USDA Economic Research Service
U.S. DEPARTMENT OF AGRICULTURE

Building A Public-Use Small-Area Panel Price Index Database Using Scanner Data

Chen Zhen
University of Georgia, Athens

Abigail Okrent
Economic Research Service

Lan Mu
University of Georgia, Athens

Gauri Datta
University of Georgia, Athens

FCSM Research & Policy Conference
October 25-27, 2022

Acknowledgment

- This work was funded under NIFA grant #2019-67023-29672.

- Any views expressed are those of the authors and should not be attributed to ERS, USDA, or IRI.

Outline of presentation

- Motivation for creating small-area panel price indexes
- Methods
- Data
- Preliminary results
- Further research

Price information is limited in many surveys of food purchase or consumption.

- Prices are important determinants of consumer food choice and knowing this relationship is fundamental to examining many food policy analyses (e.g., taxes, subsidies)
- Examples of public data sets used in food policy analyses
  - BLS's Consumer Expenditure Survey
  - NCHS's National Health and Nutrition Examination Survey
  - CDC's Behavioral Risk Factor Surveillance System
  - ERS's 2012 FoodAPS (field test for FoodAPS-2 happening now)

Current public-use price information not granular enough to conduct food policy analyses.

- BLS's Consumer Price Indexes
- BEA's Regional Price Parities
- C2ER average prices
- ERS's Quarterly Food-at-Home Price Database (discontinued)
- ERS's Monthly Food-at-Home Price Database (MFAHPD) (forthcoming)

The MFAHPD is a step in the right direction but…

- Covers only 10 MSAs
- Limited to 82 ERS Food Product Groups, including 8 nonalcoholic beverages
  - –Sweetened coffee and tea
  - –Unsweetened coffee and tea
  - –Flavored milk and other sweetened milk-based beverages
  - –Low-calorie beverages
  - –All other caloric beverages
  - –Fresh, 100% fruit/vegetable juice
  - –Frozen, 100% fruit/vegetable juice
  - –Canned/shelf-stable, 100% fruit/vegetable juice

This research provides more granular price information across geographies and products than MFAHPD.

- Roughly 5,000 Census places
- For 24 What We Eat In America nonalcoholic beverage categories

  - ✓ Citrus juice
  - ✓ Apple juice
  - ✓ Other fruit juice
  - ✓ Vegetable juice
  - ✓ Diet soft drinks
  - ✓ Diet sport and energy drinks
  - ✓ Soft drinks
  - ✓ Fruit drinks
  - ✓ Sport and energy drinks
  - ✓ Nutritional beverages
  - ✓ Smoothies and grain drinks
  - ✓ Milk, whole
  - ✓ Milk, reduced fat
  - ✓ Milk, low fat
  - ✓ Milk, no fat
  - ✓ Flavored milk, whole
  - ✓ Flavored milk, reduced fat
  - ✓ Flavored milk, low fat
  - ✓ Flavored milk, nonfat
  - ✓ Milk shakes and other dairy drinks

Similar to MFAHPD, panel (time and geography) price indexes constructed using rolling-window GEKS.

- Multilateral indexes like the Gini-Eltetö and Köves-Szulc (GEKS) price index formula allows for price indexes to be transitive:

$$P_{GEKS}^{0j} = \prod_{l=0}^{M} \left(P_F^{0l} \times P_F^{lj} \right)^{1/(M+1)}$$

Fisher-Ideal price index between base 0 and entity l

Fisher-Ideal price index between entity l and entity j

- This index also avoids chain drift but to do this, requires revision of published index numbers when new data arrive.
- Adding rolling windows to GEKs price index formula solves revision problem such that the rolling window GEKS for entity k in month $T + 1$ (T = end of base period) is:

$$P_{RWGEKS}^{0k} = P_{GEKS}^{0j} \prod_{l \in I_{T+1:T-11}} \left(P_F^{jl} \times P_F^{lk} \right)^{1/M_{T+1:T-11}}$$

