

NBER WORKING PAPER SERIES

MEASURING THE COST OF LIVING IN MEXICO AND THE US

David O. Argente
Chang-Tai Hsieh
Munseob Lee

Working Paper 27806
<http://www.nber.org/papers/w27806>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2020

We thank Robert Inklaar, Esteban Rossi-Hansberg and Ina Simonovska for helpful comments. Results are calculated based in part on data from AC Nielsen provided by the Kilts Center for Marketing Data Center at Chicago Booth. The conclusions drawn from the Nielsen 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by David O. Argente, Chang-Tai Hsieh, and Munseob Lee. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Measuring the Cost of Living in Mexico and the US
David O. Argente, Chang-Tai Hsieh, and Munseob Lee
NBER Working Paper No. 27806
September 2020
JEL No. E0,F0,O0

ABSTRACT

Cross-country price indexes are crucial to compare living standards between countries and to measure global inequality. An accurate measurement of these price indexes is a difficult task because of the lack of accurate data on the consumption patterns of different countries. We construct a unique data on prices and quantities for consumer packaged goods matched at the barcode-level across the United States and Mexico. We estimate that the Mexican real consumption relative to the United States is larger than previously estimated. We identify heterogeneity in shopping behavior, quality of products, and variety availability as important sources of bias in international price comparisons.

David O. Argente
Pennsylvania State University
Department of Economics
403 Kern Building
University Park
State College, PA 16801
dargente@psu.edu

Munseob Lee
University of California at San Diego
9500 Gilman Drive, #0519
La Jolla, CA 92093-0519
munseoble@ucsd.edu

Chang-Tai Hsieh
Booth School of Business
University of Chicago
5807 S Woodlawn Ave
Chicago, IL 60637
and NBER
chsieh@chicagoBooth.edu

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 in multiple countries, which it then aggregates into price indexes of 155 broad product categories (“basic headings”) in each country.

There are however three potential biases with the ICP’s price indexes. First, instead of collecting prices directly from consumers, the ICP collects prices from retail stores chosen based on their *total* sales. However, if consumers are more likely to purchase more of the cheaply priced items in each store, prices should be sampled based on the volume of sales of items in a store. Second, products differ in quality across countries. The ICP addresses differences in quality by finding products that are “comparable” across countries, but it is possible that the ICP matches lower quality items in one country with higher quality items in another country. Third, we know that goods available in one country are not available in other countries and the ICP makes no adjustment for potential differences in the availability of products across countries.

In this paper, we use a new data set, the Nielsen Consumer Panel data (hereafter, the Nielsen data) for the US and Mexico, to quantify the magnitude of the potential biases in the ICP due to sampling, quality, and variety. The Nielsen data tracks the shopping behavior of 40-60 thousand households in the US and 5 thousand households in Mexico. Households in the two countries use in-home scanners or diaries to record their purchases of packaged goods from each retail store and shopping trip. The Nielsen data thus has information on prices and quantities purchased by each household of specific packaged goods, identified by a 12-digit barcode, for each shopping visit and each retail store.

We use the Nielsen data to calculate price indexes for 42 items of nondurables in Mexico vs the US. The items of nondurables covered in the Nielsen data account for 60-65% of nondurable spending in the two countries. This alternative price index exploits the richness of the Nielsen data to improve upon the ICP in three ways.

First, to measure the sampling bias that comes from sampling from retail stores, we use the fact that the Nielsen data has information on prices and quantities of all purchases for a representative sample of households in each country. We thus measure the price paid for each product from a representative sample of *households* from the Nielsen data. The price gap between Mexico and the US using weights from consumer expenditures are 15% lower compared to the price gap measured using weights from a sample of retail stores. We show that this gap comes from the fact that consumers in Mexico behave differently than in

the US. Mexican households shop more frequently and are more likely to purchase only the lower priced items in each store compared to American households. Because of this, sampling prices from stores based on their *total* sales will overstate the prices paid by consumers more in Mexico compared to the US.

Second, the Nielsen data identifies products by barcodes, which we use to identify more than 5,000 identical products (with the same barcodes) in the two countries. The price gap between Mexico and the US among products with the same barcode is 35% *higher* compared to the price index of “comparable” products calculated following the ICP’s methodology. This suggests that many products that appear to be similar in the two countries may not in fact be comparable if they do not share a common barcode.

Third, since we observe all the purchases made by households, we can estimate 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 not sold in Mexico. When we take into account the differences in the availability of varieties in the two countries, effective prices in Mexico are 20% lower compared to the US. The net effect of all three adjustments, for sampling, quality, and variety, lowers Mexican prices relative to the US by 9% compared to the ICP.

There are multiple conceptual and practical challenges in making cross-country price comparisons (Deaton and Heston, 2010). Multiple revisions of the Penn World Table (PWT) tackled some of those challenges (Feenstra, Inklaar and Timmer, 2015). However, our understanding on the size of potential biases is still limited, and this paper quantifies it.

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 (2020) and Simonovska (2015) use online data and Feenstra, Xu and Antoniadou (2019) use scanner data of toothpaste, laundry detergent, personal wash items and shampoo to estimate PPPs across countries. The main advantage of our data is that it contains detailed data on *majority* of nondurables from a representative sample of *consumers*.

The rest of the paper is organized as follows. Section 2 presents the description of the data. In Section 3 we present three stylized facts about the prices in Mexico and the US. In Section 4 we develop our price index and the definition of the three biases we quantify. Section 5 presents our results. In Section 6, we discuss the potential implications of our findings and conclude.

2 Data Description

2.1 International Comparison Program (ICP)

We use the restricted ICP micro-data for Mexico and the United States from the 2011 ICP. The ICP is a worldwide statistical initiative led by the United Nations and the World Bank that collects and compares price data and GDP expenditures to estimate purchasing power parities (PPPs). The data is collected every 6 years through partnerships with statistical agencies in each country. The publicly available data contain price indexes of 155 basic headings (e.g. “Appliances, articles, and products of personal care”) that cover all the components of GDP. Approximately, 53 basic headings refer to goods of which 33 are non-durables. The restricted ICP data also includes information on the average price of the “items” (e.g. shampoo, razors, diapers) within each basic heading, but the items priced within a given basic heading is not necessarily the same in all countries. In the case of Mexico and the US, ICP 2001 collects data on the prices of 586 items in the two countries, of which 329 items are priced in both two countries.

