

Recent and Upcoming Advancements Using Alternative Collection Methods for BLS Price Programs

Susan Fleck, International Price Program

Bonnie Murphy, Producer Price Program

Alexander Paplomatas, Consumer Price Program

Data Users' Conference

April 21, 2021



Alternative Collection Methods Overview

Bonnie Murphy
Producer Price Program



Traditional survey data collection

- Economists or Economic assistants collect prices of specific items (in store or online)
 - ▶ Producer, import and export prices collected from companies via monthly internet survey after initial in person collection
- Specific items in sample rotate over time
 - ▶ Rotation intervals differ across price programs
- Allows us to make sure we are pricing the exact items or adjusting for any changes



Alternative collection methods

- Corporate supplied
- Web scraping/application program interface (API)
- Secondary source
 - Administrative
 - Private source



BLS strategic plan objectives

■ Goal 3.1 - Integrate alternative data into BLS programs

- ▶ Objective 3.1.1 – Identify alternative data sources, including administrative data
- ▶ Objective 3.1.2 – Evaluate the “fitness for use” of alternative data sources and develop methods to incorporate these data into BLS programs
- ▶ Objective 3.1.3 – Incorporate alternative data and methods into BLS programs and survey processes
- ▶ Objective 3.1.4 – Increase collaboration and information sharing within BLS and with other organizations, including international statistical organizations, on alternative data sources and their uses



Why explore alternative sources?

- Increasing availability of secondary source data
- Decline in response rates
- Lack of sample size to appropriately represent areas with very large number of transactions a month
- Potential cost efficiencies (both human and financial)



Benefits of alternative sources

- Allows for evaluation & improvements
- **May** reduce collection costs
- Reduces respondent burden
- Almost always increases sample size –expanding coverage for an individual respondent or increasing the number of units sampled
- **May** increase data quality
- **May** increase timeliness of reporting



Challenges

- Availability – differs across price programs
- Methodological - transacted vs. offer prices, representativeness of in-scope production, data descriptiveness
- Operational
 - ▶ Data lag, continuity, quality verification
 - ▶ Current system data size limits
- Legal, Policy, and Budgetary
 - ▶ Contracting for data purchase, terms of service agreements, confidentiality concerns, cost



Data uses

- Sample frames
- Benchmark sample
- Supplement collected data to support hedonic modeling
- Replace/supplement current data collection methods
- Validation of survey data



Quality/fitness for use

- Quality metrics are similar regardless of the type of alternative data



Dimensions of data quality

- Accuracy
- Timeliness
- Reliability/Durability
- Coherence/Comparability
- Accessibility and Relevance
- Cost



Accuracy considerations

- Transparency
- Sample versus universe
- Quality control
- Concordance
 - Mapping items into existing structures
 - Blending with existing survey data



Data accuracy: coverage error

- Time period/history
- Geographic (different for each price program)
- Outlet/establishment
- Item level price and characteristic detail



Timeliness

- Timing of data procurement
- Timing of data delivery - control of schedule



- Limits on long term contracts

Reliability/durability

- Will it exist going forward?
- Condition of data/ 'cleaning' or validating needed?
- Data security



Coherence/comparability

- Methodological constructs
 - Fit into existing?
 - New one needed?
- Comparability of data over time



Accessibility and relevance

■ Accessibility

- ▶ Resulting indexes will continue to be accessible to users
- ▶ Microdata
 - Survey microdata is accessible to “sworn agents” of BLS
 - Alternative source may not allow sworn agent access

■ Relevance – as long as our publication structures remain relevant, alternative data would not change their relevance

Cost

- Cost to purchase
- Cost of IT infrastructure
- Cost of processing changes to use data
- Incorporation into existing BLS structure
- Purchasing process- huge cost of switching vendors
- Staff – maintain skill set to manage/massage data
- Fall back plan



BLS index quality metrics

Current:

- Variance/standard errors for current price indexes
- Response rates
- Non-response bias studies (if necessary)
- Timely release of monthly data
- Size of Revisions (PPI)

Alternative data sources: Analyzed at more detailed level for fitness for use



Alternative Collection Methods for Consumer Prices

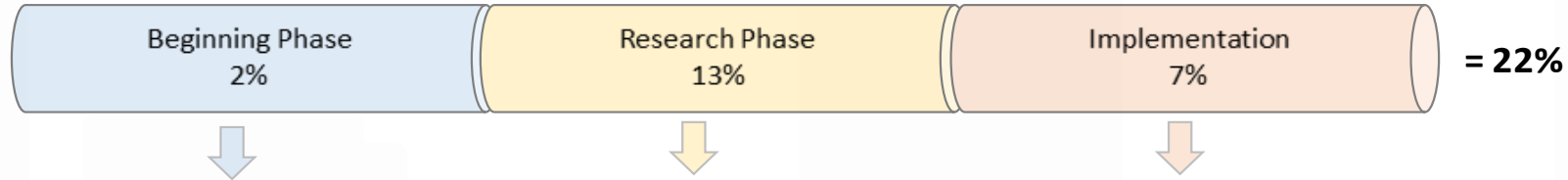
Alexander Paplomatas
Consumer Price Program



Alternative collection: current projects

CPI Alternative Data Pipeline

April 2021



1. Identify new sources
2. Collecting data 2%
 - Apparel**
Web-scraping, one footwear retailer
 - General merchandise**
Web-scraping, many item categories
 - Food away from home**
Corporate data, one fast food company

3. Developing methodology 13%
 - Medical services**
Purchased data, insurance payments to physician's services, hospital services
 - Wireless phone service**
Purchased web-scraped data, offer prices for new plans
 - Residential telecommunication service**
Purchased web-scraped data, offer prices for new plans
 - Airfare**
Web-scrape aggregator site, near full item coverage
 - Vehicle leasing**
Purchased data (JD Power), near full item coverage
 - Hotels**
Web-scrape aggregator site, near full item coverage
 - Housing**
HUD administrative data, government subsidized rental properties

4. Seeking approval 0%
5. Approved for implementation 4%
 - New vehicles**
Purchased data (JD Power), full item coverage, targeted deployment 2022
 - Airline**
Corporate data, one airline
6. In development 3%
 - Motor fuel**
Corporate data, full item coverage, targeted deployment June 2021
7. Parallel testing 0%
8. In production 3%
 - Used cars**
Purchased data, longtime source
 - Postage**
Publicly available data, longtime source
 - CorpY**
Corporate data, March 2018
 - CorpX**
Corporate data, March 2019



General steps for data sources

- Determine what to pursue
- Evaluate options
- Evaluate selected source (definition, coverage, other quality dimensions)
- Test estimation methods and evaluate results
- Transition to production?



