10

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

Don A. Fast and Susan E. Fleck

10.1 Introduction

The US Bureau of Labor Statistics (BLS) Import and Export Price Indexes (MXPI) track price changes in internationally traded merchandise goods. The indexes underpin inflation adjustment of US 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, 20,000 to 25,000 prices of unique items from thousands of companies have been collected monthly to calculate detailed and all-goods price indexes. Trade has grown and sample size has been constant

Don A. Fast is a senior economist at the US Bureau of Labor Statistics (BLS).

Susan E. Fleck is the assistant commissioner of the International Price Program at 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, Alberto Cavallo, John Haltiwanger, Marshall Reinsdorf, and our discussants, Carol Corrado and Susan Houseman, on earlier versions of this research. Use of the export trade data is subject to Agreement No. 2067-2018-001, Memorandum of Understanding (MOU) between the US Census Bureau and BLS. The BLS has received prior approval from the US Census Bureau, which affirms that the research results do not present disclosure risks and approves the publication of the results. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see https://www.nber.org/books-and-chapters/big-data-21st-century-economic-statistics/unit-values-import-and-export-price-indexes-proof-concept.

and—more recently—reduced. Both trends result in thinner item coverage, directly reducing the number of detailed indexes of publishable quality. 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 half of the most detailed Bureau of Economic Analysis (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—that is, 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 and which 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 US 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 foodstuffs, 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, which BLS also deemed as having substantial unit-value bias due to lack of detail and the inclusion of heterogeneous products (Alterman 1991).

Since that time, some experts have proposed that unit values for homogeneous goods may track prices (Mead 2014; Silver 2010). More than twenty years ago, Feenstra and Diewert (1997) proposed that BLS analyze the detailed administrative trade data that are the subject of this chapter, given the improvements in coverage, detail, and availability at that time. However, BLS had less capacity than today to address the complexity of the data and the lag in its receipt, and so did not pursue the project. More recently, Nakamura et al. (2015) set out both historic precedence and mathematical formulas to incorporate unit values into official price indexes as a viable alternative to address substitution and other biases.

The proof that unit values could be used in price indexes is in the doing, and BLS has begun research on exports to evaluate the aforementioned administrative trade transactions. The administrative trade data are reported by type of export product per exporter per vessel per day, based on the detailed Harmonized System (HS) classification with more than 5,000 merchandise good categories. The transaction records include dozens of data fields. The data provide the opportunity to evaluate whether and which grouped transactions with a range of price differences are homogeneous, essentially addressing Nakamura et al.'s "impediment 2" to the adoption of unit values—"the question of if and when auxiliary product unit attributes should be used in forming index basket product definitions" (Nakamura et al. 2015, 54).

The basic questions are (1) whether the data source can be used to calculate unit values and (2) how to select and group the attributes of these transactions into homogeneous products. The first question is more easily answered than the second. The approach we use allows for multiple transactions per product at multiple prices to calculate a unit value with current prices and quantities per time period. The second question is how to differentiate heterogeneous from homogeneous product categories—and thus unit values—with the attributes in the trade data in addition to the detailed HS product category (called here 10-digit HS). Many researchers use the trade data to calculate their own price or price index comparisons. For example, unit values are calculated for cross-country comparisons, using 10-digit HS product categories (Feenstra et al. 2009; Feenstra and Romalis 2014). Impacts of import prices on welfare in the United States group the 10-digit HS with one or two data characteristics to calculate more detailed unit values. For example, Broda and Weinstein (2006) estimate the impact of product variety changes on prices and welfares by including country of origin in their import indexes. Hottman and Monarch (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. Kamal and Monarch (2017) analyze the reliability of the trade data in the context of US–foreign supplier relations. These one-time research projects show the potential to calculate unit values and to group transactions into products. But we know of no work that evaluates the reliability of, bias in, or homogeneity of unit values calculated from the trade data. To consider the trade data as a source in official statistics, these topics must be addressed.