Link entity j in T

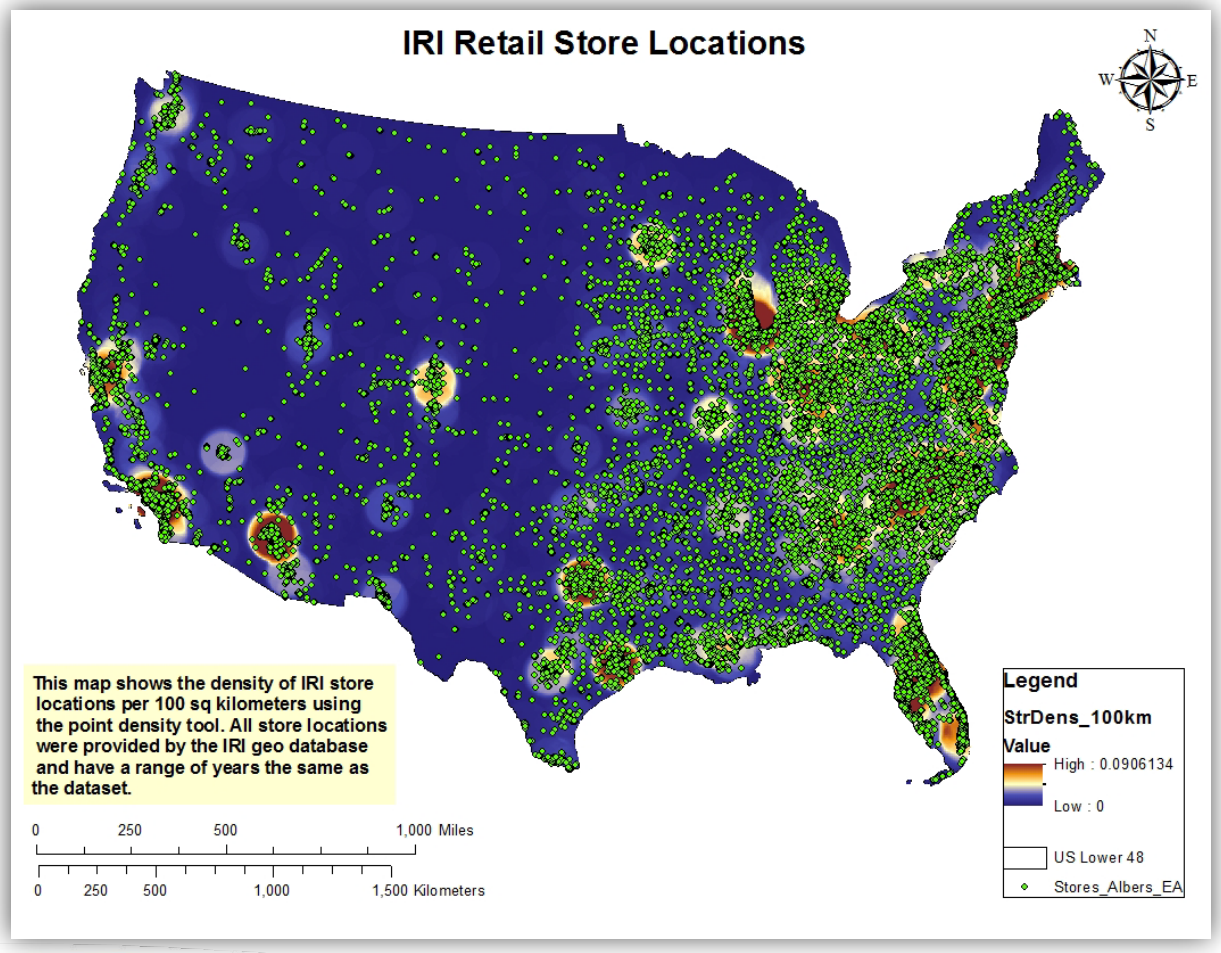
the set of all entities between T−11 and T+1 (window = one year)

