

# 강의 1: 온라인 및 공식 가격 지수: 아르헨티나의 정보 측정 (Cavallo, JME, 2013)

## 거시경제학의 빅데이터 분석

조두연

양적응용경제학과

성균관대학교

# 개요

- 이 문서는 온라인 소매업체에서 수집한 가격이 라틴 아메리카 5개국의 공식 정보 추정치와 일치하는지 여부를 연구합니다.
  - ① 아르헨티나
  - ② 브라질
  - ③ 칠레
  - ④ 콜롬비아
  - ⑤ 베네수엘라
- 온라인 가격의 가용성은  
물가 지수의 구성 및 정보 측정 세계에서.
  - 온라인 소매업체에서 수집한 가격은 공식 통계를 보완하는 일일 가격 지수를 구성하는 데 사용할 수 있습니다.

## 개요(계속)

- 데이터는 2007년 10월부터 2011년 3월까지 각 국가에서 가장 큰 슈퍼마켓에서 수집되었습니다.
  - 이제 특수 소프트웨어를 사용하여 전례 없는 양의 미시적 가격 데이터를 수집할 수 있습니다.
  - 이러한 유형의 데이터는 원격으로 수집할 수 있으며, 훨씬 더 높은 주파수에서.
- 라틴 아메리카 5개국 중 아르헨티나의 경우 최근 몇 년 동안 공식 통계가 큰 비판을 받았습니다.
  - 브라질, 칠레, 콜롬비아 및 베네수엘라의 온라인 물가 지수는 공식 정보의 수준과 주요 역학을 모두 근사합니다. 대조적으로, 아르헨티나?? 온라인 in??
  - ation 비율은 거의 3배 이상 공식 추정치보다

# 소개

- 온라인 데이터로 구성된 물가 지수는 공식 추정치가 신뢰성을 상실한 국가에서 대체 정보 추정치를 얻는 데 사용할 수 있습니다.
  - 아르헨티나 정부는 그동안 공식 정보 지수 조작 2007년부터.
- i) 온라인 가격, ii) 표준 CPI 방법론 및 iii) 공식 카테고리 가중치의 조합이 '온라인 가격 지수'를 구축하는 데 사용됩니다. 각 국가에서.
  - 그런 다음 각 온라인 지수는 식품, 음료 및 가정용품(온라인 데이터에서 사용할 수 있는 동일한 범주)의 공식 CPI 구성요소의 가중 평균으로 구성된 동등한 공식 슈퍼마켓 지수와 비교됩니다.

## 소개(계속)

- 매칭은 칠레에서 가장 좋으며, 평균 연간 유입량은 다음과 같습니다.  
3.00% 온라인 및 3.19% oine 및 연간 정보 시리즈에서 0.97의 상관관계.
  - 경기는 콜롬비아에서도 연간 유입의 수준과 역동성 모두에서 가깝습니다.
  - 브라질과 베네수엘라에서 온라인 지수는 공식 지수의 주요 정보 추세와 일치할 수 있지만 연간 정보 시리즈는 시간이 지남에 따라 덜 동기화됩니다.
- 아르헨티나의 경우 온라인과 공식 가격 지수 사이의 큰 불일치 그것은 시간이 지남에 따라 지속됩니다.
  - 3.5년 이상 동안 온라인 가격은 지속적으로 2~3배 높다 공식 통계보다 표본 기간의 평균 비율이 20.14%입니다.

공식 데이터의 8.38%.

  - 놀랍게도 정보의 수준은 더 높지만 온라인 정보의 동적 동작은 공식 데이터와 매우 잘 일치합니다.

