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:
  \[ NH^M = \Theta^M \times ICP^M \]
  \[ 0.72 = 0.90 \times 0.80 \]

- \( \Theta^M = S^M \times Q^M \times V^M \)
  - Sampling bias: \( S^M = 0.82 \)
  - Quality bias: \( Q^M = 1.45 \)
  - Engel-curve Variety bias: \( V^M = 0.75 \)
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
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 ($\text{ICP}_b$).
Nielsen data *mimics well* the prices constructed by the ICP.
Mexican households shop more frequently and visit more stores. Therefore, Mexicans buy a larger share of items at stores where they are cheaper.

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 significant presence of US brands in the Mexican market gives more variety to Mexican consumers.
Theoretical framework

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

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

where $\varepsilon_{kib}$ 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

\[ \text{NH}^M_{ib} \equiv \frac{P^M_{ib}}{P^U_{ib}} = \prod_{k \in \Omega_{ib}} \left( \frac{p^M_{kib}}{p^U_{kib}} \right)^{\omega_{kib} \frac{1}{1-\theta_{ib}}} \times \left( \frac{\lambda^M_{ib}}{\lambda^U_{ib}} \right)^{\frac{1}{\sigma_{ib}-1} \frac{1}{1-\theta_{ib}}} \times \left( \frac{E^M_{ib}}{E^U_{ib}} \right)^{\frac{\theta_{ib}}{\theta_{ib}-1}} \]

where

\[ \theta_{ib} \equiv \sum_{k \in \Omega_{ib}} \omega_{kib} \frac{e_{kib} - 1}{\sigma_{ib} - 1} \]

- **Sato-Vartia index across common barcodes**
- **Variety correction**
- **Engel-curve adjustment**
Decomposition of Non-homothetic Price Index

\[ \text{NH}^M_b = \Theta^M_b \times \text{ICP}^M_b \]

\[ \Theta^M_b \equiv S^M_b \times Q^M_b \times V^M_b \]

- \( S^M_b \): Sampling Bias
- \( Q^M_b \): Quality Bias
- \( V^M_b \): Engel-curve Variety Bias
\( S_b^M: \) Sampling Bias

\[
S_b^M \equiv \left( \prod_{i \in \Omega_b} \frac{\bar{p}_{ib}^M}{\bar{p}_{ib}} \right)^{-\frac{1}{N_b}} \times \prod_{i \in \Omega_b} \left( \frac{\bar{p}_{ib}^M}{\bar{p}_{ib}} \right)^{\omega_{ib}} \times \prod_{i \in \Omega_b} \left( \frac{\hat{p}_{ib}^M / \bar{p}_{ib}^M}{\hat{U}_{ib} / \bar{p}_{ib}^M} \right)^{\omega_{ib}}
\]

where

\[
\bar{p}_{ib}^M = \prod_{s \in \Psi^M} \left( \bar{p}_{sib}^M \right)^{\phi_{sib}^M} \quad \text{and} \quad \hat{p}_{ib}^M = \prod_{s \in \Psi^M} \left( \hat{p}_{sib}^M \right)^{\phi_{sib}^M}
\]

- Bias comes from missing expenditures for each item.
- Bias depends on covariance between expenditures and prices across items.
  - No significant difference between two countries
\( S_M^b: \text{Sampling Bias} \)

\[
S_b^M \equiv \left( \left( \prod_{i \in \Omega_b} \frac{\bar{p}_{ib}^M}{\bar{p}_{ib}} \right)^{-\frac{1}{N_b}} \times \prod_{i \in \Omega_b} \left( \frac{\bar{p}_{ib}^M}{\bar{p}_{ib}} \right)^{\omega_{ib}} \right) \times \left( \prod_{i \in \Omega_b} \left( \frac{\hat{p}_{ib}^M / \bar{p}_{ib}^M}{\hat{p}_{ib} / \bar{p}_{ib}} \right)^{\omega_{ib}} \right)
\]

where

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

- 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]
\( Q^M_b \): Quality Bias, \( \nabla^M_b \): Engel-curve Variety Bias

\[
Q^M_b = \prod_{i \in \Omega_b} \left( \left( \frac{\hat{p}^M_{ib}}{\hat{U}_{ib}} \right)^{-1} \times \prod_{k \in \Omega_{ib}} \left( \frac{p^M_{kib}}{p^M_{kib}} \right)^{\omega_{kib}} \right)^{\omega_{ib}}
\]

\[
\nabla^M_b = \prod_{i \in \Omega_b} \left( \frac{\lambda^M_{ib}}{\lambda^U_{ib}} \right)^{\frac{\omega_{ib}}{\sigma_{ib} - 1}} \times \prod_{i \in \Omega_b} \left[ \left( \frac{E^M_{ib} / E^U_{ib}}{EPI^M_{ib}} \right)^{\frac{\omega_{ib}\theta_{ib}}{\theta_{ib} - 1}} \right]
\]
Parameter Estimation

- GMM estimation for $\sigma_{ib}$ as Broda and Weinstein (2006, 2010)
  - mean 9.29, std.dev. 3.42

- Given $\sigma_{ib}$ estimates, we use the Engel curve to estimate $\varepsilon_{kib}$ as Comin et al. (2020).
  - mean 0.82, std.dev. 1.79
Conclusion

