# Measuring the Cost of Living in Mexico and the US*

**David Argente**†
Pennsylvania State University

**Chang-Tai Hsieh**‡
University of Chicago and NBER

**Munseob Lee**§
University of California San Diego

July 2021

**Abstract**

We use a dataset with prices and spending on consumer packaged goods matched at the barcode-level across the US and Mexico to measure the price index in Mexico relative to the US. Mexican prices relative to the US are 25% *lower* compared to the International Comparisons Project's (ICP) price index. We decompose the 25% gap into the biases from imputation, sampling, quality, and variety. Quality bias *increases* Mexican prices by 48%. Imputation, sampling, and variety bias *lowers* Mexican prices by 11%, 14%, and 34%, respectively.

*JEL Classification Numbers: E01, E31, O47*
*Keywords: Price Index, Purchasing Power Parity, International Price Comparison*

*We thank Robert Inklaar, Esteban Rossi-Hansberg and Ina Simonovska for helpful comments. Researchers own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. Churn Ken Lee, Jake Orchard, and Jihoon Sung provided excellent research assistance.

†Email: dargente@psu.edu. Address: 403 Kern Building, University Park, PA 16801.
‡Email: chsieh@chicagobooth.edu. Address: 5807 South Woodlawn Ave., Chicago, IL 60637.
§Email: munseoblee@ucsd.edu. Address: 9500 Gilman Drive #0519, La Jolla, CA 92093-0519.

# 1 Introduction

Indexes of prices across countries are a vital ingredient in estimates of standards of living and real output across countries. The most widely used price indexes are those by the International Comparison Program (ICP). The ICP collects prices of more than a thousand specific products ("items") in multiple countries which it then aggregates into price indexes of 155 broad product categories ("basic headings").

Despite their widespread use, it is well known that there are four potential biases in the prices provided by the ICP. The first is imputation bias. The ICP's surveyors are unable to collect the prices of many items. The ICP imputes prices for the items with missing prices using data from other countries. The second is sampling bias. The ICP calculates the price of an item as an average of store-level prices weighted by each store's *total* sales, but this procedure may not yield prices that reflect what consumers actually pay. Third, products differ in quality across countries, and it is possible that the ICP matches lower quality items in poor countries with higher quality ones in richer countries. Fourth, products available in one country are not available in other countries and no adjustment is made for potential differences in the availability of products across countries.

These potential biases have been brought up by many authors. For example, Deaton (2010) discusses how the ICP imputes missing prices, Deaton and Heston (2010) bring up the potential bias due to sampling and quality, and Feenstra, Xu and Antoniades (2020) address the issue of variety bias in the ICP. However, although these biases are well understood conceptually, we have very little evidence on their empirical magnitude. Our goal in this paper is to use a new dataset to measure these biases empirically for the specific case of the price index of nondurable goods in Mexico relative to the US.

Our data is the Nielsen Consumer Panel data for Mexico and the US (the Nielsen data). The Nielsen data collects data on spending on nondurables for 40-60 thousand households in the US and 6 thousand households in Mexico. Households in the two countries use in-home scanners or diaries to record their purchases of packaged goods. The Nielsen data includes information on prices and quantities of nondurable goods (identified by a 12-digit barcode) purchased by each household on each shopping trip and from each retail store visited. The nondurables in the Nielsen data account for 60-65% of total nondurable spending in the two countries.

The richness of the Nielsen data allows us to measure potential biases in the ICP for nondurable goods. First, the Nielsen data has better coverage of items across both countries within the basic headings it covers. As a result, we can construct a price index similar to the one produced by the ICP but that does not rely on price imputations. We find that the

ICP's imputation of missing prices overstates prices by 11% in Mexico compared to the US.

Second, the Nielsen data has information on prices and expenditures on all products for a representative sample of households in each country. The Nielsen data also identifies the retail store where each purchase was made. We can therefore construct two price indices, one using weights from household expenditures and another based on a store's total sales. Mexican prices (compared to the US) aggregated from consumer expenditure weights are 14% lower compared to prices aggregated from weights that reflect a store's total sales. This gap comes from the fact that Mexican households shop more frequently and are more likely to purchase only lower priced items in a given store compared to American households. Thus, using total sales in a store as weights overstates average prices in both countries, but more in Mexico compared to the US.

Third, the Nielsen data includes barcode information, which we use to identify more than 5,000 identical barcodes in the two countries. Mexican prices (compared to the US) among products with the same barcode are 48% *higher* compared to the price index of "comparable" products calculated following the ICP's methodology. This suggests that, as conjectured by Deaton and Heston (2010), the ICP may match low-quality products in low-income countries with high-quality products in high income countries.

Fourth, since we observe all the purchases made by households, we can measure the importance of products available to Mexican consumers but not to American consumers, and vice versa. We find that Mexican varieties missing in the US market matter more than US products missing in Mexico. When we take into account the differences in the availability of varieties in the two countries, effective prices in Mexico are 34% lower compared to the US. The net effect of all four adjustments, for imputation, sampling, quality, and variety, lowers Mexican prices relative to the US by 25% compared to the ICP.

This paper builds on recent work that use alternative micro-data to estimate price indexes across countries. Specifically, Cavallo, Diewert, Feenstra, Inklaar and Timmer (2018), Cavallo, Feenstra and Inklaar (2021), and Simonovska (2015) use online data and Feenstra, Xu and Antoniades (2020) use scanner data of toothpaste, laundry detergent, personal wash items and shampoo to estimate prices across countries. Our contribution is to bring new data to this long-standing question and to empirically measure multiple potential biases in the ICP using this new data. Specifically, our data contains detailed data on both prices and quantities on the *majority* of nondurables for a representative sample of *consumers*.

The four biases we document for the price index in Mexico relative to the US are also potentially present in measures of inflation. For example, Nordhaus (1997), Bils and Klenow (2001), and Bils (2009) document the bias in price indexes over time because quality improvements are not fully taken into account. Broda and Weinstein (2010) quantify the size of the variety bias in the US CPI over time. Coibion, Gorodnichenko and Hong (2015) studies the role of store-switching and Chevalier and Kashyap (2019) document the importance of price discrimination in the bias in the CPI's sampling weights. We are not aware of a study that measures the importance of all these biases in the price index over time taken together in one consistent dataset.[1]

The rest of the paper is organized as follows. Section 2 describes the data. In Section 3 we develop our price index. Section 4 presents our results and decomposes the gap between our price index and the ICP into the contribution of imputation, sampling, quality, and variety. The last section concludes.

# 2 Data Description

In this section, we describe the data collected by the ICP, the Nielsen data, and the matched data we constructed from the Nielsen data matched to the items covered in the ICP.

## 2.1 International Comparison Program (ICP)

The ICP is a statistical initiative that collects prices on more than thousand detailed products ("items") around the world. The prices of different items are aggregated into 155 broad product categories ("basic headings") that cover all the components of GDP. Approximately 53 basic headings refer to goods, of which 33 are non-durables. The publicly available data has price indices for the basic headings. We use the restricted 2011 ICP micro-data for Mexico and the United States, which has the prices at the item level.

The ICP specifies the characteristics of a "representative product" for each item. The specification of a representative product includes quantity and packaging (e.g. 250 milliliters of milk), source (e.g. produced domestically or imported), seasonal availability (e.g. year-round or only seasonal), product characteristics, and brand. For example, the representative product for the item "Baby Diapers" is a well-known brand, containing between 18 to 24 pieces, either classic or basic type, with a size between 4 and 9.5 kg, and with a multi-pack package.

Prices of the representative product of each item are collected from a sample of retail establishments chosen based on their total sales. Prices are collected from products in a retail establishment that meet the specifications of the representative products. The ICP

[1]There is also a literature that uses Engel's law to measure price indexes over space and time without the need of price data. See Hamilton (2001), Costa (2001), and Nakamura, Steinsson and Liu (2016), who quantify biases in the CPI of the US and China. Almås (2012) and Atkin, Faber, Fally and Gonzalez-Navarro (2020) apply this methodology to estimate welfare differences across income groups.

then calculates the average price of an item as a weighted average of the unit price of the representative product across all the sampled stores, where the weights are the total sales of each store.

There are two points to note about the ICP's sampling. First, we do not know the exact product chosen as the representative product in each store. Although the goal is to price the same product in multiple stores, the unavoidable problem is that stores differ in the products they sell. Therefore, it is likely that surveyors choose different products as the representative product of an item in different stores. Second, the weights used to aggregate prices in each store are calculated using the total sales of each store instead of the sales of an item in each store. The problem, as we will see later, is that total sales of a store as a share of total sales in all stores can be very different from weights calculated from the sales of an item in a store as a share of total sales of the item in all stores. For example, whole milk can account for a large share of sales in grocery stores but is relatively unimportant in sales for gas station convenience stores.

The ICP also collects national accounts expenditures for each basic heading but not for the items within each basic heading. These expenditures are used to aggregate basic headings into an aggregate price index. Within each basic heading, the ICP does not have expenditure weights at the item level. Instead, it classifies each item as "important" or "less-important." For the nondurables in Mexico and the US that we consider, the ICP classifies the majority of these items as "important."

## 2.2 Mexico and US Nielsen

The Nielsen data for the US tracks the shopping behavior of 40,000 to 60,000 households in 48 contiguous states plus Washington D.C. Each household uses in-home scanners to record their purchases. The US data contain slightly under one million distinct 12-digit barcodes. For each barcode, the data contains information on the brand, size, packaging, and other rich sets of product features. We combine the information on the price paid by the consumer with Nielsen's data on the products' characteristics to calculate the *unit price* of each product (e.g., price per ounce). In what follows, we use "price" as a shorthand for a product's unit price.

The data also contains information on each purchasing trip the panelist makes, including information on the retailer, the retailer's location, the date of the transaction, and the expenditures and prices of each barcode purchased in each store. Furthermore, the data have demographic variables such as age, education, annual income, marital status, and employment that are updated annually based on surveys sent to the households. Nielsen constructs projection weights that make the sample representative of the US urban population that we use in our calculations.

The Nielsen data for Mexico tracks the shopping behavior of 6,000 households for the years 2012-2013.[2] The sample is representative of all cities over 50,000 people and covers 55 cities in Mexico. Instead of using in-home scanners, households record their spending in diaries that are collected biweekly by Nielsen. Nielsen's data for Mexico contains approximately 55,000 distinct barcodes.

The Mexican data contains detailed information on each shopping trip (date, store, amount spent), transaction level information for each product purchased (quantity, price, deals, coupons), as well as detailed product level characteristics (brand, size, packaging, flavor). As in the US data, we calculate the unit price from the products' characteristics and the price paid by the consumer. The data also include demographic variables at the household level such as the occupation of the household members, education, age, and family size. As in the case of the US, Nielsen constructs projection weights that make the sample representative of the Mexican urban population that we use in our calculations.

## 2.3 Nielsen Matched to ICP Items

We use the ICP's definition of an item along with Nielsen's description of the characteristics of each barcode to assign barcodes to items in the ICP. We match the barcodes in Nielsen in Mexico and the US to 71 ICP items in 18 basic headings.[3] These 71 items account for 60% of aggregate expenditures on non-durables in Mexico and 65% in the US. The basic headings of nondurables in the ICP not covered by the Nielsen data are products without barcodes such as fresh meats and fruits.

In our matched sample of 71 ICP items, there are 45 items whose prices are collected by the ICP in Mexico and the US. Table 1 lists these 45 items along with the number of

[2]Since the ICP sampling is conducted every 6 years, and the Mexican scanner data is only available from 2012-2013, we focus on the cross-sectional patterns of the 2011 ICP estimates.

[3]There are a total 117 items in the ICP in the 18 basic headings we match with the Nielsen data. We aggregated the ICP items that we cannot distinguish from the barcode description in the Nielsen data. For example, the description of the characteristics of barcodes of breads do not distinguish between "White Bread (Not Sliced)" and "Sliced White Bread" (two distinct items in the ICP) so we aggregate these two ICP items into one. This aggregation reduces the number of ICP items we match to Nielsen to 104. The "new" items after we aggregate are: White Bread (not sliced and sliced), Dried/Instant Noodles (dried and instant noodles), Milk, un-skimmed (pasteurized and ultra-pasteurized), Cheese (cheddar, processed, camembert, and gouda), Tomato Paste (small and large), Ice Cream (cornetto-type and packed), Black Tea (bags and loose leaves), Canned Beer (domestic and imported), Coffee Roasted (Arabica and Robusta), and Detergent Powder (washing machine and hand wash). Among these 104 items, 6 of them are not covered by Nielsen and 27 are priced only in the US Nielsen data but not in the Mexican data. We end up with a total of 71 ICP items for which we have Nielsen prices in Mexico and the US. We find no significant difference in ICP's Mexican prices (compared to US prices) between matched and unmatched items.

barcodes in Nielsen we classify under each item. The 2011 ICP imputes the prices for Mexico and/or the US for the remaining 26 items. These items are listed in Table 2, again with the number of Nielsen barcodes we match to each item.

