A Preliminary Study to Evaluate Variance of Blended Index from Establishment Survey and Administrative Data

Daniel Yang, Daniell Toth
OSMR, U.S. Bureau of Labor Statistics
OPLC Team: Susan Fleck, Steven Paben, Jeffrey Blaha
Xiao Fan, Chris Grant, Larry Lang

2024 Federal Committee on Statistical Methodology (FCSM) Research & Policy Conference, Hyattsville, MD
October 24th, 2024

Disclaimer

Disclaimer: This presentation provides a summary of research results. The information is being released for statistical purposes, to inform interested parties, and to encourage discussion of work in progress. The presentation does not represent an existing, or a forthcoming new, official BLS statistical data product or production series.

Overview

I. Survey Data Source: IPP survey of ~3,000 companies.
II. Administrative Data Source: Census Trade Data (CTD).
III. Blended Import and Export Price Indexes (MXPI).
IV. Variance Estimation of Blended Index (Resampling Methods).
V. A Small Subset Study and Summary.

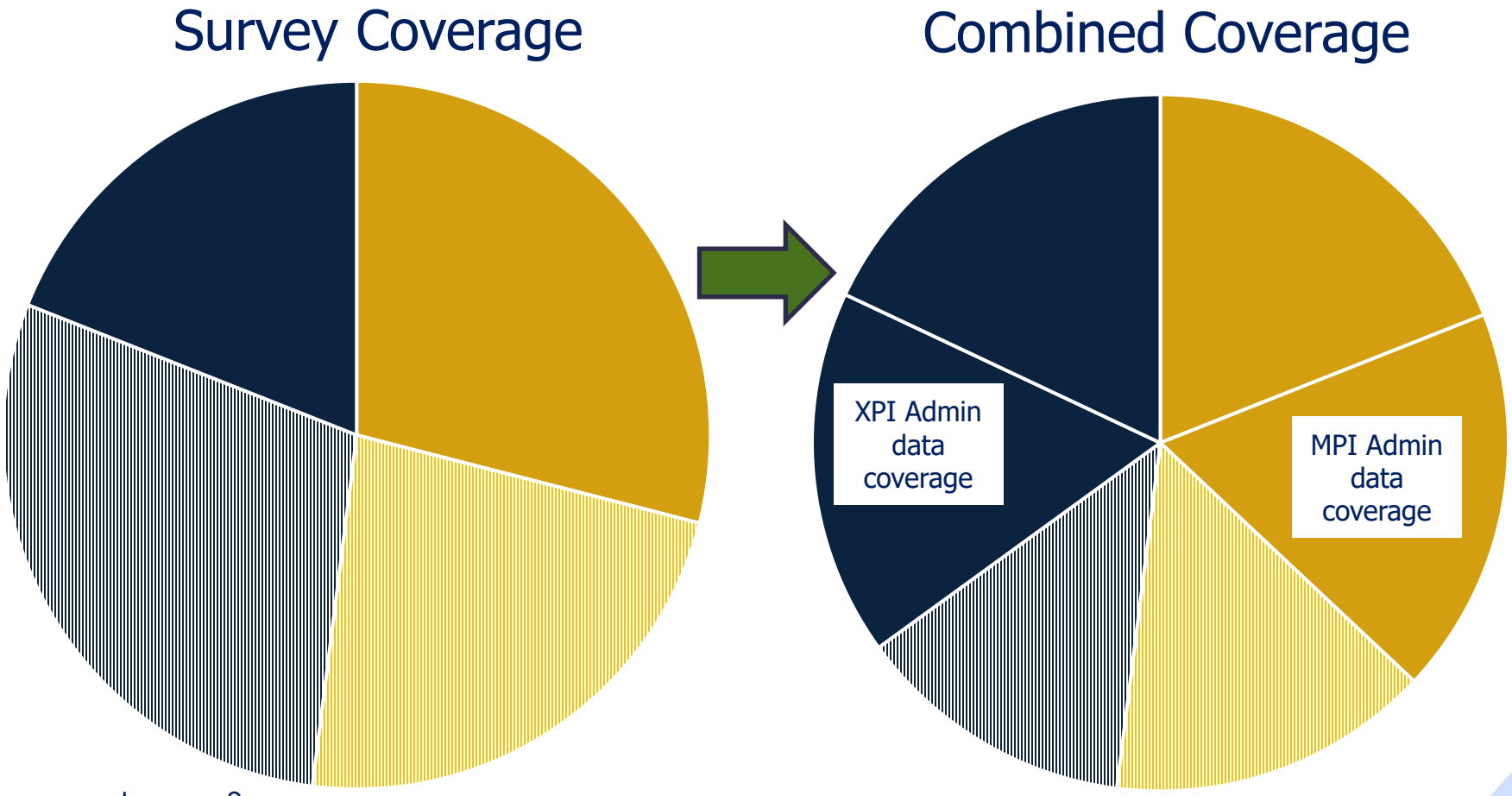
I. International Price Program Survey

- Establishment survey collects data on U. S. trade with foreign nations and produces Import and Export Price Indexes (MXPI).
- MXPI measure price changes of U.S. imports and exports.
  - Principal Federal Economic Indicator.
- The survey collected a monthly stratified sample of approximately 15,000 import and 9,000 export price quotes from roughly 4,000 establishments (McCulley and Mead 2016).

