Online Appendix
Taxing to Reduce Obesity

By Arturo Aguilar, Emilio Gutierrez, Enrique Seira

Draft: June 17, 2016

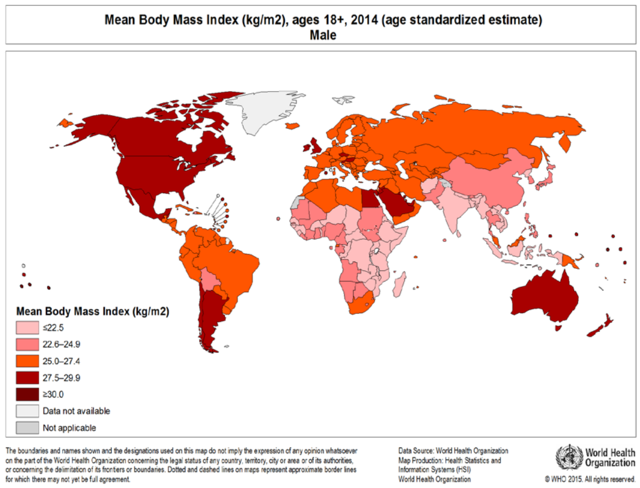
A1. BMI evolution and measure in our data

The taxes being analyzed in this paper were established as a government response to a growing trend in obesity levels in Mexico. Thus, a variable of interest in this study is the body mass index (BMI), which we use to classify households based on its anthropometric characteristics. Following WHO standards, we classify households as normal, overweight or obese according to its household's head body mass index (BMI) and explore if the effects analyzed in the paper differ for each group. In part of the analysis BMI is also used as a dependent variable itself.

Our BMI data comes from yearly surveys that the Kantar World Panel (KWP) implements to households from our main dataset. As part of the survey, KWP gathers self-reported information of weight and height from the male and female household heads.[1] The use of self-reported data might raise some concerns. We performed some checks with strongly suggest that the BMI measure has high quality. First, it respects the ranking Mexico vis-

[1]More information about KWP and data gathered are available in section **??** of this online appendix.

Figure A1. World Overweight Status



Note: Source: `http://gamapserver.who.int/gho/interactive_charts/ncd/risk_factors/bmi/atlas`

a-vis the other countries we have in our data. Second, it is consistent with time trends in BMI as reported to the WHO. Third, and very importantly, it is very close to data from the ENSANUT survey which is the most important health survey in Mexico, is representative of the country and measures height and weight in situ by trained nurses. Fourth, is displays tight relationships with demographics in an intuitive way. Finally, we note that although it would be better to have measured weight, more than half of the countries reporting to WHO use self reported statistics like the one in this study. This evidence provides support to the use of this variable in the main text.

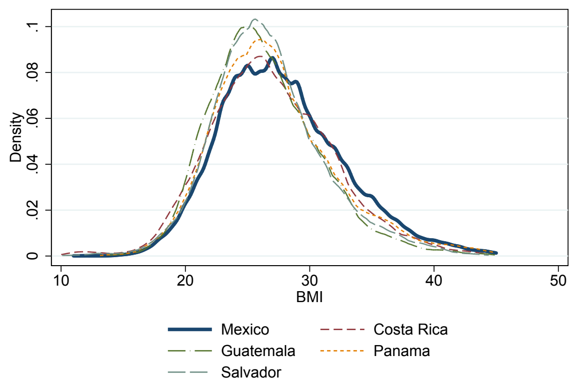
First, we show how Mexico compares to other countries in terms of BMI, including a subset of countries for which we have KWP data available. Figure **??** confirms that Mexico is one of the countries with highest average BMI. This map, taken from the WHO website maps, indicates that the average BMI in Mexico for individuals aged 18 or older is 27.5.[2] Compared with the rest of Central American countries for which we have KWP data available, Mexico ranks as the highest BMI country. According to KWP data, mean values for Costa Rica, Guatemala, Panama and El Salvador are 26.9, 26.3, 27.1 and 26.6, respectively. Mean values according to the WHO are very similar: Costa Rica, 26.7; Guatemala, 25.8; Panama, 26.4; and El Salvador, 26.8. The only difference between the KWP and WHO rankings is that KWP reports a larger mean BMI for Panama than for the rest of Central American countries. Nonetheless, Panama is also the country for which the sample of households included in KWP is smallest. KWP reports information for 4108, 4468, 3600, and 4008 households in Costa Rica, Guatemala, Panama and El Salvador, respectively. Figure **??** gives a first insight of how the distribution of BMI compares between Mexico and the Central American countries using the female head information from KWP. It shows that Mexico whole distribution is to the right of the other four countries for which we have data.

Part of the interest in obesity fighting taxes is not only the level, but the rate of growth of obesity and BMI. Figure **??** shows the striking growth for selected countries. Figure **??** mirrors this increasing trend with the five countries for which we have data from Kantar. This is important. One concern with self reported data is that respondents do not update their weight information. We are witnessing a yearly increase in BMI in our data for all the countries, implying that households *are* updating their weight.

Finding that our data respects rankings and trends is comforting. But perhaps the most comforting evidence was to find out that our self-reported

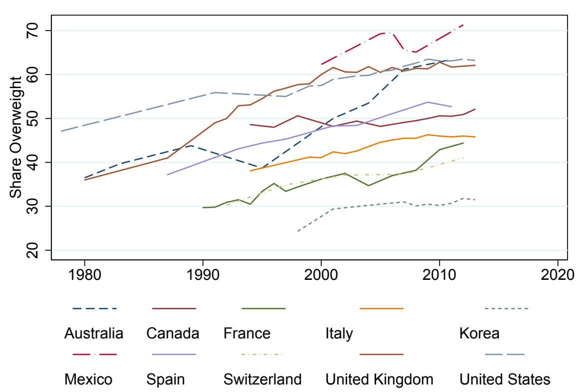
[2] `http://gamapserver.who.int/gho/interactive_charts/ncd/risk_factors/bmi/atlas.html`

Figure A2. BMI Distribution, Central America and Mexico



Note: Kernel density estimates of BMI for our five Latin America countries as reported in KWP.

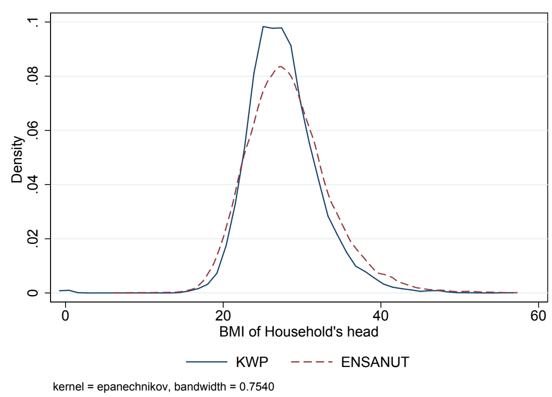
Figure A3. Overweight in Context



Note: Source: OECD Obesity Update. 2012.

BMI measures coincide substantially with those collected in-situ by the 2012 Mexican Health and Nutrition Survey (ENSANUT). Figure **??** plots BMI densities according to both data sources. The distributions have a very similar mean (see Table 1 in the paper), although ENSANUT reports a slightly larger variance.

Figure A4. BMI distributions KWP and ENSANUT

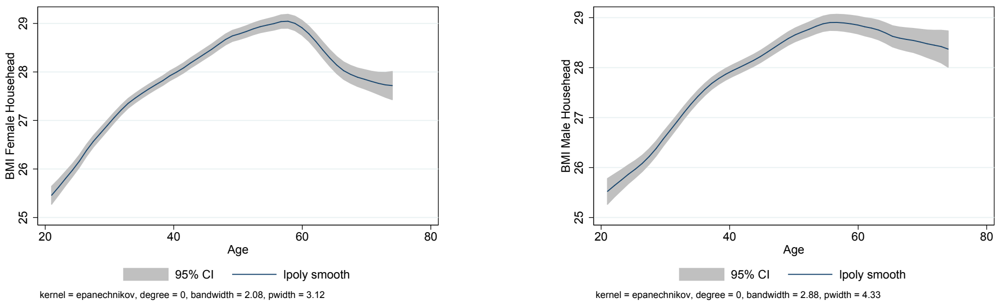


Note: Kernel density estimates of BMI according to KWP and ENSANUT.

One other consistency check one might implement is to explore correlations of BMI against variables which we know should correlate with BMI. Figure **??** shows the that BMI in our data is closely related to age. Finally, we are not very concerned about using the female household head as a measure of obesity (see Figure **??**).

Together, these checks give us confidence that BMI measures reported in KWP are good approximations one of our main variables of interest.

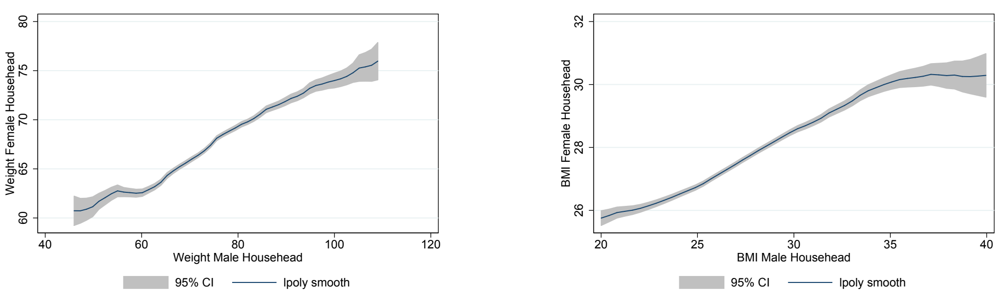
Figure A5. BMI-Age Profiles



(a) Female Profile (b) Male Profile

Note: Non-parametric estimations (local polynomials). KWP data.

Figure A6. Relationship weight and BMI of male and female household head



(a) Weight (b) BMI

Note: Non-parametric estimations (local polynomials). KWP data.

A2. Kantar World Panel: some examples

Section III in the main paper describes our data sources. Our main source comes from Kantar World Panel (KWP). KWP is a company with more than 50 years of experience measuring consumption of households in dozens of countries. Together with Nielsen they are the most important companies worldwide.

