sciendo

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 83–106, http://dx.doi.org/10.2478/JOS-2022-0005

Unit Value Indexes for Exports – New Developments Using Administrative Trade Data

Don Fast[1], *Susan E. Fleck*[1], *and Dominic A. Smith*[1]

U.S. import and export price indexes replaced unit value indexes forty years ago, given quality concerns of mismeasurement due to unit value bias. The administrative trade data underlying the unit values have greatly improved since that time. The transaction records are now more detailed, available electronically, and compiled monthly with little delay. The data are used by academic researchers to calculate price measures, and unit value indexes based on trade data are used by other national statistical offices (NSOs). The U.S. Bureau of Labor Statistics is now evaluating whether replacing price indexes with unit value indexes for homogeneous products calculated from administrative trade data could expand the number of published official import and export price indexes. Using export transactions, the research calculates detailed unit value indexes from 200+ million trade records from 2012–2017 for 123 export product categories. Results show that 27 of the 123 unit value indexes are homogeneous and closely comparable to published official price indexes. This article presents the concepts and methods considered to calculate and evaluate the unit value indexes and to select the product categories that are homogeneous. Compared to official price indexes, export unit value indexes for the 27 5-digit BEA (U.S. Bureau of Economic Analysis) end-use product categories would deflate real exports of these goods by 13 percentage points less over the period. Incorporating these 27 indexes into the top-level XPI would increase the value of real exports of all merchandise goods by 2.6 percentage points at the end of 2017.

Key words: Unit values; trade; large data sets.

1. Introduction

The U.S. Bureau of Labor Statistics (BLS) official Import and Export Price Indexes (MXPI) measure price changes of U.S. imports and exports of goods and a limited number of services. Other national statistics that calculate trade and output depend on the quality and detail of MXPI to adjust current-dollar measures to constant dollars (Cerritos 2015; Moulton 2018). Over time, the number of publishable detailed MXPI has declined as budget-related sample size reductions shrink the number of items in the market basket and the number of prices that support index quality. Nearly half of the import and export price

[1] U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Rm. 3955 Washington, DC 20212, U.S.A. Emails: fast.don@bls.gov, fleck.susan@bls.gov and smith.dominic@bls.gov
Acknowledgments: We recognize and thank these BLS staff for their research and data support for this paper: Brett Edgerton, Matthew Fisher, Michael Havlin, Laurence Lang, Christina Qiu, Daryl Slusher. Additional support was received from Antonio Caraballo, Ara Khatchadourian, Aric Schneider, and Praveenkumar Yerramareddy. 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. All results have been reviewed to ensure that no confidential information is disclosed.

indexes comprising detailed BEA end-use product categories for merchandise goods did not meet publication criteria in 2020. Furthermore, data collection has become more difficult as business respondents are less willing to participate in the voluntary business survey. Challenges to collecting data from sampled businesses are compounded by social distancing recommendations based on public health guidelines.

The BLS is considering an alternative data source to calculate monthly import and export price changes and is piloting the approach with export transactions. The data source is the official administrative trade data set collected by the U.S. Customs and Border Protection agency that are cleaned and edited by the U.S. Census Bureau for statistical purposes. This data set comprises nearly all importers' and exporters' self-reported detailed shipment records by the international Harmonized System (HS) product classification for the United States.

Unit values have been and are used in both import and export price indexes for the most homogeneous products in which the law of one-price holds, such as commodities traded on a worldwide exchange. For example, the U.S. import price index for crude petroleum is currently calculated using unit values derived from the U.S. Department of Energy (DOE) petroleum transaction import records. 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.

However, using unit values, in general, has been criticized for quality concerns. Even though values and quantities are available in the administrative trade data set and can be used to calculate average prices and unit value indexes (UVIs), previous research has shown that unit value indexes often differ significantly from the price indexes that they are meant to approximate (Alterman 1991; Bradley 2005; Silver 2009). The source of concern is that differences in the composition and characteristics of transactions comprising a unit value can lead to movement within and between unit values that are unrelated to price movements and thus can mismeasure prices. Nonetheless, other research describes how UVIs can and even should be used in place of traditional methods when homogeneous items can be defined and when units of measure are consistent (Balk 1998; Dalton and Fissel 2018; Diewert and Von der Lippe 2010; Silver 2010, 2011).

This research identifies product categories where unit value indexes (UVIs) can potentially replace price indexes for U.S. exports. The homogeneity of product categories is ranked using a coefficient of variation test, which identifies items with less price dispersion. We then calculate UVIs for all product categories and show that UVIs for the more homogeneous product categories are typically consistent with the comparable official export price index (XPI). A subset of UVIs are identified as being consistent with official XPIs.

We compare official export price indexes against all UVIs for 123 5-digit BEA end-use product categories. The approach identifies a fifth of the UVIs, accounting for USD407 billion in trade (2015), to be of sufficient quality to replace published or augment otherwise unpublished price indexes, and another fourth, accounting for another USD228 billion in trade (2015), that are potentially useable but are of indeterminate quality. On the other hand, slightly more than one half of the UVIs, accounting for USD525 billion in trade (2015), were of poor quality showing a wide range of price variability and unit value bias common among heterogeneous goods.

We find that replacing current sources with administrative data for the 27 good UVIs would have a moderate impact on XPI between January 2012 and December 2017. The export indexes created with administrative data would deflate real exports by 13 percentage points less at the end of 2017, compared with current sources. Incorporating these 27 indexes into the top-level goods XPI would increase the value of real exports by 2.6 percentage points at the end of 2017.

UVIs' current prices and quantities potentially improves the current matched model approach that uses a fixed-basket of weights that is known to overestimate price changes. Moulton (2018) describes multiples biases in the MXPI-including lower-level sourcing substitution bias and new product bias-that result from insufficient observations and lagged weights, issues that can be mitigated using the administrative data. Price experts describe how unit values could actually mitigate outlet and sourcing substitution bias (Nakamura et al. 2015, 43–45). In reviewing Alterman's (1991) critique of unit value indexes Nakamura et al. (2015) re-evaluate Alterman's conclusions. They hypothesize that the slower increase of the UVI compared to official MPI that Alterman observed for some product categories could have been an appropriate adjustment for sourcing substitution bias, rather than a mismeasurement of prices (Nakamura et al. 2015, 52). Hottman and Monarch's (2018) research validates Nakamura et al. (2015). When an all-goods import unit value index is calculated with the administrative trade data it shows a flatter price trend than the modified Laspeyres formula used in the official import price index, at a rate of about 2% yearly from 1998 to 2014. Reinsdorf and Yuskavage (2018) also calculate unit values to evaluate sourcing substitution bias of MPIs. All of the research described above supports the use of unit values to mitigate substitution bias and provides impetus to our research to calculate UVIs to replace official import and export price indexes in homogeneous product areas.

Conceptually the 'unit value indexes' in our research differ from the U.S. Census' unit value indexes criticized by Alterman, and the unit values are more narrowly defined, approximating Diewert and Von der Lippe's (2010) classifications of 'reasonably' homogeneous products. While traditional price collection methods in the international price survey typically record a dozen prices per importer or exporter for a few thousand companies, the U.S. electronic trade transaction data set contains tens of thousands of transactions for detailed product classifications by company.

Section 2 first describes the constraints and new frontiers faced when blending alternative data, and particularly the administrative trade data, into official statistics. Section 3 describes the coefficient of variation test for price dispersion. Section 4 describes the building blocks of the methodology. Section 5 describes how the quality of the UVIs was evaluated. Section 6 describes the results and how replacing official XPI with UVIs would affect measures of real exports. Section 7 concludes with a discussion of the steps to operationalize the new approach and potential future work.

2. Requirements for Integrating Alternative Data into Official Statistics

Consideration of detailed U.S. trade transaction data to measure import and export price indexes is part of an effort by U.S. Bureau of Labor Statistics' official price statistics programs to actively pursue alternative data to support the following operational objectives-measure price change more accurately, improve BLS management of respondent burden, expand item and geographic coverage, publish new products, and achieve cost savings. To move forward with alternative data, the new source is expected to meet the core measurement objective of the target population by being timely, accurate, and representative, as well as being consistent with legal and data directives. The methodology applied to the new data source should be comparable or complementary to the methodology used for the official price index. Such methods would address quality change, new and replacement products, sample representativeness, and minimum publishability standards to ensure representativeness and protect against respondent identification. Alternative data sources for price indexes would mitigate loss as we maintain, improve, or expand the components of the relevant market basket. These components include replacement or expansion of current products, introduction of new products, weights for integrating the data into the larger market basket, or more frequent prices.

The completeness of the U.S. electronic administrative trade data set and new constraints on data collection with social distancing in place provide opportunity and impetus, respectively, to set up statistical production in a way that blends unit values from the administrative trade data with the directly collected data, and is similar to other efforts to blend data sources (Reid et al. 2017). This section describes the constraints and new frontiers to meet these statistical obligations, and looks at how limiting price dispersion likely improves homogeneity.