We also match barcodes in the US and Mexico, using the fact that the two countries use the same barcode system.[4] The third row in Table 1 and 2 shows the number of matched barcodes for each ICP item. It is clear that the number of common barcodes is small share of the barcodes sold in each country. The majority of barcodes sold in Mexico are not sold in the US, and vice versa.

[4]Both countries use the Universal Product Code (UPC), which consists of 12 numeric digits. The organization that assigns these barcodes (i.e. GS1), is present in over 115 countries.

## Table 1: Items with Prices in the ICP

| | # barcodes MX | # barcodes US | # common |
|---|---|---|---|
| Cornflakes | 1312 | 5813 | 263 |
| White Bread | 159 | 1683 | 42 |
| Roll | 77 | 4662 | 16 |
| Sandwich biscuits/cookies | 568 | 709 | 3 |
| Butter biscuits | 18 | 26 | 0 |
| Flavored biscuits/cookies sweet | 1803 | 13124 | 135 |
| Spaghetti | 147 | 1971 | 21 |
| Dried/Instant Noodles | 401 | 1169 | 60 |
| Canned tuna without skin | 328 | 864 | 5 |
| Milk, un-skimmed | 763 | 1521 | 15 |
| Milk, low-fat, Pasteurized | 46 | 3486 | 9 |
| Yoghurt, plain | 1210 | 6207 | 13 |
| Sour cream | 206 | 909 | 5 |
| Cheese | 1445 | 15713 | 99 |
| Cream cheese | 72 | 1113 | 19 |
| Potato chips | 725 | 5043 | 118 |
| Tomato Paste | 104 | 1022 | 20 |
| Tinned sweet corn/Maize | 186 | 877 | 37 |
| White sugar | 360 | 375 | 3 |
| Strawberry/Apricot Jam | 164 | 901 | 12 |
| Orange marmalade | 10 | 204 | 2 |
| Chocolate bar | 882 | 3406 | 75 |
| Ice Cream | 547 | 8552 | 93 |
| Cooking salt | 235 | 1468 | 13 |
| Tomato ketchup | 167 | 627 | 31 |
| Chicken Extract (bouillon/stock cube) | 258 | 208 | 11 |
| Baby food | 45 | 1148 | 4 |
| Cocoa Powder, Tin | 201 | 558 | 8 |
| Instant coffee | 362 | 463 | 31 |
| Black Tea | 15 | 2307 | 0 |
| Mineral water | 920 | 4095 | 43 |
| Carbonated Soft Drink (Small) | 267 | 4376 | 47 |
| Carbonated Soft Drink (Large) | 1495 | 7669 | 41 |
| Apple juice | 204 | 1742 | 27 |
| Canned Beer | 231 | 1025 | 22 |
| Bottled Beer | 289 | 2196 | 23 |
| Detergent Powder | 821 | 710 | 41 |
| Liquid Window Cleaner | 70 | 356 | 4 |
| Kitchen paper roll | 481 | 1567 | 57 |
| Dishwashing detergent | 189 | 1405 | 57 |
| Tooth paste, tube | 420 | 1275 | 75 |
| Shower gel | 719 | 7466 | 133 |
| Regular sanitary pad/napkin | 554 | 1341 | 20 |
| Shampoo | 2164 | 4321 | 199 |
| Toilet paper - Multipack | 1029 | 2042 | 49 |

Note: The table reports the number of barcodes in Mexico and the US in the 45 items whose prices are collected by the ICP.

Table 2: **Items with Imputed Prices in the ICP**

| | # barcodes MX | # barcodes US | # common |
|---|---|---|---|
| Wheat Semolina (Suji) | 8 | 2 | 0 |
| Oats, rolled | 222 | 1898 | 19 |
| Whole wheat bread | 192 | 1988 | 65 |
| Pita bread | 2 | 390 | 0 |
| Salted crackers | 225 | 2961 | 32 |
| Short pasta | 210 | 50 | 0 |
| Vermicelli (Angel Hair) | 8 | 30 | 0 |
| Macaroni | 54 | 3307 | 23 |
| Milk, condensed | 56 | 132 | 2 |
| Milk, powdered | 140 | 567 | 2 |
| Green Olives (with stones) | 155 | 2354 | 22 |
| Tinned green peas | 71 | 709 | 4 |
| Tinned Button Mushrooms | 101 | 618 | 4 |
| Brown sugar | 17 | 652 | 1 |
| Pineapple Jam | 20 | 72 | 0 |
| Natural honey, Mixed blossoms | 198 | 2721 | 23 |
| Thin Soya Sauce | 73 | 337 | 13 |
| Chili sauce | 443 | 3095 | 33 |
| Coffee Roasted | 319 | 5574 | 43 |
| Tea, green | 12 | 1281 | 0 |
| Orange juice | 285 | 2634 | 33 |
| Lemonade | 52 | 1029 | 3 |
| All purposes household cleaner | 1352 | 4212 | 59 |
| Deodorant roll-on for men | 638 | 119 | 13 |
| Toilet Soap | 910 | 1850 | 71 |
| Baby diapers | 957 | 2399 | 45 |

Note: The table reports the number of barcodes in Mexico and the US in the 26 items where the ICP imputes the price of the item in Mexico or the US (or in both countries) .

# 3 An Exact Price Index

This section derives the ideal price index. We also describe how we use the Nielsen data to construct a price index that mimics the procedure followed by the ICP.

## 3.1 Exact Price Index

Following the ICP, we assume that utility is a function of consumption in the basic headings, and that basic headings are in turn an aggregate of expenditures of specific items.

Specifically, the utility of a representative household in a country is given by:

$$\mathbb{U} = \left[ \sum_b \left( \sum_i C_{ib}{}^{\frac{\eta_b - 1}{\eta_b}} \right)^{\frac{\eta_b}{\eta_b - 1} \frac{\gamma - 1}{\gamma}} \right]^{\frac{\gamma}{\gamma - 1}} \tag{1}$$

where $C_{ib}$ is consumption of item $i$ in basic heading $b$. When possible, we omit the country index in the notation. The parameters $\gamma$ and $\eta_b$ denote the elasticity of substitution across basic headings and items within basic headings, respectively. The set of basic headings and items are the same for all countries in the ICP. Therefore, the indexes for basic headings and items do not carry an index for the country.

The ICP assumes that an item consists of a single representative product, but there are multiple barcodes that satisfy the ICP definition of the representative product of an item. For example, there are 463 barcodes in the US and 362 in Mexico that meet the ICP's definition of the representative product for "instant coffee" (see Table 1). To account for the varieties in each item, we assume consumption of an item in a country (say Mexico, indexed by $m$) is itself a CES aggregate of individual *barcodes* indexed by $k$:

$$C_{ib}^m = \left( \sum_{k \in m} \left( \varphi_{kib} C_{kib}^m \right)^{\frac{\sigma_{ib} - 1}{\sigma_{ib}}} \right)^{\frac{\sigma_{ib}}{\sigma_{ib} - 1}}. \tag{2}$$

Here $C_{kib}^m$ denotes total physical units and $\varphi_{kib}$ the quality of barcode $k$, where the parameter $\sigma_{ib}$ denotes the elasticity of substitution between the barcodes. Note that the aggregation over the barcodes is only over the set available in each country.[5]

The exact price index of item $i$ in basic heading $b$ in Mexico, taking the US (indexed by $u$) as the numeraire, is then given by:

$$\text{EPI}_{ib} = \left( \frac{\lambda_{ib}^m}{\lambda_{ib}^u} \right)^{\frac{1}{\sigma_{ib} - 1}} \times \prod_{k \in \text{common}} \left( \frac{p_{kib}^m}{p_{kib}^u} \right)^{\omega_{kib}} \tag{3}$$

The first term in equation 3 is the ratio of the share of total spending on the common barcodes in Mexico relative to the US, where the spending shares are defined as:

$$\lambda_{ib}^m \equiv \frac{\sum_{k \in \text{common}} p_{kib}^m C_{kib}^m}{\sum_{k \in m} p_{kib}^m C_{kib}^m} \quad \text{and} \quad \lambda_{ib}^u \equiv \frac{\sum_{k \in \text{common}} p_{kib}^u C_{kib}^u}{\sum_{k \in u} p_{kib}^u C_{kib}^u}$$

[5]In Appendix A we extend the utility function at the item level to allow for non-homothetic preferences. There we show that the bias in the ICP price index between Mexico and the US is almost the same when we account for non-homothetic preferences as in our baseline case with homothetic preferences.

The second term is the geometric mean of the ratio of the price in Mexico relative to the US of barcode $k$ *for the barcodes common to the two countries*, weighted by the logarithmic mean of the expenditure shares of the barcode.[6] The price of a barcode in equation 3 is a weighted average of the price of the same barcode in all the stores in a country

$$p^m_{kib} = \prod_{s \in m} (p^m_{skib})^{\phi^m_{skib}} \quad \text{and} \quad p^u_{kib} = \prod_{s \in u} (p^u_{skib})^{\phi^u_{skib}}$$

where $p^m_{skib}$ is the price of barcode $k$ in store $s$ (in Mexico) and the weights are the share of spending on each barcode in the store.[7]

The exact price index of item $i$ in basic heading $b$ requires at least one common barcode between Mexico and the US. For 63 of the 71 items in the matched data, we observe at least one common barcode. For the remaining 8 items that do not have a common barcode, we account for these items with a variety correction term at the item level. Specifically, define $I_C$ as the set of items with common barcodes. The exact price index for the basic heading can be defined as:

$$\mathbb{EPI}_b = \left(\frac{\lambda^m_b}{\lambda^u_b}\right)^{\frac{1}{\sigma_b - 1}} \times \prod_{i \in I_C} \mathbb{EPI}_{ib}^{\omega^*_{ib}} \tag{4}$$

where $\omega^*_{ib}$ is logarithmic mean of the expenditure share of each item and the spending shares are defined as:

$$\lambda^m_b \equiv \frac{\sum_{i \in I_C} p^m_{ib} C^m_{ib}}{\sum_i p^m_{ib} C^m_{ib}} \quad \text{and} \quad \lambda^u_b \equiv \frac{\sum_{i \in I_C} p^u_{ib} C^u_{ib}}{\sum_i p^u_{ib} C^u_{ib}}$$

The second term in equation 4 is the weighted geometric average of the price index of the items that can be priced from equation 3. The first term in equation 4 is ratio of the spending shares on items with common barcodes in the two countries. This term captures the weighted geometric average of the price index of the items that can *not* be priced from equation 3.

Finally, the aggregate exact price index is a weighted geometric average of the exact price index of each basic item:

$$\mathbb{EPI} = \prod_b EPI_b^{\omega_b} \tag{5}$$

[6] The logarithmic mean is $\omega_{kib} \equiv \frac{s^m_{kib} - s^u_{kib}}{\ln s^m_{kib} - \ln s^u_{kib}} \Big/ \sum_{k \in \text{common}} \frac{s^m_{kib} - s^u_{kib}}{\ln s^m_{kib} - \ln s^u_{kib}}$ where $s^m_{kib} \equiv \frac{p^m_{kib} C^m_{kib}}{\sum\limits_{k \in m} p^m_{kib} C^m_{kib}}$ and $s^u_{kib} \equiv \frac{p^u_{kib} C^u_{kib}}{\sum\limits_{k \in u} p^u_{kib} C^u_{kib}}$

[7] Note that in the data we observe multiple prices for the same barcode purchased from multiple stores. Appendix B shows that the ideal price of a barcode can be derived as a CES price index from a consumer's discrete choice problem and that a first order approximation to this index is the Cobb-Douglas price index.

where $\omega_b$ is the logarithmic mean of the expenditure shares of each basic heading and $\mathbb{EPI}_b$ is given by equation 4.

## 3.2 A "Pseudo-ICP" Price Index

There are two differences between the ideal price index we will calculate and the ICP's price index. First, the underlying data (the Nielsen data) is different from the data used by the ICP. Second, there is the difference in methodology. To isolate the effect of methodology, we use the Nielsen data to construct a price index that mimics the ICP. We call this a "Pseudo-ICP" price index.

We proceed in four steps. First, for each ICP item, we identify the set of barcodes in each store that meet the ICP's specifications for the representative product.[8] From these set of barcodes, we pick the barcode with the largest volume of sales in each store as the representative product of the item in the store. The representative product of an item is therefore not the same in each store. We obviously do not know the exact product in each store surveyed by the ICP's price surveyors, but we believe this procedure is a good approximation of how the ICP chooses the representative product of an item in each store. The price of an item is then the geometric mean of the price of the store-specific representative product weighted by the store's *total* sales.

Second, the ICP does not collect prices for 26 items in Table 2. Instead it imputes the prices for these items using data from other countries. We estimate the imputed prices by comparing the basic heading price indexes reported by the ICP with the geometric mean of the item-level prices they do collect. If, for example, a basic heading price reported by the ICP is higher than the geometric mean of item-level observed prices, we infer that the imputed prices for the average item with a missing price are higher than those of other items in the same basic heading.