There is limited precedent using unit values as prices in the import price index in the international price program. A crude petroleum import price index is currently calculated using unit values derived from the US Department of Energy (DOE) petroleum transaction import records.[1] The DOE administrative data source is more reliable than survey data in the face of low company response rates and the price volatility of this heavily traded product. Furthermore, crude petroleum import records provide fairly detailed product information. In contrast, the administrative trade transaction records do not have consistently similar product and transaction information across the thousands of categories, in part because of the regulatory nature of trade. Many of the 10-digit HS product categories are composed of differentiated goods, which means that unit values grouped only by HS product are likely to be heterogeneous and not track product price trends. In the face of the uneven detail of administrative trade data, is it possible to move beyond a "special case" use of unit values, such as in crude petroleum, to a more comprehensive approach?

Key to the decision of whether and how to use unit values from the administrative trade data is having sound criteria for deciding when and how they can be applied. BLS requires a consistent and transparent approach to evaluate (1) whether a product category is homogeneous and, relatedly, (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—is not insignificant but will not be addressed here.

In this paper, we outline both concepts and methods for using administrative trade data to produce unit values and unit value indexes. Using 2015–2016 export transaction records for dairy and vegetables, we test six different ways to group characteristics in the administrative records into entry-level items (ELIs). Entry-level items are the products in the index basket for which prices are tracked across time periods, and which form the base unit of price change for price indexes. Unit values for these 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 the groupings—or item keys—that result in the least bias. ELI prices then are aggregated using a Tornqvist index formula to produce the 10-digit HS price indexes that are

1. Import crude petroleum prices are derived from the administrative records of crude petroleum imports collected by the US 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.

the building blocks for the official product price indexes (Harmonized and BEA End Use) and industry price indexes for imports and exports.

For this research, applying a modified Laspeyres index formula, we use the 10-digit HS unit value price indexes to form 5-digit BEA End Use indexes, and then compare those indexes to existing BLS official price indexes as benchmarks for quality. A natural question is how our indexes compare to BLS's published BEA End Use export price indexes. Those published price indexes are used to deflate imports and exports in GDP, meaning that differences in index values would result in revisions to GDP if the unit value indexes were adopted. The comparative analysis of the unit value indexes and the benchmark indexes leads us to propose a prototype unit value index approach. The promising first results we obtain provide a road map for comprehensively evaluating all import and export price indexes for homogeneous categories.

10.2 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 categories are homogeneous and whether a more detailed grouping of attributes is necessary to mitigate compositional effects of shipping contents on the resulting unit value. The simplest case is one in which all or some 10-digit HS unit values provide as good a measure of price change as the published import and export price indexes.

Two principles guide the methodological approaches in this research—to evaluate item homogeneity, and to improve the index where possible. The research develops and evaluates new methods to identify homogeneous products and 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).[2] We selected these two product categories for two reasons—because the 10-digit HS product groups that comprise each BEA End Use product area appear relatively homogeneous and because these indexes historically had been of uneven quality. The issues generally have stemmed from an insufficient number of representative businesses voluntarily participating in the survey, resulting in an insufficient number of prices, incomplete representation of sampled products, or inadvertent exclusion of large traders. Precisely because of the quality issues, the official XPI for these product

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

categories may be an imperfect benchmark to validate the consistency and quality of the pilot index measures.

10.2.1 Defining Homogeneity

Moving from a matched model price to homogeneous product unit values requires consistency of definition of product attributes, sufficient transactions to group by similar product attributes, and persistence over time of transactions with those same attributes.

Before using a homogeneous unit value in a price index, it is necessary to define what a homogeneous product is. Nakamura et al. (2015) consider primary attributes of products as the only necessary characteristics to define a unit value. However, in the administrative trade data, many 10-digit HS product categories include a mix of different products. Given that 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), we expect that the non-price characteristics in the administrative records can provide additional information to define products. That is, similarity of the transaction characteristics that define a sale are expected to signal similarity of products and purchasers.