the number of entities in the set $I_{T+1:T-11}$

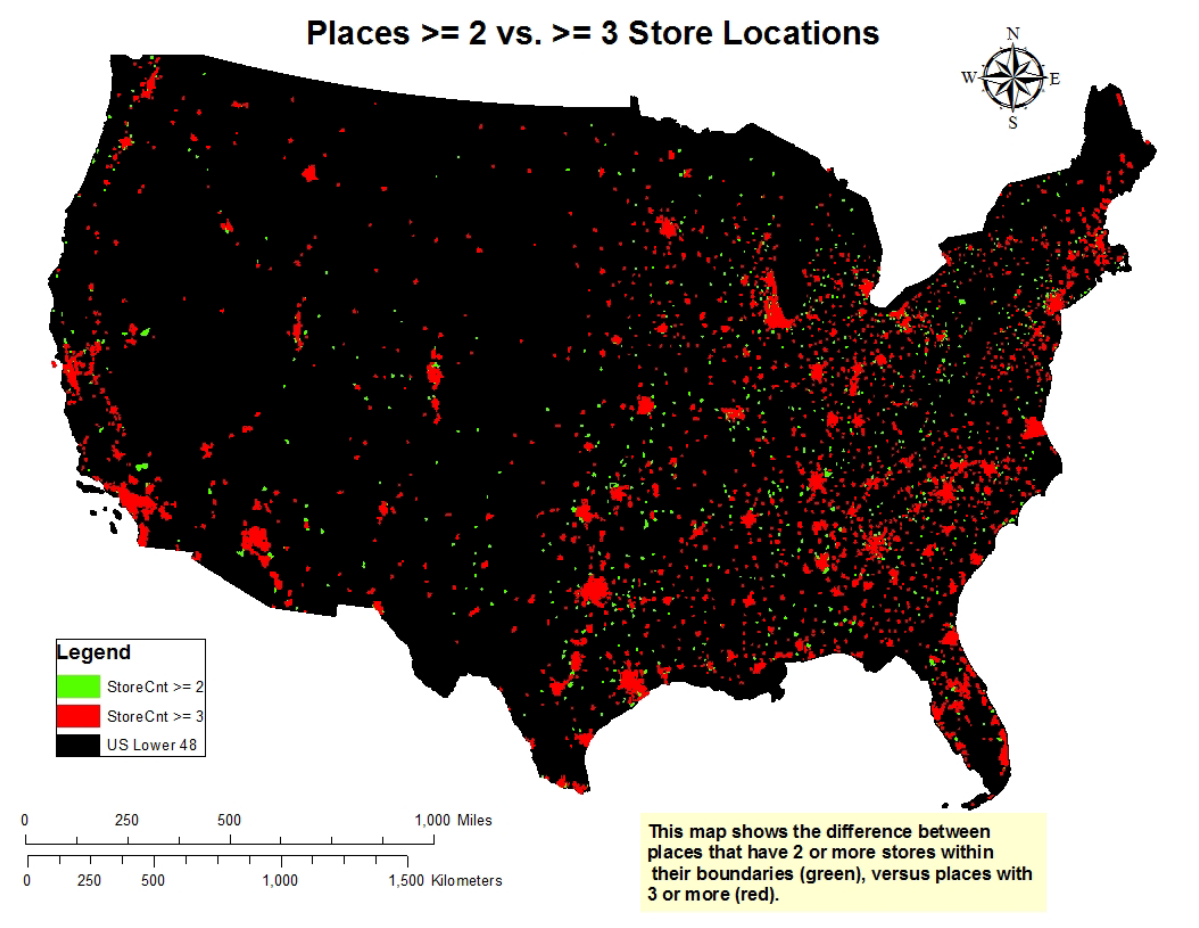
IRI Infoscan data is primary data set for constructing the rolling window-GEKS spatial price indexes.