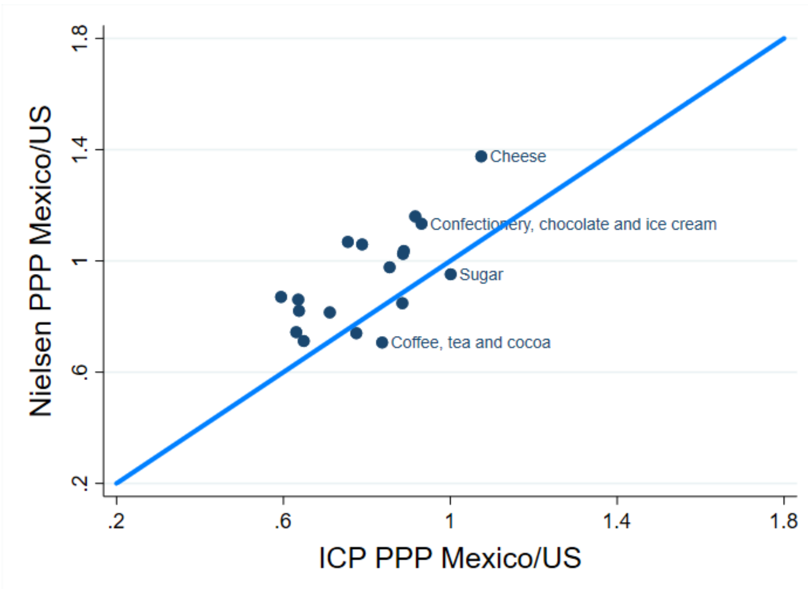
Third, the price index of a basic heading is an equally weighted geometric average of the price of the items in each basic heading.[9] Figure 1 shows the scatterplot of the Pseudo-ICP price index of a basic heading (Mexico relative to the US) vs. the price index published by the ICP. The correlation is not one but the pseudo ICP price index generally aligns with the ICP's price index. The median of the pseudo ICP price index at the basic heading level is 0.91 whereas the median of the actual ICP price index is 0.81. The regression coefficient is 1.16 and statistically significant at the 1 percent level. At the item level, the median of the

[8] We also mimic the fact that the ICP survey only covers four cities in Mexico (Mexico City, Guadalajara, Monterrey, and Puebla), while it covers all urban areas in the US.

[9] In principle the ICP classifies items as "important" and "less important" and gives "important" items more weight when aggregating to the basic heading level. However, most of the nondurable ICP items in Mexico and the US that we match to the Nielsen data are classified as "important."

pseudo ICP price index is 0.84 whereas the median of the numbers published by the ICP is 0.72. A linear regression across the prices at the item level yields a coefficient is 1.13 and is also statistically significant at the 1 percent level.

**Figure 1: Pseudo vs. Actual ICP Price Index**




Notes: The figure plots the pseudo-ICP index of each 18 basic heading against the index published by ICP.

The final step is to aggregate the price of a basic heading into an aggregate price index. The ICP calculates the geometric average of the price index of a basic heading across all basic headings. In the case of the price index of Mexico vs. the US, the price indexes of the basic headings are first averaged using Mexico's weights, then averaged using the US weights, and then the geometric mean of the two is taken. The result is a Fisher index, which facilitates multilateral comparisons since such indexes are transitive (i.e. price comparisons between two countries are the same whether it is computed directly or indirectly through a third country) and base country-invariant (i.e. price comparisons between two countries are the same regardless of the choice of base country). Since we are only comparing two countries, we aggregate the price index at the basic heading level using the logarithmic weight of each basic heading. This is the theoretically consistent way to aggregate prices with a CES utility function.

# 4 The Price of Nondurables in Mexico vs. the US

In this section, we calculate the exact price index from the Nielsen data. We decompose the gap between the exact and the pseudo-ICP price index into the bias due to imputation, sampling, quality, and variety.

We need the elasticity of substitution between barcodes within each item. We follow Feenstra (1994)'s procedure as extended by Broda and Weinstein (2006) and Broda and Weinstein (2010). The procedure consists of estimating a demand and supply equation for each barcode using the information on prices and quantities of each barcode from the Nielsen data.[10] The average of the elasticity of substitution across barcodes we obtain is 6.48 with standard deviation of 3.09.[11]

We use these estimates of the elasticity of substitution between barcodes, along with the Nielsen data, to estimate the exact price index of each item from equation 3. We then use equation 4 to aggregate the price of an item using logarithmic weights of each item in a basic heading, which we then aggregate into an aggregate price index using logarithmic weights for the basic heading as equation 5. Column 1 in Table 3 shows that the exact price index is 0.64, which implies that Mexican prices are 36% lower than in the US.

The second column shows the pseudo-ICP price index. Here we take the weighted average of the pseudo ICP price of the basic heading (shown in figure 1) using the same logarithmic weights for each basic heading that we used to aggregate the exact price index of each basic heading. The aggregate pseudo-ICP price index, shown in column 2 in Table 3, is 0.86 which suggests that Mexican prices are only 14% lower than in the US. The exact price index calculated from the same data suggests that Mexican prices are 36% lower, which indicates an aggregate bias of 25%.

**Table 3: Cost of Living in Mexico vs. the US**

| Aggregate Price Index | | Bias due to: | | | | Aggregate |
|---|---|---|---|---|---|---|
| Exact | Pseudo-ICP | Imputation | Sampling | Quality | Variety | Bias |
| 0.64 | 0.86 | 0.89 | 0.86 | 1.48 | 0.66 | 0.75 |

Notes: The table reports the aggregate exact price index, pseudo ICP price index, and the gap between the exact and ICP index due to imputation, sampling, quality, and variety. Aggregate bias is the product of the bias due to imputation, sampling, quality, and variety.

[10] See Appendix C for more details.

[11] We check the robustness of these results by calculating price indexes and biases with the common elasticity of substitution, 6. The estimated exact price index (0.64) is the same at two decimal points with the item-specific elasticity of substitution (0.64). See Appendix D for more details.

Recall that we use the same logarithmic weights at the basic heading level to aggregate the exact and pseudo ICP price indexes for each basic heading. Therefore, the difference between the aggregate exact and the ICP price index in Table 3 comes from the price index at the basic heading level. In turn, the gap between the two indexes at the basic heading level is the product of the biases due to sampling, quality, and variety for each basic heading:

$$\frac{\mathbb{EPI}_b}{\mathbb{ICP}_b} = \text{Imputation Bias}_b \times \text{Sampling Bias}_b \times \text{Quality Bias}_b \times \text{Variety Bias}_b$$

We now quantify each of these biases.

**Imputation Bias:** Out of the 71 items in our sample, the ICP does not collect prices for 26 of them from either Mexico or the US. The ICP runs a set of country-product dummy (CPD) regressions to impute the missing values for items whose prices cannot be found (Deaton, 2010). We revisit this imputation and quantify the size of the bias. As discussed in the previous section, we back out imputed prices by comparing a basic heading price reported by the ICP and geometric mean of item-level prices within a basic heading.

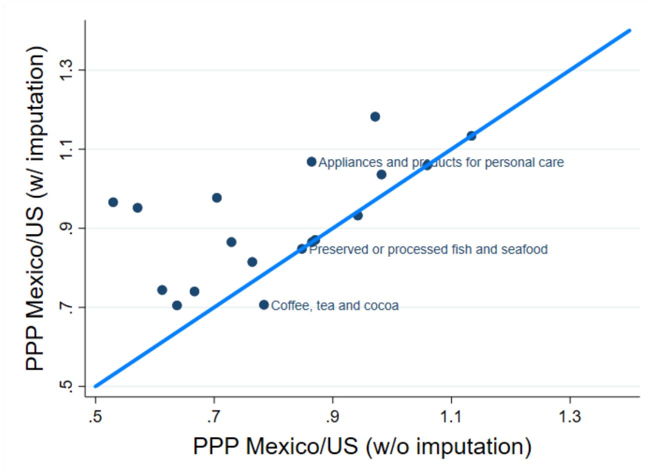
The imputation bias is defined as follows:

$$\text{Imputation Bias}_b \equiv \left[\prod_i \left(\tilde{p}^m_{ib}/\tilde{p}^u_{ib}\right)^{\frac{1}{N_b}}\right]/\mathbb{ICP}_b \tag{6}$$

where $\mathbb{ICP}_b$ is the pseudo ICP that uses imputed prices for items not collected from either Mexico or the US and $\tilde{p}_{ib}$ is the price of the item computed from the Nielsen data. There is no imputation bias if the ICP collects prices for all the items.

Figure 2 plots the ratio of US to Mexican prices of a basic heading with imputed item level prices on the y-axis against the price ratio calculated from the Nielsen data on the x-axis. The observations on the 45-degree line are those for which the ICP has prices for all items within the basic heading. For these basic headings the imputation bias is one. However, on average the majority of the observations are above the 45-degree line: the ratio of Mexican to US prices with imputed item level prices is typically higher than the price ratio calculated from actual price data.

**Figure 2: Imputation Bias for Each Basic Heading**




Notes: The figure plots the PPP with and without item-level imputation for each of the 18 basic headings.

Table 3 shows the net effect of imputation on the aggregate price index. Specifically, the third column shows the imputation bias calculated from equation 6. Imputation lowers Mexican prices by 11% compared to the ICP's index.

**Sampling Bias:** There are three sources of sampling bias. First, the ICP samples items only from the four cities in Mexico (i.e. Mexico City, Guadalajara, Monterrey, and Puebla). Second, at the item level, the ICP uses an equal-weights geometric average because they lack data on expenditures at the item level. Third, the ICP aggregates prices from each store using total sales of the store instead of the sales of the item in the store as weights. The sampling bias is the product of these three biases:

$$\text{Sampling Bias}_b \equiv \left[\prod_i \left(\frac{\bar{p}^m_{ib}/\bar{p}^u_{ib}}{\tilde{p}^m_{ib}/\tilde{p}^u_{ib}}\right)^{\frac{1}{N_b}}\right] \times \left[\frac{\prod_i \left(\frac{\bar{p}^m_{ib}}{\bar{p}^u_{ib}}\right)^{\omega_{ib}}}{\prod_i \left(\frac{\bar{p}^m_{ib}}{\bar{p}^u_{ib}}\right)^{\frac{1}{N_b}}}\right] \times \left[\prod_i \left(\frac{\hat{p}^m_{ib}/\bar{p}^m_{ib}}{\hat{p}^u_{ib}/\bar{p}^u_{ib}}\right)^{\omega_{ib}}\right] \tag{7}$$

where $\tilde{p}^m_{ib}$ is the average price of an item computed as the weighted average of store level prices, where the weights are the total sales of the store considering only the main cities in Mexico, $\bar{p}^m_{ib}$ is the same weighted average considering all cities and $\hat{p}^m_{ib}$ is the weighted average where the weights are the sales of the particular item in the store.

The first term denotes the bias from the geographical coverage of the ICP survey in Mexico. By sampling only from the four major cities in Mexico, the ICP underestimates the average price of items by approximately 1%.[12] The four cities the ICP samples in Mexico

[12]Consistent with the coverage of our data, the ICP samples items from urban areas in all US states (i.e.

are among the largest in terms of population and their average household income is close to that of the median city in the country (i.e. choosing these four cities is close to choosing a representative city for Mexico.) This, combined with the fact that the ICP only samples urban cities in the US as well, yields a small geographic bias.

The second term in equation 7 is the sampling bias discussed by Deaton and Heston (2010). It denotes the bias from using the simple average instead of the weighted average of the item level price, and its magnitude depends on the covariance of the expenditure weights of each item and its price.[13] The covariance is negative if people spend more on items with lower prices. In the Nielsen data, the difference in these covariances between Mexico and the US is essentially zero. Thus, at least in the case of Mexico relative to the US, there is little bias from not using expenditure weights at the item level.

The third term in equation 7 is the bias from aggregating store-level prices using the total sales of the store as weights instead of the sales of the item in the store. For example, a fruit vendor may also sell milk, but the vendor's total sales depend mostly on fruit sales instead of milk. The resulting bias in the price index depends on whether consumers in Mexico purchase more products at stores where these products are cheaper compared to consumers in the US.

