



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 (eg 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:

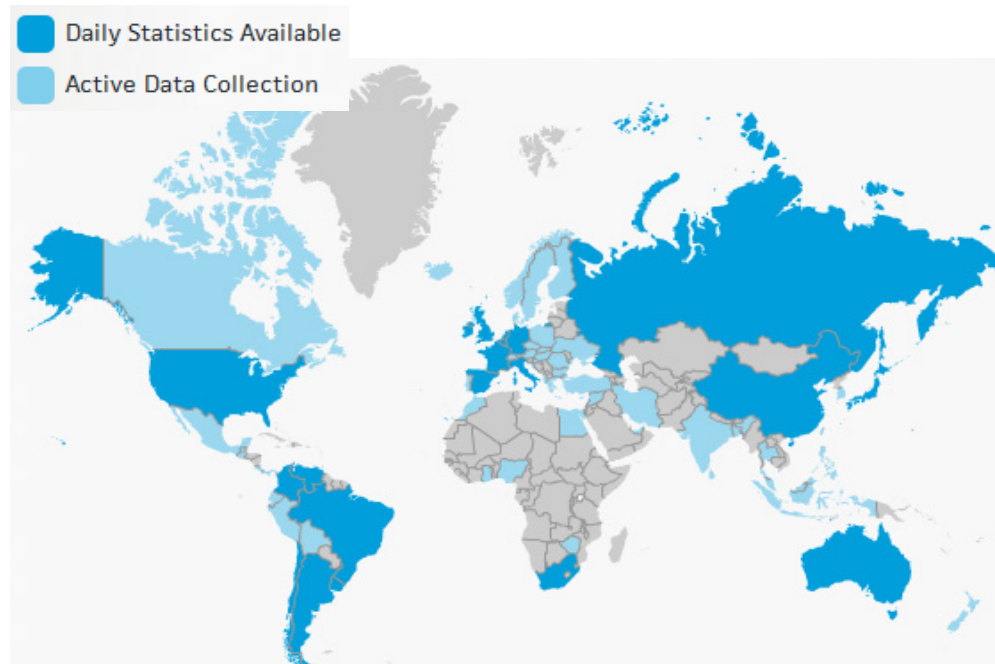
- Relatively few retailers
- No quantity data → expenditure weights?
- Only ~60% CPI categories (most services are not yet available online)

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

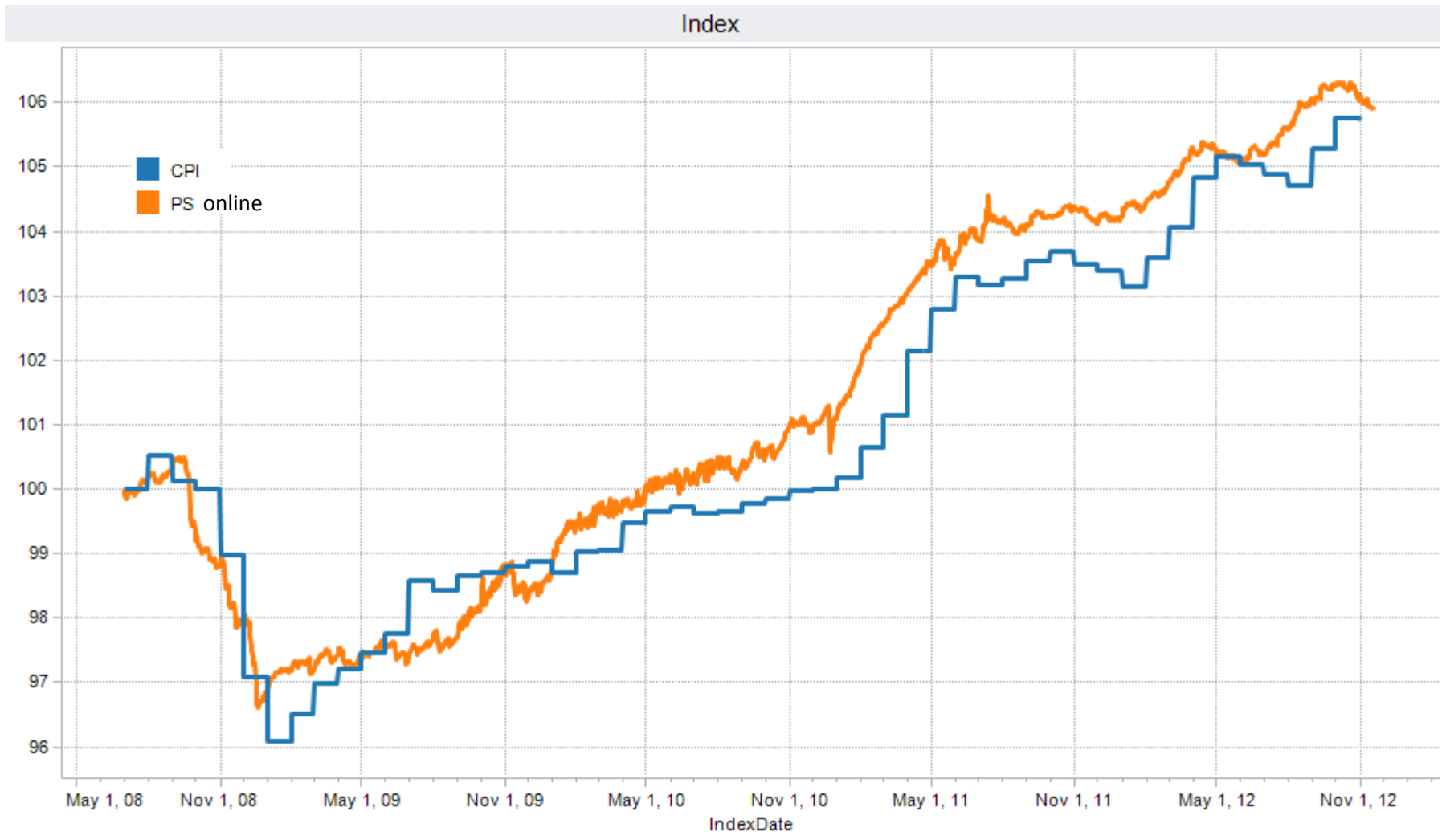
- Joint work with Roberto Rigobon
- Started collecting online prices in 2007
- Daily prices from ~1000 retailers in 50 countries



Daily Price Indices

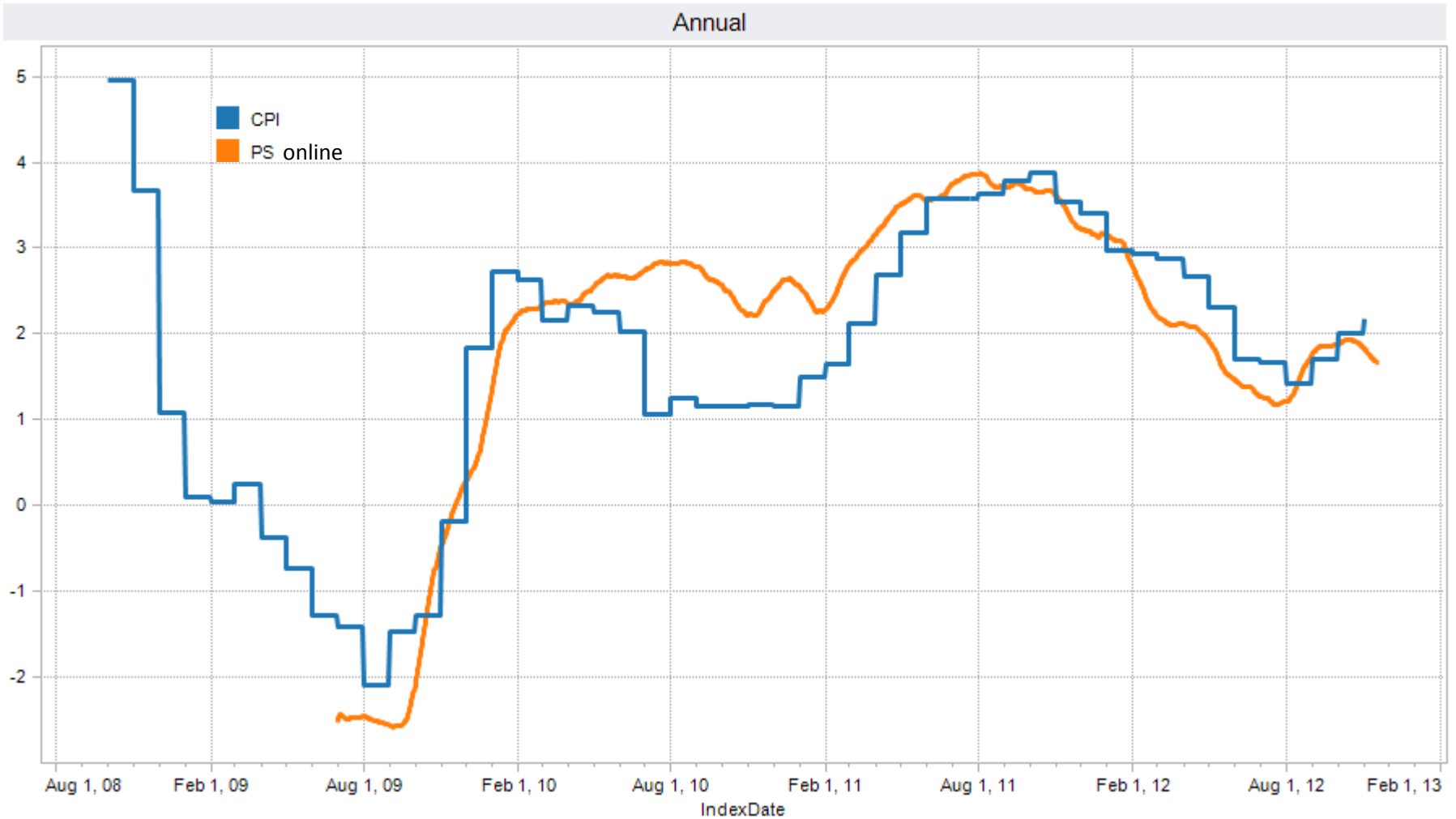
- Available for 21 countries:
 - USA, Argentina, Australia, Canada, Chile, China, Colombia, France, Germany, Greece, Ireland, Italy, Japan, Korea, Netherlands, Russia, South Africa, Spain, UK, Uruguay, and Venezuela
- US and Argentina indices are publicly available online
- Daily series published with a 3-day lag
- Use standard CPI techniques and weights, but we don't make explicit quality adjustments.

US Daily Price Index



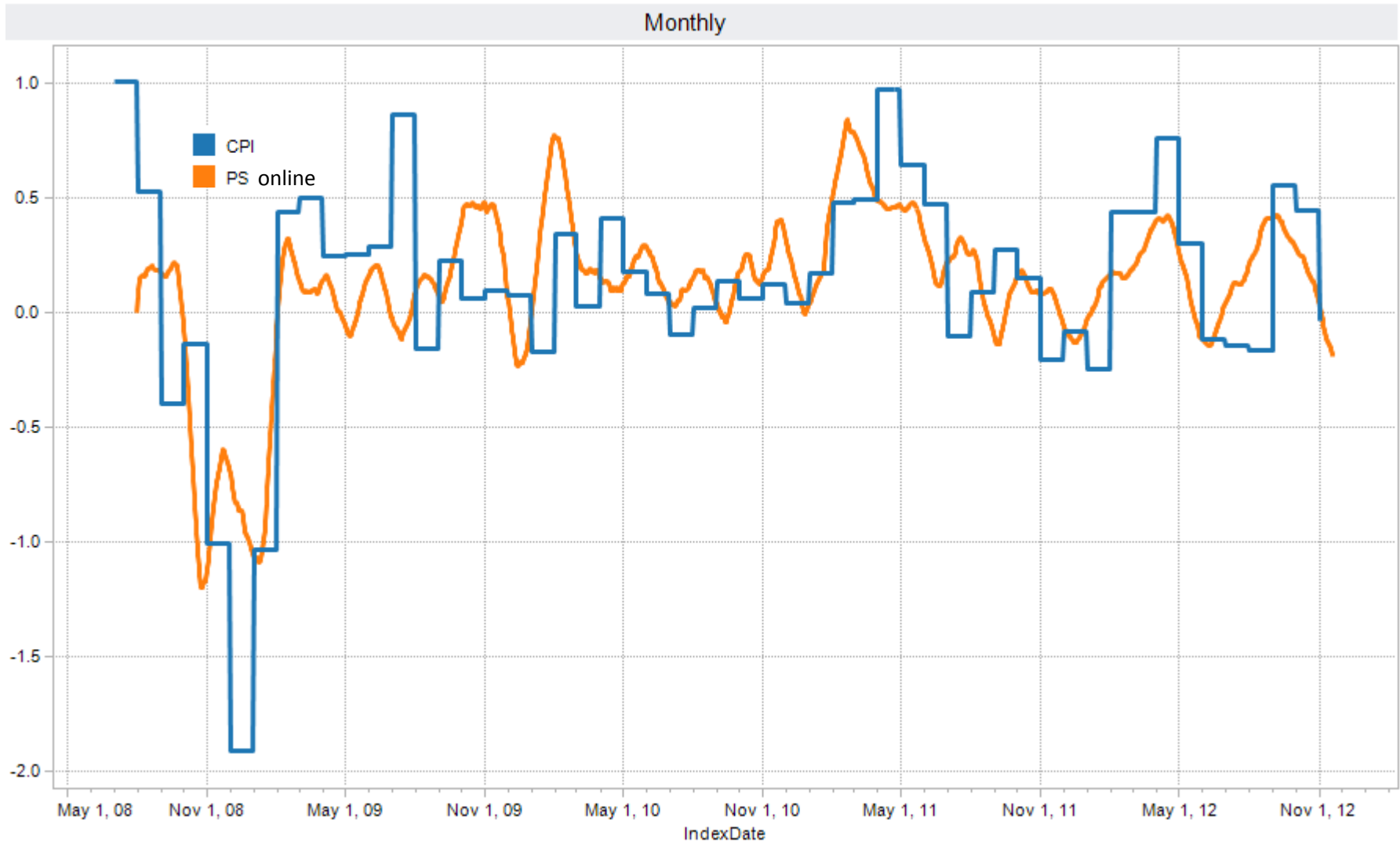
Source: BPP – PriceStats – BLS (CPI-U, US city-average, all items, NSA). Updated until 11/20/2012.