- Store- and barcode-level sales for over 40,000 food stores in contiguous United States that agree to release information to ERS.
- Additional 18,000 food stores aggregate barcode-level sales into retail marketing areas, which can range from MSA-sized areas to Census division-sized areas.
- InfoScan data represented an estimated 41.0% of retail establishments and 55.3% of retail food sales compared with Census Bureau benchmarks in 2012 (Muth et al. 2016).
- Also use ERS's Purchase to Plate Crosswalk to assign UPCs to 1 of 24 What We Eat In America nonalcoholic beverage categories.

Some areas of the United States are not covered in the IRI retail data that ERS purchases.

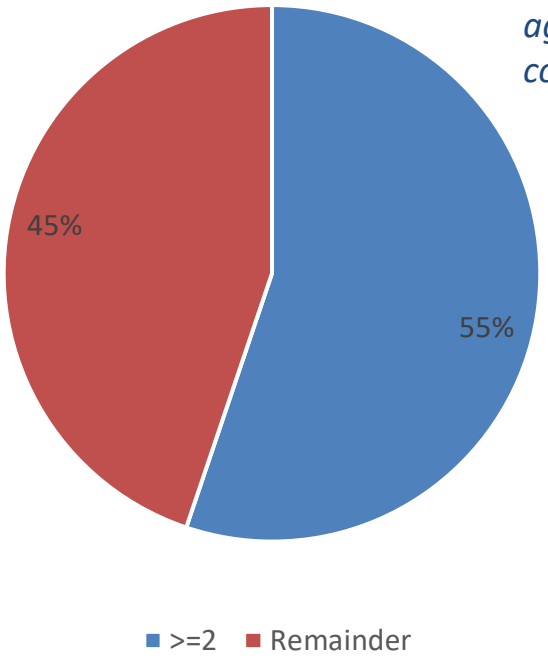


Because of disclosure purposes, rolling window GEKS prices only calculated for Census places with 3 or more stores.



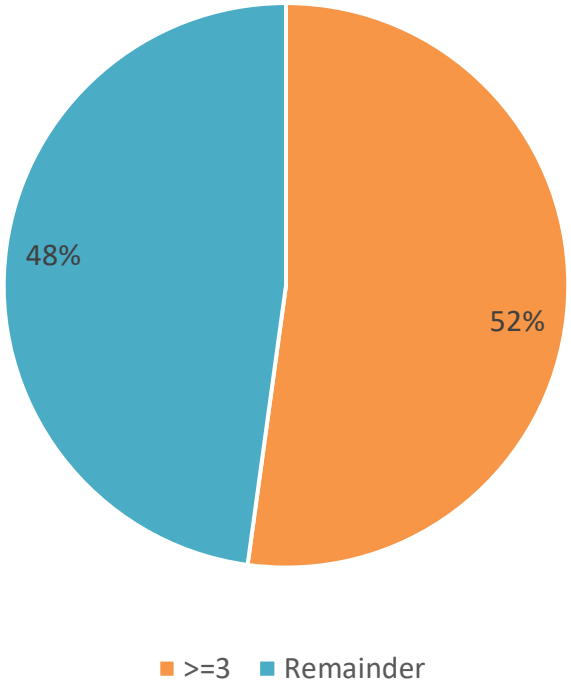
Population within Census places with more than 2 stores is 168 million (55% of total U.S. population) whereas those with more than 3 stores is 159 million (52%).