Source selection

- Initial priorities
 - ▶ Problem items: Low response rate, measurement complexity, etc.
 - ▶ Maintain respondents
- New data source selection process
 - ▶ Proposed projects
 - ▶ Evaluate data using relevant factors
 - ▶ Field assessment of cooperativeness of top respondents
- Data collection realities
 - ▶ Respondents open to alternative collection
 - ▶ Post-pandemic return to personal visit collection?



Research

■ Goals

- ▶ Standardize methods and streamline research to have a continuous flow of recommendations and improvements
- ▶ Expand researcher expertise in new price index methods

■ Challenges

- ▶ Data quality metrics
- ▶ BLS index quality metrics
- ▶ Product lifecycle, quality change, blending sources



Production considerations

- Is it as good or better than current pricing methodology?
- Does improvement in index justify any additional costs – back-up plans?
- Should we use a short-term solution while continuing to research longer-term improvements?



Gasoline secondary source

- Use in CPI planned for later this year
- About 3.0% Relative Importance
- Replace gasoline—all grades
- Also included in average price series



Gasoline secondary source (cont.)

- Monthly prices will be the arithmetic mean unit price for fuel type-station
- Price relatives as month-to-month change
- Index formula: unweighted geometric mean for all fuel type-station relatives in a county
- Geometric mean for each CPI PSU using county allocated expenditure weights
- Geometric mean for each index area using sum of county weights within each PSU in the index area
- Aggregate item-area indexes using Laspeyres (unchanged from current)



New vehicles secondary source

- Likely our next major source in production
- Purchased data from J.D. Power
- About 3.9% Relative Importance
- Transaction data
- Currently published as research indexes



New vehicles secondary source (cont.)

- Applying current methods to new source posed problems
- New method combines long-run trend with high-frequency changes
- Transaction data improves our ability to capture market changes



What can data users expect?

- The same great quality data
- Transparency
- Changes with more impact will require more notification