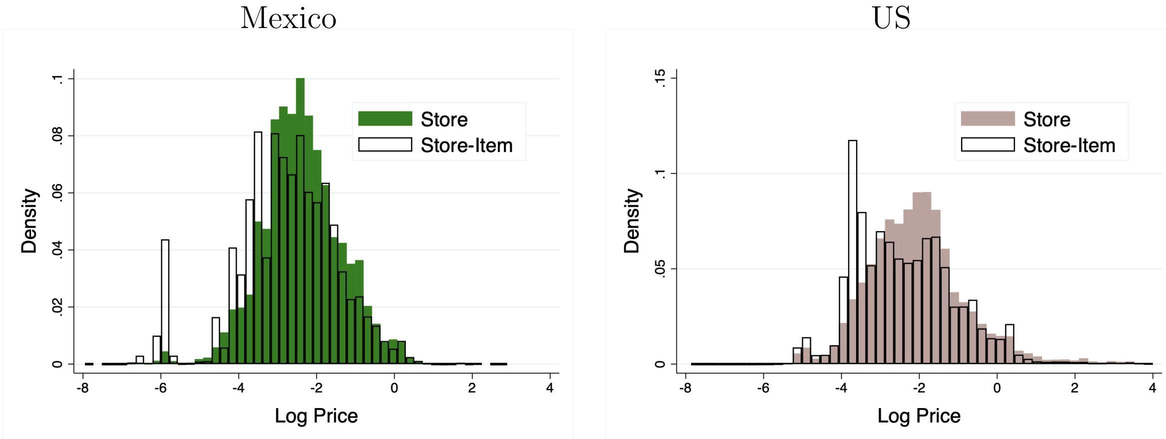
Figure 3 shows that this is indeed the case. It plots the distribution of prices for each item where the items' average price is averaged over the store specific price using the stores' total sales and the sales of each item in the store. For both countries, the distribution weighted using store-item weights is shifted to the left compared to the distribution using store level weights. Furthermore, the gap between the two distributions is larger for Mexico compared to the US. Weighting store specific prices using a store's total sales overstates the average price paid by consumers in the two countries but more so in Mexico.

---

$\tilde{p}^u_{ib} = \bar{p}^u_{ib}$).

[13] Proposition 1 in the Appendix E formally shows this relationship and empirically estimates the covariance using a reduced form approach.

**Figure 3: Distribution of Prices using Store vs. Store-Item Weights**




Notes: The figure shows the distribution of the price of an item calculated as a geometric average of the price of the item at the store level using either the store's total sales or the sales of the item in the store as weights.

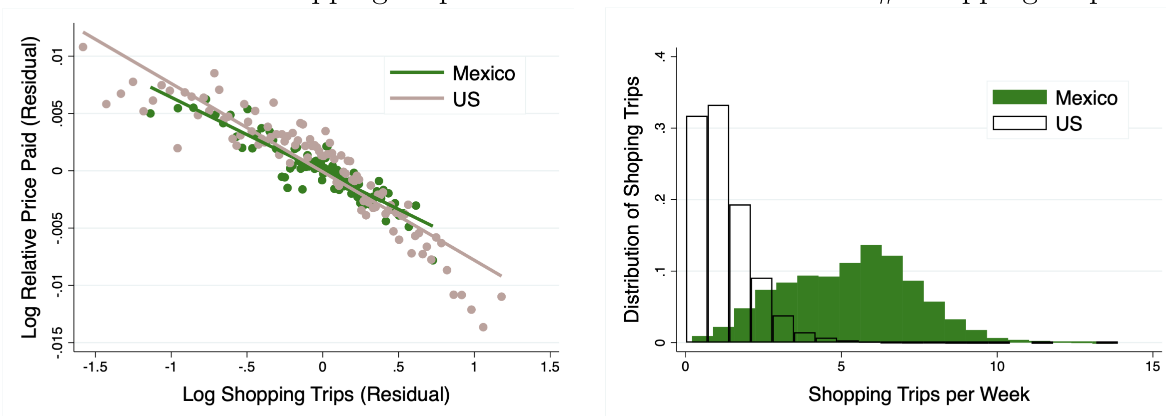
Figure 4 shows why the gap between the two price distributions is larger for Mexico. The left panel in Figure 4 plots the demeaned price of a barcode paid by a household in a shopping visit against the number of shopping visits of the household in the Nielsen data.[14] Households in Mexico and the US that shop more frequently pay lower prices for the same product compared to households that shop less frequently. The right panel in Figure 4 shows the distribution of the number of shopping trips per household. A typical US household makes one shopping trip per week whereas in Mexico the number is five. The average Mexican household shops more intensely than the typical American household, and thus buys more of the cheaper products in a given store.

Table 3 shows the effect of the difference in shopping behavior between Mexico and the US on the price index. Specifically, the third column shows the weighted mean of the sampling bias at the basic heading level calculated from equation 7, where the weights are the logarithmic share of each basic heading. The difference in shopping intensity between Mexico and the US lowers Mexican prices by 25% compared to the ICP because the ICP's sampling weights ignores the effect of shopping intensity.

[14]The number of shopping visits is the number of retail outlets a household makes purchases from in a week.

**Figure 4: Shopping Intensity and Prices in Mexico and the US**

Prices and Shopping Trips — Distribution of # Shopping Trips




Notes: Left panel plots the demeaned price paid for a product vs. the demeaned number of shopping trips. The demeaned product price is the residual from a regression of the log price of a barcode on category, store, and quarter fixed effects. The demeaned number of shopping trips is the residual of the log number of shopping trips on category, store, and quarter fixed effects. Right panel shows the distribution of the number of shopping trips per household. Number of shopping trips is the number of stores visited by a household in a week.

**Quality Bias:** In 63 out of the 71 items in our matched sample, there is at least one common barcode sold in the two countries. With at least one common barcode within an item, we can calculate quality and variety biases separately.[15] The quality bias in the ICP reflects the effect of matching high-quality goods in one country with a low-quality good in another country. We measure quality bias of an item as:

$$\text{Quality Bias}_{ib} \equiv \prod_{k \in \text{common}} \left( \frac{p^m_{kib}}{p^u_{kib}} \right)^{\omega_{kib}} \Big/ \left( \frac{\hat{p}^m_{ib}}{\hat{p}^u_{ib}} \right) \tag{8}$$

The denominator in equation 8 is the price gap between Mexico and the US at the item level computed as the average price of the “representative product” in each store. The numerator is the price gap where the “representative product” in a store is chosen from the set of barcodes sold in both Mexico and the US.[16] Quality bias is greater than one if

[15] The relative magnitude of the variety bias and the quality bias could be affected if we identify too few barcodes in common, particularly if there are a significant number of products among the unmatched barcodes that are in fact the same products. Nonetheless, the aggregate bias is the *product* of the variety and the quality biases. Therefore, as long as products with the same barcode are in fact the same product in the two countries, the aggregate bias is not affected by the possibility that some of the unmatched products may in fact be the same product.

[16] Both the numerator and the denominator in equation 8 weights the price of the item in each store by the sales of the item in the store.

the representative product of an item in a store in Mexico are of lower quality than the representative products of the same item in the US.

Table 4 shows that this is indeed the case. The first row follows the ICP's methodology and chooses one representative product per store to calculate the average price of an item. The first column shows the average percent difference between Mexico and the US in the price of an item, which is 28.0%. The second column picks the representative product of an item in the store chosen from the barcodes common to the two countries. The average price gap using products with same barcode is only 2.0%. The second row provides a similar number but where the average price of an item is calculated as a weighted average of the price of *all* the barcodes in the item. The first column considers all the barcodes in an item; the second column only considers barcodes sold in the two countries. Average prices calculated from all the barcodes are 42.1% lower in Mexico. The second column restricts to barcodes sold in the two countries. Here, the average price in Mexico is 1.5% lower than in the US.

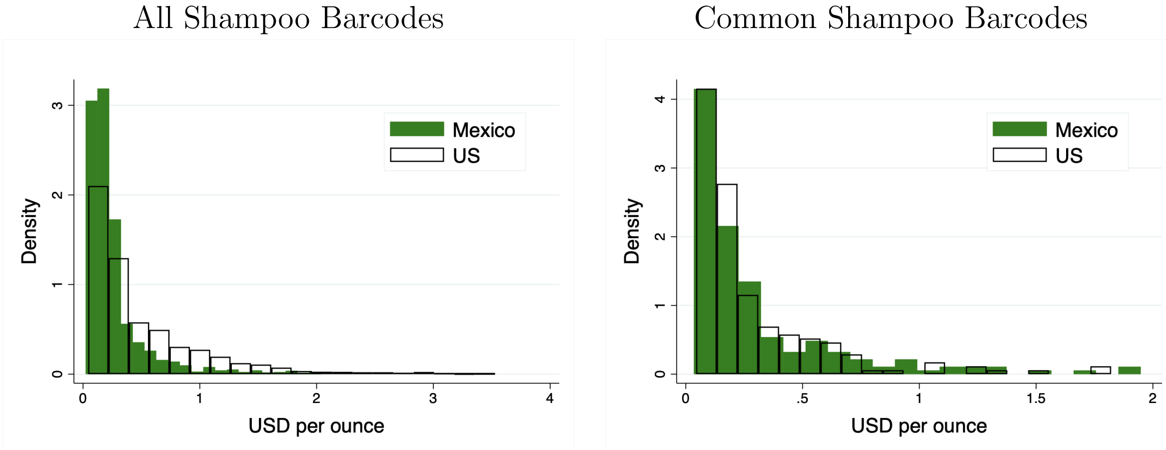
**Table 4: Price Gap between Mexico and US for All vs. Common Barcodes**

| Sampling | All barcodes | Common barcodes |
|---|---|---|
| Representative product | -28.0 % | -2.0 % |
| All products in item | -42.1 % | -1.5 % |

Notes: The table reports average percent difference in item-level prices between Mexico and the US. We take the weighted average with item-level expenditure weights. Column 1 considers all barcodes available in the two countries within each item. Column 2 restricts the sample to barcodes sold in Mexico and the US. Row 1 chooses one representative product per store. Row 2 considers all the barcodes and calculates the weighted average of the prices of chosen products.

Figure 5 shows the distribution of the price of shampoo. The left panel shows distribution of the unit price of shampoo in the US and Mexico for all shampoo barcodes. The right panel shows the distribution of the unit price of shampoo only for shampoo barcodes sold in both countries. As can be seen, the average shampoo in the US is more expensive than in Mexico, but the difference is much smaller for shampoos sold in both countries.

**Figure 5: Price Distribution for All and Common Shampoo Barcodes**




Notes: Left panel shows the price distribution of barcodes of all shampoo products in the US and Mexico. Right panel plots the price distribution of shampoo products sold in both countries.

Table 3 shows the aggregate effect of the patterns shown in Figure 5 and Table 4 on aggregate quality bias (for the 63 items for which this calculation is possible). The effect is large: quality bias *increases* Mexican prices by 48% relative to the US.

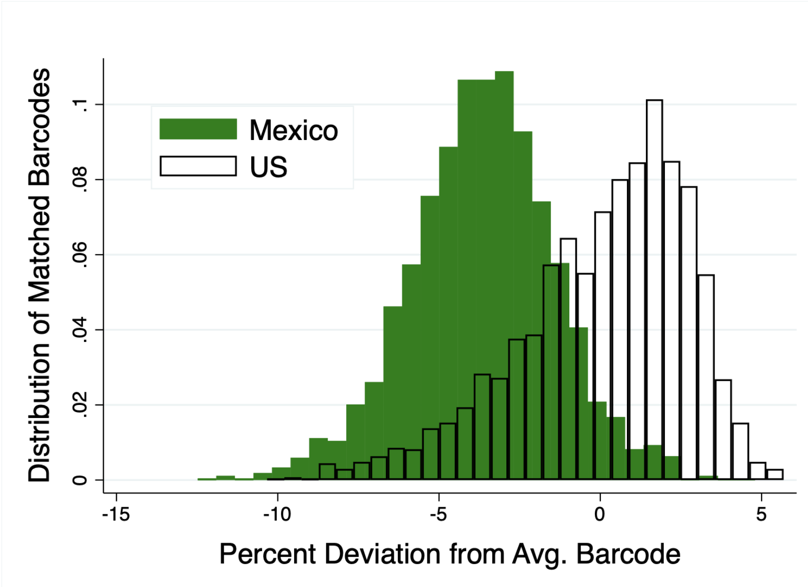
**Variety Bias:** We have already seen that the majority of Mexican barcodes are not available in the US market, and vice versa. The ICP price index is biased due to missing varieties if the importance of Mexican barcodes not available in the US is different from the importance of American barcodes not available in Mexico. We measure variety bias (of an item) by:

$$\text{Variety Bias}_{ib} \equiv \left(\frac{\lambda_{ib}^{m}}{\lambda_{ib}^{u}}\right)^{\frac{1}{\sigma_{ib}-1}}. \tag{9}$$

This is the ratio of the expenditure share of common barcodes in Mexico relative to the expenditure share of common products in US as in Feenstra (1994).

Figure 6 plots the distribution of share of spending on common barcodes relative to the spending share on the average barcode in the country. US households spend more on typical common barcodes while the opposite is true in Mexico. This means that missing Mexican varieties are more important compared to missing American varieties.

**Figure 6: Distribution of Revenue of Barcodes Sold in Both Countries**




Notes: The figure shows the distribution of revenue generated by the common barcodes relative to the average product in each country.

Barcodes that are present in Mexico and not present in the US are both very popular in Mexico (in terms of expenditures) and also very cheap relative to those sold in both countries. In the case of cereals, for instance, unmatched barcodes in Mexico are mainly produced by Mexican firms at a significantly lower price per unit. Some examples of cereal brands present in Mexico and not in the US are shown in Figure 7.[17]

[17]The first cereal on the top left corner of Figure 7 is "Glacs," which is produced by a Mexican firm. Glacs is a cereal similar to "Frosted Flakes" produced and sold by Kellogg (named "Zucaritas" in Mexico). On July 15 2021, the price of Glacs was 41.90 pesos for 720g, according to the website of the third largest retailer in Mexico. The price of Zucaritas was 69.5 pesos for 710g. As a child, one of the authors of this paper had the local Mexican cereal for breakfast instead of the more expensive US brands.

