Online Appendix: Not for Publication

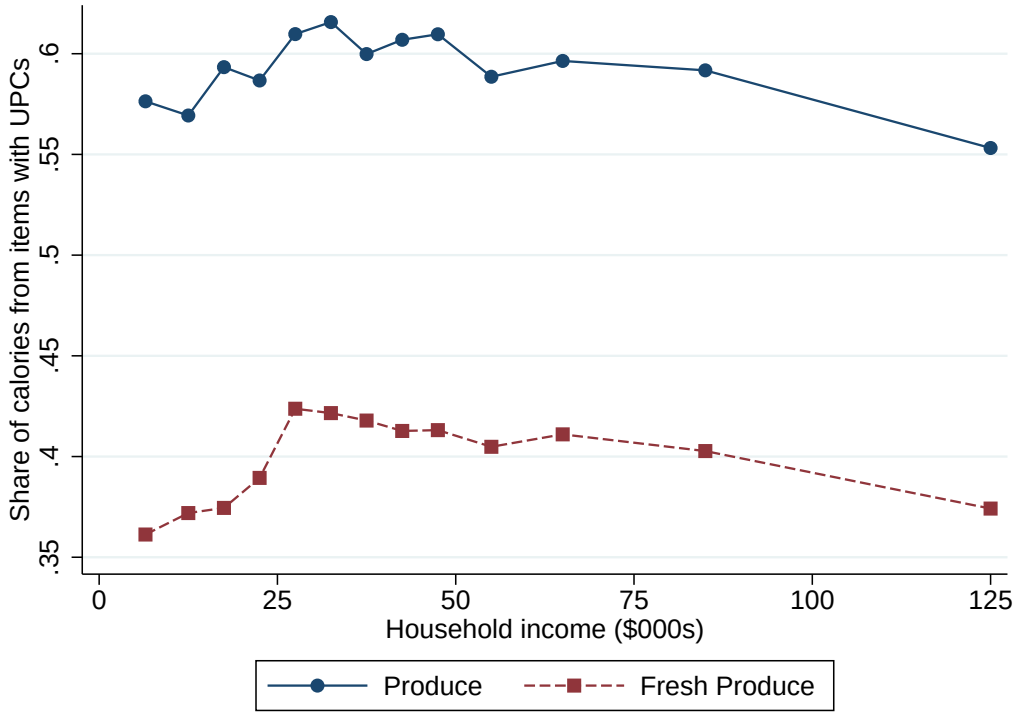
The Geography of Poverty and Nutrition: Food Deserts and Food Choices Across the United States

Hunt Allcott, Rebecca Diamond, and Jean-Pierre Dubé

A Appendix to Data Section

A.A Magnet Calorie Shares

Figure A1: **Magnet Data: Share of Produce from Packaged Items**



Notes: This figure uses the Nielsen Homescan "magnet" subsample for 2004-2006 to show the share of produce and fresh produce calories coming from items with UPCs, which are the items that we observe outside the Magnet subsample. "Produce" includes fresh, dried, canned, and frozen produce. Observations are weighted for national representativeness.

A.B Health Index

Table A1: **Health Index Function**

Nutrient	Recommendation	Recommended Daily Intake (grams)	Explanation
		Fruits and vegetables	
Fruits	Increase	320	Two cups/day (Food Patterns); 160 g/cup
Vegetables	Increase	390	Three cups/day (Food Patterns); 130 g/cup
		All other items	
Protein	Increase	51	51 grams/day (DRI)
Fiber	Increase	29.5	29.5 grams/day (DRI)
Sugar	Reduce	32.8	45% of 282 calories/day from sugar+sat. fat (Food Patterns)
Saturated fat	Reduce	17.2	55% of 282 calories/day from sugar+sat. fat (Food Patterns)
Sodium	Reduce	2.3	2300 mg/day (Dietary Guidelines)
Cholesterol	Reduce	0.3	300 mg/day (Dietary Guidelines)

Notes: Our "raw Health Index" for product n is the sum of healthy minus unhealthy nutrient contents per 1000 calories, weighting each by its recommended daily intake (RDI): $\tilde{H}_n = \sum_c G_c a_{nc}/r_c$, where a_{nc} is the grams of nutrient c per 1000 calories, r_c is the RDI for a normal adult, and G_c takes value 1 for "healthy" macronutrients to "increase" and -1 for "unhealthy" nutrients to "reduce." This table presents the increase/reduce recommendation G_c and RDI r_c used to construct the raw Health Index.