One remarkable characteristic from the KWP data is its high frequency (weekly) and the the high level of disaggregation, which allows us to make a detailed statistical analysis from different perspectives. KWP's product catalogue comprises a total of $58,721$ barcodes (SKUs); 26,101 for foods and drinks. It classifies barcodes along ten variables. The categories used to classify products include: basic ingredient, flavor, size, brand, product's characteristics (e.g. light, lactose-free, caffeine-free, etc.), among others. Table **??** below shows an example of how five products are classified along these dimensions.

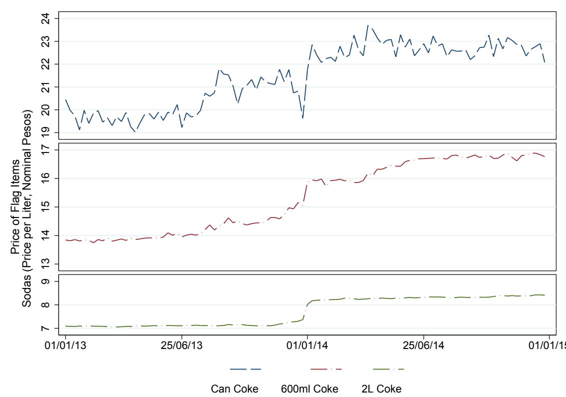
Table A1—Product classification example

id	product	subproduct	brand	clas01	clas02	clas03	clas04	clas05	content
3510	Cereals	No subproduct	Kelloggs	Corn Flakes	.	Apple w cinnamon	.	.	750 gr
105944	Cookies	Sweet	Marinela	Principe	.	Strawberry-filled	Chocolate	.	44 gr
43379	Carbonated drinks	Regular	Lift	Red apple	.	.	.	.	1000 ml
72523	Milk	Pasterurized	Alpura	.	Regular	Non lactose-free	.	.	1000 ml
133285	Mayonnaise	Regular	Kraft	Mayo	Regular	.	.	Light	443 gr

We used the above information to detect which barcodes were subject to the tax and which were exempt and to impute nutritional content as we detail in this Appendix. Before proceeding with this detail however, we would like to exemplify the richness of our data by plotting prices for some selected products. Figure **??** shows the price per liter through time for three

Coca-Cola presentations: the 355 ml. can, the 600 ml. bottle and the 2 liter bottle. We could potentially plot these graphs by storebrand, by city or by type of household.

FIGURE A7. PRICE PER LITER OF COCA COLA BY PRESENTATION SIZE

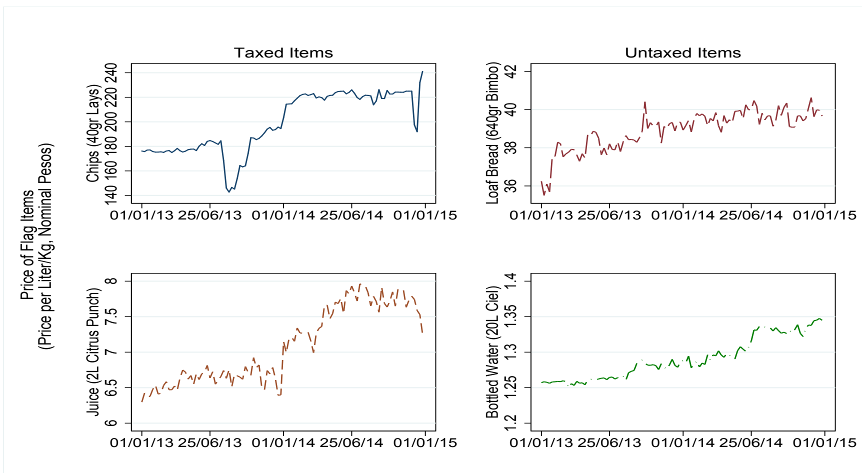


Note: Source: KWP data

Figure ?? gives another example. It plots the price per unit (liter or kilogram, depending on the product) for the following products: (i) a 40 gram bag of Lays chips (a taxable food product), (ii) a 640 gram of Bimbo bread (a not taxable food), (iii) a 2 liter bottle of Citrus Punch juice (a taxable drink), and (iv) a 20 liter bottle of Ciel natural water (a non-taxable drink).

Along with the information on purchases, KWP also administers a yearly questionnaire that captures a set of socio-economic characteristics from which SES categories are derived using standard methodologies. SES categories are derived from different measures of household assets: number of rooms, type of floor, number of bathrooms, whether the dwelling has a gas stove, number of light bulbs, number of cars, and household head's education. Following the procedure, about 21, 52 and 27 percent of households in the data are classified in the A/B/C+, C/D+ and D/E SES categories, respectively. Finally, each year KWP gathers also self-reported information on the age, weight and height of the household's male and female heads, which allows us to calculate BMI. Using BMI, we produce obesity and overweight indicators following WHO standards. The SES and BMI categories are employed in the analysis to study heterogenous effects.

Figure A8. Four products example, prices



Note: Source: KWP data

A3. Nutritional Content Data and Imputation

A. Collection of Nutritional Data

We mentioned above that we start with 26,101 barcodes for foods and drinks. However some of these are exactly the same product in the sense they have exactly the same substance in it, and just vary in the quantity per unit. For example a can of 355ml coke has the same substance as a 500ml coke. Once we remove the size differences we end up with $15,622$ barcodes. Our goal in this section is to describe how we recorded or imputed caloric and nutritional content for each of these given that Kantar does not do it.

The first step consisted of recruiting a team of more than 20 enumerators and train them to search for the items in supermarkets (physically and through their internet website[3]), manufacturer's website and local grocery stores. This work involved several months of work. Enumerators took pictures of the nutritional label and then coded such label on a predetermined format. This enabled us to do double coding for a sample of barcodes and verify that the enumerators actually went to the supermarket and found the appropriate product. Each enumerator was given a list of products by the researchers an a format to fill out and enter the nutritional content available in the product's label, mainly: serving size, calories, sugar, fat, saturated fat, iron, carbohydrates, cholesterol, and sodium. We prioritized barcodes that covered a large percentage of purchase events. Finally we performed quality check on a random sample of 5 percent of the barcodes. We found less than 1 percent of imputing errors.

Using this process we managed to collect direct information for 1,953 bar-

[3]Mainly Walmart's website was used: [`www.walmart.com.mx`]

codes which cover 72 percent of purchase events and 68 percent of household's expenditure of the weekly scanner data for 2013 and 2014. Table **??** shows the proportion of events and expenditure that were collected for each product-category that covers all of KWP data.[4]

We could find information of more than 95 percent of the products we selected, and therefore we think there is no systematic selection problem in the data. However, we excluded some products that were very local and that comprised a very low fraction of observations.

B. Caloric and nutrient imputations

After collecting the nutritional information for a very large fraction of purchase events, we had to impute the nutritional characteristics of the remaining products that appear in KWP's dataset. The next steps were followed to perform such imputation:

1) Convert sizes to the same units: We begin by using serving sizes along with calories and nutrients of the products that we gathered to generate a variable that shows the amount of calories or nutrients per 100 grams (or per 100 ml. for drinks). We will refer to this variable as our *density measure*.

2) Run an exploratory analysis to measure the variance of caloric density for different aggregations, using the variables {product,subproduct,brand,clas01, clas02,clas03, clas04,clas05} exemplified in Table **??**. We call these the *matching variables*. The highest level of disaggregation is when we use all these 8 variables to create cells. At these level which we call L8 we find tiny variances with cells. This just means for instance that

[4] Product category is a variable in KWP data.

Table A2—Information collected by type of product

Product	Expenditure (percent)	Purchase events (percent)
Cooking oil	83	81
Bottled water	38	38
Sparkling water	87	85
Baby food	45	55
Canned tuna	85	82
Cereal bar	90	89
Powdered drinks	90	90
Energy drink	95	95
Carbonated soft drink	98	98
Sports drink	89	89
Non-carbonated soft drink	74	73
Snacks	96	96
Coffee	85	87
Condensed milk candy	45	30
Seasonings and broths	78	81
Cereal	94	91
Chocolate	94	93
Tomato puree	88	88
Liquid seasoning	56	53
Creamer/Substitutes of cream	63	76
Sour cream	60	60
Cream Spreads	69	60
Breadcrumbs	96	93
Cookies	80	84
Cornflour drink (atole)	73	66
Ice creams and popsicles	57	70
Vegetable Juice	94	95
Condensed Milk	96	96
Powdered Milk	35	35
Evaporated Milk	95	95
Milk	60	60
Flavored Milk	85	83
Margarine	70	71
Mayonnaise	83	84
Jams	73	71
Honey	53	56
Flavored Milk Powder	90	92
Industrialized bread	92	90
Pastas	63	62
Refrigerated Dessert	99	100
Powdered Desserts	66	65
Tomato puree	87	87
Ketchup	86	89
Snack sauce	69	71
Homemade bottled sauce	84	85
Pasta Sauce	53	35
Instant Soups	91	90
Iced-tea	37	35
Yogurt	80	81

that Alpura Milk has the same level of calories across its presentation sizes. If we go up one level to L7, there is some variance but is tiny. This means for instance that Alpura chocolate milk has similar caloric density than Alpura strawberry milk.

3) We explored by hand each product-group[5] follows a different classification (which is reflected in the variables shown in Table **??**), and within each product-group we imputed the nutrient/caloric density to the closest category to which it belonged.

4) In other words, for each article with missing caloric and nutrient information, we search for the most similar product with information available based on the *matching variables*. We seek to establish such match with the highest level of disaggregation possible. If an exact match was not possible for a given level of disaggregation, we used a lower level of disaggregation, that is, we use a subset of the *matching variables*. The following table exemplifies this procedure. In this case, *subproduct* and *clas01* were chosen as the matching variables. As can be seen, there are 4 products with $subproduct = REGULAR$[6] and $clas01 = COLA$ with missing information for calories. In this case, these products are matched to those with the same *subproduct* and *clas01* for which information is non missing. The average value of

[5]The different types of products-groups available are: cooking oil, bottled water, sparkling water, baby food, canned tuna, cereal bars, powdered drinks, energy drinks, carbonated soft drinks, sport drinks, non-carbonated soft drinks, snacks, coffee, condensed milk candy (cajeta), seasonings and broths, cereal, beer, chocolate, tomato puree, liquid seasoning, creamer, sour cream, sweet spreads, breadcrumbs, cookies, cornflour drink (atole), ice creams and popsicles, vegetable juice, condensed milk, powdered milk, evaporated milk, regular milk, milk-based flavored drinks, margarine, mayonnaise, jams, honey, milk-based flavored powder, industrialized bread, pastas, refrigerated dessert, powdered desserts (e.g. jello), ketchup, salsas, homemade bottled salsas, pasta sauce, instant soups, iced-tea, yogurt

[6]DIET/LIGHT is the other possible value for *subproduct*

the *density measure* of those products with available information is imputed to those with missing information. The same process is followed with the two missing values for the $subproduct$ = $REGULAR$ and $clas01$ = $LEMON\_LIME$ products. However, we can see that we have in the first line a cherry-flavored drink which cannot be matched to other drink with the same *subproduct* and *clas01* values. In this case, a lower level of disaggregation is used and that product is imputed with the average *density measure* value from all drinks that have $subproduct = REGULAR$.