**Figure 7: Examples of Cereal Brands Present in Mexico and Not in the US**


Notes: The figure shows eight examples of cereal brands present in Mexico and not in the US. These brands are produced and sold by Mexican firms.

Table 3 summarizes the effect of variety bias on the price index (for items with common barcodes in the two countries). We estimate that varieties only available to Mexican consumers lowers the effective prices paid by Mexican consumers by 34% compared to the US.[18]

**Bias from Items Without Common Barcodes:** Finally, 8 of the 71 items in our sample do not have common barcodes across the two countries. For these items, we cannot separately calculate quality and variety biases, but we can measure the *product* of these two biases on

[18]Our results from comparing Mexico and the US differs from those found in studies using within-country variation (e.g. Handbury and Weinstein, 2015). Intuitively, the variety bias captures the relative share of expenditures in common goods across locations. Within countries, households residing in a low income state consume many of the same goods that are available in higher income states. Across countries, variety bias is driven by country-specific factors and, in the case of Mexico and the US, the importance of local brands in Mexican households' consumption is remarkable. As a result, the share of expenditures in common goods for Mexico is substantially lower than that of the US, even if the US has a higher average household income.

the price of a basic heading from the following equation:

$$\text{Quality and Variety Bias From Items Without Common Barcodes}_b = \left(\frac{\lambda_b^m}{\lambda_b^u}\right)^{\frac{1}{\sigma_b - 1}} \times \frac{\prod_{i \in I_C} \left(\frac{\hat{p}_{ib}^m}{\hat{p}_{ib}^u}\right)^{\omega_{ib}^*}}{\prod_i \left(\frac{\hat{p}_{ib}^m}{\hat{p}_{ib}^u}\right)^{\omega_{ib}}} \tag{10}$$

where as before $I_C$ is the set of items with common barcodes. In our data, equation 10 is very close to one for all the basic headings with at least one item that does not have a common barcode in the two countries. So at least for these 8 items for which we need to calculate equation 10 the net effect of quality and variety biases does not change the price index of Mexican goods relative to the US. Note that this is not the case for the items for which we can separately estimate the effect of quality and variety biases. For these 63 items, the net effect of these two biases decreases Mexican prices by 2% ($1.48 \times 0.66 = 0.98$) compared to the US.

**Price Index Bias for Basic Headings:** The implication from Table 3 is that the ICP overstates Mexican prices because of imputation, sampling and variety bias, but understates Mexican prices because of quality bias. The net effect – aggregate bias in Table 3 – is that the exact price index is about 25% lower compared to the pseudo-ICP index.

Table 5 shows the bias in the ICP for the basic headings.[19] The first message is that there is a large amount of heterogeneity in the aggregate bias. The exact price index is similar to the ICP for "sugar" and "coffee, tea and cocoa." At the other extreme, the ICP price index is significantly lower than the ideal price index for "food products nec" and "non-durable household goods". This heterogeneity suggests that one should be careful about drawing strong inferences from prices obtained from a narrow set of products.

The relative importance of the four biases also differs quite bit across products. Imputation bias is large for basic headings with a large number of missing prices in the ICP such as "Pasta products" where 4 out 6 items are missing either from Mexico or the US. Sampling bias is very large for "appliances and articles for personal care" and virtually one for "fresh milk." Mexican households appear to be more price sensitive in where they buy appliances and articles for personal care compared to fresh milk. Quality bias is basically one for fresh milk, sugar, and beer: the price of the average fresh milk, sugar, and beer product is about the same as fresh milk, sugar, and beer sold both in Mexico and the US. On the other hand, quality bias is very large for cereals, drinks and juices, non-durable household goods, and

[19] The product of the quality and variety terms from items without common barcodes are very close to one for all basic headings.

personal care products. Finally, Mexican varieties are more important than American varieties for Mexican households in almost every basic item (the variety bias term is less than 1), but the exception is sugar. For this basic heading, American varieties not sold in Mexico are slightly more important than Mexican barcodes not sold in the US.

**Table 5: Biases in ICP for Each Basic Heading**

| Basic Heading | Biases | | | | Aggregate |
|---|---|---|---|---|---|
| | Imputation | Sampling | Quality | Variety | Bias |
| Other cereals, flour and other products | 0.83 | 0.53 | 1.57 | 0.56 | 0.38 |
| Bread | 1.01 | 0.95 | 0.87 | 0.63 | 0.53 |
| Other bakery products | 0.65 | 1.27 | 1.77 | 0.46 | 0.67 |
| Pasta products | 0.68 | 1.02 | 1.41 | 0.83 | 0.80 |
| Preserved or processed fish and seafood | 1.00 | 0.94 | 0.94 | 0.22 | 0.19 |
| Fresh milk | 1.00 | 1.04 | 0.98 | 0.37 | 0.38 |
| Preserved milk and other milk products | 0.82 | 1.02 | 1.41 | 0.66 | 0.79 |
| Cheese | 1.00 | 0.88 | 1.46 | 0.58 | 0.75 |
| Frozen, preserved or processed vegetables | 0.82 | 0.95 | 0.84 | 0.60 | 0.39 |
| Sugar | 0.60 | 1.44 | 1.10 | 1.06 | 1.01 |
| Jams, marmalades and honey | 0.95 | 0.69 | 1.05 | 0.69 | 0.47 |
| Confectionery, chocolate and ice cream | 1.00 | 0.75 | 1.25 | 0.92 | 0.85 |
| Food products nec | 0.90 | 0.94 | 1.61 | 0.98 | 1.35 |
| Coffee, tea and cocoa | 1.11 | 0.97 | 1.74 | 0.55 | 1.03 |
| Mineral waters, soft drinks, and juices | 0.94 | 0.73 | 1.71 | 0.55 | 0.65 |
| Beer | 1.00 | 0.96 | 1.06 | 0.86 | 0.88 |
| Non-durable household goods | 0.90 | 0.99 | 2.81 | 0.54 | 1.35 |
| Appliances and products for personal care | 0.81 | 0.70 | 1.54 | 0.66 | 0.58 |

Notes: The table reports imputation bias, sampling bias, quality bias, and variety bias for each basic heading. Aggregate bias is defined as the product of the imputation, sampling, quality, and variety biases.

# 5 Conclusion

The construction of cross-country price indexes is of crucial importance to compare living standards between countries and to measure global inequality. The ICP has taken on the important and heroic exercise of measuring these prices, but faces severe data limitations. In this paper, we construct a data set for two countries that allows us to address some of these data limitations; namely, the fact that the ICP has incomplete information on item across countries, that they do not have expenditure information to weight items appropriately, that they cannot compare exactly the same item across countries, and that they do not have information on differences on the set of products available in each country. Using our alternative data, we estimate that Mexican real consumption is larger relative to the

United States than was previously estimated. We identify the imputation of prices and the heterogeneity in shopping behavior, quality of products, and variety availability as important sources of bias in international price comparisons. Overall, our results show that the real non-durable consumption inequality across the US and Mexico is 25% lower than that predicted by the ICP estimates.

There are several generalizable lessons from our study for international price comparisons. First, aggregate bias must be estimated considering all biases jointly. Biases often imply adjustments in opposite directions, so addressing them in isolation could lead to drastically different conclusions about the comparison of the standards of living across countries. Second, because average prices are correlated with income per capita, the magnitude of the quality bias increases when comparing unequal countries. Third, the relevance of the variety bias is not correlated with income, but instead with the importance of *domestic* varieties in overall consumption. For this reason, using within-country variation yields the opposite results to using cross-country variation. Fourth, the bias from the limited geographical coverage of the ICP survey in Mexico is not large, indicating that coverage is less relevant if the sampling for each country only focuses on urban areas, as it does in the case of Mexico and the US. Fifth, the bias from using the simple average instead of the weighted average of the item level price is not large; the covariance of the expenditure weights of each item and its price is not large for neither the US or Mexico. Sixth, a country's shopping intensity is correlated with the relevance of the sampling bias. This is because weighting store specific prices using a store's total sales overstates the average price paid by consumers in all countries. Finally, we find a great deal of heterogeneity in the size of aggregate bias across different product categories. As more product-level data becomes available, we will be able to get better estimates for aggregate price index beyond consumer packaged goods.

# References


Almås, Ingvild, "International income inequality: Measuring PPP bias by estimating Engel curves for food," *American Economic Review*, 2012, *102* (2), 1093–1117.

Anderson, Simon P, Andre De Palma, and Jacques-Francois Thisse, *Discrete choice theory of product differentiation*, MIT press, 1992.

Atkin, David, Benjamin Faber, Thibault Fally, and Marco Gonzalez-Navarro, "Measuring Welfare and Inequality with Incomplete Price Information," *NBER Working Paper No. 26890*, 2020.

Bils, Mark, "Do higher prices for new goods reflect quality growth or inflation?," *The Quarterly Journal of Economics*, 2009, *124* (2), 637–675.

_ and Peter J Klenow, "Quantifying Quality Growth," *American Economic Review*, 2001, pp. 1006–1030.

Broda, Christian and David E Weinstein, "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, 2006, *121* (2), 541–585.

_ and _ , "Product creation and destruction: Evidence and price implications," *American Economic Review*, 2010, *100* (3), 691–723.

Cavallo, Alberto, Robert C Feenstra, and Robert Inklaar, "Product Variety, the Cost of Living and Welfare Across Countries," Technical Report, National Bureau of Economic Research 2021.

_ , W Erwin Diewert, Robert C Feenstra, Robert Inklaar, and Marcel P Timmer, "Using Online Prices for Measuring Real Consumption across Countries," *AEA Papers and Proceedings*, 2018, *108*, 483–87.

Chevalier, Judith A and Anil K Kashyap, "Best prices: Price discrimination and consumer substitution," *American Economic Journal: Economic Policy*, 2019, *11* (1), 126–59.

Coibion, Olivier, Yuriy Gorodnichenko, and Gee Hee Hong, "The cyclicality of sales, regular and effective prices: Business cycle and policy implications," *American Economic Review*, 2015, *105* (3), 993–1029.

Comin, Diego, Danial Lashkari, and Martí Mestieri, "Structural change with long-run income and price effects," *Econometrica*, 2021, *89* (1), 311–374.

Costa, Dora L, "Estimating real income in the United States from 1888 to 1994: Correcting CPI bias using Engel curves," *Journal of Political Economy*, 2001, *109* (6), 1288–1310.

Deaton, Angus, "Price indexes, inequality, and the measurement of world poverty," *American Economic Review*, 2010, *100* (1), 5–34.

_ and Alan Heston, "Understanding PPPs and PPP-based national accounts," *American Economic Journal: Macroeconomics*, 2010, *2* (4), 1–35.

Feenstra, Robert C, "New product varieties and the measurement of international prices," *American Economic Review*, 1994, pp. 157–177.

_ , Mingzhi Xu, and Alexis Antoniades, "What is the Price of Tea in China? Goods Prices and Availability in Chinese Cities," *The Economic Journal*, 2020, *130* (632), 2438–2467.

Hamilton, Bruce W, "Using Engel's Law to estimate CPI bias," *American Economic Review*, 2001, *91* (3), 619–630.

Handbury, Jessie and David E Weinstein, "Goods Prices and Availability in Cities," *Review of Economic Studies*, 2015, *82*, 258–296.

Hanoch, Giora, "Production and demand models with direct or indirect implicit additivity," *Econometrica*, 1975, *43* (3), 395.

Hansen, Lars Peter, "Large sample properties of generalized method of moments estimators," *Econometrica*, 1982, pp. 1029–1054.

Hottman, Colin J, Stephen J Redding, and David E Weinstein, "Quantifying the sources of firm heterogeneity," *The Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364.

Kmenta, Jan, "On estimation of the CES production function," *International Economic Review*, 1967, *8* (2), 180–189.

Matsuyama, Kiminori, "Engel's Law in the Global Economy: Demand-Induced Patterns of Structural Change, Innovation, and Trade," *Econometrica*, 2019, *87* (2), 497–528.

McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press, 1974, pp. 105–142.

Nakamura, Emi, Jón Steinsson, and Miao Liu, "Are Chinese growth and inflation too smooth? Evidence from Engel curves," *American Economic Journal: Macroeconomics*, 2016, *8* (3), 113–44.

Nordhaus, William D, "Do Real-Output and Real-Wage Measures Capture Reality? The History of Lighting Suggests Not.," in Timothy F. Bresnahan and Robert J. Gordon, eds., *The Economics of New Goods*, University of Chicago Press, 1997, pp. 29–66.

Redding, Stephen J and David E Weinstein, "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences," *The Quarterly Journal of Economics*, 2020, *135* (1), 503–560.