Alternative Collection Methods for Import and Export Prices

Susan Fleck
International Price Program



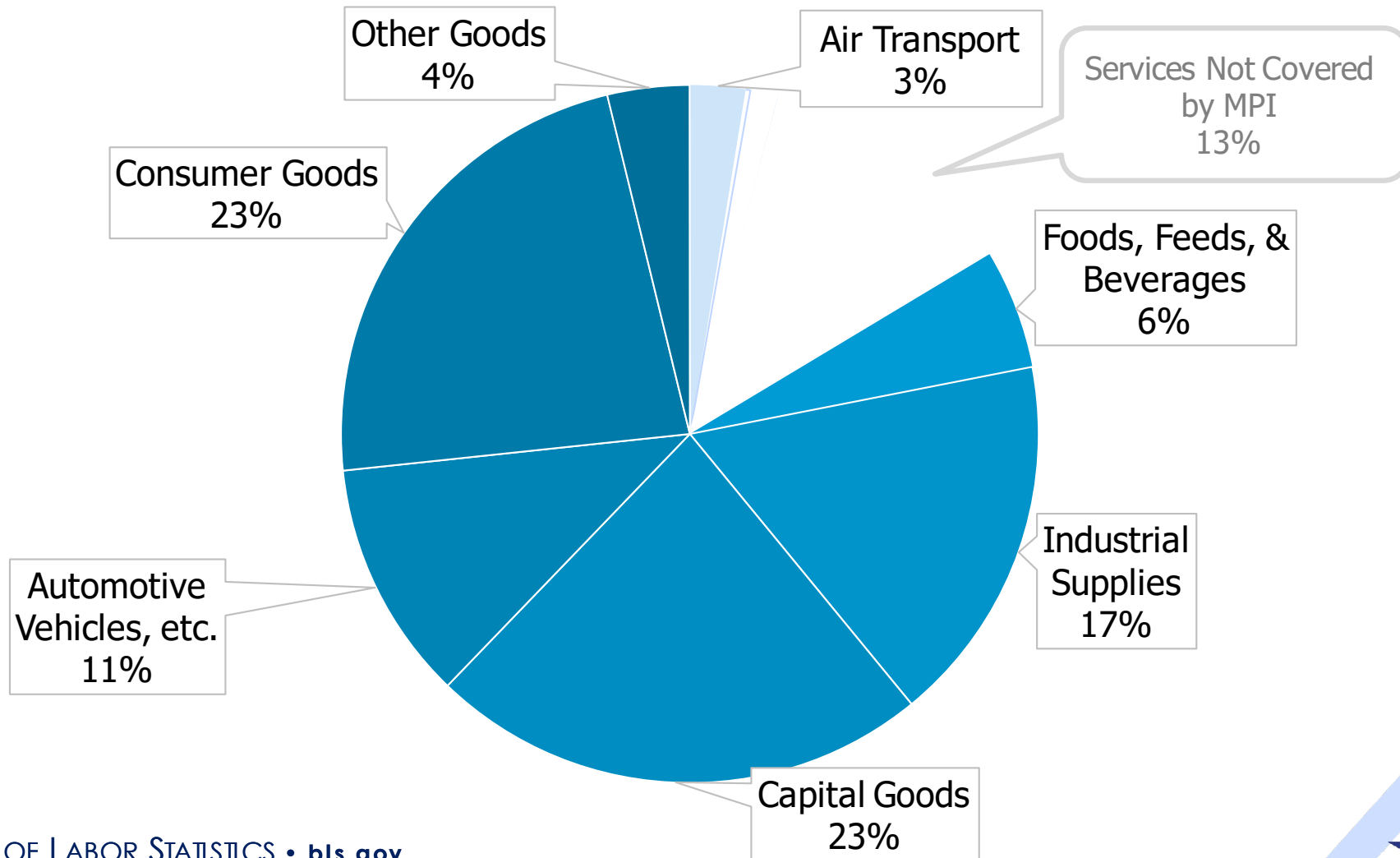
Import and Export Price Indexes

Alternative data project

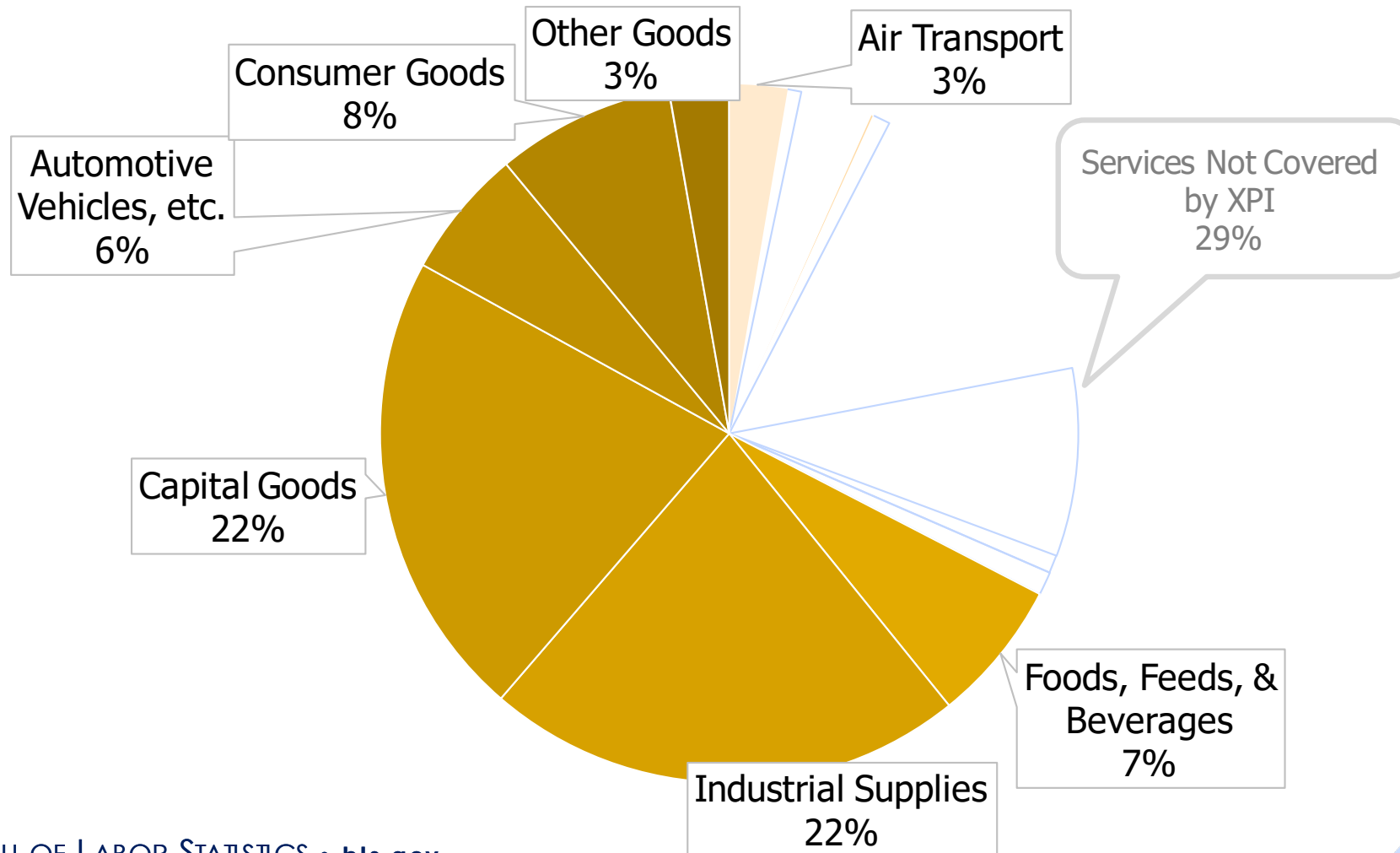
- International Trade and MXPI Basics
- Data, Data Gap and Product Gap
- Administrative Trade Data Alternative
- Next Steps