2.1. Constraints on Using Administrative Trade Data

The alternative data proposed to replace or augment direct data collection from businesses must meet core measurement objectives of the target population by being timely, accurate, and representative and must use a methodology that should be comparable or complementary to the methodology used for the official price statistics. These requirements act as constraints or limitations on the options we have to compute unit values from the administrative trade data. The high priority constraints described here relate to (1) timeliness and availability of data, (2) accuracy and representativeness of homogeneous unit values, (3) consistency of published data for data users, (4) ability to aggregate lower level UVIs to upper level published indexes.

First, the administrative trade data must meet timeliness and availability constraints to fit the production schedule of data processing. Sufficient data must be received at the beginning of each month for the previous month's data to assure timely and accurate publication of the preliminary price indexes two weeks after month's end. The first publication of the preliminary price index is updated with additional data in the subsequent three months, but the preliminary index must meet quality standards of representativeness. Collaboration with other government agencies is ongoing to evaluate how much data can be made available in time to meet current publication deadlines. Research includes assessing the quality difference between preliminary and final indexes based on partial and full month trade data, respectively. This work is ongoing and is not described in detail here.

Second, the accuracy and representativeness of unit value indexes rests on the homogeneity of the items and their unit values. We find that there is such variability of product categories that no one item key, or group of transaction characteristics, could define homogeneity across all product categories. Each product category has a different mix of goods, and different patterns of trade. The need to mitigate unit value bias while maintaining item continuity leads us to use the coefficient of variation to evaluate price variability as a key indicator of homogeneity, or at the minimum, of heterogeneity.

The common definition of unit value bias is the compositional effect on indexes when a unit value, or average price, is not adequately narrowly defined to represent one item or product. The choice of index framework will be discussed in the methodology section. As the Export and Import Price Index Manual advises, direct data collection is preferable to unit values based on Customs data (International Monetary Fund 2009). Yet of the six reasons to prefer price survey indexes over unit value indexes, only three pertain to the BLS MXPI-biases due to changes in the mix of the heterogeneous items in customs transaction records, poor quality of recorded data on quantities, and infrequent trade of 'unique' goods, such as ships and large machinery." (International Monetary Fund 2009, xiv). Of these three, the first is the issue that is addressed in our research. The second issue of missing quantities affects only 10% of U.S. trade transactions, providing millions of transactions with quantities and total value, compared to the 45% response rates for exports in the international price survey reported in July 2020 (U.S. Bureau of Labor Statistics 2020a); and the last concern regarding unique high priced products is not adequately addressed in price surveys either, given that sampling and pricing methods exclude infrequent trade of unique items. The compositional differences of product codes in scanner data have been widely discussed for consumer prices, and we borrow from this research to consider how U.S. trade transaction records could measure import and export price changes.

The literature on scanner data informs our approach to defining homogeneous items. There are differences in the way to bundle transactions to calculate unit values for trade data, which Nakamura et al. (2015, 56) think bodes well for combining producer products. Given the production- and contract-driven aspect of international trade, transactions with shared characteristics will likely be tailored to one supplier or market and will represent larger dollar amounts than consumer purchases. The broader homogeneity is also discussed by Diewert and von der Lippe when they describe the tradeoff between homogeneity and continuity. That is, the narrower the classification of the items, the more likely prices and quantities will be zero across periods given the sporadic nature of shipments (Diewert and Von der Lippe 2010, 20). This tradeoff is described mathematically by Chessa (2019), who evaluates scanner data for the Dutch CPI. A homogeneous product is defined by balancing two measures: one measure quantifies the homogeneity of product bar codes and other attributes, while the second measure expresses the degree to which products can be matched each month to ascertain continuity of trade and minimize item churn (Chessa 2019).

Because there is such fluidity between all aspects that define homogeneity – price dispersion, number of grouped characteristics to define an item, and item continuity, we depend on a qualitative and quantitative comparison of the UVIs with the official price indexes to evaluate whether a UVI is equivalent enough to the official price index to replace it.

The third constraint–that the administrative trade data should assure consistency of published data for data users–goes hand in glove with the accuracy and representativeness of both UVIs and the official MXPI. The official import and export price indexes are based on a methodical and statistically sound approach to sample, interview, and collect data from U.S. importers and exporters. As a major economic indicator, all aspects of the survey process are scrutinized with quality performance measures on a quarterly basis. Nearly a thousand price indexes are systemically reviewed for quality and publishability as part of monthly publication, and variance statistics assure representativeness.

Since most of the import and export BEA end-use price indexes are used to deflate net trade in the U.S. GDP, any major difference would need to be explained clearly as breaks in series could affect the usefulness of the indexes. While it is expected that known upward biases of lower level substitution should be dampened with the UVIs, any differences that cannot be explained would reduce the likelihood of acceptance of the new indicator. The trust of data users in the current data product limits us to refine and define homogeneity in a way that assures as much continuity as feasible with our directly collected MXPI. It is not ex-ante clear that unit value indexes are always worse, given that official price indexes may be based on a small number of price quotes and are constructed using less desirable index formulas. Matching directly collected data both gives us confidence in the quality of the UVIs and lets us publish a consistent time series when incorporating UVIs in XPI.

Finally, for consistency of methodology, the data source must be integrated into the directly collected survey data systemically. Direct data collection begins with elementary level items whose prices are reported by the respondent for each sampled product the company recently traded. So too can transaction prices be averaged for individual items, aggregated at the classification group level and then aggregated to the upper level index.

Each transaction in the data set is a shipping record that is primarily categorized by the 10-digit HS classification and includes a few dozen other transaction characteristics. We select a subset of these transaction characteristics, including the HS classification, as an item key. Each unique combination of values for the selected characteristic constitutes an item. We establish four item keys, each of which uses a different subset of characteristics to create different items and unit values. The most broadly defined item key is for the 10-digit HS classification, which produces one item and one unit value composed of all transactions for that product classification. To specify the other three item keys, we heed the Export and Import Price Manual to select price-determining characteristics of export items (International Monetary Fund 2009). These include the foreign destination, U.S. point of origin, domestic producer (U.S. principle party of interest), and other characteristics, like unit of measure and related vs arms-length trade. The variables and their abbreviations used in the figures are domestic/re-export goods (F), employer ID (E), state of origin (S), zip code of export (Z), country of destination (C), U.S. Port (D), quantity units (Q), related trade indicator (R), and 10-digit Harmonized system classification (H). Together these variables provide significant detail about trading relationships, but they provide less detail about the good being traded than would be available in the XPI survey. We use this information to define items at a level of detail that was not studied in previous applications. This approach is consistent with the suggestion of Von der Lippe and Mehrhoff (2010, 7) for improving unit values in German price indexes.

Final price indexes are created through multiple layers of aggregation. Items are combined to form classification group indexes, which are similar to HS classifications. The classification group provides continuity when HS classifications change and itself is mapped through product and industry concordances to the three different classification systems for which the MXPI are published-product indexes based on the Harmonized System and BEA end-use, and industry indexes based on the NAICS. Upper level indexes are aggregated from classification group indexes using a Lowe formula, which the BLS refers to as a modified Laspeyres formula. The trade weights used in this aggregation are updated once yearly, lagging the index by two years. A full explanation of the different classification schemes, the upper level index formula, and the trade weights used by the MXPI is available in the International Price Program Handbook of Methods (U.S. Bureau of Labor Statistics 2020b).

2.2. New Frontier of Administrative Trade Data

From constraints to frontiers, there are potential improvements to be gained from using alternative data sources. The primary reason for revisiting unit value indexes is to reinstate publication of detailed indexes that no longer meet quality standards due to insufficient prices or company coverage. The administrative trade data provides much better representativeness of trade in homogeneous goods, as well as a volume of transactions that dwarfs in-person data collection. Furthermore, the cost savings realized when administrative data replace direct data collection would be used to expand the currently thin coverage of service import and export indexes.

The data source can also mitigate biases that are difficult to address with existing practices. Access to current prices and quantities provides an opportunity to account for lower level substitution bias as described by Moulton, which for export price indexes can be described as producer-side product replacement bias, in which “price changes that occur at the time of product replacements tend to be dropped,” (Nakamura and Steinsson 2012, 3278).

As Nakamura et al. (2015) show in appendix 2b solutions to export product substitution and import sourcing substitution bias both necessarily involve averaging unit prices for different products. They, like we, look to balance homogeneity with substitution, to define an item that is not so broad as to have unit value bias, but not so narrow that price and quantity changes of two similar items are not reflected in the index. We consider groupings of items with broadly and narrowly defined item keys. We find that there is no one item key for all product categories for which the price index best matches the official price index.

