Measuring the Cost of Living in Mexico and the US

David Argente — Penn State

Chang-Tai Hsieh — University of Chicago

Munseob Lee — UC San Diego

July 2020, NBER SI CRIW

Price Measurement Across Countries

- Long-standing problems of measurement
  - Sampling: collected from stores instead of consumers
  - Quality: brain surgery in Nairobi vs Tokyo
  - Variety: product availability

This Paper

1. Nielsen US and Mexico
   - Representative panel of households with purchases of consumer goods.
   - Products matched at the barcode level across countries.
2. Quantify potential biases behind the ICP using non-homothetic price index
   - A new decomposition framework to quantify sampling bias, quality bias, and variety bias independently

Preview of Findings

- Price level ratio between Mexico and the US:

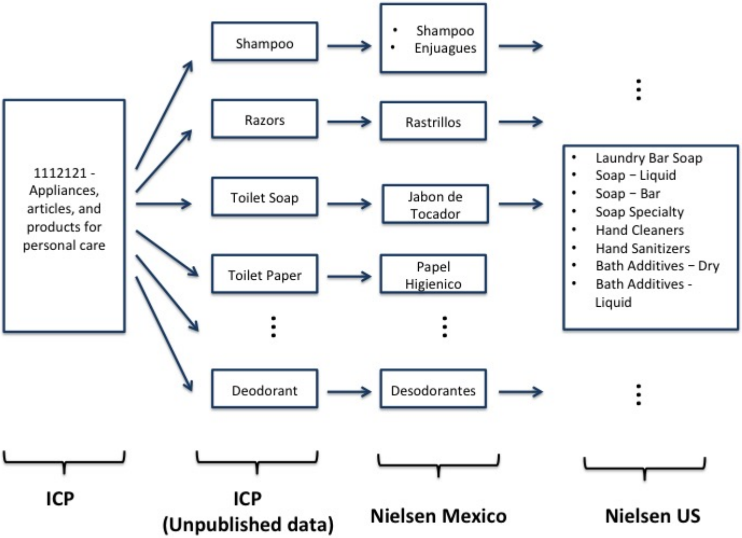
$$\mathbb{NH}^M = \Theta^M \times \mathbb{ICP}^M$$
$$0.72 = 0.90 \times 0.80$$

- $\Theta^M = \mathbb{S}^M \times \mathbb{Q}^M \times \mathbb{V}^M$
  - Sampling bias: $\mathbb{S}^M = 0.82$
  - Quality bias: $\mathbb{Q}^M = 1.45$
  - Engel-curve Variety bias: $\mathbb{V}^M = 0.75$

Data

- ICP 2011
  - 155 basic headings
  - Thousands of comparable items
- Nielsen Mexico
  - Representative sample of 5,000 households for 2012-2013.
  - Households visited biweekly report consumption diary information.
- Nielsen US
  - Representative sample of 60,000 households.
  - Panelists use in-home scanners to record their purchases.

Matched sample: Nielsen data, ICP data



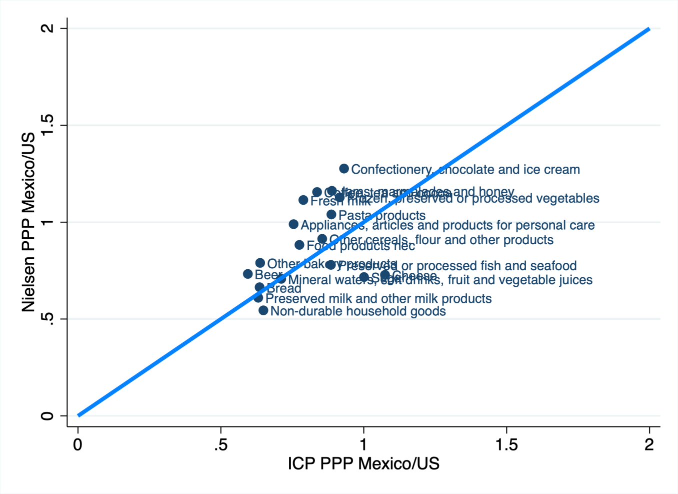
ICP Procedure: Data Validation Details

1. Among multiple barcodes for specific item in a store, pick one barcode to represent the item ($\bar{p}_{sib}$).

2. Aggregate across stores with store size weights ($\bar{p}_{ib}$).

3. Jevons index across items ($\mathbb{ICP}_b$).

Nielsen vs ICP PPP by Basic Heading



- Nielsen data mimics well the prices constructed by the ICP.

