Unit Values for Import and Export Price Indexes – A Proof of Concept

Don A. Fast and Susan E. Fleck

February 2019

Prepared for the CRIW Conference Big Data for 21st Century Economic Statistics

Abstract

The U.S. Bureau of Labor Statistics' import and export price indexes (MXPI) are published from an ever decreasing sample relative to the size of trade. The Principal Federal Economic Indicator has an opportunity to retain and regain detailed MXPI using unit values calculated from comprehensive administrative trade data. The unit values from the high-frequency, high-volume source present a dilemma for official price statistics, given that unit value indexes are known to not track price indexes. This BLS research proposes a new methodological and statistical approach to identify detailed homogeneous product categories that show minimal unit value bias to include in the MXPI. The proof of concept for identifying homogeneous items is based on an analysis of two export products – dairy and vegetables – for 2015-16. The results provide a prototype and a roadmap for a consistent and testable approach that aligns with the concepts in official MXPI measures, maximizes the use of high-frequency data, and mitigates the likelihood of unit value bias. Applying the prototype, 52 of 142 import and 50 of 129 export 5-digit BEA End Use categories are identified as homogeneous using administrative data. This coverage accounts for 35 and 39 percent of the 2016 value of imports and exports, respectively. Incorporating unit values has the potential to deepen coverage and expand publication of detailed import and export price indexes.

Don A. Fast is a Senior Economist and Susan E. Fleck is the Assistant Commissioner of the International Price Program at the U.S. Bureau of Labor Statistics (BLS). We particularly thank Christina Qiu and Daryl Slusher for their extensive contributions to the final chapter. We recognize and thank these BLS staff for their research and data support: Jeff Blaha, Antonio Caraballo, Jenny Fitzgerald, David Friedman, Michael Havlin, Ara Khatchadourian, Laurence Lang, Robert Martin, Helen McCulley, Steven Paben, Sudha Polumatla, Tamar Schmidt, Aric Schneider, Ilmo Sung, and Praveenkumar Yerramareddy. We gratefully acknowledge the conference organizers and Jon Samuels for feedback and advice. We have benefited from comments by Ana Aizcorbe, Alejandro Cavallo, John Haltiwanger, Marshall Reinsdorf, and our discussants, Carol Corrado and Susan Housman on earlier versions of this research. Author emails: fast.don@bls.gov, fleck.susan@bls.gov. Use of the export trade data are subject to Agreement No. 2067-2018-001, Memorandum of Understanding (MOU) between the U.S. Census Bureau and the Bureau of Labor Statistics (BLS). The BLS has received prior approval from the U.S. Census Bureau, which affirms that the research results do not present disclosure risks and approve the publication of the results.

Contents

INTRODUCTION 2
THE RESEARCH APPROACH 6
UNIT VALUES and UNIT VALUE BIAS 10
BENCHMARKING UNIT VALUE INDEXES WITH PUBLISHED PRICE INDEXES 15
AN INITIAL PROTOTYPE FOR UNIT VALUES and UNIT VALUE INDEXES 22
CONCLUSION 24

INTRODUCTION

BLS Import and Export Price Indexes (MXPI) track price changes in internationally traded merchandise goods. The indexes underpin inflation adjustment of U.S. net exports and trade balances from current to constant dollars. The quality of the indexes is founded on the matched model and implemented through an establishment survey. The matched model records same-good price differences at the item level and aggregates price changes weighted by product, company, and trade dollar value shares to all-goods import and export price indexes. For the past twenty years, 20 to 25 thousand prices of unique items from thousands of companies have been collected monthly to calculate detailed and all-goods price indexes. Simultaneously, trade has grown, resulting in a sample size that could not support the previously published level of detail. While the top-level MXPI - principal federal economic indicators – are of consistently high quality, measures for detailed price indexes are at risk. Symptomatic of this trend is the fact that BLS publishes only one-third of the most detailed BEA End Use goods price indexes for both imports and exports.

There exists an extensive source of administrative trade data that – up until now – has been used only as the sample frame for the international price establishment survey. The price and quantity information from these administrative records results in an average price or unit value, i.e., the total dollar value of the shipment divided by the quantity shipped. The 2.9 million monthly export records dwarf the approximately 24,000 export and import items currently in the directly collected international price survey. The question analyzed here is whether unit values can be used on a large scale to track price change to bolster the number and improve the quality of published detailed price indexes and, by extension, the top-level indexes.

Incorporating unit values on a large scale into a BLS price index is a major methodological change to existing practices, given that the BLS program was founded in response to critiques of unit value measures. The BLS established the international price program to directly collect price data, following significant research conducted by the National Bureau of Economic Research in the 1960s. The Stigler Commission (Price Statistics Review Committee 1961), a historical series of import and export price indexes for 11 commodity groups (Lipsey 1963), and an extensive study on the measurement and calculation of price measures for international trade (Kravis and Lipsey 1971) described how unit values captured compositional effects of changes in product mix and different quality of goods and did not mimic price changes. Unit value indexes at that time were calculated from average values for customs declarations that included value and quantity. The records were often incomplete, and thus unit values covered no more than a third of finished manufactured trade and slightly more than half of commodity trade (Kravis and Lipsey 1971). The ability to determine U.S. competitiveness was hampered because of the poor quality of these measures. The Census monthly unit value export and import indexes, published from July 1933 through 1990, were calculated for five broad economic commodity categories (crude materials, crude food-stuffs, manufactured foodstuffs and beverages, semimanufactures, and finished manufactures). The first BLS import and export price indexes based on an establishment survey were published in 1973 as a consequence of this high-profile research to replace the Census unit value indexes.

Since that time, some have argued that unit values for homogeneous goods may track prices (Silver 2010). But there is consensus that unit values are biased measures of heterogeneous or differentiated product prices. The first question at the heart of this research is whether the unit values of much improved trade data can replicate price movements of homogeneous goods. Beginning in the 1990s, detailed import and export transactions of 10-digit harmonized system (HS) product categories are recorded for nearly all traded goods with improving accuracy over time. Twenty years ago, Diewert and Feenstra proposed that BLS analyze detailed trade data, but at that time BLS had less capacity than today to address the complexity of and lag in the receipt of data, and so BLS did not pursue the project (Diewert and Feenstra 1997).