II. Census Trade Data (CTD)

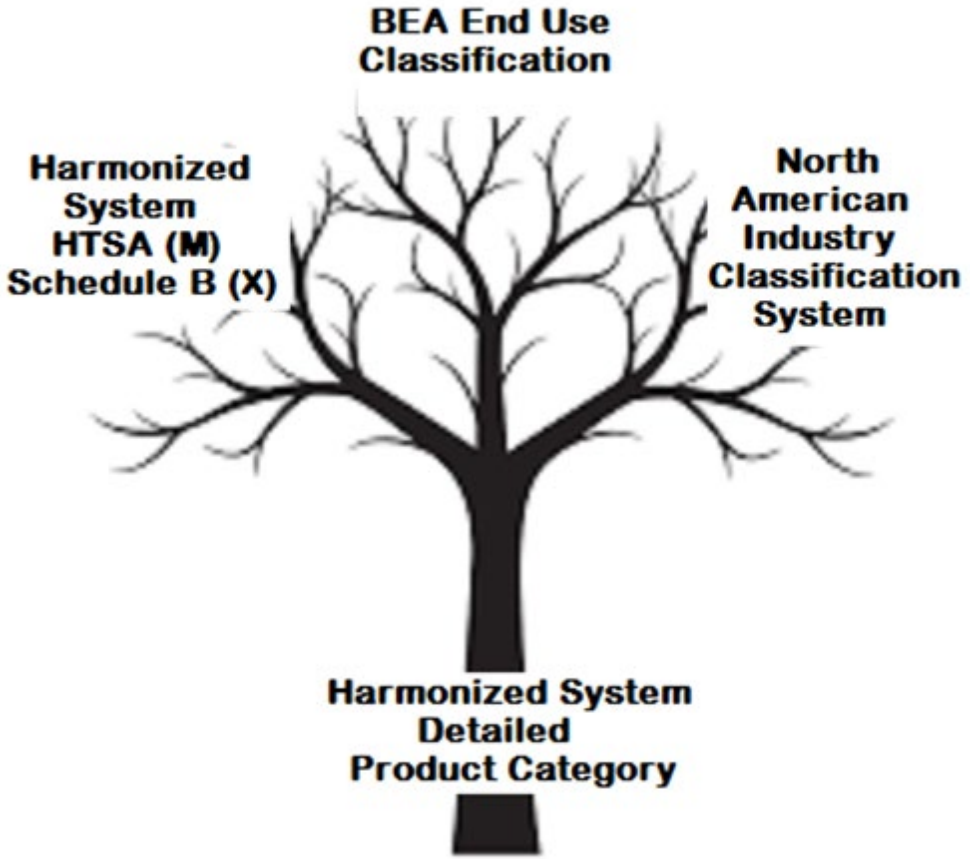
- Number of publishable detailed MXPI has declined over time.
- BLS is planning to implement an alternative data source: CTD for replacing direct data collection (in commodity product areas).
  - CTD contains detailed shipment records for imports and exports of merchandise goods.
- CTD complements direct data collection resulting in two contiguous non-overlapping sets of product categories for all goods trade.

Census Trade Data Increases Published Detailed MXPI



Harmonized System Detailed Product

Categories are Basic Classification Group to Calculate Aggregate MXPIs



III. Blended MXPI: Index from two Data Sources

- Laspeyres index formula:

$$LTR_t = \left(\frac{\sum p_{i,t}q_{i,0}}{\sum p_{i,0}q_{i,0}}\right)(100) = \left(\frac{\sum w_{i,0}r_{i,t}}{\sum w_{i,0}r_{i,t-1}}\right)(P_{t-1}) = (STR_t)(LTR_{t-1}).$$

LTR_t : long term relative of a collection of items at time i,

$p_{i,t}$: price of item i at time t,

$q_{i,0}$: quantity of item i in base period 0,

$w_{i,0} = p_{i,0} \times q_{i,0}$: total revenue of item i, in base period 0,

$r_{i,t} = p_{i,t}/p_{i,0}$: long term relative of item i at time t,

$STR_t = \frac{\sum w_{i,0}r_{i,t}}{\sum w_{i,0}r_{i,t-1}}$: short term relative of a collection of items i, at time t.

III. Blended MXPI Estimate

- A blended estimator of population total:

$$\hat{Y} = \lambda \hat{Y}^{(1)}_{\{1,2\}} + (1-\lambda) Y^{(2)}_{\{1,2\}}$$

- λ– proportion
- The blended MXPI estimator from (establishment) survey and Census:

$$\hat{P}_{sl,t}(\lambda_{sl}) = \lambda_{sl} \hat{P}^{DC}_{c,t} + (1-\lambda_{sl}) \hat{P}^{CTD}_{c,t}$$

- λ_{sl}– annually adjusted proportion (supported by MXPI trade)

IV. Variance Estimation of Blended MXPI

- The derivatives of $\hat{P}_{sl,t}(\lambda_{sl})$:

$$\frac{\partial P_{sl,t}}{\partial \lambda_{sl}} = \hat{P}_{c,t}^{DC} - \hat{P}_{c,t}^{CTD},$$

$$\frac{\partial P_{sl,t}}{\partial P_{c,t}^{MXPI}} = \lambda_{sl}, \frac{\partial P_{sl,t}}{\partial P_{c,t}^{CTD}} = 1 - \lambda_{sl}$$

- Taylor Series Linearization (TSL) variance approximation:

$$\hat{V}\left(\hat{P}_{sl,t}(\lambda_{sl})\right) = \hat{\lambda}_{sl}^2 \hat{V}\left(\hat{P}_{c,t}^{DC}\right) + \left(1 - \hat{\lambda}_{sl}\right)^2 \hat{V}\left(\hat{P}_{c,t}^{CTD}\right)$$

Error Propagation: CTD's Advantage and Disadvantage

- CTD's primary advantage: provide values and quantities for almost all U.S. trade within a month.
- CTD's primary disadvantage:
  - Prices are average transaction prices of large shipments of goods.
  - Not as unique of a price as directly collected survey data.
- However, unit values are considered viable for homogeneous product areas.

