

Big Data in the U.S. Consumer Price Index: Experiences & Plans

Crystal Konny, Brendan Williams,
and David Friedman

Federal Economic Statistical Advisory Committee
Meeting - June 14, 2019



Potential Benefits

- Transaction prices
- Larger sample sizes
- Reduced collection costs
- Reduced or eliminated respondent burden
- Data descriptiveness
- Real-time expenditures and weights



Challenges

■ Methodological

- ▶ Product life cycle, representativeness, data descriptiveness

■ Operational

- ▶ Data lag, continuity, quality verification
- ▶ Geographic structure
- ▶ System design

■ Legal, Policy, and Budgetary

- ▶ Contracting for data, webscraping agreements, confidentiality concerns

Alternative Data

- Data not collected through *traditional* field collection procedures by BLS staff
(*traditional* = in-store/on-phone/manually on website)
- Three main categories:
 - ▶ Corporate
 - ▶ Secondary Source
 - ▶ Web/Mobile app scraping data
- Decade of explorations & pilot projects – transition into production



General steps for Alt Data Projects

- Determine what to pursue
- Evaluate options
- Evaluate selected source (definition, coverage, other quality dimensions)
- Evaluate data quality over predefined time
- Methods to test
- Evaluate results
- Transition to production?



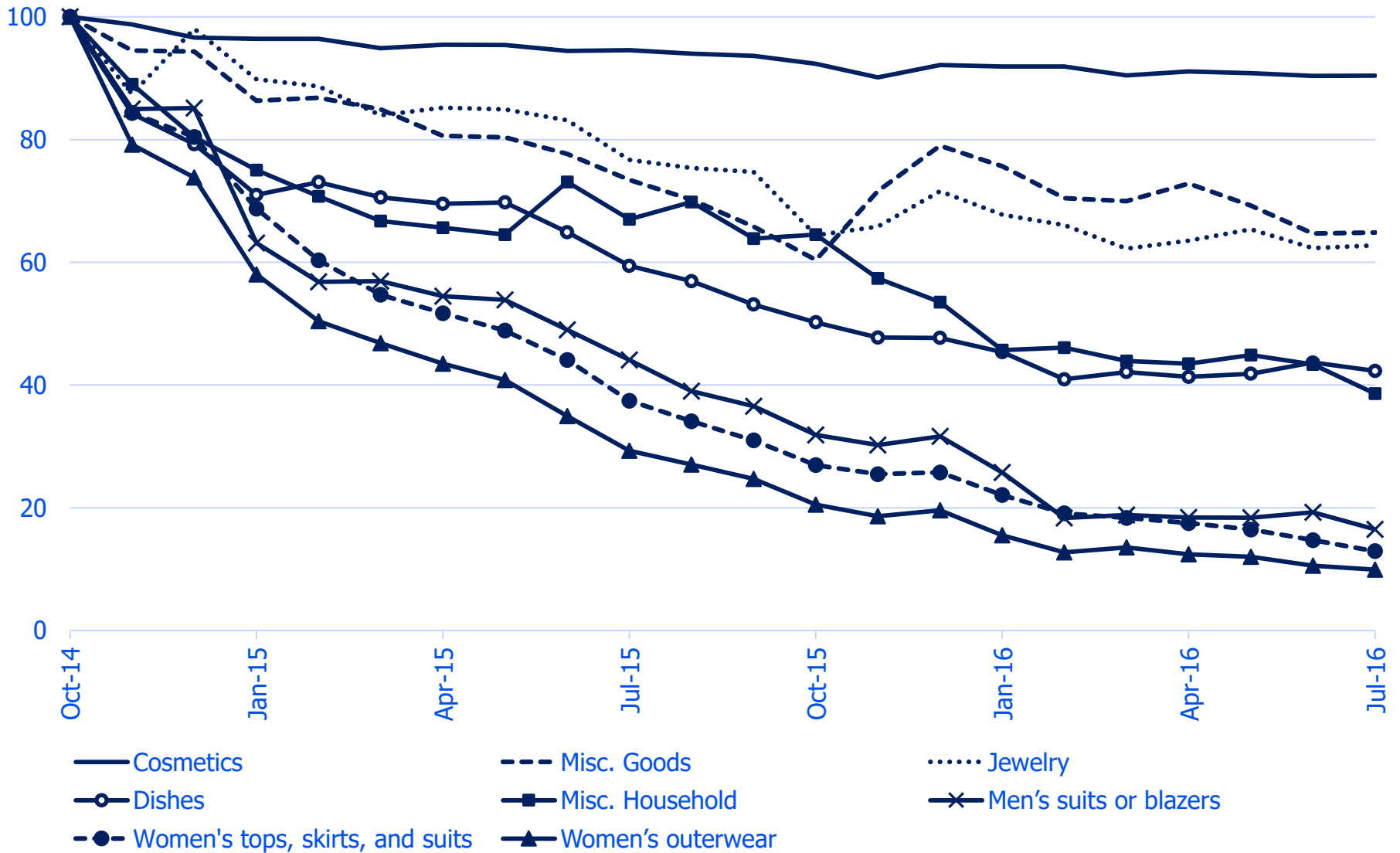
Criteria for use in production (to date)