The second question is how to define homogeneous unit values with the trade data. Unit values are calculated for cross country comparisons, despite the known bias (Feenstra et.al. 2009, Feenstra and Romalis 2014). For the United States, recent research effectively treats all 10-digit HS product categories as homogeneous across the board, but also add one or two data characteristics to differentiate unit values. For example, Broda and Weinstein (2006) estimate the impact of variety changes on prices and welfares by including country of origin in their import indexes. Kamal and Monarch (2017) analyze the reliability of the trade data in the context of U.S.-foreign supplier relations. Monarch and Hottman (2018) create an import price index that includes the foreign supplier ID and map out the welfare impacts of import price changes on select consumer profiles. However, there is no proof that all product areas are homogeneous, as many 10-digit HS product categories are composed of differentiated goods. Consequently, there is no evidence that unit values at the 10-digit HS category level track price trends.

Crude petroleum imports are currently calculated using unit values from the detailed administrative data, forming a precedent in the International Price Program.[i] These unit values are deemed to be the most reliable source in the face of low response rates and the price volatility of this heavily traded product. Furthermore, crude petroleum product information is fairly detailed. This detail stands in contrast to the type of detail and level of homogeneity reported in the administrative trade data. The administrative trade data are rich in transaction details but have only one product description – the 10-digit HS product category. More significantly, the regulatory nature of trade has created unintended differences in degrees of homogeneity and product detail across 10-digit HS category. Is it possible to move beyond a 'special case' use of unit values, such as in crude petroleum, to a comprehensive approach?

Key to the decision of whether to use unit values from the administrative trade data is when and how to use unit values. BLS requires a consistent and transparent approach to evaluate 1)whether a product category is homogeneous and 2)to what degree unit value bias exists in the entry level item and the published index level. The potential to use unit values for the MXPI statistics faces two hurdles. The first – evaluating and establishing a proof of concept to select homogeneous categories and calculate indexes accurately – is the focus of this paper. The second – whether there is a way to integrate the lagged administrative data into official monthly production – will not be addressed here, but is not insignificant.

The chapter is organized as follows. The research approach outlines both item- and index-specific concepts and methods. 2015-16 export transaction records for homogeneous product categories dairy and vegetables are the data used for this pilot. Six options for calculating unit values of entry level items (ELIs) are described and analyzed. Prices and price changes (short term ratios, or STRs) are tested for unit value bias within and across months to identify optimal dimensions, which we call item keys that result in the least bias. ELI prices are then aggregated to 10-digit HS price indexes—applying Tornqvist index formulas and addressing imputation, outliers and consistency of trade, for all six ELI options. These results are then statistically compared to published price indexes as benchmarks for quality. Based on the analysis, a prototype unit value and index approach is proposed. All 5-digit import and export product categories are evaluated to define the set of potential homogeneous products. From this list, a couple of export indexes of homogeneous goods are compared against their benchmark published indexes to gauge how well the prototype fits. These promising first results provide a road map to comprehensively evaluate all homogeneous import and export price indexes.

THE RESEARCH APPROACH

Maintaining the standard for Principal Federal Economic Indicators when considering new concepts or methodology requires thoughtful and thorough review. This research evaluates which 10-digit HS levels are homogeneous and whether a more detailed unit value than the 10-digit HS level is necessary to mitigate compositional effects on the value. The most general case is one in which all or some 10-digit HS unit values are as good a measure of price change as the published import and export price indexes.

The research develops and evaluates new methods to calculate unit value prices and indexes with administrative trade data, using a small subset of export data for two years (2015-2016) for two product areas – dairy and eggs (BEA End Use Classification 00310), and vegetables, vegetable preparations, and juices (BEA End Use Classification 00330).[ii] Indeed, the two product categories were selected because of their homogeneity as judged by the product substitutability at the BEA End Use and 10-digit HS strata, which was evaluated by the number and homogeneity of the 10-digit HS categories (see Table A) – and also because the two 5-digit indexes were not of sufficient quality to be published. The first assumption is shown to be a reasonable premise for selection, but the second created difficulties in validating the consistency and quality of the pilot measures for the two selected product areas; comparisons with additional product areas validates this approach.

Table A. Characteristics of Select BEA End Use Export Products, 2016-2017

BEA End Use Export Classification	Average number of monthly transactions	Number of 10-digit HS categories
Dairy products & eggs	6,839	41
Vegetables and vegetable preparations and juices	32,430	161

Three principles guide the methodological approaches in this research – to approximate the matched model, to evaluate characteristics of homogeneity, and to improve the measurement of the index where possible. Unlike consumer goods, wholesale goods require contracts and financing in place for a transaction to occur. Wholesale trade depends on business-to-business long-term relations; business characteristics reveal these relations. International trade transactions are more logistically complex and depend on well-defined sales contracts in order to be backed by a letter of credit from a financial institution (Amiti and Weinstein 2009). Thus, transaction characteristics that define a sale also are likely to signal similarity of products and purchasers.

Approximating the Matched Model. A matched model links the same item across time. For our research, transactions that are similar along multiple dimensions are assumed to be more likely to have the same composition of goods, since cross-border trade depends on long-term contracts between firms, and products are expected to have the same non-price characteristics across time. The detail level of each item key used to break out the administrative data into more dimensions is expected to test the compositional bias of unit values.

Characteristics of Homogeneous Goods. Approximating the matched model, homogeneous unit values could track close substitutes over time, as long as transactions can be grouped in such a way as to minimize compositional effects and to maximize substitutability. Establishing conditions of substitutability helps to define homogeneity, and vice versa. Research on price-setting informs the notion of substitutability, beginning with Rauch (2001) who separates goods into homogeneous and differentiated product categories, where homogeneous goods are reference-priced. In studies of exchange rate pass-through spanning nearly 100,000 goods in the international price survey from 1994 to 2005, Gopinath and Itskhoki (2010) and Gopinath and Rigobon (2008) demonstrate that homogeneous goods experience both more frequent and larger price changes than differentiated goods. These differences are attributed to larger elasticities of demand by consumers contributing to greater costs of price stickiness for producers. Both characteristics point to greater levels of substitution among homogeneous goods, suggesting that large fluctuations of homogeneous item prices do not consistently translate to large fluctuations in consumer spending. Thus, in the case of homogeneous goods, unit values may more accurately represent import and export prices by accounting for intra-item substitutability. Additionally, the unit value indexes calculated from the unit values are expected to not demonstrate the "product replacement bias" of matched models delineated in Nakamura et al. (2012), where frequent product turnover results in no price changes for 40% of imported items.