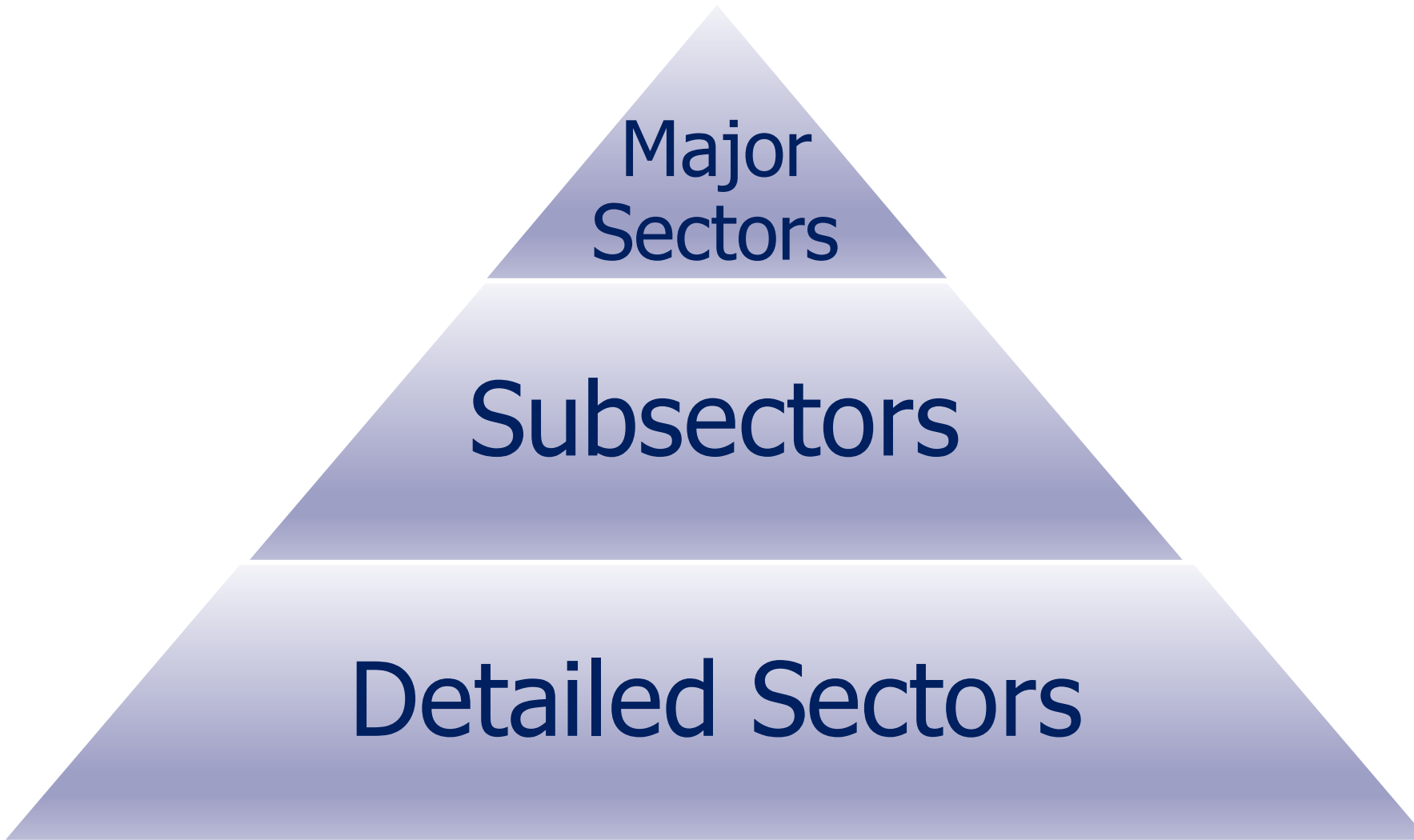
Price index coverage of 2020 imports – 87% of \$2.8 trillion



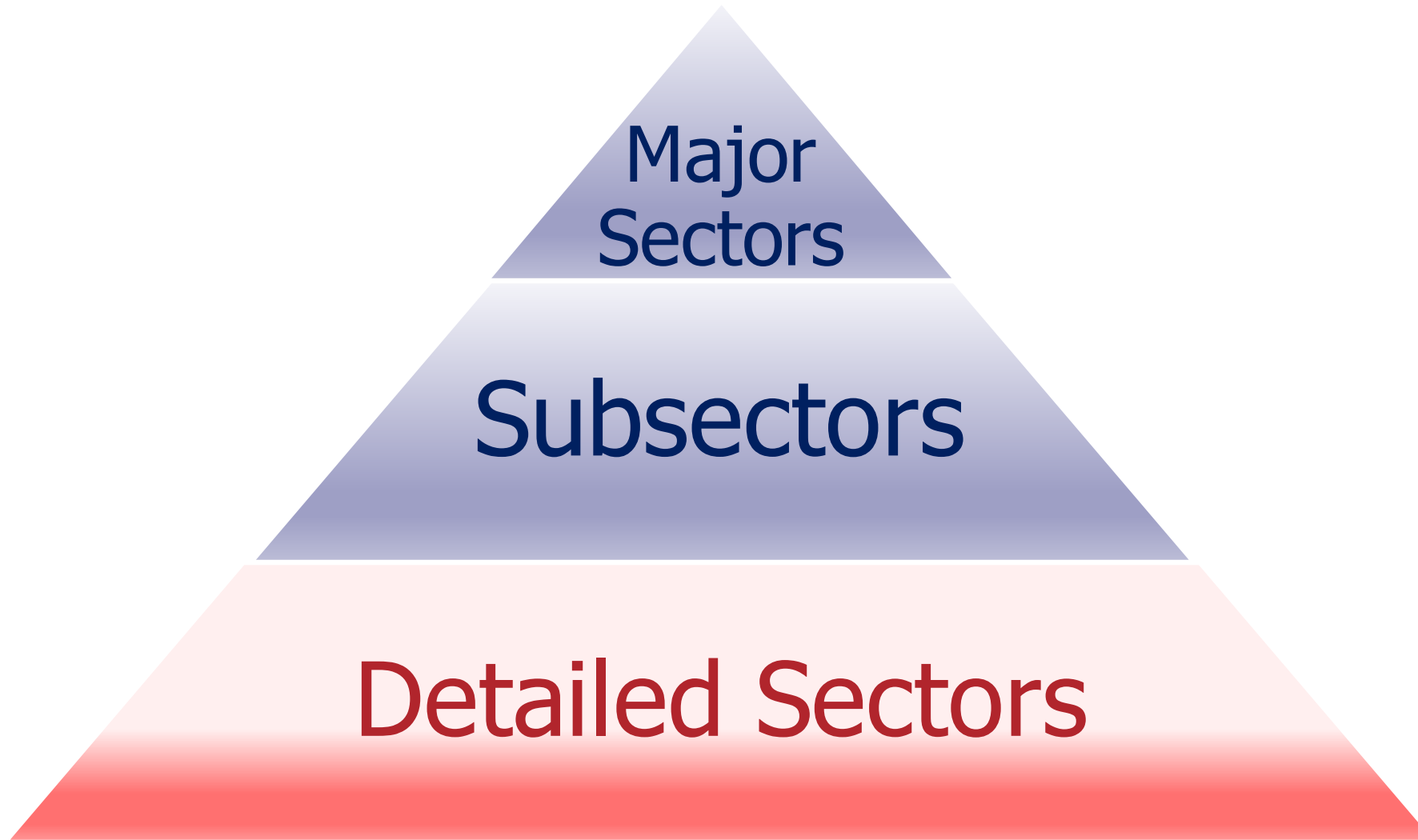
Price index coverage of 2020 exports – 71% of \$2.1 trillion