# 데이터

- 데이터는 MIT의 Billion Price Project에서 ??라는 기술을 사용하여 수집했습니다. 웹 스크래핑??아르헨티나, 브라질, 칠레, 콜롬비아, 베네수엘라에서 가장 큰 슈퍼마켓에서 2007년 10월부터 2011년 3월까지 온라인으로 판매된 모든 상품의 가격을 기록합니다.
- 웹 스크래핑
  - 대부분의 웹 페이지는 HTML(Hyper Text Markup Language)이라는 구조화된 코딩 언어를 사용하여 구축됩니다. 이 코드에는 다음과 같은 간단한 ??태그??가 있습니다. <센터> 그리고 <굵게>, 페이지에서 텍스트의 스타일과 배치를 결정합니다. 이러한 태그는 독특한 '모양과 느낌'을 제공하므로 시간이 지나도 일정하게 유지되는 경향이 있습니다. 각 페이지에. 대조적으로, 제품 가격
  - 과 같은 이러한 태그 내의 정보는 항상 변경됩니다. 스크래핑 소프트웨어는 HTML 태그를 사용하여 제품에 대한 관련 정보를 찾고 데이터베이스에 저장하도록 가르칠 수 있습니다.
  - 매일 프로세스를 반복하면 하루에 제품당 하나의 레코드가 있는 패널 데이터베이스가 생성됩니다.

## 데이터(계속)

- 모든 경우에 온라인 데이터에는 식품, 음료 및 가정용품의 조합이 포함됩니다. 카테고리 범위 ??계란?? '가전'으로, 그 중 약 3분의 1이 가정용품(청소제, 건강 및 미용 제품, 가구, 가전제품 및 책 포함)에 해당합니다.
- 이 범주는 28.44%(콜롬비아)와 이들 국가에서 CPI 가중치의 48.51%(아르헨티나).

**Table 1**  
Online data description.

	Argentina Retailer #1	Argentina Retailer #2 <sup>a</sup>	Brazil	Chile	Colombia	Venezuela
Starts	10/7/2007	23/7/2007	10/10/2007	10/24/2007	11/13/2007	04/16/2008
Ends	3/24/2011	03/20/2011	03/01/2010	03/20/2011	03/24/2011	03/01/2010
Prices P/day (mean)	11,560	4790	11,000	12,000	5000	9256
Total products	26,333	10,929	21,804	35,432	9166	20,847
Price changes	204,449	136,781	25,9875	12,0112	76,979	94,808
Category indicator	Yes	No	Yes	Yes	Yes	No
CPI weights covered	48.51%	–	27.93%	31.00%	28.44%	–
Retailer market share <sup>b</sup>	28%	n/a	15%	27%	30%	n/a

<sup>a</sup> Argentina's Retailer #2 is used only in the robustness results discussed in Section 5.3.2 and Fig. 4.

<sup>b</sup> Market shares are based on the information posted on the corporate webpages of each supermarket.

# 온라인 물가 지수

- 온라인 가격 지수는 온라인 데이터와 공식 카테고리 가중치의 조합을 사용합니다.
  - ① 일일 데이터는 온라인 가격 지수를 구성하는 데 사용됩니다.
  - ② 온라인 인덱스는 사용 가능한 모든 제품의 가격을 사용하여 작성됩니다.  
각 판매점에서 구입하십시오.
  - ③ 강제적인 제품 교체나 품질 변경을 위한 조정은 없습니다.



## CBA 지수 가중치

- CBA 바구니는 제품당 주어진 수의 그램으로 구성됩니다. 색인에는 45개의 제품이 포함되어 있으며 표 C.2에 자세히 설명되어 있습니다. 이러한 제품과 해당 그램은 30-59세 범위의 성인 남성에게 대한 최소 영양 요구 사항을 충족하도록 INDEC에서 설정합니다. INDEC 목록에 있는 제품과 일치하도록 신중하게 선택한 스크랩 데이터에서 45개 항목을 사용하여 바구니의 일일 비용을 계산합니다.