3. First Steps to Evaluate Price Dispersion and Homogeneity – Coefficient of Variation

Similar to the proof of concept for unit values described in Fast and Fleck (2022) we propose using a coefficient of variation test to select homogeneous items as a first step to evaluate UVIs and the item keys underlying them for unit value bias. The coefficient of variation test identifies whether within-month price variation of items is small relative to the mean price of that item, which we take as an indicator that the transactions for each item involve relatively homogeneous goods. The coefficient of variation is defined for an item as the standard deviation of the unit value of an item within a month divided by the mean unit value of the item within a month multiplied by 100. We calculate the coefficient of variation for each item’s unit value with multiple transactions within a month. A small coefficient of variation for an item within a month suggests that the physical goods traded in those transactions are similar. Thus, the transactions comprising the item are more likely to represent the same product. Calculating the coefficient of variation within each month minimizes the role of inflation in causing variation across transactions within an item, but it also understates variation that may be due to quality changes over time.

We calculate the coefficient of variation for each item in each 5-digit BEA end-use product category and calculate the distribution of these coefficients across items by category on a weighted basis for each month. Then, we plot the average of these monthly distributions. Figure 1 plots the cumulative distribution function (CDF) of the distribution of item level coefficient of variations by product category. More homogeneous product categories have many items with small coefficients of variation, indicating less price variability within a month, which results in a CDF that is above the other lines.

In some cases the lines corresponding to different product categories cross indicating that one product category has many items that have a small coefficient of variation, but also some items that are more variable. This makes it difficult to strictly rank the product categories according to their homogeneity. In the next section, we construct indexes for each product category and compare them to XPI indexes as a benchmark. We find that UVIs with CDFs that are generally above the CDF of the seasonal product category ‘vegetables, vegetable preparations, and juices’ almost always are close to the corresponding XPI index. Thus, we will refer to any product category with a CDF above that of ‘vegetables, vegetable preparations, and juices’ as homogeneous.

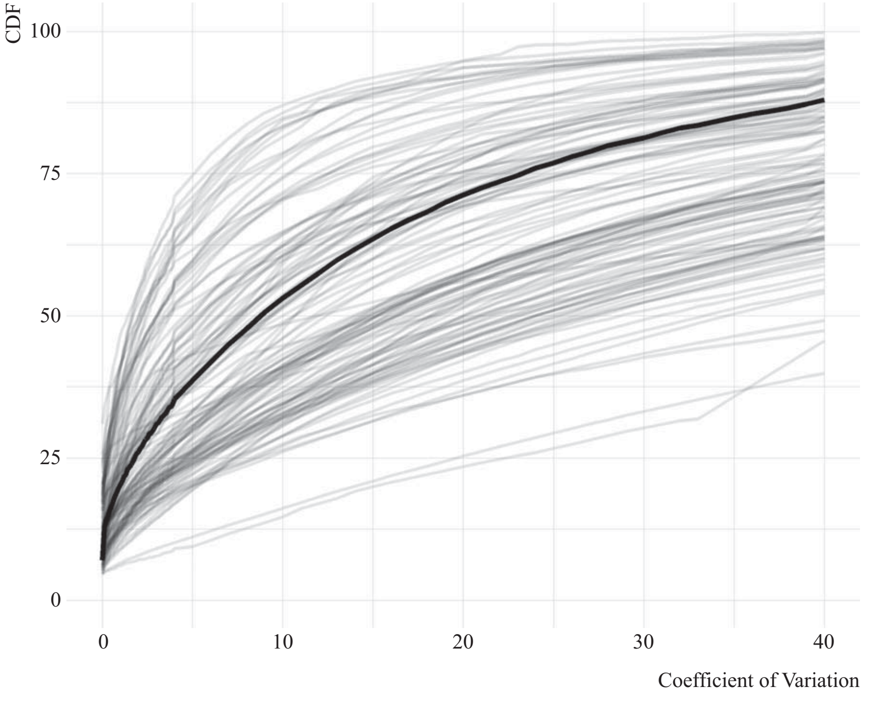

Fig. 1. CDF of Item Coefficient of Variation for 123 BEA end-use product categories. A lower C.V. value indicates less price dispersion, and the steeper and more concave lines represent product categories with more items that have less price dispersion. The dark line is the CDF for ‘vegetables, vegetable preparations, and juices.’ CDFs are calculated for individual months and then averaged across months. The x-axis is truncated at 40 for clarity.

Our results suggest that the coefficient of variation test could be used to identify indexes with well-defined items for unit values if a published index benchmark is not available. However, establishing precise thresholds may be challenging because the expected amount of variation of homogeneous items may vary across settings. Gopinath and Itshoki (2010) and Gopinath and Rigobon (2008) demonstrate that homogeneous goods experience both more frequent and larger price changes than differentiated goods in their analysis of the directly collected import and export price microdata. For our relative threshold, we choose a detailed item key for vegetables, which results in a relatively flat but concave cumulative distribution. Vegetable export prices are widely variable, even at detailed item levels. Both supply and demand are variable, as yield, quality, and availability depend on growing seasons and weather locally and worldwide. There is also significant competition in the sector; 2015–2016 records report more than 4,000 exporters monthly, and nearly 1,000 large regular exporters (Fast and Fleck 2019). Missing prices are filled in by imputation, and this can also cause variability. The combination of competition, short horizon from field to market, and imputation results in price variability, even though many of the 161 HS product categories are of nondifferentiated products, like light red kidney beans, or certified organic asparagus. The index encompasses nearly all conditions of variability that indexes face. There are 36 product categories that have CDFs that are generally above that of vegetables. We will refer to these product categories as homogeneous. There are an additional 18 product categories with CDFs that are similar to that of vegetables.

Our primary evidence of unit value bias comes from comparing UVIs to indexes created using directly collected data. The different data sources are used to create indexes that are blended using a modified Laspeyres formula to calculate overall U.S. Import and Export Price Indexes. This approach to blending data sources is currently in place for imports of crude petroleum and exports of grain, sourced from the Energy Information Agency and Department of Agriculture, respectively.

4. Computing Unit Value Indexes

This section describes our approach to constructing UVIs. The construction takes place in three steps. First, unit values for each trade transaction are computed from value and quantity. Second, unit values for transactions are combined to form item unit values using a weighted geometric mean. Items are defined by shared characteristics of transactions. Third, the price relatives of these items are aggregated to a classification group using a Tornqvist index formula. Finally, the classification group indexes are aggregated to create 123 5-digit BEA end-use indexes using a modified Laspeyres formula. We consider four different "item keys" which are sets of characteristics. Each set of items related to these item keys results in different unit values and UVIs. Thus, there are four options for each 5-digit BEA end-use index.

4.1. Calculating Unit Values

The administrative data contain information on the value of shipments and the associated quantity that can be used to form unit values. We calculate the unit value for a transaction by dividing the dollar value by the quantity. The unit value P of a transaction s for any given month t is the value of the transaction V divided by the quantity of units Q in the transaction,

$$P_{s,t} = \frac{V_{s,t}}{Q_{s,t}}. \tag{1}$$

The transactions, S_k, associated with some item k, defined as a unique combination of characteristics, are aggregated using a weighted geometric mean to form a unit value for each item. The unit value of an item k at some month t is

$$P_{k,t} = \left[\prod_{s \in S_k} P_{s,t}^{V_{s,t}}\right]^{\frac{1}{\sum_{j \in S_k} V_{j,t}}}. \tag{2}$$

4.2. Classification Group Unit Value Indexes

The items are then aggregated to the classification group using a Törnqvist formula. The unit value index for classification group c is calculated by aggregating the unit value indexes for the set of items K_c that belong to classification group c. The weights, $w_{k,t} = \frac{V_{k,t}}{\sum_{j \in K_c} V_{j,t}}$, are created using the value of trade associated with an item in some month, $V_{k,t} = \sum_{s \in S_k} V_{s,t}$.

$$R_{c,t} = \prod_{k \in K_c} \left[\frac{P_{k,t}}{P_{k,t-1}}\right]^{\frac{W_{k,t-1} + W_{k,t}}{2}} \tag{3}$$

Equation (3) describes the month-to-month changes in the classification group UVI. This is converted to an index level by starting the index at 100 in January of 2012 and advancing it using R_{ct}. Therefore, the index level for a classification group in a month is

$$P_{c,t} = P_{c,t-1} R_{c,t} \tag{4}$$

with $P_{c,0} = 100$.