Population within Census places with 2 or more stores



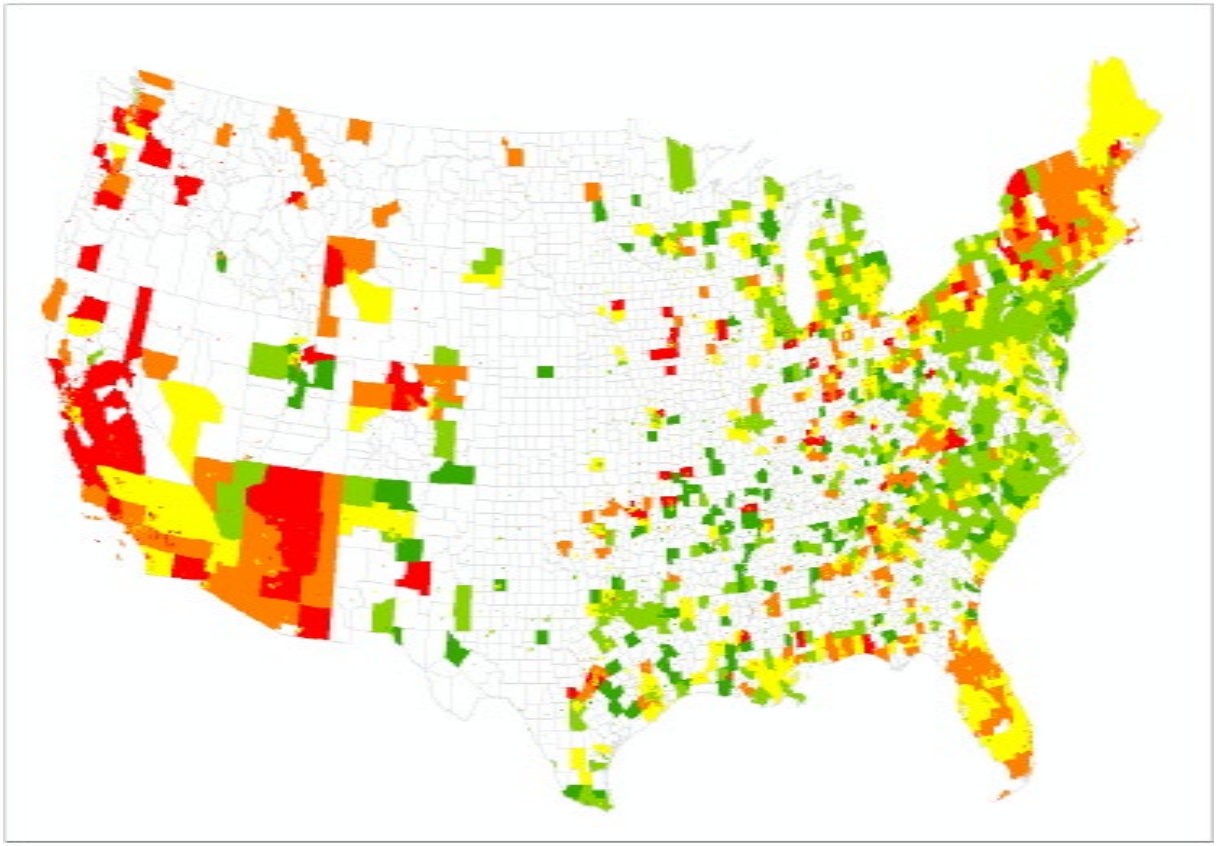
Remainder stores are aggregated into counties

