

# 소비자물가지수 생산을 위한 스캐너 데이터 활용방안 연구

—  
2017. 11





통계청장 귀하

본 연구결과 보고서를 “소비자 물가지수 생산을 위한 스캐너 데이터 활용방안 연구”의 최종 연구보고서로 제출합니다.

2017년 11월 24일

주관연구기관 : 한국통계학회



연구 책임자 : 이금희

공동 연구자 : 장영재

손종철

연구 보조원 : 강정일

하혜진





# 요 약 문

1990년대부터 유통업체들이 판매시점 정보관리 시스템(Point of Sales, POS 시스템)을 도입하여 바코드를 이용한 상품거래를 하면서 개별 거래들의 내역인 고객별 거래 데이터를 축적하고 있다. 이와 같이 바코드를 이용한 상품 거래 데이터를 스캐너 데이터(scanner data)라고 한다. 여기서 바코드는 GTIN(Global Trade Item Number) 또는 EAN(European Article Number)를 따라 작성되는데 이 코드는 국내 또는 국외로 유통되는 상품을 식별하는 국제표준상품 코드이다.

대한상공회의소 유통물류진흥원이 다양한 유통사로부터 스캐너 데이터를 수집, 축적하고 있다. 스캐너 데이터는 “각 유통매장의 일 마감 프로세스 이후 정리된 각 상품별 판매수량, 판매총액, 판매단가의 형태”를 띠고 있다. 이번 연구에서는 [그림 1]과 같이 대한상공회의소가 제공한 유통매장별 일마감 데이터를 주단위로 통합한 스캐너 데이터를 이용하였다.

[그림 1] 대한상공회의소 통계청 제공 스캐너 데이터

출처	일 마감 일자	판매수량	판매총액	판매단가
대한상림,매출단가_20170929.zip	2017-10-12 오전 10:48			
대한상림,매출단가_20171001.xls	2017-10-19 오후 10:30			
대한상림,매출단가_20170904.xls	2017-10-12 오전 10:44			
대한상림,매출단가_20170915.xls	2017-10-12 오전 10:40			
대한상림,매출단가_20170918.xls	2017-10-12 오전 10:37			
대한상림,매출단가_20170903.xls	2017-09-29 오후 11:34			
대한상림,매출단가_20170928.xls	2017-09-27 오전 10:37			
대한상림,매출단가_20170827.xls	2017-09-07 오후 9:34			
대한상림,매출단가_20170804.xls	2017-09-07 오전 9:31			
대한상림,매출단가_20170818.xls	2017-09-07 오전 9:27			
대한상림,매출단가_20170806.xls	2017-09-07 오후 9:34			
대한상림,매출단가_20170701.xls	2017-08-09 오후 11:06			
대한상림,매출단가_20170707.xls	2017-08-09 오전 11:34			
대한상림,매출단가_20170723.xls	2017-08-09 오전 11:06			
대한상림,매출단가_20170716.xls	2017-08-09 오전 11:03			
대한상림,매출단가_20170709.xls	2017-08-09 오전 11:00			
대한상림,매출단가_20181226.xls	2017-08-09 오후 11:00			
대한상림,매출단가_20170625.xls	2017-07-20 오전 9:30			
대한상림,매출단가_20170702.xls	2017-07-20 오전 9:30			
대한상림,매출단가_20170618.xls	2017-07-20 오전 9:37			
대한상림,매출단가_20181204.xls	2017-07-20 오전 9:34			
대한상림,매출단가_20170606.xls	2017-07-19 오후 9:30			
대한상림,매출단가_20170613.xls	2017-07-19 오후 9:46			
대한상림,매출단가_20170626.xls	2017-07-19 오후 10:30			
대한상림,매출단가_20170628.xls	2017-07-06 오후 10:27			

스캐너 데이터를 이용하는 경우 가격 및 거래량 정보를 빠르게 고빈도로 파악할 수 있고, 새로운 상품의 발생에 대한 정보도 조기에 파악할 수 있어서 스캐너 데이터는 소비자물가 조사에서 나타나는 어려움과 고정 가중치의 라스파이레스 지수의 한계를 동시에 극복할 수 있는 유용한 정보가 된다. 그러나 현재 대한상공회의소로부터 제공받은 스캐너 데이터는 대형 유통사가 포함되지 않는 등 포괄범위에 제약이 있고 거래 데이터에 반품, 묶음판매 등이 포함되어 통계작성에 스캐너 데이터를 통계 작성에 조정 없이 이용하는 데에는 제약이 있을 수밖에 없다. 또한 통계청이 스캐너 데이터를 즉시적,

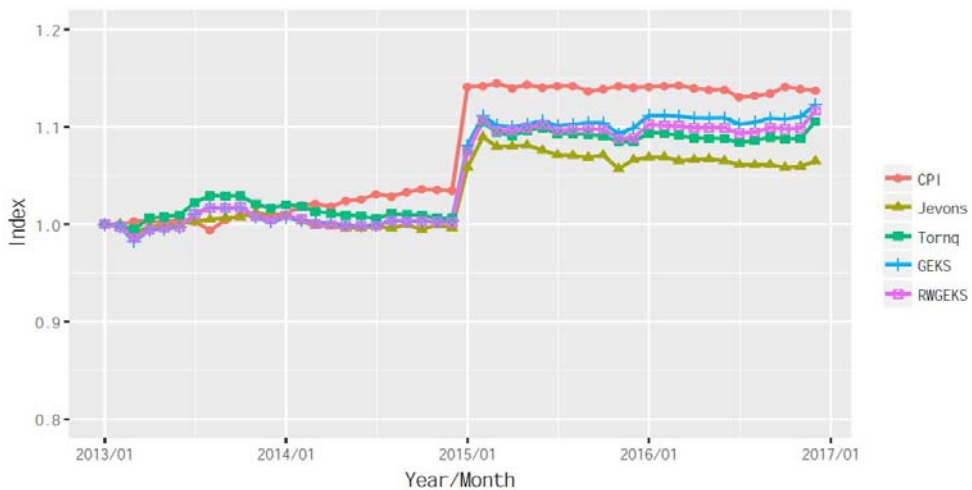
안정적으로 확보할 수 있는 제도적 환경이 마련되지 않아서 통계 작성 관련 지속성의 제약도 있다.

네덜란드, 스웨덴, 노르웨이, 스위스 등 유럽의 일부 국가에서 스캐너 데이터를 활용하여 일부 품목의 소비자물가지수를 작성하고 있다. 이들 국가에서 이용되고 있는 품목은 스캐너 데이터의 지속적 확보 가능성을 감안하여 음식료품, 주류, 에너지와 일부 가전제품이다. 또한 각국에서 스캐너 데이터를 이용하여 품목별 소비자물가지수를 작성할 때 각 나라의 환경에 맞추어 제본스(Jevons) 지수, 비가중 기하평균, 품질조정 단위가격지수, RWGEKS 지수 등을 이용하고 있다.

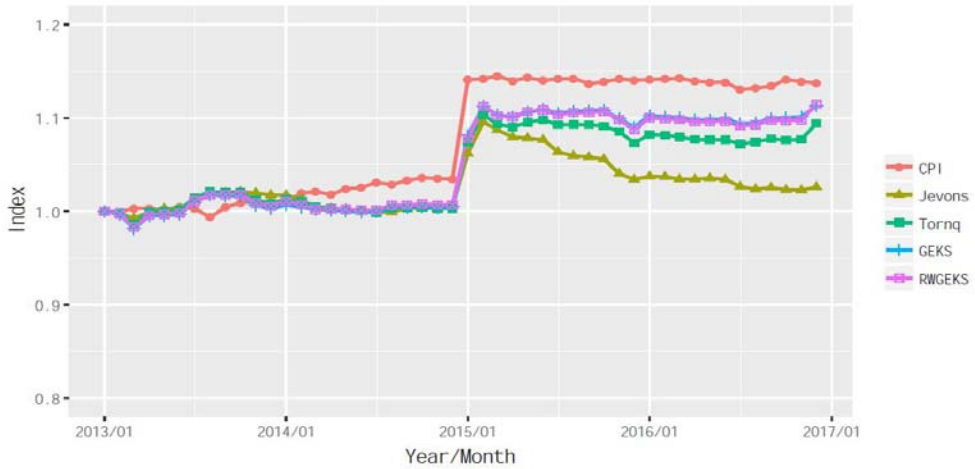
본 연구에서는 대한상의에서 제공한 주별 스캐너 데이터를 바탕으로 지수 산식을 달리하여 81개 품목별 월별 소비자물가지수를 시산하였다. 지수형태로는 업태별, 지역별 합산한 톨크비스트(Törnqvist), 제본스, GEKS, RWGEKS 지수를 이용하였으며 이상치 제거, 4년 간 지속 상품 등으로 상품 범위를 좁혀서도 각종 지수를 시산하였다. 스캐너 데이터를 이용하여 작성된 개별 품목지수에 통계청 소비자물가지수 작성시 이용되는 품목별 가중치를 적용하여 라스파이레스 식 ‘대한상의 스캐너 데이터 총지수’를 계산하였다.

[그림 2, 3, 4]은 2013년 1월을 기준치로 하여 대한상의 스캐너 데이터 총지수를 산식을 달리하여 상대적인 크기로 소비자물가지수와 같이 나타낸 것이다. 지수 작성 방법을 비교해 보면 물량 정보를 반영한 톨크비스트, GEKS 및 RWGEKS 지수가 단순 가중치를 반영한 제본스 지수에 비해 안정적이고 일관된 것으로 나타났다.

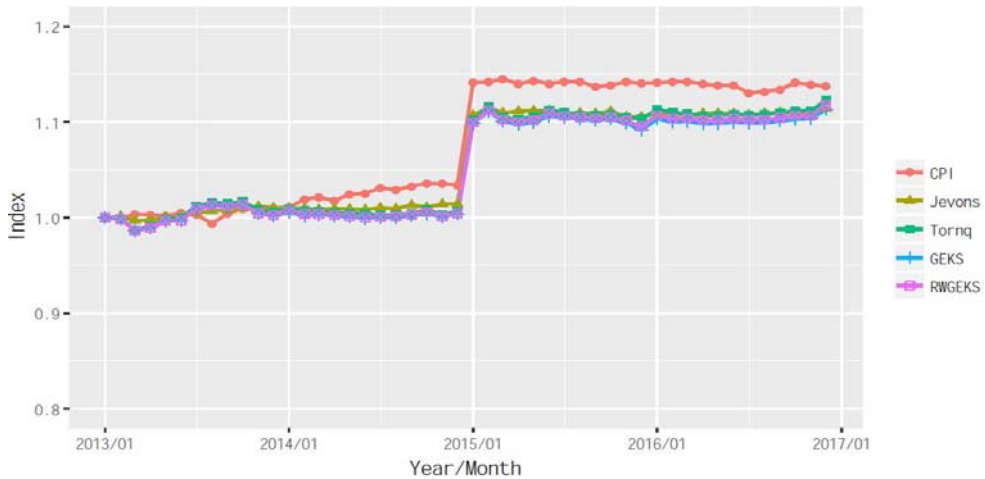
[그림 2] 대한상의 스캐너 데이터 총지수(원자료, 업태별 합산)



[그림 3] 대한상의 스캐너 데이터 총지수(원자료, 지역별 합산)



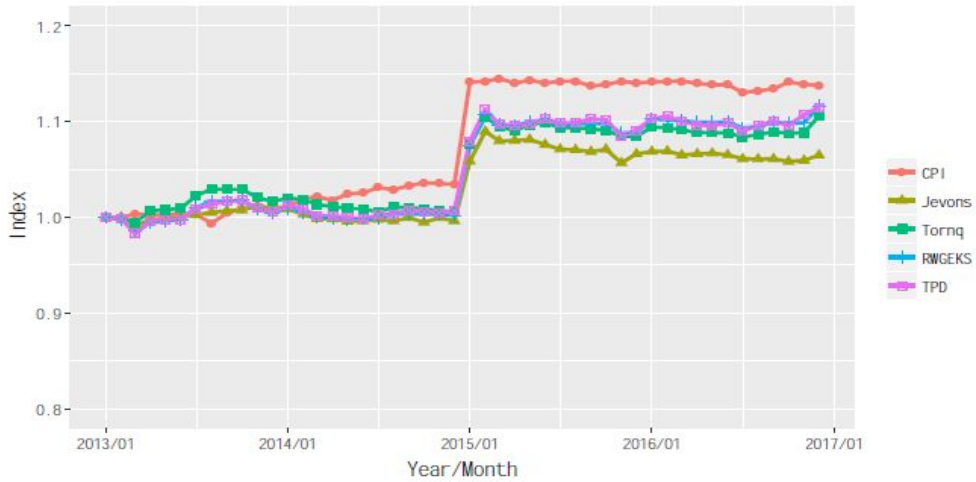
[그림 4] 대한상의 스캐너 데이터 총지수(이상치 제거, 4년 지속 상품, 업태별 합산)



[그림 5]에서 품질조정을 실시한 지수까지 산출하여 그 필요성을 검토해 보았는데 실제 하위 품목지수나 총지수의 경우 모두 톱크비스트 또는 RWGEKS와 매우 유사한 움직임을 보이면서 품질조정에 따른 추가적인 효과가 극명하게 드러나지 않는 것으로 나타났다.

<표 1>은 각 방식으로 지수를 산출한 결과를 요약한 표이다. 이를 보면 분석 대상 세부 상품을 4년 지속 상품으로 제한할 경우 모든 지수가 CPI와 매우 유사한 행태를 나타냈다. 또한 지역별 합산 방식보다는 업태별 합산 방식을 통해 산출된 지수가 보다 안정적인 것으로 나타났다. 지수의 행태와 실무적인 면을 함께 고려해보면 톱크비스트나 RWGEKS 지수로 품목 지수를 작성하는 것이 유용하다고 판단된다.

[그림 5] 대한상의 스캐너 데이터 총지수(원자료, 품질조정 포함)



<표 1> CPI와 각 지수의 비교

		제본스	통크비스트	GEKS	RWGEKS
원자료	업태별 합산	다소 격차	대체로 CPI와 유사하나 GEKS지수의 수준이 CPI에 가장 근접		
	지역별 합산	점차 격차 확대	대체로 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			

스캐너 데이터를 통계 작성기관이 이용하려면 품질이 우수한 스캐너 데이터를 안정적으로 확보해야 한다. <표 2>는 나라별 스캐너 데이터 입수방법을 정리한 것인데 이를 감안하여 우리나라 통계청 나름의 스캐너 데이터 입수 방법을 마련해야 한다. 우리나라도 다른 나라와 마찬가지로 통계법을 적용하거나, 구매, 업무협약 등을 통해 스캐너 데이터를 확보하는 방안을 정할 필요가 있다. 우리나라 통계법 상 가능한 방법으로는 스캐너 데이터를 이용

한 통계를 국가승인통계로 지정하고 동 통계를 관리하여 스캐너데이터를 얻는 것이다.

<표 2> 해외 주요국 스캐너 데이터 자료 접근 근거

국가	CPI 구축시 스캐너 자료 포함 여부	스캐너 자료 접근 및 확보 근거
영국	활용방안 연구중	향후 안정적인 확보 방안에 대해서 지속적인 검토를 해나갈 계획
프랑스	활용방안 연구중	시장 리서치 기관으로부터 구매
스웨덴	실제 활용중	유통산업과의 자발적 업무협약
노르웨이	실제 활용중	통계법에 따라 무료로 접근 가능
스위스	실제 활용중	통계법에 따라 무료로 접근 가능
네덜란드	실제 활용중	슈퍼마켓 체인점이 직접 통계청에 제공
뉴질랜드	실제 활용중	시장조사기관인 GfK로부터 매월 제공받음
벨기에	실제 활용중	FTP 방식으로 매주 자료를 취합

추가로 소비자물가지수 개편과 지수 작성방법와의 관계를 살펴보았다. 소비자물가지수는 각종 계약, 임금협상 등의 기초가 되므로 소비자물가지수의 잦은 개정은 이용자의 혼란을 초래한다. Lowe지수는 공리적 속성을 가지고 있으면서 잦은 소비자물가지수 개정을 줄이고 있고 많은 나라에서 이용하고 있으므로 2018년 소비자물가지수 개편시 기존 국가통계위원회 보고(통계청 2014)에서 제시한 바와 같이 Lowe지수로 소비자물가지수 중간년 기준년 개편을 실시할 필요가 있다. 다만 Lowe지수 월중첩 지수접속법을 적용하는 경우 특정 월(12월)의 가격을 이용하게 된다. 일부 품목 가격이 계절성을 가질 때 특정 월 가격 적용은 가중치의 대표성을 상대적으로 낮게 하므로 이를 처리하는 방안을 모색할 필요가 있다.

스캐너 데이터를 소비자물가지수 작성에 활용하기 위해서는 질 높은 스캐너 데이터 확보, 소비자물가지수 작성 과정에서 활용성 검토 등에 대한 중기적이면서 치밀한 검토가 필요한데 이를 구체적으로 살펴보면 다음과 같다. 첫째, 스캐너 데이터의 포괄범위 확대, 상품 코드 개선, 상품목록체계 정교화와 정제과정의 자동화·정교화 등을 통해 스캐너 데이터의 품질을 보다 높여야 한다. 둘째, 스캐너 데이터는 품목별(이번 연구 제공 대상 품목은 81개) 소비자물가지수 작성 과정에서 보조정보로 이용될 수 있다. 예를 들면 소비자물가지수조사에서 품목별 대표 상품 선정, 신규 상품 파악, 실제 거래 가격 파악, 무응답 대체, 조사 횟수 증가, 가중치 조정 등에서 보조정보로 이용할 수

있다. 그러나 스캐너 데이터와 기존 조사결과와의 차이 등을 면밀히 살펴보고 중기적으로 순차적으로 진행할 필요가 있다. 스캐너데이터 정보가 소비자물가지수 작성 과정에 활용되어 신뢰를 얻게 된다면 일부 품목의 가격조사는 중기적으로 표본조사가 아닌 스캐너 데이터 이용방식으로 대체할 수도 있을 것으로 판단된다.

스캐너 데이터를 이용한 지수에 대한 수요가 급격히 증가한다면 통계청 외의 기관(예: 대한상공회의소)이 소비자물가지수와 다른 주기로 공표하는 것이 통계 이용자의 혼란을 방지할 수 있을 것으로 판단된다. 스캐너 데이터를 이용한 지수를 가격지수와 물량지수로 나누어 작성한다면 동 지수를 통해 가격의 변화뿐만 아니라 물량의 변화도 파악할 수 있다. 만약 일부 품목의 가격 또는 물량의 변화는 경기변동, 위기변동 등의 지표로도 활용될 수 있다. 또한 스캐너 데이터를 바탕으로 통크비스트 지수 또는 피셔 지수 등을 작성하여 라스파이레스 방식 소비자물가지수의 한계를 극복할 수 있다.

스캐너 데이터를 소비자물가 등 국가통계에 제대로 이용하려면 데이터산업 육성법(또는 데이터진흥법) 등 관련 제도적 환경을 만들고 관련 인력도 육성해야 한다. 그리고 국가통계 작성을 표본 설계와 많은 조사원을 이용한 표본조사에만 의존하기 보다는 표본조사와 스캐너 데이터, 카드 데이터, 온라인 데이터 등 빅데이터 정리결과를 결합한 형태로 중기적으로 전환할 필요가 있다.

# 목 차

<b>I. 연구추진 배경 및 범위</b> .....	<b>1</b>
1. 연구 배경 .....	1
2. 연구 범위 .....	2
<b>II. 소비자물가지수 산정의 문제점, 이론 및 주요국 사례</b> .....	<b>5</b>
1. 소비자물가지수 산정의 문제점 .....	5
2. 소비자물가지수 편제 이론 .....	6
제1절 직접지수 .....	6
가. 라스파이레스지수 .....	7
나. 로우지수 .....	7
다. 파세지수 .....	8
라. 피셔지수 .....	8
마. 톱크비스트지수 .....	9
제2절 연쇄지수 .....	9
제3절 연쇄편의 문제 .....	10
제4절 GEKS지수 .....	11
제5절 Rolling Window GEKS지수 .....	12
제6절 품질조정 기법 .....	14
가. 시간-더미 헤도닉 기법 .....	14
나. 시간-상품-더미 헤도닉 기법 .....	15
3. 주요국 소비자물가지수 편제 현황 .....	16
제1절 서론 .....	16
제2절 CPI에 스캐너 데이터를 이용하는 국가 .....	17
가. 스위스 .....	17
나. 네덜란드 .....	18
다. 노르웨이 .....	19
라. 벨기에 .....	19
마. 스웨덴 .....	20
바. 덴마크 .....	21
4. 소결 .....	22
<b>III. 스캐너 데이터를 이용한 기존 연구 및 사례</b> .....	<b>24</b>
1. 스캐너 데이터 활용의 장단점 .....	24
제1절 장점 .....	24
제2절 도전 과제 .....	24
가. 비통계적인 부문 .....	24
나. 통계적인 부문 .....	25
다. 물가지수와 관련한 스캐너 데이터 활용 .....	25

2. 국내 기존 연구 및 적용 사례 .....	26
제1절 서론 .....	26
제2절 음식료 품목군 적용 사례 .....	26
가. 개관 .....	26
나. 지수편제 결과 .....	27
제3절 냉장고 수출 품목군 적용 사례 .....	29
3. 해외 기존 연구 및 적용 사례 .....	30
제1절 서론 .....	30
제2절 CPI에 스캐너 데이터 활용 국가 .....	30
가. 스위스 .....	30
나. 네덜란드 .....	31
다. 노르웨이 .....	34
라. 벨기에 .....	34
마. 스웨덴 .....	36
바. 뉴질랜드 .....	37
제3절 스캐너 데이터 활용 방안 연구중인 국가 .....	38
가. 영국 .....	38
나. 프랑스 .....	39
다. 일본 .....	41
4. 소결 .....	42

#### IV. 우리나라 스캐너 데이터를 이용한 품목별 소비자물가지수 작성 .....

1. 대한상공회의소의 스캐너 데이터 .....	46
제1절 스캐너 데이터의 수집, 분석 과정의 요약 .....	46
제2절 스캐너 데이터의 수집, 분석 과정의 요약 .....	46
제3절 스캐너 데이터의 수집 .....	46
제4절 스캐너 데이터의 분류와 특징 .....	48
제5절 스캐너 데이터의 특이항 처리 .....	49
제6절 스캐너 데이터의 보완사항과 활용확대 .....	50
2. 스캐너 데이터의 입수 및 가공 .....	51
제1절 데이터 입수 및 가공 과정 .....	51
제2절 데이터의 원천 .....	51
제3절 데이터의 속성 .....	52
제4절 데이터의 코드 매칭 .....	53
가. 대한상의 코드(KAN)와 통계청 코드(PUMID)의 대응 .....	53
나. 키워드를 이용한 KAN 코드의 분류 및 대응 .....	54
다. 가공식품 지수를 산출하기 위한 코드 매칭 .....	55
제5절 이상치의 탐색 .....	56
가. 이상치 추정 사례 .....	56
나. 이상치 탐색 및 제거 방법 .....	57
3. 스캐너 데이터 기반 물가지수 작성 방법 .....	59



제1절 적절한 지수의 선택 .....	59
제2절 품질조정 기법의 적용 .....	60
4. 스캐너 데이터를 이용한 지수 시산 결과 .....	60
제1절 대한상의 스캐너 데이터 지수 .....	60
가. 원자료 기준 총지수 .....	61
나. 이상치 제거 후 산출 총지수 .....	63
다. CPI 전기대비 증감률과의 상관계수 .....	65
라. 일부 품목 CPI 대체 후 총지수 .....	66
마. 품목별 지수 .....	67
제2절 가공식품 총지수 (통계청 품목 기준) .....	71
제3절 품질조정 지수 .....	73
5. 소결 .....	75
<b>V. 소비자물가지수 개편과 지수 산식 검토 .....</b>	<b>77</b>
1. 소비자물가지수 개편현황과 문제점 .....	77
2. 소비자물가지수 개편 방향 .....	78
<b>VI. 스캐너 데이터 확보 방안 검토 .....</b>	<b>81</b>
1. 해외 주요국의 스캐너 데이터 접근 근거 .....	81
2. 스캐너 데이터 확보 관련 고려 사항 .....	83
3. 스캐너 데이터 확보 방안 .....	84
<b>VII. 결론 및 향후 과제 .....</b>	<b>86</b>
1. 스캐너 데이터 품질 개선과 안정적 확보 .....	86
2. 스캐너 데이터의 소비자물가지수 작성에서의 활용 .....	87
3. 스캐너 데이터를 이용한 지수 작성 .....	88
4. 스캐너 데이터를 활용하기 위한 환경 확보 .....	89
<b>&lt;참고 문헌&gt; .....</b>	<b>91</b>
<b>&lt;별 첨&gt; .....</b>	<b>93</b>
1. 코드 분류에 사용된 키워드 일람표 .....	95
2. 가공식품 지수 산출을 위한 코드 매칭표(종합) .....	98
3. CPI 전기대비 증감률과의 상관계수 .....	101
<b>&lt;부 록&gt; 스캐너 데이터 기반 품목별 소비자물가지수 그래프 .....</b>	<b>113</b>

## 표 목 차

<표 III-1> 21개 품목군의 기하평균 - 점포 종합화 적용 .....	27
<표 III-2> 21개 품목군의 기하평균 - 점포 종합화 비적용 .....	27
<표 III-3> GEKS, RWGEKS 및 피셔 연쇄(변동 바스켓) 지수(21개 품목군의 기하평균) .....	28
<표 III-4> 서울의 통계청 소비자 물가지수 및 GEKS지수(18개 품목군의 기하평균) .....	29
<표 III-5> 물가지수 시뮬레이션 결과 : 직접지수와 GEKS지수 .....	29
<표 III-6> 물가지수 시뮬레이션 결과 : 연쇄지수, GEKS지수 및 RWGEKS지수 .....	30
<표 III-7> 스캐너 데이터를 이용하는 COICOP 그룹 .....	32
<표 III-8> 네덜란드의 스캐너 데이터에 대한 지수 작성 방법 .....	32
<표 III-9> 클리닝 필터 적용 전후 지수 구축결과 비교 .....	39
<표 III-10> CPI에 스캐너 데이터 활용 현황 .....	44
<표 III-11> CPI에 스캐너 데이터 활용 방안 연구 현황 .....	45
<표 IV-1> 대한상공회의소 스캐너 데이터 구조 .....	47
<표 IV-2> 비정상 값 유형별 사유 및 처리 방안 .....	49
<표 IV-3> 대한상공회의소 제공 스캐너 데이터의 예시 .....	52
<표 IV-4> KAN과 PUMID의 다대일 대응 .....	54
<표 IV-5> KAN과 PUMID의 일대다 대응 .....	54
<표 IV-6> 키워드를 이용한 KAN코드 분류 결과 .....	55
<표 IV-7> 이상치 제거 기준에 따른 비중 비교 .....	58
<표 IV-8> CPI와 각 지수의 비교 .....	61
<표 IV-9> CPI 전기대비 증감률과의 상관계수 .....	66
<표 VI-1> 해외 주요국 스캐너 데이터 접근 근거 .....	81
<표 VI-2> 스캐너 데이터 접근 관련 통계법 조문 .....	82

# 그 립 목 차

[그림 I -1] 바코드 규격 .....	2
[그림 III-1] COICOP 5 단계에서의 지수 산정 .....	35
[그림 III-2] COICOP 6 단계에서의 스캐너 데이터 집계 .....	36
[그림 IV-1] 대한상공회의소 스캐너 데이터 .....	47
[그림 IV-2] 대한상공회의소 통계청 제공 스캐너 데이터 .....	48
[그림 IV-3] 이상치 제거 후 세부 상품의 가격 분포 사례 .....	59
[그림 IV-4] 대한상의 스캐너 데이터 총지수(원자료, 업태별 합산) .....	62
[그림 IV-5] 대한상의 스캐너 데이터 총지수(원자료, 지역별 합산) .....	62
[그림 IV-6] 대한상의 스캐너 데이터 총지수(원자료, 4년 지속 상품, 업태별 합산) .....	63
[그림 IV-7] 대한상의 스캐너 데이터 총지수(이상치 제거, 업태별 합산) .....	64
[그림 IV-8] 대한상의 스캐너 데이터 총지수(이상치 제거, 지역별 합산) .....	64
[그림 IV-9] 대한상의 스캐너 데이터 총지수(이상치 제거, 4년 지속 상품, 업태별 합산) .....	65
[그림 IV-10] 대한상의 스캐너 데이터 총지수(이상치 제거, 4년 지속 상품, 업태별 합산, 일부 품목 CPI 대체 시) .....	67
[그림 IV-11] 품목별 지수 - 우유 .....	68
[그림 IV-12] 품목별 지수 - 스낵과자 .....	69
[그림 IV-13] 품목별 지수 - 담배 .....	70
[그림 IV-14] 가공식품 총지수(원자료) .....	71
[그림 IV-15] 가공식품 총지수(원자료, 4년 지속 상품) .....	72
[그림 IV-16] 가공식품 총지수(이상치 제거) .....	73
[그림 IV-17] 가공식품 총지수(이상치 제거, 4년 지속 상품) .....	73
[그림 IV-18] 대한상의 스캐너 데이터 총지수(원자료, 품질조정 포함) .....	74
[그림 IV-19] 대한상의 스캐너 데이터 총지수(원자료, 4년 지속 상품, 품질조정 포함) .....	74
[그림 IV-20] 대한상의 스캐너 데이터 총지수(이상치 제거, 품질조정 포함) .....	75
[그림 IV-21] 대한상의 스캐너 데이터 총지수(이상치 제거, 4년 지속 상품, 품질조정 포함) .....	75
[그림 V-1] 계절변동이 있다고 판단되는 일부 소비자물가지수 품목 .....	80



# I. 연구추진 배경 및 범위

## 1. 연구 배경

빅데이터 시대에 국가통계 작성을 위한 현장 조사의 어려움이 가중되면서 행정통계, 금융거래 등 빅데이터를 활용한 국가통계를 작성하려는 노력이 꾸준히 진행되어 왔다. 또한 무응답, 조사항목 증대, 복잡성 증대 등으로 현장 조사의 효율성과 정확성이 과거보다 낮아지고 있다. 물가조사도 조사대상 업체들이 주요 영업비밀 중 하나인 가격정보를 공개하기 꺼려하고, 물가조사 관련 품목의 생멸주기가 짧아져서 물가지수 구성 품목의 대표성이 낮아지는 등의 문제에 노출되고 있다. 현재 소비자물가지수는 고정된 품목과 가중치를 바탕으로 기준시점 고정 가중산술평균법인 라스파이레스(Laspeyres) 산식을 이용하여 작성되고 있는데 이는 상향 편의를 발생시킬 수 있다. 현행 라스파이레스 방식 소비자물가지수의 편의를 해소하기 위해서는 지수 작성 시점마다 주요 소비 품목과 관련 가중치를 새롭게 산출·변경해야 한다. 이때 유통업체의 바코드를 이용하여 축적된 상품거래 데이터인 스캐너 데이터(scanner data)가 주요 정보원이 될 수 있다. 왜냐하면 스캐너데이터에는 소매점에서 거래되고 있는 모든 상품에 대한 거래 가격 및 거래량 정보, 고빈도(시간별, 일별) 정보, 신상품에 대한 정보가 있어서 기존의 고정 가중식 라스파이레스 지수보다 현실 반영도가 높다고 판단되는 톨크비스트(Törnqvist) 지수 또는 피셔(Fisher) 지수 등을 작성할 수 있기 때문이다.

1990년대부터 유통업체들은 판매시점 정보관리 시스템(Point of Sales, POS 시스템)<sup>1)</sup>을 도입하여 바코드를 이용한 상품거래를 하면서 동 거래 데이터를 축적하고 있다. 이와 같이 바코드를 이용한 상품거래 데이터를 스캐너 데이터(Scanner data)라고 부른다. 여기서 바코드는 GTIN(Global Trade Item Number) 또는 EAN (European Article Number)를 따라 작성되는데 이 코드는 국내 또는 국외로 유통되는 상품을 식별하는데 사용하는 국제표준상품 코드이다.<sup>2)</sup> 소매상품에 일반적으로 되고 있는 바코드는 [그림 I-1]과

1) 유통산업발전법 제2조에 따르면 POS 시스템은 광학적 자동관독방식에 따라 상품의 판매, 매입 또는 배송 등의 정보가 수록된 것을 말한다.

2) 유통산업발전법 제2조 제10항에 있는 유통표준코드를 의미한다. 유통산업발전법 제2조 제10항 : "유통표준코드"란 상품·상품포장·포장용기 또는 운반용기의 표면에 표준화된 체계에 따라 표기된 숫자와 바코드 등이다. 유통발전법 시행규칙 제3조에 따르면 유통표준코드로는 공통상품코드용 바코드심별(KS X 6703), 유통상품코드용 바코드심별(KS X 6704), 물류정보시스템용 응용식별자와 UCC/EAN-128바코드심별(KS X 6705)가 있다.

같은 표준형 상품식별코드(GTIN-13, EAN-13)이다. 이를 보면 상품, 회사, 국가를 식별하는 12자리와 검토하는 1자리 총 13자리로 구성되어 있다.

[그림 I-1] 바코드 규격



2017년 현재 소비재 유통매장의 90% 이상, 대형 유통매장 거래의 100%에 대해 POS 시스템을 도입하고 있는 것으로 판단된다.<sup>3)</sup> 이는 신용카드의 결제 승인을 관리하는 VAN 사들이 POS 터미널(PC형) 및 바코드 스캐너, 사인패드, 카드 단말기 일체를 공급하는데 기인한다. 유통업체는 POS 시스템에서 판매상품의 상품명과 판매가격을 등록하여 상품을 거래하고, 할인행사, 매대 관리 및 상품 구매도 하고 있다. 유통업체는 POS 시스템을 통해 판매 상품의 바코드를 스캔하여 상품명과 가격을 영수증에 인쇄하고 구매품 총액을 계산하여 거래를 완성한다. 그리고 유통업체는 영업시간 종료 후 POS 시스템을 통해 해당일의 모든 거래 정보를 종합하여 최종 매출액, 상품별 총 판매 개수, 총 판매액을 정리한다. 이때 POS 시스템으로부터 얻는 데이터를 스캐너 데이터라 부르는데 이 데이터는 유통매장에서 일마감 후 정리된 상품별 판매수량, 판매총액, 판매 단가로 구성된 데이터이다.

통상 스캐너 데이터는 데이터 수집 비용을 낮추고 조사항목을 늘려 표본 분산과 측정 오차를 감소시킬 뿐만 아니라 조사대상처의 응답부담도 감소시킨다는 장점을 가지고 있다. 벨기에, 덴마크, 네덜란드, 스웨덴, 스위스 등의 통계작성기관들은 이미 소비자물가지수 작성에 스캐너 데이터를 활용하고 있다. 따라서 우리나라에서도 스캐너 데이터를 소비자물가지수 작성에 어떻게 활용할 수 있는지 면밀히 검토할 필요가 있다.

## 2. 연구 범위

본 연구에서는 스캐너 데이터를 이용한 소비자물가지수 작성 국내외 사례

3) 이 패러그래프는 대한상공회의소의 자문을 바탕으로 작성되었다.

를 수집·정리하고, 대한상공회의소의 스캐너 데이터를 이용하여 품목별 소비자물가지수를 최상위 지수를 포함한 지수로 작성한다. 작성된 스캐너 데이터 관련 품목별 소비자물가지수와 스캐너 데이터를 이용한 국내외 사례를 바탕으로 우리나라에서 스캐너 데이터를 소비자물가지수 작성 또는 소비자물가지수 조사에 어떻게 활용할 수 있는지 살펴본다. 아울러 스캐너 데이터를 안정적으로 확보할 수 있는 방안과 소비자물가지수 개편과 관련된 지수 적용 방안을 검토한다.

연구 범위를 구체적으로 다시 정리하면 다음과 같다. 첫째, 본 연구에서는 가중치의 고정 문제 등 기존 라스파이레스식 소비자물가지수 산정의 문제점에 대해서 개괄적 논의를 정리하고, 이와 관련하여 네덜란드, 스위스, 스웨덴, 뉴질랜드 등 주요국의 소비자물가지수 작성 사례에 대해서도 전반적으로 정리해본다. 사례를 바탕으로 스캐너 데이터가 부상하게 된 배경과 이와 관련된 지수 산식을 정리한다. 지수 산식에서는 기존의 라스파이레스 지수 외에도 로우 지수, 파쉐 지수, 톱크비스크 지수 등을 살펴본다.

둘째, POS 시스템으로부터 축적된 고객별 거래 데이터를 바탕으로 스캐너 데이터를 정의하고, 대한상공회소가 축적하여 통계청에 제공한 스캐너 데이터의 특성을 정리한다. 그리고 반품, 묶음 판매 등에 따라 나타나는 스캐너 데이터의 이상치를 조정·정제 방법을 정리한다.

셋째, 스캐너 데이터에 대한 국내외 활용 및 연구를 사례를 정리하여 스캐너 데이터가 사용되는 품목, 스캐너데이터를 소비자물가지수 적용시 발생하는 장점과 단점을 정리한다. 특히 스캐너 데이터를 실제 품목별 소비자물가지수에 적용하고 있는 사례를 정리한다.

넷째, 스캐너 데이터를 이용하여 GEKS, Rolling Window GEKS 등 새로운 형태의 지수를 작성해보고 기존 방식의 소비자물가지수와 비교, 분석하여 스캐너데이터가 현행 소비자물가지수 작성에 도움이 될 수 있는 방안을 정리한다.

다섯째, 스캐너 데이터를 이용하여 품목별 소비자물가지수를 작성하려면 스캐너 데이터를 안정적으로 주기적으로 확보하는 것이 중요하다. 본 연구에서는 스캐너 데이터를 안정적으로 확보할 수 있는 방안에 대한 해외 사례를 정리하고 이를 바탕으로 관련 법적, 제도적 쟁점사항을 검토하고, 스캐너 데이터 보유기관과의 협력 방안을 합리적으로 마련한다.

여섯째, 스캐너 데이터를 소비자물가지수 작성과 소비자물가지수 조사에 활용할 수 있는 방안을 조사과정 및 지수작성과 연결하여 정리한다. 다른 나라의 스캐너 데이터를 소비자물가지수에 활용하는 방법을 벤치마킹하여 우리나라

에서 스캐너 데이터를 소비자물가지조사에 적용하는 방안을 마련한다.

일곱째, 소비자물가지수의 개편주기를 단축하면서 소비자물가지수의 개정의 빈도가 잦아졌다. 이러한 잦은 개정에 따른 이용자의 혼란을 만들기 때문에 이를 방지하기 위해서 지수 작성 방법중 로우(Lowe) 지수를 소비자물가지수 개편에 활용할 수 있는지 살펴본다.



## II. 소비자물가지수 산정의 문제점, 이론 및 주요국 사례

### 1. 소비자물가지수 산정의 문제점

우리나라 소비자물가지수는 가구에서 일상생활을 영위하기 위하여 구입하는 상품과 서비스의 가격변동을 측정하기 위하여 작성된다. 소비자물가지수 조사대상은 조사권역 내 상권, 매출액 규모 등을 감안하여 그 지역의 가격을 대표할 수 있으며, 계속적으로 조사 가능한 대상처를 유의추출한다. 대상처는 25,000개, 휴폐업, 업종변경, 상품 취급 중단 등으로 3.2% 표본대체 및 무응답으로 처리된다. 대상 품목은 2015년 기준 품목수가 460개이고 조사 상품수는 999개이다. 1인 이상 전국가구(농어가제외)의 월평균 소비지출총액의 10,000분의 1 이상이면서 동종 상품군의 가격을 대표할 수 있으며, 시장에서 계속적으로 조사가 가능한 품목이어야 한다. 지출목적별 분류(COICOP)체계에 따라 품목부호가 부여되고 조사방법은 조사대상처를 직접 방문하여 면접 조사를 시행한다.

소비자물가지수는 고정된 품목과 가중치를 바탕으로 기준시점고정 가중산술평균법인 라스파이레스(Laspeyres) 산식을 사용하여 작성한다. 가중치는 가계동향조사의 소비지출액 자료를 기초로 산출한다.

$$I_t = \frac{\sum P_t Q_o}{\sum P_o Q_o} = \frac{\sum P_o Q_o \cdot \frac{P_t}{P_o}}{\sum P_o Q_o} = \frac{\sum W_o \cdot \frac{P_t}{P_o}}{\sum W_o} \quad (1)$$

\* I: 물가지수, P: 가격, Q: 거래량, W: 가중치, o: 기준년도, t: 비교시점

라스파이레스 산식은 품목과 가중치가 기준시점에 고정되어 있어 지수계산이 간편하지만 기준시점에서 멀어질수록 소비구조의 변화를 물가지수에 반영하지 못해서 물가지수에 상향편의가 발생한다는 문제를 가지고 있다.

1996년 미국 보스킨위원회 (Boskin et al., 2001)는 라스파이레스 산식을 이용하는 미국의 소비자 물가지수에 품목 대체 편익, 판매점 대체 편익, 품질 변화 편익, 신제품 편익, 기본 지수 편익으로 인해 소비자물가지수에 상향편의가 있음을 정리하였다. 이상적 지수는 지수 편익시 소비 품목과 가중치를 바꾸어야 하는데 비용과 시간이 많이 들고 통계작성의 안정성을 담보하는데

제약이 있기 때문에 대부분의 국가 통계작성기관이 라스파이레스 산식을 이용하여 소비자물가지수를 작성하고 있다. 현재 우리나라에서는 소비자물가지수의 현실 반영도를 제고하기 위해 매 5년에서 매 2년~3년 마다 조사대상 품목 및 가중치 등을 재조정하는 한편 연쇄지수를 작성하고 있다. 소비자물가지수에 나타나는 상향 편이의 문제를 근본적으로 규명하기 위해서는 기준시점과 비교시점에서의 가격과 물량자료를 필요로 한다. 최근에는 소매점에 바코드와 스캐너 장비가 사용됨에 따라 가격과 판매량에 관한 대용량의 실시간 판매 정보 획득이 용이해졌는데, 이에 따라 유럽 및 호주 등에서는 스캐너 데이터를 이용하여 소비자물가지수 작성에 활용하는 방법에 관한 다양한 연구가 진행되고 있다 (김기석, 2012; Ivancic, 2005; Hawkes and Piotrowski, 2003). 스캐너 데이터를 이용하는 경우 기간별로 변동하는 가중치를 적용하여 개별 품목 단위에서는 최상위 지수(superlative index)에 가까운 지수 편제가 가능하다. 가령 스캐너 데이터를 이용하여 연쇄지수(chained index)를 작성하는 경우 개별 품목 단위에서 피셔지수(Fisher index) 및 톨크비스트지수(Törnqvist index)와 같은 최상위 지수의 산출이 가능한데, 이처럼 스캐너 데이터를 소비자물가지수에 활용하는 경우 적어도 개별 품목 단위에서는 라스파이레스 산식에 얽매이지 않고 보다 진전된 지수 구축이 가능하다. 이에 따라 다음 절에서는 스캐너 데이터를 적용한 기존 연구 및 지수 구축에 앞서서 라스파이레스 산식을 포함하여 물가지수 이론에 대해서 전반적으로 검토·정리한다.

## 2. 소비자물가지수 편제 이론

### 제1절 직접지수

직접지수(direct index)는 기준시점과 비교시점에 동시에 존재하는 제품들만을 대상으로 지수를 산출하기 때문에 시간이 지날수록 대응되는 제품의 수가 감소하게 된다. 이에 반해 스캐너 데이터는 기존의 조사 자료와 비교하여 표본기간 중 사라지거나 추가되는 제품들이 많기 때문에 직접지수를 이용하면 시간이 지날수록 대응되는 제품의 수가 크게 줄어드는 경향이 있다. 따라서 스캐너 데이터를 이용하여 지수를 산출하는 경우에는 연쇄지수(chained index)를 이용하여 인접한 두 시점 간 대응되는 제품의 수를 극대화할 수가 있다. 직접지수에서 판매액 또는 지출액 등을 가중치로 활용하는 대표적인 품목지수 산식으로는 라스파이레스식, 로우식, 파세식, 피셔식, 톨

크비스트식 등이 있다. 아래에서는 강규호 외 (2015) 논문 등을 주로 참고하여 각 지수를 정의하고 해당 특징을 살펴보고자 한다.

### 가. 라스파이레스지수

라스파이레스식은 아래 식에서 볼 수 있듯이 기준시점의 가중치를 적용하는 방식으로 현행 소비자물가지수 작성 시 활용되는 방법이다.

$$Laspeyres_t = \sum_i^N \left( \frac{p_i^t}{p_i^0} s_i^0 \right) \quad (2)$$

$p_i^0$ 는 품목  $i$ 의 기준시점 가격이고,  $p_i^t$ 는 품목  $i$ 의  $t$ 시점 가격( $t=1, \dots, T$ ),  $s_i^0 = (p_i^0 \times q_i^0) / \sum_{i=1}^N (p_i^0 \times q_i^0)$ 로서 기준시점의 전체 소비지출에서 차지하는 품목  $i$ 의 비중을 각각 의미한다. 즉, 라스파이레스식은 기준시점의 소비 바스켓으로 소비량을 고정하는 경우 비교시점에 동 소비 바스켓을 구입하기 위해서 얼마나 더 지출하였는지를 측정하는 방식이다.

### 나. 로우지수

한편 로우식은 기본적으로 과거의 소비지출 바스켓을 이용한다는 점에서 라스파이레스식과 유사하나, 기준시점의 소비량( $q_i^0$ ) 대신 특정 시점의 소비량( $q_i^b$ )을 이용하는 면에서 차이가 있다.

$$Lowe_t = \sum_i^N \left( \frac{p_i^t}{p_i^b} s_i^b \right) \quad (3)$$

$s_i^b = (p_i^b \times q_i^b) / \sum_{i=1}^N (p_i^b \times q_i^b)$ 로서 특정 시점의 전체 소비지출에서 차지하는 품목  $i$ 의 비중을 의미한다.

#### 다. 파셰지수

파셰식은 식(4)에서 볼 수 있듯이 기준시점 대신 비교시점 또는 해당시점의 소비 바스켓을 가중치로 적용하는 방식이다.

$$Paasche_t = \left[ \sum_i^N \left( \frac{p_i^0}{p_i^t} s_i^t \right) \right]^{-1} \quad (4)$$

$s_i^t = (p_i^t \times q_i^t) / \sum_{i=1}^N (p_i^t \times q_i^t)$ 는 t시점 전체 소비지출에서 차지하는 품목 i의 비중을 의미한다. 즉, 파셰식은 해당시점의 바스켓을 소비한다면 기준시점의 가격으로 동 바스켓을 구입하기 위해 지출한 금액(기준시점 지출액)을 계산한 후 해당시점의 지출액과 비교함으로써 물가지수를 산출하는 방식이다.

현행 실제 물가지수 산출시 파셰식 보다는 라스파이레스식이 주로 이용되는데 이는 라스파이레스식이 과거 소비 지출액을 이용하여 가중치를 산출이 용이한 반면 파셰식은 실시간으로 소비 지출액을 파악해야 하므로 작성이 용이하지 않기 때문이다.

#### 라. 피셔지수

피셔식은 식(5)에서 볼 수 있듯이 라스파이레스식과 파셰식으로 산출된 지수를 기하평균하는 방식으로 산출된다.

$$Fisher_t = [Laspeyres_t \times Paasche_t]^{\frac{1}{2}} \quad (5)$$

피셔지수는 어떤 상품의 상대가격 상승(하락) 시 해당 상품의 상대수요도 감소(증가)하는 일반적인 경제 환경 하에서 최상위 지수로 간주된다. 상품 간에 대체관계가 존재하고 교차탄력성이 1이라고 가정한다면 기하평균은 생계비 변화를 정확하게 반영하는 산출방식이 되는 반면, 대체효과를 고려하지 않는 산술평균은 일종의 상향 편의를 갖게 된다 (한국은행, 2014). 이는 해당 상품의 상대가격이 가령 10% 상승하는 경우 상대수요량도 10% 감소하는 경우를 의미한다.

## 마. 톱크비스트지수

톱크비스트식도 아래 식에서 볼 수 있듯이 상품 간의 대체관계를 반영하면 서도 기준 및 비교 시점의 소비지출 가중치를 모두 적용하는 최상위지수로 간주된다.

$$\prod_{i=1}^N \left[ \frac{p_i^t}{p_i^0} \right]^{\frac{s_i^0 + s_i^t}{2}} \quad (6)$$

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

$p_i^0$  : 제품  $i$ 의 기준시점 단가       $p_i^t$  : 제품  $i$ 의 비교시점 단가

$q_i^0$  : 제품  $i$ 의 기준시점 판매량       $q_i^t$  : 제품  $i$ 의 비교시점 판매량

다만 현실에서 판매액 등에 대한 자료가 존재하지 않아 가중치를 활용할 수 없는 경우에는 모든 제품들의 지출비중을 동일하게 설정( $s_i^0 = s_i^t = 1/N$ )한 제본스지수(Jevons index)를 적용할 수가 있다.

$$\prod_{i=0}^N [p_i^t / p_i^0]^{1/N} \quad (7)$$

## 제2절 연쇄지수

시장에서 판매되는 제품들의 조합이 안정적이지 못하고 시점에 따라 변동하는 스캐너 데이터의 경우에는 직접지수 대신 연쇄지수를 활용하여 물가지수를 산출한다. 연쇄지수는 우선 인접한 두 기간 사이의 대응된 거래들에 대해 단기지수(short-term index)를 구한 후 이들의 등락률을 누적하여 최종적으로 지수를 산출한다. 한편 단기지수 산출에 적용된 지수산식에 따라 연쇄지수는 직접지수와 같이 라스파이레스, 파세, 톱크비스트 연쇄지수 등으로 분류될 수 있다.

· 라스파이레스 단기지수 등락률

$$SI_t = \sum_{i \in U_{t-1,t}} \left[ \frac{p_i^t}{p_i^{t-1}} \times s_i^{t-1} \right] \quad (8)$$

· 파세 단가지수 등락률

$$SI_t = \sum_{i \in U_{t-1,t}} \left[ \frac{p_i^t}{p_i^{t-1}} \times s_i^t \right] \quad (9)$$

· 톱크비스크 단가지수 등락률

$$SI_t = \prod_{i \in U_{t-1,t}} \left[ \frac{p_i^t}{p_i^{t-1}} \right]^{\frac{s_i^{t-1} + s_i^t}{2}} \quad (10)$$

· 연쇄지수=전기지수\*당기 등락률

$$P_{0,t} = P_{0,t-1} \times SI_t \quad (11)$$

$U_{t-1,t}$  : 시점  $t-1$ 과  $t$ 에 매칭되는 제품의 집합

$p_i^{t-1}$  : 제품  $i$ 의  $t-1$ 시점 가격

$p_i^t$  : 제품  $i$ 의  $t$ 시점 가격

$s_i^{t-1}$  : 전기의 가격으로 전기 바스켓을 구입했을 때, 제품  $i$ 의 지출비중

$$s_i^{t-1} = (p_i^{t-1} \times q_i^{t-1}) / \left( \sum_{t \in U_{t-1,t}} [p_i^{t-1} \times q_i^{t-1}] \right)$$

$s_i^t$  : 당기 가격으로 당기 바스켓을 구입했을 때, 제품  $i$ 의 지출비중,

$$s_i^t = (p_i^t \times q_i^t) / \left( \sum_{t \in U_{t-1,t}} [p_i^t \times q_i^t] \right)$$

### 제3절 연쇄편의 문제

연쇄편의는 모든 제품들의 가격이 기준시점 수준으로 회귀하더라도 해당 품목의 물가지수가 기준시점의 값인 1(unity)로 돌아오지 않는 현상을 의미한다. 스캐너 데이터와 같이 소모율(attrition rate)이 높은 자료를 대상으로 물가지수를 산출할 때는 직접지수 대신 연쇄지수가 이용되는데 가격 및 수량 변동성이 클 경우 연쇄편의를 유발할 수 있다는 문제가 있다.

<참고>

연쇄편의 문제에 대한 직관적 이해<sup>4)</sup>

2개 제품에 대해 3개 시점(0, 1, 2)을 가정하는 경우 0시점에서 2시점까지의 톱크비스트 연쇄지수는 다음과 같다.

$$P_{0,2} = \prod_{i=1}^2 (p_i^1/p_i^0)^{(s_i^0+s_i^1)/2} \times \prod_{i=1}^2 (p_i^2/p_i^1)^{(s_i^1+s_i^2)/2}$$

제품 1이 시점 1에서 가격을 할인해서 가격이 크게 하락한 후 시점 2에서 원래 수준을 회복하는 반면( $p_1^1 < p_1^0 = p_1^2$ ), 제품 2의 가격은 변함이 없는 경우( $p_2^1/p_2^0 = p_2^2/p_2^1 = 1$ )를 가정하면 톱크비스트 연쇄지수는 다음과 같다.

$$P_{0,2} = (p_1^1/p_1^0)^{(s_1^0-s_1^2)/2}$$

제품 1의 판매비중이 가격 할인 전 시점(0)과 가격 할인 후 시점(2)간 차이가 없다면 물가지수( $P_{0,2}$ )는 1이 된다. 하지만 소비자들이 가격 할인 기간인 시점 1에 구매를 늘리고 시점 2에서는 제품 1의 소비량을 큰 폭으로 감소시킬 가능성이 크기 때문에  $p_1^1/p_1^0 < 1$ 이고  $s_1^0 > s_1^2$ 이 된다. 따라서  $P_{0,2} < 1$ 이 되어 물가지수에 하향편의가 발생하게 된다. 한편 직접지수인 경우에는 시점 0과 시점 2의 가격이 동일하기 때문에 라스파이레스에서 톱크비스트 지수까지 물가지수에 편의가 발생하지 않는다.

#### 제4절 GEKS지수

이러한 연쇄편의 문제를 극복하기 위해 Ivancic et al (2011)에 의해 제시된 방법이 바로 GEKS지수 (Gini, 1931; Elter and Kves, 1964; Szulc, 1964)이다. GEKS지수는 우선 비교가능한 모든 두 시점을 대상으로 최상위지수(피셔지수 또는 톱크비스트지수)를 만든 후, 표본기간 내에 존재하는 모든 시점을 접속시점(linking period)으로 하여 자료 개수만큼의 지수를 산출한다. 그런 다음 동 지수들을 기하평균함으로써 해당 시점(t)의 GEKS지수( $P_{0,t}^G$ )를 산출한다.

4) 참고 내용은 강규호 외(2012)를 참조하여 작성되었다.

$$P_{0,t}^G = \prod_{i=1}^2 (P_{0,j} \times P_{j,t})^{\frac{1}{T+1}} \quad (12)$$

$$P_{0,j} = \prod_{i \in U_{0,j}} \left( \frac{p_i^j}{p_i^0} \right)^{\frac{s_i^0 + s_i^j}{2}} \quad (13)$$

$$P_{j,t} = \prod_{i \in U_{j,t}} \left( \frac{p_i^t}{p_i^j} \right)^{\frac{s_i^j + s_i^t}{2}} \quad (14)$$

$P_{0,t}^G$ : 시점  $t$ 의 GEKS지수       $P_{0,j}$ : 0과  $j$ 사이의 지수(통크비스크 산식)

$P_{j,t}$ : 시점  $j$ 과  $t$  사이의 지수(통크비스크 산식)

$U_{j,t}$ : 시점  $j$ 와  $t$ 에 매칭되는 제품의 집합

$p_i^0$ : 제품  $i$ 의 시점 0에서의 가격       $s_i^0$ : 제품  $i$ 의 시점 0에서의 가중치

$p_i^j(p_i^t)$ : 제품  $i$ 의 시점  $j(t)$ 에서의 가격

$s_i^j(s_i^t)$ : 제품  $i$ 의 시점  $j(t)$ 에서의 가중치

GEKS지수는 비교 가능한 모든 두 시점의 재화 조합을 이용함으로써 직접 지수와 비교하여 대응되는 제품의 수를 크게 늘릴 수 있는 장점이 있다. 또한 스캐너 데이터를 이용할 때 직접지수의 대안으로 사용되는 연쇄지수가 연쇄편의에 취약한 반면, GEKS지수는 동 편의로부터 자유로운데 이는 피셔 또는 통크비스트 산식을 통해 작성된 GEKS지수가 이행성(transitivity:

$$P_{0,t}^G \equiv \prod_{l=0}^t P_{j-1,j}^G)을 만족하기 때문이다.$$

## 제5절 Rolling Window GEKS지수

GEKS지수는 연쇄편의로부터 자유로운 점 등 우월한 측면이 있지만 현실적으로 물가지수 편제에 직접 적용할 수는 없는데, 이는 새로운 시계열 자료가 추가되면 과거 모든 시점의 GEKS지수를 수정(revision)해야 하기 때문이다. 예를 들어 T시점에서는 GEKS지수인  $P_{0,T}^G$ 을 계산하기 위해 0부터 T시점까지의 자료를 사용한다.

$$P_{0,T}^G \equiv \prod_{l=0}^T (P_{0,l} \times P_{l,T})^{\frac{1}{T+1}} \quad (15)$$

한 시점이 추가된 T+1시점에서는 새로 추가된 자료를 이용하여 T시점의



GEKS지수  $P_{0,T}^{G(0,T+1)}$ 를 회귀적으로(retrospectively) 수정한다.

$$P_{0,T}^{G(0,T+1)} \equiv \prod_{l=0}^{T+1} (P_{0,l} \times P_{l,T})^{\frac{1}{T+2}} \quad (16)$$

Ivancic et al(2011)은 과거 지수에 대해 수정과정을 거치지 않는 RWGEKS지수를 제안하였다. 동 지수의 산출과정을 살펴보면 우선, 피셔 또는 톱크비스트 GEKS지수는 이행성(transitivity)을 만족하기 때문에 T+1 시점의 GEKS지수를 식(17)과 같이 표현할 수 있다.

$$\begin{aligned} P_{0,T+1}^G &= P_{0,T}^G \times P_{T,T+1}^G \\ &= \prod_{l=0}^{T+1} (P_{0,l} \times P_{l,T})^{\frac{1}{T+2}} \times \prod_{l=0}^{T+1} (P_{T,l} \times P_{l,T+1})^{\frac{1}{T+2}} \end{aligned} \quad (17)$$

RWGEKS지수를 구하기 위해서는 우선 데이터 또는 윈도우의 크기(window size)를 결정해야 하는데 여기서는 T+1개로 자료의 크기를 고정한 후 T+1 시점의 RWGEKS지수를 산출한다. 지수수정 문제를 피하기 위하여, T 시점에서 산출된 GEKS지수( $P_{0,T}^G$ )에 T 시점과 T+1 시점 사이의 물가변동을 나타내는 링크(link)지수인  $P_{T,T+1}^{G(1,T+1)}$ 를 곱해줌으로써 T+1 시점의 RWGEKS지수를 식(18)과 같이 구할 수 있다.

$$\begin{aligned} P_{0,T+1}^{RG} &= P_{0,T}^G \times P_{T,T+1}^{G(1,T+1)} \\ &= \prod_{l=0}^T (P_{0,l} \times P_{l,T})^{\frac{1}{T+1}} \times \prod_{j=1}^{T+1} (P_{T,j} \times P_{j,T+1})^{\frac{1}{T+1}} \end{aligned} \quad (18)$$

만약 T+2 시점의 RWGEKS지수를 산출하기 위한 링크지수 산출을 위해서는 윈도우가 기존 (1, T+1)에서 (2, T+2)로 한 시점 앞으로 순차적으로 이동하게 된다. RWGEKS지수는 GEKS지수에서 약간의 변형이 이루어졌다는 점에서 연쇄편의로부터 완전히 자유롭지는 않지만, 새로운 자료가 추가되더라도 과거 지수를 수정할 필요가 없다는 측면에서 실무적 활용성이 높다고 할 수 있다.

## 제6절 품질조정 기법

물가지수 편제를 위한 가격을 조사할 때, 신규 조사규격이 변경됨에 따라 조사상품 및 서비스의 가격이 달라지는 경우가 많다. 이러한 가격 변화에는 품질변화에 기인한 것과 순수한 가격변동에 의한 것이 혼재되어 있다고 볼 수 있다. 따라서 가격 변동을 물가지수 편제에 반영할 때에는 순수한 가격의 변동에 의한 부분만 물가지수에 반영해 주어야 하는데 이러한 방법들을 품질조정 기법이라고 한다. 빠른 기술발전으로 신규상품의 품질 차이가 크게 발생하는 경우에, 가격차이 모두를 품질에 기인한다고 판단한다면 가격변동이 지수에 반영되지 않는 문제점이 발생하는데, 이를 해결하고자 제안된 방법으로 헤도닉(Hedonic) 기법이 있다. 본 연구에서는 상품제품별로 다양한 가격자료를 수집하고 통계회귀모형을 구하여 변화된 품질요소별로 가격을 계량화하는 헤도닉(Hedonic) 방법을 적용하여 품질조정을 적용해 보았다. 이 기법에 사용되는 회귀모형은 더블로그, 세미로그, 일반화선형모형 등이 있으며 소비자 물가지수 산출시 적용되는 모형은 종속변수만 로그로 변환하는 세미로그 형태를 취하는 것으로 알려져 있다. 헤도닉 기법을 적용하기 위해서는 품질을 정의할 수 있는 요인별로 다양한 가격대의 모델이 존재해야 하며 관측치도 충분히 많아야 한다. 헤도닉 기법은 기술혁신이 품질 변화로 이어지는 상품에 주로 적용되는데, 실제로 통계청의 소비자물가지수 산출 시 컴퓨터본체, 노트북컴퓨터, 모니터, TV 등에 활용되고 있다.

### 가. 시간-더미 헤도닉 기법

강규호 외 (2015)에서 소개한 스캐너타입 자료의 품질조정을 위해 개발된 방법은 시간-더미 헤도닉(Time-Dummy Hedonic method ; TD) 기법의 특징을 살펴보면 다음과 같다.

일반적인 TD 기법 적용 절차는 크게 두 단계로 구분해 볼 수 있다. 1 단계는 상품의 품질을 결정하는 특성치들과 시간더미(time dummy)를 설명변수로 설정하고 상품가격의 로그값을 종속변수로 하는 선형회귀식을 가중최소제곱방법(weighted least squares ; WLS)으로 추정한다. 1 단계에서 적합한 모형을 간략하게 기술해 보면 식 (19)와 같다.

$$\ln p_i^t = \alpha + \delta_t D^t + \sum_{k=1}^K \beta_k z_{ik} + \frac{\sigma}{\sqrt{s_i^t}} \epsilon_i^t \quad (19)$$

단,  $\epsilon_i^t$  : 표준정규분포를 따르는 오차항

$p_i^t$  : 제품  $i$ 의  $t$  시점 가격

$z_{ik}$  : 제품  $i$ 의 특성치

$s_i^t$  : 제품  $i$ 의  $t$  시점에서의 지출 비중

$D_i^t$  : 기준시점에는 0,  $t$  시점에는 1의 값이 주어지는 시간 더미변수

모형적합 후 2 단계에서는 1 단계에서 구한 추정치  $\delta_t$ 를 이용하여 0 시점과  $t$  시점 사이의 물가지수 변동분을 식 (20)과 같이 계산한다.

$$P_{0,t} = \exp(\hat{\delta}_t) \quad (20)$$

종속변수인 가격변수에 로그를 취한 뒤 시간 더미변수를 포함한 선형 회귀모형을 적합하였기 때문에 시간 더미변수의 회귀계수값에 지수(exponential)함수를 취하여 역산을 해줌으로써 가격 변동분을 계산하게 된다. 여타 개별 상품의 특성치들의 변화로 설명되지 않는 부분이 가격 변동분으로 간주되어 품질조정 효과를 나타내게 되는 것이다.

#### 나. 시간-상품-더미 헤도닉 기법

앞서 살펴본 모형은 직관적이고 이해하기 쉬운 반면, 실무적인 제약이 많이 따른다. 스캐너 데이터의 경우 상품의 품질에 영향을 미치는 특성치 변수에 대한 정보를 얻기가 실제로는 매우 어렵다고 할 수 있다. 이러한 실무적 어려움을 극복하고자 de Haan and Krsinich (2014)는 바코드(barcode) 자체를 하나의 품질 특성치들의 집합으로 간주하는 TPD (Time Product Dummy) 회귀식을 제안하였다. 즉, 특정 개별 상품의 특성치 변수 대신 바코드로 설명되는 품질 집단을 전제하고 모든 제품에 동시에 존재하는 고정시간효과(fixed-time effect)가 품질변동과 무관한 순수한 가격변동이라는 가정을 반영하여 모형을 설정하게 된다. 이 모형을 간략하게 표현하면 식 (21)과 같다.

$$\ln p_i^t = \alpha + \delta_i D^t + \sum_{i=1}^{N-1} \gamma_i D_i + \frac{\sigma}{\sqrt{s_i^t}} \epsilon_i^t \quad (21)$$

단,  $\epsilon_i^t$  : 표준정규분포를 따르는 오차항

$p_i^t$  : 제품  $i$ 의  $t$  시점 가격

$D^t$  : 기준시점에는 0,  $t$  시점에는 1의 값이 주어지는 더미변수

$D_i$  : 제품  $i$ 에 해당하면 1, 아니면 0의 값이 주어지는 더미변수

$s_i^t$  : 제품  $i$ 의  $t$  시점에서의 지출 비중

동 모형에 의하면 회귀계수  $\hat{\gamma}_i$ 가 제품별 특성 차이에 의한 가격차를 설명하며  $\hat{\delta}_i$ 가 순수한 가격변동을 설명하게 된다.

모형적합 후 2 단계에서는 1 단계에서 구한 회귀계수  $\hat{\alpha}$ ,  $\hat{\delta}_i$ ,  $\hat{\gamma}_i$ 을 이용해서 식 (22) 및 (23)와 같이 기준시점과 비교시점의 가격변수를 추정하게 된다.

$$\ln p_i^0 = \hat{\alpha} + \hat{\gamma}_i \quad (22)$$

$$\ln p_i^t = \hat{\alpha} + \hat{\delta}_i + \hat{\gamma}_i \quad (23)$$

최종적으로 기준시점 0 시점에 대한  $t$  시점의 품질조정 물가지수는 앞서 사용했던 식 (22)와 동일한 방식으로 계산된다.

### 3. 주요국 소비자물가지수 편제 현황

#### 제1절 서론

본 장은 IMF의 SDDS(Special Data Dissemination Standard) 자료를 기초로 스캐너 데이터를 실제 소비자물가지수 편제에 활용하고 있는 나라를 중심으로 스캐너 데이터와의 관련성을 중심으로 동 물가지수 편제 현황을 정리해 보았다. 현재 Chessa (2016)에 따르면 전 세계적으로 스캐너 데이터를 실제 소비자물가지수에 활용하고 있는 나라는 네덜란드, 노르웨이, 스웨덴, 스위스, 벨기에, 덴마크, 아이슬란드 및 뉴질랜드로 파악된다. 본 장에서는 이 중 인구가 작은 아이슬란드와 IMF SDDS 상에 해당 통계정보가 제공되어 있지 않은 뉴질랜드를 제외한 6개 나라의 물가지수 현황을 살펴본다.

## 제2절 CPI에 스캐너 데이터를 이용하는 국가

### 가. 스위스

#### 1) 대상 범위

인구 범위로는 스위스에 영구 거주하는 모든 가구가 대상이 되며, 지역적으로는 스위스 전체를 대상으로 한다. 한편 스위스 영내에서 거주자 및 비거주자의 소비지출을 모두 포괄하는 HICP (Harmonized Index of Consumer Prices) 지수를 별도로 편제하고 있다. CPI와 HICP는 12개의 주요 상품군 및 122개의 하위 품목군으로 구성되며 267개의 하위 지수와 1,100개의 설문 조사에 기초하고 있다. 상품 분류 체계는 COICOP(목적별 개인 소비 분류)에 정의를 따른다.

대부분의 가격은 매월 첫 2주 동안 한 달에 한번 수집된다. 연료 및 휘발유 가격은 시작과 중간 등 한 달에 두 번 수집되며 가장 중요한 소매점의 식품 및 유사 식품(near food) 가격은 스캐너 데이터를 통해 매월 2주 동안 수집된다. 과일, 의류 및 기타 제품의 경우 계절별로 자료 수집 월이 별도로 정의되며 일부 서비스의 가격은 분기별로 수집된다.

#### 2) 가격의 집계 및 지수 산정

가격 정보에 대해서는 다양한 수집 방법이 적용되고 있다. 우선, 약 27,000개의 가격이 태블릿 컴퓨터를 통해 소매 판매점 및 서비스 업체로부터 직접 수집된다. 또한 4개의 주요한 소매체인 업체는 식품 및 유사 식품 가격 정보를 스캐너 데이터의 형태로 제공한다. 본부에서 가격 책정 정책을 수행하고 있는 일부 주요 소매 업체는 인터넷 등을 통해 해당 가격을 직접 FSO에 제공하고 있다. 이 외에도 인터넷, 이메일, 인쇄된 양식, 전화 등으로 가격이 조사되고 있다.

각 지역 및 소매 판매점 유형에 따라서 별도의 지수가 계산되며 각 지수는 현재와 기준 시점의 가격 비율에 대한 기하 평균으로 작성된다. 전국 소비자물가지수는 전국의 11개 지역 및 소매 판매점 유형별 지수에 대한 가중 산술 평균으로 구해진다.

## 나. 네덜란드

### 1) 대상 범위

일반적으로 네덜란드 소비자 물가 지수의 소비 지출 분류는 ECOICOP (European COICOP) 기준이지만 국제 표준으로 규정되지 않는 두 개의 분류가 추가된다. 이는 오염세, 자동차세 및 강아지세 등 구입이 아니라 소유에 따른 세금이며 다른 하나는 네덜란드 거주자가 해외에서 소비지출한 항목 두 가지이다. 두 개의 추가적인 항목은 전체 CPI의 약 8%를 점유하고 있다.

### 2) 가격의 집계 및 지수 산정

가격 자료는 월 단위로 수집되는 반면 연료 가격은 매일 수집된다. 거래 가격 정보와 관련하여 정가(list) 가격 및 스캐너 데이터가 일부 소매점 및 본부 부서로부터 제공된다. 할인 가격은 특별한 차별이 없고 별다른 어려움 없이 획득될 수 있는 경우에만 해당 거래 가격으로 기록된다.

가격 조사의 체계와 관련하여 대다수의 상품과 서비스가 광범위하게 포함된다. 슈퍼마켓의 경우 거의 모든 가격이 스캐너 데이터를 통해 수집되며, 향후 점점 더 스캐너 데이터와 같은 등록 데이터(register data)의 비중이 높아지고 있으며 면담자(interviewer)를 통한 가격 수집의 비중은 지속적으로 축소되고 있다. 소매점의 표본은 소매점 유형에 따라 분류된다. 즉 지역 데이터 수집은 스캐너 데이터, 등록 데이터 수집 및 인터넷 로봇으로 점점 더 많이 대체되고 있다. 스캐너 데이터 및 인터넷 로봇 데이터를 제외한 약 1,100종의 제품이 지역 CPI를 편제를 위한 자료에 포함된다. 대부분의 제품에 대해 소매점별로 가격 정보가 분류되며, 제품 및 소매점별 가격 지수는 내부 편제에만 이용되며 별도로 공표되지는 않는다.

슈퍼마켓 체인점의 가격은 관련 상점의 모든 거래와 관련된 스캐너 자료를 이용하여 수집한다. 이러한 스캐너 데이터에는 유럽 표준 번호(European Article Number EAN)에 의해 지정된 품목별 판매량 및 수량에 대한 데이터가 포함되어 있다. 또한 항목 유형을 식별하는 문자 필드와 소매점 단계의 집계 코드가 포함될 수 있다. 앞서 언급되었듯이 CPI 집계에서 스캐너 데이터 사용이 확대되고 있으며 2016년 기준 CPI 편제와 관련한 모든 슈퍼마켓 거래 기록은 스캐너 데이터를 이용하고 있다.

기초 지수를 상위 수준으로 집계하는 방법은 국제적으로 허용되는 라스파

이레스 수식을 이용한다. 지수 산출시 적용되는 지출 가중치는 매년 업데이트(영 지수, Young index)되며 고정된 값이 적용된다.

## 다. 노르웨이

### 1) 대상 범위

소비자물가지수 대상 품목은 약 800개의 서비스 및 제품군이 있으며 해당 표본은 가구 서베이 조사 및 지역본부의 정보, 시장조사 자료 등의 정보를 기초로 설계된다. 또한 소비자물가지수는 가게에 제공되는 물품 및 서비스의 수수료, 수수료 및 보조금을 포함하여 실제 판매 가격을 측정한다.

### 2) 가격의 집계 및 지수 산정

소비자물가 품목의 가격 정보는 설문 조사를 통해 주로 수집되고 있지만 스캐너 데이터를 이용하는 비중을 지속적으로 늘려가고 있다. 현재 노르웨이 통계청은 식료품점, 의약품점 및 주유소로부터 매월 스캐너 데이터를 제공받고 있는데 여기서 약 14,000개의 상품에 대한 스캐너 데이터가 식품 및 비주류 음료에 대한 하위 물가지수 산정에 이용된다.

체인 소매점은 노르웨이 통계청의 비즈니스 등록부에서 선택하는데, 산업 및 지역별로 인구를 계층화한 후에 이루어진다. 선택 될 확률은 상점의 재고 회전율(turnover)의 크기에 비례하며 조사대상 아울렛 중 1/6이 매년 교체된다. 이는 각 아울렛이 평균 6년 동안 조사대상으로 남아 있음을 의미한다. 가장 하위 수준의 가격 지수는 비가중 기하 평균을 사용하여 지역 차원에서 각 항목에 대해서 계산된다. 지역 가중치는 각 상품의 국가지수들로 지역 지수들을 결합하는 데 사용된다. 노르웨이 소비자 물가 지수는 모든 대표 품목군에 대해 라스페이레스 산식으로 집계되며 매년 1월에 가중치가 바뀌는 연쇄 방식으로 작성되고 있다.

## 라. 벨기에

### 1) 대상 범위

소비자물가지수 대상 품목의 경우 가게에 의해 소비되는 모든 재화 및 서

비스를 포함하며 ECOICOP (European COICOP)의 분류체계에 따라서 5 레벨에는 227개의 상품군, 4 레벨에는 98 상품군, 3 레벨에는 40 상품군 그리고 2 레벨에는 12개의 상품군이 각각 포함되어 있다. 지수는 매월 셋째 주까지의 조사된 후에 마지막 주에 공표되고 있다.

## 2) 가격의 집계 및 지수 산정

약 60%의 가격 정보는 지역에 기반한 조사를 통해서 수집되며 나머지 40%는 중앙 기반의 조사를 통해서 수집된다. 또한 슈퍼마켓의 스캐너 데이터 또한 가격 수집이 이용되고 있다. 새로운 상품의 도입은 매년 가능하며 가중치에 대한 업데이트는 매 2년마다 이루어지고 있다.

가장 하위의 가격지수 산정에는 제본스 지수가 활용되고 있으며 가장 상위의 소비자물가지수는 라스페이레스 산식이 적용되고 있다. 중간 단계의 각 상품지수는 비가중 기하평균이 적용되며, 서비스의 경우 상위 단계에서는 지역별 가중치에 따라 집계되고 있다.

## 마. 스웨덴

### 1) 대상 범위

소비자물가지수 대상 품목의 경우 시설수용 가계(institutional households) 및 외국의 방문객을 포함하여 가계에 의해 소비되는 모든 재화 및 서비스가 포함된다. 다만, 적절한 측정 방법이 개발되지 않은 아이돌보미, 노인돌보미, 생명보험 수수료 등의 일부 서비스 항목은 지수 산정에서 제외된다. COICOP 분류체계에 따라서 1,212개의 주요 품목군과 그 아래에 90여개의 하위 품목군을 두고 있다. 지수 산정을 위한 해당 정보는 매월 15일 전후의 셋째 주까지 통계조사원 또는 스캐너 데이터를 통해서 수집된다.

### 2) 가격의 집계 및 지수 산정

매년 5월에 제품에 대한 샘플링 작업을 수행하며 평균 20% 정도의 제품이 교체된다. 그리고 10% 정도의 제품은 모집단의 변경으로 또한 교체됨에 따라 약 70%의 제품만이 남아 있게 된다. 포장음식, 세탁세제 및 기타 일상용품은 3개의 소매점 그룹의 스캐너 데이터를 이용하여 가격 정보가 수집되고



있다. 즉, 80개의 소매점 체인으로부터 90개의 품목군을 구성하는 800개의 대표 제품에 스캐너 데이터를 수집하고 있다. 800개의 대표 제품 구성을 위한 샘플링은 해당 소매점의 스캐너 데이터(cash register data)에 기초하고 있다. 스웨덴은 2005년 이후 최상위 지수의 하나인 Walsh 지수기법을 적용하여 전국 소비자물가지수를 산정하고 있으며 기초 단계의 하위 지수 산정에는 제본스 지수를 이용하고 있다. Walsh 지수는 기준 시점 및 해당 시점의 가중치를 산정할 때 두 기간 가중치의 기하평균을 적용하여 물가지수를 산정하는 기법이다. 한편 스웨덴의 소비자물가지수는 매년 이러한 Walsh 산식에 따라서 가중치가 변경되는 연쇄지수 기법을 적용하고 있다.

## 바. 덴마크

### 1) 대상 범위

소비자물가지수 대상 품목은 COICOP의 분류체계에 따라서 매월 1,500개의 소매점으로부터 약 25,000개의 가격이 수집되고 있다. 스웨덴과 마찬가지로 시설수용 가계(institutional households) 및 외국의 방문객을 포함하여 가계에 의해 소비되는 모든 재화 및 서비스가 대상 품목이 된다. 다만 생명보험 및 도박 등은 제외된다. 수집 가격은 세금이 포함되고 보조금 등으로 차감된 금액이 포함된 실제 소비자가 지급하는 것을 대상으로 한다.

### 2) 가격의 집계 및 지수 산정

가격의 수집은 방문 조사, 설문 조사 및 스캐너 데이터를 통해서 이루어진다. 이 중에서 음식료품, 주류, 및 담배는 슈퍼마켓의 스캐너 데이터를 통해서 입수되고 있으며 해당 품목군의 약 20% 정도의 비중을 차지하고 있다. 가중치는 매년 가구 서베이 조사결과를 기초로 업데이트되고 있다. 가장 기초 단계의 물가지수(Elementary Aggregate)는 월 연쇄지수의 형태로 기하평균 기법을 적용하여 산정하고 있다. 기초 단계 물가지수는 라스파이레스 산식을 이용하여 상위 및 최종적인 물가지수 산정에 합산되고 있다.

#### 4. 소결

본 장에서는 우리나라 소비자물가지수 편제의 문제점, 물가지수 편제 이론 및 스캐너 데이터를 활용하고 있는 주요 국가의 소비자물가지수 편제 현황을 대략적으로 살펴보았다.