US Annual Inflation



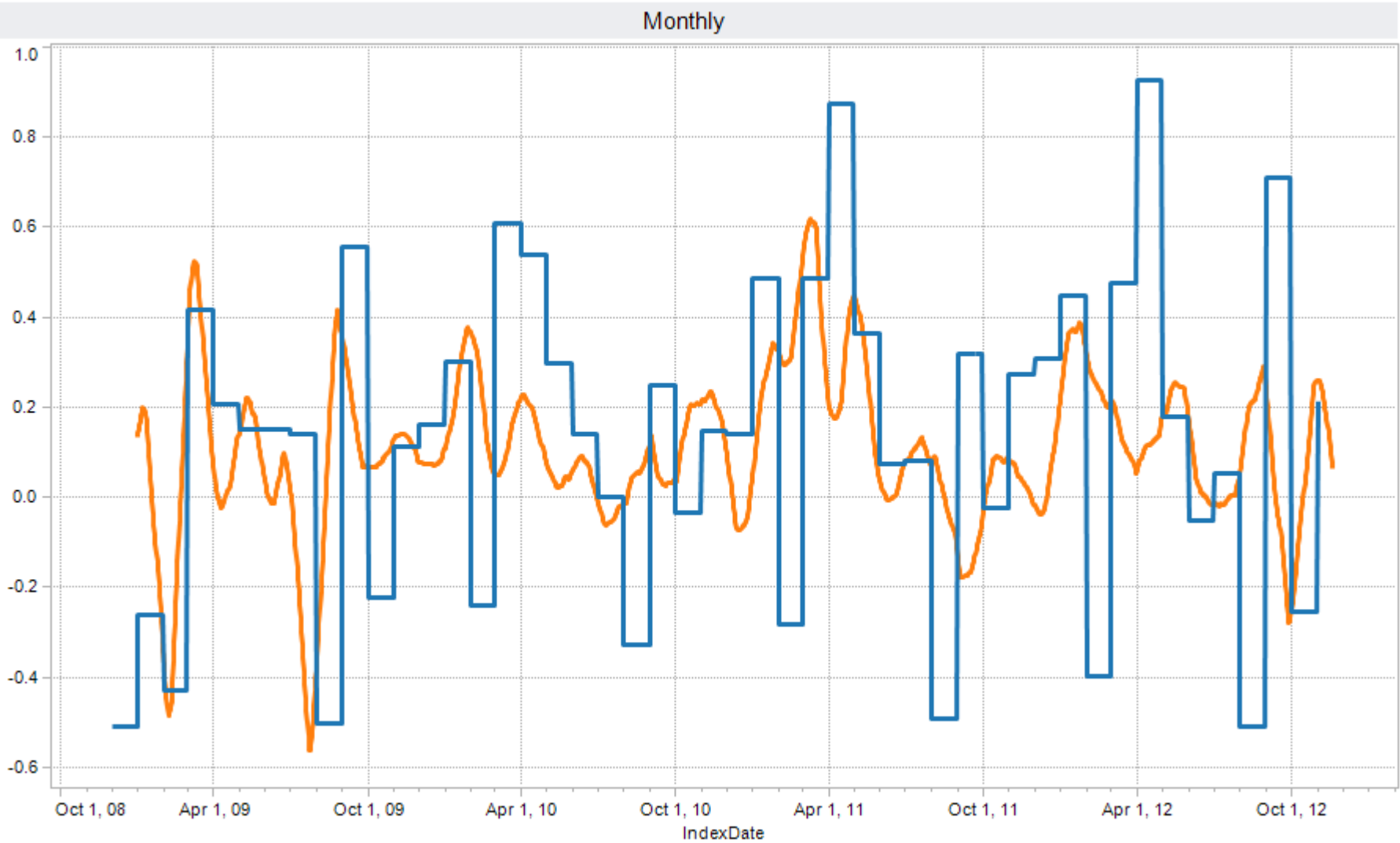
Source: BPP – PriceStats – BLS (CPI-U, US city-average, all items, NSA)

US Monthly Inflation



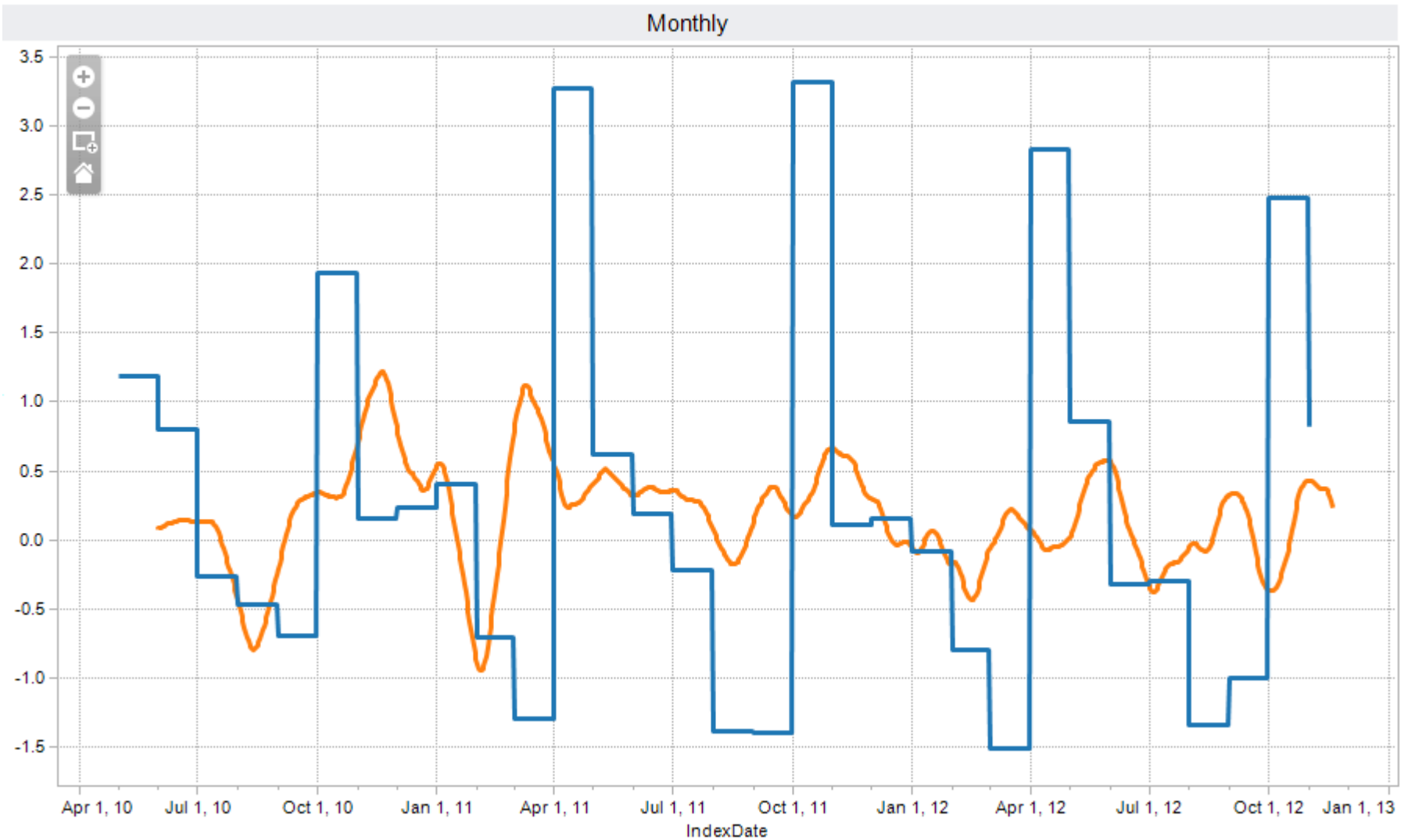
Source: BPP – PriceStats – BLS (CPI-U, US city-average, all items, NSA)

