ONLINE APPENDIX FOR

The Causal Effect of Competition on Prices and Quality: Evidence from a Field Experiment*

By MATIAS BUSSO AND SEBASTIAN GALIANI

Appendix I: Price Variable Construction

A. Data Sources

Price data were obtained from the responses to three questions (sources). First, the retail survey questionnaire included a question (Question Q1) about 15 products. Retailers were asked about the brand and price of the cheapest brand that is normally available at their stores. This question pre-specified the unit of measurement. Second, in Question Q2, retailers were asked to identify the three products that they sell the most of to program beneficiaries and to provide information about the price, brand, variety and unit of measurement for three different versions of these three products. Finally, in Question Q3, consumers were asked about their weekly expenditure and the physical amount that they bought of each of the 15 products in the last 7 days.

B. Coding Varieties and Brands

In order to code all possible combinations of brand-variety (barcodes) for each product, we pooled all three sources of information. A unique code was assigned to each combination of brand-variety for each of the 15 products. Q1 and Q3 were intended to only deal with brands. In some instances, however, survey respondents mixed brands with varieties. For some products, information about the variety could be recovered from the question even when the respondent did not identify the variety, since in some cases the brand is associated with a particular variety. This imputation of missing information was based on data obtained from the webpages for each product. Two issues warrant discussion. First, the variety of the products is often not associated with a single characteristic. This is more frequently the case for some products than for others. For instance, the variety of eggs

* Busso: Research Department, Inter-American Development Bank; 1300 New York Ave. NW, Washington, DC, 20577 (email: mbusso@iadb.org). Galiani: Department of Economics, University of Maryland; 3114 Tydings Hall, College Park, MD, 20852 (email: galiani@econ.umd.edu).

could differ because of their size, yolk quality, etc. So in those cases, varieties were grouped together even though the relevant attributes differ. Second, neither retailers nor consumers provided information about varieties of bread. The previous table showed the complete list of brands and varieties for each product in our sample.

C. Measures

Average Price (retailers). For each retailer *i* at time *t* (t=baseline, endline), we computed the average over all 15 products (*k*):

$$P_{it} = \sum_{k=1}^{15} W_k * p_{itk}$$

In the case of the weighted average price, W_k is the share of expenditure on product *k* (see below). In the case of the unweighted average price, W_k=1/15 for all *k*.

Average Price (consumer). For each consumer *i* at time *t* (t=baseline, endline), we computed the average (relative) price over all 15 products (*k*):

$$P_{it} = \sum_{k=1}^{K} W_k * \left[\frac{p_{itk}}{\overline{P_{kt}}}\right]$$

In the case of the weighted average price, W_k is the share of expenditure on product *k* (see below). In the case of the unweighted average price, W_k=1/K for all *k*. Many consumers did not report spending for all 15 products. To avoid differences in average prices due to bundle composition, we standardized the price of each product using its average price in the sample.

D. Weights

The weights W_k for the 15 products were created using the household survey. The weights represent the share of monthly expenditure on product *k* made by all the surveyed households at baseline. In all measures, the weights add up to 1.

The weights W_k were compared with the results of a nationally representative survey of program beneficiaries, the Evaluation Survey of Social Protection (EEPS), which was conducted in 2010/2011. In this survey, households were queried about their expenditure on a broader set of products. Appendix Table A1 indicates the results of this comparison. The first column shows the product and the second column, the sample size. The third column shows the percentage of households that reported having consumed a given product in the previous week. The fourth column shows the average share of expenditure on each product. Panel A gives the corresponding information for the 15 products that were covered in our survey. Panel B summarizes the information about other non-perishable products that may