Sato, Ryuzo, "The most general class of CES functions," *Econometrica*, 1975, *43* (5-6), 999.

Simonovska, Ina, "Income differences and prices of tradables: Insights from an online retailer," *Review of Economic Studies*, 2015, *82* (4), 1612–1656.

# APPENDIX

# A Results with Non-homothetic CES Framework

## A.1 Non-homothetic CES Preference and Price Index

We introduce non-homotheticities using the non-separable class of CES functions in Sato (1975), Comin, Lashkari and Mestieri (2021), Matsuyama (2019) and Redding and Weinstein (2020), which satisfy implicit additivity in Hanoch (1975). The non-homothetic CES consumption index for item $i$, $C^m_{ib}$, is defined by the following implicit function:

$$\sum_{k\in m}\left(\frac{\varphi^m_{kib}C^m_{kib}}{(C^m_{ib})^{(\epsilon_{kib}-\sigma_{ib})/(1-\sigma_{ib})}}\right)^{\frac{\sigma_{ib}-1}{\sigma_{ib}}}=1 \tag{11}$$

where $C^m_{kib}$ denotes total consumption of barcode $k$; $\sigma_{ib}$ is the elasticity of substitution between barcodes; $\epsilon_{kib}$ is the constant elasticity of consumption of barcode $k$ with respect to the consumption index ($C^m_{ib}$) which controls the income elasticity of demand for that barcode. Assuming that barcodes are substitutes ($\sigma_{ib}>1$), it is required $\epsilon_{kib}<\sigma_{ib}$ for the consumption index to be globally monotonically increasing and quasi-concave, and therefore to correspond to a well-defined utility function. When $\epsilon_{kib}=1$ for all $k\in m$, the utility function becomes homothetic.

We solve the expenditure minimization problem for a given barcode within an item and basic heading to obtain the following expressions for the price index ($P^m_{ib}$) dual to the consumption index ($C_{kib}$) and the expenditure share for a individual barcode $k$ ($s^m_{kib}$):

$$P^m_{ib}=\left(\sum_{k\in m}(p^m_{kib}/\varphi^m_{kib})^{1-\sigma_{ib}}(C^m_{ib})^{\epsilon_{kib}-1}\right)^{\frac{1}{1-\sigma_{ib}}} \tag{12}$$

$$s^m_{kib}=\frac{(p^m_{kib}/\varphi^m_{kib})^{1-\sigma_{ib}}(C^m_{ib})^{\epsilon_{kib}-1}}{\sum_{l\in m}(p^m_{lib}/\varphi^m_{lib})^{1-\sigma_{ib}}(C^m_{lib})^{\epsilon_{lib}-1}}=\frac{(p^m_{kib}/\varphi^m_{kib})^{1-\sigma_{ib}}(E^m_{ib}/P^m_{ib})^{\epsilon_{kib}-1}}{(P^m_{ib})^{1-\sigma_{ib}}} \tag{13}$$

Taking ratios of the shares of Mexico and the United States and rearranging, we obtain the following expression for the difference in the cost of living, which holds for each common barcode available in two countries (common):

$$\frac{P^m_{ib}}{P^u_{ib}}=\frac{p^m_{kib}/\varphi^m_{kib}}{p^u_{kib}/\varphi^u_{kib}}\left(\frac{E^m_{ib}/P^m_{ib}}{E^u_{ib}/P^u_{ib}}\right)^{\frac{\epsilon_{kib}-1}{1-\sigma_{ib}}}\left(\frac{s^m_{kib}}{s^u_{kib}}\right)^{\frac{1}{\sigma_{ib}-1}},\quad k\in\text{common} \tag{14}$$

Summing expenditures across common barcodes, we obtain the following expression for the aggregate share of common barcodes in total expenditure for Mexico and the United

States ($\lambda_{ib}^{m}$ and $\lambda_{ib}^{u}$):

$$\lambda_{ib}^{m} \equiv \frac{\sum_{k\in \text{common}} p_{kib}^{m} C_{kib}^{m}}{\sum_{k\in m} p_{kib}^{m} C_{kib}^{m}} \quad \text{and} \quad \lambda_{ib}^{u} \equiv \frac{\sum_{k\in \text{common}} p_{kib}^{u} C_{kib}^{u}}{\sum_{k\in u} p_{kib}^{u} C_{kib}^{u}} \tag{15}$$

Using this expression, the share of an individual barcode in total expenditure ($s_{kib}^{m}$) in equation (13) can be re-written as its share of expenditure on common barcodes ($s_{kibt}^{m*}$) times this aggregate share of common barcodes in total expenditure ($\lambda_{ib}^{m}$):

$$s_{kib}^{m} = \lambda_{ib}^{m} s_{kib}^{m*}, \quad k \in \text{common} \tag{16}$$

Taking logs to equation (14) and using equation (16), we obtain the following equation:

$$\log\left(\frac{P_{ib}^{m}}{P_{ib}^{u}}\right)^{1+\frac{\epsilon_{kib}-1}{1-\sigma_{ib}}} = \log\frac{p_{kib}^{m}/\varphi_{kib}^{m}}{p_{kib}^{u}/\varphi_{kib}^{u}} + \log\left(\frac{E_{ib}^{m}}{E_{ib}^{u}}\right)^{\frac{\epsilon_{kib}-1}{1-\sigma_{ib}}} + \log\left(\frac{s_{kib}^{m*}}{s_{kib}^{u*}}\right)^{\frac{1}{\sigma_{ib}-1}} + \log\left(\frac{\lambda_{ib}^{m}}{\lambda_{ib}^{u}}\right)^{\frac{1}{\sigma_{ib}-1}} \tag{17}$$

We define the ideal log-difference weights ($\omega_{kib}$), the logarithmic mean of common variety expenditure shares, as follows:

$$\omega_{kib}^{M} = \frac{\dfrac{s_{kib}^{m*} - s_{kib}^{u*}}{\ln s_{kib}^{m*} - \ln s_{kib}^{u*}}}{\sum_{k\in \text{common}} \dfrac{s_{kib}^{m*} - s_{kib}^{u*}}{\ln s_{kib}^{m*} - \ln s_{kib}^{u*}}} \tag{18}$$

where

$$s_{kib}^{m*} = \frac{p_{kib}^{m} C_{kib}^{m}}{\sum_{k\in \text{common}} p_{kib}^{m} C_{kib}^{m}} \quad \text{and} \quad s_{kib}^{u*} = \frac{p_{kib}^{u} C_{kib}^{u}}{\sum_{k\in \text{common}} p_{kib}^{u} C_{kib}^{u}}$$

We introduce an assumption that tastes are the same between two countries for each common barcode ($\varphi_{kib}^{m} = \varphi_{kib}^{u}$ for all $k \in$ common).

By multiplying the ideal log-difference weights ($\omega_{kib}$) to equation (17), under the assumption that tastes are the same between two countries for each common barcode ($\varphi_{kib}^{m} = \varphi_{kib}^{u}$ for all $k \in$ common), we can obtain the non-homothetic CES price index for item $i$ by taking the arithmetic mean across common barcodes ($\Omega_{it}$) and exponents on both sides:

$$\text{Non-homothetic Price Index}_{ib} \equiv \frac{P_{ib}^{m}}{P_{ib}^{u}} = \left(\prod_{k\in \text{common}} \left(\frac{p_{kib}^{m}}{p_{kib}^{u}}\right)^{\omega_{kib}} \times \left(\frac{\lambda_{ib}^{m}}{\lambda_{ib}^{u}}\right)^{\frac{1}{\sigma_{ib}-1}}\right)^{\frac{1}{1-\theta_{ib}}} \left(\frac{E_{ib}^{m}}{E_{ib}^{u}}\right)^{\frac{\theta_{ib}}{\theta_{ib}-1}} \tag{19}$$

where

$$\theta_{ib} \equiv \sum_{k \in \text{common}} \omega_{kib} \frac{\epsilon_{kib} - 1}{\sigma_{ib} - 1}$$

and the ratio of $\lambda_{ib}^m$ and $\lambda_{ib}^u$ represents the conventional variety correction term that accounts for different sets of goods available in two countries as in Feenstra (1994) and Broda and Weinstein (2006, 2010).

Define $I_C$ as the set of items with common barcodes. Then, an *aggregate* exact price index can be defined as:

$$\textit{Aggregate} \text{ Non-homothetic Price Index} = \prod_b \left( \prod_{i \in I_C} \mathbb{NH}_{ib}^{\omega_{ib}^*} \times \left( \frac{\lambda_b^m}{\lambda_b^u} \right)^{\frac{1}{\sigma_b - 1}} \right)^{\omega_b} \tag{20}$$

where $\omega_{ib}^*$ and $\omega_b$ are the logarithmic mean of the expenditure shares of each item and each basic heading, respectively. The spending shares are defined as:

$$\lambda_b^m \equiv \frac{\sum_{i \in I_C} p_{ib}^m C_{ib}^m}{\sum_i p_{ib}^m C_{ib}^m} \quad \text{and} \quad \lambda_b^u \equiv \frac{\sum_{i \in I_C} p_{ib}^u C_{ib}^u}{\sum_i p_{ib}^u C_{ib}^u}$$

## A.2 Decomposition of Non-homothetic CES Price Index

The non-homothetic CES price index at the basic heading level can be written as a function of the price index developed by the ICP. The relationship between the two indexes can be written as:

$$\frac{\text{Non-homothetic Index}_b}{\mathbb{ICP}_b} = \text{Imputation}_b \times \text{Sampling}_b \times \text{Quality}_b \times \text{Engel-curve Variety}_b \tag{21}$$

where imputation bias, sampling bias, and quality bias are the same as the homothetic case. We define "Engel curve variety bias" to replace "variety bias" in equation 9 from homothetic case.

**Engel Curve Variety Bias:** This bias measures both the cross-country differences in availability of barcodes and the differences in real consumption across countries and is defined as follows:

$$\text{Engel Curve Variety Bias}_i = \text{Variety Bias}_{ib} \times \left( \frac{E_{ib}^m / E_{ib}^u}{\mathbb{EPI}_{ib}} \right)^{\frac{\theta_{ib}}{\theta_{ib} - 1}} \tag{22}$$

where

$$\text{Variety Bias}_{ib} \equiv \left( \frac{\lambda_{ib}^m}{\lambda_{ib}^u} \right)^{\frac{1}{\sigma_{ib} - 1}}. \tag{23}$$

We called the second term Engel curve adjustment since it captures the differences in real consumption across the two countries and depends on the income elasticity of demand of the common barcodes across the two countries. When not all barcodes are common across countries, this term serves as an adjustment to the standard variety bias since the share of common barcodes across countries naturally depends on their income differences. However, even if all barcodes across the two countries are common, the Engel Curve adjustment corrects the price index for the relative importance of each barcode as the relative income of the countries change. If the elasticities of consumption of each barcode with respect to the consumption index equal to one, we are back to the homothetic preferences case. In this case, the Engel-curve variety bias becomes the variety bias as homothetic case.

## A.3 Parameter Estimation

Taking estimates of the elasticity of substitution as given, we estimate the constant elasticity of consumption ($\epsilon_{kib}$) for each barcode $k$ with respect to the consumption index ($C^m_{kib}$) as in Comin, Lashkari and Mestieri (2021):

$$\ln \frac{s^h_{kibt}}{s^h_{\mathbf{K}ibt}} - (1-\sigma_{ib}) \ln \frac{p^h_{kibt}}{p^h_{\mathbf{K}ibt}} = (\epsilon_{kib} - 1) \left( \ln \frac{E^h_{ibt}}{p^h_{\mathbf{K}ibt}} + \frac{1}{(1-\sigma_{ib})} \ln s^h_{\mathbf{K}ibt} \right) + \psi^h_t + \epsilon^h_{kibt} \tag{24}$$

where $\mathbf{K}$ is the benchmark barcode, which corresponds to the largest selling barcode in each item, and $\psi^h_t$ is the set of fixed effects. We aggregate households into seven groups by their annual household income. With the US Nielsen data, equation 24 is estimated with quarter×Census region fixed effects. Note that because barcodes are substitutes within all items ($\sigma_{ib} > 1$), it is required $\epsilon_{kib} < \sigma_{ib}$ for the consumption index to be globally monotonically increasing and quasi-concave, and therefore to correspond to a well-defined utility function.[20]

[20] In less than one percent of the cases, $\epsilon_{kib} >= \sigma_{ib}$. In these cases, we impute $\epsilon_{kib} = \sigma_{ib} - 0.01$.