After an item is selected for pricing, each country collects prices for the item from a sample of retail establishments. The ICP provides specifications about the products to be sampled for each item by the price collectors. These specifications include: 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, for the item “Baby Diapers”, the price collector must sample products belonging to a well-known brand, containing between 18-24 pieces, either classic or basic type, with a size between 4 and 9.5 kg, and with a multi-pack package.

Price collectors collect prices of each item from a sample of outlets chosen based on the total sales of the establishment. Importantly, prices of all the products that meet the specifications for the items to be surveyed in an establishment are collected regardless of the volume of sales of that product in the particular retail store. The average price of an item is the simple average of the specified product in the establishments that are surveyed.

The data contains the national accounts expenditures for each basic heading but not for each item 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.” The price index of a basic heading then is a geometric mean of the item’s price, where the weight of an “important” item is three times that of an “less important” item.

2.2 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 panelist uses in-home scanners to record their purchases. A 12-digit universal product code (UPC) identifies the items the panelists purchase. The data contain several million distinct UPCs grouped into 1,235 “product modules” defined by Nielsen.

For each UPC, the data contain information on the brand, size, packaging, and a rich set of product features. If the panelist purchases the good at a store covered by Nielsen, the price is set automatically to the average price of the good at the store during the week when the purchase was made. If not, the panelist directly enters the price. Nielsen reports detailed transaction information for each product purchased (e.g., UPC code, quantity, price, deals, and coupons). We combine this information with the weight and volume of the product to compute unit values.

The data also contain information about each purchasing trip the panelist makes, including information on the retailer, the location, and the date of the transaction. Further, 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 panelists.

The Nielsen data for Mexico tracks the shopping behavior of 6,000 households for the years 2012-2013. 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. Just as in the US, a UPC identifies each product. Nielsen’s data for Mexico contain around 55,000 distinct UPCs grouped into 100 Nielsen “product categories.”

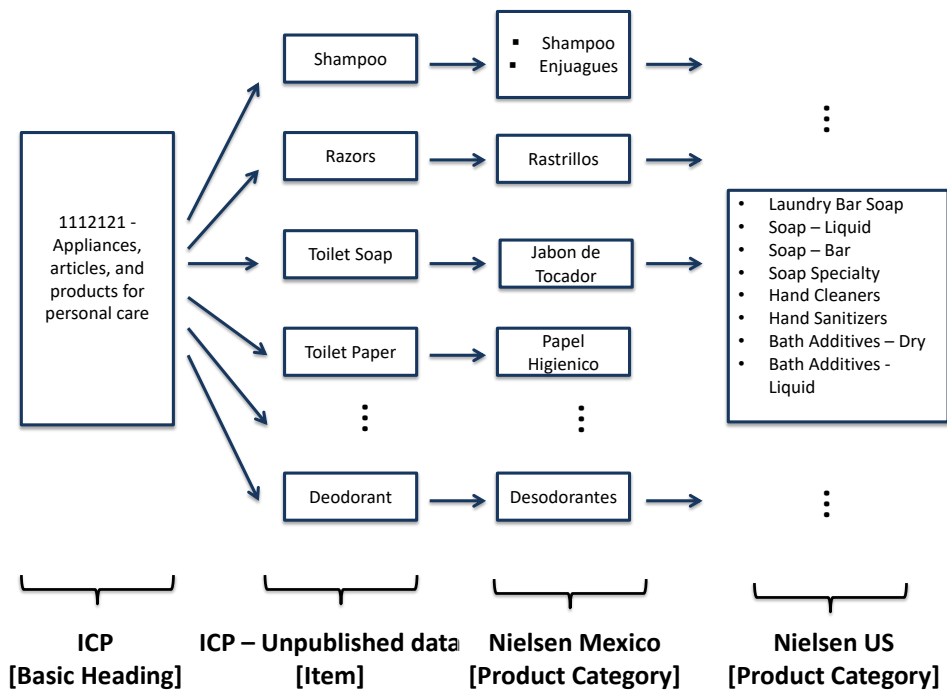
As in the US data, the Mexican data contains detailed information of each shopping trip (e.g. date, store, amount spent), transaction level information for each product purchased (e.g. quantity, price, deals, coupons), as well as detailed product level characteristics (e.g., brand, size, packaging, flavor) so that unit values can be computed. Most of the sales in the data take place at corner stores (43%) or department stores (40%); yet, since the data set is collected through diaries, it also contains information of purchases from street vendors (2.5%), tortillerias/bakeries (2%), pharmacies (1.5%) and other types of stores (11%). The data also include demographic variables at the household level such as the occupation of the household members, education, age, and family size.

2.3 Matched Data

The narrowest definition of a product in the Nielsen data are barcodes, grouped into “product categories.” In the ICP, the most disaggregated product is an item, grouped into “basic headings.” We manually link barcodes in the Nielsen data to 42 items in the ICP available in both Mexico and the US. For these items, we match each item (e.g. Diapers) to the closest Nielsen product category. The items in our linked data represents 60% of non-durable goods expenditures in Mexico and 65% in the US. The non-durable consumption categories that we are unable to match are those with products without barcodes such as fresh meats and fruits.

Figure 1 illustrates how we match the ICP item toilet soap. The item is first matched with “Jabon de Tocador” in the Mexican Nielsen data, then with multiple product modules in US Nielsen data, such such as “Soap- Liquid,” “Soap- Bar,” and “Soap Specialty”. Appendix A reports statistics about the 42 items matched across the Nielsen data of US and Mexico.

