

스캐너 데이터, 소비자물가지수
작성에 어떻게 이용될 수 있을까?



목차

I. 연구 배경과 범위

II. 스캐너데이터를 이용한 연구 및 사례

III. 스캐너데이터를 이용한 우리나라 지수 시산

IV. 스캐너데이터 자료 확보방안 연구

V. 향후 과제

연구 배경

연구배경

◆ 조사환경 악화

조사 대상 업체들이 가격정보 공개하기 꺼려하고, 조사 품목의 생멸주기가 짧아져서 물가지수 구성 품목의 대표성이 낮아짐

◆ 라스파이레스 방식(고정가중)의 소비자물가지수 편 지수 작성 시점마다 주요 소비 품목과 관련 가중치를 새롭게 산출할 필요

➡ 스캐너 데이터가 소비자물가지수 작성 문제 해결의 대안

연구배경

- ◆ 스캐너 데이터 : 바코드를 이용한 상품거래 데이터
 - 바코드는 GTIN 또는 EAN 코드 : GTIN-13, EAN-13
(국가, 회사, 상품 식별 12자리와 검토표 1자리 총 13자리)



연구배경

◆ 통계청 제공 스캐너 데이터 : 데이터를 추출하여 SQL로 제공

The image shows a file explorer window on the left and a text editor window on the right. The file explorer displays a list of files with names like '대한상의.매출집계_201709.zip' and dates. The text editor shows a large dataset of scanner data with columns for various attributes and values.

FILE	수집한 날짜
대한상의.매출집계_201709.zip	2017-10-12 오전 10:48
대한상의.매출집계_20171001.txt	2017-10-19 오후 2:50
대한상의.매출집계_20170924.txt	2017-10-12 오전 10:44
대한상의.매출집계_20170917.txt	2017-10-12 오전 10:40
대한상의.매출집계_20170910.txt	2017-10-12 오전 10:37
대한상의.매출집계_20170903.txt	2017-09-20 오후 1:18
대한상의.매출집계_201708.zip	2017-09-07 오전 10:37
대한상의.매출집계_20170827.txt	2017-09-07 오전 9:34
대한상의.매출집계_20170820.txt	2017-09-07 오전 9:31
대한상의.매출집계_20170813.txt	2017-09-07 오전 9:26
대한상의.매출집계_20170806.txt	2017-09-07 오전 9:24
대한상의.매출집계_20170730.txt	2017-08-08 오후 1:08
대한상의.매출집계_201707.zip	2017-08-08 오전 11:14
대한상의.매출집계_20170723.txt	2017-08-08 오전 11:06
대한상의.매출집계_20170716.txt	2017-08-08 오전 11:03
대한상의.매출집계_20170709.txt	2017-08-08 오전 11:00
대한상의.매출집계_20161225.txt	2017-08-03 오후 1:30
대한상의.매출집계_20170625.txt	2017-07-20 오전 9:10
대한상의.매출집계_20170702.txt	2017-07-20 오전 9:08
대한상의.매출집계_20170618.txt	2017-07-20 오전 9:07
대한상의.매출집계_20161204.txt	2017-07-20 오전 8:54
대한상의.매출집계_20170604.txt	2017-07-19 오후 6:50
대한상의.매출집계_20170611.txt	2017-07-19 오후 6:48
대한상의.매출집계_201706.zip	2017-07-19 오전 10:50
대한상의.매출집계_2013.zip	2017-07-06 오후 6:27

FILE	수집한 날짜
대한상의.매출집계_20171001.txt	2017-10-12 오전 10:44
크기: 74.7MB	

FILE	수집한 날짜
대한상의.매출집계_20171001.txt	2017-10-12 오전 10:44
크기: 74.7MB	

FILE	수집한 날짜
대한상의.매출집계_20171001.txt	2017-10-12 오전 10:44
크기: 74.7MB	

연구배경

- ◆ 일부 국가 통계작성기관들은 스캐너 데이터를 품목 소비자물가지수 작성에 적용
 - 벨기에, 덴마크, 네덜란드, 스웨덴, 스위스 등
- ◆ 대한상공회의소가 POS시스템을 운영하고 있는 국내 유통업체들 일부로부터 상품의 판매데이터를 수집, 축적

연구배경

- ◆ 대한상공회의소 : 유통업체로부터 txt 또는 excel 수집
(주 1회 또는 월 1회)

	0	10	20	30	40	50	60	70
1	M3RS5G1709038809422354225	1	1800	1800				
2	UHB4311709038809175304058	1	5500	5500				
3	M1TG4V1709038801114112649	1	1400	1400				
4	NKL1861709038806002006437	1	1500	1500				
5	M1TG7V1709038801119264206	1	2700	2700				
6	LLO7111709034054500132224	1	2500	2600				
7	M1TG6V1709038410525180200	1	500	500				
8	LLO7111709038801116007257	12	54000	4500				
9	XLA1741709030000020434014	1	2000	2000				
10	PQO9871709038801128503211	1	950	1050				
11	M1TG4V1709038809042091968	1	300	300				
12	XKU1641709038809398628764	2	1800	900				
13	ASK7531709038801116000456	2	9000	4500				
14	M3RS5G1709038801094082604	2	3800	1900				
15	XKI8621709038801166027236	1	800	800				
16	M3RS5G1709038801077683408	1	1900	1900				
17	M6RS7H1709038801056610012	1	1200	1200				
18	MVW8661709038809398621840	1	3800	3800				
19	M1TG7V1709038801048953301	9	25200	2800				

연구배경

- ◆ 스캐너 데이터는 소매점에서 거래되고 있는 모든 상품에 대한 거래 가격 및 거래량 정보, 고빈도(시간별, 일별) 정보, 신상품에 대한 정보를 보유 : 낮은 수집 비용, 측정 오차 감소
 - ➡ 라스파이레스 지수보다 현실반영도가 높은 톨크비스트(Törnqvist) 지수 또는 피셔(Fisher) 지수 등 작성가능

연구배경

- ◆ 우리나라에서도 소비자물가지수 생산에 스캐너 데이터를 활용 필요 → 조사의 어려움과 상품가격의 대표성, 가중치 선정 등의 문제를 보완
 - 해외사례의 분석, 자료의 입수 가능성, 대체 가능성 및 활용
 - 스캐너 데이터를 소비자물가지수 작성에 활용하는 방안 마련

스캐너데이터를 이용한 연구 및 사례

CPI에 스캐너 데이터 활용 국가(1/2)

	적용 품목	스캐너 데이터의 비중	주요 결과
스위스	<ul style="list-style-type: none"> - 음식료품 - 유사 음식료품 	<ul style="list-style-type: none"> - 4개의 소매 체인점 - 해당 품목 시장점유율 60~70% 	<ul style="list-style-type: none"> - 향후 품목과 소매점을 늘려갈 계획
네덜란드	<ul style="list-style-type: none"> - 슈퍼마켓에서 판매하는 대부분의 품목 - 유류(에너지) - 휴대폰 - 여행상품 	<ul style="list-style-type: none"> - 10개의 모든 슈퍼마켓 체인점 - 슈퍼마켓 해당 품목의 경우 서베이 자료를 100% 대체 - CPI 전체의 약 20%에 해당 	<ul style="list-style-type: none"> - 제품의 재론칭(relaunch) 등에 따른 통계 왜곡을 완화하기 위해 QU지수의 적용범위를 확대할 계획 - 일정한 필터링을 통과한 스캐너 데이터 이용
노르웨이	<ul style="list-style-type: none"> - 식료품, 의약품, 유류, 주류 등 	<ul style="list-style-type: none"> - 3개 대형마트(시장 점유율 96%) - CPI 전체 약 30%에 해당 - 품목별 85%~100% 시장점유율 	<ul style="list-style-type: none"> - 식품 및 비주류 음료에 대한 하위 물가지수 산정

CPI에 스캐너 데이터 활용 국가(2/2)