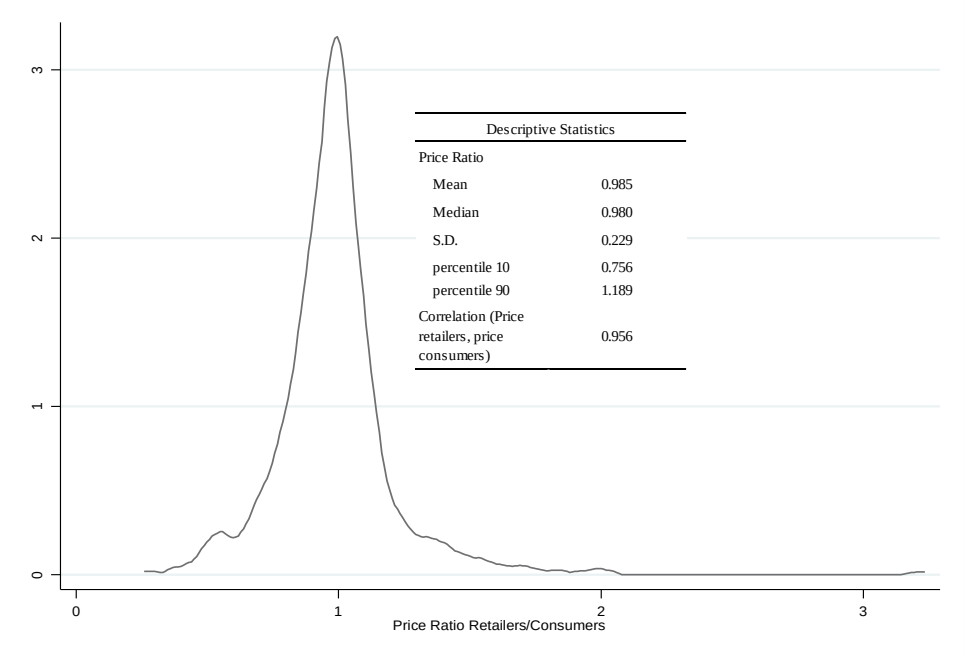
be sold by small-scale retailers. Panel C shows the measures for other fresh or perishable products typically not sold by the retailers in question.

Several facts are worth mentioning here. First, the 15 products included in our survey account for 60% of total food expenditure. Second, the other products that are sold by the retailers under analysis represent 12% of total food expenditure. Third, most households bought these 15 products. Fourth, the weights calculated in our sample are very close to those observed in the EEPS.

E. Price Validation

In order to assess the validity of our price measures, we compare price measures obtained using retailer data with those obtained using beneficiary data (an independent source of information). For each product and brand in all the districts, we calculated an average price based on the prices reported by the retailers and by the beneficiaries.

Appendix Figure 1: Distribution of ($\bar{P}_{ks}^{R}/\bar{P}_{ks}^{C}$)



Note: The figure plots the ratio between the (simple) average price of product j in district s as reported by retailers and the (simple) average price of product j in district s as reported by consumers. The table reports statsitics that describe the distribution of that ratio and, in the last row, the correlation between those two prices.

Let $\bar{P}_{ks}^{R}$ be the average price in district *s* of product *k* computed using retailer information R, which corresponds to the cheapest available option for each product. Similarly, let $\bar{P}_{ks}^{C}$ be the average price in district *s* computed using consumers' information C which corresponds to the products actually bought by consumers. The average district relative price ($\bar{P}_{ks}^{R}/\bar{P}_{ks}^{C}$) is a useful statistic for assessing how close these two measures are. Note that, without measurement error in the measures of prices, this statistic is bounded from above at 1. The next figure shows a kernel density estimation of that price ratio. We find that the average relative price over all products and districts is 0.99.

Appendix II: Tables

TABLE A1. SAMPLE SIZE

	At baseline	At endline
Universe of retailers in area under study	432	425
Universe of entrant retailers	61	61
Sample size: Retailers (in surveys)	401	400
By type		
Incumbent	350	341
Entrant	51	59
Located in targeted neighborhood	257	254
Incumbent in targeted neighborhood	215	212
Sample size: Beneficiaries (in surveys)	2250	2118
By type		
Shop in incumbent retailers	1620	1563
Located in targeted neighborhood	2250	2118
Number of districts	72	72