Error Propagation Types from Data Sources

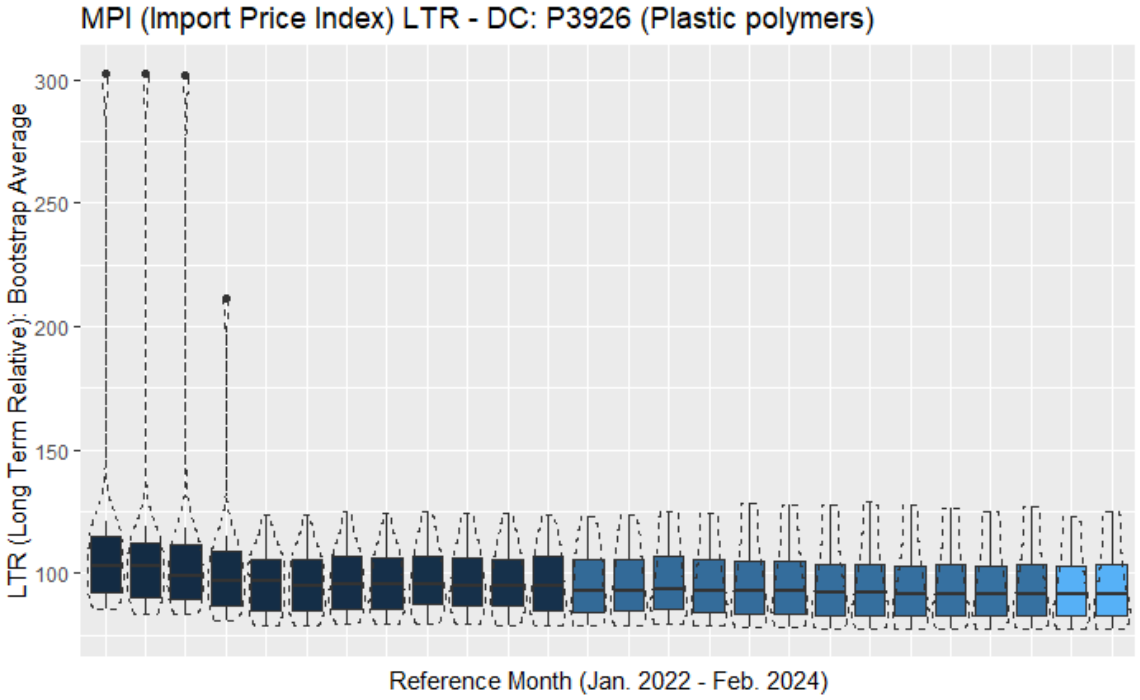
- MXPI Survey Data represents a share of trade across U.S. borders:
  - Sampling error.
- CTD represents more coverage on trade across U.S. borders:
  - Missing data error (some records do not have complete entries).
  - Outlier exclusion error (some records are excluded because their price deviates far from the mean).
  - Census means there is no sampling error, however, all sample transactions with 10-digit Harmonized System (HS) numbers are excluded for CTD areas.

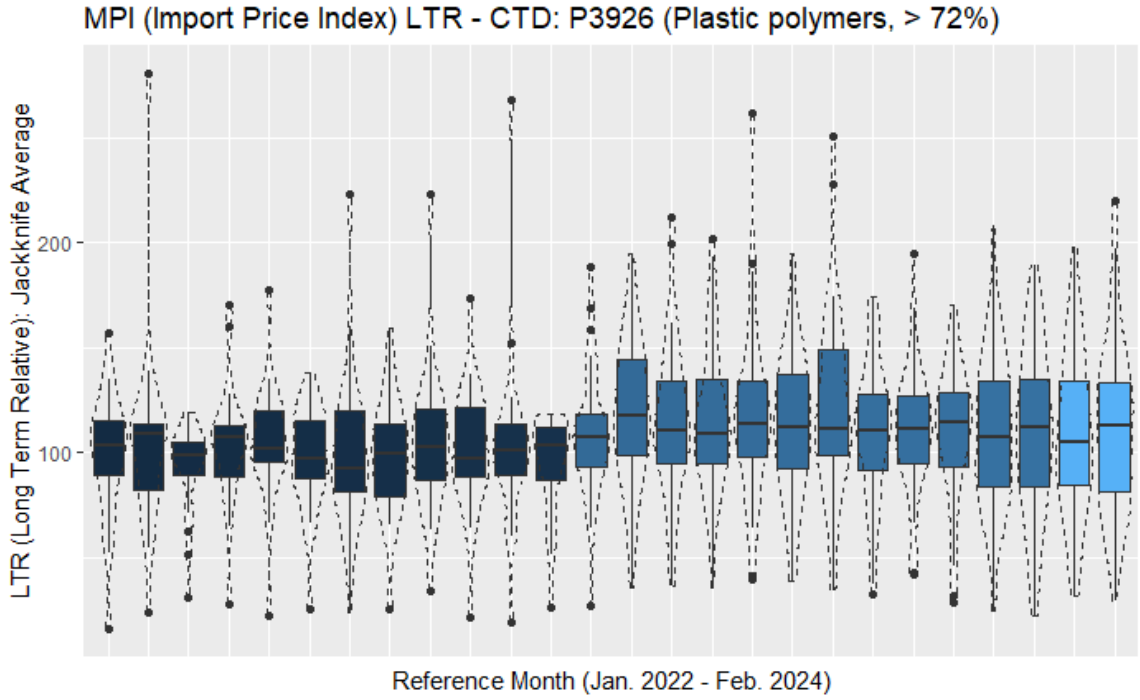
Error Propagation: Capture via Resampling Methods

- Error propagation in CTD is not in the sense of sampling errors in the directly collected (DC) survey.