Transactions should be grouped to minimize differences in product attributes and also maximize substitutability among the products in the included set. Price-setting research tells us that the prices of homogeneous products vary over time. 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. They attribute these differences to larger elasticities of demand by consumers contributing to greater costs of price stickiness for producers. Thus, in the case of homogeneous goods, unit values allow for substitutability among similar products with different prices. As Nakamura et al. (2015) propose, such unit values may more accurately represent import and export prices than a single price observation for the product from one sampled establishment. Additionally, the unit value indexes calculated from the unit values are expected not to demonstrate the "product replacement bias" of matched models delineated in Nakamura and Steinsson (2012), where frequent product turnover results in no price changes across months for 40 percent of imported items.

What are the shared attributes that help define homogeneity? 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. Furthermore, Clausing (2003) describes how intra-firm trade and country impact price setting. This research leads us to suspect that 10-digit HS product categories on their own are likely to be too broad for unit value indexes to demonstrate the characteristics of homogeneous products. To group transactions with a greater level of specificity than the 10-digit HS product categories, we take into account price and nonprice trade characteristics that separate goods into unique bins or groups of substitutable products. 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 what we call intra-item substitutability by grouping transactions by as many attributes that define the purchaser-seller relationship while assuring persistence over time of transactions with those same attributes. To objectively evaluate the different groupings of products and their price dispersion, we use the coefficient of variation (described below) to compare the different product groupings.

10.2.2 Better 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). Nakamura et al. (2015) and Bridgman (2015) also describe sourcing substitution and trade cost biases, especially for import price indexes, arguing that official price indexes are upwardly biased.

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. More specifically, 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 expand 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. Furthermore, the use of multiple transactions at multiple prices addresses the criticism of Nakamura et al. (2015) that single items may not be representative of a product when multiple prices are present in a population.

The prices and indexes calculated and presented here are based on the two 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 homogeneous unit values and items, (2) test item homogeneity, (3) identify appropriate BLS price indexes as 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.

10.3 Unit Values and Unit Value Bias

10.3.1 Defining Unit Values

The point of departure for the research is to establish the 10-digit HS product category as the starting point for evaluating unit values. This level of detail is naturally occurring in the administrative trade data, as records are HS-specific.[3] Given the fact that the 10-digit HS are also the strata 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. Our research tests the premise that the 10-digit HS product categories are homogeneous, and products grouped with more attributes are more homogeneous, thus establishing a range of homogeneity from fewer products with fewer attributes to more products with more attributes. Unit values are then calculated for this range of products within each 10-digit HS product, in which each entry-level item is actually a product group, and each entry-level item price is a unit value.

Whereas the simplest case occurs when the item key—the list of price-determining characteristics that defines the item—contains only the 10-digit HS code (H), other item keys include additional attributes that are similar to price-determining characteristics in the international price survey. The attributes used in the item keys are: 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 arms-length trade,[4] and unit of measure. The data fields for HS, EIN, and zip code correspond with the sampling unit (multistage sampling for the directly collected international price survey allocates price quotes across establishments at the 10-digit HS product category level). The data fields for state of origin, port of export, country of destination, and related or arms-length transaction correspond to production and/or market relations between exporter and foreign consumer. Most of these descriptors also are collected in the survey as price-determining characteristics. For measurement consistency, the unit of measure (e.g., gross, piece, ton) also is included. Each item key specification results in a different set of unique items, or ELIs, with the same attributes grouped by the same shared characteristics.

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

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

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

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

$$(1) \qquad p_{K_i}^{(j,t),H} = \frac{\sum_{z\in i} p_{K_{i,z}}^{(j,t),H}}{z},$$

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 weighted geometric mean of unit value transaction prices, which yields

$$(2) \qquad p_{(j,t)}^{H} = \exp\left(\frac{\sum_{i\in j}[w_{K_i}^{(j,t),H}\ln(p_{K_i}^{(j,t),H})]}{\sum_{i\in K} w_{K_i}^{(j,t),H}}\right),$$

where normalized transaction-level weights are represented as

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

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

$$(3) \qquad q_{(j,t)}^{H} = \sum_{i\in K} q_{K_i}^{(j,t),H}.$$

Taking an experimental approach to test different specifications of items supports the objective to identify the best unit value measure. For the unit value tests, we use the price changes of actual transactions based on attributes for six item key specifications.

10.3.2 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 multiple transactions of one product, 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 that 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 types of test 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 two price clustering tests—the first shows the most promise.