Rauch (2001) notes that business networks linking country of origin and country of destination play an important role in market share, price, and trade volume of goods. This leads us to assume that 10-digit HS product categories on their own are too broad to see sufficient levels of intra-item substitutability for which unit value indexes demonstrate the above benefits. Matching transactions to a greater level of specificity than the 10-digit HS product categories takes into account price- and non-price determining trade characteristics that separate goods into unique bins of substitutable items. Given the high frequency of transactions in trade data, each bin is likely to have more than one transaction. In other words, we aim to increase intra-item substitutability by enforcing stricter category inclusion criteria. This approach mirrors the quality-adjustment method of Feenstra and Romalis (2014) for industry-level World Development Indicators that displays less sensitivity to assumptions on the extensive margin of firms.

Addressing Criticisms and Improving Measures. Mismeasurement of trade impacts other indicators such as real GDP and productivity. The matched model has been criticized for measuring price changes of the same good only, and missing prices for new goods and different quality goods (Feldstein 2017). The focus of this criticism has been on an expected upward bias of price indexes for differentiated goods. Even though quality adjustments are not characteristic of homogeneous goods, the proposed method potentially accounts for consumer substitution related to quality change because all items and transactions are weighted using their current trade share.

The ability to account for new products and disappearing products and product varieties is a benefit of the new method because the current values for all items are available and can be integrated into a superlative unit value index. Particularly, the Tornqvist index is known to adequately address substitution bias and can be implemented with the proposed unit value indexes (Diewert 1976). It is important to note that the lag in collection of new goods and the lack of current weights to account for changing tastes and trading patterns are not inherent in the matched model method, but are related instead to the resources available for timely data collection. The administrative data expands the ability to account for new goods, to exclude products that are no longer traded, and to use current weights in a superlative index to account for substitution.

The prices and indexes calculated and presented here are based on the three principles described above. They are tested and evaluated for the degree of homogeneity and the existence of unit value bias. Basic parameters are established as a result of this research to 1)define a homogeneous unit value and item, 2)test homogeneity of an item, 3)identify appropriate published benchmarks for comparison, and 4)propose the concepts and methods to use for survey production. These parameters provide the roadmap to systemically evaluate homogeneity at the item and index levels.

UNIT VALUES and UNIT VALUE BIAS

Defining Unit Values. The point of departure for the research is to establish the 10-digit HS product category as the general case for evaluating unit values. This level of detail is naturally occurring in the administrative trade data – as records are HS-specific.[iii] Given the fact that 10-digit HS is also the stratum from which MXPI indexes are sampled and calculated, this level of detail provides the most convenient entry point to blend the unit values into the statistical production process. By beginning with the general case, the BLS research tests the premise that the 10-digit HS products are homogeneous, and proceeds with testing unit values of items that add more dimensions to approach the matched model. To align with the entry point for blending data, these unique items are considered to be entry level items that comprise a 10-digit HS stratum.

Whereas the general case occurs when the item key contains only the 10-digit HS code (H), all other item key specifications include dimensions that have been determined to be similar to price-determining characteristics in the survey. The data fields used in the item keys include: HS commodity classification, EIN (establishment ID number) for the exporting company, zip code, state of origin, domestic port of export, country of destination, related or non-related trade[iv] and unit of measure. The data fields – HS, EIN, and zip code– correspond with the sampling unit and the method of collection. Sampling for the directly collected international price survey is carried out at the 10-digit HS product category and at the establishment and company level by location of exporter. The data fields – state of origin, port of export, country of destination, related or arms-length transaction – correspond to production and/or market relations between exporter and foreign consumer. Most of these descriptors are also collected in the survey as price-determining characteristics. For measurement consistency, the unit of measure, e.g. gross, piece, ton, is also included. Each item key specification results in a different set of unique items, or ELIs, with the same dimensions grouped by the same shared characteristics. A total of six variations of item keys are tested.

The unit value is calculated at the level of the transaction. The unit value can be represented as a transaction *i* of a unique item *j* in month *t*, where *j* is composed of a 10-digit HS code *H*, and is further defined by an array of price characteristics, item key *K*. Transaction *i* involves the trade of *z* actual items, where $|z|$ is the number of actual items traded in transaction *i*. The unit value price of a transaction *i* is the average of prices for actual items traded in *I*, or

$$\text{(i)} \quad p_{K_i}^{(j,t),H} = \frac{\sum_{z \in i} p_{K_{i,z}}^{(j,t),H}}{|z|},$$

where $|z|$ can alternatively be represented as $q_{K_i}^{(j,t),H}$.

For all like-transactions of a given *K* that comprise the unique item *j*, the price of item *j* is represented as a geometric mean of unit value transaction prices which yields:

$$\text{(ii)} \quad p_{(j,t)}^{H} = \prod_{i \in (j,t)} \exp\left[w_{K_i}^{(j,t),H} \cdot ln\left(p_{K_i}^{(j,t),H}\right)\right]$$

Where normalized transaction-level weights are represented as

$$w_{K_i}^{(j,t),H} = \frac{\sum_{z\in i} p_{K_{i,z}}^{(j,t),H}}{\sum_{i\in K}\sum_{z\in i} p_{K_{i,z}}^{(j,t),H}}.$$