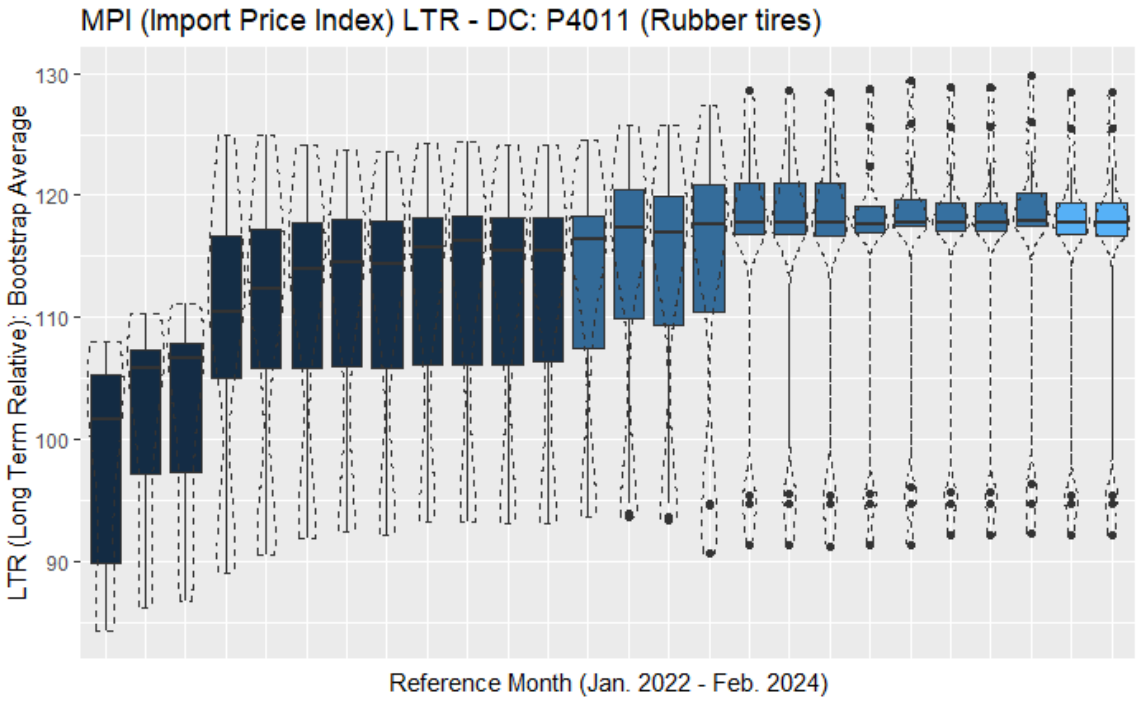
- Borrowing information from Census improves the estimate.
- Bootstrap resampling: conducted for DC monthly sample.
- Jackknife resampling: a feasible solution to gauge error propagation in CTD.

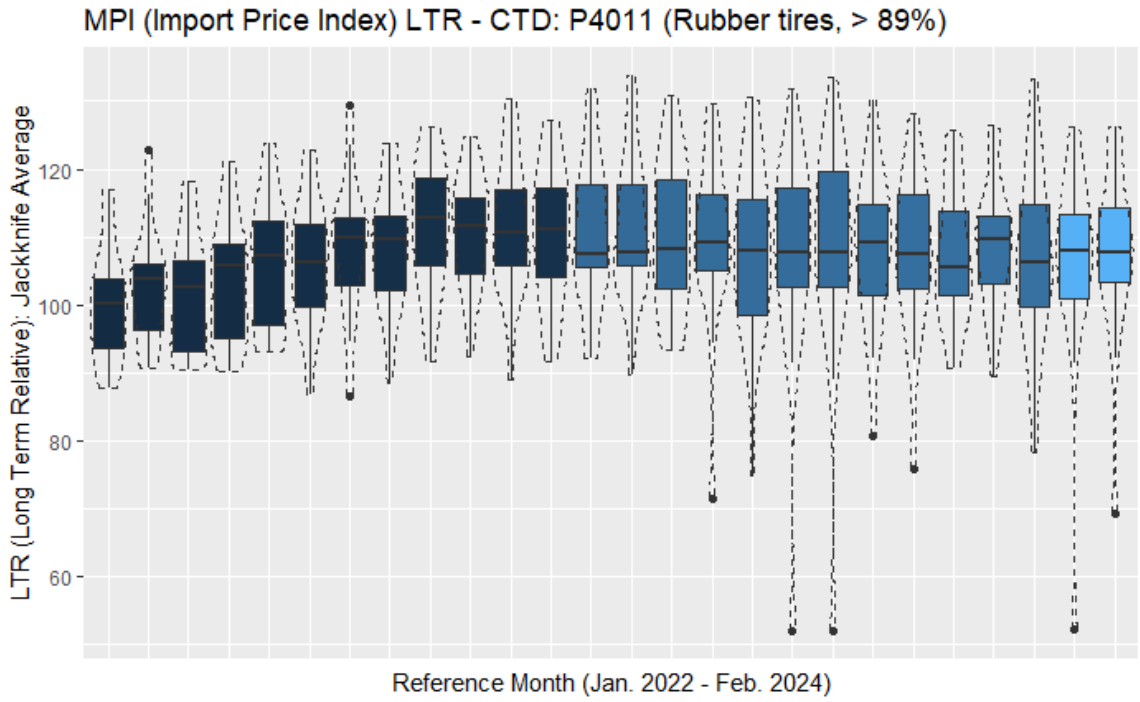
Plastic polymers (P3926) Violin Plots: DC and CTD (A Small Subset)



