Alternative Data Sources and Hedonic Price Indexes for High-Tech Products in the U.S. Consumer Price Index

Craig Brown

Senior Economist | Bureau of Labor Statistics | Consumer Price Index

Jeremy Smucker

Economist | Bureau of Labor Statistics | Consumer Price Index

AEA · 2024 ASSA Annual Meeting · SGE Hosted Session

San Antonio, TX · January 7, 2024

Background – Non-Traditional Data

BLS has a goal of transitioning a portion of the CPI market basket from traditional data collection to non-traditional data sources and collection modes

Non-traditional data often contains a universe of prices, an improvement upon CPI's sample

Non-Traditional data often includes more detailed, more accurate specification data, which can be used in hedonic indexing techniques

Using non-traditional data reduces both respondent burden and data collection costs

Non-traditional datasets often collect multiple prices per month, an improvement upon BLS's pricing strategy

Background – Methodology Choice

CPI Methodology

Tracks actual prices of the same item over time until it leaves the market. Then, calculates a quality-adjusted relative comparing the entering and exiting items.

Hedonic Imputation

Tracks predicted prices for items over time using a hedonic regression model. Can also impute relatives for entering and exiting goods.

Time Dummy

Estimates a single hedonic model where time periods are included as dummy variables.

Data Summary – Prices & Characteristics

Televisions

- Data collected from API of national retailer
- Prices collected 3 times per month
- Approximately 400 televisions per month
- No geographic information, but outlet has national pricing
- Each row a unique observation based on UPC/SKU
- 94 Unique Specification Variables

Wireless Telephone Services

- Data purchased from a 3rd party vendor
- Collected via web-scraping and non-automated methods
- Average of 170 plans per month
- High geographic and provider overlap w/ CPI
- Each row is a unique service plan package
- Detailed information about offer prices and service plan features

Data Summary – Weights

Televisions

- Panel survey market data
- Over 60,000 reported television purchases per year
- Purchases categorized by product specification including screen size, resolution, and brand
- Geographic and demographic information included but not used

Wireless Telephone Services

- Household survey data
- Approximately 30,000 surveys per year
- Service plans categorized by product specifications including type of account, number of serviced line per plan, amount of included data, and provider
- Geographic and demographic information included but not used

Hedonic Imputation Process

Both projects have the same core indexing process

01	**CLEAN DATA AND ADD WEIGHTS**	Datasets are cleaned (log of price, entering/exiting goods identified, ineligible obs removed) and weights are derived and added at this step.
02	**BUILD WEIGHTED REGRESSIONS FOR T & T-1**	Weighted regressions are built for t and t-1 datasets with log(price) as the dependent variable and product specification variables as independent variables.
03	**PREDICT LOG(PRICE) FOR ALL OBS IN T & T-1**	Prices are predicted for observations in t and t-1. Prices are predicted for entering, exiting, & continuing obs in both periods. Residual change predicted for TVs
04	**CALCULATE AND AGGREGATE RELATIVES**	Calculate monthly obs-level relatives by comparing t and t-1 predicted prices. Aggregate obs-level relatives (Tornqvist aggregation) to create an index.

Aggregation Formula

Used in Step 4

$$\frac{P_t}{P_{t-1}} = \prod_{t-1}^{n} \left(\frac{p_{it}}{p_{i,t-1}}\right)^{\frac{1}{2}\left[\frac{p_{i,t-1}q_{i,t-1}}{p_{t-1}q_{t-1}} + \frac{p_{i,t}q_{i,t}}{p_t q_t}\right]}$$

Where:

i = an individual observation

p = predicted price

t = the current month

t-1 = the previous month

q = estimated expenditure

Weighting Scheme

Weights are estimated from expenditures

- Obtain 3rd party survey data which segments expenditures by product specifications
- Identify price-determining specifications and calculate expenditure shares for groups of specifications:
  - For example, of all TVs purchased, 2.5% of expenditures were for 4k, Medium Size, Brand Z TVs.
- The wireless telephone services project allocates spec group expenditure equally to all observations within a spec group
- Instead of equally weighting within a spec group, the televisions project allocates spec group expenditure to individual televisions by each observation's share of monthly reviews

$$Televisions\colon\ q_{i,t} = group\_expend * \left(\frac{review\_count_{i,t}}{review\_count_{group,t}}\right)$$

$$Wireless\ Telephone\colon\ q_{i,t} = group\_expend * \left(\frac{1}{observation\_count_{group,t}}\right)$$

