

소비자동향조사 순환표본 도입에 관한 연구

박인호*, 조천희**

한국은행의 소비자동향조사는 소비자들의 경기나 생활 형편 등에 대한 주관적 판단과 전망, 미래 소비지출 계획 등을 설문 조사를 통해 지수화하여, 가계의 소비 동향 및 전망 등을 정확하게 파악하는 것을 목적으로 한다. 소비자동향조사는 현재 5년 고정표본방식으로 운영되고 있는데 최근 들어 급격히 증가하고 있는 무응답으로 인해 표본운영의 어려움을 겪을 것으로 예상되고 있어 향후 순환표본방식으로 변경을 검토 중에 있다.

본 연구는 소비자동향조사에 순환표본을 도입하기 위해 고려해야 할 것들을 살펴보고, 이론과 실무적 측면에서 고려한 순환모형, 표본운영, 표본추출 등을 제시하였다. 제안한 순환모형은 응답부담을 낮추어 표본대상의 중간탈락을 최소화할 수 있는 1/G 모형을 고려하되 조사의 운영관리 업무부담을 고려하여 월 단위가 아닌 연 단위 순환집단의 교체를 고려한 블록순환 방식이다. 제안된 순환표본은 연도별로는 80%의 표본이 중복되며 특정연도내에서는 100%의 표본이 중복되어 시점 간 순변동이 안정적으로 추정될 것으로 기대된다.

I. 서론

II. 반복조사를 위한 순환표본

1. 반복조사
2. 순환표본

III. 소비자동향조사를 위한 순환표본

1. 현행 조사설계 및 소비자동향지수
2. 순환표본 도입과 관련된 기존연구 및 시사점
3. 순환표본 제안

IV. 결론

* 부경대학교 통계학과 (e-mail: ipark@pknu.ac.kr, phone: 051-629-5542)

** 한국은행 경제통계국 (e-mail: skyflier@bok.or.kr, phone: 02-759-5673)

※ 본 연구의 내용은 집필자들의 개인 의견으로 한국은행의 공식견해를 나타내는 것은 아님.

I. 서론

소비자동향조사는 소비자들의 경기나 생활 형편 등에 대한 주관적 판단과 전망, 미래 소비지출 계획 등을 설문 조사를 통해 지수화하여, 가계의 소비 동향 및 전망 등을 보다 정확하게 파악하는 데 목적이 있다. 소비자동향조사는 현재 5년 고정표본방식으로 운영되고 있다. 이는 5년 주기로 이루어지는 통계청 인구주택총조사(이하 인총) 결과를 반영하기 위해 인총 결과의 발표 시점에 맞추어 표본을 개편함은 물론 표본설계와 조사관리를 용이하게 하기 위함이다.

최근 들어 무응답이 급격히 증가하여 기존의 고정표본방식의 소비자동향조사 표본운영은 어려움을 겪을 것으로 예상된다. 또한, 통계청은 기존의 인총 집계를 2015년 이후 행정정보를 활용한 등록센서스로 변경하였고 행정정보를 활용하여 매년 인총 집계를 수정하고 있다.

경기판단을 위한 표본조사의 주된 관심은 특정 시점별 추정값(횡단면 분석)과 시점 간 추정값 변동(순변동 분석)을 살펴보는 것이다. 따라서 소비자동향조사를 위한 표본추출과 조사운영은 조사자료의 분석 목적에 맞도록 시점별 동향은 물론 시점 간 변동을 잘 반영할 수 있어야 한다. 고정표본을 이용한 반복적 조사방식은 표본운영이 매우 단순하다는 이점이 있지만, 무응답 누적에 따른 표본탈락이 심각해질 수 있다. 반면, 시점별로 독립적인 표본을 추출하여 조사하는 방식은 횡단면 대표성을 높일 수는 있지만, 시점 간 변동 측정에는 바람직하지 않다. 따라서 절충적 표본운영인 기존 표본의 일부만을 포함하고 나머지는 신규표본을 투입하는 방식을 통해 시점별 통계량의 대표성을 강화하는 것은 물론 시점 간 순변동을 잘 대표할 수 있을 것이다.

한국은행 경제통계국은 현재 소비자동향조사의 표본설계를 고정표본에서 순환표본으로 변경할 것을 검토 중에 있다. 본 연구는 소비자동향조사에 적합한 순환표본에 대해 살펴보았다. 2장에서는 반복조사를 위한 표본중복의 역할을 살펴보고 국내외의 적용사례를 살펴보았다. 3장에서는 소비자동향조사에서 고려할 순환표본에 대한 기존 연구결과를 살펴보고 시사점을 도출하였으며 실무적인 측면에서 가장 적합하다고 판단되는 적용방안을 제시하였다. 4장은 본 연구에 대한 간단한 결론을 포함한다.

II . 반복조사를 위한 순환표본

1. 반복조사

반복조사(repeated survey)란 여러 시점에 걸쳐 반복적으로 수행하는 횡단면조사(cross-sectional survey)이다. 반복조사는 특정 시점 간 전체 수준(aggregate level)에서의 특성변동인 순변동(net change)¹⁾ 혹은 트렌드(trend) 분석을 가능하게 한다. 예로, 지난 2년간 가구당 연간 평균소득의 증감, 1인 가구 비율의 변화 등이 있다.