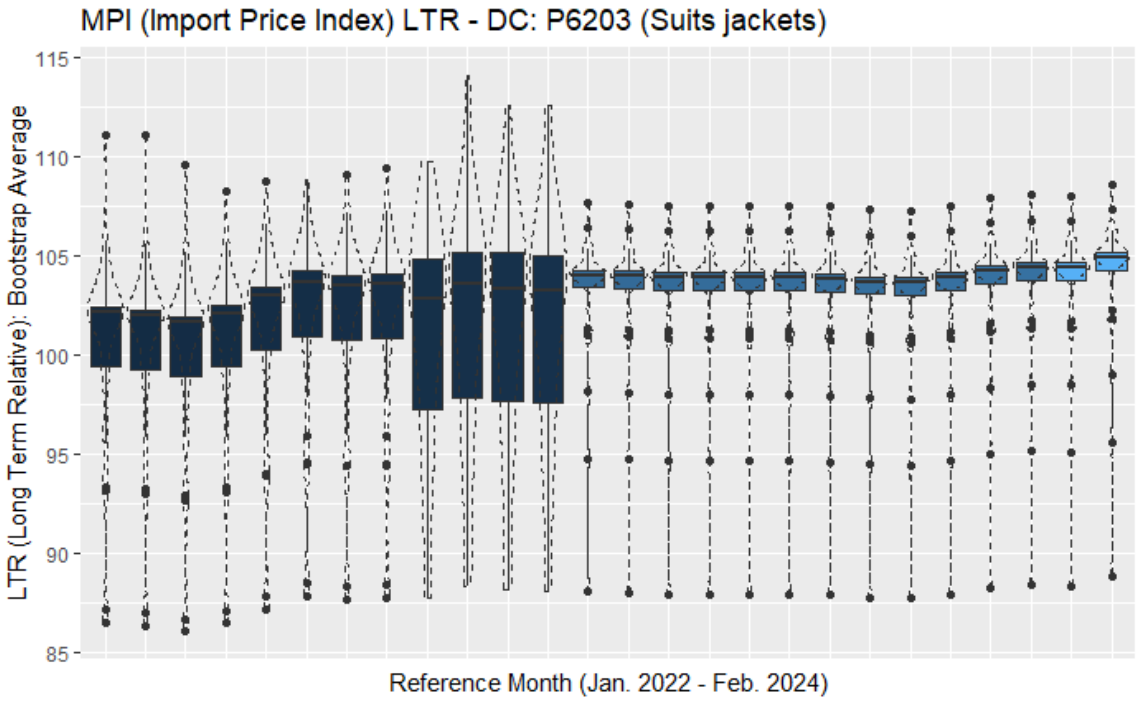


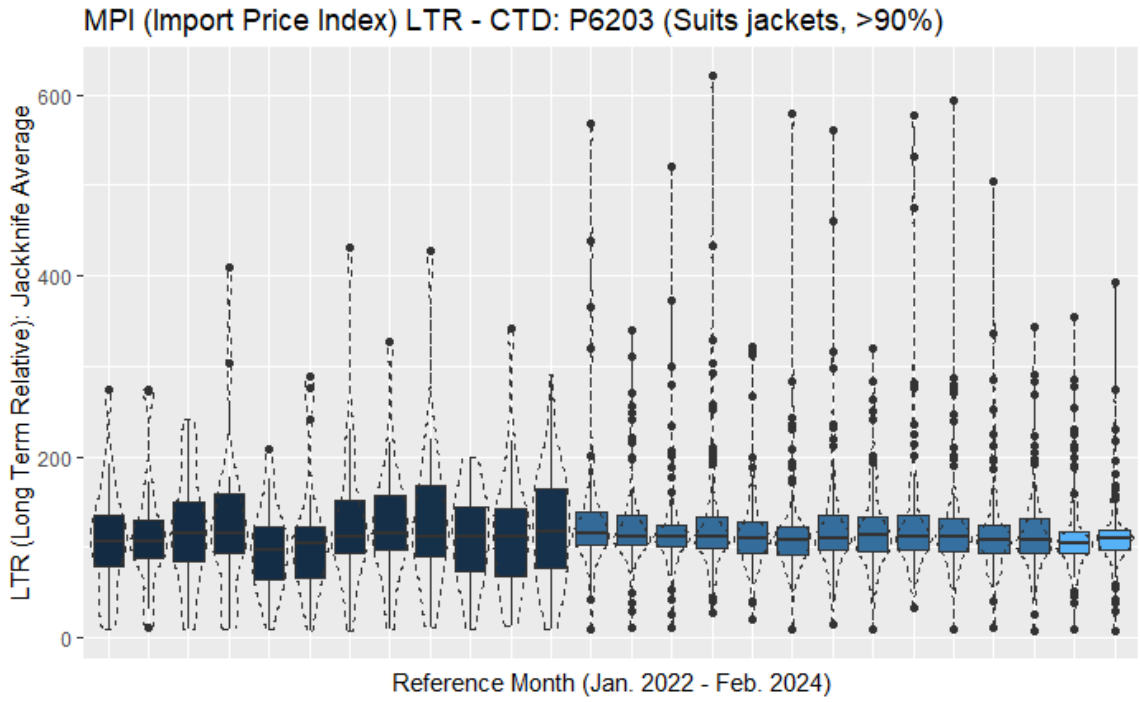
Rubber Tires (P4011) Violin Plots: DC and CTD