5) For barcodes for which we were not reasonably sure about the quality of the imputation we send enumerators to get the information on the supermarket.

TABLE A3—IMPUTATION EXAMPLE

num	product	subproduct	brand	clas01	kcal_100ml
124	CARB. DRINK	REGULAR	GARCI CRESPO	CHERRY	
48	CARB. DRINK	REGULAR	CABALLITOS	VANILLA / ROOT BEER	29
2	CARB. DRINK	REGULAR	PEPSI	COLA	45.3
7	CARB. DRINK	REGULAR	BIG COLA	COLA	42
4	CARB. DRINK	REGULAR	RED COLA	COLA	
90	CARB. DRINK	REGULAR	CHIVA	COLA	
133	CARB. DRINK	REGULAR	SMART	COLA	
39	CARB. DRINK	REGULAR	BIG COLA DOBLE	COLA	43
1	CARB. DRINK	REGULAR	COCA COLA	COLA	42
52	CARB. DRINK	REGULAR	AURRERA	COLA	
3	CARB. DRINK	REGULAR	BIG COLA MEGA	COLA	42
115	CARB. DRINK	REGULAR	COCA COLA LIFE	COLA	18
11	CARB. DRINK	REGULAR	7 UP	LEMON-LIME	34
8	CARB. DRINK	REGULAR	SPRITE	LEMON-LIME	36
81	CARB. DRINK	REGULAR	BIG FRESH	LEMON-LIME	
97	CARB. DRINK	REGULAR	AURRERA	LEMON-LIME	

To assess the quality of our imputation procedure, we performed the following exercise:

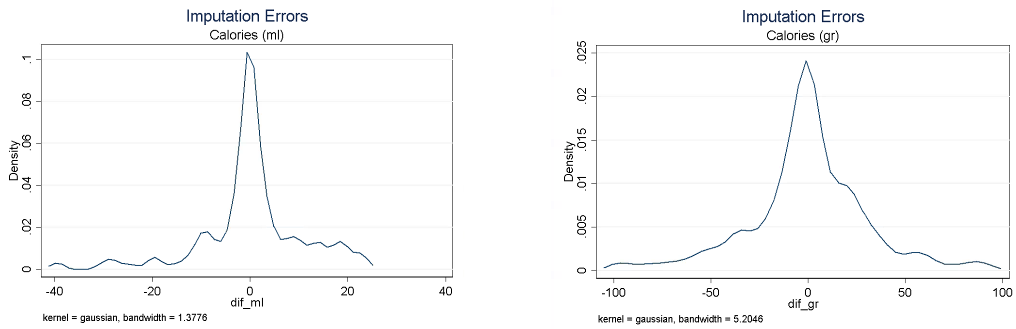
1) We randomly drop the observed values for 10 percent of the products for which we *have* available caloric and nutrient information.

2) We implement the imputation algorithm described above for the product whose information was deleted in the previous step. We compare the imputed versus the observed values.

3) We repeated this exercise 5 times.

4) Figures **??** to **??** plot the density of difference between the observed and imputed values for calories, sugar and fat. Similar graphs for the rest of the nutrients are available upon request. As the graphs display, there is a big concentration in zero which speaks of the adequacy of the process implemented to input the missing information. The imputation errors for calories are relatively small: for liquids we obtain from the out of sample exercise that the absolute value of the imputation error is, on average, 8.2 calories per 100 ml, while on average caloric density is 104.7 calories per 100ml. For solid foods, the corresponding average value of the imputation error is 61.6 calories per 100 gr, while on average caloric density is 298.4 calories per 100 gr.

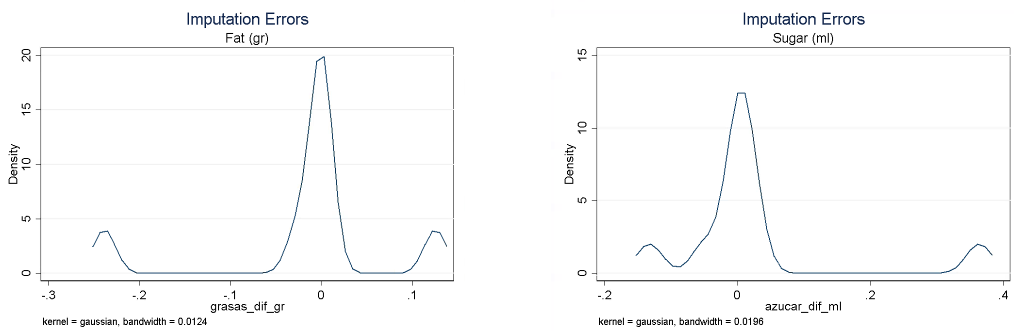
Figure A9. Outsample imputation errors for Calories



(a) Liquids (b) Solids

Note: Gaussian kernel density estimations shown

Figure A10. Outsample Imputation Errors Fat and Sugar



(a) Fat (b) Sugar

Note: Gaussian kernel density estimations shown

A4. Tax variable

As described in the main text, two different tax designs were implemented beginning on January 1st, 2014: (i) one peso per liter of sugar added drinks, and (ii) an 8 percent ad-valorem tax for non-basic food items with a caloric density above 275 kilocalories per 100 grams.

To have an accurate identification of the products in KWP's dataset that would be subject to the sugary drink tax, we started by identifying the tax exempt products. For drinks these include bottled water, sparkling water, beer, milk, powdered milk, evaporated milk, atole, etc. We then classified as taxable drinks those that fell into powdered drinks, energy drinks, carbonated and non-carbonated soft drinks, sports drinks, vegetable juice, flavored milk and iced-tea. For the products that are considered for the tax, the only exception considered is when they are sub-classified as "diet, light" or "non-sugar added" in KWP's categories. Then we went inside these broad clarifications to find cases of barcodes that had no added sugar and therefore should be exempted. We did this barcode by barcode, and for those where we were not sure we asked industry experts.

As for the ad-valorem tax, non-basic food products with higher caloric density than 275 kilocalories per 100 grams with the following product classification are classified as tax eligible: cereal bars, snacks, condensed milk candy (cajeta), seasonings and broths, cereal, chocolate, sweet spreads (e.g. nutella, peanut butter), breadcrumbs, cookies, ice creams and popsicles, margarine, mayonnaise, jams, honey, industrialized bread, pastas, refrigerated dessert, powdered desserts (e.g. jello), ketchup, salsas, homemade bottled salsas, pasta sauce, and instant soups, etc.

Tables **??** and **??** show, by product-group, the percentage of items clas-

Table A4—Tax variable: sugary drinks

Product type	Barcodes with tax (percent)
Bottled water	0
Sparkling water	0
Powdered drinks	74
Energy drink	94
Carbonated soft drink	92
Sports drink	100
Non-carbonated soft drink	70
Coffee	10
Beer	0
Cornflour drink (atole)	0
Vegetable Juice	100
Powdered Milk	0
Evaporated Milk	0
Drinking Milk	0
Flavored Milk	100
Iced-tea	86

sified as being subject to either the sugary drink (*tax_beverage*)or the high caloric-dense food tax (*tax_food*).

Table A5—Tax variable: caloric dense food

Product type	Barcodes with tax (percent)
Cooking oil	0
Baby food	0
Canned tuna	0
Cereal bar	100
Snacks	100
Condensed milk candy	98
Seasonings and broths	5
Cereal	98
Chocolate	96
Tomato puree	0
Liquid seasoning	17
Creamer/Substitutes of cream	0
Sour cream	0
Sweet spreads	100
Breadcrumbs	100
Cookies	100
Ice creams and popsicles	0
Condensed Milk	0
Margarine	99
Mayonnaise	99
Jams	0
Honey	76
Flavored Milk Powder	97
Industrialized bread	63
Pastas	100
Refrigerated Dessert	10
Powdered Desserts	6
Tomato puree	0
Ketchup	4
Snack sauce	0
Homemade bottled sauce	0
Pasta Sauce	0
Instant Soups	36
Yogurt	0

A5. Price Imputation

Barcode selection: We first look at barcodes that exist (i.e. were sold to someone, in some store, in some city in the Kantar data) in *all* years (i.e. at least sold in one week of 2012, 2013, 2014). We only study those barcodes.

The problem of missing values: we are interested in the prices that would be paid for the same 2013 basket of consumption for each household[7]. We keep the basket fixed at 2013 since we want to isolate any changes on quantities in 2014 derived from higher taxes. We therefore need to price the 2013 basket for each week of 2014 and 2013 (and for some exercises for 2012 as well), but it is very common that a household does not buy the same barcode every week, and therefore in that week we have to impute the price that *that* household would face for that barcode in the that storebrand. It turns out that for 2012, 2013 and 2014 we have to impute 52,096,512 values of price.

Creation of "cells" for imputation: Next we create (the most dissagregated) cells as a product of combinations of city-storebrand-barcode-time to be able to impute mean cell prices for the case the price is missing for a barcode b which household i purchased in 2013 in a given storebrand j[8] in its city c, in a week t. We could create a full city-store-barcode-week Cartesian product, but that would generate 6,462,260,532 cells for SDs and

[7]A basket of consumption for household i is a duple D:=(barcode, storebrand), i.e. it not only involves what good the household buys but also where it buys it. A duple D is in the 2013 basket for household i if in any week of 2013 the household bought that barcode is a storebrand j. For example if I only buy coke 500ml in Wallmart and coke 500ml in Costco my basket only has two elements. Differentiating across storebrands eliminates price changes due to pure substitution across stores

[8]A storebrand is for example Wallmart, 7 eleven, Oxxo, Superama, etc.