MXPI data: price index coverage

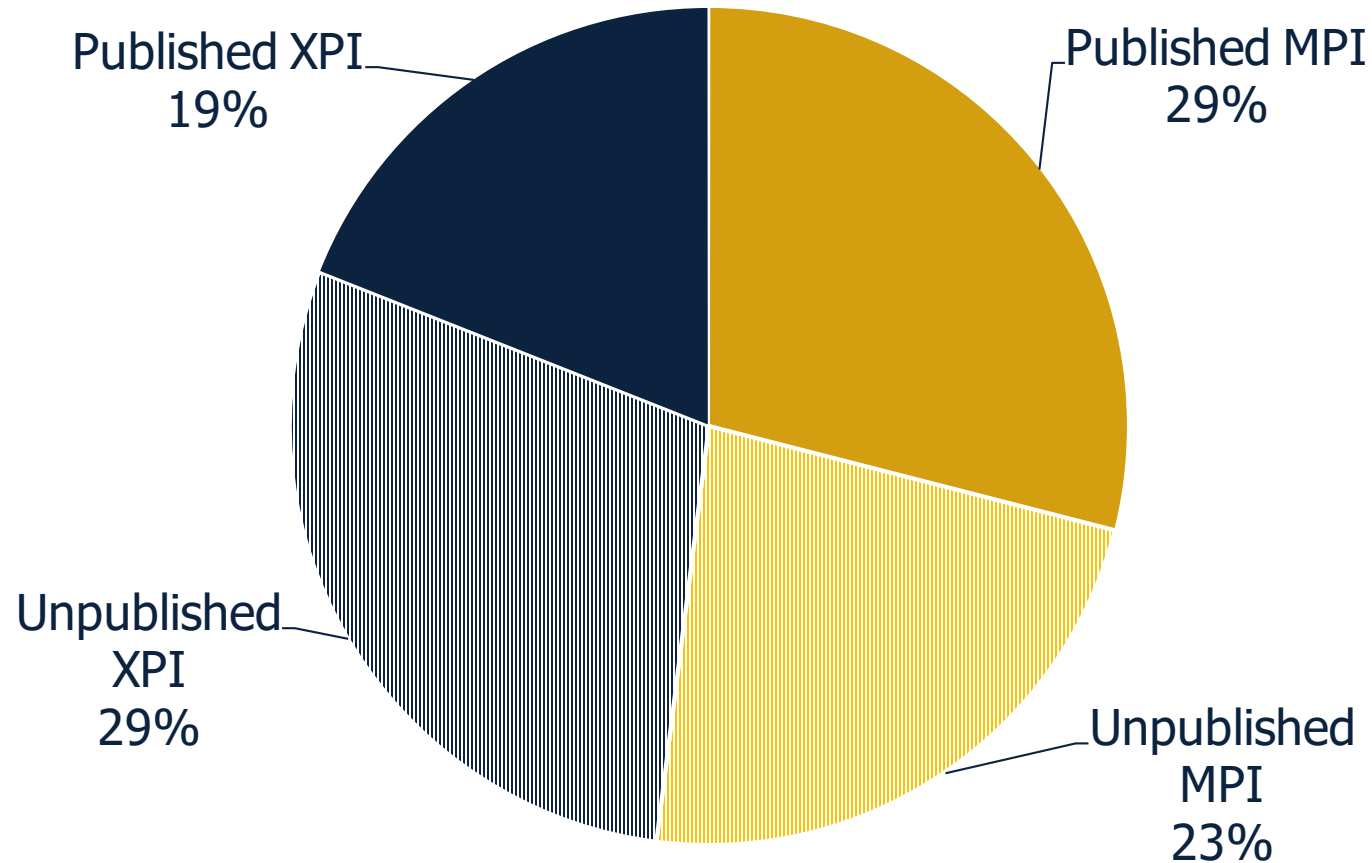


MXPI data: price index coverage (cont.)



MXPI data gap creates a product gap

5-digit Detailed BEA End Use MXPI, N = 264



MXPI data and products today

■ Monthly

- ▶ Directly collect ~20,000 item prices
- ▶ From ~3,000 companies
- ▶ Publish ~1,000 indexes

■ Yearly

- ▶ Contact 1,500 companies as older companies fall out of the sample

Administrative trade data

- Electronically provided by importers and exporters
- Collected by U.S. Customs
- Cleaned by U.S. Census