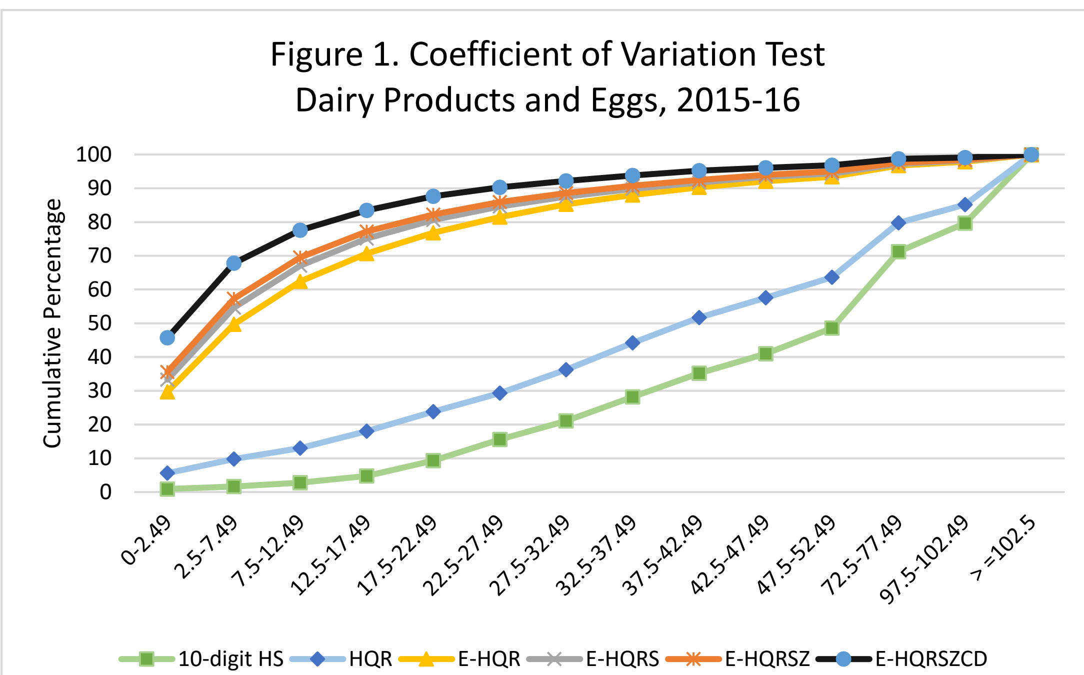
The quantity of item j is represented as a sum of transaction quantities:

(iii) $$q_{(j,t)}^{H} = \sum_{i\in K} q_{K_i}^{(j,t),H}.$$

Taking this experimental approach to test different specifications of items supports the objective to identify the best unit value measure. The price changes of actual transactions based on dimensions for six item key specifications are used for the unit value tests.

Testing Unit Value Bias. To test for unit value bias, one must consider the price characteristics of a homogeneous item. Homogeneous items are close, if not perfect, substitutes. Thus, in a competitive market, they would be expected to have similar price levels and be affected by the same market conditions over time. For one ELI with multiple transactions, we call this condition intra-item substitutability. If there is no supply or demand shock or large exchange rate fluctuation, one would expect a homogeneous product's within-month prices to group close to a mean, and its cross-month prices to show smoothness. For an item which faces a market shock, prices may cluster around more than one mean price. Although some HS 10-digit product categories experience more variable prices both within and across months, the large majority of items display little price change between months. Efforts to define homogeneity in a consistent way lead us to apply three tests to the prices and price changes of items for the six item key specifications. Of these tests – the price dispersion test, an across month item percentage change test, and price clustering tests – the first shows promise.

The price dispersion test was conducted on the actual unit values for dairy and vegetables transactions. The coefficient of variation is the ratio of the weighted standard deviation to the weighted mean; lower percentages indicate less variability in the ELI. The greater the variability of prices within a month, the lower the level of intra-item substitutability and the less likely a good is homogeneous. This test fits with the trade literature that states that similar products from a producer are priced similarly. The intra-month intra-item unit values for each of the six item keys were evaluated for all 24 months. Results are shown for dairy unit values only, as vegetables trend similarly. In Figure 1 bins specify ranges of C.V.s. The least detailed item keys that exclude the company identifier (E) result in a concave cumulative distribution, in which the vast majority of ELIs present with high variability of within-month prices, which implies poor intra-item substitutability. About 60 percent of dairy products had a C.V. of less than 52.5 percent for the two item keys that exclude E. When the company identifier is added to the ELI specification, prices cluster closer to the mean – 60 percent of the ELIs that include the company identifier had a C.V. less than 12.5 percent. Furthermore, the most detailed item key, which



NOTE: Letters correspond to these non-price transaction characteristics: EIN (E), 10-digit HS (H), Unit of Measure (Q), Related Transaction (R), State of Origin (S), Zip Code of shipper (Z), Country of Destination(C), Domestic Port Code (D) includes company identifier and country of destination, experiences the least price dispersion for each good. The wide dispersion and variability shown in the item keys that exclude the EIN demonstrate more unit value bias than those that include that characteristic.

Another test of homogeneity looks at the month over month percentage change. Monthly price changes are grouped into price variability bins for all months. Following on past price-setting research that price variability across months is not expected to be large, any such price change across months for item keys could indicate that the ELI may not represent the same good. Looking at the cumulative results for dairy and vegetables, both show 75-85 percent of ELIs with less than 22.5 percent monthly price changes. These results do not reveal intra-item substitutability improvements with additional item key dimensions and are not informative for item key selection or unit value bias.

Two types of price cluster tests are applied to the price data for the ELIs—the first method minimizes the variance in the price cluster created (Ward Minimum Variance Method) and the second method minimizes the distance in the price clusters created (SAS Clustering Method 1). Assuming no price shocks and no unit value bias, the optimal number of clusters for each ELI should be one, as the item's price should reflect intra-item substitutability. Ward Minimum Variance Method was applied to price clusters for all ELI that had 100 or more transactions during the two year period. The clustering results show that all item keys for both vegetables and dairy saw around 80 percent of their ELIs falling within one cluster. However, results are sensitive to price cluster distance when using SAS Clustering Method 1. When EIN is included in the item key, the ELIs fall in one cluster around 60-63 percent of the time, compared to 31-40 percent of the time when it is excluded. These results suggest that including EIN in the item key increases intra-item substitutability. Yet when outliers are removed at the second standard deviation from the mean, ELIs had one cluster around 78-91 percent of the time, demonstrating no definitive difference from the most general case of the 10-digit HS unit values.

The results of the coefficient of variation test align with the expectations of intra-item substitutability, showing that the more detailed ELIs have similar within month unit values. This test has strong explanatory power and is used to evaluate item homogeneity.

BENCHMARKING UNIT VALUE INDEXES WITH PUBLISHED PRICE INDEXES

In addition to establishing an approach to identify the best ELIs with the least biased unit values, we consider the options for calculating the least biased unit value indexes, and then compare them to BLS price indexes. Given the voluminous data and the existence of current prices and quantities, we are interested in implementing improvements to the index where feasible. Precisely because the data are so voluminous and the options for ELIs are broad, the options for index calculation, imputations, and outliers can have widely different results. Comparing the BEA 5-digit unit value indexes is first, an important step in validating the difference between directly collected survey data and administrative data, and second, a necessary step to evaluate how and whether the unit value indexes will impact the top-level indexes if incorporated. As the international price survey samples items and weights prices based on patterns of trade, published XPI indexes most accurately reflect trade prices, but the initial conditions in selecting the pilot product areas make this comparison tentative at best.