13,854,472,632 cells for HCFs.[9] Many of these cells would not be useful however; for example if barcode A is only sold in one city, it does not help to define cells with that barcode for other cities. Similarly, if only 7 Eleven stores sell barcode X it is not useful to generate a cell of what this barcode would cost at Wallmart as that cell would always be empty. For parsimony we first define an event-unit (EU) as a triplet (city, store, barcode). Some cities then may have more EUs since they may have more storebands or more barcodes. Then we expand EUs to the 156 weeks of 2012-2014. This is our finest level of disaggregation. We do the imputation separately for sugary drinks (see Table **??**) and for High Caloric Foods (Table **??**).

Lets us explain Table **??** for SDs. We begin with the EUs × 156 which generates 24,326,328 cells.[10] Many of these are empty of course. In fact, only 4.2 percent are non-empty; i.e. 1,021,706 cells. This might seem like a lot of missing information, however these 1 million plus non-empty cells are sufficient to impute 33.1 percent of the all the missing *household ibjt* prices however. One way to measure the imputation "precision" is to have an idea of the within cell price variation, i.e. the variation not explained by the 1,021,706 cell dummies. It turns out that regressing actual prices (i.e. before imputation) on these dummies returns an R^2 of 0.98.

To impute prices for the remaining 76.9 percentof missing observations we are forced to go to a higher level of aggregation. We could for example aggregate across cities but keep separate cells for storebrand-barcode-week, or alternatively we could first aggregate across storebrands and keep separate cells for city-barcode-week. It turned out that even if the second

[9]There are 91 cities, 2863 SD barcodes, 6138 HCF barcodes, 159 storebrands, and 156 weeks.

[10]Note that we are note that cells do not directly have household in their 'cartesian' product.

Table A6—Sequential price imputation for SDs

Imputation of Prices for SDs					
		Prices Panel			Fixed Baskets Panel
Step	Aggregation Level	R^2	Cells	Fill. Cells (percent)	Imp. Observ. (percent)
1	City, store, week, article	0.98	24326328[1]	4.2	33.1
2	Store, week, article	0.94	6485700	29.6	30.8
3	City, week, article	0.94	8092188	16.6	10.3
4	Week, article	0.85	303264	36.1	17.6
5	Month, article	0.84	69984	8.0	4.7
6	Quarter, article	0.83	23328	2.9	1.6
7	Year, article	0.82	5832	2.5	1.5

Both panels are balanced in the sense that for each price or fixed basket we have 156 weeks of observations. [1] Imputed cells of the first aggregation level.

aggregation generates more cells[11], which of these two we used first in the sequential imputation was not very important.[12] Using the storebrand-barcode-week aggregation meant that we could fill-out an additional 29.6 percent of the original 24,326,328 cells.[13] Now turning to household level imputation, which is what we care about, this second step imputes an additional 30.8 percent of the original missing values. This means that until this point we have imputed 63.9 percent of the original missing values.

We proceed in an analogous fashion to higher and higher levels of aggregation, but note that we do not need to aggregate much to get almost all of the imputations: by the 3rd step we have 3/4 of the imputation and by the 4th step we have more than 9/10. Note actually that even the coarser cell dummies have substantial explanatory power: the R^2 in step 7 is 82 percent. This means that barcodes are very good predictors of prices and that prices at the barcode level for SDs do not vary too much by city, time

[11] 8,092,188 vs 6,485,700.

[12] In fact they have the same R^2 of 0.94 in a regression of unimputed price against their cells.

[13] So at this point we have filled out 4.2 percent + 29.6 percent of the original cells.

and store.

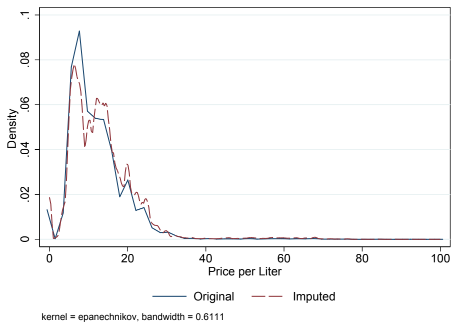
Table A7—Sequential price imputation for HCFs

Imputation of Prices for HCFs					
		Prices Panel			Fixed Baskets Panel
Step	Aggregation Level	R^2	Cells	Fill. Cells (percent)	Imp. Observ. (percent)
1	City, store, week, article	0.97	54001272[1]	4.3	33.3
2	Store, week, article	0.72	14202864	28.0	29.7
3	City, week, article	0.89	17483232	15.2	10.3
4	Week, article	0.41	641472	35.7	17.5
5	Month, article	0.3	148032	9.2	4.8
6	Quarter, article	0.27	49344	3.8	1.9
7	Year, article	0.19	12336	3.8	2.5

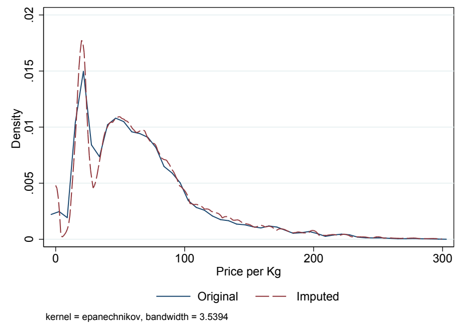
Both panels are balanced in the sense that for each price or fixed basket we have 156 weeks of observations. [1] Imputed cells of the first aggregation level.

The procedure is analogous for HCFs –but we use HCF barcodes instead of SD barcodes and the Table is interpreted analogously so we wont explain it verbally here. Let us just say that beyond the high R^2s we have reported, three extra facts make us confident that these imputations are reasonable. First, if we estimate a *barcode-level* regression with week, household and barcode fixed effects (Table **??**), the estimated increase in SD and HCF prices are 14.3 and 5.5 percent, respectively, which are very close to our estimates presented in the main text. This means that our imputation+aggregation is not distorting the effect of the tax on price. The second is that the distribution of prices before and after imputing is very similar, as shown in Figure **??**. Third, note that we follow the same imputation protocol for all years. Since our strategy is to compare across years, the imputation should not have important influence on our estimates.

Figure A11. Price distributions before and after imputations



(a) SDs



(b) HCFs

Note: Price densities in KWP before and after imputation.

Table A8—Price Effects at the Barcode Level

Robustness: Impact of the Tax on Prices						
	Price per Liter		Price per Kg		Price per 100 Cal	
	(1)	(2)	(1)	(2)	(1)	(2)
Tax	0.143	0.139	0.055	0.046	0.044	0.046
	[0.001]	[0.002]	[0.001]	[0.002]	[0.001]	[0.001]
Barcode×Storebrand×Week FE	Yes	No	Yes	No	Yes	No
Barcode×Storebrand FE	No	Yes	No	Yes	No	Yes
Observations	2,651,936	2,651,936	3,471,284	3,471,284	3,471,110	3,471,110
R-squared	0.616	0.657	0.567	0.621	0.795	0.825

All regressions include a quadratic time trend. This table reports RD regression results using an specification analogous to equation (2), except that an observation is a purchase event (i.e. it is dissagregated by barcodes, etc.). We use weights proportional to liters or proportional to kilograms to have a meaningful comparison with the regressions in the paper. Robust standard errors clustered at the date level in brackets.

A6. Additional Descriptive Statistics

In the main text we show how the main descriptive statistics compare between KWP and other national representative surveys. In this section, we add some descriptive statistics that may be useful to better understand the composition of households' consumption bundles in KWP and look at this statistics by SES and BMI groups. Table **??** presents additional descriptive statistics regarding the composition of households' consumption bundle according to KWP. Food and drinks comprise 68.9 percent of total expenditures in KWP, while 14.2 and 13.5 percent of total expenditures correspond to HCF and SD, respectively. In terms of taxing calories, 42.6 percent of the calories consumed on average come from a food or drink product that is now taxed (25.8 percent correspond to HCFs and 16.8 percent to SDs). Interestingly, neither the fraction of total expenditures devoted to HCF or SD nor the fraction of total calories purchased from HCF and SD vary considerably across SES and BMI categories.

TABLE A9—DESCRIPTIVE STATISTICS

	Weekly Exp (2013 mxn pesos)	Share of Total Expenditures Food	HCFs	SDs
Total				
Total	392.2	68.9	14.2	13.5
Socio-Economic Status				
ABC+	441.2	68.9	14.1	12.8
CD+	402.6	68.7	14.2	13.5
DE	350.9	69.3	14.3	14.0
Body Mass Index				
Slim/Ideal	364.5	68.4	14.7	12.6
Overweight	407.0	69.4	13.9	14.3
Obese	391.5	68.6	14.3	13.2
	Mean Total Calories	Share of Total Calories	HCFs	SDs
Total				
Total	16912.8		25.8	16.8
Socio-Economic Status				
ABC+	17285.3		26.8	16.5
CD+	15839.2		25.8	17.0
DE	15756.0		25.3	16.8
Body Mass Index				
Slim/Ideal	15756.0		26.4	15.5
Overweight	17697.7		25.3	17.8
Obese	16737.6		26.1	16.4

Table based on our classification of barcodes into taxable and non-taxable, as well as the imputation of calories described in section A2.

A7. Main Results

The main findings in the paper indicate that the tax caused a 14.6 and 5.3 percent price increase in SDs and HCFs, respectively. Also, the quantities consumed were somewhat affected in the case of SDs where the consumption decreased 6.7 percent, while HCFs had a non significant 0.3 percent change in terms of quantities.

Tables **??** and **??** show the results of regressions analogous to those that produce the previous results, this time using the log of the price per 100 calories and the log of calories purchased from SD and HCF, respectively. All the estimated coefficients are very similar in magnitude to those we present when using liters of SD or kilograms of HCF as our unit measure.[14] The results obtained from these specifications find a 14.5 and 4 percent price increase for SDs and HCFs, respectively. Meanwhile, quantities of SDs deacrease 6.3 percent and those of HCFs have a non-significant increase of 0.5 percent. In summary, the changes found are very similar when using calories instead of liters and kilograms.

[14]Given that magnitudes of liters and kilograms are different, the results presented here normalize the caloric content to that of one average liter and kilogram of SD and HCF to make the coefficients easily comparable across the tables.