- As good or better than current pricing methodology
- Does improvement in index justify any additional costs – cost effective?
- In general, is it a good fit for CPI?
- Use of short-term solution while continuing to research longer-term improvements

Corporate Data



CorpX



Impact of Incorporating CorpX



CorpY

- February 2012 refused to initiate new prescription drug sample
- March 2015 agreement to supply data corporately
- May 2015 first use in index



CorpY

	CorpY	In-store
Item Selection	Probability Proportional to Size (PPS) over the past year nationally by sales excluding lowest 10% of transactions	PPS based on price of the last 20 prescriptions sold
Geography	National	Outlet Specific
Price	Average price of at least 100 transactions	Single price
	Insurance prices	Mostly cash prices
	National price	Outlet specific price
	Per pill price	Per prescription price
Patent Loss	Unit prices averaged across brand and generic	Based on analyst monitoring of patents for an NDC
Data Frequency	Bimonthly odd collection	Monthly and bimonthly odd/even collection

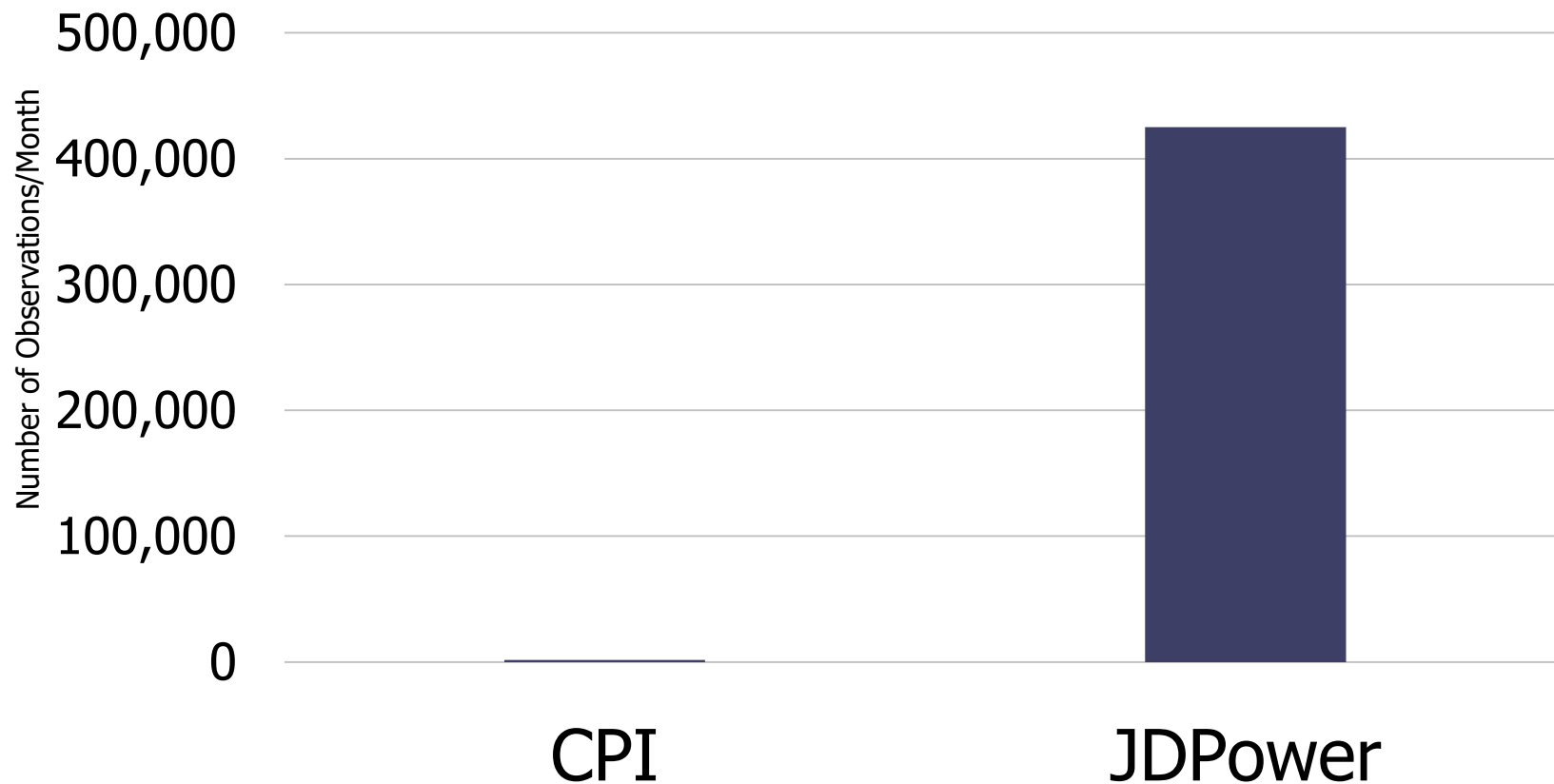
Secondary Source Data



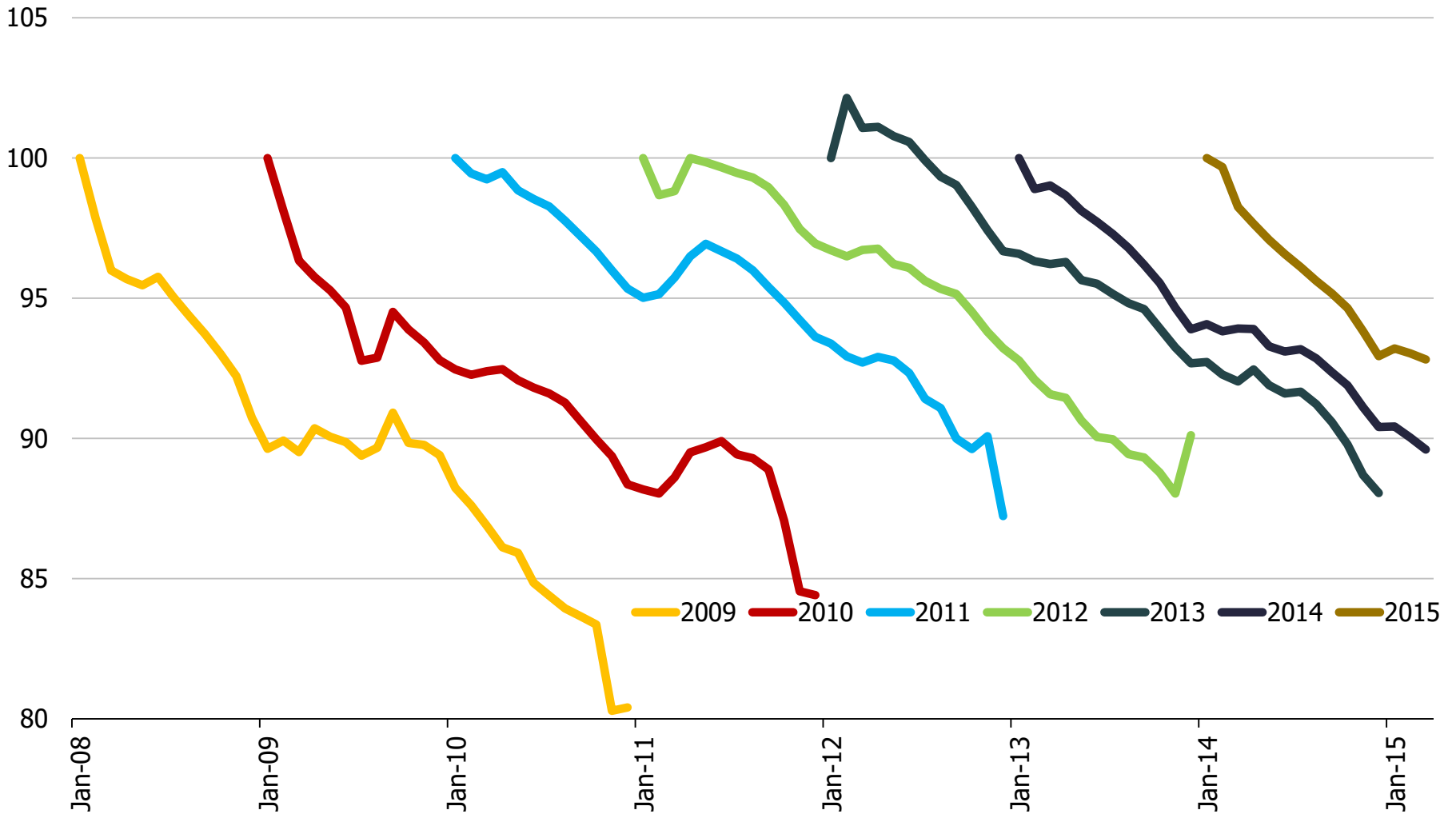
Hospitals and Physicians' Services

- Relative Importance 4.04%; response rate for Medical Care is 48.1%
 - 4,116 price quotes
 - Cash price overrepresented
 - High respondent burden
 - High collection costs
 - Difficult collection methodology
- ➔ Researching use of medical claims datasets