Table A2: **Correlations Between Health Index and Its Components in Homescan**

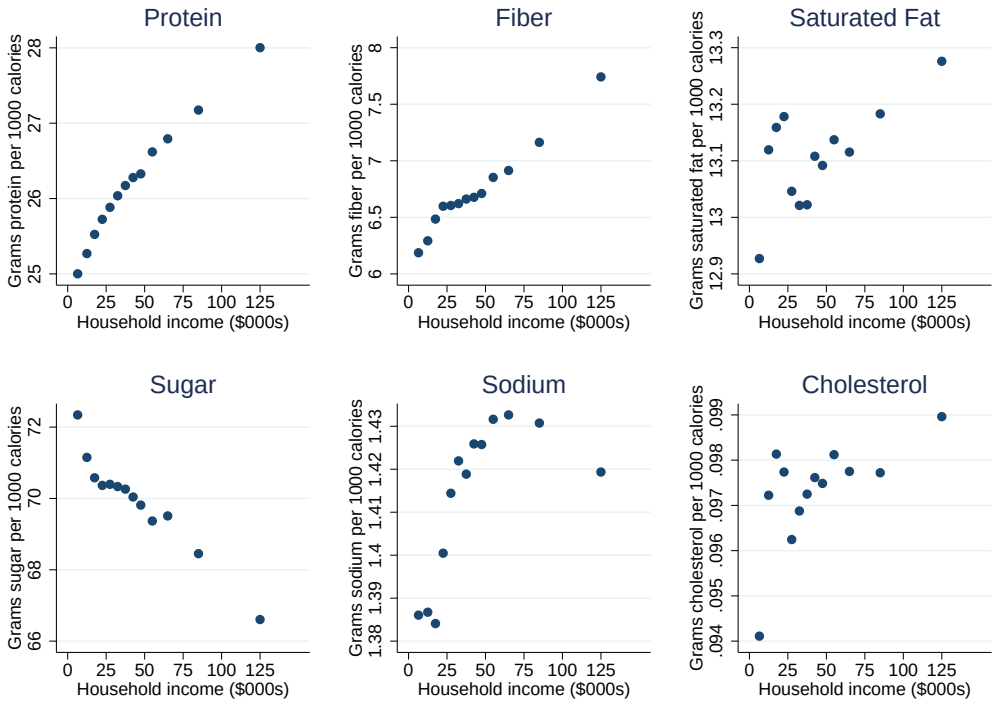
Attribute	Correlation with Health Index
Fruits and vegetables	
Fruits	0.33
Vegetables	0.34
All other items	
Protein	0.52
Fiber	0.64
Sugar	-0.75
Saturated fat	-0.05
Sodium	-0.19
Cholesterol	-0.008

Notes: Using Homescan household-by-year data for 2004-2015, this table presents the correlation coefficients between Health Index and its components, using data in units of grams per 1000 calories consumed. Observations are weighted for national representativeness.

B Appendix to Stylized Facts Section

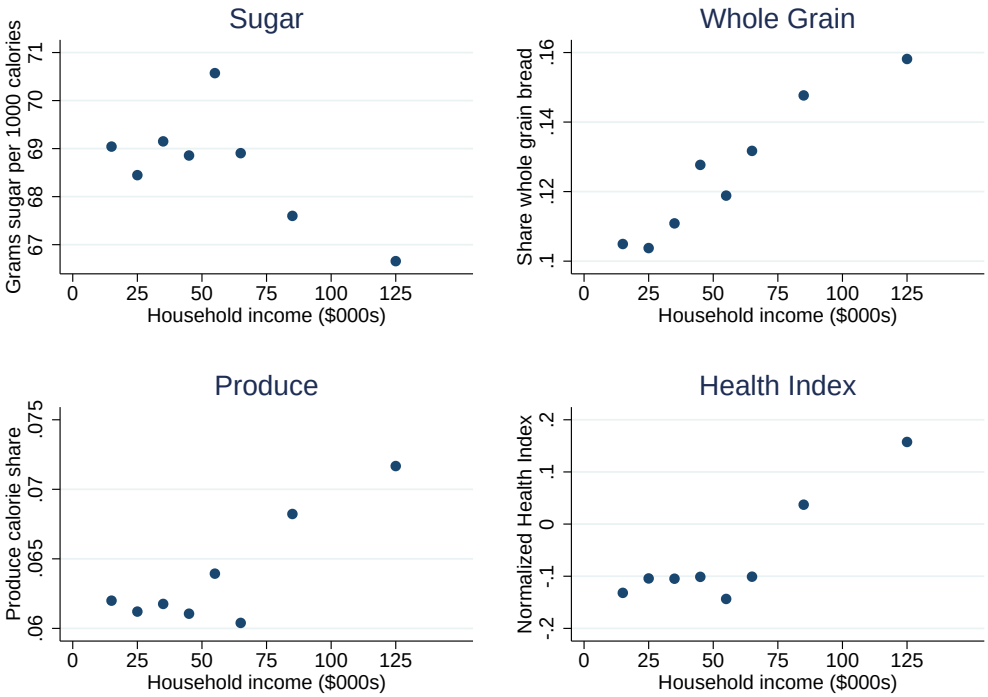
B.A Additional Figures and Tables

Figure A2: **Macronutrient Purchases by Household Income**



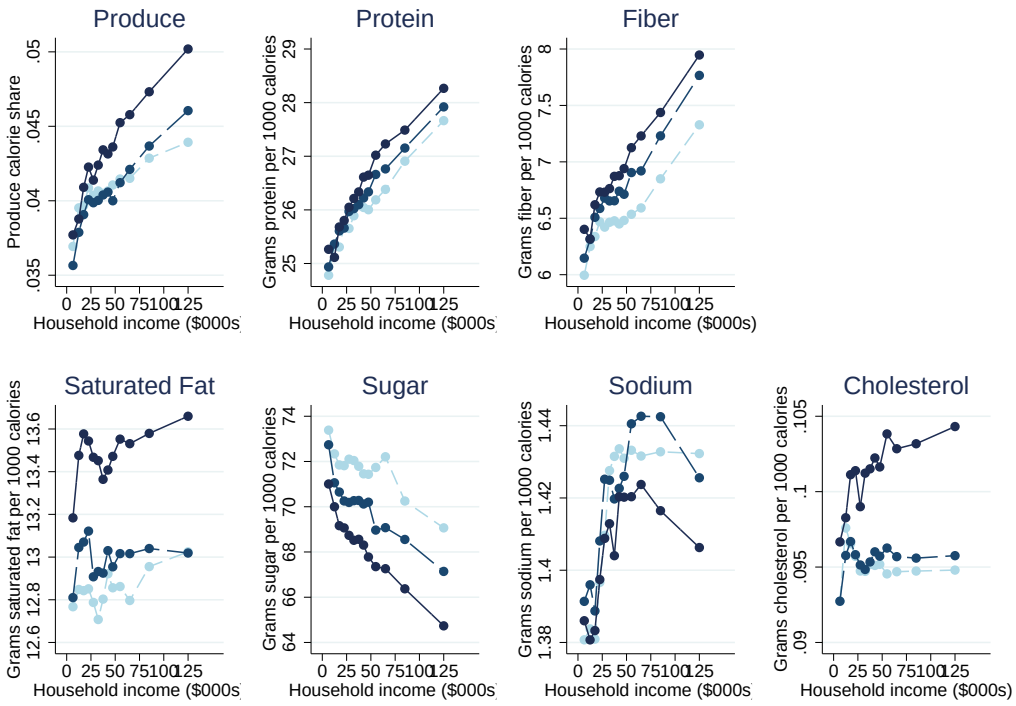
Notes: This figure presents calorie-weighted average macronutrient contents of purchases using Nielsen Homescan data for 2004-2015. The x-axis presents nominal income bins; household incomes larger than $100,000 are coded as $125,000. Observations are weighted for national representativeness.

