APPENDIX

A   List of Matched Items

Table A.I: List of Matched Items

Item Name	US modules	US barcodes	MX barcodes	common barcodes
Beer	6	7451	521	51
Biscuits	1	13135	2622	189
Bread	2	605	910	0
Canned Fish	1	864	328	5
Canned Milk	1	365	43	3
Canned Soup	1	4822	216	0
Cereal	5	7922	1903	287
Cheese	16	16863	1581	118
Chicken Soup	1	373	345	12
Chocolate	4	13691	1164	128
Crackers	28	31901	3024	274
Dairy Cream	1	1029	168	1
Desserts	12	4759	165	4
Detergents	6	4564	2348	178
Ground Coffee	1	5575	371	43
Hair Rinse & Conditioners	1	2755	1066	74
Ice Cream	3	9582	854	91
Instant Coffee	3	1253	474	37
Instant Meals	1	215	53	0
Instant Tea	2	95	86	1
Jams and Spreads	12	6924	711	64
Juice	16	8916	2287	118
Ketchup	1	627	167	31
Liquid Cleaners	11	3501	1365	43
Milk	5	8236	809	31
Modified Milk	3	1915	288	16
Paper Towels	1	4642	481	35
Pasta	4	7362	1300	84
Powdered Milk	1	166	492	1
Preserved Vegetables	47	13997	726	82
Salt	2	1471	236	13
Sanitary Products	2	2079	676	26
Sauce	30	15028	181	48
Shampoo	3	4329	2283	197
Soap	9	9996	1715	200
Soda	1	8189	1855	68
Soup Bases	2	1807	126	1
Spicy Sauce	4	3569	442	27
Sugar	4	1144	318	8
Tomato Puree	1	112	108	1
Toothpaste	3	232	453	0
Yogurt	2	6219	1210	12

Note: The table reports statistics about the 42 items matched across the Nielsen data of US and Mexico. This covers 260 US modules, 238,280 US barcodes, 36,471 MX barcodes, and 2,602 common barcodes.

B   Sampling Bias

B.1   Proofs of Propositions

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

Proof. The bias on the first term for country M is:

$$\prod_{i \in \Omega_b} \frac{(\bar{p}_{ib}^m)^{\omega_{ib}}}{(\bar{p}_{ib}^m)^{\frac{1}{N_b}}}$$

This ratio is greater than one if and only if:

$$\sum_{i \in \Omega_b} \left(\omega_{ib} - \frac{1}{N_b} \right) \ln(\bar{p}_{ib}^m) > 0 \tag{17}$$

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

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

Using that

$$\lim_{N_b \to \infty} \frac{1}{N_b - 1} \sum_{i \in \Omega_b} \left(\omega_{ib} - \frac{1}{N_b} \right) \left(\frac{1}{N_b - 1} \sum_{i \in \Omega_b} \ln(\bar{p}_{ib}^m) \right) = 0$$

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

PROPOSITION 2.   If the number of stores $S \to \infty$, the second 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 second term of the sample bias for a country c is:

$$\frac{\hat{p}_{ib}^c}{\bar{p}_{ib}^c} = \frac{\prod\limits_{s\in\Psi^c} (\bar{p}_{sib}^c)^{\phi_{sib}^c}}{\prod\limits_{s\in\Psi^c} (\bar{p}_{sib}^c)^{\phi_s^c}}$$

This ratio is larger than 1 for country c if

$$\sum_{s\in\Psi^c} \phi_{sib}^c \ln\left(\bar{p}_{sib}^c\right) > \sum_{s\in\Psi^c} \phi_s^c \ln\left(\bar{p}_{sib}^c\right) \tag{18}$$

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 18 can be written as:

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

Using the definition of covariance on both sides

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

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

B.2 Empirical Tests

To quantify the importance of the first term of the sampling bias, we rely on Proposition 1 and test whether $\text{cov}(\omega_{\mathbf{b}}, \ln\left(\bar{\mathbf{p}}_{\mathbf{b}}^{\mathbf{m}}\right)) > \text{cov}(\omega_{\mathbf{b}}, \ln\left(\bar{\mathbf{p}}_{\mathbf{b}}^{\mathbf{u}}\right))$. To do so we realy 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 are the Sato-Vartia weights for each item. The coefficient

of interest is β which indicates whether there is a difference between the covariance between the weights and the prices of items across the two countries; Table B.I shows that we do not find a significant difference indicating that the first term of the sampling bias is close to 1.

