Overlapping Quality Adjustment using Online Data

Preliminary Results

Alberto Cavallo
MIT
Outline

• Scraped Online Data
  • Introduction
  • Advantages

• Overlapping Quality Adjustments
  • CPI data vs Online Prices
  • Example: US Televisions
Goods & services online

- Food and Beverages
- Electronics
- Household Products
- Apparel
- Health, Beauty & Drugs
- Transportation (airline, trains, taxis)
- Gas/Fuel/Electricity
- Real Estate
- Labor

➢ ~ 60% of CPI weights can now be covered with online data
Online prices behave like offline prices

- In some countries (e.g., US) online and offline markets are closely integrated → similar price levels & pricing strategies
- Often online prices have a markup over offline data, but it tends to be constant over time → implies similar inflation rates
- In countries where online markets are very small, online and offline prices can be identical because firms do not have a differentiated pricing strategy online (they simply show their offline database in their website)
Online prices as a new source of Price Index data

Disadvantages:
• Relatively few retailers
• No quantity data → expenditure weights?
• Only ~60% CPI categories (most services are not yet available online)

Advantages:
• Low cost data collection
• Prices are available in real-time
• Data can be collected remotely in ~70 countries
• Many product details: brand, package size, sale indicator, price control, etc.
• High-frequency (daily)
• Information on *all* products sold in each retailer (i.e. census within retailer)
• New products sampled automatically from the day they are introduced until they disappear from the store
• Retailers group products into narrow categories or “urls” → close substitutes are easy to identify
Billion Prices Project at MIT

- Joint work with Roberto Rigobon
- Started collecting online prices in 2007
- Daily prices from ~1000 retailers in 50 countries
Daily Price Indices

- Available for 21 countries:
  - USA, Argentina, Australia, Canada, Chile, China, Colombia, France, Germany, Greece, Ireland, Italy, Japan, Korea, Netherlands, Russia, South Africa, Spain, UK, Uruguay, and Venezuela

- US and Argentina indices are publicly available online

- Daily series published with a 3-day lag

- Use standard CPI techniques and weights, but we don’t make explicit quality adjustments.
US Daily Price Index

US Annual Inflation

Source: BPP – PriceStats – BLS (CPI-U, US city-average, all items, NSA)
US Monthly Inflation

Source: BPP – PriceStats – BLS (CPI-U, US city-average, all items, NSA)
France - Monthly Inflation

Source: BPP – PriceStats – BLS (CPI-U, US city-average, all items, NSA)
Greece – Monthly Inflation

Source: BPP – PriceStats – BLS (CPI-U, US city-average, all items, NSA)
The Problem of Quality Adjustment

Traditional CPI Data

Ohta and Griliches (76): “...what the hedonic approach attempted was to provide a tool for estimating ‘missing’ prices.”

Turvey (1999): Advocates a chained geometric index of matched observations, incorporating entrys and exits immediately, with a sample size that varies over time.
Overlapping Quality Adjustment

From the ILO/IMF/OECD/… CPI Manual:

“When there is overlap, simple linking...may provide an acceptable solution to the problem of ... quality change”.

“in practice, however, this method is not used very extensively because the requisite data are seldom available”

“the information needed for this...will never be available if price collectors ... introduce a new quality when an old one is dropped”
Traditional CPI data
Online Data

The graph shows the price of three models over 12 months. The price of each model decreases each month, with some variation in the rate of decline. Model 1 has the steepest decline, while Model 3 has a more gradual decline. The graph indicates that by the end of the 12th month, all models have a price below the starting price.
Online Data

![Online Data Chart]

- **Price** vs **Month**
- Models: Model 1, Model 2, Model 3, Model 5, Model 6, Model 7

Legend:
- Model 1: Blue Line
- Model 2: Red Line
- Model 3: Green Line
- Model 5: Purple Line
- Model 6: Teal Line
- Model 7: Orange Line
How online data can help

• New products are immediately priced in the data
• Large number of overlapping products
• Retailers automatically identify close substitutes (via url)
• High frequency:
  – Detect new products sooner
  – Easier to identify sales / spikes / anomalies
Literature on Matched-Model vs Hedonic Indices

• Silver & Heravi (99): Scanner data. “Matched approach as a special case of the theoretically based SEHI approach. Caution is however advised when the loss of data in matching is severe. “