Unit Value Index Calculation Methods. Unit value indexes are calculated at the level of 10-digit HS strata. This procedure generally provides an opportunity to incorporate current weights. The problem of missing prices is addressed both for the regular continuation of an ELI in the index and also as it relates to consistency of establishments' trade. The likely problem of outliers that arises with high-frequency low-detail data is also addressed.

Tornqvist index formula. The long term relative (LTR) of the 10-digit stratum is the entry point for blending data. For official indexes, company weights are used to aggregate ELI price changes to the 10-digit HS product category, and then trade dollar weights - lagged two years - for 10-digit categories are used to aggregate the LTRs and map them into the BEA End Use price index and other classifications. Because current period weights are available, the unit value ELIs can be aggregated into their corresponding 10-digit strata. The 10-digit HS unit value Tornqvist indexes are then aggregated into the BEA 5-digit index using official estimation procedures. The Tornqvist index is superior to Laspeyeres because it accounts for the introduction of new goods, disappearing goods, and trade volumes (Diewert 1976, Triplett 1992). At the 10-digit HS stratum, the general case is to use the unit value as the entry level item.

Using the current period weights, the 10-digit HS stratum is represented by a Tornqvist index comprising all unique items *j*:

(iv) $$R_H = \prod_{(j,t)\in H} \left(\frac{p^H_{(j,t)}}{p^H_{(j,t-1)}} \right)^{\frac{W^H_{(j,t-1)} + W^H_{(j,t)}}{2}}$$

where $W^H_{(j,t)} = \frac{p^H_{(j,t)} q^H_{(j,t)}}{\sum_{(j,t)\in H} p^H_{(j,t)} q^H_{(j,t)}}$.

These calculations differ from existing methodology, not only for the use of unit values, but also for the use of current weights to account for item turnover. The opportunity to apply the Tornqvist index to the unit values addresses a common criticism of the official indexes – that they do not sufficiently account for substitution of new items[v].

*Missing prices, consistency of trade and outliers***.** In order to evaluate the unit value indexes, more advanced methodological issues must be addressed to treat missing prices, consistency of trade and outliers.

Two months of actual prices establish an item in the index. Imputation fills in the gaps when an item is not traded and an item's price is of questionable quality.[vi] Even though 80 percent of the dairy and vegetable establishments in the two-year dataset are traded every month at the 5-digit BEA product level, the items traded each month vary considerably, resulting in missing prices. Missing prices were even more prevalent as dimensions were added to the item key, because each ELI had fewer transactions and experienced more turnover. Imputation is used to maintain items in the index, but there is a point at which imputation negatively impact index quality. To minimize the negative impact that continuing imputed prices over time has on the mean of the ELI, imputation - and thus items – are eliminated from the calculation if no transaction is recorded after three months. Beyond that point, the price imputations overwhelmed the count of unit values calculated from transaction records by more than two to one.

The decision whether or not to include inconsistently traded items affects the number of ELIs used in unit value index calculation. Including inconsistently traded items increase the use of imputation. However, excluding items that are not consistently traded because item keys have more dimensions could bias unit values by excluding too many items. Thus, two variations of unit value calculation are tested: retaining all items regardless of consistency of trade and truncating items that are traded less than half the year. Both approaches preserve the 3-month imputation rule set above.

The decision whether or not to eliminate outliers is of particular importance for unit value index calculation. In the official MXPI, an outlier price is flagged to evaluate the validity of price change. However, an outlier in the unit value of the transaction cannot be evaluated in the same way and may represent an error, another product, or a better product. Three unit value index calculations are considered: retaining the outlier, and recalculating the unit value with an imputed price when the price change falls outside both the weighted 2nd and 3rd standard deviations (note that the 2nd SD is more restrictive than the 3rd SD).

We nest outlier treatment within the two conditions of restrictions on consistent trade. Combined, these variations create six alternatives to calculating unit value indexes. Table 1 shows the index calculation methods from the least constrained to most constrained options regarding truncation of ELIs. All methods use the Tornqvist index formula and impute missing prices for up to three months. The first three calculation methods include all items, and the last three calculation methods truncate items that are not consistently traded.

Benchmark Comparisons. The comparison of the unit value indexes against a benchmark of BLS price indexes helps narrow down the proof of concept – of six different item keys that define the ELI and six different methodological approaches to calculate the unit value indexes – to a prototype. The 5-digit BEA End Use unit value indexes for dairy and vegetables are calculated from the 10-digit HS strata with the methods used for the official MXPI, and these indexes are then compared with a BLS price index as a benchmark. Holding all else equal, the company identifier significantly improves the correlation and reduces the root mean squared error. More detailed item keys show a closer fit than the general case of the 10-digit HS ELI. The differences between the index calculation methods of including or excluding consistent trade and treatment of outliers are not as clear-cut.

Export Price Indexes (XPI), spot prices, Consumer Price Indexes (CPI) (U.S. city average, all urban consumers, not seasonally adjusted), and Producer Price Indexes (PPI) were considered as possible benchmarks for unit value indexes. As MXPI samples and weights items based on patterns of trade, we expect published XPI to most accurately reflect trade prices. However, since the two product groups were chosen due to quality concerns, the XPI for dairy and vegetables for this time period were respectively unpublished and thinly priced. Additionally, for vegetables, fixed annual weights on XPI do not account for volumes of trade of the 140 10-digit HS strata in that index, almost all of which change seasonally. Whereas the unpublished XPI was chosen as a benchmark for dairy, the CPI was chosen for vegetables due to seasonal weighting concerns of the vegetable XPI. Although the CPI makes a better benchmark than the XPI for the vegetable unit value index, the CPI is generally a second-best comparative benchmark, since consumer prices systematically vary from export prices.

Correlation coefficient comparison. Correlation coefficients assess how closely indexes calculated from administrative data track changes in benchmark price indexes, where an estimate of 1 suggests perfect alignment. We apply the six variations of unit value index calculations for the six selected item keys. Though the benefits of unit value indexes are realized at higher-dimensional item key specifications than the 10-digit HS level, item key specifications that are too detailed may be "over-fitted"—failing to account for intra-item substitution and impacting too heavily price changes of high-volume item keys. Additionally, truncating outliers may introduce bias if outliers represent real price shocks.