The price dispersion test was conducted on the actual unit values for dairy and vegetables transactions. The coefficient of variation (CV) is the ratio of the weighted standard deviation of prices within a month to the weighted mean; lower percentages indicate less variability in the ELI. Even though findings from the trade literature report price variability in homogeneous products, we assume there is a degree of within-month price variability for an item beyond which an item is not homogeneous. The CV test allows us to identify a frontier of price variability beyond which a group of transactions comprising an item should not be considered homogeneous. This test fits with findings from the trade literature 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. The bins in figure 10.1 specify ranges of CVs. The least detailed item keys that exclude the company identifier (EIN, or "E" in the legend) 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 CV of less than 52.5 percent for the two item keys that exclude EIN. When the company identifier is added to the

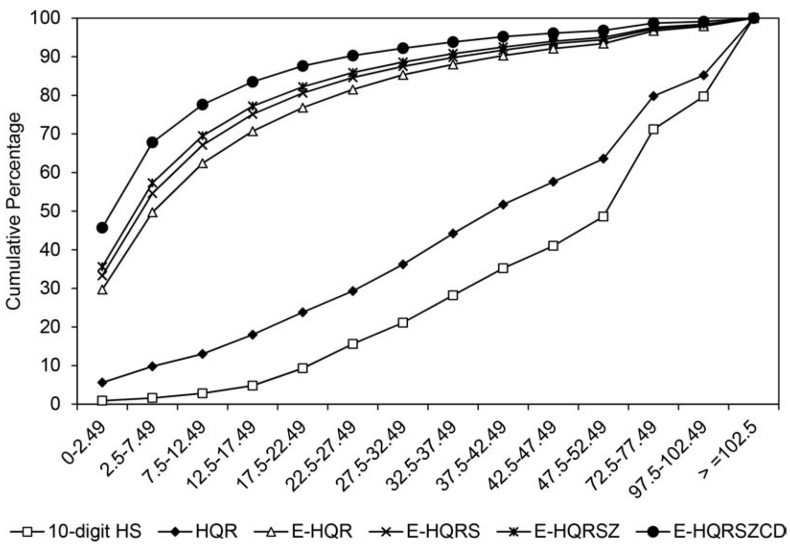

Fig. 10.1 Coefficient of variation test, dairy products and eggs, 2015–2016

Note: Letters correspond to these nonprice 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).

ELI specification, prices cluster closer to the mean—60 percent of the ELIs that include the company identifier had a CV less than 12.5 percent. Furthermore, the most detailed item key, which 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 for the item keys that include that characteristic.

Another test of homogeneity looks at the month-over-month percentage change in price. 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 attributes 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 unit price should reflect intra-item substitutability. The 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. When using SAS Clustering Method 1, results are sensitive to price cluster distance. 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 simplest case of 10-digit HS unit values.

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

10.4 Benchmarking Unit Value Indexes with BLS Price Indexes

Having selected ELIs that have intra-item substitutability and established an index methodology, we consider the options for calculating the least biased unit value indexes and then compare the resulting indexes to existing BLS price indexes. As set out in the introduction, we compare the unit value indexes for 5-digit BEA End Use categories to appropriate price index benchmarks in order to evaluate the potential impact of their adoption on GDP revisions. The data we analyze are voluminous and many choices must be made in producing the unit value indexes. We apply different assumptions for index calculation, imputations, and outliers to produce a wide range of results, then compare the resulting unit value indexes for dairy and vegetables with official benchmarks. The most obvious benchmarks for the unit value indexes would be the official export price indexes based on the BEA End Use classification, but we have selected two product areas whose official export price indexes are not of the highest quality. For this reason, we consider other benchmarks.

