Variance Estimation of Blended Import and Export Price Indexes with Census Trade Data

Daniel K. Yang

Office of Survey Methods Research

U.S. Bureau of Labor Statistics

2023 Federal Committee on Statistical Methodology (FCSM) Research & Policy Conference, Hyattsville, MD

October 26th, 2023

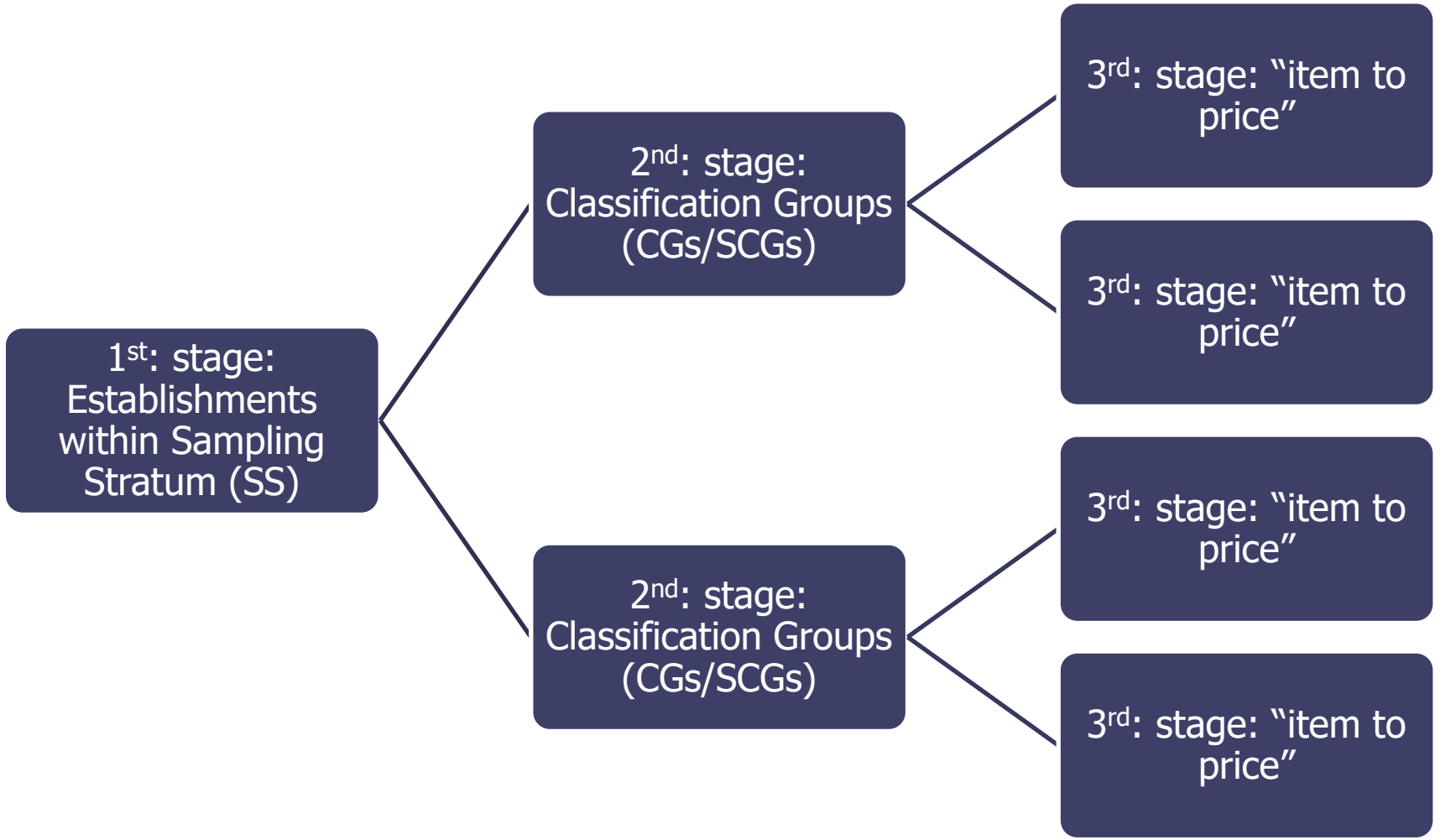
Overview

I. International Price Program (IPP) Survey
II. Census Trade Data (CTD).