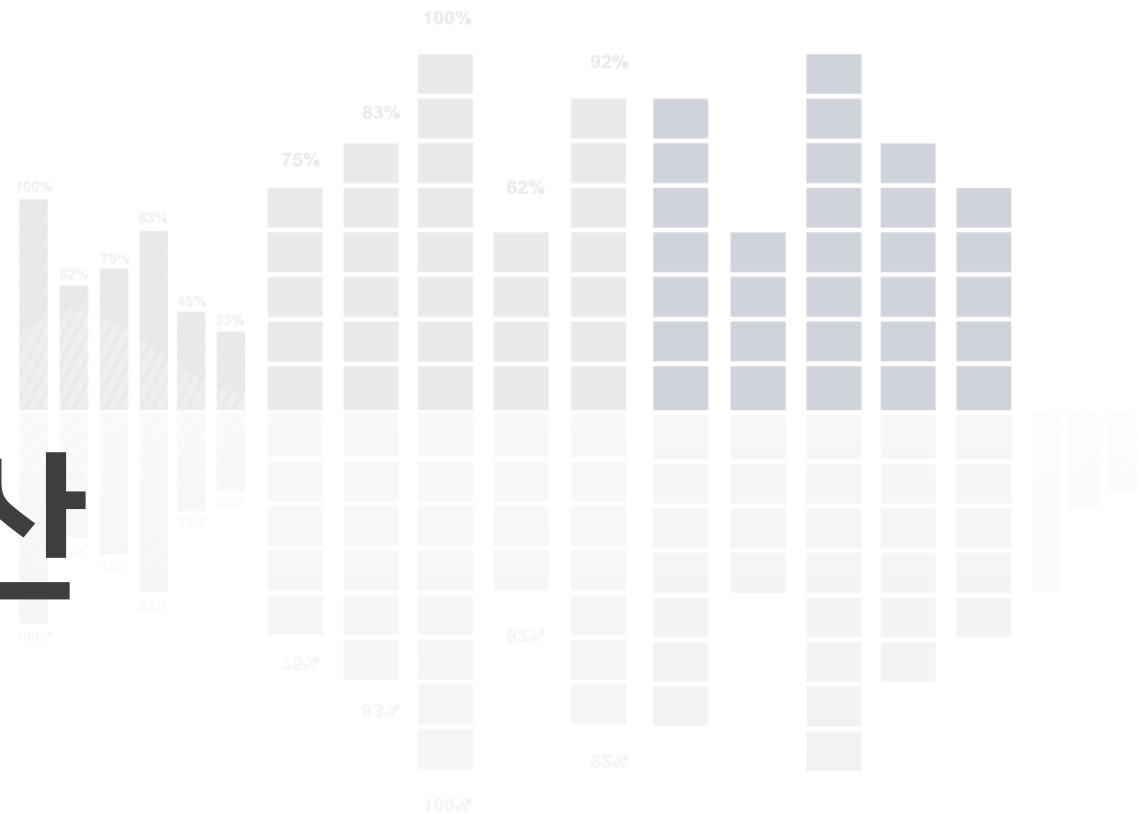
	적용 품목	스캐너 데이터의 비중	주요 결과
벨기에	- 슈퍼마켓 판매 대부분 품목	<ul style="list-style-type: none"> - 3대 슈퍼마켓 체인점(시장점유율 75%) - 스캐너 데이터가 COICOP 5단계 9개 품목군에 이용 	<ul style="list-style-type: none"> - 스캐너 데이터와 서베이 방식의 물가정보는 COICOP 5단계에서 결합 - 2개월에 걸쳐 시장점유율이 일정한 기준점 이하인 경우 제외됨 - 표본 제외 제품과 신규 진입 제품을 재점검
뉴질랜드	- 전자제품(가전제품)	- 품목별로 90%~20% (CPI 분기마다 공표)	<ul style="list-style-type: none"> - 가전제품 품목은 RWGEKS지수 산출 - 스캐너 데이터를 시장에서 구매
스웨덴	<ul style="list-style-type: none"> - 식료품, 맥주, 담배, 세제 및 개인 위생품 등 생필품 - 800개 대표 제품 - 90개의 제품군 선정 	<ul style="list-style-type: none"> - 3개 소매점 80개 체인점 - CPI 전체의 약14% 	<ul style="list-style-type: none"> - 매년 가중치가 변경되는 연쇄지수 기법을 적용 - CPI 전체적으로 매년 30% 정도의 샘플링이 교체

스캐너 데이터 활용방안 연구중인 국가

	적용 품목	스캐너 데이터의 비중	스캐너 자료 활용 방식/지수 구축 방식 등	주요 결과
영국	치약 및 샴푸	- 해당 품목	<ul style="list-style-type: none"> - 다양한 클리닝 필터(극단적 변동률 및 판매액이 미미한 품목 제외 등)를 적용 - 스캐너 데이터를 이용하여 제본스지수 산정 시 필터별로 9~45%의 자료가 제외됨 	<ul style="list-style-type: none"> - 클리닝 필터전 제본스지수가 기존 품목지수보다 20~30% 정도 낮아짐 - 제본스지수 보다는 톱크비스트 및 라스파이레스 지수가 클리닝필터 전후로 보다 안정적
프랑스	커피, 샐러드오일, 쌀, 요거트, 달걀, 초코렛바, 과일주스 및 치즈	- 해당 품목	<ul style="list-style-type: none"> - 6개의 슈퍼마켓 체인 10개 품목군 - 일년간 평균 대체율은 45%판매량에 비례한 무작위 샘플링으로 소비 바스켓을 줄였음 	<ul style="list-style-type: none"> - 서베이 및 스캐너 데이터의 연단위 연쇄 라스파이레스 지수를 산출한 결과 두 지수간 차이가 거의 없었음 - 스캐너 데이터 27%를 포함시켜 작성한 혼합 품목군 지수 또한 기존의 서베이 지수와 큰 차이가 없었음
일본	<ul style="list-style-type: none"> - 300개 슈퍼마켓 모든 상품 - 200개 대분류 - 1800개 소분류 	- 해당 품목	<ul style="list-style-type: none"> - 각 점포별로 대분류 200개 상품의 가격 변동을 가중 평균치 산출 - 대분류 상품의 가격 변동을 평균치 산출 - 종합지수가격 변동률 평균치 산출 - 판매액 가중치 고려하는 톱크비스트 방식을 적용 - 일별물가지수 	<ul style="list-style-type: none"> - 도쿄대에서 작성 및 공표 - 공식적인 CPI에 비해서 변동성이 더 크고 물가상승률은 약 1%p 낮은 것으로 나타났음 - 두 시점(이틀) 모두에 판매기록이 나타나지 않으면 샘플에서 제외

03

스캐너데이터를 이용한 우리나라 소비자물가지수 시산



1. 데이터 개요

◆ 스캐너 데이터의 입수 및 가공 프로세스

- ① 2013년부터 2016년까지의 4년에 걸친 분석 대상 데이터를 입수
'13년 1,262만건, '14년 1,235만건, '15년 1,295만건,
'16년 2,001만건
- ② 대한상공회의소(대한상의) 데이터 코드와 통계청 품목 코드의 매칭
- ③ 이상치의 탐색 및 제거

1. 데이터 개요

◆ 스캐너 데이터의 속성

- 대한상공회의소가 POS시스템을 운영하고 있는 국내 유통업체들로부터 상품 의 판매데이터를 수집 분석
- 기간: 2013.1월부터 2016.12월 (주별, 업태별, 지역별 취합된 물량 및 가격-월별 전환)

1. 데이터 개요

◆ 스캐너 데이터의 속성

- 품목 : 총 81개 품목
- 상품 수 (개별 상품 x 업태별 조합 수)
 - 2013년 : 87,684개, 2014년 : 86,616개,
2015년 : 103,276개, 2016년 : 112,926개

1. 데이터 개요

◆ 스캐너 데이터의 속성

- 총 분석 대상 상품 중 이상치 제거 : 단가(가격/물량) 기준 상, 하위 25% 분위수 초과 또는 하회 상품 제외
- 해당 상품 제외 후 (개별 상품 x 업태별 조합 수)
2013년 : 83,207개, 2014년 : 81,855개, 2015년 : 97,869개, 2016년 : 107,670개
- 분석 대상 기간 내 일관 된 상품 선정 방법도 고려
: 2013년부터 2016년까지 4년 간 상존 상품만 추출
- 4년 지속 상품 수 : 45,516개