10.4.1 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 HS stratum is the entry point for blending data. For official price indexes, company weights are used to aggregate ELI price changes to the 10-digit HS product category, and then trade dollar weights for 10-digit HS categories, lagged two years, 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 in the administrative trade data, the unit value ELIs can be aggregated into their corresponding 10-digit strata. The 10-digit HS unit value Tornqvist indexes then are aggregated into the BEA 5-digit index using official estimation procedures. The Tornqvist index is superior to a Laspeyres index because it accounts for the introduction of new goods, disappearing goods, and changes in trade volumes (Diewert 1976; Triplett 1992). The baseline case is to use the 10-digit HS stratum 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:

$$(4) \qquad R_{H,t} = \prod_{j \in H} \left(\frac{p^H_{(j,t)}}{p^H_{(j,t-1)}} \right)^{(W^H_{(j,t-1)} + W^H_{(j,t)})/2},$$

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

These calculations differ from existing methodology, not only because we are using unit values, but also in 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.[5]

Missing prices, consistency of trade and outliers. In order to evaluate the unit value indexes, methods must be adopted to address the problems of missing prices, inconsistent trading, and outlier observations.

Index calculation requires two months of actual prices to establish an item in the index. Once an item is established, imputation fills in the gaps when the item is not traded or its price is of questionable quality.[6] 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 many missing prices. Missing prices become even more prevalent as attributes are added to the item key, because each ELI has fewer transactions and experiences more turnover. Imputation is used to maintain items in the index, but there is a point at which imputation negatively impacts index quality. To minimize the negative impact that continuing imputed prices over time has on the indexes for the 10-digit HS strata, imputation is suspended for items that have no transaction recorded after three months. Beyond that point, the price imputations overwhelmed the count of unit values calculated directly from transaction records by more than two to one.

Establishments with inconsistent trade are excluded from the sample for the official MXPI to focus on respondents that can provide monthly prices. Inconsistent trade manifests itself in the administrative trade data in the form of a trade-off at the item level between defining the item more precisely and experiencing more missing prices. The decision whether to include inconsistently traded items in the 10-digit HS unit value indexes has implications for index quality. Including inconsistently traded items increases the use of imputation but excluding items that are not consistently traded could bias unit values by not accounting for new goods. Thus, two variations are tested for the unit value calculations—retain all items regardless of consistency of trade and exclude items that are traded less than half the year. Both approaches preserve the three-month imputation rule set above.

The decision whether 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 monthly price change, but an outlier in the unit value of the transaction cannot be evaluated in the same way. It may represent an error, or a different product being traded. Three unit value index

5. 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).

6. 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.

calculations are considered—retain the outlier; recalculate the unit value with an imputed price when the price change falls outside the two-standard-deviation band; or recalculate the unit value with an imputed price when the price change falls outside the three-standard-deviation band.

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 10.1 shows the index calculation methods from the least constrained to most constrained options regarding truncation of ELIs, and the statistical comparison of these alternative indexes against BLS price indexes. 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 exclude items that are not consistently traded.

10.4.2 Benchmark Comparisons

The comparison of the unit value indexes against BLS official price indexes as benchmarks 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 baseline 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.

Because the two product groups were chosen due to quality concerns, the XPI for dairy and vegetables for this time period were respectively unpublished and had low coverage. Thus, the best benchmark against which to measure the unit value indexes was not necessarily the XPI. Export Price Indexes, spot prices, the relevant Consumer Price Indexes for all urban consumers, and the relevant Producer Price Indexes (PPI) were considered as possible benchmarks for unit value indexes. The unpublished XPI was chosen as a benchmark for dairy—even though the index was unpublished due to insufficient company representativeness, there were a sufficient number of prices in the index. Although consumer prices are systematically different from export prices, meaning that the CPI is generally not the best comparative benchmark, it was chosen as the benchmark for vegetables due to seasonal weighting concerns with the official vegetable XPI.

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 the unit value index calculations for each of the six selected item keys. The benefits of unit value indexes are realized with more detailed item key specifications than the 10-digit HS level, but there is a possibility that item key specifications with too much detail may be "overfitted"—understating intra-item substitution and missing price changes of high-volume or price-variable products. Additionally, truncating outliers may introduce bias if outliers represent real price shocks.