Table C.2: CBA Index Weights

INDEC		Scraped Data		
Product (1)	Grams (2)	Product (3)	Package size in grams (4)	Index Weight (2)/(4)
Pan	6,060	Pan	500	12.12
Galletitas Saladas	420	Galletas Saladas	1000	0.42
Galletitas Dulces	720	Galletas Dulces	160	4.50
Arroz	630	Arroz Largo Fino	500	1.26
Harina De Trigo	1,020	Harina De Trigo	1000	1.02
Otras Harinas (Maz)	210	Harina Maiz	500	0.42
Fideos	1,290	Fideos	1000	1.29
Papa	7,050	Papa Negra	1000	7.05
Batata	690	Batata	1000	0.69
Azcar	1,440	Azucar	1000	1.44
Dulce De Leche	80	Dulce De Leche	250	0.32
Dulce De Batata	80	Dulce De Batata	1000	0.08
Mermeladas	80	Mermelada De Frutilla	454	0.18
Legumbres Secas				
Lentejas	80	Lentejas	500	0.16
Porotos	80	Porotos	500	0.16

# CBA 지수 가중치(계속)

Arvejas	80	Arvejas Verdes	350	0.23
Cebolla	655	Cebolla	1000	0.66
Lechuga	655	Lechuga Francesa	1000	0.66
Tomate	655	Tomate	1000	0.66
Zanahoria	655	Zanahoria	1000	0.66
Zapallo	655	Zapallo	1000	0.66
Tomate En Lata	655	Tomate Perita	400	1.64
Manzana	2,010	Manzana	1000	2.01
Naranja	2,010	Naranja De Jugo	1000	2.01
Asado	896	Asado Centro Novillito	1000	0.90
Carne Picada	896	Carne Picada	1000	0.90
Carnaza	896	Carnaza Comun	1000	0.90
Cuadril	896	Colita De Cuadril	1000	0.90
Nalga	896	Nalga	1000	0.90
Paleta	896	Paleta De Cerdo	1000	0.90
Pollo	896	Pollo Sin Piel	1000	0.90
Huevos	630	Huevo	300	2.10
Leche	7,950	Leche Entera	1000	7.95
Fresco	90	Queso Crema	200	0.45
De Rallar	90	Queso Rallado	210	0.43
Crema	90	Queso Cremoso	1000	0.09
Aceite	1,200	Aceite Mezcla	1000	1.20
Bebidas Edulcoradas	4,050	Coca Cola	2000	2.03
Bebidas Gaseosas Sin Edulcorar	3,450	Soda	2000	1.73
Sal Fina	150	Sal Fina	500	0.30
Sal Gruesa	90	Sal Gruesa	500	0.18
Vinagre	90	Vinagre Alcohol	1000	0.09
Caf	60	Cafe Molido	500	0.12
T	60	Te	50	1.20
Yerba	600	Yerba	500	1.20

# 인덱스 계산

- 가격 변동은 제품 수준에서 계산된 다음 가중되지 않은 기하 평균을 사용하여 범주 내에서 평균을 낸 다음 가중 산술 평균을 사용하여 범주 전반에 걸쳐 집계됩니다.
- 첫 번째 단계는 가격의 가중되지 않은 기하학적 평균을 구하는 것입니다.  
카테고리 변경 제*i*/매일  $NS$ :

$$NS_{제i/NS, NS1} = \prod_{NS} \frac{NS_{제i/NS, NS1}^{!1}}{NS_{NS1}}$$

- 어디  $NS_{NS}$  좋은 가격이다  $NS \sim$ 에 시각  $NS$ , 그리고  $N$ 의 수입니다  
카테고리의 제품 제*i*/그 날 샘플에 존재합니다.

## 인덱스 계산(계속)

- 두 번째 단계는 범주 수준 지수를 계산하는 것입니다.  $NS$ :

$$NS_{제이} = \frac{NS_{제이,0}}{NS_{제이,1} \dots NS_{제이,NS-1}}$$

- 마지막으로 슈퍼마켓 지수는 모든 범주 지수의 가중 산술 평균입니다.

$$NS_{NS} = \sum \frac{승_{제이}}{NS_{제이}}$$

- 어디  $승_{제이}$  해당 카테고리에 대한 공식 CPI 가중치이며  $제이$  샘플에 포함된 모든 가중치의 합입니다.
- CPI 통계와 비교할 때 이러한 온라인 지수는 다음과 같은 측면에서 이점이 있습니다. 빈도 및 항목 수 각 범주 내에서 샘플링됩니다.