Facts on Sampling, Quality and Variety

1. Mexican households shop more frequently and visit more stores. Therefore, Mexicans buy a larger share of items at stores where they are cheaper. Details
2. The distribution of prices in the US has a higher mean and a longer right tail, but these patterns are attenuated when we compare common goods. Details
3. A significant presence of US brands in the Mexican market gives more variety to Mexican consumers. Details

Theoretical framework ▸ Engel curves

- basic headings b, items i, barcodes k, stores s
- CES aggregation across basic headings
- CES aggregation across items
- Non-homothetic CES aggregation across barcodes

$$1 = \sum_{k \in \Omega_{ib}^M} \left(\frac{\varphi_{kib}^M C_{kib}^M}{(C_{ib}^M)^{(\varepsilon_{kib} - \sigma_{ib})/(1-\sigma_{ib})}} \right)^{\frac{\sigma_{ib}-1}{\sigma_{ib}}}$$

  where ε_{kib} is the elasticity of a barcode k with respect to item-level consumption C_{ib}^M

- Cobb-Douglas Aggregation across stores.

Non-homothetic Price Index

$$\mathbb{NH}_{ib}^{M} \equiv \frac{P_{ib}^{M}}{P_{ib}^{U}} = \prod_{k \in \Omega_{ib}} \left(\frac{p_{kib}^{M}}{p_{kib}^{U}} \right)^{\omega_{kib} \frac{1}{1-\theta_{ib}}} \times \left(\frac{\lambda_{ib}^{M}}{\lambda_{ib}^{U}} \right)^{\frac{1}{\sigma_{ib}-1} \frac{1}{1-\theta_{ib}}} \times \left(\frac{E_{ib}^{M}}{E_{ib}^{U}} \right)^{\frac{\theta_{ib}}{\theta_{ib}-1}}$$

where

$$\theta_{ib} \equiv \sum_{k \in \Omega_{ib}} \omega_{kib} \frac{\varepsilon_{kib} - 1}{\sigma_{ib} - 1}$$

- Sato-Vartia index across common barcodes
- Variety correction
- Engel-curve adjustment

Decomposition of Non-homothetic Price Index

$$\mathbb{NH}_b^M = \Theta_b^M \times \mathbb{ICP}_b^M$$
$$\Theta_b^M \equiv \mathbb{S}_b^M \times \mathbb{Q}_b^M \times \mathbb{V}_b^M$$

- $\mathbb{S}_b^M$: Sampling Bias
- $\mathbb{Q}_b^M$: Quality Bias
- $\mathbb{V}_b^M$: Engel-curve Variety Bias

$\mathbb{S}_b^M$: Sampling Bias

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

where

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

- Bias comes from missing expenditures for each item.
- Bias depends on covariance between expenditures and prices across items.
  - No significant difference between two countries ▸ Details

$\mathbb{S}_b^M$: Sampling Bias

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

where

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

- Bias comes from missing expenditures for each item at each store.
- Bias depends on covariance between expenditures and prices across stores.
  - Significant difference between two countries ▸ Details

$\mathbb{Q}_b^M$: Quality Bias, $\mathbb{V}_b^M$: Engel-curve Variety Bias

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

$$\mathbb{V}_b^M = \prod_{i \in \Omega_b} \left(\frac{\lambda_{ib}^M}{\lambda_{ib}^U} \right)^{\frac{\omega_{ib}}{\sigma_{ib}-1}} \times \prod_{i \in \Omega_b} \left[\left(\frac{E_{ib}^M / E_{ib}^U}{\mathbb{EPI}_{ib}^M} \right)^{\frac{\omega_{ib}\theta_{ib}}{\theta_{ib}-1}} \right]$$