Generally, correlation coefficients for dairy unit value indexes are higher than correlation coefficients for vegetable unit value indexes—that is, dairy unit value indexes do a better job of tracking the price trends in the benchmark index. For dairy, correlation coefficients remain consistent across different treatments of outliers and trade consistency. Correlation coefficients vary more for vegetables, pointing to a less consistent time series. Dairy correlation coefficients significantly improve after including company identifier in the item keys, with correlation coefficients being on average 0.090 higher than correlation coefficients of indexes excluding the company identifier, or EIN. Adding other attributes to define products resulted in correlation coefficients that were 0.002 lower on average. The large increase in dairy correlation coefficients in item keys that include the EIN implies that product differentiation may occur at the firm level for items in the dairy category. This pattern, however, is not reflected for vegetables. Comparing vegetable products with item keys that include and exclude the EIN, the correlation coefficients are on average 0.012 *lower* than correlation coefficients excluding the EIN. This statistic is of a smaller magnitude than the average 0.020 correlation coefficient increase with the addition of non-EIN attributes in vegetable item keys.

Our assessment of the impact of index calculation methods on the correlation coefficient is less informative. Dairy unit value indexes mirror the unpublished XPI benchmark, no matter the index calculation method, when the EIN attribute becomes part of the item key. The vegetable unit value indexes do not track the CPI benchmark to any large degree.

Root mean squared error/mean absolute error comparison. Root mean squared error and mean absolute error measure differences between calculated and benchmark price indexes. We interpret these measures as an indication of 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 10.1, across index calculation variations the dairy unit value indexes display larger error than the vegetable unit value indexes compared to their respective benchmarks. For both indexes, error measures trend downward as item keys become more detailed, implying that accuracy increases when more attributes are used to create items, regardless of index calculation methods.

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

Table 10.1 **Unit value index comparison to BLS price indexes, dairy and vegetables, 2015–2016**

	Exclude company identifier		Include company identifier (EIN)			
	10-digit HS	+ transfer price + unit of 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/3-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/3-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/3-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/3-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/3-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/3-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/3-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/3-month imputation + consistent trade	2.50 / 2.04	2.07 / 1.57	1.92 / 1.53	1.84 / 1.40	1.79 / 1.4	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

for vegetables. Mirroring the previous correlation coefficient analysis, root mean squared error decreases by 0.555 points on average after inclusion of EIN into the dairy item key, compared to a decrease of 0.029 points on average for inclusion of a non-EIN attribute. For vegetables, root mean squared error decreases on average by 0.126 points after EIN inclusion into item keys, compared to a decrease of 0.047 points on average for inclusion of a non-EIN characteristic. For dairy, the lowest level of error is found using the most detailed item key with the least restrictive 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 the unit value vegetable index tracks its benchmark, the vegetable index comparison has smaller errors, indicating greater accuracy. Both correlation coefficient and error analysis point to similar methodologies to optimize accuracy and mirroring of benchmarks; most especially, for both indexes, the inclusion of EIN in the item key but also the stronger treatment of outliers for the vegetable index.

10.5 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. The coefficient of variation results show the product prices based on the most detailed item key are the least variable in price and the most homogeneous. Results including the EIN but not necessarily other attributes were robust across the correlation coefficient, root mean squared error, and mean absolute error analyses.

Regarding the index calculation methods, results are not as clear cut. Because neither of the benchmark indexes was a published export price index, it is possible that results are not consistent 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. It is possible that the differing success of particular methods reflects differing market forces for the two cases. In particular, price and quantity changes are more variable with seasonal items like vegetables, making price outliers less informative of general price trends (see table 10.2).

To proceed with a prototype index calculation method, we make two strong assumptions in order to test other BEA 5-digit export indexes composed of homogeneous products that also have published XPI benchmarks. First, we assume that the three-month imputation rule sufficiently addresses

Table 10.2 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 byproducts	0.9116	2.927	2.349
Animal feeds	0.9519	0.918	0.744

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 proceed on the basis that it is prudent to treat price outliers, assuming that they likely are due to differences in product mix in the shipment or incorrect transaction records. Thus, we apply the Tornqvist index to a dataset with no more than three months' imputation for missing prices and additionally replace outlier prices outside the third standard deviation band with imputed values.