# 일별 데이터를 사용한 월간 및 연간 정보(계속)

- 마찬가지로, 연간 유입률은 같은 기간 동안 지난 30일 동안 지수의 백분율 변화로 계산됩니다.

며칠 전에,

$$\begin{array}{rcl}
 & 2 & \text{삼} \\
 & \frac{1}{30} \sum_{NS=0}^{29} NS_{\text{텍사스}} & \\
 \frac{\phi_i / \phi_0}{NS} = 100 & \frac{6641}{30} \sum_{NS=365}^{394} NS_{\text{텍사스}} & \\
 & \frac{175}{5} & 
 \end{array}$$

- 매년 마지막 날,  $\phi_i / \phi_0$   $NS$ 와 비교된다  
공식 통계에 보고된 연간 도입률.

## 결과: 브라질, 칠레, 콜롬비아, 베네수엘라

- 온라인 데이터러를 표준 공식 범주로 분류할 수 없는 베네수엘라를 제외하고 모든 지수는 공식 CPI 가중치를 사용하여 가중치가 부여됩니다.
  - 이 경우 가중치가 없는 지수는 매일 관찰된 모든 가격 변동의 단순 기하 평균으로 구성됩니다.
- 그림 1은 온라인 인덱스의 놀라운 능력을 보여줍니다. 주요 정보 동향 추적 오랜 기간 동안.
- 표 2는 해당 기간의 평균 연간 주입률이 모든 국가에서 거의 동일함을 보여줍니다.
- 국가 간 온라인-공식 매칭의 차이는 다음과 관련이 있는 것으로 보입니다. 소매상의 대표성

온라인 데이터에 사용됩니다.

## 결과: 4개국(계속)

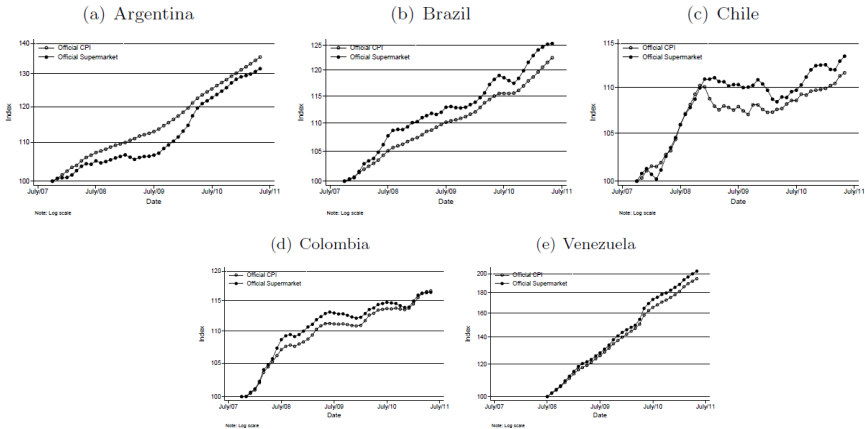
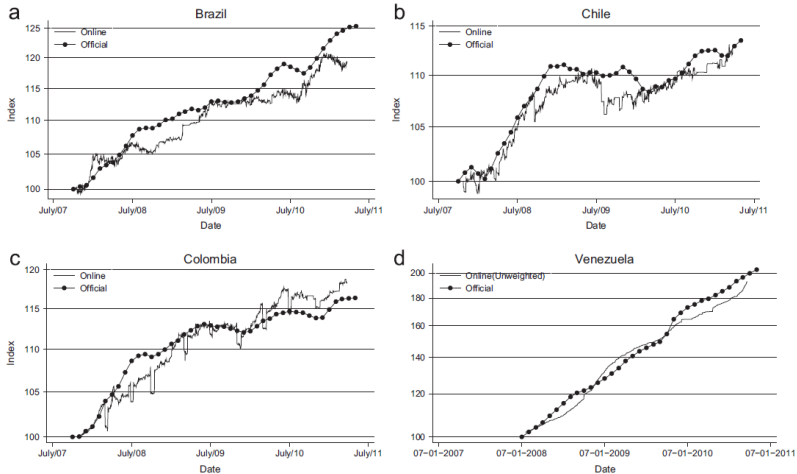