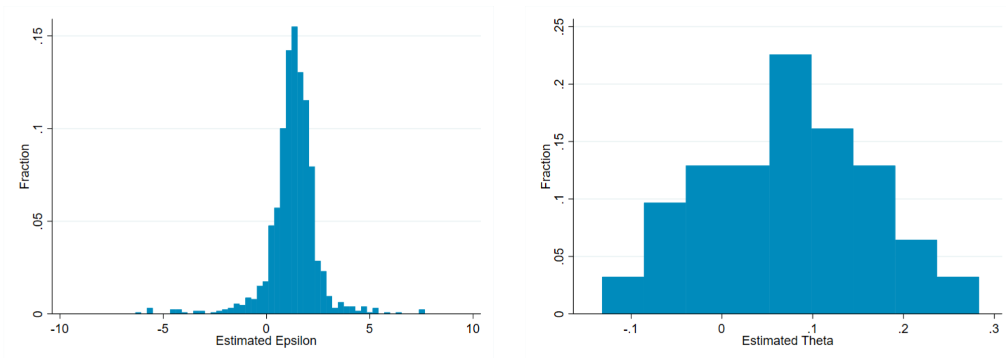
**Table A.I: Descriptive Statistics of Estimated Parameters**

| | Mean | Std. Dev. | 10th-Percentile | Median | 90th-Percentile |
|---|---|---|---|---|---|
| $\sigma_{ib}$ | 6.48 | 3.09 | 4.32 | 5.53 | 10.82 |
| $\epsilon_{kib}$ | 1.27 | 1.20 | 0.22 | 1.37 | 2.31 |
| $\theta_{ib}$ | 0.07 | 0.09 | -0.04 | 0.08 | 0.18 |
| $E^m_{ib}/E^u_{ib}$ | 0.81 | 0.69 | 0.23 | 0.65 | 1.60 |

Note: This table reports descriptive statistics for the elasticity of substitution ($\sigma_{ib}$), the elasticity of consumption of barcode $k$ with respect to the consumption index ($\epsilon_{kib}$), parameter in the non-homothetic CES price index ($\theta_{ib}$) and nominal expenditure ratio ($E^m_{ib}/E^u_{ib}$).

The first row of Table A.I reports descriptive statistics for $\sigma_{ib}$. The average of the elasticity of substitution we use is 6.48 with standard deviation of 3.09. The second row of Table A.I reports the descriptive statistics for $\epsilon_{kib}$. Our estimates for this parameter have a mean of 1.27 and a standard deviation of 1.20. The third row of Table A.I reports the descriptive statistics for $\theta_{ib}$. Recall that this parameter is the equally weighted average of $\frac{\epsilon_{kib}-1}{\sigma_{ib}-1}$ across the common barcodes within an item across the two countries. Our estimates of this parameter have a mean of 0.07 and a standard deviation of 0.09. Note that when $\theta_{ib}$ is close to zero, the expenditure ratio plays a small role in the price index. Lastly, we report other informative moments for the quantification of the Engel-curve variety bias such as the nominal expenditure ratio. As the last row of Table A.I shows, the ratio varies across items. It has a median of 0.65 and a mean of 0.81 (with standard deviation of 0.69), which indicates that the distribution is skewed to the right. Figure A.1 shows the distribution of estimated $\epsilon_{kib}$ and $\theta_{ib}$. The mean of $\epsilon_{kib}$ is close to one. As a result, the mean of $\theta_{ib}$ is around zero.

**Figure A.1: Distribution of Estimated $\epsilon_{kib}$ and $\theta_{ib}$**




$\epsilon_{ki}$ $\theta_i$

Note: Panel (a) and (b) show the distribution of estimated $\epsilon_{kib}$ and $\theta_{ib}$, respectively.

## A.4 Decomposition Results

Equation 21 indicates that the gap between the two price indexes can be decomposed into the imputation bias, the sampling bias, the quality bias, and the quality-variety bias. The aggregate bias is estimated to be 0.75, which is the same at two decimal points with the homothetic preference. This is mainly because, for the common barcodes, the average elasticity of consumption is estimated to be close to 1.

**Table A.II: Decomposition Results for the Non-Homothetic CES Price Index**

| Aggregate Price Index | | Bias due to: | | | Aggregate |
|---|---|---|---|---|---|
| Exact | Pseudo-ICP | Imputation | Sampling | Quality-Variety | Bias |
| 0.64 | 0.86 | 0.89 | 0.86 | 0.98 | 0.75 |

Notes: The table reports the aggregate non-homothetic CES price index, pseudo ICP price index, and the gap between the exact and ICP index due to imputation, sampling, and quality-variety. Aggregate bias is the product of the bias due to imputation, sampling, and quality-variety.

# B Logit Specification of Barcode-level Price Aggregation

In the main text, we use Cobb-Douglas aggregation as the first-order approximation of CES aggregation (Kmenta, 1967). In this section, we show that CES aggregation can be derived from the aggregation of the choices of individual consumers with extreme-value-distributed idiosyncratic amenities (net of costs) from different stores.

Changing the notion of McFadden (1974) slightly, we suppose that the utility of an individual consumer $i$ who consumes $q_{is}$ units of barcode $k$ from store $s$ is given by:

$$U_i = u_s + a_{is}, \quad u_s \equiv \ln q_{is} \tag{25}$$

where $a_{is}$ captures idiosyncratic amenities to visit store $s$ that are drawn from an independent Type-I Extreme Value distribution:

$$G(a) = e^{-e^{(-a/\nu+\kappa)}} \tag{26}$$

where $\nu$ is the scale parameter of the extreme value distribution and $\kappa$ is the Euler-Mascheroni constant.

Each consumer has the same expenditure on barcode $k$, $E_k$, and chooses their preferred store given the observed realization for idiosyncratic amenities. Therefore, the consumer's budget constraint implies:

$$q_{is} = \frac{E_k}{p_s} \tag{27}$$

The probability that individual $i$ choose store $s$ is:

$$\begin{aligned} x_{ikt} &= \text{Prob}[u_{is} + c_{is} > u_{il} + c_{il}, \forall l \neq s] \\ &= \text{Prob}[c_{il} < c_{is} + u_{is} - u_{il}, \forall l \neq s] \end{aligned} \tag{28}$$

Using the distribution of idiosyncratic amenities in equation 26, we have:

$$x_{is}|a_{is} = \prod_{l \neq k} e^{-e^{-(a_{is}+u_{is}-u_{il})/\nu+\kappa}} \tag{29}$$

Once we integrate it across the probability density function for $a_{is}$ and use the change of variable technique as in Anderson, De Palma and Thisse (1992), the probability that individual $i$ chooses store $s$ becomes:

$$s_{is} = s_s = \frac{p_s^{-1/\nu}}{\sum_{l \in \Psi} p_l^{-1/\nu}} \tag{30}$$

The expected utility of consumer $i$ is:

$$E[U_i] = E[\max\{u_{i1} + a_{i1}+, ..., u_{iN} + a_{iN}\}] = \nu \ln \left[ \sum_{l \in \Psi} \exp\left( \frac{u_{il}}{\nu} \right) \right] \tag{31}$$

Using the definition of $u_{is}$ in equation 25 and the budget constraint in equation 27, expected utility can be written as:

$$E[U_i] = E_k / P \tag{32}$$

where $P$ is the unit expenditure function:

$$P = \left[ \sum_{s \in \Psi} p_s^{-1/\nu} \right]^{-\nu} \tag{33}$$

Therefore, theoretically consistent aggregation of prices of a specific barcode across stores is the CES aggregation.

# C Parameter Estimation

In order to obtain the elasticity of substitution, $\sigma_{ib}$, for each item, we rely on the method developed by Feenstra (1994) and extended by Broda and Weinstein (2006) and Broda and Weinstein (2010). The procedure consists of estimating a demand and supply equation for each barcode by using only the information on prices and quantities. For this estimation, we face the standard endogeneity problem for a given barcode. Although we cannot identify supply and demand, the data do provide information about the joint distribution of supply and demand parameters.

We first model the supply and demand conditions for each barcode within an item. Specifically, we estimate the demand elasticities by using the following system of differenced demand and supply equations as in Broda and Weinstein (2006):

$$\Delta^{\underline{k},t}\ln S_{kibt} = (1-\sigma_{ib})\Delta^{\underline{k},t}\ln P_{kibt} + \iota_{kibt} \tag{34}$$

$$\Delta^{\underline{k},t}\ln P_{kibt} = \frac{\delta_{ib}}{1+\delta_{ib}}\Delta^{\underline{k},t}\ln S_{kibt} + \kappa_{kibt} \tag{35}$$

Note that when the inverse supply elasticity is zero (i.e. $\delta_{ib}$=0), the supply curve is horizontal and there is no simultaneity bias in $\sigma_g$. Equations 34 and 35 are the demand and supply equations of barcode $k$ in an item $i$ differenced with respect to a benchmark barcode in the same item. The $k^{th}$ good corresponds to the largest selling barcode in each item. The $k$-differencing removes any item level shocks from the data.

The identification strategy relies on two important assumptions. First, we assume that $\iota_{kibt}$ and $\kappa_{kibt}$, the double-differenced demand and supply shocks, are uncorrelated (i.e., $\mathbb{E}_t(\iota_{kibt}\kappa_{kibt}) = 0$). This expectation defines a rectangular hyperbola in ($\delta_{ib}$, $\sigma_{ib}$) space for each barcode within an item, which places bounds on the demand and supply elasticities. Because we already removed any item level shocks, we are left with within item variation that is likely to render independence of the barcode-level demand and supply shocks within an item. Second, we assume that $\sigma_{ib}$ and $\omega_{ib}$ are restricted to be the same over time and for all barcodes in a given item.

To take advantage of these assumptions, we define a set of moment conditions for each item $i$ in a basic heading $b$ as below:

$$G(\beta_{ib}) = E_T[\nu_{kibt}(\beta_{ib})] = 0 \tag{36}$$

where $\beta_{ib} = [\sigma_{ib}, \delta_{ib}]'$ and $\nu_{kibt} = \iota_{kibt}\kappa_{kibt}$.

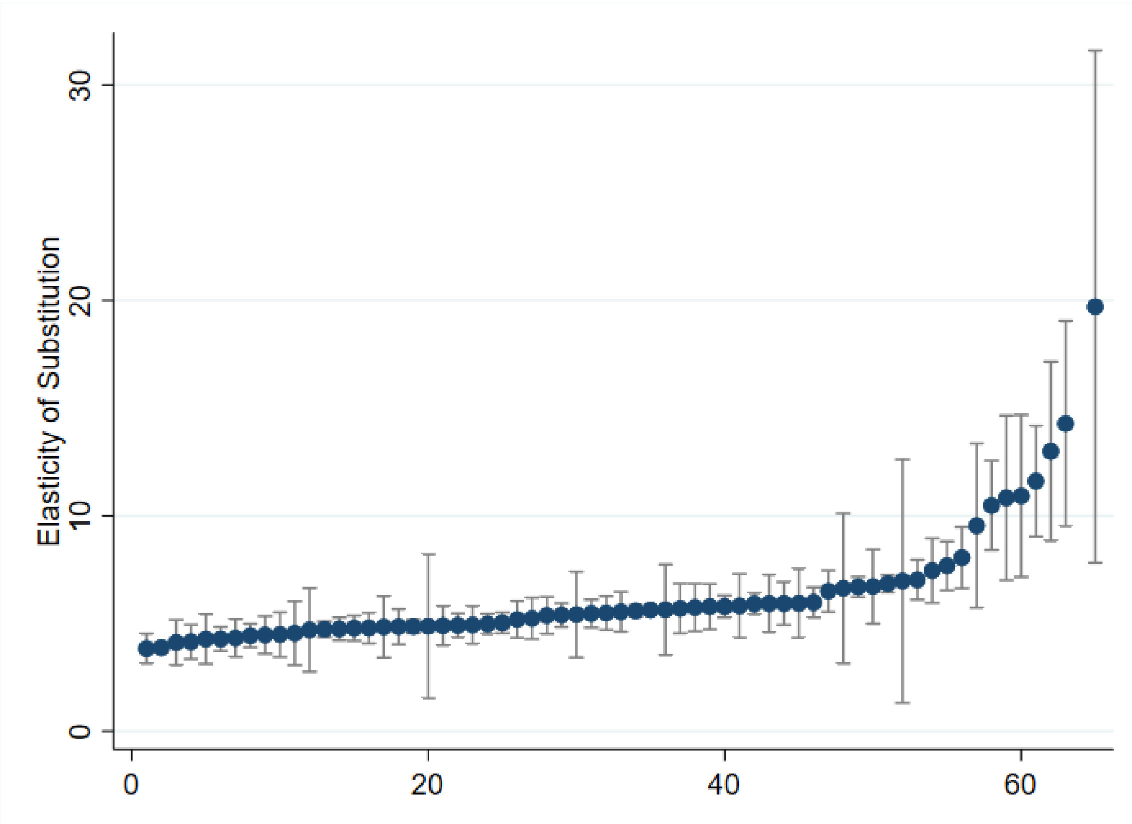
For each item $i$, all the moment conditions that enter the GMM objective function can be combined to obtain Hansen (1982)'s estimator:

$$\hat{\beta_{ib}} = \arg\min_{\beta_{ib} \in B} G^*(\beta_{ib})' W G^*(\beta_{ib}) \quad \forall\, i \in \omega_b \tag{37}$$

where $G^*(\beta_{ib})$ is the sample analog of $G(\beta_{ib})$, $W$ is a positive definite weighting matrix, and $B$ is the set of economically feasible $\beta_{ib}$ (i.e., $\sigma_{ib} > 0$). Our estimation procedure follows Redding and Weinstein (2020) using he Nielsen Homescan data from 2004-2019. The elasticities are estimated using data at the quarterly frequency. Households are aggregated using sampling weights to make the sample representative of each country's population. We weight the data for each barcode by the number of raw buyers to ensure that our objective function is more sensitive to barcodes purchased by larger numbers of consumers. We consider barcodes with more 10 or more observations during the estimation. If the procedure renders imaginary estimates or estimates of the wrong sign, we use a grid search to evaluate the GMM objective function above. The average of the elasticity of substitution we obtain is 6.48 with standard deviation of 3.09.[21] Figure C.1 shows the distribution of elasticities sorted by their magnitude and the 95% confidence interval of the point estimates.