Parameter Estimation

- GMM estimation for σ_{ib} as Broda and Weinstein (2006,2010) ▸ Details
  - mean 9.29, std.dev. 3.42
- Given σ_{ib} estimates, we use the Engel curve to estimate ε_{kib} as Comin et al. (2020). ▸ Details
  - mean 0.82, std.dev. 1.79

Conclusion

- Price level ratio between Mexico and the US:

$$\mathbb{NH}^M = \Theta^M \times \mathbb{ICP}^M$$
$$0.72 = 0.90 \times 0.80$$

- $\Theta^M = \mathbb{S}^M \times \mathbb{Q}^M \times \mathbb{V}^M$
  - Sampling bias: $\mathbb{S}^M = 0.82$
    - Mexicans buy a larger share of items at stores where they are cheaper.
  - Quality bias: $\mathbb{Q}^M = 1.45$
    - Low quality products in Mexico matched to high quality products in US.
  - Engel-curve Variety bias: $\mathbb{V}^M = 0.75$
    - A significant presence of US brands in the Mexican market gives more variety to Mexican consumers.

- Real non-durable consumption in Mexico relative to US is 10 percent higher than previously estimated.

Nielsen vs ICP Back

- Follow the procedures followed by ICP 2011
- Use the categories matched between Nielsen and ICP
  1. Select a single item i in country j
  2. Aggregate across stores using expenditure weights
  3. Estimate:

$$\log \bar{p}_{ib}^{c} = \eta_i^c + \eta_b^c + \varepsilon_{ib}^c$$

     where $\bar{p}_{ib}^{c}$ is the price of item i belonging to heading b in country c
  4. The estimated PPP for basic heading b and country is: $\bar{p}_b^c = exp(\eta_b^c)$.

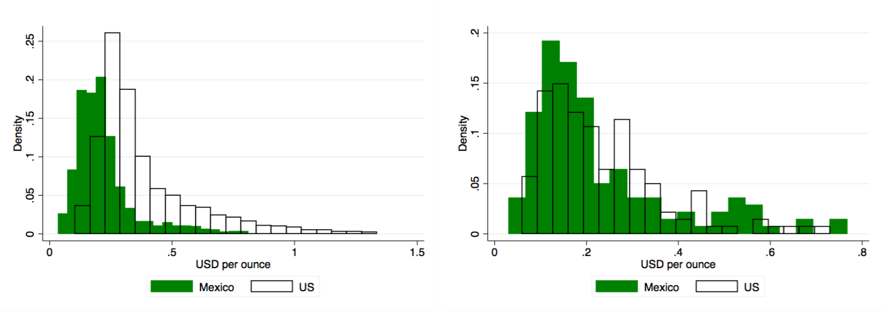
Fact 1: Sampling ◂ Back

Average number of shopping trip per week is 5 in Mexico and 1 in the US.

Fact 2: Quality, example of cheese

Back

The distribution of prices in the US has a higher mean and a longer right tail, but these patterns are attenuated when we compare common goods.