Where:
i = an individual observation
group_expend = expenditure in the observation's spec group from a base period
review_count_i,t = number of website reviews for the observation in time period t
review_count group,t = sum of website review in the observation's spec group
observation_count_group_t = number of observations in a spec group

Methodological Variations

Televisions

- Uses methodology outlined in Erickson and Pakes (2011)
- Includes an adjustment for predicted residual change to capture the effect of unobserved characteristics

$$Relative = e^{(\text{predicted\_logprice}_t - \text{predicted\_logprice}_{t-1} + \text{predicted\_residualchange})}$$

Wireless Telephone Services

- Uses methodology outlined in Pakes (2003)
- Calculates relatives through a simple comparison of predicted prices

$$Relative = e^{(\text{predicted\_logprice}_t - \text{predicted\_logprice}_{t-1})}$$

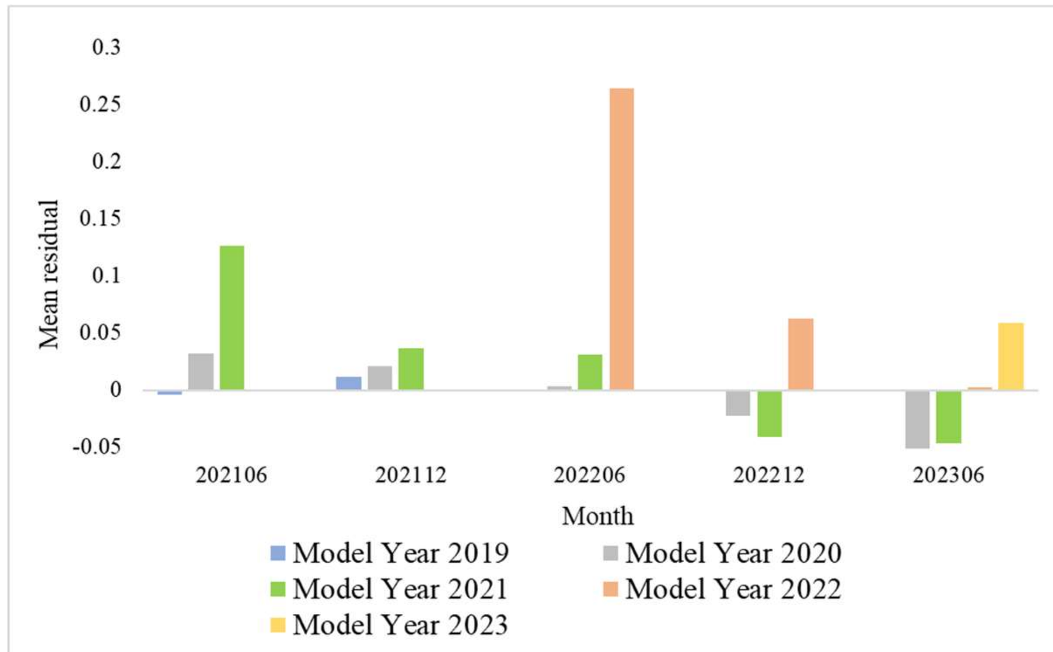
Both projects impute **entering goods** as well as exiting goods.

Neither Pakes (2003) or Erickson and Pakes (2011) imputes the prices of entering good.

For Televisions, like Ehrlich, et al (2023), we assume the t-1 residual is 0 for entering goods.

Methodological Choice – Televisions

- Predicted television prices tend to fall at a slower rate than actual offer prices. This leads to declining residuals throughout the product life cycle.



Methodological Choice – Televisions

- Predicted television prices tend to fall at a slower rate than actual offer prices. This leads to declining residuals throughout the product life cycle.
- As discussed in Erickson and Pakes (2011), this is due to unobserved characteristics in the TVs market.
- We can simulate this effect by building hedonic indexes with more limited regressor sets and analyzing the results.
- To control for this effect, we introduce the 2nd-stage residual change adjustment.