France - Monthly Inflation



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

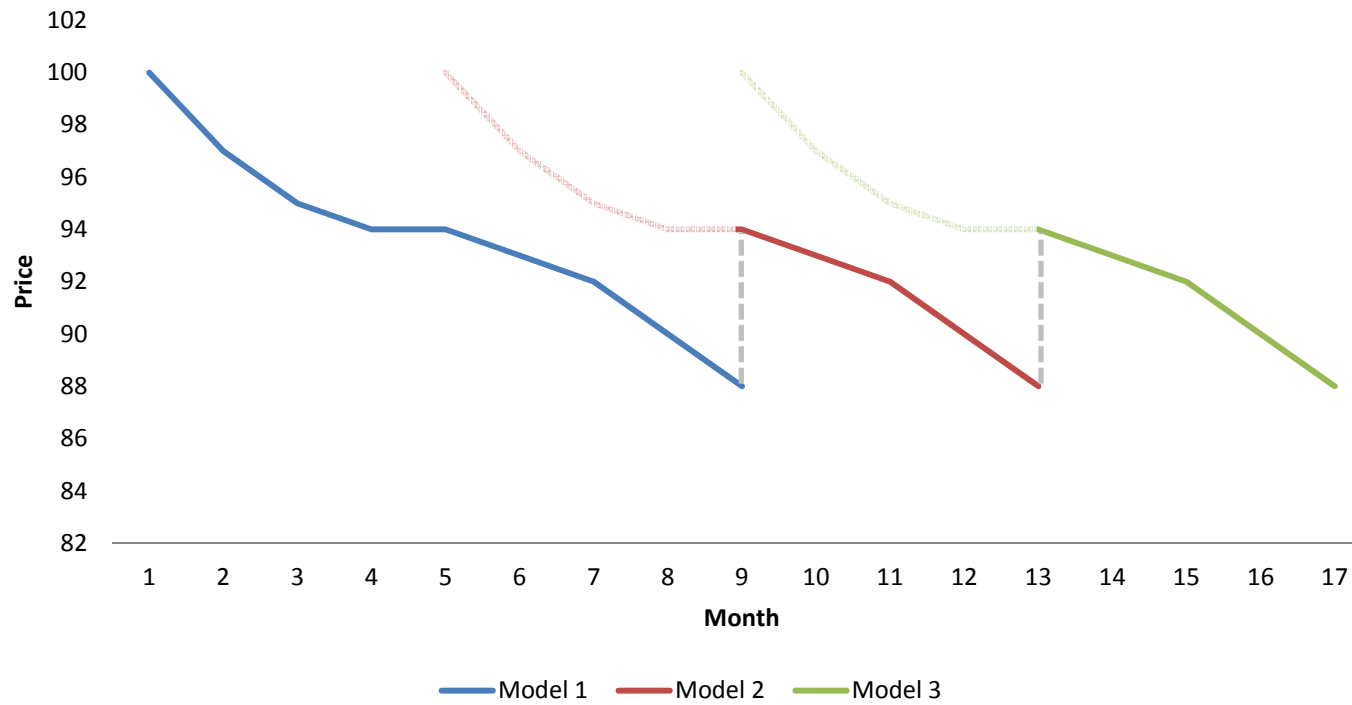
Greece – Monthly Inflation



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

The Problem of Quality Adjustment

Traditional CPI Data



Overlapping Quality Adjustment

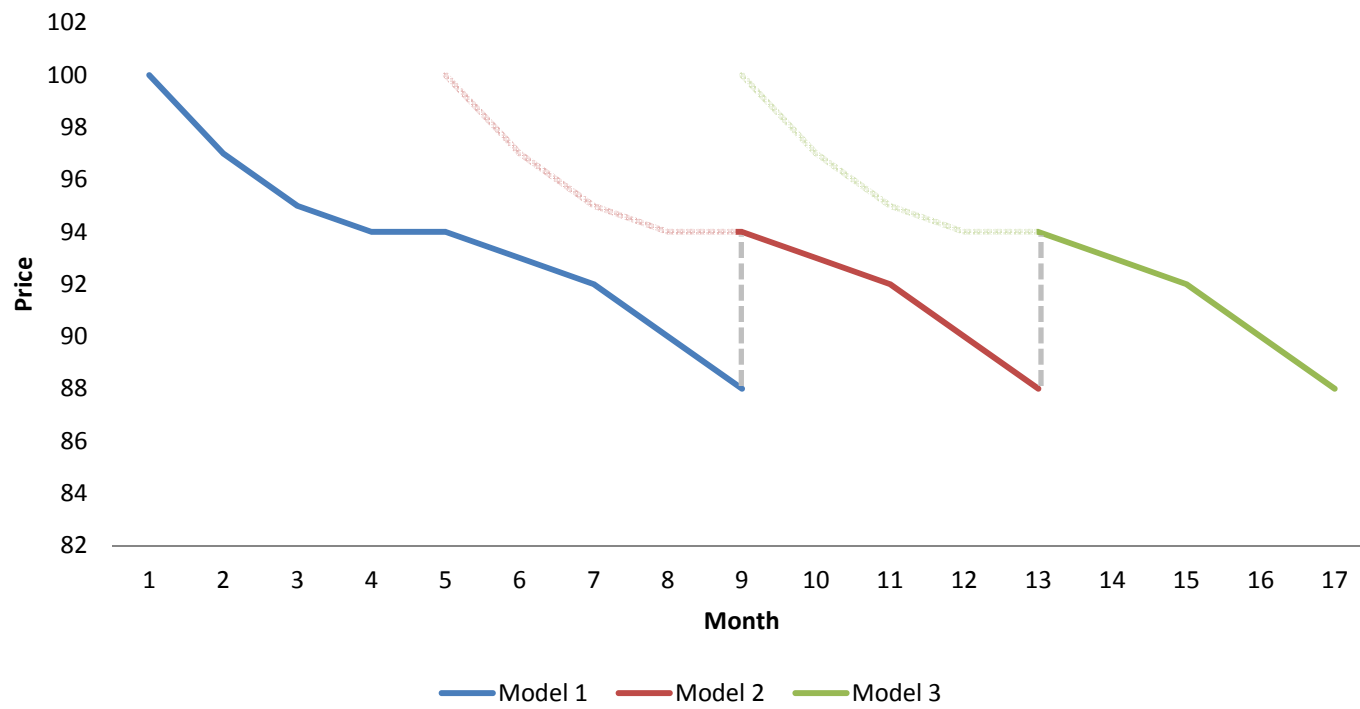
From the ILO/IMF/OECD/... CPI Manual:

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

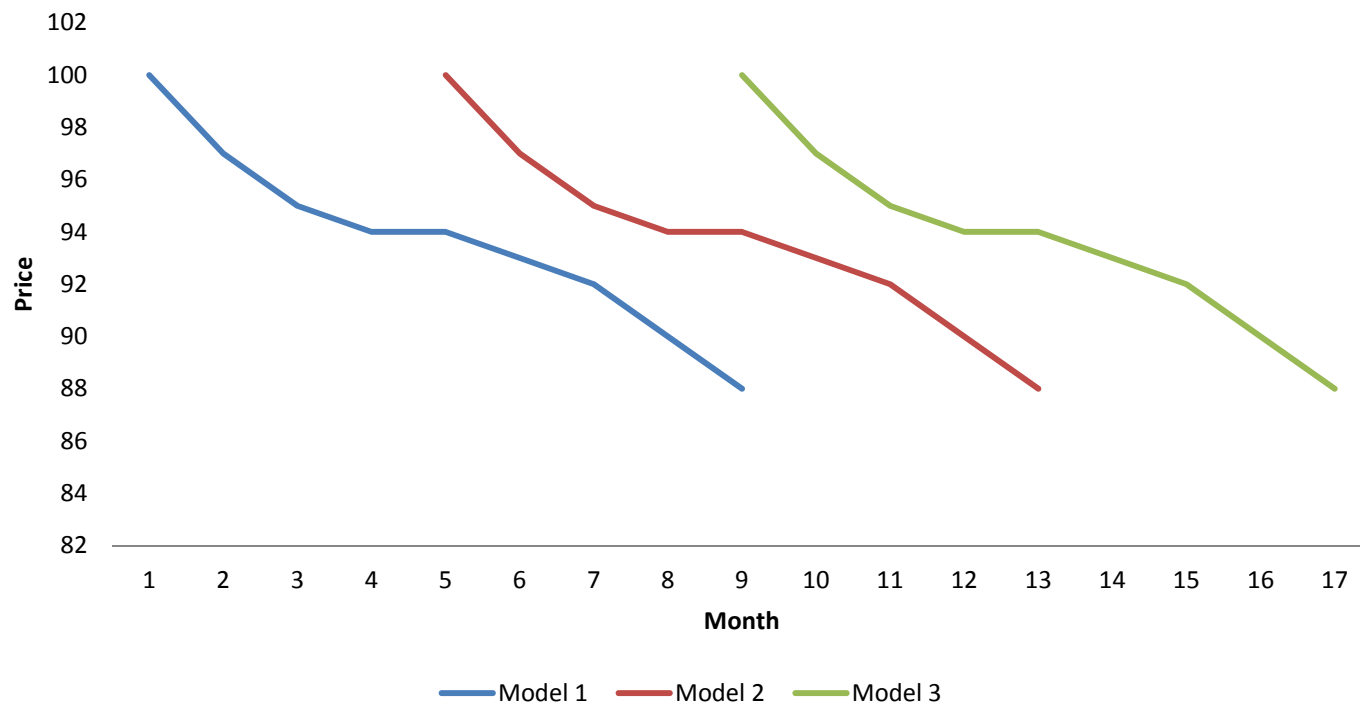
“in practice, however, this method is not used very extensively because the requisite data are seldom available”

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

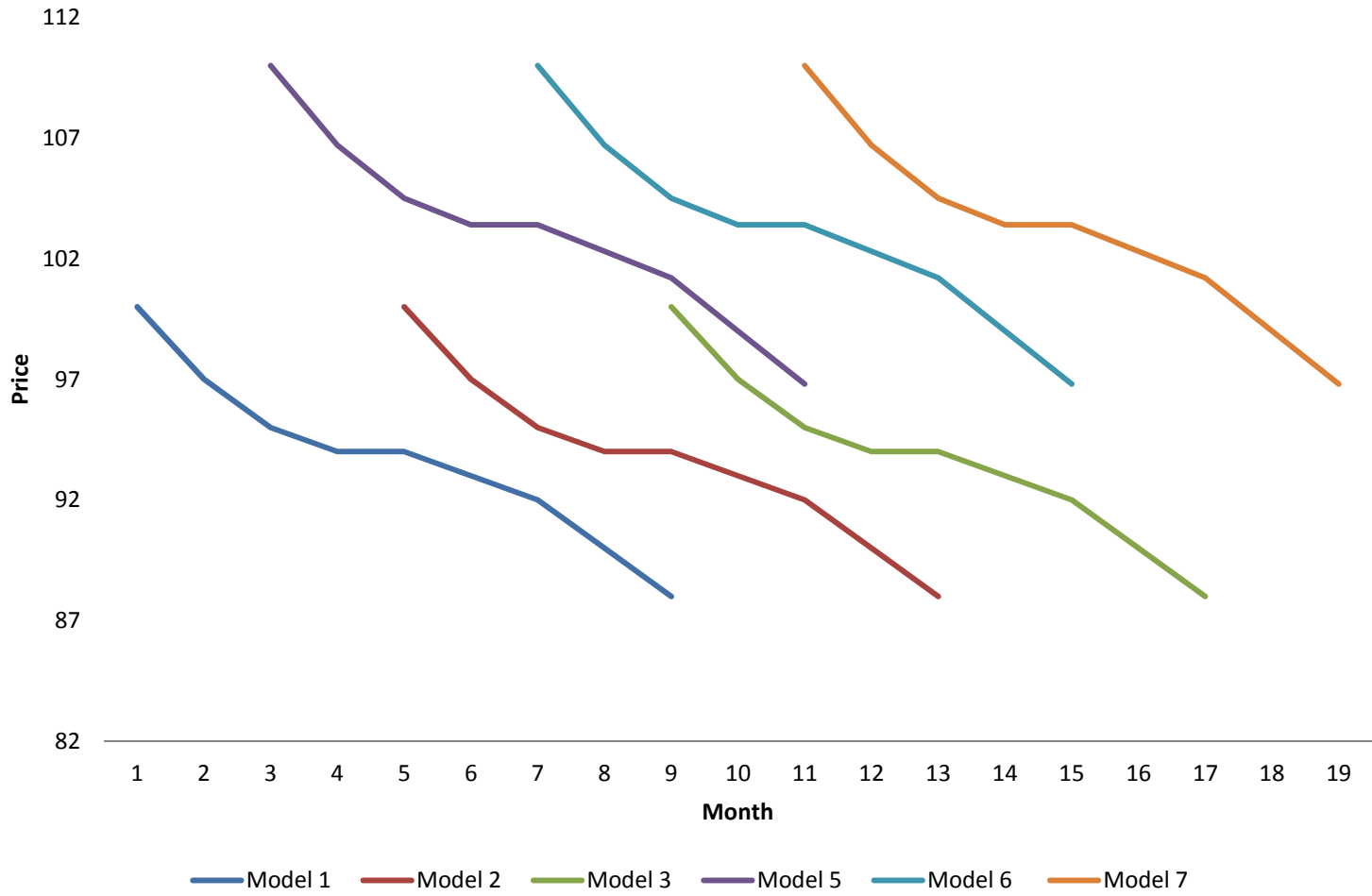
Traditional CPI data



Online Data



Online Data



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

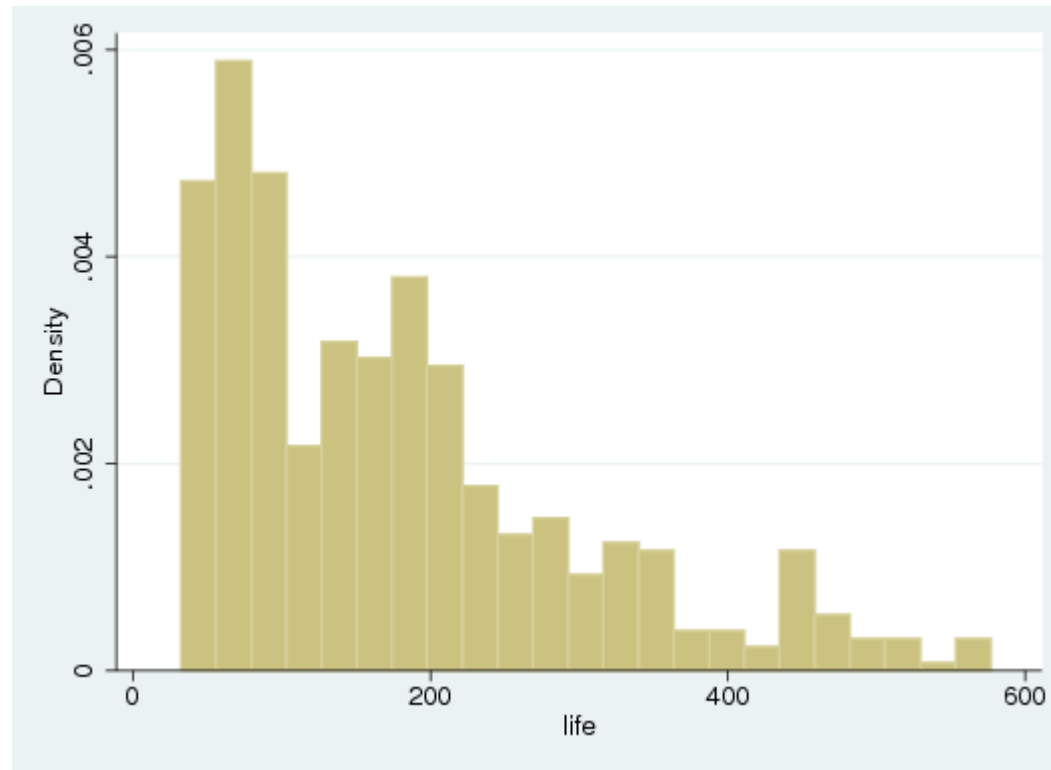
Literature on Matched-Model vs Hedonic Indices

- 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)