Generally, correlation coefficients for dairy are higher than correlation coefficients for vegetables, i.e. the dairy unit value index does a better job of tracking the price trends in the benchmark index. For dairy, correlation coefficients remain consistent across different treatments of outliers and consistent trade. Correlation coefficients vary more for vegetables, pointing to a less consistent time series. Dairy correlation coefficients significantly improve after including company identifier into item keys. Dairy correlation coefficients for E-HQR item keys are on average 0.090 higher than correlation coefficients for HQR item keys—contrasting to correlation coefficients lowering by 0.002 on average after inclusion of a non-EIN dimension into item keys.

The large increase in dairy correlation coefficients after inclusion of EIN in item keys implies that product differentiation may occur at the firm-level for items in the dairy category. This pattern, however, is not reflected for vegetables. Using the same HQR and E-HQR item key comparison, the inclusion of EIN into vegetable item keys results in correlation coefficients that are on average 0.012 lower than HQR correlation coefficients. This statistic is a smaller magnitude than the average 0.020 correlation coefficient increase for the inclusion of all other non-EIN dimension into vegetable item keys.

The impact of index calculation methods on the correlation coefficient is less informative. The dairy correlation coefficient is largest at 0.61 using the most detailed item key with the least constrained index calculation method. The vegetable correlation coefficient is largest at 0.48 using the HQR item key with the least constrained index calculation method. However, the vegetable correlation coefficient also reaches consistent comparable levels for item keys that include EIN with the most constrained index calculation method. EIN inclusion into item keys improves dairy unit values' mirroring of the unpublished XPI benchmark, no matter the index calculation method, whereas – except for the case mentioned above – vegetable unit value indexes track the CPI benchmark most closely with EIN inclusion and for index calculation treatment of imputations and outliers.

Root mean squared error/mean absolute error comparison. Root mean squared error and mean absolute error measure differences between calculated and benchmark price indexes, indicating accuracy. Large differences are more heavily weighted in root mean squared error than in mean absolute error. An error value of 0 implies perfect similarity between unit value and benchmark price indexes. As can be seen in Table 1, across index calculation variations, the dairy unit value index displays larger error than the vegetable unit value index compared to their respective benchmarks. For both indexes, error measures trend downwards as item keys become more detailed, implying that accuracy increases with higher-dimensional identification regardless of index calculation methods.

Table 1. Unit Value Index Comparison to Published Price Indexes, Dairy and Vegetables, 2015-2016

	Exclude Company Identifier		Include Company Identifier (EIN)			
	10-digit HS	+ transfer price + unit measure	+ company identifier	+ state of origin	+ zip code of shipper	+ country of destination + U.S. port
Dairy U.V.Index	**Correlation Coefficient**					
Tornqvist index w/ 4 month imputation	0.48	0.5	0.58	0.6	0.59	0.61
+ exclude outliers 3rd Std.	0.5	0.51	0.6	0.62	0.6	0.59
+ exclude outliers 2nd Std.	0.5	0.52	0.57	0.6	0.6	0.57
Tornqvist index w/ 4 month imputation + consistent trade	0.48	0.5	0.61	0.6	0.58	0.59
+ exclude outliers 3rd Std.	0.5	0.53	0.62	0.58	0.57	0.57
+ exclude outliers 2nd Std.	0.5	0.52	0.64	0.6	0.53	0.57
	Root Mean Squared Errors / Mean Absolute Errors					
Tornqvist index w/ 4 month imputation	2.71 / 2.16	2.61 / 2.07	2.00 / 1.57	1.91 / 1.45	1.90 / 1.35	1.82 / 1.44
+ exclude outliers 3rd Std.	2.61 / 2.10	2.55 / 2.06	2.02 / 1.50	1.90 / 1.43	1.96 / 1.50	1.88 / 1.50
+ exclude outliers 2nd Std.	2.61 / 2.10	2.58 / 2.09	2.07 / 1.53	2.00 / 1.47	1.97 / 1.50	1.96 / 1.60
Tornqvist index w/ 4 month imputation + consistent trade	2.72 / 2.18	2.59 / 2.10	1.99 / 1.53	2.04 / 1.52	2.03 / 1.48	1.96 / 1.54
+ exclude outliers 3rd Std.	2.61 / 2.11	2.56 / 2.11	2.05 / 1.53	2.08 / 1.63	2.08 / 1.67	2.07 / 1.58
+ exclude outliers 2nd Std.	2.61 / 2.11	2.56 / 2.10	1.99 / 1.52	2.07 / 1.52	2.22 / 1.65	2.04 / 1.57
Vegetable U.V.Index	**Correlation Coefficient**					
Tornqvist index w/ 4 month imputation	0.37	0.48	0.24	0.23	0.29	0.35
+ exclude outliers 3rd Std.	0.32	0.30	0.35	0.34	0.40	0.39
+ exclude outliers 2nd Std.	0.32	0.31	0.35	0.37	0.35	0.39
Tornqvist index w/ 4 month imputation + consistent trade	0.26	0.37	0.32	0.35	0.37	0.33
+ exclude outliers 3rd Std.	0.33	0.32	0.38	0.38	0.43	0.41
+ exclude outliers 2nd Std.	0.33	0.33	0.40	0.45	0.46	0.47
	Root Mean Squared Errors / Mean Absolute Errors					
Tornqvist index w/ 4 month imputation	2.37 / 1.94	1.92 / 1.51	2.07 / 1.67	2.13 / 1.68	2.02 / 1.60	1.86 / 1.34
+ exclude outliers 3rd Std.	2.02 / 1.56	2.02 / 1.49	1.82 / 1.41	1.86 / 1.49	1.79 / 1.42	1.82 / 1.39
+ exclude outliers 2nd Std.	2.02 / 1.56	2.03 / 1.50	1.82 / 1.45	1.82 / 1.45	1.84 / 1.41	1.79 / 1.34
Tornqvist index w/ 4 month imputation + consistent trade	2.50 / 2.04	2.07 / 1.57	1.92 / 1.53	1.84 / 1.40	1.79 / 1.41	1.92 / 1.43
+ exclude outliers 3rd Std.	2.00 / 1.55	1.98 / 1.46	1.83 / 1.45	1.82 / 1.42	1.75 / 1.42	1.84 / 1.44
+ exclude outliers 2nd Std.	2.00 / 1.55	1.99 / 1.47	1.79 / 1.44	1.73 / 1.40	1.67 / 1.31	1.69 / 1.33

Similar to correlation coefficient trends, error decreases most significantly for dairy when EIN is added into the item key, a trend that is not observed for vegetables. Mirroring previous analysis on correlation coefficients, root mean squared error decreases by 0.555 points on average after inclusion of EIN into the dairy item key while decreasing by 0.029 points on average after inclusion of non-EIN characteristics into dairy item keys. For vegetables, root mean squared error decreases on average by 0.126 points after EIN inclusion into item keys while decreasing by 0.047 points on average after the item key inclusion of a non-EIN characteristic. For dairy, the lowest level of error is found using the most detailed item key in the most general index calculation method; for vegetables, the lowest level of error is found using the most detailed key with the most constrained index calculation method. Both findings corroborate those based on the correlation coefficient analyses.