Figure A3: **Magnet Subsample: Healthful Purchases by Household Income**



Notes: Nielsen Homescan data, magnet subsample, for 2004-2006. The x-axis presents nominal income bins; household incomes larger than $100,000 are coded as $125,000. This parallels Figure 2, except using the magnet subsample which also records purchases of non-UPC items such as bulk produce. Sugar is the grams of sugar per 1000 calories purchased, whole grain is the calorie-weighted average share of bread, buns, and rolls purchases that are whole grain, produce is the share of calories from fresh, canned, dried, and frozen fruits and vegetables, and Health Index is our overall measure of the healthfulness of grocery purchases, normalized to mean zero, standard deviation one across households. Observations are weighted for national representativeness.

Figure A4: **Trends in Macronutrient Purchases by Household Income**



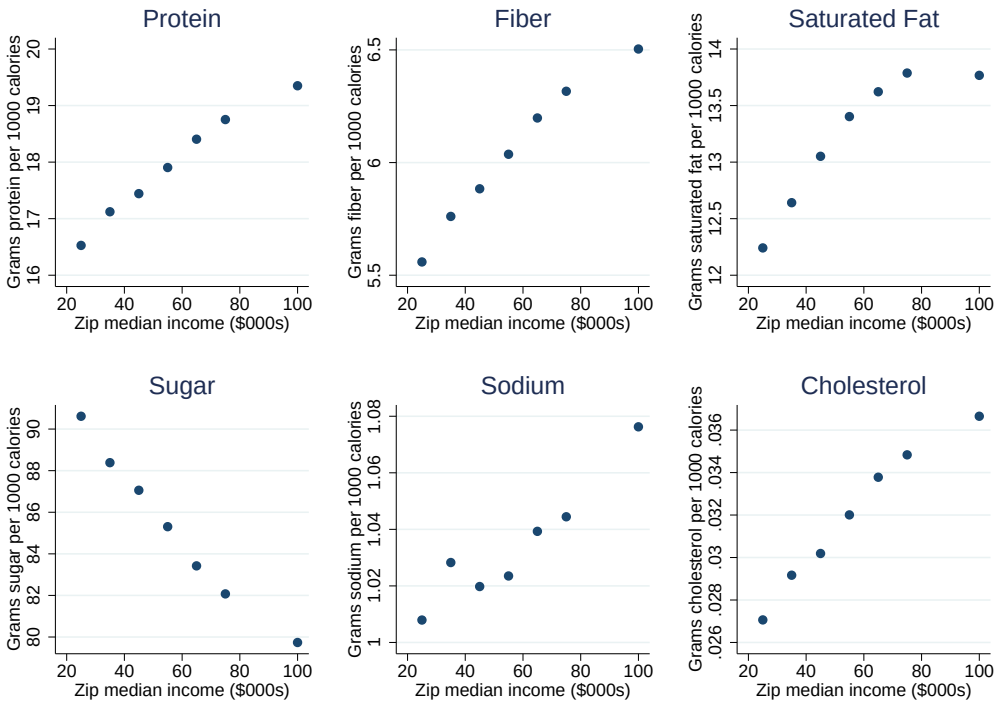
Notes: This figure presents calorie-weighted average macronutrient contents of purchases using Nielsen Homescan data for 2004-2015. The x-axis presents nominal income bins; household incomes larger than $100,000 are coded as $125,000. On each plot, the three lines plot 2004-2007, 2008-2011, and 2012-2015 averages, respectively, in light, medium, and dark lines. Observations are weighted for national representativeness.

Table A3: **Pooled OLS versus Within-Household Income Variation**

	(1)	(2)	(3)	(4)
Sample:	Households Observed in Two or More Years			Exclude Income>$100k
ln(Household income)	0.130	0.0198	0.0160	0.0134
	(0.00506)***	(0.00484)***	(0.00492)***	(0.00580)**
Household-by-Census tract fixed effects	No	Yes	Yes	Yes
Household demographics	No	No	Yes	Yes
Observations	603,230	603,230	603,230	516,170
Income coefficient/column 1 coefficient	1	0.15	0.12	0.10

Notes: This table presents regressions of Health Index on natural log of household income and year indicators using Nielsen Homescan data for 2004-2015. Columns 2-4 also include household-by-Census tract fixed effects, and columns 3 and 4 also include household demographics (natural log of years of education, age indicators, an indicator for whether the household includes children, race indicators, employment status, weekly work hours, and total calorie need). The sample is restricted to households observed in two or more years; column 4 additionally excludes observations with household income above $100,000. Health Index is our overall measure of the healthfulness of grocery purchases, normalized to mean zero, standard deviation one across households. Observations are weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