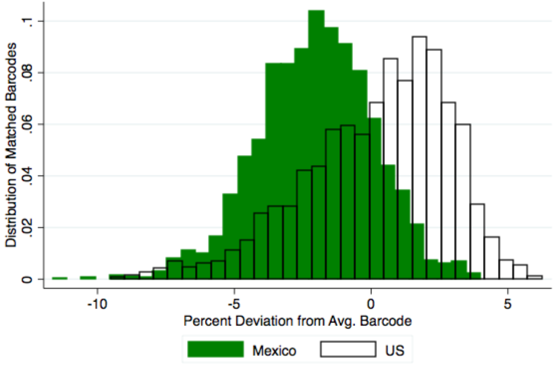


(a) All Products

(b) Overlapping Products

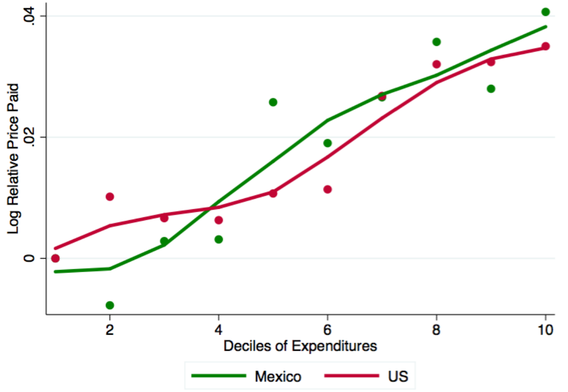
Fact 3: Variety ◂ Back

Mexican households spend less on overlapping products.



Non-homotheticity ◂ Back

Within category of products, richer households buy more expensive products.



Sampling Bias: $\mathrm{cov}(\omega_{\mathbf{b}}, \ln(\bar{\mathbf{p}}_b^c))$ Back

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

	(1)	(2)	(3)	(4)
$\ln(\bar{p})$	-0.010	-0.044	-0.010	-0.039
	(0.071)	(0.051)	(0.071)	(0.059)
$\ln(\bar{p}) \times$ Mexico	-0.002	0.004	-0.002	-0.006
	(0.095)	(0.023)	(0.095)	(0.056)
Observations	58	58	58	58
R-squared	0.001	0.775	0.001	0.775
Basic Heading	N	Y	N	Y
Country	N	N	Y	Y

Sampling Bias: $\text{cov}(\boldsymbol{\phi}^{\mathbf{c}}_{\mathbf{ib}}, \ln(\bar{\mathbf{p}}^{c}_{ib}))$ Back

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

	(1)	(2)	(3)	(4)
$\ln(\bar{p})$	-0.000	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(\bar{p}) \times$ Mexico	-0.002***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	764,419	764,419	761,751	761,750
R-squared	0.028	0.030	0.212	0.212
Store	N	N	Y	Y
Country	N	Y	N	Y

Estimation of σ_{ib} Back

- Feenstra (1994), Broda and Weinstein (2006, 2010)
- Double-difference log UPC expenditure shares and UPC pricing rule over time and relative to the largest UPC within each firm.

$$\Delta^{\underline{u},t} \ln S_{kibt} = (1 - \sigma_{ib}) \Delta^{\underline{u},t} \ln P_{kibt} + \omega_{kibt}$$

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

$\omega_{kibt} = [\Delta^t \ln \varphi_{kibt} - \Delta^t \ln \varphi_{\underline{k}ibt}]$ and $\kappa_{kibt} = [\Delta^t \ln a_{kibt} - \Delta^t \ln a_{\underline{k}ibt}]$

Estimation of σ_{ib} Back

- Orthogonality of the double-differenced demand and supply shocks defines a set of moment conditions:

$$G(\beta_g) = E_T[v_{kibt}(\beta_g)] = 0$$

where $\beta_g = [\sigma_{ib}, \delta_{ib}]'$ and $v_{kibt} = \kappa_{kibt}\omega_{kibt}$.

- We proceed with GMM.

Parameter Estimation ◂ Back

- Given σ_{ib} estimates, we use the Engel curve to estimate ε_{kib} as Comin et al. (2020):

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

  where $\mathbf{K}$ is the benchmark barcode, which corresponds to the largest selling barcode in each item, and ψ^h_t is the set of fixed effects and controls.

Parameter Estimation ◂ Back

	mean	std. dev.	10th-percentile	median	90th-percentile
σ_i	9.29	3.42	5.61	8.73	12.49
ε_{kib}	0.82	1.79	-1.58	0.89	2.99
θ_{ib}	0.02	0.13	-0.13	0.03	0.16
E_{ib}^M / E_{ib}^U	0.72	0.60	0.14	0.53	1.47