Figure 1: Structure of Matched Data



Notes: The figure describes how we match the ICP item toilet shop between ICP, Nielsen Mexico and Nielsen US.

We also match products with the same barcodes in the US and Mexico. The two countries use the barcode system. To obtain a UPC code, firms in both countries must first obtain a

Global Standards One (GS1) company prefix. GS1 is the single official source of UPC codes and it has member organizations in over 100 countries, including Mexico and the US. GS1 US and GS1 Mexico issue authorized GS1 barcodes for businesses in the US and Mexico beginning with 00-139 and 750 respectively.¹ In our data, more than that 5,000 barcodes are consumed by households both in the US and Mexico. Approximately, 80% of the matched barcodes have prefixes authorized by GS1 US.

3 Stylized Facts

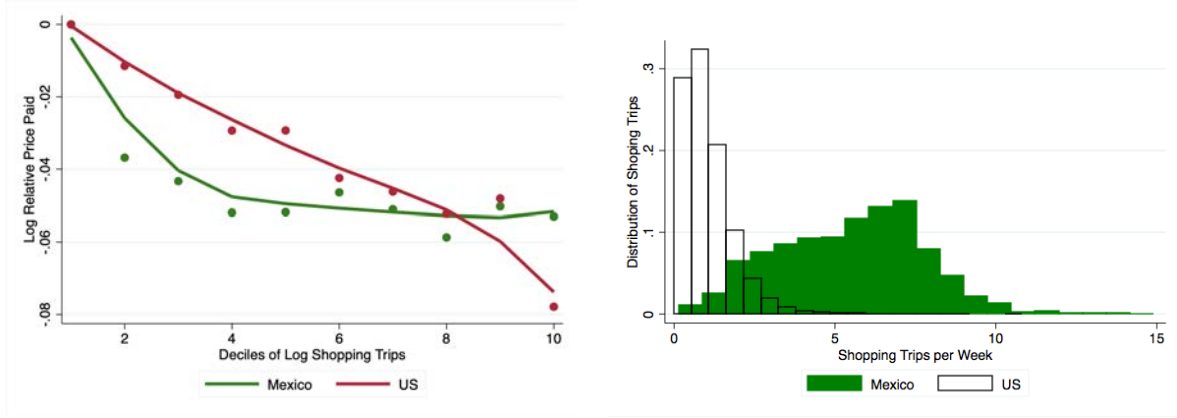
In this section, we present three stylized facts from the matched Nielsen data. These facts will be important in interpreting the estimated biases in the ICP we will document in the next section.

Fact 1: Mexican households shop more frequently and buy cheaper products in the stores they visit.

Panel (a) in Figure 2 plots the demeaned price of a product (barcode) paid by a household in a shopping visit and the number of shopping visits of the household, where the latter is defined as the number of retail outlets a household makes purchases from in a week. 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 panel (b) shows the distribution of number of shopping trips per household in Mexico and in the US. A typical US household makes one shopping trip per week whereas in Mexico the number is five. So the typical Mexican household makes more shopping visits and thus buy more of the cheaper products in a given store.

¹The first digits of the GS1 prefixes identify only where the barcode was issued and not the country of origin for a given product. The prefixes simply provide number capacity to different countries for assignment from that location to companies who apply. Those companies in turn may manufacture products anywhere in the world. For example, if a Mexican company imports an item from a different country, then packaged and shipped that item to the US, the country code in the GS1 prefix would likely correspond to Mexico.

Figure 2: Shopping Trips - Mexico and the US



(a) Prices and Shopping Trips

(b) Distribution of Shopping Trips

Notes: Panel (a) plots the demeaned price paid for a product and the number of shopping trips. Relative prices are measured in a regression of the log unit price paid against dummies for the deciles of log shopping trips and category, store, and quarter fixed effects. Panel (b) shows the distribution of shopping trips per household and week. A shopping trip is defined as a date-store combination. The distributions are plotted using the projection factor provided by the data to make the sample of household representative to that of each country.

Fact 2: Prices of comparable products are more expensive in the US compared to Mexico, but this pattern is attenuated for products with the same barcode.

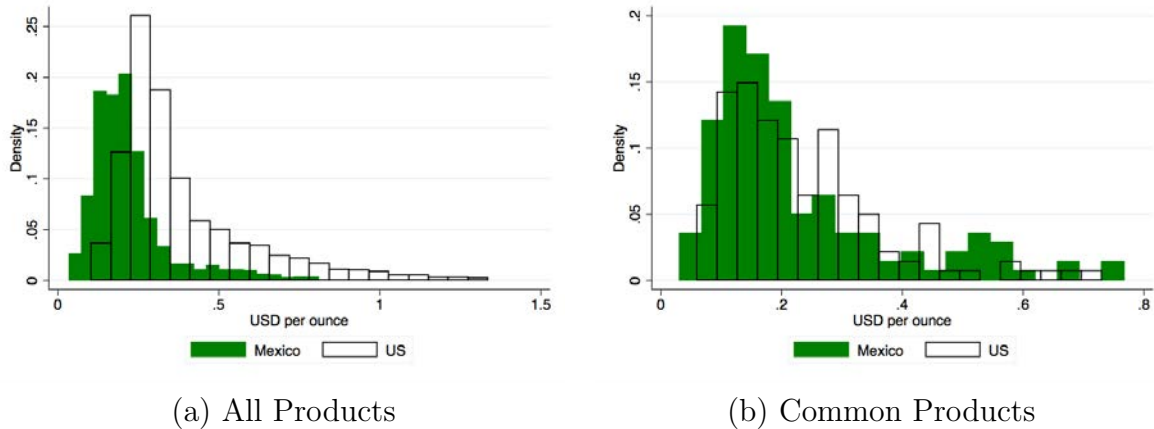
Next, we compare the distribution of prices across the two countries within an item. We compare the unit prices of products across the two countries in each category. Table 1 shows the distribution of the percent difference in prices of products (barcodes) in Mexico vs. the US within a Nielsen product category. It shows that, if we consider all products available in both countries, the average price in the US is 23% higher. The second row shows the distribution of the price gap only for the products with the same barcodes. When we consider only products that are sold in both countries, the difference reduces to 8%.