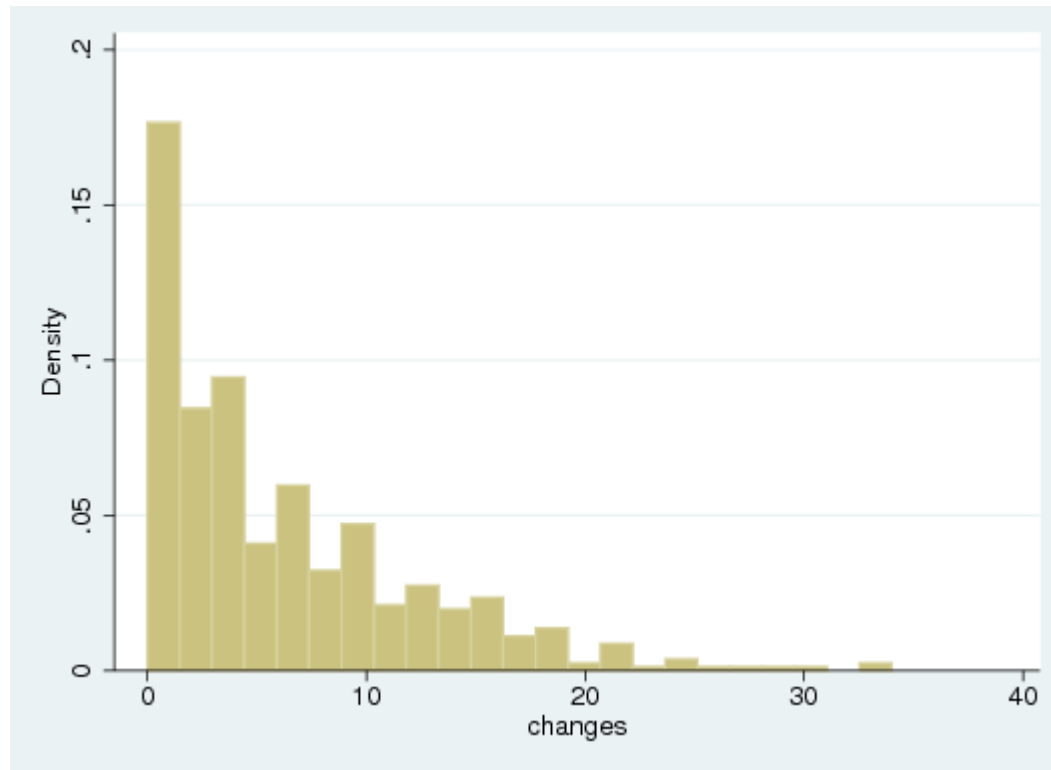
Product Life (duration)



Mean Product life : 178 days (median 154)

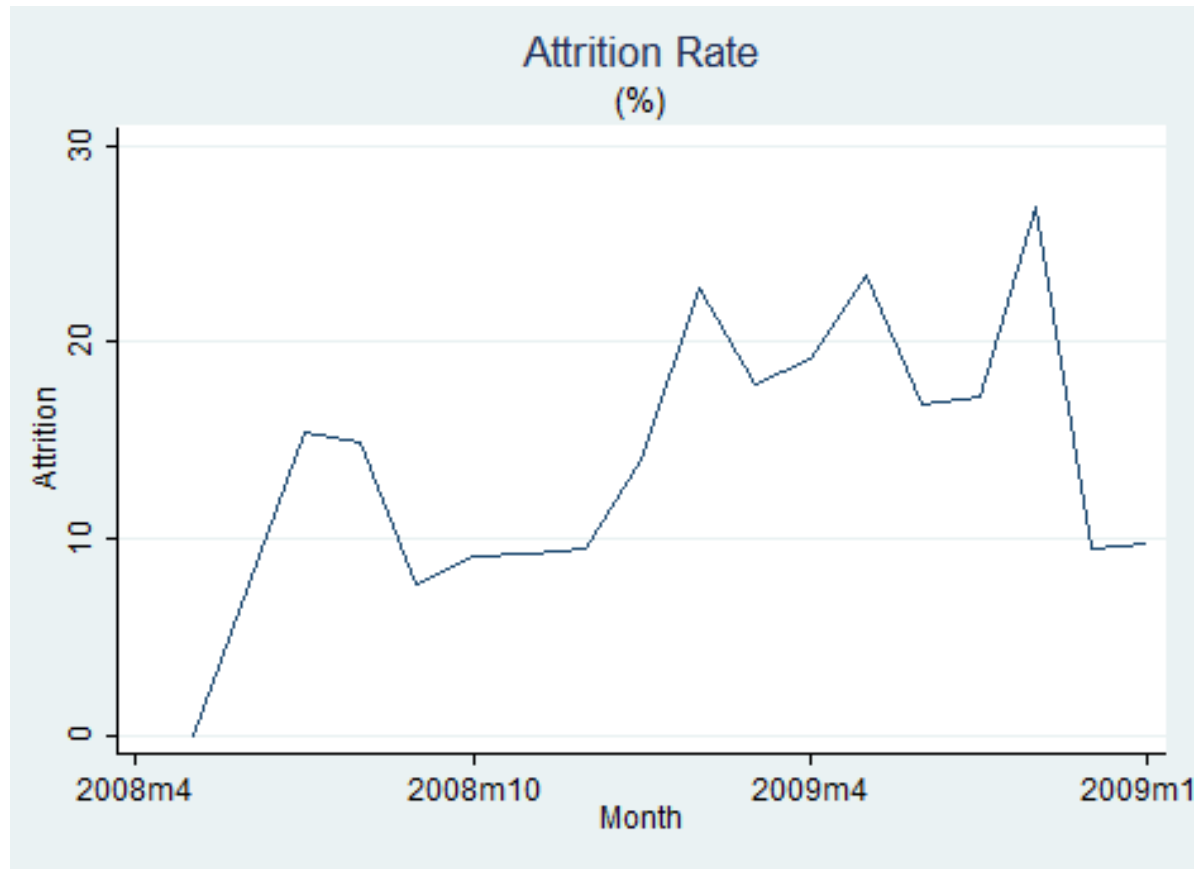
Note: Excluding products lasting less than 30 days (refurbished or open box)

Changes per Product



Mean Price Changes per id: 6 (median 4)

Share of “Exits”



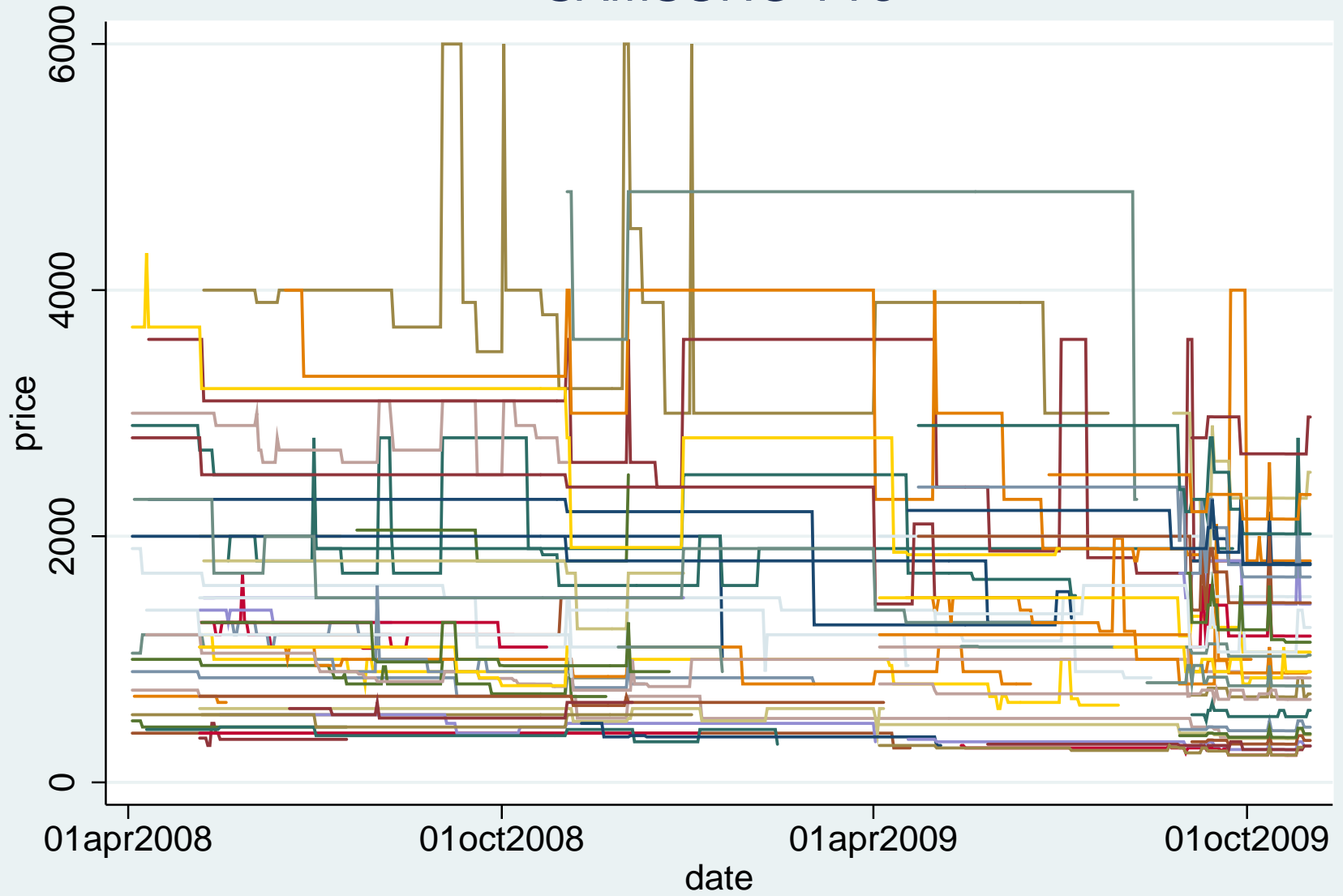
Alternative Price Indices

- 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) → all Q change)
- Online Index - Small sample
- Online Index
- Online Index with daily data

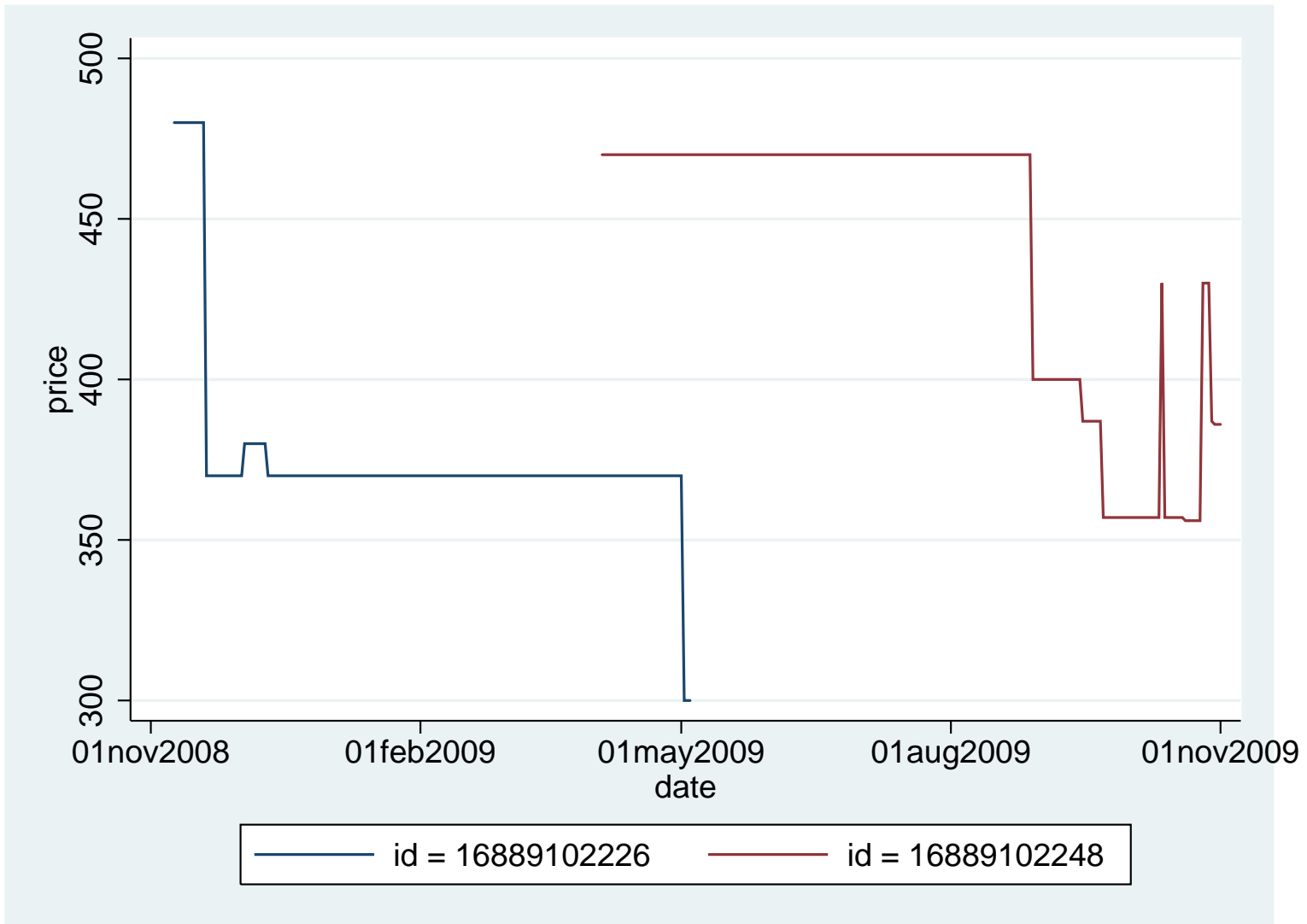
Online Index - Comparables

- 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)