Methodological Choice – Wireless Telephone Service

- Television prices drop during their product life cycle as they become obsoleted by higher-tech entrants.
- Wireless telephone service plans tend to be replaced by new plans with no overlap. Unique service plans tend to not change in price during the product life cycle.
- Since Erickson and Pakes (2011) relies on intertemporal price change, Erickson and Pakes (2003) was selected for wireless telephone services.

Direction of 1 Month Price Change (w/ Actual Prices)

Type of Price Change (1 Month Relative)	Televisions	Wireless Telephone Services
Positive	24.3%	0.8%
None	34.7%	97.6%
Negative	41.0%	1.6%

Hedonic Model – Wireless Telephone Services

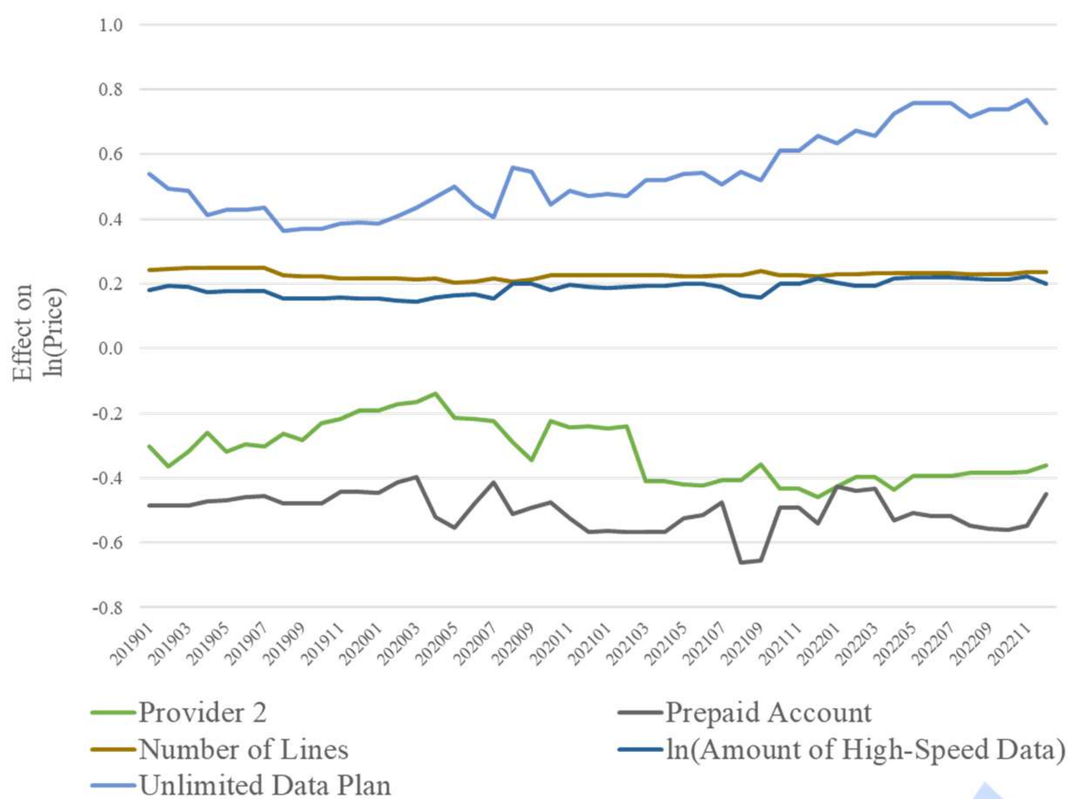
Monthly Pricing data from Jan 2019 - Dec 2022