APPENDIX TABLE A2. EEPS 2010 - SHARE OF EXPENDINTURE ON ALL PRODUCTS

Product	*EEPS 2010*				
	N	Percentage consumption	Share of expenditure	Share of expenditure in price index	Survey weightings
Fifteen survey products			*0.601*	*1.000*	*1.000*
Rice	6783	0.962	0.157	0.262	0.293
Chicken	6784	0.784	0.089	0.148	0.170
Oil	6786	0.936	0.059	0.099	0.094
Milk	6786	0.338	0.045	0.075	0.062
Sugar	6785	0.955	0.045	0.075	0.052
Beans	6785	0.849	0.043	0.072	0.063
Salami	6786	0.758	0.039	0.064	0.048
Eggs	6785	0.792	0.030	0.051	0.050
Bread	6785	0.755	0.028	0.046	0.074
Pasta	6786	0.771	0.019	0.032	0.017
Onion	6785	0.886	0.018	0.030	0.020
Cod	6785	0.192	0.011	0.018	0.018
Sardines	6786	0.216	0.009	0.014	0.014
Chocolate	6784	0.366	0.007	0.011	0.015
Flour	6786	0.278	0.002	0.003	0.010
Other non-perishable products			*0.121*		
Powdered chicken bouillon	6786	0.874	0.025	-	-
Coffee	6785	0.708	0.023	-	-
Water	6786	0.485	0.017	-	-
Tomato paste	6786	0.715	0.017	-	-
Soda	6786	0.296	0.012	-	-
Smoked cutlets	6785	0.142	0.008	-	-
Powdered juice	6786	0.287	0.007	-	-
Ice	6786	0.329	0.005	-	-
Pigeon peas	6785	0.123	0.004	-	-
Dried coconut	6785	0.085	0.002	-	-
Canned green beans	6785	0.026	0.001	-	-
Fresh or persihable products			*0.264*		
White cheese	6785	0.336	0.017	-	-
Milk	6784	0.237	0.007	-	-
Yellow cheese	6786	0.113	0.005	-	-
Butter	6786	0.255	0.004	-	-
Orange juice	6786	0.072	0.003	-	-
Plantains	6785	0.723	0.037	-	-
Avocados	6784	0.787	0.022	-	-
Garlic	6785	0.900	0.022	-	-
Beef	6785	0.240	0.020	-	-
Pork	6785	0.232	0.019	-	-
Yucca	6784	0.526	0.014	-	-
Green bananas	6785	0.650	0.014	-	-
Chili peppers	6782	0.749	0.009	-	-
Fresh fish	6782	0.096	0.008	-	-
Potatoes	6785	0.252	0.007	-	-
Other vegetables	6784	0.604	0.006	-	-
Eggplants	6785	0.303	0.005	-	-
Squash	6785	0.399	0.005	-	-
Peas	6786	0.134	0.005	-	-
Clupea (fish)	6785	0.147	0.005	-	-
Lemons	6783	0.401	0.004	-	-
Tomatoes	6785	0.243	0.004	-	-
Chayote	6785	0.237	0.003	-	-
Cabbage	6784	0.194	0.003	-	-
Bananas	6786	0.271	0.003	-	-
Carrots	6786	0.175	0.003	-	-
Sweet potatoes	6785	0.114	0.002	-	-
Yautia	6785	0.073	0.002	-	-
Other fruits	6786	0.095	0.002	-	-
Beetroot	6785	0.064	0.001	-	-
Oranges	6786	0.115	0.001	-	-
Mangos	6786	0.055	0.001	-	-

Note: The products in each of the three product groups are listed in descending order of share of expenditure.

APPENDIX TABLE A3. VARIABLES

Variable	Description	Source
District Characteristics		
Log (total beneficiaries - 2010)	Number of beneficiaries in January 2010 at the district level	Administrative
Change in log (total beneficiaries -2009/2010)	Change in the number of beneficiaries at the district level from January 2009 to January 2010	Administrative
Log (sales -2010)	Total sales from January to May 2010 at the district level	Administrative
Change in log (sales -2009/2010)	Change in total sales from January-May 2009 to January-May 2010 at the district level	Administrative
Number of incumbent retailers 2010	Number of active retailers per district as of February 2011	Administrative
Number of brands	Average number of brands available in each distict $Brands_{s,t} = \sum_{k=1}^{K} N_{k,S,t}/15$	Retailer survey
Change in log (number of retailers 2009/2010)	$\frac{\#\ retailers\ 2010 - \#\ retailers\ 2009}{0.5 * (\#\ retailers\ 2010\ + \#\ retailers\ 2009)}$	Administrative
% Solidaridad program beneficiaries / population	Solidaridad program beneficiaries as a percentage of the total population (above 18 years)	Administrative
Average household monthly income (US$)	Average household income in the district (above 18 years)	Household survey
% of population with completed primary education	Percentage of beneficiaries with incomplete primary education or lower (above 18 years)	Household survey
% of population with incomplete secondary	Percentage of beneficiaries with incomplete secondary education	Household survey
% Population with secondary complete or higher	Percen of beneficiariies with secondary complete or higher education	Household survey
Urban	1 (if district is urban)	Administrative
District includes non-targeted neighborhoods	1 (if district includes a non-targeted neighborhood)	Administrative
Retailer Characteristics		
Average Price (weighted)	$\mathrm{Log}(P_{i0})$, where: $P_{it} = \sum_{k=1}^{K} W_k * p_{itk}$ p_{ik} Price of product k in retailer i W_k Weight computed from the household survey $W_k = \frac{w_k}{\sum_{k=1}^{K} w_k}$ K is the number of products available at the store	Household and retailer surveys
Average Price (unweighted)	$P_{it} = \sum_{k=1}^{K} W_k * p_{itk}$ $W_k = \frac{1}{K}$ K is the number of products available at the store	Retailer survey
Log (sales)	Log (self-reported sales)	Retailer survey
Log (employees)	Log (self-reported total number of employees)	Retailer survey
Share of CCT beneficiary customers	Percentage of customers who, according to the retailer, are program beneficiaries	Retailer survey
Number of customers - best day	Number of customers on the best day for sales	Retailer survey
Store cleanliness	Hygienic conditions in the store - scale of 1 to 10	Retailer survey
Retailer's gender	Gender of retailer's owner	Retailer survey
Retailer's ownership	1 (owns the retail store)	Retailer survey
Retailer's education	1 (if retailer has more than a completed primary education)	Retailer survey
Share of retailers in targeted neighborhood	1 (If retailer is in a targeted neighborhood)	Retailer survey