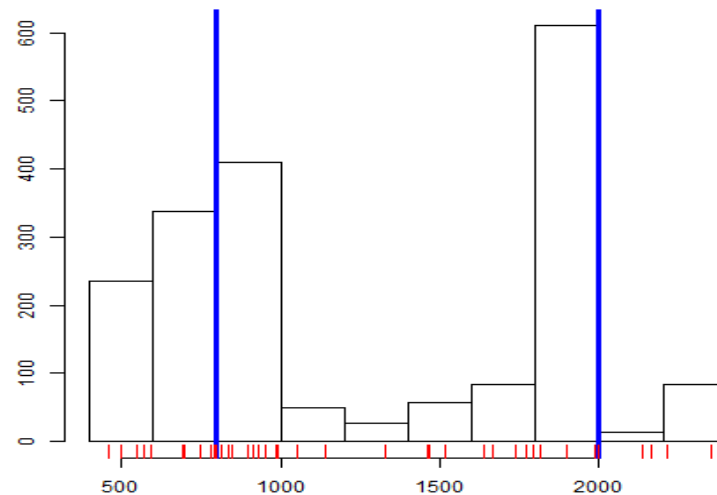
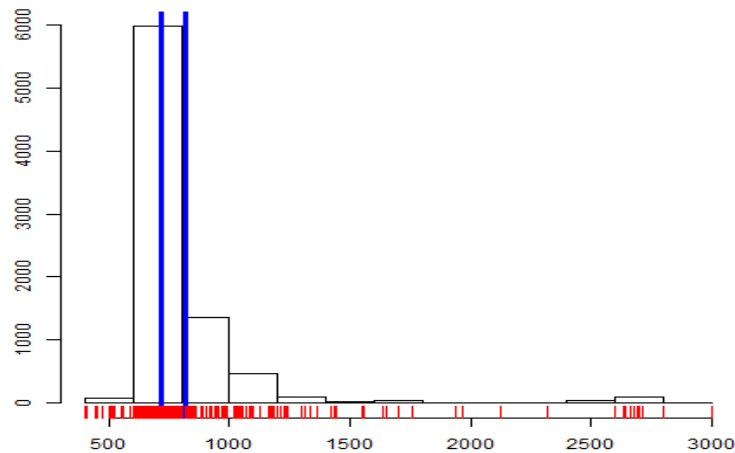
1. 데이터 개요

◆ 대한상의 코드와 통계청 코드의 매칭

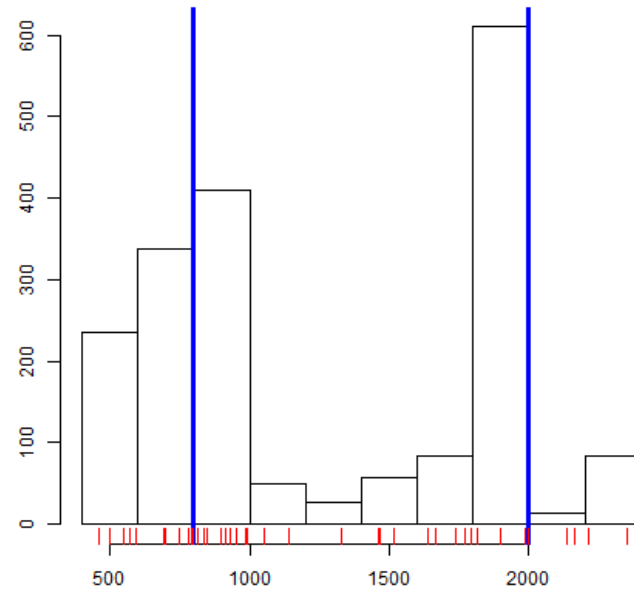
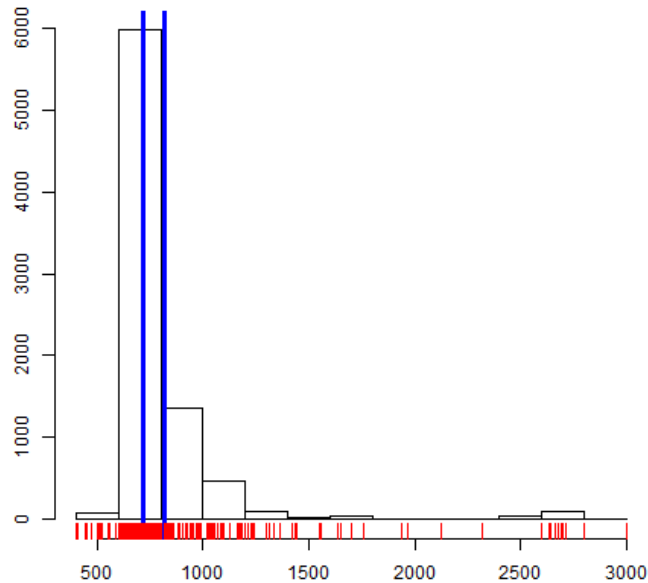
- 대한상의 스캐너 데이터의 KAN 코드를 통계청 품목 코드로 대응 후 통계청 기준에 품목(PUMID)별 가격지수를 계산
- KAN : PUMID 대응 : 대부분 일대일 대응 관계를 이루지만 일부 코드 사이에 다대일, 일대다 대응이 존재
 - 다대일 대응의 경우에는 다수의 KAN 품목을 하나의 PUMID 품목으로 간주하여 해결
- 일대다 대응 : 세부 상품들을 수작업으로 여러 코드에 나누어 대응
세부 상품명에 속한 키워드를 이용해 자동으로 분류 방법을 이용

2. 이상치 제거

- ◆ 단가 기준 이상치 제거를 위해 절사한 표본을 사용
 - 좌측 그림 : 이상치가 적절히 제거, 우측 그림 : bimodal 분포
- ◆ 이상치 제거 이후에도 bimodal 분포를 보이는 상품들은 분석에서 제외



2. 이상치 제거



2. 이상치 제거

◆ 이상치 제거를 위한 세 가지 방법

- ① 개별 세부 상품의 단가(가격/물량)를 계산하고 4 개년에 걸친 단가 자료의 표준편차를 계산하여 2배의 표준편차를 초과하는 단가를 이상치로 정의
- ② 개별 세부 상품의 단가(가격/물량)를 계산하고 4 개년에 걸친 단가 자료의 왜도(Skewness)를 계산하여 일정 기준을 초과하는 단가를 이상치로 정의
- ③ 개별 세부 상품의 단가(가격/물량)를 계산하고 4 개년에 걸친 단가 자료의 상위 25%를 초과하거나 하위 25%를 하회하는 단가를 이상치로 정의

3. 지수의 산출

◆ 라스파이레스(Laspeyres) 지수 (가공식품 품목 상위지수 산출 시 사용)

- 기준시점고정 가중산술평균법

$$\sum_{i=1}^N \left[\left(\frac{p_i^t}{p_i^0} \right) s_i^0 \right]$$

p_i^0 : unit price of product i at time 0

p_i^t : unit price of product i at time t

q_i^0 : sales quantity of product i at time 0

q_i^t : sales quantity of product i at time t

N: # of all products,

$$s_i^0 = (p_i^0 \times q_i^0) / \sum_{i=1}^N (p_i^0 \times q_i^0)$$

$$s_i^t = (p_i^t \times q_i^t) / \sum_{i=1}^N (p_i^t \times q_i^t)$$

3. 지수의 산출

◆ 통크비스트(Törnqvist) 지수

- 기준년도 및 비교년도 지출 비중(s) 고려

$$\prod_{i=1}^N \left(\frac{p_i^t}{p_i^0} \right)^{\frac{s_i^0 + s_i^t}{2}}$$

p_i^0 : unit price of product i at time 0

p_i^t : unit price of product i at time t

q_i^0 : sales quantity of product i at time 0

q_i^t : sales quantity of product i at time t

N: # of all products,

$$s_i^0 = (p_i^0 \times q_i^0) / \sum_{i=1}^N (p_i^0 \times q_i^0)$$

$$s_i^t = (p_i^t \times q_i^t) / \sum_{i=1}^N (p_i^t \times q_i^t)$$

3. 지수의 산출

◆ 제본스(Jevons) 지수

- 기준년도 및 비교년도 지출 균등 배분

$$\prod_{i=1}^N \left(\frac{p_i^t}{p_i^0} \right)^{\frac{1}{N}}$$

p_i^0 : unit price of product i at time 0

p_i^t : unit price of product i at time t

q_i^0 : sales quantity of product i at time 0

q_i^t : sales quantity of product i at time t

N: # of all products,

3. 지수의 산출

◆ 연쇄지수의 활용 (=전기지수 x 당기 등락률)

- 스캐너데이터 : 제품 조합이 시점에 따라 변동

$$I_t = \prod_{i=1}^N \left(\frac{p_i^t}{p_i^0} \right)^{\frac{s_i^0 + s_i^t}{2}}$$

$$P_{0,t} = P_{0,t-1} \times I_t$$

p_i^0 : unit price of product i at time 0
 p_i^t : unit price of product i at time t
 q_i^0 : sales quantity of product i at time 0
 q_i^t : sales quantity of product i at time t
 N: # of all products,

$$s_i^0 = (p_i^0 \times q_i^0) / \sum_{i=1}^N (p_i^0 \times q_i^0)$$

$$s_i^t = (p_i^t \times q_i^t) / \sum_{i=1}^N (p_i^t \times q_i^t)$$

3. 지수의 산출

◆ GEKS 지수

- 비교 가능한 모든 두 시점 대상으로 지수를 산출한 뒤 접속시점 기준 기하평균 (T+1) 시점 추가 시 과거 자료 수정이 필요

$$P_{0,T} = \prod_{l=0}^T (P_{0,l} \times P_{l,T})^{\frac{1}{T+1}}$$

$$P^{0,T(0,T+1)} = \prod_{l=0}^{T+1} (P_{0,l} \times P_{l,T})^{\frac{1}{T+2}}$$

3. 지수의 산출

◆ RWGEKS 지수

- 지수 수정의 문제를 피하기 위해 윈도우 크기를 (T+1)로 고정하여 이동시키며 산출

$$P_{0,T} = \prod_{l=0}^T (P_{0,l} \times P_{l,T})^{\frac{1}{T+1}} \times \prod_{j=0}^{T+1} (P_{T,j} \times P_{j,T+1})^{\frac{1}{T+1}}$$