Figure A5: **Store Average Healthfulness by Zip Code Median Income**

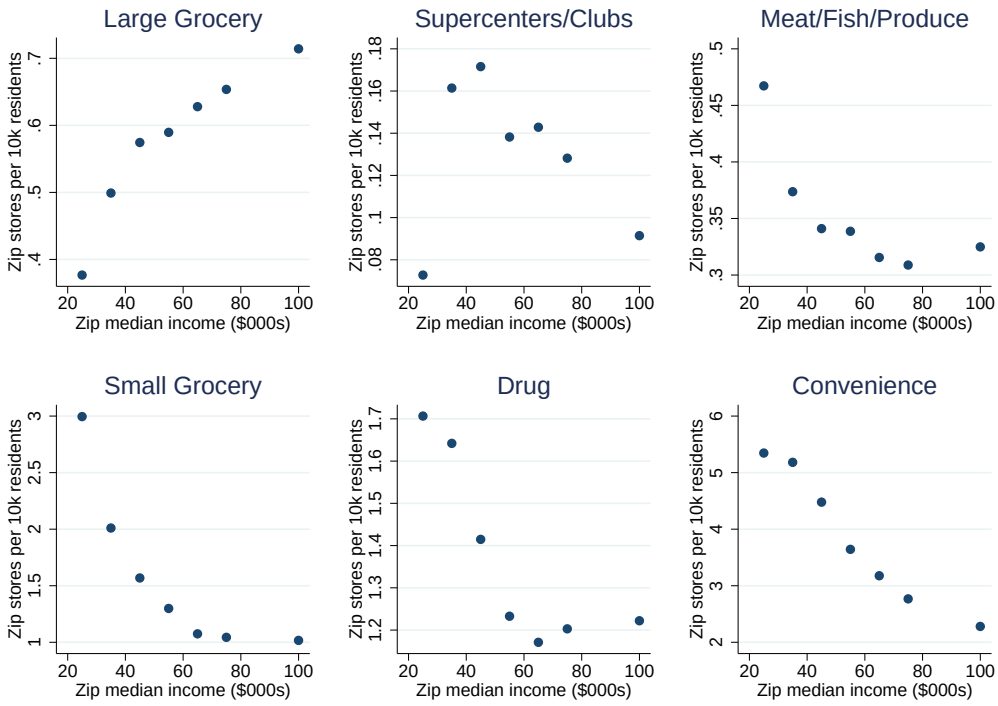


Notes: Using Nielsen RMS data for year 2012, we constructed calorie-weighted mean macronutrient content across all UPCs offered in each store. This figure presents the means of these variables within categories of zip code median income. This parallels Figure 2 in the text.

B.B Low-income neighborhoods have relatively more unhealthful store types

Using the Zip Code Business Patterns data for 2004-2015, Appendix Figure A6 plots the average count of stores by channel type for zip codes by median income category. Zip codes vary substantially in area and population, so this figure normalizes store counts per 10,000 residents; the mean zip code has 12,000 residents. Lower-income zip codes have more stores per capita of all channel types, with two exceptions. First, the concentration of large grocery stores per capita is sharply monotonically increasing in median income, consistent with Powell et al. (2007). Second, the concentration of supercenters and club stores takes an inverted-U shape, with many fewer per capita in the very lowest-income zip codes.

Figure A6: **Store Counts by Zip Code Median Income**

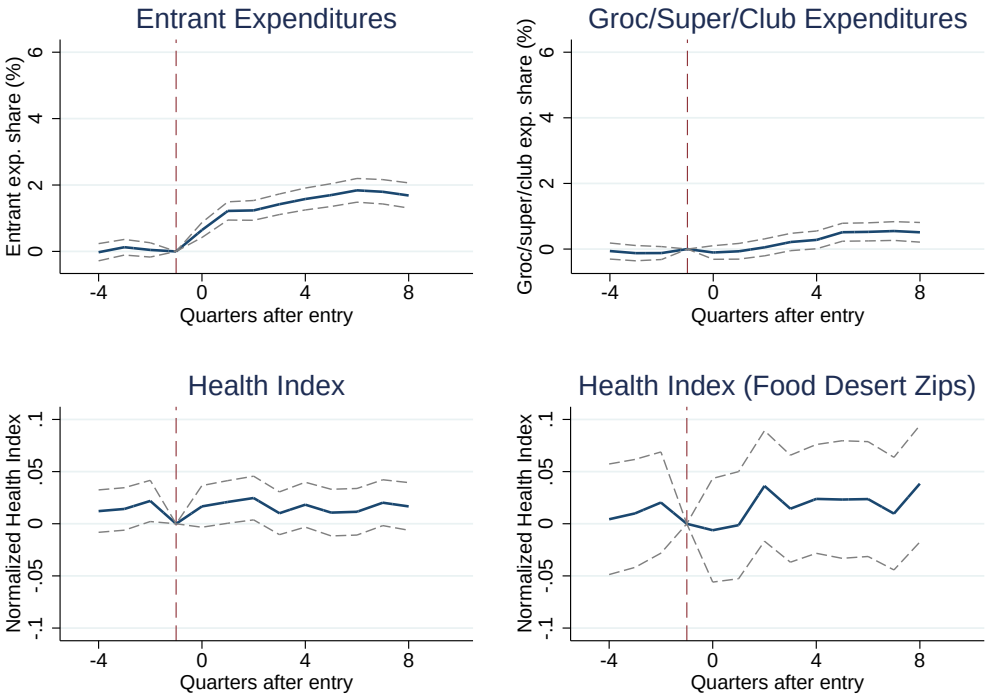


Notes: This figure presents mean store counts per 10,000 residents by zip code income category using data from Zip Code Business Patterns, averaged over 2004-2015. Large (small) grocers are defined as those with 50 or more (fewer than 50) employees.