Expenditure data:
- 2018 used in 2019/20
- 2019 used in 2021/22

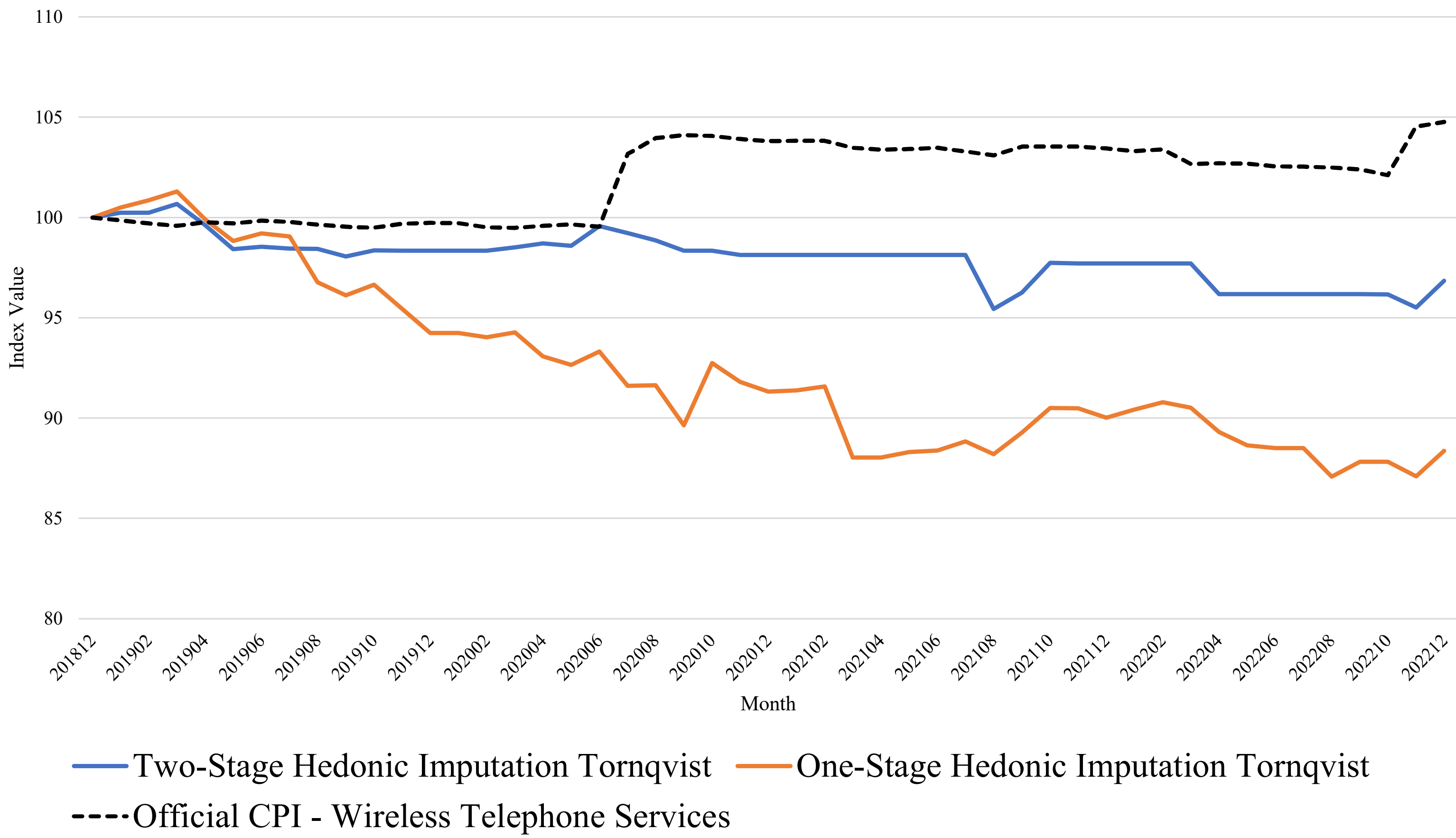
Hedonic Model is a monthly weighted regression on the log of plan price

Indicator Variables: Prepaid Plans, Unlimited Data Plans, Provider Dummy
Continuous Variables: Number of Serviced Lines, Amount of High-Speed Data

One parsimonious model specification used over the entire four-year period
- Average adjusted R-squared = 0.843
- Average root mean squared error = 0.158