III. Blended Import and Export Price Indexes (MXPI) Estimation.
IV. Proposed Bootstrap Approach for CTD.
V. Summary.

I. International Price Program (IPP) Survey

- IPP 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 (a Principal Federal Economic Indicator or PFEI).
- IPP Survey: 3-stage sampling design.

IPP Survey: 3-stage sampling design



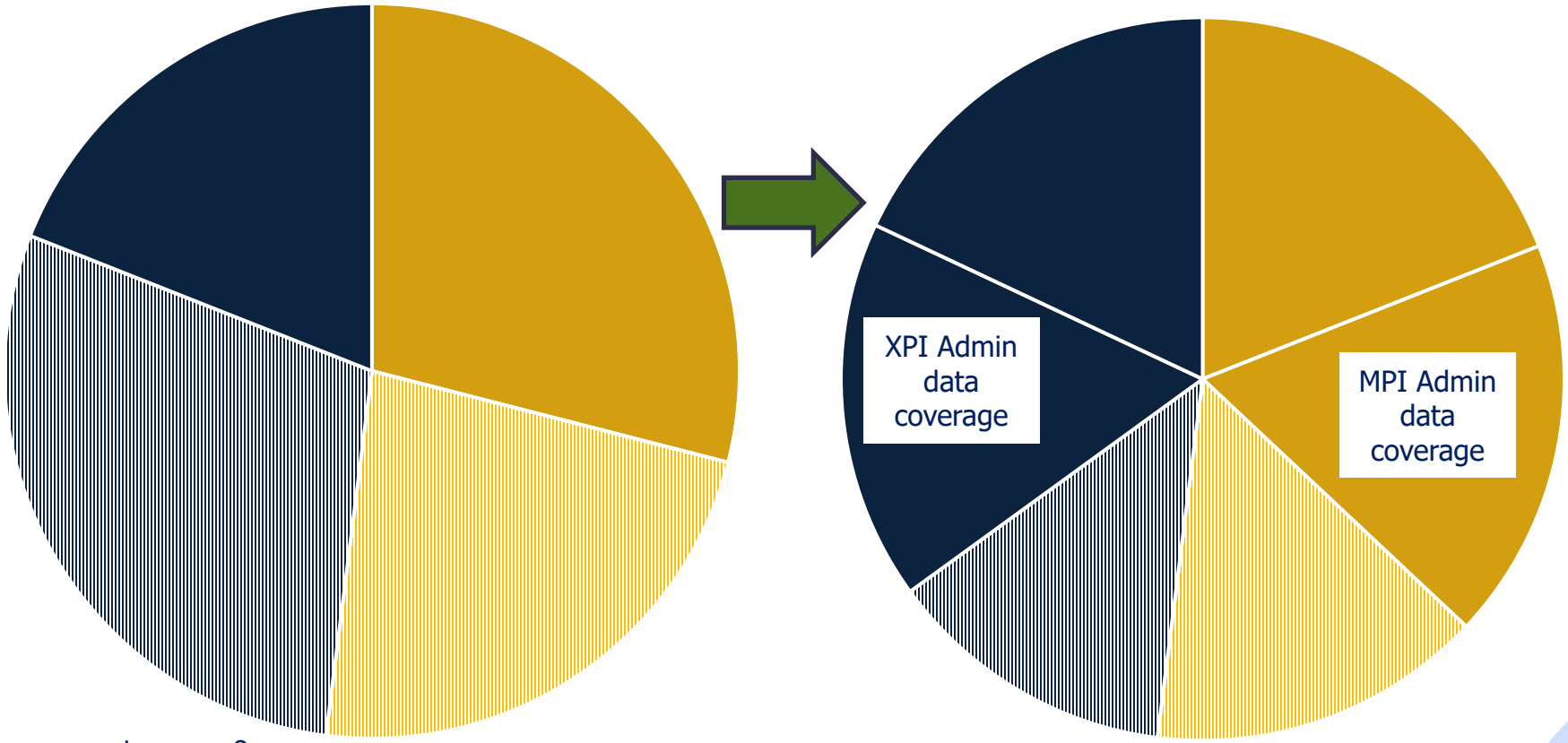
II. Census Trade Data (CTD)