Figure B.3: Official CPI and Official Supermarket Indexes

Notes: The official supermarket index is constructed as a weighted average of the Food and Beverage and Household Product official price indexes in each country.

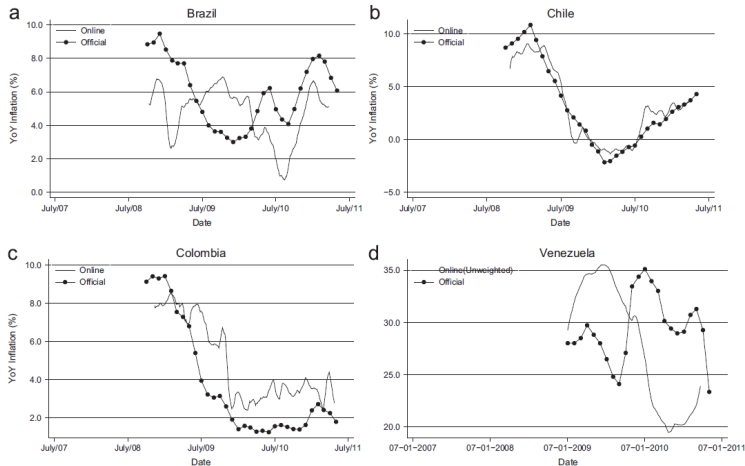


## 결과: 4개국(계속)



**Fig. 1.** Online and official price indexes in four Latin American countries. *Notes:* The daily online supermarket index is constructed with an online prices and official CPI category weights. In Venezuela, the online index is built as a geometric average of all price changes observed each day. The official supermarket index is an equivalent indicator constructed as a weighted average of the “Food and Beverages” and “Household Products” official price indexes in each country.

## 결과: 4개국(계속)



**Fig. 2.** Online and official annual inflation rates. *Notes:* The annual online inflation rate is a daily time series computed as the percentage change in the average of the index during the previous 30 days with respect to the average of the index in the same period a year before. The annual official inflation rate is a monthly time series computed as the percentage change in the index in the previous 12 months.

# 결과: 4개국(계속)

**Table 2**

Online vs. official series.

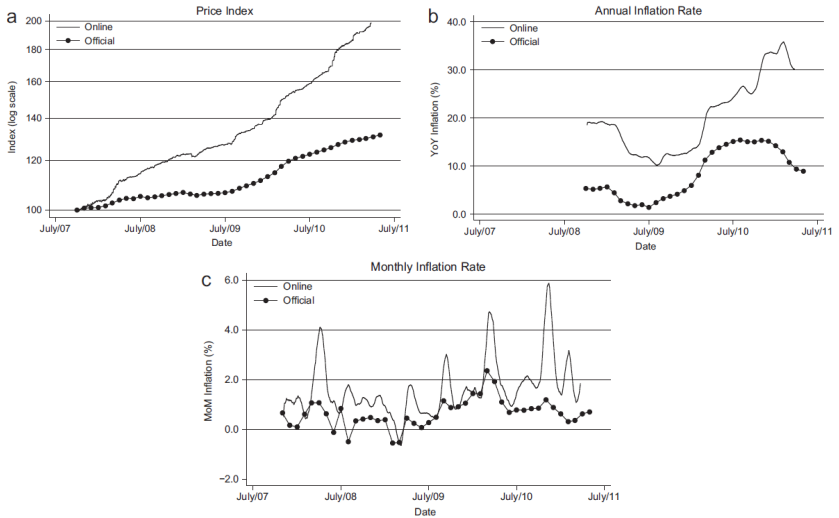
	Argentina	Brazil	Chile	Colombia	Venezuela
<i>Mean annual inflation (%)</i>					
Official CPI index	8.53	5.28	2.44	3.79	27.37
Official supermarket index	8.38	5.91	3.19	3.73	29.38
Online supermarket index	20.14	4.72	3	4.88	27.43
<i>Correlations between online and official supermarket series</i>					
Price index	0.98	0.96	0.97	0.95	0.92
Annual inflation	0.84	0.09	0.97	0.89	-0.08
Monthly inflation	0.6	0.5	0.38	0.43	0.18
<i>Regression – official supermarket on online monthly inflation rates (12 lags)</i>					
Constant	0.84	-0.54	0.14	0.03	-1.96
Constant p-value	0.000	0.19	0.17	0.86	0.23
R2	0.9	0.55	0.66	0.59	0.66
<i>Monthly inflation rate volatility (standard deviation)</i>					
Official supermarket index	0.57	0.51	0.58	0.48	0.98
Online supermarket index	1.11	0.73	0.62	0.76	0.86