Table A10—Results using calories as unit measure for SD

Impact of the Tax on the Price per 100 Calories of Sugary Drinks							
		By Body Mass Index			By Socio-Economic Status		
	Total	BMI ⩽ 25	Overweight	Obese	ABC+	CD+	DE
Tax	0.145	0.141	0.145	0.148	0.141	0.146	0.147
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Observations	762,706	143,255	329,988	289,277	148,408	348,672	265,440
R-squared	0.985	0.987	0.985	0.985	0.984	0.986	0.986
Impact of the Tax on the Purchase of Calories of Sugary Drinks							
		By Body Mass Index			By Socio-Economic Status		
	Total	BMI ⩽ 25	Overweight	Obese	ABC+	CD+	DE
Tax	-0.063	-0.058	-0.063	-0.068	-0.078	-0.059	-0.061
	[0.006]	[0.007]	[0.007]	[0.006]	[0.007]	[0.006]	[0.006]
Observations	762,706	143,255	329,988	289,277	148,408	348,672	265,440
R-squared	0.780	0.782	0.776	0.779	0.792	0.776	0.776

The *tax* variable indicates if the observation is captured after January 1st 2014. All estimations include household-week of the year fixed effects and a quadratic time trend. Each column shows the result for a different subsample. Data employed for this estimation comes from the weekly KWP Mexico chapter for 2013 and 2014. The unit of observation is the household-week pair. Standard errors clustered at the week level in brackets.

Also, Table **??** shows the result from using a regression discontinuity design as the one illustrated in Figure 4 in the main text. The estimates shown in the table correspond to estimating the following specification:

$$\ln\left(P_b^{SD}\right) = \beta_0 + \beta_1 \ I(Den_b > 275) + f(Den_b) + \epsilon_b \tag{1}$$

where b denotes barcode and Den_b is the caloric density of barcode b. In the estimates presented, we use a non-parametric estimation in the case of column 1 and a linear trend with different slopes before and after the cutoff in column 2. To make the results comparable to those presented with the household-week as unit of observation, each observation used (which corresponds to a barcode) is weighted by the number of transactions done

Table A11—Results using calories as unit measure for HCF

	Total	BMI ⩽ 25	Overweight	Obese	ABC+	CD+	DE
Impact of the Tax on the Price per 100 Calories of HCF							
		By Body Mass Index			By Socio-Economic Status		
	Total	BMI ⩽ 25	Overweight	Obese	ABC+	CD+	DE
Tax	0.040	0.039	0.041	0.038	0.030	0.040	0.045
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Observations	762,706	143,255	329,988	289,277	148,408	348,672	265,440
R-squared	0.866	0.848	0.884	0.855	0.866	0.870	0.862
Impact of the Tax on the Purchase of Calories from HCF							
		By Body Mass Index			By Socio-Economic Status		
	Total	BMI ⩽ 25	Overweight	Obese	ABC+	CD+	DE
Tax	0.005	0.019	0.003	0.000	-0.054	0.025	0.010
	[0.021]	[0.024]	[0.022]	[0.022]	[0.029]	[0.023]	[0.017]
Observations	762,706	143,255	329,988	289,277	148,408	348,672	265,440
R-squared	0.650	0.664	0.649	0.644	0.658	0.639	0.659

The *tax* variable indicates if the observation is captured after January 1st 2014. All estimations include household-week of the year fixed effects and a quadratic time trend. Each column shows the result for a different subsample. Data employed for this estimation comes from the weekly KWP Mexico chapter for 2013 and 2014. The unit of observation is the household-week pair. Standard errors clustered at the week level in brackets.

with that barcode.[15]

The results found with this specification give a 7.7 to 8.7 percent change in prices, which is around the 8 percent tax rate, meaning something close to a 100 percent pass through. The results of using this specification are slightly larger than the 5.3 percent estimated in the main regressions. It should be indicated though that as a traditional RD result, the estimate is valid for the value of density at the cutoff. Also, it is important to recall that in the paper, no evidence of strategic decisions around the discontinuity is found (see Figure 7).

[15]The purpose of this is to represent that a product with very high demand should capture more importance that a product that is seldom consumed.

Table A12—RD for HCF Prices

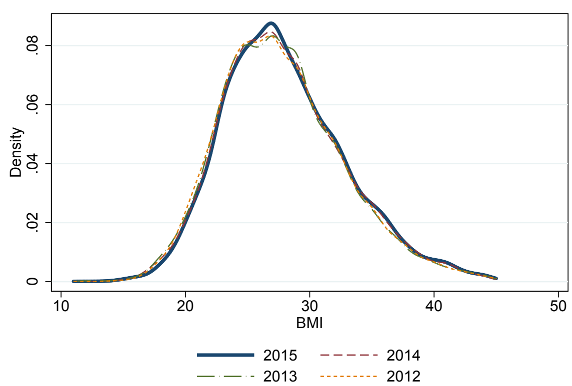
Regression Discontinuity for HCF Prices		
	NP (bw=5)	Linear
Density>275	0.087	0.077
	[0.037]	[0.035]
Observations	8,966	3,322
R-squared	-	0.008

The first column presents the results of a non parametric (NP) RD estimate. The second presents a parametric estimate adjusting a linear trend on each side of the discontinuity. Each article is weighted according to the acts of buying registered in our dataset. For the linear specification, we restrict the estimation to the domain of [180, 400] kcal per 100 grs.

A8. Estimates of effects on BMI

The taxes introduced were aimed at combatting obesity. For this reason, individuals' BMI may be the most relevant outcome to analyze. If the intake of total calories did not change as a result of the taxes, we should not observe significant changes in BMI. Before presenting a differences in differences analysis, Figure **??** just plots kernel densities of BMI for several years in Mexico, from two years before and two years after the tax. The resemblance is striking. There was no change in BMI in any part of the distribution.

Figure A12. Overweight in Context

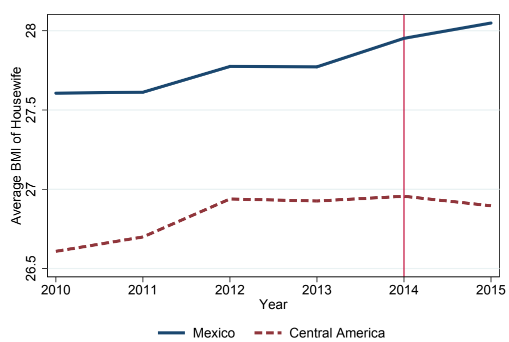


Note: Source: Kernel density estimates of BMI as reported in KWP.

Figure KWP **??** of course does not contain a counterfactual comparison. We would need a control group for which there was no tax in 2014. Our empirical strategy next is to compare Mexico vs its Latin American neighbors. This is reasonable since (a) we find the same pre-trends in BMI, (b) the tend to consume similar baskets of goods, (c) they were not subject to the tax and we were able to get BMI data at the household level from the same source (KWP) using the same methodology.

We use yearly self-reported information on the weight and height of female household heads for participating households in Costa Rica, Guatemala, El Salvador and Panama as a control group[16] for Mexican households. Our identification assumption is that in the absence of the taxes households in Mexico would have experienced a similar change in BMI as households in the these four countries. As shown below this assumption is reasonable; these countries have similar pre-trends in BMI than Mexico.

Figure A13. BMI Trends in Mexico vs Central American Countries



Note: The graph simply plots BMI for our sample of female household heads in Mexico, and for the rest of Central American countries, pooling all samples together.

Figure **??** shows that, consistent with WHO's data [17], (average across households) BMI has increased both in Mexico and the rest of countries in Central America during the period of analysis. This adds confidence that the BMI data is not noise with the same mean each year. Moreover, although

[16]KWP collects BMI for a subset of 89 percent of male households heads in Mexico, but unfortunately this information is only available for 2 years in Central America. We therefore focus on BMI of the female head. We have 54,143, 4,442, 4,918, 4,026 and 4,315 household-year observations from Mexico, Costa Rica, Guatemala, El Salvador and Panama, respectively on the female heads BMI.

[17]See section **??** of the online appendix.

BMI is higher in Mexico than in the rest of Central America, the time trends showed by both Mexico and the rest of Central America are relatively similar before the enactment of the tax, supporting the identification assumption of parallel pre-treatment trends.

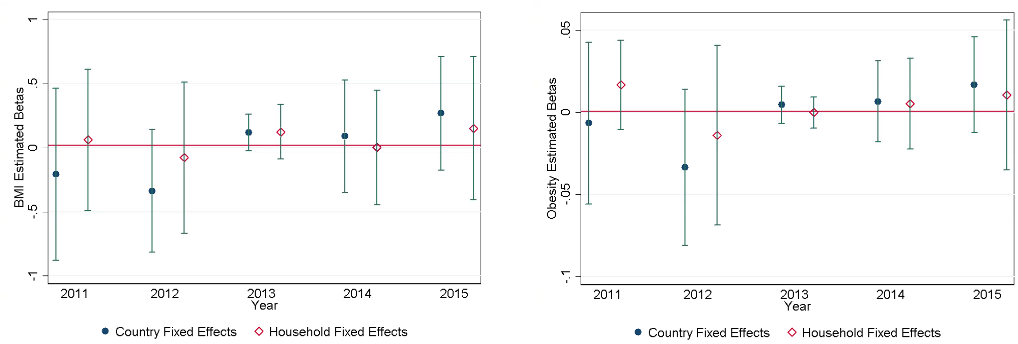
Graphical evidence suggest that there was not change in the evolution of BMI over time for Mexico after the taxes were introduced. In order to provide statistical evidence we estimate the following difference-in-difference (DD) specification:

$$BMI_{ijy} = \alpha_j + \theta_y + \sum_{t=2011}^{2015} \beta_t Mexico_j * I(y = t) + \epsilon_{ijy} \tag{2}$$

Where i, j and y denote individual, country and year. The equation includes dummies for 5 years 2011-2015 (θ_y), dummies for each Central American country (α_j) to allow for country-specific BMI levels, and an interaction of the Mexico (*treatment*) dummy with the year dummies. We cluster standard errors at the country level. The β_t's are the coefficients of interest, which measure the differential change in BMI (respect 2010) experienced in Mexico versus other Central American countries each year.