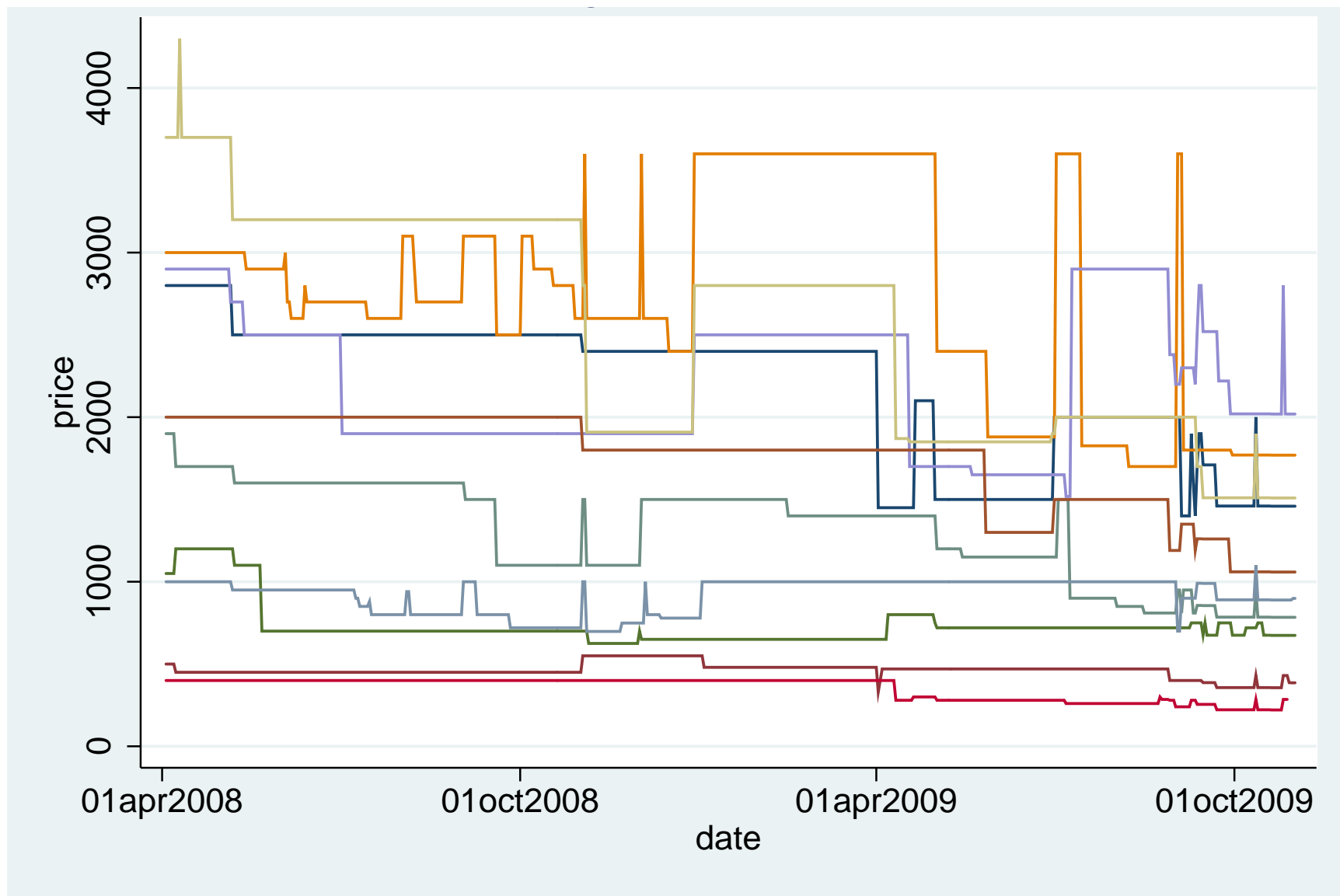
SAMSUNG TVs



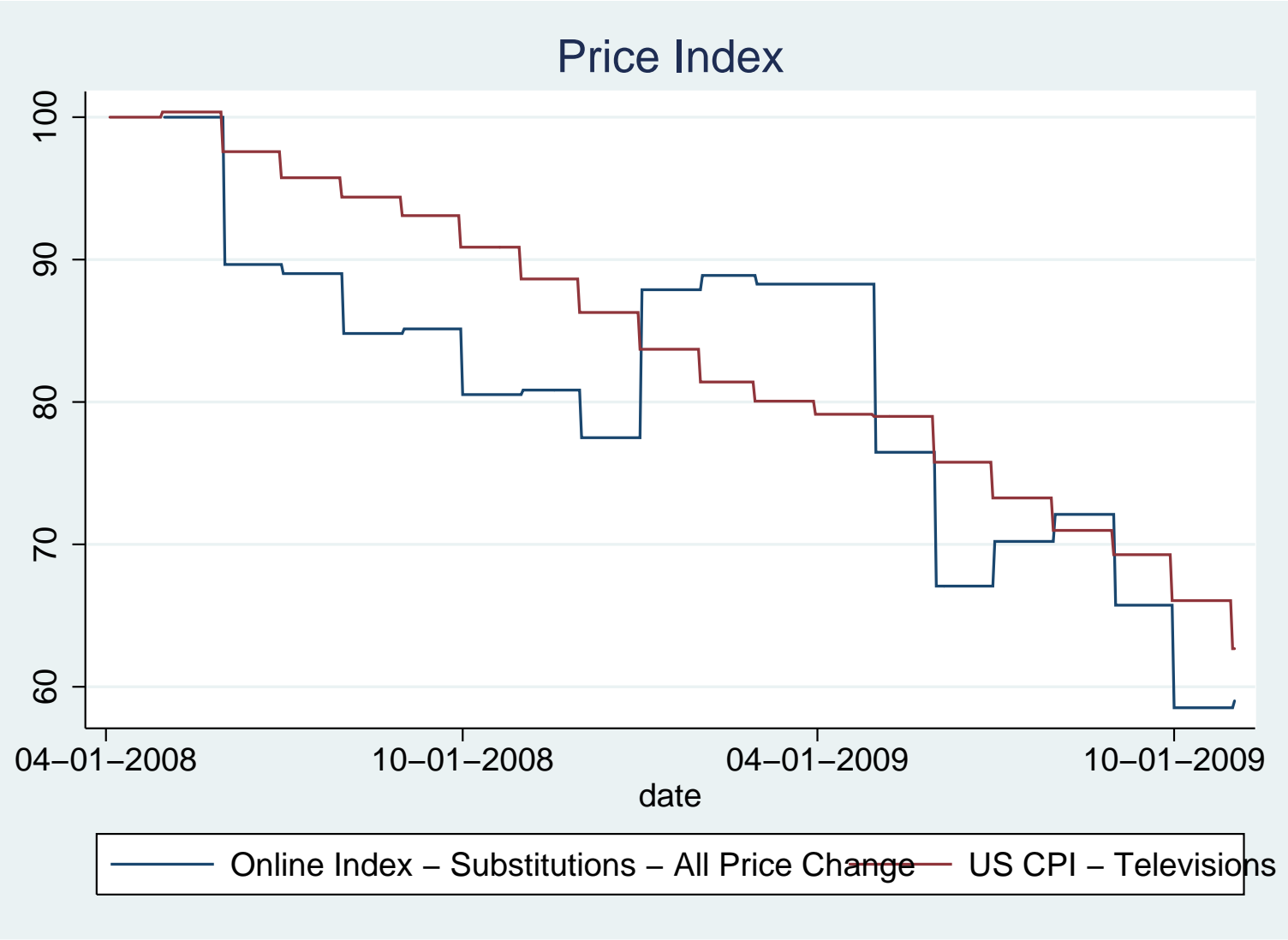
Example of a matched model



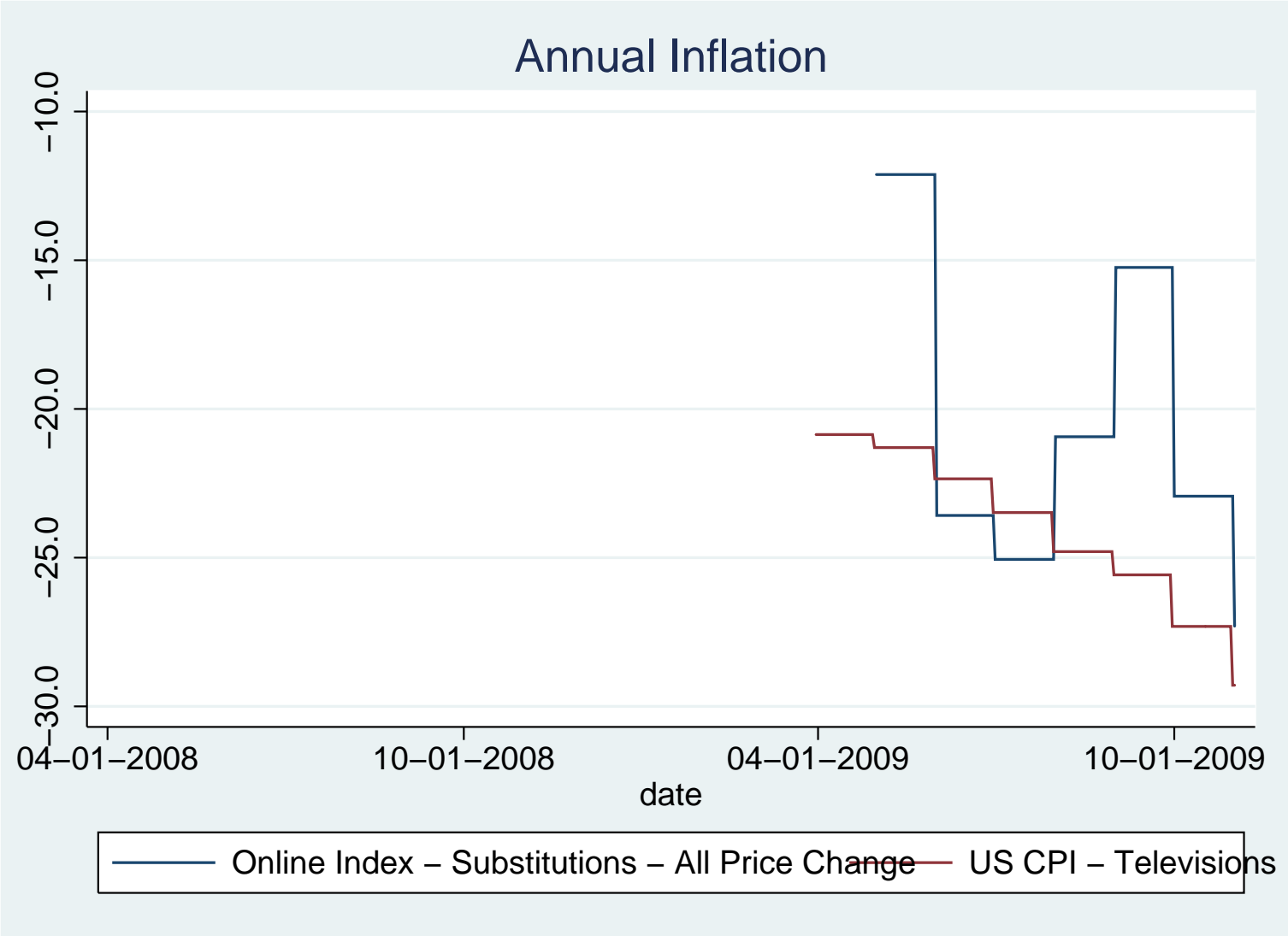
Matched models



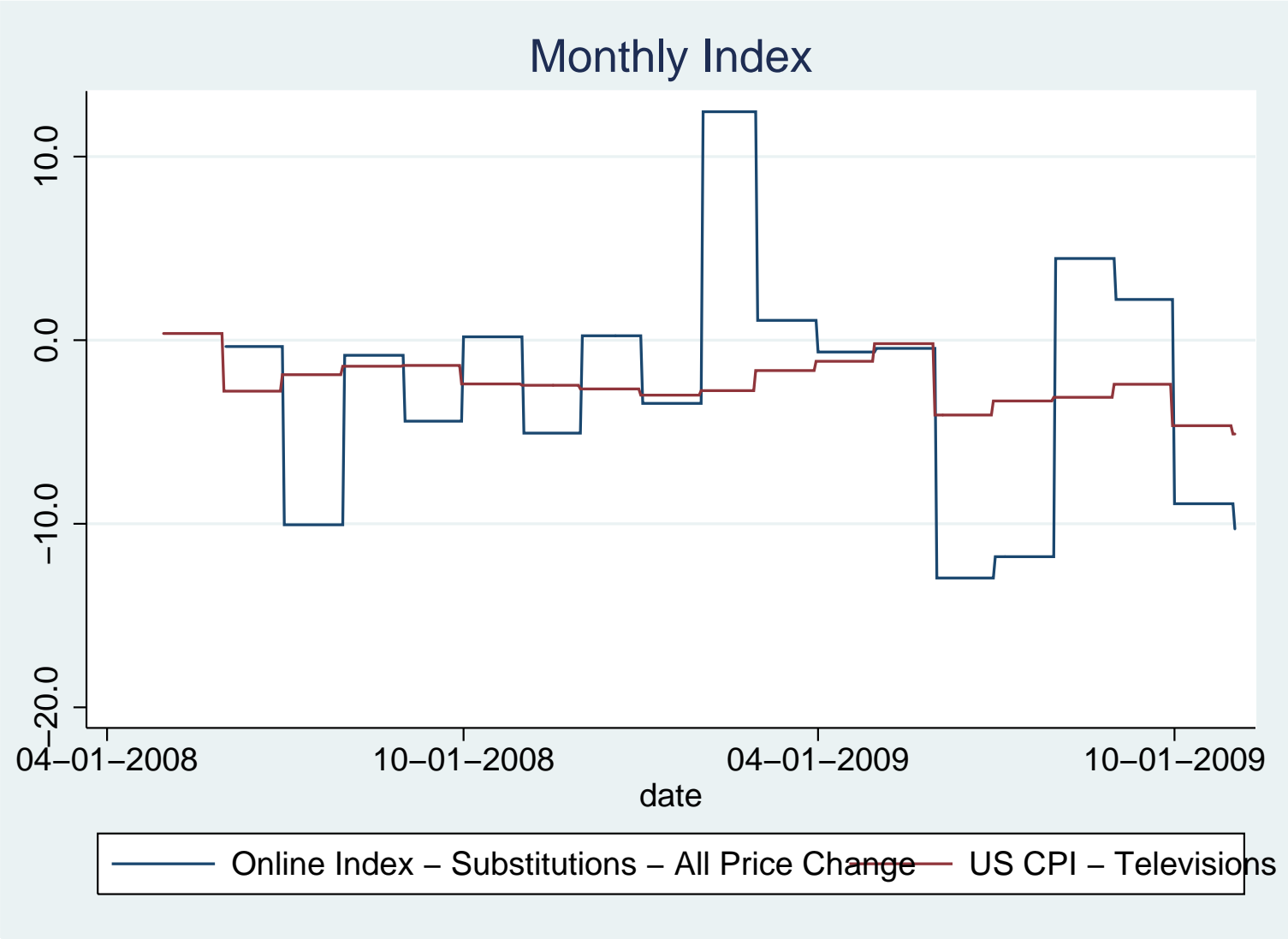
Online Index – Comparables