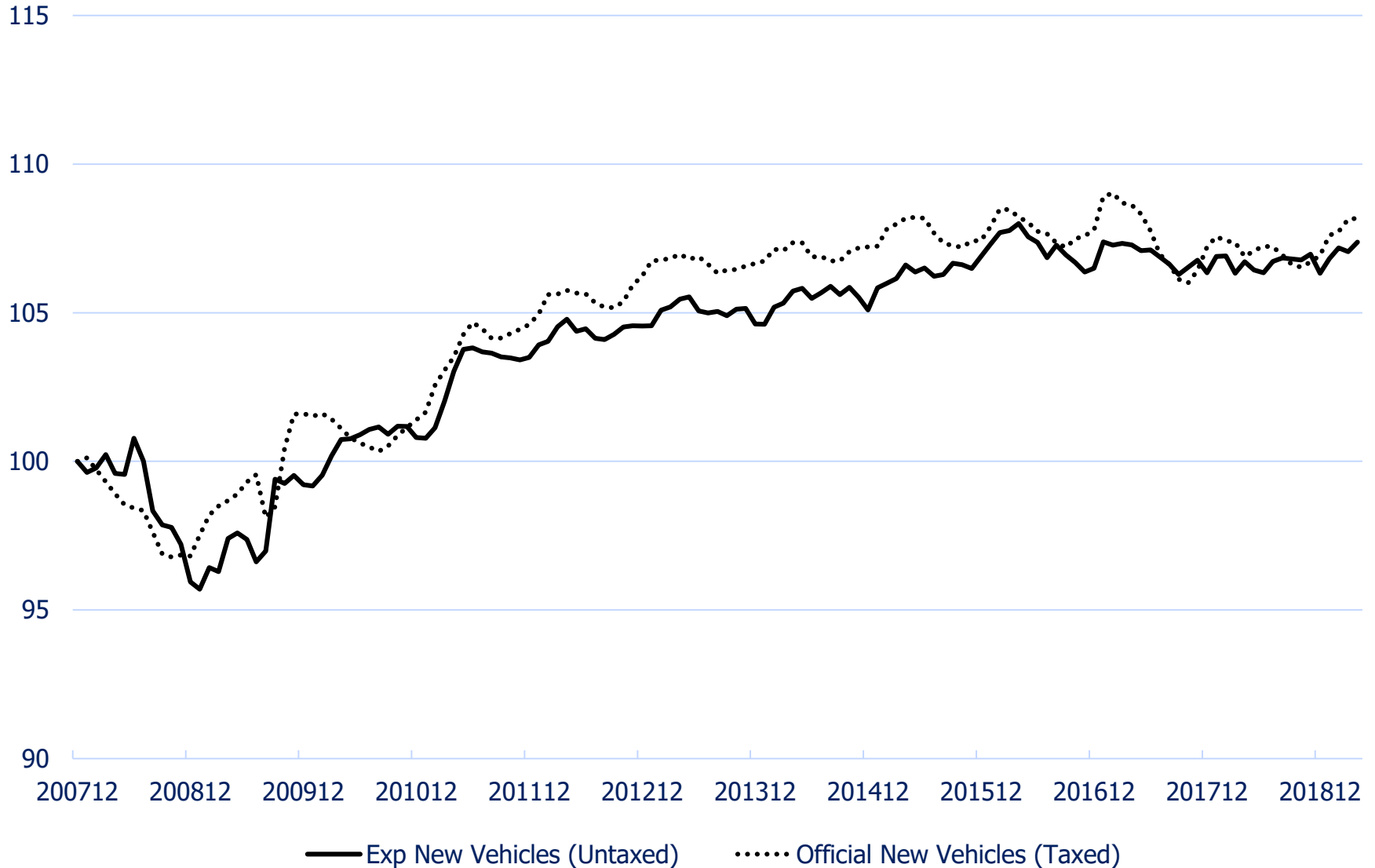
New Vehicle Observations



Model Year Price Indexes



Experimental Index for New Vehicles

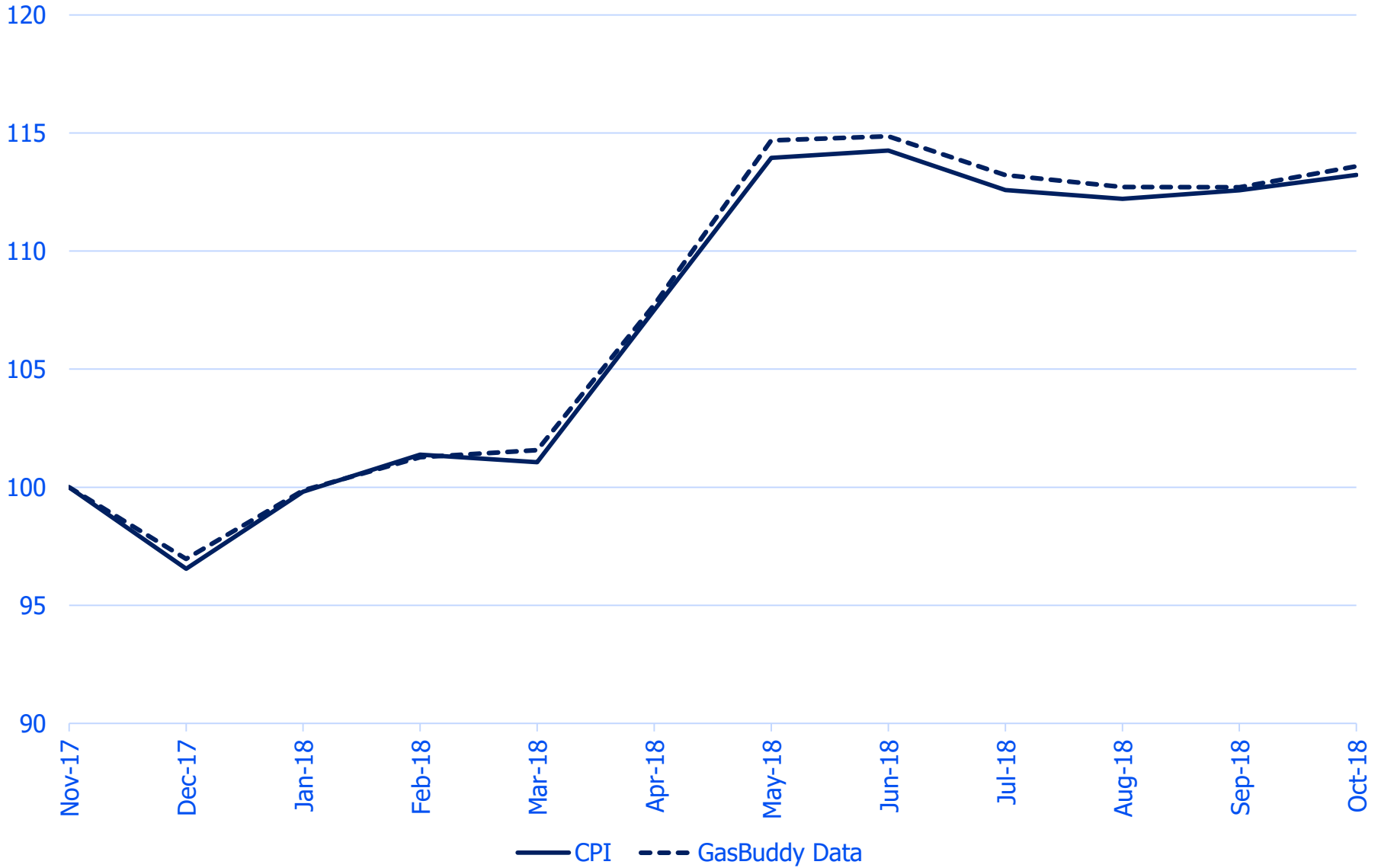


Web/Mobile app scraping data



Crowd Sourced Motor Fuels

Regular Unleaded Gasoline



Plans



Establishing Priorities

- Relative importance of the item
- Number of quotes replaced
- Cost of collection relative to cost of alternative data
- Respondent relationship with BLS
- Concentration of respondents in the sample
- Ease of implementation
- Accuracy issues in the current index...



In the works

Item	RI	# quotes	concentration	issues	priority	Source of data	% sample
Gasoline (all types)	4.344	3,778	M	L	H	scrape	100
Other motor fuels	0.094	830	M	L	H	scrape	90
New vehicles	3.695	1,900	L	H	H	sec	100
Physicians' services	1.728	1,993	L	H	H	sec	75