Table B.I: Sampling Bias First Term: Expenditure Weights and Prices

	(1)	(2)	(3)	(4)
$\ln(\bar{p})$	-0.059	-0.004	-0.059	-0.017
	(0.058)	(0.027)	(0.058)	(0.033)
$\ln(\bar{p}) \times$ Mexico	0.056	0.005	0.056	0.030
	(0.079)	(0.015)	(0.079)	(0.037)
Observations	84	84	84	84
R-squared	0.013	0.836	0.013	0.837
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 second term of the sampling bias, we rely on Proposition 2.

In order to compare the magnitude of $\mathrm{cov}(\phi_{\mathbf{ib}}^{\mathbf{m}}, \ln(\bar{\mathbf{p}}_{\mathbf{ib}}^{\mathbf{m}}))$ relative to $\mathrm{cov}(\phi_{\mathbf{ib}}^{\mathbf{u}}, \ln(\bar{\mathbf{p}}_{\mathbf{ib}}^{\mathbf{u}}))$ we estimate the following specification:

$$\phi_{sib}^{c} = \alpha + \beta \ln(\bar{p}_{sib}^{c}) \times \mathbb{1}\left\{c = \text{Mexico}\right\} + \theta^{c} + \lambda_{i} + \epsilon_{sib}^{c}$$

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 B.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.

4

Table B.II: Sampling Bias Second Term: Expenditure Weights and Prices at the Store $\times$ Item Level

	(1)	(2)	(3)
$\ln(\bar{p})$	-0.0043	-0.0081	-0.0022
	(0.005)	(0.004)	(0.010)
$\ln(\bar{p}) \times$ Mexico	-0.0771***	-0.0550*	-0.0662*
	(0.000)	(0.005)	(0.008)
Observations	340,381	340,381	336,641
R-squared	0.016	0.026	0.126
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

C   Results with Non-homothetic CES Framework

C.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 (forthcoming), Matsuyama (2019) and Redding and Weinstein (2020), which satisfy implicit additivity in Hanoch (1975). The non-homothetic CES consumption index for item i, C_{ib}^m, is defined by the following implicit function:

$$\sum_{k \in \Omega_{ib}^m} \left(\frac{\varphi_{kib}^m C_{kib}^m}{(C_{ib}^m)^{(\epsilon_{kib} - \sigma_{ib})/(1-\sigma_{ib})}} \right)^{\frac{\sigma_{ib}-1}{\sigma_{ib}}} = 1 \tag{19}$$

where C_{kib}^m denotes total consumption of barcode k; Ω_{ib}^m is the set of barcodes in Mexico M within an item i and a basic heading b; σ_{ib} is the elasticity of substitution between barcodes; ϵ_{kib} is the constant elasticity of consumption of barcode k with respect to the consumption index (C_{ib}^m) 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 $k \in \Omega_{ib}^m$, 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_{ib}^m) dual to the consumption index (C_{kib}) and the expenditure share for a individual barcode k (s_{kib}^m):

$$P_{ib}^m = \left(\sum_{k \in \Omega_{ib}^m} (p_{kib}^m / \varphi_{kib}^m)^{1-\sigma_{ib}} (C_{ib}^m)^{\epsilon_{kib}-1} \right)^{\frac{1}{1-\sigma_{ib}}} \tag{20}$$

$$s_{kib}^m = \frac{(p_{kib}^m / \varphi_{kib}^m)^{1-\sigma_{ib}} (C_{ib}^m)^{\epsilon_{kib}-1}}{\sum_{l \in \Omega_{ib}^m} (p_{lib}^m / \varphi_{lib}^m)^{1-\sigma_{ib}} (C_{lib}^m)^{\epsilon_{lib}-1}} = \frac{(p_{kib}^m / \varphi_{kib}^m)^{1-\sigma_{ib}} (E_{ib}^m / P_{ib}^m)^{\epsilon_{kib}-1}}{(P_{ib}^m)^{1-\sigma_{ib}}} \tag{21}$$