Index Results – Wireless Telephone Services



Hedonic Model – Televisions

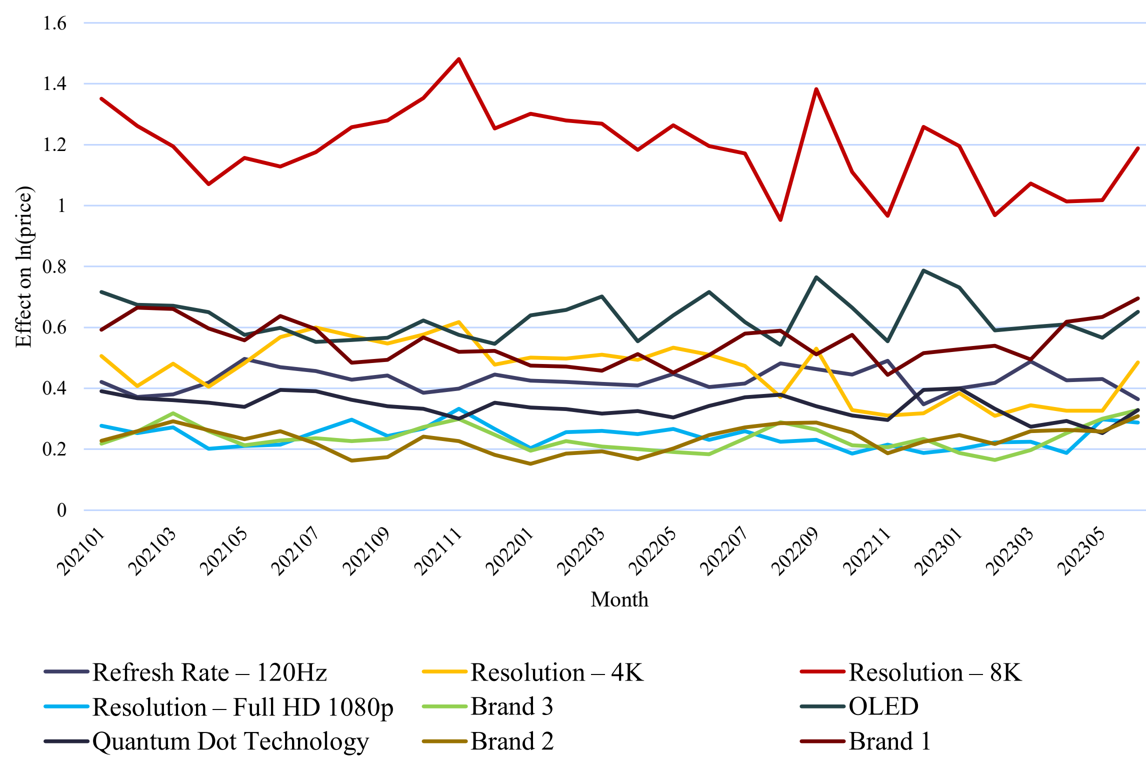
Monthly Pricing data from Jan 2020 - Jun 2023