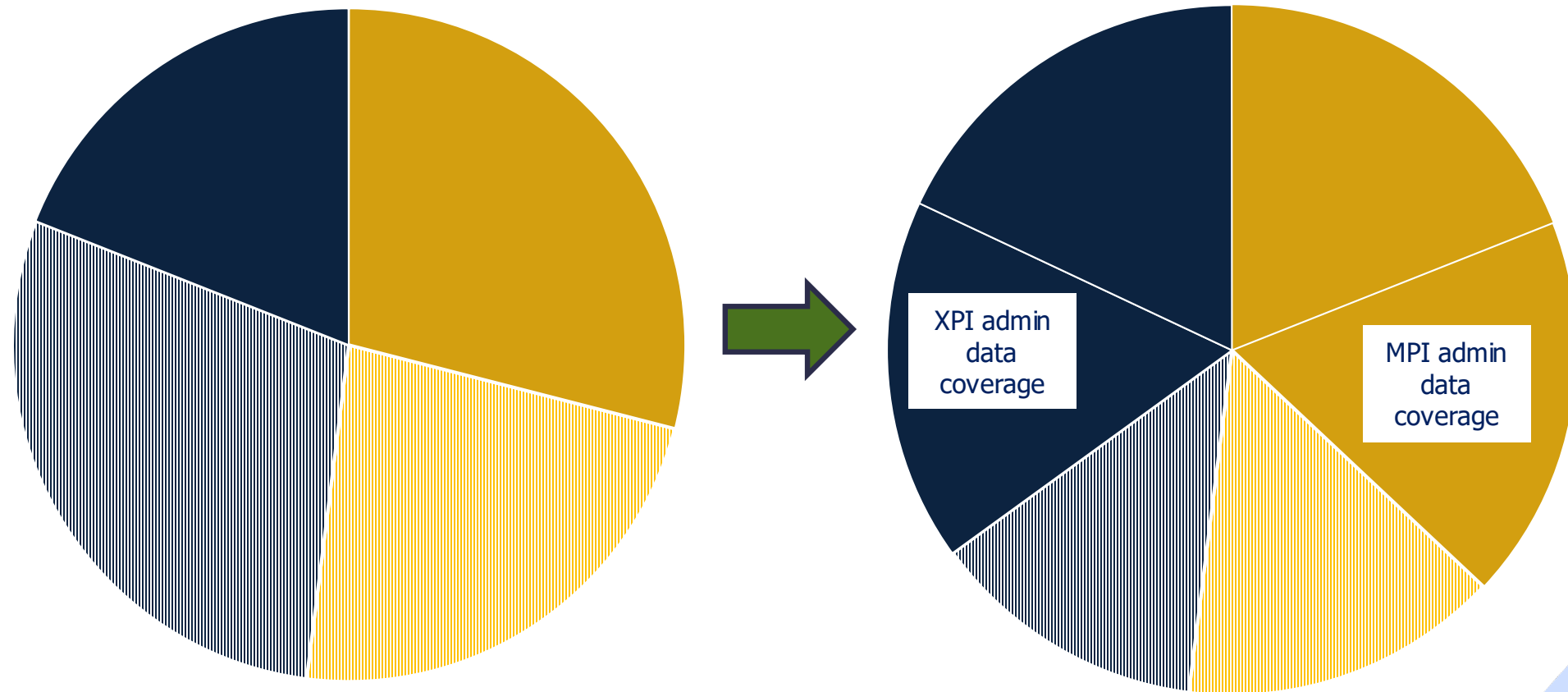


- Sent to International Price Program to select sample of companies currently trading
- Now being researched to calculate prices and indexes

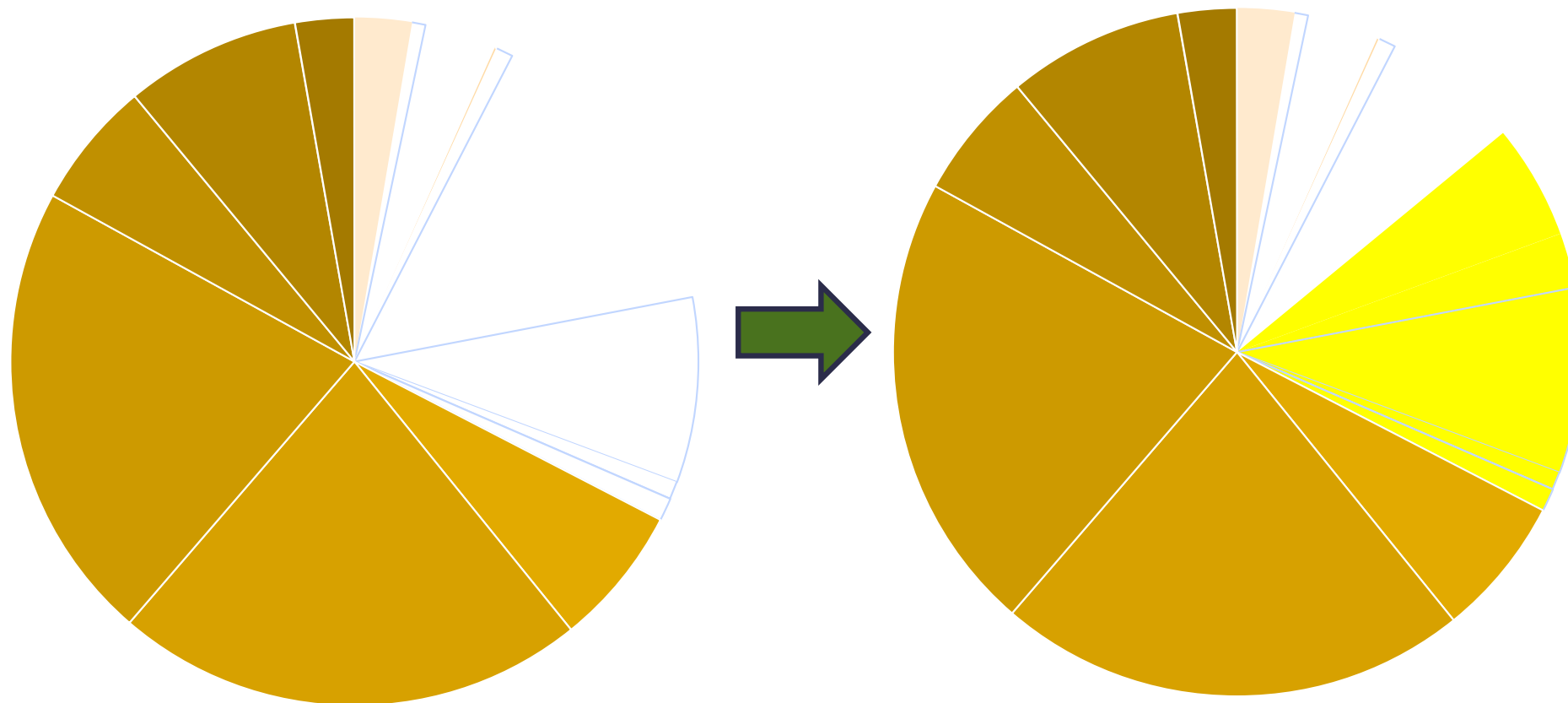
Closing the MXPI data gap with administrative trade data