• Aizcorbe, Corrado & Doms (2003): matched indices can yield similar numerical results to dummy-variable hedonics when:
  • Panel of prices with constant quality
  • “High frequency” to minimize importance of “unmatched” models
Example With US TVs

• Largest electronics retailer in the US in 2009
• Daily panel data for 544 LCD televisions
• Sold from April 2008 to November 2009 (577 days)
Product Life (duration)

Mean Product life: 178 days (median 154)

Note: Excluding products lasting less than 30 days (refurbished or open box)
Changes per Product

Mean Price Changes per id: 6  (median 4)
Share of “Exits”
Alternative Price Indices

- CPI – Televisions (BLS, with hedonic adjustment)
- Online Index – with “traditional CPI sampling”
  - Small sample (few varieties)
  - Censored price spells (missing introduction prices)
    1. Comparable (treated as same good) → all P change
    2. Non-Comparable (treated as different goods) → al Q change
- Online Index - Small sample
- Online Index
- Online Index with daily data
Online Index - Comparables

- Un-weighted geometric mean of price relatives (Jevons index)
- Monthly data (1st day of each month)
- Censored Spells
  - Comparables: start with all products available on the first month, then replace them as they disappear with the closest model (price and characteristics).
  - Treated as same good → all price change
- Small sample: only “Samsung” brand (64 TVs)
Example of a matched model

![Graph showing price changes over dates with two lines, one for id = 16889102226 and another for id = 16889102248.]
Matched models
Online Index – Comparables

Annual Inflation

- Online Index – Substitutions – All Price Change
- US CPI – Televisions
Online Index – Comparables

Monthly Index

- Online Index – Substitutions – All Price Change
- US CPI – Televisions
Online Index with Non-comparables

- Un-weighted geometric mean of price relatives (Jevons)
- Monthly data (1st day of each month)
- Censored Spells
  - **Non-comparables**: start with all products available on the first month, then substitute them as they disappear with the closest model (price and characteristics).
  - Treated as different goods → all quality change
- Small sample: only “Samsung” brand (64 TVs)
Online Index with Non-comparables

Annual Inflation

- Online Index – Substitutions – All Quality
- US CPI – Televisions
Online Index with Non-comparables

Monthly Index

- Online Index – Substitutions – All Quality
- US CPI – Televisions
Online – Small Sample

• Un-weighted geometric mean (Jevons Index)
• Monthly data
• Complete times series for each good
• Small sample:
  – Only “Samsung” brand - 64 tvs
  – Randomly select 10% of ids
Online – Small Sample

Annual Inflation

-30.0
-25.0
-20.0
-15.0
-10.0
-5.0
0
5
10
15
20
25
30

04-01-2008
10-01-2008
04-01-2009
10-01-2009

Online Index – Small Sample
US CPI – Televisions
Online Index – Small Sample

Monthly Index

Date

Online Index – Small Sample

US CPI – Televisions
Online Index

• Un-weighted geometric mean (Jevons)
• Monthly data
• **Complete times series** for each good
• **Large sample**: all products available (577)
Online Index

Price Index

- Online Index
- US CPI – Televisions
Online Index

Monthly Index

-8.0
-6.0
-4.0
-2.0
0.0
2.0
4.0
6.0
8.0

04-01-2008
10-01-2008
04-01-2009
10-01-2009

date

Online Index
US CPI – Televisions
Do we need daily data?

• Online index with **daily** prices
Online Index – Daily Data

Price Index

Date

Online Index – Daily

Online Index
Online Index – Daily Data

Annual Inflation

-35.0
-30.0
-25.0
-20.0
-15.0
-10.0
-5.0
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0

04–01–2008
10–01–2008
04–01–2009
10–01–2009

date

Online Index – Daily
Online Index
Online Index – Daily Data

Monthly Index

Date


Online Index – Daily
Online Index
Preliminary Results

- An online index with simple overlapping quality adjustments can approximate the BLS TV index with hedonic adjustments.
- Requires both uncensored price spells and a large number of closely comparable goods.
- Daily data is not required, but it provides anticipation and can be used to detect price anomalies.
Other Advantages

• No “new good bias”: Silver (98) → the launch of a new model affects the behavior of the old model
Next Steps

• Expand the analysis to other retailers, sectors, and countries

• What happens in sectors like Apparel, where there is little overlap?