Population within Census places with 3 or more stores



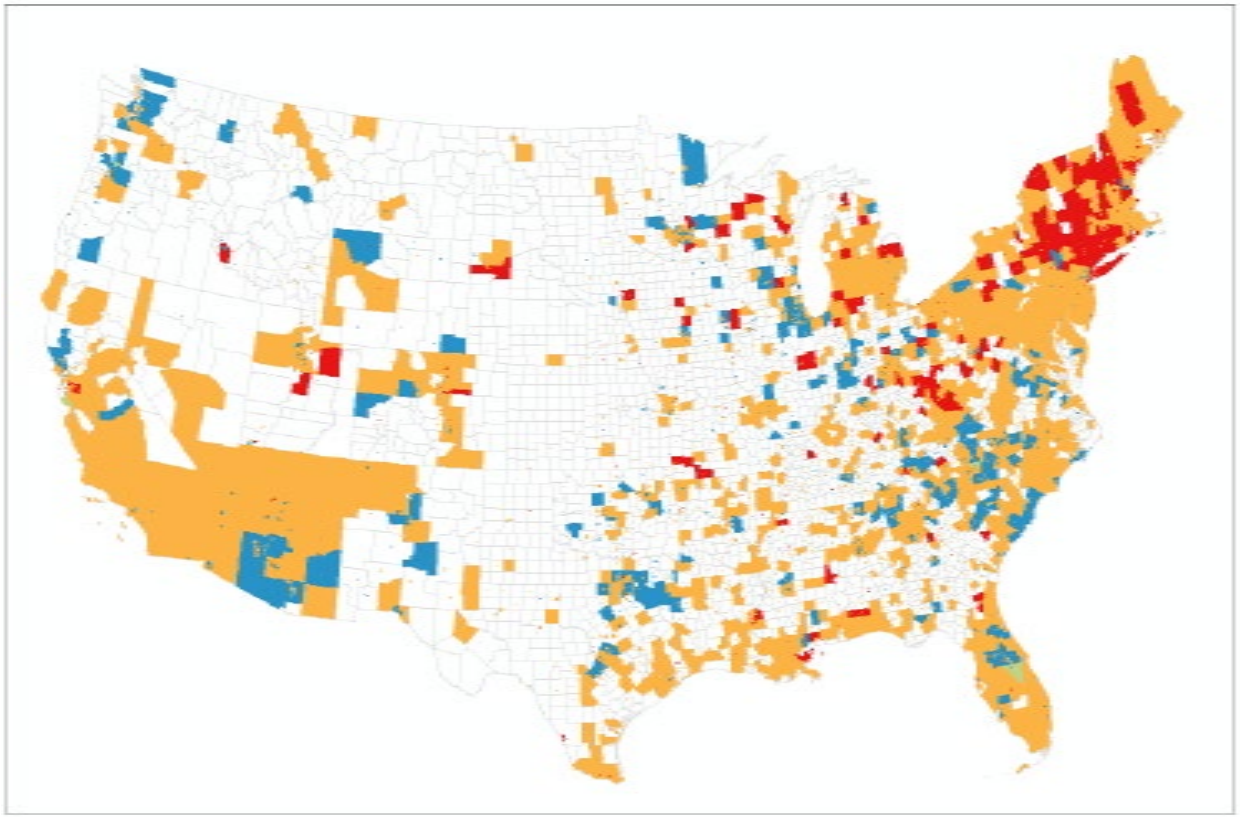
Preliminary results show prices for fruit-flavored soft drinks (caffeine free) in Census places on the West coast are 8 to 33% higher than the US average.

Mean GEKS index for Census places with 3 or more stores (excluding RMA stores), 12/31/2008-1/10/2010