C Appendix to Reduced-Form Event Studies

C.A Additional Figures and Tables for Entry Event Study

Figure A7: **Event Study of Supermarket Entry Between 10 and 15 Minutes from Home**



Notes: This figure presents the $\tau_{[10,15)q}$ parameters and 95 percent confidence intervals from estimates of Equation (3): the effects of entry by several large supermarket chains, using 2004-2015 household-by-quarter Homescan data. All regressions control for household demographics (natural log of income, natural log of years of education, age indicators, an indicator for whether the household includes children, race indicators, employment status, weekly work hours, and total calorie need), Census division-by-quarter of sample indicators, and household-by-Census tract fixed effects. The top two panels present effects on expenditure shares, in units of percentage points. The bottom two panels present effects on Health Index, our overall measure of the healthfulness of grocery purchases, normalized to mean zero, standard deviation one across households. The dashed vertical line is the last quarter before entry. Observations are not weighted for national representativeness.

Table A4: **Effects of Supermarket Entry**

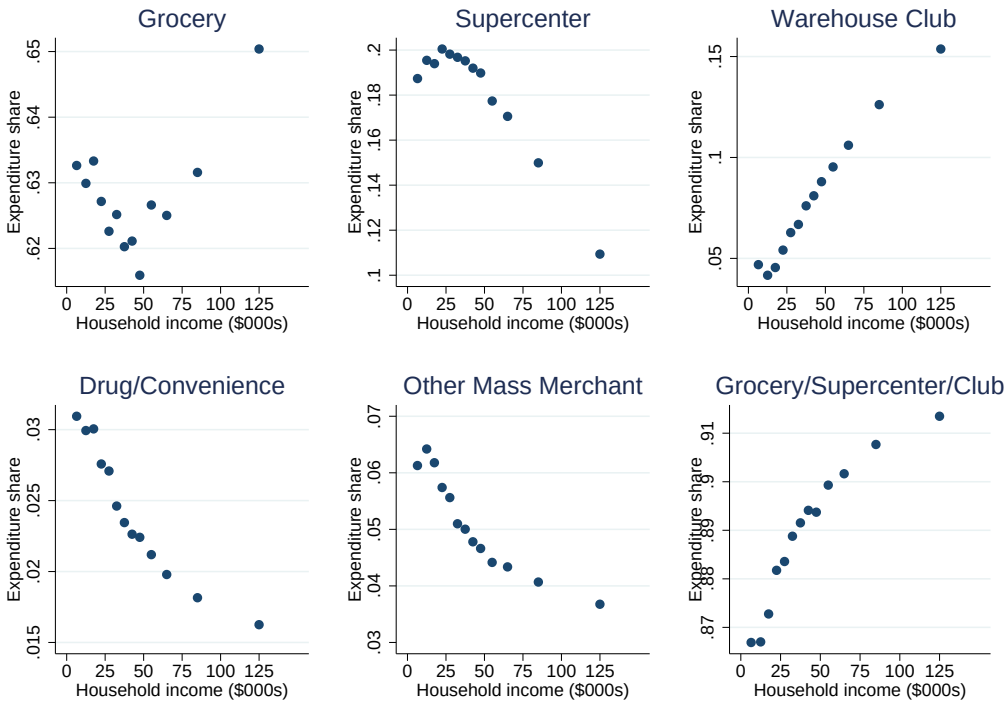
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Effects on Expenditure Shares at Other Store Types						
Sample:	Full Sample		Income < $25,000		"Food Desert" Zip Codes	
Expenditure shares at store type:	Conv/ Drug Stores	Other Mass Merchants	Conv/ Drug Stores	Other Mass Merchants	Conv/ Drug Stores	Other Mass Merchants
Post entry: 0-10 minutes	-0.0888	-0.515	-0.265	-0.856	-0.106	-0.598
	(0.0352)**	(0.0642)***	(0.125)**	(0.223)***	(0.129)	(0.164)***
Post entry: 10-15 minutes	-0.0541	-0.113	-0.120	-0.0331	-0.116	-0.233
	(0.0254)**	(0.0434)***	(0.0904)	(0.143)	(0.0782)	(0.107)**
Observations	2,627,947	2,627,947	404,868	404,868	640,498	640,498
Dependent var. mean	2.6	5.3	3.5	7.1	2.4	5.6

Panel B: Effects on Health Index Using Alternative Food Desert Definitions

	(1)	(2)	(3)
Sample:	< 1000 Produce UPCs	No Medium Groceries	Three-Mile Radius
Post entry: 0-10 minutes	-0.00968	-0.0129	0.00564
	(0.0182)	(0.0192)	(0.0201)
Post entry: 10-15 minutes	0.0338	0.0304	0.0463
	(0.0109)***	(0.0115)***	(0.0133)***
Observations	408,160	380,868	487,646