Table 1: Percent Difference in Prices: Mexico vs US

Sample	Mean	25%	Median	75%
All	-23	-48	-26	-8
Matched	-8	-23	-3	16

Notes: The table reports the mean percent difference in prices between Mexico and the US. We compute the average unit price for products in each category. Then we take the average across categories. The first row considers all products available in the two countries within a category. The second row considers only the products matched at the barcode level. Negative numbers indicate that the prices in the US are higher. The differences are reported in percent.

An example of these patterns can be seen in the product category “Cheese”. Panel (a) in Figure 3 shows distribution of product prices in the US and Mexico. It shows that, on average, cheese in the US is more expensive. Panel (b) shows that, for common products (same barcode) within the category of cheese, the distributions of product prices of the two countries are very similar and that the difference in mean prices is reduced.

Figure 3: Price Distribution for All and Common Products: Cheese

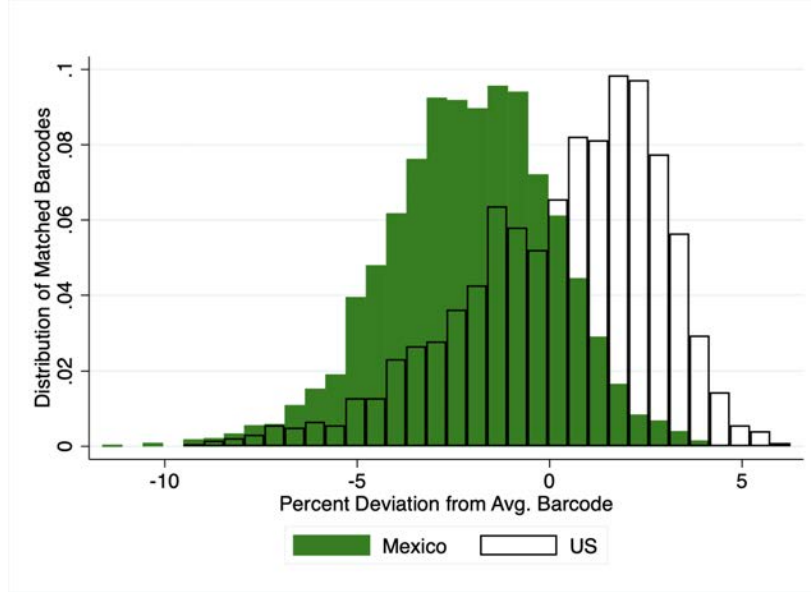
Notes: Panel (a) shows price distribution of products in cheese category for all products in the US and Mexico. Panel (b) describes price distribution of common products that are available in both countries.

Fact 3: Mexican households spend less on products available in Mexico and the US.

Lastly, we document the importance of the common products with the same barcodes in the consumption basket of households in the two countries. Figure 4 shows the distribution

of total household spending on the products present in Mexico and the US relative to the average product in each country. US households spend relatively more on the typical common products, but the opposite is true in Mexico.

Figure 4: Distribution of Revenue - Common Products



Notes: The figure shows the distribution of revenue generated by the common products relative to the average product of each country. The total revenue generated by each product is computed using the projection factor provided by the data in order to make the sample of household representative to that of each country.

4 Theoretical Framework

In this section, we derive the ideal price index. We then decompose the difference between our ideal price index and that provided by the ICP into the contribution of sampling, quality, and variety bias.

4.1 Preferences and Price Index

Aggregate utility of a representative household in Mexico is given by:

$$U^m = \left[\sum_{b \in \Omega} \left(\sum_{i \in \Omega_b} C_{ib}^m \frac{\eta_b - 1}{\eta_b} \right)^{\frac{\eta_b - 1}{\eta_b - 1} \frac{\gamma - 1}{\gamma}} \right]^{\frac{\gamma}{\gamma - 1}} \quad (1)$$

where C_{ib}^m denotes consumption of item i in basic heading b (superscript m stands for Mexico), γ is the elasticity of substitution across basic headings, η_b is the elasticity of substitution

across items in basic heading b , Ω is the set of all basic headings common across the two countries, and Ω_b is the set of all items within a basic heading b are common across countries. Note that Ω and Ω_b do not have superscripts for country because basic headings and items are common across countries in the ICP methodology.

Aggregate consumption of an *item* is then a CES aggregate of consumption of individual barcodes indexed by k :

$$C_{ib}^m = \left(\sum_{i \in \Omega_{ib}^m} (\varphi_{kib} C_{kib}^m)^{\frac{\sigma_{ib}-1}{\sigma_{ib}}} \right)^{\frac{\sigma_{ib}}{\sigma_{ib}-1}} \quad (2)$$

where C_{kib}^m denotes total physical units and φ_{kib} the quality of barcode k and σ_{ib} is the elasticity of substitution between barcodes in basic heading b and item i . In contrast to the set of basic headings and items which are common to both countries, the set of barcodes within each item Ω_{ib}^m is country specific and thus has a superscript for the country. Finally, to allow for the fact that the same product (barcode) is sold in different stores, we assume:

$$C_{kib}^m = \sum_{s \in \Omega_{kib}^m} (C_{skib}^m)^{\phi_{skib}^m} \quad (3)$$

Here Ω_{kib}^m denote the set of stores that sell barcode k .