우선 우리나라 소비자물가지수 편제는 품목과 가중치가 기준시점에 고정되어 있어 라스파이레스 산식을 적용하고 있다. 이는 지수 계산이 간편하다는 장점이 있는 반면, 기준시점에서 멀어질수록 소비구조의 변화가 물가지수에 반영되지 못함에 따라 물가지수에 상향 편의가 생기는 문제를 가지고 있다. 다만 우리나라에서는 소비자물가지수의 현실 반영도를 제고하기 위해 매 5년에서 매 2년~3년으로 조사대상 품목 및 가중치를 재조정하고 있으며 연쇄지수 또한 작성하고 있다.

최근 유럽의 여러 나라에서는 소매점을 통한 스캐너 데이터를 소비자물가지수 산정에 활용하고 있다. 이를 통해 가격과 판매량에 관한 대용량의 실시간 판매 정보 획득이 용이해졌기 때문이다. 또한 이러한 실시간 가격 및 수량 정보는 기간별로 변동하는 가중치를 적용하여 개별 품목 단위에서는 라스파이레스 산식에 얽매이지는 않는 최상위 지수(superlative index)에 가까운 지수 편제가 가능할 수 있다. 이에 따라 이어지는 본 섹션에서는 다양한 소비자물가지수 편제 이론을 정리해 보았다.

우선 라스파이레스, 로우, 파쎌, 피셔 및 톱크비스트 지수 등에 대해 직접 지수와 연쇄지수를 각각 살펴보았다. 또한 가격과 수량 정보가 실시간으로 포착됨에 따라 가장 이상적인 최상위 지수로 알려져 있는 피셔 및 톱크비스트 지수를 살펴보았다. 한편 스캐너 데이터와 같이 소모율(attrition rate)이 높은 자료를 대상으로 물가지수를 산출할 때는 직접지수 대신 연쇄지수가 이용되는데 가격 및 수량 변동성이 클 경우 연쇄편의를 유발할 수 있다는 문제가 있다. 통상 연쇄편의는 모든 제품들의 가격이 기준시점 수준으로 회귀하더라도 해당 품목의 물가지수가 기준시점의 값인 1(unity)로 돌아오지 않는 현상을 의미한다.

이에 따라 가격과 물량이라는 정보를 모두 활용하는 최상위 지수이면서도 연쇄편의 문제를 완화한 GEKS 및 RWGEKS 두 지수를 살펴보았다. 특히 RWGEKS 지수의 경우 연쇄편의 문제를 거의 해소하는 가운데 매기 새로운 정보가 업데이트 되더라도 기존의 물가지수 값을 변동시키지 않기 때문에 통계당국에는 보다 매력적인 지수로 활용될 수 있다.

마지막으로 소비자물가 편제시에 스캐너 데이터를 활용하고 있는 유럽의

여러 나라에서 실제로 지수 편제가 어떻게 이루어지는지 살펴보았다. 스캐너 데이터는 현재로서는 설문조사 방식으로 수집되는 가격 정보에 비해서는 여전히 상대적으로 작은 비중을 차지하고 있지만 이들 나라들은 대체로 향후 그 활용 비중을 높여나갈 계획이다. 현재 이들 나라에서 스캐너 데이터는 개별 상품의 가격 정보를 중심으로 활용되고 있으며 상위의 품목군 지수로 집계되는 경우에는 고정된 가중치를 적용하는 라스파이레스 산식이 적용되고 있다. 다만 가중치의 재조정 주기는 주로 1~2년으로 상대적으로 빠른 편이라고 판단된다.

### Ⅲ. 스캐너 데이터를 이용한 기존 연구 및 사례

#### 1. 스캐너 데이터 활용의 장단점

##### 제1절 장점

스캐너 데이터를 소비자물가지수 작성에 활용하는 방안에 대한 오랜 연구를 진행해 온 서구 유럽 국가의 경우 대체로 다음과 같이 스캐너 데이터의 주요 이점을 제시하였다. 첫째, 스캐너 데이터는 기준 시점과 비교 시점 모두에 관한 가격과 수량 정보를 제공함에 따라 동 자료를 이용하는 경우 최상위지수의 작성이 보다 용이해진다. 이는 스캐너 데이터를 활용함으로써 소비자가 재화의 구입을 위해 실제로 지불한 금액에 대해 정확한 정보를 확보함에 따라 보다 정확한 물가지수의 작성이 가능하다는 것을 시사한다. 둘째, 스캐너 데이터는 기존의 표본조사 방식에 따른 가격조사 자료에 비해 보다 광범위한 제품에 관한 판매액 정보를 포함하고 있다. 모집단에 관한 전수조사에 가까운 정보로 실제 표본의 크기를 크게 증대시키는 효과로 일부 제품을 표본으로 추출하여 조사하는데 따른 표본오차를 크게 축소할 수 있게 된다. 마지막으로, 스캐너 데이터에 포함된 판매 수량 정보를 통해 새로운 제품에 대한 파악 및 모니터링이 용이함에 따라 품목별 구성비에 대한 정보 획득에 용이하다는 점이다.

##### 제2절 도전과제

###### 가. 비통계적인 부문

이처럼 다양한 활용 가능성을 가지고 있는 스캐너 데이터에 대해서 유럽 여러 나라의 기존 연구에 따르면 대체로 다음과 같은 비통계적 및 통계적인 도전 과제가 제시되고 있다. 비통계적인 부문에서는 스캐너 자료의 IT 기술 의존성과 스캐너 자료 제공자에 대한 의존성이라는 도전과제가 있다. 스캐너 자료는 소매점이 자사의 물품판매와 마케팅을 위해 개발한 각자의 개별적인 시스템에 의존함에 따라 지수 구축을 위해서는 각각의 시스템을 유효하게 통합할 수 있는 IT 기술을 적시에 개발하는 것이 필요하다는 점이다. 또한 자료에 대한 안정적인 접근과 확보를 위해서는 소매 판매점과의 원활한 업무협조(또는 협약)가 필수적이다.

## 나. 통계적인 부문

통계적인 부문을 보면 스캐너 코드매칭 문제가 있다. 유통점에서는 동일한 제품임에도 스캐너 코드가 시간에 걸쳐 변화하거나, 동일한 스캐너 코드에 판매점에서 소멸한 제품을 대체하여 새로운 제품에 부여되는 경우가 있다. 스캐너 코드 변화를 정확히 식별하는 것은 상당히 어려운 작업으로서 유통점과의 긴밀한 업무협력이 요구된다. 또한 스캐너 데이터의 경우 방법론적인 문제로 인해 물가지수 구축시 하위 합산 단계(elementary level)에서만 제한적으로만 이용되고 있으며 보다 상위 단계에서는 고정 바스켓(fixed basket, 또는Lowe) 방식으로 합산되고 있다. 기존의 라스파이레스 산식의 경우 산출인플레이션 분포의 상단으로 도출되는 것(상향편의)과 반대로, 최상위지수 구축시 인플레이션 수준이 기조적으로 저하(하향편의)될 가능성이 있다.

앞서 살펴보았듯이 최상위지수 구축시 연쇄편의 문제에 노출될 수 있으며 스캐너 데이터는 제품 스펙(item specification)에 대한 정보가 부족하기 때문에 품질조정이 어려울 가능성도 존재한다. 현행 표본조사 방식에서는 업체 담당자와의 의견 교환 등을 통해 가격 변화분 중 제품스펙 변화에 따른 품질 변화분에 대한 추론이 일정 정도 가능하다. 평균 집계 시점 또한 매우 민감한 문제인데 스캐너 자료의 경우 주, 월, 분기 시점 등 어느 시점에서 평균을 산출하느냐에 매우 민감할 가능성이 있다.

## 다. 물가지수와 관련한 스캐너 데이터 활용

기존 연구에 의하면 스캐너 데이터를 집계할 때 주 단위 보다는 월 또는 분기 기준의 집계 방식이 보다 선호된다. 물가지수 통계와 관련한 스캐너 데이터의 활용 가능성은 다음과 같이 4가지로 정리될 수 있다. 첫째, 기존의 표본조사 방식으로 수집된 자료를 대체하며, 스캐너 데이터를 이용해 초기 단계의 합산 과정을 통해 제본스 지수 등을 구축하는 경우 적은 비용으로 표본오차를 크게 줄일 수가 있다. 또한 모든 제품에 대한 스캐너 데이터가 활용 가능하다면 소비자물가지수에 대해 피셔지수와 같은 최상위지수로 구축하는 것도 가능하다. 단, 유통점은 통상 가격은 유지한 채 제품 부피를 줄여서 실질적인 가격 상승을 유도하려는 경향 등 통계당국이 유통점의 제품 단위 변경에 대해 적절한 대응을 취하는데 제약이 있다는 점은 염두에 둘 필요가 있다. 마지막으로 또 다른 정보원으로서 스캐너 데이터를 활용할 수가 있는데 스캐너 데이터를 이용한 지수와 기존 방식의 지수의 차이를 비교

분석하는 것이다. 스캐너 데이터를 물가지수 작성에 활용하지 않고 기존 지수의 품질관리와 감사를 위해 보조적으로 활용하는 것이다.

## 2. 국내 기존 연구 및 적용 사례

### 제1절 서론

본 절에서는 스캐너 데이터를 이용해서 우리나라의 물가지수 산정한 김기석 (2012) 및 강규호 외 (2015)의 논문을 살펴본다. 두 논문 모두 스캐너 데이터를 이용해서 다양한 물가지수를 편제하고 기존의 방식으로 편제되는 지수와 비교분석하였다. 또한 스캐너 데이터의 장점인 가격과 거래량 정보를 활용하여 최상위 지수 및 연쇄편의 문제에 강점이 있는 GEKS 및 RWGEKS 지수 등을 편제하였다.

### 제2절 음식료 품목군 적용 사례

#### 가. 개관

김기석 (2012)은 21개 음식료 품목군을 중심으로 한 스캐너 데이터를 이용하여 품목군별 물가지수를 작성하고 라스파이레스, 파쇄 및 피셔 지수 값을 비교분석하였다. 동 스캐너 데이터에는 2000년 1주~2012년 13주에 걸쳐서 19개 백화점, 16개 대형 할인점, 82개 편의점 및 39개 슈퍼마켓 등 총 156개 점포에서 수집된 131백만 건이 넘는 관측치가 포함되는데 품목군에 포함되는 품목수는 비스킷 품목군의 2,002개 품목수에서 밀가루 품목군의 67개 품목수에 이르기까지 모두 9,974개의 품목수가 포함된다. 21개 각 품목군별 라스파이레스, 파쇄 및 피셔 지수를 각각 작성하였으며 직접지수, 고정 바스켓 (fixed basket) 연쇄지수 및 변동 바스켓 (flexible basket) 연쇄지수를 각각 추정하였다. 이와 함께 월별 및 분기별 시간 종합화 및 점포 종합화 적용/비적용 등 다양한 종합화 수단을 적용하였으며 산정된 21개 품목의 각 지수는 기하평균으로 합산되었다.

## 나. 지수편제 결과

### 1) 기존 지수

지수 추정 결과를 살펴보면 우선 월별 및 분기별 시간 종합화 방법에 따른 차이는 크지 않은 것으로 나타났는데 다만 점포 종합화 적용/비적용 여부와 상관없이 세 지수에서 모두 분기별 종합화한 기하평균이 월별 보다 추세적으로 높게 나타났다.

<표 III-1> 21개 품목군의 기하평균 - 점포 종합화 적용

각 품목군별 산정 방식	직접지수		연쇄지수 (Fixed basket)		연쇄지수 (Flexible basket)	
	분기	월	분기	월	분기	월
라스파이레스	113.78	112.31	113.80	112.41	102.39	100.93
파쉐	106.79	105.81	106.85	105.81	101.95	98.99
피셔	110.23	109.01	110.27	109.06	102.17	99.96

출처 : 김기석 (2012)

<표 III-2> 21개 품목군의 기하평균 - 점포 종합화 비적용

각 품목군별 산정 방식	직접지수		연쇄지수 (Fixed basket)		연쇄지수 (Flexible basket)	
	분기	월	분기	월	분기	월
라스파이레스	113.72	112.89	113.64	112.53	102.77	101.51
파쉐	107.75	105.22	107.93	106.89	98.93	98.74
피셔	110.63	108.99	110.75	109.67	100.83	100.12

출처 : 김기석 (2012)

한편 ILO (2004)의 소비자물가지수 작성편람을 보면 피셔지수를 최상위 지수로 판단할 때 라스파이레스지수는 상향편의 그리고 파쉐지수는 하향편의를 보인다고 정리하고 있다. 이러한 라스파이레스지수의 상향편의 크기는 스캐너 데이터를 이용한 호주의 연구 (IDF, 2011)에서 나타난 상향편의보다 다소 크게 나타났다. 다만 분기별 지수와 월별 지수 간의 차이는 김기석 (2012)의 추정 결과가 호주의 결과보다 다소 작게 나타났다. 한편 동일한 품

목에 대한 점포 종합화의 적용 여부는 지수 작성 결과에 별다른 영향을 미치지 않는 것으로 나타났다.

마지막으로 고정 바스켓 연쇄지수와 변동 바스켓 연쇄지수를 비교해 보면, 대체로 고정 바스켓 연쇄지수가 라스파이레스 지수와 유사한 것으로 나타난 반면, 변동 바스켓 연쇄지수의 경우 세 가지 지수 산정에서 모두 해당 지수 값이 낮아지는 것으로 나타났다. 최상위 지수인 피셔지수의 경우 고정 및 변동 바스켓 연쇄지수 간 차이가 여타 지수에 비해 상대적으로 작은 것으로 나타났으며 이와 같은 결과는 대체로 호주의 사례 연구 (IDF, 2011) 결과와 일치하는 것이다.

## 2) GEKS 및 RWGEKS 지수

스캐너 데이터를 이용하여 지수를 작성할 때 Rolling Window 길이는 계절성 등을 고려하여 13개월로 정하여 연쇄편의로부터 자유로운 GEKS지수 및 RWGEKS지수 등을 산정하고 있다. 우선 연쇄편의의 크기를 확인하기 위해 피셔지수와 비교해 본 결과, 두 GEKS지수에 비해 최상위지수인 피셔지수에서 상당한 크기의 하향 편의가 발생한다.

한편 GEKS지수와 13개월의 Rolling Window를 적용한 RWGEKS 지수 간에는 큰 차이가 없는 것으로 나타났으며 이는 스캐너 데이터에 의한 연쇄지수 작성에서 발생하는 연쇄편의 현상이 주로 가격할인 및 그에 따른 물량 증가에 기인하며, 이와 같은 가격할인 판매가 주로 매년 같은 기간에 반복적으로 나타나는 계절성을 가지고 있기 때문으로 분석된다. 이와 같은 13개월 Rolling Window의 적절성은 Haan and Grient (2011)의 연구 결과와도 일치한다.

<표 III-3> GEKS, RWGEKS 및 피셔 연쇄(변동 바스켓) 지수  
(21개 품목군의 기하평균)

	점포 종합화 적용			점포 종합화 비적용		
	GEKS	RWGEKS	피셔	GEKS	RWGEKS	피셔
<b>월간</b>	109.14	108.48	101.25	109.64	108.63	101.73
<b>분기</b>	110.62	110.78	102.17	110.70	110.72	100.83

출처 : 김기석 (2012)



한편 비교가 가능한 18개 품목군에 대해 GEKS지수와 상응하는 통계청 품목별 지수를 비교해 본 결과, 월별 기준으로 통계청 해당 지수에서 대체로 5% 내외의 상향편의가 발생하는 것으로 나타났다. 분기 기준으로도 통계청 해당 지수가 대략 4% 정도의 상향편의를 갖는 것으로 나타났다.

<표 III-4> 서울의 통계청 소비자 물가지수 및 GEKS지수  
(18개 품목군의 기하평균)

	RWGEKS 지수		Official CPI figures, Seoul
	점포 종합화 적용	점포 종합화 미적용	
월간	109.45	109.85	114.62
분기	111.00	110.91	115.15

출처 : 김기석 (2012)

### 제3절 냉장고 수출 품목군 적용 사례

강규호 외(2015)는 스캐너 데이터에 해당하는 관세청의 냉장고 수출 품목을 대상으로 연쇄지수를 산출하고 연쇄편의 정도를 측정해 보았는데 측정 결과 RWGEKS지수의 연쇄편의 정도가 GEKS지수에 비해 크지 않아 적합한 물가지수 산출방법으로 추천되었다. 즉, 아래 <표 III-5>를 보면, 품목내 제품들의 가격이 모두 기준시점 수준으로 회귀함에 따라 직접지수의 경우 어떤 산식을 적용하더라도 물가지수가 모두 원래수준(1.0)으로 회귀한 것을 확인할 수 있었다. 한편 <표 III-6>에서 볼 수 있듯이 라스파이레스, 파세 및 통크비스트 연쇄지수는 모두 연쇄편의를 유발하나 RWGEKS지수의 연쇄편의 정도는 상대적으로 매우 미미한 것으로 나타났다.

<표 III-5> 물가지수 시물레이션 결과 : 직접지수와 GEKS지수

시점	라스파이레스	파세	통크비스크	GEKS
0	1.0000	1.0000	1.0000	1.0000
1	0.6625	0.5420	0.5971	0.7253
2	0.8815	0.8738	0.8776	0.9029
3	0.9964	0.9826	0.9895	0.9855
4	1.0000	1.0000	1.0000	1.0000

출처 : 강규호 외 (2015)

<표 III-6> 물가지수 시물레이션 결과 : 연쇄지수, GEKS지수 및 RWGEKS지수

시점	라스파이레스	파세	통크비스크	GEKS	RWGEKS
0	1.0000	1.0000	1.0000	1.0000	1.0000
1	0.6625	0.5420	0.5971	0.7253	0.7541
2	1.0856	0.5900	0.7898	0.9029	0.9080
3	1.1973	0.6601	0.8774	0.9855	0.9847
4	1.2185	0.6625	0.8868	1.0000	0.9992

출처 : 강규호 외 (2015)

한편 동 논문에서는 해당 스캐너 데이터를 이용하여 de Haan and Krsinich (2014)가 제안한 TPD(Time Product Dummy)기법을 활용하여 품질조정 물가지수를 추정했다. 그러나 동 추정치 불확실성이 너무 크게 나타나는 등 품질 특성치에 대한 정보가 충분치 않은 스캐너 데이터의 경우 적절한 품질조정을 실시하는 데 현실적인 한계가 있는 것으로 나타났다.

### 3. 해외 기존 연구 및 적용 사례

#### 제1절 서론

스위스, 노르웨이, 네덜란드 및 뉴질랜드 등에서는 실제 스캐너 데이터를 통해 입수된 가격 정보 등을 소비자물가지수 작성 시 일부 하위 품목의 편성에 활용하고 있다. 또한 영국, 프랑스, 스웨덴 등의 유럽 국가들은 스캐너 데이터를 활용하기 위한 방안에 대한 연구를 꾸준히 수행해 오고 있는 중이다. 본 절에서는 유럽의 해당 나라별로 스캐너 데이터에 대한 연구결과 및 실제 활용 사례 등을 정리하였다.

#### 제2절 CPI에 스캐너 데이터 활용 국가

##### 가. 스위스

##### 1) 적용 품목 및 비중

Muller(2010)에 따르면 스위스 통계청은 2008년 7월부터 물가지수 구축 시 음식료품(food) 및 유사 음식료품(near-food) 품목에 대해서 스캐너 데이터

를 이용한 품목별 지수를 산정하기 시작하였다. 스위스 소비자물가지수는 매년 가중치가 업데이트되는 로우 산식을 적용하고 있으며 최하위 단계에서는 제본스 지수를 적용하여 하위 품목의 지수를 작성한다. 한편 지역별 가격체계(regional price structure)에 포함되어 있는 상품의 가격 정보는 독립적인 시장조사 기관으로부터 구매하고 있는 반면 국가 가격체계(nationwide price structure)에 포함되어 있는 상품의 가격 정보는 통계청에 의해서 수집되고 있다. 현재 음식료품 및 유사 음식료품의 품목별 지수는 스캐너 데이터를 이용하여 작성되고 있으며 동 스캐너 데이터는 음식료품과 관련하여 스위스에서 60~70%의 시장점유율을 가지고 있는 두 개의 가장 큰 소매판매 체인점을 통해서 확보되고 있다.

## 2) 지수편제 결과 활용

현 단계 스위스에서의 스캐너 데이터 활용은 음식료품 및 유사 음식료품 등 제한적으로 두 품목별 지수작성에 필요한 가격 정보를 확보하는데 이용되고 있다. 동 품목별 지수의 작성은 기존의 표본조사 자료의 지수 작성에 적용된 것과 동일한 방식, 즉 제본스 산식이 적용되고 있고 향후 점차적으로 품목과 소매점을 늘려가며 대안적인 지수 작성 및 소비자물가지수 개선 등을 위한 활용 방안에 대한 연구를 지속해나갈 계획이다.

## 나. 네덜란드

### 1) 적용 품목 및 비중

Chessa et al (2015) 및 Chessa (2016)에 따르면 네덜란드 통계청은 2002년 6월부터 슈퍼마켓의 스캐너 데이터를 소비자물가지수 작성에 활용해오고 있다. 네덜란드 통계청에서는 2013년 1월부터 10개의 모든 슈퍼마켓 체인점으로부터 스캐너 데이터를 제공받음에 따라 동 슈퍼마켓 체인에 대해 더 이상의 방문 및 설문 조사는 이루어지지 않고 있다. 이 외에도 여행사, 에너지 및 휴대폰 등의 품목에 대해서도 스캐너 데이터가 활용됨에 따라 현재 소비자물가지수 가중치의 20% 이상이 스캐너 데이터를 통한 가격 정보에 의존하고 있다. 아래 <표 III-7>에서는 COICOP 그룹별로 소비자물가지수 편제에 이용되는 스캐너 데이터 품목군 그룹을 보여주고 있다.

<표 III-7> 스캐너 데이터를 이용하는 COICOP 그룹

해당 코드	품목군
010000	식품 및 무알콜 음료
021200	와인
021300	맥주
055000	주거, 정원용 도구 및 장비
056000	일상적인 가정용 유지·보수용 상품 및 서비스
061000	의약품
093400	애완용품, 사료
131300	가전제품, 개인 위생용품, 생활 용품

출처 : Chessa et al (2015)

## 2) 지수편제 결과 및 활용

슈퍼마켓 스캐너 데이터에 대한 지수 작성 방법은 <표 III-8>에서 볼 수 있듯이 하위 품목에 대해서는 가격 정보만을 활용하는 제본스 지수를 이용해서 품목별 지수를 산출하고 있는 반면, 상위 단계에서는 라스파이레스 산식을 적용하고 있다. 한편 2016년 1월 이후 모든 가격 정보가 스캐너 데이터를 통해 입수되는 백화점 및 휴대폰 품목군에 대해서 라스파이레스 산식을 적용하기 전에 QU지수(Quality adjusted Unit value index) 기법이라는 새로운 지수편제 방식이 적용되고 있다.

<표 III-8> 네덜란드의 스캐너 데이터에 대한 지수 작성 방법

	Version 0	Version 1	Version 2
개발 및 이용 시기	1990후반	2002-2009	2010-현재
샘플	모든 데이터	바스켓(소매점 당 +-10,000개)	특정 필터를 만족하는 모든 EAN
동일 상품군 기초	EANs	EANs	EANs
상품 교체	없음	있음	없음
지수 방법론	피셔(월 기준 연쇄)	라스파이레스(연간 고정되는 가중치)	제본스지수(월 기준 연쇄, 허용된 EANs에 대해 동일한 가중치)
CPI 활용 여부	안함	활용	활용

출처 : Chessa et al (2015)

백화점 및 휴대폰 품목군의 스캐너 데이터를 통해 작성되는 QU지수는 식 (24)에서와 같이 정의될 수 있다. 만약  $v_i = p_{i,t}$ 이라면 모든 제품 i에 대해 라스파이레스 산식을 적용하는 것이 되며, 만약  $v_i = p_{i,0}$ 이라면 모든 제품 i에 대해 파쉐 산식을 적용하게 된다.

$$P_t = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t} / \sum_{i \in G_0} p_{i,0} q_{i,0}}{\sum_{i \in G_t} v_i q_{i,t} / \sum_{i \in G_0} v_i q_{i,0}} \quad (24)$$

0 = 기준 월

t = 해당 월

G = 동일 제품의 집합(set of consumption segment)

i = 동일 제품(homogeneous product)

한편 동일 제품의 집합인 G는 시간에 걸쳐서 변화할 수 있다. 만약 하나의 동일 제품군에 포함되는 모든 제품이 동일하다면 식 (24)는 단위값 지수(unit value index)를 의미하게 되며, 동일하지 않다면 해당 단위값 지수는 조정된다.  $v_i$  선택과 관련하여 통상 12월을 기준월로 통계 및 민감도 분석을 진행한 후 매년 변경될 수 있다. 식 (24)는 식 (25)과 같은 축약된 형태로 다시 표시될 수 있다.

$$P_t = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t} / \sum_{i \in G_0} p_{i,0} q_{i,0}}{\sum_{i \in G_t} v_i q_{i,t} / \sum_{i \in G_0} v_i q_{i,0}} = \frac{\overline{p_t / p_0}}{\overline{v_t / v_0}} \quad (25)$$

만약 높은 물량 가중치를 갖고 있는 제품의 소비가 증가하게 된다면, 이는 가격과 함께 물량 가중치가 동시에 증가하면서 물가지수를 크게 상승시킬 수 있다. 하지만 가중치가 높은 제품의 소비는 또한 식 (20)의 분모값을 증가시키게 된다. 이러한 조정 과정을 통해서 해당 물가지수에 보완적인 하향 안정 효과(complementary downward effect)가 나타날 수 있다.

네덜란드에서 스캐너 데이터는 제품의 재론칭(relaunch)에 따른 해당 정보의 왜곡을 완화해 주기 위해서 도입되었다. 어떤 제품이 재론칭되는 경우 실제적으로 질적으로 유사한 제품이지만 약간의 포장 및 디자인 재구성을 통해서 전혀 다른 EAN 코드가 부여되면서 높은 가격으로 판매가 된다. 이에

반해 기존의 유사한 제품은 재론칭 제품의 등장으로 세일 등을 통해 가격이 크게 하락한 상태에서 판매가 이루어지게 된다. 하지만 EAN 코드가 다리 적용된 스캐너 데이터 자료에서는 두 제품이 서로 다른 별개의 제품으로 인식되는 문제가 나타난다.

네덜란드 통계청은 향후 다양한 검정을 거친 후에 이러한 QU지수의 적용 범위를 보다 확대해나갈 계획이다.

#### 다. 노르웨이

##### 1) 적용 품목 및 비중

Johannessen (2016)에 따르면 노르웨이는 2001년부터 스캐너 데이터를 소비자물가지수 작성 시 공식적으로 활용하고 있다. 노르웨이의 스캐너 데이터 적용 품목 및 비중을 보면, 2001년 주류 자료를 이후 순차적으로 주유소, 약국 그리고 음식 및 비알콜 음료 관련 자료로 확대되어 유통업체와의 업무협약을 확대하고 있는 중이다.

##### 2) 지수편제 결과 및 활용

소비자물가지수 구축 시 스캐너 자료가 약 30% 정도의 비중으로 이용되고 있으며 노르웨이 시장 점유율의 96%를 차지하는 3개 대형마트의 200개 점포로부터 매주 70만 건의 스캐너 자료가 이용된다. 국가 독점의 주류체인은 100% 스캐너 데이터를 확보하고 있으며, 약국체인은 4개 중 3개에서 스캐너 데이터를 받고 있다. 주유체인점은 6개 중 4개에서 스캐너 데이터를 받고 있다. 노르웨이는 스캐너 데이터를 이용하여 점포 종합화를 수행하지는 않으나 월별 연쇄지수를 작성하고 있다.

#### 라. 벨기에

##### 1) 적용 품목 및 비중

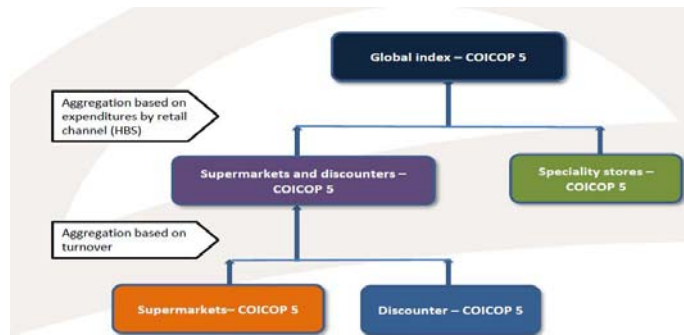
Valenduc and Van Loon (2015)에 따르면 벨기에 통계청은 2012년~2013년 중 소매업체와의 자료 접근에 대한 협상을 통해 2013년 말부터 시장 점유율 75%를 차지하는 3대 슈퍼마켓 체인으로부터 스캐너 데이터를 제공받기 시작했다. 2016년부터는 소비자물가지수 편제와 관련하여 동 스캐너 데이터

가 COICOP 5 단계 9개의 품목군에 이용되고 있다.

## 2) 지수편제 결과 및 활용

스캐너 데이터 및 전통적인 방식으로 수집된 물가 정보는 [그림 III-1]에서와 같이 항상 COICOP 5 단계에서 결합된다. 이러한 자료의 결합을 위해 층위 모형(stratification model)이 이용된다.

[그림 III-1] COICOP 5 단계에서의 지수 산정

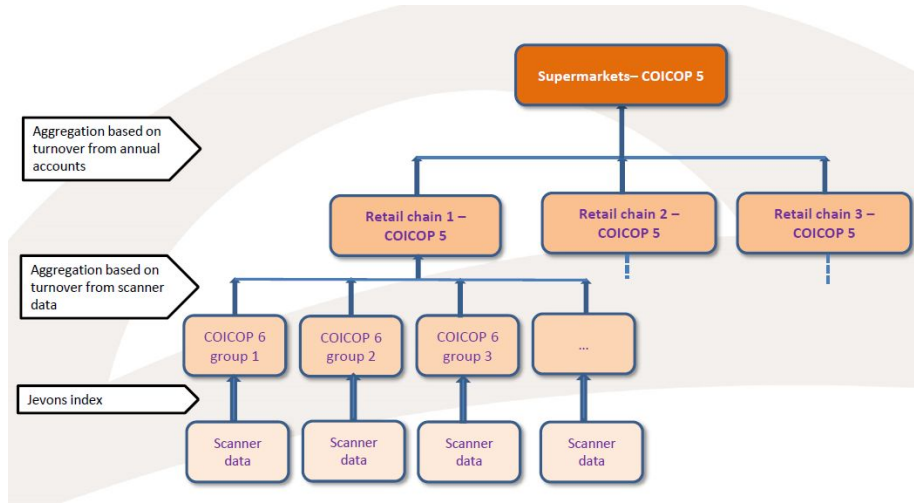


주: Valenduc and Van Loon(2015)에서 재인용

또한 [그림 III-2]에서 볼 수 있듯이 벨기에의 스캐너 데이터 집계 방식은 네덜란드 및 스위스의 방법론과 유사하다고 할 수 있다. 즉, COICOP 6 단계에서 월별로 연쇄 체본스지수를 작성한다. 그리고 아래와 같은 제품 필터링 기법을 적용하고 있다. 즉,  $\frac{s_m + s_{m+1}}{2} > \frac{1}{1.25 \times n}$ 인 경우 해당 제품이 지수 산정 샘플에 포함될 수 있다. 여기서 n은 제품의 수,  $s_m, s_{m+1}$ 은 해당 제품의 2개월에 걸친 시장 점유율을 의미하는 것으로서 일정한 시장 점유율을 하회하는 경우 제외된다. 또한 일시적으로 누락된 제품 및 덤핑 제품의 과도한 대표성을 통제하기 위해 가격 귀속계산(price imputation) 및 이상치 필터링 등을 적용하고 있다.

한편 제품 재론칭(relaunch)에 따른 가격 수준의 왜곡을 완화하기 위해서 샘플에서 제외되는 제품과 새로 샘플에 진입한 제품을 비교하여 실질적으로 동일한 제품으로 판정되면 두 제품은 하나의 동일한 제품으로 간주된다. 현재 이러한 작업은 수동적인 접근을 통해서 이루어지고 있으며 COICOP 1 단계 기준으로 EAN 코드의 상이함으로 서로 다른 제품으로 인식되었던 80개 정도의 제품이 다시 연결(linking)되고 있다.

[그림 III-2] COICOP 6 단계에서의 스캐너 데이터 집계



주: Valenduc and Van Loon(2015)에서 재인용

## 마. 스웨덴

### 1) 적용 품목 및 비중

스웨덴에서는 스캐너 데이터(cash register data)를 2012년부터 생필품(daily necessities) 가격 지수를 작성하는데 활용하여 오다가 최근에는 소비자물가지수 편제에도 활용하고 있다. 스웨덴 통계청은 식료품, 맥주 및 담배, 세제 및 개인 위생품 등에 대한 스캐너 데이터를 3개 소매점 80개 체인점으로부터 제공받고 있다. 각 체인점은 800개의 대표 제품을 선정하고 동 제품을 기초로 90개의 제품군이 생성된다. 대표 제품은 총판매액에 비례한 확률에 기초하여 임의 추출(random sampling) 과정을 통해 선발된다.

이러한 스캐너 데이터 관련 품목의 비중은 소비자물가지수 구성중 약 14%를 차지한다. 샘플링을 통해 제품과 매장이 선택되며 해당 스캐너 데이터는 주 단위로 취합된다. 주 단위의 가격이 해당 매장별 각 제품의 월 가격으로 취합되는데 각 제품의 가격을 취합할 때 해당 제품 판매가격에 더해져 있는 보증금(deposit)은 제외된다.

### 2) 지수편제 결과 및 활용

Sammar et al (2013)는 하나의 아울렛 체인점을 대상으로 소비자물가지수



의 품목과 관련하여 스캐너 데이터와 기존의 설문조사 방식으로 수집된 물가 자료를 비교한 결과 85%에서는 차이가 없는 것으로 나타났다. 15%의 차이를 분석해 보면 7%~9%는 두 물가 자료 간 알 수 없는 이유로 차이가 발생했으며 약 5%는 해당 품목이 주중에 판매되지 않았기 때문에 발생하였다.

한편 스캐너 코드의 변화를 살펴보면 2009년에 존재했던 EAN 코드(생필품 포함 전체 스캐너 데이터 기준) 중 약 70% 정도가 2010년 말에 남아있었다. 어떤 제품은 기존 제품과 비교하여 전혀 바뀌지 않았는데도 생산자에 의해 새로운 EAN 코드가 부여될 수 있으므로 스캐너 코드의 대체율을 낮추기 위해서는 지속적인 모니터링이 필요하다. 2012년에는 스캐너 데이터를 이용한 생필품 지수의 경우 538개의 코드 중 35개(6.5%)의 코드를 수정하였다. 스캐너 데이터와 관련된 또 다른 논의사항은 할인판매(discounted sales)에 대한 판단이다. 한 달에 3주씩 39주에 걸쳐 주당 평균 8,000거래를 포함하고 있는 스캐너 데이터에 대해 가격을 분석한 결과 이 중 49.2%가 가격할인을 포함하고 있는 것으로 나타났다.

## 바. 뉴질랜드

### 1) 적용품목 및 비중

2014년 3/4분기부터 소매업 거래 데이터 또는 스캐너 데이터를 CPI에 통합하여 소비자 전자제품 품목의 가격 변화를 측정하였다. 뉴질랜드에서는 소비자물가지수가 분기마다 공표되며, 스캐너 데이터를 사용하는 가전제품 품목은 냉난방장치, 데스크탑 컴퓨터, 노트북 컴퓨터, 태블릿 컴퓨터, 다기능 장치(multi-function devices), 휴대폰 단말기, 디지털 카메라, 디지털 카메라 메모리 카드, TV 셋톱박스, TV, DVD, 블루레이 플레이어, 홈시어터, 오디오 시스템 등이 있다.

스캐너 데이터가 지출액 가중치 측면에서의 품목별 기여도는 다음과 같다. 냉난방장치의 경우 2014년 6월 기준 스캐너 데이터가 주요 가전제품 품목군의 20%(동 품목 군이 소비자물가지수에서 차지하는 가중치는 0.71%)를, 휴대폰 단말기의 경우 통신 장비 품목군의 90% 이상(동 품목 군이 소비자물가지수에서 차지하는 가중치는 0.29%)을, 시청각 및 컴퓨터 장비 품목군은 80%(소비자물가지수에서 차지하는 가중치는 1.16%)를 차지한다.

## 2) 지수편제 결과 및 활용

뉴질랜드는 스캐너 데이터를 이용하여 해당 품목의 물가지수를 산출하기 위해 일종의 Rolling Window GEKS지수인 ITRYGEKS(통크비스트 회계연도 GEKS)지수를 작성하고 있다. 가전제품은 수명주기가 짧을 수 있기 때문에 새로운 제품이 지수에 적절히 반영되는 것이 중요하다.

### 제3절 스캐너 데이터 활용 방안 연구중인 국가

#### 가. 영국

##### 1) 적용 품목 및 비중

현재 영국에서는 소비자물가지수 작성에 스캐너 데이터를 공식적으로 활용하고 있지는 않지만 활용 방안 등에 대한 연구는 꾸준히 진행해오고 있다. 영국의 스캐너 데이터 적용 품목 및 비중은 Bird et al (2014)에 의하면 한 개의 유통체인점을 대상으로 2012년 1년 간 치약과 샴푸에 대한 스캐너 데이터를 활용하여 스캐너 지수와 기존의 해당 품목 물가지수를 비교·분석하였다. 동 연구에서 스캐너 데이터는 해당 품목을 100% 대체하고 있으며 스캐너 데이터에는 동일한 제품이더라도 정규 가격 및 다양한 이유로 할인 판매된 제품의 가격과 수량이 모두 기록되어 있다. 각 제품별 단위 가격 산정시 판매량 비중에 따라서 가중평균을 적용하여 산출한다.

##### 2) 지수편제 결과

샴푸와 치약에 대해 스캐너 데이터를 이용한 제본스 지수와 기존 CPI 상의 품목 지수를 비교해보면 제본스 지수가 기존 CPI 상의 품목 지수보다 수준이 20~30% 낮게 나타났으며 변동성은 크게 나타났다. 이는 스캐너 데이터가 한 개의 유통체인점 자료를 이용한 반면 CPI 품목 지수는 전국의 유통체인점 자료를 이용했다는 점 등 다양한 이유에 기인하는 것으로 나타났다. 한편 CPI 품목 지수에 비해 매우 높은 스캐너 데이터의 대체율을 조절하기 위해서 일년중 재고가 모두 소진되거나 EAN 코드가 변한 제품은 지수 구축시 제외한다. Eurostat 보고서에 따르면 통상적인 1년중 스캐너 데이터상 대체율은 30%를 기록한다고 알려져 있다.

스캐너 데이터에 다양한 클리닝 필터를 적용한 후의 물가지수와 클리닝 전 모든 스캐너 자료를 적용한 물가지수를 작성하고 비교분석하였고 물가지수 작성에는 라스파이레스, 제본스 및 톱크비스트 산식을 각각 적용하였다.

<표 III-9> 클리닝 필터 적용 전후 지수 구축결과 비교

클리닝 필터	삼푸의 제외 비율	치약의 제외 비율	지수 구축 결과 비교
불연속 제품 제외	41%	40%	-
전월대비 가격상승 분포에서 상위 절반평균의 2.5배 이상 및 하위 절반평균의 2.5배 이하 제외	16%	20%	-
전월대비 가격이 33% 이상 및 이하로 변동한 경우 제외	8%	9%	- 톱크비스트 및 라스파이레스 지수는 클리닝 전후 큰 변화가 없었음 - 제본스 지수는 필터링 후 물가지수 수준이 상승함
판매액이 1000파운드 이하인 경우 제외	17%	25%	- 필터링 후 제본수 지수의 물가지수 수준이 크게 상승한 반면 톱크비스트 지수의 상승폭은 상대적으로 작았음
제품의 시장점유율이 0.1% 이하인 경우 제외	45%	34%	- 필터링 후 제본수 지수의 물가지수 수준이 크게 상승한 반면 톱크비스트 지수의 상승폭은 상대적으로 작았음

대체로 극단치의 가중치가 고려되지 않는 제본스 지수의 경우 동 극단치가 필터링 후 제외되면서 물가지수 수준이 상승하는 경향을 보였으며 이에 비해 라스파이레스 및 톱크비스트 지수는 상대적으로 필터링 전후로 덜 민감한 것으로 나타났다.

## 나. 프랑스

### 1) 적용품목 및 비중

Faivre (2012)에 따르면 프랑스 통계청은 2012년부터 6개의 슈퍼마켓 체인점의 2007년~2009년간 주별 스캐너 데이터 표본을 이용하여 다양한 활용방안에 대한 연구를 진행해 오고 있다. 스캐너 데이터 적용 품목 및 비중을 보

면 커피, 샐러드오일, 쌀, 요거트, 달걀, 초콜릿바, 과일주스 및 치즈 등 8개 품목군에서 스캐너 데이터를 이용하고 있다. 각 품목군에는 평균 약 1,050개의 제품이 있으며 동 스캐너 데이터 표본은 약 1억3천만 개의 관측치를 가지고 있다. 전통적인 표본조사 방식에서 A 매장에 있는 B 품목을 소비 바스켓의 기초 단위(elementary unit)라고 할 때, A 매장의 EAN 코드가 여기에 대응되는 기초 단위이다. 기초 단위는 3천만 개가 있으며 2008년 말과 2009년 말 사이 평균 대체율(또는 소모율)은 약 45%이다. 따라서 실제 분석에서는 이러한 대체율을 낮추기 위해 모든 시리즈를 사용하기 보다는 판매량에 비례한 무작위 샘플링 시리즈를 사용하여 소비 바스켓을 줄였으며 이 과정에서 대체율이 종전의 45%에서 28%로 하락하였다. 가령 초콜릿바의 경우 총 1,388개의 EAN 코드가 있으나 100개의 EAN 코드가 전체 판매량의 56%를 차지하며 3년에 걸쳐 계속 존재했던 약 55% EAN 코드가 전체 매출액의 72%를 차지한다.

한편 표본조사 데이터의 경우 연간 대체율은 17%인 것으로 나타나 스캐너 데이터(28%)보다 낮게 나타났는데 이 차이 11%p는 다음과 같은 이유에 따른 것이다. 첫째, 7%p는 프로모션 또는 생산자 특가 등 새로운 형태의 판매 단위가 등장하는 경우 스캐너 데이터에는 포착되지만 표본조사에서는 해당 판매 단위가 포착되지 않을 가능성이 있다. 둘째, 4%p는 포괄 상품의 차이에서 비롯될 수 있다. 표본조사에서는 대량 판매된 상품의 정보를 주로 조사하지만, 스캐너 데이터에서는 소량 판매된 상품의 정보도 모두 포함한다. 이러한 소량 판매 상품은 상대적으로 생존 기간이 짧아 결과적으로 대체율을 높이게 된다.

## 2) 지수편제 결과

표본조사 및 스캐너 데이터를 이용하여 연 단위 연쇄 라스파이레스 지수를 각각 작성하고 비교해보면 두 지수 간 큰 차이를 보이지 않았으며 또한 표본조사 데이터에 스캐너 데이터 약 27%를 포함시켜 작성한 혼합 품목군 지수 또한 기존의 표본조사 데이터로 작성된 지수와 크게 다르지 않은 것으로 나타났다.

## 다. 일본

### 1) 적용품목 및 비중

정태훈 (2016)에 따르면 일본 도쿄대학은 POS(point of sale) 데이터, 즉 스캐너 데이터를 이용하여 일별물가지수(UTokyo Daily Price Project)를 측정하고 공표하고 있다. 동 POS 데이터는 일본경제신문 디지털 미디어 회사(닛케이)에서 제공하는 일별 판매시점정보관리 데이터를 통해 1988년 4월부터 수집되기 시작했으며, 2016년 1월부터는 새로이 설립된 주식회사 나우캐스트(NOWCAST)가 데이터의 수집과 가공 그리고 배포를 담당하고 있다.

2015년말 기준으로 약 300개의 슈퍼마켓 점포로부터 약 35만여 종의 상품 판매정보가 기록되고 있다. 일별물가지수에는 대분류코드에 따라서 상품이 약 200개로 분류되고 소분류코드로는 모든 상품이 약 1,800개로 분류된다. 이에 따라 일별물가지수의 재화와 서비스는 일본의 공식적인 소비자물가지수 편제 상의 재화와 서비스 품목과 거의 겹친다고 평가될 수 있다.

### 2) 지수편제 결과

동 스캐너 데이터를 적용한 지수 구축 과정은 3단계를 거치게 된다. 우선, 개별 상품의 일간 변동률에 판매액을 가중치로 고려하여, 즉 톱크비스트 방식으로, 점포별 그리고 상품의 대분류별로 변동률 가중 평균치를 산출한다. 평균치 산출과정에서 두 시점 모두에서 판매기록이 나타나지 않는 경우 상품 집합에서 제외된다. 두 번째 단계에서는, 각 점포별 대분류 가격 변화율을 이용하여 각 점포별 판매액 가중치를 적용하여 대분류별 가격 변화율을 산출한다. 이 때 실제 판매액이 가중치로 고려되는 톱크비스트 방식을 이용한다. 세 번째 단계에서는 200개의 대분류별 가격 변화율에 해당 판매액 가중치를 고려한 톱크비스트 방식을 적용하여 종합적인 물가 상승률을 산출하게 된다.

이와 같이 산출된 일별 물가지수는 기존의 공식적인 소비자물가지수에 비해서 변동성이 크게 나타났으며 물가상승률은 평균적으로 1.0%p 낮게 나타났다. 이와 같은 추정결과는 유럽의 스캐너 데이터를 이용한 지수 산출 결과에 대체로 부합하는 것이다. 이는 일본도 스캐너 데이터가 상품의 높은 교체율 및 할인 상품의 정보를 포함하는데 기인한다.

#### 4. 소결

본 장에서는 스캐너 데이터 활용의 장점과 도전과제를 우선 정리해 보고 국내외 연구사례를 정리해 보았다. 특히 해외 기존 연구와 관련해서는 실제 스캐너 데이터를 소비자물가지수 작성에 활용하고 있는 나라를 중심으로 활용 방안과 자료 수집 방법, 스캐너 데이터의 기존 자료의 대체 정도 및 실제 지수 산식 적용법에 대해서 살펴보았다. 이와 더불어 통계청을 중심으로 스캐너 데이터를 소비자물가지수에 적용하는 방안을 연구하고 있는 나라와 일본처럼 도쿄대학교의 일별물가지수 사례처럼 별도의 민간부문에서 물가지수를 작성하고 있는 경우도 정리해 보았다.

우선 스캐너 데이터에 대해서 오랜 연구를 진행해 온 서구 유럽 국가의 경험에 비추어 스캐너 데이터의 장점을 정리하면 다음과 같다. 첫째, 스캐너 데이터를 이용하는 경우 최상위 물가지수의 작성이 보다 용이해진다. 둘째, 스캐너 데이터는 해당 모집단에 대한 전수조사에 가깝기 때문에 실제 표본의 크기를 크게 증대시키는 효과로 샘플링과 관련된 표본오차를 크게 줄일 수 있다. 마지막으로 스캐너 데이터를 통해 제품의 생멸에 관한 모니터링을 보다 쉽게 할 수 있다.

스캐너 데이터는 다양한 활용 가능성을 가지고 있지만 다음과 같은 비통계적 또는 통계적인 도전 과제가 있다. 비통계적인 부문에서 보면 스캐너 데이터는 IT 기술과 스캐너 자료 제공자에 의존한다는 단점이 있다. 즉, 스캐너 데이터는 소매점이 자사의 물품판매와 마케팅을 위해 개발한 각자의 개별적인 시스템에 의존함에 따라 소비자물가지수 산정에 활용하기 위해서는 소매점별로 개별적인 시스템을 유효하게 통합할 수 있는 IT 기술을 개발하는 것이 필요하다. 또한 자료에 대한 안정적인 접근과 확보를 위해서는 통계법 근거의 마련 및 소매 판매점과의 원활한 업무협약이 필수 전제조건이다. 통계적 도전과제를 살펴보면, 우선 제품 재론칭(relaunch) 등에 따른 스캐너 코드 매칭 문제가 상존하고 있다. 소매점의 경영 전략적인 판단에 따라 실질적으로 동일한 제품임에도 시간에 따라 별도의 스캐너 코드가 부여될 수 있다. 또한 스캐너 데이터를 이용한 물가지수 합산시 하위 합산 단계(elementary level)에서만 제한적으로만 이용되고 있으며 보다 상위 단계에서는 대체로 라스파이레스 산식으로 합산되고 있다. 이는 스캐너 데이터를 이용하여 최상위지수를 구축하는 경우 인플레이션 수준이 기초적으로 저하(하향편의)될 가능성이 있기 때문이다. 이 외에도 최상위 지수 구축시 연쇄편의 문제에 노출될 수 있으며 스캐너 데이터는 제품 스펙(item specification)에 대한 정보가

부족하기 때문에 품질조정이 어려울 수 있다. 또한 스캐너 데이터를 집계할 때 집계시점 및 점포별 집계 여부에 따라 평균 물가지수가 민감하게 변할 수 있는 문제가 있다.

스캐너 데이터를 이용한 국내의 기존 연구를 살펴보면, 우선 김기석(2012)은 21개 음식료 품목군을 중심으로 라스파이레스, 파쉐 및 피셔 물가지수를 산정하고 비교분석을 수행하였다. 해당 자료는 19개의 백화점, 16개 대형 할인점, 82개 편의점 및 39개의 슈퍼마켓으로부터 2000년~2012년 기간의 실시간 거래가 기록되어 있다. 21개의 품목군중에서 9,974개 상품의 1억 3천백만 건의 관측치가 포함되어 있다. 추정 결과를 보면, 점포 종합화 및 비종합화 여부는 물가지수 결과에 큰 영향을 미치지 않는 것으로 나타났다. 또한 대체로 직접지수와 고정 바스켓 연쇄지수 간의 차이는 크지 않았으나 변동 바스켓 연쇄지수는 해당 물가지수 값이 직접지수에 비해 비교적 크게 하락하는 것으로 나타났다. 또한 연쇄편의 문제로부터 자유로운 GEKS 및 RWGEKS 지수를 산정하여 대응되는 통계청의 서울시 물가지수와 비교한 결과, 통계청의 공식 지수에 5% 내외의 상향 편이가 존재하는 것으로 나타났다. 한편 강규호 외 (2015)의 경우 관세청의 냉장고 수출 품목 스캐너 데이터를 활용하여 연쇄지수 및 연쇄편의를 측정해본 결과, RWGEKS 지수의 연쇄편의 정도는 매우 미미한 것으로 나타났다.

실제로 스캐너 데이터를 물가지수 산정에 활용하고 있는 유럽 등 여러 나라의 사례 조사 결과는 다음과 같다. 대체로 스캐너 데이터는 자료의 확보 가능성 등으로 음식료품, 주류, 에너지 및 휴대폰 등 일부 가전제품 등에 집중되어 있다. 둘째로 개별 하위 품목별 지수 산정시 제본스 지수, 비가중 기하평균, 품질조정 단위가격지수, RWGEKS 지수 등 각 나라의 자료 및 소비패턴 환경에 맞추어 작성되고 있었다. 그러나 상위 지수 및 종합 소비자물가지수로 합산될 때에는 대체로 라스파이레스 산식이 적용되었다. 다만 가중치의 변경은 1년~3년 정도 빠르게 재조정하고 있는 것으로 나타났다. 스웨덴의 경우는 앞서 살펴보았듯이 2005년 이후 종합 소비자물가지수 산정시 라스파이레스 산식이 아니라 최상위 지수의 하나인 Walsh 지수 기법을 적용하고 있다. 앞서 논의한 것과 같이 소비자물가지수 편제에 스캐너 데이터를 활용하고 있는 나라의 현황은 <표 III-10>와 같다.

<표 III-10> CPI에 스캐너 데이터 활용 현황

	적용 품목	CPI에서 스캐너 데이터의 비중	스캐너 데이터 활용 방식/ 지수 구축 방식 등	주요 결과
스위스	- 음식료품 - 유사 음식료품(near food)	- 4개의 소매 체인 - 해당 품목의 시장 점유율 약 60~70%	- 최하위 품목은 제본스 - CPI는 로우 산식(가중치는 매년 업데이트)	- 향후 품목과 소매점을 늘려갈 계획
네덜란드	- 슈퍼마켓에서 판매하는 대부분의 품목 - 유류(에너지) - 휴대폰 - 여행상품	- 10개의 모든 슈퍼마켓 체인점 - 슈퍼마켓 해당 품목의 경우 서베이 자료를 100% 대체 - CPI전체의 약 20%에 해당	- 하위 품목은 제본스지수 - 일부 하위 품목은 QU(quality-adjusted unit value)지수 - CPI는 라스파이레스(가중치는 매년 업데이트)	- 제품의 재론칭(relaunch) 등에 따른 통계 왜곡을 완화하기 위해 QU지수의 적용범위를 확대할 계획 - 일정한 필터링을 통과한 스캐너 데이터 이용
노르웨이	- 식료품, 의약품, 유류, 주류 등	- 3개 대형마트(시장 점유율 96% 차지) - CPI 전체의 약 30%에 해당 - 품목별로 100% ~ 85%의 시장점유율	- 최하위 품목은 비가중 기하평균 - CPI는 라스파이레스 - 매년 1월에 가중치가 조정되는 연쇄방식	- 스캐너 데이터를 활용하여 식품 및 비주류 음료에 대한 하위 물가지수 산정 - 무료 접근으로 물가지수 작성의 효율성이 제고되었다고 판단
벨기에	- 슈퍼마켓에서 판매하는 대부분의 품목	- 3대 슈퍼마켓 체인점 (시장 점유율 75%) - 스캐너 데이터가 COICOP 5단계 9개 품목군에 이용	- COICOP 6단계 월별 연쇄 제본스 지수 - 중간단계는 비가중 기하평균 - CPI는 라스파이레스(가중치는 2년마다 업데이트)	- 스캐너 데이터와 서베이 방식의 물가정보는 COICOP 5단계에서 결합 - 2개월에 걸쳐 시장점유율이 일정한 기준점 이하인 경우 제외됨 - 샘플에서 제외되는 제품과 신규로 진입한 제품을 재점검
뉴질랜드	- 전자제품(가전제품)	- 품목별로 90% ~ 20% (CPI 분기마다 공표)	- 스캐너 데이터로부터 가전제품의 소비지출 가중치 산출	- 가전제품 품목은 RWGEKS 지수 산출 - 시장에서 스캐너 데이터 구매
스웨덴	- 식료품, 맥주, 담배,세제 및 개인 위생품 등 생필품 - 800개 대표 제품 - 90개의 제품군 선정	- 3개 소매점 80개 체인점 - CPI 전체의 약 14%	- 하위 지수는 제본스지수 - CPI는 2005년 이후 최상위지수인 Walsh지수를 편제 - Walsh지수는 기준 및 해당 시점 판매비중의 기하평균을 가중치로 적용	- 매년 가중치가 변경되는 연쇄 지수 기법을 적용 - CPI 전체적으로 매년 30% 정도의 샘플링이 교체

또한 스캐너 데이터를 향후 소비자물가지수에 활용하기 위한 연구를 수행하고 있는 영국 및 프랑스의 사례 조사에 따르면, 스캐너 데이터는 극단적인 증감률 및 판매 비중 등을 고려하여 적절한 필터링 작업이 필요하며 이 경우 해당 제품의 연평균 교체율은 약 30% 수준으로 나타났다. 또한 스캐너 데이터를 이용한 지수는 대체로 설문 및 방문 조사 방식을 통한 물가지수에



비해서 대체로 낮은 수준을 보이는 것으로 나타났다. 이는 기존 표본조사로부터 산정한 물가지수는 대체로 대량 판매 및 지속성을 보이는 상품 중심으로 가격 정보가 획득되는 반면, 스캐너 데이터는 소량 판매된 상품의 가격도 지수 산정에 편입되게 된다. 이러한 소량 판매 상품의 경우 제품 생존주기가 짧아지면서 판매 종료시점까지 지속적으로 할인판매에 노출될 수 있으며 스캐너 데이터의 높은 상품 교체율을 낳는 주요 요인으로 작용하고 있다. 일본의 경우는 민간기관에서 스캐너 데이터를 이용하여 일별 물가지수를 발표하고 있다. 동 지수는 하위에서 종합 단계까지 톱크비스트 산식이 적용되며 산출된다. 동 지수의 변동성은 기존의 공식적인 물가지수 보다 높으며 물가 지수 수준도 평균적으로 1%p 낮은 것으로 나타났다. <표 III-11>은 상기에서 논의한 스캐너 데이터 활용 방안을 연구중인 나라의 현황을 요약한 것이다.

<표 III-11> CPI에 스캐너 데이터 활용 방안 연구 현황

	적용 품목	CPI에서 스캐너 데이터의 비중	스캐너 데이터 활용 방식/ 지수 구축 방식 등	주요 결과
미국	- 치약 및 샴푸	- 해당 품목 100%	- 다양한 클리닝 필터(극단적 변동률 및 판매액이 미미한 품목 제외 등)를 적용 - 스캐너 데이터를 이용하여 제본스지수 산정시 필터별로 9~45%의 자료가 제외됨	- 클리닝 필터전 제본스지수가 기존 품목지수보다 20~30% 정도 낮아짐 - 제본스지수 보다는 톱크비스트 및 라스파이레스 지수가 클리닝필터 전후로 보다 안정적인
프랑스	- 커피, 샐러드 오일, 쌀, 요거트, 달걀, 초코렛 바, 과일주스 및 치즈	- 해당 품목 100%	- 6개의 슈퍼마켓 체인 10개 품목군, 일년간 평균 대체율은 45% - 판매량에 비례한 무작위 샘플링으로 소비 바스켓을 줄였음: 45% --> 28% - 일년 기준 고정 바스켓 라스파이레스 지수	- 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	- 300개 슈퍼마켓 모든 상품 - 200개 대분류 - 1800개 소분류	- 해당 품목 100%	- 각 점포별로 대분류 200개 상품의 가격 변동률 가중 평균치 산출 - 대분류 상품의 가격 변동률 평균치 산출 - 종합지수가격 변동률 평균치 산출 - 판매액 가중치 고려하는 톱크비스트 방식을 적용 - 일별물가지수	- 도쿄대에서 작성 및 공표 - 공식적인 CPI에 비해서 변동성이 더 크고 물가상승률은 약 1%p 낮은 것으로 나타났다음 - 두 시점(이틀간) 모두에 판매 기록이 나타나지 않으면 샘플에서 제외

## IV. 우리나라 스캐너 데이터를 이용한 품목별 소비자물가지수 작성

### 1. 대한상공회의소의 스캐너 데이터<sup>5)</sup>

#### 제1절 스캐너 데이터의 특징

스캐너 데이터는 다음과 같은 특징을 가진다. 첫째, 스캐너 데이터 상의 판매 단가는 1+1 증정행사, 기간제/시간대 할인행사 적용 등으로 인해 상품의 정가와 다를 수 있다. 둘째, 과거 판매 상품이 반품되어 해당일 판매실적이 음수 값(-)이 될 수 있다. 셋째, 묶음 상품의 매출 실적을 개별 상품의 매출로 처리할 수도 있고, 개별 상품과 다른 별개의 상품으로 처리할 수도 있다. 대한상공회의소 유통물류진흥원이 동 데이터를 유통사로부터 수집하고 있어서 이번 연구에 제공한 스캐너 데이터는 유통매장별 일 마감 데이터를 주단위로 통합한 것이다.

#### 제2절 스캐너 데이터의 수집, 분석 과정의 요약

대한상공회의소의 스캐너데이터 수집·분석 과정은 데이터 수집, 데이터 정제, 상품목록수정, 데이터 처리, 데이터 분석의 과정으로 구분된다. 첫째, 데이터 수집과정은 유통업체에서 POS시스템을 통해 수집된 표본점포 매출데이터를 대한상의 분석시스템으로 전송하는 과정이다. 둘째, 데이터정제 과정은 수집데이터의 오류, 데이터의 중복 또는 누락 등을 처리하는 과정이다. 셋째, 상품목록수정 과정은 신규 출시된 상품 정보를 확보하고 동일 속성정보 추적하는 과정이다. 넷째, 데이터 처리 과정은 상품별 판매액, 시장점유율, 판매증가율을 집계하는 과정이다. 마지막으로 데이터 분석과정은 분야별 보고서를 작성하여 인터넷에 게시하거나 이메일 전송하는 과정이다. 대한상공회의소의 스캐너데이터 수집 및 분석과정중 추가 설명이 필요한 부분은 절을 달리하여 정리한다.

#### 제3절 스캐너 데이터의 수집

유통사 스캐너 데이터(POS 데이터)는 유통사의 영업 비밀에 속하므로 대한

5) 이 내용은 대한상공회의소 유통물류진흥원의 자문을 바탕으로 연구진이 재서술한 것이다.

상공회의소에서는 데이터를 유통사 시스템을 통해 직접 수집하지 않고, 유통사가 자체 제작한 엑셀 또는 텍스트 형식의 데이터 파일을 대한상공회소가 지정된 웹사이트 특정 계정에 업로드하는 방식으로 월별 수집 일정에 따라 수집한다. 농협하나로유통의 경우 전용 시스템과 직접 연계하여 1주일 1회 스캐너 데이터를 제공하고 있고, 나머지 유통사들은 한 달에 1번 데이터를 제공하고 있다. 매일 유통사로부터 수집되는 스캐너 데이터는 <표 IV-1>과 같은 양식을 가지고 있다. 대한상공회의소는 스캐너 데이터 자료들을 텍스트(txt) 파일 기준으로 데이터를 정리하고 있다. 이 때 농협, 롯데백화점에서 제공되는 Excel 파일 형태의 데이터는 수작업으로 파일 형태를 텍스트로 변경하고 있다.

<표 IV-1> 대한상공회의소 스캐너 데이터 구조

매장코드	주간일자	바코드	주간매출 수량	주간매출 총액	주간매출 단가
7자리	6자리	13자리	7자리	11자리	7자리

- 주 : 1) 매장코드 : 유통사코드 4자리 대한상의 지정 + 매장코드 3자리 유통사 임의 지정  
 2) 주간일자 : 해당 주의 마지막 일자 YYMMDD  
 3) 바코드 : GS1에서 발급하는 국제표준바코드 8자리(GTIN-8), 12자리(UPC-A), 13자리(GTIN-13) ·주간매출단가 = 주간매출총액 ÷ 주간매출수량

수집된 데이터는 오류체크 프로그램을 통해 검토 및 확인하여 오류가 있는 경우 해당 데이터를 삭제하거나 수정한다. 첫째, 유통사별 총 점포수, 신규점포 포함여부 확인하고, 신규점포인 경우 매장마스터에 동 정보를 추가한다. 둘째, 주간 일자별 데이터 수량과 판매액을 검토하여 정상적인 수량 증가 및 감소를 확인한다. 셋째, 과거 3개월 데이터와의 매출액 비교하여 비정상적으로 수량이 증가 또는 감소했는지 확인한다. 넷째, 필드 자릿수, 공백, Tab, 특수문자 등이 포함되었는지 확인한다.

[그림 IV-1] 대한상공회의소 스캐너 데이터

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		

대한상공회의소는 정제된 스캐너 데이터를 프로그램을 통해 데이터베이스 서버에 업로드하고 데이터 확인 및 수정을 진행한다. 월간 스캐너 데이터를 벌크(Bulk) 형태로 활용하는 사업자들에게 전달하기 위해 매장코드는 시스템에서 암호화시켜 [그림 IV-1]과 같이 저장한다. 이번 연구에 이용된 스캐너데이터는 [그림 IV-2]의 형태로 대한상공회의소에서 SQL을 이용하여 추출한 것이다.

[그림 IV-2] 대한상공회의소 통계청 제공 스캐너 데이터

The screenshot shows a software interface for managing scanner data. On the left, there is a list of transactions with columns for '이름' (Name), '수정일 날짜' (Update Date), and '수정일 시간' (Update Time). On the right, a detailed view of a transaction is shown, including fields for 'Barcode', 'Product Name', 'Price', and 'Quantity'. The data is presented in a structured, tabular format typical of a database application.

#### 제4절 스캐너 데이터의 분류와 특징

기본적인 스캐너 데이터는 매장에서 발생하는 “고객별 판매데이터”이다. 고객에게 상품을 판매하고 영수증 발급한 개별 거래들의 내역이 POS 시스템에 저장·추적된다. 이 데이터는 한 고객이 어떠한 상품을 구매하였고 총 구매액이 얼마인지 알 수 있게 하는 정보이다. 고객에게 물건 판매가 끝날 때 마다 생성되는 이 데이터는 “영수증 데이터” 또는 “장바구니 데이터”라고 불린다. 만약 과거에 구매한 상품에 대한 거래 번호를 안다면 POS 시스템을 통해 동 상품을 환불 또는 교환 처리할 수 있다.

POS 시스템의 “고객별 판매 데이터”는 일정한 처리를 통해 “상품별 판매 데이터”로 전환할 수 있다. 상품별 판매데이터는 상품명, 판매단가, 매출총액, 판매시간, 할인조건 등으로 구성된다. 이를 통해 상품 단위로 판매실적을 파악할 수 있으며 이를 통해 상품의 재고 관리 및 공급망 관리가 효율적으로 이루어 질 수 있다.

POS 시스템이 매장 관리를 위한 종합 시스템으로 변화된 이후 유통업체는 스캐너 데이터를 이용하여 상품의 매출분석과 고객관계관리(Customer Relationship Management, CRM)를 하고 있다. POS 시스템에서의 CRM 기능은 매장에서의 고객 마일리지 서비스를 제공하기 위한 목적으로 운영되고

있지만, 일부 유통 매장의 경우, 고객 분석 차원에서 고객의 성별과 연령대만 판매 시점에서 등록하는 방식으로 고객 데이터를 관리하고 있다.

고객별 상품판매 데이터를 살펴보면 한 고객 장바구니에 어떤 상품이 같이 있는지 분석하면 특정 상품들의 묶임 빈도수를 측정하여 상품 간 연관분석이 가능하다. 고객별 상품 판매 데이터는 고객식별자, 성별, 연령대, 거주지, 소득수준, 거래 실적 등으로 구성되어 있으며 고객 마일리지 카드나 모바일 앱, 상품권 행사 등을 통해 수집된다. 이 데이터는 고객의 특성에 따른 구매 패턴 연구와 고객 특성 분석에 중점을 둔 상권 분석, 고객 특성 분석에 따른 상품 구색 및 매장 진열 관리에 필요한 정보가 된다. 고객 특성이 비슷한 지역 매장의 마케팅 전략을 서로 벤치마킹할 수 있다. 고객의 구매 패턴을 분석하여 자동으로 추천하는 기능을 제공할 수 있다.

### 제5절 스캐너 데이터의 특이항 처리

유통매장에서의 상품 판매와 관련하여 1:1 정상 판매, 묶음 상품 판매, 가격 할인 행사 판매, 증정품 행사 판매, 상품 반품 등 다양한 거래가 있다. 이 거래에서 정상적이지 않은 값이 발생한다. 특히 상품의 판매단위를 구분하는 바코드 번호가 잘못 사용되는 경우 데이터 자체가 의미 없으므로 스캐너 시스템에 등록된 상품정보의 정확도를 확보하는 것과 유통사 별 일관된 관리 방침을 정하는 것이 스캐너 데이터의 품질 향상을 위해서 중요하다. 스캐너 데이터의 비정상 값의 대표적 유형은 가격 또는 수량이 0 또는 0 이하인 경우와 동일 상품의 가격이 크게 다른 경우 등이다. 비정상 값 유형별 사유 및 처리 방안은 <표 IV-2>와 같다.

<표 IV-2> 비정상 값 유형별 사유 및 처리 방안

유형사례	발생 원인	처리 방안
판매가격이 0	증정품 행사 시, 일반 판매상품을 증정품으로 처리하고 판매량과 가격을 0원으로 임의 처리	레코드 삭제
판매수량이 0 이하	집계 기간 동안 정상 판매 실적은 없으나 고객 반품이 발생하는 경우 재고 수량 조정을 위한 판매수량 (-) 등록	레코드 삭제
동일 상품 가격이 기간 별로 크게 상이	(묶음 상품 행사 시 상품정보 관리 문제) 단일 상품과 묶음 상품의 바코드를 동일하게 관리하여 하나의 상품이 기간에 따라 판매가가 크게 다르게 보이게 됨	묶음 상품 행사 기간을 알 수 있는 경우에는 해당 기간의 판매수량을 단품의 판매수량으로 역산하여 판매 가격과 수량을 보정함

## 제6절 스캐너 데이터의 보완사항과 활용확대

스캐너 데이터는 유통사 인식 전환과 보안 체계 강화를 통해 다음과 같이 보완될 필요가 있다. 첫째, 스캐너 데이터의 수집과 통합을 쉽게 할 수 있도록 유통사 POS 시스템을 표준화 하고, 데이터 양식과 수집방식을 표준화하는 작업이 필요하다. 예를 들면 국내 유통업계에서 공통으로 활용할 수 있는 상품정보 마스터 데이터베이스를 구축하고 이를 바탕으로 스캐너 데이터를 수집하여 POS 시스템에 등록된 상품정보와 행사 상품 관리 등에 일관성을 유지하는 것이다. 둘째, 현재 주요 대형마트의 스캐너 데이터가 포함되지 않아서 대표성에 한계가 있으므로 이들 유통사 정보를 조사 등으로 통해 얻고 이를 결합하는 방안을 마련할 필요가 있다. 셋째, 데이터 수집의 시의성을 확보하는 것이다. 스캐너 데이터 수집을 위한 연계시스템을 유통사에 보급하여 수집 주기를 단축하는 것이다. 또한 파일 수집 방식이 아닌, API 연계 방식으로 전환할 필요가 있다. 넷째, 고객별 거래 단위의 스캐너 데이터를 추가하는 확보 하는 것이다. 이를 통해 인구 구조 별 소비 형태 분석 등, 분석 범위를 확대할 수 있다.

스캐너 데이터는 다양한 분야로 활용될 수 있다. 첫째, 상품별 매출 동향을 파악할 수 있다. 기존의 단순 매출 분석 뿐 아니라, 날씨, 기온, 특정기간에 대한 상품별 매출 동향을 분석하여 상품 별 팔림세의 특성이나 외부 환경에 의한 구매 탄력도가 높은 상품 군과 같은 의미 있는 분석도 가능하다. 둘째, 상품별 가격 동향을 파악할 수 있다. 개별 상품 또는 특정 상품군으로 그룹핑한 상품들의 가격의 변동 여부와 가격대 형성 분포를 분석할 수 있다. 상품군 분석을 위한 기준 상품분류체계를 갖추고, 수요기업/기관에서 필요로 하는 상품군을 스캐너 데이터 기준 상품분류와 대응시켜 원하는 상품군에 대한 동향 파악이 용이하도록 하는 체계가 필요하다. 셋째, 고객별 구매행태를 분석할 수 있다. POS 시스템에 고객 정보를 추가 관리하여 고객 특성별 구매 내역을 집계하고, 구매상품과 고객의 연관도를 분석할 수 있다. 고객 이름이나 주소 같은 개인정보가 아니어도, 고객의 성별, 연령대, 주거지역, 현금/카드(결제 수단) 등으로도 충분히 고객 특성과 관련된 구매 행태에 대한 분석가능하다. 넷째, 상품 간 구매 연관도를 분석할 수 있다. 거래 당 상품 구매 내역을 분석하면 이종의 상품들이 동시에 팔리는 현상을 확인할 수 있고, 그 결과 상품 간 연관도 분석이 가능하다. 다섯째, 상품별 소비심리 탄력도를 분석할 수 있다. 국내 경기에 따라 소비가 변화하는 상품들을 조사하여 해당 상품 판매액의 변화를 통해 체감경기 예측을 시도할 수 있다.

## 2. 스캐너 데이터의 입수 및 가공

### 제1절 데이터 입수 및 가공 과정

본 연구는 대한상공회의소(이하 대한상의)의 협조를 통해 POS시스템 기반의 스캐너 데이터를 입수하여 진행되었다. 스캐너 데이터를 이용한 품목지수를 산출하기 앞서 다음과 같이 스캐너 데이터의 입수 및 가공 단계를 거치게 된다.

- ① 2013년부터 2016년까지의 4년에 걸친 분석 대상 데이터를 입수
- ② 대한상의 데이터 코드와 통계청 품목 코드의 매칭
- ③ 이상치의 탐색 및 제거

이 장에서는 스캐너 데이터의 가공 과정을, 다음 장에서는 지수 시산 과정을 살펴본다.

### 제2절 데이터의 원천

연구진은 대한상으로부터 일별 자료를 취합한 스캐너 데이터를 주별 자료로 가공한 데이터를 통계청을 통해 수집하였다. 이 데이터에는 각 상품코드별로 지역변수와 업태변수가 정의되어 있다. 지역변수는 16개 시도별로 구분되어 있고, 업태는 대형마트, 독립수퍼, 수퍼체인 등 3가지로 구분되어 있다. 대한상의는 소비자물가지수 편제의 편의성을 제고하기 위하여 개별 상품의 상위 코드인 KAN 코드를 생성하여 함께 제공하고 있다. 이 코드는 대한상의의 스캐너 데이터를 통계청의 상품분류 코드와 대응하기 위해 만들어졌다.

<표 IV-3>은 대한상의에서 제공한 스캐너 데이터의 가상 표본인데 그 의미는 다음과 같다. KAN은 통계청의 상품코드와 대응하기 위한 목적으로 생성한 대한상의의 코드로서 ITEM\_CD라는 개별상품을 나타내는 코드의 상위 분류다. 또한, KAN\_NM은 해당 KAN코드에 적절한 이름을 부여한 변수이고 ATTR1과 ATTR2는 각각 업태를 나타내는 변수와 지역을 나타내는 변수이다. 업태와 지역에 관한 변수는 향후 상위 총지수 산출에 있어서 기준이 되는 변수이다. WEEK\_END\_DT는 데이터가 입수된 시점을 의미하며 ITEM\_NM은 개별 세부 상품명을, SUM\_QTY는 해당 세부 상품의 판매량 총계를, SUM\_AMT는 해당 세부 상품의 매출액 총계를 나타낸다.

<표 IV-3> 대한상공회의소 제공 스캐너 데이터의 예시

KAN	KAN_NM	ATTR1	ATTR2	WEEK_END_DT	ITEM_CD	ITEM_NM	SUM_QTY	SUM_A MT
10102	소금	대형마트	경기도	201705**	00702471*****	기타 팝랜드저염454G - VINYL	10	*****
10102	소금	대형마트	강원도	201705**	33050400*****	Guerande 게랑드토판천일염500G	1	*****
10102	소금	대형마트	경상남도	201705**	33050400*****	Guerande 게랑드토판천일염500G	1	*****
10102	소금	독립수퍼	서울	201705**	33050400*****	Guerande 게랑드토판천일염500G	3	*****
10102	소금	대형마트	인천광역시	201705**	33050400*****	Guerande 게랑드토판천일염500G	1	*****
10102	소금	대형마트	전라남도	201705**	33050400*****	Guerande 게랑드토판천일염500G	1	*****
10102	소금	대형마트	전라북도	201705**	33050400*****	Guerande 게랑드토판천일염500G	1	*****
10102	소금	수퍼체인	경기도	201705**	33050411*****	지이디통상 가는토판천일염250G - CAN	6	*****
10102	소금	독립수퍼	광주광역시	201705**	33050411*****	지이디통상 가는토판천일염250G - CAN	4	*****
10102	소금	수퍼체인	경상남도	201705**	33050414*****	효성인터내셔널 해초소금250G - CAN	1	*****
10102	소금	대형마트	경상북도	201705**	33050414*****	효성인터내셔널 해초소금250G - CAN	1	*****
10102	소금	수퍼체인	광주광역시	201705**	33050414*****	효성인터내셔널 해초소금250G - CAN	1	*****
10102	소금	대형마트	경상남도	201705**	33050440*****	기타 플레르드셀소금의꽃폴리250G - VINYL	1	*****
10102	소금	수퍼체인	경상북도	201705**	33050440*****	기타 플레르드셀소금의꽃폴리250G - VINYL	1	*****

\* 표에 수록된 세부 상품의 코드와 가격 자료는 masking한 상태임

### 제3절 데이터의 속성

똑같은 상품이라 하더라도 개별 점포마다 판매가격이 다를 수 있다. 대한상의 자료에는 개별 점포의 정보 대신 이의 상위 분류로 볼 수 있는 업태별로 취합된 총판매량과 매출액 총계가 수록되어 있다. 또한 일별 자료를 취합하여 주별 단위 기준으로 가공이 되어 있다. 이렇게 대한상의가 가공된 자료를 제공한 것은 개별 점포별 기준으로 상품을 분류하였을 경우 첫째, 세부 상품 단위에서 볼 때, 상품 판매의 일관성이 확보되기 어렵고 둘째, 필요이상으로 데이터 크기가 증가하여 효율적인 분석이 어렵다고 판단되었기 때문이다. 또한 기본적으로 물가지수를 산출하기 위해서는 특정 상품의 가격정보가 지속적으로 관측되어야 하는데, 개별 점포단위나 일별 단위에서 이러한 경향을 보이는 상품을 다수 확보하기에 현실적으로 제약이 있기 때문이다. 따라서 본 연구에서는 개별 점포 대신 업태별 구분, 주별로 취합한 자료



대신 월별 기준 변환 자료를 이용하여 지수를 산출하였다. 이렇게 자료를 변환한 뒤 대한상의의 코드를 기준으로 개별 세부 상품명과 각 상품이 판매되고 있는 개별 상점이 속하는 업체의 조합을 계산해 보면, 2013년에는 87,684개, 2014년에는 86,616개, 2015년에는 103,276개, 2016년에는 112,926개로 점차 증가하는 것으로 나타났다. 주의할 점은 업체별 세부 상품 수는 증가하고 있으나 실제 이 자료 내에서는 개별 세부 상품의 생명이 반복되고 있다는 점이다. 즉, 분석 대상 기간인 2013년부터 2016년까지 계속 판매되고 있는 상품이 있는 반면, 해당 기간 중 판매가 중지된 상품이나 새롭게 시장에 선보인 상품 등 여러 가지 경우의 수가 존재한다는 것이다. 이러한 점을 고려하여 합리적인 지수산출을 위해 적절한 코드매칭과 이상치 제거 등 가공단계가 반드시 뒤따라야 한다.

#### 제4절 데이터의 코드 매칭

대한상으로부터 입수한 상품의 품목을 대한상의 자료 기준인 KAN 코드 기준으로 81개이다. 앞서 살펴보았지만, KAN 코드는 통계청 분류 기준에 맞추어 개별 세부 상품을 상위품목으로 묶어 새롭게 생성한 것이다. 따라서 KAN 코드 81개는 원칙적으로 통계청 분류(편의상 PUMID 코드로 명명)와 정확히 일치해야 하지만 다음에 서술한 바와 같이 해결해야 할 몇 가지 문제점을 가지고 있다.