- Number of publishable detailed MXPI has declined over time.
- BLS is planning to implement an alternative data source: CTD for many commodity product areas.
- CTD contains detailed shipment records for nearly all imports and exports.

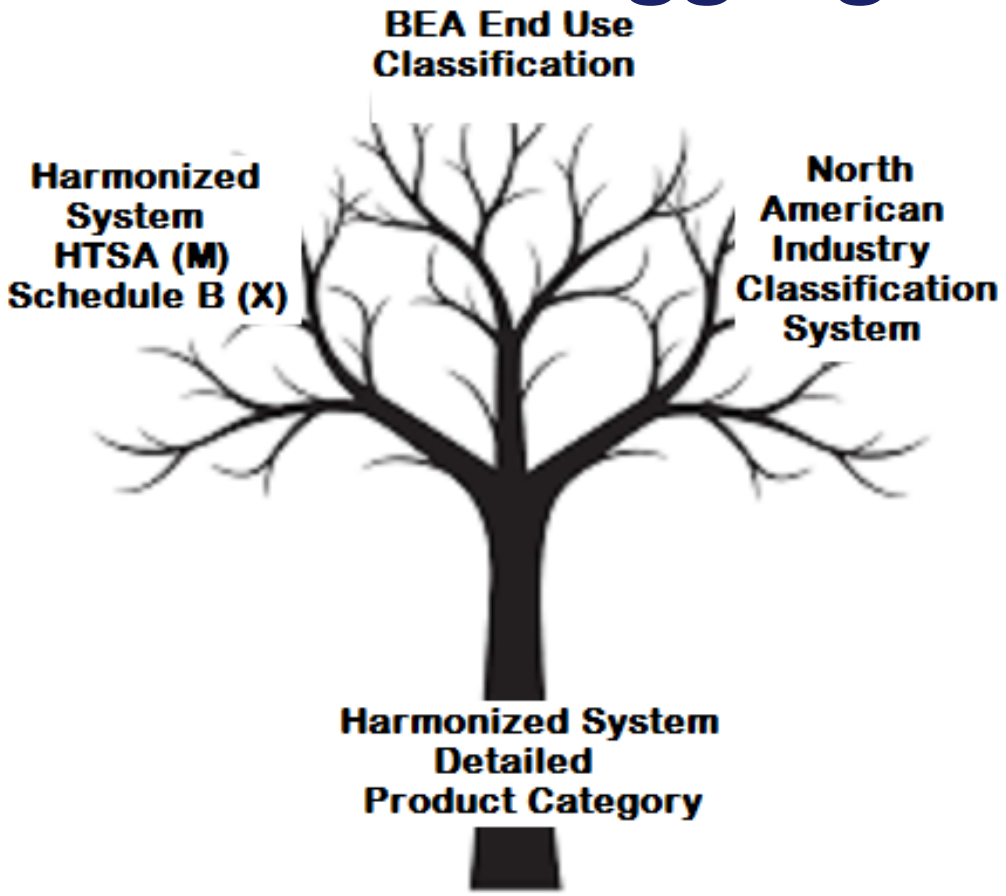
Impact of Census Trade Data Replacing Direct Data Collection by Panel (2019 Trade $ Values—in Billions)