3. 지수의 산출

◆ 품질조정 지수(시간-더미 헤도닉 기법)

- 상품의 품질을 결정하는 특성치들과 시간더미를 설명변수로 설정

$$\ln p_i^t = \alpha + \delta_t D^t + \sum_{k=1}^K \beta_k Z_{ik} + \frac{\sigma}{\sqrt{s_i^t}} e_i^t \quad P_{0,t} = \exp(\hat{\delta}_t)$$

3. 지수의 산출

◆ 품질조정지수(시간-상품-더미 헤도닉 기법)

- 모든 제품에 동시에 존재하는 고정시간효과(fixed-time effect)가 품질변동과 무관한 순수한 가격변동이라는 가정

$$\ln p_i^t = \alpha + \delta_t D^t + \sum_{i=1}^{N-1} \gamma_i D_i + \frac{\sigma}{\sqrt{s_i^t}} e_i^t$$

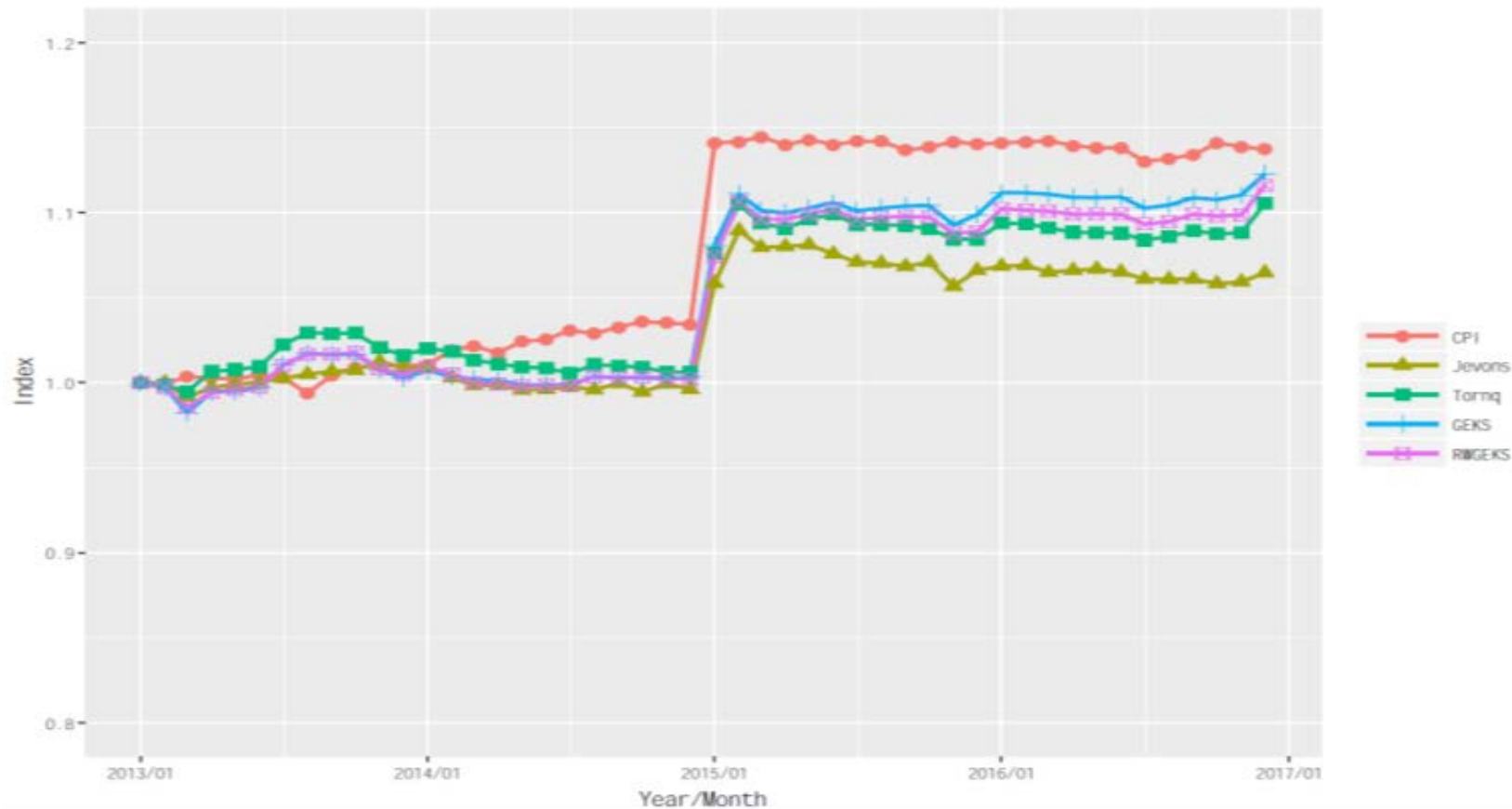
$$\ln \hat{p}_i^t = \hat{\alpha} + \hat{\gamma}_i \quad \ln \hat{p}_i^t = \hat{\alpha} + \hat{\delta}_t + \hat{\gamma}_i$$

3. 지수의 산출

- ◆ 스캐너데이터는 소매점에서 거래되고 있는 모든상품에 대한거래
가격 및 거래량 정보,고빈도(시간별, 일별)정보, 신상품에 대한정보를 보유
→ 라스파이레스 지수보다 현실반영도가 높은 톨크비스트(Törnqvist)
지수 또는 제본스(Jevons) 지수를 작성 (연쇄지수)
- ◆ 연쇄지수를 작성함에 따라 발생 가능한 연쇄편의의 문제를 해결하기
위해 GEKS 및 RWGEKS 지수도 함께 작성

4. 시산 결과

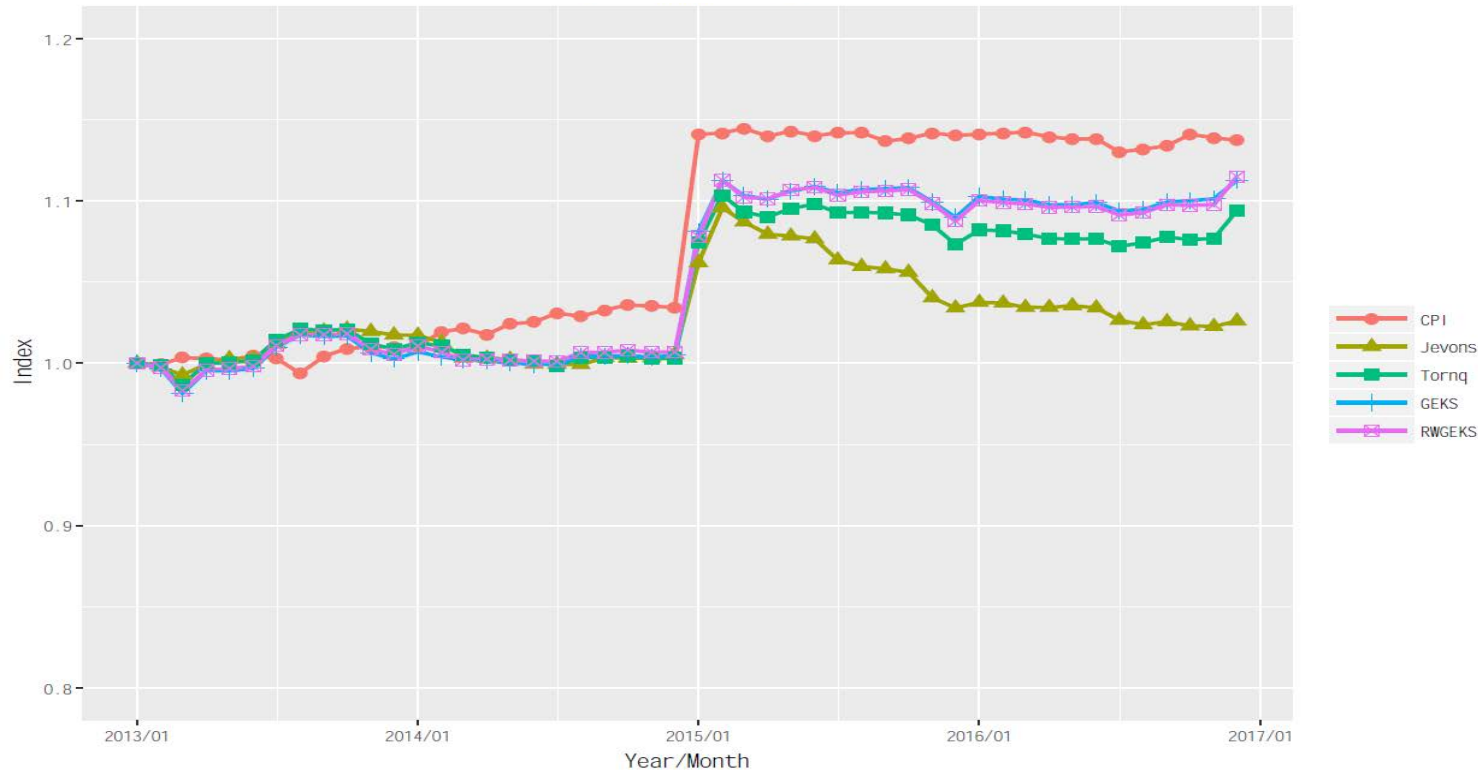
◆ 대한상의 스캐너 데이터 총지수 (원데이터, 업태별)