[21] Unlike Redding and Weinstein (2020), Broda and Weinstein (2010) and Hottman, Redding and Weinstein (2016) include an additional brand/firm-level layer within a product category. In order to check the robustness of our estimates, we estimate elasticity of substitution within an item-firm level and compare it to our elasticity of substitution within an item level. We use the GS1 data to identify a firm for every barcode. The estimated elasticities of substitution within an item-firm have slightly higher mean (6.91), but both are highly correlated. Correlation is 0.65 and statistically significant at the 1 percent level. Therefore, using elasticity of substitution within an item-firm does not quantitatively affect our results.

**Figure C.1: Elasticity of Substitution ($\sigma_{ib}$)**




Note: The figure reports the estimated elasticity of substitution for each item, sorted by its magnitude. The gray lines indicate the 95% confidence intervals.

# D Decomposition Results with Common Elasticity of Substitution

In this section, we report decomposition results with common elasticity of substitution across items, $\sigma_{ib} = 6$, which is chosen from a mean of $\sigma_{ib}$ (6.48). Table D.I reports decomposition results for the exact price index. The aggregate bias (0.75) is estimated to be the same with the item-specific elasticity of substitution case (0.75).

**Table D.I: Decomposition Results for the Exact Price Index under** $\sigma_{ib} = 8$

| Aggregate Price Index | | Bias due to: | | | Aggregate |
|---|---|---|---|---|---|
| Exact | Pseudo-ICP | Imputation | Sampling | Quality-Variety | Bias |
| 0.64 | 0.86 | 0.89 | 0.86 | 0.98 | 0.75 |

Notes: The table reports the aggregate exact price index, pseudo ICP price index, and the gap between the exact and ICP index due to imputation, sampling, and quality-variety under $\sigma_{ib} = 6$. Aggregate bias is the product of the bias due to imputation, sampling, and quality-variety.

# E Sampling Bias

## E.1 Proofs of Propositions

PROPOSITION 1. If the number of basic headings $N_b \to \infty$, the second term of the sampling bias is larger than 1 if $\text{cov}(\omega_{\mathbf{b}}, \ln(\bar{\mathbf{p}}_{\mathbf{b}}^{\mathbf{c}})) > \text{cov}(\omega_{\mathbf{b}}, \ln(\bar{\mathbf{p}}_{\mathbf{b}}^{\mathbf{u}}))$.

**Proof.** The bias on the second term for country $c$ is:

$$\prod_i \frac{(\bar{p}_{ib}^c)^{\omega_{ib}}}{(\bar{p}_{ib}^c)^{\frac{1}{N_b}}}$$

This ratio is greater than one if and only if:

$$\sum_i \left(\omega_{ib} - \frac{1}{N_b}\right) \ln(\bar{p}_{ib}^c) > 0 \tag{38}$$

We want to show that this term is equivalent to $\text{cov}(\omega_{\mathbf{b}}, \ln(\bar{\mathbf{p}}_{\mathbf{b}}^{\mathbf{c}}))$ where $\omega_{\mathbf{b}}$ is a vector of weights in basic heading $b$ and $\ln(\bar{\mathbf{p}}_{\mathbf{b}}^{\mathbf{c}})$ is the vector of log prices. By definition:

$$\text{cov}(\omega_{\mathbf{b}}, \ln(\bar{\mathbf{p}}_{\mathbf{b}}^{\mathbf{c}})) = \lim_{N_b \to \infty} \frac{1}{N_b - 1} \sum_i \left(\omega_{ib} - \frac{1}{N_b}\right) \left(\ln(\bar{p}_{ib}^c) - \frac{1}{N_b - 1} \sum_i \ln(\bar{p}_{ib}^c)\right)$$

Using that

$$\lim_{N_b \to \infty} \frac{1}{N_b - 1} \sum_i \left(\omega_{ib} - \frac{1}{N_b}\right) \left(\frac{1}{N_b - 1} \sum_i \ln(\bar{p}_{ib}^c)\right) = 0$$

Then $\text{cov}(\omega_{\mathbf{b}}, \ln(\bar{\mathbf{p}}_{\mathbf{b}}^{\mathbf{c}}))$ is equivalent to $\frac{1}{N_b - 1} \sum_i \left(\omega_{ib} - \frac{1}{N_b}\right) \ln(\bar{p}_{ib}^c)$.

PROPOSITION 2. If the number of stores $S \to \infty$, the third term of the sampling bias is larger than 1 if $\text{cov}(\phi_{\mathbf{ib}}^{\mathbf{m}}, \ln(\bar{\mathbf{p}}_{\mathbf{ib}}^{\mathbf{m}})) - \text{cov}(\phi_{\mathbf{ib}}^{\mathbf{u}}, \ln(\bar{\mathbf{p}}_{\mathbf{ib}}^{\mathbf{u}})) > \text{cov}(\phi^{\mathbf{u}}, \ln(\bar{\mathbf{p}}_{\mathbf{ib}}^{\mathbf{u}})) - \text{cov}(\phi^{\mathbf{m}}, \ln(\bar{\mathbf{p}}_{\mathbf{ib}}^{\mathbf{m}}))$

**Proof.** Before aggregating across items, the third term of the sample bias for a country $c$ is:

$$\frac{\hat{p}_{ib}^c}{\bar{p}_{ib}^c} = \frac{\prod_{s\in\Psi^c} \left(\bar{p}_{sib}^c\right)^{\phi_{sib}^c}}{\prod_{s\in\Psi^c} \left(\bar{p}_{sib}^c\right)^{\phi_{s}^c}}$$

where $\phi_s^c$ are total sales weights at the store level and $\phi_{sib}^c$ are total sales weights at the item-store level. This ratio is larger than 1 for country $c$ if

$$\sum_{s\in\Psi^c} \phi_{sib}^c \ln\left(\bar{p}_{sib}^c\right) > \sum_{s\in\Psi^c} \phi_{s}^c \ln\left(\bar{p}_{sib}^c\right) \tag{39}$$

Let $\phi_{\mathbf{ib}}^{\mathbf{c}}$ be the vector of expenditure weights that vary by item and basic heading, (i.e. $\{\phi_{sib}^c\}_{s\in\Psi^c}$), $\phi^{\mathbf{c}}$ be the vector of weights that only vary at the store level (i.e. $\{\phi_s^c\}_{s\in\Psi^c}$), and $(\bar{\mathbf{p}}_{\mathbf{ib}}^{\mathbf{c}})$ be the vector of log prices (i.e. $\{\ln\left(\bar{p}_{sib}^c\right)\}_{s\in\Psi^c}$). Then equation 39 can be written as:

$$\lim_{S\to\infty} \frac{1}{S-1} \sum_{s\in\Psi^c} \phi_{sib}^c \ln\left(\bar{p}_{sib}^c\right) > \lim_{S\to\infty} \frac{1}{S-1} \sum_{s\in\Psi^c} \phi_{s}^c \ln\left(\bar{p}_{sib}^c\right)$$

Using the definition of covariance on both sides

$$\begin{aligned}\lim_{S\to\infty} \frac{1}{S-1} \sum_{s\in\Psi^c} \phi_{sib}^c \times \frac{1}{S-1} \sum_{s\in\Psi^c} \ln\left(\bar{p}_{sib}^c\right) + \frac{1}{S-1} \sum_{s\in\Psi^c} \left(\phi_{sib}^c - \bar{\phi}_{ib}^c\right) \left(\ln\left(\bar{p}_{sib}^c\right) - \overline{\ln\left(\bar{p}_{sib}^c\right)}\right) > \\ \lim_{S\to\infty} \frac{1}{S-1} \sum_{s\in\Psi^c} \phi_{s}^c \times \frac{1}{S-1} \sum_{s\in\Psi^c} \ln\left(\bar{p}_{sib}^c\right) + \frac{1}{S-1} \sum_{s\in\Psi^c} \left(\phi_{s}^c - \bar{\phi}^c\right) \left(\ln\left(\bar{p}_{sib}^c\right) - \overline{\ln\left(\bar{p}_{sib}^c\right)}\right)\end{aligned}$$

Taking the limit as $\lim_{S\to\infty}$ on both sides, we find that $\text{cov}(\phi_{\mathbf{ib}}^{\mathbf{c}}, \bar{\mathbf{p}}_{\mathbf{ib}}^{\mathbf{c}}) > \text{cov}(\phi^{\mathbf{c}}, \bar{\mathbf{p}}_{\mathbf{ib}}^{\mathbf{c}})$.

## E.2 Empirical Tests

To quantify the importance of the second term of the sampling bias, we rely on Proposition 1 and test whether $\text{cov}(\omega_{\mathbf{b}}, \ln(\bar{\mathbf{p}}_{\mathbf{b}}^{\mathbf{m}})) > \text{cov}(\omega_{\mathbf{b}}, \ln(\bar{\mathbf{p}}_{\mathbf{b}}^{\mathbf{u}}))$. To do so we rely on the following specification:

$$\omega_{ib} = \alpha + \beta \ \ln(\bar{p}_{ib}^c) \times \mathbb{1}\left\{c = \text{Mexico}\right\} + \lambda^c + \theta_b + \epsilon_{ib}^c$$

where the dependent variable is the Sato-Vartia weight for each item. The coefficient of interest is $\beta$ which indicates whether there is a difference between the covariance between the weights and the prices of items across the two countries; Table E.I shows that we do not find a significant difference indicating that the second term of the sampling bias is close to 1.

**Table E.I: Sampling Bias Second Term: Expenditure Weights and Prices**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $\ln(\bar{p})$ | -0.026 | -0.000 | -0.026 | -0.007 |
| | (0.018) | (0.019) | (0.018) | (0.019) |
| $\ln(\bar{p}) \times$ Mexico | 0.032 | -0.019 | 0.032 | 0.023 |
| | (0.032) | (0.012) | (0.032) | (0.029) |
| Observations | 165 | 165 | 165 | 165 |
| R-squared | 0.041 | 0.328 | 0.041 | 0.339 |
| Basic Heading | N | Y | N | Y |
| Country | N | N | Y | Y |

Note: The table shows the relationship between the expenditure weights at the item level and the prices of items within a basic heading. Column (2) includes basic heading effects, Column (3) includes country effects, and Column (4) both.

To quantify the size of the third term of the sampling bias, we rely on Proposition 2. In order to compare the magnitude of $\text{cov}(\phi^{\mathbf{m}}_{\mathbf{ib}}, \ln(\bar{\mathbf{p}}^{\mathbf{m}}_{\mathbf{ib}}))$ relative to $\text{cov}(\phi^{\mathbf{u}}_{\mathbf{ib}}, \ln(\bar{\mathbf{p}}^{\mathbf{u}}_{\mathbf{ib}}))$ we estimate the following specification:

$$\phi^c_{sib} = \alpha + \beta \ln(\bar{p}^c_{sib}) \times \mathbb{1}\{c = \text{Mexico}\} + \theta^c + \lambda_i + \epsilon^c_{sib}$$

where the dependent variable are the country-specific expenditure weights at the store-item level and the independent variable are the log prices at the same level. We include country and item effects in the specification. Table E.II presents the results. It shows that the covariance of expenditure weights and prices is strongly negative for items and stores in Mexico. The results are robust after controlling for store, item, and country effects simultaneously.

**Table E.II: Sampling Bias Third Term: Expenditure Weights and Prices at the Store $\times$ Item Level**

| | (1) | (2) | (3) |
|---|---|---|---|
| $\ln(\bar{p})$ | -0.0056 | -0.0050 | 0.0007 |
| | (0.002) | (0.001) | (0.005) |
| $\ln(\bar{p}) \times$ Mexico | -0.0839*** | -0.0766** | -0.0799* |
| | (0.000) | (0.006) | (0.007) |
| Observations | 420,501 | 420,501 | 417,290 |
| R-squared | 0.021 | 0.031 | 0.129 |
| Store | N | N | Y |
| Item | N | Y | Y |
| Country | N | Y | Y |

Note: The table shows the results of estimating the relationship between the country-specific expenditure weights at the store-item level and the log prices at the same level. The dependent variable is multiplied times $10^4$. Column (2) includes item and country effects, Column (3) includes the same controls in addition to store effects. The standard errors are clustered at the country level.