4.3. Forming 5-Digit BEA End Use Indexes

For this article we focus on aggregating the indexes to the 5-digit BEA end-use product category because that is the level at which decisions about whether to use unit value indexes for a given set of classification group indexes will be made. Eventually the classification group unit value indexes will be used as the building blocks for HS, BEA end-use, and NAICS based indexes. These indexes will be combined with price indexes from directly collected data using a modified Laspeyres index formula. The weights used to aggregate from classification group to BEA end-use product category are lagged values for consistency with directly collected data. The weights used in the MXPI are lagged two years due to delays in the availability of real-time weights. The formula for the unit value index for BEA end-use product category e is

$$R_{e,t} = \sum_{c \in C_e} \frac{V_{c,b}}{\sum_{i \in C_e} V_{i,b}} \frac{P_{c,t}}{P_{c,t-1}} \tag{5}$$

This is converted to an index level by starting the index at 100 in January of 2012 and advancing it using $R_{e,t}$. Therefore, the index level for a product category in a month is

$$P_{e,t} = P_{e,t-1} R_{e,t} \tag{6}$$

with $P_{e,0} = 100$.

4.4. Missing and New Items

Our approach to missing and new items is similar to the approach used in the official MXPI. When an item is missing, we impute a price for the item using cell mean imputation for up to three months before discontinuing the series. For directly collected data, there are attempts to replace discontinued items and contact respondents for missing prices, whereas for administrative data a missing price for an item indicates that no trade occurred. An item naturally occurs as trade transactions begin, and we pre-impute a price in the month before the item enters using cell mean imputation.

4.5. Possibilities for Improvement and Caveats

There are a number of ways in which unit value indexes with administrative data may improve on the official XPI. The administrative data eliminate issues with representativeness of trade and response rate concerns. The data also allow us to address changes in trade in a timelier manner.

The administrative data could potentially expand the number of detailed published indexes, because two major hurdles with directly collected data are avoided. Representativeness of trade and nonresponse are no longer problems when transaction records comprise the universe of trade. Administrative trade data permit detailed indexes to be published at whatever level that trade occurs, while the directly collected data are based on a voluntary survey. Indexes must meet quality review standards of sample representativeness and index quality. The number of sample units is distributed across all merchandise goods categories by their trade weight to assure representativeness for the aggregate MXPI, but this sacrifices more detailed index representativeness. Even though item prices are collected, detailed indexes are not representative of the trade if the trade dollar value is below the cut-off. For product categories meeting the trade dollar cut-off, nonresponse negatively affects publishability as measures for representativeness of companies, number of companies, and number of items are all used as publishability standards. Nonresponse is a nonissue for administrative data.

Using current values and quantities with administrative data allows for the incorporation of new goods and replacement of substitutes. New and exiting goods are accurately counted because the complete data set accounts for all current trade. Furthermore, substitution with classification groups can be addressed because the availability of current trade values allows us to use a superlative index formula.

5. Comparison of UVIs to XPI Benchmarks

After calculating UVIs for each item key and product category, we must determine which are of sufficient quality to replace the product category's XPI counterpart. Our first criteria is agreement with the relevant XPI when the XPI is of sufficient quality to be used as a benchmark. We evaluate agreement for each of the four item keys under consideration and manually choose the best key for each product category and rate the quality of the fit. This analysis used graphs of index levels, month-to-month changes, and statistics measuring the quality of the fit. Then, we evaluate the performance of this "best" item key against XPI using both the manual ratings and multiple statistics that measure the quality of fit. Our focus in evaluating the quality of fit is on long-term agreement between the two index levels. Finally, we discuss what we do when the official XPI may not be a reliable benchmark.

5.1. Identifying "Best" Item Keys and Visual Analysis

Our first step of determining which unit value indexes were potentially able to replace official price indexes was to have research team members manually review each index. Each member studied each of the BEA end-use product categories individually using graphs of index levels, month-to-month changes, and statistics comparing the fit of each index. They identified the item key that was the best fit and rated the fit of that key as either "good", "undecided", or "poor". Examples of each type of index are presented in Figures 2, 3, and 4. In each figure, we show the index levels and month-to-month changes for each item key and the XPI benchmark. All three examples involve published XPI indexes that are considered to be high quality. Figure 2 is an example of a "good" index. The best item key is HECQR, a key with a medium amount of detail. The UVI tracks the XPI index very closely. The UVI generated using just information on HS product codes, the H key, has large spikes and performs much worse. Figure 3 shows an "undecided" index. The best item key is H. In this case, the unit value index is less close to the XPI benchmark. There are multiple times where the UVI increases significantly more than the XPI benchmark, but the levels are generally close until the final year. In this case, there is some hope that with further improvements to our methodology, such as using a different item key or handling outliers different, a UVI could be used. Figure 4 shows a "poor" index where there is little reason to believe a unit value index could be used for this BEA end-use product category. The more detailed keys generally perform better, but none of the keys performs very well.

Table 1 shows the rating of the indexes and groups the number of BEA end-use product categories by rating and publication status of official XPI. When there was disagreement between team members, the more common rating was used. Disagreements were relatively rare and only involved whether an index was "good" or "undecided" or whether an index was "undecided" or "poor".

Manually reviewing the indexes allows us to consider multiple pieces of information simultaneously when rating indexes, but the process is time consuming as we evaluate the impact of alternative methodologies. Additionally, manual ratings can vary across individuals and time. In the following subsection, we describe our approach to develop acceptance rules based on statistical comparisons. We search for a set of acceptance

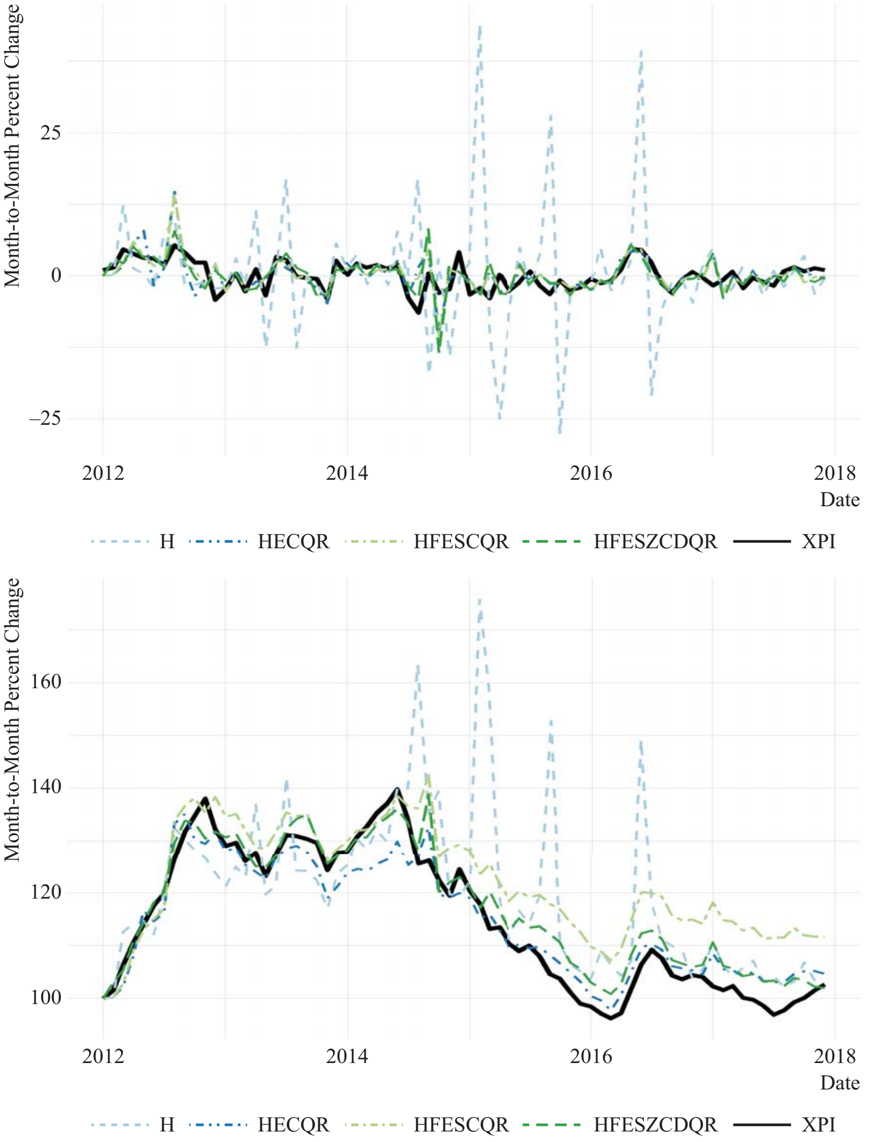

Fig. 2. Example of product category with a "good" UVI. The best key is HECQR. The product category is "Other animal feeds, not elsewhere classified." The top panel depicts month-to-month percent changes. The bottom panel depicts index levels.

criteria that can be applied to our data. The goal of these criteria is to approximate the manual review. We focus on a set of criteria that can differentiate between "good" and "poor" indexes.