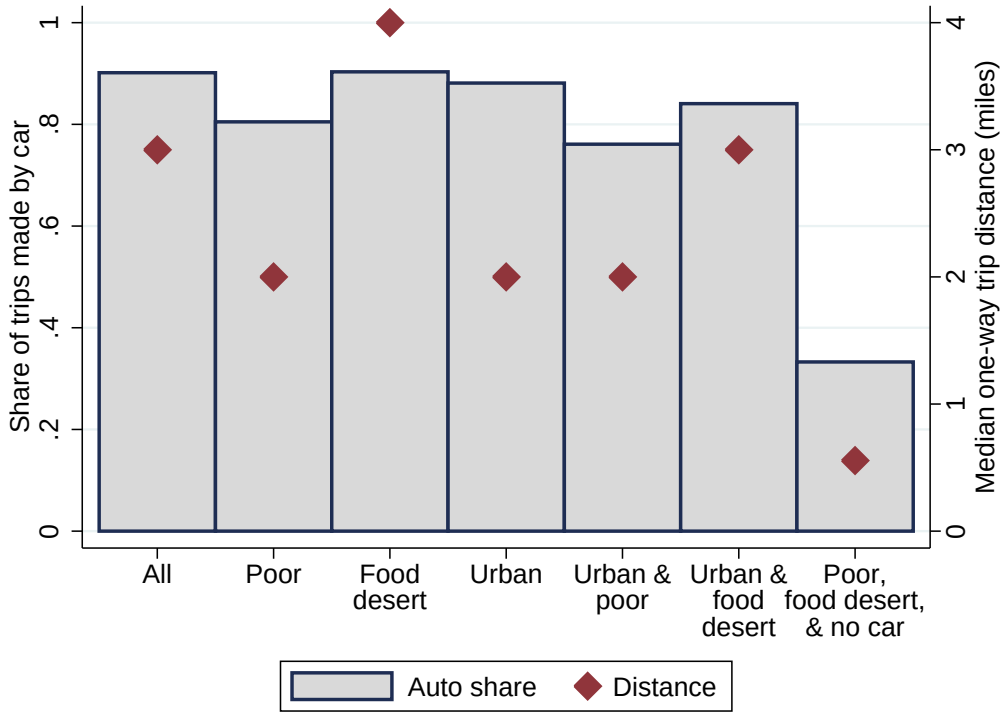
Notes: This table uses 2004-2015 Nielsen Homescan data at the household-by-quarter level. The table parallels Table 3, except Panel A presents effects on expenditure shares at alternative channel types, and Panel B uses alternative definitions of a "food desert." In Panel B, columns 1 and 2 limit the sample to zip codes with fewer than 1000 produce UPCs available in 2003, as predicted by applying RMS data from Table 2 to Zip Code Business Patterns data; columns 3 and 4 also exclude any zip codes with grocery stores employing between 10 and 49 employees in 2003; columns 5 and 6 define a zip code as a food desert only if all zip codes with centroids within three miles have no grocery stores with 50 or more employees, supercenters, or club stores in 2003. Expenditure shares are in units of percentage points. Health Index is our overall measure of the healthfulness of grocery purchases, normalized to mean zero, standard deviation one across households. Reported independent variables are indicators for whether a specific retailer has entered within a 0-10 or 10-15 minute drive from the household's Census tract centroid. All regressions control for household demographics (natural log of income, natural log of years of education, age indicators, an indicator for whether the household includes children, race indicators, employment status, weekly work hours, and total calorie need), Census division-by-quarter of sample indicators, and household-by-Census tract fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

Figure A8: **Channel Type Expenditure Shares by Household Income**



Notes: This figure uses Nielsen Homescan data for 2004-2015. The x-axis presents nominal income bins; household incomes larger than $100,000 are coded as $125,000. Another 5-6 percent of expenditures are at channels not plotted, including bakeries, butchers, candy stores, liquor stores, fruit stands, and fish markets; this proportion is fairly constant by income. Observations are weighted for national representativeness.

Figure A9: **Median Shopping Trip Distances by Household Income**



Notes: Data are from the 2009 National Household Travel Survey. Diamonds represent the median one-way trip distance for trips beginning or ending in "buying goods: groceries/clothing/hardware store." "Poor" means household income less than $25,000. "Food desert" means that the household is in a zip code with no grocery stores with 50 or more employees, supercenters, or club stores. "Urban" includes urbanized areas or urban clusters of at least 2500 people, using the U.S. Census Bureau definition. "No car" means that the household does not own a car.

C.B Entry by All Retailers Using Zip Code Business Patterns

To complement the event study estimates in the body of the paper, we present alternative specifications that measure entry using the Zip Code Business Patterns (ZBP) data.

Panel A of Appendix Table A5 shows that the Zip Code Business Patterns data date openings of specific supercenters in the correct year 50 to 80 percent of the time, although they are sometimes recorded a year later and sometimes in a broader "general merchandise" NAICS code (452) instead of the specific "supercenter and club store" NAICS code (452910).

The entry event study regression is analogous to Equation (2). Define S_{zt} and G_{zt}, respectively, as the count of supercenters/club stores and large (at least 50 employee) grocery stores in zip code z in year t. Using household-by-year data and now denoting μ_{dt} as Census division-by-year indicators, the regression is:

$$Y_{izt} = \tau_S S_{zt} + \tau_G G_{zt} + \beta \boldsymbol{X}_{it} + \mu_{dt} + \phi_{ic} + \varepsilon_{izt} \tag{34}$$

Standard errors are again clustered by household, and observations are again all weighted equally.

Appendix Table A6 presents results. The structure is similar to that of Table 3: Panel A presents effects on expenditure shares, while Panel B presents effects on healthful eating.

Columns 1-3 present estimates for the full sample. Columns 1 and 2 confirm that the ZBP data contain meaningful information. Column 1 shows that conditional on household fixed effects, a larger count of large grocery stores and/or a smaller count of supercenters and club stores in the zip code are both strongly positively associated with higher expenditure share at chain groceries. Column 2 shows the opposite: fewer grocery stores and more supercenters are strongly positively associated with higher expenditures at supercenters and club stores. Column 3 presents effects on combined expenditure shares for all grocery stores, supercenters, and club stores. Columns 4-6 and 7-9 present estimates for the low-income and food desert subsamples. As in Table 3, effects of entry on expenditures generally larger in food deserts. Also as in Table 3, we see that entry by a large grocery retailer substantially diverts sales from other supermarkets, so the effects on combined expenditures at grocery stores, supercenters, and club stores are limited. Appendix Table A7 shows that most of this diversion is from other mass merchants; there is no statistically significant diversion from drug and convenience stores.

The bottom panel shows no statistically significant effect of the number of large grocers and supercenters/clubs on Health Index. With 95 percent confidence, we can bound the effects on low-income households' Health Index at less than 0.011 standard deviations per large grocery store and 0.052 standard deviations per supercenter or club store. Appendix Table A7 shows that under all alternative definitions of "food deserts," the number of local large grocers, supermarkets, and club stores has no statistically or economically significant effect on Health Index.