##### 가. 대한상의 코드(KAN)와 통계청 코드(PUMID)의 대응

데이터 가공에 있어서 일차적인 목표는 대한상으로부터 입수한 스캐너 데이터의 KAN 코드를 PUMID 코드로 대응시켜 통계청 기준에 부합하는 품목 코드인 PUMID별 가격지수를 계산하는 것이다. KAN 코드와 PUMID 코드의 관계는 대부분 일대일 대응 관계이다. 이 경우 해당 품목의 가격지수를 PUMID 코드 기준으로 산출하는 것은 큰 문제가 되지 않는다. 하지만, 두 코드의 대응 관계가 다대일 또는 일대다일 경우 조금 더 면밀하게 검토해 볼 필요가 있다. <표 IV-4>와 <표 IV-5>는 KAN 코드와 PUMID 코드의 다대일 대응, 일대다 대응을 정리한 것이다.

<표 IV-4> KAN과 PUMID의 다대일 대응

	KAN코드	PUMID코드
이대일 대응	011499, 011599	A019130(스프)
	011502, 011599	A019180(즉석식품)
	012016, 012099	B011040(막걸리)
	021809, 021810	E061070(가정용비닐용품)
삼대일 대응	012001, 012002, 012004	A021010(커피)
	021908, 021909, 021910	E061050(청소용세제)

<표 IV-5> KAN과 PUMID의 일대다 대응

	KAN코드	PUMID코드
일대다 대응	010701	A012050(소시지), A012060(햄 및 베이컨)
	010802	A017280(김), A017290(맛김)
	011599	A019130(스프), A019180(즉석식품)
	011902	A018050(비스킷), A018070(파이)
	012011	A022040(기능성음료), A022060(혼합음료)
	012016	B011040(막걸리), B011060(약주)
일대삼 대응	010199	A018110(물엿), A019070(양념소스), A019120(혼합조미료)

실제로 KAN 코드를 PUMID 코드로 대응시킬 때에, KAN 코드에서 PUMID 코드로 다대일 대응의 경우는 문제되지 않는다. 다수의 KAN 코드 분류를 무시하고 하나의 코드처럼 간주하여 PUMID 코드 기준에서 세부 상품들을 병합하여 가격지수를 산출해 주기만 하면 되기 때문이다. 그러나 일대다로 대응되는 경우 KAN 코드에 속한 물품들을 PUMID 코드로 일일이 나누어 대응시켜야하기 때문에 보다 어려운 문제에 직면하게 된다. 한 개의 KAN 코드에 속한 세부 개별 상품의 수는 수천 개에 해당하기 때문에 이를 일일이 수작업으로 직접 분류하는 것은 현실적으로 불가능하다. 또한 새로운 상품이 시장에 출현한다면 이때마다 상품명을 파악하여 일일이 적절한 코드로 분류해야 하는 어려움도 뒤따르게 된다.

#### 나. 키워드를 이용한 KAN 코드의 분류 및 대응

일대다 대응의 문제를 해결하기 위해 본 연구에서는 세부 상품명에 속한 키워드를 이용해 자동으로 분류하는 방법을 이용하였다. 키워드를 설정할 때에는 통계청 CPI 물품 분류표에 포함된 BASENAME을 참고하였다. BASENAME은 PUMID의 간략한 정의이다. 예를 들어, KAN코드 '010701'을 PUMID 코드 'A012050(소시지)'와 'A012060(햄 및 베이컨)'으로 분류할 때

키워드를 소시지, 햄으로 설정하여 BASENAME에 소시지를 포함한다면 ‘A012050’으로, 햄을 포함한다면 ‘A012060’으로 분류하는 방식이다. 이러한 방식을 이용하여 KAN 코드와 PUMID 간의 일대다 대응의 문제를 해결한 결과가 <표 IV-6>에 요약되어 있다. <표 IV-6>은 2013년부터 2016년까지의 스캐너 데이터 품목 중 키워드를 이용해 KAN 코드를 분류한 결과, 전체 개별 상품 중 어느 정도가 분류되고 어느 정도가 분류되지 못하였는지를 각 KAN 코드별로 살펴본 것이다.

<표 IV-6> 키워드를 이용한 KAN코드 분류 결과

KAN코드	010199	010701	010802	011599	011902	012011	012016
전체 상품 수(개)	3,020	3,258	5,599	3,645	6,015	3,338	1,493
미분류 상품 수(개)	121	228	6	1737	1407	457	148
미분류 상품 거래 비중	0.13%	7.8%	0.0028%	31%	14%	2.7%	1.6%

위의 표를 보면 KAN 코드 ‘011599’와 ‘011902’에서 분류되지 못한 물품 수와 거래 규모가 상당 부분을 차지하는 것을 확인할 수 있는데 이는 키워드 분류 방법의 한계점으로 보인다. ‘011599’는 ‘A019130(스프)’와 ‘A019180(즉석식품)’으로, ‘011902’는 ‘A018050(비스킷)’과 ‘A018070(파이)’로 나뉘어 분류되어야 한다. 그러나 해당 품목의 키워드를 선택하는 데에는 상당한 어려움이 뒤따른다. 첫 번째 경우 즉석식품의 분류 기준이 되는 기준 품명이 ‘1) 즉석국(블럭포장 또는 레토르트) 2) 즉석밥(플라스틱용기)’인데 키워드 분류 방법으로는 적절히 구분하기 어려운 상품들이 많이 존재하기 때문이다. 예를 들면, ‘반찬류’, ‘샐러드’, ‘삶은 계란’과 같은 상품들은 키워드 분류로 식별하기 매우 어렵다. 두 번째 경우도 마찬가지로 비스킷과 파이 각각의 기준 품명으로 세부 상품을 적절히 구분하기에는 상당한 제약이 있다. 전반적으로 키워드에 의한 분류 방법은 설정 방법이 다소 작위적이고 분류에도 한계가 있었지만 현실적으로 프로그램을 통한 자동화 분류 방법이 가장 최적의 선택이었다고 판단된다. <별첨 2>에는 KAN 코드를 PUMID 코드에 매칭시키기 위해 분류에 사용된 키워드의 예시가 수록되어 있다.

#### 다. 가공식품 지수를 산출하기 위한 코드 매칭

대한상의에서 제공한 데이터를 통계청 분류에 맞추어 ‘가공식품’ 상위 지

수를 산출하기 위한 코드 매칭도 고려해 보았다. 대한상의의 데이터에는 가공식품 하위 품목지수에 속하는 품목뿐 아니라 기타 품목 지수도 포함되어 있으므로 통계청 분류에 따른 가공식품 상위지수를 산출하기 위해서는 적절한 품목만 선별하여 상위지수 편제 시 산입해 주어야 한다. 대한상의 스캐너 데이터에 포함되어 있어 매칭이 가능한 통계청 기준 가공식품 분류 품목지수는 총 58개였으며 동 데이터에 포함되어 있지 않은 가공식품 분류 품목지수는 12개였다. 따라서 가공식품 총지수를 산출하기 위해서는 스캐너 데이터를 바탕으로 산출한 품목별 지수 58개와 통계청 CPI 품목별 지수 12개를 병합하여 이용해야 한다. <별첨 2>에는 앞서 살펴본 코드별 대응관계를 포함하여 대한상의의 KAN 코드를 통계청 기준 PUMID 코드에 대응시키기 위한 기준이 종합되어 있다.

## 제5절 이상치의 탐색

기초 데이터를 탐색하는 과정에서 코드 매칭과 더불어 중요한 문제로 부각된 것이 이상치 제거에 관한 것이다.

### 가. 이상치 추정 사례

이상치의 원인으로 추정되는 사례는 다음과 같이 크게 두 가지로 구분할 수 있다.

#### 1) 가격의 정합성

대한상으로부터 제공받은 스캐너 데이터를 정리하면서 가격 정보가 비정상적인 것으로 판단되는 관측치를 발견하였다. 첫째는 가격, 수량이 0이거나 음의 값을 나타내는 거래이다. 이러한 관측치가 발생하는 이유는, 스캐너 데이터가 유통사의 POS 데이터를 원천적으로 가공한 데이터로서 유통사별 매장 POS 에서 반품이나 증정품을 처리하는 방식에 따라 0이나 음으로 기록된 관측치가 포함될 수 있다는 것이다. 즉, 특정 세부 상품에 있어서 일반적인 반품처리가 발생하고 해당 주에 그 특정 상품의 정상 판매가 없는 경우, 가격이 (-)로 기록될 수 있다는 것이다. 이는 재고량 조정을 위해 임의로 반품처리를 하는 경우에도 동일하게 적용된다. 또한 증정품 행사 시 증정품의 판매량을 0개, 가격을 0원으로 처리하여 이러한 비정상적인 가격정보가 기록될 수도 있다.

## 2) 동일 상품의 가격이 크게 차이가 나는 경우

세부 상품별로 살펴보면 상식적인 수준을 벗어나 가격의 범위가 상당히 넓게 분포하고 있는 상품들이 많이 관측되었다. 상품별 총매출액을 총판매량으로 나눈 단가를 기준으로 살펴보면 동일한 상품인데도 불구하고 낮은 금액 수준의 수 배에 달하는 높은 금액 수준이 상당 수 포착되었다. 예를 들어 1000원짜리 상품의 빈도가 수백 개 관측되면서도 8000원짜리 상품의 빈도도 역시 수백 개가 관측되는 품목이 이러한 경우에 해당한다. 이러한 현상은 특히 업체 간 가격 차이로 주로 나타난다. 예를 들면 낮은 금액에 한 업체의 가격이 주로 분포하고 높은 가격에 다른 한 업체의 가격이 주로 분포하는 것이다. 물론 업체 간 가격 격차는 어느 정도 예상할 수 있는 것이지만 실제 데이터를 살펴보니 그 격차가 상식적인 수준을 벗어나는 경우가 관측되었다. 일부 상품의 경우, 단품과 묶음 상품의 바코드를 동일하게 등록하여 특정기간 동안 단품 상품 바코드를 묶음 상품의 바코드로 사용하는 경우가 발생한 데 따른 것이라 추정된다. 참고로 대한상의의 데이터를 이용하는 외부 마케팅 업체의 경우에는 묶음상품 정보를 파악하여 해당 가격을 낮게 상품의 판매실적으로 역산하여 처리한다고 한다. 이상의 사례를 참고하여 효과적으로 이상치를 제거하는 방법을 비교해 보도록 한다.

### 나. 이상치 탐색 및 제거 방법

이상치 제거를 위해 다음과 같은 세 가지 방법을 고려하였다.

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

위와 같은 방법을 적용해 본 결과 첫 번째 이상치 제거 방법인 표준편차를 고려한 방법은 가격 분포가 쌍봉 분포일 경우, 즉 낮은 가격대와 높은 가격대에서 상대적으로 높은 빈도수를 나타낼 경우 이상치를 효과적으로 제거하기 어렵다. 또한 왜도(Skewness)를 고려하여 한쪽 방향으로 치우친 가격의 이상치를 제거하려 할 경우, 의도치 않게 정상적인 가격대가 제거되는 현상이 나타났다. 따라서 세 가지 방법 중 세 번째 각 세부 제품에 대한상의 데이터를 위와 같은 방법을 적용하여 각 개별 상품마다 4 개년에 걸친 단가 자료를 기준으로 상위 25%를 초과하거나 하위 25%를 하회하는 단가는 이상치로 간주하여 절사하여 데이터를 정제하였다. 총 분석 대상 상품 중 이러한 방법으로 이상치를 제거하였더니 각 연도별로 정제된 이후의 조합의 수가 원자료와 비슷하게 점차 증가하는 것으로 나타났다. 세부 개별 상품과 업태의 조합 수를 기준으로 살펴보면 2013년에는 79,967개, 2014년에는 79,008개, 2015년에는 92,492개, 2016년에는 102,310개로 나타나 모두 원데이터의 90% 정도를 차지하였다. 참고로 <표 IV-7>와 같이 이상치 제거의 기준을 10%로 낮추어 데이터를 정제한 이후의 결과를 살펴보았는데, 이 경우에는 원데이터의 95% 정도를 차지하는 것으로 나타났다. 검토 결과 25% 기준의 절사 방법이 이상치를 보다 효과적으로 제거하는 것으로 판단되었다.

<표 IV-7> 이상치 제거 기준에 따른 비중 비교

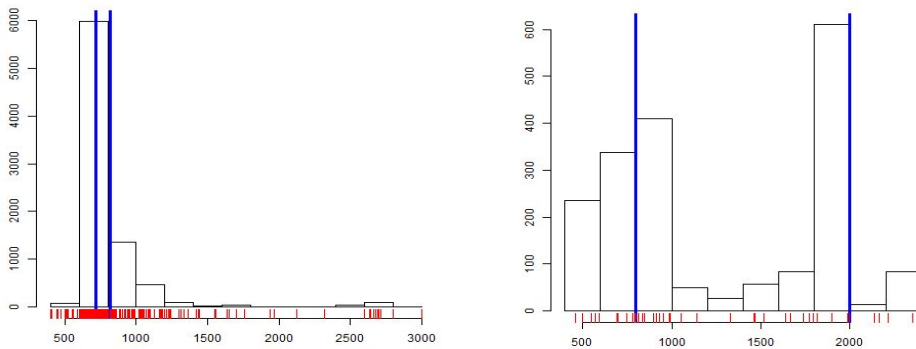
기준	상위 10% 초과 또는 하위 10% 하회 기준		상위 25% 초과 또는 하위 25% 하회 기준	
	이상치 제거 후(B)	비율(B/A)	이상치 제거 후(C)	비율(C/A)
원데이터(A)				
87,684	83,747	0.9551	79,967	0.9120
86,616	82,821	0.9562	79,008	0.9122
103,276	97,612	0.9452	92,492	0.8956
112,926	107,014	0.9476	102,310	0.9060

이상치 제거와는 별도로 분석 대상 기간 내 일관된 상품 선정 방법도 고려하였다. 즉, 2013년부터 2016년까지 4년 간 소멸하지 않고 계속 존재하였던 상품만을 추출해 보았다. 이 경우에는 해당 기간 중 새롭게 시장에 출시된 상품은 고려하지 않고 4년 간의 기간 동안 지속적으로 존재하는 상품만을 추출하는 것인데, 결과적으로 다른 방법에 비해 일관된 기준 하에서 가격의 비교를 할 수 있다는 특징이 있다. 다만, 4년 상존 상품만 추출하므로 해당 조합 수가 45,516개에 그친다. 참고로 상기 두 가지 조건을 모두 만족하는 상품과 업태의 조합 수는 42,640개로 나타났는데 이렇게 4년 지속 상품으

로 한정하고 이상치를 제거한 경우는 최종 분석 대상 상품 조합 수 162,367 개의 26.26%에 해당한다.

실제로 단가 기준으로 상, 하위 25%를 기준으로 삼아 이상치를 제거한 사례를 살펴보자. [그림 IV-3]은 대표적인 사례로서 좌측 그림은 이상치가 적절히 제거된 특정 상품의 경우이지만 우측 그림은 이상치를 제거하고도 쌍봉분포(bimodal distribution)를 나타내는 상품의 사례이다. 우측과 같은 형태의 가격 분포를 나타내는 세부 상품의 경우 옳은 가격을 찾기 어려우므로 분석에서 제외하였다.

[그림 IV-3] 이상치 제거 후 세부 상품의 가격 분포 사례



### 3. 스캐너 데이터 기반 물가지수 작성 방법

#### 제1절 적절한 지수의 선택

스캐너 데이터는 소매점에서 거래되고 있는 모든 상품에 대한 거래 가격 및 거래량 정보를 가지고 있다. 고빈도(시간별, 일별) 정보로서 신상품에 대한 정보까지 포함되어 있음을 고려하면 가급적 현실을 잘 반영할 수 있는 편제 방법을 이용해야 한다. 이러한 경우에는 라스파이레스 지수보다 톨크비스트(Tornqvist) 지수 또는 제본스(Jevons) 지수를 작성하는 것이 바람직하다. 또한 연쇄지수를 작성할 경우 연쇄지수를 작성함에 따라 발생 가능한 연쇄편의의 문제를 해결하기 위해서 GEKS 및 RWGEKS 지수의 작성도 고려해 볼 수 있다. 각 지수의 정의와 자세한 속성은 2장을 참고하기 바란다. 본 연구에서는 스캐너 데이터의 속성 상 시장에서 판매되는 제품들의 조합이 안정적이지 못하고 시점에 따라 변동한다는 점을 고려하여 연쇄지수를 활용하였으며 주별 단위 거래를 월별로 누적하여 물가지수를 산출하였다.

## 제2절 품질조정 기법의 적용

물가지수 편제를 위한 가격을 조사할 때, 조사규격이 변경됨에 따라 조사 상품 및 서비스의 가격이 달라지는 경우가 있다. 이러한 가격 변동에는 품질 변화와 순수한 가격변동이 혼재되어 있다. 따라서 이 가격 변동을 물가지수 편제에 반영할 때에는 순수한 가격 변동만 물가지수에 반영해 주어야 하는데 이러한 방법을 품질조정 기법이라고 한다. 빠른 기술발전으로 신구상품의 품질 차이가 크게 발생하는 경우 가격차이 모두를 품질에 기인한다고 판단한다면 가격변동이 지수에 반영되지 않는 문제점이 발생된다. 이를 해결하고자 제안된 방법으로는 헤도닉(Hedonic) 기법이 있다.

본 연구에서는 상품제품별로 다양한 가격자료를 수집하고 회귀모형을 구하여 변화된 품질요소별로 가격을 계량화하는 헤도닉(Hedonic) 방법을 적용하여 품질조정을 실시했다. 이 기법에 사용되는 회귀모형은 더블로그, 세미로그, 일반화선형모형 등이 있으며 소비자물가지수 산출 시 적용되는 모형은 종속 변수만 로그로 변환하는 세미로그형태를 취하는 것으로 알려져 있다. 헤도닉 기법을 적용하기 위해서는 품질을 정의할 수 있는 요인별로 다양한 가격대의 모델이 존재해야 하며 관측치도 충분히 많아야 한다. 헤도닉 기법은 기술 혁신이 품질 변화로 이어지는 상품에 주로 적용되는데, 실제로 통계청의 소비자물가지수 산출 시 컴퓨터본체, 노트북컴퓨터, 모니터, TV 등에 활용되고 있다. 품질조정에 관한 이론적 검토는 제2장을 참고하기 바란다.

### 4. 스캐너 데이터를 이용한 지수 시산 결과

#### 제1절 대한상의 스캐너 데이터 지수

본 연구에서는 대한상의에서 제공한 스캐너 데이터를 바탕으로 각종 물가지수를 시산하였다. 기초 데이터로 제공된 기준 품목 지수 81개 각각에 대해 톱크비스트 및 제본스 연쇄지수와 GEKS, RWGEKS 지수를 편제하되 업태를 기준으로 합산하여 최종 상위지수를 산출하는 방법과 16개 시도 지역을 기준으로 합산하여 최종 상위지수를 산출하였다. 또한 이상치를 제거한 후의 데이터를 가지고 시산해 보았으며, 4년 간 지속 상품으로 범위를 좁히고 이상치를 제거한 이후의 데이터를 기준으로 지수를 시산해 보았다. 시산한 각 개별 품목의 최상위 합산지수를 ‘대한상의 스캐너 데이터 총지수’라고 정의하였다. 총지수는 통계청 소비자물가지수 작성시 이용되는 품목별 가중치를



이용하여 라스파이레스 산식으로 계산하였다.

<표 IV-8>은 각 방식으로 지수를 산출한 결과를 요약한 표이다. 분석 대상 세부 상품을 4년 지속 상품으로 제한할 경우 모든 지수가 CPI와 매우 유사한 행태를 나타내었으며 지역별 합산 방식보다는 업태별 합산 방식을 통해 산출된 지수가 보다 안정적인 것으로 나타났다. 지수의 행태와 실무적인 면을 함께 고려해보면 톱크비스트나 RWGEKS 지수로 상위 지수를 작성하는 것이 유용하다고 판단된다.

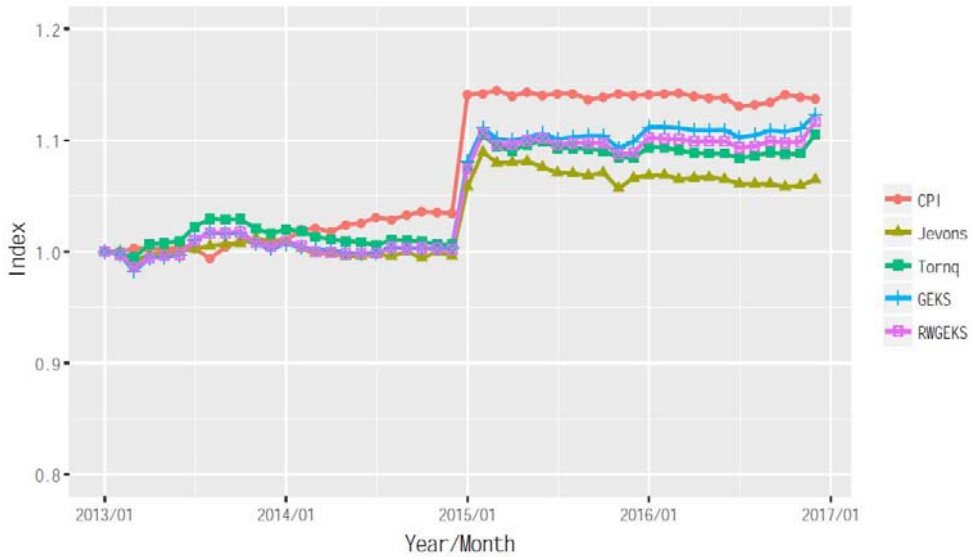
<표 IV-8> CPI와 각 지수의 비교

		제본스	톱크비스트	GEKS	RWGEKS
원자료	업태별 합산	다소 격차	대체로 CPI와 유사하나 GEKS지수의 수준이 CPI에 가장 근접		
	지역별 합산	점차 격차 확대	대체로 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			

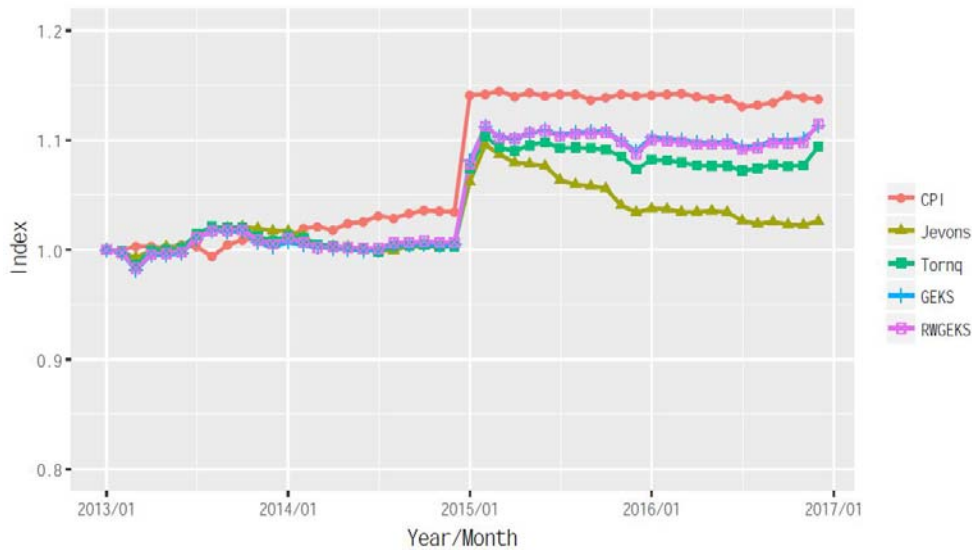
### 가. 원자료 기준 총지수

[그림 IV-4]는 대한상의 스캐너 데이터 총지수를 2013년 1월 지수를 기준으로 하여 상대적인 크기를 나타낸 것으로 업태별 기준으로 품목별 지수를 산출한 뒤 총지수를 산출한 것이다. 최초 제공받은 데이터에서 이상치를 제거하지 않고 원자료를 그대로 이용하여 지수를 산출하였다. 지수의 적절성을 평가해 보기 위해 CPI 지수도 같은 기준으로 환산하여 포함하였다. 이 그림에서는 모든 지수가 2015년 1월을 기점으로 급격히 상승한 모습을 보이고 있으며 다른 지수들에 비해 제본스 지수가 낮은 수준을 나타내고 있다. 톱크비스트, GEKS, RWGEKS 지수는 비슷한 수준을 나타내며 조금 더 CPI에 가까운 추세를 보이고 있다.

[그림 IV-4] 대한상의 스캐너 데이터 총지수(원자료, 업태별 합산)



[그림 IV-5] 대한상의 스캐너 데이터 총지수(원자료, 지역별 합산)

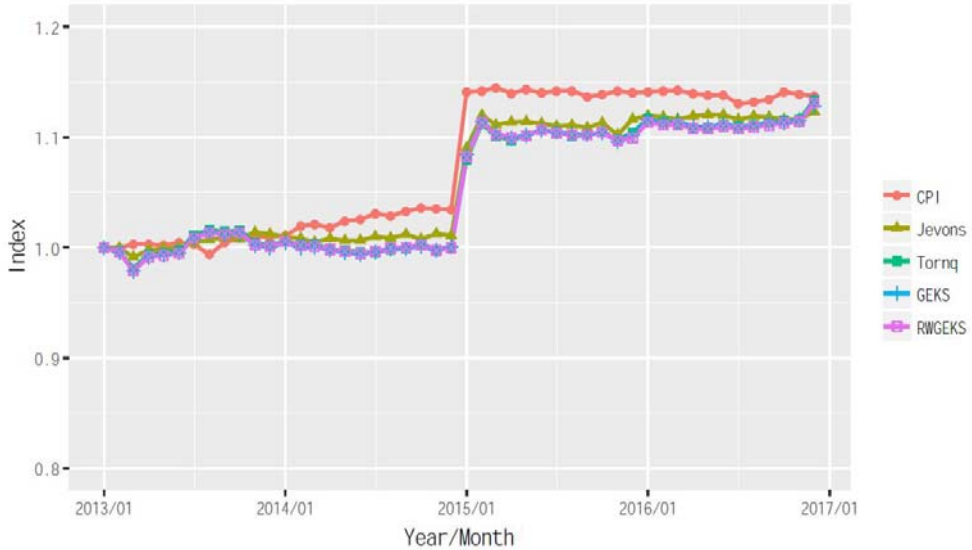


[그림 IV-5]는 시도 지역별 기준으로 품목별 지수를 산출한 뒤 총지수를 산출한 것이다. 앞의 시산 결과와 차이가 나는 점은 제본스 지수가 여타 지수에 비해 급격히 하락하고 있다. 이러한 편차는 표본이 매우 작은 지역의 가격 이상치에 기인하는 것으로 분석되었다. 예를 들어 제주 지역과 같이 스캐너 데이터에 포함된 세부 상품 수가 충분히 많지 않은 지역에서는 제본스 지수처럼 물량에 대한 고려가 없이 모든 상품에 동일 가중치가 부여될 경우, 한두 개 상품의 가격 급변동이 전체 지수에 크게 영향을 미칠 수 있기 때문

이다. 특히 이상치가 제거되지 않을 경우에는 이러한 영향이 더 크게 나타날 수 있음을 시사하고 있다.

[그림 IV-6]은 원데이터를 기준으로 하되, 4년 간 소멸하지 않고 계속 지속되었던 상품만을 가지고 업태별 기준으로 품목별 지수를 산출한 후 총지수를 산출한 것이다.

[그림 IV-6] 대한상의 스캐너 데이터 총지수(원자료, 4년 지속 상품, 업태별 합산)



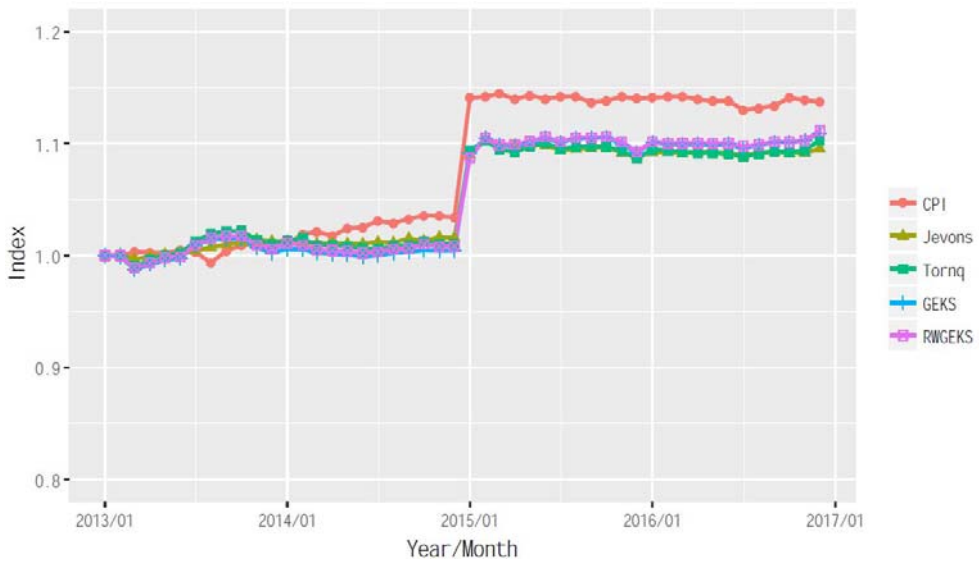
4년 간 지속된 데이터는 상대적으로 이상치가 많이 포함되어 있지 않는데다가 가격 변화도 크지 않고 안정적이어서 모든 지수의 움직임이 유사하게 안정적으로 나타났다. 4년 지속 상품으로 데이터를 축소 제한하여 지수를 산출할 경우에는 지역별 해당 상품 수가 너무 적어 지수 편제가 불가능하므로 이 방식의 총지수를 산출하지 않았다.

#### 나. 이상치 제거 후 산출 총지수

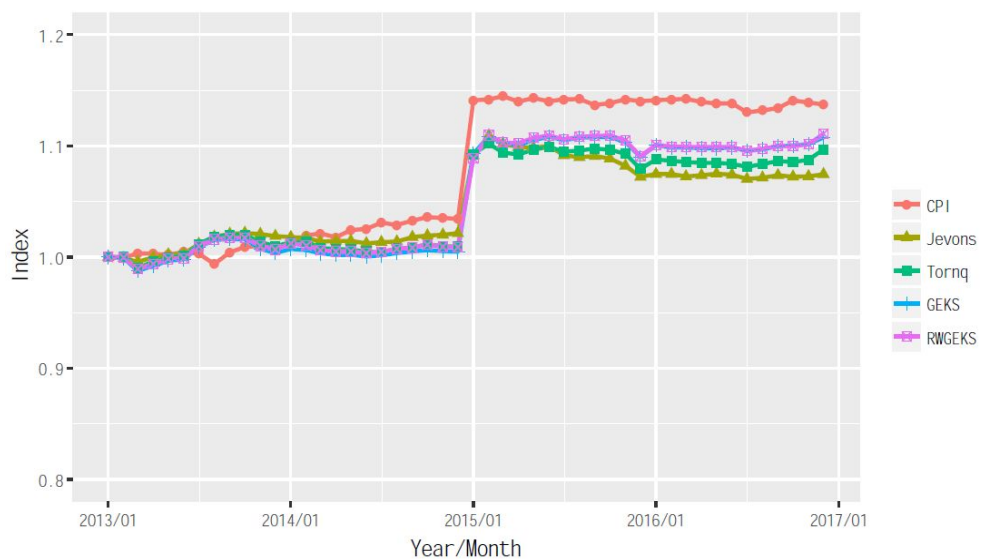
본 연구에서는 이상치 제거 방법으로 각 세부 상품별로 단가기준 상위 25%를 초과하거나 하위 25%를 하회하는 관측치를 절사하는 방식을 채택하였다. 이러한 방식으로 이상치를 제거한다는 전제하에서 총지수 산출 방법을 다음과 같이 구분하여 살펴보기로 한다. 업태별로 품목지수를 산출하고 합산, 지역별로 품목지수를 산출하고 합산, 4년 간 소멸하지 않고 계속 지속되었던 상품만을 가지고 업태별 품목지수를 산출하고 합산 등 세 가지 경우가

다. [그림 IV-7]부터 [그림 IV-9]를 통해서 각 지수의 움직임을 살펴볼 수 있다. 원자료를 그대로 이용하고 총지수를 산출하는 경우와 비교하여 볼 때, 이상치를 제거하였을 경우 상대적으로 지수 간 격차도 크지 않고 움직임도 더 안정적인임을 알 수 있다. 특히 [그림 IV-6]과 [그림 IV-9]를 비교해 보면 지역별로 품목지수를 산출하고 총지수를 산출하는 경우에는 이상치를 제거하였을 때가 훨씬 안정적인 움직임을 나타냈다.

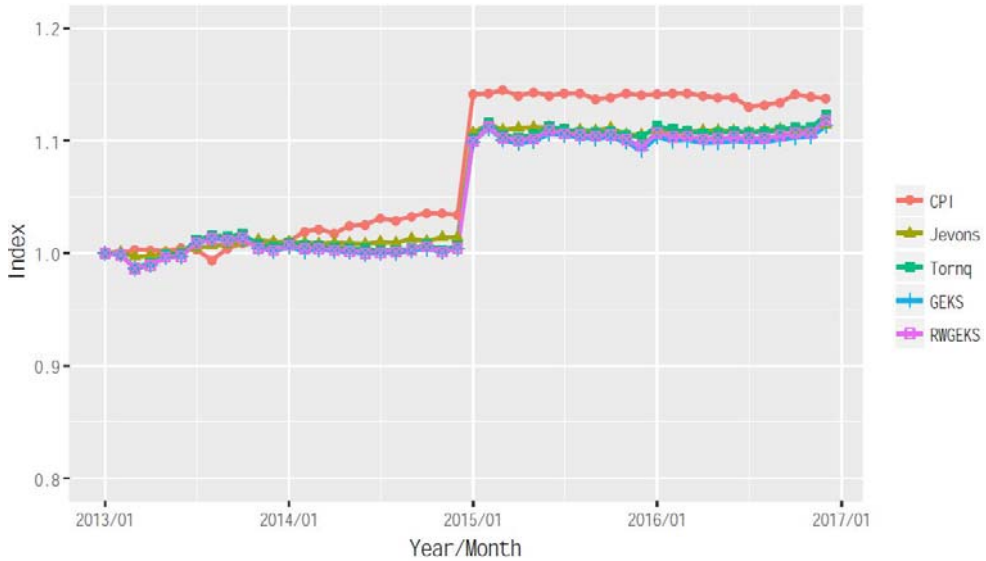
[그림 IV-7] 대한상의 스캐너 데이터 총지수(이상치 제거, 업태별 합산)



[그림 IV-8] 대한상의 스캐너 데이터 총지수(이상치 제거, 지역별 합산)



[그림 IV-9] 대한상의 스캐너 데이터 총지수(이상치 제거, 4년 지속 상품, 업태별 합산)



#### 다. CPI 전기대비 증감률과의 상관계수

앞서 살펴본 대로 원자료를 이용할 경우와 이상치를 제거할 경우 각 편제 방법에 따른 지수는 대체로 CPI와 유사한 움직임을 나타내고 있다. 각 지수가 CPI와 얼마나 유사한 움직임을 나타내는지 종합적으로 평가해 보기 위해 해당 기간 중 CPI의 전기대비 증감률과 각 지수의 전기대비 증감률 간의 상관계수를 계산해 보았다. <표 IV-9>에는 CPI의 전기대비 증감률과 데이터의 가공 방식별, 각 지수별 전기대비 증감률과의 상관계수가 제시되어 있는데, 대체로 높은 수준을 나타내고는 있으나 원지수에 비해서는 이상치를 제거한 경우가 더 높은 것으로 나타났다. 특히, 4년 동안 시장에서 소멸하지 않고 지속적으로 판매되고 있는 상품만으로 데이터를 제한한 경우의 지수가 CPI와의 상관계수가 더 높게 나타났다. 이는 대표 상품에 대해 표본조사를 통해 지속적으로 가격을 조사하고 이를 반영하여 CPI 산출 방식과 스캐너 데이터 상 지속 상품만으로 지수를 산출하는 방법이 개념적으로 유사하기 때문이다.

<표 IV-9> CPI 전기대비 증감률과의 상관계수

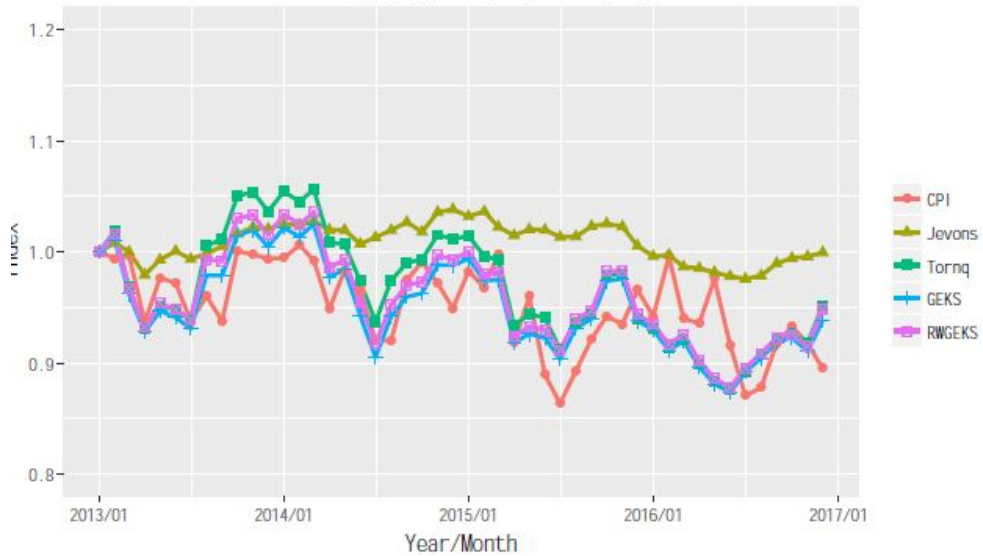
	제본스	통크비스트	GEKS	RWGEKS
원자료(지역별 합산)	0.745	0.791	0.800	0.766
원자료(업태별 합산)	0.796	0.800	0.806	0.785
이상치 제거지수(지역별 합산)	0.894	0.897	0.901	0.874
이상치 제거지수(업태별 합산)	0.937	0.902	0.902	0.881
4년 지속 상품지수(업태별합산)	0.856	0.798	0.822	0.812
4년 지속 상품, 이상치 제거지수 (업태별합산)	0.951	0.905	0.906	0.902

한편, 대한상의 스캐너 데이터에 포함되어 있는 세부 품목별 지수와 CPI와의 상관계수는 <별첨 3>에 제시되어 있다. 원자료를 이상치 제거 없이 이용하되 업태별로 합산한 경우, 지역별로 합산한 경우, 4년 지속 상품만을 대상으로 지수를 산출한 경우로 나누어 상관계수를 계산하였다. 또한 이상치를 제거한 뒤, 같은 방식으로 업태별로 합산한 경우, 지역별로 합산한 경우, 4년 지속 상품만을 대상으로 지수를 산출한 경우로 나누어 상관계수가 수록되어 있다. 스캐너 데이터의 세부 품목 중 CPI와의 상관계수의 값이 0.5를 상회하여 향후 현행 서베이 방식을 대체하거나 보완할 수 있는 품목 후보군으로 검토할 수 있는 품목으로서는 담배, 양념소스, 당면, 소주, 우유, 물엿 등으로 분석되었다.

#### 라. 일부 품목 CPI 대체 후 총지수

대한상의 스캐너 데이터 총지수를 산출함에 있어서 가장 많은 제약을 준 경우에도 동 총지수가 CPI의 움직임과 다소 상이하게 나타나는 점을 개선하기 위해 하위 품목 중 CPI와의 괴리가 두드러진 일부 품목을 선택하여 CPI로 대체하여 총지수를 산출해 보았다. 즉, 이상치를 제거하고 4년 지속 상품의 업태별 지수 산출하되 CPI와 괴리가 두드러진 일부 품목을 CPI 지수로 대체한 것이다. 괴리가 두드러진 품목으로는 소시지, 햄 및 베이컨, 우유, 비스킷, 스낵과자, 파이, 기초화장품 등으로 전체 상품 중 23.4%의 비중을 차지하는 품목들이다. [그림 IV-10]은 이러한 방식으로 가공한 총지수를 여러 가지 방법으로 산출한 것이다. 모든 지수의 경우 CPI와 격차가 거의 없게 나타났다.

[그림 IV-10] 대한상의 스캐너 데이터 총지수(이상치 제거, 4년 지속 상품, 업태별 합산, 일부품목 CPI 대체 시)



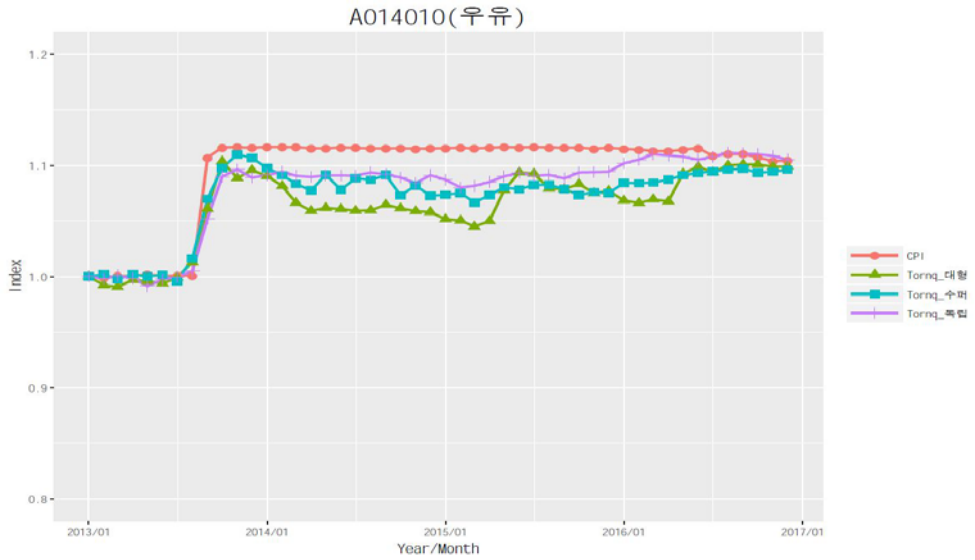
## 마. 품목별 지수

이상치를 제거하고 4년 지속상품으로 제한한 스캐너 데이터로 하위 품목별 지수를 산출하였을 때, 각 품목지수의 움직임은 대체로 CPI와 유사했지만 몇 가지 품목의 경우에는 격차를 나타내었다. 대표적 사례를 품목별로 살펴보면 다음과 같다.

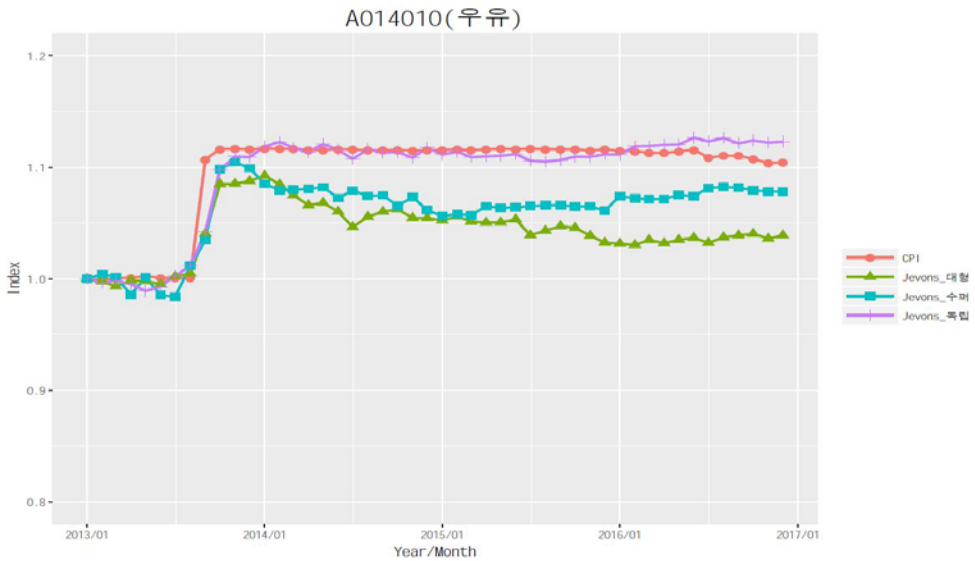
[그림 IV-11]은 우유 품목의 지수를 업태별로 나누어 산출한 것으로 위 그림은 톱크비스트 방법을 적용한 지수들이고 아랫 그림은 제본스 방법을 적용한 지수들이다. 그림에서 보듯이 위 그림의 독립수퍼의 제본스 지수가 CPI와 가장 유사한데, 이는 유통에 따른 가격 차이가 반영된 사례로 볼 수 있다. 즉, 아랫 그림의 제본스 지수를 살펴보면 대형마트나 수퍼체인의 경우 할인판매 정책 등에 의한 가격 하락분이 많이 반영된 데 비해 독립수퍼는 상대적으로 가격정책 변화폭이 크지 않은 데 기인하는 것으로 추정된다. 제본스 지수가 물량을 고려하지 않는 동일 가중치 부여 방식이라는 점을 고려하면 우측 그림에서 이러한 현상이 더욱 뚜렷한 것을 이해할 수 있다.

[그림 IV-11] 품목별 지수 - 우유

(a) 톱크비스트 지수



(b) 제본스 지수

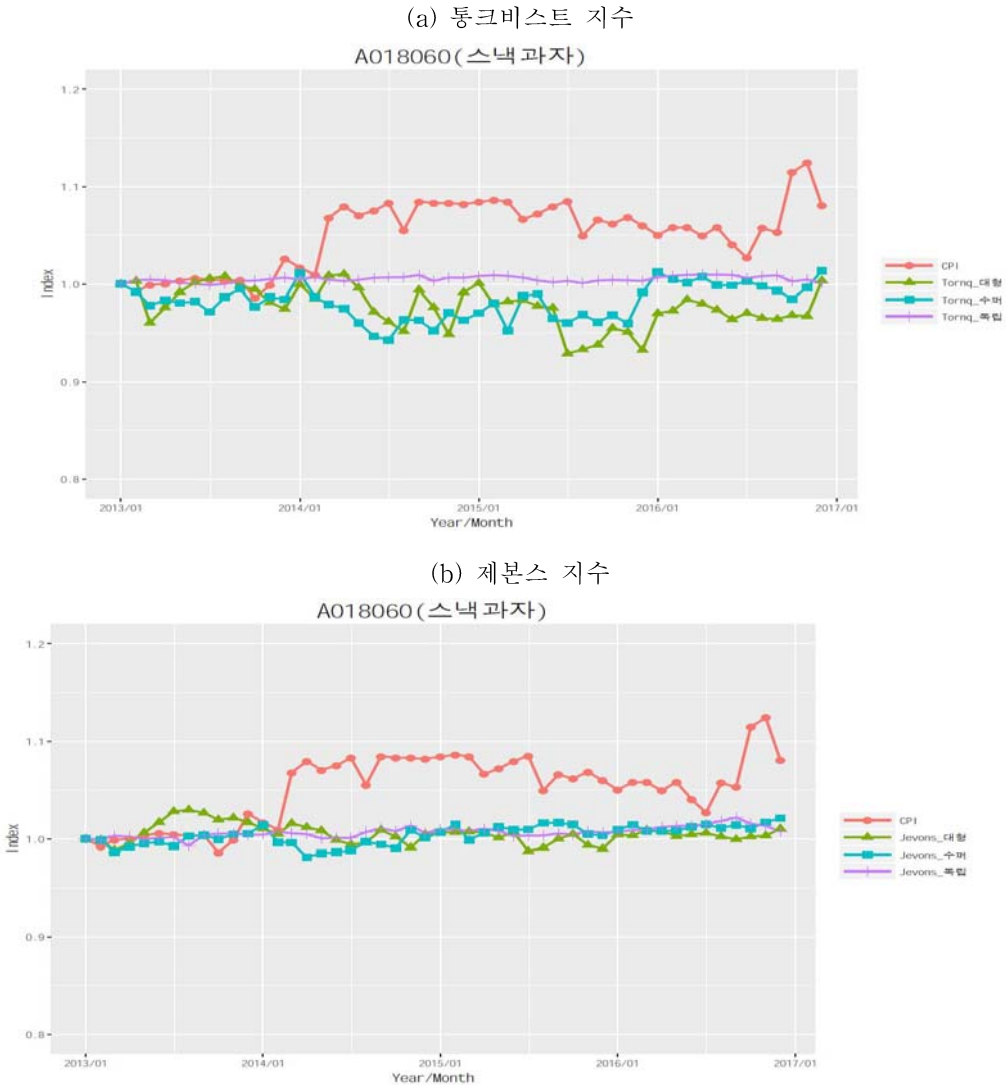


[그림 IV-12]는 스낵과자 품목의 지수를 업태별로 나누어 산출한 것으로 위 그림은 톱크비스트 방법을 적용한 지수들이고 아래 그림은 제본스 방법을 적용한 지수들이다. 이 품목은 스캐너 데이터를 기반으로 산출한 지수들의 움직임이 상당히 안정적인 데 비해, CPI의 가격이 다소 큰 폭으로 변동하여 괴리



가 발생하는 사례 중 한 가지이다. 스낵과자 외에 키워드 분류로 상품을 분류해서 품목별 지수를 산출하는 경우에는 상대적으로 가격 변동폭이 미미했는데, 이는 키워드로 분류한 상품이 대체로 4년 동안 상대적으로 많이 판매된 상품일 가능성이 높고 이에 따라 가격도 안정적인데 기인한다.

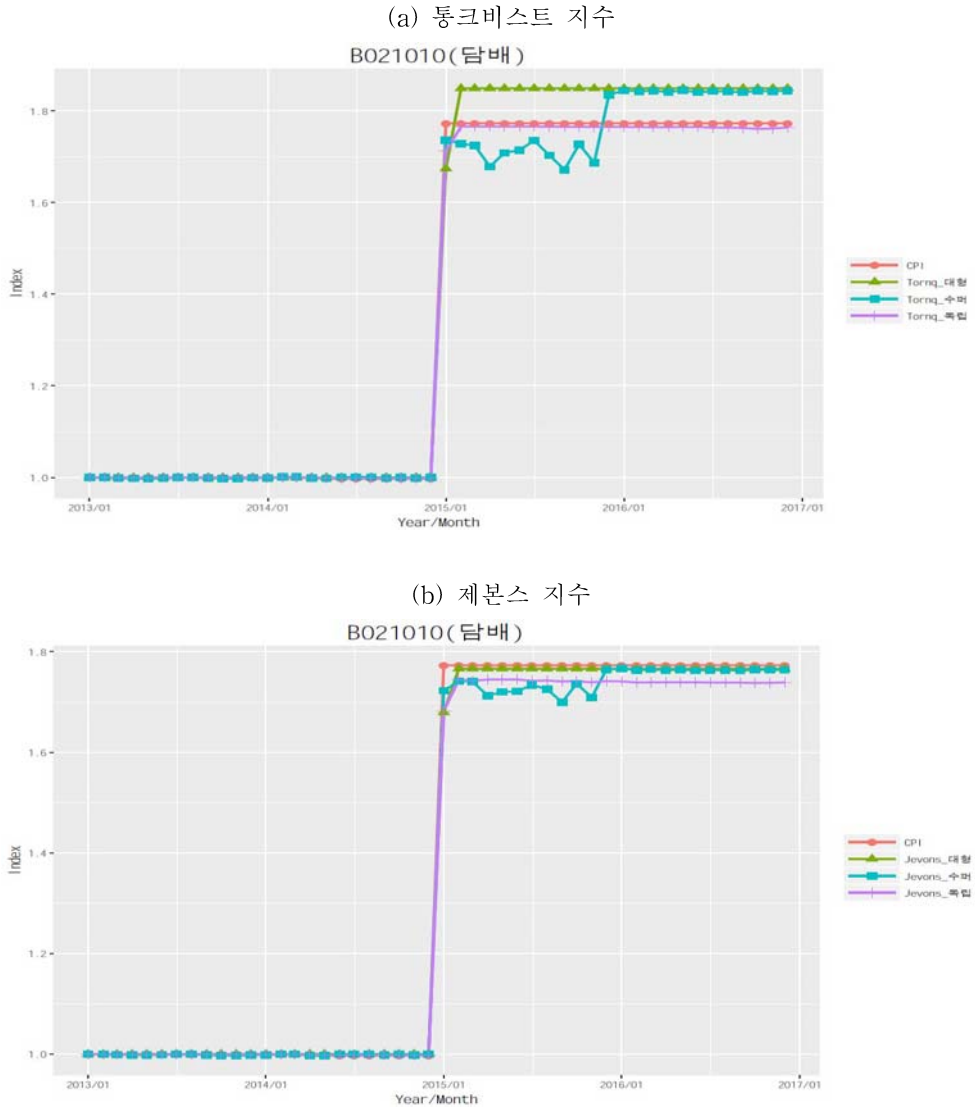
[그림 IV-12] 품목별 지수 - 스낵과자



[그림 IV-13]은 담배 품목의 지수를 업태별로 나누어 산출한 것으로 위 그림은 톱크비스트 방법을 적용한 지수들이고 아래 그림은 제본스 방법을 적용한 지수들이다. 담배 가격은 2015년에 접어들면서 정책적으로 가격을 인상

하였는데 스캐너 데이터에는 업태별로 가격 인상폭이 다소 다르게 반영되는 것으로 나타났다.

[그림 IV-13] 품목별 지수 - 담배



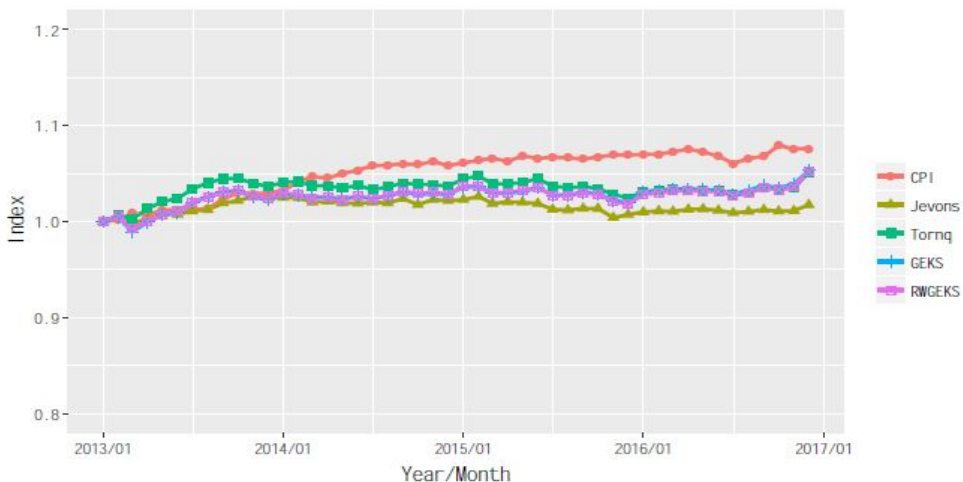
[그림 IV-13]의 윗 그림의 물량을 고려한 통크비스트 지수의 경우 대형마트의 수준이 CPI보다 높게 나타났는데 이는 업태별 주력상품 차이에 기인한다. 즉, 대형마트에서는 수입담배의 판매가 거의 없거나 미미하다는 점을 고려하면 일률적인 가격 인상에 의해 계산되는 인상률은 저가의 담배의 경우 더 크게 반영되어 결과적으로 대형마트의 인상폭이 큰 것처럼 나타나게 되는 것이다. 다만, 물량 고려 없이 동일 가중치를 반영한 체본스 지수의 경우

에는 업태별 차이도 상대적으로 미미한 것으로 나타났다. 한편, 슈퍼 체인의 경우 비정상적인 가격 변동이 나타났는데, 해당 데이터를 세부적으로 살펴본 결과 부정확한 가격이 다수 포착되었다. 담배는 정책적인 가격 변동 외에 크게 가격에 영향을 미치는 외부 요인이 없는 품목이므로 왜곡된 가격 정보를 쉽게 포착할 수 있다. 이러한 점은 스캐너 데이터와 같은 빅데이터 생성과 입수, 가공 단계 등에서 데이터 신뢰도 제고를 위해 조금 더 관심을 기울여야 한다는 점을 시사한다.

## 제2절 가공식품 총지수 (통계청 품목 기준)

제1절에서는 대한상의에서 제공받은 스캐너 데이터 품목 81개 전체를 이용하여 총지수를 산출하되 여러 가지 기준과 편제 방법을 적용하여 비교 분석해 보았다. 본 연구에서는 통계청의 상위지수와 직접적으로 비교하기 위해 가공식품 총지수를 산출해 보았다. 다만, 대한상의 스캐너 데이터의 품목 중 통계청 기준의 가공식품 하위분류에 해당하는 총 품목수 70개 중 58개에 그쳤다. 따라서 58개 지수는 스캐너 데이터를 이용해 산출한 품목지수를 이용하고 나머지 12개 품목지수는 통계청의 CPI지수를 그대로 이용하여 가공식품 지수를 산출하였다.

[그림 IV-14] 가공식품 총지수(원자료)

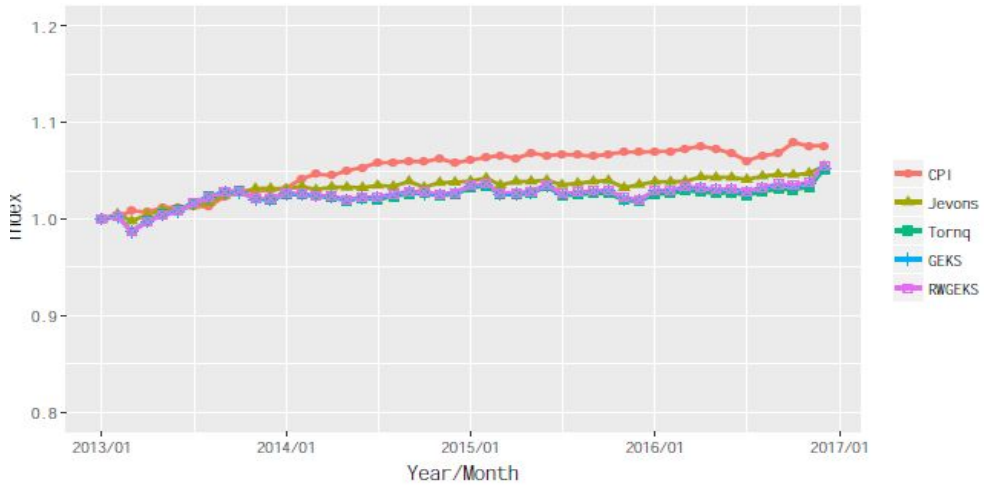


가공식품 상위지수를 산출할 때에는 통계청의 가중치를 이용하여 라스콕이레스 방식으로 합사하여 산출하였다. [그림 IV-14]는 스캐너 데이터 원자료를 이용하여 가공식품 총지수를 산출한 것이다. 원자료를 이상치 제거 없이

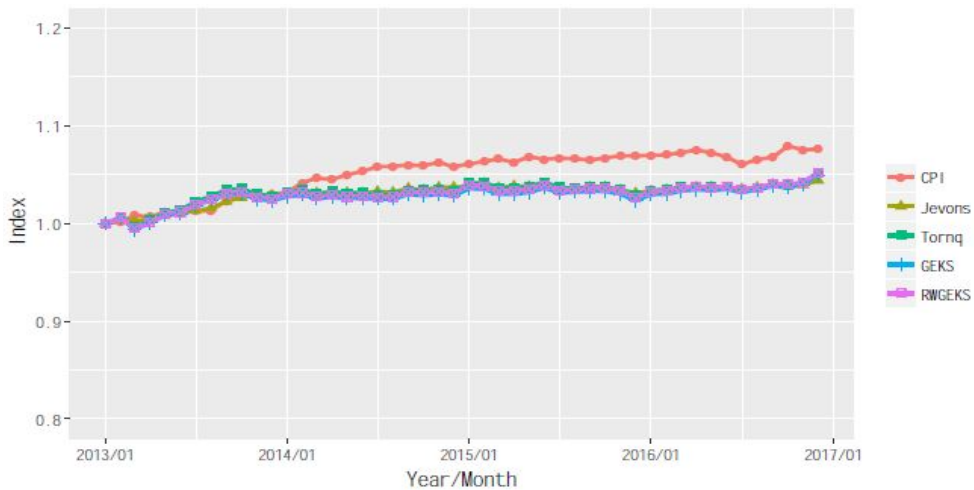
그대로 사용한 경우인데도 불구하고 담배와 같은 기타 품목이 빠져있으므로 앞서 살펴 본 스캐너 데이터 총지수에 비해 가격 변동 폭이 크지 않고 지수 간 격차도 크지 않은 것으로 나타났다.

[그림 IV-15]는 스캐너 데이터 원자료 중 4년 동안 소멸하지 않고 지속적으로 판매된 세부 상품들만을 이용하여 가공식품 총지수를 산출한 것이다.

[그림 IV-15] 가공식품 총지수(원자료, 4년 지속 상품)



[그림 IV-16] 가공식품 총지수(이상치 제거)

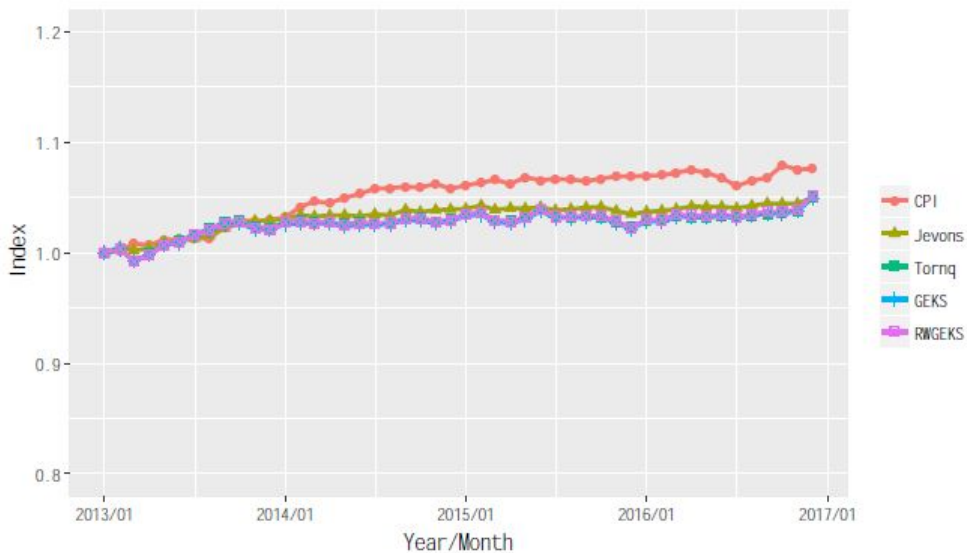


[그림 IV-16]은 스캐너 데이터 원자료 중 세부 상품들의 단가를 기준으로 상위 25%를 초과하거나 하위 25%를 하회하는 상품들을 이상치로 간주하여 제거하고 가공식품 총지수를 산출한 것이다. 제본스, 톱크비스트, GEKS,

RWGEKS 등 각 방법을 적용한 지수 간 격차는 크게 줄어들었으나 CPI와의 격차는 줄어들지 않았다. 이는 스캐너 데이터는 업태별, 상품별 다양한 가격 정보의 종합이고 CPI는 대표성이 뚜렷한 대표 가격이라는 태생적인 차이에 기인한다고 볼 수 있다.

[그림 IV-17]은 스캐너 데이터 원자료 중 이상치를 제거하고 4년 지속 세부 상품들만으로 제한하여 가공식품 총지수를 산출한 것인데, [그림 IV-17]과 마찬가지로 CPI와의 격차는 줄어들지 않았다. 다만, 앞서 분석하였듯이 증감률 기준으로 볼 때에는 CPI와 유사한 행태를 보인다는 것을 예상할 수 있다.

[그림 IV-17] 가공식품 총지수(이상치 제거, 4년 지속 상품)

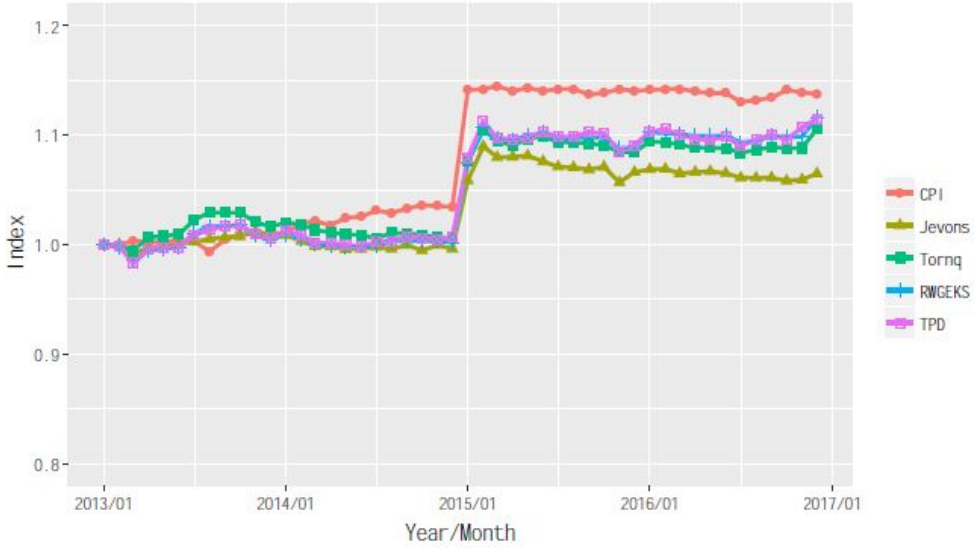


### 제3절 품질조정 지수

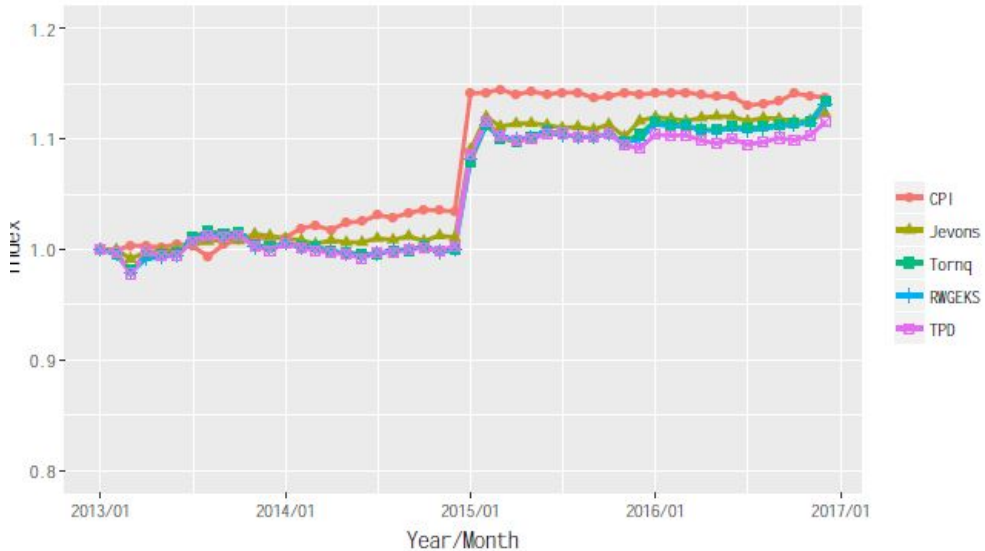
앞서 정리해 본 de Haan and Krsinich(2014)의 TPD 품질조정 방법을 이용하여 대한상의 스캐너 데이터의 모든 품목에 대해 품질조정을 실시하고 총지수를 산출한 결과, 거의 모든 품목에서 RWGEKS 방식으로 산출한 지수와 매우 유사하게 나타났다. 분석 대상 품목의 경우 품질의 변화가 극명하지 않아 여타 방법과 큰 차이를 나타내지 않았을 가능성과 더불어 시간 상품 더미변수를 포함하고 있는 회귀모형에 의거한 추정 방법이 적절한 윈도우를 설정하여 이를 이동시키면서 지수를 산출하는 RWGEKS 방법과 결과적으로 비슷한 평활화 효과를 반영하게 되었을 가능성이 있다. [그림 IV-18]부터 [그

림 IV-21]는 각각 원자료, 4년 지속 상품 제한할 경우, 이상치 제거, 이상치를 제거한 뒤 4년 지속 상품으로 제한할 경우의 대한상의 스캐너 데이터 총지수가 품질조정 방법까지 반영하여 제시되어 있다. 총지수와 관련하여 여러 가지 경우에도 품질조정 방법을 반영한 지수는 여타 지수와 큰 차이를 보이지 않음을 알 수 있다.

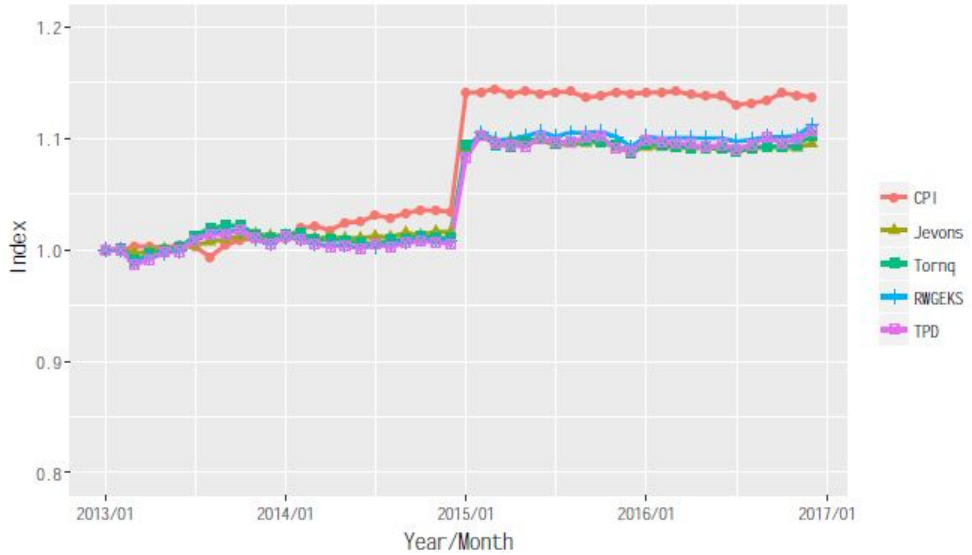
[그림 IV-18] 대한상의 스캐너 데이터 총지수(원자료, 품질조정 포함)



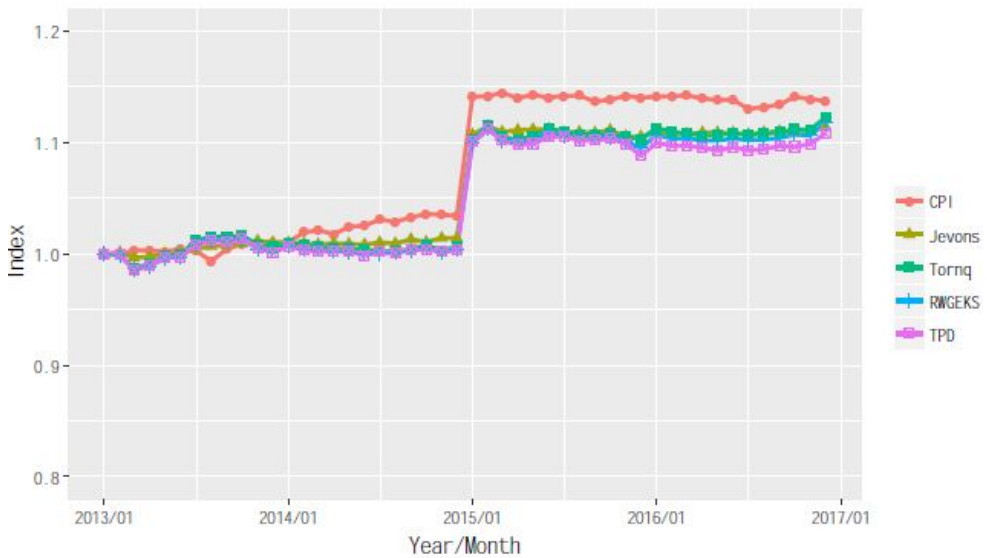
[그림 IV-19] 대한상의 스캐너 데이터 총지수(원자료, 4년 지속 상품, 품질조정 포함)



[그림 IV-20] 대한상의 스캐너 데이터 총지수(이상치 제거, 품질조정 포함)



[그림 IV-21] 대한상의 스캐너 데이터 총지수(이상치 제거, 4년 지속 상품, 품질조정 포함)



## 5. 소결

본 연구는 대한상공회의소의 협조를 바탕으로 POS시스템 기반의 스캐너 데이터를 입수하여 진행한 것으로서 빅데이터를 이용한 가격 지수의 편제라

는 점에 있어서 매우 유용하고 의미 있는 연구다. 특히 무엇보다도 데이터를 입수하고 가공하는 과정에서 양질의 데이터의 확보가 전제되어야 한다는 점을 확인할 수 있었다. 이에 따라 데이터 정제 과정의 중요성이 더욱 부각된다. 실제로 본 연구에서도 이상치를 적절하게 제거하고, 4년 지속 상품 제한하였을 때, CPI와 가장 유사한 패턴의 지수를 산출할 수 있었다. 다만, 실제로 어느 정도까지의 정제과정을 거쳐야 하느냐는 많은 실무적인 경험을 통해 결정되어야 할 것으로 보인다. 지수 산출 방법을 비교해 보면 물량 정보를 반영한 톱크비스트, GEKS 및 RWGEKS 지수가 단순 가중치를 반영한 제본스 지수에 비해 안정적이고 일관된 행태를 보이는 것으로 분석이 되었다. 더불어 품질조정을 실시한 지수까지 산출하여 그 필요성을 검토해 보았는데 실제 하위 품목지수나 총지수의 경우 모두 톱크비스트 또는 RWGEKS와 매우 유사한 움직임을 보이면서 품질조정에 따른 부가적인 효과가 극명하게 드러나지 않는 것으로 나타났다. 스캐너 데이터를 이용하여 분석한 품목군이 전자제품과 같이 기술혁신의 영향으로 품질의 변화가 짧은 시간에 뚜렷하게 나타나는 품목이 아니라는 것도 한 가지 원인으로 추정되었다.

지표의 활용 측면에서는 본 연구를 통해 산출한 다양한 지수들을 CPI를 대체하는 데 직접 활용할 것인지, 아니면 보조지표로서 활용할 것인지 선택의 문제에 당면할 수 있는데 데이터의 성격이나 포괄범위 등을 고려할 때 당장 CPI를 대체하여 활용하기에는 다소 무리가 따를 수 있다고 판단된다. 최대한 정제한 스캐너 데이터를 이용하더라도 스캐너 데이터를 기반으로 한 가공식품 총지수가 여전히 CPI와 다소 차이를 나타내는 점은 이를 뒷받침한다. 특히 개념적으로 볼 때에도 CPI는 대표 가격으로 인식되며, 스캐너 데이터는 실제 시장에서 거래되는 다양한 가격의 종합으로 인식되기 때문에 실제 지수공표의 측면에서는 혼동을 줄이기 위해 신중하고 지속적인 검토가 필요하다는 점도 중요한 고려사항이라고 할 수 있다. 더불어 스캐너 데이터의 특성상 세부 상품의 생멸이 거듭되므로 기준년 개편처럼 고정 상품 선정 시기 및 기준 설정이 필요하다.



## V. 소비자물가지수 개편과 지수 산식 검토

통계청은 소비자물가지수의 개편주기를 단축하여 소비자물가지수의 현실 반영도를 높였으나 소비자물가지수의 소급 수정 빈도가 증가하여 지수의 안정성이 저해되는 문제가 발생하였다. 소비자물가지수는 각종 계약, 임금협상 등의 기초가 되므로 소비자물가지수의 잦은 개정은 이용자의 혼란을 초래한다. 캐나다 통계청의 경우 소비자물가지수를 공표한 후 이를 수정하지 않고 있다 (Huang, Wimalaratne and Pollard 2015). 본 절에서는 소비자물가지수의 개편과 연관된 지수 산식을 검토한다.

### 1. 소비자물가지수 개편현황과 문제점

통계청은 소비자의 기호, 구매패턴 등 소비자물가지수의 현실 반영도를 보다 높이고자 2013년 12월 통계 발표시부터 기준년 개편의 주기를 5년에서 2~3년 주기로 단축하였다.<sup>6)</sup> 품목 조정을 포함한 전반적인 지수 개편은 5년(0, 5년) 주기로 하면서 중간년도(2, 7년)에 가중치만 별도 개편하는 방식을 채택하였다. 지수의 산식으로는 전과 동일하게 순수 Laspeyres식을 적용하고 있다.

소비자물가 품목의 가중치는 가계동향조사 소비지출항목 지출액을 이용하여 산출된다. 구체적으로 살펴보면 소비자물가지수 품목과 가계동향조사 항목을 대응시켜서 해당 지출액을 적용하여 가중치가 산정된다. 소비자물가지수 품목과 가계동향조사 항목간 대응이 불완전한 경우 관련 정보를 이용하여 배분하거나 유사품목에 포함시킨다. 또한 지역별·연도별 편차가 크거나 소비구조 변화가 심각한 경우 특성을 반영하여 가중치를 조정한다.

소비자물가지수에 대해 기준년 개편을 실시하면 과거 자료를 소급 개편하게 된다. 2013년 말 진행된 기준년 개편을 중심으로 살펴보면 다음과 같다. 2013년 12월 기준년 개편시 2010년 기준 가중치를 사용한 2012년 1월 ~ 2013. 11월까지의 구지수를 2012년 기준 가중치를 사용한 신지수로 변경한다. 즉 중간년도(T) 기준년 개편으로 T-1년, T년의 통계가 동시에 개편된다. 이와 같이 2~3년 주기로 기준년을 개편하면 소비자물가지수의 소급 개편되는 기간이 10년중 8년이 되어 소비자물가지수의 이용자가 불편하게 된다. 이와 같이 소비자물가지수가 빈번하게 소급 개편되는 주요 이유 중 하나는 Laspeyres 지수를 이용하기 때문이다.

6) 2013년 캐나다 통계청도 개편주기를 4년에서 2년으로 변경하였다.