반복조사의 조사대상은 시점마다 서로 독립적으로 새롭게 추출하거나 시점 간 중복하여 포함할 수도 있다. 전자는 시점별 모집단을 잘 반영할 수 있고 특정 기간 내 평균적인 모집단 특성을 잘 파악할 수 있는 장점이 있으나, 시점별로 표본추출과 대상접촉에 필요한 조사수행의 시간 및 비용 소모가 크다는 단점도 갖는다. 반면, 후자는 중복대상에 대해서는 표본추출과 대상접촉에 드는 새로운 노력을 피할 수 있어서 조사수행의 시간 및 비용을 절감할 수 있는 장점이 있으나, 시점별 모집단 특성을 좀 더 잘 반영할 수 없는 단점도 갖는다.

시점별 조사대상의 중복규모에 따라 반복조사의 순변동 추정량이 갖는 정도수준도 영향을 받게 된다. 예로, 시점별 추정량을 $\hat{\theta}_t$ 로 정의하면, 연속된 두 시점 간 순변동은 $\hat{\delta}_{t1} = \hat{\theta}_t - \hat{\theta}_{t-1}$ 로 정의되며 해당 분산은 다음과 같이 정의된다.

$$V(\hat{\delta}_{t1}) = V(\hat{\theta}_t) + V(\hat{\theta}_{t-1}) - 2Cov(\hat{\theta}_t, \hat{\theta}_{t-1}) \quad (2.1)$$

여기서 $V(\hat{\theta}_t)$ 는 t 시점 추정량 $\hat{\theta}_t$ 의 분산이고 $Cov(\hat{\theta}_t, \hat{\theta}_{t-1})$ 는 연속된 두 시점 추정량 $\hat{\theta}_t$ 와 $\hat{\theta}_{t-1}$ 의 공분산을 나타낸다. 만약 두 시점 추정량 간에 양의 상관관계($Cov(\hat{\theta}_t, \hat{\theta}_{t-1}) > 0$)를 갖는다면, 표본중복률을 높여 식 (2.1)의 순변동 추정량의 분산을 축소시킬 수 있다. 예를 들면, 시점별 표본크기와 조사특성의 모분산이 동일하다면, 단순추출에서 표본평균간

1) 총변동(gross change)은 개별 수준에서의 변화량(예, 개인 평가가 긍정에서 부정으로 또는 그 반대로 얼마나 자주 변화하는지)을 의미하며, 순변동(net change)은 전체수준에서의 변화(예, 전체 지지율이 더 높거나 혹은 더 낮는지 여부)를 의미한다.

차이 $\bar{d}_{t1} = \bar{y}_t - \bar{y}_{t-1}$ 의 분산은 다음과 같이 유도된다(Gambino and Silva, 2009).

$$V_{srs}(\bar{d}_{t1}) = \frac{2S_y^2}{n} \left(1 - \frac{n_c}{n} \rho_{y1} \right) \quad (2.2)$$

여기서 S_y^2 는 조사특성의 모분산, n 은 표본크기, $n_c (\leq n)$ 는 시점 간 중복표본의 크기, ρ_{y1} 는 두 시점별 조사특성의 상관계수를 나타낸다. 따라서, 두 시점별 조사특성 간 양의 상관계수 $\rho_{y1} > 0$ (혹은 평균 추정량 \bar{y}_t 와 \bar{y}_{t-1} 간 양의 공분산 $Cov(\bar{y}_t, \bar{y}_{t-1}) = S_y^2 \rho_{y1} n_c / n^2 > 0$)를 갖는다면, 중복이 없는 표본($n_c = 0$)보다는 중복이 많은 표본($n_c > 0$)이 순변동 \bar{d}_{t1} 을 좀 더 정확히 추정할 수 있게 된다.

한편, 두 시점 추정량의 평균 $\hat{\mu}_{t1} = (\hat{\theta}_t + \hat{\theta}_{t-1})/2$ 의 경우, 추정량 간에 양의 상관관계 ($Cov(\hat{\theta}_t, \hat{\theta}_{t-1}) > 0$)가 존재한다면 중복을 적게 할수록 정도수준(precision level)을 높일 수 있다. 이는 기간평균 $\hat{\mu}_{t1}$ 의 분산이 다음과 같기 때문이다.

$$V(\hat{\mu}_{t1}) = V(\hat{\theta}_t) + V(\hat{\theta}_{t-1}) + 2Cov(\hat{\theta}_t, \hat{\theta}_{t-1}) \quad (2.3)$$

순변동과 기간평균을 모두 추정하는 것을 목적으로 한다면, 추정량의 상대적 중요도에 따라 적절한 중복수준을 정해야 할 것이다. 예로, 주요국의 노동력조사(labor force survey, LFS)는 매월 혹은 매분기와 같은 짧은 주기의 반복조사를 수행하는데, 시점별 추정량을 구축함에 있어 표본중복 여부와 주요 비교시점 등을 고려하여 다음과 같은 AK 합성추정량(composite estimator)을 사용한다.

$$\hat{\theta}_t^{AK} = (1-K)\hat{\theta}_t + K(\hat{\theta}_t^{AK} + \hat{\delta}_{t1}) + A(\hat{\theta}_{ut} - \hat{\theta}_{ct}) \quad (2.4)$$

여기서 A 와 K 는 적절히 선택된 상수이고, $\hat{\theta}_{ct}$ 와 $\hat{\theta}_{ut}$ 는 각각 t 시점의 중복표본과 비중복표본만을 이용한 추정량을 나타낸다(Caldwell and Cantwell, 1996). 만일 합성추정치 $\hat{\theta}_t^{AK}$ 의 가중치 $A=0$ 와 $K=0$ 을 사용한다면, AK 합성추정량은 전 시점과 중복에 대한 절충을 제거하여 원래의 t 시점 추정량 $\hat{\theta}_t$ 이 된다.