4. 시산 결과

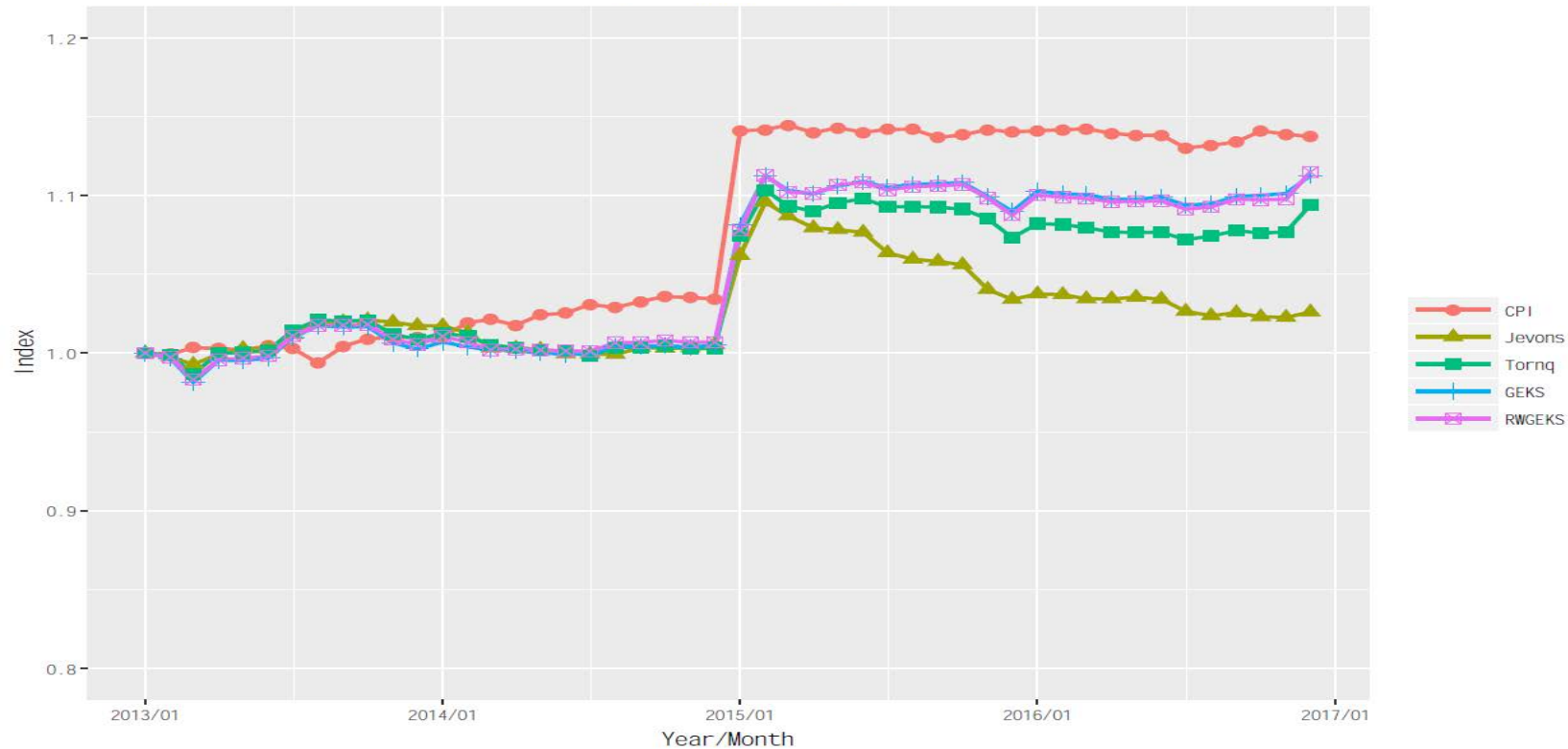
◆ 대한상의 스캐너 데이터 총지수(원데이터,지역별)

- Jevons 지수 : 표본이 매우 작은 지역의 가격 이상치에 기인



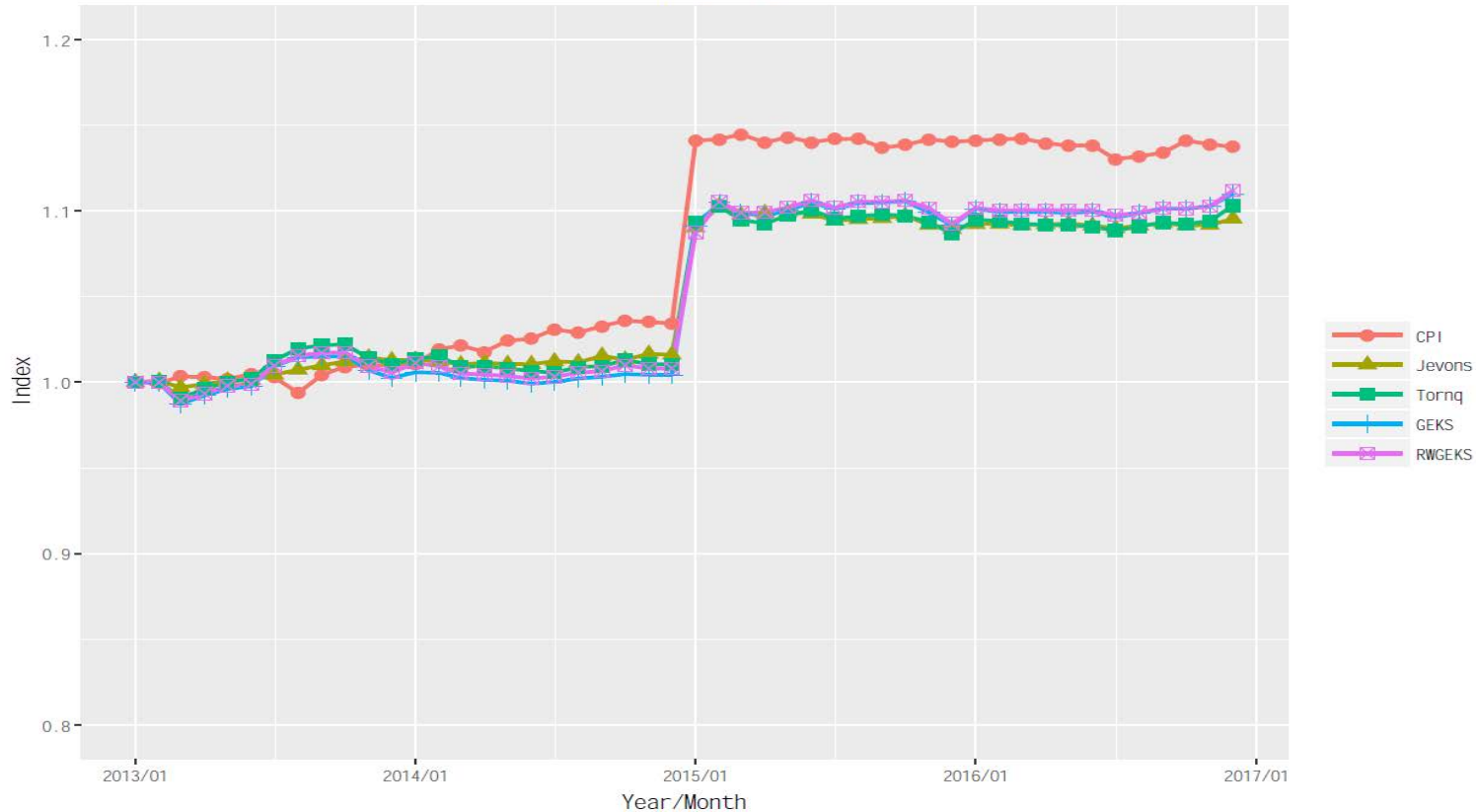
4. 시산 결과

◆ 대한상의 총지수 (원데이터, 4년 지속 상품 업태별)