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 (Ω_{ib}):

$$\frac{P_{ib}^m}{P_{ib}^u} = \frac{p_{kib}^m / \varphi_{kib}^m}{p_{kib}^u / \varphi_{kib}^u} \left(\frac{E_{ib}^m / P_{ib}^m}{E_{ib}^u / P_{ib}^u} \right)^{\frac{\epsilon_{kib}-1}{1-\sigma_{ib}}} \left(\frac{s_{kib}^m}{s_{kib}^u} \right)^{\frac{1}{\sigma_{ib}-1}}, \quad k \in \Omega_{ib} \tag{22}$$

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 (λ_{ib}^m and λ_{ib}^u):

$$\lambda_{ib}^m \equiv \frac{\sum_{\Omega_{ib}} p_{kib}^m C_{kib}^m}{\sum_{\Omega_{ib}^m} p_{kib}^m C_{kib}^m} \quad \text{and} \quad \lambda_{ib}^u \equiv \frac{\sum_{\Omega_{ib}} p_{kib}^u C_{kib}^u}{\sum_{\Omega_{ib}^u} p_{kib}^u C_{kib}^u} \tag{23}$$

where Ω_{ib} is the set of common barcodes between the two countries and Ω_{ib}^m and Ω_{ib}^u represent the set of total barcodes in Mexico and the United States respectively.

Using this expression, the share of an individual barcode in total expenditure (s_{kib}^m) in equation (21) 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 (λ_{ib}^m):

$$s_{kib}^m = \lambda_{ib}^m s_{kib}^{M*}, \quad k \in \Omega_{ib} \tag{24}$$

Taking logs to equation (22) and using equation (24), we obtain the following equation:

$$\log \left(\frac{P_{ib}^m}{P_{ib}^u} \right)^{1+\frac{\epsilon_{kib}-1}{1-\sigma_{ib}}} = \log \frac{p_{kib}^m/\varphi_{kib}^m}{p_{kib}^u/\varphi_{kib}^u} + \log \left(\frac{E_{ib}^m}{E_{ib}^u} \right)^{\frac{\epsilon_{kib}-1}{1-\sigma_{ib}}} + \log \left(\frac{s_{kib}^{M*}}{s_{kib}^{U*}} \right)^{\frac{1}{\sigma_{ib}-1}} + \log \left(\frac{\lambda_{ib}^m}{\lambda_{ib}^u} \right)^{\frac{1}{\sigma_{ib}-1}} \tag{25}$$

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

$$\omega_{kib}^M = \frac{\dfrac{s_{kib}^{M*} - s_{kib}^{U*}}{\ln s_{kib}^{M*} - \ln s_{kib}^{U*}}}{\displaystyle\sum_{k \in \Omega_{ib}} \dfrac{s_{kib}^{M*} - s_{kib}^{U*}}{\ln s_{kib}^{M*} - \ln s_{kib}^{U*}}} \tag{26}$$

where

$$s_{kib}^{M*} = \frac{p_{kib}^M C_{kib}^M}{\sum_{k \in \Omega_{ib}} p_{kib}^M C_{kib}^M} \quad \text{and} \quad s_{kib}^{U*} = \frac{p_{kib}^U C_{kib}^U}{\sum_{k \in \Omega_{ib}} p_{kib}^U C_{kib}^U}$$

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

By multiplying the ideal log-difference weights (ω_{kib}) to equation (25), under the assumption that tastes are the same between two countries for each common barcode ($\varphi_{kib}^m = \varphi_{kib}^u$ for all $k \in \Omega_{ib}$), we can obtain the non-homothetic CES price index for item i by taking the arithmetic mean across common barcodes (Ω_{it}) and exponents on both sides:

$$\mathbb{NH}_{ib}^m \equiv \frac{P_{ib}^m}{P_{ib}^u} = \left(\prod_{k \in \Omega_{ib}} \left(\frac{p_{kib}^m}{p_{kib}^u} \right)^{\omega_{kib}} \times \left(\frac{\lambda_{ib}^m}{\lambda_{ib}^u} \right)^{\frac{1}{\sigma_{ib}-1}} \right)^{\frac{1}{1-\theta_{ib}}} \left(\frac{E_{ib}^m}{E_{ib}^u} \right)^{\frac{\theta_{ib}}{\theta_{ib}-1}} \tag{27}$$