The ideal price of barcode k in Mexico is given by:

$$p_{kib}^m = \prod_{s \in \Omega_{kib}^m} (p_{skib}^m)^{\phi_{skib}^m} \quad (4)$$

Note that the weights are barcode-store specific. The exact price index (EPI) of an item in Mexico, taking the US as the numeraire, is then given by:

$$\text{EPI}_{ib} = \left(\frac{\lambda_{ib}^m}{\lambda_{ib}^u} \right)^{\frac{1}{\sigma_{ib}-1}} \times \prod_{k \in \Omega_{ib}} \left(\frac{p_{kib}^m}{p_{kib}^u} \right)^{\omega_{kib}} \quad (5)$$

where $\frac{p_{kib}^m}{p_{kib}^u}$ is the ratio of the price in Mexico relative to the US of barcode k of item i in basic heading b (superscript u stands for the US), λ_{ib}^m and λ_{ib}^u are the share of spending on the barcodes common to the US and Mexico in total spending in Mexico and the US, defined as:

$$\lambda_{ib}^m \equiv \frac{\sum_{\Omega_{ib}} p_{kib}^m C_{kib}^m}{\sum_{\Omega_{ib}^m} p_{kib}^m C_{kib}^m} \quad \text{and} \quad \lambda_{ib}^u \equiv \frac{\sum_{\Omega_{ib}} p_{kib}^u C_{kib}^u}{\sum_{\Omega_{ib}^u} p_{kib}^u C_{kib}^u} \quad (6)$$

where Ω_{ib} is the set of barcodes available in the two countries and Ω_{ib}^m and Ω_{ib}^u are the set of *all* barcodes in Mexico and the United States, and ω_{kib} is the logarithmic mean of the

expenditure shares of the common barcodes in the two countries.² The exact price index (EPI) of each item can then be aggregated into an aggregate exact price index:

$$\text{EPI} = \prod_{b \in \Omega_b} \prod_{i \in \Omega_{ib}} \text{EPI}_{ib}^{\omega_{ib} \omega_b} \quad (7)$$

where ω_{ib} is the logarithmic mean of the expenditure share of item i of basic heading b and ω_b is the same for basic heading b .

4.2 Decomposition of Exact Price Index

We can now decompose the gap between the exact price index and the one constructed by the ICP into the contribution of sampling, quality, and variety biases. We first use the Nielsen data to construct a price index that mimics that of ICP. First, for each ICP item, we pick the product in a given store that meets the specifications provided by the ICP and with the largest volume of sales in each store; \bar{p}_{sib}^m denotes the price of this item. Since the ICP aggregates store specific prices using the store's *total* sales, we calculate the weighted average of the price of the product chosen in each store, where Ψ^m is the set of stores sampled in Mexico and the weights ϕ_s^m are the total sales of the store.

$$\bar{p}_{ib}^m = \prod_{s \in \Psi^m} (\bar{p}_{sib}^m)^{\phi_s^m}$$

Lastly, we take the geometric average across items within a basic heading using equal weights, since the ICP only has expenditures at basic heading level and not at the item level. The ICP price index for basic heading b (ICP_b^m) is therefore:

$$\text{ICP}_b = \left(\prod_{i \in \Omega_b} \frac{\bar{p}_{ib}^m}{\bar{p}_{ib}^u} \right)^{\frac{1}{N_b}} \quad (8)$$

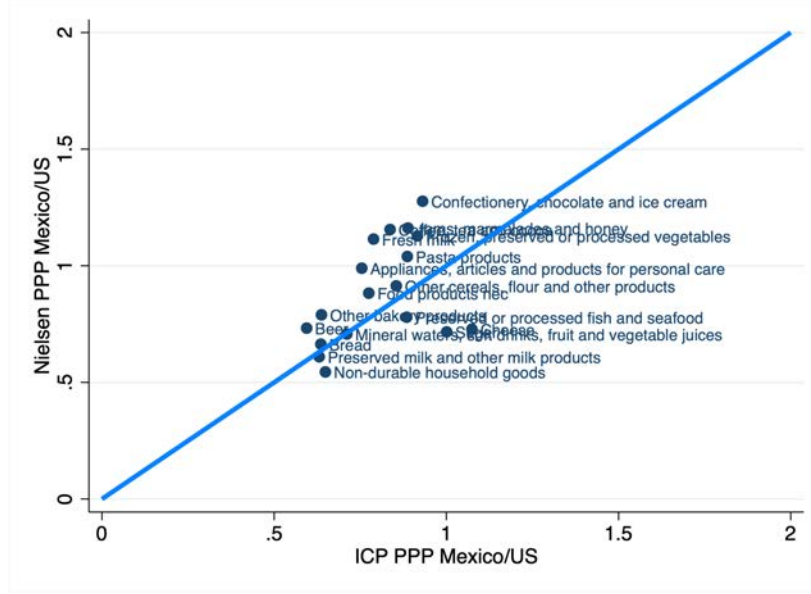
where N_b is the number of items within a basic heading b .³ Figure 5 shows the scatterplot of the ratio of the price index of a basic heading calculated from equation 8 vs. the number published by the ICP. Although the underlying data are different (from the Nielsen data and from the ICP price surveys), the Nielsen-ICP price index aligns well with the price index

² The logarithmic mean is defined as $\omega_{kib} = \frac{s_{kib}^m - s_{kib}^u}{\ln s_{kib}^m - \ln s_{kib}^u} / \sum_{k \in \Omega_{ib}} \frac{s_{kib}^m - s_{kib}^u}{\ln s_{kib}^m - \ln s_{kib}^u}$ where $s_{kib}^m = \frac{p_{kib}^m C_{kib}^m}{\sum_{k \in \Omega_{ib}} p_{kib}^m C_{kib}^m}$ and $s_{kib}^u = \frac{p_{kib}^u C_{kib}^u}{\sum_{k \in \Omega_{ib}} p_{kib}^u C_{kib}^u}$

³ Since we are comparing the price indexes of only two countries, this procedure is isomorphic to using the Country Product Dummy (CPD) method estimated by the ICP.

published by the ICP.

Figure 5: Nielsen vs ICP PPP



Notes: The figure plots the ICP for Mexico and the United States against the Nielsen-ICP, which is an index generated using the same methodology as the ICP but using the Nielsen data.