Consumer Characteristics			
Weighted demeaned price	$\text{Log}(P_{i0})$, where: $P_{it} = \sum_{k=1}^{K} W_k * \left[\frac{p_{itk}}{\overline{P_{kt}}}\right]$ p_{ik} Price of product k for household i (computed by dividing the amount of money spent on product i in the last week by the physical amount acquired). Units used in questions were homogenous. $\overline{p_{kt}}$ is the average price of product k at time t. W_k Weight computed from the household survey $W_k = \frac{w_k}{\sum_{k=1}^{K} w_k}$ K is the number of products reported by each beneficiary	Household survey
Unweighted demeaned price	$P_{it} = \sum_{k=1}^{K} W_k * \left[\frac{p_{itk}}{\overline{P_{kt}}}\right]$ p_{ik} Price of product k for household i (computed by dividing the amount of money spent on product i in the last week by the physical amount acquired). Units used in questions were homogenous. $\overline{p_{kt}}$ is the average price of product k at time t. W_k Weight computed from the household survey $W_k = \frac{1}{K}$	Household survey
Service quality	Quality scale (1- 10)	Household survey
Delivery	1 (retail has delivery)	Household survey
Switch to entrant retailer	1 (household change to entrant retailer between baseline and endline)	Household survey
Time shopping	Average minutes the household needs to shop	Household survey
Household head or spouse working	Head of household or spouse is working	Household survey
Head of household's gender	Head of household's gender	Household survey
Percentage of head of household married	1 (Head of household is married)	Household survey
Head of household's age	Head of household's age	Household survey
Household log-income	Household's income	Household survey

APPENDIX TABLE A4. AVERAGE CHARACTERISTICS OF ENTRANT VS INCUMBENT RETAILERS AT BASELINE

	Entrants	Incumbents	p-value of difference	Number of observations
	[1]	[2]	[3]	[4]
Log-price index - pre-treatment (weighted)	-0.343	-0.332	0.443	400
	[0.094]	[0.080]		
Log-price index - pre-treatment (unweighted)	-0.258	-0.248	0.388	400
	[0.077]	[0.082]		
1 (retailer does special sales/promotions)	0.431	0.386	0.527	401
	[0.500]	[0.487]		
Log (sales)	8.989	9.117	0.371	388
	[0.904]	[0.821]		
Log (total employees)	1.399	1.509	0.028	401
	[0.378]	[0.484]		
Percent male	0.804	0.849	0.494	401
	[0.401]	[0.359]		
1 (if the surveyed person is the retailer's owner)	0.627	0.643	0.822	401
	[0.488]	[0.480]		
1 (if has more than complete primary education)	0.686	0.623	0.318	401
	[0.469]	[0.485]		
% Solidaridad Clients	49.25	48.037	0.8424	347
	[26.570]	[23.842]		

Note: Columns [1] and [2] report the mean and standard deviation (in square brackets) of each variable for the entrant retailers and incumbent retailers at baseline. Column [3] reports the p-value of a t-test of the difference between the two samples (using clustered standard errors at the district level). Column [4] shows the number of observations used.