5.2. *Statistical Comparison*

We considered multiple statistics to create a set of criteria to rate unit value indexes. The criteria use information on both index levels and month-to-month changes. From the initial

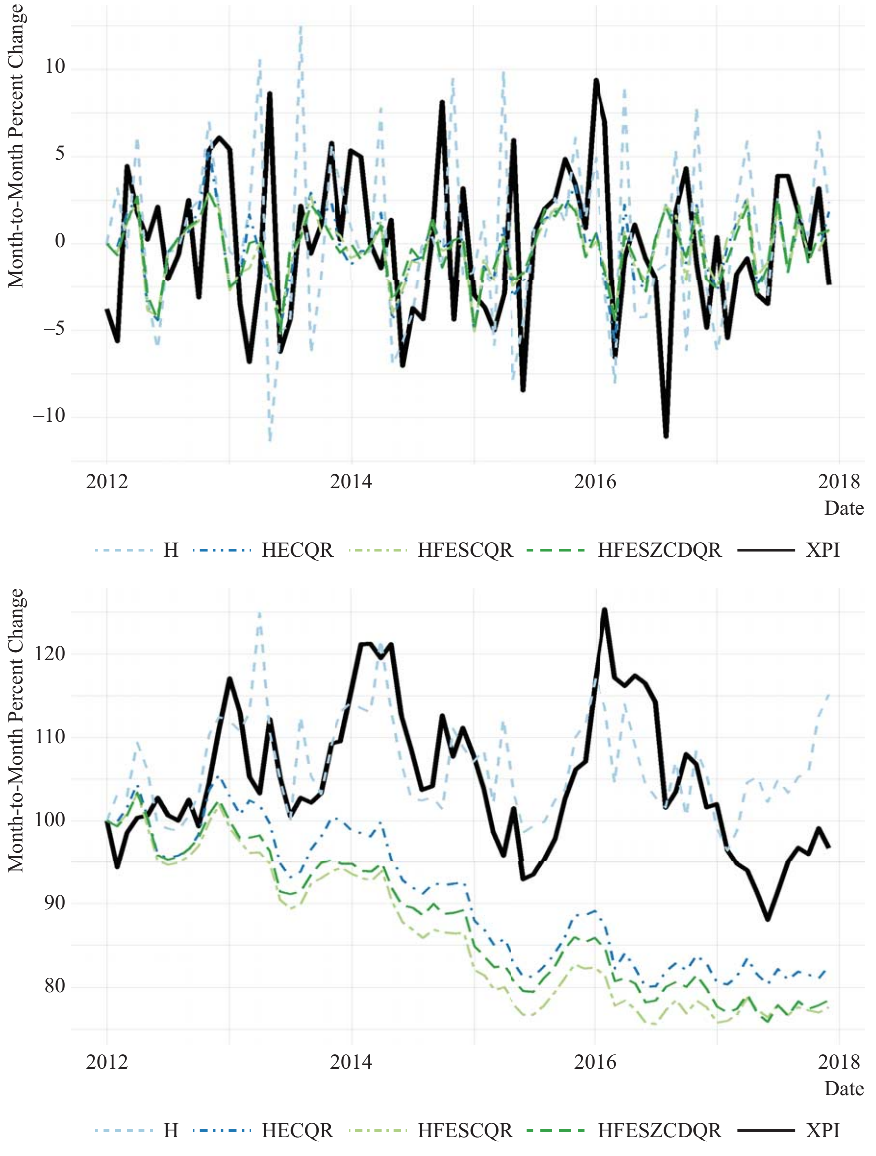

Fig. 3. Example of a product category with an "undecided" UVI. The best key is H. The product category is "Fruit and fruit preparations, including fruit juices." The top panel depicts month-tomonth percent changes. The bottom panel depicts index levels.

set of statistics we identified three that jointly had significant predictive power for whether an index would be classified as "good" by the manual review. We describe the three statistics that had significant predictive power and what values of these statistics generally indicated a good index.

Previous work has used a wide range of statistics to evaluate whether price indexes generated with alternative data are similar to benchmark indexes, but there is significant disagreement about which tests should be used and what the appropriate thresholds are (Fitzgerald and Shoemaker 2013). The appropriate threshold will depend on the needs of

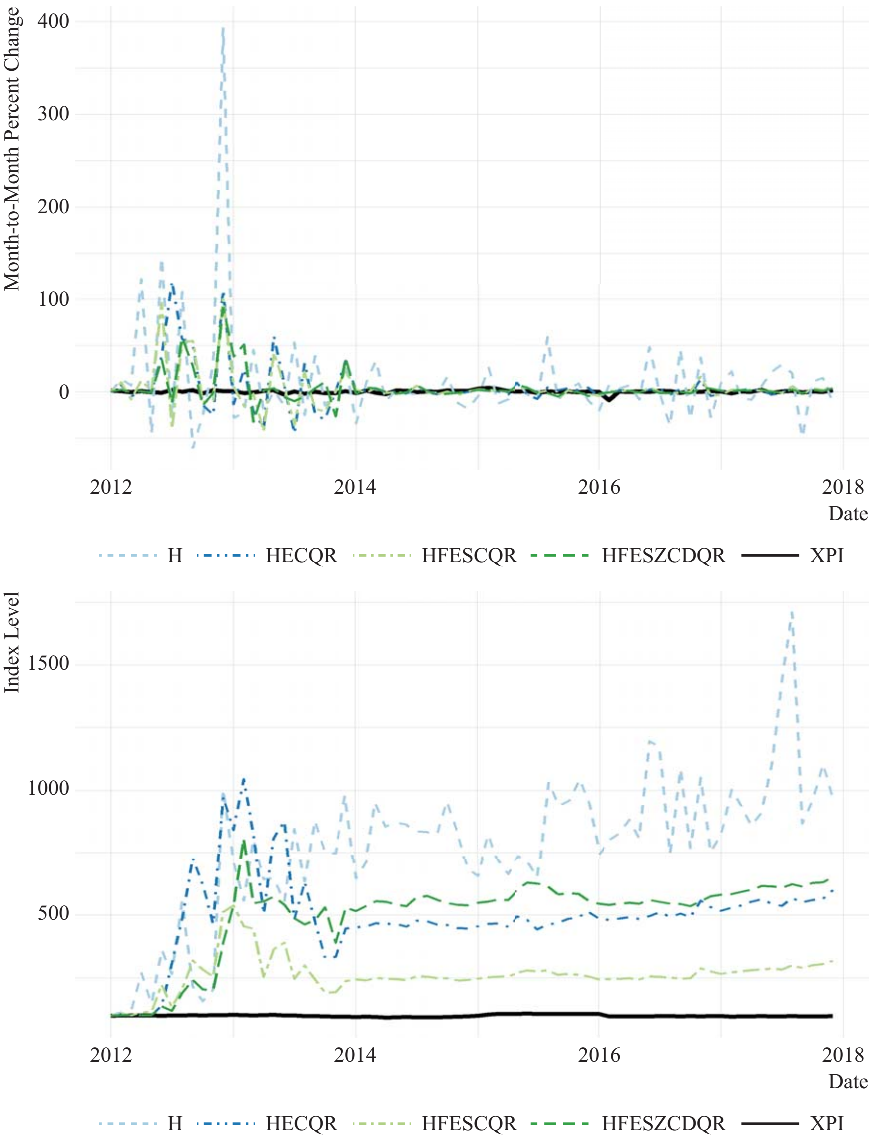

Fig. 4. Example of a product category with a "poor" UVI. The best key is HFESCQR, but that is not a close approximation to the official XPI. The product category is "Industrial inorganic chemicals." The top panel depicts month-to-month prercent changes. The bottom panel depicts index levels.

data users and potential benefits of replacing official indexes with alternative data. It will also depend on whether it is most important for the UVIs and benchmark to have similar month-to-month changes or longterm trends. Certain statistics, such as the root mean squared error between the month-to-month changes in indexes, can indicate agreement over the short term, but lead to large long-term differences if the monthly misses are all in a certain direction.

We identified three statistics that together seemed to have the most predictive power. They are the fifth largest absolute difference between index levels, the intercept of a

Table 1. Number of BEA end-use product categories receiving a rating by published status.

Rating	Published	Unpublished	Total
Good	15	12	27
Undecided	11	21	32
Poor	24	40	64

Notes: Numbers are the count of BEA end-use product categories manually assigned a rating. Columns differentiate between product categories that are published and those that are not.

regression of index levels of the unit value index on XPI price index levels, and a t-test of whether the month- to-month changes have the same mean. For each statistic, we suggest an approximate cutoff value to differentiate the "good" and "poor" UVIs. These cutoff values were chosen jointly as the values such that satisfying all three cutoffs indicated a "good" UVI and failing one or more indicated a "poor" UVI. We experimented with multiple combinations of statistics and with treating the "undecided" UVIs as either "good" or "poor". The set of three statistics proposed here and their associated cutoffs represent the choices that consistently performed well at minimizing misclassification of "good" and "poor" UVIs. They should be treated as guidelines to quickly evaluate the quality of generated indexes before a more rigorous review.