We apply the prototype ELI—the most detailed item key—to evaluate homogeneity of all 5-digit BEA End Use export product 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. To limit the presence of problematic outliers, we use the coefficient of variation for prototype vegetable unit values as an upper bound on the coefficient of variation for a homogeneous category. Using this criterion, we identify 50 export and 52 import 5-digit BEA End Use unit value indexes as homogeneous. We calculate three 5-digit BEA end use export indexes—meat, soybeans, and animal feed—based on the prototype and evaluate the results against published XPIs with extensive price quotes. The indexes for soybeans and animal feed show a high degree of accuracy when assessed using correlation coefficients, and the indexes for meat and animal feed closely track published XPI benchmark indexes.

10.6 Conclusion

Our findings hold the promise that it may be possible to blend unit value indexes with directly collected survey data to calculate MXPI. Defining homogeneity and addressing unit value bias are essential to this approach. We establish that the best approach to defining homogeneous items involves adding attributes to the 10-digit HS product grouping to create more detailed items and limiting the price dispersion allowable for an item to be considered homogeneous. We identified an inverse relationship between the number of attributes used to define an item and the price variability among the transactions that comprise the item's unit value. While having more attributes and less price variability means that items are more homogeneous, it also means that there is a greater risk of the items not being traded consistently, as the number of transactions that comprise that item's unit value for a month is lower and the prevalence of missing prices across months is greater.

Establishing an index methodology that works with unit values also is essential to blending unit value indexes into the MXPI. The availability of prices and quantities allowed us to use a Tornqvist index to address substitution bias. We established imputation to account for missing prices and addressed outliers. These new methods were tested by comparing the unit value indexes against benchmark price indexes to evaluate their similarities and differences. The three tests we conducted 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 nonprice characteristics more precisely define a homogeneous good. The most detailed item key shows the least price dispersion, most accuracy, and best benchmark tracking. There was no clear result for which index formula provided the most comparable index, but the groundwork has been laid for the next round of comparisons.

Future research will assess unit value indexes from 2012 to 2017 for all 50 export and 52 import 5-digit BEA End Use categories that have sufficiently low within-category price dispersion as to be considered homogeneous. The results will be used to validate a prototype for ELI specification and index calculation that consistently provides strong results. As part of this research, options for systematically identifying overfitted and underfitted indexes will be explored. Indexes' impact on net trade and GDP, as well as on top-level price indexes, also will be evaluated. Much work remains to be done, but we are encouraged by the results obtained thus far.

References

Alterman, W. 1991. "Price Trends in U.S. Trade: New Data, New Insights." In *International Economic Transactions: Issues in Measurement and Empirical Research*, Studies in Income and Wealth, vol. 55, edited by P. Hooper and J. D. Richardson, 109–43. Chicago: University of Chicago Press.

Amiti, M., and D. E. Weinstein. 2009. "Exports and Financial Shocks." NBER Working Paper No. 15556, National Bureau of Economic Research, Cambridge, MA. http://www.nber.org/papers/w15556.

Bridgman, B. 2015. "Specific Trade Costs, Quality, and Import Prices." In *Measuring Globalization: Better Trade Statistics for Better Policy*, vol. 1, *Biases to Price, Output, and Productivity Statistics from Trade*, edited by S. N. Houseman and M. J. Mandel, 121–48. https://www.jstor.org/stable/j.ctvh4zfcn.7?seq=1#metadata_info_tab_contents.