Online Index – Comparables



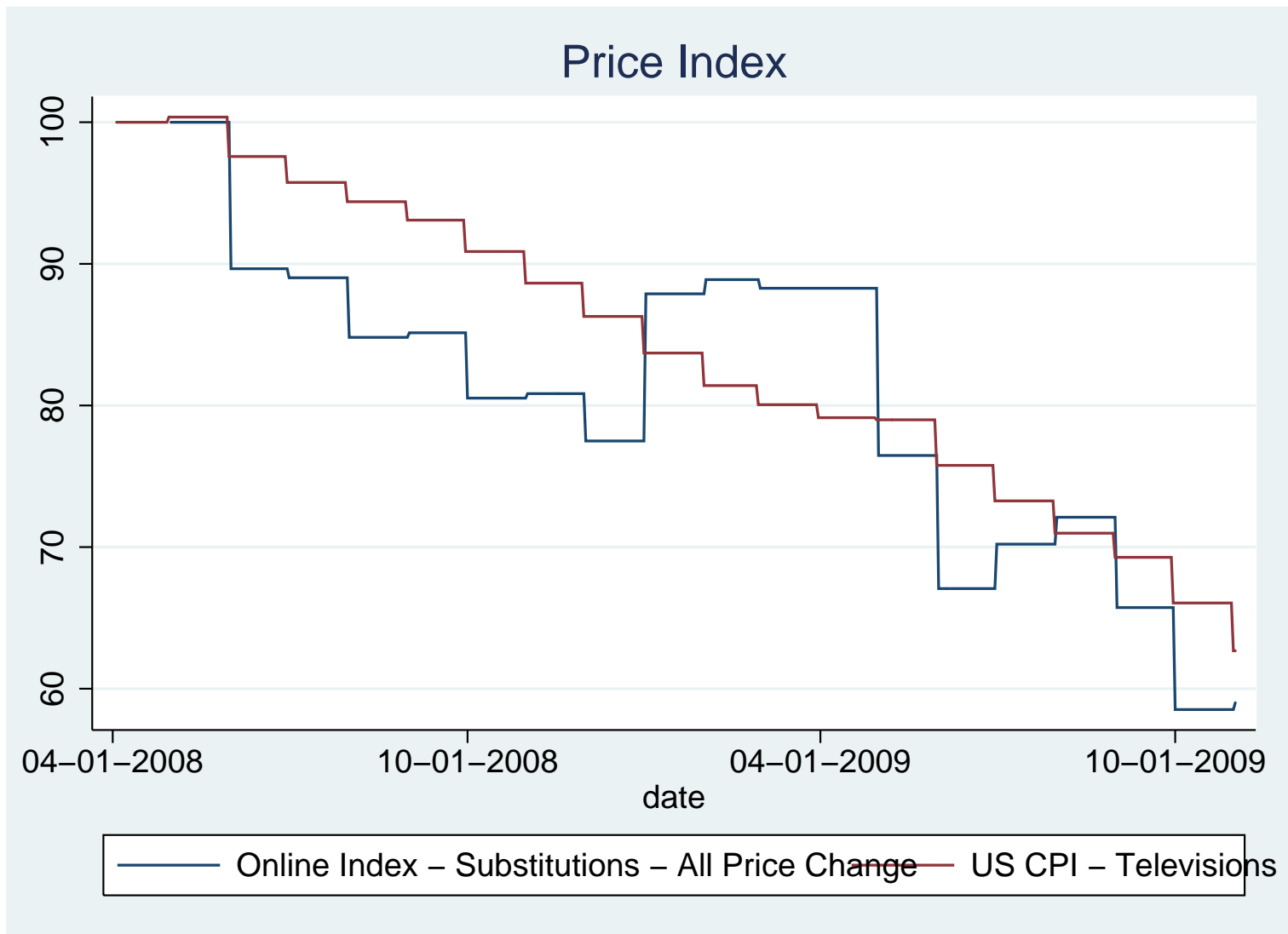
Online Index – Comparables



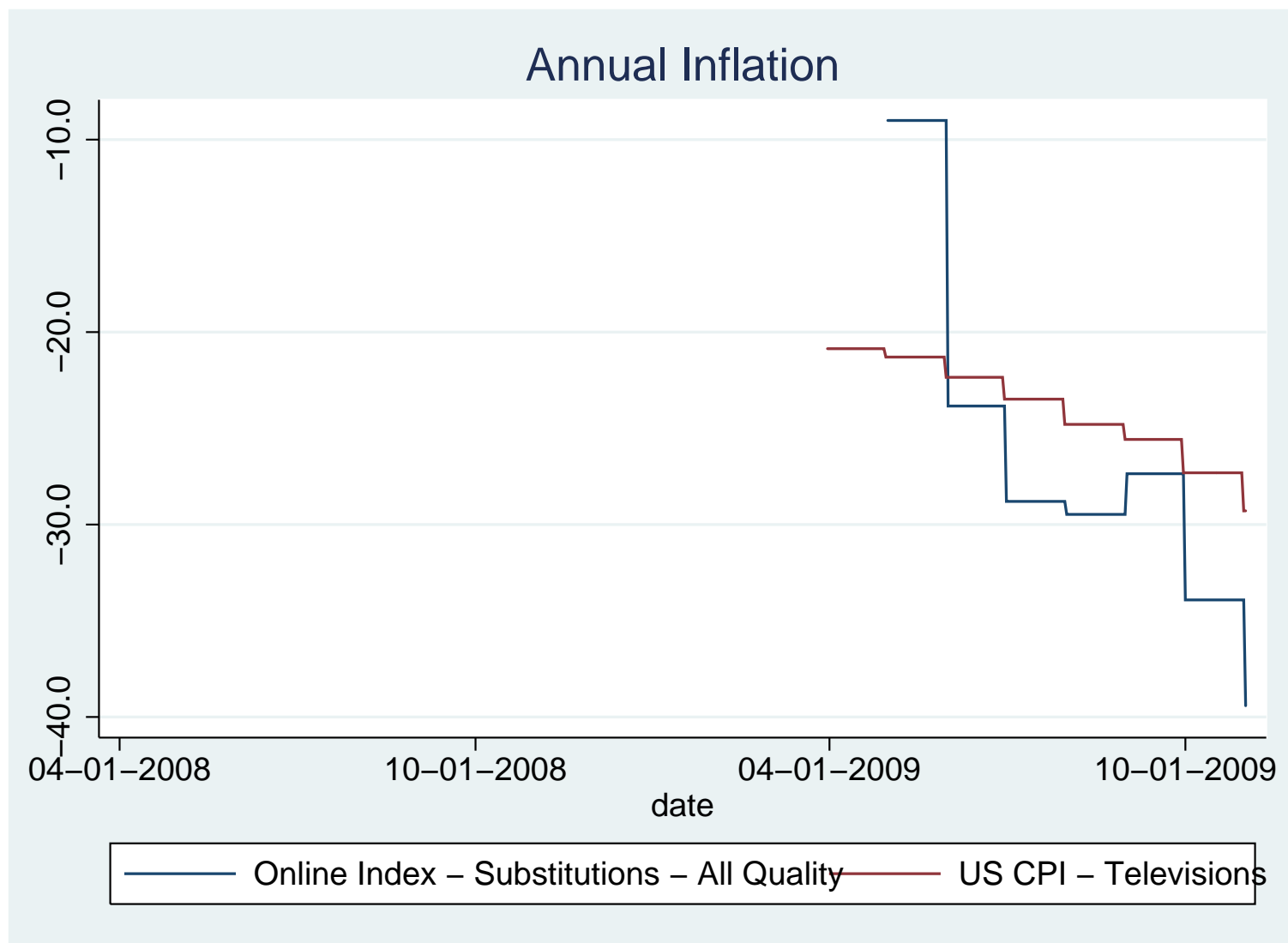
Online Index with Non-comparables

- 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)