Trade Type	Panel	Area	Sample-Eligible $ Value	Percentage Sample-Eligible $ Value That Is Replaced	Percentage Sample-Eligible $ Value That Is Replaced (within Panel)	Percentage Sample-Eligible $ Value That Is Replaced (within Trade Type)
Export	A	Crude Materials	170.4	57%	61%	33%
Export	A	Food & Beverages	108.8	91%		
Export	A	Minerals, Chemicals, & Rubber	333.9	72%		
Export	A	Miscellaneous	113.1	4%		
Export	B	Machinery	431.8	3%	2%	
Export	B	Vehicles	190.6	0%		
Import	A	Crude Materials	356.3	45%	40%	33%
Import	A	Food & Beverages	158.0	84%		
Import	A	Vehicles	332.9	20%		
Import	A	Miscellaneous	203.8	28%		
Import	B	Machinery	713.7	17%	28%	
Import	B	Minerals, Chemicals, & Rubber	397.1	46%		

Census Trade Data Increases Published Detailed MXPI



Harmonized System Detailed Product Categories are Basic Classification Group to Calculate Aggregate MXPIs



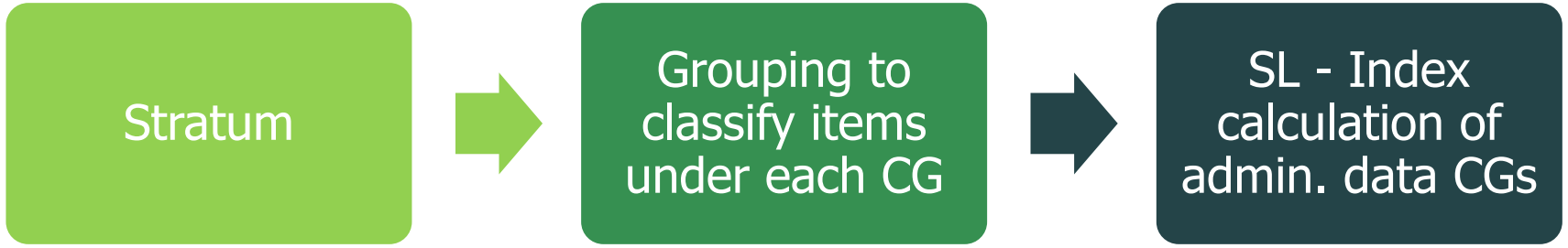
Census Trade Data Source as Share of Harmonized Product Categories

Import End Use Index Harmonized Commodity Classification Codes Representation			
Index Weight Year	Total Number of Harmonized Codes in Index	Number of Harmonized Codes Represented by Census Trade Data	Percentage of Harmonized Codes Represented by Census Trade Data
2019	17,686	8,034	45%
2020	17,627	7,988	45%
2021	17,771	8,108	46%
…	…	…	…
Export End Use Index Harmonized Commodity Classification Codes Representation			
Index Weight Year	Total Number of Harmonized Codes Used in Indexes	Number of Harmonized Codes Represented by Census Trade Data	Percentage of Harmonized Codes Represented by Census Trade Data
2019	9,006	4,177	46%
2020	8,983	4,180	47%
2021	9,000	4,183	46%
…	…	…	…

Data Source Differences

- IPP Survey Data:

- CTD Administrative Data:



Data Source Error Type

- IPP Survey Data represents a share of trade across U.S. borders
  - Sampling error

- CTD Administrative Data represents all 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)

III. Blended MXPI Estimate

- Lohr (2021) blended estimator proposal.
- MXPI will be estimated from dual data sources: IPP Survey and CTD admin data.