Laspeyres 지수의 대안적 지수로는 Lowe지수와 Young지수 등이 있는데 이들 지수를 적용하는 경우 지수 개편에 따른 소급 개편의 빈도를 최소화할 수 있다. 다음은 각종 산술형(arithmetic) 지수인데  $P$ 는 가격,  $Q$ 는 수량,  $w$ 는 가중치,  $t$ 는 시점,  $i$ 는 품목이다. 0는 기준년 개편 시점이고,  $b$ 는 변경된 기준년이다. 이를 보면 Lowe 지수의 경우 수량이 고정된 지수인 반면 Young 지수의 경우 지출 비가 고정된 지수이다.

$$\text{Laspeyres 지수} = \sum_{i=1}^n \left( \frac{P_t^i}{P_0^i} \times w_0^i \right) = \frac{\sum_{i=1}^n (P_t^i \times Q_0^i)}{\sum_{i=1}^n (P_0^i \times Q_0^i)} * w_0^i = \frac{P_0^i Q_0^i}{\sum_{i=1}^n P_0^i Q_0^i} \quad (26)$$

$$\text{Lowe 지수} = \sum_{i=1}^n \left( \frac{P_t^i}{P_0^i} \times w_b^i \right) = \frac{\sum_{i=1}^n (P_t^i \times Q_b^i)}{\sum_{i=1}^n (P_0^i \times Q_b^i)} * w_b^i = \frac{P_0^i Q_b^i}{\sum_{i=1}^n P_0^i Q_b^i} \quad (27)$$

$$\text{Young 지수} = \sum_{i=1}^n \left( \frac{P_t^i}{P_0^i} \times w_b^i \right) * w_b^i = \frac{P_b^i Q_b^i}{\sum_{i=1}^n P_b^i Q_b^i} \quad (28)$$

여기서 소개된 지수는 기하적(geometric) 지수로도 전환되어 검토되고 있다.

$$\text{기하지수} = \sum_{i=1}^n \left( \frac{P_t^i}{P_0^i} \right)^{w_b^i} \quad (29)$$

지수는 최상위 지수와 비교되고 있으며 여러 가지 변형도 제시되고 있다. Lent and Dorfman (2009)는 미국 데이터로 최상위(superlative) 지수에 근접한 지수로 산술 및 기하 Laspeyres 식의 가중평균을 제시하였다. Armknecht and Silver (2012)는 기하 Young 지수와 Lowe 지수 또는 Young 지수와 기하 Lowe 지수를 기하평균한 것이 최상위지수에 근접하다고 판단하고 있다.

## 2. 소비자물가지수 개편방향

가계동향조사 등 소지지출 관련통계가 시차를 두고 작성되어 각국 통계청은 소비자물가지수 개편에 따른 문제를 가지고 있다. 통계청 (2014)은 2014년 국가통계위원회에 보고된 “2015년 기준 소비자물가지수 개편계획”을 통

해 현행 중간년도 소비자물가지수 가중치 개편시 사용된 계산식의 문제점을 인식하고 과거 시계열이 수정하는 문제를 개선하는 방안을 제시하였다. 그 방안은 중간년 가중치의 상위단계 개편시 가격보정 가중치와 월중첩 지수접속법을 적용하는 Lowe 방식으로 소비자물가지수 작성하는 것이다. 이 방법을 적용하는 경우 과거 시계열 소급 적용 없이 소비자물가지수를 작성할 수 있다.

2007년 9월 시행된 UNECE/ILO 조사에 따르면 47개국중 미국(2년), 캐나다(2년), 호주(5~6년) 등 32개국이 Lowe 지수(price-update weights)를 이용하고 15개국이 Young지수(original weights)를 쓰는 것으로 나타났다. 한편 독일, 일본은 5년 주기로 Laspeyres 지수를 이용하여 소급적용을 하고 있다(Hansen 2007, Armknecht and Silver 2012).<sup>7)</sup> 우리나라의 경우 2~3년 주기로 Laspeyres 지수를 이용하여 소급적용을 하고 있다. 한편 우리나라 생산자 및 수출입물가지수의 경우 Lowe 지수를 적용하고 있다.

Lowe지수와 Young 지수의 특성을 비교해 보면 다음과 같다. ILO(2004)는 “가격에 장기적 추세가 있다면 Lowe 지수가 Young 지수보다 크게 나타난다.”고 정리하였다. Hansen (2007)은 Lowe 지수와 Young 지수를 덴마크 데이터에 대해 비교하였는데 비교 결과 Lowe 지수가 Young 지수에 비해 상향편의가 있는 것으로 나타났다. 또한 Lowe지수와 Young지수를 비교하였는데 이를 보면 품목의 대체 탄력성이 1에 근접하면 Young지수가 우수하지만 대체 탄력성이 낮아지면 Lowe지수가 우수한 것으로 나타났다. Hansen(2007)은 지수간 비교 결과를 바탕으로 Lowe지수를 권고하고 있다. 이는 Lowe지수는 Young지수에 비해 상향편의가 있지만 순환성, 시점 역전 테스트 등을 만족하여 공리적 속성을 만족하고 있고 다수의 국가가 Lowe지수를 이용하므로 국제비교가 가능하기 때문이다. 한편 Young지수는 Lowe지수에 비해 상향편의는 적지만 지수의 공리적 속성을 가지지 못한다는 한계가 있다.

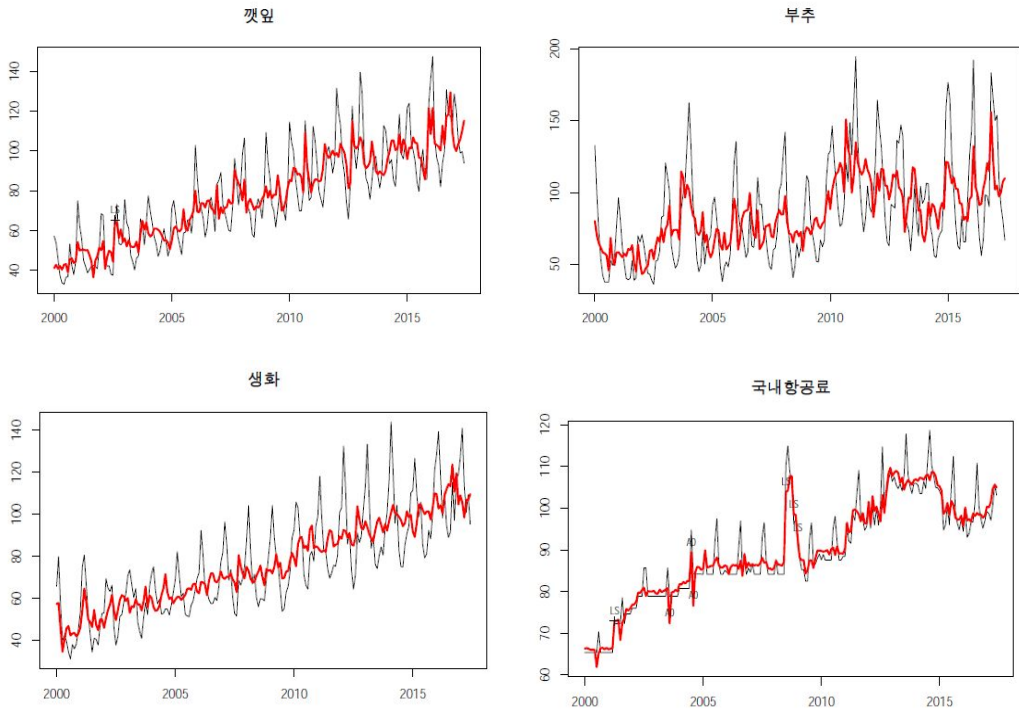
따라서 통계청의 경우도 기존 보고(통계청 2014)에서 제시한 바와 같이 2018년 완료 예정인 2017년 개편에서 Lowe지수를 이용하여 중간년 기준년 개편을 실시하여 이용자의 편의를 높일 필요가 있다. 다만 Lowe지수에서 월중첩 지수접속법을 적용하는 경우 특정 월의 가격을 이용하는데 일부 품목에서 계절성이 나타나 가중치의 대표성이 낮아질 수 있으므로 이를 보정할 필요가 있다.

통상 물가는 계절성이 없다고 판단하고 계절조정을 하지 않고 있다. 일부

7) 지수 선택은 장단점이 혼재하므로 어느 지수가 우수하다고 판단하기 어렵다. 또한 각국 통계청에서 지수 선택의 이유를 제시하지 않고 있다.

계절 품목의 경우 조사에서 계절성을 처리를 하지만 품목별로 보면 계절성이 나타나고 있다.<sup>8)</sup> <그림 V-1>은 껌, 부추, 생화, 국내항공료 등의 소비자물가지수의 원계열과 계절조정계열이다. 이들 품목지수의 경우 지수내 계절변동이 존재하는 것으로 나타나 계절조정을 실시했다.<sup>9)</sup> 여기서 붉은 적색선은 X-13ARIMA-SEATS로 구한 계절조정계열이고, 흑색선은 원계열이다. 만약 특정월의 월중첩법으로 지수를 작성할 때 계절변동이 포함된 가격을 그대로 이용하는 경우 가중치의 문제가 발생할 소지가 있으므로 이들 품목 가격 적용시 보정이 필요하다.

[그림 V-1] 계절변동이 있다고 판단되는 일부 소비자물가지수 품목



통계청은 2000년 기준 지수개편부터 연쇄방식 소비자물가지수가 보조지표로 연간으로 작성되고 있다. 이 경험을 바탕으로 월별 소비자물가지수도 연쇄방식으로 작성하여 가중치 개편주기를 축소할 필요가 있다. 현재 수출입물가지수와 생산자물가지수는 연쇄지수로 작성되고 있다.

8) BSI 등도 조사과정에서 계절성을 처리하고 있지만 계절조정계열을 별도로 작성하여 발표하고 있다.  
 9) X-13ARIMA-SEATS를 이용하여 계절변동 존재 여부를 안정 계절성 검정과 이동계절성 검정을 통해 파악하고 있다. 두 검정을 종합하여 경제통계에 식별가능한 계절성이 있는지 검토하고 있다. 미국, 일본의 경우 계절조정계열을 작성하여 공표하고 있으므로 중기적으로 계절조정계열 작성을 검토할 필요가 있다.

## VI. 스캐너 데이터 확보 방안 검토

본장에서는 우리나라의 스캐너 데이터 확보 방안을 검토한다. 이를 위해 우선 해외 주요국의 스캐너 데이터 자료 확보 근거를 정리한다. 이후에 우리나라에서 스캐너 데이터 확보와 관련하여 우선적인 검토 사항을 짚어보고 실제 데이터 확보 방안을 논의한다.

### 1. 해외 주요국의 스캐너 데이터 접근 근거

<표 VI-1>는 영국과 프랑스를 제외하고 모두 실제 소비자물가지수 산정할 때 스캐너 데이터를 활용하는 국가들의 스캐너 데이터 접근 근거를 정리한 것이다. 이를 보면 스캐너 데이터를 연구 중인 영국의 경우 아직 별도의 안정적인 확보 방안에 대한 구체적인 논의가 이루어지지 않았으며 프랑스의 경우는 시장 리서치 기관을 통해 해당 스캐너 데이터를 구매하여 필요한 연구를 진행하고 있다.

<표 VI-1> 해외 주요국 스캐너 데이터 접근 근거

국가	CPI 구축시 스캐너 자료 포함 여부	스캐너 자료 접근 및 확보 근거
영국	활용방안 연구중	향후 안정적인 확보 방안에 대해서 지속적인 검토를 해나갈 계획
프랑스	활용방안 연구중	시장 리서치 기관으로부터 구매
스웨덴	실제 활용중	유통산업과의 자발적 업무협약
노르웨이	실제 활용중	통계법에 따라 무료로 접근 가능
스위스	실제 활용중	통계법에 따라 무료로 접근 가능
네덜란드	실제 활용중	슈퍼마켓 체인점이 직접 통계청에 제공
뉴질랜드	실제 활용중	시장조사기관인 GfK로부터 매월 제공받음
벨기에	실제 활용중	FTP 방식으로 매주 자료를 취합

한편 스캐너 데이터를 실제 소비자물가지수 편제에 이용하고 있는 나라의 경우 대체로 통계법적 근거 또는 소매점과의 자발적인 업무협약을 통해 스캐너 데이터를 무료로 제공받고 있다. 뉴질랜드의 경우는 시장조사기관인 GfK로부터 해당 스캐너 데이터를 구매하고 있다. GfK는 소매점에서 수집한 데이터에 제품 특징, 특성에 대한 정보를 추가하고 소매점에서 수집한 데이

터는 월간 총 판매액, 판매량 등 거래액의 80~95% 정도가 포함되어 있는 것으로 알려져 있다. 대체로 여타 통계법적인 근거를 가지고 스캐너 데이터를 확보하고 있는 유럽의 나라에 비해서 해당 통계의 양과 질이 다소 떨어진다고 평가할 수 있다.

여타 유럽 국가의 스캐너 데이터 확보 근거를 좀 더 자세히 살펴보면, 스웨덴 통계청은 해당 업체와 업무협약을 통해 무료로 스캐너 데이터를 확보하고 있다. 동 업무협약에는 판매업체는 스캐너 데이터의 질적 담보에 대해서 책임을 지지 않으며 통계청은 해당 자료를 공적인 통계 용도에 한정해서 활용한다는 내용이 포함되어 있다. 노르웨이의 경우 1989년 통계법에 따라 공적 용도의 통계 작성으로 인정되는 경우 해당 사업자에게 강제 보고의무가 부여된다.

<표 VI-2> 스캐너 데이터 접근 관련 통계법 조문

국가	해당 조항	해당 통계법 조문
덴마크	통계청법 제8조	(1) 모든 사업주는 덴마크 통계청의 요청이 있는 경우, 다음 각 호에 관한 정보를 제공하여야 한다. 1. 기업의 성격, 소재지 및 소유관계, 토지사용, 총 매출 및 가격, 인건비... 2. 전국적 사업주 단체에 제(1)항에서 정하는 통계정보 이외에 정보를 제공할 것을 요청하는 경우, 경제사무부장은 이사회의 권고에 따라, 그와 같은 정보를 덴마크 통계청에 제공하도록 지시할 수 있다.
네덜란드	통계청법 제5부 3항	청장은 통계목적에 위하여 국가통계위원회(CCS)와 상의하여 행정명령에서 정하는 데이터를 같은 행정명령에서 정하는 회사, 독립적 전문가, 기관 및 법인으로부터 요청할 수 있다.
스위스	연방통계법 제4조 데이터 수집 원칙	1. 연방이 필요한 데이터를 입수하거나 연방법의 이행 시 동 법률의 적용을 받는 기구에 해당 정보가 제공될 경우(연방 행정 정보), 연방 통계 기구는 별도로 조사를 실시하지 않아도 된다(관찰 및 측정을 기반으로 하는 직접 조사, 간접 조사) 3. 직접 조사는 동 법에 국한된 목적을 위해 자연인과 법인을 면담하여 원점에서 신규 데이터를 수집해야 한다. 조사의 횟수 및 형식은 필요한 범위로 엄격하게 제한한다.

자료 : 해외 주요국 통계법령에서 발췌

한편 덴마크, 네덜란드, 스위스의 경우 <표 VI-2>에서 볼 수 있듯이, 소매점 체인 등의 스캐너 데이터를 직접 확보할 수 있는 통계법적 근거를 갖추고 있

다. 즉, 공적 기관 및 금융기관뿐만 아니라 일반 민간기업에 대해서도 공공의 통계목적을 위해서 적절한 자문기구의 권고를 바탕으로 해당 자료를 직접 요구할 수 있도록 되어 있다. 이에 따라 스위스의 경우, 스캐너 데이터 입수 초기에는 자발적인 업무협약 등을 통해 다양한 판매 체인점이 적극적으로 스캐너 데이터 제공에 나섰으나 이후 점차 자료 제공에서 이탈하는 체인점에 나타났다. 이에 따라 해당 통계청은 관련 연방통계법을 적용하여 해당 기업이 스캐너 데이터를 의무적으로 제공토록 하고 있다. 또한 덴마크 및 네덜란드도 관련 통계법에 따라 슈퍼마켓 체인점 등이 스캐너 데이터를 정기적으로 제공함에 따라 해당 데이터의 사용을 확대하고 있는 중이다.

## 2. 스캐너 데이터 확보 관련 고려 사항

본 절에서는 우리나라의 스캐너 데이터 확보 방안과 관련하여 해외 주요국의 사례에 비추어 우선적으로 검토가 필요한 사항을 정리해 본다. 첫째, 스캐너 데이터를 실제 소비자물가지수 편제 활용하고 있는 나라의 경우 대부분 통계법적인 근거 또는 자발적인 업무협약 등을 통해 무료로 해당 자료를 확보하고 있다. 이는 스캐너 데이터의 경우 비교적 많은 시간과 비용이 소요되는 전통적인 방문조사 방식을 대체함에 따라 통계수집 비용을 낮출 수 있다는 이점과 관련하여 중요한 시사점을 제시하고 있다. 가령, 시장조사기관으로부터 스캐너 데이터를 구매하는 경우 막대한 비용이 소요됨에 따라 스캐너 데이터 활용의 장점 중 하나인 방문 및 설문조사(서베이) 조사비용의 절감 효과가 저하될 수 있기 때문이다. 뉴질랜드의 경우 해당 데이터를 구매하고 있지만 동 나라의 경우 스캐너 데이터의 활용이 유럽 국가에 비해서 제한적인 것으로 평가된다.

둘째, 우리나라에서 스캐너 데이터를 이용하는 경우 원 스캐너 데이터를 확보할 것인지 아니면 가공된 스캐너 데이터를 확보할 것인지에 대해서도 일정한 사전 정지작업이 필요하다. 가령, 스캐너 데이터를 실제 소비자물가지수에 활용하고 있거나 활발한 연구를 진행하고 있는 모든 나라에서 통상 가공되지 않은 스캐너 데이터를 확보하고 이후 통계당국이 해당 나라의 여건에 맞추어 적절한 필터링을 직접 수행하고 있다. 본 보고서에서 연구를 수행한 우리나라의 스캐너 데이터는 대한상의의 1차 가공을 거친 데이터라고 할 수 있다. 즉, 대한상의의 스캐너 데이터는 가장 낮은 단계의 해당 상품에 대해 판매점별 그리고 지역별로 합산한 값이라고 할 수 있다. 우리나라의 경우 대한상의의 1차 가공된 데이터를 이용할 것인지, 또는 대한상의의 원자료

를 확보할 것인지, 아니면 별도의 스캐너 데이터 확보를 위한 시스템을 구축할 것인지에 대해서도 향후 장기적인 관점에서 비용 및 편익 등에 대한 사전 논의가 필요하다.

셋째, 우리나라의 인구 크기 및 유통구조에 대한 고려가 필요하다. 스캐너 데이터를 적극적으로 활용하고 있는 나라는 대체로 북유럽, 스위스 및 뉴질랜드이다. 이들 나라들은 대체로 인구가 1,000만 명 내외로 작으며 대형 체인점의 시장 점유율 비중이 높아 스캐너 데이터의 활용이 용이한 것으로 평가된다. 반면 이들 나라에 비해 보다 인구도 많으며 보다 복잡한 유통구조를 가지고 있는 영국 및 프랑스의 경우 소비자물가지수에 활용을 위해서 여전히 다양한 사전 연구를 진행하고 있다. 우리나라의 경우도 전자보다는 후자에 가까운 인구 및 유통구조를 가지고 있는 것으로 평가되는 만큼 향후 실제 스캐너 데이터를 소비자물가지수 산정에 활용하기 위해서는 향후 보다 많은 연구와 시간을 필요하다는 것을 시사한다.

### 3. 스캐너 데이터 확보 방안

스캐너 데이터를 통계 작성기관이 이용하려면 품질이 우수한 스캐너 데이터를 주기적으로 그리고 안정적으로 확보해야 한다. 현재로서는 통계 작성기관이 스캐너 데이터를 수집하는 것이 사실상 불가능하므로 이를 지속적으로 확보할 수 있는 방안을 해당기관과 협의하여 마련할 필요가 있다.

빅데이터 수집·제공자(데이터 사업자)와 통계작성자 간 협의를 통해서 다음과 같은 방안을 고려해 볼 수 있다. 첫째, 데이터 사업자의 해당 스캐너 데이터를 구매하는 방식이다. 둘째, 통계법적인 근거를 가지고 강제 보고를 통해 확보하는 방안이다. 셋째, 데이터 사업자가 정부 공식 통계로 승인 받은 후 작성 및 공표하고 통계 작성자는 이를 이용하여 국가 통계작성에 이용하는 방안이다. 마지막으로, 데이터산업 육성법 등을 통해 데이터 산업을 육성하고 이 과정에서 적절한 협약 등을 통해 국가가 기초 스캐너 데이터를 확보하는 방안이다.

이러한 방안 중에서 우선 첫째의 경우는 앞서 논의했듯이, 스캐너 데이터를 구매하는 경우 그 비용을 가늠하기가 쉽지 않으며 스캐너 데이터 활용의 장점 중 하나인 방문 및 설문조사(서베이) 조사비용의 절감 효과가 크게 저하될 수 있다. 둘째의 경우는 우리나라 통계법상 통계작성과 관련한 법적 강제력을 행사하려면 통계조사대상 업체여야 하는데, 대한상의의 경우 통계조사대상 업체가 아니기 때문에 직접적인 통계법적 근거를 적용하기는 쉽지



않은 것으로 판단된다. 다만 국가승인통계로 인정을 받는다면, 공식적인 데이터 접근을 위한 법적인 근거가 보다 확충된다고 판단된다. 이는 셋째의 경우와 맞닿아 있다고 할 수 있다. 마지막으로 데이터산업 육성법 등을 활용하는 것은 통계당국과 대한상의회가 협력할 수 있는 상호이익의 근거를 제공한다는 점에서 보다 현실적으로 보이지만 다소간의 시간이 필요해 보인다.

이러한 논의를 배경으로 우리나라에서 스캐너 데이터를 조기에 확보하기 위한 현실적인 방안은 다음과 같이 정리될 수 있다. 첫째, 유통시장 점유율이 높은 대형마트 3사의 스캐너 데이터 또한 대한상의회로 집중시킬 필요가 있다. 둘째, 통계청은 데이터산업 육성법 등을 활용하여 상호이익의 근거를 마련하면서 대한상의회와 업무협약을 통해 스캐너 데이터의 원자료 또는 가공된 자료를 확보해나갈 필요가 있다. 셋째, 처음에는 자발적인 업무협약으로 시작했으나 이후 유통업체가 해당 데이터 제공에 소극적으로 변하자 스위스 통계당국은 통계법적인 근거를 적용하여 데이터 확보의 지속성을 확보한 사례를 비추어 현재 상태에서 스캐너 데이터의 확보가 지속성을 가질 수 있기 위해서는 향후 국가승인통계로의 지위 획득 등을 통해 일정한 법적 근거를 마련해 나가는 노력이 필요하다.

## VII. 결론 및 향후 과제

본 연구는 스캐너 데이터를 소비자물가지수 작성에 어떻게 활용할 수 있는지 살펴보기 위해 진행되었다. 스캐너 데이터를 이용하는 경우 상품 가격 및 거래량 정보를 빠르게 알 수 있고, 새로운 상품의 출현 정보도 조기에 파악할 수 있어서 소비자물가 조사에 발생하는 표본조사의 어려움과 라스파이레스 지수의 한계를 동시에 극복할 수 있다. 그러나 현재 스캐너 데이터의 포괄범위가 넓지 않고 통계청에서 스캐너 데이터를 안정적으로 확보할 수 있는 제도적 환경이 마련되지 않아 스캐너 데이터를 소비자물가지수 작성에 직접 활용하는 데에는 제약이 있다. 이 장에서는 조사단계, 지수작성 단계, 지수검토 단계로 나누어 스캐너 데이터를 소비자물가지수 작성에 활용하는 방안을 모색하고, 향후 과제를 스캐너 데이터의 품질 개선과 안정적 확보, 스캐너 데이터의 소비자물가지수 작성에서의 활용과 스캐너 데이터를 이용한 지수 작성 등으로 나누어 살펴본다.

### 1. 스캐너 데이터의 품질 개선과 안정적 확보

스캐너 데이터는 표본조사로 얻을 수 없는 다양하고 유용한 정보를 제공하는 일종의 빅데이터이다. 본 연구를 통해서 살펴본 스캐너 데이터는 경제통계 작성에 매우 유용하게 활용될 수 있는 빅데이터이지만 다른 빅데이터와 마찬가지로 통계작성 목적에 맞춰 이용하려면 데이터 수집 및 정제 과정에 많은 노력이 필요하다.

대한상공회의소에 만든 스캐너 데이터를 소비자물가지수 작성에 활용하려면 다음을 감안해야 한다. 첫째, 현재 대한상공회의소에서 만들고 있는 스캐너 데이터는 대형 마트 3사의 거래 데이터와 점차 그 규모가 커지고 있는 온라인 판매업체의 거래 데이터가 포함되어 있지 않는 포괄범위의 제약이 있다. 통상 상품의 가격은 유통사별로 거의 같게 움직이므로 큰 문제가 되지 않을 수 있다. 그러나 많은 거래 데이터가 포함되지 않아서 스캐너 데이터의 거래량의 대표성에 의문을 제기할 가능성이 높다. 따라서 통계청에서 대한상공회의소뿐만 아니라 다른 유통업체의 스캐너 데이터를 확보하는 방안을 마련할 필요가 있다. 통계청이 직접 해당 유통업체를 조사하거나 품목별 정보를 자동으로 얻을 수 있는 시스템을 구축할 필요가 있다. 예를 들면 카드사와 유통사가 결합된 조사 전용 카드를 만들고 이를 통해 구매 및 가격정보를 얻고 이에 대한 비용을 지급하는 것이다. 이를 통해 가격과 지출 데이터

를 동시에 얻을 수 있고 동 데이터를 조사에 반영하는 것이다. 둘째, 스캐너 데이터와 관련된 상품 바코드에는 상품 분류체계가 포함되어 있지 않고 기업에서 부여하기 때문에 스캐너 데이터를 통계청 품목지수로 전환하는데 제약이 있다. 스캐너 데이터를 통계로 전환하기 위해서는 바코드 숫자 부여 방식을 통계 작성자 측면에서 검토하고 이를 국제표준화에 반영할 필요가 있다. 셋째, 스캐너 데이터에는 반품, 교환, 묶음판매 등이 특별한 기준 없이 기록되어 데이터 오류로 나타난다. 이러한 데이터 오류는 유통업체의 경영활동 파악에도 제약요인으로 작용할 것으로 판단된다. 유통업체도 데이터 기반으로 경영하기 위해서는 기초 거래 데이터의 분석이 필요하므로 유통업체의 POS시스템에서 이러한 특수한 거래를 별도의 코드로 처리할 필요가 있다. 넷째, 스캐너 데이터의 정제 과정을 보다 정교화할 필요가 있다. 예를 들면 동일 상품에 대한 가격들 간 차이가 지나치게 큰 경우 그 원인을 범주화할 수 있는 알고리즘을 마련하여 자동으로 문제를 처리하는 방안을 마련할 필요가 있다. 그리고 별도의 상품명 데이터베이스를 축적하고 여러 가지 학습 알고리즘을 적용하여 스캐너 데이터의 특이항 등을 적절히 처리할 필요가 있다. 다섯째, 통계청이 스캐너 데이터를 통계작성에 활용하려면 스캐너 데이터를 안정적으로 확보해야 한다. 가장 쉬운 스캐너 데이터 입수 방법은 통계법을 적용하여 확보하거나 구매하는 것이다. 스캐너데이터를 구매하는 데에는 비용이 많이 소요되지만 안정성을 담보하기 어려우므로 결국 통계법으로 데이터를 확보할 수 있는 방안을 마련해야 한다. 현행 통계법을 기반으로 스캐너 데이터를 얻기 위한 방안은 대한상공회의소가 동 데이터를 지수를 작성하고 국가통계로 승인받는 것이다. 통계청은 국가통계 승인과정에서 통계청 기준의 분류체계의 결과를 작성토록하여 이를 소비자물가지수 작성에 활용하는 것이다.<sup>10)</sup>

## 2. 스캐너 데이터의 소비자물가지수 작성에서의 활용

벨기에, 덴마크, 네덜란드, 스웨덴, 스위스 등의 통계작성기관과 연구기관들은 스캐너 데이터가 소비자물가지수 작성의 효율성과 정확성을 높일 수 있다고 판단하고 스캐너 데이터를 소비자물가지수 작성에 부분적으로 활용하거나 활용하는 방안을 연구하고 있다. 연구결과 스캐너 데이터는 데이터 수집 비용을 낮추고 조사항목을 늘려 표본 분산과 측정오차를 감소시킬 뿐

10) 대한상공회의소에서 스캐너데이터를 이용하여 통계를 작성하고 이를 통계청이 기초자료로 이용할 경우 대한상공회의소의 활동에 대해 대외 신뢰도가 높아져서 다른 유통기관의 스캐너 데이터 수집이 용이해지고 사회적 영향력도 증가할 수 있을 것으로 판단된다.

만 아니라 조사대상처의 응답부담도 감소시키는 것으로 나타났다. 우리나라에서도 품목별 소비자물가지수 작성 과정에서 스캐너데이터가 아래와 같이 다양하게 이용될 수 있다.

첫째, 소비자물가지수 과정의 보조정보로 유용하게 이용할 수 있다. 2015년 기준으로 보면 소비자물가지수는 38개 도시 460개 대표품목의 소매점 및 서비스업체에 대해서 이루어진다. 이 때 품목별 대표 상품을 선정하는데 물량 데이터를 알 수 있는 스캐너 데이터는 중요한 정보가 될 수 있다. 스캐너 데이터를 통해 조사대상 품목의 대표 상품을 선정할 수 있다. 둘째, 현재 소비자물가지수에서는 일시적인 제조업체와 유통업체의 할인판매 관련 가격지수에 반영되지 않는 한계가 있다. 그 동안은 조사의 어려움으로 할인 등은 가격조사에는 반영하지 않았지만 스캐너 데이터를 활용한다면 이를 반영하여 물가지수의 체감성을 높일 수 있다. 셋째, 무응답 대체 등에 스캐너 데이터를 활용할 수 있다. 현재 무응답이 있을 경우 그 값을 통상 전월의 값 또는 평균값으로 대체하는데 이 때 스캐너 데이터로 가장 비슷한 지역과 업체의 상품 가격으로 대체할 수 있다. 넷째, 스캐너 데이터는 조사된 가격이 적절한지 여부를 사전에 검토할 수 있는 자료로 활용될 수 있다. 또한 스캐너 데이터는 주별로 측정되므로 월별 1회 조사 상품의 가격의 신뢰성을 파악하는 데에도 이용될 수 있다. 스캐너 데이터에서 월별 변동이 있다면 소비자물가지수의 조사횟수를 늘리든지 스캐너 데이터와 결합해서 가격을 정할 수 있다. 다섯째, 스캐너 데이터를 이용하여 가중치 관련 정보를 보완할 수 있다. 이제까지 가중치 정보는 가격정보와 별도로 측정되었는데 스캐너 데이터에서는 물량과 가격의 두 정보가 동시에 있으므로 가계동향조사의 가중치를 그대로 이용하기 보다는 스캐너 데이터의 물량 데이터를 바탕으로 가중치가 검토되거나 수정될 필요가 있다.

이와 같이 스캐너데이터 정보가 소비자물가지수 작성에 활용되어 신뢰를 얻게 된다면 일부 품목의 가격조사는 표본조사가 아닌 스캐너 데이터를 이용한 가격조사로 대체될 수도 있다. 스캐너 데이터에 대한 무조건적 거부감과 지나친 신뢰 모두 경계하면서 스캐너 데이터를 소비자물가지수의 중요한 정보로 활용해야 한다.

### 3. 스캐너 데이터를 이용한 지수 작성

스캐너 데이터로 직접 만들어진 소비자물가지수 결과는 조사방식, 가중치, 지수산식 등의 차이로 통계청에서 발표된 공식 소비자물가지수 결과와 다르

게 나타날 수밖에 없다. 통계청이 두 지수를 동시에 만든다면 이용자가 혼란에 빠지고 이미 공표하고 있는 소비자물가지수의 신뢰성이 크게 낮아질 가능성이 높다. 따라서 통계청이 스캐너 데이터를 이용한 소비자물가지수를 만드는 것은 기존 일부 품목 조사를 스캐너 데이터로 전환하는 것을 의미할 수 밖에 없다.

스캐너 데이터를 이용한 지수를 만든다면 통계청 외의 기관(예 : 대한상공회의소)이 동 지수를 소비자물가지수 작성 주기와 다르게 공표하는 것이 바람직하다. 동 지수는 가격지수와 물량지수로 구분할 수 있으므로 동지수를 통해 상품 가격의 변화뿐만 아니라 물량의 변화도 파악할 수 있다. 특정 품목의 가격 또는 물량의 변화는 경기변동, 위기변동<sup>11)</sup> 등의 지표로도 활용될 수 있다. 스캐너 데이터로 작성된 지수는 소비자물가를 파악할 수 있는 또 다른 자원이 될 뿐만 아니라 소비자물가지수를 검토할 수 있는 자원도 될 수 있다. 또한 스캐너 데이터는 온라인 물가 데이터와 결합되어 보다 대표성 높은 소비자물가지수 작성에도 활용될 수 있을 것으로 판단된다. 스캐너 데이터를 이용할 경우 물량 데이터를 이용한 톨크비스트(Törnqvist) 지수 또는 피셔(Fisher) 지수 등을 작성하여 라스파이레스 방식 소비자물가지수 작성의 한계를 극복할 수 있다.

#### 4. 스캐너 데이터를 활용하기 위한 환경 확보

스캐너 데이터를 소비자물가 등 국가통계 작성에 제대로 이용하려면 빅데이터 수집 관련 제도적 환경을 조성하고 관련 인력을 육성해야 한다. 통계청 등 국가통계 작성기관이 표본조사를 중시하는 데에서 벗어나 다양하게 측정되고 있는 빅데이터를 국가통계에 활용하는 방안을 적극 모색할 필요가 있다. 첫째, 제도적 환경으로는 데이터 기업이 영업적 목적으로 수집한 빅데이터를 바탕으로 영업활동을 원활히 할 수 있도록 하는 한편 국가통계 작성 목적으로 동 빅데이터의 요약정보를 수집할 수 있도록 하는 데이터진흥법을 제정할 필요가 있다. 이를 통해 카드사 데이터, 소매 거래 데이터 등 수많은 국가통계 작성과 관련된 정보를 수집·축적할 수 있다. 둘째, 국가통계작성을 표본 설계와 많은 조사원이 필요한 표본조사 중심에서 데이터 자동수집과 조사를 결합한 형태로 전환할 필요가 있다. 그리고 관련 인력도 표본조사 중심에서 프로그램이 가능한 데이터 분석 전문가를 적극 육성하여 통계 작성

11) 예를 들면 사람들이 지진, 태풍, 전쟁, 전염병 등의 위협을 느낀다면 관련된 품목을 사재기할 가능성이 있고 이는 스캐너 데이터로 조기에 파악할 수 있다.

에 활용할 필요가 있다. 셋째, 빅데이터를 이용한 국가통계작성이 신뢰를 얻으려면 오랜 시간이 소요되므로 관련 빅데이터 연구를 지속적으로 진행할 수 있도록 빅데이터, 관련 알고리즘 등을 연구 목적으로 공개하여 빅데이터 국가통계 작성 인프라를 조성할 필요가 있다.

## <참고문헌>

- Armknrecht, P. and M. Silver (2012). Post-Laspeyres: The Case for a New Formula for Compiling Consumer Price Indexes, IMF Working Paper.
- Boskin, M., Dullberger, E., Gordon, R., Griliches, Z. and Jorgensen, D. (2001), Final Report of the Commission to Study the Consumer Price Index, U.S. Senate, Committee on Finance (Washington, D.C.: U.S. Government Printing Office)
- Chessa, A. (2015), "A New Methodology for Processing Scanner Data in the Dutch CPI," Eurostat Review on National Accounts and Macroeconomic Indicators.
- Chessa, A., S. Boumans and J. Walschots, (2015), "Towards a New Methodology for Scanner Data in the Dutch CPI," manuscript for Scanner Data Workshop in Rome.
- De Haan, J., Willenborg, L., Chessa, A., &Verburg, J. (2016, May). An overview of price index methods for scanner data. In Paper for the UNECE-ILO meeting of the group of experts on consumer price indexes (pp. 2-4).
- FAIVRE, S., SILLARD, P., &VARLET, G. Would scanner data improve the French CPI?.
- FitzGerald, J., &Shoemaker, O. Evaluating the Consumer Price Index Using Nielsen's Scanner Data October 2013.
- Hansen, Carsten Bolden, (2007). "Recalculations of the Danish CPI 1996 - 2006," *Proceedings of the Tenth Meeting of the International Working Group on Price Indexes*, (Ottawa, Canada).
- Huang, N. W. Wimalaratne and B. Pollard (2015). Choice of index number formula and the upper-level substitution bias in the Canadian CPI, *Proceedings of 14th Ottawa Group Meeting* (Tokyo, Japan).
- Imai, S., &Watanabe, T. (2015). Replicating Japan's CPI Using Scanner Data.
- International Labour Office. (2004). "Consumer Price Index Manual: Theory and Practice." Geneva.
- Johannessen, R. (2016), "Scanner Data in CPI/HICP," ESS Modernization Workshop Bucharest.

- Lent, Janice and Alan H. Dorfman (2009). “Using a weighted average of base period price indexes to approximate a superlative index.” *Journal of Official Statistics*. Vol. 25, no. 1. p. 139-149.
- Melser, D. (2011, May). Constructing High Frequency Price Indexes Using Scanner Data. In twelfth meeting of the Ottawa Group (Vol. 46).
- United Nations Economic Commission for Europe (UNECE), ILO, IMF, OECD, Eurostat, World Bank, Office for National Statistics United Kingdom, (2009), *Practical Guide to Producing Consumer Price Indices* (New York and Geneva: United Nations): <http://www.imf.org/external/data.htm#guide>.
- Valenduc, F. and K. Van Loon, (2015), “Belgian Scanner Data Project - Methodology and Results,” manuscript for Scanner Data Workshop in Rome.
- 김기석, (2012), “다변 지수법을 이용한 스캐너 데이터 기반 소비자 물가 지수,” 석사학위 논문, 성균관대학교 통계학과
- 박현정 (2009). 소비자물가지수 현실반영도 제고를 위한 지수산식 연구, 통계개발원, 연구보고서.
- 정태훈, (2016), “POS 데이터를 이용한 일별 물가지수의 측정: 일본의 UTokyo 일별 물가지수를 중심으로,” 『통계연구』, 특별호, pp. 52-71.
- 통계청 (2013). 2012년 「소비자물가지수 가중치」 개편결과.
- 통계청 (2014). 소비자물가지수 개편계획(안), 경제통계2분과위원회 회의 자료.
- 통계청 (2016). 2015년 기준 소비자물가지수 개편결과.



- 
1. 코드 분류에 사용된 키워드 일람표
  2. 가공식품 지수 산출을 위한 코드 매칭표(종합)
  3. CPI 전기대비 증감률과의 상관계수
-



<별 첨 1>

코드 분류에 사용된 키워드 일람표

KAN코드	PUMID코드	키워드
010199	A018110(물엿)	"물엿", "조리엿", "올리고당", "요리당", "조청", "쌀엿", "맥아엿", "이온엿", "맥아", "황엿", "엿"
	A019070 (양념소스)	"양념", "된장", "고추장", "육수", "전통장", "간장", "볶음장", "식초", "자연초", "신송", "쌈표", "홍초", "소스", "시럽", "비빔장", '브로스', "비빔장", '넥타', '매실청', "청주한큰술", "맛술", '맛내기장', '참치액'
	A019120 (혼합조미료)	"조미료", "미원", "다시다", "스톡", "다시", "치킨파우더", "후리가께", "후리가케", "나카타니엔", "밥이랑", "밥친구", "미림", "미향", "연두", "맛나", "감치미", "보크라이스", "맛선생", "주먹밥", "멸치가루", "오차즈께", "차향노모토", "미풍", "쇠고기스틱", "멸치스틱", "산들애", "아이미", "한식예찬", "밥사랑", "맛있는밥", "사골분말", "양파가루", "카라멜", "황태가루", "새우가루", "멸치가루", "홍합가루", "참맛", "버섯가루", "시즈닝", "올가", "시즈닝", "비빌락", "맘스키친", "멸치분", "빅마마시푸드", "조미 분말", "카라멜", "미주", "핵산", "아미노산", "빅마마씨푸드", "디포리가루", "신앙촌", "분말", "가루", "쇠고기맛", "마시다", "소금", "스프", "염", "미정", "백설", "글루", "MSG", "파우더", "국물", "나트륨", '밥대장', '나물&엔', '큐브', '멸치', '구시다', '밥짜꿍', '뿌리고', '맛보고'
010701	A012050(소시지)	"소세지", "소시지", "비엔나", "후랑크", "프랑크", "에센보독", "부어스트", "빅팜", "키스틱", "줄줄이", "맥스봉", "켄터키", "소시 지", "엔네이처", "휠터치", "부르거", "초리조", "초리소", "윈너", "만시지", "싱켄", "페퍼로니", "뿌까", "후랑트"
	A012060 (햄및베이컨)	"마늘햄", "김밥햄", "슬라이스햄", "슬라이스 햄", "스팸", "로스팜", "뚝심", "양파햄", "김밥속햄", "베이컨", "마늘 햄", "김밥속 햄", "리챔", "갈릭햄", "우리팜", "살라미", "스모크 햄", "사각햄", "라운드 햄", "목우촌", "야채햄", "스모크햄", "런천미트", "햄", "팜", "팸", "하몽", "델립", "스테이크", "떡갈비", "동그랑땡", "동그랑 땡", "툴립", "숯불", "챔", "프로슈토", "햄", "바베큐", "슬라이스", "경단", "프라", "크랩"
010802	A017280(김)	"김밥김", "김전장", "전장", "김밥용", "100매", "BOX", "추석", "세트", "호", "50매", "30매", "20매", "전지"

	A017290(맛김)	"자반김", "맛김", "자른김", "과래김", "도시락김", "재래김", "가루", "8절", "자반", "9G", "7G", "6G", "5G", "4G", "3G", "2G", "*8", "*16", "*12", "식탁김", "9매", "10매", "조미", "과래", "김"
	A019130(스프)	"스프", "수프", "가루", "믹스", "만들기", "부꾸미", "브라우니", "만드는"
011599	A019180 (즉석식품)	"즉석밥", "헛반", "오투기밥", "삼각김밥", "곰탕", "사골국", "북어국", "미역국", "갈비탕", "도가니탕", "설렁탕", "우거지", "어묵탕", "죽", "탕", "국", '육개장', '찌개', '파스타', '스파게티', '까르보나라', '짜장', '사골', '닭갈비', '찜닭', '삼겹', '퐁치', '고등어', '메로', '3분', '초밥', '닭개장', '갈치', '장어', '닭발', '치즈김밥', '참치김밥', '불고기', '덮밥'
011902	A018050(비스킷)	"비스킷", "비스킷", "크래커", "쿠키", "크래커", "웨하스", "시리얼", "웨이퍼", "스틱", "구움", "예감", "빼빼로", "칸초", "건빵", "고소해", '와퍼', '프레즐', '그라놀라', '크리스피', '샌드', '오트밀', '브레첼', '센베이', '와플', '오레오', '브레이크', '버터링', '에이스', '누룽지', '사루비아', '샌드위치', '까메오', '고래밥', '쿠크다스', '누네띠네', '빈츠', '에너지마', '빠다코코넛', '깨', '제크', '구운감자', '틴틴', '폴리', '뽕또', '하임', '다이제', '전병', '라이스바', '라이트바', '아이비'
	A018070(파이)	"몽셀", "후레쉬베리", "오에스", "파이", "후레쉬블루베리", '브라우니', '초코', '타르트', '홈런볼', '크레페', '페스트리', '케익', '카스타드', '가또', '마가렛트', '롱스', '오뜨', '페스트리', '페이스트리', '칼로리바란스'
012011	A022040 (기능성음료)	"드링크", "박카스", "홍삼", "굿베이스", "한삼인", "비타", "정관장", "한삼인", "바이오하우스", "고려인삼", "레모나", "아로니아", "즙", "아사이베리", "산삼", "더덕", "인삼", "진액", "엑기스", "한뿌리", "도라지", "녹십자", "구론산", "웅녹정", "오가피", "흑마늘", "복분자", "고로쇠", "사자발쑀", "천마", "녹용", "천호", "오로나민", "인협", "그린비아", "사슴", "특산", "다미나", "원기", "천지양", "쌍화", "천제명", "동충하초", "천지인", "옥녹정", "우루", "헬스원", "열매", '레드불', '몬스터', '에너지', '컨디션', '팻다운', '에네르', '헛개', '술'

	A022060 (혼합음료)	"포카리", "게토레이", "파워에이드", "스무디", "음료", "미에로", "콜라", '식이섬유', '에이드', '이온', '알로에', '화이버'
012016	B011040(막걸리)	"막걸리", "탁주", "생탁", "탁배기", "막걸이", "아이스", "쌀탁", '탁테일', '우국생'
	B011060(약주)	"문배주", "불휘", "홍주", "백주", "소주", "동동주", "샘주", "풍년주", "국화주", "돌배주", "문경바람백자", "고현정백주", "과화주", "금복주", "송이주", "소곡주", "민속제주", "소곡화주", "대나무통주", "통주", "가시오가피주", "알밤술", "약주", "아락", "솔잎주", "홍구기주", "화요", "구기주", "산사춘", "이강주", "과하주", "백세주", '모태주', '오가피', '송이', '대나무', '버섯', '모주', '오가자', '오디', '뽕', '매화', '오메기', '대추', '복분자', '맑은', '더덕', '흑마늘', '오미', '오곡', '국화', '백일'

<별첨 2>

가공식품 지수 산출을 위한 코드 매칭표(종합)

KAN	KAN 품목명	PUMID	PUMID 품목명	통계청분류
010301	간장	A019050	간장	가공식품
010303	고추장	A019080	고추장	가공식품
012017	과실주	B011020	과실주	가공식품
011099	과실통조림	A016160	과일가공품	가공식품
010901	국수	A011090	국수	가공식품
012011	기능성음료	A022040	기능성음료	가공식품
011101	김치	A019150	김치	가공식품
011903	깍	A018030	깍	가공식품
011299	냉동식품	A019170	냉동식품	가공식품
011199	단무지	A017270	단무지	가공식품
010902	당면	A011110	당면	가공식품
010302	된장	A019060	된장	가공식품
011301	두부	A011120	두부	가공식품
012012	두유	A022020	두유	가공식품
010299	드레싱	A019110	드레싱	가공식품
010903	라면	A011100	라면	가공식품
012016, 012099	막걸리	B011040	막걸리	가공식품
010802	맛김	A017290	맛김	가공식품
011399	맛살	A013160	맛살	가공식품
012013	맥주	B011030	맥주	가공식품
010199	물엿	A018110	물엿	가공식품
010602	발효유	A014040	발효유	가공식품
011401	부침가루	A011140	부침가루	가공식품
011701	분유	A014020	분유	가공식품
011902	비스킷	A018050	비스킷	가공식품
011904	사탕	A018020	사탕	가공식품
011001	생선통조림	A013170	생선통조림	가공식품
012010	생수	A022030	생수	가공식품
010103	설탕	A018080	설탕	가공식품
010102	소금	A019040	소금	가공식품
010701	소시지	A012050	소시지	가공식품
012014	소주	B011010	소주	가공식품
011901	스낵과자	A018060	스낵과자	가공식품
011499, 011599	스프	A019130	스프	가공식품

<b>010904</b>	씨리얼식품	A011130	시리얼	가공식품
<b>010502</b>	식용유	A015020	식용유	가공식품
<b>011201</b>	아이스크림	A018040	아이스크림	가공식품
<b>012016</b>	약주	B011060	약주	가공식품
<b>010199</b>	양념장	A019070	양념소스	가공식품
<b>012015</b>	위스키	B011050	양주	가공식품
<b>011302</b>	어묵	A013150	어묵	가공식품
<b>010601</b>	우유	A014010	우유	가공식품
<b>011799</b>	이유식	A019140	이유식	가공식품
<b>011003</b>	잼	A018090	잼	가공식품
<b>011102</b>	젓갈	A013180	젓갈	가공식품
<b>012009</b>	과일주스	A022010	주스	가공식품
<b>011502, 011599</b>	즉석식품	A019180	즉석식품	가공식품
<b>012006</b>	차	A021020	차	가공식품
<b>010501</b>	참기름	A015010	참기름	가공식품
<b>011910</b>	초콜릿	A018010	초콜릿	가공식품
<b>010603</b>	치즈	A014030	치즈	가공식품
<b>011501</b>	카레	A019090	카레	가공식품
<b>012001, 012002, 012004</b>	커피	A021010	커피	가공식품
<b>012008</b>	탄산음료	A022050	탄산음료	가공식품
<b>011902</b>	초코파이	A018070	파이	가공식품
<b>010701</b>	햄	A012060	햄릿베이컨	가공식품
<b>012011</b>	혼합음료	A022060	혼합음료	가공식품
<b>010199</b>	혼합조미료	A019120	혼합조미료	가공식품
<b>021809, 021810</b>	가정용비닐용품	E061070	가정용비닐용품	기타품목
<b>010401</b>	고춧가루	A019010	고춧가루	기타품목
<b>020199</b>	구강세정제	L012130	구강세정제	기타품목
<b>020302</b>	화장지	L012090	기초화장품	기타품목
<b>010802</b>	김	A017280	김	기타품목
<b>011801</b>	꿀	A018100	꿀	기타품목
<b>012198</b>	담배	B021010	담배	기타품목
<b>020201</b>	면도기	L012010	면도기	기타품목
<b>021403</b>	모발염색약	L012120	모발염색약	기타품목
<b>010899</b>	미역	A017300	미역	기타품목
<b>021905</b>	방향제	E061090	방향제	기타품목
<b>021808</b>	부엌용세제	E061040	부엌용세제	기타품목

<b>020702</b>	화장비누	L012050	비누	기타품목
<b>021101</b>	살충제	E061060	살충제	기타품목
<b>020602</b>	샴푸	L012060	샴푸	기타품목
<b>022003</b>	섬유유연제	E061020	섬유유연제	기타품목
<b>022001</b>	세탁세제	E061010	세탁세제	기타품목
<b>021907</b>	습기제거제	E061100	습기제거제	기타품목
<b>021908, 021909, 021910</b>	청소용세제	E061050	청소용세제	기타품목
<b>020101</b>	치약	L012040	치약	기타품목
<b>020102</b>	칫솔	L012030	칫솔	기타품목
<b>020305</b>	키친타월	E061080	키친타월	기타품목
<b>020301</b>	화장지	L012080	화장지	기타품목



(업태별 합산)

	Jevons	Tornq	GEKS	RWGEKS
전체	0.796	0.800	0.806	0.785
국수	0.045	0.277	0.383	0.335
라면	-0.088	0.046	0.031	0.044
당면	0.756	0.779	0.790	0.775
두부	0.013	0.076	0.165	0.137
시리얼	0.060	0.335	0.359	0.291
부침가루	-0.019	0.281	0.220	0.253
소시지	-0.120	-0.157	-0.173	-0.164
햄및베이컨	0.303	0.104	0.241	0.252
어묵	0.050	0.282	0.330	0.267
맛살	0.254	0.143	0.124	0.117
생선통조림	-0.088	0.020	0.195	0.119
젓갈	0.124	0.384	0.365	0.354
우유	0.462	0.621	0.665	0.660
분유	-0.026	-0.068	-0.101	-0.028
치즈	0.012	0.074	0.058	0.075
발효유	0.325	0.476	0.453	0.427
참기름	-0.030	0.166	0.184	0.140
식용유	-0.111	0.089	0.041	0.057
과일가공품	0.058	0.236	0.278	0.333
단무지	0.014	0.083	0.337	0.144
김	0.352	0.245	0.134	0.190
맛김	0.304	0.203	0.251	0.240
미역	0.083	0.149	0.096	0.159
초콜릿	-0.273	-0.338	-0.305	-0.297
사탕	0.287	0.501	0.493	0.496
껌	-0.010	0.176	0.125	0.153
아이스크림	0.442	0.398	0.453	0.423
비스킷	-0.110	0.088	0.030	-0.028
스낵과자	-0.154	0.240	0.209	0.187
파이	-0.150	0.119	0.162	0.170
설탕	0.212	0.147	0.160	0.186
잼	-0.178	0.292	0.332	0.268
꿀	0.177	0.021	0.090	0.054
물엿	0.205	0.607	0.586	0.596
고춧가루	0.683	0.208	0.184	0.193
소금	-0.020	0.098	0.065	0.081
간장	0.298	0.185	0.143	0.165
된장	0.133	-0.172	-0.140	-0.179
양념소스	0.635	0.727	0.772	0.729
고추장	-0.099	-0.221	-0.165	-0.155
카레	-0.169	-0.085	-0.134	-0.126
드레싱	0.288	0.131	0.118	0.161

혼합조미료	-0.054	0.105	0.042	0.069
스프	0.250	0.059	0.047	0.052
이유식	0.181	0.148	0.046	0.093
김치	0.058	0.354	0.477	0.434
냉동식품	0.348	0.100	0.062	0.069
즉석식품	0.082	0.078	0.080	0.043
커피	-0.166	0.081	0.001	0.016
차	0.206	0.306	0.334	0.349
주스	0.086	-0.114	-0.128	-0.063
두유	-0.237	-0.184	-0.215	-0.157
생수	0.086	0.537	0.371	0.454
기능성음료	0.051	0.046	0.085	0.088
탄산음료	-0.002	-0.024	-0.025	-0.031
혼합음료	-0.086	-0.159	-0.086	-0.112
소주	0.703	0.809	0.778	0.775
과실주	-0.105	0.110	0.096	0.102
맥주	0.078	0.089	0.061	0.042
막걸리	0.018	0.019	0.020	-0.013
양주	0.026	0.331	0.392	0.362
약주	0.126	0.513	0.383	0.469
담배	0.941	0.957	0.957	0.943
세탁세제	-0.130	-0.209	-0.090	-0.177
섬유유연제	0.367	0.246	0.199	0.277
부엌용세제	-0.231	0.005	-0.071	-0.057
청소용세제	0.128	-0.106	0.074	-0.029
살충제	0.098	0.192	0.159	0.168
가정용비닐용품	-0.066	-0.162	-0.069	-0.141
키친타월	0.016	-0.031	-0.225	-0.156
방향제	-0.067	-0.220	-0.128	-0.217
습기제거제	-0.150	-0.061	-0.086	-0.029
면도기	-0.252	-0.124	-0.170	-0.077
칫솔	-0.330	-0.083	-0.127	-0.066
치약	-0.173	-0.065	-0.135	-0.178
비누	0.020	0.266	0.254	0.261
샴푸	-0.070	0.041	0.087	0.074
화장지	0.104	0.225	0.277	0.263
기초화장품	-0.318	-0.041	-0.003	-0.021
모발염색약	0.026	0.321	0.322	0.321
구강세정제	0.044	0.250	0.164	0.249

## (지역별 합산)

	Jevons	Tomq	GEKS	RWGEKS
전체	0.745	0.791	0.800	0.766
국수	-0.043	0.301	0.415	0.349
라면	-0.130	0.037	0.021	0.041
당면	0.673	0.784	0.798	0.785
두부	0.141	0.063	0.181	0.137
시리얼	0.023	0.323	0.373	0.281
부침가루	0.176	0.292	0.267	0.295
소시지	-0.038	-0.149	-0.154	-0.144
햄및베이컨	0.341	0.078	0.232	0.171
어묵	0.302	0.259	0.286	0.267
맛살	0.288	0.139	0.103	0.107
생선통조림	0.065	0.052	0.179	0.104
젓갈	0.368	0.390	0.369	0.370
우유	0.433	0.614	0.659	0.637
분유	-0.114	-0.062	-0.124	0.009
치즈	0.165	0.078	0.086	0.114
발효유	0.410	0.451	0.467	0.437
참기름	-0.007	0.152	0.192	0.135
식용유	0.099	0.074	0.051	0.058
과일가공품	0.214	0.248	0.289	0.345
단무지	0.264	0.117	0.374	0.169
김	0.251	0.256	0.176	0.206
맛김	0.228	0.215	0.274	0.253
미역	0.167	0.140	0.115	0.146
초콜릿	-0.282	-0.333	-0.298	-0.301
사탕	0.443	0.478	0.497	0.479
껌	0.202	0.168	0.136	0.140
아이스크림	0.458	0.415	0.481	0.468
비스킷	-0.036	0.072	0.026	-0.006
스낵과자	0.178	0.249	0.197	0.166
파이	0.132	0.163	0.177	0.180
설탕	0.102	0.186	0.192	0.210
잼	-0.093	0.304	0.347	0.287
꿀	0.055	0.015	0.170	0.066
물엿	0.440	0.584	0.582	0.583
고춧가루	0.481	0.170	0.141	0.157
소금	0.083	0.094	0.061	0.044
간장	0.162	0.178	0.139	0.161
된장	-0.026	-0.163	-0.146	-0.174
양념소스	0.632	0.724	0.769	0.714
고추장	-0.114	-0.223	-0.150	-0.161
카레	-0.212	-0.083	-0.135	-0.131
드레싱	0.236	0.131	0.116	0.136
혼합조미료	0.181	0.114	0.077	0.081
스프	0.202	0.070	0.073	0.046
이유식	0.221	0.144	0.053	0.096

김치	0.265	0.343	0.426	0.391
냉동식품	0.190	0.074	0.085	0.084
즉석식품	0.130	0.088	0.118	0.068
커피	-0.195	0.061	-0.007	-0.007
차	0.191	0.297	0.361	0.381
주스	-0.132	-0.117	-0.118	-0.077
두유	-0.229	-0.167	-0.225	-0.151
생수	0.243	0.562	0.371	0.454
기능성음료	0.102	0.044	0.069	0.058
탄산음료	-0.152	-0.027	-0.026	-0.023
혼합음료	-0.215	-0.142	-0.077	-0.073
소주	0.301	0.726	0.696	0.698
과실주	0.146	0.115	0.118	0.107
맥주	-0.124	0.154	0.103	0.095
막걸리	0.068	-0.004	0.029	-0.019
양주	0.202	0.365	0.424	0.395
약주	0.494	0.539	0.418	0.482
담배	0.938	0.960	0.956	0.937
세탁세제	-0.135	-0.208	-0.094	-0.185
섬유유연제	0.352	0.246	0.233	0.267
부엌용세제	0.131	0.044	-0.003	0.008
청소용세제	0.025	-0.109	0.005	-0.019
살충제	0.161	0.220	0.202	0.195
가정용비닐용품	-0.073	-0.155	-0.145	-0.184
키친타월	0.072	-0.047	-0.148	-0.096
방향제	-0.195	-0.237	-0.144	-0.217
습기제거제	-0.121	-0.028	-0.108	-0.119
면도기	-0.095	-0.108	-0.139	-0.078
칫솔	-0.317	-0.102	-0.119	-0.095
치약	0.148	-0.006	-0.064	-0.069
비누	0.227	0.252	0.238	0.247
샴푸	0.031	0.057	0.119	0.084
화장지	0.232	0.241	0.250	0.269
기초화장품	-0.107	-0.036	-0.022	-0.031
모발염색약	0.244	0.314	0.331	0.323
구강세정제	0.175	0.223	0.144	0.210

(4년 지속 상품, 업체별 합산)

	Jevons	Tornq	GEKS	RWGEKS
전체	0.857	0.799	0.822	0.812
국수	0.230	0.399	0.418	0.419
라면	-0.105	0.003	0.002	0.001
당면	0.737	0.769	0.763	0.754
두부	0.231	0.194	0.207	0.199
시리얼	0.221	0.405	0.413	0.397
부침가루	0.114	0.321	0.251	0.319
소시지	-0.033	-0.037	-0.118	-0.110
햄및베이컨	0.388	0.296	0.332	0.289
어묵	0.054	0.377	0.378	0.379
맛살	0.410	0.117	0.114	0.101
생선통조림	0.199	0.142	0.159	0.131
젓갈	0.280	0.336	0.342	0.328
우유	0.481	0.636	0.673	0.659
분유	0.084	-0.108	-0.152	-0.165
치즈	-0.007	0.030	0.034	0.043
발효유	0.330	0.508	0.474	0.488
참기름	0.111	0.171	0.186	0.144
식용유	-0.068	0.068	0.046	0.061
과일가공품	0.001	0.244	0.259	0.258
단무지	0.026	0.359	0.376	0.378
김	0.465	0.169	0.176	0.169
맛김	0.273	0.320	0.328	0.329
미역	-0.010	0.008	-0.019	-0.022
초콜릿	-0.442	-0.215	-0.207	-0.201
사탕	0.350	0.485	0.481	0.487
껌	0.113	0.220	0.193	0.199
아이스크림	0.506	0.420	0.463	0.453
비스킷	-0.157	-0.002	0.001	-0.016
스낵과자	-0.173	-0.036	-0.020	-0.011
파이	-0.056	0.124	0.141	0.138
설탕	0.338	0.156	0.161	0.183
잼	0.057	0.375	0.383	0.371
꿀	0.045	0.194	0.232	0.224
물엿	0.375	0.612	0.605	0.602
고춧가루	0.650	0.121	0.122	0.106
소금	0.038	0.070	0.021	0.005
간장	0.213	0.195	0.153	0.177
된장	0.264	-0.105	-0.071	-0.112
양념소스	0.757	0.806	0.815	0.810
고추장	0.023	-0.135	-0.086	-0.081
카레	-0.256	-0.101	-0.149	-0.141
드레싱	0.197	0.171	0.158	0.163
혼합조미료	-0.166	0.012	-0.002	0.007
스프	0.078	0.144	0.148	0.167

이유식	0.032	0.036	0.045	0.039
김치	0.188	0.436	0.433	0.430
냉동식품	0.324	0.184	0.076	0.099
즉석식품	0.220	0.238	0.228	0.216
커피	-0.112	0.039	0.021	0.019
차	0.120	0.273	0.281	0.289
주스	0.159	-0.123	-0.083	-0.083
두유	-0.201	-0.158	-0.232	-0.183
생수	0.346	0.406	0.342	0.379
기능성음료	0.074	0.006	0.045	0.020
탄산음료	-0.046	0.044	-0.001	0.011
혼합음료	0.076	-0.056	-0.042	-0.050
소주	0.795	0.809	0.782	0.777
과실주	-0.022	0.072	0.079	0.068
맥주	0.171	0.083	0.052	0.056
막걸리	0.104	0.055	0.077	0.068
양주	0.309	0.569	0.546	0.553
약주	0.279	0.269	0.258	0.258
담배	0.967	0.959	0.969	0.965
세탁세제	-0.046	-0.119	-0.027	-0.098
섬유유연제	0.294	0.160	0.099	0.136
부엌용세제	-0.135	-0.018	-0.020	-0.018
청소용세제	0.005	0.029	0.048	0.029
살충제	-0.131	-0.121	-0.117	-0.120
가정용비닐용품	-0.269	-0.076	-0.192	-0.172
키친타월	0.004	-0.187	-0.211	-0.219
방향제	-0.110	-0.094	-0.119	-0.145
면도기	0.027	-0.029	-0.054	-0.046
칫솔	-0.207	-0.049	-0.056	-0.029
치약	-0.171	0.001	0.008	0.009
비누	0.062	0.242	0.231	0.246
샴푸	-0.077	0.156	0.189	0.175
화장지	0.237	0.251	0.241	0.262
기초화장품	-0.111	0.067	0.058	0.037
모발염색약	0.026	0.309	0.309	0.321
구강세정제	-0.151	0.112	0.107	0.115

## (이상치 제거, 업태별 합산)

	Jevons	Tornq	GEKS	RWGEKS
전체	0.937	0.902	0.902	0.881
국수	0.153	0.281	0.350	0.308
라면	-0.128	0.028	0.009	0.027
당면	0.762	0.754	0.749	0.745
두부	0.074	0.056	0.122	-0.004
시리얼	-0.047	0.252	0.329	0.216
부침가루	0.156	0.326	0.230	0.314
소시지	0.028	-0.133	-0.167	-0.139
햄및베이컨	0.341	0.144	0.274	0.233
어묵	0.044	0.191	0.265	0.212
맛살	0.612	0.247	0.238	0.216
생선통조림	-0.090	0.080	0.223	0.155
젓갈	0.298	0.358	0.367	0.365
우유	0.452	0.637	0.656	0.638
분유	0.074	-0.067	-0.070	-0.030
치즈	0.011	0.121	0.112	0.128
발효유	0.421	0.465	0.493	0.491
참기름	0.028	-0.002	0.087	0.033
식용유	0.027	0.084	0.074	0.066
과일가공품	-0.044	0.118	0.176	0.233
단무지	0.317	0.288	0.427	0.358
김	0.475	0.224	0.188	0.223
맛김	0.250	0.060	0.113	0.094
미역	0.132	0.157	0.109	0.163
초콜릿	-0.252	-0.255	-0.244	-0.256
사탕	0.268	0.493	0.508	0.495
껌	-0.081	0.163	0.113	0.139
아이스크림	0.435	0.409	0.445	0.430
비스킷	-0.112	0.081	0.008	-0.039
스낵과자	-0.015	0.204	0.153	0.132
파이	-0.074	0.173	0.159	0.165
설탕	0.239	0.157	0.142	0.198
잼	0.064	0.265	0.320	0.293
꿀	-0.008	-0.013	0.134	0.025
물엿	0.417	0.610	0.600	0.602
고춧가루	0.745	0.310	0.298	0.283
소금	0.069	0.127	0.063	0.057
간장	0.413	0.189	0.170	0.187
된장	0.009	-0.191	-0.150	-0.201
양념소스	0.743	0.765	0.799	0.771
고추장	0.060	-0.125	-0.098	-0.099
카레	-0.146	-0.120	-0.188	-0.161
드레싱	0.299	0.113	0.068	0.100
혼합조미료	-0.032	0.177	0.118	0.147
스프	0.098	0.050	0.040	0.041
이유식	0.301	0.214	0.111	0.152

김치	0.331	0.335	0.419	0.397
냉동식품	0.257	0.137	0.133	0.172
즉석식품	0.166	0.097	0.100	0.064
커피	-0.188	-0.002	-0.040	-0.051
차	0.197	0.294	0.305	0.337
주스	0.105	-0.133	-0.165	-0.111
두유	-0.262	-0.189	-0.226	-0.170
생수	0.128	0.442	0.235	0.323
기능성음료	0.050	0.039	0.070	0.063
탄산음료	-0.037	-0.029	-0.030	-0.046
혼합음료	-0.190	-0.200	-0.073	-0.099
소주	0.258	0.660	0.644	0.630
과실주	0.005	0.142	0.128	0.118
맥주	0.021	0.176	0.188	0.132
막걸리	0.134	0.115	0.028	0.066
양주	0.112	0.313	0.368	0.341
약주	0.141	0.574	0.367	0.532
담배	0.994	0.996	0.991	0.984
세탁세제	-0.008	-0.216	-0.082	-0.176
섬유유연제	0.233	0.259	0.190	0.232
부엌용세제	-0.035	0.003	-0.028	-0.005
청소용세제	0.268	-0.030	0.044	-0.018
살충제	0.132	0.224	0.197	0.188
가정용비닐용품	-0.104	-0.196	-0.068	-0.155
키친타월	0.113	-0.028	-0.043	-0.080
방향제	0.127	-0.091	0.038	-0.090
습기제거제	-0.191	-0.194	0.004	0.019
면도기	0.072	-0.058	0.017	0.087
칫솔	-0.344	-0.125	-0.093	-0.108
치약	-0.105	0.158	0.144	0.155
비누	-0.055	0.147	0.178	0.180
샴푸	-0.077	0.142	0.092	0.111
화장지	0.247	0.262	0.243	0.226
기초화장품	-0.204	-0.102	-0.026	-0.072
모발염색약	0.125	0.370	0.386	0.379
구강세정제	0.023	0.162	0.079	0.165



(이상치 제거, 지역별 합산)

	Jevons	Tornq	GEKS	RWGEKS
전체	0.894	0.897	0.901	0.874
국수	-0.033	0.307	0.392	0.34
라면	-0.147	0.03	0.009	0.029
당면	0.72	0.773	0.779	0.775
두부	0.119	0.031	0.124	0.046
시리얼	0.039	0.227	0.3	0.19
부침가루	0.201	0.318	0.256	0.314
소시지	0.01	-0.147	-0.151	-0.13
햄및베이컨	0.387	0.222	0.316	0.277
어묵	0.16	0.163	0.215	0.193
맛살	0.377	0.263	0.22	0.23
생선통조림	0.117	0.117	0.24	0.166
젓갈	0.353	0.384	0.355	0.358
우유	0.431	0.614	0.648	0.619
분유	-0.106	0.005	-0.023	0.055
치즈	0.148	0.131	0.122	0.14
발효유	0.37	0.445	0.467	0.453
참기름	-0.061	-0.001	0.099	0.036
식용유	0.053	0.027	0.033	0.025
과일가공품	-0.026	0.139	0.195	0.227
단무지	0.313	0.314	0.447	0.391
김	0.29	0.239	0.196	0.216
맛김	0.043	0.071	0.112	0.09
미역	0.117	0.147	0.09	0.152
초콜릿	-0.312	-0.244	-0.224	-0.257
사탕	0.446	0.418	0.451	0.449
껌	0.219	0.161	0.128	0.136
아이스크림	0.431	0.429	0.465	0.46
비스킷	-0.033	0.074	0.015	-0.009
스낵과자	0.237	0.201	0.169	0.151
파이	0.206	0.196	0.187	0.196
설탕	0.087	0.2	0.182	0.243
잼	0.021	0.27	0.318	0.291
꿀	-0.003	-0.02	0.136	0.029
물엿	0.489	0.586	0.58	0.577
고춧가루	0.508	0.302	0.297	0.283
소금	0.096	0.144	0.076	0.067
간장	0.219	0.181	0.157	0.181
된장	-0.125	-0.178	-0.155	-0.196
양념소스	0.643	0.763	0.793	0.768
고추장	0.01	-0.137	-0.103	-0.118
카레	-0.283	-0.123	-0.197	-0.168
드레싱	0.18	0.126	0.098	0.11
혼합조미료	0.176	0.168	0.127	0.142
스프	0.082	0.04	0.045	0.028
이유식	0.323	0.198	0.092	0.141

김치	0.3	0.325	0.405	0.389
냉동식품	0.124	0.109	0.142	0.15
즉석식품	0.171	0.118	0.151	0.096
커피	-0.177	0.015	-0.051	-0.051
차	0.18	0.299	0.302	0.336
주스	-0.158	-0.137	-0.162	-0.118
두유	-0.213	-0.178	-0.215	-0.164
생수	0.242	0.436	0.211	0.32
기능성음료	0.114	0.039	0.068	0.052
탄산음료	-0.253	-0.029	-0.035	-0.041
혼합음료	-0.224	-0.162	-0.058	-0.052
소주	0.164	0.587	0.57	0.564
과실주	0.179	0.138	0.134	0.118
맥주	-0.151	0.166	0.144	0.117
막걸리	0.073	0.049	0.024	0.029
양주	0.195	0.339	0.379	0.355
약주	0.414	0.551	0.424	0.524
담배	0.991	0.996	0.991	0.979
세탁세제	-0.164	-0.215	-0.085	-0.187
섬유유연제	0.227	0.233	0.195	0.221
부엌용세제	0.12	0.014	-0.004	-0.001
청소용세제	0.144	-0.03	0.031	-0.01
살충제	0.191	0.268	0.233	0.242
가정용비닐용품	-0.161	-0.197	-0.163	-0.176
키친타월	0.044	-0.053	-0.085	-0.106
방향제	-0.019	-0.118	0.016	-0.095
습기제거제	-0.024	-0.156	-0.034	-0.019
면도기	0.046	-0.035	0.038	0.114
칫솔	-0.332	-0.131	-0.084	-0.12
치약	0.175	0.183	0.12	0.136
비누	0.161	0.136	0.167	0.172
샴푸	0.113	0.119	0.127	0.108
화장지	0.108	0.235	0.194	0.221
기초화장품	-0.073	-0.101	-0.028	-0.077
모발염색약	0.343	0.356	0.377	0.369
구강세정제	0.08	0.13	0.054	0.114

(이상치 제거, 4년 지속 상품, 업태별 합산)

	Jevons	Tomq	GEKS	RWGEKS
전체	0.951	0.905	0.906	0.902
국수	0.335	0.390	0.404	0.406
라면	-0.019	-0.016	-0.014	-0.017
당면	0.756	0.744	0.732	0.728
두부	0.233	0.362	0.459	0.394
시리얼	0.226	0.338	0.372	0.338
부침가루	0.281	0.340	0.263	0.322
소시지	0.043	-0.099	-0.176	-0.193
햄및베이컨	0.347	0.378	0.399	0.381
어묵	0.191	0.317	0.321	0.322
맛살	0.622	0.246	0.251	0.224
생선통조림	0.274	0.171	0.187	0.167
젓갈	0.342	0.333	0.344	0.332
우유	0.506	0.664	0.679	0.671
분유	0.160	-0.191	-0.257	-0.261
치즈	0.043	0.072	0.085	0.094
발효유	0.302	0.522	0.490	0.523
참기름	0.095	-0.095	-0.034	-0.077
식용유	0.127	0.070	0.094	0.095
과일가공품	0.079	0.350	0.372	0.362
단무지	0.220	0.369	0.373	0.376
김	0.466	0.163	0.169	0.157
맛김	0.295	0.164	0.184	0.168
미역	-0.063	0.036	0.011	0.003
초콜릿	-0.364	-0.114	-0.112	-0.135
사탕	0.346	0.491	0.499	0.497
껌	0.174	0.195	0.164	0.173
아이스크림	0.482	0.410	0.444	0.436
비스킷	-0.235	-0.236	-0.255	-0.283
스낵과자	-0.127	-0.023	-0.026	-0.019
파이	0.003	0.161	0.165	0.169
설탕	0.265	0.170	0.141	0.197
잼	0.213	0.333	0.350	0.348
꿀	-0.047	0.119	0.141	0.142
물엿	0.538	0.624	0.614	0.605
고춧가루	0.625	0.175	0.189	0.177
소금	0.147	0.072	0.028	0.023
간장	0.435	0.223	0.187	0.212
된장	0.221	-0.117	-0.084	-0.127
양념소스	0.814	0.806	0.813	0.808
고추장	0.113	-0.135	-0.127	-0.147
카레	-0.260	-0.125	-0.200	-0.170
드레싱	0.245	0.179	0.159	0.168
혼합조미료	-0.142	0.062	0.053	0.068
스프	-0.027	0.070	0.103	0.111
이유식	0.085	0.069	0.070	0.069

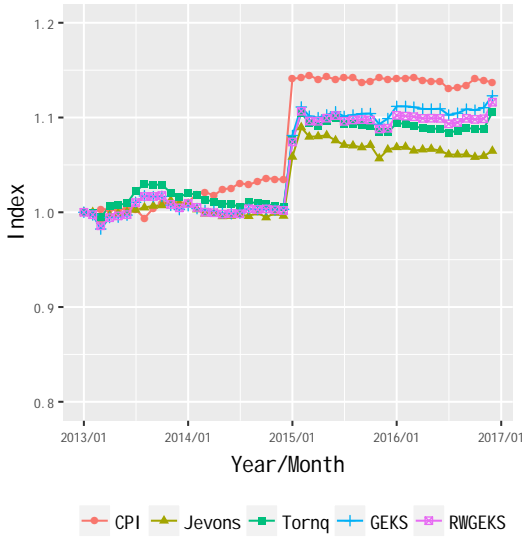
김치	0.425	0.393	0.423	0.422
냉동식품	0.304	0.261	0.163	0.178
즉석식품	0.205	0.200	0.211	0.217
커피	-0.038	-0.011	-0.018	-0.035
차	0.078	0.188	0.197	0.207
주스	0.044	-0.146	-0.106	-0.111
두유	-0.174	-0.106	-0.137	-0.107
생수	0.280	0.184	0.083	0.129
기능성음료	0.015	0.090	0.110	0.101
탄산음료	-0.077	0.062	0.048	0.061
혼합음료	-0.012	-0.121	-0.060	-0.097
소주	0.380	0.679	0.663	0.660
과실주	0.115	0.140	0.133	0.127
맥주	0.096	0.249	0.209	0.199
막걸리	0.189	0.192	0.203	0.210
양주	0.192	0.468	0.434	0.439
약주	0.226	0.285	0.257	0.264
담배	0.999	0.996	0.997	0.996
세탁세제	-0.014	-0.155	-0.037	-0.126
섬유유연제	-0.001	0.227	0.141	0.176
부엌용세제	-0.146	0.153	0.194	0.171
청소용세제	0.035	0.041	0.040	0.041
살충제	-0.116	-0.061	-0.058	-0.064
가정용비닐용품	-0.318	-0.095	-0.228	-0.184
키친타월	-0.080	-0.225	-0.239	-0.251
방향제	-0.043	0.026	0.001	-0.053
면도기	0.158	0.047	0.041	0.048
칫솔	-0.100	-0.067	-0.059	-0.042
치약	-0.150	0.226	0.160	0.192
비누	-0.026	0.114	0.147	0.145
샴푸	-0.174	0.075	0.080	0.081
화장지	0.325	0.244	0.234	0.249
기초화장품	-0.093	-0.093	-0.032	-0.101
모발염색약	0.166	0.351	0.368	0.369
구강세정제	-0.151	0.047	0.038	0.043