더불어 AK 합성추정량의 효율성을 기준으로 다양한 종류의 순환표본설계(rotation sample design 혹은 연동표본설계)를 고려할 수 있다. 관심 모수가 개별 시점과 시점 간 모평균 $\theta_t = \bar{Y}_t$ 들의 함수라면, 두 시점의 표본중복 수준을 50~80%로 하는 것이 효율적일 수 있음이 알려져 있다(Steel and McLaren(2009), Holt and Skinner(1989)).

2. 순환표본

가. 순환모형

순환표본조사(rotation sample survey)는 정기적인 반복조사에서 표본을 다수의 부표본(subsample) 혹은 순환집단(rotation group)으로 구성하고 시점마다 일정 비율의 부표본을 미리 정해진 규칙 혹은 순환모형에 따라 투입하고 제거하는 조사이다. 다양한 순환모형이 제시되고 있는데 응답자 부담이나 조사비용과 같은 현실적인 문제와 연결되어 선택되고 있다. 순환표본의 중복률이 높을수록 표본추정량의 안정성이 높아질 수 있지만, 대상자 조사 참여 기간이 길어져 패널 조건화(panel conditioning)²⁾와 같은 비표본오차(non-sampling error)가 발생할 가능성이 커진다. 주요국들에서는 일반적으로 현실적인 문제들을 고려한 순환표본방안을 채택하며, 주어진 방법에서 효율성을 높일 수 있는 추정량을 연구 개발하여 사용하는 전략을 취한다(박민규·김영원, 2011).

순환모형은 기본적으로 횡단면 통계가 불편추정(unbiased estimation)이 되도록 하되, 조사비용과 주요한 변화 추정량(change estimator)의 분산이 최소화될 수 있도록 고려한다. 분산 축소는 시점 간 관심변수의 상관관계와 표본중복 정도에 따라 결정된다. 표본중복이 클수록 순변동 추정량의 분산이 작아진다. 순환모형 채택시 조사횟수는 제한되어야 하는데, 이는 응답률과 조사품질 제고를 가능하게 하기 위함이다.

순환모형에는 크게 a-b-a 모형과 1/G 모형의 두 가지로 분류된다. 먼저, a-b-a 모형은 표본 개체를 a 시점 동안 조사에 포함한 후 b 시점 동안 조사에서 제외하고 다시 a 시점 동안 조사에 포함시키는 패턴을 반복한다. 예로, 미국 조사국의 Current Population Survey(이하 CPS)는 4-8-4 모형을 채택하고 있다. 순환모형 a-b-a에 따라 표본가구들이 참여하게 되는 총 조사 시점 수 m 을 강조하여 a-b-a(m)으로도 표기한다. 예로, 4-8-4 모형에서는 표본개체가 조사에 참여하는 총 시점 수가 8이므로 4-8-4(8)로도 표기한다(Steel and McLaren, 2009). 또한

2) 패널 조건화란 이전 조사 참여가 이후 조사결과에 영향을 주는 응답효과를 뜻한다. 즉, 이전 조사에 참여한 표본과 처음 참여한 표본의 응답 간에 체계적 차이가 존재함을 뜻한다.

조사에 참여하는 a 시점과 제외되는 b 시점에 더불어 참여와 제거가 반복되는 횟수를 표기하는 a-b(r) 모형으로도 표기가 가능하다. 예로, 앞서의 484 모형은 48(2)로도 표기할 수 있다.

반면 1/G 모형은 표본개체에 대해 G 시점 동안 조사에 포함된 후에 조사에서 완전히 제외한다. 예로, 1/6 모형은 표본 개체를 6개 시점 간 조사에 참여시킨 후 조사에서 완전히 제외시킨다. 앞서 a-b-a(G)에서 b=0와 G=a로도 표기할 수 있고 또는 a-b(r)에서 b=0와 r=1로도 표기할 수 있다. 예로, 1/6 모형은 6(6)이나 6-0(1)로 표기할 수 있다.

나. 국내외 적용사례

본 절에서는 순환모형을 적용한 반복조사의 국내외 사례를 월 단위와 연 단위 조사주기 별로 살펴본다.

1) 월 단위 조사 사례

미국 조사국의 CPS는 484 모형을 사용하는데, 월간 변동추정량과 연간 변동추정량의 분산을 최소화하며 연속적 조사 기간에 대한 응답 부담을 분산시키는 것을 목표로 한다. 484 혹은 48(2) 시스템을 채택하여 월별 중복과 연별 중복을 각각 75%와 50% 확보시키고 8회에 걸친 조사 부담을 총 16개월로 분산시키고자 하였다.

캐나다 통계청의 노동력조사(LFS)는 1/6 모형을 채택하고 있는데, 연속된 두 달 간 변동추정량의 분산을 최소화하며 응답 부담의 경감을 목표로 한다. 1/6 혹은 6-0(1) 모형을 채택한 이유는 50%의 높은 월간 중복률을 확보하기 위한 것으로 판단된다.

미국 미시간대학 소비자연구센터의 소비자동향조사(Surveys of Consumers)는 순환표본의 특수한 방식을 채택하고 있다. 일반적 순환모형은 아니며, 월별 조사를 위한 표본대상의 60%는 신규로 40%는 6개월 전 조사대상 가구 중에서 선정한 가구로 구성한다(Curtine, 2005). 이는 반복조사로 인한 탈락(attrition)을 감안하여 신규표본의 구성비를 높게 한 것으로 판단된다.