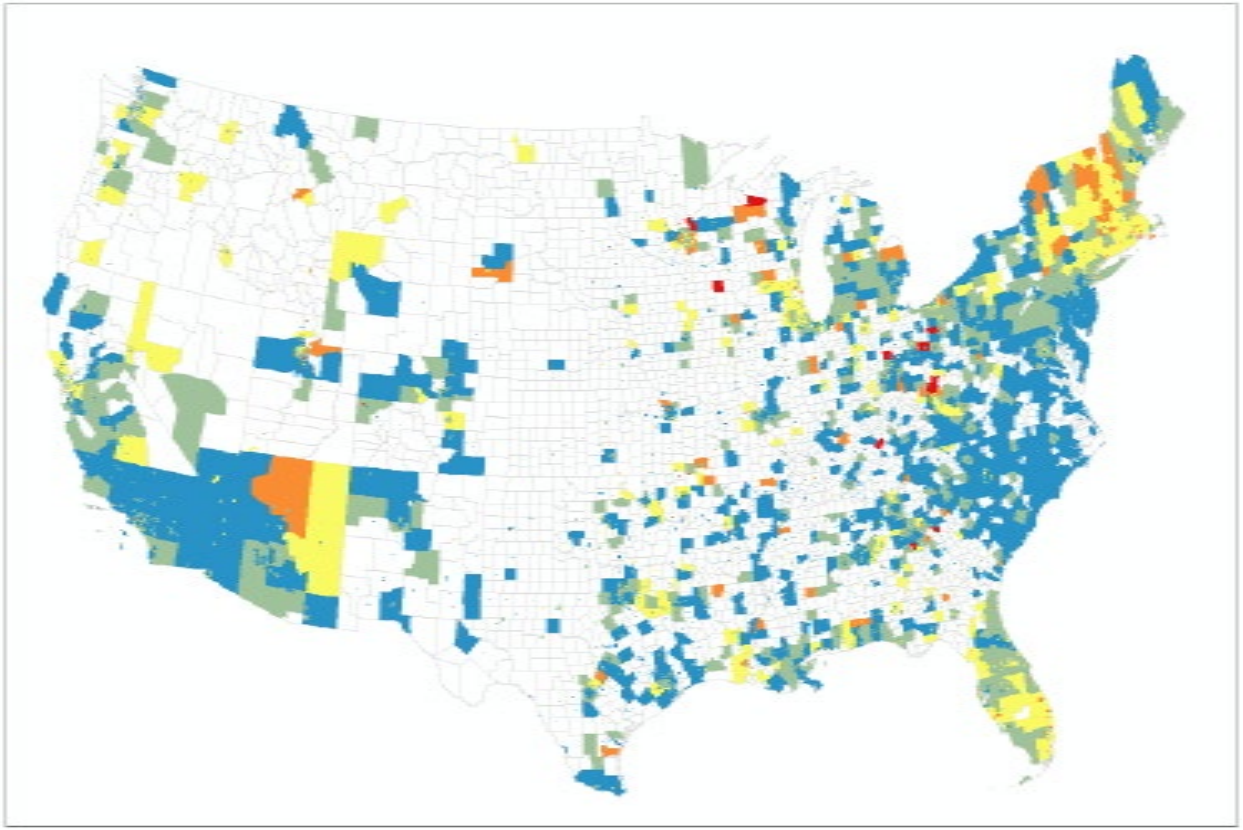
The Census places with the largest price increases between 12/31/08 to 1/10/10 are in the New England area and a handful of places in Midwest.

Change in GEKS index for Census places with 3 or more stores (excluding RMA stores), 12/31/2008-1/10/2010



The mean standard deviation across all Census places is 0.05, which implies 95% of all observations are within 0.1 points around the mean index for a particular place.

Standard deviation of GEKS indexes for Census places with 3 or more stores (excluding RMA stores), 12/31/2008-1/10/2010



The next steps will be:

- Include stores for chains that only provide information at their retail marketing area
  - Assume barcode level prices are the same across stores of a particular chain in a particular retail marketing area (DellaVigna and Gentzkow 2019; Dong 2022)
  - Impute store-level sales for these chains using TDLinx sales information
    - TDLinx has information for close to a census of food retailers, including *total sales,* square footage, number of employees, number of checkout registers.
    - One way to do this is proportionately assign *food and beverage* retail marketing sales volume to stores within them based on store-level *total sales* in TDLinx.
- Calculate GEKS for remaining 23 nonalcoholic beverage products
- Calculate GEKS for other food subgroups, e.g., processed foods, fruits, vegetables

Limitations

- Stores are nonrandom subset of all stores in the contiguous United States.
- Unlike MFAHPD, no weighting of stores to be more representative at regional level.
- UPCs sold at different chains are treated as the same product. This could introduce bias if shoppers see it as different products depending on where it is sold.

Like, Share, & Follow ERS

www.ers.usda.gov

www.ers.usda.gov/data-products/charts-of-note

@USDA_ERS

linkedin.com/company/usda-economic-research-service

Subscribe to Weekly E-mail Notifications: www.ers.usda.gov/subscribe

Learn About Careers at ERS: www.ers.usda.gov/about-ers/careers-at-ers