## 스캐너 데이터 기반 품목별 소비자물가지수 그래프

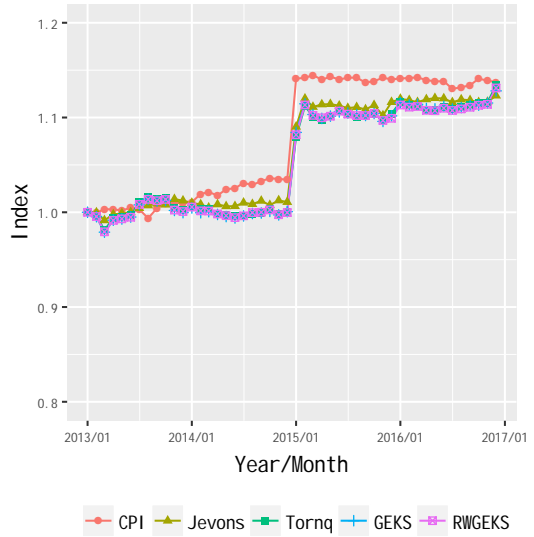


# 전체

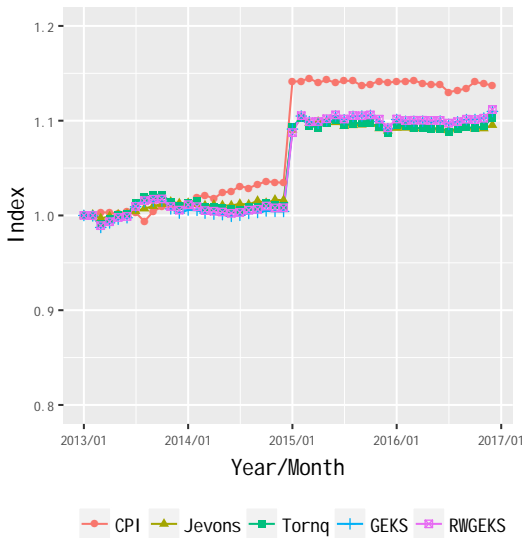
## 원데이터



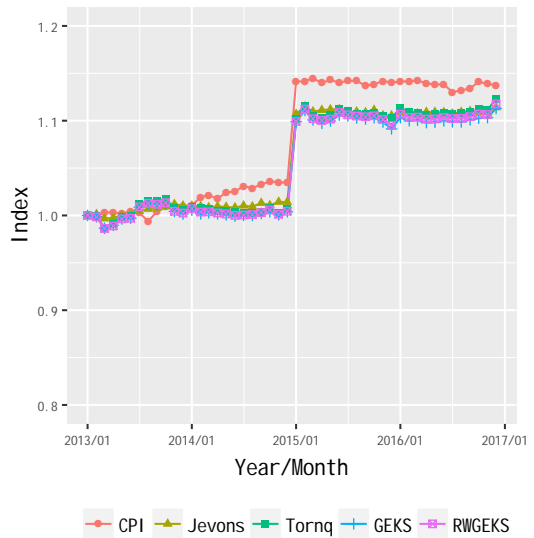
## 공통상품



## 특이항제거

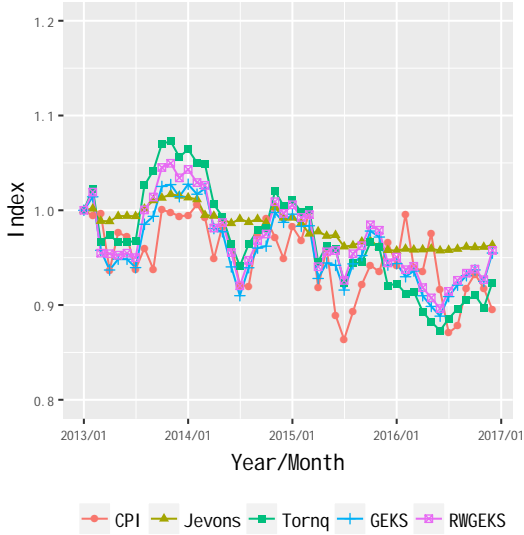


## 공통상품, 특이항제거

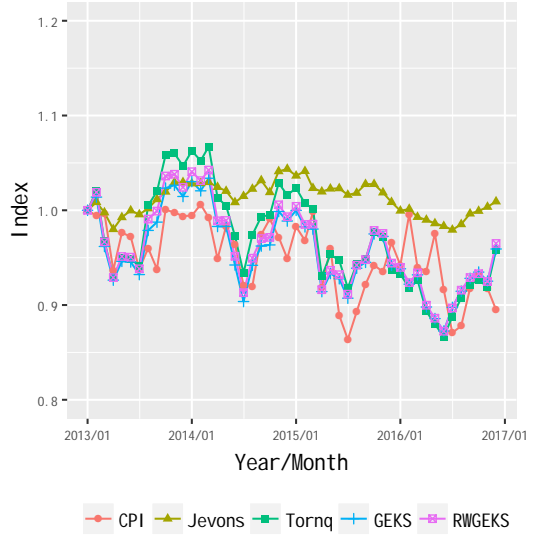


# A011090(국수)

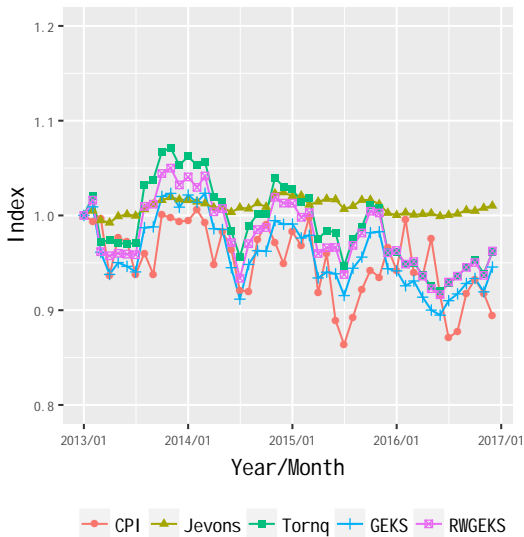
## 원데이터



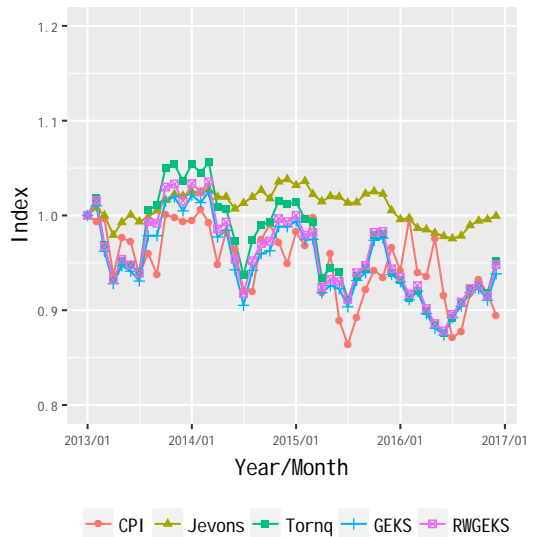
## 공통상품



## 특이항제거



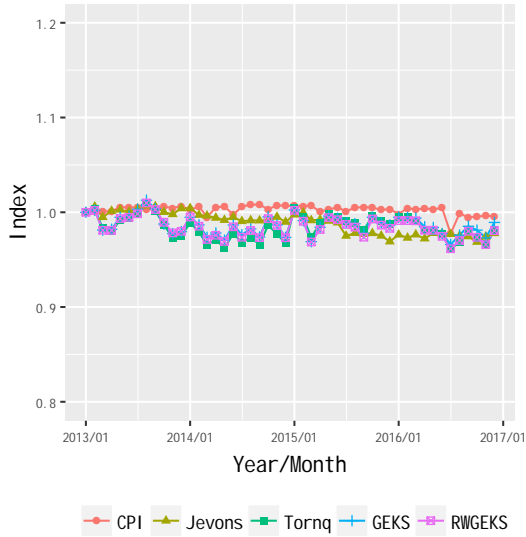
## 공통상품, 특이항제거



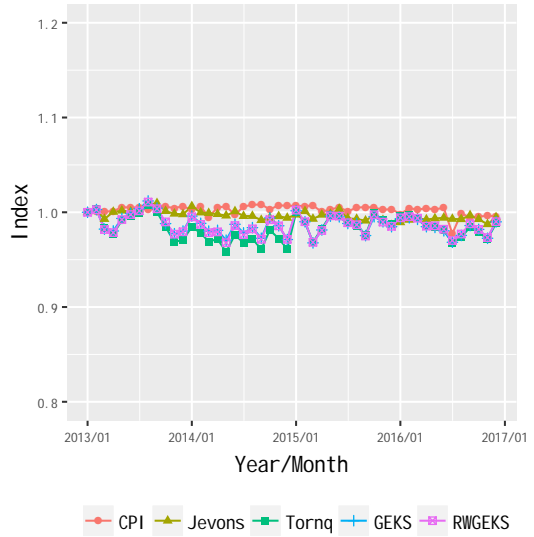


# A011100(라면)

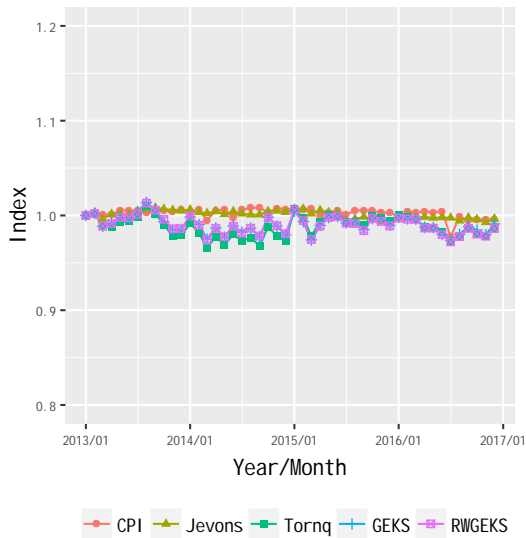
## 원데이터



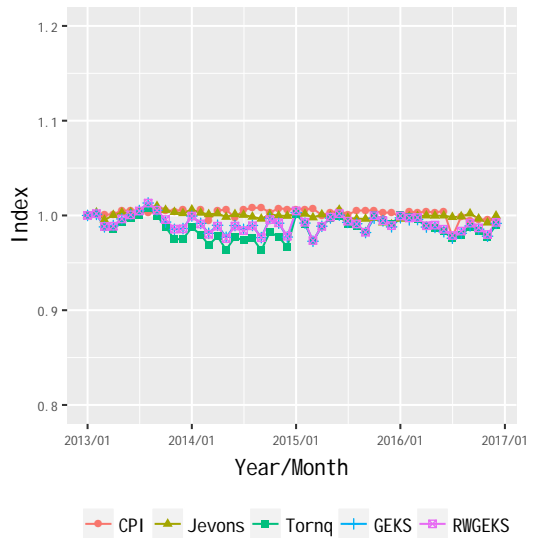
## 공통상품



## 특이항제거

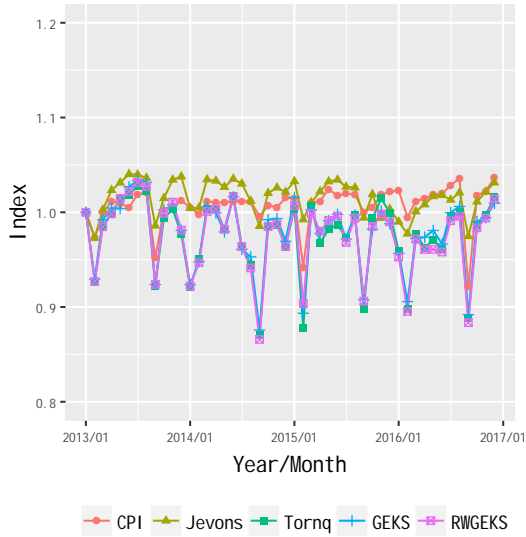


## 공통상품, 특이항제거

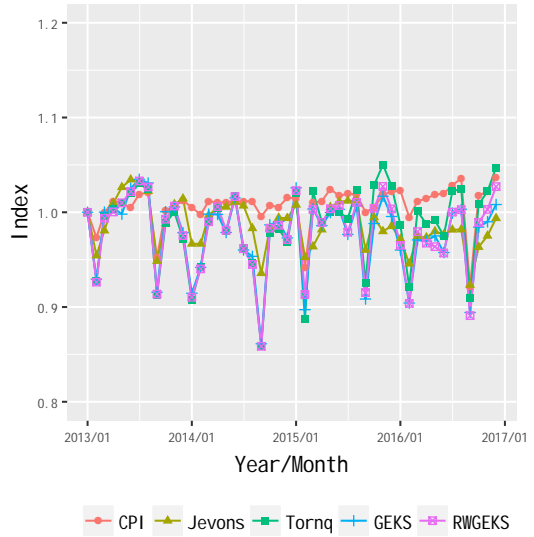


# A011110(당면)

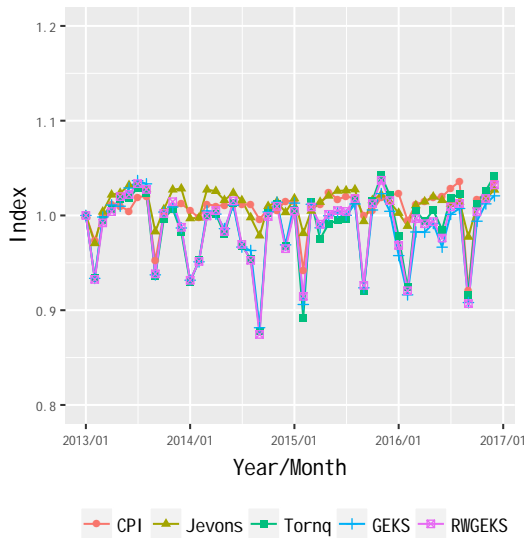
## 원데이터



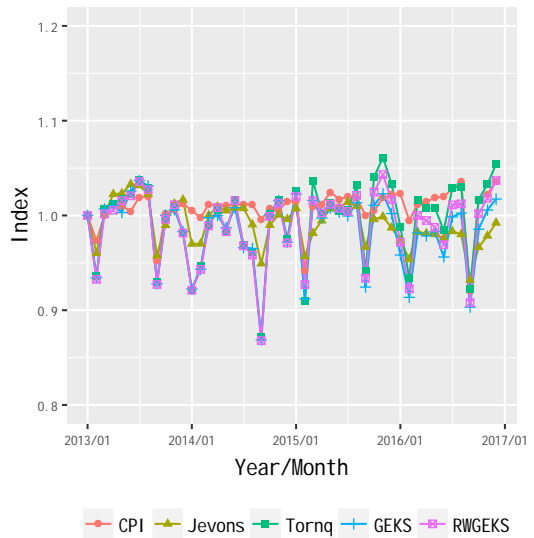
## 공통상품



## 특이항제거

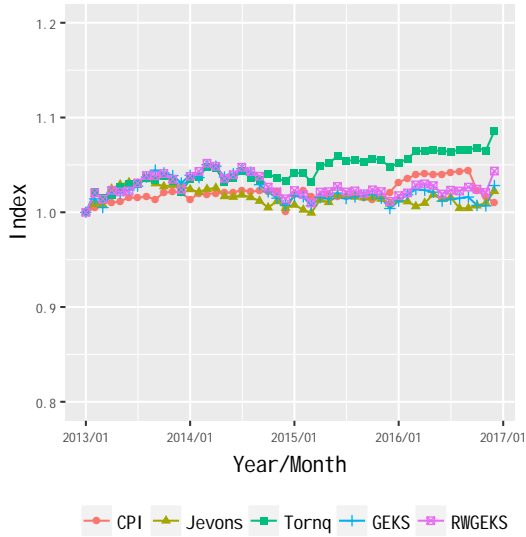


## 공통상품, 특이항제거

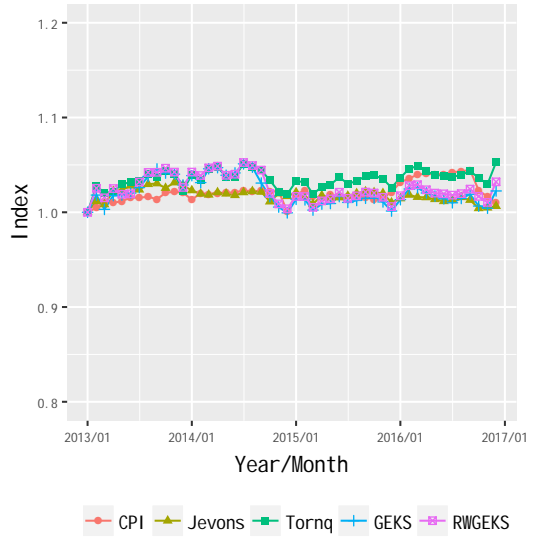


# A011120(두부)

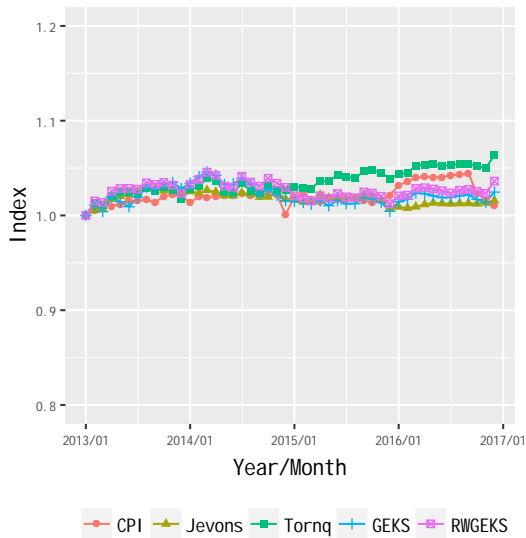
## 원데이터



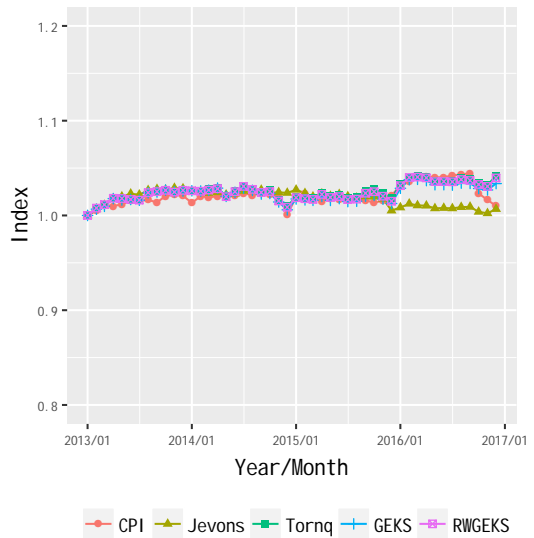
## 공통상품



## 특이항제거

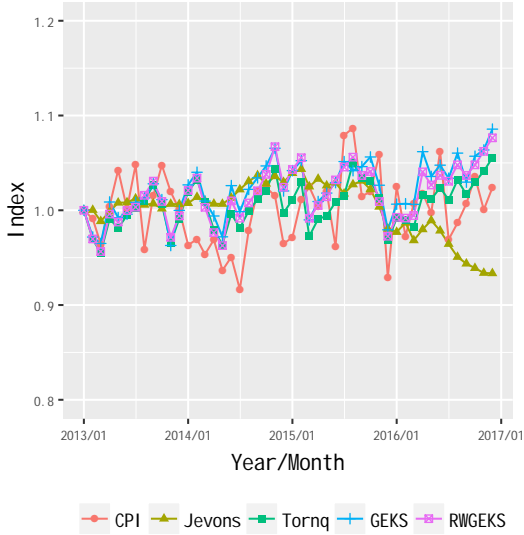


## 공통상품, 특이항제거

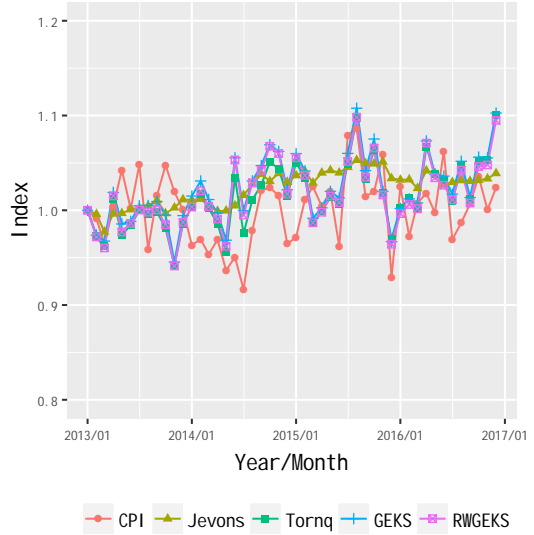


# A011130(시리얼)

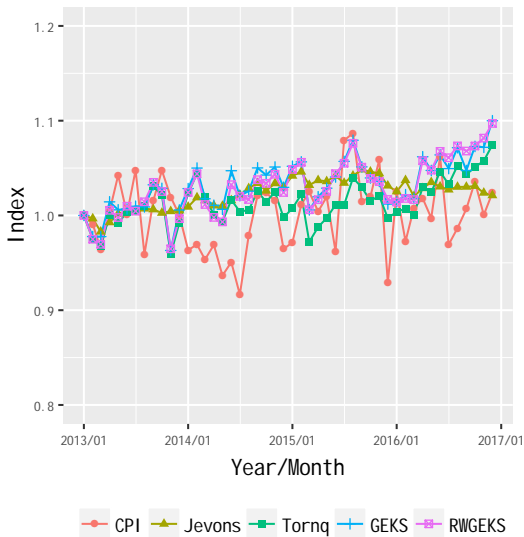
## 원데이터



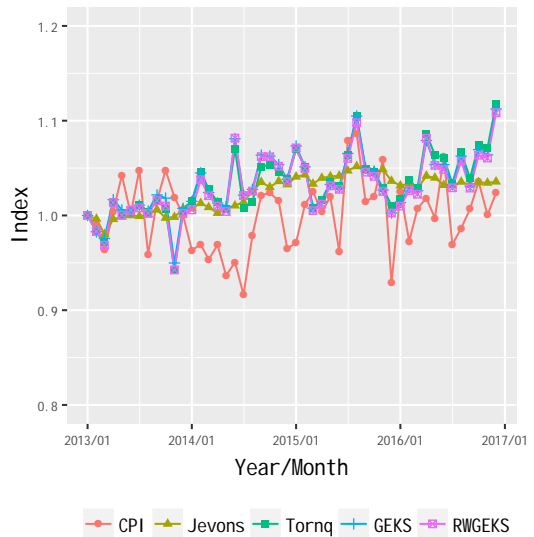
## 공통상품



## 특이항제거

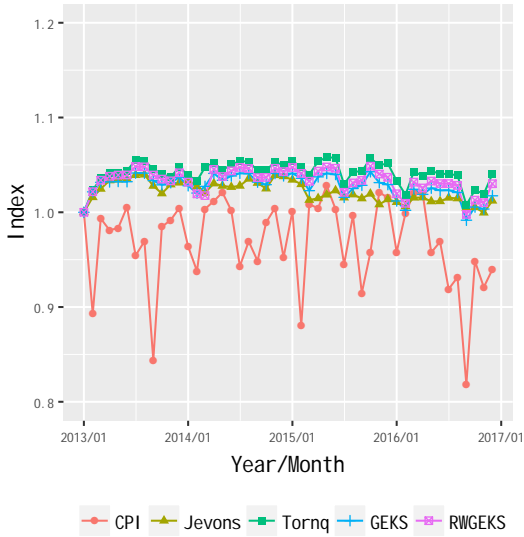


## 공통상품, 특이항제거

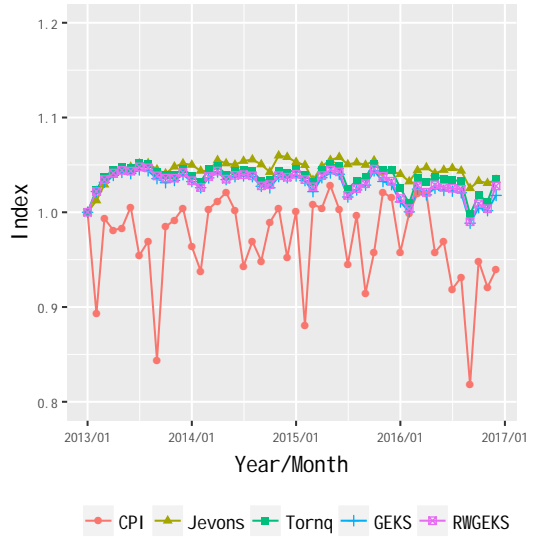


# A011140(부침가루)

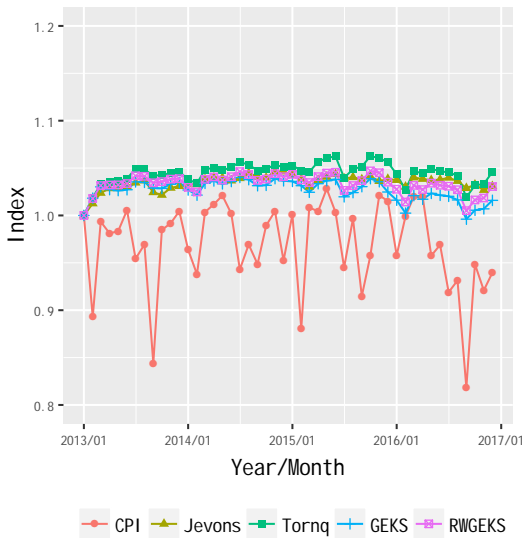
## 원데이터



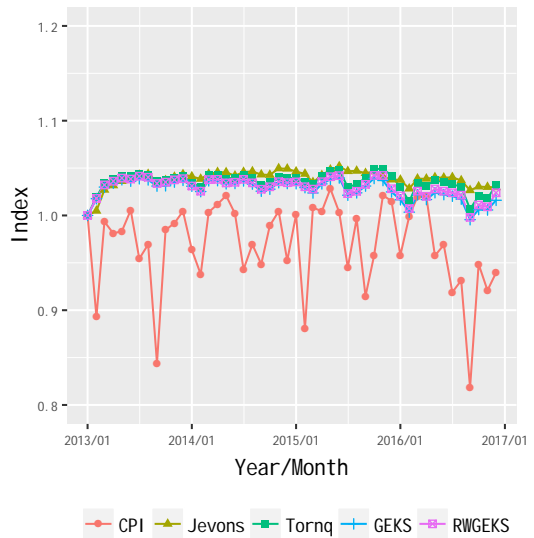
## 공통상품



## 특이항제거

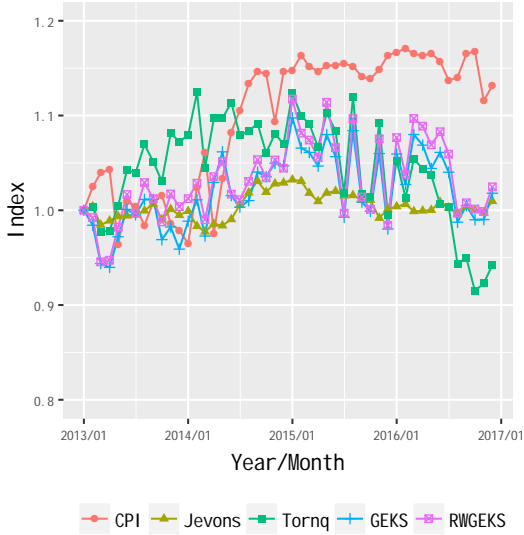


## 공통상품, 특이항제거

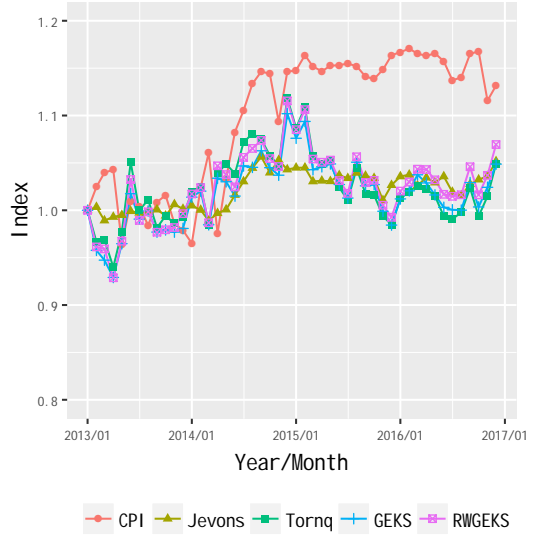


# A012050(소시지)

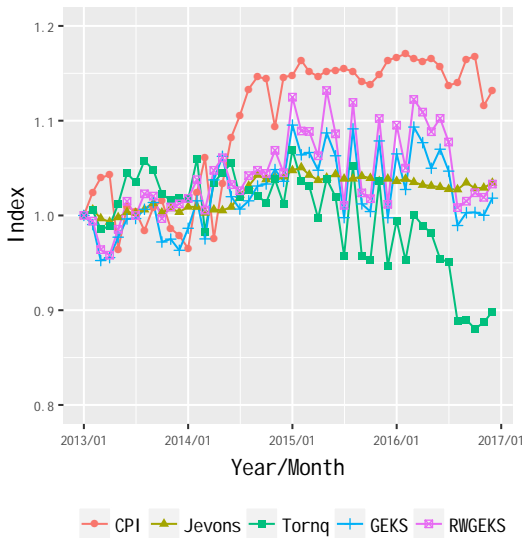
## 원데이터



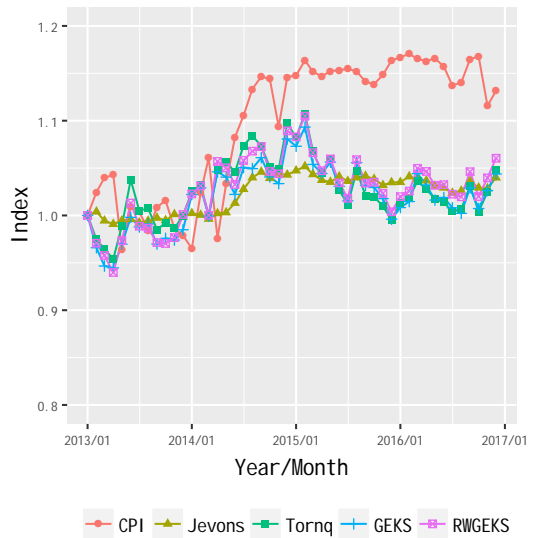
## 공통상품



## 특이항제거

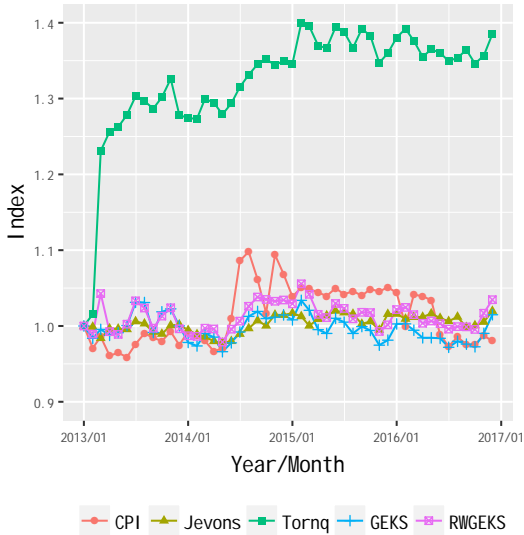


## 공통상품, 특이항제거

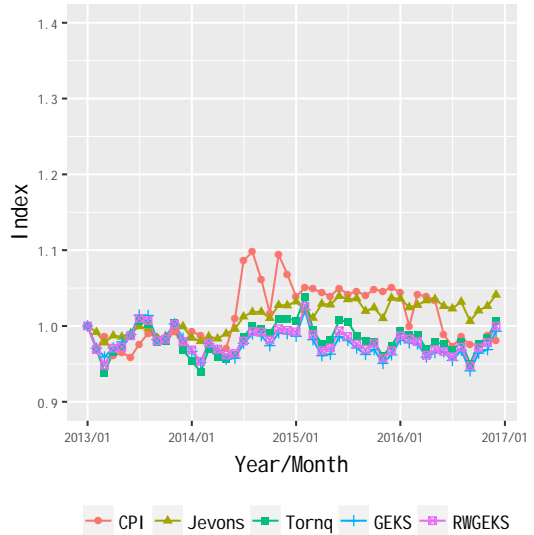


# A012060(햄 및 베이컨)

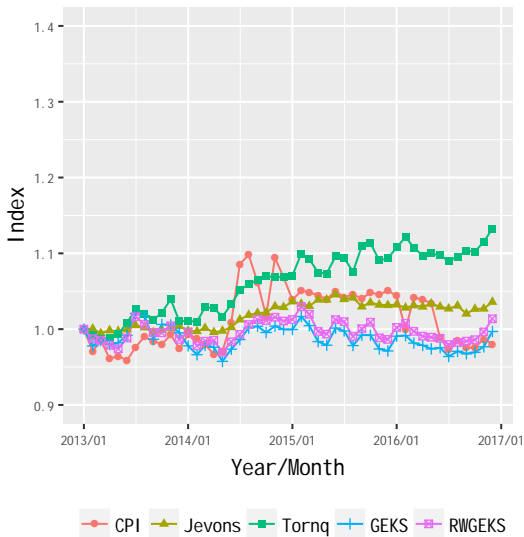
## 원데이터



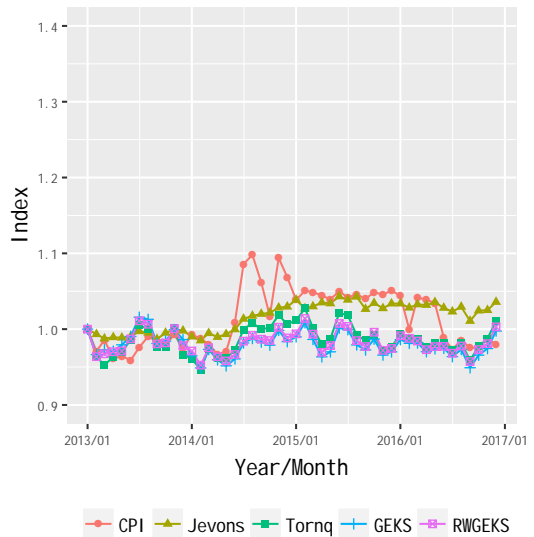
## 공통상품



## 특이항제거

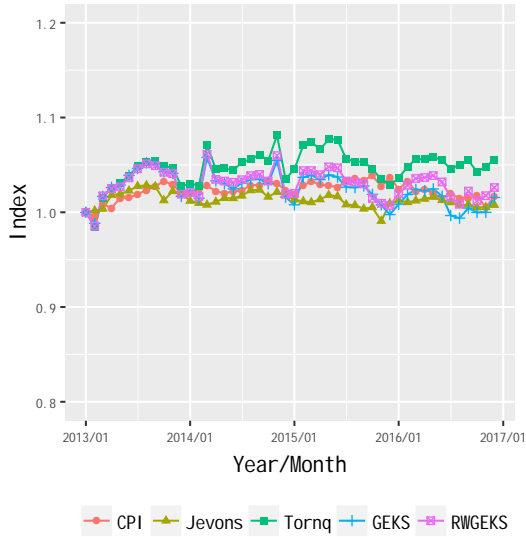


## 공통상품, 특이항제거

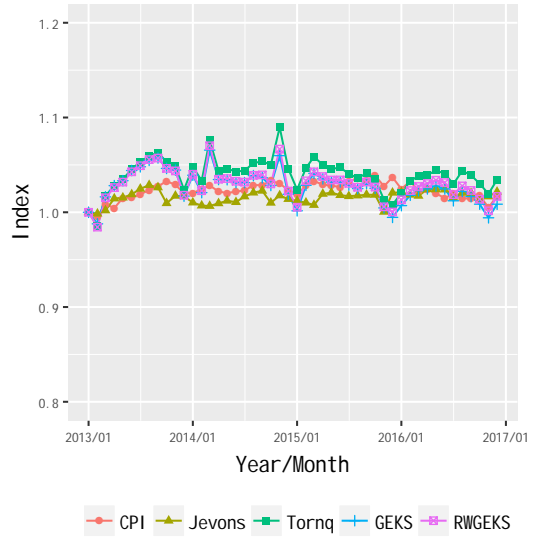


# A013150(어묵)

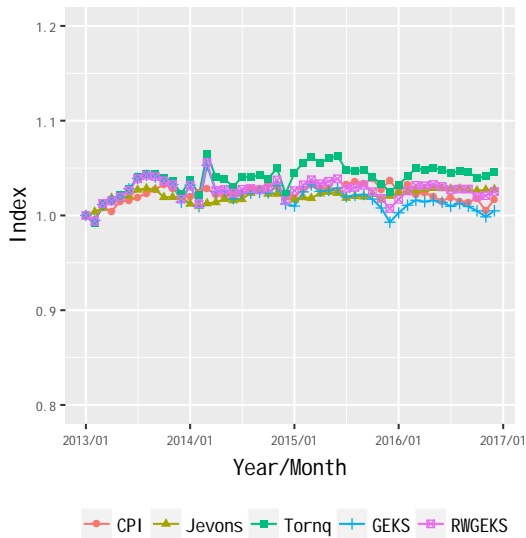
## 원데이터



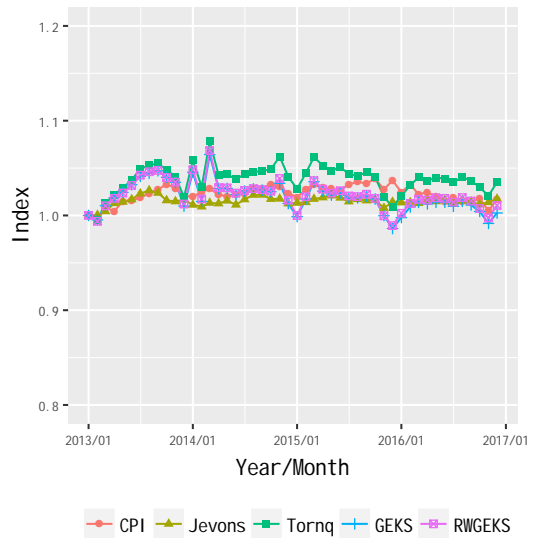
## 공통상품



## 특이항제거



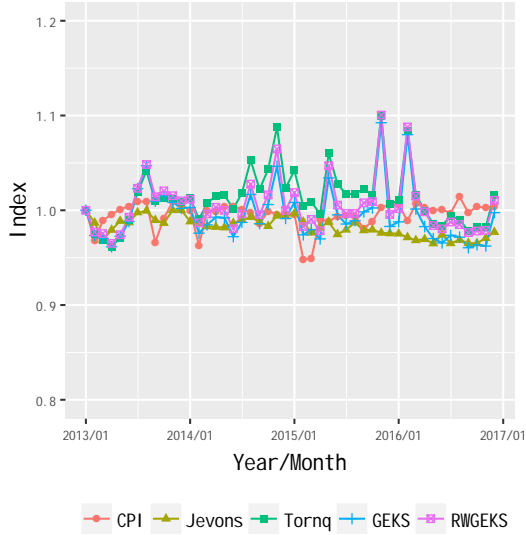
## 공통상품, 특이항제거



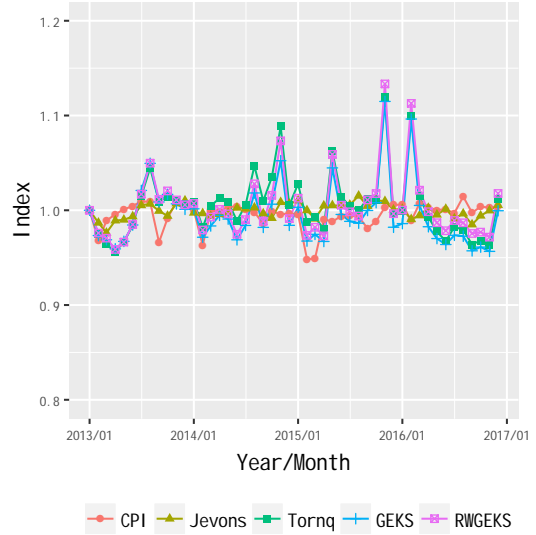


# A013160(맛살)

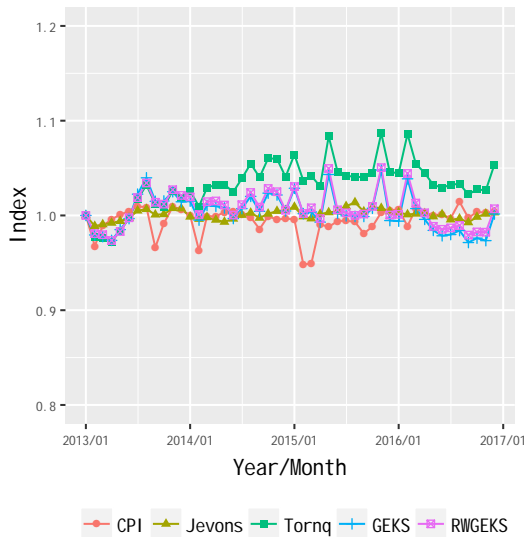
## 원데이터



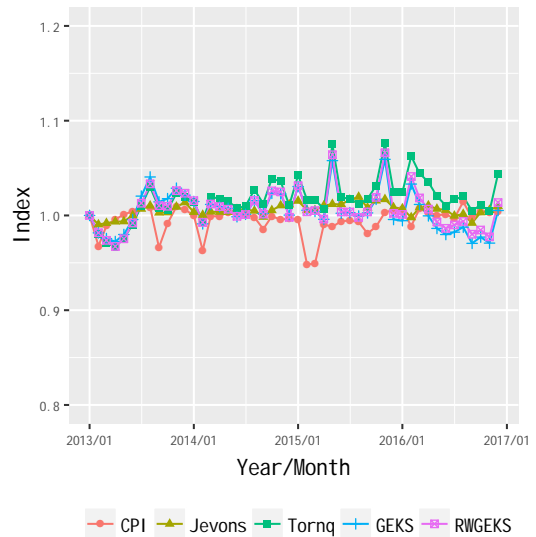
## 공통상품



## 특이항제거

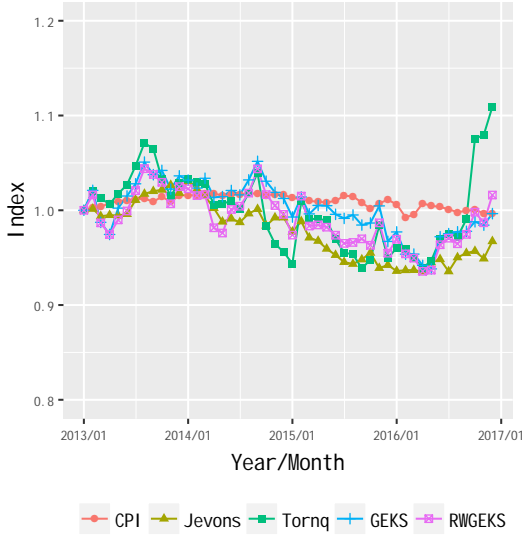


## 공통상품, 특이항제거

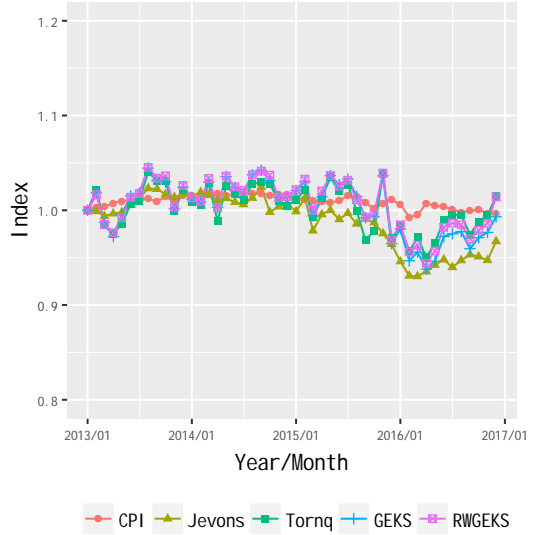


# A013170(생선통조림)

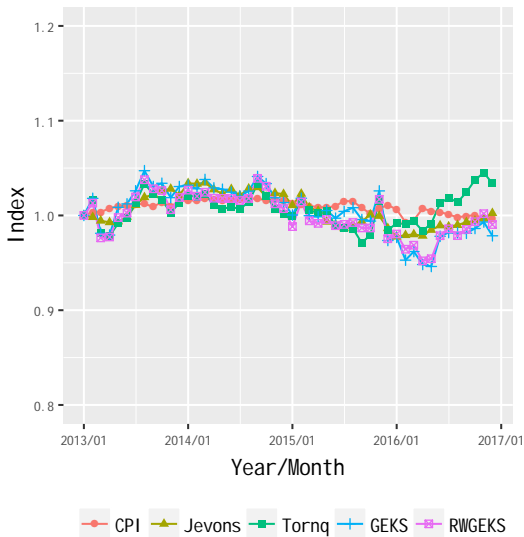
## 원데이터



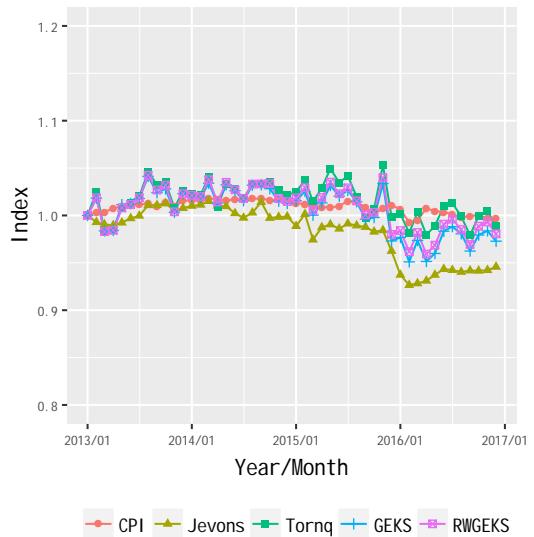
## 공통상품



## 특이항제거

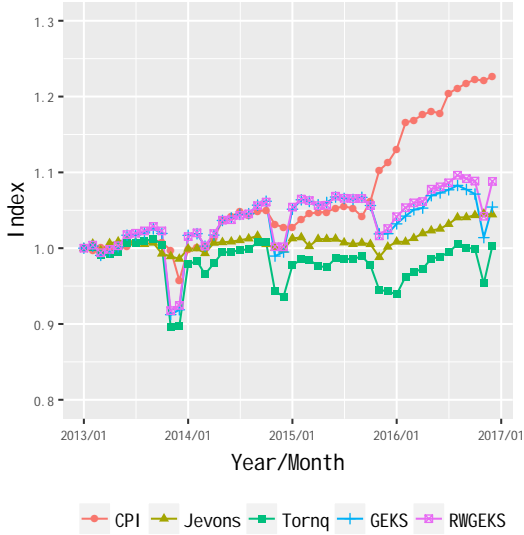


## 공통상품, 특이항제거

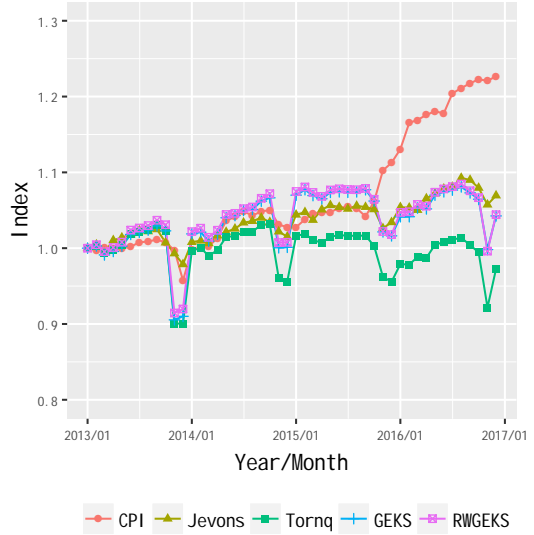


# A013180(젓갈)

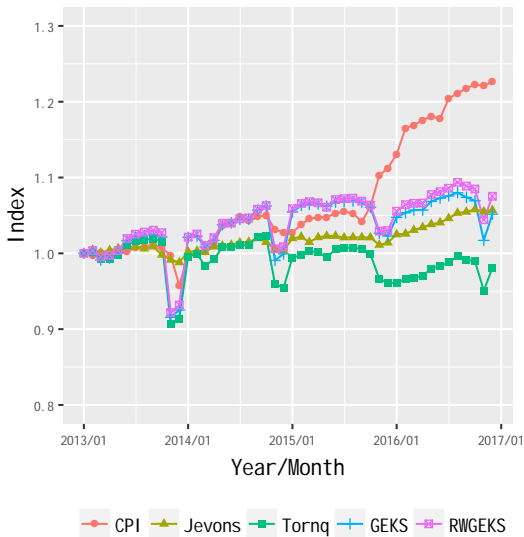
## 원데이터



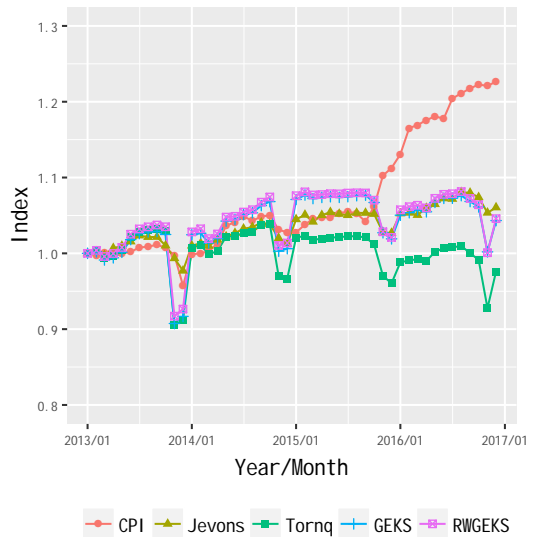
## 공통상품



## 특이항제거

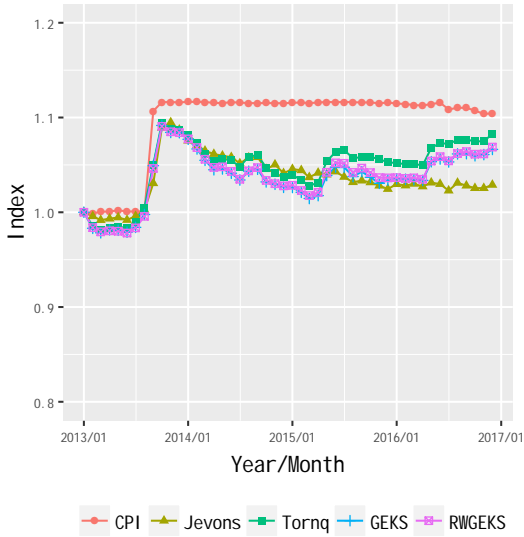


## 공통상품, 특이항제거

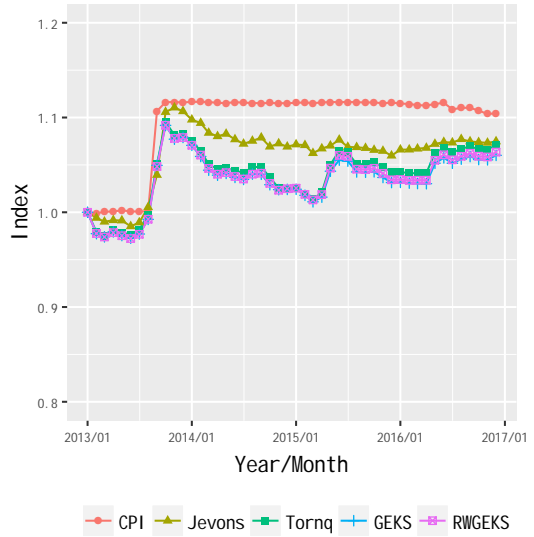


# A014010(우유)

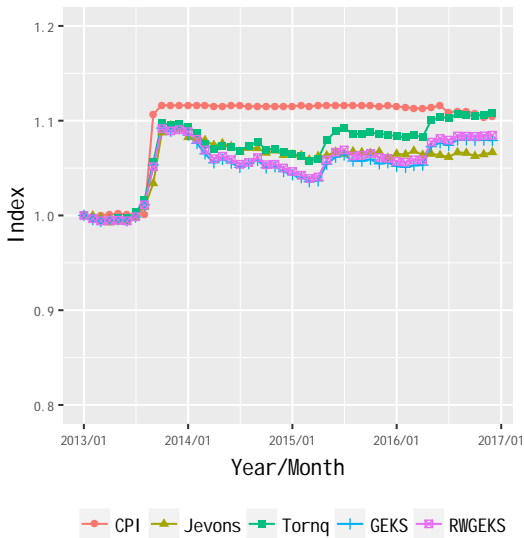
## 원데이터



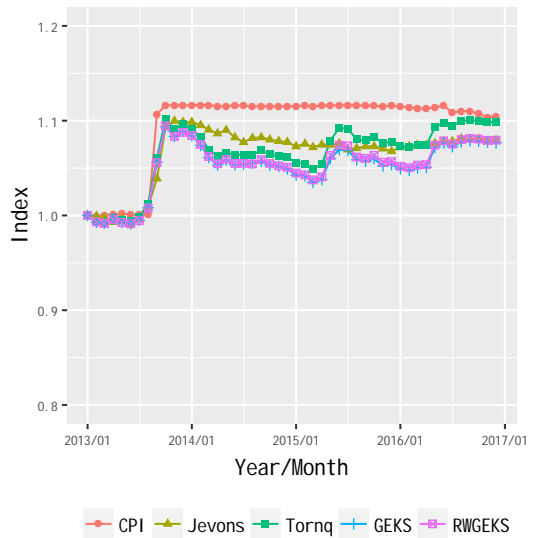
## 공통상품



## 특이항제거

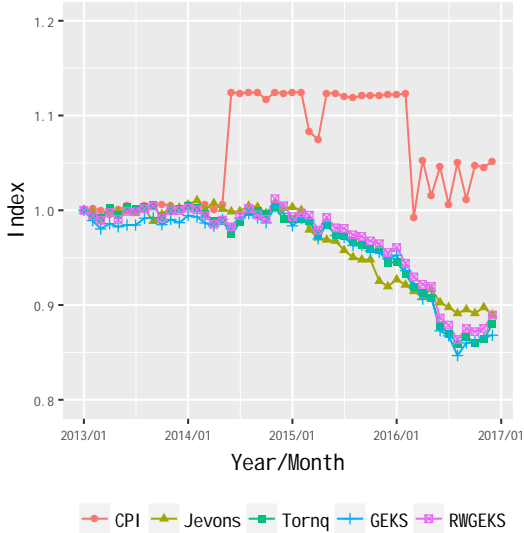


## 공통상품, 특이항제거

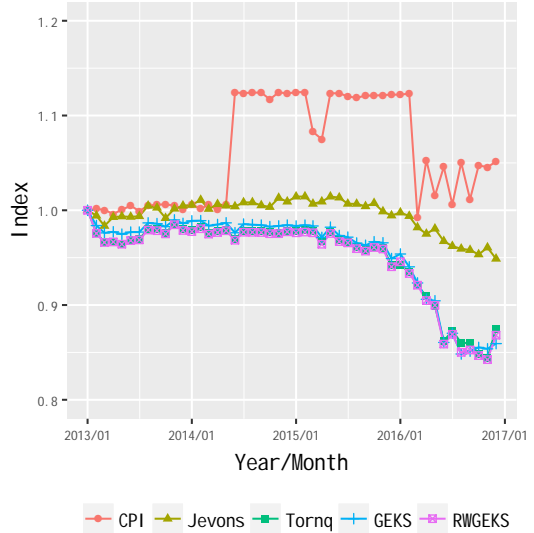


# A014020(분유)

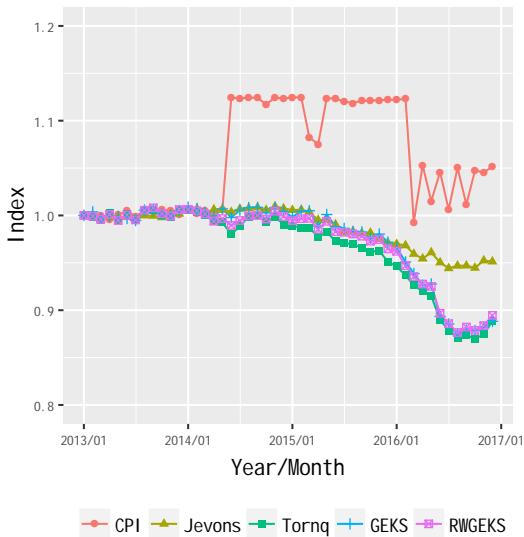
## 원데이터



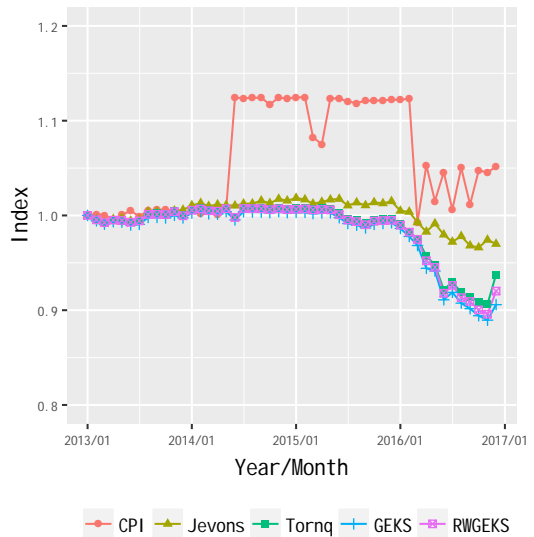
## 공통상품



## 특이항제거

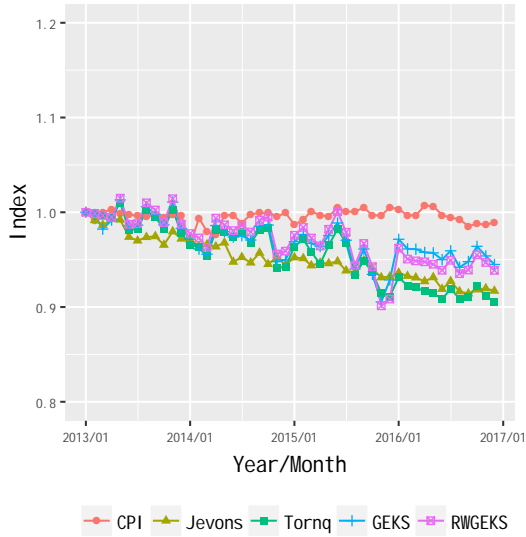


## 공통상품, 특이항제거

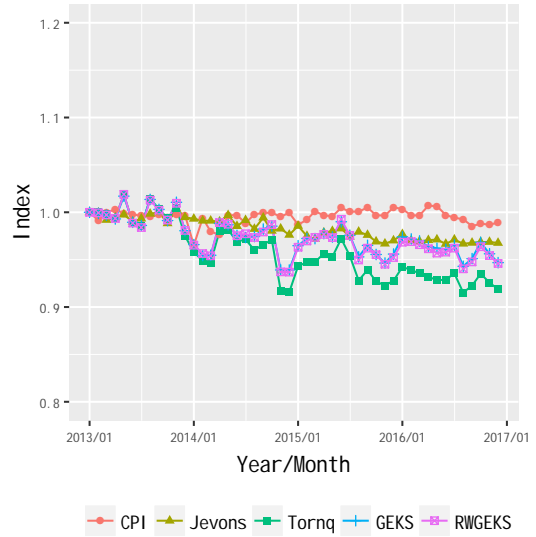


# A014030(치즈)

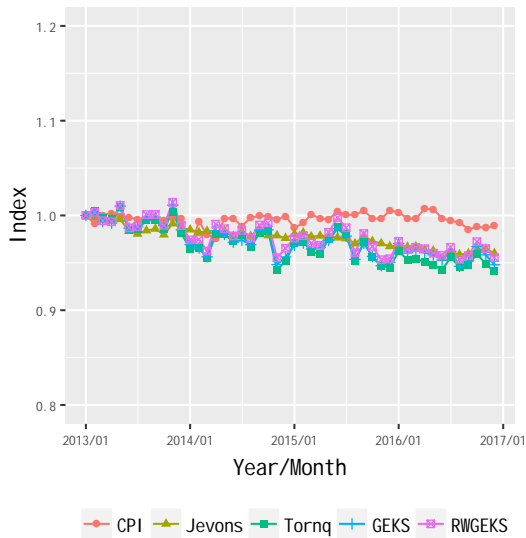
## 원데이터



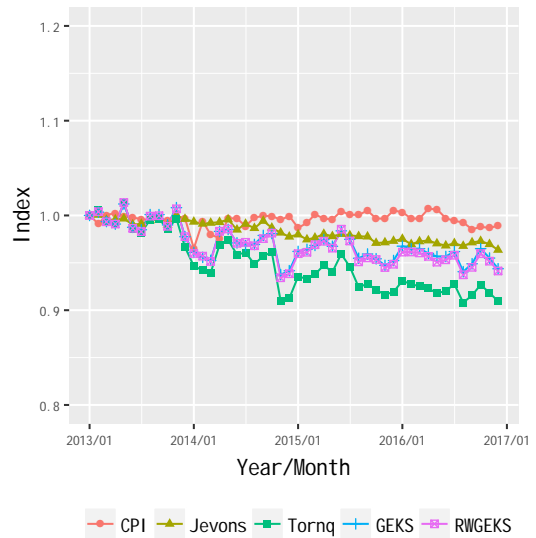
## 공통상품



## 특이항제거

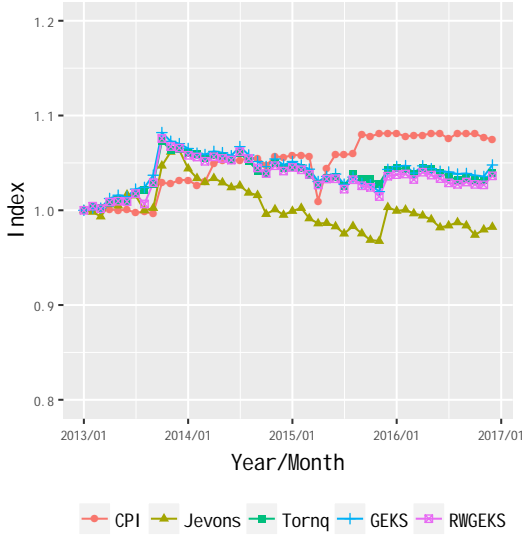


## 공통상품, 특이항제거

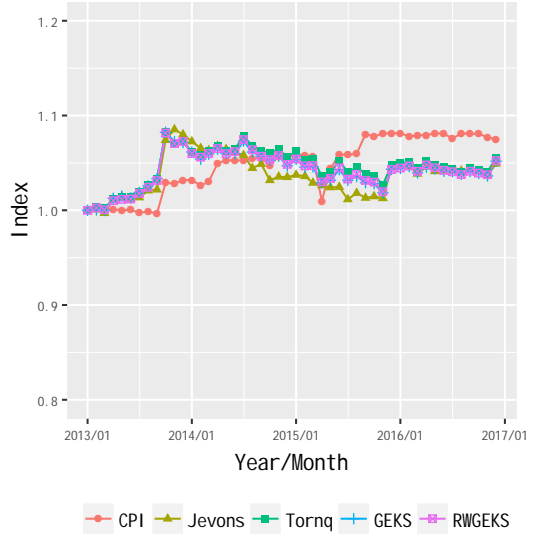


# A014040(발효유)

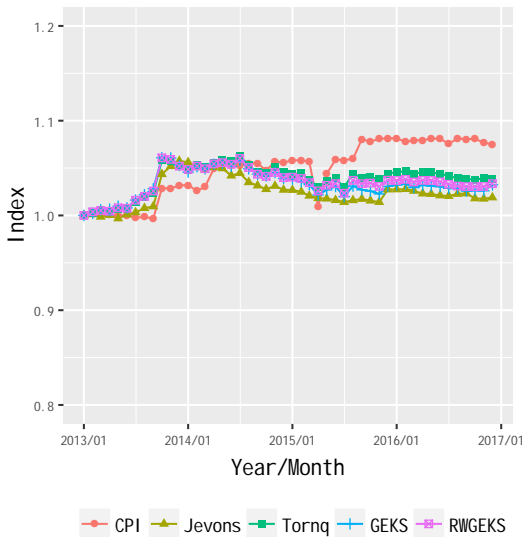
## 원데이터



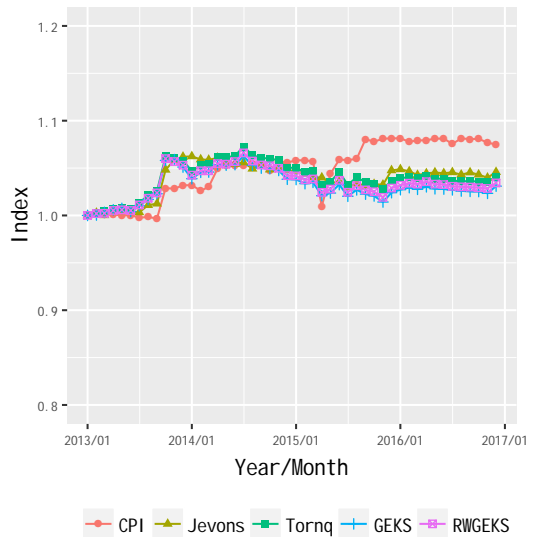
## 공통상품



## 특이항제거

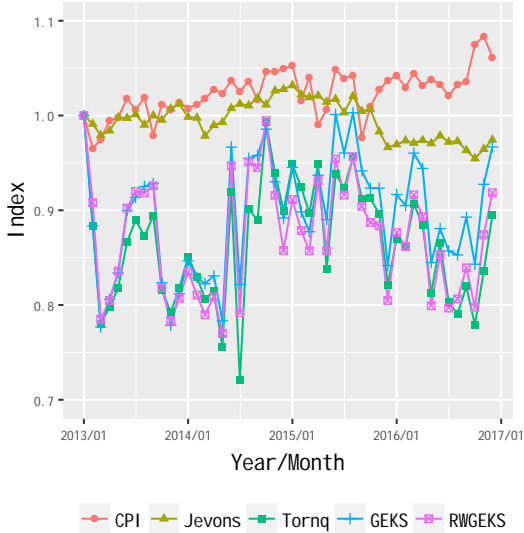


## 공통상품, 특이항제거

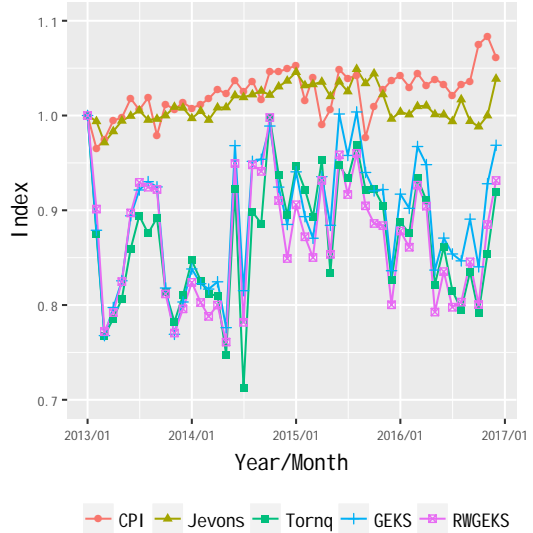


# A015010(참기름)

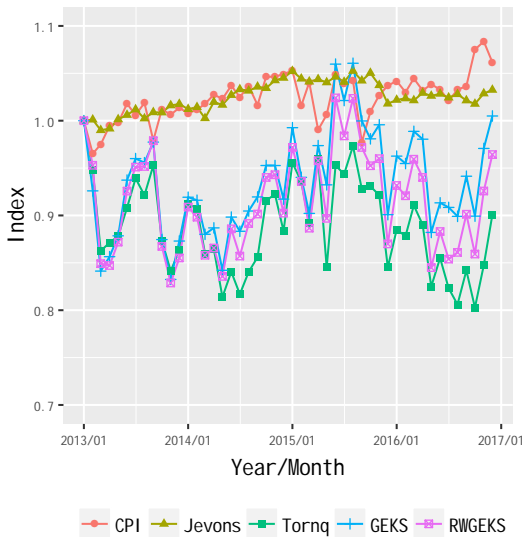
## 원데이터



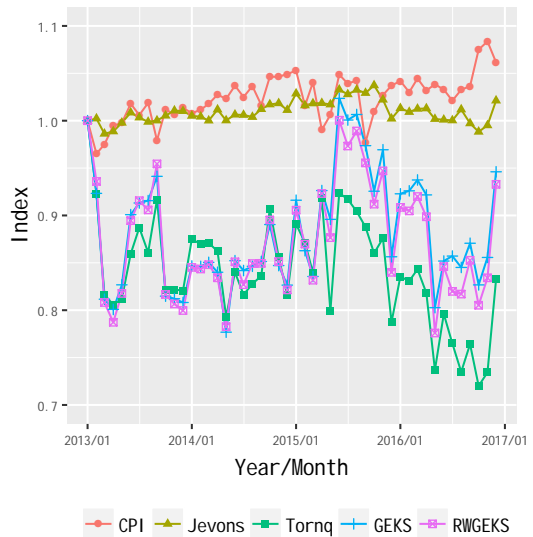
## 공통상품



## 특이항제거



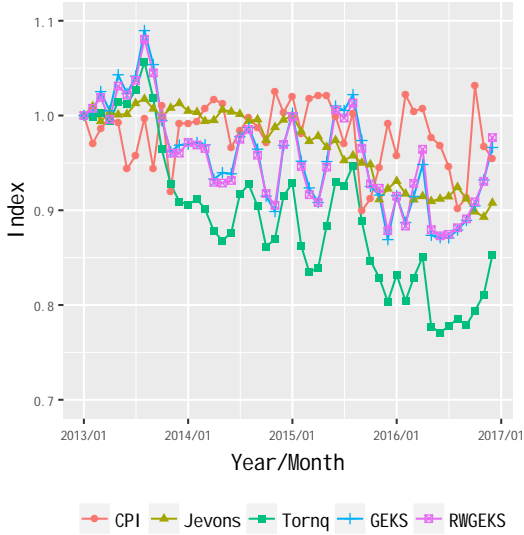
## 공통상품, 특이항제거



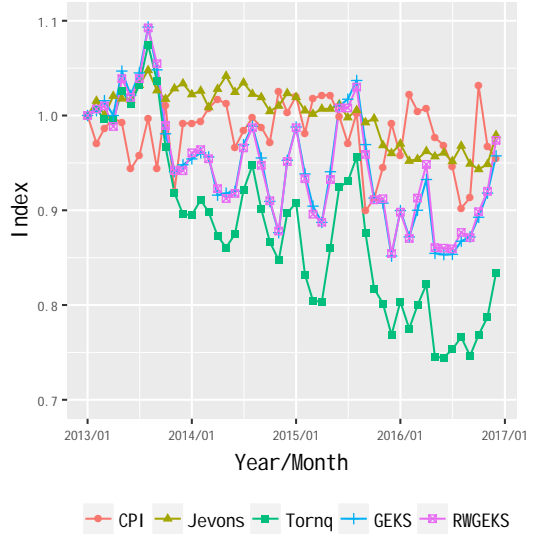


# A015020(식용유)

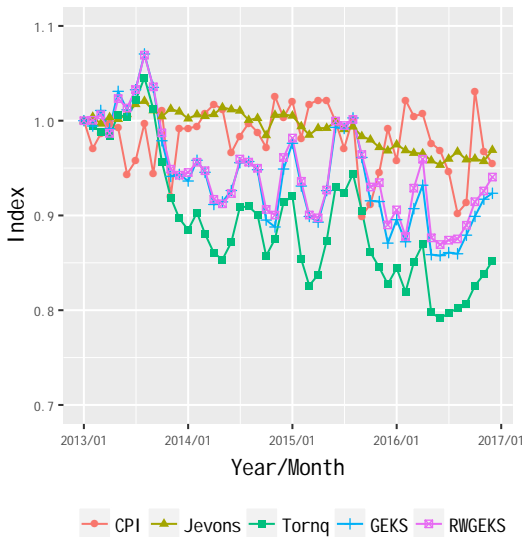
## 원데이터



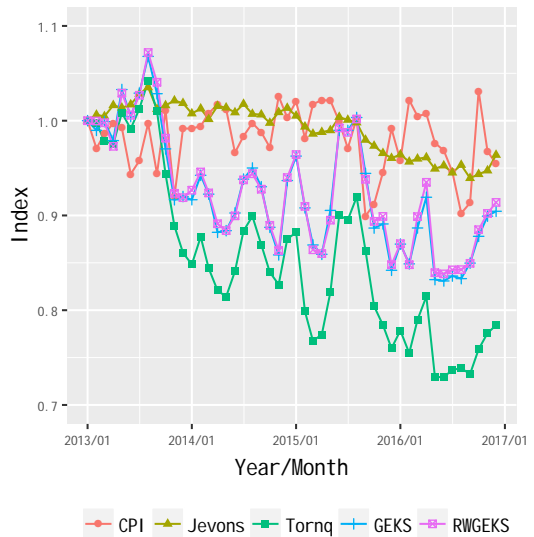
## 공통상품



## 특이항제거

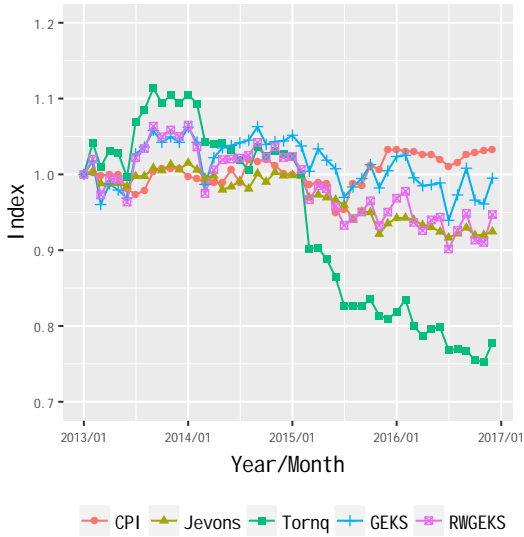


## 공통상품, 특이항제거

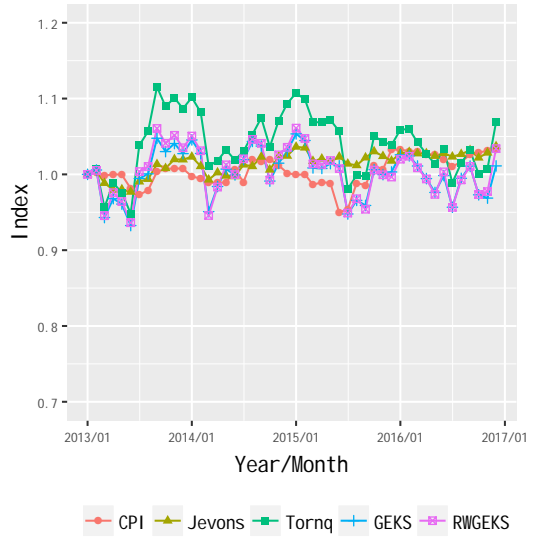


# A016160(과일가공품)

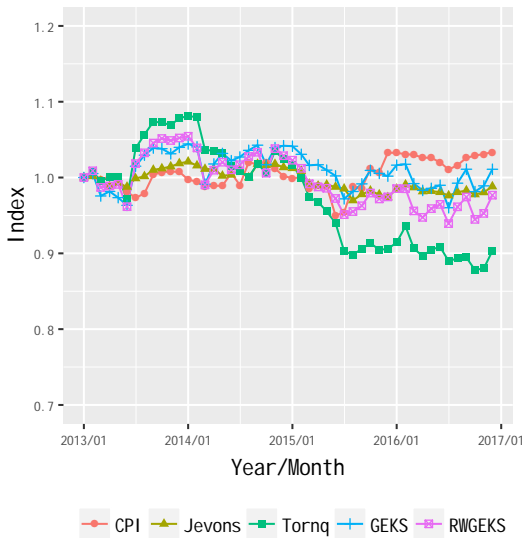
## 원데이터



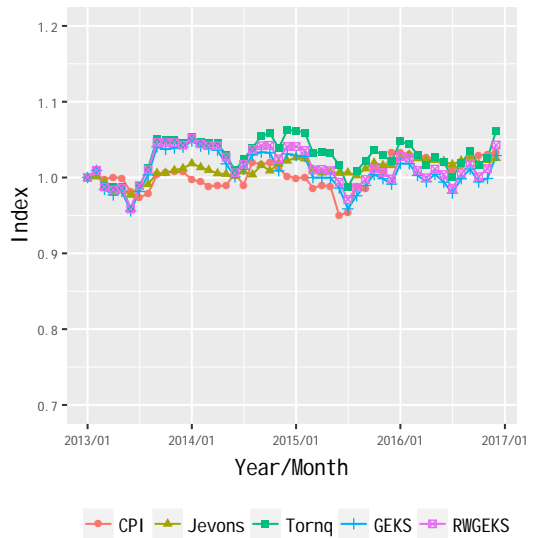
## 공통상품



## 특이항제거

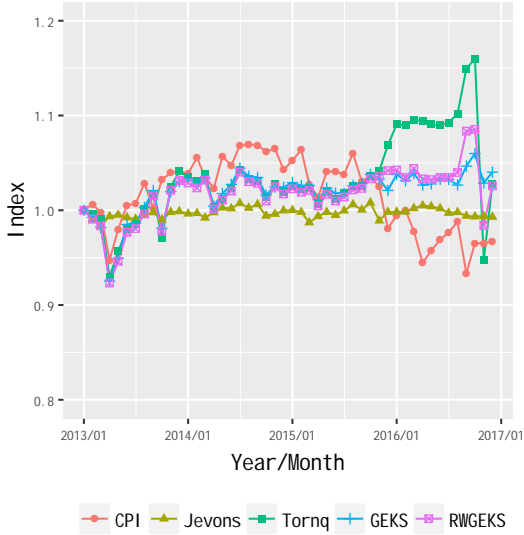


## 공통상품, 특이항제거

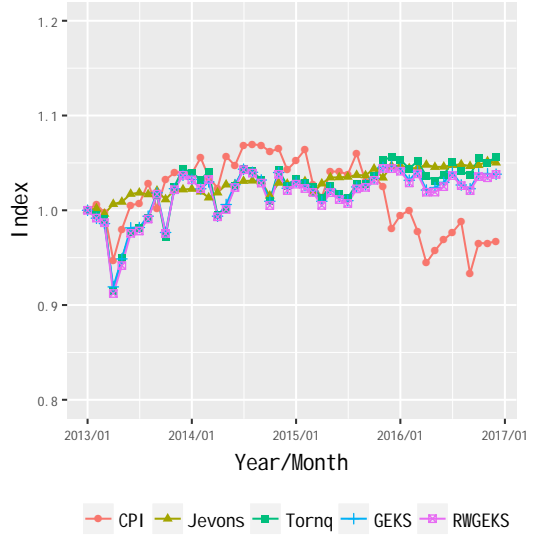


# A017270(단무지)

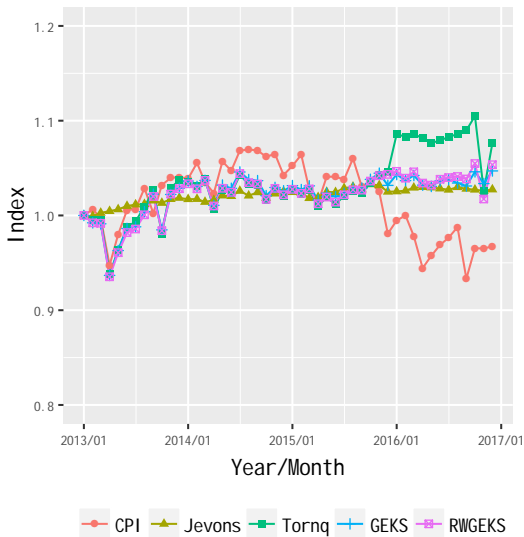