The fifth largest absolute difference tests whether index values stay close to each other over time. Using the fifth largest difference allows for temporary differences between indexes if the indexes come back together before five months have passed. We find there are cases where indexes have large differences for a month or two, but otherwise match well. Therefore, we prefer the fifth largest difference over a test such as the maximum difference. We also considered using the percentage difference accounts for differences in index levels due to different growth rates across BEA end-use product categories, but the % difference is not symmetric. We find that a fifth largest absolute difference of less than 22 index points is a necessary condition for a unit value index to receive a "good" rating.

The second test comes from running a regression of the index values of the unit value index on the index value of the XPI. If the values of the two indexes were always in agreement the estimated parameters would be an intercept of zero and a slope of one. These tests are similar to testing for a strong positive correlation between the two series. However, this test avoids cases where series are correlated, but diverge because in those cases either the slope or intercept will be far from its target value. We have generally found that an estimated intercept of less than 50 is indicative of a "good" UVI. Recall that the indexes are 100 in their base period which helps explain the rather large cutoff. Figure 5 illustrates this test using the "good" index from Figure 2. The points are generally near the 45-degree line, but the UVI tends to exceed the XPI index level when both levels are low. Then, when index levels have grown, the UVI is consistently less than the XPI index level.

Table 2 The third test is a paired t-test. The test checks whether the means of the month-to-month changes of the two indexes are different. The null hypothesis of this test is that the month-to- month changes from the two indexes come from the same distribution. We use the p-value of the paired t-test as the value we track. A large p-value indicates that the null hypothesis cannot be rejected which is the goal of this test. We find that a p-value greater than 0.5 is typically indicative of a "good" UVI.

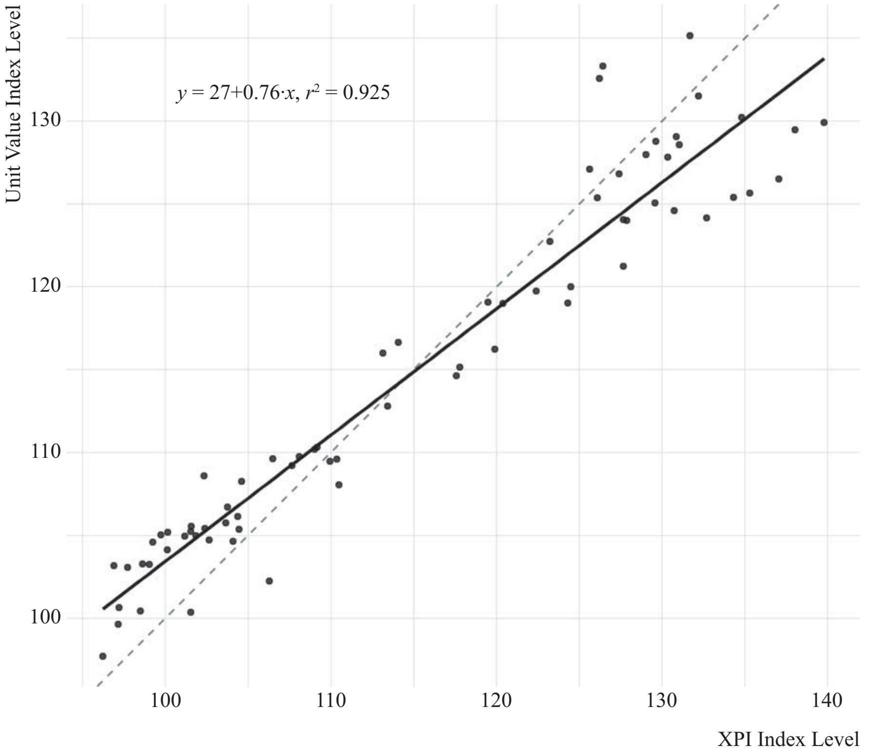

Fig. 5. Regression of Unit Value Index Levels on XPI Index Levels. Each point represents the index levels for a month. The 45 degree line is included as a dashed line. The solid line is a linear regression line and the equation for the line is at the top left of the figure along with the R-squared of the regression. The product category is "Other animal feeds, not elsewhere classified." The UVI for this product category is plotted in Figure 2.

Table 2. Average of three main statistics by manual UVI rating and published status.

	Published			Unpublished		
	Good	Undecided	Poor	Good	Undecided	Poor
Difference	8.91	23.75	172.76	13.25	26.44	492.90
Intercept	34.11	124.14	2162.52	29.97	101.37	2804.93
T-Test	0.80	0.30	0.14	0.83	0.57	0.21

Notes: Average of each statistic for indexes in each group. Difference is the fifth largest absolute difference. Intercept is the intercept of a regression of the index values of the unit value index on the index value of the XPI. T-test is the p-value from a paired t-test using the month-to-month changes of both indexes. Lower values indicate more agreement for the difference and intercept. Higher value indicate more agreement for the t-test. Published indicates the BEA end-use product category is published by the XPI, indicating it is a more reliable benchmark.

Table 3 presents results for additional statistics to show that these follow the same trends with the "good" product categories performing the best. The results of these statistics are broadly consistent with the results presented above. However, in each of these cases, we found instances where the statistic indicated significant agreement, but the manual review indicated the UVI was not a good match with the XPI benchmark. The table shows that mean absolute difference in month-to-month changes between the UVIs in published product categories that received a "good" rating and their corresponding XPI

Table 3. Average of three Main Statistics by manual UVI rating and published status.

	Published			Unpublished		
	Good	Undecided	Poor	Good	Undecided	Poor
RMSE	3.35	15.66	36.64	4.30	5.58	60.55
MAD	2.50	4.15	12.03	3.10	3.84	25.97
Sign Agreement	67.81	58.18	48.63	55.71	48.03	44.54

Notes: Average of each statistic for indexes in each group. RMSE is the root mean-squared error between the month-to-month % changes. MAD is the mean absolute difference between the month-to-month % changes. Sign agreement is the percentage of months where the sign of the month-to-month changes is the same. Published indicates the BEA end-use product category is published by the XPI, indicating it is a more reliable benchmark.

index is 2.5 percentage points. The sign of the month- to-month changes agreed 67.81% of the time. The "undecided" indexes performed slightly worse with a difference of 4.15 percentage points, agreeing only 58.18% of the time. The "poor" indexes performed the worst by far with a difference of 12.03 percentage points, but the sign of the change still agreed over 48% of the time.

6. Results

Table 4 shows that 28.7% of the share of goods export trade could be represented by good quality unit value indexes valued at USD407 billion of 2015 trade, counting both published and unpublished XPI indexes. If the undecided unit value indexes were to have a better fit with different assumptions, an additional 16% of export trade could use unit value indexes to measure price change. Of the undecided unit value indexes, 21 (representing 9% of trade) are not published. The discrepancy between unit value indexes and XPI could be because XPI is not a good benchmark. Seven of the 21 product categories are more homogeneous than "vegetables" so using unit value indexes for at least these product categories may be possible. There are 64 unit value indexes that are considered poor, representing 56.6% of the value of trade. Forty of these come from unpublished indexes, but only two of the 40 product categories are more homogeneous than "vegetables" so it appears likely that many of the poor fitting unit value indexes should not replace the XPI index. Most of the remaining product categories are heterogeneous goods such as machinery that are not suitable for UVIs.

Table 4. Manual rating by published status.

Published	Rating	Number of Indexes	Trade USD Value	Share of Indexes	Share of Trade USD
Yes	Good	15	271	12.1	19.1
No	Good	12	136	9.8	9.6
Yes	Undecided	11	100	8.9	7.1
No	Undecided	21	128	17.1	9.0
Yes	Poor	24	307	19.5	21.6
No	Poor	40	218	32.5	15.4

Notes: Published indicates whether XPI publishes a detailed index for a given product category. Rating is the consensus rating from manual review. The trade USD value is in 2015 billions.