Expenditure data:
- 3rd party survey data from concurrent quarter

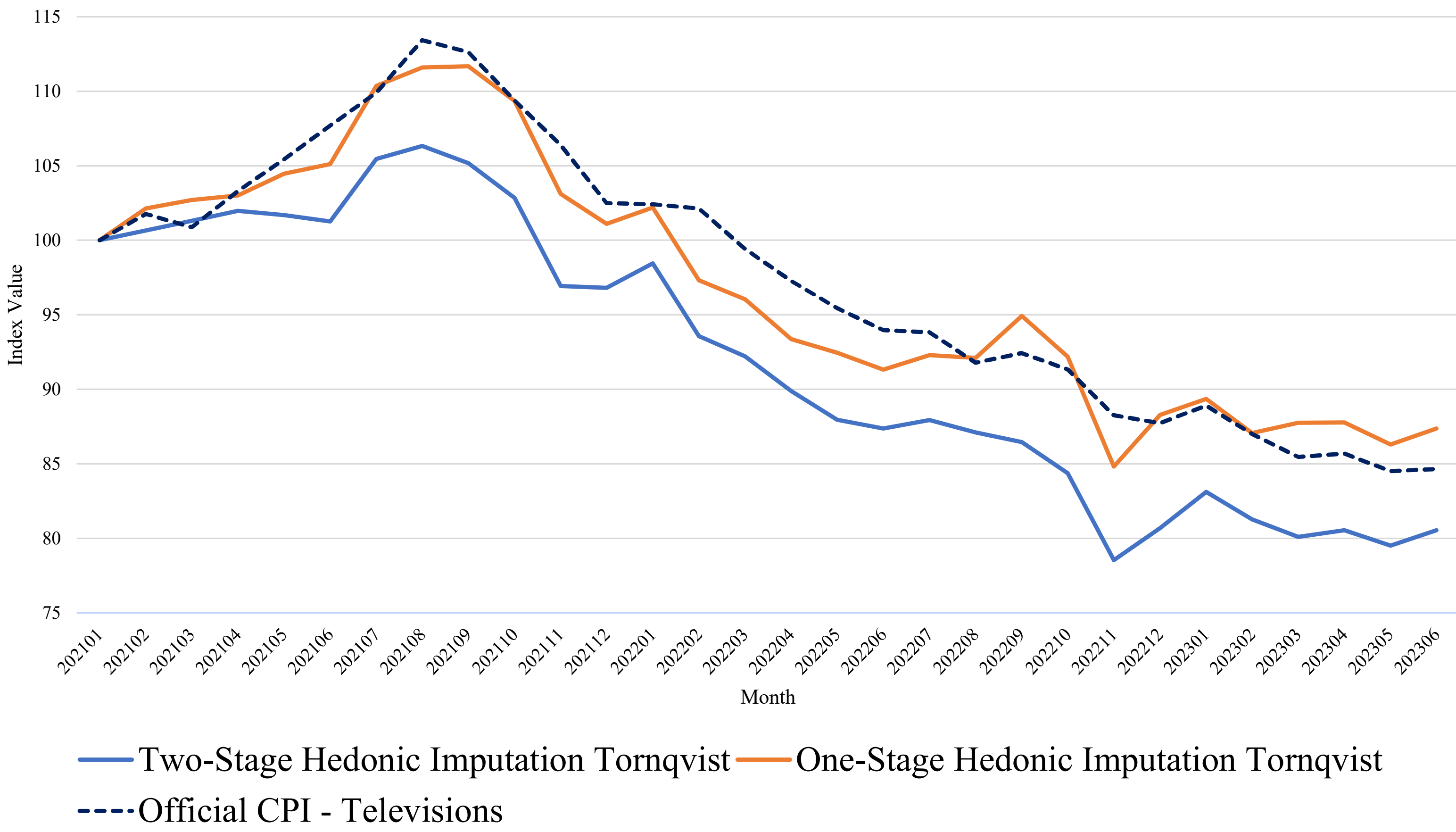
Hedonic Model is a monthly weighted regression on the log of television price

Indicator Variables: Quantum Dot, OLED, 120 Hz Refresh Rate, 8k, 4k, 1080p, Brand Variables
Continuous Variables: Screen Area, Screen Area Squared, Depth

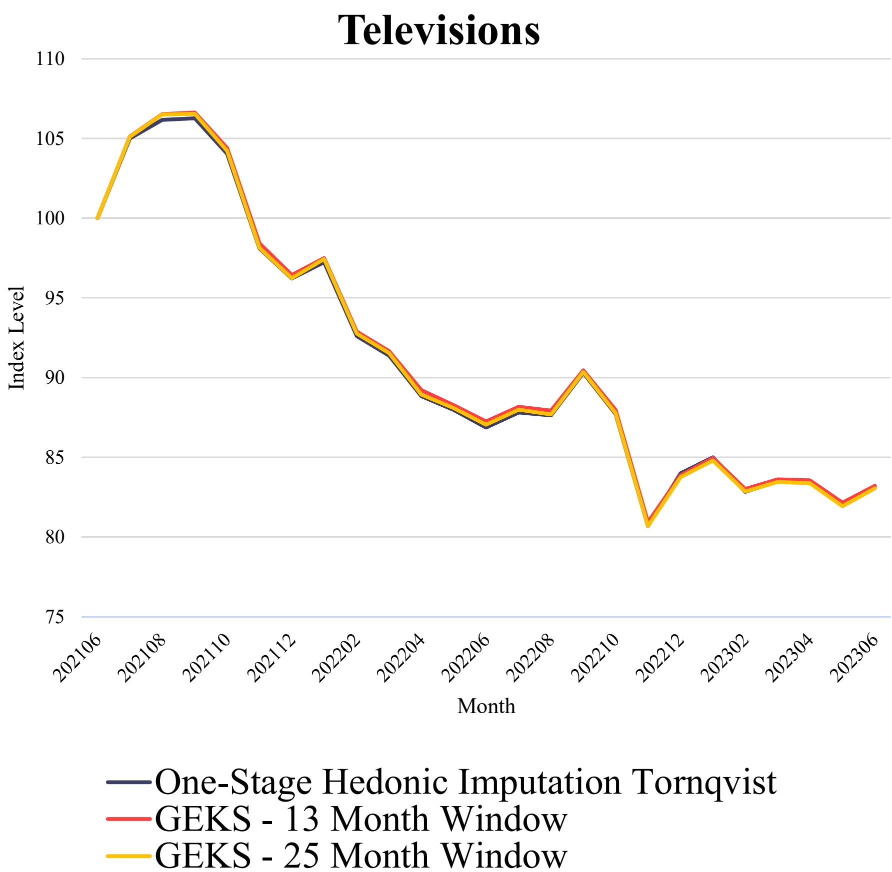
One parsimonious model specification used over the entire four-year period
- Average adjusted R-squared = 0.958
- Average root mean squared error = 0.056