Because the panels are unbalanced, we present regression results with and without female household head fixed effects. Figure **??** plots the estimated β_t's along with their 95 percent confidence intervals. The joint F-test for $\beta_j = 0 \quad \forall\, j \in \{2011, 2012, 2013\}$ is not rejected (p-value=0.17), suggesting that the parallel trends assumption holds. The coefficients for 2014 and 2015 show no effect of the tax on BMI for Mexico. The effect of the tax on weight may take time, but the fact that β_{2015} is larger than β_{2014} is discouraging. Panel B of figure **??** uses the fraction of obese population

Figure A14. BMI and Obesity Differences in Differences Analysis



(a) BMI (b) Obesity

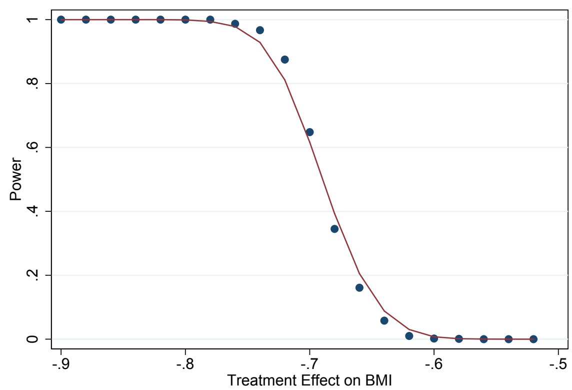
Note: This Figure plots the coefficient estimates from equation **??**. Results from years 2011, 2012, and 2013 show that the parallel trends assumption is satisfied. Results from 2014 and 2015 show that there is no effect of the tax.

as a dependent variable and reaches similar conclusions. Given the strong relationships between household members BMIs found in the literature, we believe our results may apply more broadly to other household members.[18]

Figure **??** displays a power simulation for BMI using our exact same specification **??** and our same data base. This gives us an exact power calculation that takes into account the whole structure of our data and specification. The simulation proceeds as follows: we take 2013 for Mexico as our placebo tax year and define the DID dummy = 1 for a household if the country is Mexico and the year is $\geq$ 2013. For each obervation for which DID=1, we add a placebo effect on the BMI dependent variable. The placebo effect is drawn randomly from a normal distribution with mean μ and standard

[18]We actually estimate a coefficient of 0.46 (t-stat=95) in a regression of male and female household heads BMI's. Section **??** of this online appendix reports quality checks on the weight and BMI measures. Height in the data does not change –as expected– but weight does change by reasonable magnitudes with mean changes between 1 and 2 kilos. In power simulations that use our exact empirical specification we obtain that we are able to detect effect sizes of about 0.75 BMI points and 1.3 kilos with 90 percent power and 95 percent confidence.

Figure A15. Power graph



Note: Figure ?? displays a power simulation for BMI using our exact same specification ?? and our same data base.

deviation σ. The standard deviation is taken from the empirical distrbution of the female head BMI year-to-year change, while μ runs from 0 to -1 BMI points, in steps of 0.02.

For each μ we run 400 simulations. In each simulation for each household we draw a placebo effect from the normal distribution described above. Then we estimate the regresion in equation ??, giving us 400 estimated effects. Finally we count the fraction of the placebo effects which are statistically different from zero at 95 percent confidence. This fraction is our measure of power for each μ.

Figure ?? shows that we can detect an effect of about 0.75 BMI points with above 80 percent power. For a person with 1.60 meters in height, this is equivalent to about 2kg. We think this is a reasonable effect to detect even on average in 2 years, to decrease obesity people would need to decrease oftentimes more than 10kg. Not finding effects of 2kg is informative indeed.

A9. Mechanisms

A. Change in Consumption Decisions

The results presented in section VI.A of the paper show interesting patterns in the consumers decision making process. In particular, the substitution towards cheaper goods within the high caloric density foods calls for further analysis and attention. Given that this substitution happens within the HCF group it can mean that consumers are switching towards cheaper and lower quality products within an homogeneous category (e.g. choosing a cheaper brand of the same product) or changing consumption among HCF categories toward a cheaper group (e.g. changing peanuts or pistachios for chips). To further investigate this, in Table **??** we show the change in the 2013 value of the chosen consumption basket for different categories of product (as in columns 1 and 4 of Table 6 in the main paper). This means that the different estimates will only capture within product category substitution. The evidence from the table shows that *within product category* substitution is an important driver of the results presented in section VI.A. The case of snacks and cookies is notable for its large decrease. The rest of the categories included in this table show lower decreases that range from 4.5 to 0.4 percent for categories such as industrialized bread, crackers, cereals and spreadable creams.

B. Substitution Towards Smaller Presentations

In the main text, we show that products containing more liters of sugary drinks increased prices in a higher proportion with respect to smaller presentations within the same category. Conversely, the price of barcodes in the HCF category that contain more calories per unit increased relatively less

Table A13—Substitution within a product category

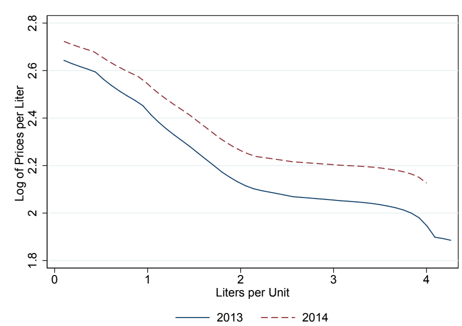
Substitution Patterns Within Food Categories						
	Log 2013 PC					
	Snacks	Bread	Cookies	Crackers	Cereal	Spreadables
Tax	-0.115	-0.033	-0.075	-0.045	-0.004	-0.008
	[0.007]	[0.002]	[0.002]	[0.002]	[0.002]	[0.001]
Observations	762,706	762,706	762,706	762,706	762,706	762,706
R-squared	0.669	0.644	0.626	0.638	0.640	0.647

The *tax* variable indicates if the observation is captured after January 1st 2014. All estimations include household-week of the year fixed effects and a quadratic time trend. Each column shows the result for baskets of different categories of products. Data employed for this estimation comes from the weekly KWP Mexico chapter for 2013 and 2014. The unit of observation is the household-week pair. Standard errors clustered at the week level in brackets.

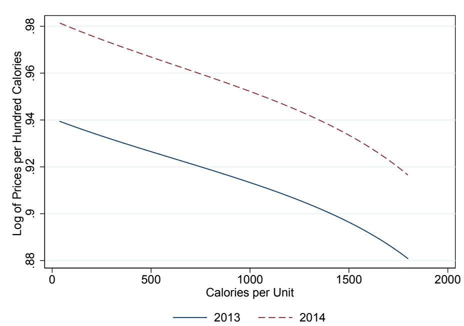
with respect to barcodes with less caloric content within the same category. This finding is illustrated in the main text in Figure 5. Figure **??** gives more detail about this contrasting fact by plotting in two graphs the changes in prices, one for SDs and another for HCFs. Figure **??** shows that both, in the case of SDs and HCFs, products with more liters (or calories) per unit have a lower price per liter (or calorie). This is shown by the negative slope in the graphs that relate price per liter and liters per unit (or price per calorie and calories per unit). The contrasting effect of the tax between SDs and HCFs is evidenced by the fact that this negative relation shifted to the right in both cases as a result of taxes. However, this shift is more pronounced at higher values of liters per unit for SDs and contrarily, in lower values of calories per unit for HCFs.

In addition, in the main text, we argue that the larger increase in prices for large SD presentations may have incentivized individuals to substitute towards smaller presentations. Table **??** provides additional evidence in this respect. It presents the results of our main specification, this time using the total liters of SD consumed, but for different categories of SDs, divided by their presentation size. This table shows that the decrease in liters of SDs purchased is driven by a large decrease in SD presentations with more than 1.5 liters in content (10.7 percent decrease). This decrease is even partially compensated with increases in the midsize presentations: 1.7 and 1.8 percent increase in the (355ml,600 ml] and (600ml.,1.5l.] sizes. In addition to the results in quantities consumed by presentation size, the table also contains effects in prices. The results from the estimations in prices confirm that the bigger presentation exhibits the largest increase (16.3 percent) followed by the medium sizes and the smallest group has the lowest effect (5.6 percent).

Figure A16. Differentiated Impact of the tax on SDs and HCFs Prices



(a) SDs



(b) HCFs

Note: The left hand side graph plots the average of the log price of SD barcodes in 2013 and 2014 against the content in liters of each barcode. The right hand side graph plots the average of the log price per 100 calories of HCF barcodes in 2013 and 2014 against the content in calories of each barcode.

Table A14—Effect of SD Tax By Presentation Size

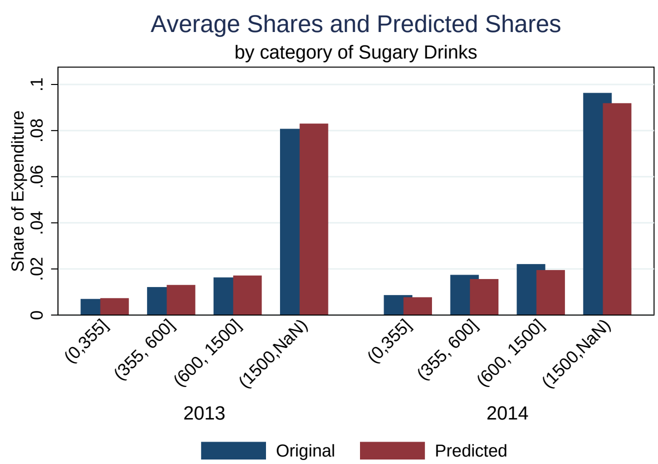
Log price per liter of SD (by presentation size categories)				
	Presentation			
	(0,355]	(355,600]	(600,1500]	(1500,NaN)
Tax	0.056	0.113	0.099	0.163
	[0.001]	[0.001]	[0.001]	[0.002]
Observations	762,706	762,706	762,706	762,706
R-squared	0.902	0.920	0.929	0.959
Log liters of SD (by presentation size categories)				
	Presentation			
	(0,355]	(355,600]	(600,1500]	(1500,NaN)
Tax	0.002	0.017	0.018	-0.107
	[0.0015]	[0.0017]	[0.0021]	[0.0051]
Observations	762,706	762,706	762,706	762,706
R-squared	0.716	0.740	0.730	0.789

The *tax* variable indicates if the observation is captured after January 1st 2014. All estimations include household-week of the year fixed effects and a quadratic time trend. Each column shows the result for baskets of different groups of products. Data employed for this estimation comes from the weekly KWP Mexico chapter for 2013 and 2014. The unit of observation is the household-week pair. Standard errors clustered at the week level in brackets.