Broda, C., and D. Weinstein. 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics* 121 (2). https://doi.org/10.1162/qjec.2006.121.2.541.
Clausing, K. A. 2003. "Tax-Motivated Transfer Pricing and US Intrafirm Trade Prices." *Journal of Public Economics* 87 (9–10): 2207–23. https://doi.org/10.1016/S0047-2727(02)00015-4.
Diewert, W. E. 1976. "Exact and Superlative Index Numbers." *Journal of Econometrics* 4 (2): 115–45.
Feenstra, R. C., and W. E. Diewert. 1997. "Contract Report on the Methodology for the U.S. Export and Import Price Indexes." Unpublished paper, US Bureau of Labor Statistics, Washington, DC.
Feenstra, R. C., A. Heston, P. M. Timmer, and H. Deng. 2009. "Estimating Real Production and Expenditures across Nations: A Proposal for Improving the Penn World Tables." *Review of Economics and Statistics* 91 (1): 201–12. http://dx.doi.org/10.1162/rest.91.1.201.
Feenstra, R. C., and J. Romalis. 2014. "International Prices and Endogenous Quality." *Quarterly Journal of Economics* 129 (2): 477–527. https://doi.org/10.1093/qje/qju001.
Feldstein, M. 2017. "Underestimating the Real Growth of GDP, Personal Income, and Productivity." *Journal of Economic Perspectives* 31 (2): 145–64. http://dx.doi.org/10.1257/jep.31.2.145.
Fitzgerald, J. 2017. "Assessing Product Downsizing, Upsizing, and Product Churn Using Nielsen Scanner Data." Paper presented at the 2017 Joint Statistical Meetings, Session on Measuring the Economy: Economic and Workforce Statistics, Government Statistics Section, Baltimore, MD, July 30, 2017.
Gopinath, G., and O. Itskhoki. 2010. "Frequency of Price Adjustment and Pass-Through." *Quarterly Journal of Economics* 125 (2): 675–727. http://www.jstor.org/stable/27867493.
Gopinath, G., and R. Rigobon. 2008. "Sticky Borders." *Quarterly Journal of Economics* 123 (2): 531–75. https://www.jstor.org/stable/25098909.
Hottman, C. J., and R. Monarch. 2018. "Estimating Unequal Gains across US Consumers with Supplier Trade Data." Federal Reserve Issue Paper 1220, January 17. https://www.federalreserve.gov/econres/ifdp/files/ifdp1220.pdf.
Kamal, Fariha, and Ryan Monarch. 2017. "Identifying Foreign Suppliers in U.S. Import Data." International Finance Discussion Paper 1142r. https://doi.org/10.17016/IFDP.2017.1142r.
Kravis, I. B., and R. E. Lipsey. 1971. "Wholesale Prices and Unit Values as Measures of International Price Competitiveness." In *Price Competitiveness in World Trade*. New York: Columbia University Press.
Lipsey, R. E. 1963. *Price and Quantity Trends in the Foreign Trade of the United States*. Princeton, NJ: Princeton University Press. https://www.nber.org/books/lips63-1.
Mead, D. 2014. "Analyzing Alternatives to Export Price Indexes." *Beyond the Numbers: Global Economy* 3 (27). US Bureau of Labor Statistics, November 2014. https://www.bls.gov/opub/btn/volume-3/analyzing-alternatives-to-export-price-indexes.htm.
Nakamura, A. O., W. E. Diewert, J. S. Greenlees, L. I. Nakamura, and M. B. Reinsdorf. 2015. "Sourcing Substitution and Related Price Index Biases." In *Measuring Globalization: Better Trade Statistics for Better Policy*, edited by S. N. Houseman and M. J. Mandel, 21–89. https://research.upjohn.org/cgi/viewcontent.cgi?article=1250&context=up_press.
Nakamura, E., and J. Steinsson. 2012. "Lost in Transit: Product Replacement Bias

and Pricing to Market." *American Economic Review* 102 (7): 3277–316. https://doi.org/10.1257/aer.102.7.3277.

Price Statistics Review Committee. 1961. Appendix A, "Export and Import Indexes," in *The Price Statistics of the Federal Government*. Cambridge, MA: National Bureau of Economic Research. https://www.nber.org/chapters/c6487.

Rauch, J. E. 2001. "Business and Social Networks in International Trade." *Journal of Economic Literature* 39 (4): 1177–203.

Silver, M. 2010. "The Wrongs and Rights of Unit Value Indices." *Review of Income and Wealth*, Series 56, Special Issue 1 (June): 206–23. https://doi.org/10.1111/j.1475-4991.2010.00391.x.

Triplett, J. E. 1992. "Economic Theory and BEA's Alternative Quantity and Price Indexes." *Survey of Current Business* (April): 49–52. https://apps.bea.gov/scb/pdf/national/nipa/1992/0492trip.pdf.