## 원데이터



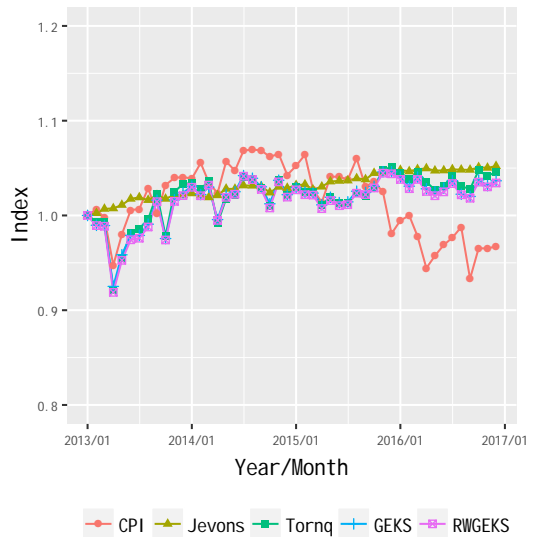
## 공통상품



## 특이항제거

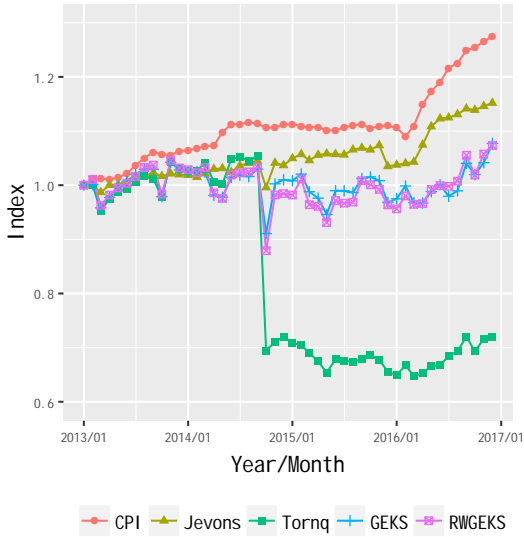


## 공통상품, 특이항제거

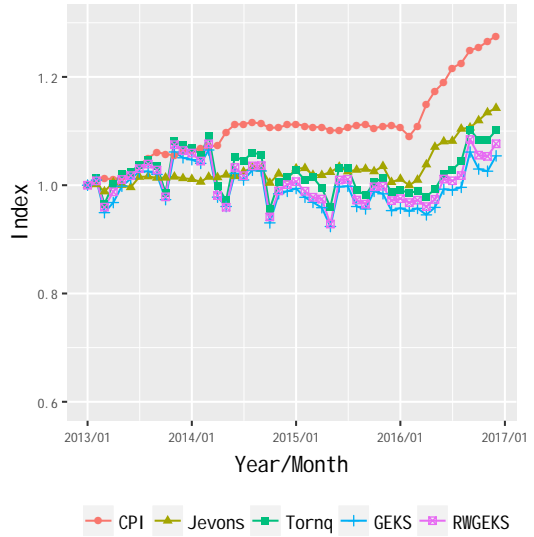


# A017280(김)

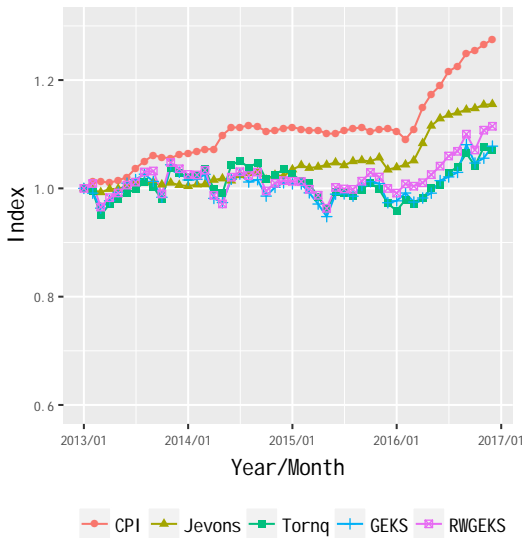
## 원데이터



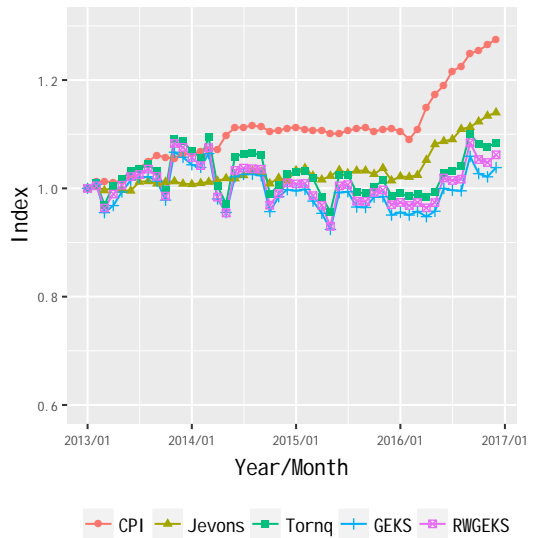
## 공통상품



## 특이항제거

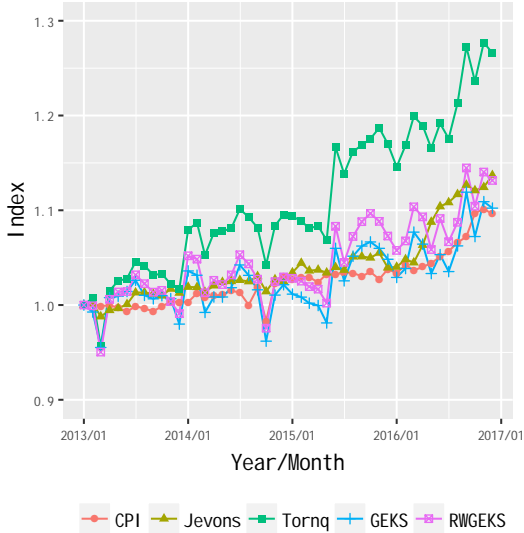


## 공통상품, 특이항제거

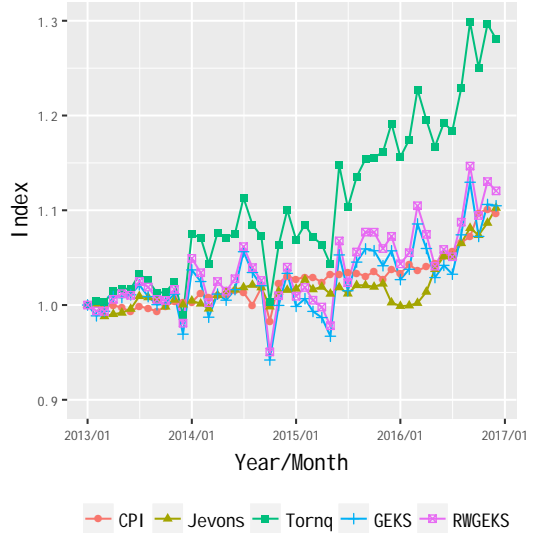


# A017290(맛김)

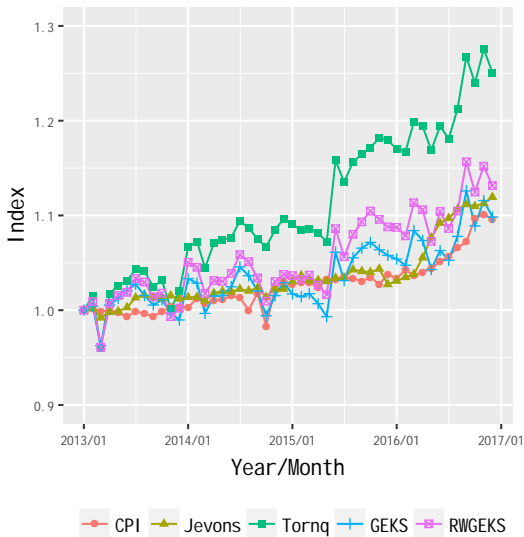
## 원데이터



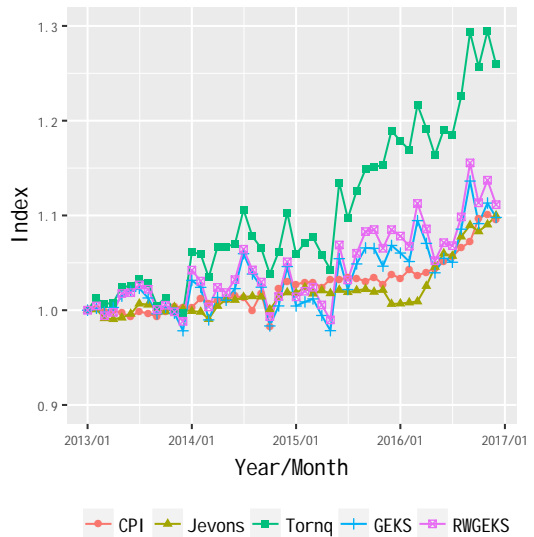
## 공통상품



## 특이항제거

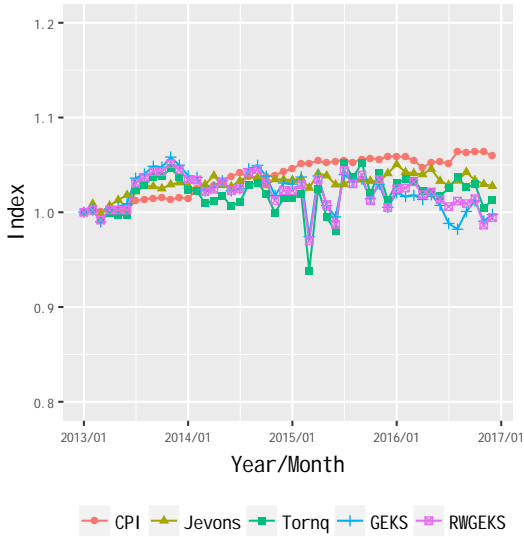


## 공통상품, 특이항제거

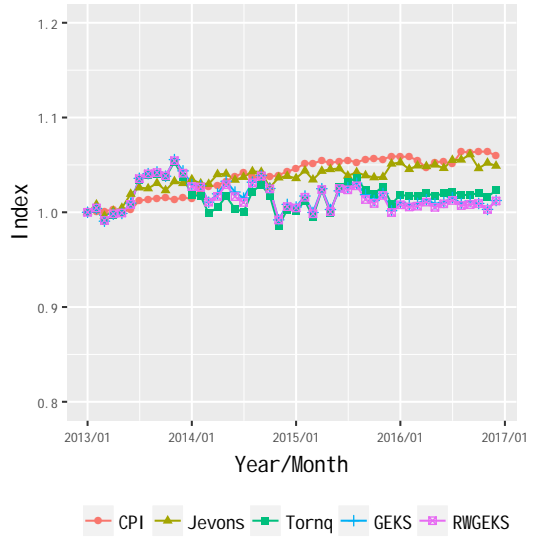


# A017300(미역)

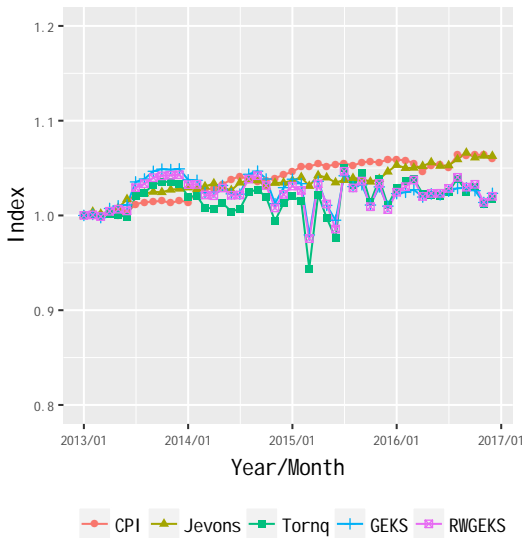
## 원데이터



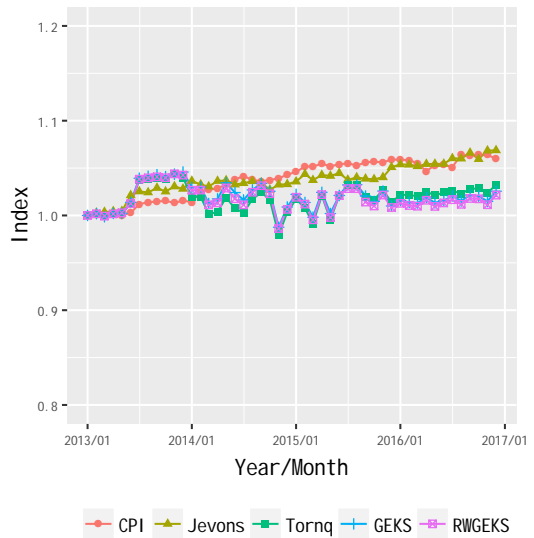
## 공통상품



## 특이항제거

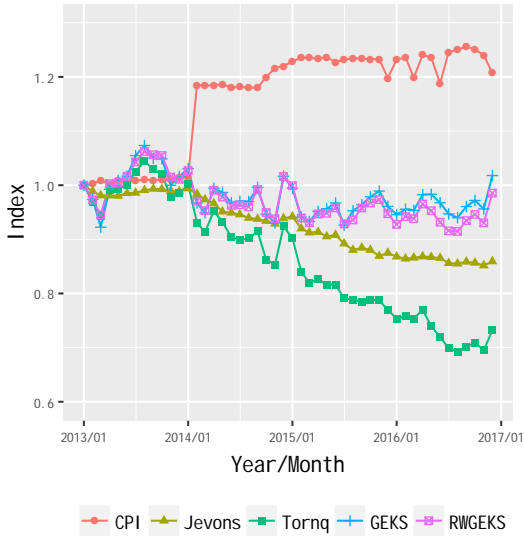


## 공통상품, 특이항제거

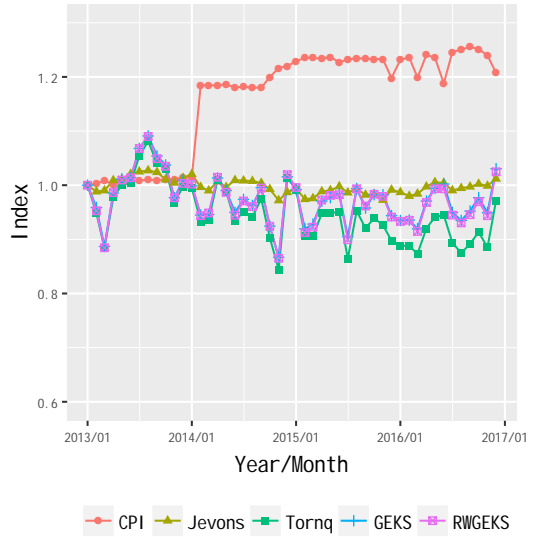


# A018010(초콜릿)

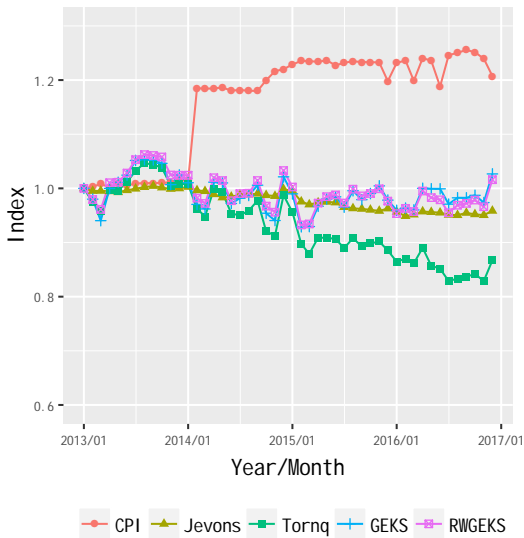
## 원데이터



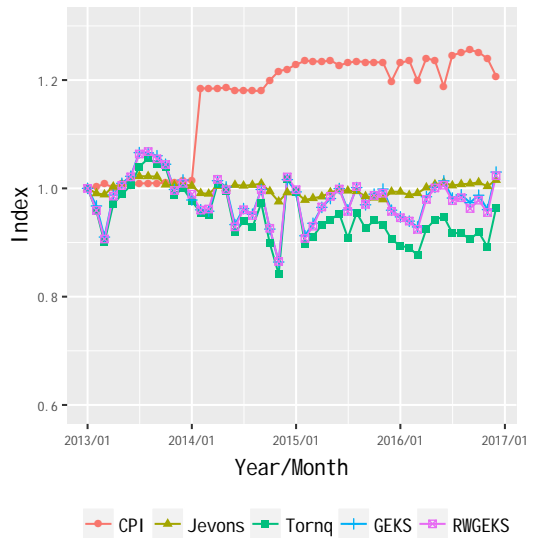
## 공통상품



## 특이항제거

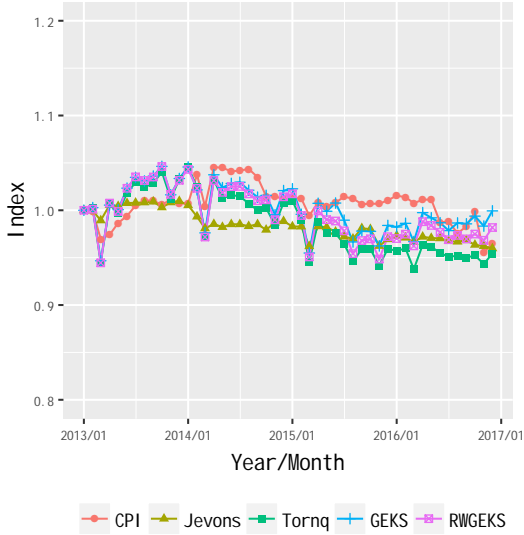


## 공통상품, 특이항제거

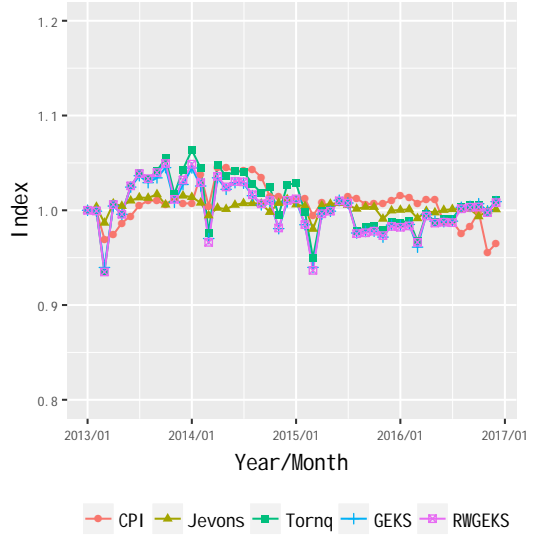


# A018020(사탕)

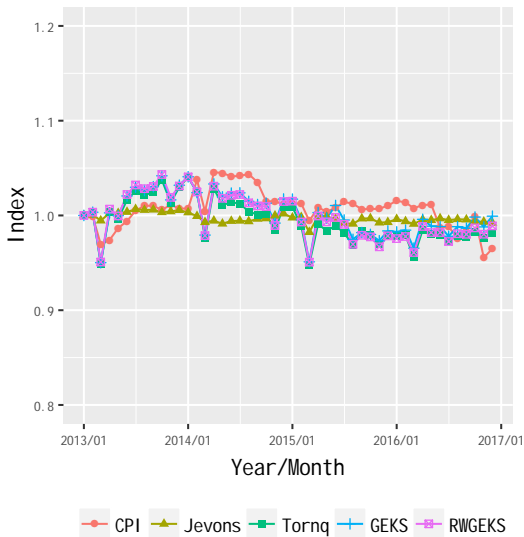
## 원데이터



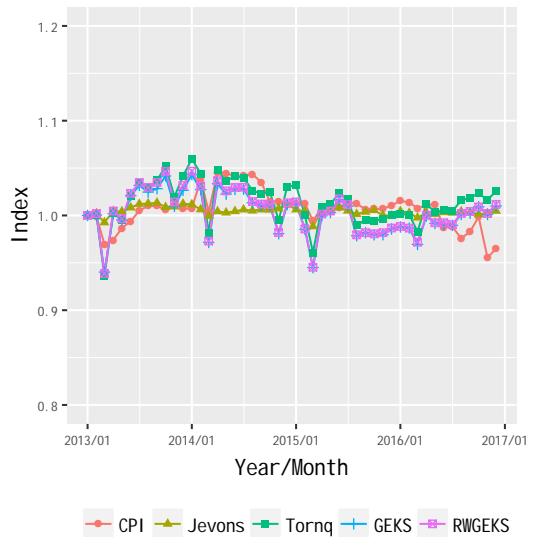
## 공통상품



## 특이항제거



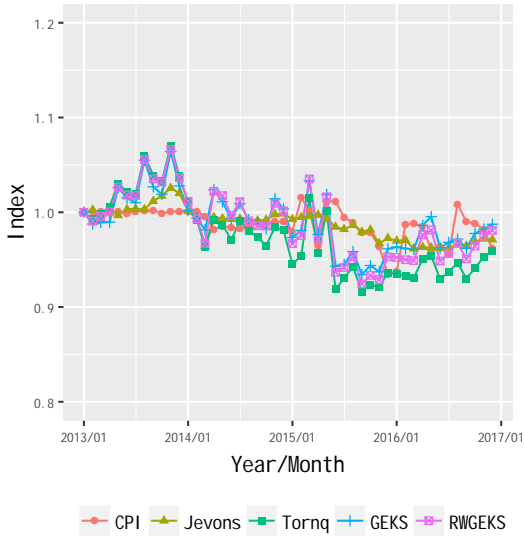
## 공통상품, 특이항제거



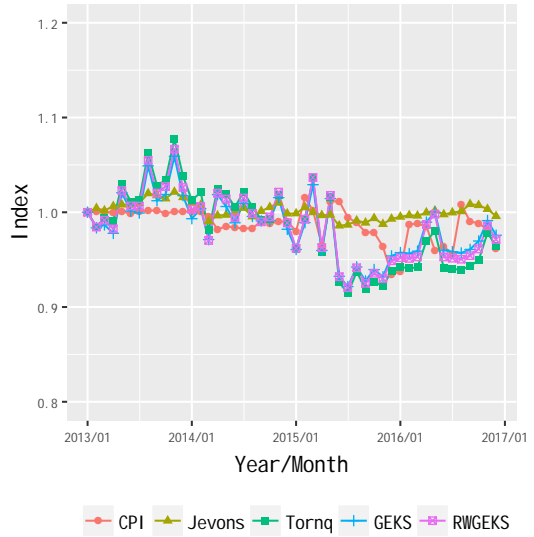


# A018030(کم)

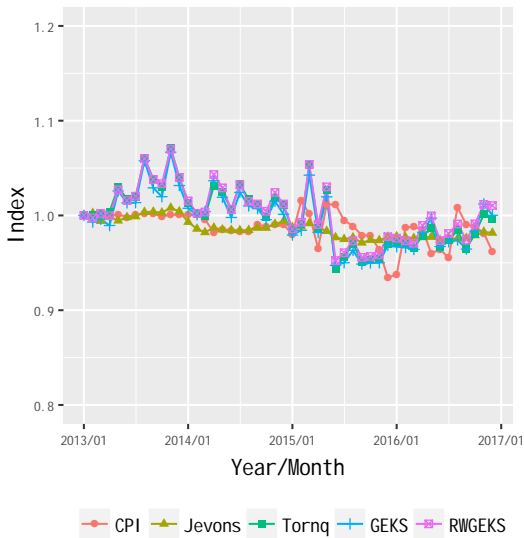
## 원데이터



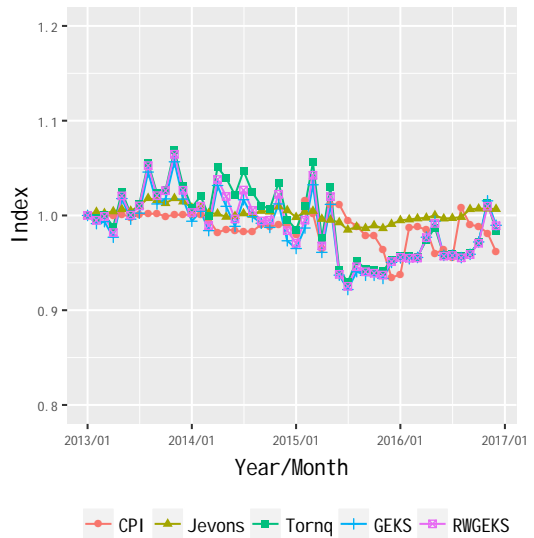
## 공통상품



## 특이항제거

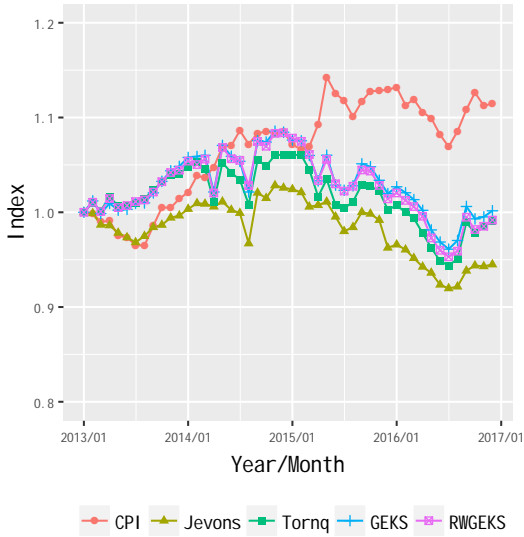


## 공통상품, 특이항제거

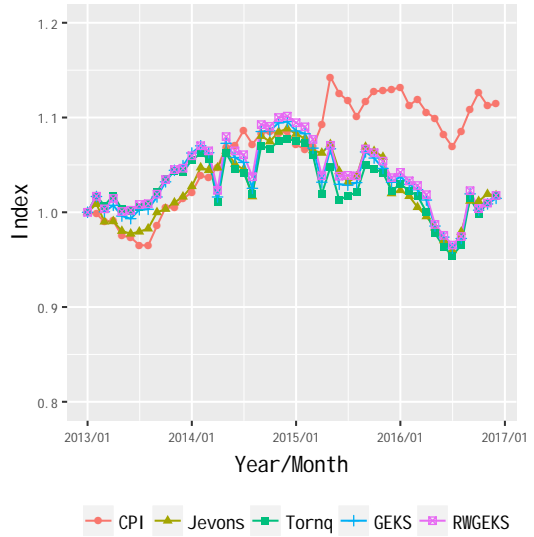


# A018040(아이스크림)

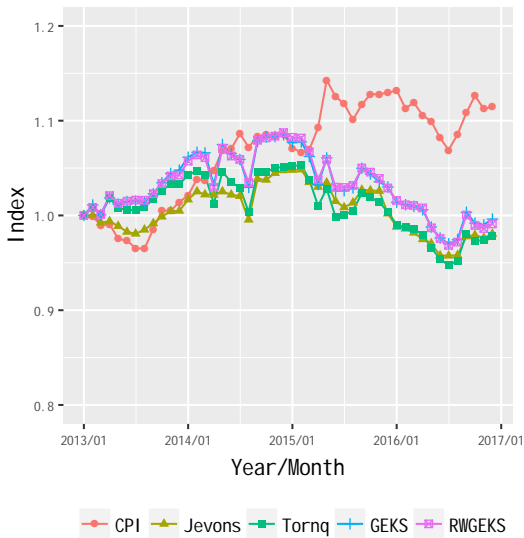
## 원데이터



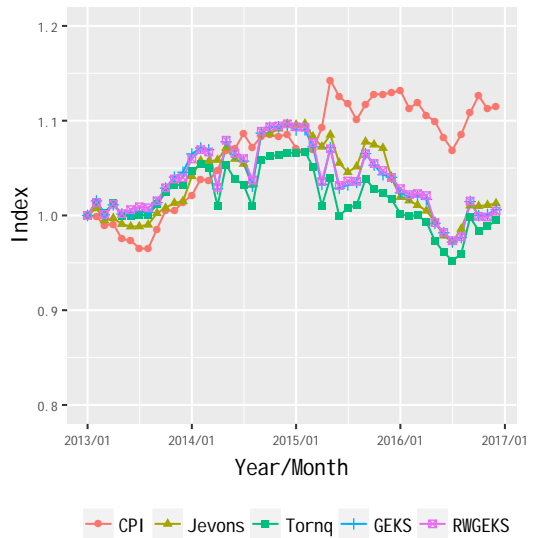
## 공통상품



## 특이항제거

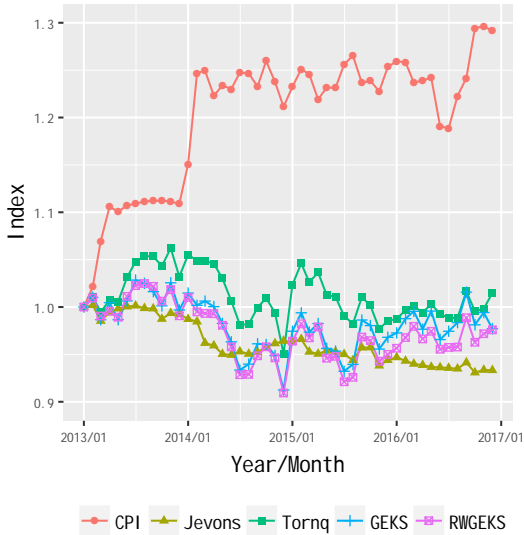


## 공통상품, 특이항제거

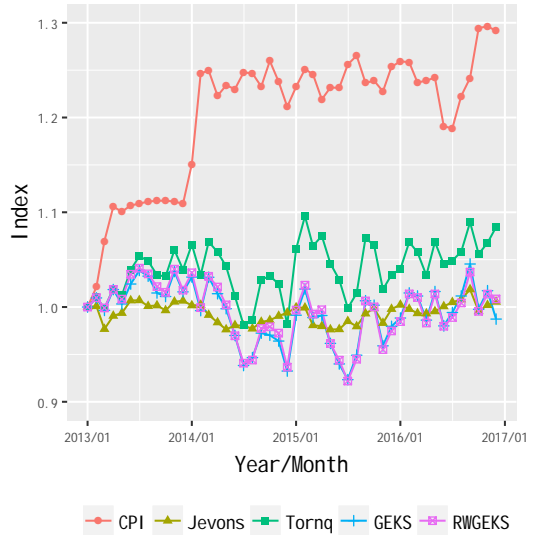


# A018050(비스킷)

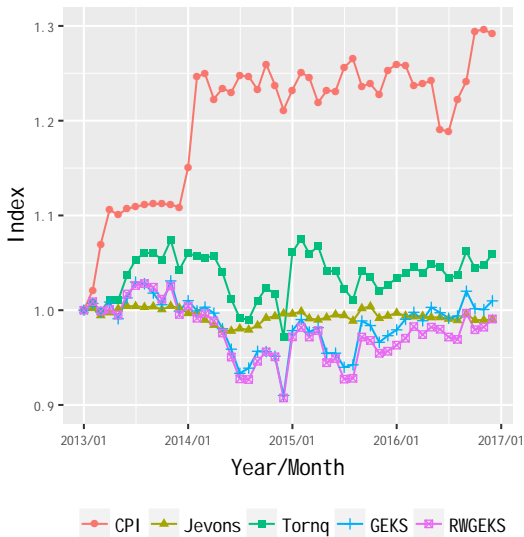
## 원데이터



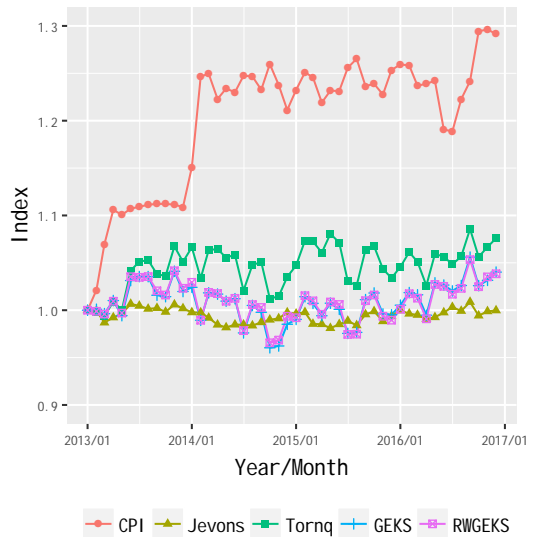
## 공통상품



## 특이항제거

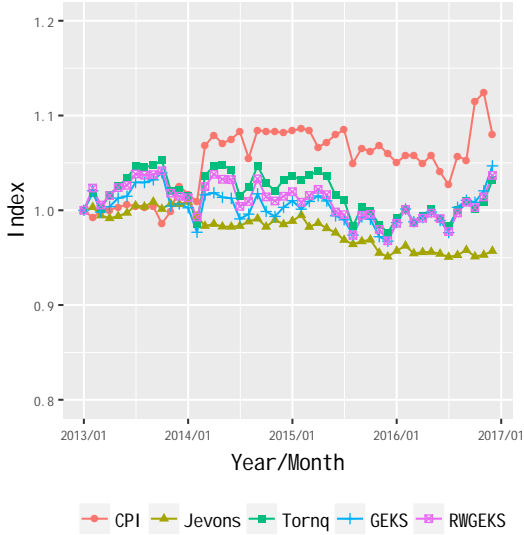


## 공통상품, 특이항제거

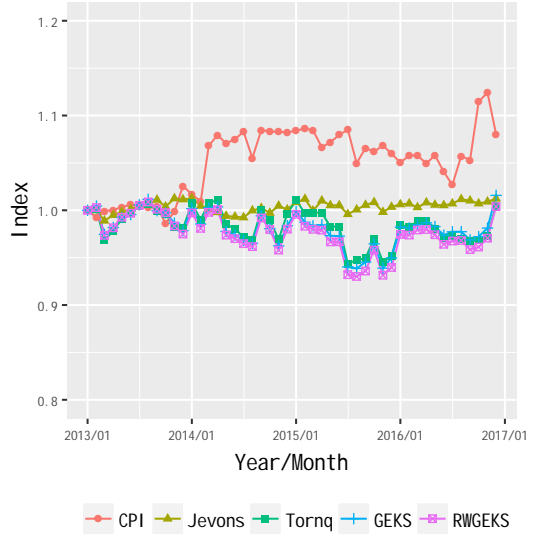


# A018060(스낵 과자)

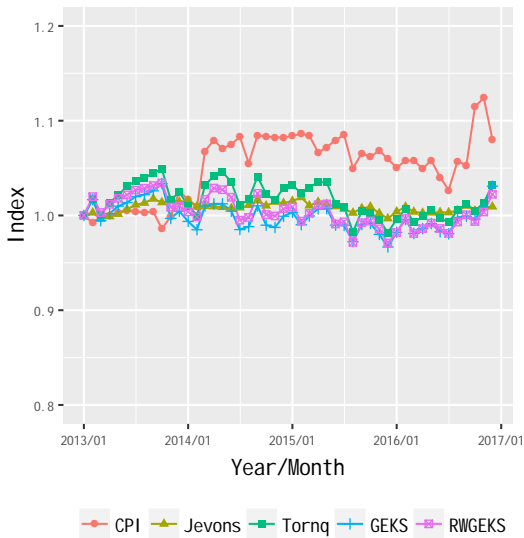
## 원데이터



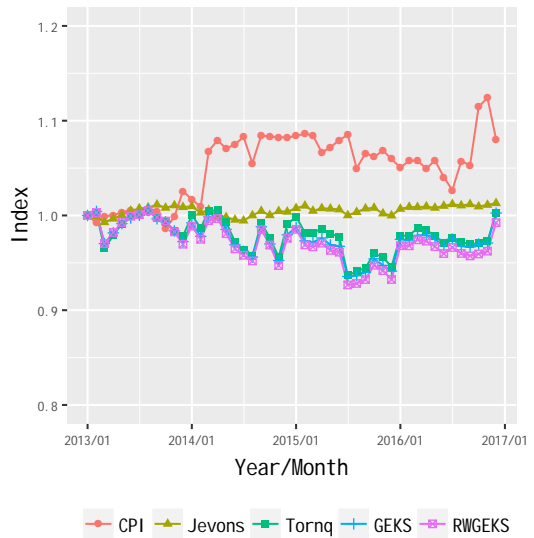
## 공통상품



## 특이항제거

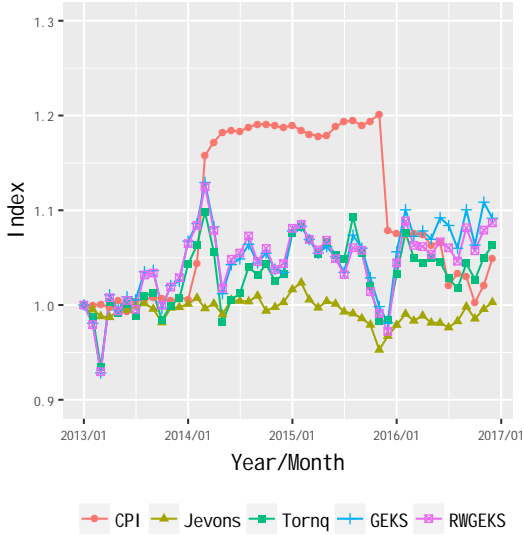


## 공통상품, 특이항제거

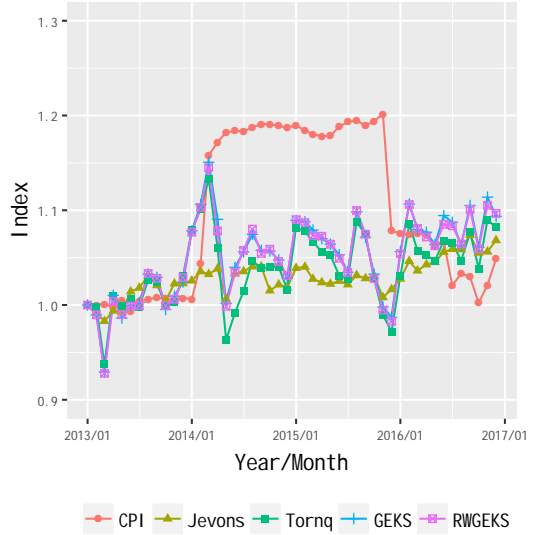


# A018070(파이)

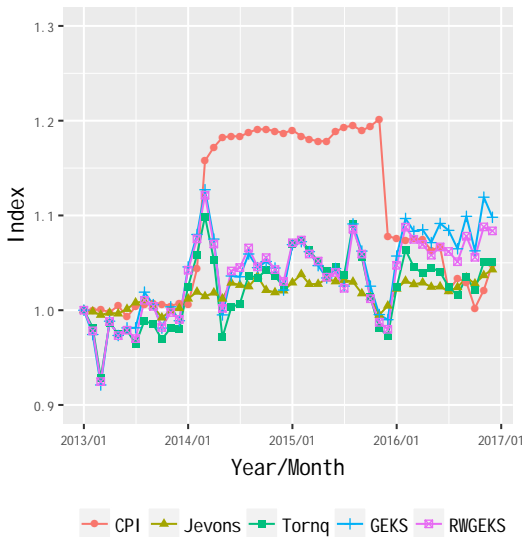
## 원데이터



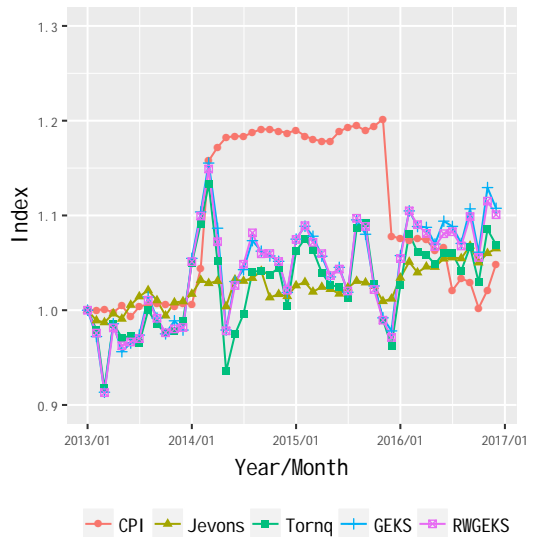
## 공통상품



## 특이항제거

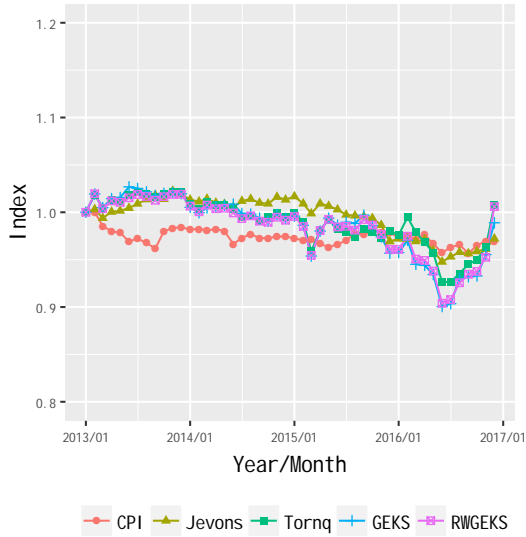


## 공통상품, 특이항제거

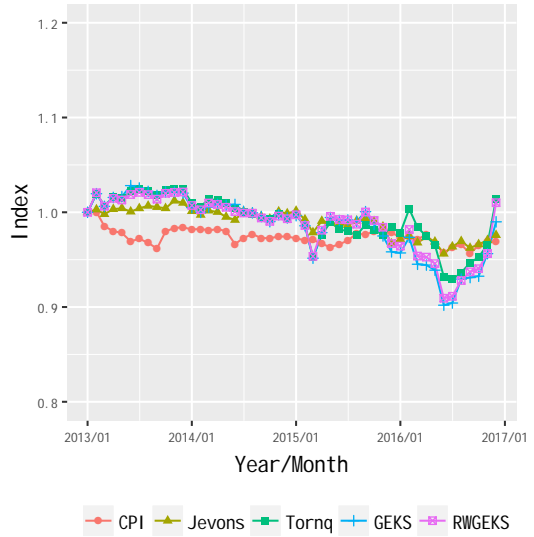


# A018080(설탕)

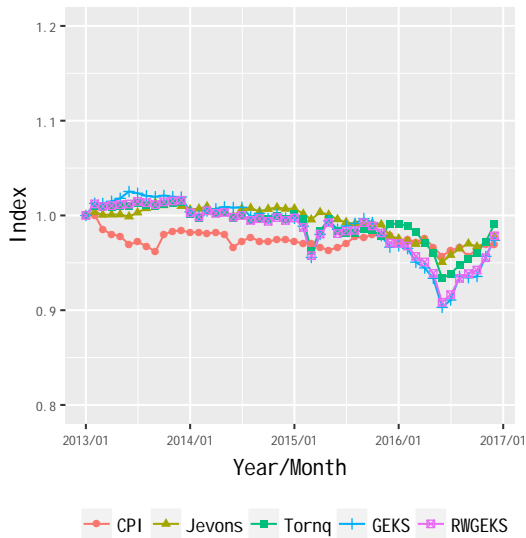
## 원데이터



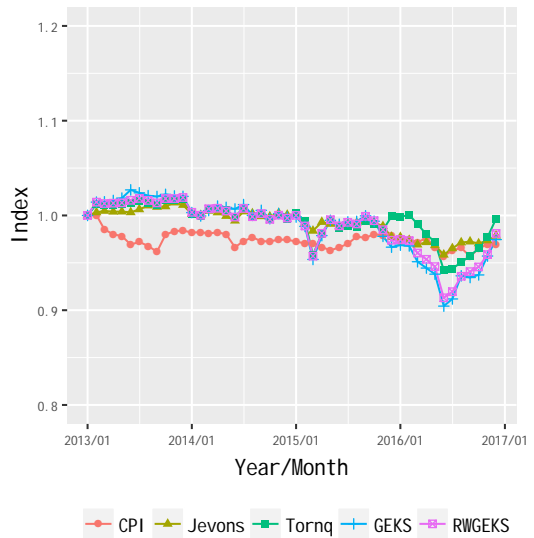
## 공통상품



## 특이항제거

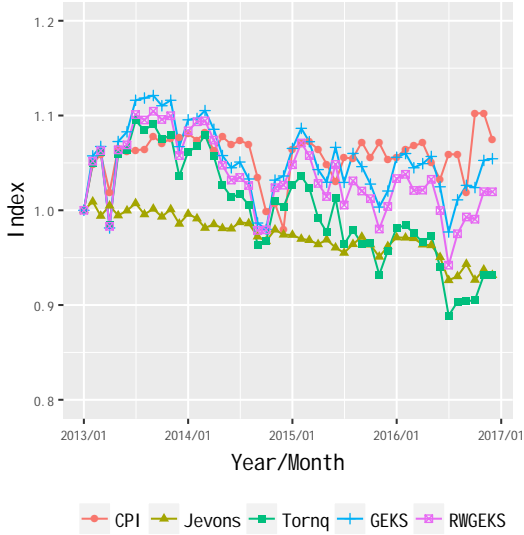


## 공통상품, 특이항제거

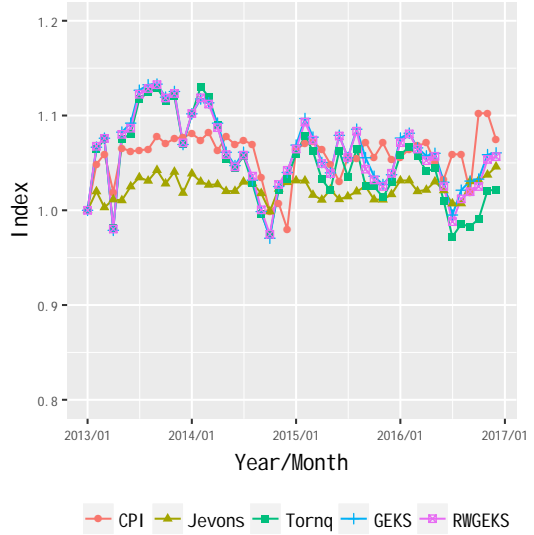


# A018090(잼)

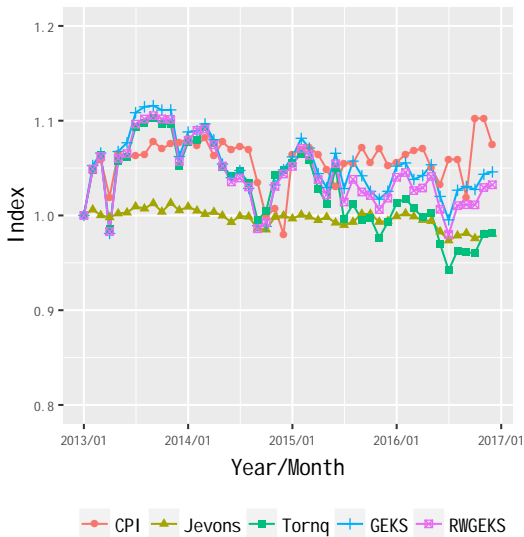
## 원데이터



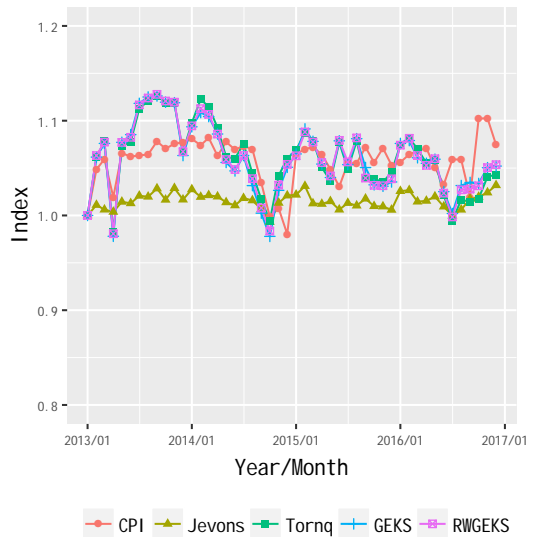
## 공통상품



## 특이항제거

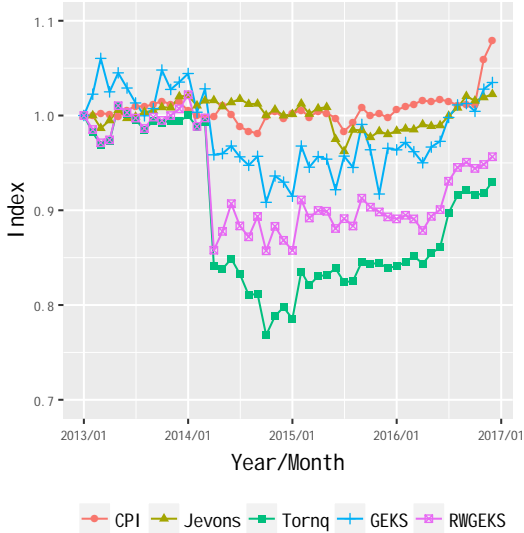


## 공통상품, 특이항제거

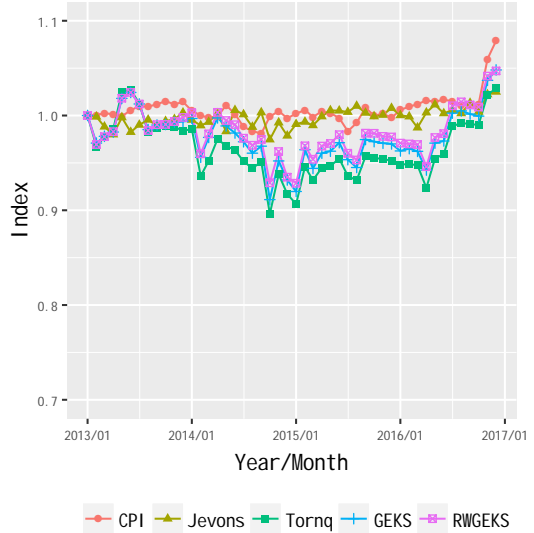


# A018100(꿀)

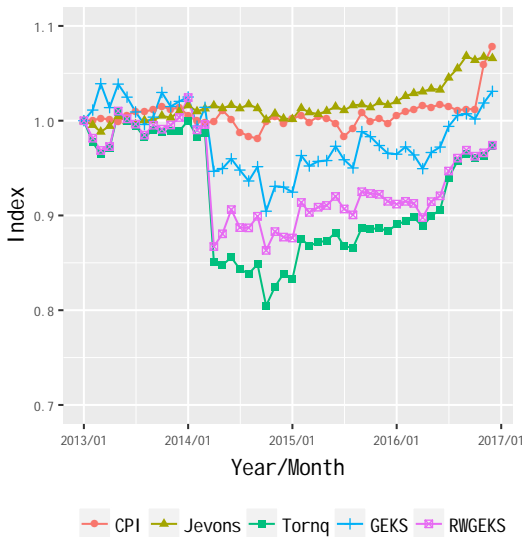
## 원데이터



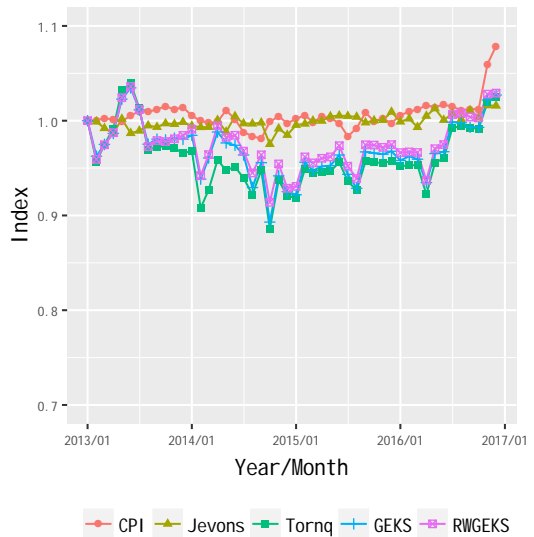
## 공통상품



## 특이항제거



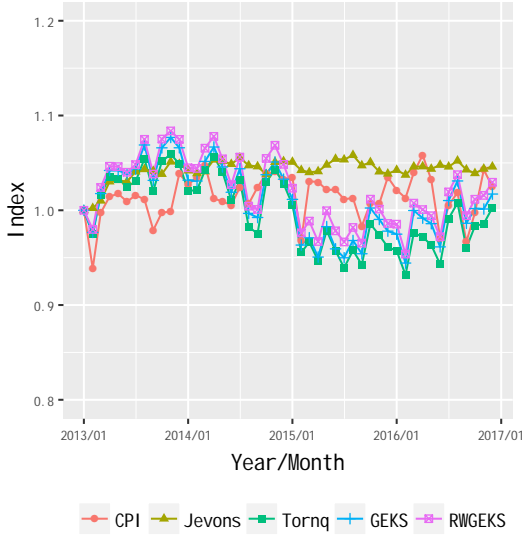
## 공통상품, 특이항제거



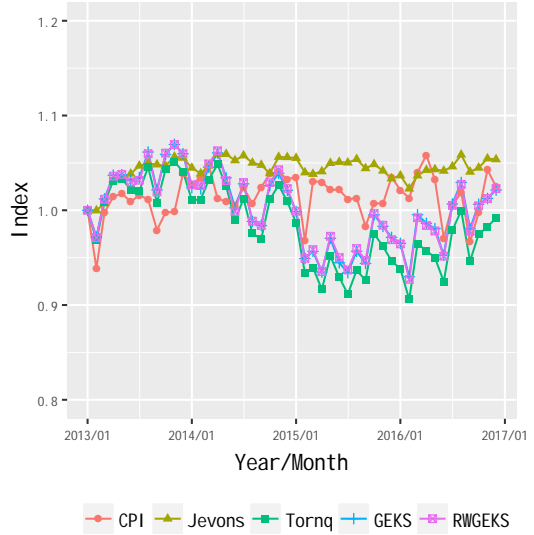


# A018110(물엿)

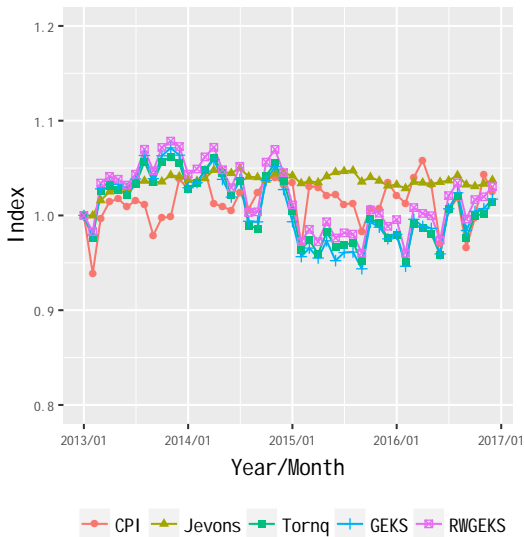
## 원데이터



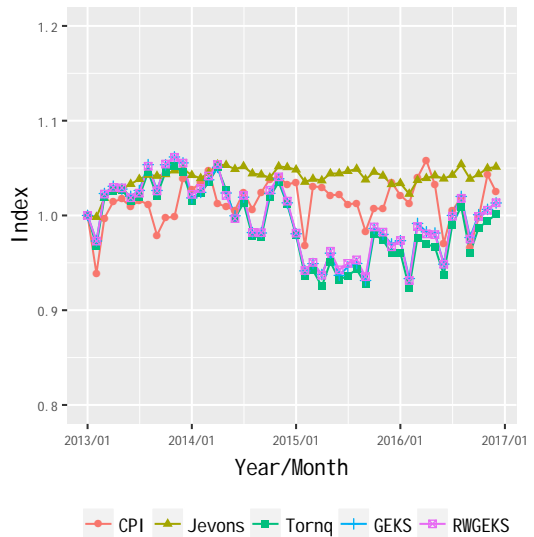
## 공통상품



## 특이항제거

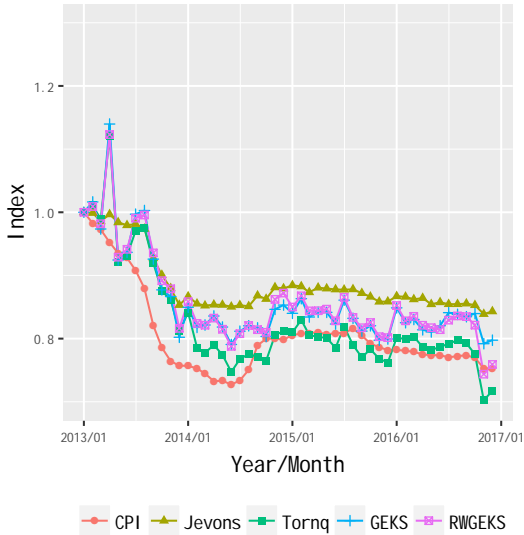


## 공통상품, 특이항제거

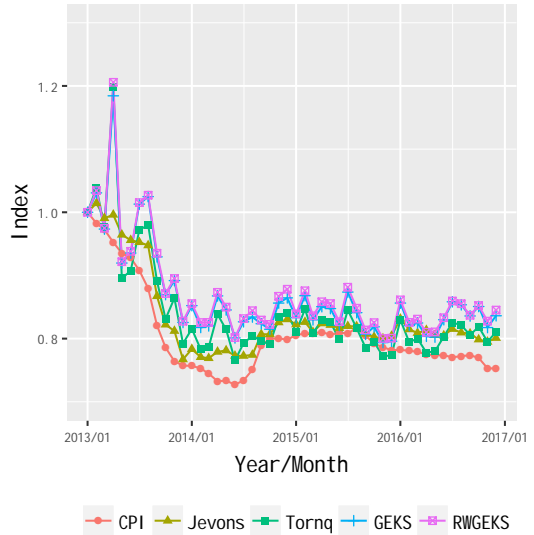


# A019010(고춧가루)

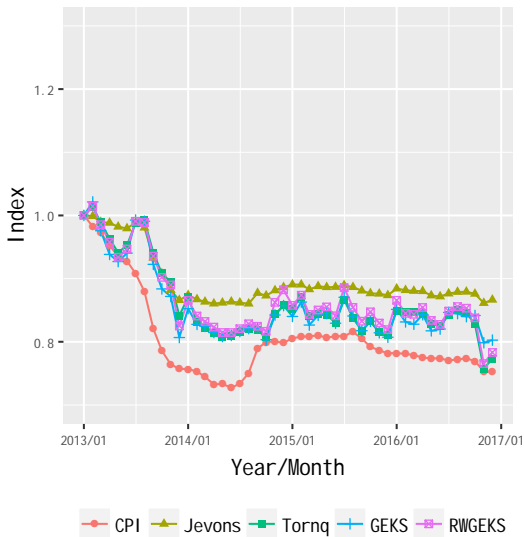
## 원데이터



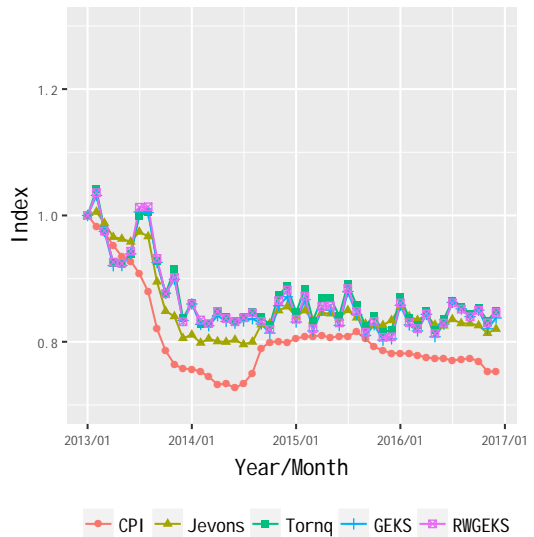
## 공통상품



## 특이항제거

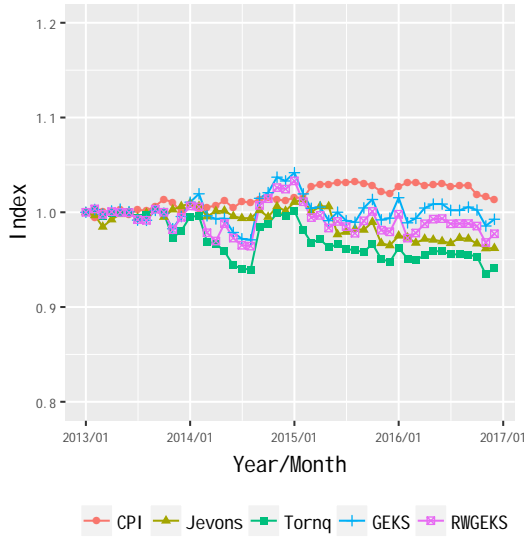


## 공통상품, 특이항제거

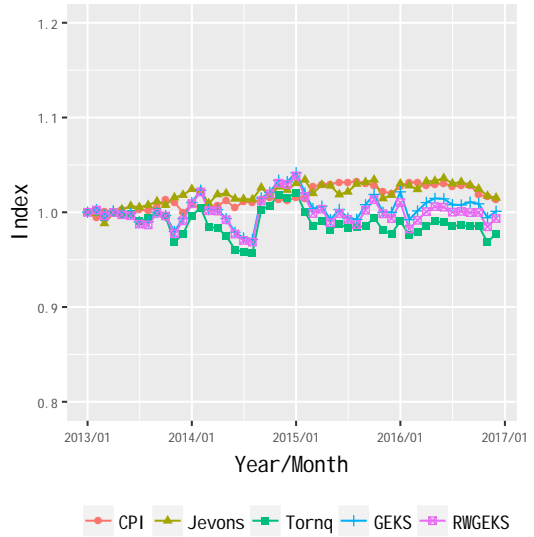


# A019040(소금)

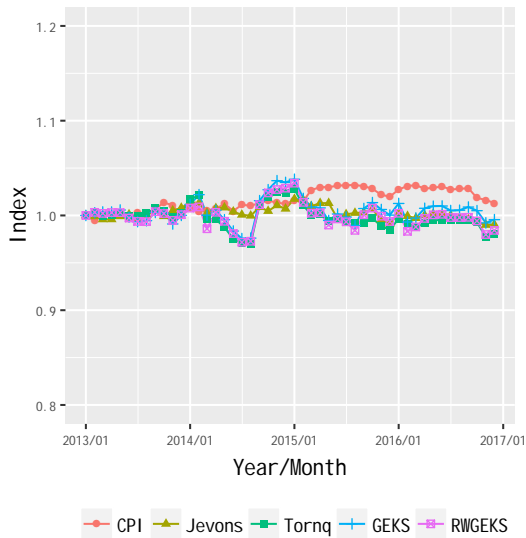
## 원데이터



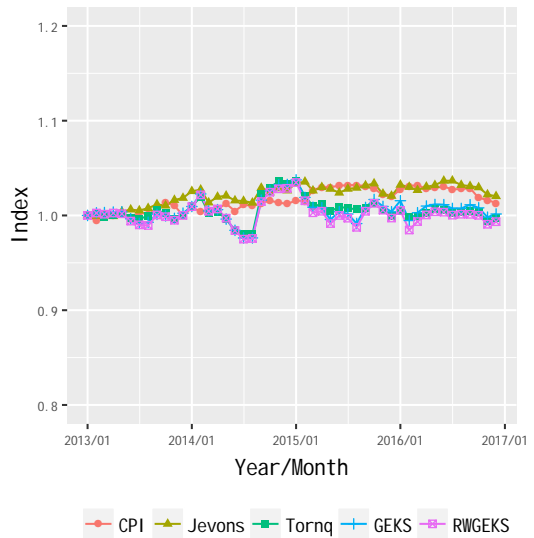
## 공통상품



## 특이항제거

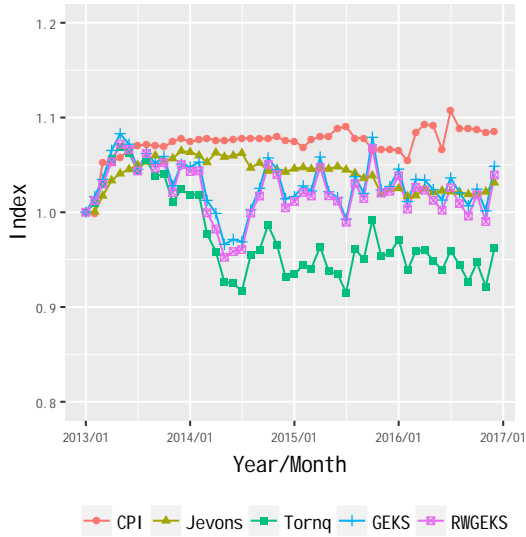


## 공통상품, 특이항제거

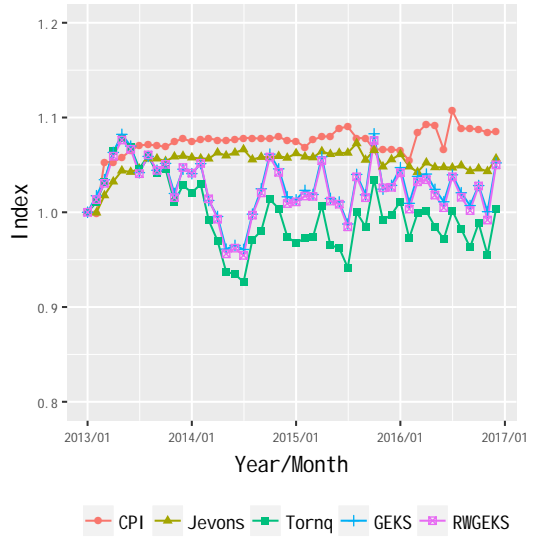


# A019050(간장)

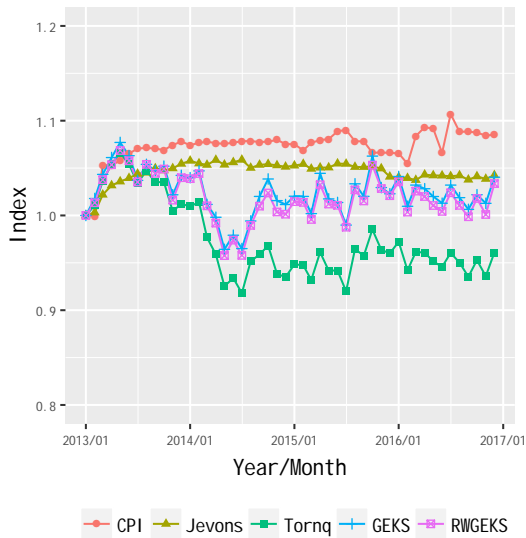
## 원데이터



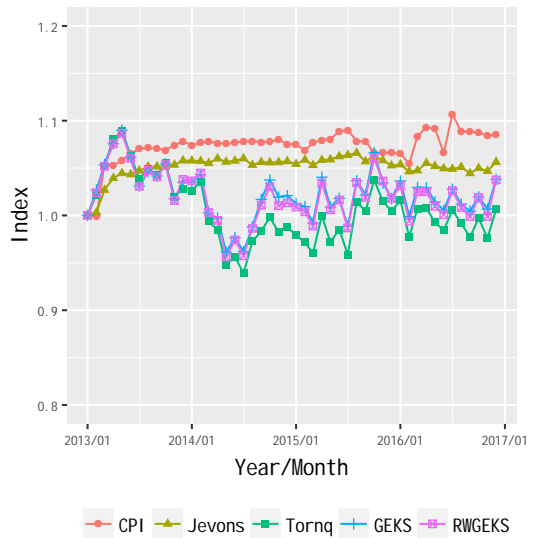
## 공통상품



## 특이항제거

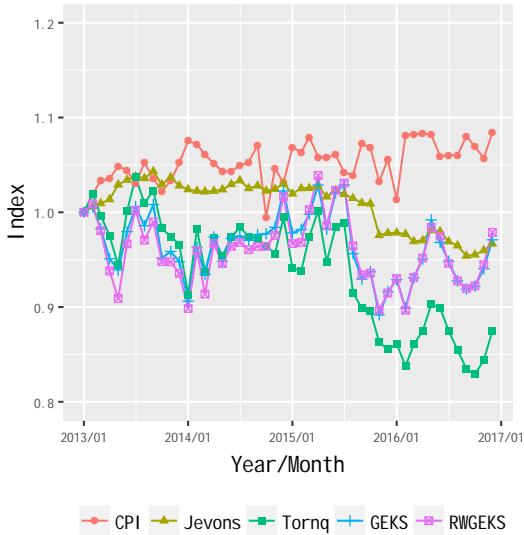


## 공통상품, 특이항제거

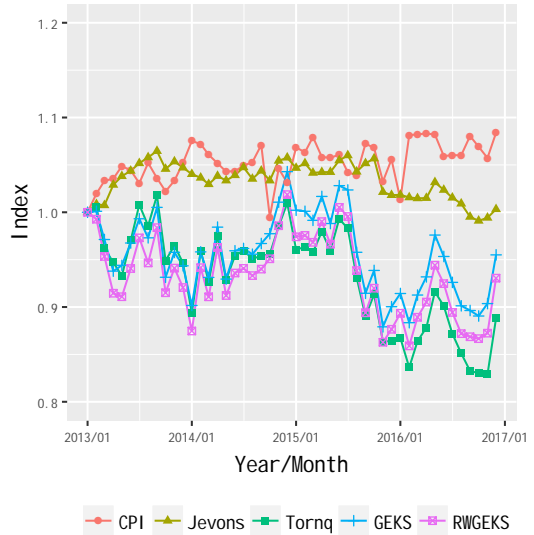


# A019060(된장)

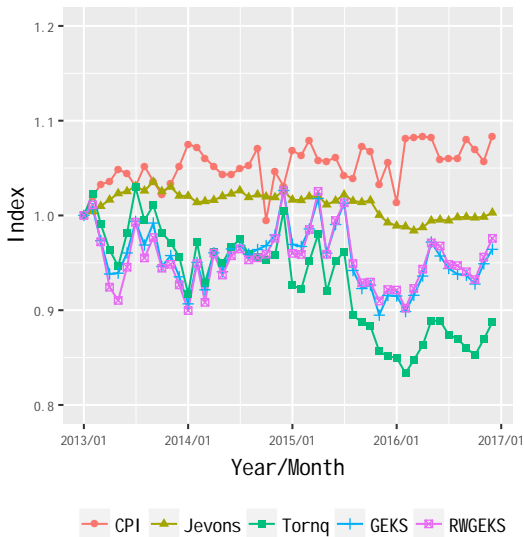
## 원데이터



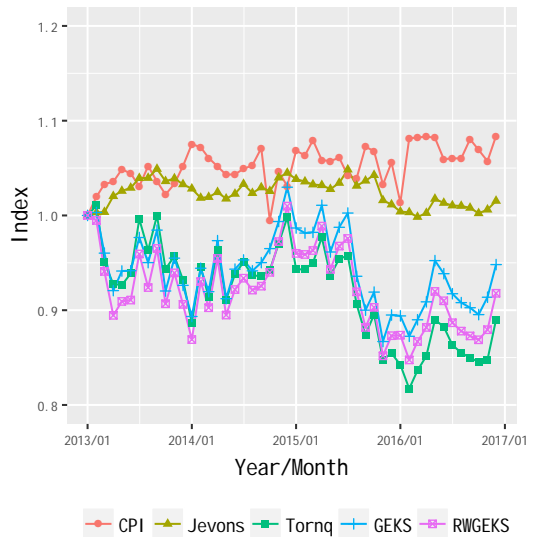
## 공통상품



## 특이항제거

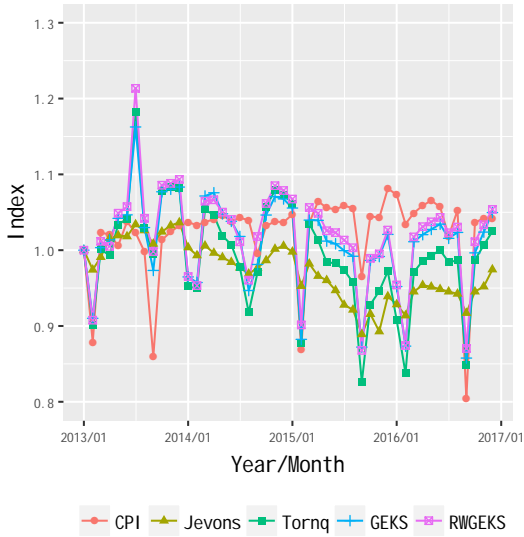


## 공통상품, 특이항제거

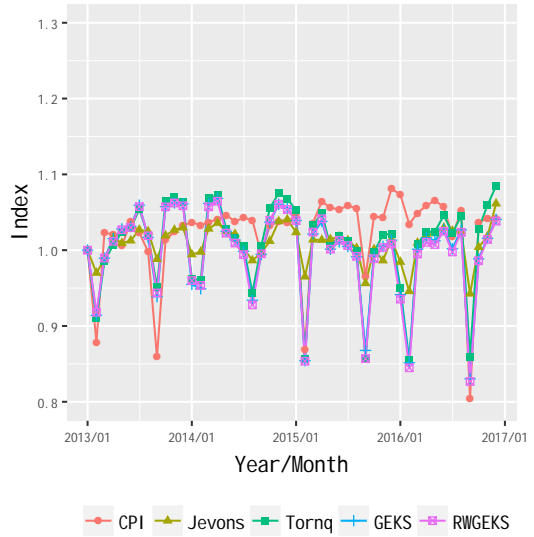


# A019070(양념소스)

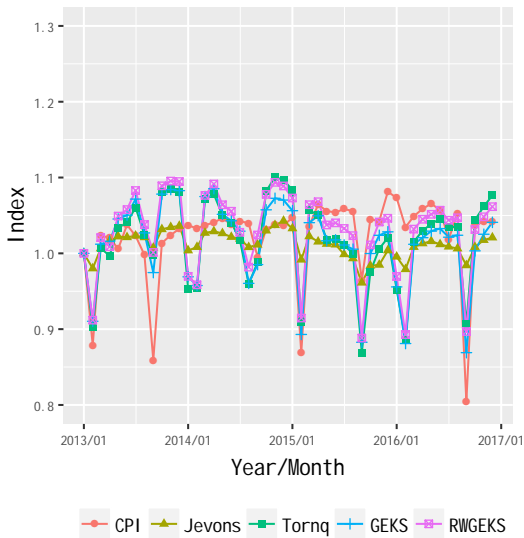
## 원데이터



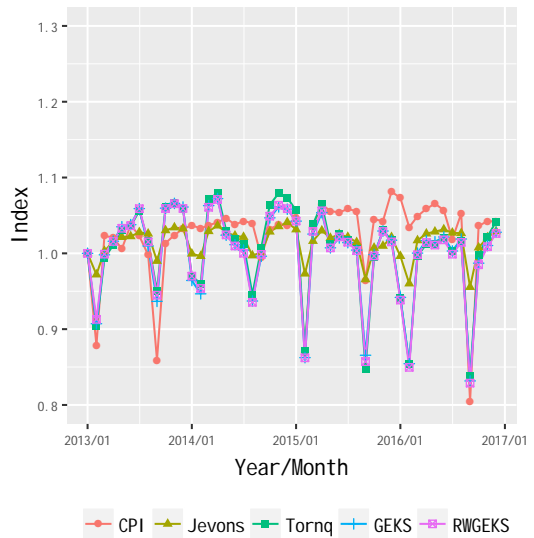
## 공통상품



## 특이항제거

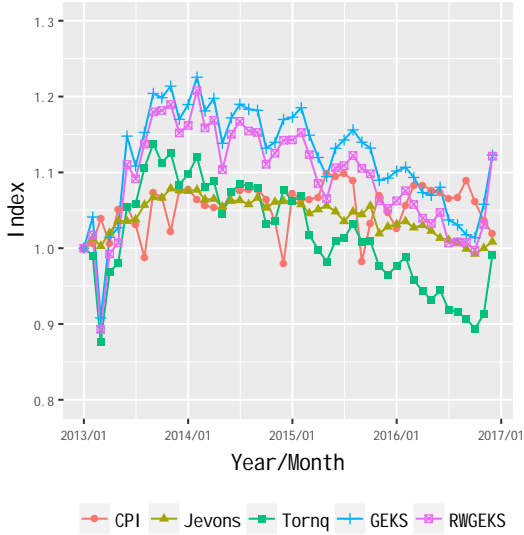


## 공통상품, 특이항제거

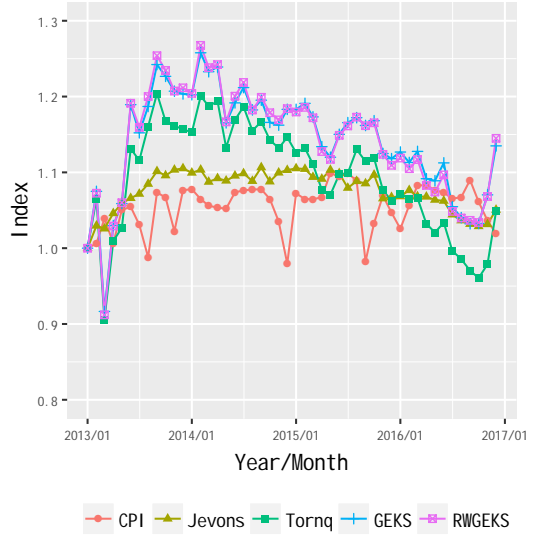


# A019080(고추장)

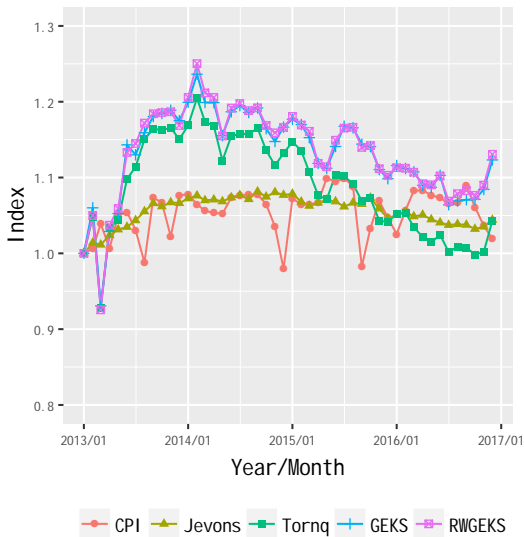
## 원데이터



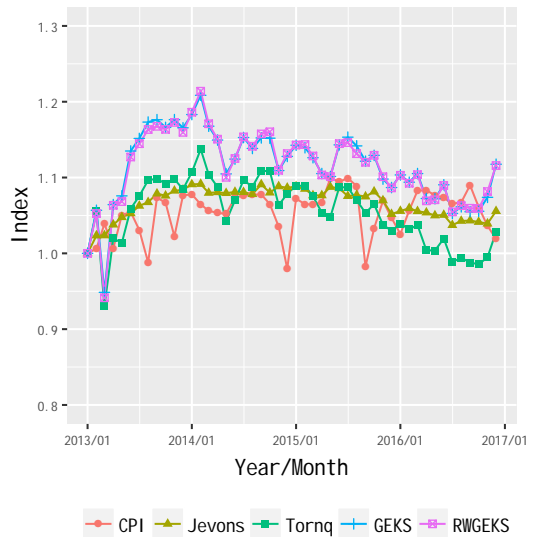
## 공통상품



## 특이항제거

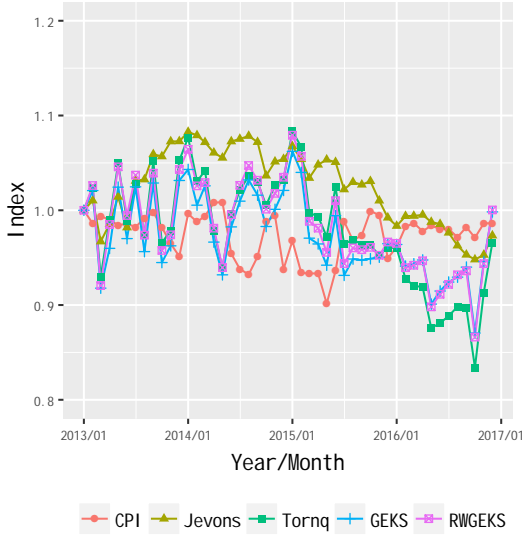