where

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

and the ratio of λ_{ib}^m and λ_{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). $\mathbb{NH}_{ib}^m$ can be aggregated to the basic heading b level ($\mathbb{NH}_b^m$) and across basic headings by taking the weighted geometric mean and using the ideal log-difference weights from our CES specification defined as:

$$\omega_{ib} = \frac{\dfrac{s_{ib}^M - s_{ib}^U}{\ln s_{ib}^M - \ln s_{ib}^U}}{\sum_{i \in \Omega_b} \dfrac{s_{ib}^M - s_{ib}^U}{\ln s_{ib}^M - \ln s_{ib}^U}} \quad \text{and} \quad \omega_b = \frac{\dfrac{s_b^M - s_b^U}{\ln s_b^M - \ln s_b^U}}{\sum_{b \in \Omega} \dfrac{s_b^M - s_b^U}{\ln s_b^M - \ln s_b^U}}$$

C.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:

$$\mathbb{NH}_b^m = \Theta_b^m \times \mathbb{ICP}_b^m \tag{28}$$

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

where Θ_b^m is the aggregate bias correction term and $\mathbb{S}_b^m$, $\mathbb{Q}_b^m$ and $\widetilde{\mathbb{V}}_b^m$ represent the sampling bias, the quality bias, and the Engel-curve variety bias, respectively. The sampling bias and the quality bias are the same as the homothetic case.

Engel Curve Variety Bias, $\widetilde{\mathbb{V}}_b^m$: This bias measures both the cross-country differences in availability of products and the differences in real consumption across countries and is defined as follows:

$$\widetilde{\mathbb{V}}_b^m = \mathbb{V}_b^m \times \prod_{i \in \Omega_b} \left[\left(\frac{E_{ib}^m / E_{ib}^u}{\mathbb{EPI}_{ib}^m} \right)^{\frac{\omega_{ib} \theta_{ib}}{\theta_{ib} - 1}} \right] \tag{29}$$

where

$$\mathbb{V}_b^m \equiv \prod_{i \in \Omega_b} \left(\frac{\lambda_{ib}^m}{\lambda_{ib}^u} \right)^{\frac{\omega_{ib}}{\sigma_{ib} - 1}} \tag{30}$$

where $\mathbb{V}_b^m$ is variety bias in the homothetic case.

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 products 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 product 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, $\widetilde{\theta}_{ib}$ is equal to zero and the Engel-curve variety bias becomes $\mathbb{V}_b^m$.

C.3 Parameter Estimation

Taking estimates of the elasticity of substitution as given, we estimate the constant elasticity of consumption (ϵ_{kib}) for each barcode k with respect to the consumption index (C_{kib}^m) as in Comin, Lashkari and Mestieri (forthcoming):

$$\ln \frac{s_{kibt}^h}{s_{\mathbf{K}ibt}^h} - (1 - \sigma_{ib}) \ln \frac{p_{kibt}^h}{p_{\mathbf{K}ibt}^h} = (\epsilon_{kib} - 1) \left(\ln \frac{E_{ibt}^h}{p_{\mathbf{K}ibt}^h} + \frac{1}{(1 - \sigma_{ib})} \ln s_{\mathbf{K}ibt}^h \right) + \psi_t^h + \epsilon_{kibt}^h \quad (31)$$

where $\mathbf{K}$ is the benchmark barcode, which corresponds to the largest selling barcode in each item, and ψ_t^h is the set of fixed effects. We aggregate households into seven groups by their annual household income. With the US Nielsen data, equation 31 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.[10]

---

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

Table C.I: Descriptive Statistics of Estimated Parameters

	mean	std. dev.	10th-percentile	median	90th-percentile
σ_{ib}	8.94	2.95	5.36	8.78	13.31
ϵ_{kib}	0.98	3.41	-1.26	1.28	3.30
θ_{ib}	0.07	0.09	-0.05	0.06	0.20
E_{ib}^m/E_{ib}^u	0.44	0.35	0.10	0.35	0.91

Note: This table reports descriptive statistics for the elasticity of substitution (σ_{ib}), the elasticity of consumption of barcode k with respect to the consumption index (ϵ_{kib}), parameter in the non-homothetic CES price index (θ_{ib}) and nominal expenditure ratio (E_{ib}^m/E_{ib}^u).