Though the unit value dairy index tracks the benchmark index better than that of the unit value vegetable index, the second has smaller errors indicating greater accuracy. Both correlation coefficient and error analysis point to similar methodologies to optimize accuracy and mirroring of benchmarks, such as EIN item key inclusion for both indexes and stronger treatment of outliers for the vegetable index.

AN INITIAL PROTOTYPE FOR UNIT VALUES and UNIT VALUE INDEXES

Coefficient of variation, correlation coefficient, and error analysis yield a prototype for unit value specification and unit value index calculation. Regarding the best specification for the ELI, the most prominent result is the importance of company identifier in the item key. Furthermore, the most detailed item key shows the most intra-item substitutability at the item level and solid results compared to benchmark indexes. Results were robust across correlation coefficient, root mean squared error, and mean absolute error analyses.

Regarding the index calculation methods, results are not as clear-cut. Given initial selection of product categories without reliable benchmark indexes, it comes as no surprise that index calculation methods do not produce consistent results when unit value indexes are compared to the benchmarks. Whereas the least constrained index method calculation - retaining outliers and not truncating ELIs that are inconsistently traded – provides a best fit for dairy, vegetables require a more rigorous treatment of outliers and consistency in trade. There exists a possibility that the different methods belie divergent market forces rather than poor benchmark selection. In particular, price and quantity changes are more variable with seasonal items like vegetables, making price outliers less informative of general price trends.

To proceed with a prototype index calculation method, we make a couple of strong assumptions to test other BEA 5-digit export indexes composed of homogeneous products that also have published XPI benchmarks. First, we determine that the three-month imputation rule sufficiently addresses any inconsistencies in trade, and thus, do not impose limits on ELIs that are inconsistently traded. Second, though dairy unit value indexes are most accurate without elimination of outliers, we determine that it is prudent to treat price outliers, as they are likely due to compositional abnormalities or incorrect transaction records. Thus, we apply the Tornqvist index with no more than three months imputation for missing prices and additionally replace outlier prices outside the third standard deviation with imputed values.

We apply the prototype ELI—the most detailed item key—to evaluate homogeneity of all 5-digit BEA End Use categories, based on the homogeneity of their ELIs. We then calculate select unit value indexes with the prototype calculation method and compare then with published XPI benchmarks. Homogeneity is evaluated as the level of intra-item substitutability, where less price-dispersion indicates more homogeneity. Price dispersion is calculated through the coefficient of variation test. Using the coefficient of variation for prototype vegetable unit values as an upper bound to limit the presence of non-impactful outliers, we identify 50 export and 52 import 5-digit BEA End Use unit value indexes as homogeneous by ranking categories with lower coefficients of variation than the vegetables estimate. Calculating three 5-digit BEA end use export indexes based on the prototype and evaluating results against published XPIs with extensive price quotes – meat, soybeans, and animal feed – soybeans and animal feeds show a high degree of accuracy with correlation coefficients, and meat and animal feeds strongly track published XPI benchmark indexes.

Table B. Unit Value Index Comparison to Published Export Price Indexes, 2016

BEA End Use Export Classification	Correlation Coefficient	RMSE	MAE
Meat, poultry, and other edible animal products	0.1657	1.677	1.128
Soybeans and soybean by-products	0.9116	2.927	2.349
Animal feeds	0.9519	0.918	0.744

CONCLUSION

This research shows promise that unit value indexes may be blended with directly collected survey data to calculate MXPI. Addressing unit value bias is essential to this approach. We determine the prevalence of unit value bias by assuming that items of similar price levels may rank more similarly in consumer preferences by minimizing income effects, and thus, we assume that less price dispersion within an ELI defines a more homogeneous item. We incorporate current quantities using a Tornqvist index to address substitution bias. The three tests we conduct to determine unit value index accuracy and tracking of benchmarks with 36 variations of item key and index calculation method show that EIN and other non-price characteristics more precisely define a homogeneous good. The most detailed item key shows the least price dispersion, most accuracy, and best tracking of benchmarks.

As a result of this research, the unit value indexes offer a promising supplement to current price index methodologies. Future research will assess unit value indexes from 2012 to 2017 for all 50 export and 52 import homogeneous 5-digit BEA End Use categories to validate a prototype for ELI specification and index calculation that consistently provides strong results. As part of this research, options for systematically identifying over-fitted and under-fitted indexes will be explored. Indexes' impact on net trade and GDP as well as on top-level price indexes also will be evaluated. The path to be taken is not yet clearly defined, and the road to implementation is not expected to be easy, but the first steps are solid.

Bibliography

Alterman, William. 2015. "Producing an Input Price Index." In Measuring Globalization: Better Trade Statistics for Better Policy - Volume 1. Biases to Price, Output, and Productivity Statistics from Trade, Susan N. Houseman and Michael Mandel, eds. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, pp. 331-358. https://doi.org/10.17848/9780880994903.vol1ch10.

Alterman, William. 1991. "Price Trends in U.S. Trade, New Data, New Insights." Chapter in NBER book International Economic Transactions: Issues in Measurement and Empirical. Edited by Peter Hooper and J. David Richardson. National Bureau of Economic Research, Studies in Income and Wealth. Volume 55. University of Chicago Press 109-143 (January).

Amiti, Mary and David E. Weinstein. 2009. Exports and Financial Shocks. Working Paper 15556. National Bureau Of Economic Research. Cambridge, MA. December. URL: http://www.nber.org/papers/w15556

Bridgman, Benjamin. 2015. "Specific Trade Costs, Quality, and Import Prices." In Measuring Globalization: Better Trade Statistics for Better Policy – Volume 1. Biases to Price, Output, and Productivity Statistics from Trade, Susan N. Houseman and Michael Mandel, eds. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, pp. 121-148. https://www.jstor.org/stable/j.ctvh4zfcn.7?seq=1#metadata_info_tab_contents

Broda, Christian and Weinstein, David. 2006. "Globalization and the Gains from Variety," Quarterly Journal of Economics, May. 121(2).

