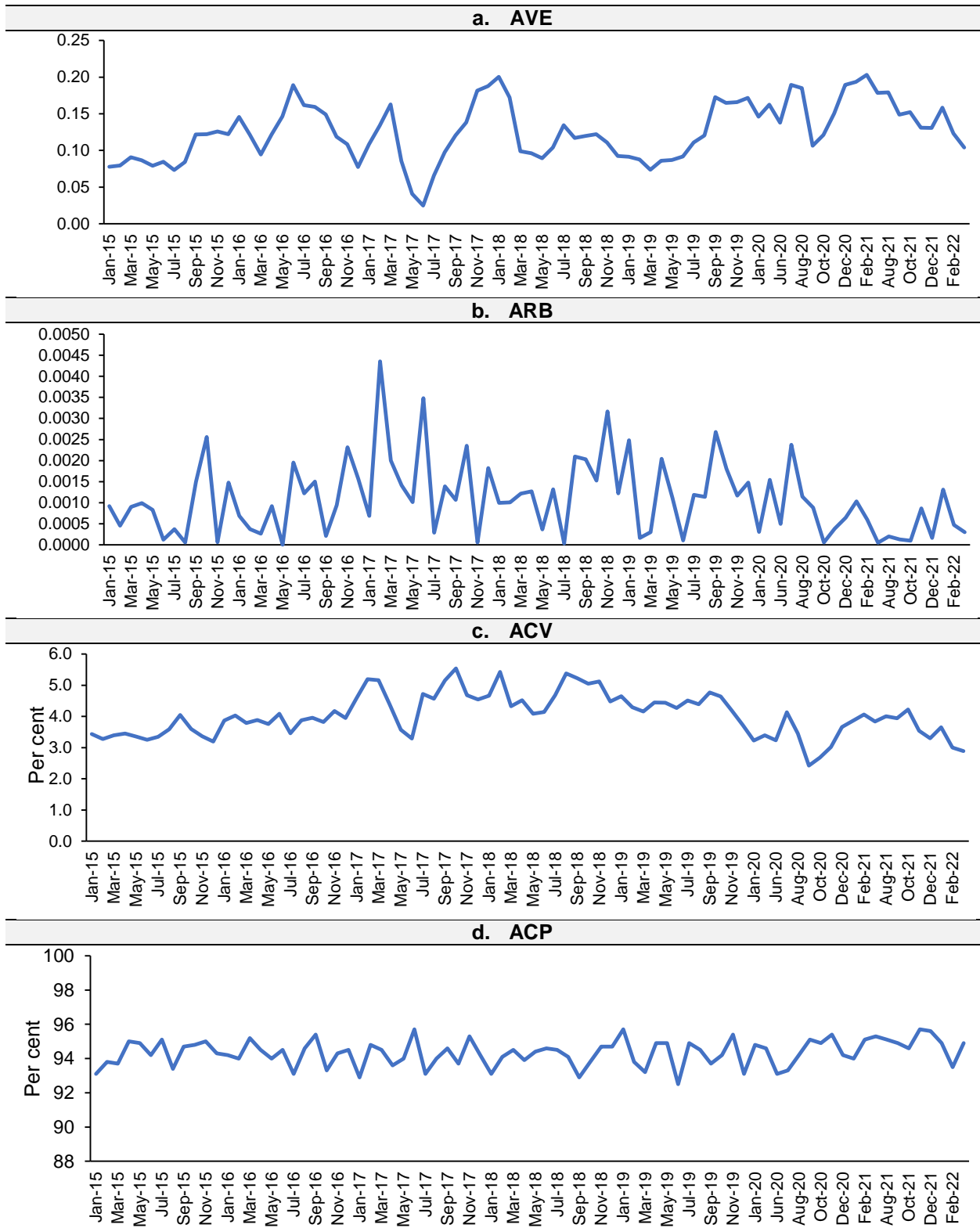
Appendix

Chart A1. Performance of Estimates Computed Using Samples Drawn From the ‘Logic Population’ – SRSWR