One reason to prefer the earlier regressions with specific known retailers is that we have high

confidence that entry dates are correctly measured. We can also imagine using the true supercenter entry dates as an instrument for ZBP data, which are measured with error. Panel B of Appendix Table A5 shows that the "first stages" of such a regression have coefficients around 0.9 and 0.66 for two different supercenter chains. If the average retailer in ZBP is measured with equal or perhaps somewhat more error than the less well-measured supercenter chain, this suggests that our bounds in the paragraph above should be increased by 50 to 100 percent due to measurement error. Even after this adjustment, however, our results in Tables 3 and A6 suggest that having a supermarket nearby explains at most only a small share of the differences in nutritional decisions between low- and high-income households.

Table A5: **Zip Code Business Patterns Accuracy Check with Known Entry Dates**

	(1)	(2)
Dependent variable is count of channel type:	Supercenter	General Merchandise
Panel A: Difference Estimator		
Supercenter Chain 1:		
2-year lead	0.00478	0.00629
	(0.00622)	(0.0166)
1-year lead	0.0375	0.0577
	(0.00978)***	(0.0164)***
Entry year	0.562	0.821
	(0.0199)***	(0.0210)***
1-year lag	0.208	0.0821
	(0.0169)***	(0.0191)***
2-year lag	0.0777	0.0133
	(0.0127)***	(0.0145)
Supercenter Chain 2:		
2-year lead	0.0158	-0.0172
	(0.0298)	(0.0421)
1-year lead	0.0133	-0.0451
	(0.0222)	(0.0462)
Entry year	0.0621	0.480
	(0.0327)*	(0.0623)***
1-year lag	0.172	0.0701
	(0.0413)***	(0.0594)
2-year lag	0.133	0.0918
	(0.0514)***	(0.0599)
Observations	264,734	264,734
Panel B: Fixed Effects Estimator		
Post entry: chain 1	0.902	0.932
	(0.0138)***	(0.0227)***
Post entry: chain 2	0.667	0.665
	(0.0365)***	(0.0659)***
Observations	297,966	297,966

Notes: Data are at the zip code-by-year level. All regressions include year indicators; fixed effects regressions have zip code fixed effects. Robust standard errors, clustered by zip code, are in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

Table A6: **Effects of Supermarket Entry Using Zip Code Business Patterns**

Panel A: Effects on Expenditure Shares

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sample:		Full Sample			Income < $25,000			"Food Desert" Zip Codes	
Expend. shares at store type:	Chain Grocers	Super/ Club	Grocery/ Super/Club	Chain Grocers	Super/ Club	Grocery/ Super/Club	Chain Grocers	Super/ Club	Grocery/ Super/Club
Large grocers	0.329	-0.382	-0.0259	0.256	-0.430	-0.0772	0.522	-0.705	-0.129
	(0.0718)***	(0.0615)***	(0.0414)	(0.207)	(0.173)**	(0.136)	(0.299)*	(0.242)***	(0.167)
Supers/clubs	-1.896	2.903	0.692	-2.023	3.553	1.057	-2.472	3.452	0.629
	(0.172)***	(0.160)***	(0.0939)***	(0.480)***	(0.451)***	(0.311)***	(0.591)***	(0.570)***	(0.291)**
Observations	664,302	664,302	664,302	102,462	102,462	102,462	163,747	163,747	163,747
Dep. var. mean	58	26	88	56	23	86	53	28	88

Panel B: Effects on Health Index

	(1)	(2)	(3)
Sample:	Full Sample	Income < $25,000	"Food Desert" Zip Codes
Large grocers	0.00102	-0.00303	-0.00518
	(0.00254)	(0.00738)	(0.0104)
Supers/clubs	0.00667	0.0217	0.0184
	(0.00551)	(0.0155)	(0.0168)
Observations	664,302	102,462	163,747

Notes: This table uses 2004-2015 Nielsen Homescan data at the household-by-year level. "Food Desert" zip codes are those with no grocery stores with 50 or more employees, supercenters, or club stores in 2003. Expenditure shares are in units of percentage points. Health Index is our overall measure of the healthfulness of grocery purchases, normalized to mean zero, standard deviation one across households. Reported independent variables are the count of stores by channel type in the household's zip code. All regressions control for household demographics (natural log of income, natural log of years of education, age indicators, an indicator for whether the household includes children, race indicators, employment status, weekly work hours, and total calorie need), Census division-by-year indicators, and household-by-Census tract fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

Table A7: **Effects of Supermarket Entry Using Zip Code Business Patterns**

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Effects on Expenditure Shares at Other Store Types						
Sample:	Full Sample		Income < $25,000		"Food Desert" Zip Codes	
Expenditure shares at store type:	Conv./ Drug Stores	Other Mass Merchants	Conv./ Drug Stores	Other Mass Merchants	Conv./ Drug Stores	Other Mass Merchants
Large grocers	-0.0145	0.0737	-0.0352	0.168	-0.00259	0.116
	(0.0173)	(0.0278)***	(0.0613)	(0.0953)*	(0.0643)	(0.106)
Supercenters/clubs	-0.0216	-0.717	-0.172	-1.079	0.142	-0.507
	(0.0348)	(0.0641)***	(0.130)	(0.213)***	(0.116)	(0.187)***
Observations	664,302	664,302	102,462	102,462	163,747	163,747
Dependent var. mean	2.6	5.2	3.5	7.0	2.4	5.6