APPENDIX TABLE A5. SUMMARY STATISTICS OF OUTCOME MEASURES

Outcome	All districts		Targeted neighborhoods		Incumbent retailers in targeted neighborhoods	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Prices						
Price index (weighted)	0.719	0.058	0.719	0.059	0.720	0.058
Price index (weighted) - Non-barcode change	0.603	0.224	0.593	0.234	0.597	0.235
Price index (unweighted)	0.782	0.063	0.783	0.062	0.784	0.062
Price index (weighted) - Random entry=1	0.719	0.058	0.719	0.058	0.719	0.058
Price index (weighted) - Random entry=2 or 3	0.719	0.058	0.719	0.058	0.719	0.058
Price index - weighted (Consumers)	-	-	0.028	0.002	0.028	0.003
Price index (weighted) of non-CCT retailers in experimental districts	-	-	0.727	0.060	-	-
Price index (weighted) of CCT retailers in non-experimental districts	-	-	0.719	0.056	-	-
*Product Availability**						
Percent of products that changed barcode	0.481	0.124	0.484	0.126	0.484	0.126
Percent of products that changed to a cheaper barcode	0.228	0.100	0.228	0.102	0.228	0.102
Percent of products that changed brand	0.446	0.156	0.452	0.169	0.452	0.169
Percent of products that changed variety	0.018	0.064	0.012	0.050	0.012	0.050
Service quality, Clients and Spillovers						
Store cleanliness	7.474	2.023	7.417	1.992	7.362	2.011
Time shopping (minutes)	-	-	61.0	47.6	62.7	48.9
Delivery	-	-	0.454	0.498	0.464	0.499
Service-quality rating	-	-	8.982	1.467	8.987	1.449
Number of customers on best day	106.0	117.9	113.2	124.9	113.8	123.8
Share of customers CCT beneficiaries	48.1	24.0	47.9	24.2	47.6	24.2
Switch to entrant retailer*	-	-	0.041	0.199	-	-

Note: For most variables summary statistics are measured at baseline and correspond to the samples located in treated and control districts. In the case of variables related to product availability (marked with *), which can only be defined at endline, we report the summary statistics for the control group.

APPENDIX TABLE A6. DIFFERENCES IN NON-COMPLIERS AND COMPLIERS

	Compliers	Non-compliers	p-value of difference	Number of obs.
	[1]	[2]	[3]	[4]
A. District characteristics				
Log (total beneficiaries - 2010)	6.417	6.556	0.584	72
	[0.977]	[0.626]		
Change in log (total beneficiaries - 2009/2010)	0.168	0.232	0.184	72
	[0.146]	[0.236]		
Log (sales - 2010)	11.285	11.295	0.978	69
	[1.229]	[1.337]		
Change in log (sales -2009/2010)	0.971	1.774	0.327	67
	[2.577]	[3.703]		
Number of incumbent retailers - 2010	5.945	6.294	0.852	72
	[6.753]	[6.469]		
Change in log (number of retailers - 2009/2010)	0.402	0.579	0.316	72
	[0.596]	[0.744]		
% Solidaridad program beneficiaries / population	0.393	0.290	0.099	72
	[0.234]	[0.183]		
Average monthly household income (US$)	491.088	495.965	0.842	72
	[86.411]	[93.582]		
% of population with completed primary education or lo	0.618	0.615	0.887	72
	[0.080]	[0.067]		
% of population with incomplete secondary education	0.208	0.215	0.615	72
	[0.052]	[0.048]		
% of population with completed secondary education or	0.174	0.170	0.809	72
	[0.065]	[0.045]		
1 (if district is urban)	0.745	0.882	0.216	72
	[0.413]	[0.332]		
District includes non-targeted neighborhoods	0.400	0.176	0.093	72
	[0.494]	[0.393]		
B. Retailer characteristics				
Log-price index - pre-treatment (weighted)	-0.336	-0.324	0.235	400
	[0.084]	[0.074]		
Percentage male	0.837	0.864	0.543	401
	[0.370]	[0.345]		
1 (if the surveyed person is the retailer's owner)	0.642	0.636	0.897	401
	[0.480]	[0.484]		
1 (if has more than a completed primary education)	0.623	0.659	0.496	401
	[0.485]	[0.477]		
Log (total employees)	1.486	1.529	0.397	401
	[0.461]	[0.513]		
Log (sales)	9.083	9.164	0.429	388
	[0.822]	[0.865]		
Share of retailers in targeted neighborhood	0.601	0.784	0.117	401
	[0.491]	[0.414]		