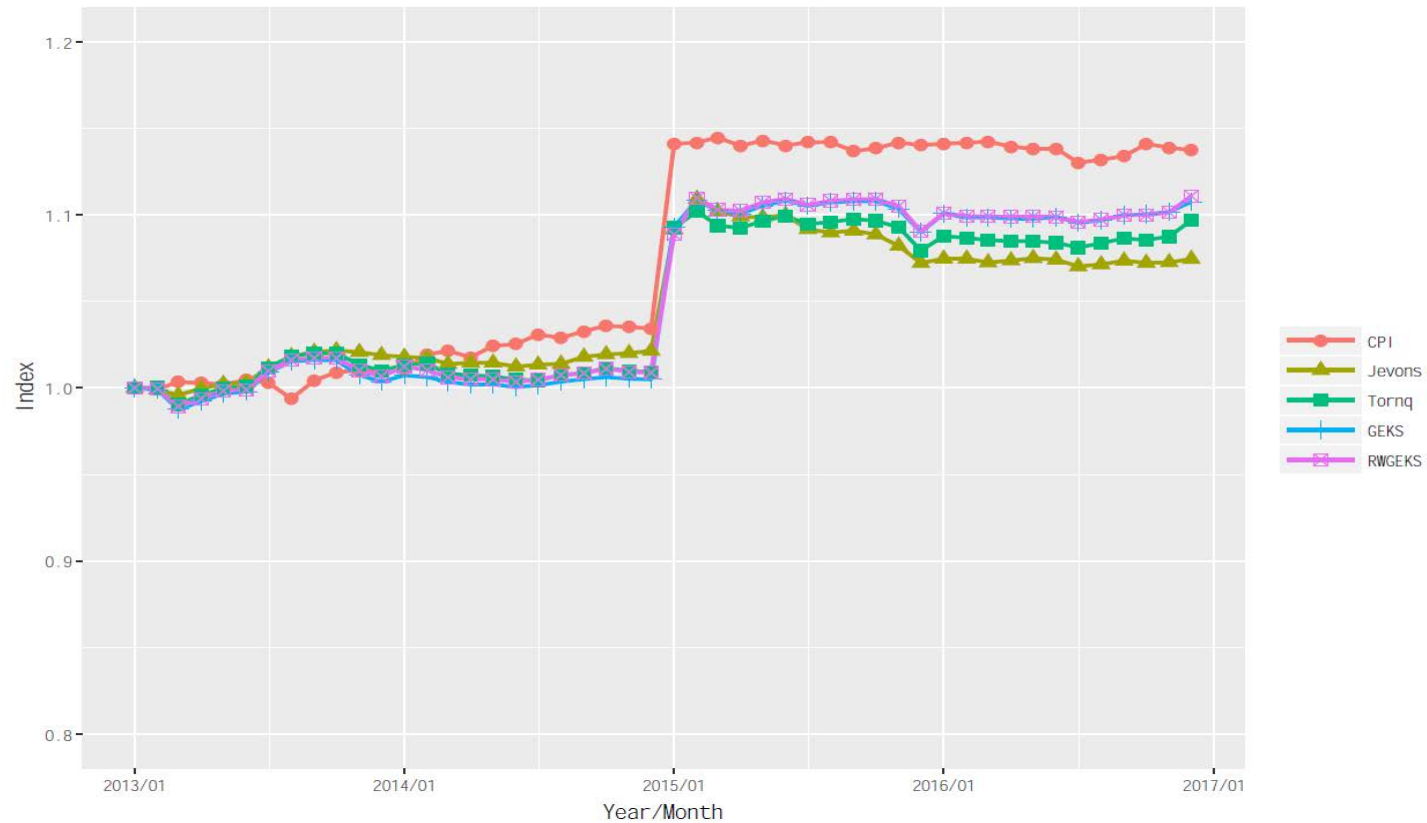
4. 시산 결과

◆ 대한상의 총지수 (이상치 제거, 업태별)



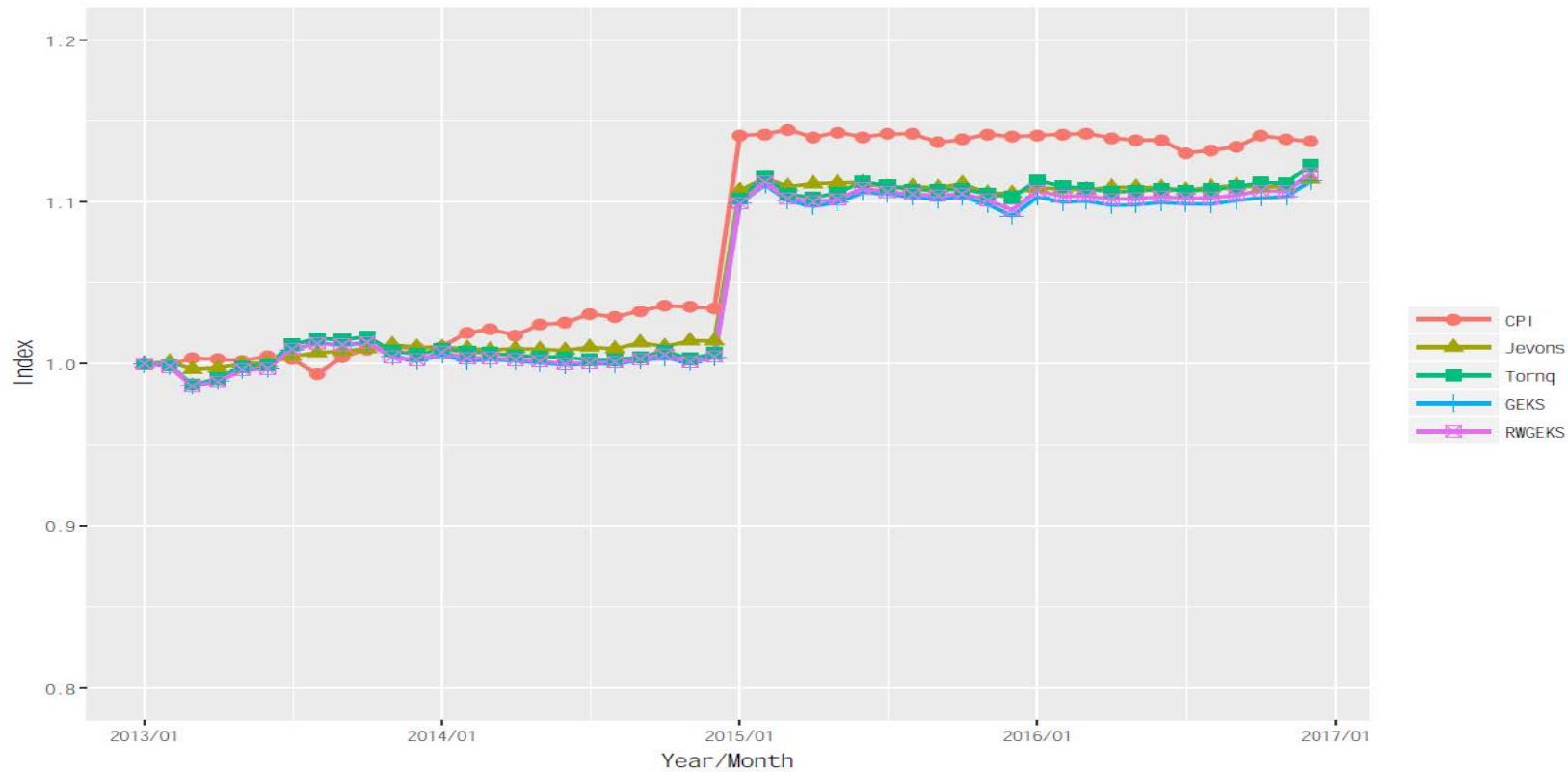
4. 시산 결과

◆ 대한상의 총지수 (이상치 제거, 지역별)



4. 시산 결과

◆ 대한상의 총지수 (이상치 제거, 4년 지속 상품의 업태별)