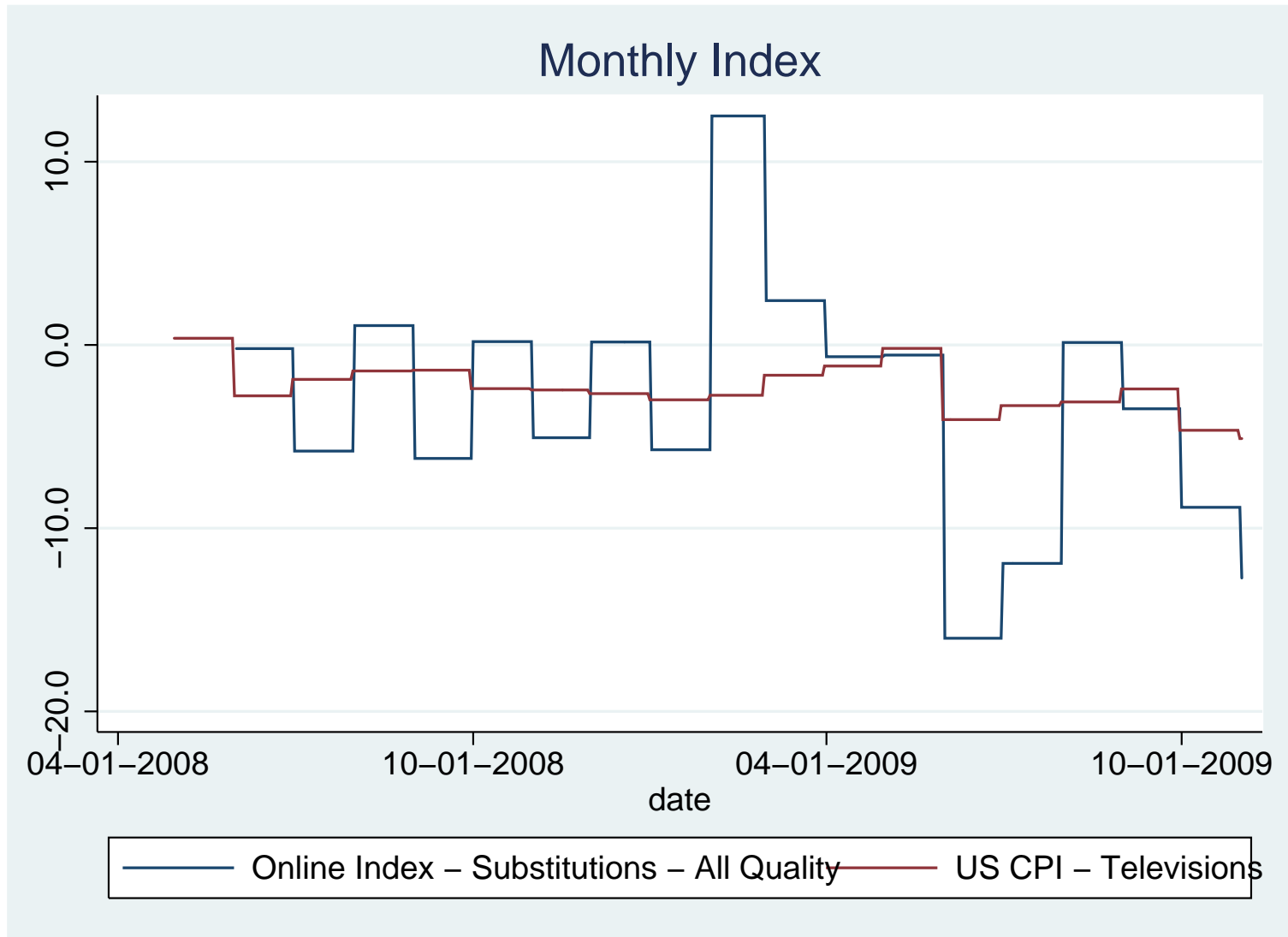
Online Index with Non-comparables



Online Index with Non-comparables



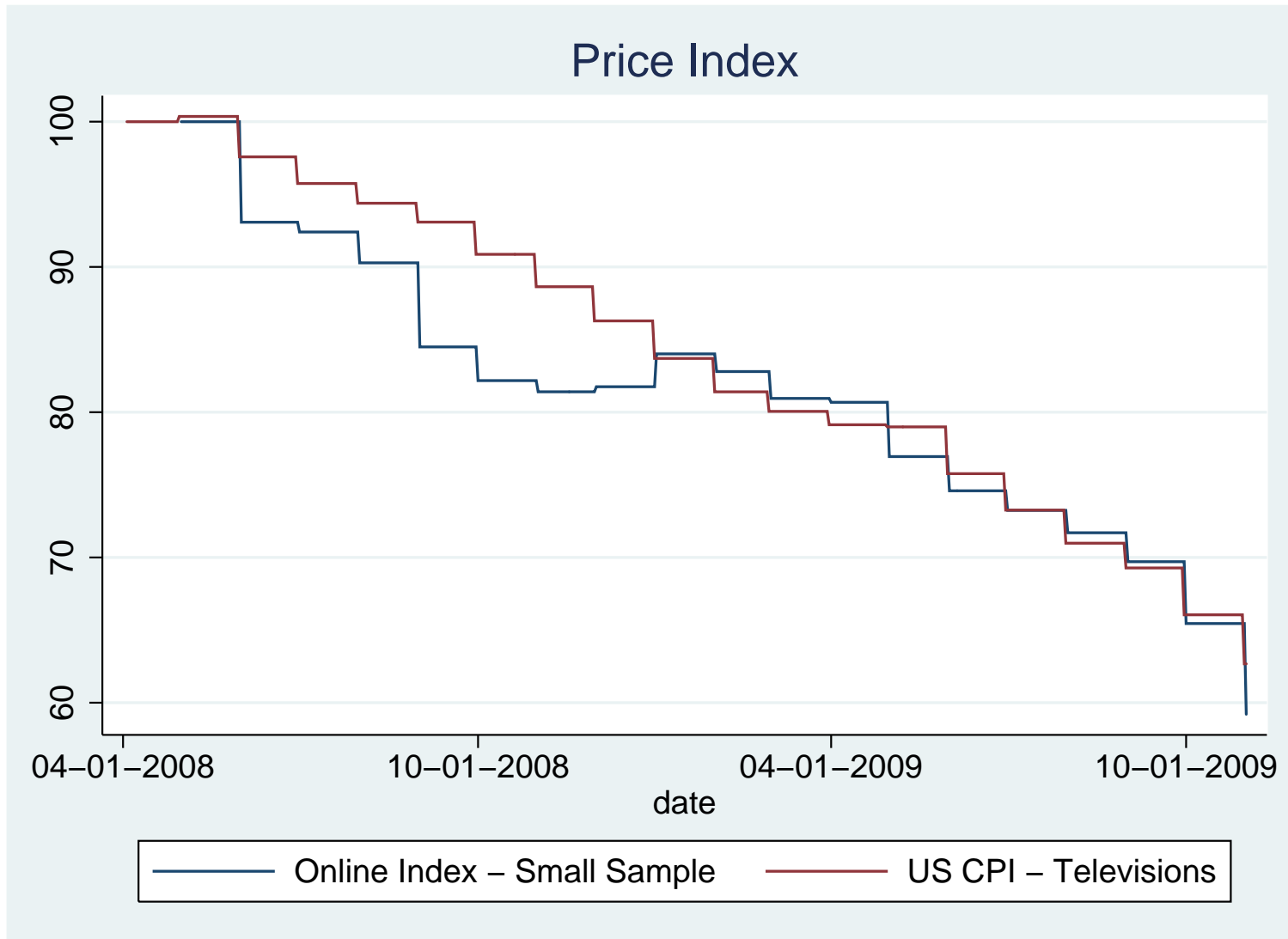
Online Index with Non-comparables



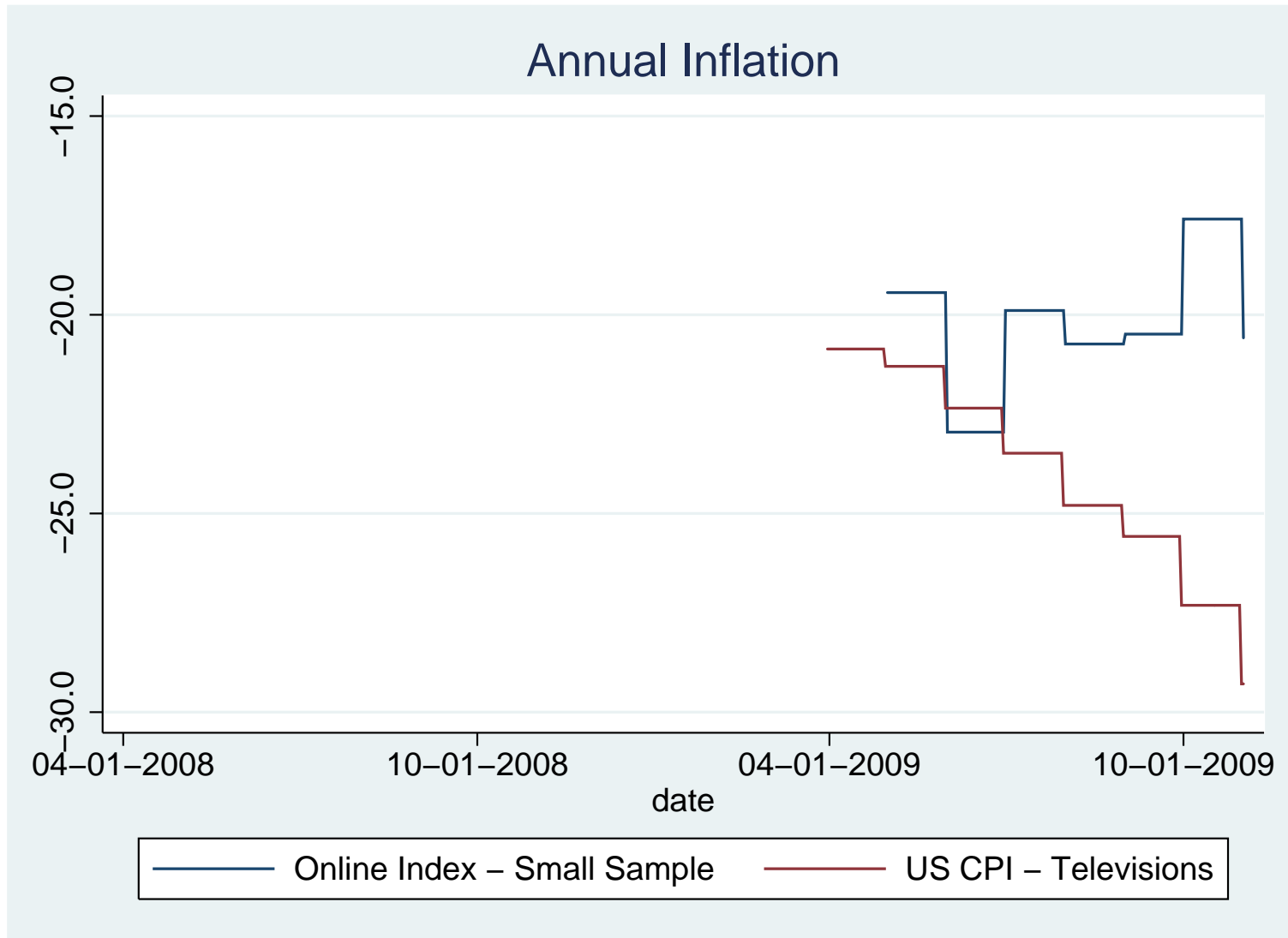
Online – Small Sample

- 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

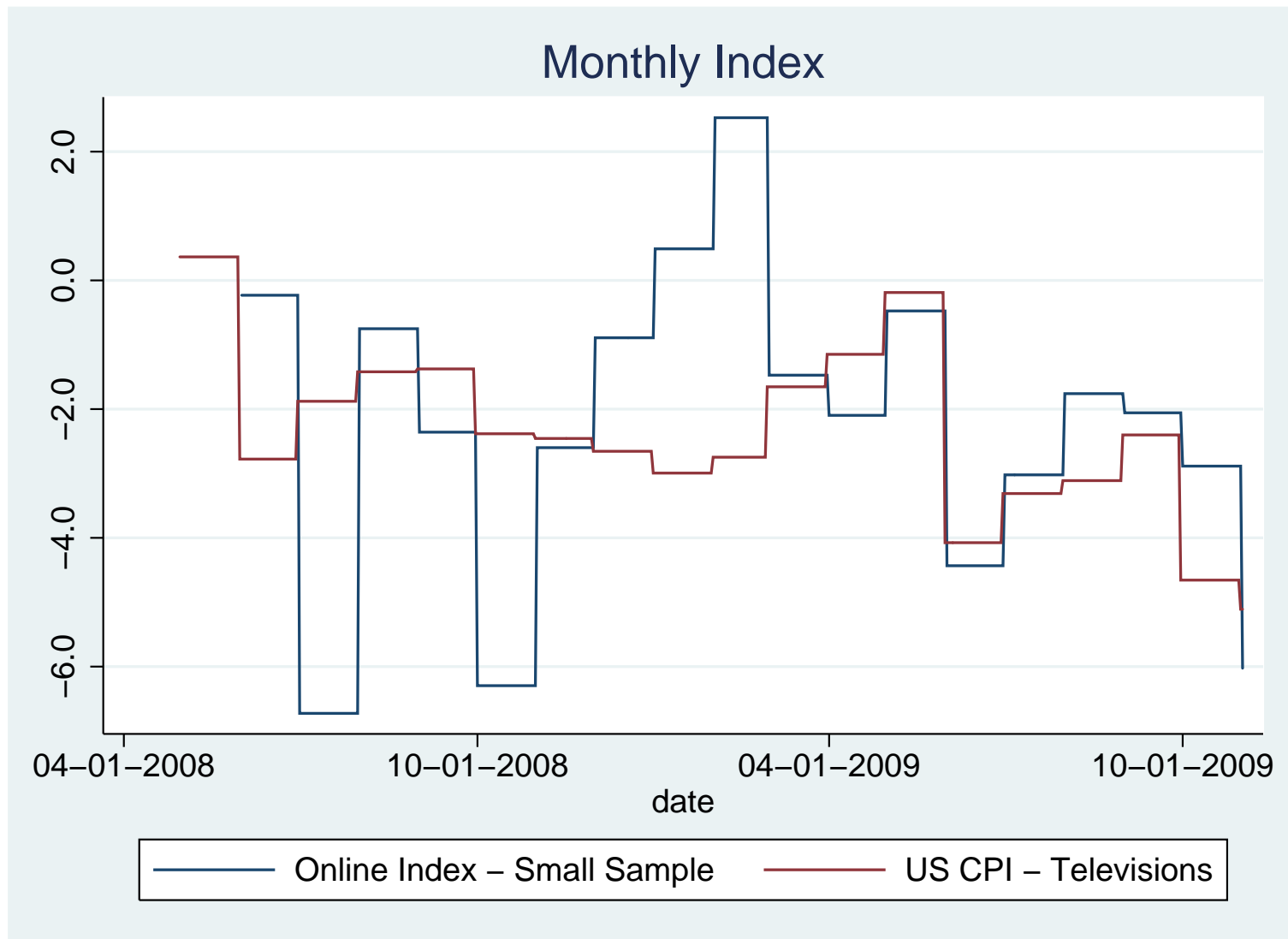
Online – Small Sample



Online – Small Sample



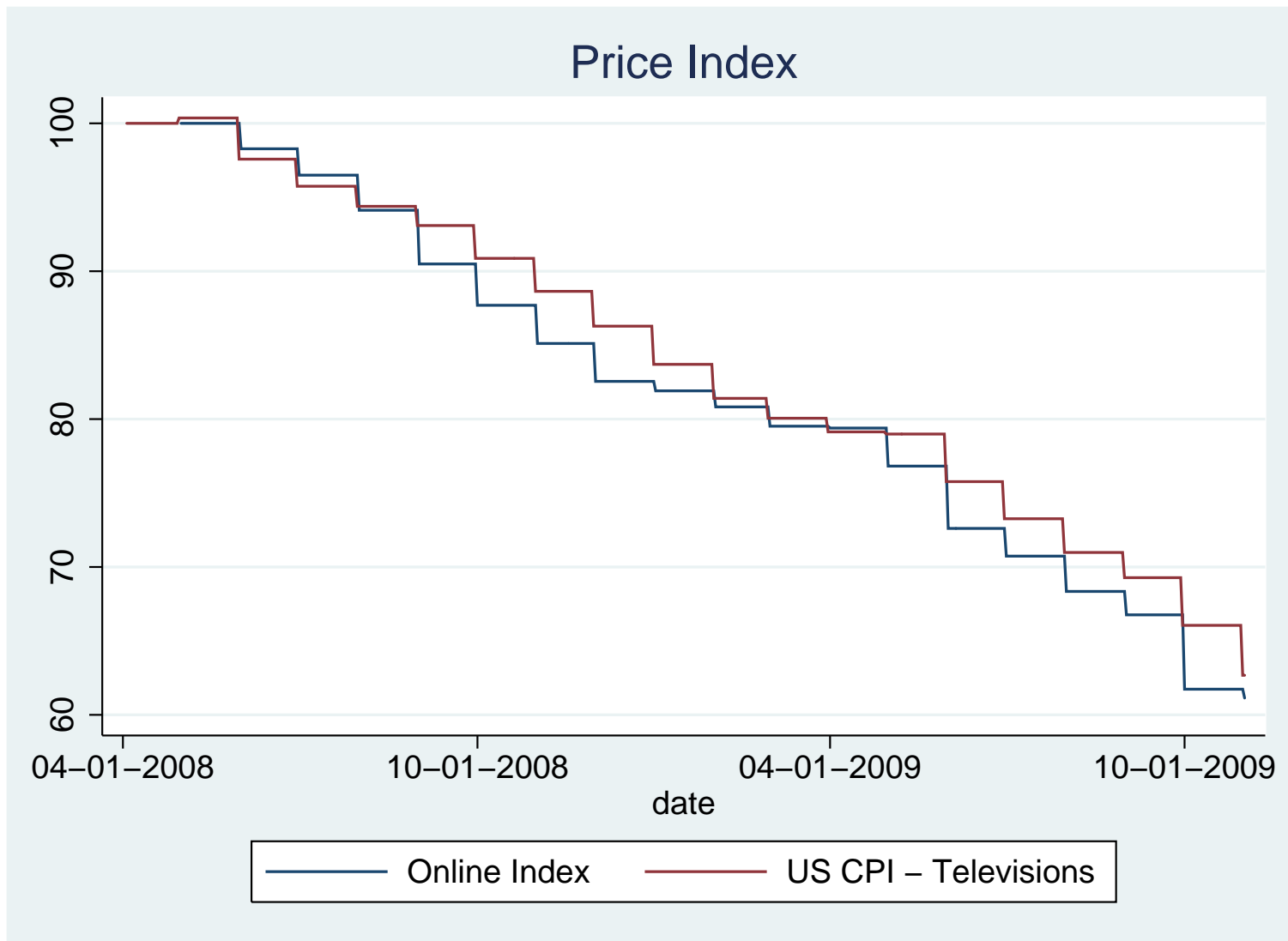