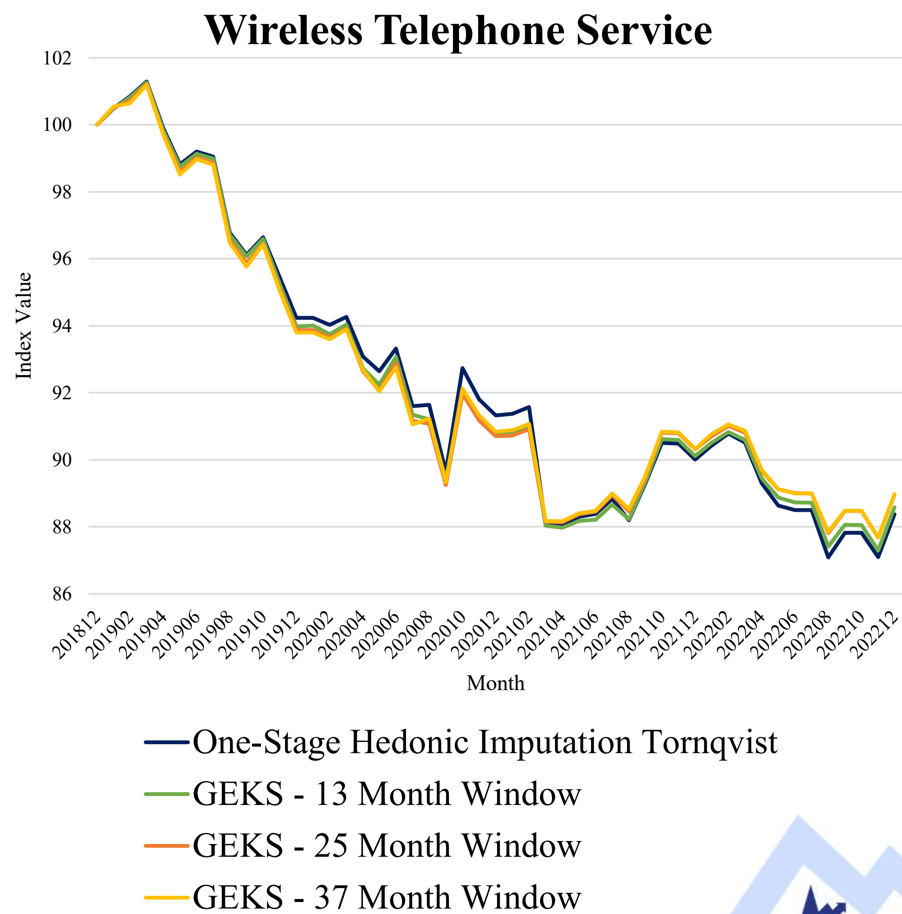


Index Results - Televisions



Multilateral Comparisons





Conclusion

Hedonic Indexes built with non-traditional data improve component indexes for the CPI in many ways

- These non-traditional data sources, which contain a universe of prices collected frequently, are more representative than the CPI's sample
- Non-traditional data sources, in conjunction with hedonic methods, incorporate new goods immediately
- Hedonic methods better capture quality change, especially for unobserved characteristics
- Incorporating 3rd party weights is an improvement over our traditional disaggregation process used in the initiation of goods

Thank you

brown.craig@bls.gov

smucker.jeremiah.@bls.gov