- Price level ratio between Mexico and the US:
  \[ \text{Nh}^M = \Theta^M \times \text{ICP}^M \]
  \[ 0.72 = 0.90 \times 0.80 \]

- \[ \Theta^M = S^M \times Q^M \times V^M \]
  - Sampling bias: \( S^M = 0.82 \)
    - Mexicans buy a larger share of items at stores where they are cheaper.
  - Quality bias: \( Q^M = 1.45 \)
    - Low quality products in Mexico matched to high quality products in US.
  - Engel-curve Variety bias: \( V^M = 0.75 \)
    - 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.
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:
   \[\log \bar{p}_{ib}^c = \eta_i^c + \eta_b^c + \epsilon_{ib}^c\]
   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: $\bar{p}_b^c = \exp(\eta_b^c)$. 
Average number of shopping trip per week is 5 in Mexico and 1 in the US.
Fact 2: Quality, example of cheese

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

Mexican households spend less on overlapping products.
Within category of products, richer households buy more expensive products.
Sampling Bias: $\text{cov}(\omega_b, \ln(\tilde{p}_b^c))$

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

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln(\bar{p})$</td>
<td>-0.010</td>
<td>-0.044</td>
<td>-0.010</td>
<td>-0.039</td>
</tr>
<tr>
<td>(0.071)</td>
<td>(0.051)</td>
<td>(0.071)</td>
<td>(0.059)</td>
<td></td>
</tr>
<tr>
<td>$\ln(\bar{p}) \times \text{Mexico}$</td>
<td>-0.002</td>
<td>0.004</td>
<td>-0.002</td>
<td>-0.006</td>
</tr>
<tr>
<td>(0.095)</td>
<td>(0.023)</td>
<td>(0.095)</td>
<td>(0.056)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.001</td>
<td>0.775</td>
<td>0.001</td>
<td>0.775</td>
</tr>
<tr>
<td>Basic Heading</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Country</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>
Sampling Bias: \( \text{cov}(\phi_{ib}^c, \ln(\bar{p}_{ib}^c)) \)

\[
\phi_{sib}^c = \alpha + \beta \ln(\bar{p}_{sib}^c) \times 1\{c = \text{Mexico}\} + \theta_s^c + \varepsilon_{sib}^c
\]

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ln(\bar{p}) )</td>
<td>-0.000</td>
<td>-0.000***</td>
<td>-0.000***</td>
<td>-0.000***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>( \ln(\bar{p}) \times \text{Mexico} )</td>
<td>-0.002***</td>
<td>-0.001***</td>
<td>-0.001***</td>
<td>-0.001***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Observations</td>
<td>764,419</td>
<td>764,419</td>
<td>761,751</td>
<td>761,750</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.028</td>
<td>0.030</td>
<td>0.212</td>
<td>0.212</td>
</tr>
<tr>
<td>Store</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Country</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

Double-difference log UPC expenditure shares and UPC pricing rule over time and relative to the largest UPC within each firm.

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

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

\[
\omega_{kibt} = [\Delta^t\ln \varphi_{kibt} - \Delta^t\ln \varphi_{kibt}] \text{ and } \kappa_{kibt} = [\Delta^t\ln a_{kibt} - \Delta^t\ln a_{kibt}]
\]
Orthogonality of the double-differenced demand and supply shocks defines a set of moment conditions:

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

where \( \beta_g = [\sigma_{ib}, \delta_{ib}]' \) and \( \nu_{kibt} = \kappa_{kibt} \omega_{kibt} \).

We proceed with GMM.
Given $\sigma_{ib}$ estimates, we use the Engel curve to estimate $\varepsilon_{kib}$ as Comin et al. (2020):

$$\ln \frac{s_{kibt}^{h}}{s_{Kibt}^{h}} - (1 - \sigma_{ib}) \ln \frac{p_{kibt}^{h}}{p_{Kibt}^{h}} = (\varepsilon_{kib} - 1) \left( \ln \frac{E_{ibt}^{h}}{p_{Kibt}^{h}} + \frac{1}{1 - \sigma_{ib}} \ln s_{Kibt}^{h} \right) + \psi^{h}_{t} + \varepsilon_{kibt}^{h}$$

where $K$ is the benchmark barcode, which corresponds to the largest selling barcode in each item, and $\psi^{h}_{t}$ is the set of fixed effects and controls.
## Parameter Estimation

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std. dev.</th>
<th>10th-percentile</th>
<th>median</th>
<th>90th-percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_i$</td>
<td>9.29</td>
<td>3.42</td>
<td>5.61</td>
<td>8.73</td>
<td>12.49</td>
</tr>
<tr>
<td>$\varepsilon_{kib}$</td>
<td>0.82</td>
<td>1.79</td>
<td>-1.58</td>
<td>0.89</td>
<td>2.99</td>
</tr>
<tr>
<td>$\theta_{ib}$</td>
<td>0.02</td>
<td>0.13</td>
<td>-0.13</td>
<td>0.03</td>
<td>0.16</td>
</tr>
<tr>
<td>$E_{ib}^M / E_{ib}^U$</td>
<td>0.72</td>
<td>0.60</td>
<td>0.14</td>
<td>0.53</td>
<td>1.47</td>
</tr>
</tbody>
</table>