Clausing, Kimberly A. 2003. "Tax-motivated transfer pricing and US intrafirm trade prices." Journal of Public Economics. Volume 87, Issues 9-10, 2207-2223 (September).

Diewert, W.E. 1976. “Exact and Superlative Index Numbers. Journal of Econometrics” Volume 4, Issue 2. North-Holland Publishing Company (May).
Diewert and Feenstra 1997
Feenstra, Robert C. 1994. “New Product Varieties and the Measurement of International Prices.” American Economic Review, March. 84(1), pp. 157-77.
Feenstra, R.C., & Heston, A., Timmer, P.M & Deng, H., (2009). Estimating real production and expenditures across nations: A proposal for improving the Penn World tables," The Review of Economics and Statistics, MIT Press, vol. 91(1): 201-212. http://dx.doi.org/10.1162/rest.91.1.201
Feenstra, R., and Romalis, J., (2014). International prices and endogenous quality. Quarterly Journal of Economics, 129 (2): 477–527. http://dx.doi.org/10.1093/qje/qju001
Fitzgerald, Jenny. 2017. Assessing Product Downsizing, Upsizing, and Product Churn using Nielsen Scanner Data, Presented at the 2017 Joint Statistical Meetings, Session on Measuring the Economy: Economic and Workforce Statistics, Government Statistics Section.
Gopinath, Gita. 2010. “Currency Choice and Exchange Rate Pass-through,” American Economic Review 100 (1): 304-336.
Gopinath, G., & Itskhoki, O. (2010). FREQUENCY OF PRICE ADJUSTMENT AND PASS-THROUGH. The Quarterly Journal of Economics, 125(2), 675-727. Stable URL: http://www.jstor.org/stable/27867493
Gita Gopinath and Roberto Rigobon. Sticky Borders. The Quarterly Journal of Economics, Vol. 123, No. 2 (May, 2008), pp. 531-575. Stable URL: https://www.jstor.org/stable/25098909
Government Price Statistics Hearings Before the Subcommittee on Economic Statistics Of the Joint Economic Committee Congress of the United States Eighty-Seventh Congress First Session Pursuant to Sec. 5(a) of Public Law 304 (79th Congress) Part 1 January 24, 1961
Hottman, Colin J. & Monarch, Ryan. 2018. Estimating Unequal Gains across US Consumers with Supplier Trade Data. Federal Reserve Issue Paper 1220. January 17, 2018. (URL: https://www.federalreserve.gov/econres/ifdp/files/ifdp1220.pdf)
Houseman, Susan, Christopher Kurz, Paul Lengermann, and Benjamin Mandel 2011. “Offshoring Bias in U.S. Manufacturing.” Journal of Economic Perspectives. Volume 25, Number 2, 111-132 (Spring).
Kamal, Fariha & Monarch, Ryan. 2017. Identifying Foreign Suppliers in U.S. Import Data. International Finance Discussion Paper. 2017. 1-36. (URL https://doi.org/10.17016/IFDP.2017.1142r)
Kravis, Irving B., Robert E. Lipsey 1971. “Wholesale Prices and Unit Values as Measures of International Price Competiveness” Chapter in Price Competitiveness in World Trade. National Bureau of Economic Research. Columbia University Press. New York.
Lipsey, Robert E. 1963. Price and Quantity Trends in the Foreign Trade of the United States. National Bureau of Economic Research. Princeton University Press. URL: https://www.nber.org/books/lips63-1
Mead, David. 2014. “Analyzing alternatives to export price indexes,” Beyond the Numbers: Global Economy, vol. 3, no. 27. U.S. Bureau of Labor Statistics, November, https://www.bls.gov/opub/btn/volume-3/analyzing-alternatives-to-export-price-indexes.htm
Moulton, Brent R. 2018. “The Measurement of Output, Prices, and Productivity: What’s Changed Since the Boskin Commission?” Hutchins Center on Fiscal & Monetary Policy at Brookings, The Brookings Institution (July). https://www.brookings.edu/research/the-measurement-of-output-prices-and-productivity.
Nakamura, Emi and Jón Steinsson. 2012. Lost in Transit: Product Replacement Bias and Pricing to Market. The American Economic Review, Vol. 102, No. 7 (DECEMBER 2012), pp. 3277-3316.
Price Statistics Review Committee. 1961. The Price Statistics of the Federal Government. National Bureau of Economic Research. Cambridge, MA. Appendix A. URL: http:/www.nber.org/chapters/c648979.

Silver, Mick. 2010. “The Wrongs and Rights of Unit Value Indices.” Review of Income and Wealth. Series 56, Special Issue 1, 206-223 (June).

Triplett, Jack E. 1992. “Economic Theory and BEA’s Alternative Quantity and Price Indexes.” Survey of Current Business. 49-52 (April). https://www.bea.gov/scb/pdf/national/nipa/1992/0492trip.pdf.

---

[i] Import crude petroleum prices are derived from the administrative records of crude petroleum imports collected by the U.S. Department of Energy. Detailed product categories are grouped by product and transaction characteristics (i.e. gravity, crude stream, and country of origin) and average weighted prices are incorporated into the price index.

[ii] The administrative trade data are collected through an electronic interface that exporters and importers use to directly enter data on trade transactions. The U.S. Census Bureau collects and cleans the export data to calculate official international trade measures, after which the data are transferred to the BLS.

[iii] For a given shipment, each company must submit an individual record for each product as defined by the 10-digit harmonized schedule classification (Schedule B for exports, and HTSUSA for imports). Thus, each record pertains to only one Employer Identification Numbers (EIN) and one shipment. The record includes total dollar value, quantity, company, transportation, and geographic information on provenance and destination of goods and shipper.

[iv] Related trade is an intra-firm transaction that takes place between a parent and an affiliate.

[v] BLS research has previously proposed using the Tornqvist index to blend secondary data sources with the matched model where current period weights are available (Fitzgerald 2017).

[vi] Missing item price values are imputed by applying the percent change of the item’s parent 10-digit stratum to the item’s price in the previous month. However, the actual month-to-month price percent change for an item may not be the same as the month-to-month price percent change for its parent classification level, which is an estimation error associated with imputation.