우리나라 통계청의 경제활동인구조사(이하 경활)는 2007년 가구조사의 표본개편을 통해 표본조사방식을 5년 주기의 고정표본(fixed sample)에서 1/36 모형으로 변환하였다(통계청, 2007). 이는 1) 인구주택총조사(이하 인총)의 개편 주기 5년(5년×12개월)에 맞춘 60개월 지속적인 조사라는 기존 표본운영방식으로 인한 응답자 부담을 36개월(3년)으로 축소하기 위함인

다. ii) 더불어 인종 개편에 맞춘 표본재설계에 따른 표본교체로 인해 발생할 수 있는 시계열의 단절 가능성을 순환표본방식을 통해 줄여주기 위함이다. 1/36 모형은 전월 대비 97.2%와 전년 대비 66.7%의 높은 중복률을 갖는다.

<표 1>은 국가별 대표적인 순환모형들의 내역을 비교하고 있다. 일반적으로 채택되는 순환모형에 비해 1/36 모형은 조사 기간도 길고 순환집단 수도 매우 많다. 미국 CPS가 채택하는 4-8-4 모형은 8개의 순환집단을 포함하며 전월(년) 대비 중복률이 75%(50%)이다. 캐나다 LFS가 채택하는 1/6 모형은 6개 순환집단을 갖고 전월(년) 대비 중복률이 83%(0%)이다. <표 2>는 한국, 미국, 캐나다 3국의 노동력조사에서 채택한 순환모형의 시스템, 순환집단수, 집단규모, 총표본수를 비교하고 있다.

<표 1> 대표적 노동력조사 순환모형

내역	국가별조사					
	미국 CPS	캐나다 LFS	호주 LFS	영국 LFS	일본 LFS	브라질 LFS
조사빈도	월별	월별	월별	분기별	월별	월별
수집기간	1주	1주	2주	13주	2주	4주
순환모형	4-8(2)	6-0(1)	8-0(1)	1-2(5)	2-10(2)	4-8(2)
월중복(%)	75	83	88	0	50	75
분기중복(%)	25	50	63	80	0	25
연중복(%)	50	0	0	20	50	50
표본가구수	60,000	53,000	29,000	20,000	40,000	41,600

출처: Gambino and Silva (2009)

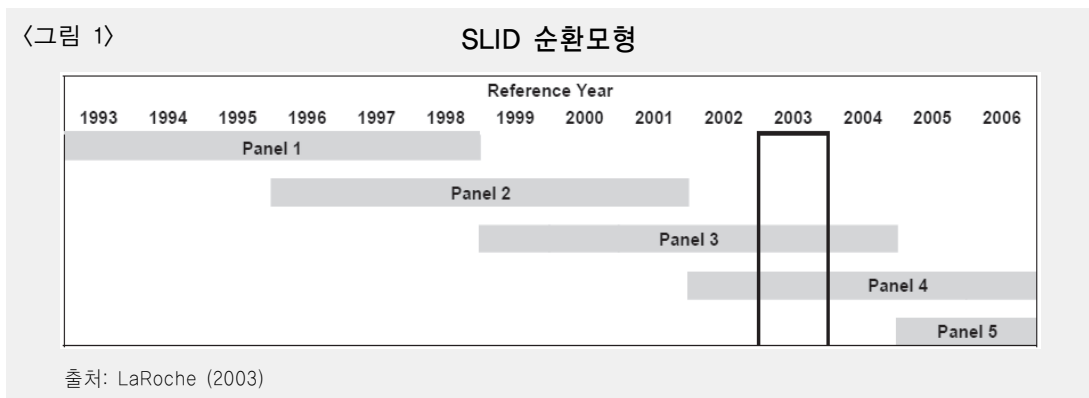
<표 2> 한국·미국·캐나다 노동력조사 순환모형 비교

조사	순환모형	순환집단수	집단규모	총표본수
한국 경찰	1/36	36	905가구	32,580가구
미국 CPS	4-8-4	8	9,000가구	72,000가구
캐나다 LFS	1/6	6	8,775가구	52,653가구

2) 연단위 조사 사례

캐나다 통계청의 **Survey of Labour Income Dynamics**(이하 **SLID**)는 개인의 경제적 복지(well-being)의 변화와 그 변화에 영향을 미치는 요인을 측정하기 위해 1993년(기준년)부터 시작한 종단면조사(longitudinal survey)이다. 표본 크기는 약 35,000개 가구이며, 동 기관의 노동력조사(LFS)로부터 부표본을 표집한다. SLID의 순환 방식은 부표본, 즉 패널별로 조사 지속기간이 6년이며, 전체 표본은 크기가 유사한 2개 패널로 나누어지고, 각 패널은 약 17,000가구(34,000 가구원)로 구성된다. 2개 패널은 항상 중복되게 운영되며, 3년마다 기존 패널 중 하나를 탈락시키고 새로운 패널을 선정하여 교체한다(<그림 1> 참조).

SLID 패널은 LFS 표본 중 순환조사주기를 마치고 제외되는 응답자들만 포함시킨다. LFS 마지막 달 조사는 SLID 조사를 위한 소개접촉으로 활용되며 SLID의 첫 번째 조사를 위한 접촉과 관련한 기본정보를 수집할 기회로 사용된다.



가계금융복지조사(이하 **가금복**)는 통계청, 금융감독원, 한국은행이 공동으로 2012년부터 매년 약 2만 가구에 대해 실시하는 가구 대상 면접조사이다. 가구의 자산, 부채, 소득, 지출 등을 종합적으로 파악하여 가구의 경제적 수준 및 변화요인 등을 분석하고, 경제사회 및 복지와 관련한 정책 연구에 활용하는 것을 목적으로 한다.