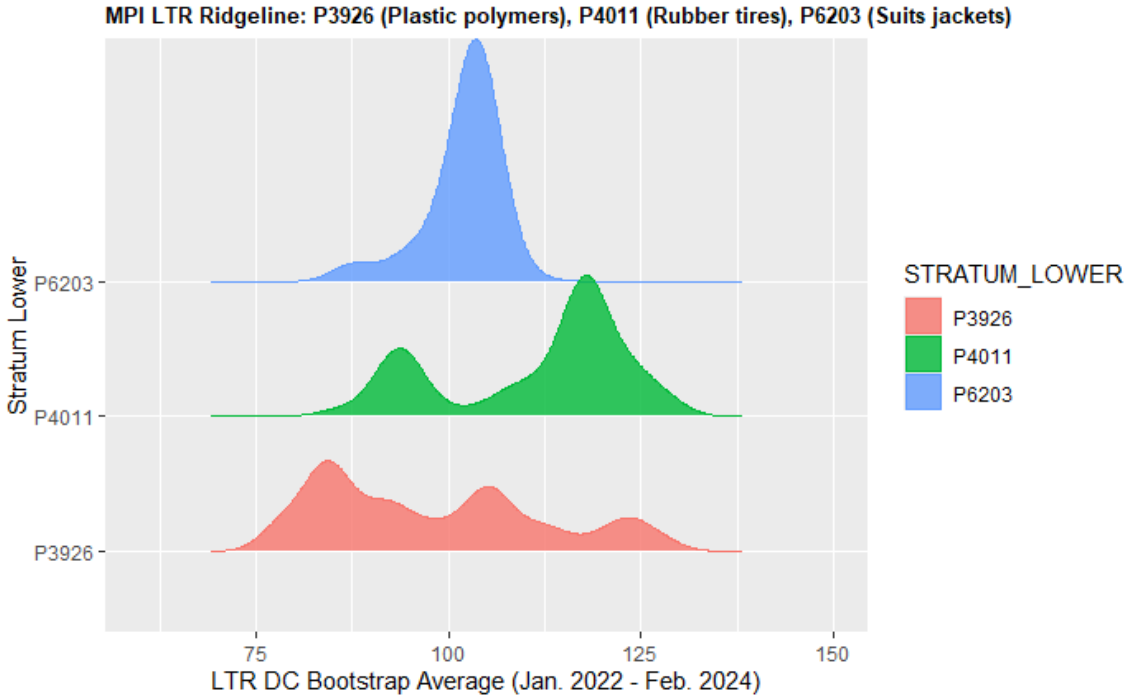


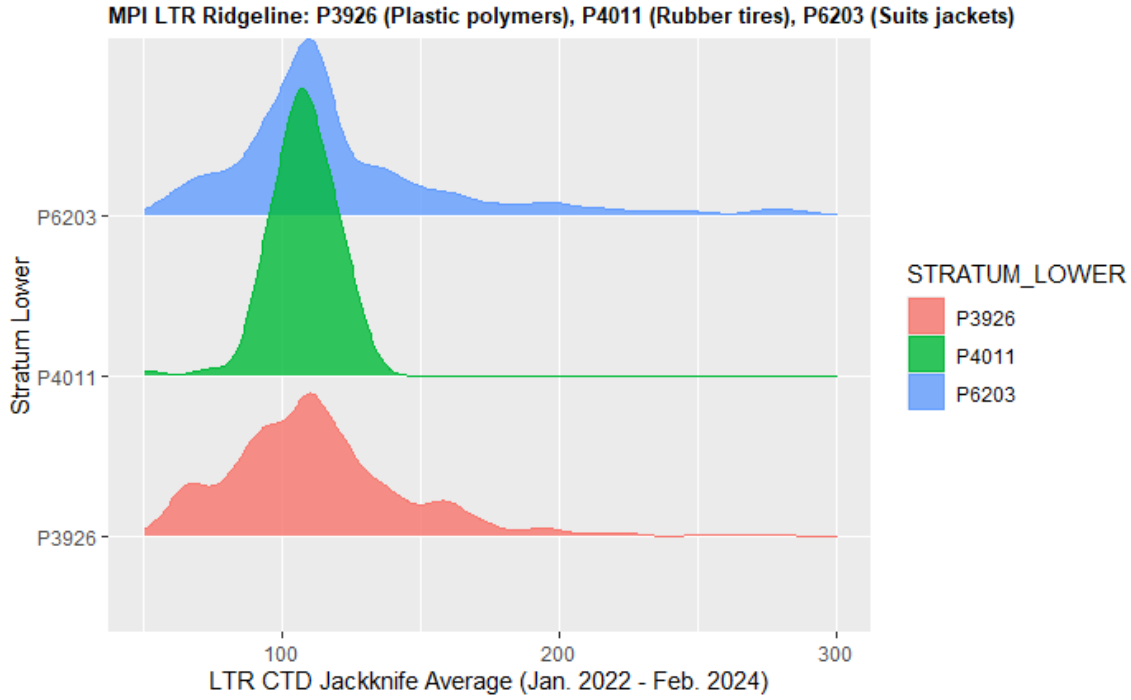
Suits jackets (P6203) Violin Plots: DC and CTD