Online Index – Small Sample



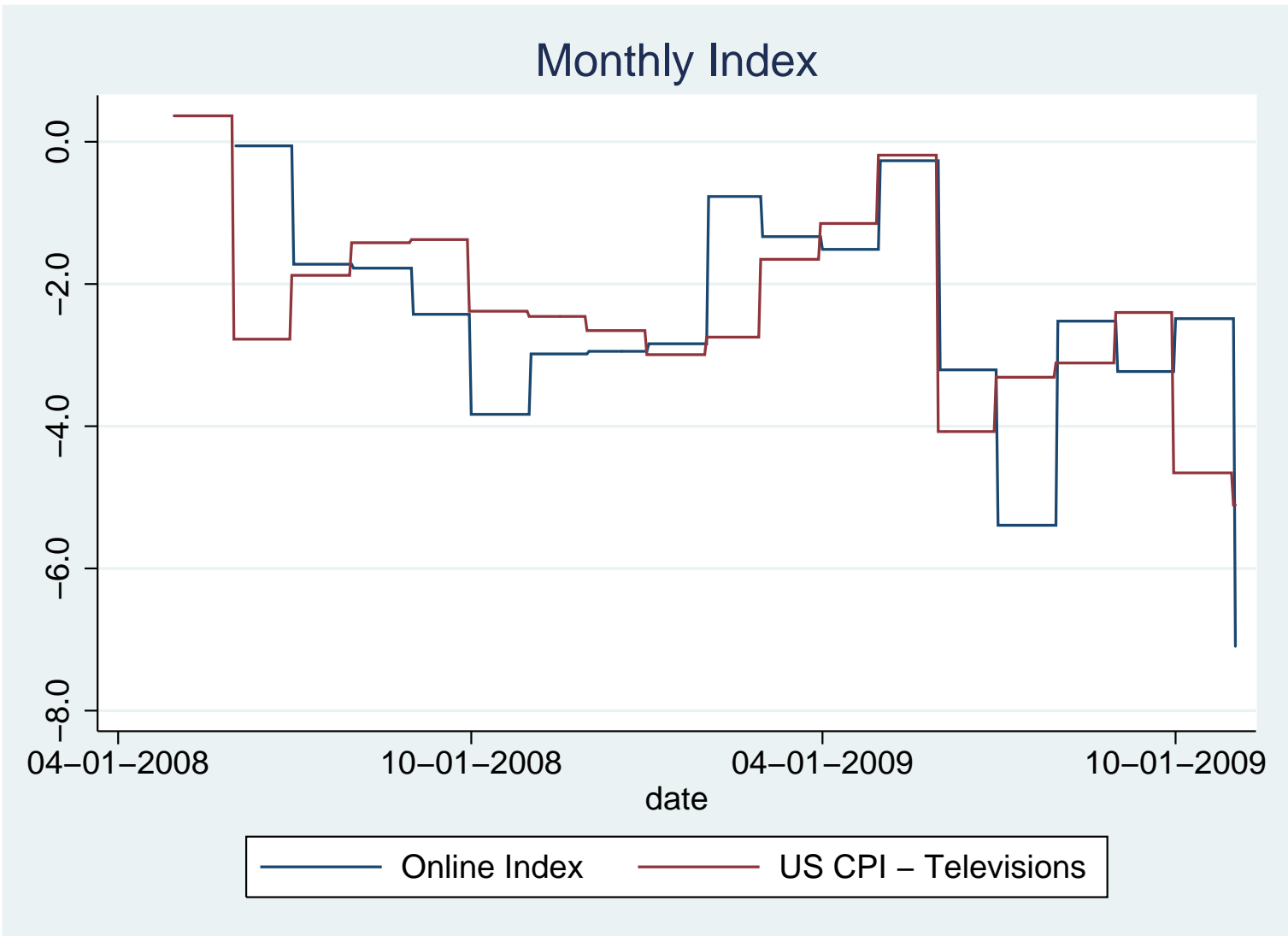
Online Index

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

Online Index



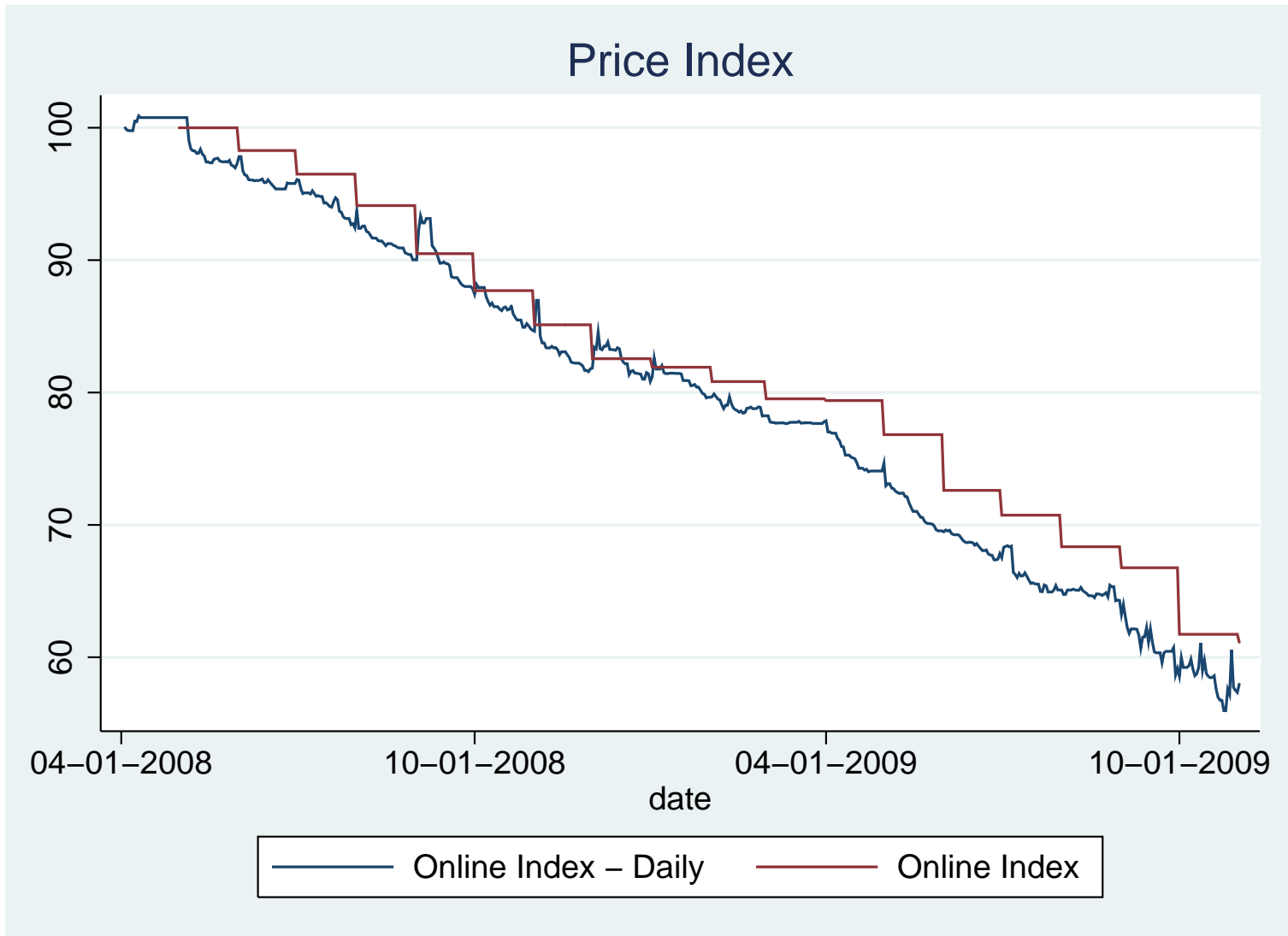
Online Index



Do we need daily data?

- Online index with **daily** prices

Online Index – Daily Data

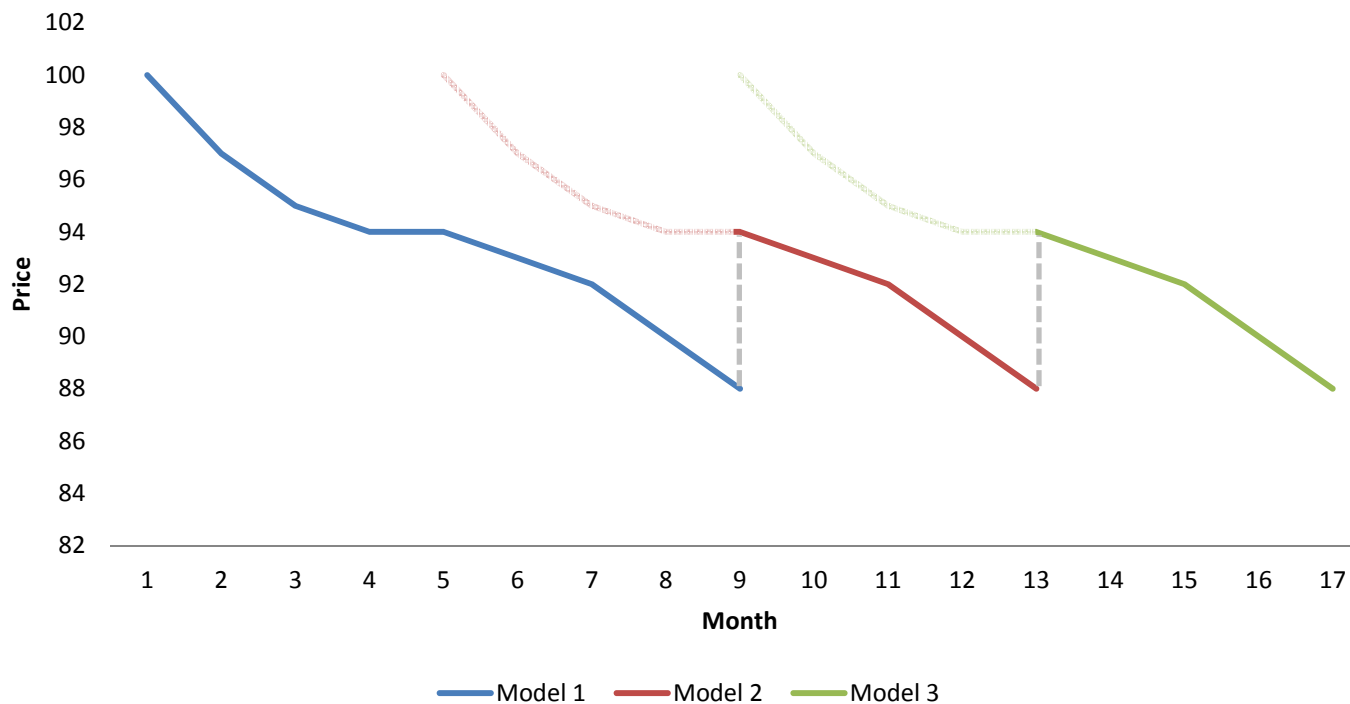


Preliminary Results

- 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?