2012년의 1차 조사부터 2014년의 3차 조사까지 고정패널(fixed panel) 형태로 조사를 수행해 오다가, 2015년의 4차 조사부터는 순환패널조사 방식(rotation panel survey)으로 전환되었다. 응답자 부담으로 인한 조사거부 및 패널탈락을 지양하고, 더불어 조사목적, 비용 및 분석에 대한 추가적 고려가 이러한 전환의 주요 이유인 것으로 파악된다. 기존 고정패널을 약 4,000여 가구로 구성한 5개의 순환집단으로 나눈 후 2015년부터 매년 한 개의 순환집단

을 표본에서 탈락시키고 동시에 약 4,000가구의 새로운 표본을 추출하여 조사에 포함하는 연단위 1/5 순환모형으로 적용하였다(백지선 외 3인, 2014). 5년 주기의 고정패널에서 4년차에 순환패널로 변경함에 따라 2018년까지 구표본의 패널 유지 기간은 3-7년으로 달라진다. <그림 2>는 2012-2019년 가금융복지패널의 순환모형을 요약하고 있다.

<그림 2> 2012~2019년 가금융복지패널의 순환모형*

	1차	2차	3차	4차	5차	6차	7차	8차
	2012	2013	2014	2015	2016	2017	2018	2019
	(고정패널)			(5년 연동패널)				
구표본			[1]					
			[2]	[2]				
			[3]	[3]	[3]			
			[4]	[4]	[4]	[4]		
			[5]	[5]	[5]	[5]	[5]	
신표본				[1]	[1]	[1]	[1]	[1]
					[2]	[2]	[2]	[2]
						[3]	[3]	[3]
							[4]	[4]
								[5]

* 출처: 백지선 외 (2014, 그림 3-1, 78쪽)

III. 소비자동향조사를 위한 순환표본

1. 현행 조사설계 및 소비자동향지수

가. 표본설계

소비자동향조사는 우리나라 소비자의 경기에 대한 인식 및 가계재정상황, 소비지출행태 등에 대한 항목을 조사하며 성별, 연령별, 직업별, 소득수준별, 주거상황별, 거주지역별로도 구분하여 조사한다. 조사대상은 통계청의 인구주택총조사에서 파악한 일반가구로 한다. 최근의 표본개편은 2018년에 이루어졌으며 사용된 표본추출틀은 2015년 인총 결과로 구축된 조사구 명부이다.

목표응답수는 본점 조사를 위한 2,500가구와 지역본부 조사를 위한 4,050가구를 합하여 총 6,550개 가구이다. 목표응답수를 달성하기 위해 무응답 및 기타 사유의 탈락을 고려하여 최종 목표 표본가구수는 이의 10배수인 총 65,500개로 정하였다. 표본추출틀로 인총 조사구 명부를 사용함에 따라 일차추출단위로 조사구를 추출하고 추출된 표본조사구 내에서 표본가구를 순차적으로 추출하는 이단추출방식을 적용하였다. 표본조사구는 총 13,100개로 지역별로 배분된 표본가구수의 1/5 규모로 선정하였는데, 이는 조사구 당 5개 표본가구를 목표로 하였기 때문이다.

표본추출에 앞서 지역 및 조사구별 가구주 평균연령을 기준으로 모집단 층화(stratification)를 하였다. 1차 층화는 시도 기준의 18개 지역(17개 특광역시도를 기초하되 강원도는 영서와 영동으로 세분)을 고려하였고, 2차 층화는 조사구 내 가구주 평균연령을 5개 연령그룹(30대 이하, 40대, 50대, 60대, 70대 이상)으로 나누었다. 지역층별로 조사구 특성(아파트, 일반), 행정구역(시군구), 가구주 평균연령, 가구평균 공시지가, 가구평균 주거용 연면적 등의 순서로 정렬한 뒤, 가구수에 비례하여 계통확률추출로 표본조사구를 선정하였다. 단, 응답률이 현저히 낮은 30대 이하 가구는 과대추출하였다. 표본조사구 내 전체 376,195개 가구 중 총 128,270개 가구를 표본가구로 추출하였다.

나. 소비자동향지수(CSI)

소비자동향조사의 소비자 응답내용에 따라 개별문항별 소비자동향지수(consumer survey index, CSI)를 산출하고, 생활형편 등 소비자심리의 종합적 판단에 유용한 중요지수 6개(현재생활형편, 생활형편전망, 현재경기판단, 향후경기전망, 소비지출전망, 가계수입전망)를 합성하여 소비자심리지수(composite consumer sentiment index, CCSI)를 작성한다. T 시점 중의 시계열 자료를 활용하여 구성 지수의 수준 및 진폭을 일정하게 표준화하여 합성한다. 간단히 요약하면 다음과 같다. 먼저, 개별문항지수는 개별항목의 5점 척도를 0-200점 스케일(scale)로 변환해 주는 다음의 식으로 산출한다.

$$\widehat{CSI}_w(q) = 100 \times \left(1 + \frac{\widehat{Y}_w(q)}{\widehat{N}_w} \right)$$

여기서 응답자 k 에 대한 문항 q 의 조사값은 $y_{qk}=1$ (매우 긍정), 0.5(다소긍정), 0(보통), -0.5(다소부정), -1.0(매우부정)이고, w_k 는 표본가중치, $\widehat{Y}_w(q) = \sum_{k \in s} w_k y_{qk}$ ($q = 1, 2, \dots, 6$)와 $\widehat{N}_w = \sum_{k \in s} w_k$ 는 각각 조사특성(항목) y 의 가중총합과 모집단 총수 추정량이다.