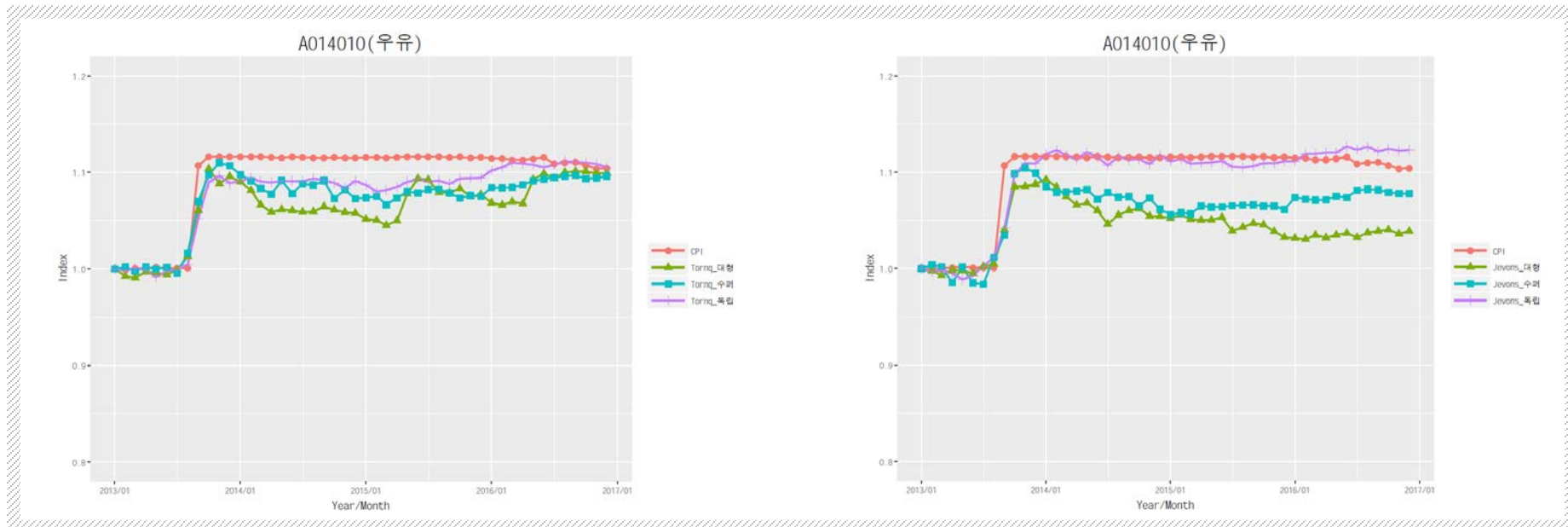
4. 시산 결과

◆ CPI 전기대비 증감률과의 상관계수

	제본스	통크비스트	GEKS	RWGEKS
원지수 (지역별 합산)	0.745	0.791	0.800	0.766
원지수 (업태별 합산)	0.796	0.800	0.806	0.785
이상치 제거지수 (지역별 합산)	0.894	0.897	0.901	0.874
이상치 제거지수 (업태별 합산)	0.937	0.902	0.902	0.881
4년 지속 상품 지수 (업태별 합산)	0.856	0.798	0.822	0.812
4년 지속 상품, 이상치 제거 지수 (업태별 합산)	0.951	0.905	0.906	0.902

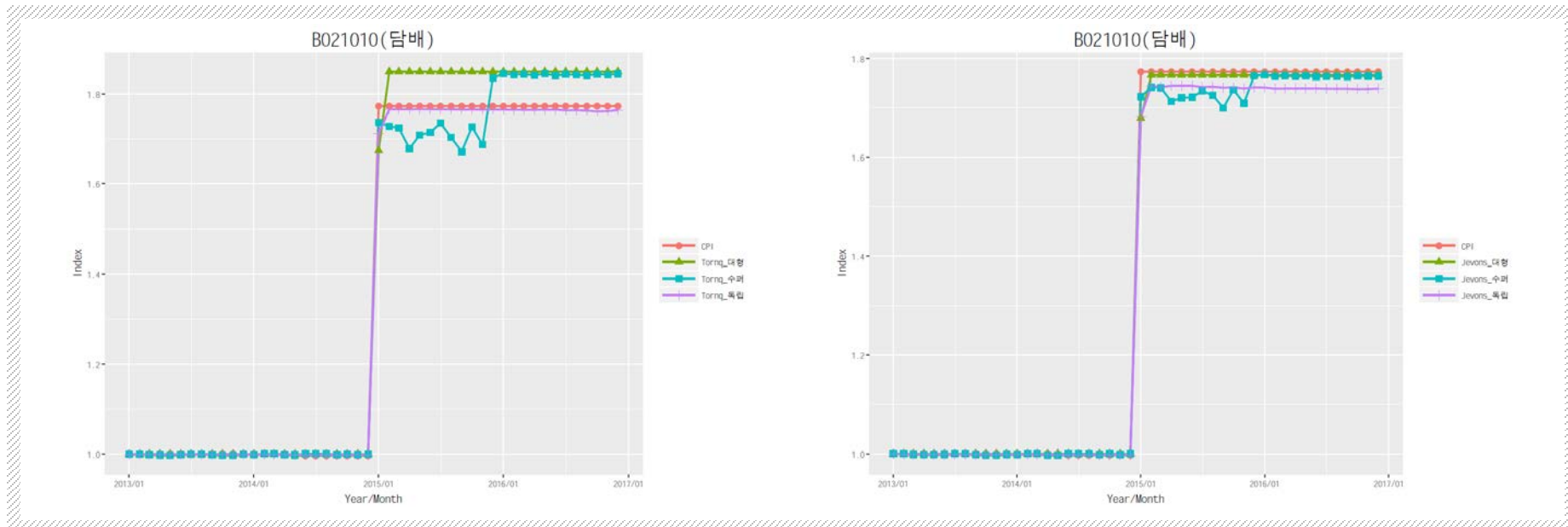
4. 시산 결과(품목예시)

- ◆ 우유 - 유통에 따른 가격 차이가 반영된 사례로 독립수퍼의 Jevons 지수가 CPI와 가장 유사 (좌측 톱크비스트, 우측 제본스)



4. 시산 결과(품목예시)

- ◆ 담배 (좌측 톰크비스트, 우측 제본스) - 업태별 주력상품 차이 (대형마트에서는 수입담배를 판매하지 않음) 및 데이터 신뢰도 문제(수퍼체인의 부정확한 가격)



4. 시산 결과

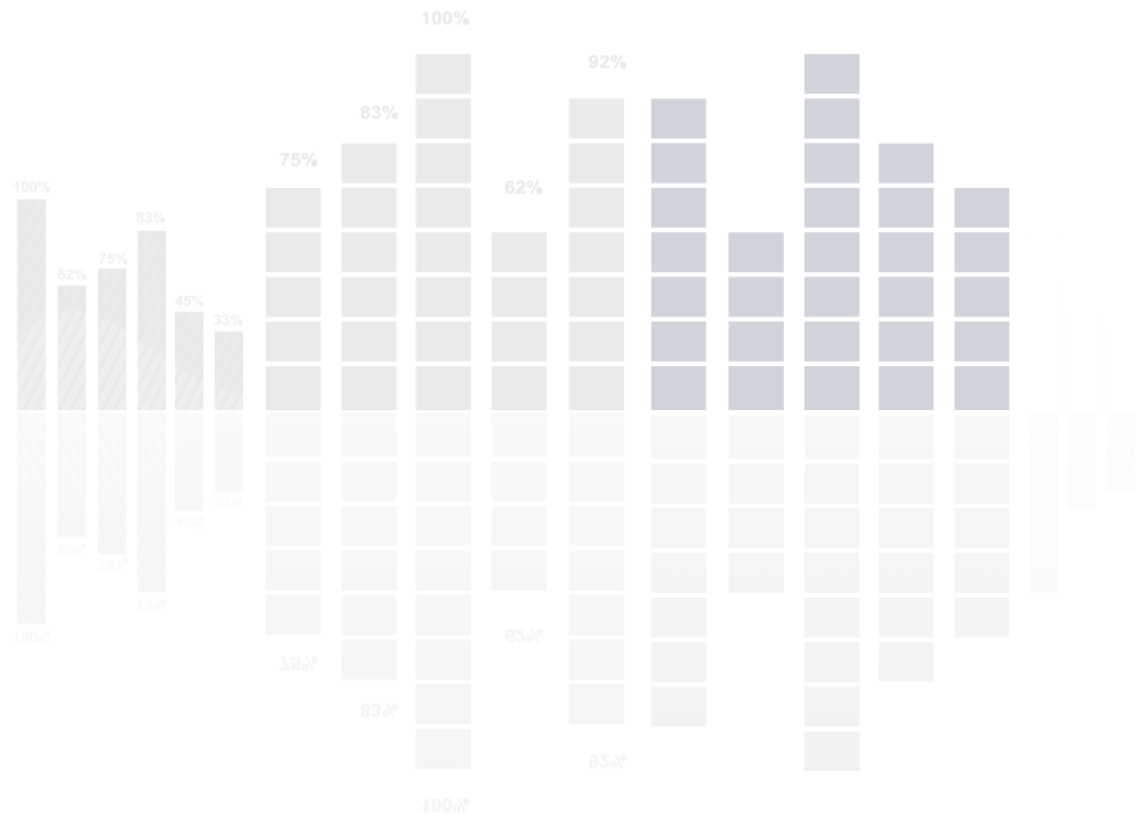
		제본스	통크비스트, GEKS, RWGEKS
원자료	업태별 합산	다소 격차	CPI와 유사하나 GEKS지수의 수준이 CPI에 가장 근접
	지역별 합산	점차 격 차 확대	CPI와 유사하나 GEKS와 RWGEKS 지수가 거의 같아짐
	4년 지속 상품 지수(업태별합산)	모든 지수가 CPI와 매우 근접	
이상치 제거	업태별 합산	모든 지수가 서로 매우 유사한 행태를 보임. CPI와 다소 격차는 보이나 행태는 CPI와 유사	
	지역별 합산	업태별 합산의 경우에 비해 지수 간 격차와 CPI와의 격차 모두 확대	
	4년 지속 상품 지수(업태별합산)	모든 지수가 CPI와 매우 근접	