C. AIDS model fit

In this section of the appendix we provide some evidence of the quality of the estimated AIDS model to fit the data and to predict the changes in expenditure out of sample. This is important since the AIDS models have several assumptions which may bias their predictions if they turn out to be wrong. The paper provides one measure of fit by predicting the 2013 liters consumed, and a measure of predictive accuracy by projecting the liters consumed out of sample for the year 2014. The results show a reasonable fit. Here Figure **??** shows that the actual 2013 and predicted 2014 *share of expenditures* for the four categories of SDs are very similar. Given that the AIDS estimation was done with information from 2013 finding a good 2013 fit is not very surprising. Nonetheless, the model is then employed to forecast the levels of expenditure shares that would result from the tax effects and they closely match the actual 2014 shares in the data. The evidence from the graph gives a robust support to our estimates since the observed and predicted levels look very similar despite the fact that this is an out of sample estimation.

Figure A17. AIDS in sample and out of sample fit



Note: This Figure uses the AIDS model estimated on 2013 data and shows the model predicted "in-sample" shares of 2013. It also shows how the model fares in predicting 2014 shares. This later prediction is an "out-of-sample" prediction as the model was not estimated with 2014.

A10. Robustness checks

A. *Anticipation*

An alternative behavioral response to taxes is to keep large stocks as the tax implementation approaches. With this response, household can substitute consumption just after the tax implementation, particularly with non perishable products. We perform a robustness test in section VII.A of the main text, where we keep out of the sample a window of time around the beginning of the tax implementation and obtain similar results. Following the marketing literature, in this section we perform two additional exercises to support that dynamic considerations are not biasing our estimates. We explore if (a) there were more purchases of SDs and HCFs in December 2013 vis-à-vis December 2012 to accumulate inventories; (b) whether the probability of buying was lower in January 2014 than in January 2012 or 2013; and (c) if households spend more weeks without making purchases of SD and HCF in January 2014 versus January 2012 and 2013. The results of this exercise are presented in Table **??**. No evidence of stockpiling behavior from the households is found. The only statistically significant coefficient is an increase of 0.1 weeks in time to buy after January 1st 2014. This is economically very small.

B. *Falsification tests*

Our second exercise is a set of falsification tests following the logic and output format of a Fisher exact test[19]. Our aim is to show that the effect captured with the tax variable in our main estimations is indeed a quite

[19] We do not call it a Fisher exact test as that test requires sampling from the same distribution and here we are sampling from different months.

Table A15—Evidence of Inventory Behavior

Robustness: Storage of SDs and HCFs						
	SDs			**HCFs**		
	Log Liters	Buy	Weeks	Log Kgs	Buy	Weeks
Dec 2012	0.066			0.031		
	[0.105]			[0.041]		
Jan 2012		0.004	-0.105		-0.007	-0.100
		[0.009]	[0.033]		[0.012]	[0.034]
Jan 2013		-0.000	-0.055		-0.008	-0.036
		[0.003]	[0.021]		[0.004]	[0.031]
Mean	1.68	0.86	1.31	0.65	0.90	1.18
SD	[0.96]	[0.34]	[1.26]	[0.43]	[0.3]	[0.75]
Observations	708,461	1,071,550	1,053,177	708,461	1,071,550	1,057,024
R-squared	0.473	0.204	0.282	0.223	0.097	0.090

This table estimates regressions for different measures of SD and HCF purchases: quantities purchased, the likelihood of purchase, and the number of weeks with no purchase. Columns 1 and 4 examine whether there were more purchases of SDs and HCFs in December 2013 vis-à-vis December 2014 and 2012. Columns 2 and 5 estimate if the probability of buying was lower in January 2014 vis-à-vis January 2013 and 2012. Columns 3 and 6 estimate if the household spends a larger spell without purchases of SDs and HCFs in January 2014 vis-à-vis January 2013 and 2012.
All regressions include household and month of the year fixed effects and a quadratic time trend. Clustered standard errors at the week level in brackets.

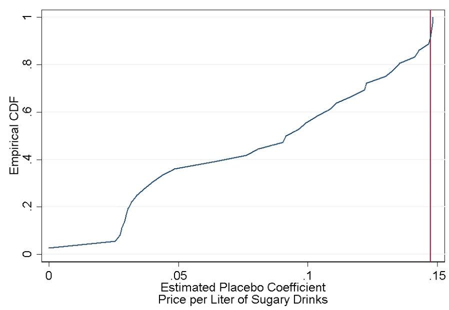
an unusual event in terms of size. This test consists in replicating our main specification (equation 2) several times and using in each replication a different 'placebo' date to generate the tax variable. In total 34 replications were done, and in each replication a different month between February 2012 and November 2014 was employed to create the tax variable. In particular, the replication that uses January 2014 is exactly our main specification. The other replications establish artificially the tax at dates where it did not happen and as such are placebo estimates.

If say end of quarter months were particularly good at SDs sales, this would show up in their respective month effects. So that these falsification tests can potentially detect if our results are due to seasonality for example.

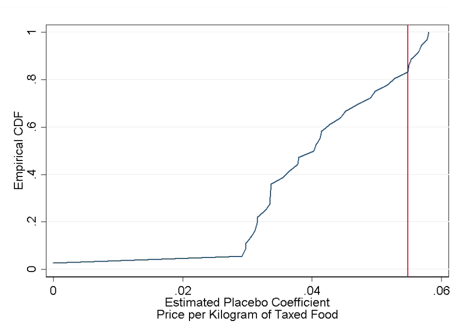
We proceeded to estimate the 34 regressions and collected the resulting $\{\beta^1, \ldots, \beta^{34}\}$ from our estimations. We do this for each dependent variable separately. If our equation is correctly specified and if in reality the tax in January 2014 did have an effect, we would expect all betas other than the one for January 2014 to be statistically zero, except perhaps for sampling variance. With 90 percent confidence we should therefore observe that the coefficient for January 2014 is either in the top 10 percent of the distribution when the effect is positive (e.g. price) or the lowest 10 percent when the effect is negative (e.g. quantities), and to be outside this range when there is no effect of the tax (e.g. calories). This test is more precise when we have more coefficients/years. Unfortunately we just have 3 years of data. We still think they are informative though.

Figure **??** uses the set of 34 coefficients and plots an empirical cumulative distribution function (CDF) for each different dependent variable used. We add a vertical red line to indicate the position of the January 2014 coefficient (the true tax). We find that for the price of SDs our true tax effect is in the top 90 percent of coefficients as expected from our previous results, whereas for HCF it is above the 80th percentile. For kilos of food it is in around the 30th percentile (strongly indicating a zero effect), while for SD the reduction is below the bottom 20th percentile. For total calories the increase is close to the 70th percentile. The results of these tests are consistent wit the results of statistical significance we got in the paper.

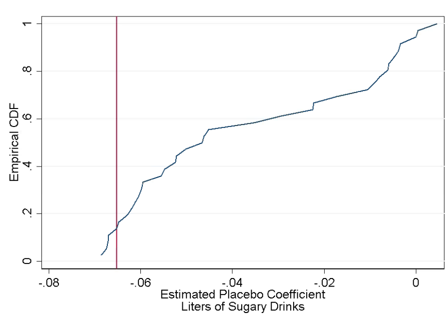
Figure A18. Fisher Exact Placebo Test



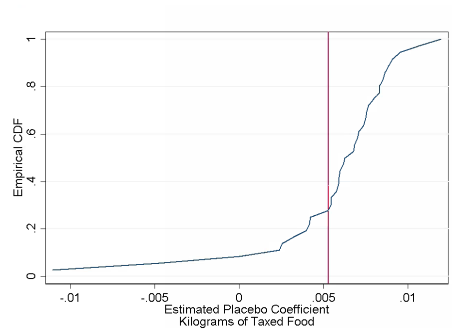
(a) Price per Lt



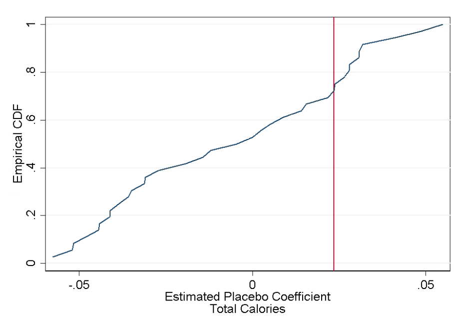
(b) Price per Kg



(c) Liters



(d) Kgs



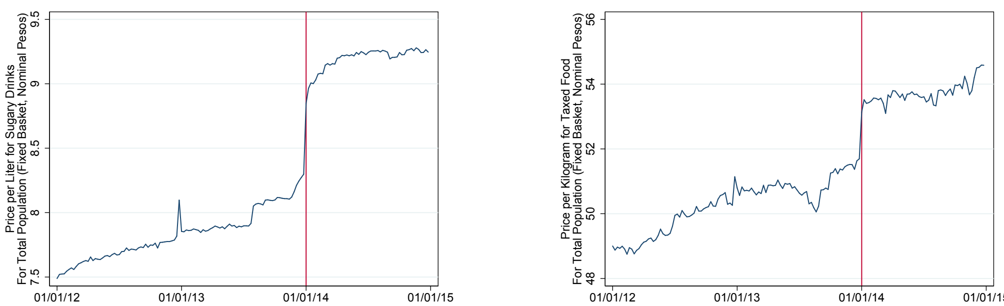
(e) Calories

Note: This figure estimates 34 regressions with placebo months to mimic the tax month. It then ranks the estimated coefficient and plots the cumulative density of these, i.e. the fraction that fall below a particular value. The vertical red line shows where the coefficient for the January 2014 lies. We expect true effects to lie at the extremes of the distribution. This is akin to a Fisher exact test.

C. Is the Change in Prices Attributable to the Taxes?

Figure ?? presents additional evidence in support of our identifying assumption: that the price changes observed in January 2014 are attributable to the taxes introduced. The figures and evidence presented here are akin to that presented in the paper but including an extra year to the analysis. This helps us to corroborate that the changes in prices observed in January 2014 is not common to other Januaries.

Figure A19. Evolution of Prices per Unit of Quantity 2012-2014



(a) Price per Lt (b) Price per Kg

Note: Average consumption bundles are defined from all observed purchases during 2013 for each household, and fixed. For calculating their cost, we multiply the number of units of each barcode in each households' bundle by their current price. The assignment of current prices to barcodes is explained in detail in section ??.

D. Fraction of Total Expenditures in KWP

One potential criticism of our results is that we capture only a subset of food expenses. In order to explore how the estimates presented vary with respect to the fraction of total purchases captured in our dataset, we perform two exercises. First, we categorize households into quintiles of the total consumption of calories. Second, we classify households by a proxy for the expenditure that KWP captures. We then estimate the main regressions of the paper by these quintiles and show that results do not vary by quintile. This lends credence to our estimates not being biased according to the share of expenditure we capture.