## 공통상품, 특이항제거

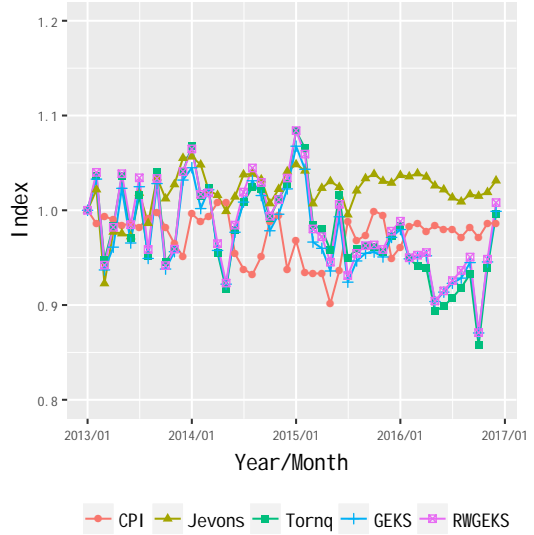


# A019090(카레)

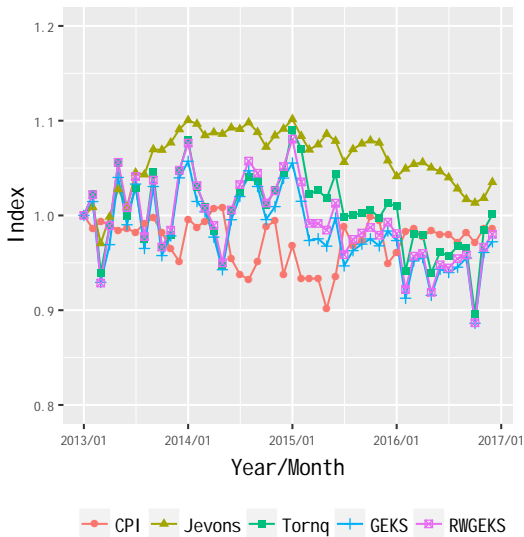
## 원데이터



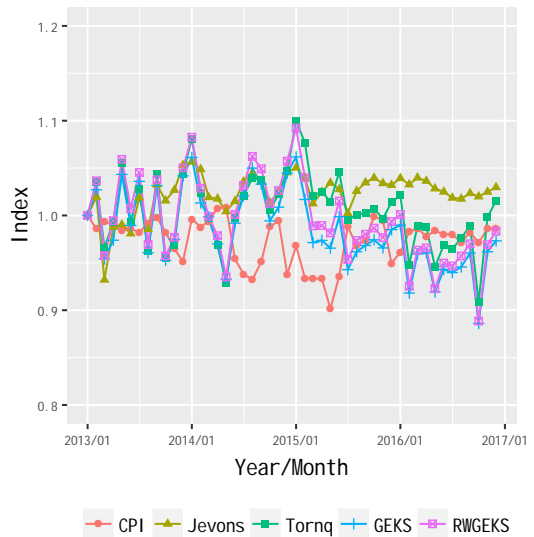
## 공통상품



## 특이항제거



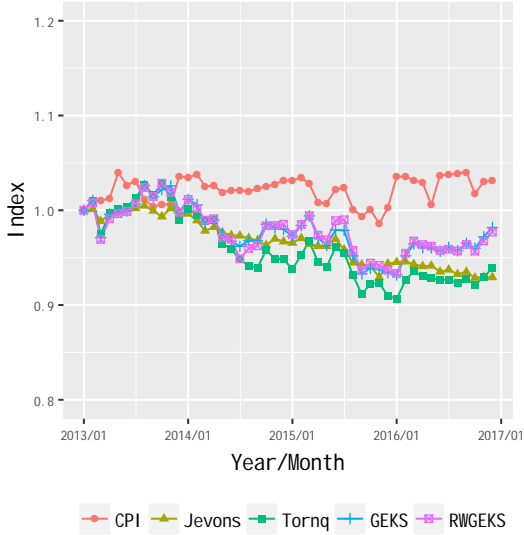
## 공통상품, 특이항제거



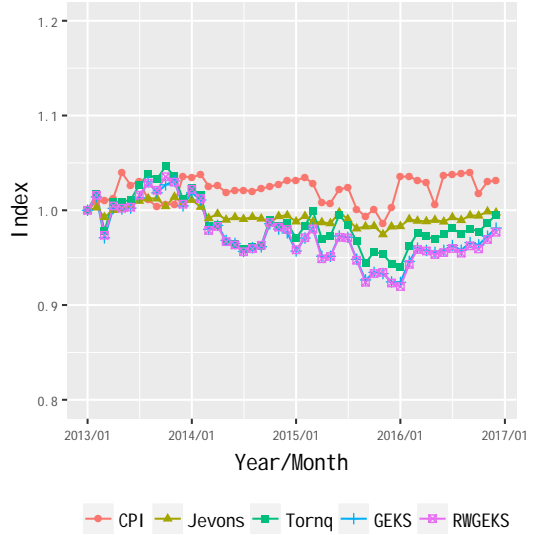


# A019110(드레싱)

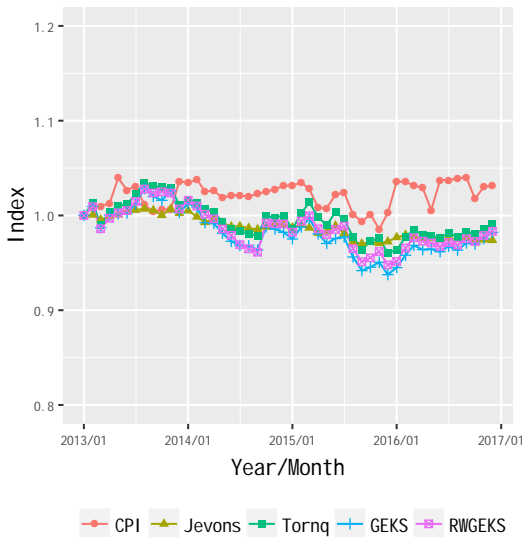
## 원데이터



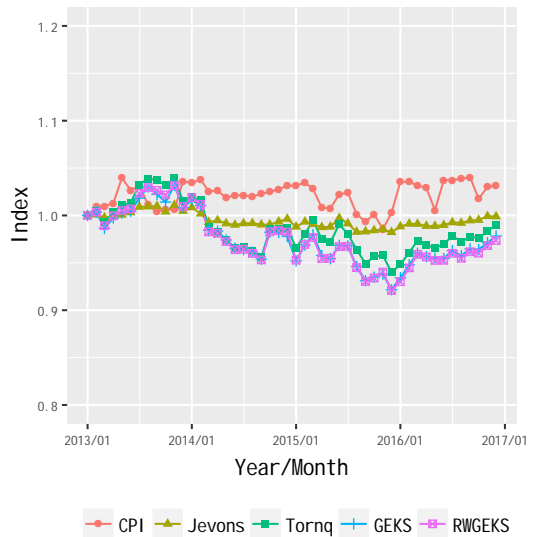
## 공통상품



## 특이항제거

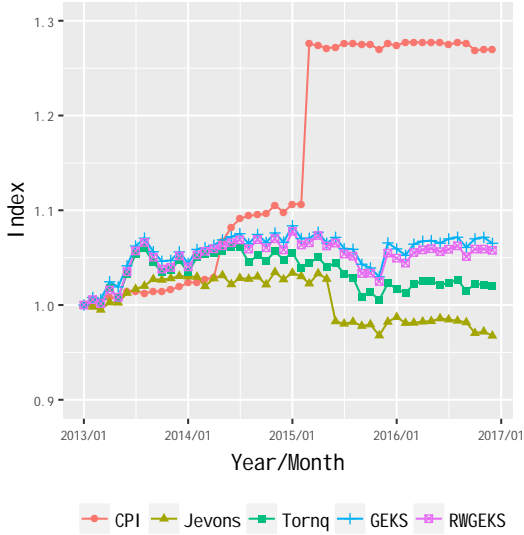


## 공통상품, 특이항제거

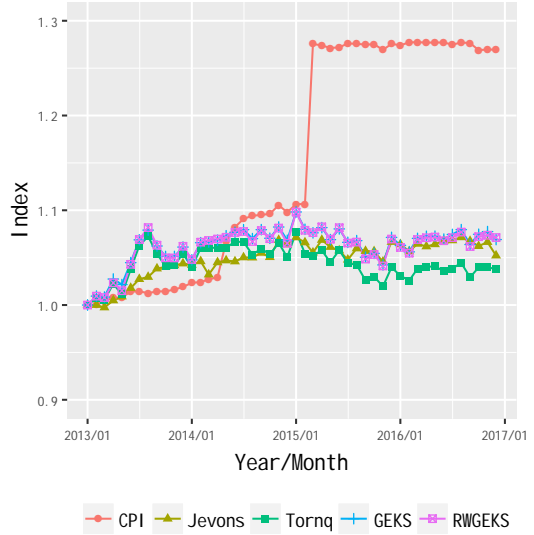


# A019120(혼합조미료)

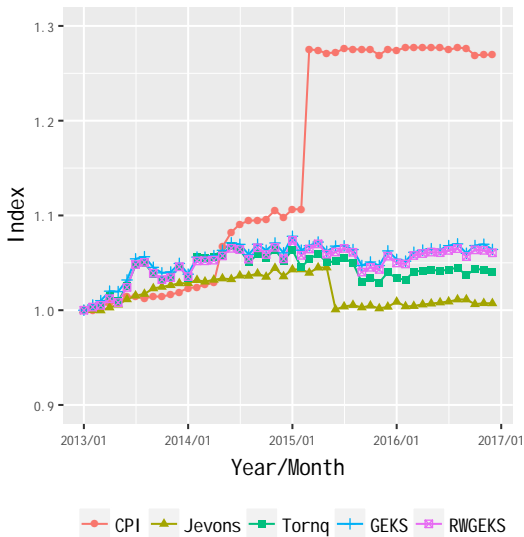
## 원데이터



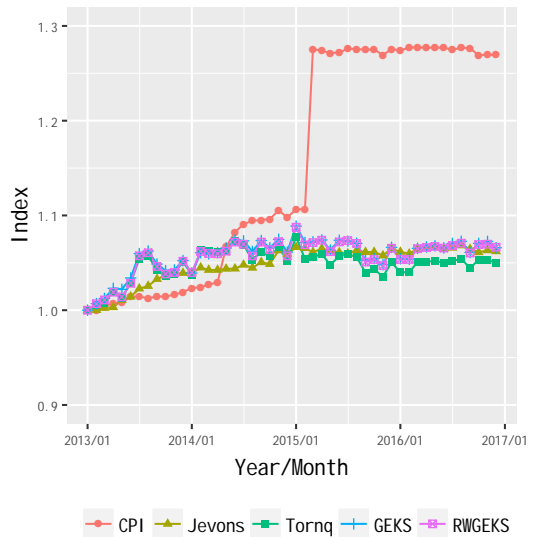
## 공통상품



## 특이항제거

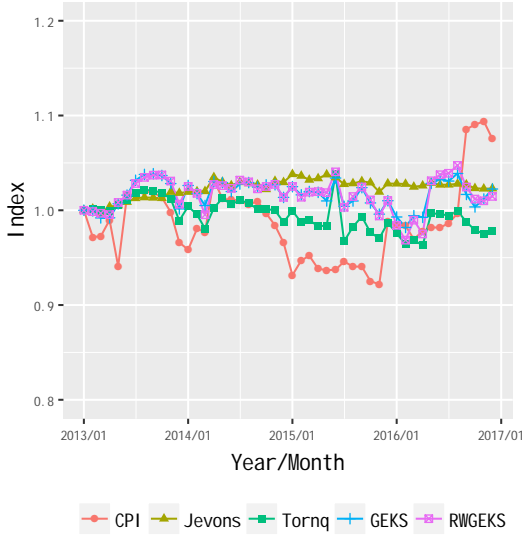


## 공통상품, 특이항제거

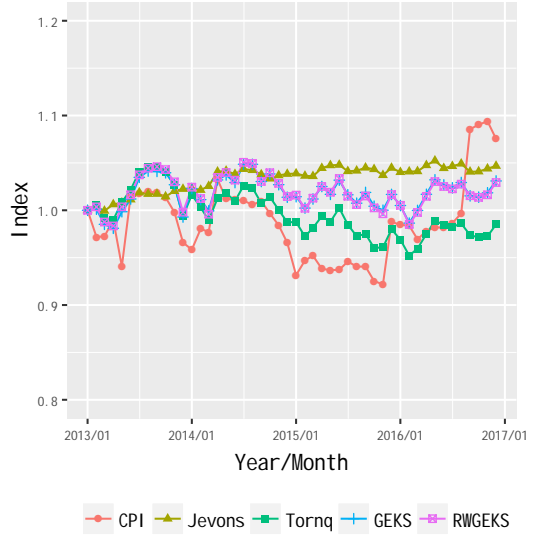


# A019130(스프)

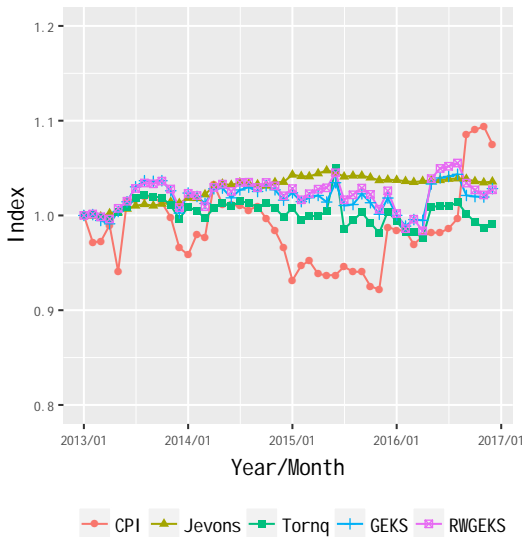
## 원데이터



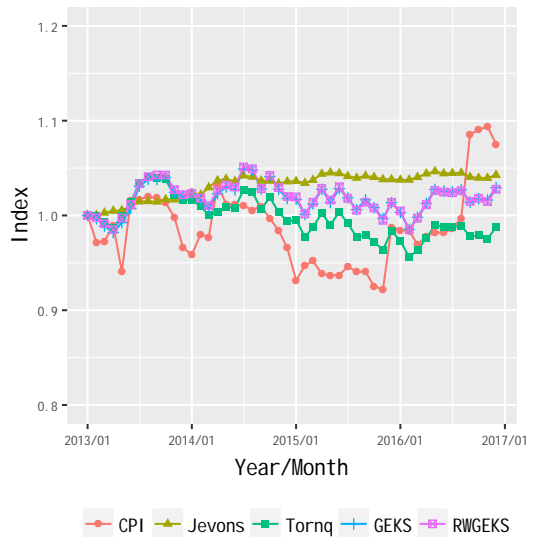
## 공통상품



## 특이항제거

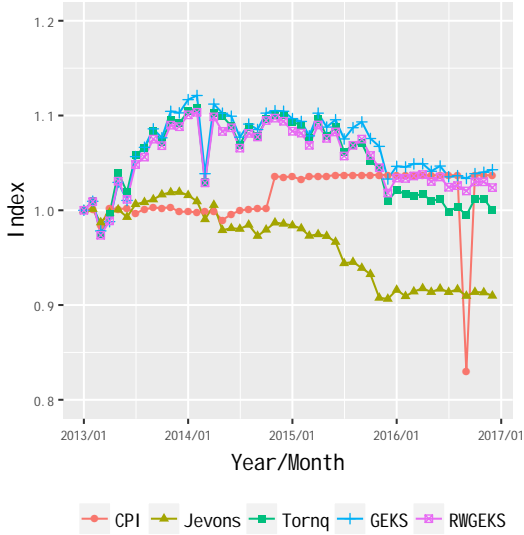


## 공통상품, 특이항제거

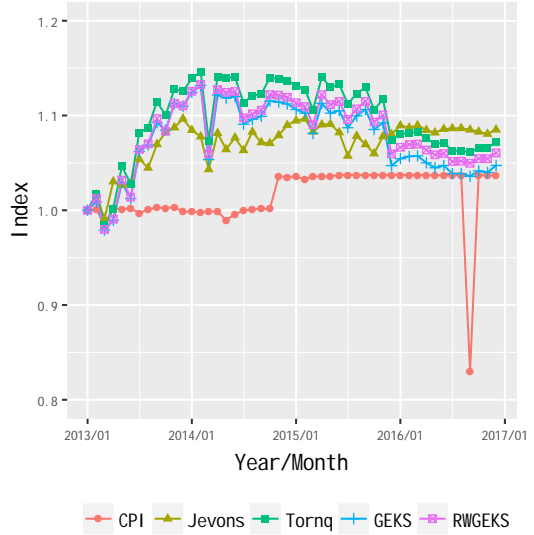


# A019140(이유식)

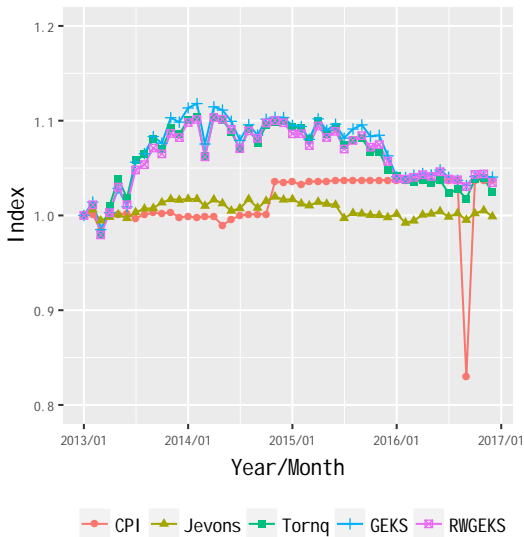
## 원데이터



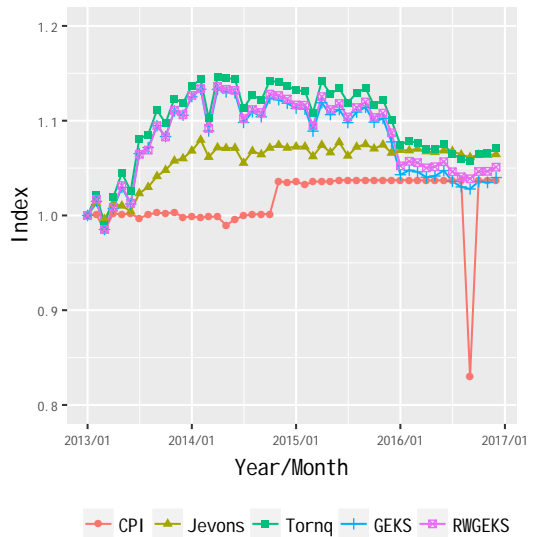
## 공통상품



## 특이항제거

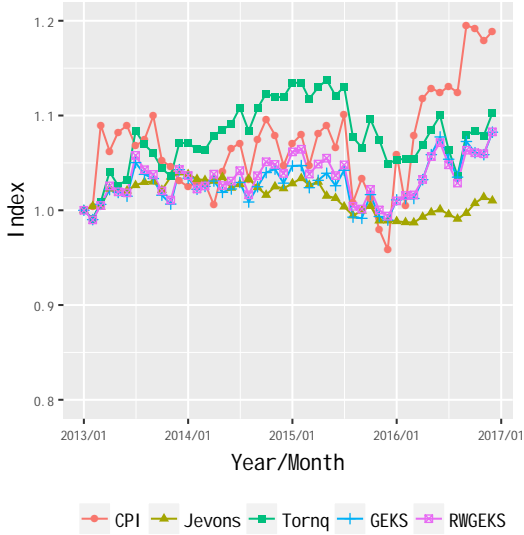


## 공통상품, 특이항제거

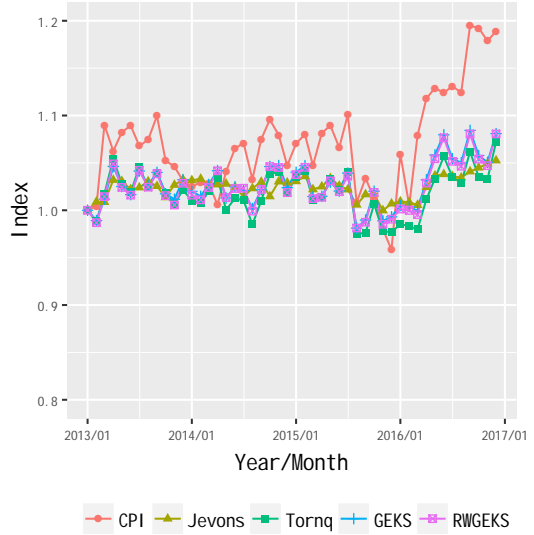


# A019150(김치)

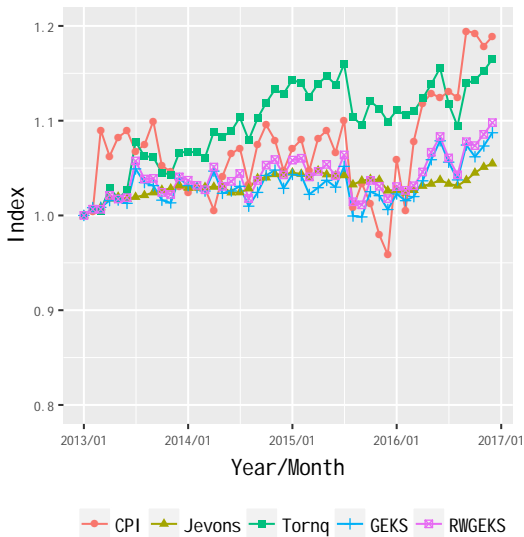
## 원데이터



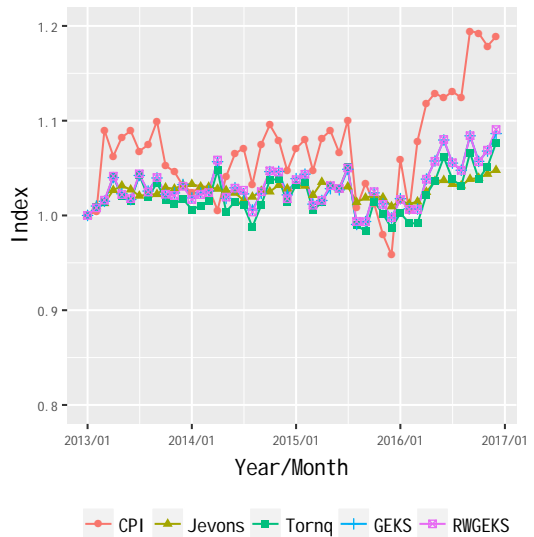
## 공통상품



## 특이항제거

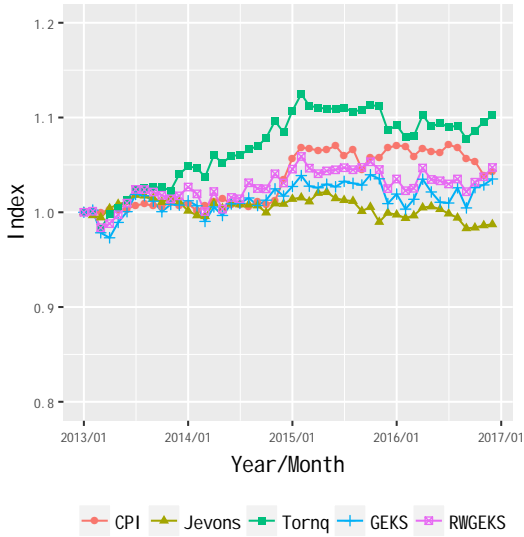


## 공통상품, 특이항제거

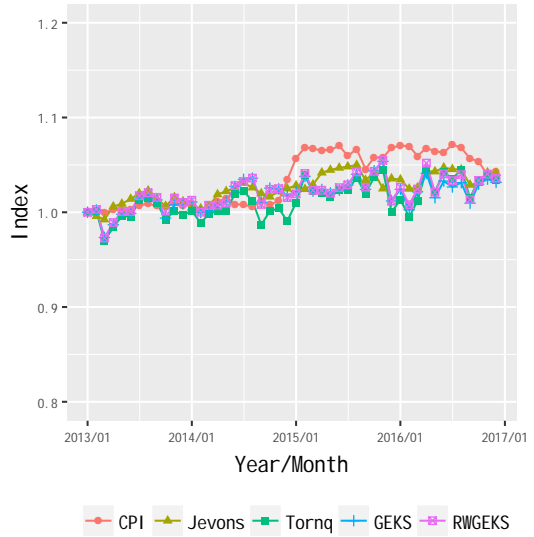


# A019170(냉동식품)

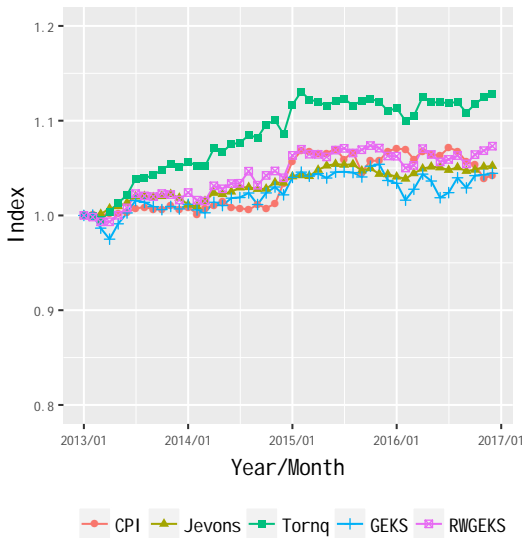
## 원데이터



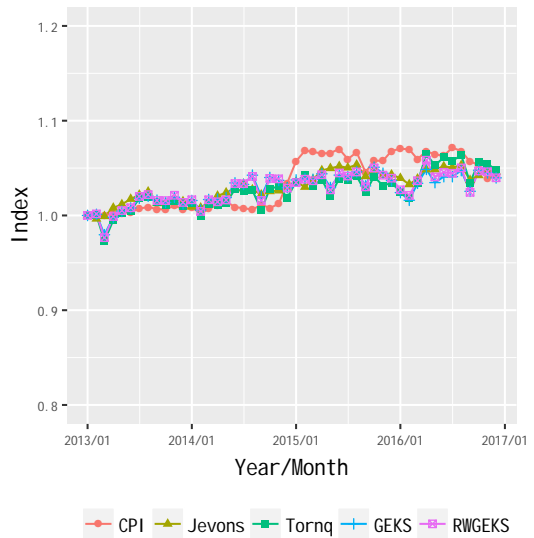
## 공통상품



## 특이항제거

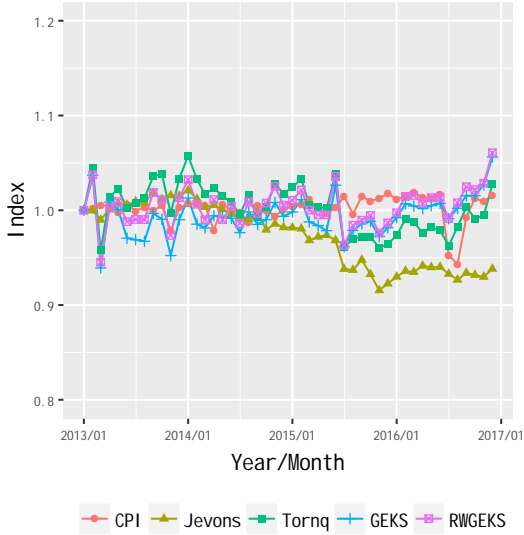


## 공통상품, 특이항제거

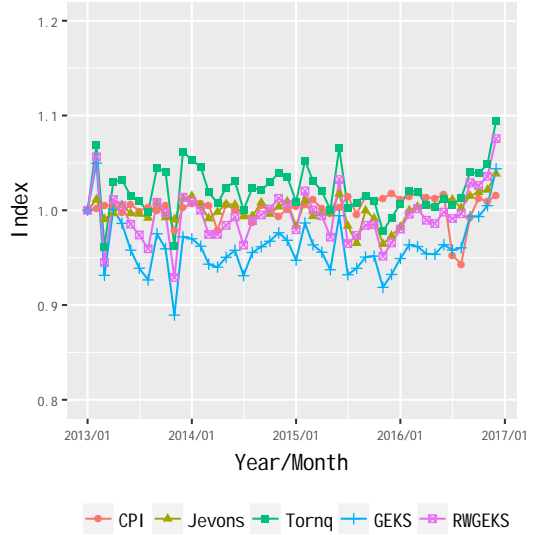


# A019180(즉석식품)

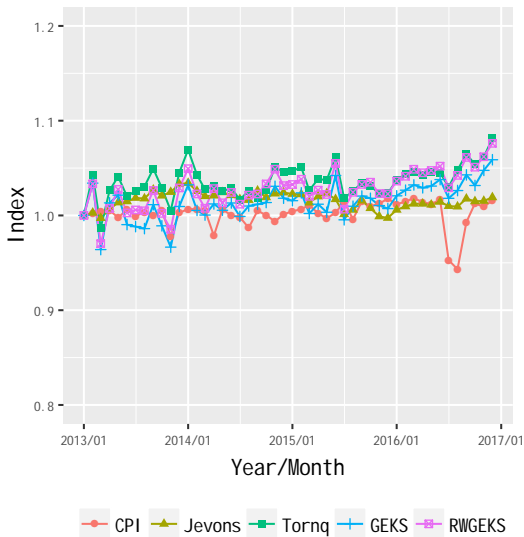
## 원데이터



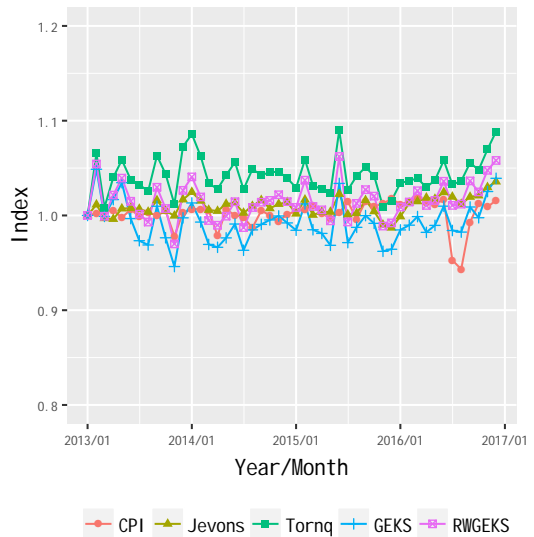
## 공통상품



## 특이항제거

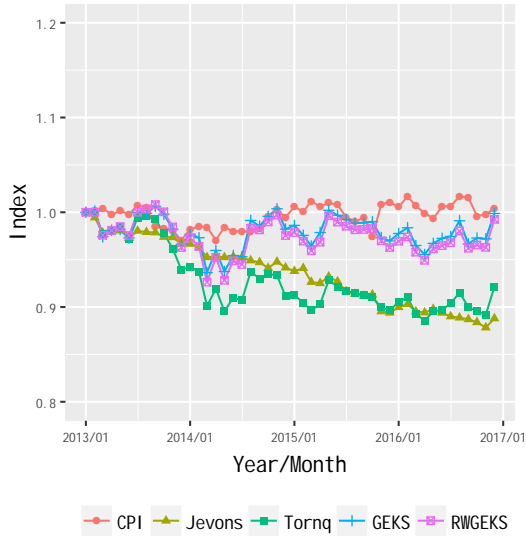


## 공통상품, 특이항제거

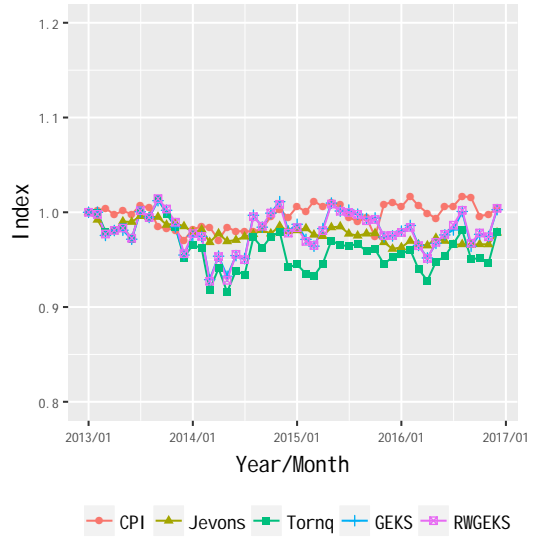


# A021010(커피)

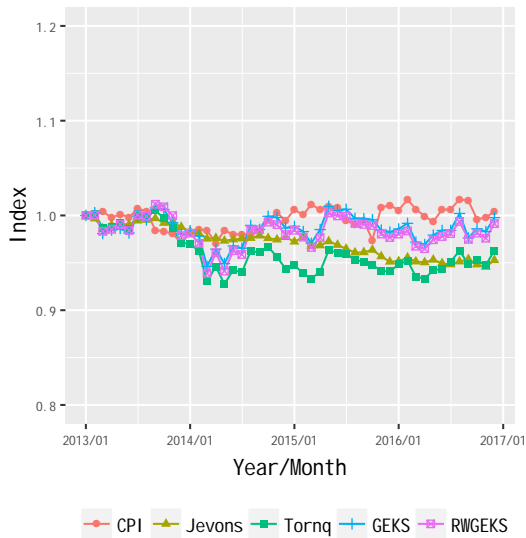
## 원데이터



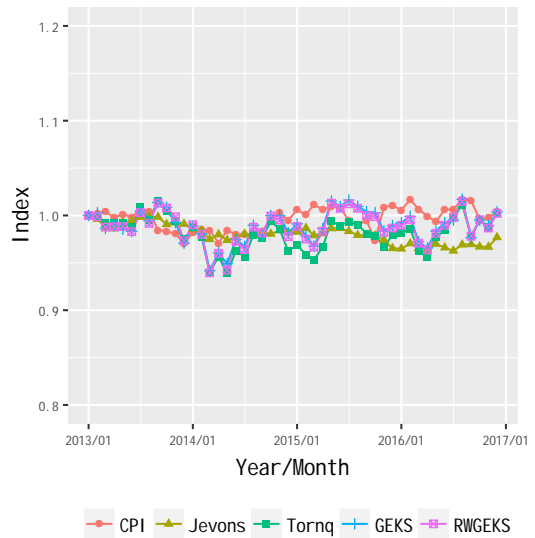
## 공통상품



## 특이항제거



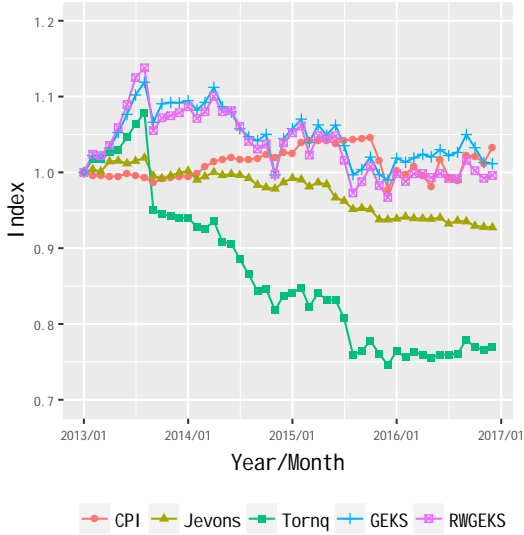
## 공통상품, 특이항제거



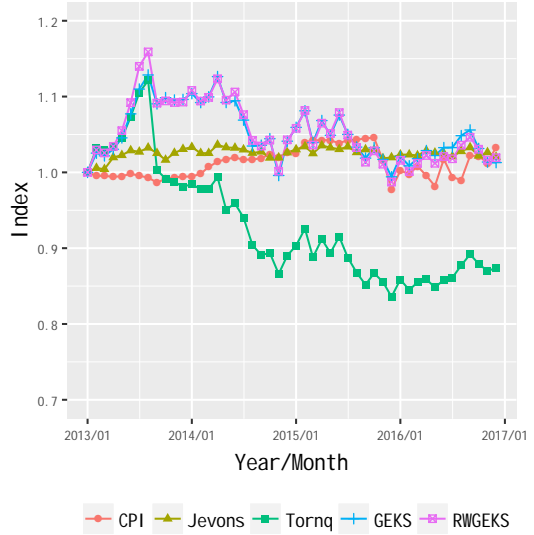


# A021020(차)

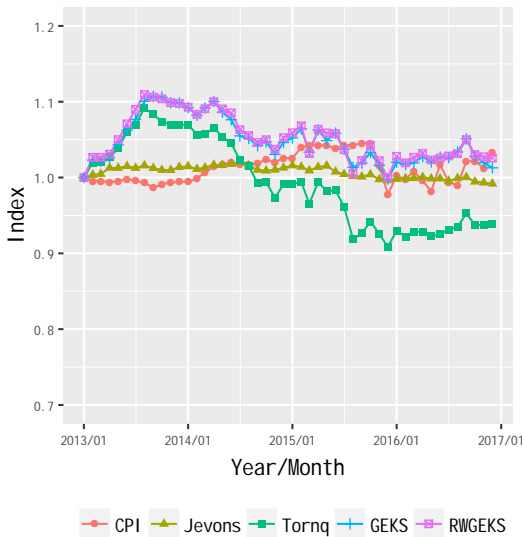
## 원데이터



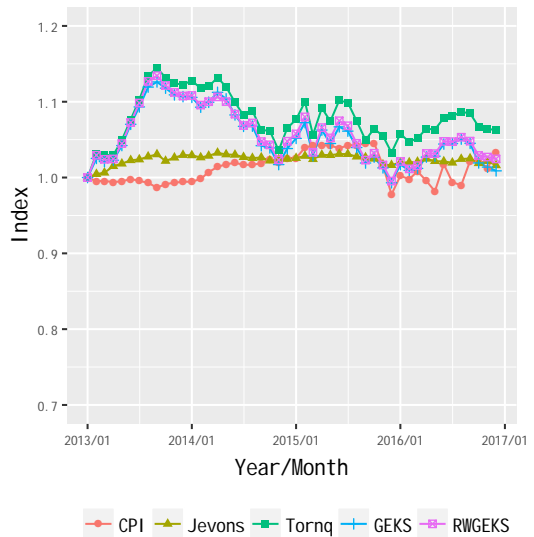
## 공통상품



## 특이항제거

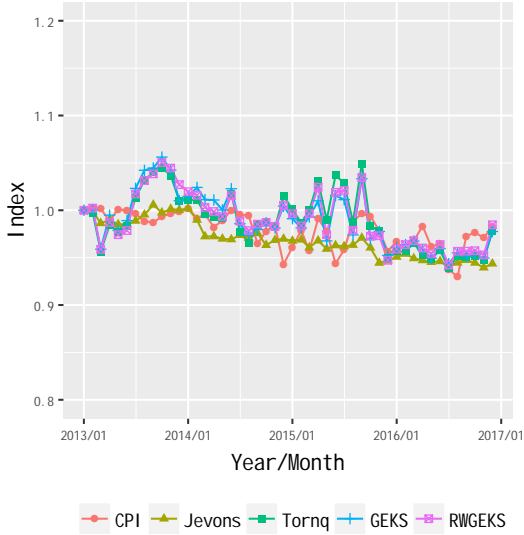


## 공통상품, 특이항제거

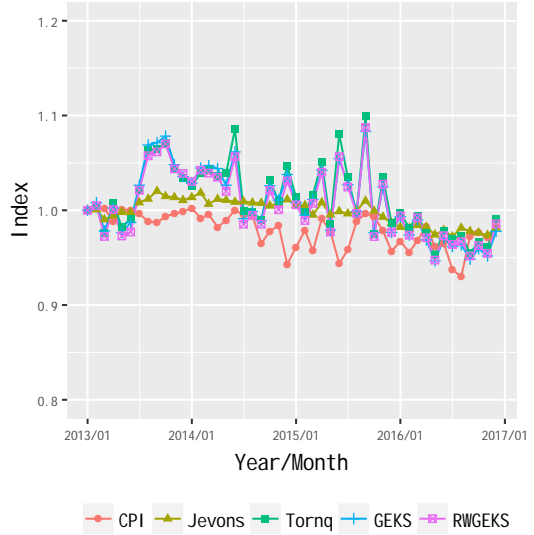


# A022010(주스)

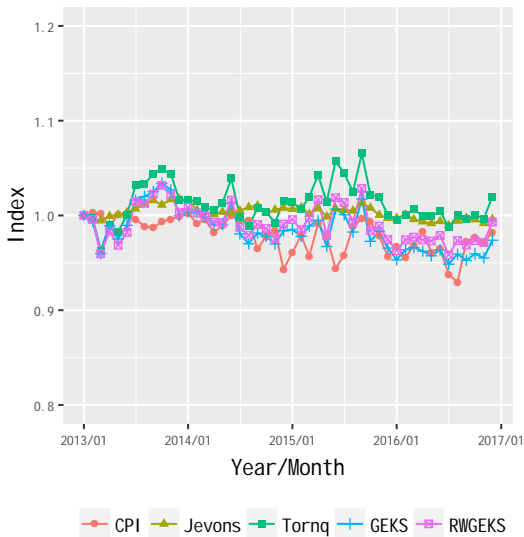
## 원데이터



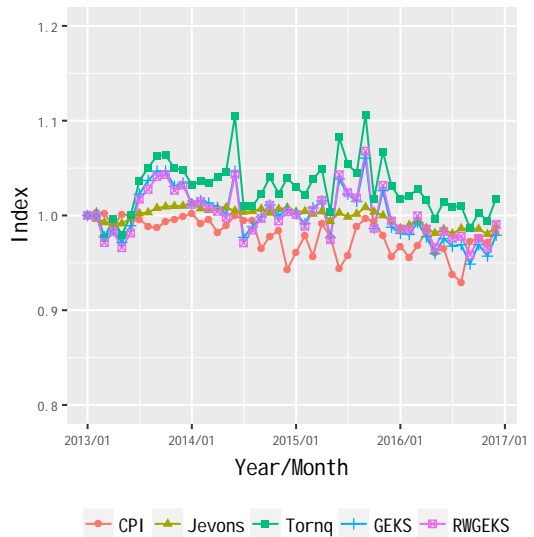
## 공통상품



## 특이항제거

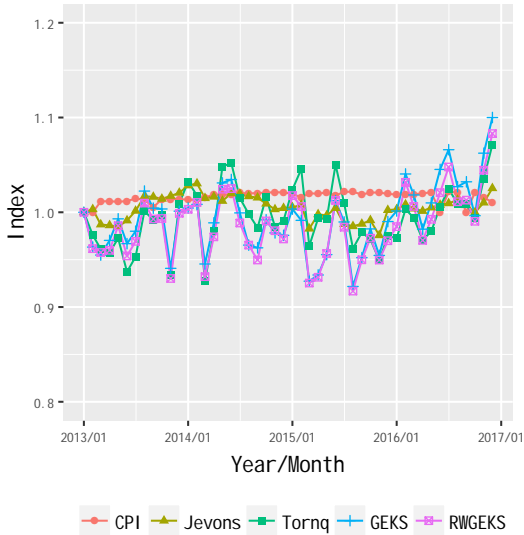


## 공통상품, 특이항제거

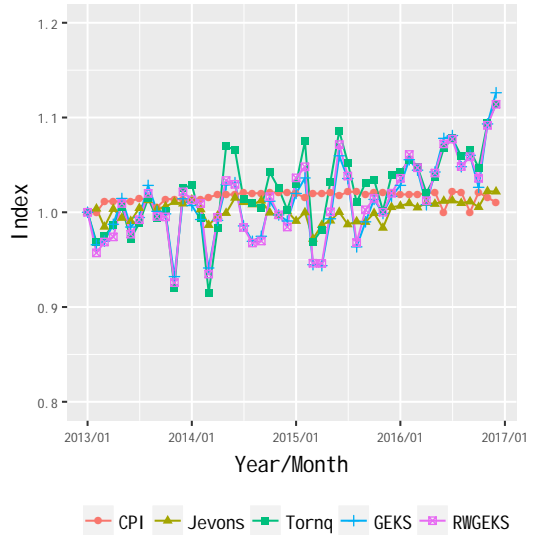


# A022020(두유)

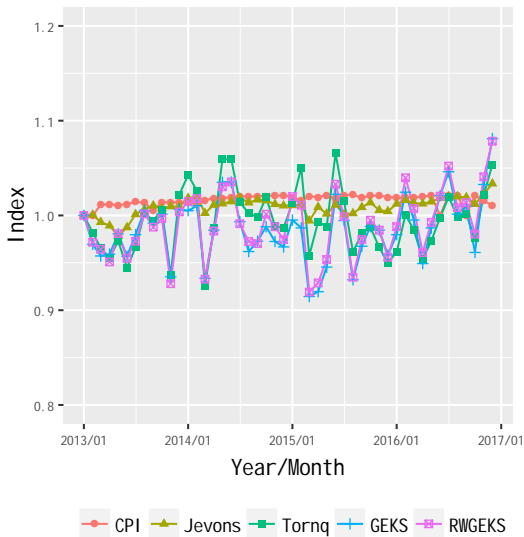
## 원데이터



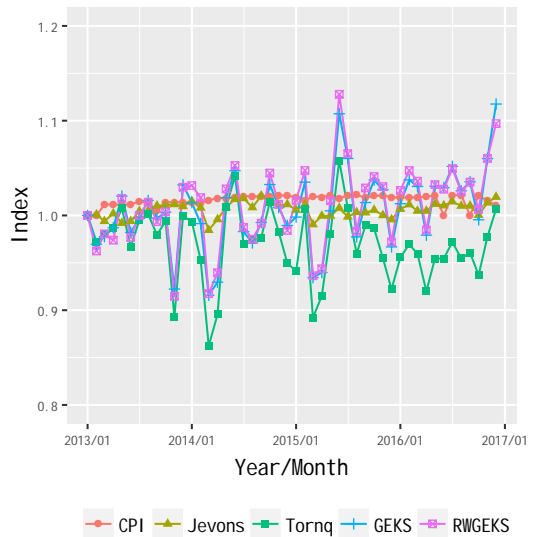
## 공통상품



## 특이항제거

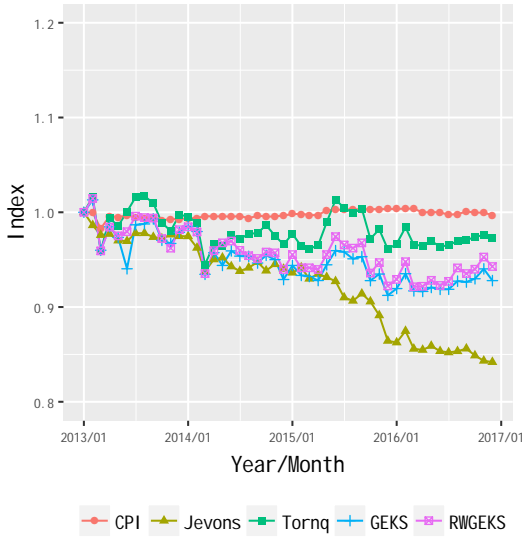


## 공통상품, 특이항제거

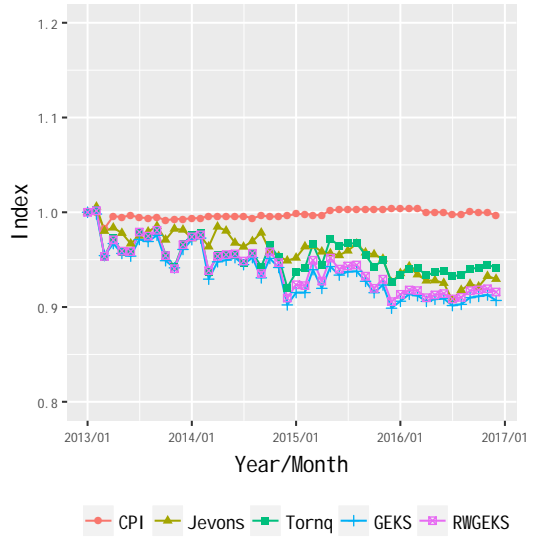


# A022030(생수)

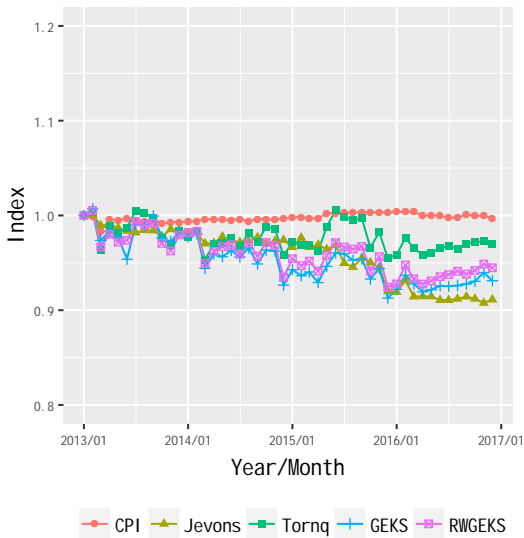
## 원데이터



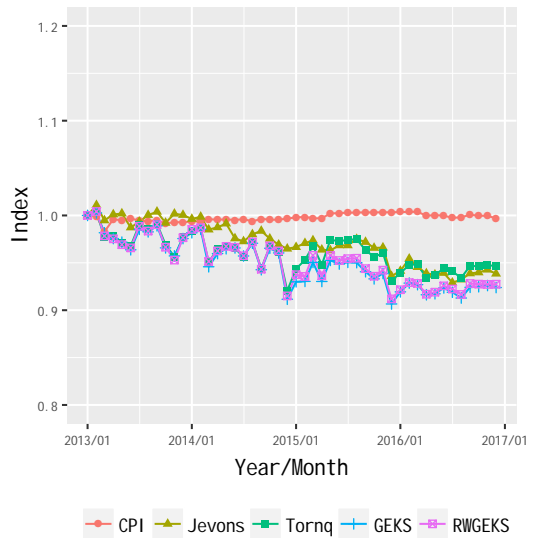
## 공통상품



## 특이항제거

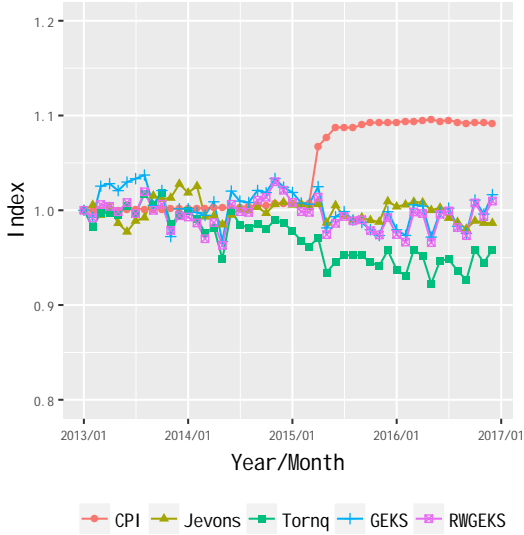


## 공통상품, 특이항제거

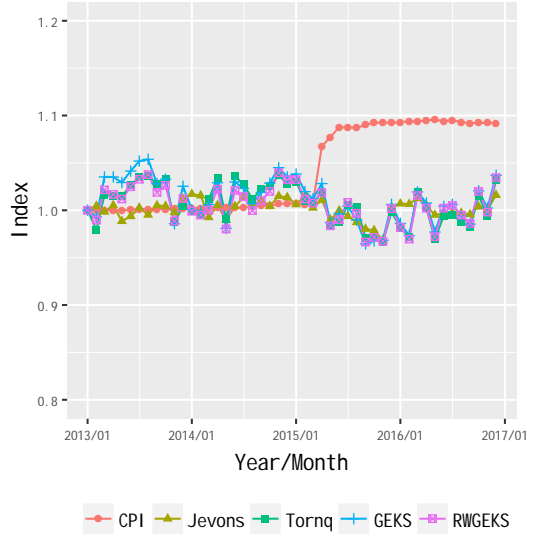


# A022040(기능성음료)

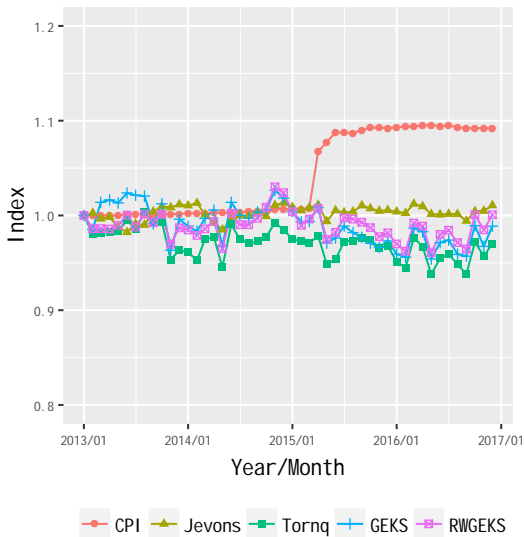
## 원데이터



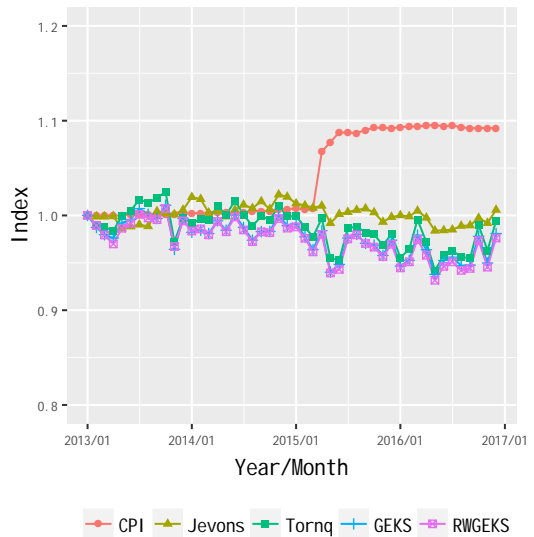
## 공통상품



## 특이항제거

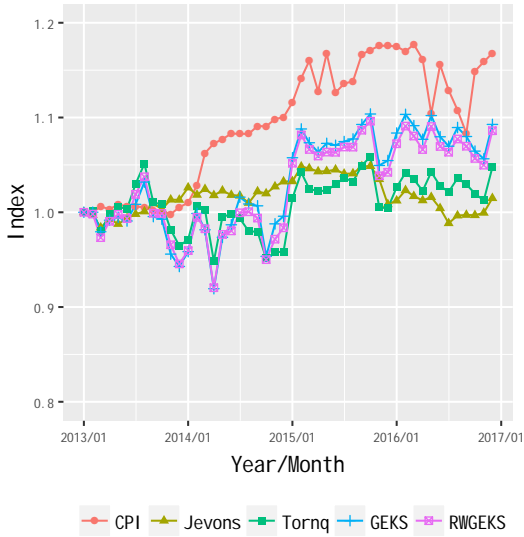


## 공통상품, 특이항제거

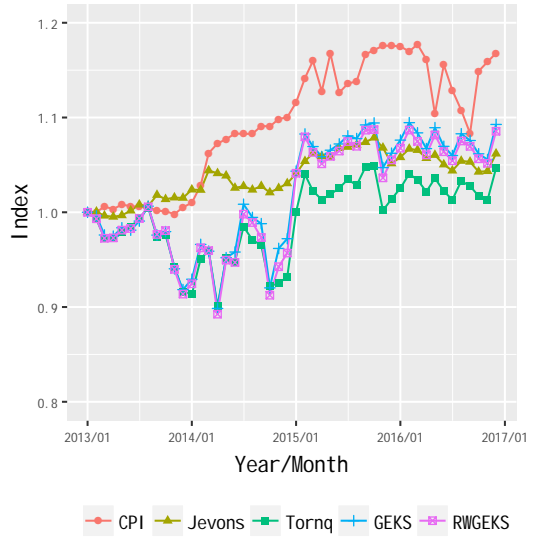


# A022050(탄산음료)

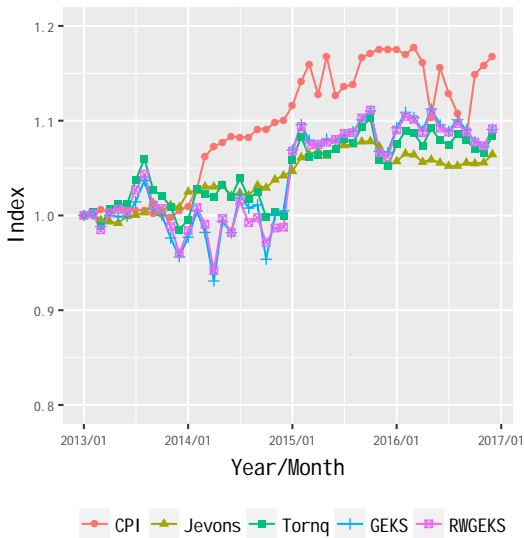
## 원데이터



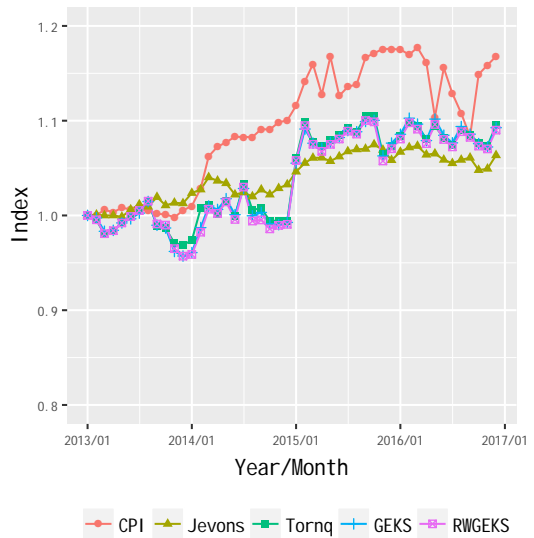
## 공통상품



## 특이항제거

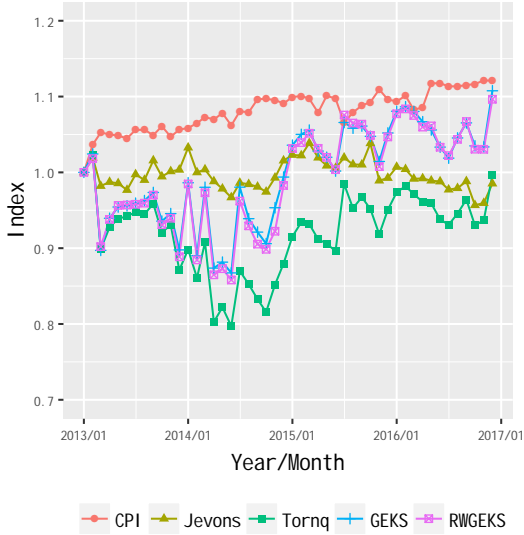


## 공통상품, 특이항제거

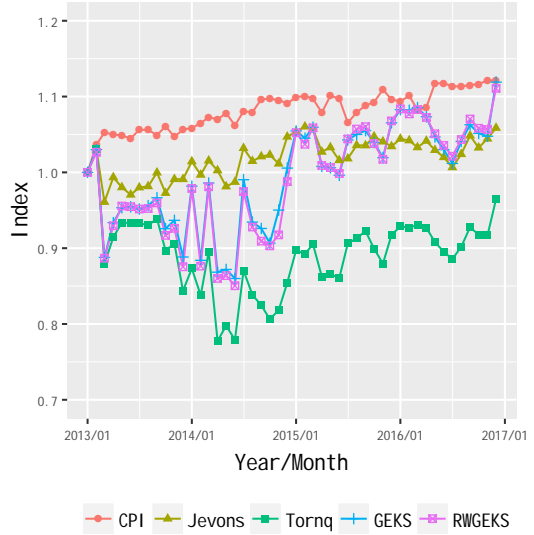


# A022060(혼합음료)

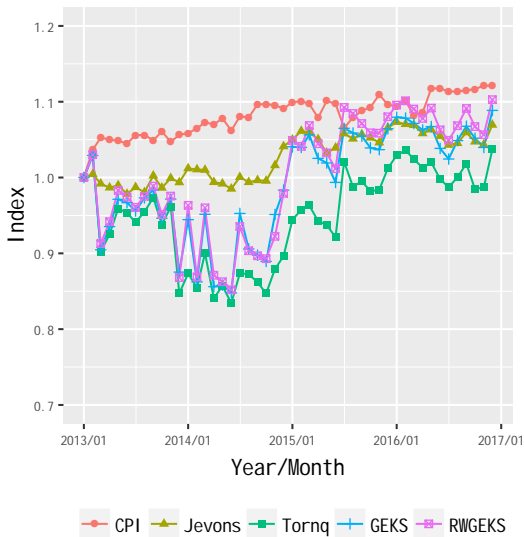
## 원데이터



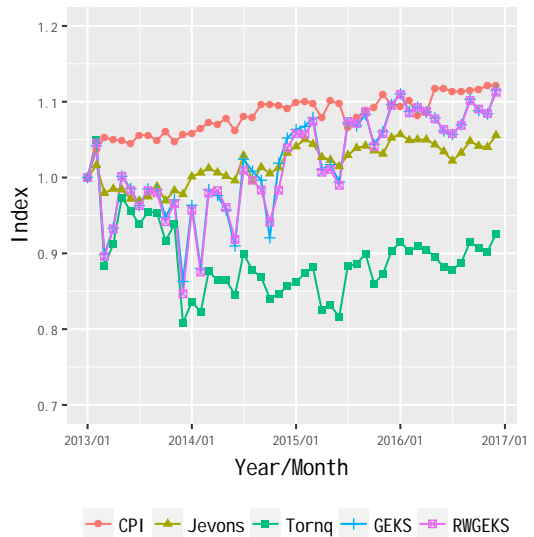
## 공통상품



## 특이항제거

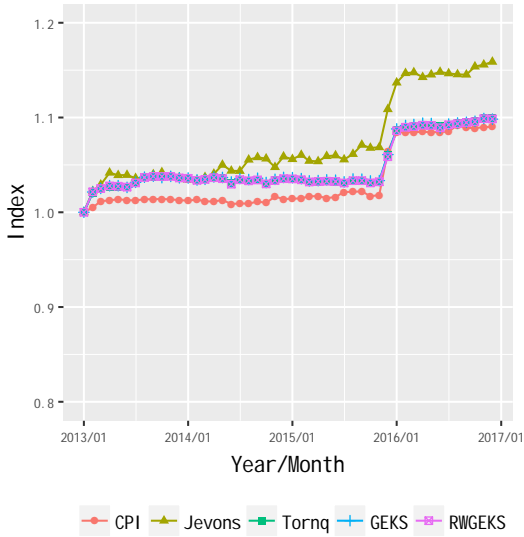


## 공통상품, 특이항제거

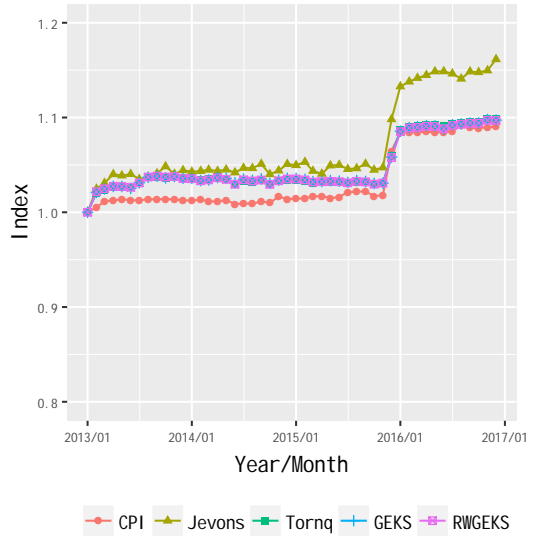


# B011010(소주)

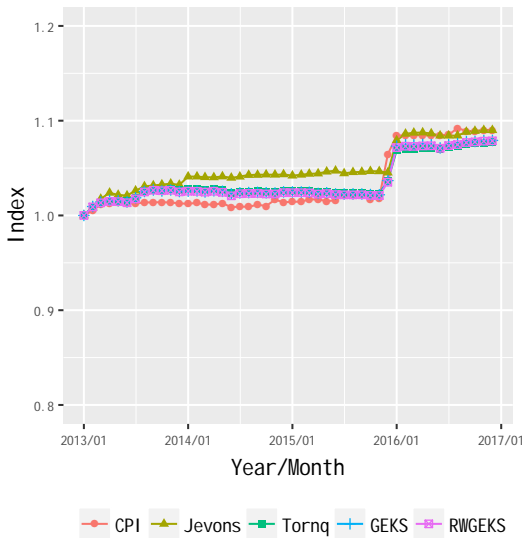
## 원데이터



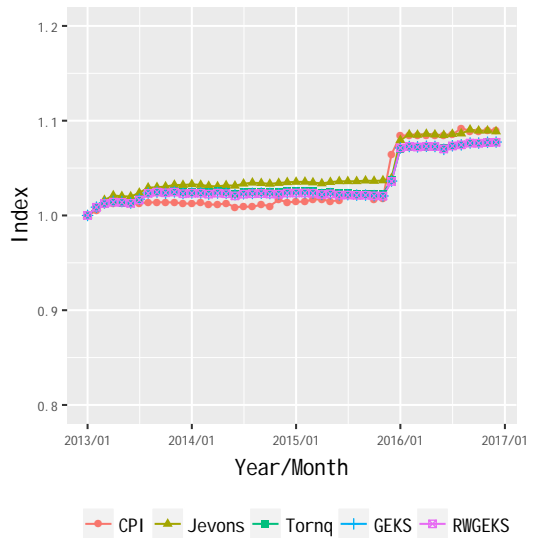
## 공통상품



## 특이항제거



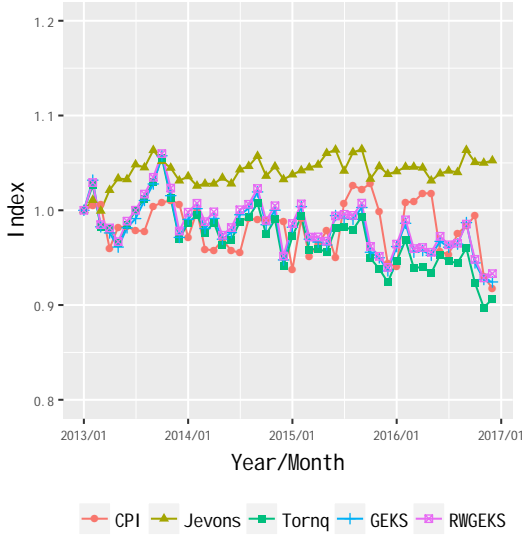
## 공통상품, 특이항제거



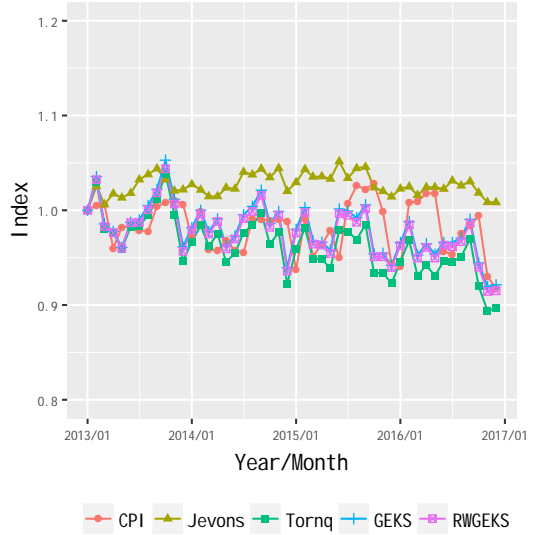


# B011020(과실주)

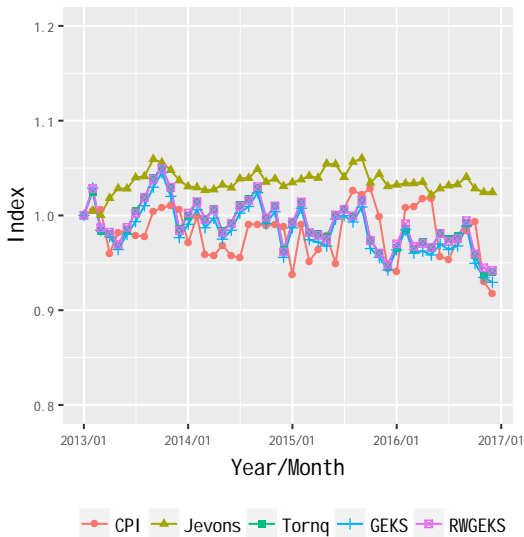
## 원데이터



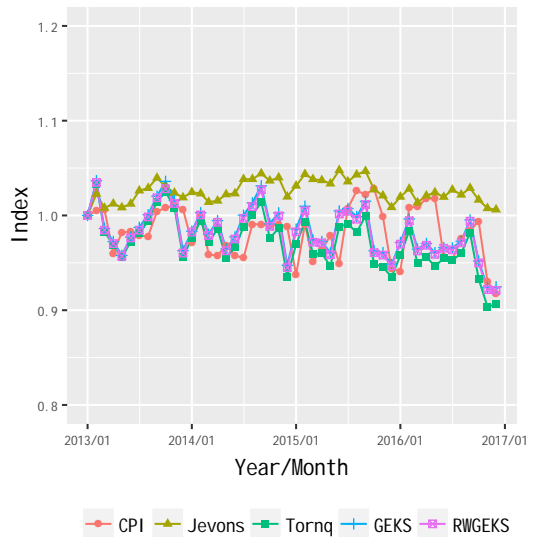
## 공통상품



## 특이항제거

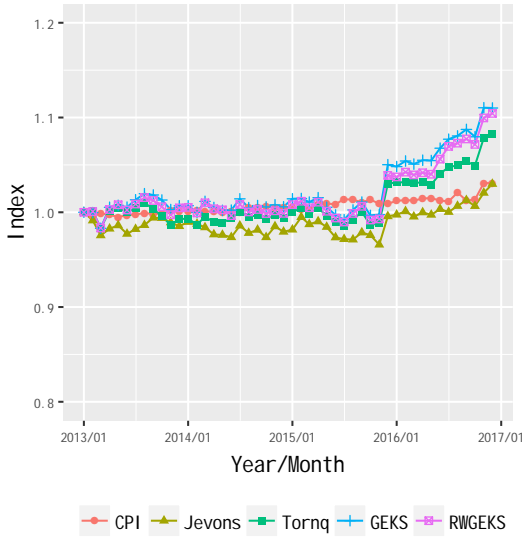


## 공통상품, 특이항제거

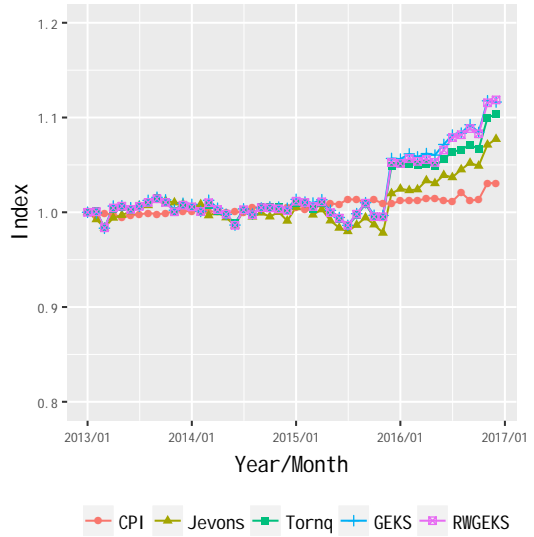


# B011030(맥주)

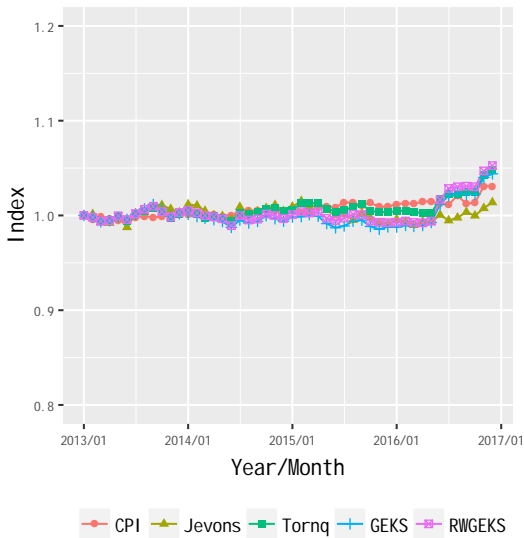
## 원데이터



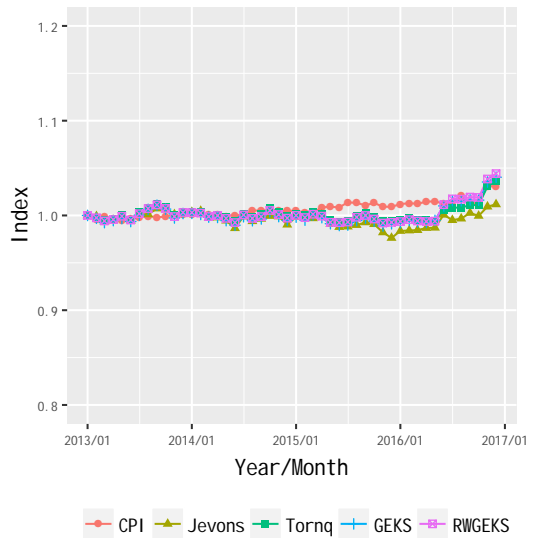
## 공통상품



## 특이항제거

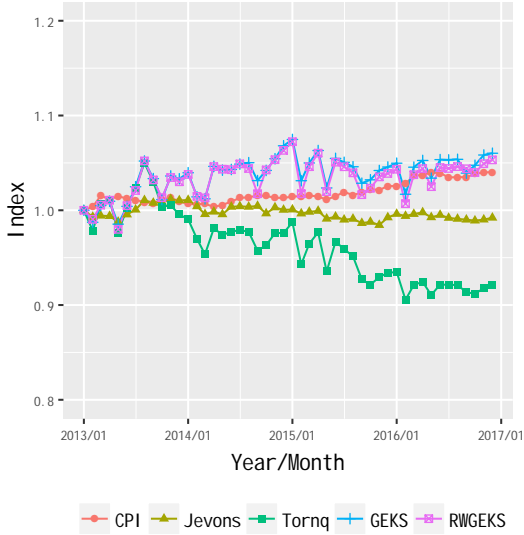


## 공통상품, 특이항제거

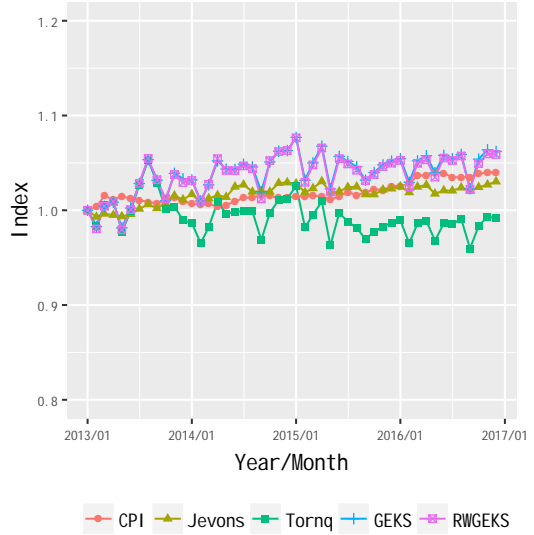


# B011040(막걸리)

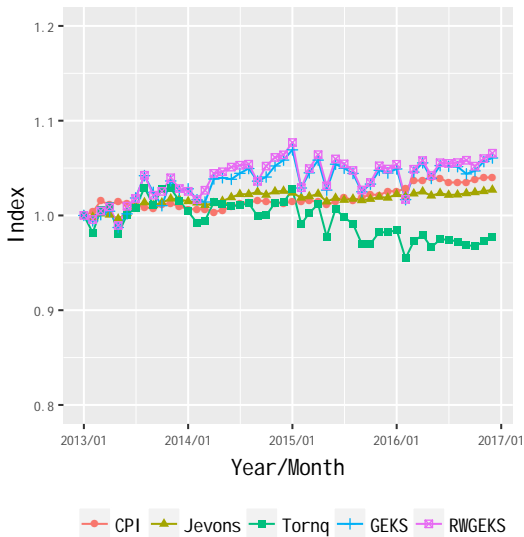
## 원데이터



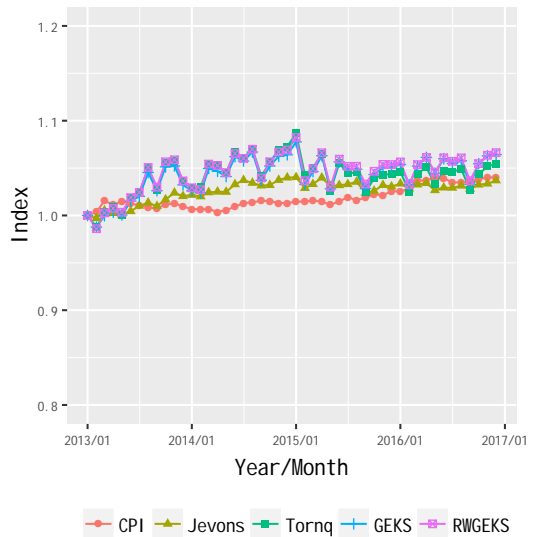
## 공통상품



## 특이항제거

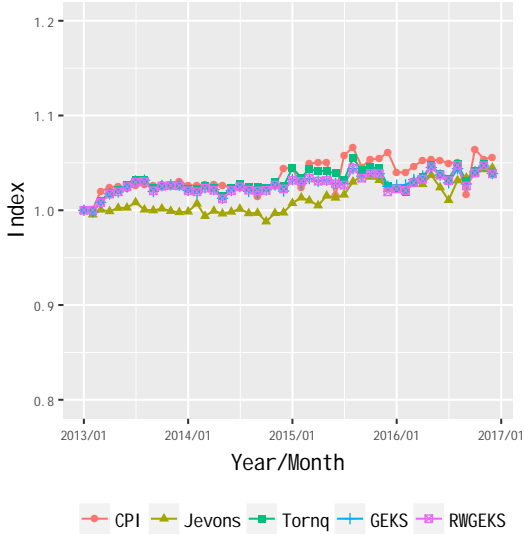


## 공통상품, 특이항제거

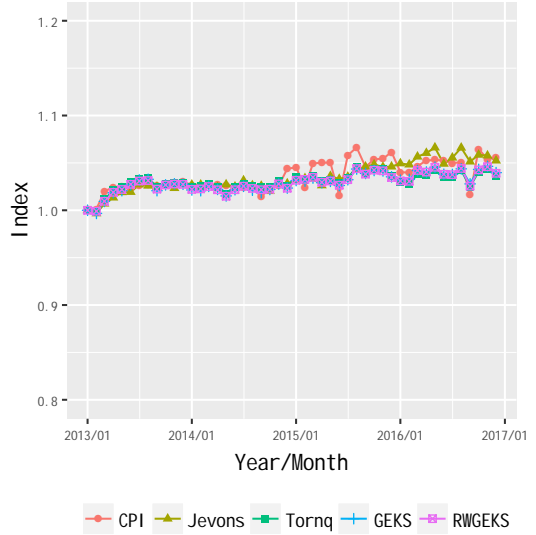


# B011050(양주)

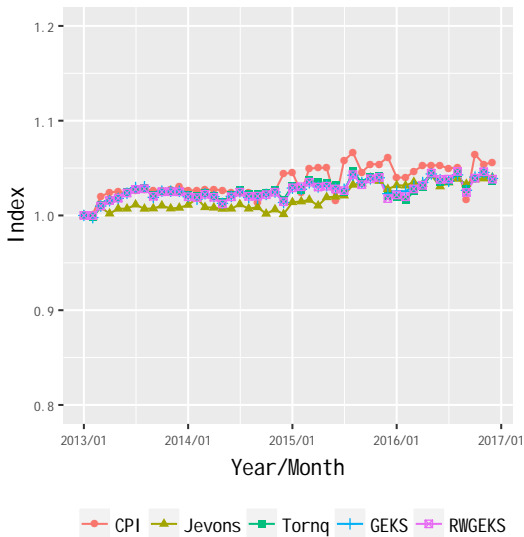
## 원데이터



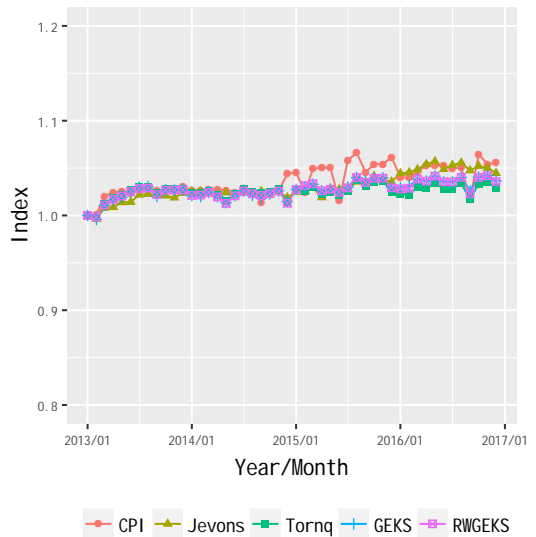