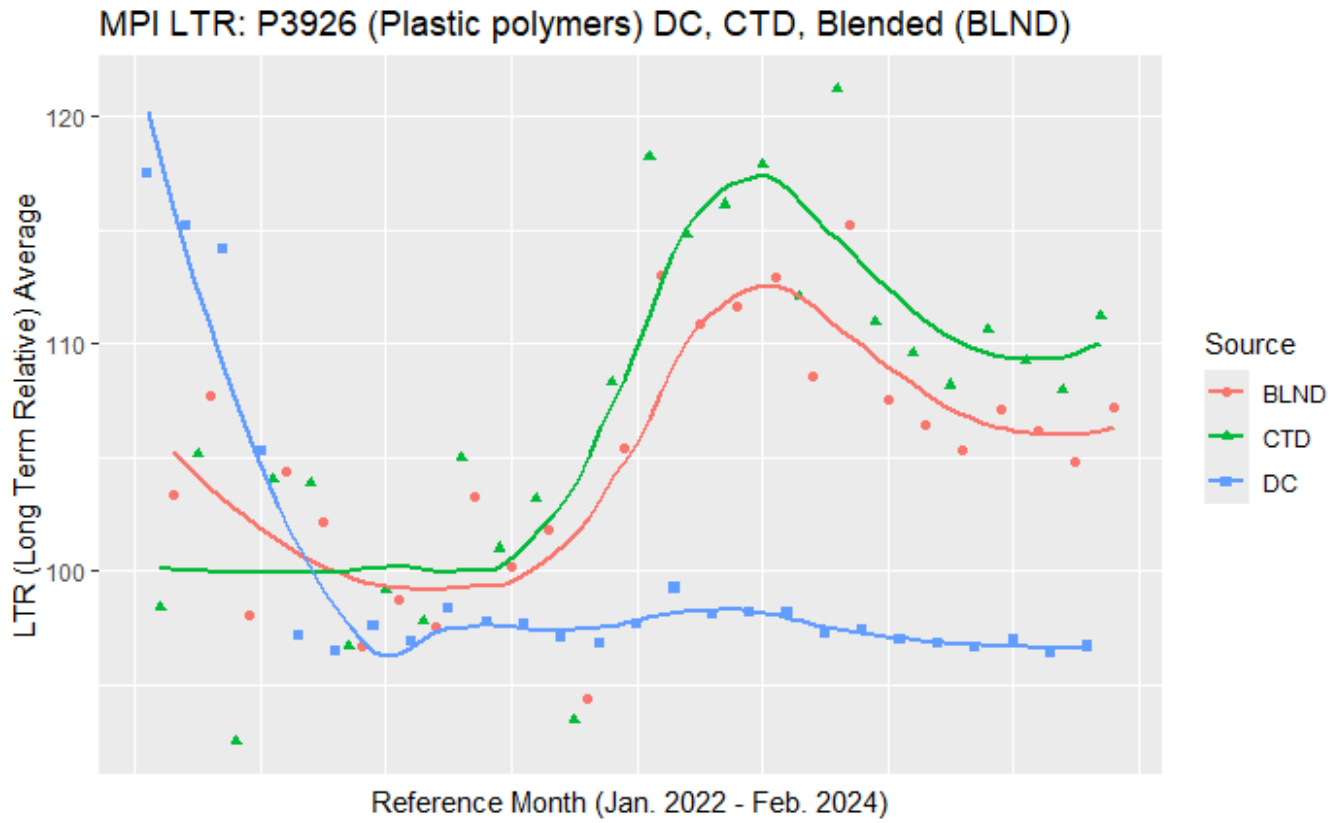
Ridgelines of Stratum Lower (SLs): DC and CTD





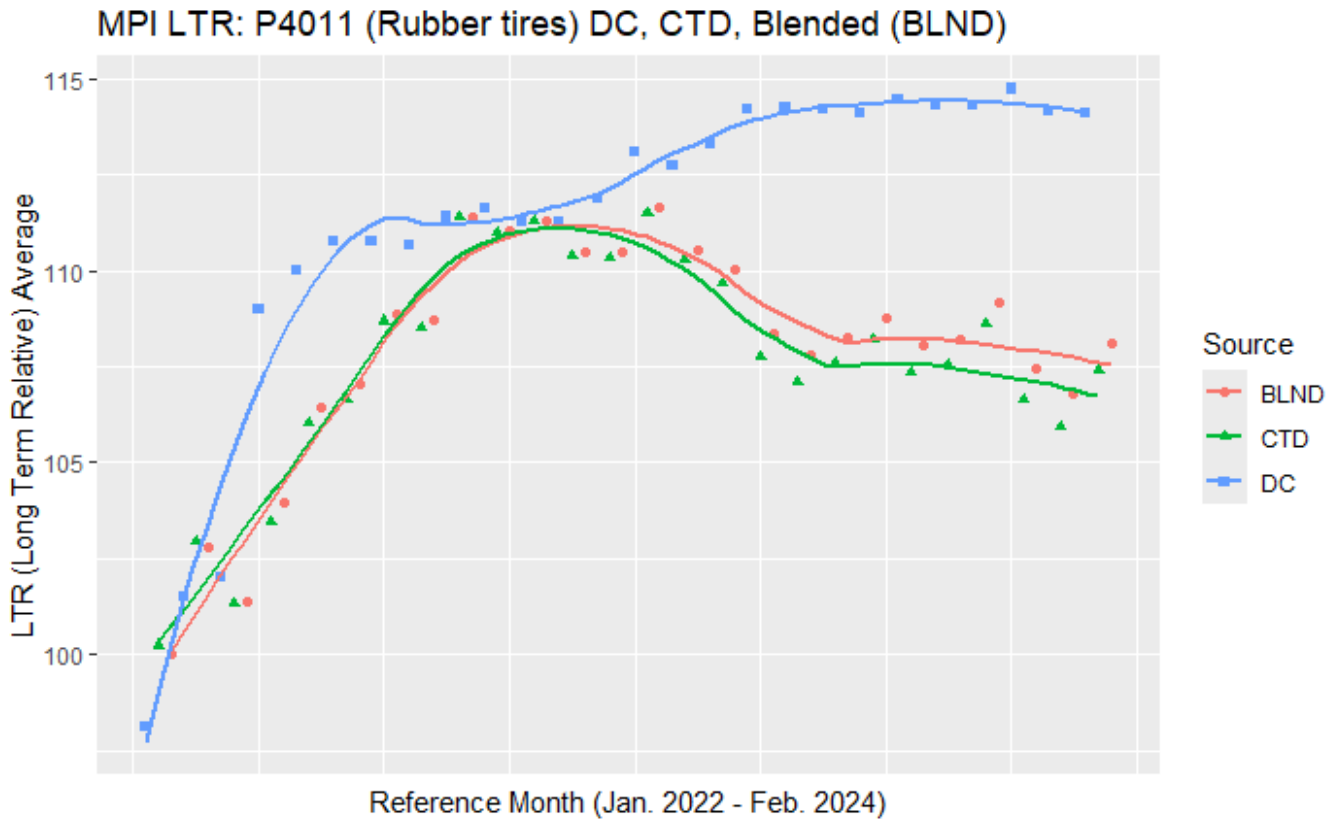
P3926: Plastic polymers, P4011: Rubber tires, P6203:Suits jackets