The exact 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:

$$\text{EPI}_b^m = \Theta_b^m \times \text{ICP}_b^m \quad (9)$$

$$\Theta_b^m \equiv \mathbb{S}_b^m \times \mathbb{Q}_b^m \times \mathbb{V}_b^m$$

where Θ_b^m is the aggregate bias correction term and \mathbb{S}_b^m , \mathbb{Q}_b^m and \mathbb{V}_b^m represent the sampling bias, the quality bias, and the variety bias, respectively.

Sampling Bias, \mathbb{S}_b^m : Due to data limitations, the ICP faces two potential biases in its sampling procedure. First, at the item level, the ICP uses an equal-weights geometric average because they lack data on expenditures at the item level. Second, the sample probabilities are proportionate to the store volume of sales, and sales may not reflect the sales of individual items. After a store is selected to be sampled, it is good practice that the price collectors gather the prices of as many items are available in the store to reduce the collection costs even if the sales of an individual item are not well represented in the stores sampled. Therefore, the sampling bias (\mathbb{S}_b^m) is the product of these two biases and is defined as follows:

$$\mathbb{S}_b^m \equiv \left(\left(\prod_{i \in \Omega_b} \frac{\bar{p}_{ib}^m}{\bar{p}_{ib}^u} \right)^{\frac{-1}{N_b}} \times \prod_{i \in \Omega_b} \left(\frac{\bar{p}_{ib}^m}{\bar{p}_{ib}^u} \right)^{\omega_{ib}} \right) \times \left(\prod_{i \in \Omega_b} \left(\frac{\hat{p}_{ib}^m / \bar{p}_{ib}^m}{\hat{p}_{ib}^u / \bar{p}_{ib}^u} \right)^{\omega_{ib}} \right) \quad (10)$$

where

$$\bar{p}_{ib}^m = \prod_{s \in \Psi^m} (\bar{p}_{sib}^m)^{\phi_s^m} \quad \text{and} \quad \hat{p}_{ib}^m \equiv \prod_{s \in \Psi^m} (\bar{p}_{sib}^m)^{\phi_{sib}^m}$$

as before \bar{p}_{ib}^m is the item level price computed using the ICP methodology (using ϕ_s^m as weights) and \hat{p}_{ib}^m is computed aggregating prices from multiple stores using weights defined at the item \times store level (ϕ_{sib}^m).

The first component of the sampling bias captures the bias due to the fact that the ICP does not have information of the expenditures of each item. The size of this term, for a country c , depends on the covariance of the expenditure weights of each item and their prices $\text{cov}(\omega_b, \ln(\bar{\mathbf{p}}_b^c))$ where ω_b is a vector of item-specific expenditure weights in basic heading b and $\ln(\bar{\mathbf{p}}_b^c)$ is the vector of log prices.⁴ The covariance is negative if people spend more on items with lower prices.

The second component of the sampling bias captures the importance of aggregating prices using weights reflecting the importance of each item in each store. For example, a fruit market may also sell milk, but its volume of sales is more reflective of the fruit's sales. This bias is larger than 1 for a country c if $\text{cov}(\phi_{ib}^c, \ln(\bar{\mathbf{p}}_{ib}^c)) > \text{cov}(\phi^c, \ln(\bar{\mathbf{p}}_{ib}^c))$ where ϕ_{ib}^c is the vector of expenditure weights that vary by item and basic heading, (i.e. $\{\phi_{sib}^c\}_{s \in \Psi^c}$), ϕ^c is the vector of weights that only vary at the store level (i.e. $\{\phi_s^c\}_{s \in \Psi^c}$), and $\ln(\bar{\mathbf{p}}_{ib}^c)$ is the vector of log prices (i.e. $\{\ln(\bar{p}_{sib}^c)\}_{s \in \Psi^c}$).⁵ Given that we expect covariances to be negative, the second term of the sampling bias, all else equal, will be less than 1 if the consumers in Mexico tend to buy a larger share of their products at stores where they are cheaper relative to the consumers in the US.

Quality Bias, Q_b^m : Lower quality products in Mexico might get matched with higher quality products in the United states due to the fact that the distribution of prices in the US within an item has a longer right tail. This could lead to an understatement of the price levels in Mexico and comes from the fact that price collectors identify products in the outlets that match the product specifications given by the ICP but these specification are often on purpose not very narrow in order to increase the overlap of items across countries. We measure the extent of the quality bias by comparing the prices collected by the ICP (under

⁴Propositin 1 in the Appendix shows formally this relationship.

⁵See Proposition 2.

the aggregation that already corrects for the sampling weights) and the average price of an item after considering all the barcodes available in a given country:

$$\mathbb{Q}_b^m \equiv \prod_{i \in \Omega_b} \left(\left(\frac{\hat{p}_{ib}^m}{\hat{p}_{ib}^u} \right)^{-1} \times \prod_{k \in \Omega_{ih}} \left(\frac{p_{kib}^m}{p_{kib}^u} \right)^{\omega_{kib}} \right)^{\omega_{ib}} \quad (11)$$

The finer level of disaggregation in our data allows us to compare the price differences across the two countries comparing products with the exact same physical attributes. This is because any change in an attribute of a good (e.g. form, size, package, formula) results in a new barcode. Note that the quality bias would be bigger than one if the ICP is matching lower quality items in Mexico to higher quality items in the US.

Variety Bias, \mathbb{V}_b^m : This last bias measures both the cross-country differences in availability of products and is defined as follows:

$$\mathbb{V}_b^m \equiv \prod_{i \in \Omega_b} \left(\frac{\lambda_{ib}^m}{\lambda_{ib}^u} \right)^{\frac{\omega_{ib}}{\sigma_{ib} - 1}} \quad (12)$$

which is the ratio of the share of common products in Mexico relative to the share of common products in US as in [Feenstra \(1994\)](#). In our context, if this term is less than one, it indicates that Mexican consumers spend a lower share of their expenditures in goods that are common in the two countries.

5 Results

In this section, we first describe the procedure to estimate the elasticity of substitution between barcodes within each item. We then use our decomposition to quantify the importance of each the biases.