6.1. Homogeneity Matters

Even though evaluating homogeneity was not the primary test of index comparability, the use of the coefficient of variation is supported by the results. Figures 6, 7, and 8 show the CDFs of the coefficient of variation for the product categories with good UVIs (Figure 6), undecided quality UVIs (Figure 7), and poor quality UVIs (Figure 8). The product categories with good indexes have coefficient of variation CDFs that are usually above the

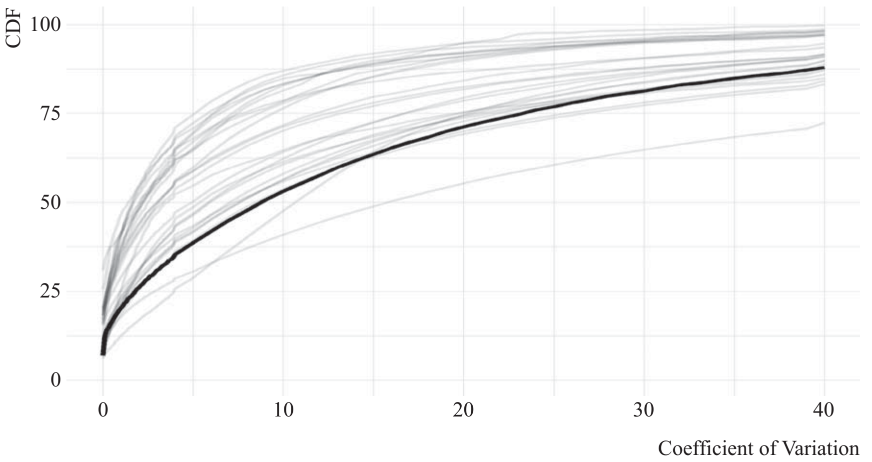

Fig. 6. CDF of Item Coefficient of Variation for 27 BEA end-use product categories that received a "good" rating. Lines represent the fraction of items in a product category that have a coefficient of variation below a given number. CDFs are calculated for individual months and then averaged across months. The darker line is the CDF for 'vegetables, vegetable preparations, and juices.' The x-axis is truncated at 40 for clarity.

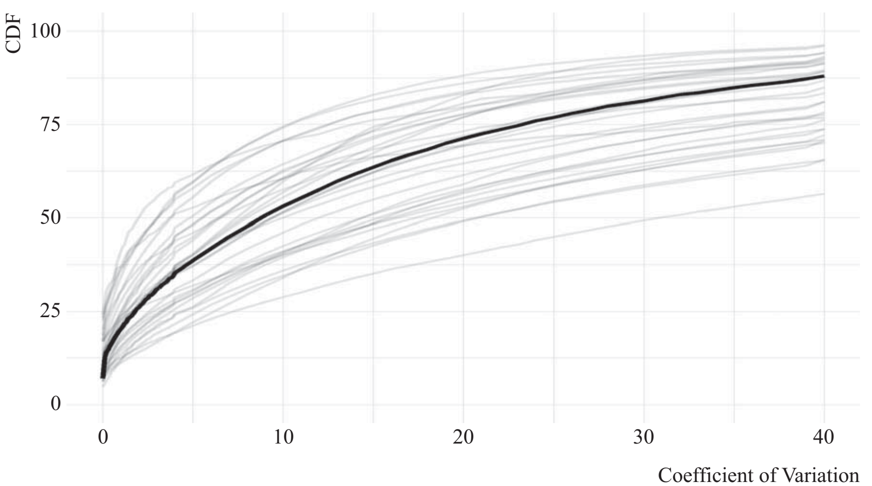

Fig. 7. CDF of Item Coefficient of Variation for 32 BEA end-use product categories that received an "undecided" rating. Lines represent the fraction of items in a product category that have a coefficient of variation below a given number. CDFs are calculated for individual months and then averaged across months. The darker line is the CDF for 'vegetables, vegetable preparations, and juices.' The x-axis is truncated at 40 for clarity.

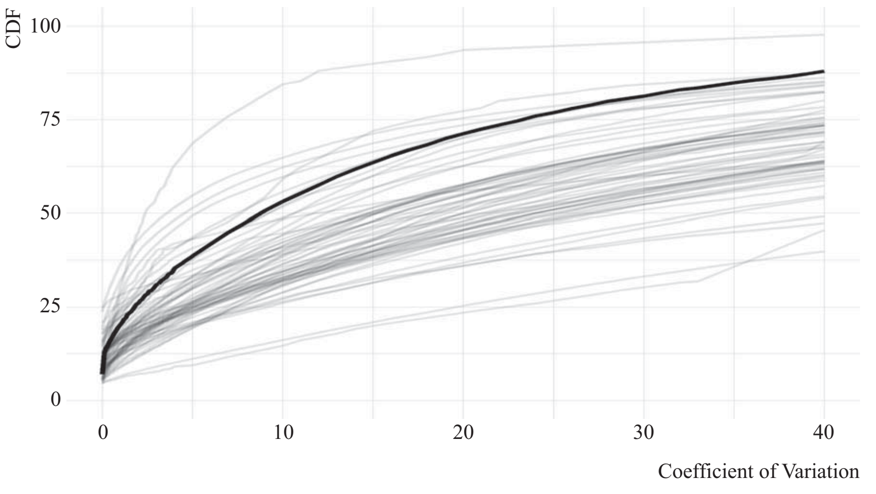

Fig. 8. CDF of Item Coefficient of Variation for 64 BEA end-use product categories that received a "poor" rating. Lines represent the fraction of items in a product category that have a coefficient of variation below a given number CDFs are calculated for individual months and then averaged across months. The darker line is the CDF for 'vegetables, vegetable preparations, and juices.' The x-axis is truncated at 40 for clarity.

vegetables CDF. The product categories with undecided quality UVIs are more mixed, while the product categories with poor quality UVIs are almost always below the CDF of vegetables.

6.2. *Impact on Gross Domestic Product*

MXPI at a detailed level are used as deflators for net exports by both the US Census and the Bureau of Economic Analysis. Net exports are a component of Gross Domestic Product. If the administrative data actually provide more and better quality statistics with the administrative trade data, they will contribute to improving the measure of GDP.

To simulate the impact of replacing the historic BEA end-use 5-digit export price indexes for 2012–2017, price indexes for each subgrouping of the "good", "undecided", and "poor" price indexes are aggregated and the difference between the trade-weighted unit value index endpoint of December 2017 is aggregated for each quality group relative to that of the corresponding official export price indexes – first for the partial aggregation and then to the top-level index, holding all else equal. As can be seen in Table 5, the 27 5-digit BEA end-use "good" unit value indexes show a December 2017 index value of 75.62. This compares to the estimate of the corresponding official XPIs of 83.87. The total dollar value of these indexes is 29% of export trade. Calculating the impact on all goods' export prices is done by assuming that the other 71% of trade value does not change prices. Applying official MXPI methodology, the 5-digit indexes are aggregated to an all-goods measure, which results in a top-level aggregate unit value index of 93.00 and the official XPI of 95.37. The price change gap in the "good" price indexes is 8.3 percentage points over the six years. For the top-level index comparison, the price gap between the "good" unit value indexes versus the historic official price indexes is 2.4 percentage points from 2012 to 2017. That is, if the "good" unit value indexes were to be incorporated into the all-

Table 5. Effect of UVIs on XPI price indexes.

Index classification	Average of detailed XPI price indexes	Average of detailed unit value indexes	Constant all-goods official price index	Constant all-goods unit value index	% of total trade USD value
Good (N=27)	83.87	75.62	95.37	93.00	29
Undecided (N=32)	100.22	119.52	100.04	103.13	16
Poor (N=64)	99.43	312.09	99.68	220.20	57

Notes: Columns 2 and 3 calculate a price index by combining the BEA end-use product categories that receive a given manual rating. Columns 4 and 5 compare the overall goods index when the UVIs with a given rating are used instead of the XPI indexes for those product categories.

goods XPI, export prices would have risen 2.4 percentage points more. In addition, when applied as deflators, real export prices would have fallen by 2.4 percentage points more. The same thought experiment comparing the "undecided" unit value indexes with their comparable 5-digit official price indexes results in a gap of 3.1 percentage points in the other direction. This gap is only slightly larger than the 2.4 percentage point different for the "good" indexes, but this fact is because the "undecided" indexes account for only 16% of the value of exports, much less than the 29% for good indexes. The "poor" unit value indexes show an extreme price gap of over 100 percentage points.

This range of price gaps and the extreme variability for the "poor" quality grouping validates the basic tenet of minimizing price variability and maximizing substitutability that guides this research and qualitatively affirms that the three quality bins are reasonable.

The effort to measure the impact on net trade and thus GDP will depend on the results from the import comparison. Hypothetically, if there were a commensurate adjustment upward to the import index prices, and thus downward impact on real goods imports, the impact on net trade may be small, and thus have a minimal impact on GDP. The direction of the impact is sensitive to the choice and number of indexes determined to be "good". The different composition of imports, with a larger share of heterogeneous products, must be calculated before the possible impact on the real trade balance and GDP can be measured.

7. Conclusion

This research on a large-scale transition to using administrative data develops an exhaustive approach to evaluate product variety, price variability, substitutability, and index comparability for potential replacement in the official MXPI. The results show that not all unit value indexes are the same and that unit value indexes most likely to replace official MXPI measures are homogeneous and should closely align with the index they replace. The ambiguity of some of the indeterminate quality indexes may lie in the fact that there is no comparable published official XPI, or that the product variety itself is broad.