Hospital services	2.312	2,123	L	H	H	sec	85
Cable and satellite television service	1.501	1,906	H	H	H	sec	95
Wireless telephone services	1.693	1,279	H	H	H	sec	98
Land-line telephone services	0.572	874	H	H	H	sec	95
Internet services & electronic info providers	0.780	773	H	H	H	sec	95



In pursuit

	RI	# quotes	concentration	issues	priority	Source of data	Experience	% sample
Prescription drugs	1.316	4,641	H	H	H	corp	some	
Limited service meals and snacks	2.542	2,808	M	L	M	corp	pursue	
Delivery services	0.014	231	H	L		corp	pursue	
Airline fares	0.683	1,745	H	L	M	scrape, corp	research	
Used cars and trucks	2.329	4,537	H	H	H	sec	Prod, seek	100
Postage	0.094	230	H	L		sec	prod	
Leased cars and trucks	0.655	265	L	H	M	sec	research	100
Electricity	2.655	1,406	M	M	H		seek	
Utility (piped) gas service	0.747	1,404	M	M	H		seek	
Rent and OER	31.548						seek	



Conclusions

- Significant portion of the CPI based on alternative data within 5 years
- Substantial R&D on methodology needed
- Alternative data introduced incrementally alongside monthly publication

Questions for FESAC

- Do you have any reactions to the general criteria CPI has used to date for determining fitness for use? Are we missing anything, etc.?
- Do our criteria for establishing priorities in moving forward make sense to you?
- Any advice for meeting the methodological challenges BLS faces with some of the alternative data sources?



Contact Information

Crystal Konny

Branch Chief

Branch of Consumer Prices

Konny.crystal@bls.gov

Brendan Williams

Senior Economist

Branch of Consumer Prices

Williams.Brendan@bls.gov

David Friedman

Associate Commissioner

Prices and Living Conditions

Friedman.david@bls.gov