합성지수는 각각 다른 특성을 지닌 개별지수를 표준화하여 다음의 식을 통해 합성하여 소비자 심리에 대한 종합적인 판단이 가능하도록 하고 있다.

$$CCSI_t = \left(\frac{Z_t - \bar{Z}}{s_Z} \right) \times 10 + 100$$

여기서 Z_t 는 t 시점의 6개 개별 표준화 지수의 합이고, \bar{Z} 는 Z_t 의 표준화 구간 중 평균이며, s_Z 는 Z_t 의 표준화구간 중 표준편차를 각각 나타낸다.

2. 순환표본 도입과 관련한 기존연구 및 시사점

가. 기존연구

순환표본설계와 관련한 기존 한국은행의 연구로는 김영원·이수희(2001)와 신기일·이상은

(2012)이 있다. 먼저, 김영원·이수희(2001)는 일반론적 관점에서 순환표본설계의 효율성을 연구하고 한국은행의 표본조사 적용시 고려할 점들에 대해 다루고 있다. 먼저, 순환표본조사의 정의와 순환모형별 사례를 간단히 소개하고, 사례별 AK 합성추정량과 분산식을 유도하였다. 고정된 비용 하에서 응답횟수에 따른 조사비용을 고려하여 각 순환모형별 표본수 및 효율성을 논의하였다. 또한 우리나라 순환표본조사를 위한 순환모형을 도입하는 경우에 월 단위 1/6 모형의 채택이 바람직할 것으로 판단하였는데, 특히 소비자동향조사에 적용하는 것이 바람직할 것으로 보고하고 있다. 하지만, 연속조사에 있어 더욱 심각한 무응답의 증가는 심각한 편향을 초래할 수 있음도 지적하였다.

신기일·이상은(2012)은 캐나다와 미국의 순환표본조사에 적용하는 AK 합성추정량을 CSI 항목 응답값(1-5점)에 대해 시산하였다. 모의실험을 통해 소비자동향조사에 대한 AK 합성추정량의 효율성을 평가하였는데, 두 시점 간 조사특성의 상관관계가 낮은 경우에는 식 (2.4)의 가중치 A와 K에 “0”을 대입하는 것과 큰 차이를 보이지 않아, 순환표본을 통한 정도수준 향상은 없다고 평가하였다. 특히 5점 척도로 측정함에 따른 다양한 자료분석의 한계를 지적하였다. 조사 현실을 반영하여 중간탈락이 없는 월 단위 1/6 순환모형이 최적의 방법이라고 판단하였다. 하지만 적용의 이점이 없어 결론적으로는 AK 합성추정량이라는 추정량 선택보다는 순환모형도입에 따른 표본관리 가능성에 대한 연구가 필요함을 지적하였다.

나. 조사운영 관점의 고려사항

순환표본 조사설계 시 야기되는 구조적 복잡성을 감안하면, 조사수행 기관의 조사여건이 이를 담당하기에 충분한지를 우선적으로 고려하여야 할 것이다. 신기일·이상은(2012)은 순환표본을 도입하면 표본관리를 위한 업무량 및 비용이 증가하고, AK 합성추정량의 산출과 이에 따른 복잡한 분산추정을 담당해야 하는 업무량 증가를 초래할 수 있음을 지적하였다.

소비자동향조사는 주로 우편·팩스를 이용하는 용이한 조사모드로 운영되고 있으나, 기존의 고정표본방식에서 순환표본의 적용에 따른 복잡한 순환집단의 선택과 관리 그리고 추정방식과 조사운영으로 변경하기에는 현재 담당인력으로는 다소 무리가 있을 것으로 판단된다. 적절한 표본관리 전략과 필요 인력의 보강 및 조사체계의 개발과 관련한 추가 연구가 필요할 것으로 판단된다.

그럼에도 불구하고 순환표본 도입을 검토해야 할 이유는 최근의 무응답 증가로 인해 기존의 고정표본 방식으로는 표본 유지의 어려움이 예상된다. 더불어 5년 주기로 이루어지는 표본개편은 시계열의 단절을 초래할 수 있어 가능한 일정규모 이상의 중복을 허용하는 표본운영방식으로서의 전환이 필요해 보인다. 순환집단의 조사 포함과 이탈을 반복적으로 허용

하는 a-b-a 모형보다는 김영원·이수희(2001)와 신기일·이상은(2012)이 제안한 것과 같이 한번 포함된 순환집단을 연속적으로 조사하고 이탈시키는 1/G 모형을 고려하는 것이 순환집단의 안정적인 응답 유지에 좀 더 바람직하며 최근 모집단을 잘 반영할 수 있을 것으로 판단된다.

3. 순환표본 제안

가. 순환모형

기본적인 순환모형의 선택은 김영원·이수희(2001)와 신기일·이상은(2012)이 논의한 바와 같이 1/G 모형이 a-b-a 모형보다 연속조사에 따른 무응답을 줄일 수 있을 것으로 판단된다. 하지만, 담당 인력이 많지 않고 매달 조사·편제 업무가 진행되므로, 매 조사 시점, 즉 매달 신규표본을 추출하여 조사에 포함하고 기존 표본 중 일부를 탈락시키는 작업을 반복하기에는 업무부담이 매우 클 것이다. 따라서 1/G 모형을 선택하되 월 단위보다는 연 단위로 운영하는 것을 제안한다. 즉, 순환집단을 연도별로 순환시키되 연도 내 매월 동일대상에 대해 반복적으로 조사하는 블록순환 방식(block-rotation sampling)의 혼합적(hybrid) 접근이 필요할 것으로 판단된다.