Before we go to the results, let us briefly describe how we calculate the proxy for the fraction of total food expenditure captured by Kantar. First we noted that KWP contains a set of socio-economic variables also captured in the Mexican Income and Expenditure Survey (ENIGH) which has total expenditure in food. Second we estimated a regression of total expenditures against these socioeconomic variables on the ENIGH sample.[20] We then impute total expenditures to each household in our KWP dataset using the predicted values from the estimated OLS regression, given the coefficients obtained in the regression using ENIGH and the socio-economic variables captured in KWP. By doing this, we then obtain a (noisy) estimate of total expenditures for each household in our dataset, and are then able to compute, for each household, what fraction of total expenditures is captured in the KWP dataset. We then categorize households into quintiles of the fraction of total expenditures that are captured in KWP and run our main specification for each of these subsamples, using the price per liter of SD, the price per 100 calories of HCF, total liters of SDs, total calories from HCFs, and total calories purchased.

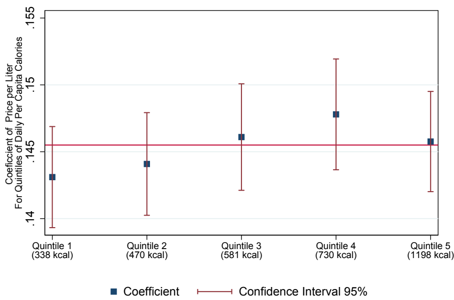
Results are presented graphically in Figures **??** and **??**, which focus on quintiles of total calories and total expenditures captured in KWP, respectively. As can be seen, the estimated increase in the price of sugary drinks is (slightly) increasing with the fraction of expenditures and calories captured

[20]We also tried with includying some variables of food expenditures that were in KWP and ENIGH and got similar results on the imputations.

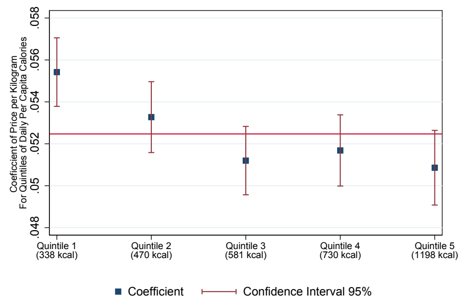
in our dataset, while the drop in consumption of liters of sugary drinks is decreasing in magnitude. As for calories from taxed foods, the log of price is also increasing with the fraction of total expenditures captured, while the change in purchases of calories from these products is negative for the lowest quintile, and positive for the highest. In sum, households for which we capture a larger fraction of their total food expenditures exhibit a smaller decrease in calories from SDs and even a slightly larger increase of calories from HCFs. This is reflected on the fact that total calories increase slightly more for this group compared to our results (i.e. the mean), although this increase is also not statistically significant. A similar result is obtained if quintiles are based on total calories consumed. Taken at face value, this means that if anything the decrease (increase) in calories would be less (more) if we could capture more of the food expenditures.[21]

[21] A caveat from this exercise is that the classification used to form the quintiles might not necessarily reflect groups for which a larger portion of the expenditure is captured, but rather households that have a diet more based on packaged processed foods, which is the bulk of the dataset. In this case the positive relation between effect on consumption and proportion of expenditures captured might only reflect the fact that those with a diet more based on packaged products are less likely to decrease their consumption and thus will see total calories less affected. However, we do not feel this alternative explanation is a threat to our results for the following reasons: (i) the positive relation between the effect on consumption and proportion of expenditures is non monotonic; (ii) even though we only capture a portion of the total calories consumed, given that packaged goods are the ones affected by the tax and, as such, the main source of calories, the effect on the calories not captured would need to be substantial to overturn the total calories result.

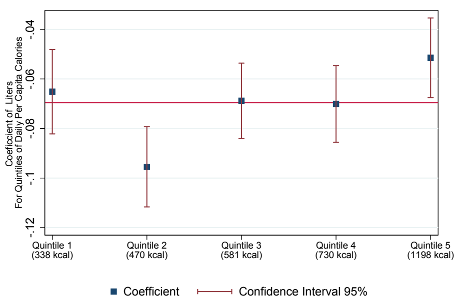
Figure A20. Main Regressions with Caloric Consumption Interaction



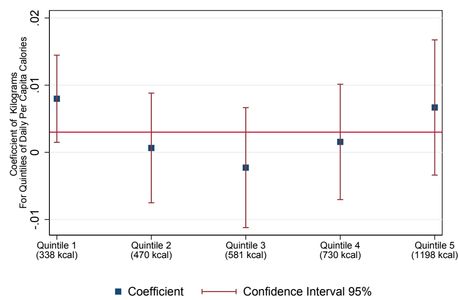
(a) Price per Lt



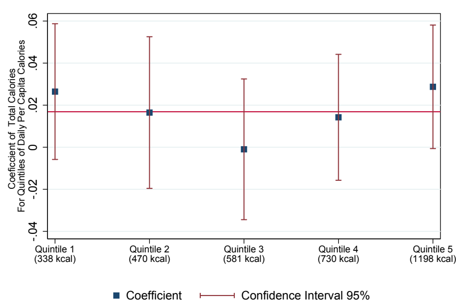
(b) Price per Kg



(c) Liters



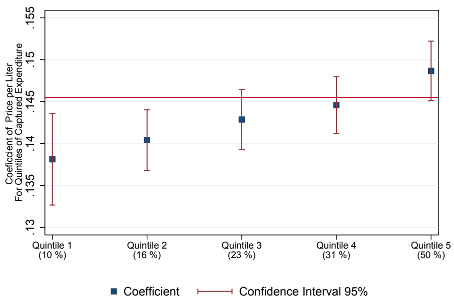
(d) Kgs



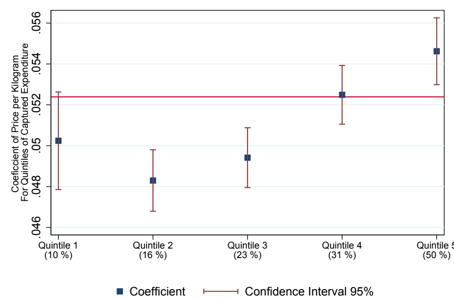
(e) Calories

Note: These graphs plot the coefficient β in equation 2, for different subsamples. In this case, we classify households by quintiles of their consumption of total calories. All regressions include household-week of the year dummies and a quadratic time trend as controls. Each point in the graph shows the results of a different regression for the subsample specified. 95 percent confidence intervals from robust standard errors clustered at the week level are also shown.

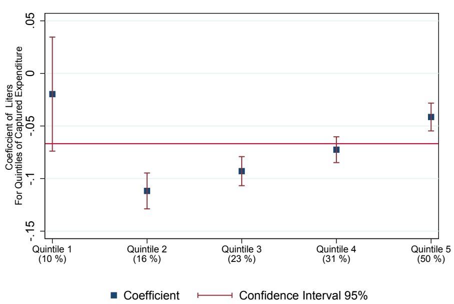
FIGURE A21. MAIN REGRESSIONS WITH EXPENDITURE PERCENTAGE INTERACTION



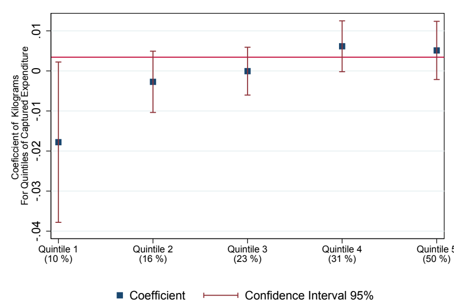
(a) Price per Lt



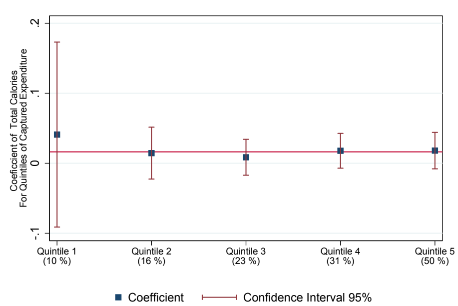
(b) Price per Kg



(c) Liters



(d) Kg



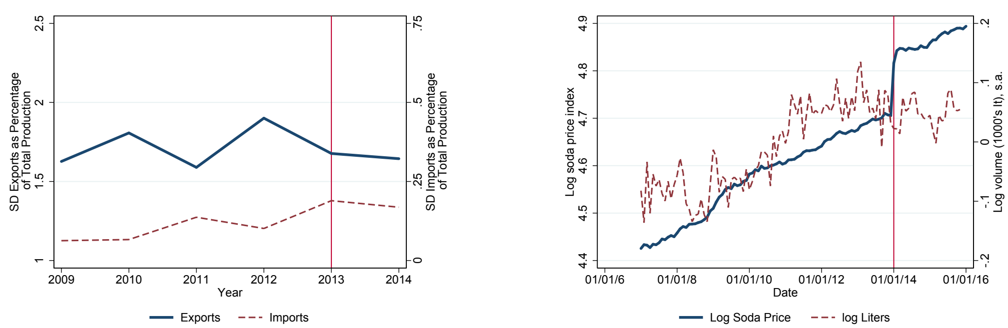
(e) Calories

Note: These graphs plot the the coefficients β in equation 2, for different subsamples. In this case, we classify households by quintiles of the predicted fraction of expenditures in ENIGH that KWP captures, according to their socio-economic characteristics. All regressions include household-week of the year dummies and a quadratic time trend as controls. Each point in the graph shows the results of a different regression for the subsample specified. 95 percent confidence intervals from robust standard errors clustered at the week level are also shown.

E. Industry level production

To strengthen the evidence from section VII.C, here we present two additional plots. Figure ?? gives evidence that soda exports and imports are not a concern in terms of validity of our estimates given that the proportion of sodas imported and exports is negligible and did not exhibited an important change after the tax was introduced. It also gives graphical evidence of the evolution of prices and quantities of sodas nationally with the data used in Table 12. As the figure shows, prices suddenly increased their level at the date in which the tax began to be implemented and consumption exhibited a slight decrease.

Figure A22. Commerce and National Production of Soda



(a) International Commerce (b) National Production

Note: Source: INEGI