5.1 Parameter Estimation

In order to obtain the elasticity of substitution, σ_{ib} , for each items, 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 a product category. Specifically, we estimate the demand elasticities by using the following system of differenced demand and supply equations as in [Broda and Weinstein \(2006\)](#):

$$\Delta^{k,t} \ln S_{kibt} = (1 - \sigma_{ib}) \Delta^{k,t} \ln P_{kibt} + \iota_{kibt} \quad (13)$$

$$\Delta^{k,t} \ln P_{kibt} = \frac{\delta_{ib}}{1 + \delta_{ib}} \Delta^{k,t} \ln S_{kibt} + \kappa_{kibt} \quad (14)$$

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 σ_g . Equations 13 and 14 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 ι_{kibt} and κ_{kibt} are uncorrelated (i.e., $\mathbb{E}_t(\iota_{kibt}\kappa_{kibt}) = 0$). 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 σ_{ib} and ω_{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:

$$G(\beta_{ib}) = E_T[\nu_{kibt}(\beta_{ib})] = 0 \quad (15)$$

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:

$$\hat{\beta}_{ib} = \arg \min_{\beta_{ib} \in B} G^*(\beta_{ib})' W G^*(\beta_{ib}) \quad \forall i \in \omega_b \quad (16)$$

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 β_{ib} (i.e., $\sigma_{ib} > 0$).

As [Argente, Lee and Moreira \(2020\)](#), we estimate σ_{ib} and δ_{ib} jointly with Nielsen Retail Scanner data from 2006 to 2016, which covers approximately 90 participating retail chains across all US markets.⁶ 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 8.94 with standard deviation of 2.95.

⁶For more details on the Nielsen Retail Scanner data, refer to [Argente, Lee and Moreira \(2018\)](#) and [Argente and Lee \(forthcoming\)](#).

5.2 Decomposition Results

We estimate the exact price index and decompose it using equation 9. We use 30 items where there are more than five common barcodes between US and Mexico.⁷ Table 2 reports the price index developed by the ICP (ICP^m) and the exact price index (EPI^m) after aggregating across basic headings. Recall that the ICP aggregates across basic headings using a Fisher index. In the table, we report both the results after aggregating basic headings using our theoretical framework which implies a CES aggregation and using a Fisher index in the first layer. Equation 9 indicates that the gap between the two price indexes can be decomposed into the sampling bias (\mathbb{S}^m), the quality bias (\mathbb{Q}^m), and the variety bias (\mathbb{V}^m). The aggregate bias (Θ^m) is estimated to be 0.91.⁸

Table 2: Decomposition Results for the Exact Price Index

	Price Indexes		Biases			Agg. Bias
	ICP^m	EPI^m	\mathbb{S}^m	\mathbb{Q}^m	\mathbb{V}^m	Θ^m
CES	0.84	0.76	0.85	1.35	0.80	0.91
Fisher	0.88	0.86	0.88	1.38	0.81	0.98

Notes: The table reports ICP-style price index (ICP^m) and the exact price index (EPI^m). By equation 9, the gap between two price indexes can be decomposed into sampling bias (\mathbb{S}^m), quality bias (\mathbb{Q}^m), and variety bias (\mathbb{V}^m). Θ^m is the aggregate bias defined as $\mathbb{S}^m \times \mathbb{Q}^m \times \mathbb{V}^m$. We report two different way of aggregations at the top level of aggregation: CES aggregation consistent to the model and Fisher aggregation as a robustness check.

The sampling bias is 0.85. We do not find a significant difference across the two countries in the first term of the sampling bias, which is related to the covariance of the expenditure weights at the item level and their prices within a basic heading. This is due to the fact that there are not many items within a basic heading, so the difference between aggregating with expenditure weights or with equal weights within a basic heading is not significantly different across the two countries.

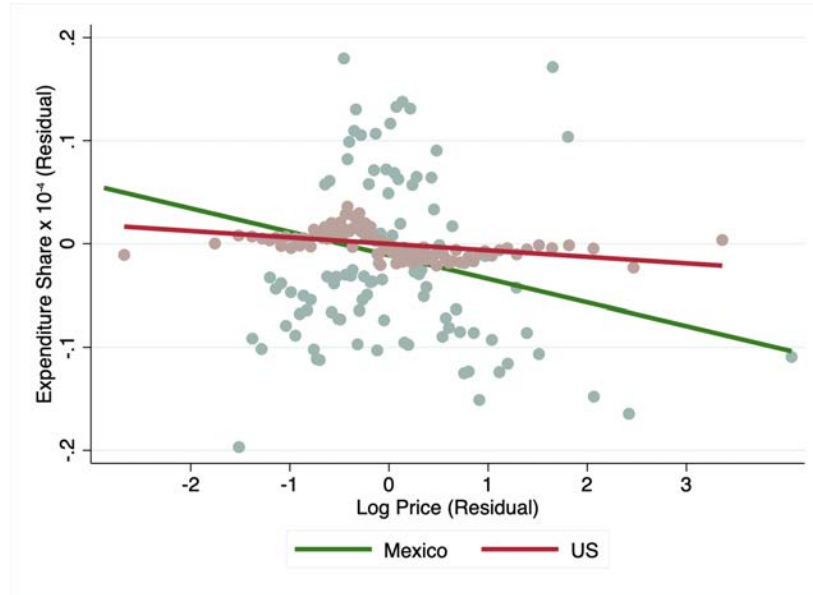
The direction of the sampling bias is mainly due to the importance of the second term, which is related to the fact that the sample probabilities are proportionate to the store volume of sales and do not reflect the sales of individual products. This is particularly relevant in Mexico, where we find that consumers tend to spend more at stores where specific items

⁷This covers 16 basic headings. Table A in the Appendix reports the number of common barcodes for each product category.