The test run of unit value indexes for 123 detailed product categories over six years shows that defining homogeneity matters, and that one can develop statistical tests and create cutoffs to evaluate differences and make judgments on the consistency, reliability, and comparability of unit value indexes relative to the official export price indexes. Product categories representing 43% of the value of exports could potentially be replaced by unit value indexes if future research is able to convert all "undecided" indexes into "good" indexes. Ongoing approaches attempt to improve the quality of these unit value indexes, by evaluating the product variety key, the bias introduced by the new methodology, and/or the nonresponse bias in the official price index.

Research continues to be evaluated and refined. Some efforts to estimate hedonic linear regressions on the complete data set to develop a systematic method of identifying the best item key combination for each strata are constrained by IT capacity. Exploring time-dummy hedonic models may reduce the specification constraints of grouping data variables and calculating indexes for items. At the margins, improvements to the quality of the "good" set of indexes and additionally "undecided" indexes possibly can move to the "good" category. The Tornqvist index formula helps with substitution bias at the classification group level, but it introduces new concerns because it is a monthly-chained index. Frequent chaining has been determined to exacerbate chain drift (Ivancic et al. 2011; De Haan and Van der Grient 2011). Thus, work is being done to measure chain drift in the UVIs and investigate alternative aggregation methods, such as the base construction strategy described by Statistics Finland (Nieminen and Montonen 2018).

In addition, work will be done to analyze and compare import unit value indexes with official import price indexes and to calculate partial-month measures with low variance compared to full-month data. These two large projects must be addressed before the project to blend administrative data into official import and export price measures can begin.

8. References

Alterman, W. 1991. "Price trends in US trade: New data, new insights." In *International economic transactions: Issues in measurement and empirical research*, edited by P. Hooper and J.D. Richardson, 109–143. University of Chicago Press. Available at: https://www.nber.org/system/files/chapters/c8429/c8429.pdf (accessed April 2021)

Balk, B.M. 1998. "On the Use of Unit Value Indices as Consumer Price Subindices." Proceedings of the Fourth Meeting of the Ottawa Group, International Working Group on Price Indices. Washington, D.C., April 22–24, 1998. Available at: http://www.ottawagroup.org (accessed April 2021).

Bradley, R. 2005. "Pitfalls of Using Unit Values as a Price Measure or Price Index." *Journal of Economic and Social Measurement* 30: 39–61. DOI: https://10.3233/JEM-2005-0243.

Cerritos, G. 2015. "The Role of BLS Import and Export Price Indexes in the Real GDP," *Monthly Labor Review*. DOI: https://10.21916/mlr.2015.19.

Chessa, A. 2019. "MARS: A Method for Defining Products and Linking Barcodes of Item Relaunches." Proceedings of the Sixteenth Meeting of the Ottawa Group, International

Working Group on Price Indices. Rio de Janeiro, Brazil, May 8–10, 2019. Available at: http://www.ottawagroup.org (accessed April 2021).

Dalton, M., and B. Fissel. 2018. "A Unified Framework for Calculating Aggregate Commodity Prices from a Census Dataset," *Journal of Economic and Social Measurement* 43: 85–104. DOI: https://10.3233/JEM-180453.

De Haan, J. and H.A. van der Grient. 2011. "Eliminating Chain Drift in Price Indexes Based on Scanner Data." *Journal of Econometrics* 161: 36–46. DOI: https://10.1016/j.jeconom.2010.09.004.

Diewert, W.E., and P. von der Lippe. 2010. *Notes on Unit Value Index Bias.* Discussion Paper 10-08, Department of Economics, University of British Columbia. October. Available at: https://economics.ubc.ca/files/2013/06/pdf_paper_erwin-diewert-10-8-notes-unit-value.pdf (accessed April 2021)

Fast, D., and S.E. Fleck. 2019. *Measuring Export Price Movements with Administrative Trade Data.* Working Paper 518. U.S. Bureau of Labor Statistics. June. Available at: https://www.bls.gov/osmr/research-papers/2019/pdf/ec190080.pdf (accessed April 2021).

Fast, D., and S.E. Fleck. 2022. "Unit Values for Import and Export Price Indexes – A Proof of Concept." in *Big Data for Twenty-First-Century Economic Statistics. University of Chicago Press*, edited by K.G. Abraham, R.S. Jarmin, B.C. Moyer, and M.D. Shapiro. Available at: https://www.nber.org/books-and-chapters/big-data-twenty-first-century-economic-statistics (accessed April 2021).

Fitzgerald, J., and O. Shoemaker. 2013. *Evaluating the Consumer Price index Using Nielsen's Scanner Data.* Office of Survey Methods Research Paper. U.S. Bureau of Labor Statistics. Available at: https://www.bls.gov/osmr/research-papers/2013/st130070.htm (accessed April 2021).

Gopinath, G., and O. Itskhoki. 2010. "Frequency of Price Adjustment and Pass-Through." *The Quarterly Journal of Economics* 125 (2): 675–727. DOI: https://10.1162/qjec.2010.125.2.675.

Gopinath, G., and R. Rigobon. 2008. "Sticky Borders." *The Quarterly Journal of Economics* 123 (2): 531–575. DOI: https://10.1162/qjec.2008.123.2.531.

Hottman, C.J., and R. Monarch. 2018. "Estimating Unequal Gains across US Consumers with Supplier Trade Data." *International Finance Discussion Papers* 1220. DOI: https:// 10.17016/IFDp. 2018.1220.

International Monetary Fund. 2009. *Export and Import Price Index Manual: Theory and Practice*. Washington, D.C. Available at: https://www.imf.org/external/np/sta/xipim/pdf/xipim.pdf (accessed April 2021).

Ivancic, L., W.E. Diewert, and K.J. Fox. 2011. "Scanner Data, Time Aggregation and the Construction of Price Indexes." *Journal of Econometrics* 161 (1): 24–35. DOI: https://10.1016/j.jeconom.2010.09.003.

Moulton, B. 2018. "The measurement of output, prices, and productivity." Available at: https://www.brookings.edu/research/the-measurement-of-output-prices-and-productivity/ (accessed April 2021).

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. Michael, 1: 21–89. DOI: https://10.17848/9780880994903.

Nakamura, E., and J. Steinsson. 2012. "Lost in Transit: Product Replacement Bias and Pricing to Market." *American Economic Review* 102 (7): 3277–3316. DOI: https://10.1257/aer.102.7.3277.

Nieminen, K., and S. Montonen. 2018. "The Foundation of Index Calculation." *Helsinki: Statistics Finland*. Available at: http://www.stat.fi/static/media/uploads/meta_en/menetelmakehitystyo/foundation_of_index_calculation_nieminenmontonen_final.pdf (accessed April 2021).

Reid, G, F. Zabala and A. Holmberg. 2017. "Extending TSE to Administrative Data: A Quality Framework and Case Studies from Stats NZ." *Journal of Official Statistics* 33 (2). DOI: https://10.1515/jos-2017-0023.

Reinsdorf, M., and R. Yuskavage. 2018. "Offshoring, Sourcing Substitution Bias and the Measurement of US Import Prices, GDP and Productivity." *The Review of Income and Wealth* 64 (1): 127–146. DOI: https://10.1111/roiw.12263.

Silver, M. 2009. "Do Unit Value Export, Import, and Terms of Trade Indices Represent or Misrepresent Price Indices?" *IMF Staff Papers* 56 (2). 297–322, June. DOI: https://10.1057/imfsp. 2008.24.

Silver, M. 2010. "The Wrongs and Rights of Unit Value Indices." *Review of Income and Wealth* 56: S206–S223. DOI: https://10.1111/j.1475-4991.2010.00391.x.

Silver, M. 2011. "An Index Number Formula Problem: The Aggregation of Broadly Comparable Items." *Journal of Official Statistics* 27 (4): 1–17. Available at: https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/an-index-number-formula-problem-the-aggregation-of-broadly-comparable-items.pdf (accessed April 2021).

U.S. Bureau of Labor Statistics. 2020a. *Response Rate Report.* Office of Survey Method Research. Online only. Available at: https://www.bls.gov/osmr/response-rates/ (accessed April 2021).

U.S. Bureau of Labor Statistics. 2020b. "International Price Program." *Handbook of Methods*. Available at: https://www.bls.gov/opub/hom/ipp/home.htm (accessed April 2021).

Von der Lippe, P., and J. Mehrhoff, 2010. "Unit Value Bias Reconsidered." Contributions to the annual meeting of the Association for Social Policy. Okonomie der Familie – Session: How Reliable are German Aggregate Price Indices? No. F2-V3 Association for Social Policy, Frankfurt a.M. Available at: http://hdl.handle.net/10419/37526 (accessed April 2021).

Received May 2020
Revised January 2021
Accepted May 2021