The first row of Table C.I reports descriptive statistics for σ_{ib}. The average of the elasticity of substitution we use is 8.94 with standard deviation of 2.95. The second row of Table C.I reports the descriptive statistics for ϵ_{kib}. Our estimates for this parameter have a mean of 0.98 and a standard deviation of 3.41. The third row of Table C.I reports the descriptive statistics for θ_{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 θ_{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 C.I shows, the ratio varies across items. It has a median of 0.35 and a mean of 0.44 (with standard deviation of 0.35), which indicates that the distribution is skewed to the right. Figure C.1 shows the distribution of estimated ϵ_{kib} and θ_{ib}. The mean of ϵ_{kib} is close to one. As a result, the mean of θ_{ib} is around zero.

Figure C.1: Distribution of Estimated ϵ_{kib} and θ_{ib}

| ϵ_{ki} | θ_i |
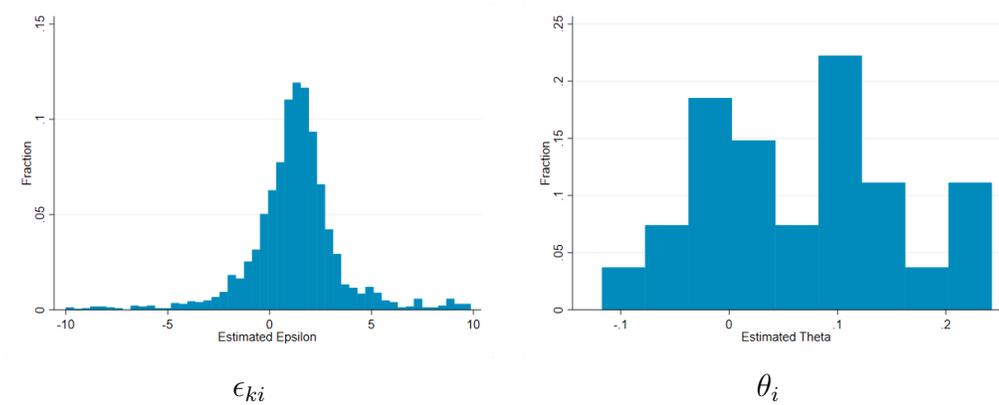
Note: Panel (a) and (b) show the distribution of estimated ϵ_{kib} and θ_{ib}, respectively.

C.4 Decomposition Results

Equation 28 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 0.94, which is close to the case with homothetic preference (0.91). This is mainly because, for the common products, the average elasticity of consumption is estimated to be close to 1.

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

	Price Indexes		Biases			Agg. Bias
	$\mathbb{ICP}^m$	$\mathbb{NH}^m$	$\mathbb{S}^m$	$\mathbb{Q}^m$	$\widetilde{\mathbb{V}}^m$	$\widetilde{\Theta}^m$
CES	0.84	0.79	0.85	1.35	0.82	0.94
Fisher	0.88	0.90	0.88	1.38	0.84	1.02

Note: The table reports ICP-style price index ($\mathbb{ICP}^m$) and the non-homothetic CES price index ($\mathbb{NH}^m$). By equation 28, the gap between two price indexes can be decomposed into sampling bias ($\mathbb{S}^m$), quality bias ($\mathbb{Q}^m$), and Engel-curve variety bias ($\widetilde{\mathbb{V}}^m$). $\widetilde{\Theta}^m$ 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.

D   Decomposition Results with Common Elasticity of Substitution

In this section, we report decomposition results with common elasticity of substitution across items, $\sigma_{ib} = 9$, which is chosen from a mean of σ_{ib} (8.94). Table D.I reports decomposition results for the exact price index. The aggregate bias is estimated to be the same with the item-specific elasticity of substitution case (0.91).

Table D.I: Decomposition Results for the Exact Price Index with $\sigma_{ib} = 9$

	Price Indexes		Biases			Agg. Bias
	$\mathbb{ICP}^m$	$\mathbb{EPI}^m$	$\mathbb{S}^m$	$\mathbb{Q}^m$	$\mathbb{V}^m$	Θ^m
CES	0.84	0.76	0.85	1.35	0.79	0.91
Fisher	0.88	0.86	0.88	1.38	0.80	0.98

Note: The table reports ICP-style price index ($\mathbb{ICP}^m$) and the exact price index ($\mathbb{EPI}^m$). 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.