⁸Appendix D reports decomposition results with common elasticity of substitution, 9. The aggregate bias is estimated to be the same.

have lower prices. In other words, we find that $\text{cov}(\phi_{ib}^m, \ln(\bar{p}_{ib}^m)) < \text{cov}(\phi_{ib}^u, \ln(\bar{p}_{ib}^u))$ after controlling for item and country effects. This is shown in Figure 6 where we plot, for each country, the relationship between the expenditure shares, ϕ_{ib}^m , and prices, $\ln(\bar{p}_{ib}^u)$, at the item-store level. The figure shows that, after controlling for item fixed effects, the relationship is negative for both countries but more so for Mexico.⁹ This is related to the first fact we document where we show that households in Mexico shop more frequently. By increasing the number of shopping trips and stores they visit, Mexican households are, on average, better able to find items where they are cheaper than households in the US.

Figure 6: Sampling Bias: Covariance between Expenditure Weights and Prices



Notes: The figure shows, for each country, the relationship between the expenditure shares and the logarithm of prices at the item-store level. The green line reports the relationship between these variables for Mexico and the red line for the US. The variables are plotted after controlling for item fixed effects over 100 bins of log prices, each representing a percentile in the distribution.

The quality bias (1.35) constitutes a downward bias to the original ICP estimates; an understatement of the price level in Mexico. This is because low quality varieties in Mexico are matched to high quality varieties in the United States, even after sampling items that meet the same ICP specifications across the two countries. The direction and size of this bias is related to the second fact we document; it arises from the fact that prices of comparable products are more expensive in the US compared to Mexico; within the same item, products in the US have a higher mean and a longer right tail. As a result, a price collector is

⁹In Appendix B.2 we provide more evidence of this relationship using a reduced form approach.

more likely to sample a higher quality product in the US even if it matches the product specifications defined by the ICP program. Given that the differences in prices of products across the two countries are attenuated for products with the same barcode, the quality bias is likely to be larger for items where the ICP specification is more broadly defined.

Lastly, Mexico has important gains from variety (0.80) relative to the US. The importance of the variety bias reflects the fact that, as shown in our third fact, the consumption shares of the common goods across the two countries is lower in Mexico than in the United States. In other words, Mexican households value more varieties that are not present in the US market. In fact, in our data, the majority of the products that are common in both countries come from firms based in the US and that have higher market shares in the US than in Mexico.

Table 3 reports all price indexes: (i) ICP (\mathbb{ICP}^m), (ii) Sato-Vartia price index (\mathbb{SV}^m) and (iii) the exact price index (\mathbb{EPI}^m). Compared to the ICP index, the Sato-Vartia price index is higher after correcting sampling and quality biases jointly. Both biases partially upset each other given that the sampling bias is less than 1 and the quality bias is larger than 1. Recall that the Sato-Vartia index does not include the variety bias. As a result, it shows that omitting the estimation of this bias could lead to the opposite conclusions regarding the overall bias of the ICP index. This can be seen in the overall estimates of the exact price index presented in the table. After accounting for the variety bias, the ICP index overestimates the prices of Mexico relative to those in the US. These results are similar when we extend our framework to consider the non-separable class of CES functions developed by [Hanoch \(1975\)](#) and recently extended by [Comin, Lashkari and Mestieri \(forthcoming\)](#). In Appendix C we show that, when we account for an Engel-curve variety bias, which measures both the cross-country differences in availability of products and the differences in real consumption across countries, the size of the aggregate bias is similar to that in the benchmark homothetic case.

Table 3: Multiple Price Indexes

	Price Indexes		
	\mathbb{ICP}^m	\mathbb{SV}^m	\mathbb{EPI}^m
CES	0.84	0.96	0.76
Fisher	0.88	1.06	0.86

Notes: The table reports ICP-style price index (\mathbb{ICP}^m), Sato-Vartia price index (\mathbb{SV}^m) and the exact price index (\mathbb{EPI}^m). We report two different way of aggregations at the top level of aggregation: CES aggregation consistent to the model and Fisher aggregation as a robustness check.

Overall, our results show that the real non-durable consumption inequality across the US and Mexico is 9% lower than that predicted by the ICP estimates. They also highlight the importance of addressing the sampling, quality, and variety biases of the ICP estimates jointly. Addressing them in isolation could lead to drastically different conclusions about the comparison of the standards of living across Mexico and the US.

6 Conclusion

The construction of cross-country price indexes is a difficult task and yet it is of crucial importance to compare living standards between countries and to measure global inequality. The ICP faces several challenges both of data availability and also methodology. In this paper, we construct a data set for two countries that allows us to address some of the data limitations faced by the ICP; namely, the fact that they cannot compare exactly the same item across countries and that they do not have expenditure information to weight items appropriately. We also develop a exact price index to be able to account for differences in the availability of varieties across countries. Using this index we estimate that the Mexican real consumption is larger relative to the United States than previously estimated. We identify heterogeneity in shopping behavior, quality of products, and variety availability as important sources of bias in international price comparisons.

References

- Argente, David and Munseob Lee, “Cost of Living Inequality during the Great Recession,” *Journal of the European Economic Association*, forthcoming.
- , —, and Sara Moreira, “Innovation and product reallocation in the great recession,” *Journal of Monetary Economics*, 2018, *93*, 1–20.
- , —, and —, “The Life Cycle of Products: Evidence and Implications,” 2020.
- 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, Trade Costs and the Standard of Living Across Countries,” 2020.
- , 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.
- Comin, Diego A, Danial Lashkari, and Martí Mestieri, “Structural Change with Long-run Income and Price Effects,” *Econometrica*, forthcoming.
- Deaton, Angus 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 Antoniadou, “What is the Price of Tea in China? Goods Prices and Availability in Chinese Cities,” *The Economic Journal*, 2019.
- , Robert Inklaar, and Marcel P Timmer, “The next generation of the Penn World Table,” *American Economic Review*, 2015, *105* (10), 3150–82.
- 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.

- Matsuyama, Kiminori, “Engel’s Law in the Global Economy: Demand-Induced Patterns of Structural Change, Innovation, and Trade,” *Econometrica*, 2019, *87* (2), 497–528.
- 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,” *The Review of Economic Studies*, 2015, *82* (4), 1612–1656.