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

- θ – proportion
- At Stratum-Lower level (SL),

$$\hat{Y}_{sl}(\theta_{sl}) = \theta_{sl} \hat{Y}^{(1)}_{sl} + (1-\theta_{sl}) \hat{Y}^{(2)}_{sl}$$

Variance Approximation for Blended MXPI Estimate

- The derivatives of $\hat{Y}_{sl}(\theta_{sl})$:

$$\frac{\partial Y_{sl}}{\partial \theta_{sl}} = Y_{sl}^{(1)} - Y_{sl}^{(2)}, \quad \frac{\partial Y_{sl}}{\partial Y_{sl}^{(1)}} = \theta_{sl}, \quad \frac{\partial Y_{sl}}{\partial Y_{sl}^{(2)}} = 1 - \theta_{sl}$$

- Taylor Series Linearization (TSL) variance approximation:

$$\hat{V}\left(\hat{Y}_{sl}(\theta_{sl})\right) = \left(\hat{Y}_{sl}^{(1)} - \hat{Y}_{sl}^{(2)}\right)^2 \hat{V}(\hat{\theta}_{sl}) + \hat{\theta}_{sl}^2 \hat{V}\left(\hat{Y}_{sl}^{(1)}\right) + \left(1 - \hat{\theta}_{sl}\right)^2 \hat{V}\left(\hat{Y}_{sl}^{(2)}\right) + \left(1 - \hat{\theta}_{sl}\right)^2 \hat{V}\left(\hat{Y}_{sl}^{(2)}\right) +$$
$$2\left(\hat{Y}_{sl}^{(1)} - \hat{Y}_{sl}^{(2)}\right)\hat{\theta}_{sl} Cov\left(\hat{Y}_{sl}^{(1)}, \hat{\theta}_{sl}\right) + 2\hat{\theta}_{sl}\left(1 - \hat{\theta}_{sl}\right) Cov\left(\hat{Y}_{sl}^{(2)}, \hat{\theta}_{sl}\right)$$

Propagation of Uncertainty

- Groves and Lyberg (2010):

- Presence of potential error propagation may not be ignored in the CTD administrative data.
- Borrowing information from Census improves the estimate.

IV. A Bootstrap Approach for IPP Sample

- Select only directly collected items from CGs and determine that they have not been discontinued as of the month being calculated.
- Further select items that contribute weights to the MXPI.
- Maintain consistency of sampling population.
- Use simple random sampling with replacement (SRSWR).

IV. Proposed Bootstrap Approach for CTD

- Resampling at CG level for CTD.
- Select only “usable” CTD items from CGs.
- Maintain consistency of sampling population.
- Use simple random sampling with replacement (SRSWR)

Bootstrap Estimate of Blended Indexes

- Blended SL level estimate at Bootstrap run b:

$$\hat{\theta}_{bsl} = (1 - \hat{p}_{bsl})\hat{\theta}_{bsl}^{IPP} + \hat{p}_{bsl}\hat{\theta}_{bsl}^{CTD}$$

- Bootstrap mean and variance estimates of θ_{sl} :

$$\hat{\theta}_{sl} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{bsl}\,, \qquad var(\hat{\theta}_{sl}) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_{bsl} - \hat{\theta}_{sl}\right)^2$$

- Bootstrap standard error (SE) and relative SE (RSE) of $\hat{\theta}_{sl}$:

$$SE(\hat{\theta}_{sl}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_{bsl} - \hat{\theta}_{sl}\right)^2}\,, \qquad RSE(\hat{\theta}_{sl}) = \frac{SE(\hat{\theta}_{sl})}{\hat{\theta}_{sl}} \times 100$$

Summary

- Alternative data source (frame) like CTD administrative data is here to stay.
- Blended estimates present challenges to establishment surveys and opportunities to improve.
- Researchers need evaluate the complex structure of administrative data and the means to access it.
- Be adaptive and ...

THANK YOU!

yang.daniel@bls.gov

Office of Survey Methods Research (OSMR)

Bureau of Labor Statistics

COPAFS

Council of Professional Associations on Federal Statistics