Note: Columns [1] and [2] report the mean and standard deviation (in square brackets) of each variable at both the district and retailer level for compliers and non-compliers. Column [3] reports the p-value of a t-test of the difference between the two samples (using clustered standard errors at the district level). Column [4] shows the number of observations used.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Source: Author's calculations

APPENDIX TABLE A7. IMPACT ON INDIVIDUAL PRODUCT PRICES (ROBUSTNESS)

Outcome Log(Product Price)	Weighting	All		Targeted		Incumbents	
		ITT	IV	ITT	IV	ITT	IV
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Rice (lb.)	0.293	-0.008	-0.015	-0.010	-0.022	-0.009	-0.022
		[0.013]	[0.023]	[0.017]	[0.033]	[0.018]	[0.037]
Cooking oil (lb.)	0.094	-0.030**	-0.057	-0.050***	-0.110**	-0.052***	-0.129**
		[0.015]	[0.038]	[0.015]	[0.046]	[0.016]	[0.059]
Sugar (lb.)	0.052	-0.001	-0.002	-0.003	-0.007	-0.008	-0.019
		[0.011]	[0.020]	[0.009]	[0.020]	[0.010]	[0.023]
Pasta (lb.)	0.017	-0.027**	-0.051**	-0.048***	-0.102**	-0.048**	-0.113*
		[0.013]	[0.024]	[0.015]	[0.047]	[0.016]	[0.058]
Eggs (unit)	0.050	-0.022	-0.042	-0.025	-0.055	-0.025	-0.059
		[0.026]	[0.044]	[0.023]	[0.046]	[0.022]	[0.052]
Powdered milk (125 gr.)	0.062	0.032	0.060	0.019	0.040	0.006	0.015
		[0.023]	[0.042]	[0.025]	[0.054]	[0.022]	[0.053]
Chocolate (unit)	0.015	0.002	0.004	-0.008	-0.017	-0.009	-0.022
		[0.011]	[0.021]	[0.014]	[0.028]	[0.014]	[0.030]
Sardines (unit)	0.014	0.028	0.053	0.015	0.032	0.017	0.040
		[0.032]	[0.062]	[0.044]	[0.097]	[0.042]	[0.100]
Green beans (lb.)	0.063	-0.005	-0.009	-0.005	-0.011	-0.003	-0.007
		[0.006]	[0.011]	[0.008]	[0.017]	[0.008]	[0.020]
Onions (lb.)	0.020	-0.013	-0.024	-0.047**	-0.104*	-0.038*	-0.092
		[0.022]	[0.044]	[0.022]	[0.062]	[0.022]	[0.066]
Salami (lb.)	0.048	-0.051*	-0.096*	-0.060	-0.132	-0.046	-0.111
		[0.028]	[0.054]	[0.039]	[0.091]	[0.040]	[0.099]
Chicken (lb.)	0.170	-0.014	-0.023	-0.008	-0.016	-0.008	-0.016
		[0.009]	[0.017]	[0.014]	[0.025]	[0.014]	[0.028]
Cod (lb.)	0.018	-0.010	-0.019	-0.020**	-0.045	-0.023**	-0.057
		[0.009]	[0.016]	[0.010]	[0.031]	[0.010]	[0.039]
Flour (lb.)	0.010	-0.038**	-0.066**	-0.042**	-0.086*	-0.040*	-0.092
		[0.015]	[0.031]	[0.020]	[0.051]	[0.021]	[0.060]
Bread (unit)	0.074	0.088*	0.151	0.021	0.039	0.035	0.073
		[0.052]	[0.107]	[0.057]	[0.107]	[0.057]	[0.121]

Note: Each entry shows an estimate of the impact of an increase in competition on the price of different products. Column [1] shows the weighting of each product in the final retailer price. Columns [2]-[3] use all the retailers; columns [4]-[5] use retailers in targeted neighborhoods; and columns [6]-[7] use incumbent retailers in targeted neigborhoods. All columns report the estimations while controlling for the baseline log(price).

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.