본 연구에서 제안하는 연 단위 1/G 블록순환모형은 매년 전체 표본의 1/G 크기의 부표본, 즉 순환집단을 추가한 후 매월 반복적으로 조사를 수행하는 것이다. 통계청의 인구주택총조사는 5년 주기로 개편하고 있어 연 단위 모형하에서는 G=5가 바람직하다. <표 3>은 5년 주기 블록순환모형을 도식화하고 있다. 제안된 모형은 인접한 연도별로는 80%의 표본이 중복되며 연도 내에서는 100%의 표본이 중복된다. 신규 투입되는 순환집단은 G년 주기별로 혹은 매년 새롭게 추출하는 다음의 두 가지 방안을 고려할 수 있다.

방안 1은 5년 단위 표본개편 및 주기 내 순환표본을 운영하는 방법으로 주기 내 첫해에 표본을 개편하고 추출된 순환집단들에 대한 조사관리만을 수행한다. 이 방안은 표본추출 이후 시점에서 순환집단의 정보 노후화³⁾ 가능성이 있고, 부처 내 순환보직에 따른 담당자 변경과 업무 인수인계로 인해 발생할 수 있는 운영관리 미비 등 추가적인 정보 손실의 가능성도 존재한다.

3) 소비자동향조사의 조사대상 추출은 통계청 인총자료를 이용한다. 통계청은 2015년 이후 기존의 모든 가구를 방문하던 전통적 조사방법에서 전체 가구에 대해 행정정보를 활용하되 20% 가구만을 방문하여 세부정보를 수집하는 등록센서스로 변경하였다. 따라서 매년 갱신되는 행정정보를 활용하여 5년 주기로 정의되는 조사구에 대한 보조적 정보도 매년 갱신하고 있다.

방안 2는 5년 단위 표본개편의 틀에서 1/G 순환집단을 매년 새롭게 추출하는 것이다. 이 방안은 순환집단을 추출하면서 최신 정보를 반영할 수 있고 조사 투입 시점에 신규 조사대상자를 접촉함에 따라 응답자 충성도를 확보할 수 있고 신규표본에 대한 초기 접촉 노력을 연도별로 분산시킬 수 있는 장점이 있다. 하지만, 표본추출 및 자료구축 작업을 매년 반복적으로 수행해야 하는 업무부담이 증가할 수 있다.

〈표 3〉 5년 주기 블록-순환 모형

조사시점		표본	패널 1					패널 2				
연	월	s	A	B	C	D	E	A	B	C	D	E
t	1	$s_{t,1}$	A_1	B_1	C_1	D_1	E_1					
	2	$s_{t,2}$	A_2	B_2	C_2	D_2	E_2					
	⋮	⋮	⋮	⋮	⋮	⋮	⋮					
	12	$s_{t,12}$	A_{12}	B_{12}	C_{12}	D_{12}	E_{12}					
t+1	1	$s_{t+1,1}$		B_1	C_1	D_1	E_1	A_1				
	2	$s_{t+1,2}$		B_2	C_2	D_2	E_2	A_2				
	⋮	⋮		⋮	⋮	⋮	⋮	⋮				
	12	$s_{t+1,12}$		B_{12}	C_{12}	D_{12}	E_{12}	A_{12}				
t+2	1	$s_{t+2,1}$			C_1	D_1	E_1	A_1	B_1			
	2	$s_{t+2,2}$			C_2	D_2	E_2	A_2	B_2			
	⋮	⋮			⋮	⋮	⋮	⋮	⋮			
	12	$s_{t+2,12}$			C_{12}	D_{12}	E_{12}	A_{12}	B_{12}			
t+3	1	$s_{t+3,1}$				D_1	E_1	A_1	B_1	C_1		
	2	$s_{t+3,2}$				D_2	E_2	A_2	B_2	C_2		
	⋮	⋮				⋮	⋮	⋮	⋮	⋮		
	12	$s_{t+3,12}$				D_{12}	E_{12}	A_{12}	B_{12}	C_{12}		
t+4	1	$s_{t+4,1}$					E_1	A_1	B_1	C_1	D_1	
	2	$s_{t+4,2}$					E_2	A_2	B_2	C_2	D_2	
	⋮	⋮					⋮	⋮	⋮	⋮	⋮	
	12	$s_{t+4,12}$					E_{12}	A_{12}	B_{12}	C_{12}	D_{12}	
t+5	1	$s_{t+5,1}$						A_1	B_1	C_1	D_1	E_1
	2	$s_{t+5,2}$						A_2	B_2	C_2	D_2	E_2
	⋮	⋮						⋮	⋮	⋮	⋮	⋮
	12	$s_{t+5,12}$						A_{12}	B_{12}	C_{12}	D_{12}	E_{12}

나. 표본추출

순환모형의 선택과 더불어 표본추출은 다음의 2가지 방안을 고려할 수 있다. 먼저, 1안은 기존 방식과 동일하게 통계청의 최신 인총자료를 이용하여 조사구를 추출하고 표본조사구 내 가구를 추출하는 이단계확률추출(two-stage probability sampling)이다. 2안은 기존 표본조사와 연계하여 순환집단을 추출하고 구축하는 방법으로 II장 2절에서 살펴본 캐나다 통계청의 SLID 표본설계 사례를 참고할 수 있다. 예로, 통계청·금융감독원·한국은행이 공동으로 수행하는 가계금융복지조사나 통계청의 가계동향조사 등과 같이 소비자(혹은 가구)를 대상으로 한 기존 표본조사와 연계하는 것이다. 단, 2안은 기존 응답 표본가구에 대한 기관 혹은 업무팀 간 정보공유가 가능할 때 고려할 수 있다.