*Note:* The top panel shows that Argentina is the only country where online data does not approximate the average official annual inflation rates. However, the second panel shows that the correlation between the monthly inflation rates is higher in Argentina than in any of the other countries. The discrepancy is therefore in the level of inflation reported, not its dynamic behavior over time. The third panel reinforces this idea with a simple OLS regression of the official monthly rate and 12 lags of the online monthly rate. The R2 is highest in Argentina, which is also the only country where the constant is statistically significant. The fourth panel shows that the official monthly inflation rate is surprisingly stable in Argentina compared to both the online index and the volatility observed in a high-inflation country like Venezuela.

# 결과: 아르헨티나

- 통계청에 대한 정부 개입
  - 2003년 이후, Argentina's inflation은 소비를 자극하고 통화의 평가를 피하기 위해 고안된 확장 통화 정책의 결과로 꾸준히 성장했습니다.
  - Inflation이 되었다 정치적으로 민감한 문제 2006년, 연간 inflation 비율이 12% 이상 증가할 때.
  - 아르헨티나 정부는 2007년 1월부터 국가통계연구소(INDEC)에 개입해 왔다.
- 공식 데이터와의 큰 차이
  - 아르헨티나의 경우 온라인 물가 지수는 그림 3과 같이 공식 지수와 완전히 다른 추세를 따릅니다. 2007년 10월과 2011년 3월 사이에 온라인 지수는
  - 100% 이상 증가한 반면 official 지수는 35%만 성장했습니다.
- 주요 결과
  - ① 첫째, 온라인 정보는 지속적으로 2번과 3번 더 높은 공식 정보보다.
  - ② 둘째, 온라인 및 공식 견적 점유율 의외로 비슷한 패턴 시간이 지남에 따라.

## 결과: 아르헨티나(계속)



**Fig. 3.** Online supermarket index in Argentina. *Notes:* The monthly online inflation rate is a daily time series computed as the percentage change in the average of the index in the last 30 days with respect to the average of the index in the same period a month before. The monthly official inflation rate is a monthly time series computed as the percentage change in the index over the previous month.

# 견고성: 셀 상대적 대치 및 가중치 없음 색인

- 첫 번째 연습에서는 가격 계열 내에서 결측값을 대치하기 위해 다른 접근 방식을 사용합니다.
  - 논문에서 사용된 방법은 누락된 가격을 각 제품에 대한 마지막 사용 가능한 가격으로 찾는 것입니다. 이 접근 방식은 온라인 데이터의 격차가 며칠만 지속되기 때문에 합리적입니다.
  - 그러나 공식 통계 부서는 월별 시리즈의 결측값을 다른 방식으로 처리합니다. 아르헨티나에서도 사용되는 표준 접근 방식은 누락된 가격을 유사한 제품의 평균 가격 변동으로 귀속시키는 것입니다.
- 대신, 우리는 노동 통계국(BLS)에서 사용하는 '세포 기준' 접근 방식을 따릅니다. 특정 날짜에 제품이 누락된 경우 해당 날짜를 계산하는 데 해당 제품을 사용하지 않습니다. 단, 이전 가격에 당일 동일한 카테고리의 제품 평균 가격 변동을 곱한 값과 동일한 가격으로 귀속됩니다.