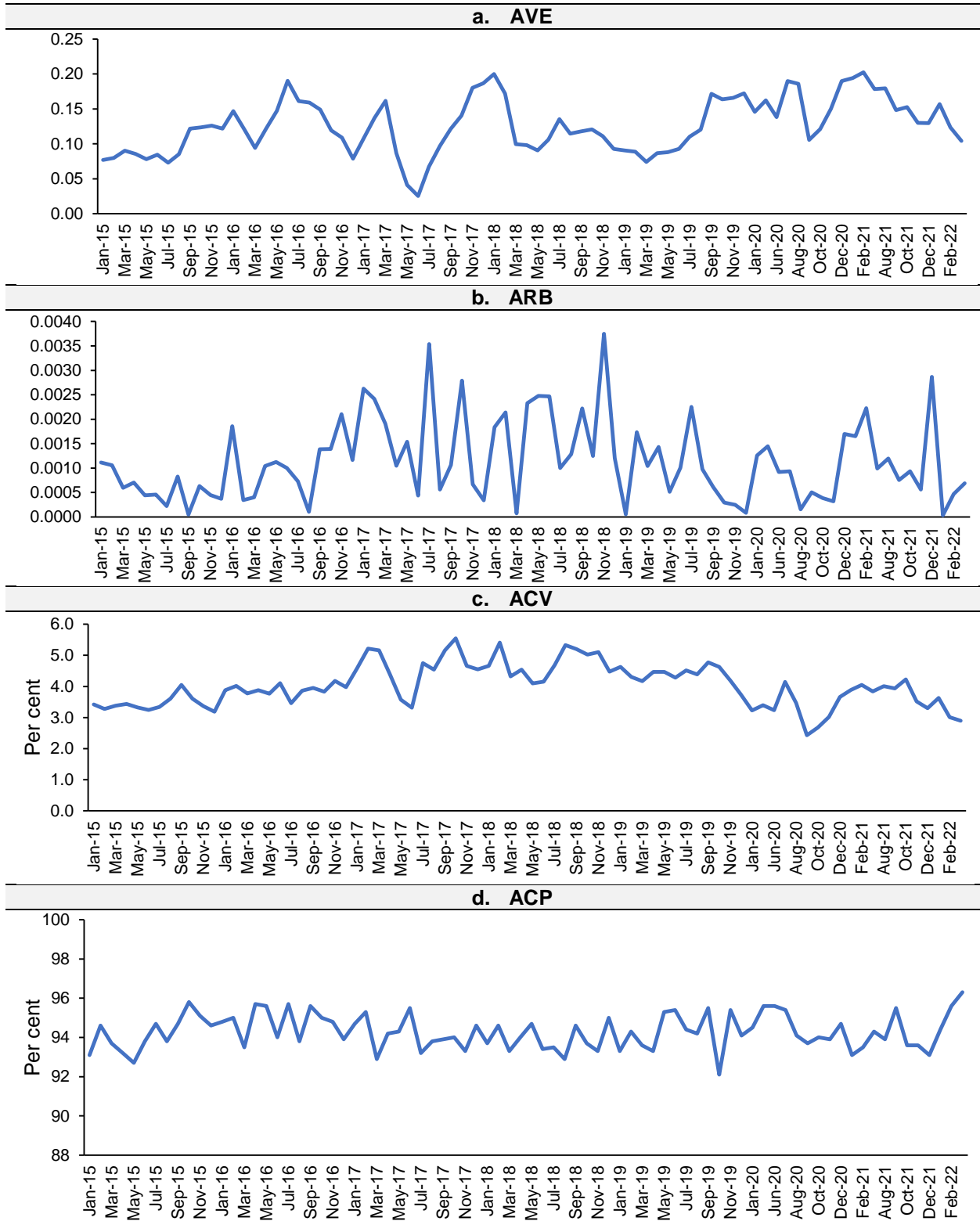
Source: Author's calculation.

Chart A2. Performance of Estimates Computed Using Samples Drawn From the 'Logic Population' – SRSWOR



Source: Author's calculation.

Chart A3. Performance of Estimates Computed Using Samples Drawn From the 'Logic Population' – Midzuno (1952)



Source: Author's calculation.