표본규모는 최초 접촉 및 표본투입($t=1$)과 이후 시점($t=2, \dots, G$)별 응답률 RR_t 를 기준으로 역산하여 필요한 초기투입 표본규모를 다음과 같이 산출할 수 있다.

$$n_{G\text{-시점(목표)응답수}} = n_{\text{최초표본수}} \times \prod_{t=1}^G RR_t$$

〈표 4〉 순환집단별 예상 응답수 예측 예시

연도	예상 응답률 (전년대비)	예상응답수								예상 응답수 합계
		순환집단	A1	B1	C1	D1	E1	A2	B2	
		최초 표본수	800	800	800	800	800	800	800	
		연도								
1	80%	1	640							:
2	90%	2	576	640						:
3	90%	3	518	576	640					:
4	90%	4	467	518	576	640				:
5	90%	5	420	467	518	576	640			2,621
		6		420	467	518	576	640		2,621
		7			420	467	518	576	640	2,621
		8				420	467	518	576	:
		9					420	467	518	:
		10						420	467	:
		11							420	:

조사환경은 지속적으로 변화하고 있고 응답률은 점차 낮아지고 있기 때문에 다소 보수적인 예측표본의 확보가 필요하다. <표 4>는 앞서 표본추출 2안인 연계추출하에서 본부 조사 기준의 순환집단별 시점별 응답수에 대한 예시를 보여준다. 예를 들면, 순환집단별 최초 표본수를 800개로 하고 연도별 응답률은 최초 접촉에서 80%, 이후 연도별로 90%이라면 매년 응답수는 2,621개로 예상될 수 있다. 물론, 1안의 표본추출을 적용한다면 최초연도 접촉에 따른 응답률은 30% 이하로 낮아질 수도 있을 것이다.

IV. 결론

본 연구는 소비자동향조사에 순환표본을 도입하기 위해 고려해야 할 사항을 살펴보고, 이론 및 실무적 측면을 고려한 순환모형, 표본개편방안, 표본추출방안 등을 제시하였다. 제안한 순환모형은 표본대상의 중간탈락을 최소화할 수 있도록 G-시점을 연속해서 조사하는 1/G 모형이 적절할 것으로 판단된다. 더불어 순환표본의 운영관리 부담을 최소화할 수 있도록 연도별 순환집단 교체를 고려하는 블록-순환 방식이 바람직하다. G-연도 단위 표본개편을 고려할 수도 있으나, 가능하다면 담당 인력의 업무부담을 참고하여 최신 모집단 정보를 반영할 수 있도록 매년 신규 부표본을 추출하여 추가시키는 방안의 검토도 필요하다.

통계청과의 협력 여부에 따라 기존 방식의 인종자료를 이용한 표본추출을 고려하거나, 가계금융복지조사 혹은 가계동향조사의 표본가구와 연계한 표본추출을 고려할 수도 있을 것이다. 후자는 초기 응답자와의 라포(rapport) 형성은 물론 조사 참여자들의 다양한 경제·사회·인구학적 보조정보 수집에도 유리할 수 있다.

소수의 담당 인력으로 순환표본의 전환을 감당하는 것은 다소 무리일 수는 있지만, 최근 들어 급격히 증가하는 무응답으로 인한 표본손실은 순환표본의 도입에 대한 충분한 동기가 될 수 있다. 향후 순변동 추정량의 효율성에 대한 추가적인 연구가 필요할 것이다. 본 연구결과가 소비자동향조사의 조사운영과 추정결과의 신뢰성 제고에 도움이 되길 기대한다.

참고문헌

- 김영원·이수희(2001), “순환표본설계의 효율성에 관한 연구”, 한국은행 계간국민계정, 제4호, pp.28~45.
- 백지선·김재광·민경아·이상기(2014), “가계금융·복지조사 연동패널 표본설계와 횡단가중치 작성 연구”, 2014년 하반기 연구보고서, 통계개발원.
- 박민규·김영원 (2011), “가구부문 연동모형 및 추정방법 개선연구”, 한국통계학회.
- 신기일·이상은 (2012), “소비자동향조사의 순환표집방법에 대한 연구,” 계간국민계정, 2권, 1-32.
- 통계청. (2007). “가구부문 표본개편보고서”.
- Caldwell, C. V. and Cantwell, P. J. (1996), “Comparing a Fixed-Panel Sample Design with a Rotating-Panel Sample Design for Monthly Trade Surveys,” Bureau of Census Annual Research.
- Curtine, R. T. (2005), “Survey of Consumers.” University of Michigan, www.sca.isr.umich.edu/documents.php?c=i
- Gambino, J. G. and Silva, P. L. D. N. (2009). “Sampling and Estimation in Household Surveys,” *Sample Surveys: Inference and Analysis*, 29A, pp. 407-439, Elsevier.
- Holt, D., and Skinner, C. J. (1989). “Components of Change in Repeated Surveys,” *International Statistical Review*, 57, 1-18.
- LaRoche, S. (2003), “Longitudinal and Cross-Sectional Weighting of the Survey of Labour and Income Dynamics,” *Income Research Paper Series*, Statistics Canada, 75F0002MIE.
- Steel, D. and McLaren, C. (2009). “Design and Analysis of Surveys Repeated over Time.” *Sample Surveys: Inference and Analysis*, 29B, pp. 289-313, Elsevier.