# 견고성: 셀 상대적 대치 및 가중치 없음

## 색인(계속)

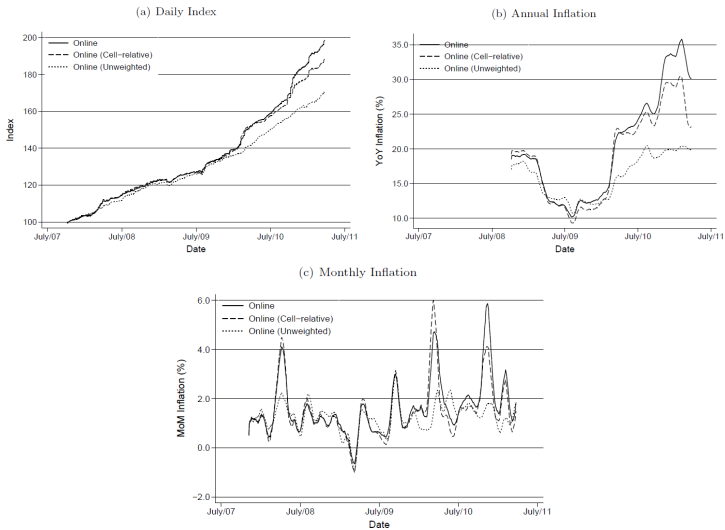


Figure B.1: Robustness: Cell-Relative Imputation

## 견고성: 가장 낮은 항목

- 이 섹션은 극단적인 가정을 사용합니다. 정부는 좁게 정의된 범주 내에서 가장 낮은 도입율 이 전체 기간 동안.
- 어떤 상품이 가장 낮은 금리를 가질지 사전에 알기가 어렵지만 적어도 그것은 INDEC에 대한 더 낮은 금리를 제공할 수 있다는 점을 감안할 때 이것은 INDEC에 대한 현실적인 대안이 될 수 없습니다. 낮은 가격의 브랜드나 아이টে을 선호하는 전략.