Plastic Polymers LTR: CTD, DC and Blended (BLND)

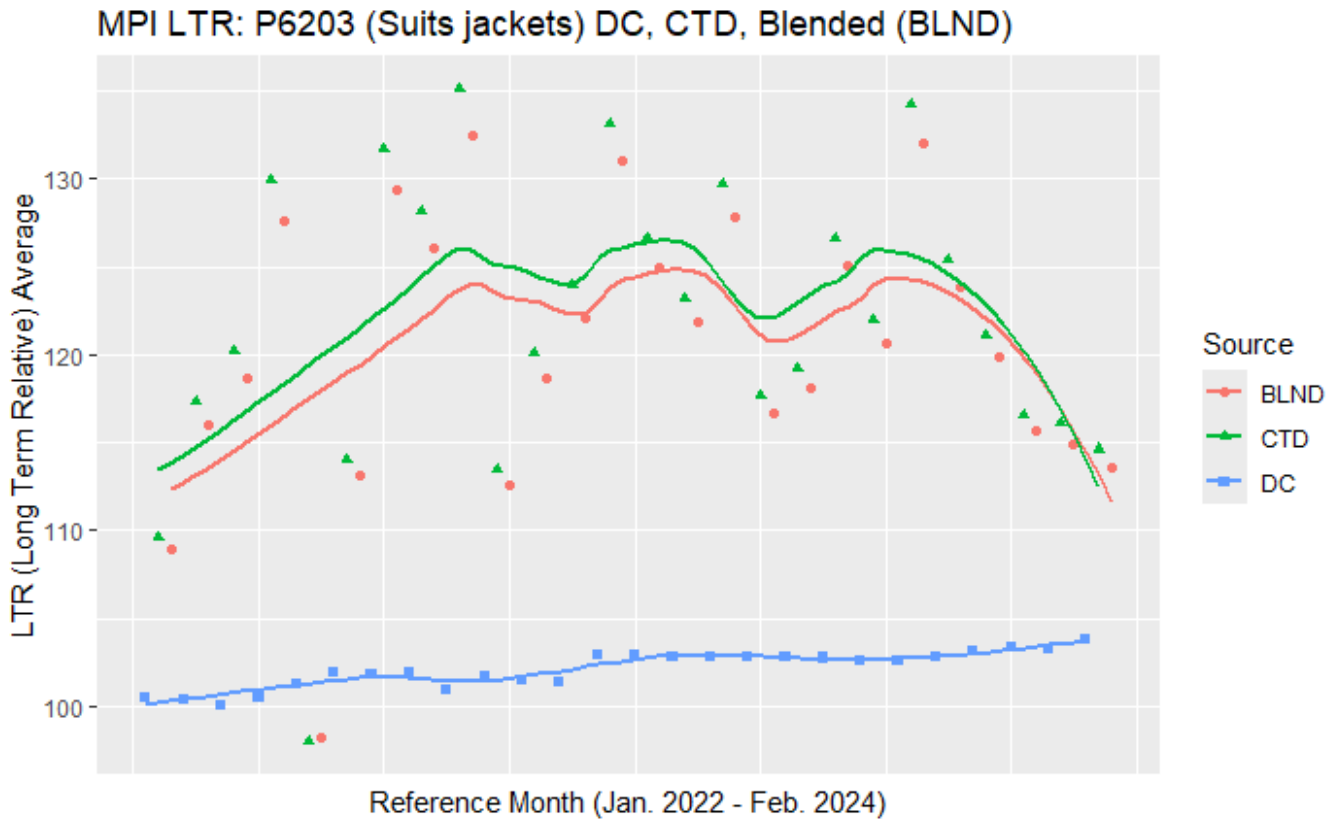


Note: two separate measurements from DC and CTD are being put together (blending) to produce an index estimate for the whole combined set of CGs which constitute the SL.

Rubber tires LTR: CTD, DC and Blended (BLND)



Suits jackets LTR: CTD, DC and Blended (BLND)



V. Summary

- Administrative data source like CTD is the new norm.
- Blended estimates present challenges to establishment surveys and opportunities to improve.
- Our resampling methods are working and captured large variations.
- Preliminary blending step results are consistently close to CTD …

THANK YOU!

yang.daniel@bls.gov

Office of Survey Methods Research (OSMR)

Bureau of Labor Statistics

Acknowledgements:

Office of Prices and Living Conditions (OPLC)

Division of International Prices (DIP): Susan Fleck, Larry Lang

Division of Price Statistical Methods (DPSM): Steven Paben, Jeffrey Blaha, Xiao Fan and Chris Grant

Office of Survey Methods Research (OSMR): Daniell Toth