- Cover ~40% of all product areas
- Add 1.5 million transaction prices
- Reduce # of companies contacted for survey to just consumer and high-tech goods
- Double number of detailed and subsector price indexes to 2,000+

Closing the MXPI product gap – increase published detailed goods MXPI



Closing the MXPI product gap – begin publishing services MXPI



Concepts, sources, and methods

Current Price Indexes

- Transaction price of one product
- Sample of companies and products
- Near 1st of month
- Prices only

Alternative Price Indexes

- Average price of shipments for product varieties
- All companies and transactions
- All-month
- Price and Quantity



Concepts, sources, and methods (cont.)

Current Price Indexes

- Solid quality for major and top level price indexes
- **Unable** to publish detailed price indexes due to insufficient number of prices for representativeness

Alternative Price Indexes

- **Able** to publish detailed price indexes with millions of shipping transaction records
- Strong results for **homogeneous** products
- Weak results for consumer and high tech **heterogeneous** products



Next steps

Continue research work to integrate into monthly price indexes with no expectation of additional funding.

- FY2020 – Validated approach
- FY2021 – Publish new XPI historic time series
- FY2022 – Publish new MPI historic time series
- FY2023 – Plan implementation & communicate with public
- FY2024 – Implement into monthly publication

Alternative Collection Methods for Producer Prices

Bonnie Murphy
Producer Price Program



PPI alternative data use

- Agriculture prices collected from administrative source since inception
- Alternative source data used to build computer hedonic models since model development in the 1990s
- Alternative sources purchased as frames for service industries for decades
- Corporate data used for select services industries
- Secondary source data is limited for producers – estimated less than 10% of industries



Corporate data

- Generally a few larger companies with very large number of heterogeneous transactions
 - ▶ Car Rental and Leasing
 - ▶ Couriers
 - ▶ Scheduled Air Passenger
 - ▶ Retail trade – may be able to share with CPI
 - ▶ Hospital system
- Currently conducting a pilot for large grocery wholesalers

Secondary source:

Investment banking and securities dealing industry

Index Code	Index Title
523110	Investment banking and securities dealing
523110P	Primary services
5231102	Dealer transactions
523110201	Dealer transactions, equities
523110202	Dealer transactions, debt securities and all other trading
5231103	Investment banking services
5231104	Other securities dealing services
523110SM	Other receipts

Secondary Source:

Investment banking and securities dealing industry (cont.)

Broad Service Category	Definition	Examples of Securities Included
"Dealer transactions, equity securities"	Includes transactions in which the firm acts as a principal in buying or selling equity securities for the purpose of executing trades.	Stocks and Exchange traded funds
"Dealer transactions, Treasury securities"	Includes transactions in which the firm acts as a principal in buying or selling US Treasury securities for the purpose of executing trades.	Treasury bills, notes, and bonds; any other security issued by the US Treasury
"Dealer transactions, all other trading"	Includes transactions in which the firm acts as a principal in buying or selling debt (except US Treasuries) and derivative securities for the purpose of executing requested trades.	Corporate bonds, municipal bonds, agency bonds, mortgage-backed securities, asset-backed securities, collateralized mortgage obligations, commercial paper, certificates of deposit, Yankee bonds, foreign debt, options, warrants, futures, forwards, swaps
"Investment banking services, advisory services"	Includes mergers and acquisitions and other advisory services. Advice and assistance are provided to firms that are merging, acquiring other firms, or being acquired, leveraged buyouts, corporate restructuring, and the reorganization of bankrupt and troubled companies.	
"Investment banking services, underwriting services"	Includes all services related to the process of distributing new securities to investors, either through the public markets or to a private pool of investors.	
"Reverse repurchase agreement services"	Includes reverse repurchase agreement services, typically for a treasury note, agency bond, agency mortgage-backed security, or an investment grade corporate bond.	
"Securities loan services and all other securities dealing services"	Includes securities lending services and all other securities dealing services.	



Selection of a corporate bond

"Dealer transactions, all other trading"	Includes transactions in which the firm acts as a principal in buying or selling debt (except US Treasuries) and derivative securities for the purpose of executing requested trades.	Corporate bonds, municipal bonds, agency bonds, mortgage-backed securities, asset-backed securities, collateralized mortgage obligations, commercial paper, certificates of deposit, Yankee bonds, foreign debt, options, warrants, futures, forwards, swaps
---	--	---

Corporate Bonds -→ Highly traded CUSIP

Challenges with traditional sampling method

- Resource, sample size and respondent burden
- Limited number of sampled transactions
- Tremendous representativeness burden on each transaction