# 견고성: 가장 낮은 항목(계속)

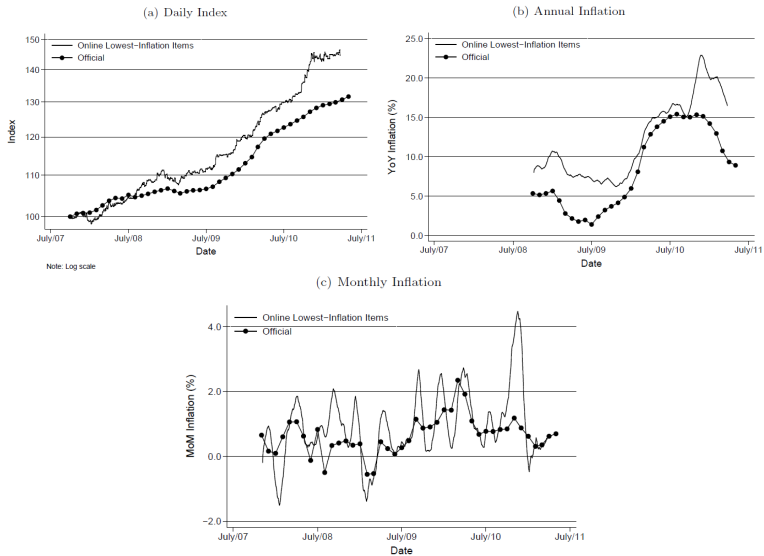


Figure B.2: Lowest-Inflation Items within URLs

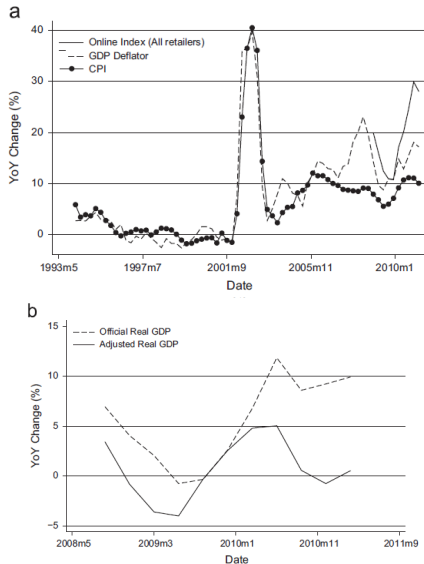
## 빈곤에 대한 시사점

- 정보 추정치의 편향은 다음과 같은 다른 통계에도 영향을 미칩니다.  
빈곤과 실질 GDP 추정치.
- 분기마다 INDEC는 CBA 바스켓 비용을 사용하여 극빈 상태에 있는 개인 수를 확인합니다. 2008년 1분기의 공식 바스켓 비용을 취하여 CBA 계산으로 조정 온라인에서 관찰된 환율로 2011년 7월 바구니의 가격은 \$259.5 아르헨티나 페소입니다.
- INDEC의 소득 조사를 사용하여 이것은 다음을 의미합니다. 6.69% 당시 인구의 100%가 극빈층이었다.  
공식 통계에 보고된 2.5%.
  - 마찬가지로 CBA를 조정하여 더 넓은 범위의 ??Canasta Basica Total?? (CBT) 장바구니에 비식량 품목을 추가하면 빈곤 수준이 됩니다. 25.9% 공식적으로 보고된 9.9% o4와 비교됩니다.

# 실질 GDP에 대한 시사점

- 실질 GDP에 대한 영향을 추정하기 위해 CPI와 GDP deflator가 지난 10년 동안 어떻게 행동했는지 살펴보는 것으로 시작합니다.
- 1994년과 2006년의 데이터는 두 시리즈가 밀접하게 상관되어 있음을 보여줍니다. 이는 정상적인 조건에서 예상되는 동작입니다.
  - 그러나 2007년 이후 GDP deflator는 CPI보다 훨씬 빠르게 증가했습니다. 이것은 정부가 GDP deflator에 있는 더 높은 inflation을 인식했다는 것을 의미합니다.
  - 온라인 지수를 이용한 '수정 실질 GDP'의 연간 성장률은 현저히 낮아짐 공식 추정치보다 (참고로 성장률은 온라인 지수와 같은 속도로 디에이터가 증가했다고 가정)

# 실질 GDP에 대한 시사점(계속)



**Fig. 7.** Implications for real GDP growth: (a) GDP deflator, CPI, and online index – annual change and (b) Real GDP – annual growth rate. *Notes:* The GDP deflator and CPI series in (a) co-moved closely together from 1994 to 2006, but started to deviate from 2007 onwards. Although higher than the CPI, the GDP deflator still has less inflation than the online index in the past 3 years. Assuming the deflator had increased at the same rate as the online index, then we can compute an “Adjusted Real GDP” with a growth rate that is significantly lower than in official estimates.

## 결론

- 온라인 데이터와 공식 방법의 조합을 사용하여 구성된 온라인 가격 지수는 다음을 수행할 수 있습니다. 공식 정보의 수준과 주요 역학 모두 일치 브라질, 칠레, 콜롬비아 및 베네수엘라에서.
- 매칭은 연간 빈도에서 가장 잘 이루어지며 시장 점유율이 큰 슈퍼마켓과 국가 전체를 더 많이 대표합니다.
- 아르헨티나의 경우 실제로 온라인에서 2배 3배 더 크다 3년 이상 공식 추정치보다
- 온라인 물가 지수는 다음을 제공할 수 있습니다. 실제 비율에 대한 좋은 근사치 나라 안에서.