5. 시사점

- ◆ 양질의 데이터의 확보가 긴요
- ◆ 데이터 정제 과정의 중요성
 - 이상치 제거, 4년 지속 상품 제한 시 CPI와 가장 유사한 패턴
- ◆ 지수의 선택
 - 물량 정보를 반영한 톱크비스트, GEKS 및 RWGEKS 지수 vs 제본스 지수
- ◆ 지표의 활용
 - CPI를 대체 vs 보조지표 활용 (ex 생활물가장바구니 지수)
 - (CPI: 대표 가격, 스캐너 데이터: 실제 시장 거래 가격)
- ◆ 기준년 개편처럼 고정 상품 선정 시기 및 기준 설정이 필요

04

스캐너데이터 데이터 확보 방안



1. 해외 사례

◆ 해외 주요국 스캐너 데이터 자료 접근 근거

국가	CPI에 스캐너 자료 활용 여부	스캐너 자료 접근 및 확보 근거
영국	활용방안 연구중	안정적인 확보 방안에 대해서 검토중
스웨덴	실제 활용중	유통산업과의 자발적 업무협약
프랑스	활용방안 연구중	시장조사기관으로부터 구매
스위스	실제 활용중	통계법에 따라 무료로 접근 가능
네덜란드	실제 활용중	통계법에 따라 슈퍼마켓 체인점이 직접 통계청에 제공
뉴질랜드	실제 활용중	시장조사기관인 GfK로부터 3년 계약 등 통해 구입
노르웨이	실제 활용중	통계법에 따라 무료로 접근 가능
벨기에	실제 활용중	통계법에 따라 FTP 방식으로 매주 자료를 취합

2. 스캐너 데이터 확보 관련 검토 사항

- ◆ 스캐너 데이터를 실제 CPI에 활용하는 나라의 경우 대부분 통계법 또는 자발적인 업무협약 등을 통해 무료로 확보

- ◆ 원 스캐너 데이터의 확보

- ◆ 소규모 인구 중심의 북유럽 vs. 인구가 많은 일본/영국/프랑스

3. 스캐너 데이터 향후 확보 방향

◆ 빅데이터 수집, 제공자(데이터 사업자)와 통계 작성자간 협의 또는 제도를 통해 다음 방안을 고려해 볼 수 있음

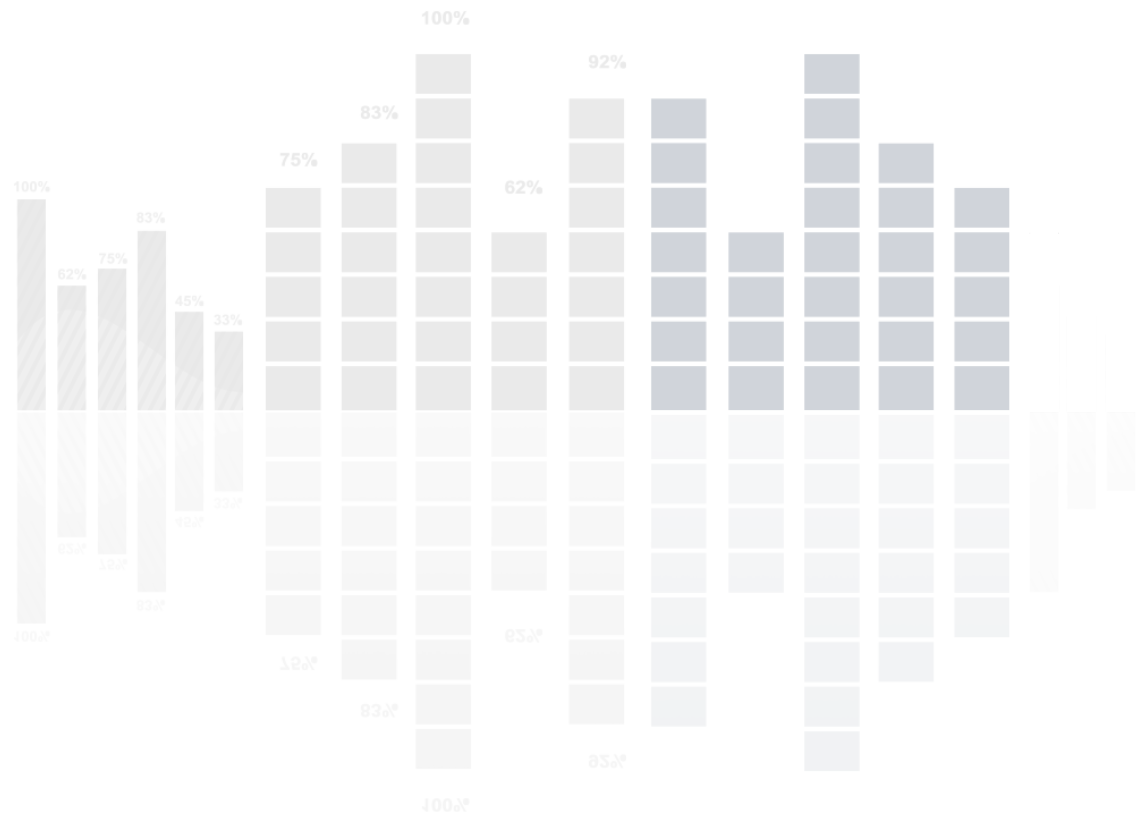
- 제1안 : 데이터 사업자(시장조사기관)의 데이터를 구매
- 제2안 : 통계법에 기초하여 확보
- 제3안 : 데이터 사업자가 공적 통계로 승인 받은 후 작성, 공표하고
통계작성자는 이를 받아서 국가통계 작성에 이용
- 제4안 : 데이터산업육성법 등을 통해 데이터 산업을 육성하고 이 과정에서
국가가 기초 데이터를 확보

3. 스캐너 데이터 향후 확보 방향

- ◆ 우리나라의 경우 시장점유율이 높은 대형 마트 3사의 스캐너 데이터를 집중시키고 통계청은 이 기관과 업무협약을 통해 가공되지 않은 스캐너 데이터를 점진적으로 확보할 필요
- 스위스의 경우 자발적인 업무협약으로 시작했으나 이후 소매 체인점이 데이터 제공에 소극적이라 통계법적인 근거를 적용

05

향후 방향



1.향후과제

◆스캐너 데이터의 품질 제고

- 포괄범위 확대, 상품 코드 개선, 상품목록체계 정교화와 정제과정의 자동화·정교화, 품목 확대 등

◆소비자물가지수로 단계적, 중기적 활용

- 품목별 대표 상품 선정, 신규 상품 파악, 실제 거래 가격 파악, 무응답 대체, 조사 횟수 증가, 가중치 조정 등
 - 신뢰를 얻게 된다면 일부 품목의 가격조사는 표본조사가 아닌 스캐너 데이터로 대체

1.향후과제

◆ 스캐너 데이터를 이용한 지수 작성 : 별도 기관 작성

- 통계청에서 작성하는 경우 이용자 혼란
- 가격지수와 물량지수로 나누어 작성
- 승인 통계로 관리하고 기초자료 입수

◆ 빅데이터 활용 환경 마련

- 제도적 노력, 인력 확보
 - 중기적으로 표본조사, 행정자료 중심에서 표본조사+빅데이터(행정 자료 포함)로 전환 필요



감사합니다.