## 공통상품



## 특이항제거

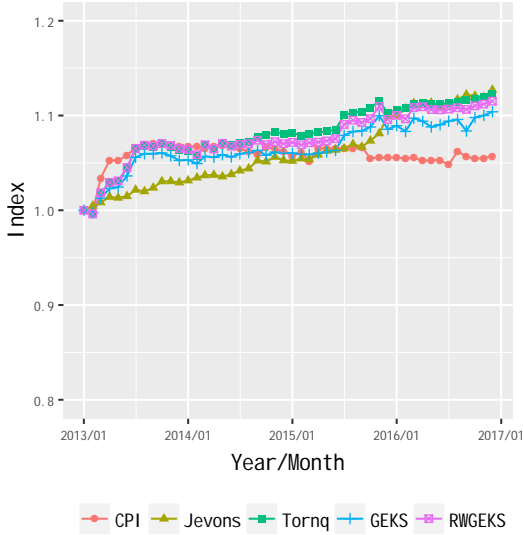


## 공통상품, 특이항제거

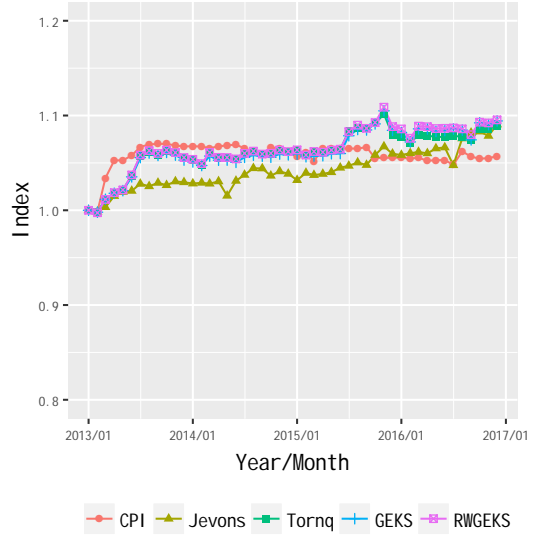


# B011060(약주)

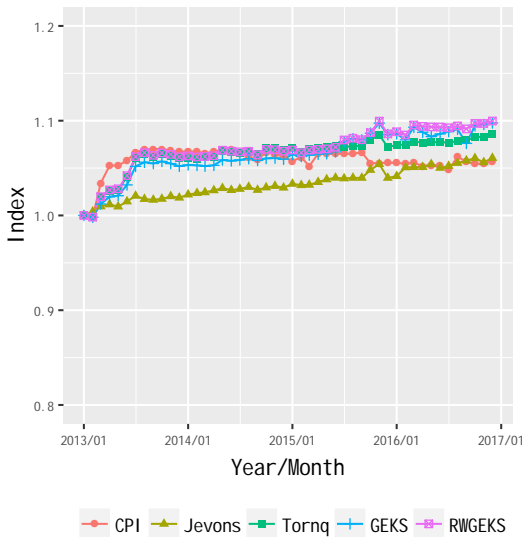
## 원데이터



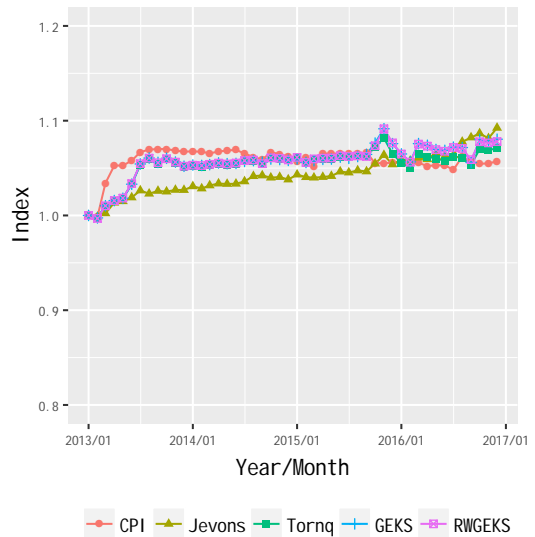
## 공통상품



## 특이항제거

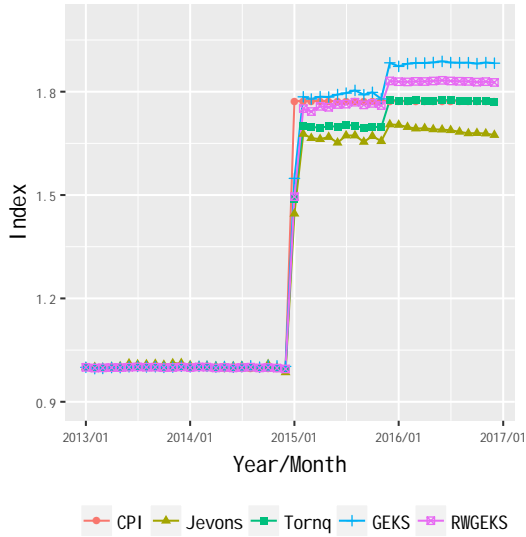


## 공통상품, 특이항제거

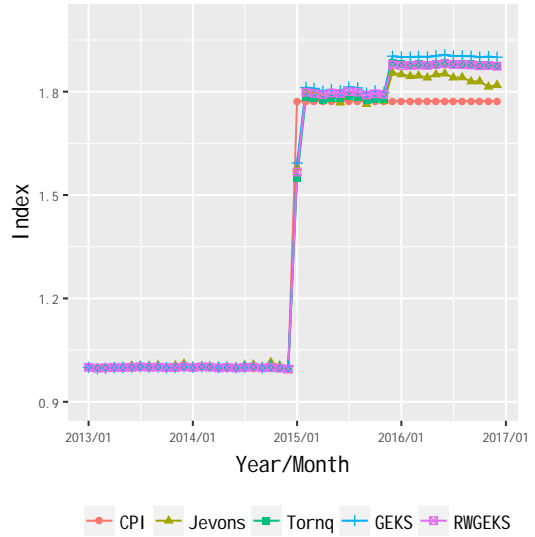


# B021010(담배)

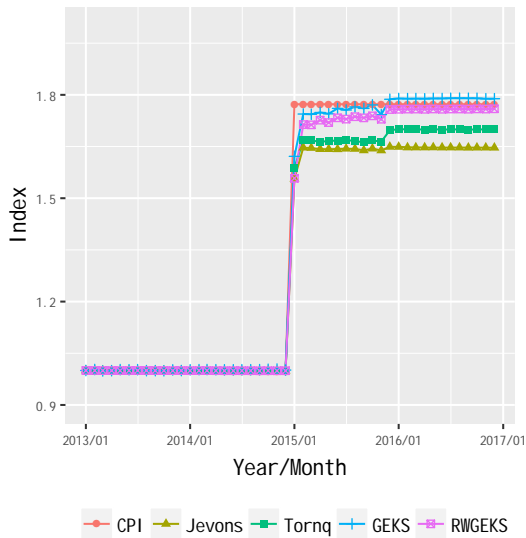
## 원데이터



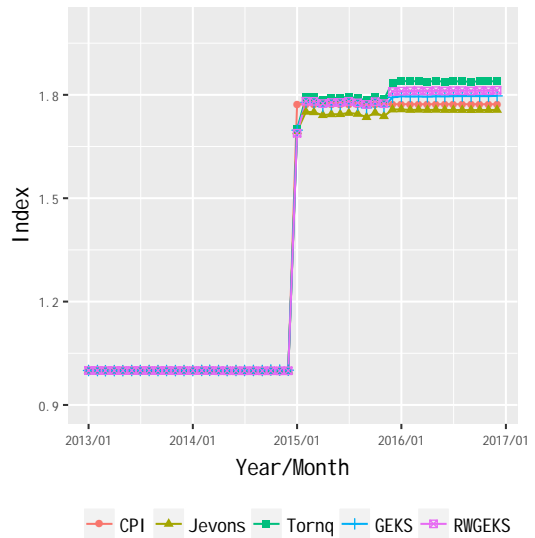
## 공통상품



## 특이항제거

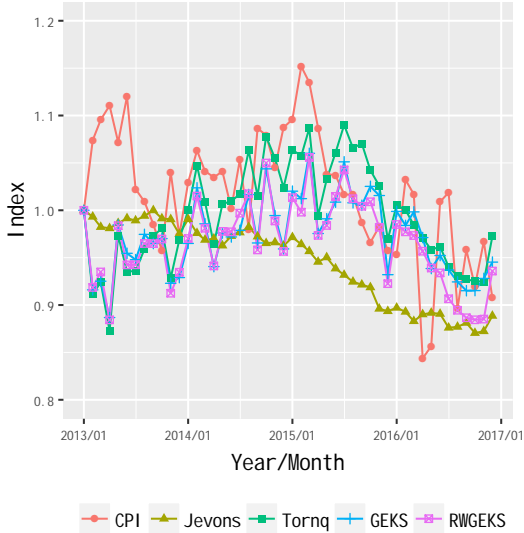


## 공통상품, 특이항제거

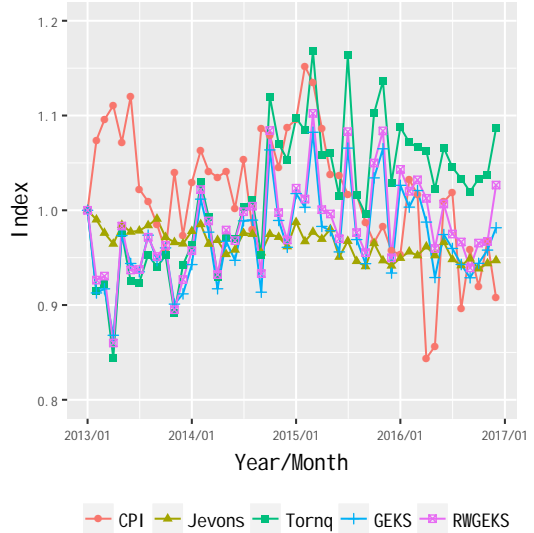


# E061010(세탁세제)

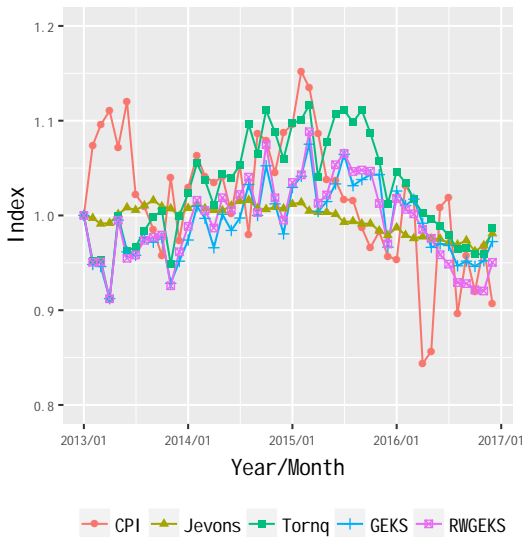
## 원데이터



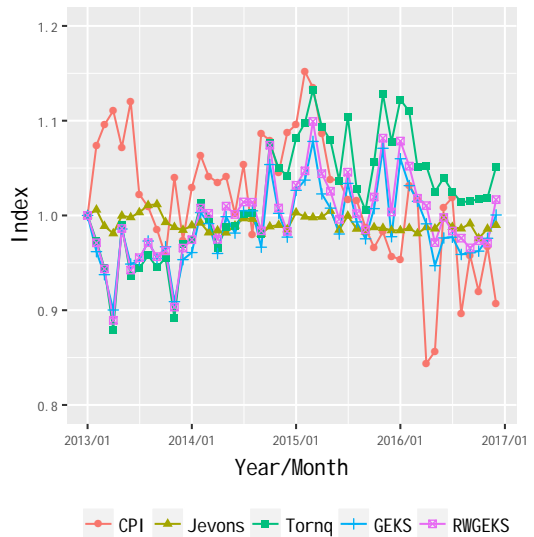
## 공통상품



## 특이항제거

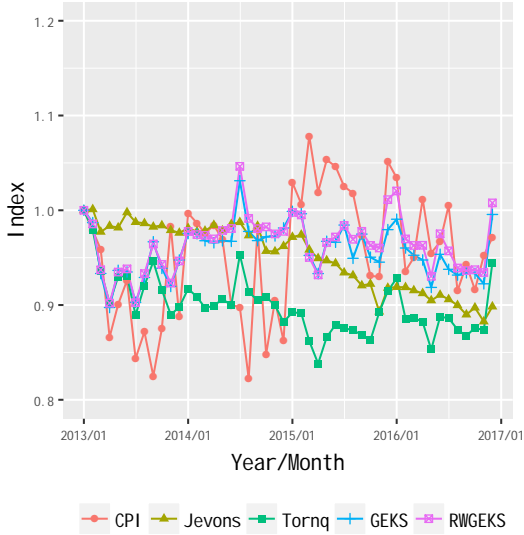


## 공통상품, 특이항제거

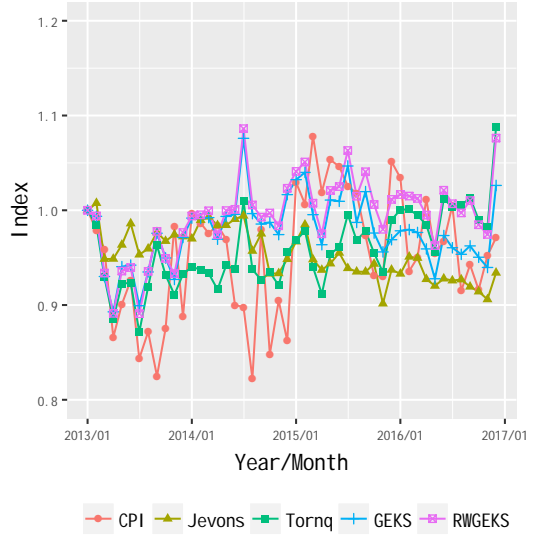


# E061020(섬유유연제)

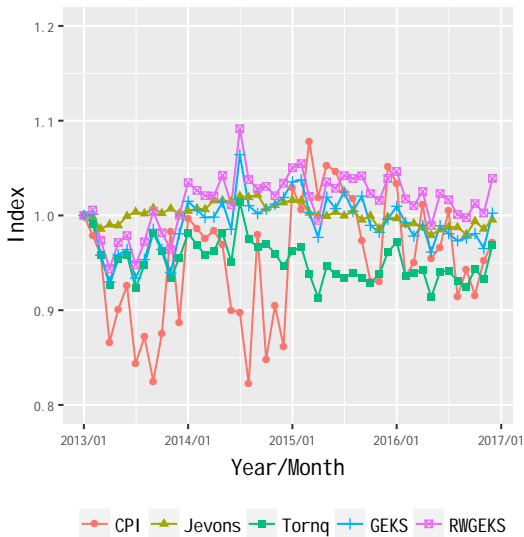
## 원데이터



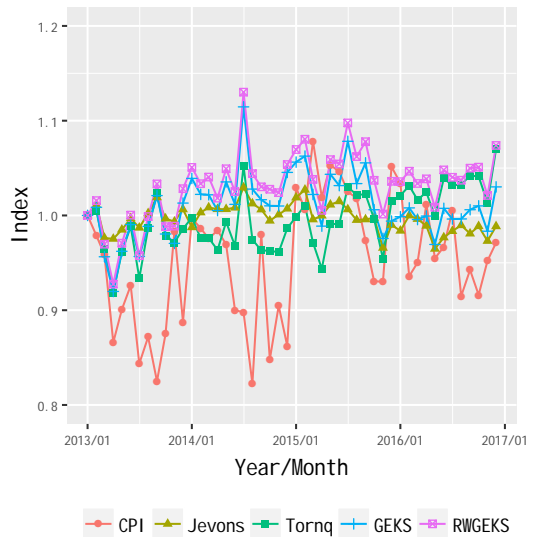
## 공통상품



## 특이항제거



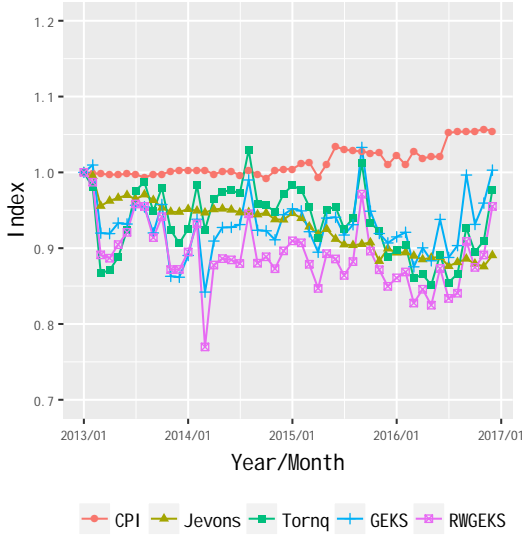
## 공통상품, 특이항제거



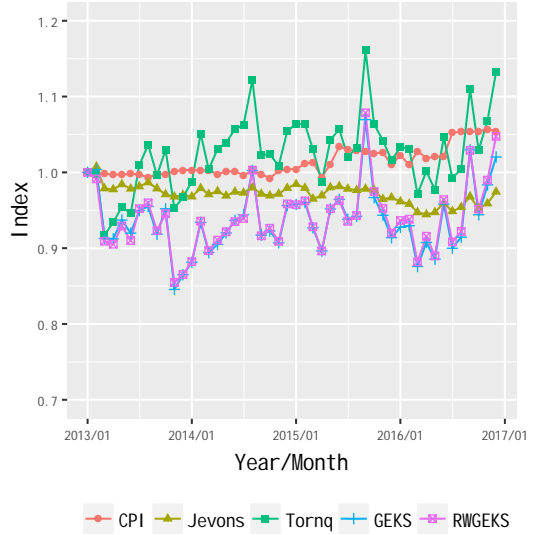


# E061040(부업용세제)

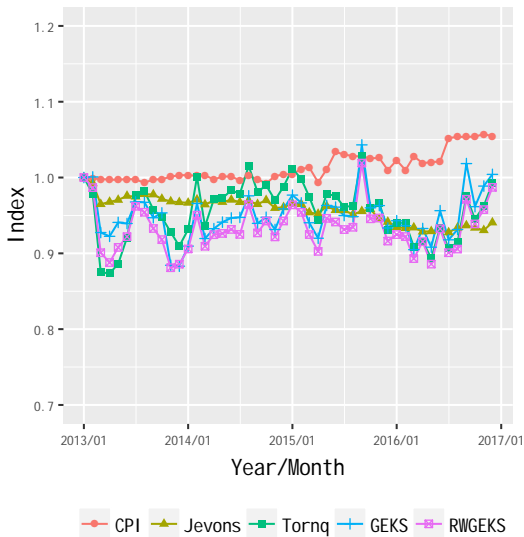
## 원데이터



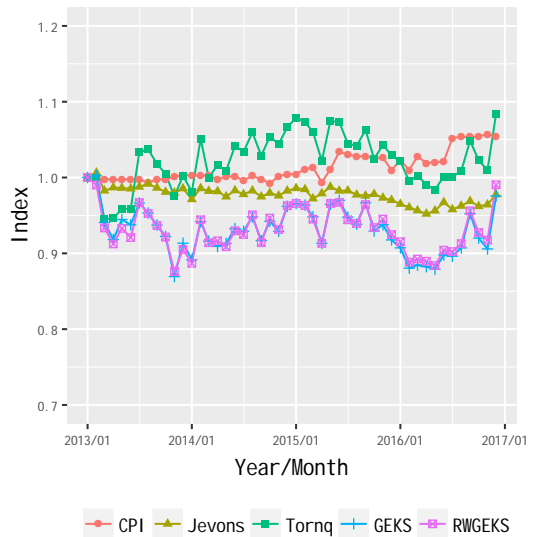
## 공통상품



## 특이항제거

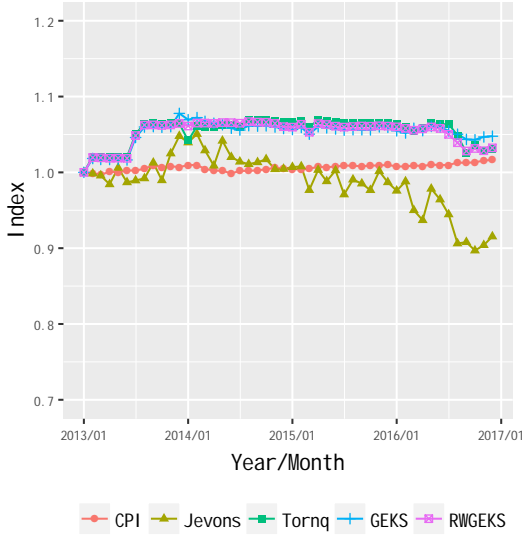


## 공통상품, 특이항제거

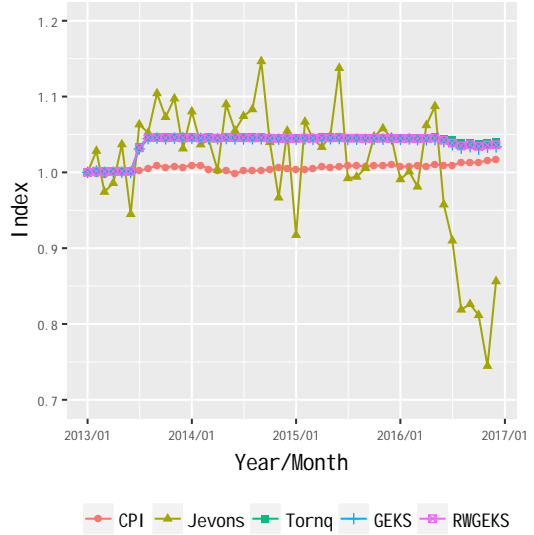


# E061050(청소용세제)

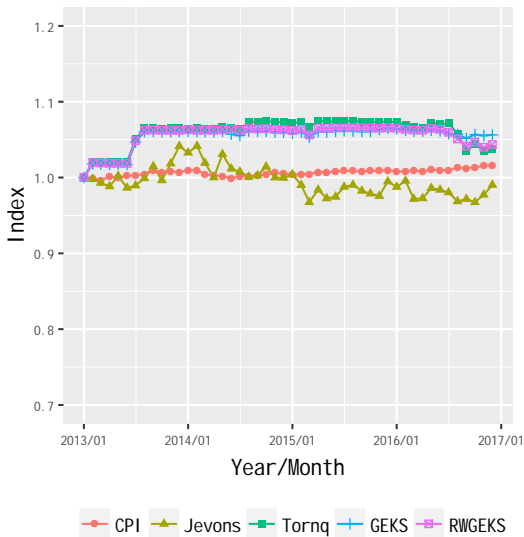
## 원데이터



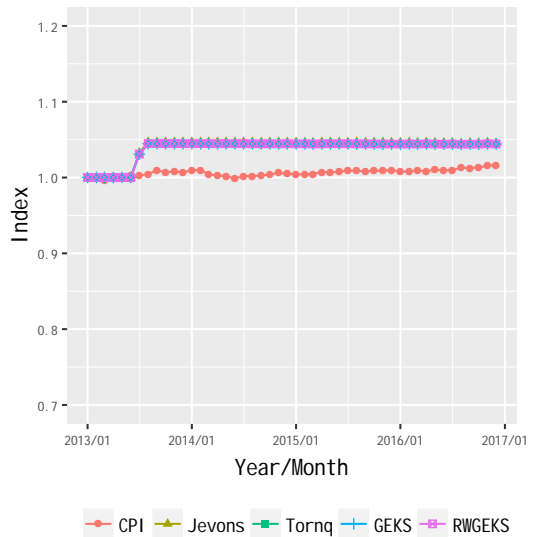
## 공통상품



## 특이항제거

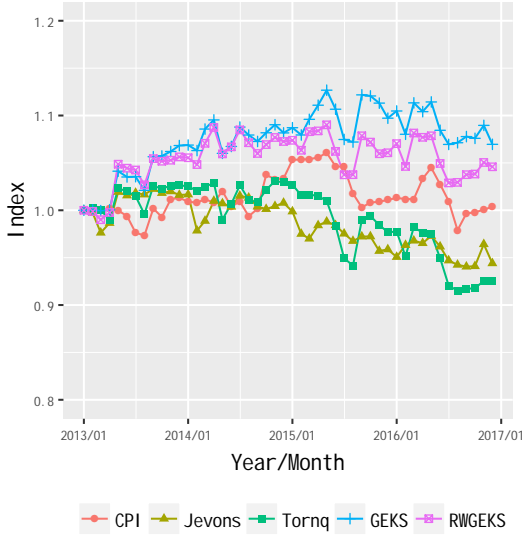


## 공통상품, 특이항제거

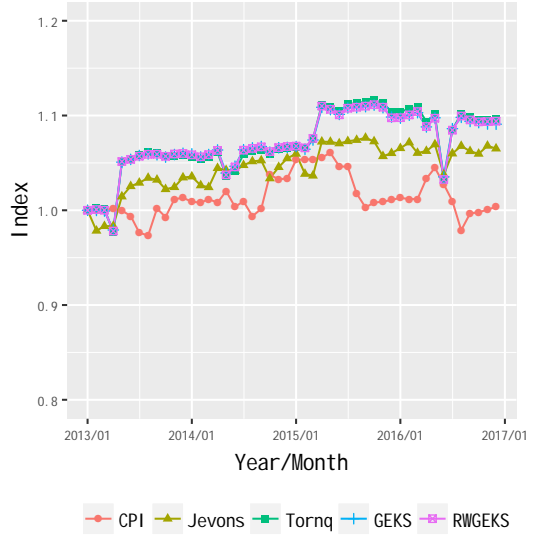


# E061060(살충제)

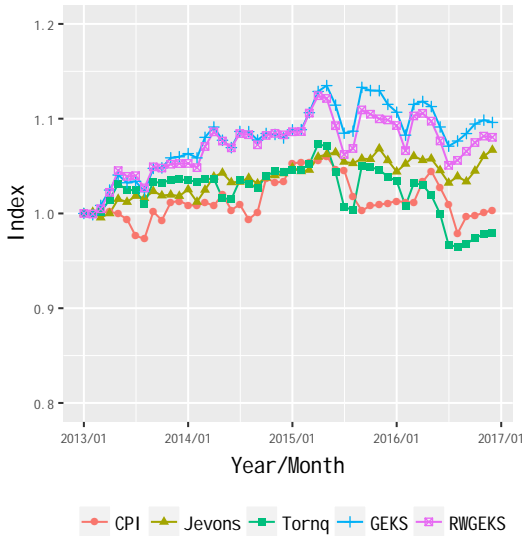
## 원데이터



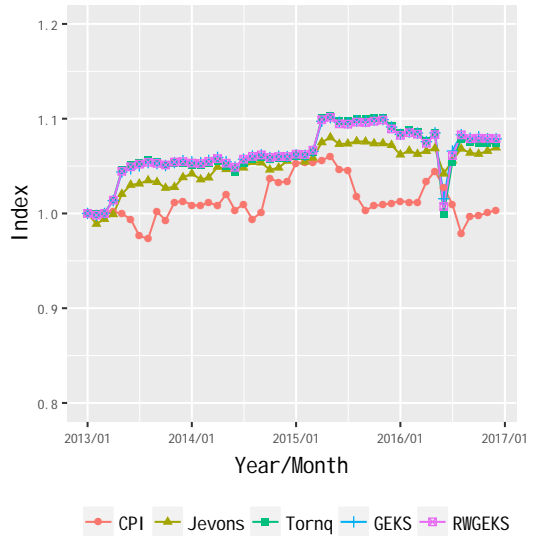
## 공통상품



## 특이항제거

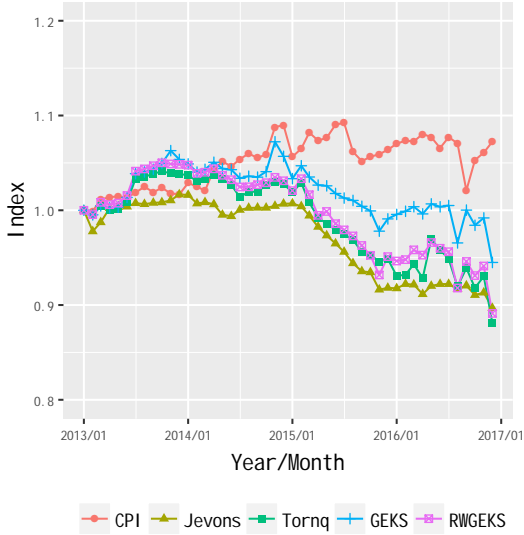


## 공통상품, 특이항제거

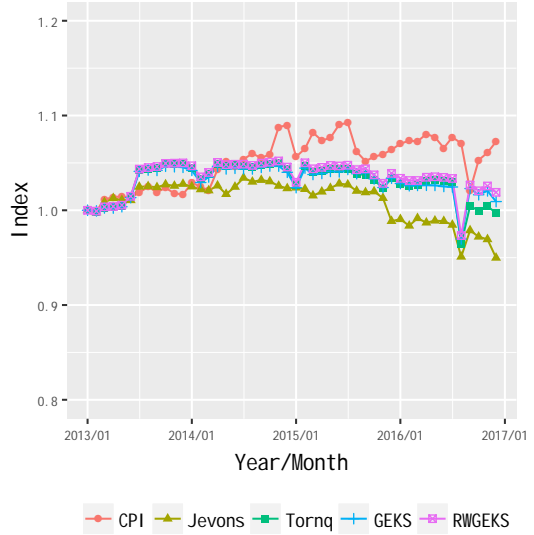


# E061070(가정용비닐용품)

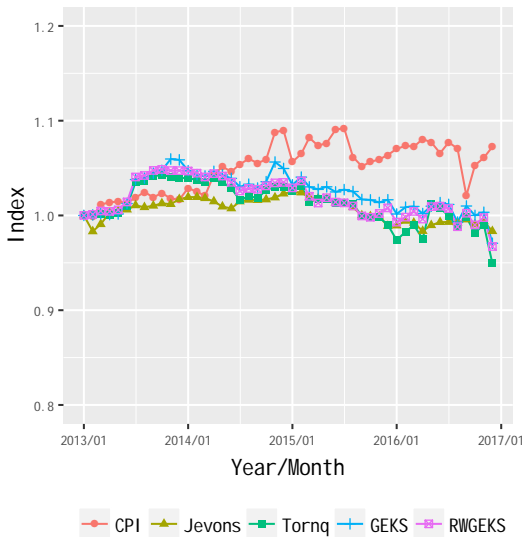
## 원데이터



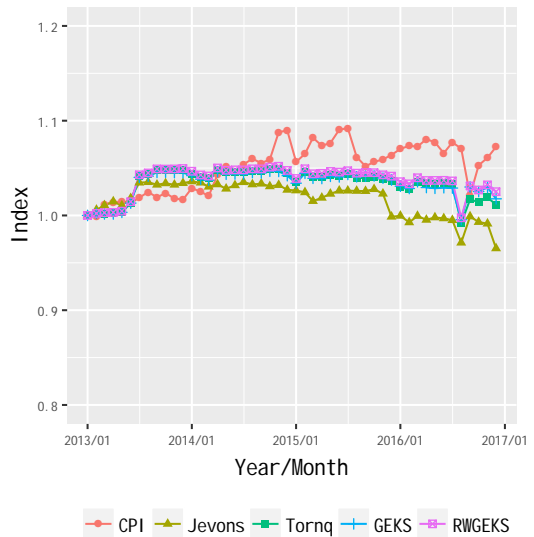
## 공통상품



## 특이항제거

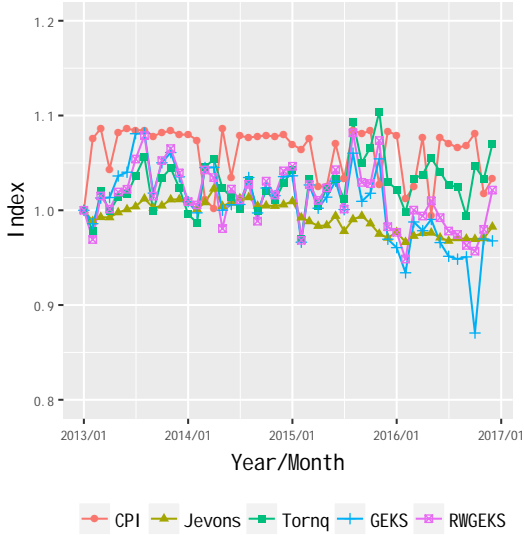


## 공통상품, 특이항제거

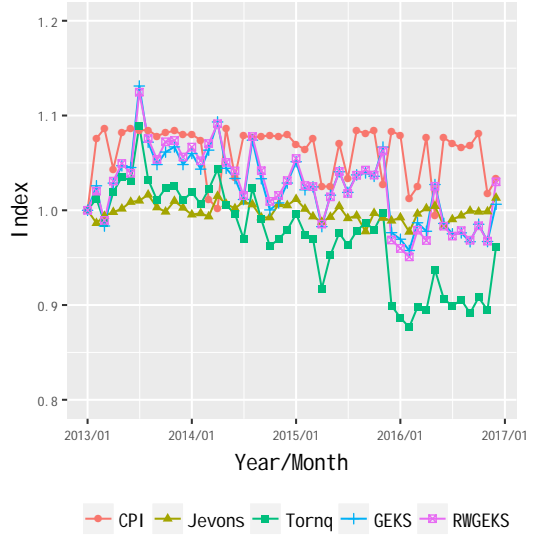


# E061080(키친타월)

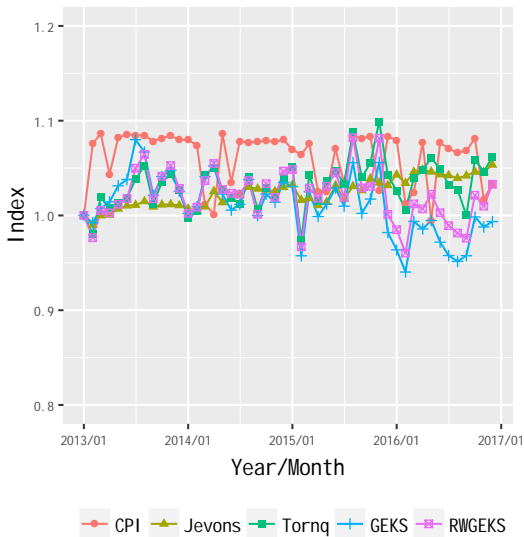
## 원데이터



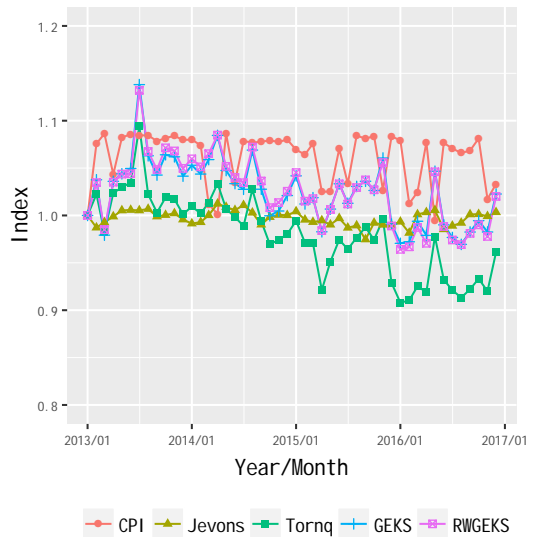
## 공통상품



## 특이항제거

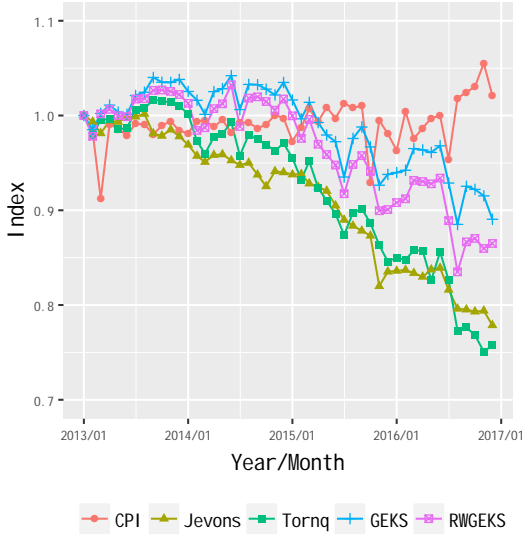


## 공통상품, 특이항제거

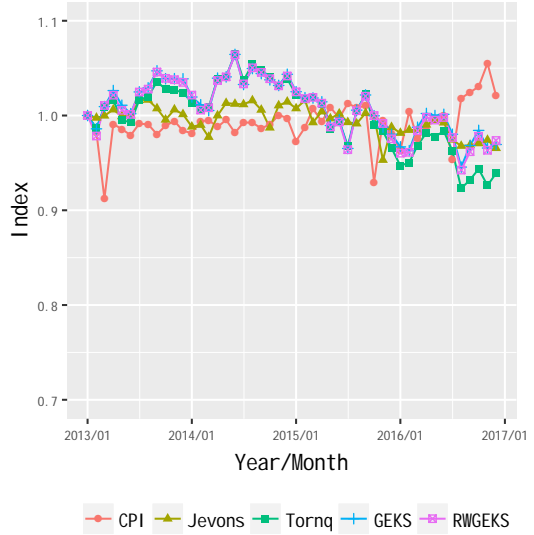


# E061090(방향제)

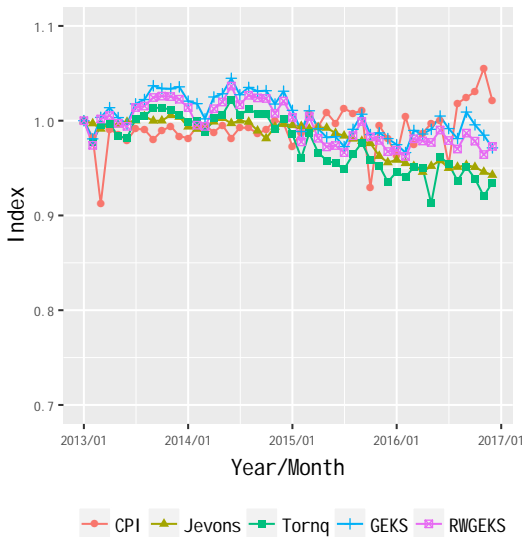
## 원데이터



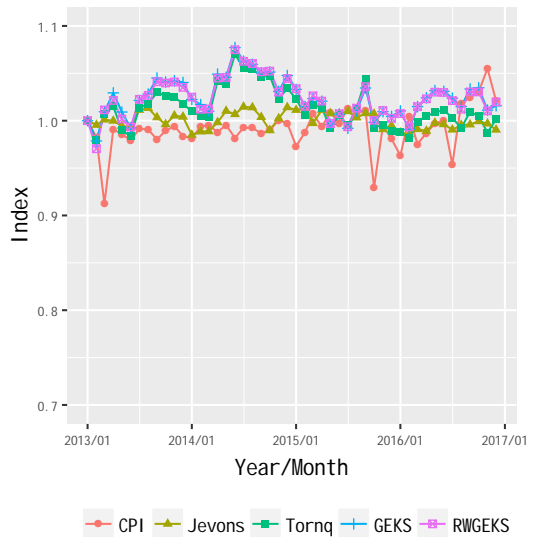
## 공통상품



## 특이항제거

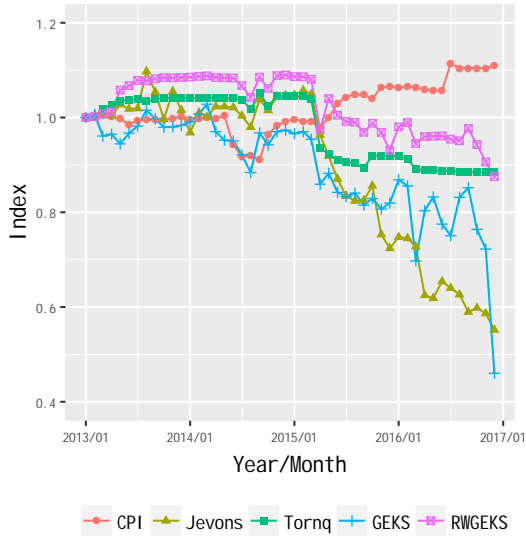


## 공통상품, 특이항제거

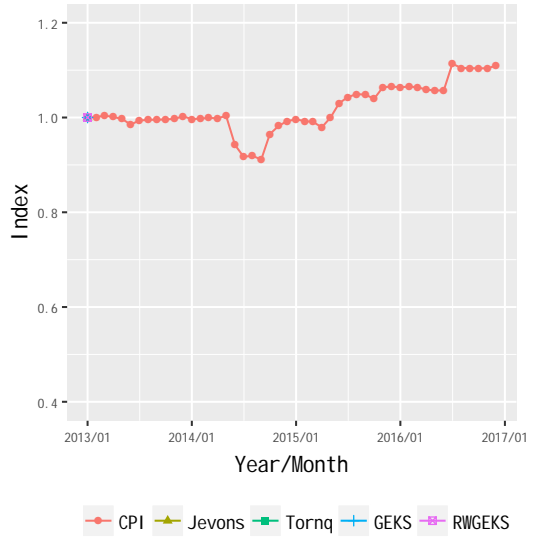


# E061100(습기제거제)

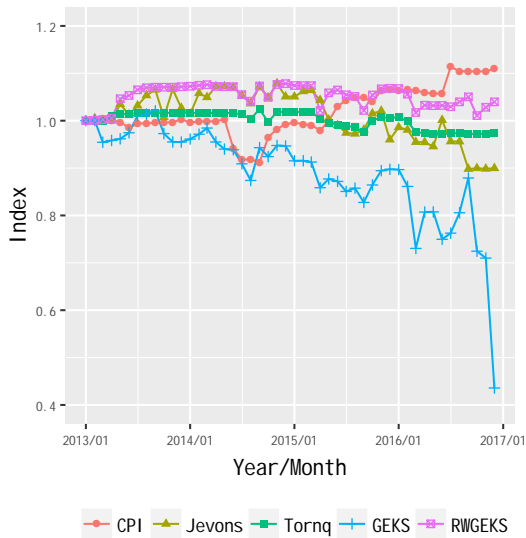
## 원데이터



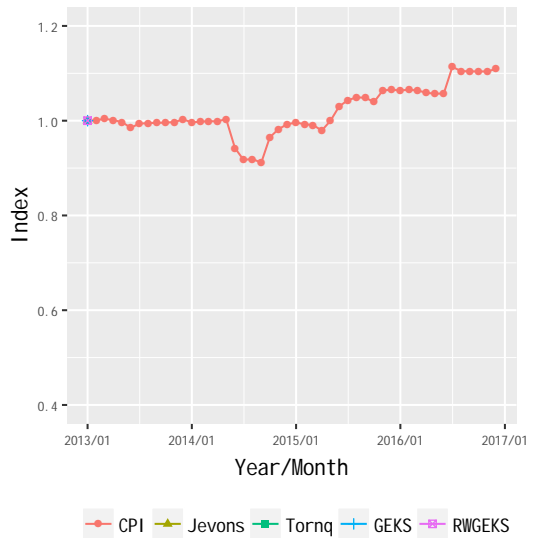
## 공통상품



## 특이항제거

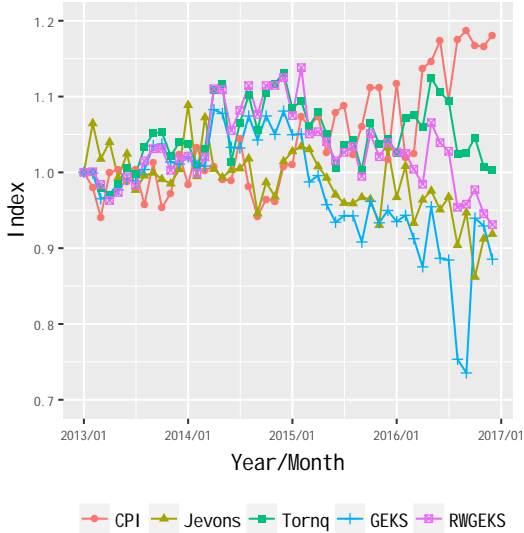


## 공통상품, 특이항제거

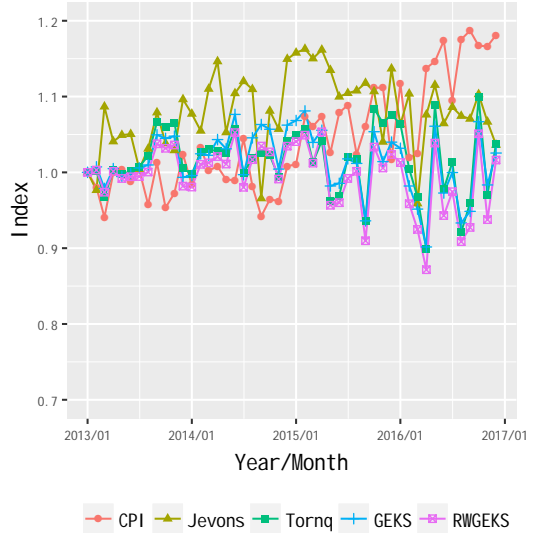


# L012010(면도기)

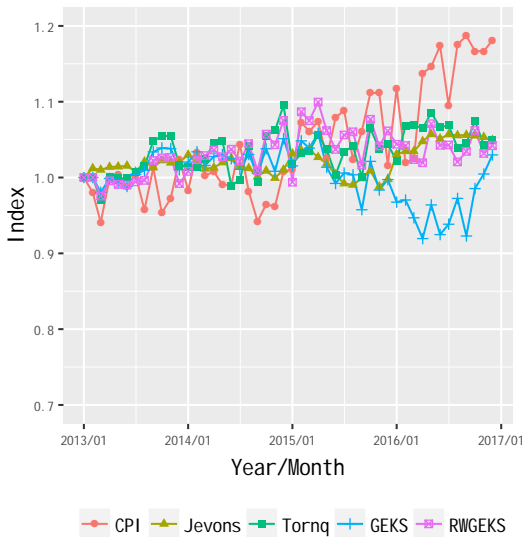
## 원데이터



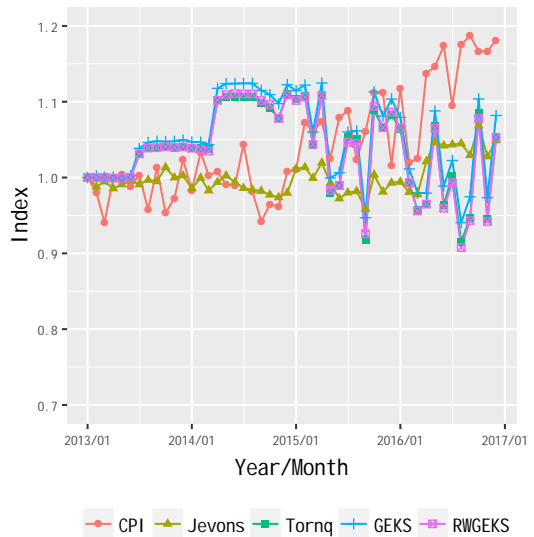
## 공통상품



## 특이항제거



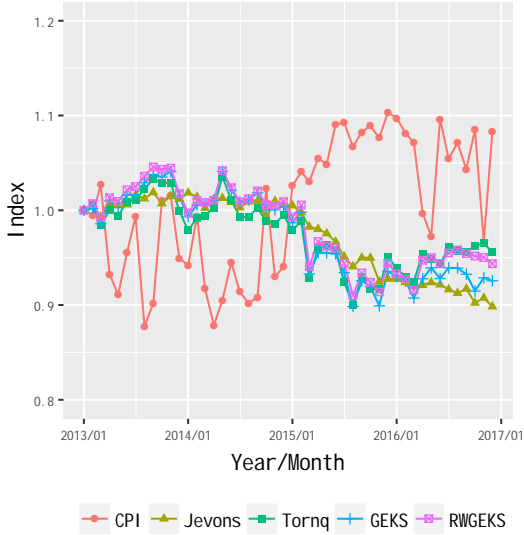
## 공통상품, 특이항제거



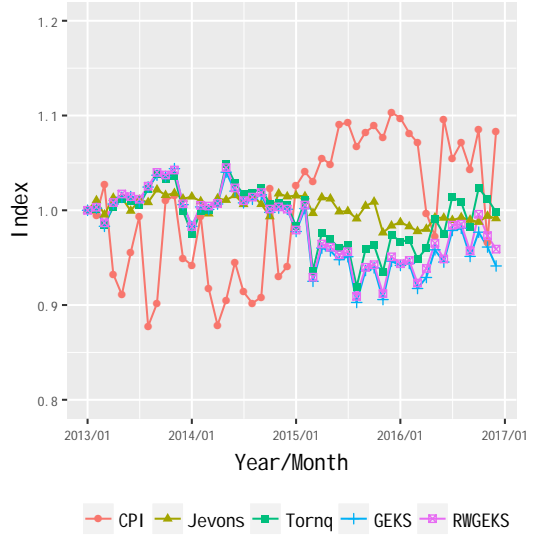


# L012030(치솔)

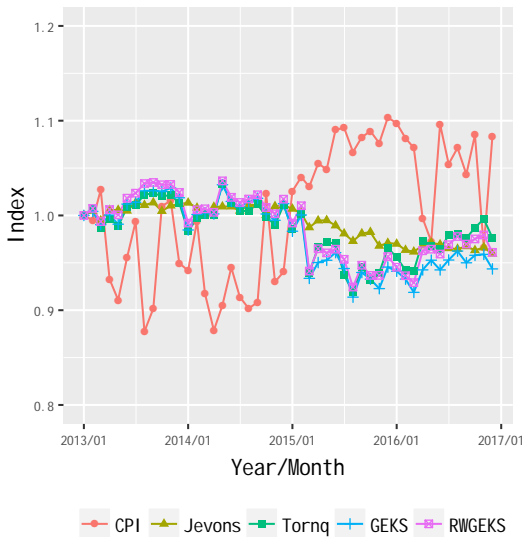
## 원데이터



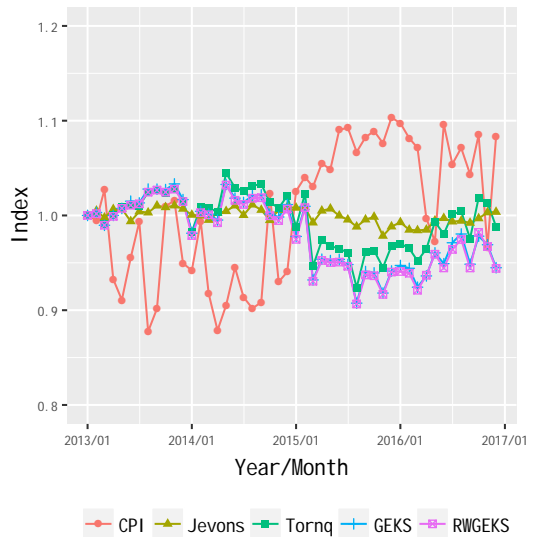
## 공통상품



## 특이항제거

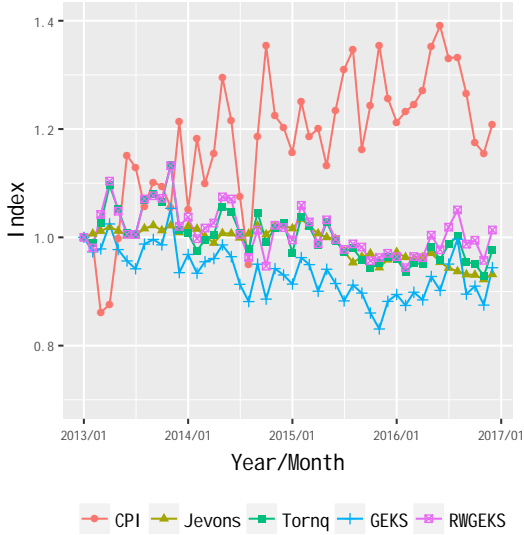


## 공통상품, 특이항제거

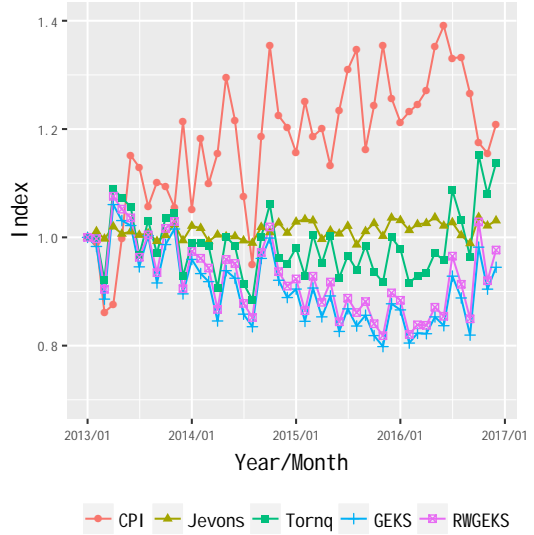


# L012040(치약)

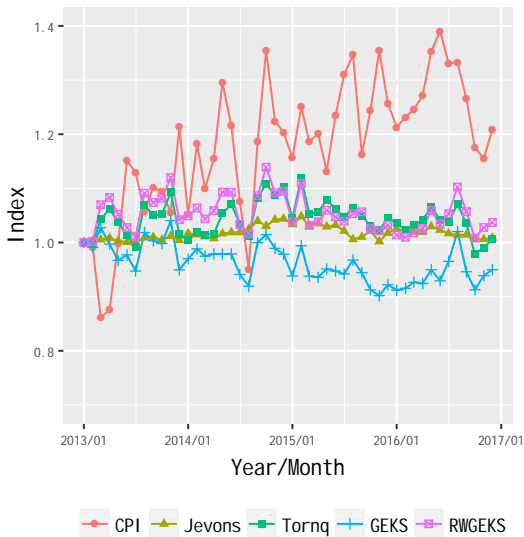
## 원데이터



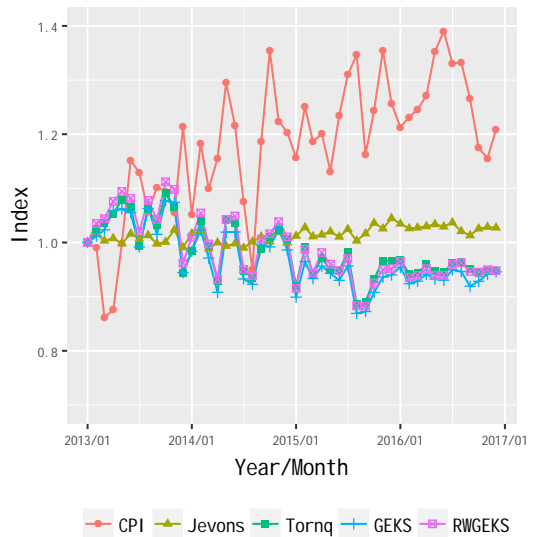
## 공통상품



## 특이항제거

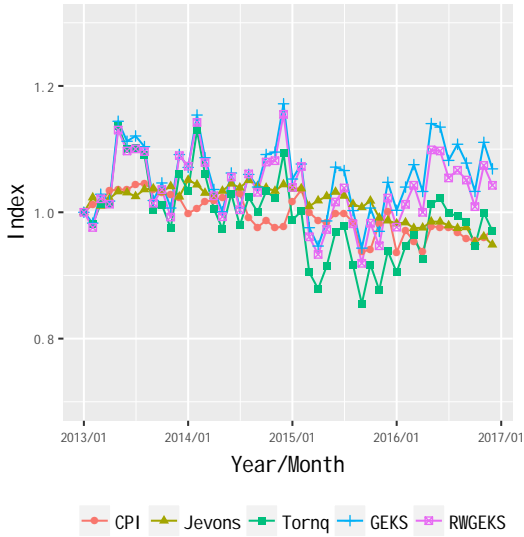


## 공통상품, 특이항제거

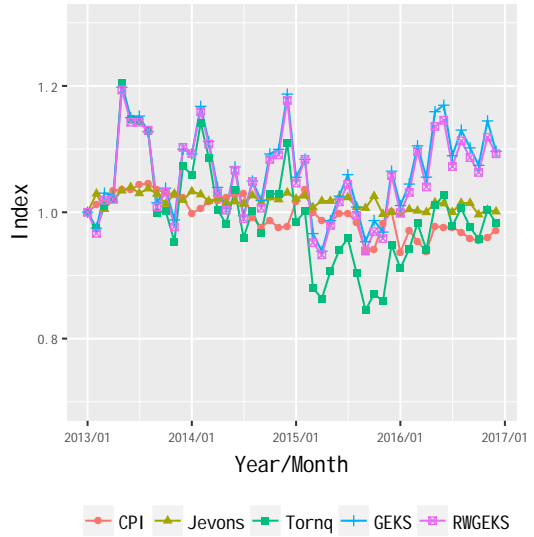


# L012050(비누)

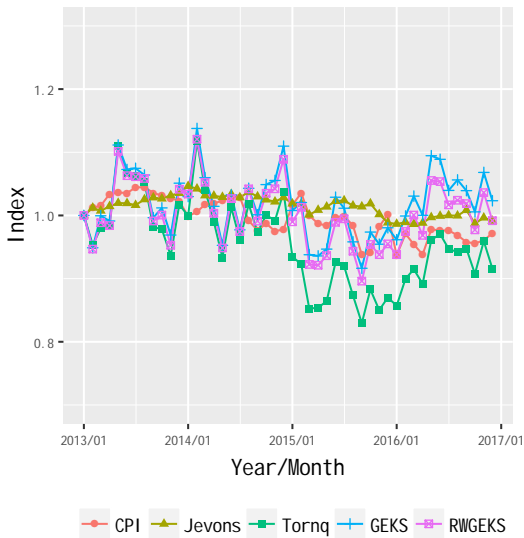
## 원데이터



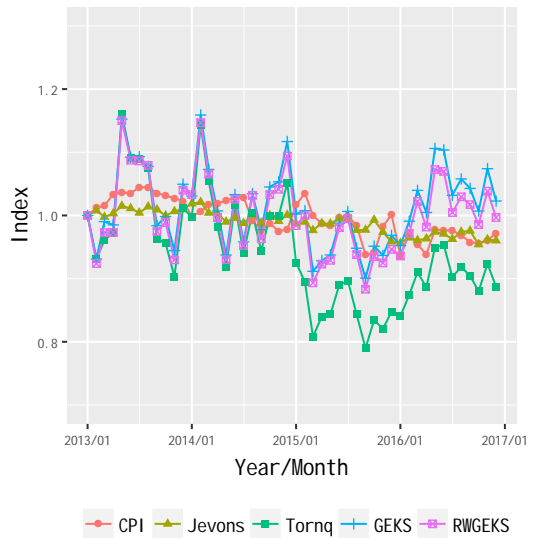
## 공통상품



## 특이항제거

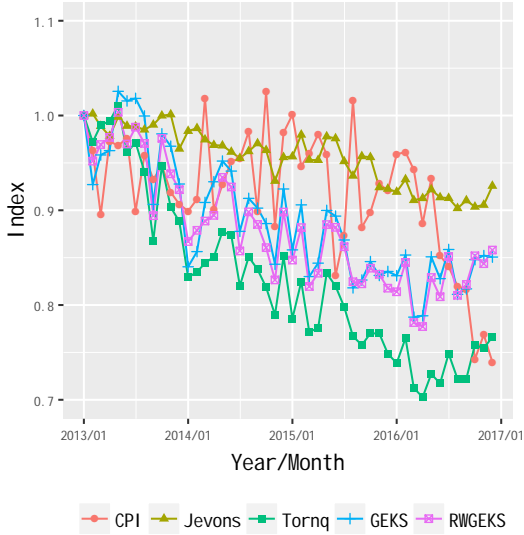


## 공통상품, 특이항제거

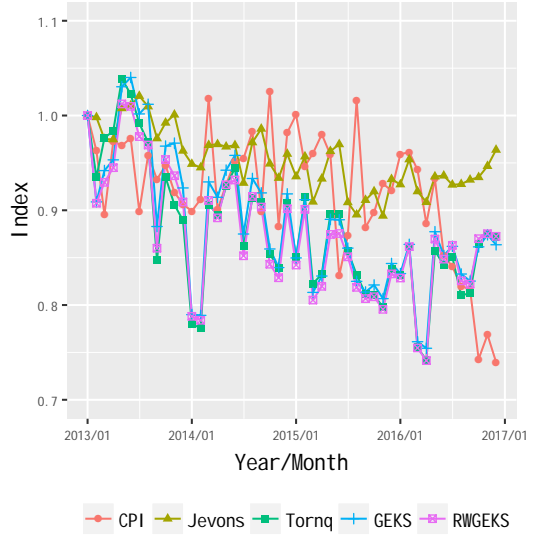


# L012060(삼푸)

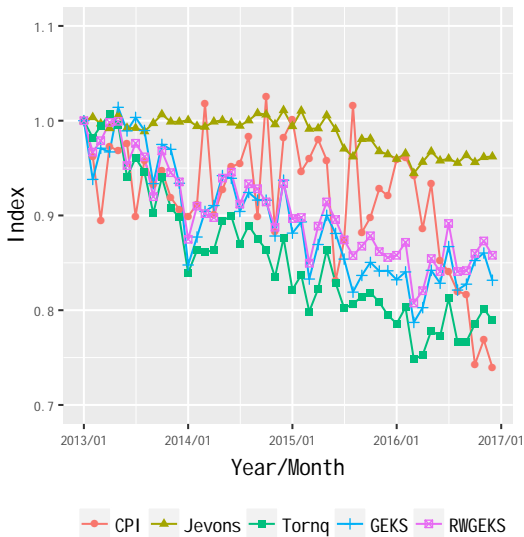
## 원데이터



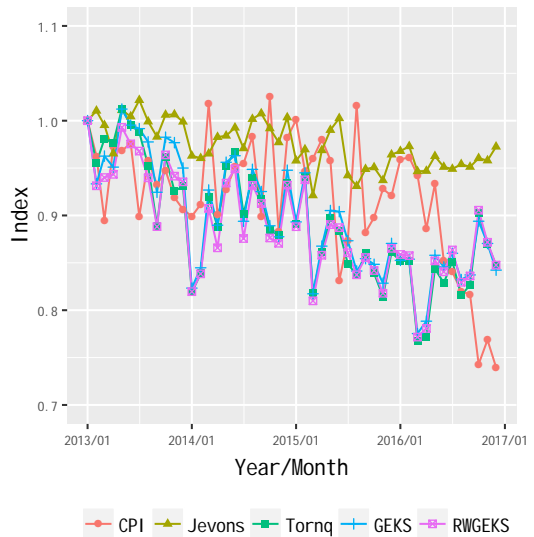
## 공통상품



## 특이항제거

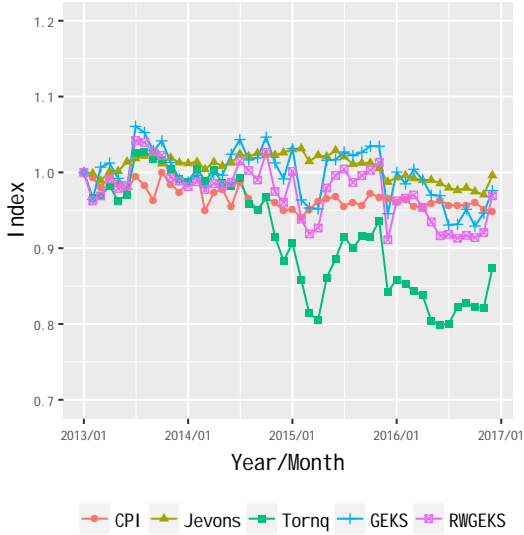


## 공통상품, 특이항제거

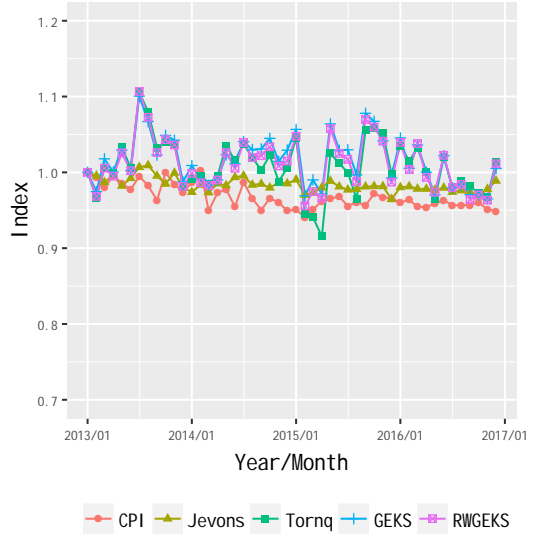


# L012080(화장지)

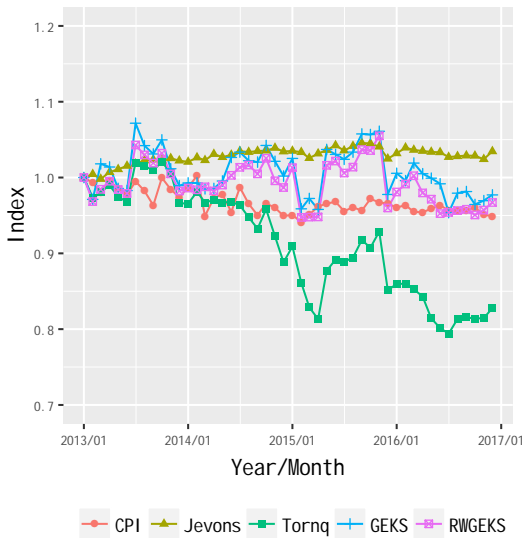
## 원데이터



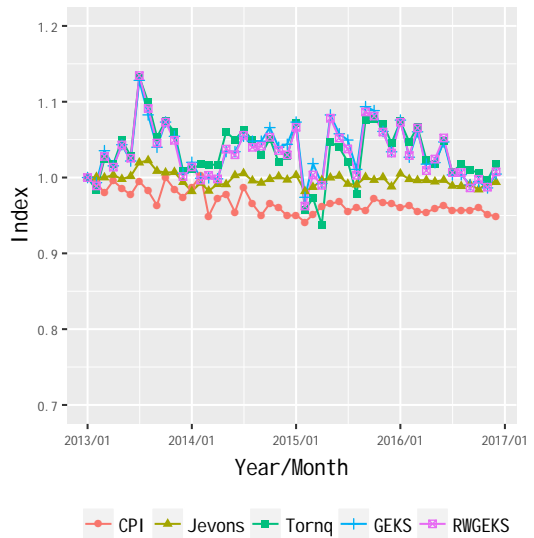
## 공통상품



## 특이항제거

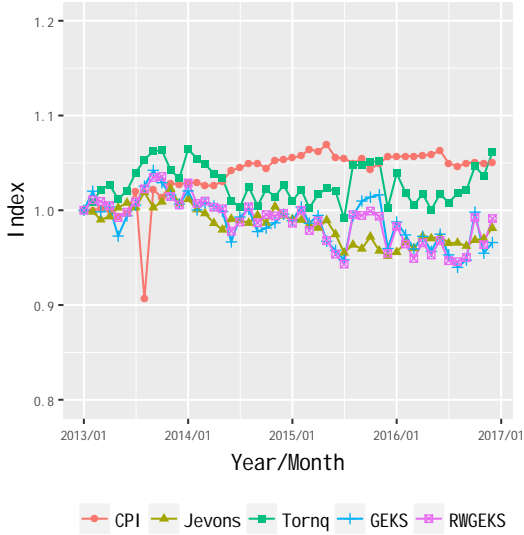


## 공통상품, 특이항제거

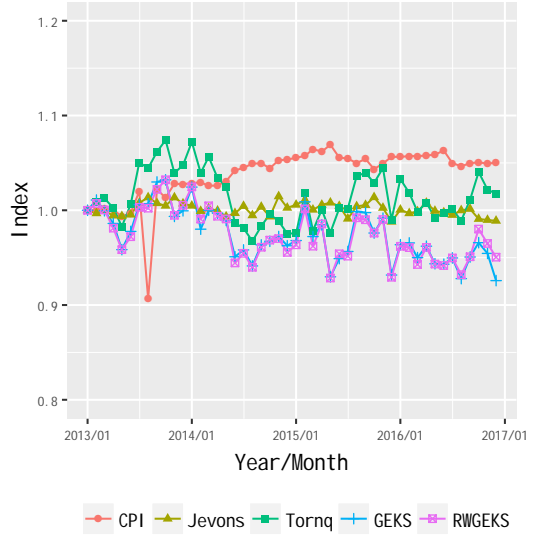


# L012090(기초화장품)

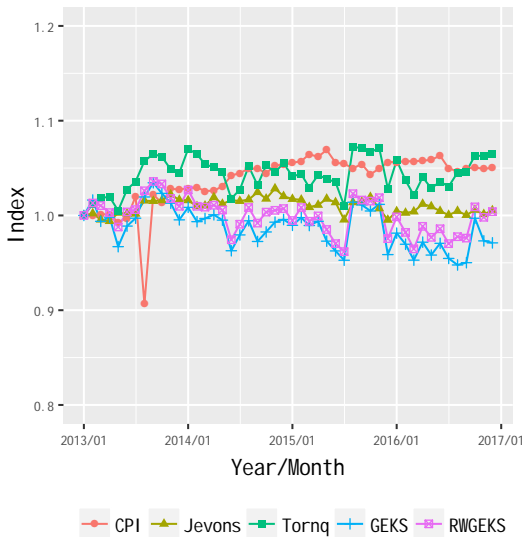
## 원데이터



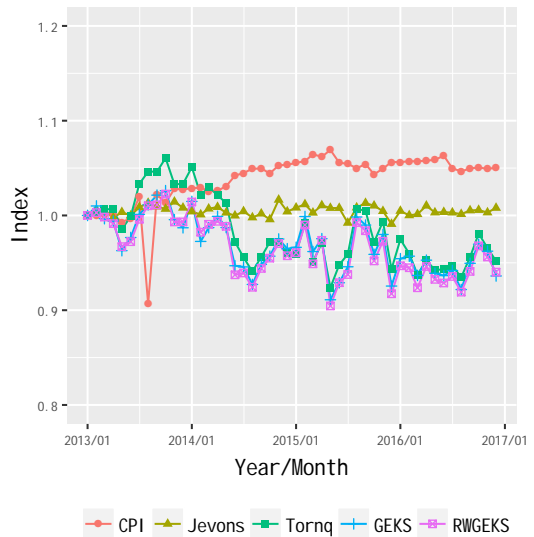
## 공통상품



## 특이항제거

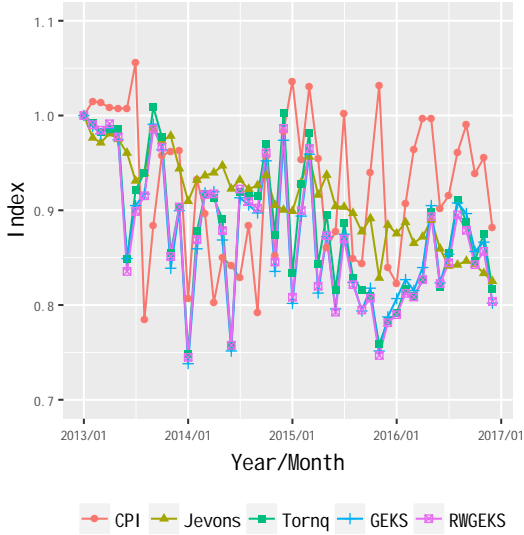


## 공통상품, 특이항제거

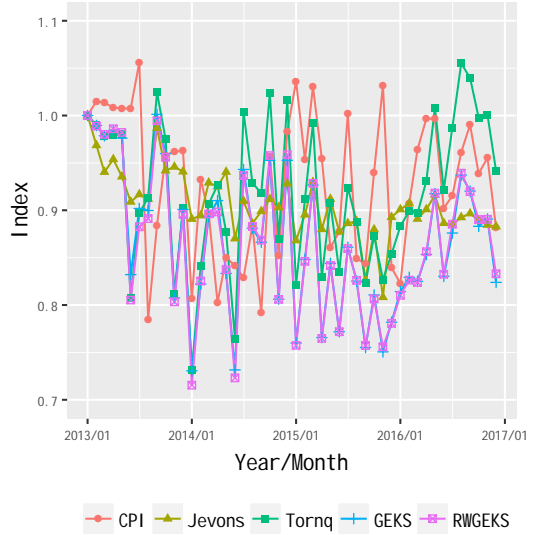


# L012120(모발염색약)

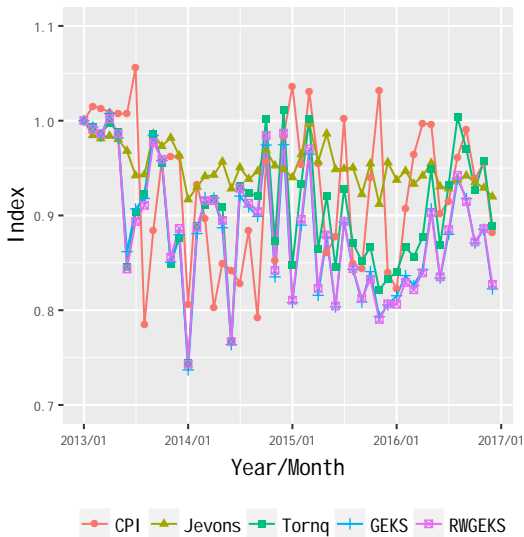
## 원데이터



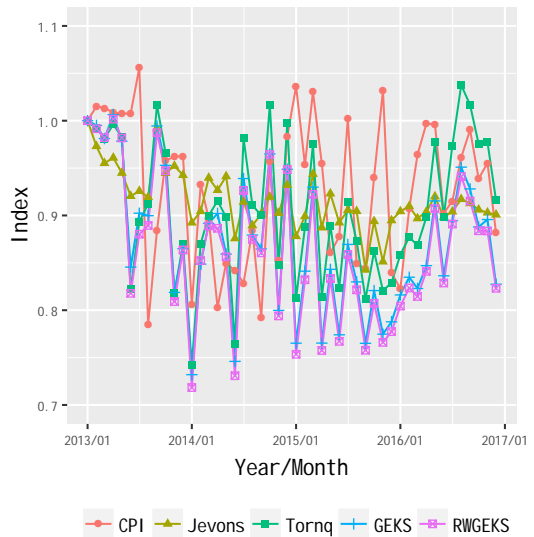
## 공통상품



## 특이항제거

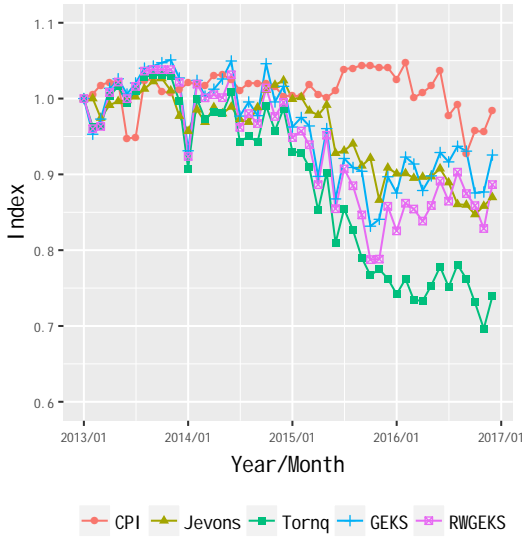


## 공통상품, 특이항제거

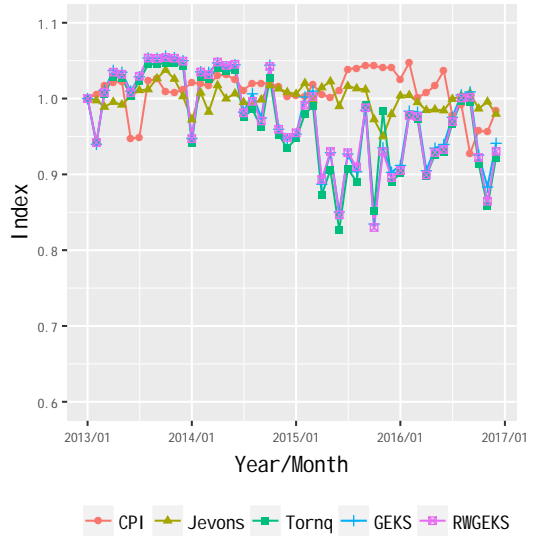


# L012130(구강세정제)

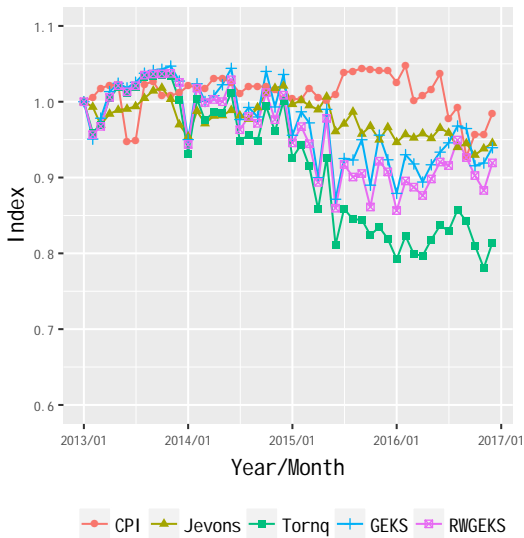
## 원데이터



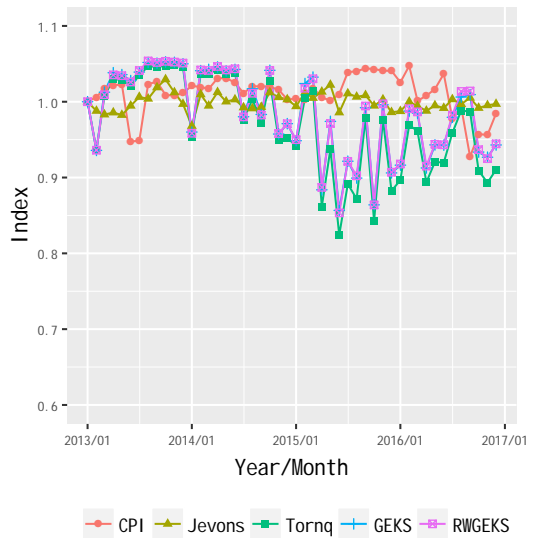
## 공통상품



## 특이항제거



## 공통상품, 특이항제거







**통계청**

(우)35208 대전광역시 서구 청사로 189(둔산동, 정부대전청사 3동)