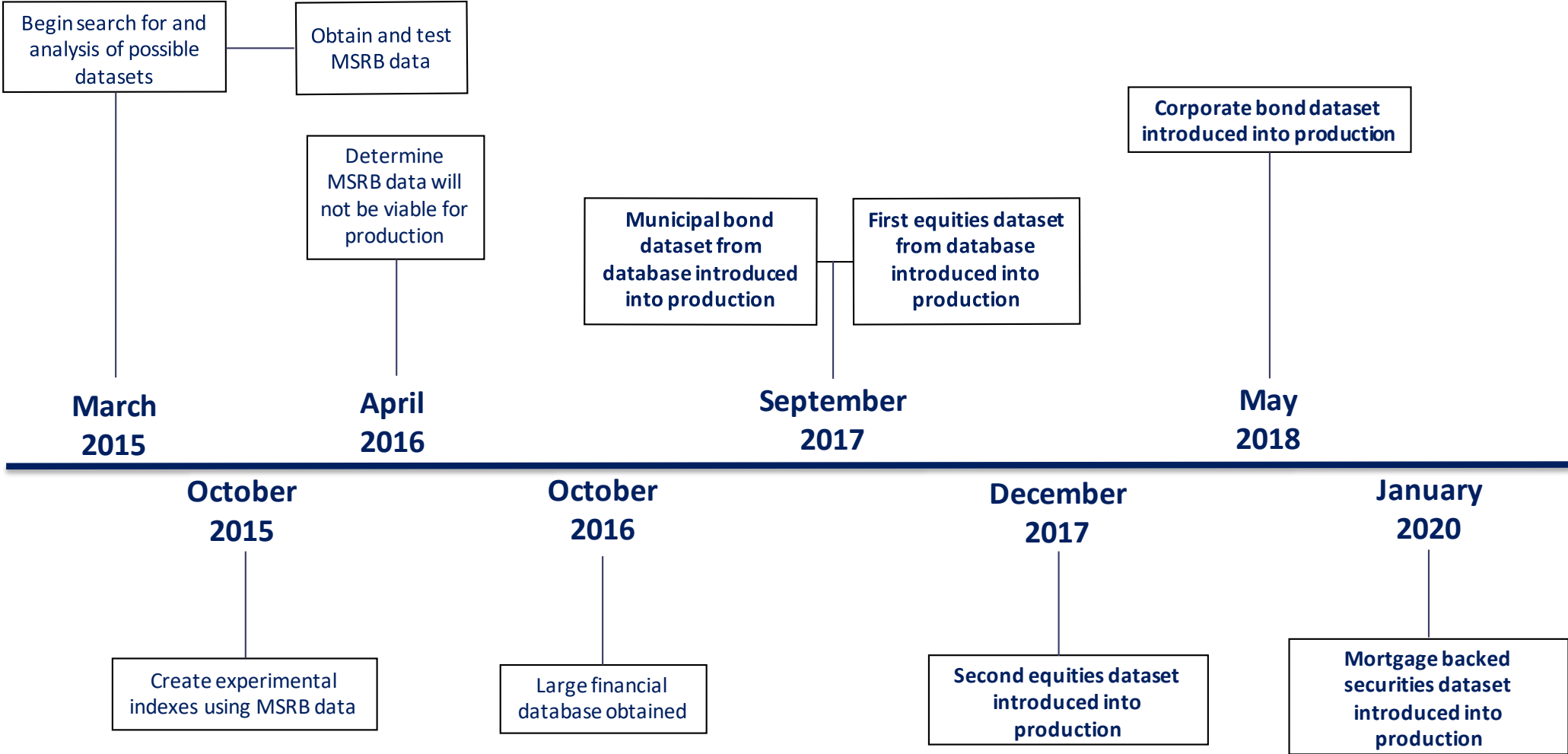


Search for alternative sources

- Exploration and evaluation of multiple sources
- Requirements:
 - ▶ Large number of transactions
 - ▶ Accurate pricing data for transactions
 - ▶ BLS legal and budget
 - ▶ Timeliness



Alternative data timeline



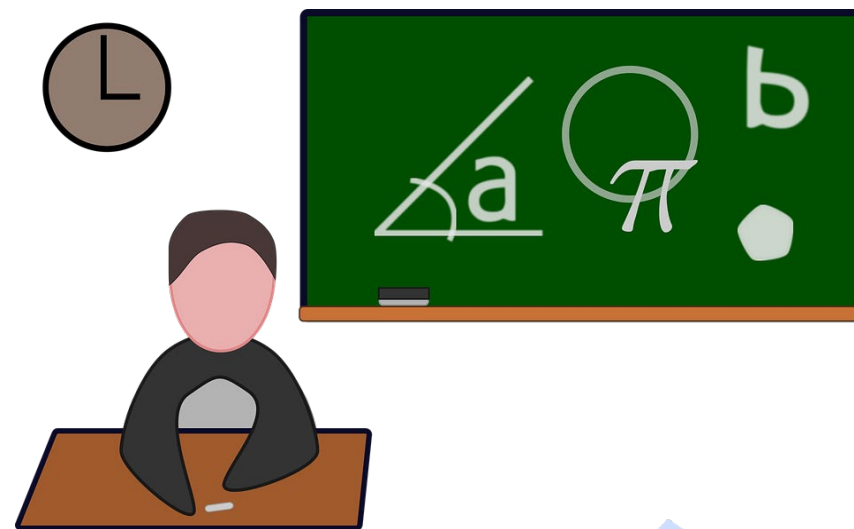
Solution: single source database

- Select products (not an entire lower level index) blended or replaced
 - ▶ Equities dealing prices
 - ▶ Debt dealing prices
- For more specific information:

[Implementing an Alternative Data Source to Estimate Producer Price Indexes for Selected Financial Services](#)

Secondary source data plans

- Explore adding new commodities for the financial service industry
- Evaluate identified alternative data sources for other industries



Contact information

Susan Fleck, Assistant Commissioner
International Price Program
fleck.susan@bls.gov

Bonnie Murphy, Branch Chief
Producer Price Program
murphy.bonnie@bls.gov

Alexander Paplomatas, Section Chief
Consumer Price Program
paplomatas.alexander@bls.gov