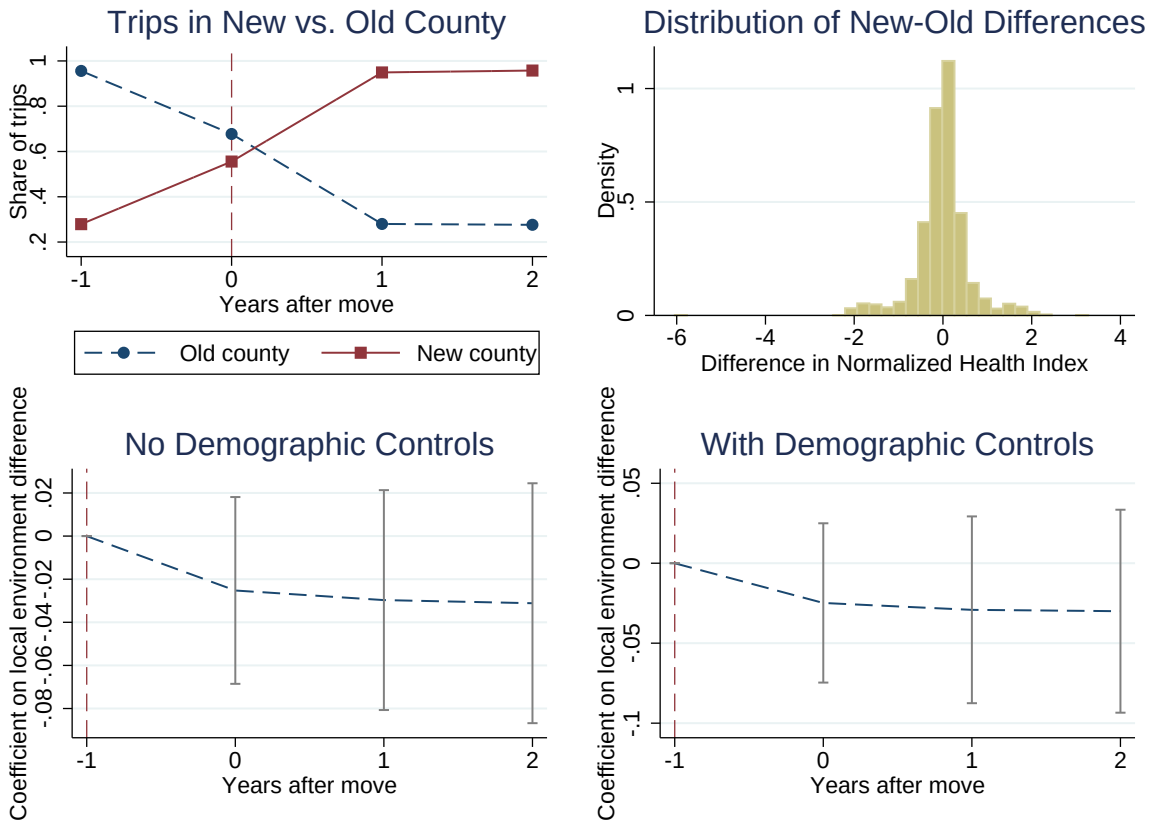
Panel B: Effects on Health Index Using Alternative Food Desert Definitions

	(1)	(2)	(3)
Sample:	< 1000 Produce UPCs	No Medium Groceries	Three-Mile Radius
Large grocers	-0.00617	-0.00660	-0.0144
	(0.0148)	(0.0161)	(0.0129)
Supers/clubs	0.0158	0.0133	0.0192
	(0.0200)	(0.0204)	(0.0211)
Observations	104,451	98,256	125,399

Notes: This table uses 2004-2015 Nielsen Homescan data at the household-by-year level. The table parallels Table A6, except Panel A presents effects on expenditure shares at alternative channel types, and Panel B uses alternative definitions of a "food desert." In Panel B, columns 1 and 2 limit the sample to zip codes with fewer than 1000 produce UPCs available in 2003, as predicted by applying RMS data from Table 2 to Zip Code Business Patterns data; columns 3 and 4 also exclude any zip codes with grocery stores employing between 10 and 49 employees in 2003; columns 5 and 6 define a zip code as a food desert only if all zip codes with centroids within three miles have no grocery stores with 50 or more employees, supercenters, or club stores in 2003. Expenditure shares are in units of percentage points. Health Index is our overall measure of the healthfulness of grocery purchases, normalized to mean zero, standard deviation one across households. Reported independent variables are the count of stores by channel type in the household's zip code. All regressions control for household demographics (natural log of income, natural log of years of education, age indicators, an indicator for whether the household includes children, race indicators, employment status, weekly work hours, and total calorie need), Census division-by-year indicators, and household-by-Census tract fixed effects. Observations are not weighted for national representativeness. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

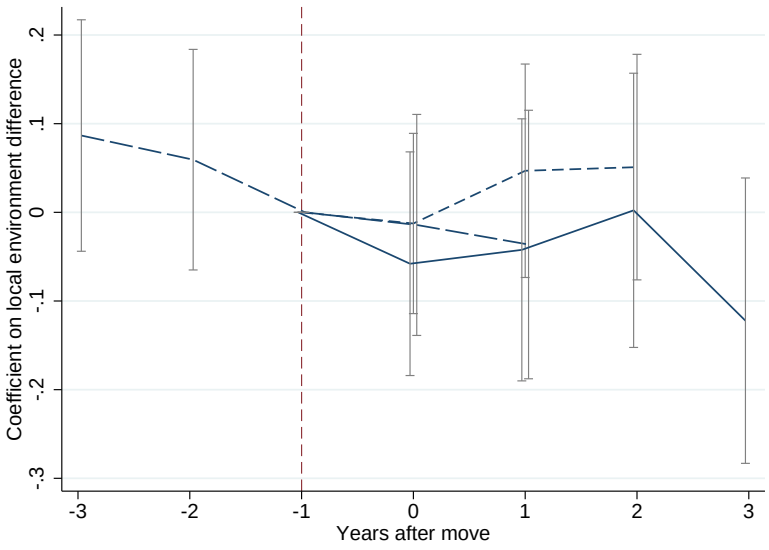
C.C Appendix to Movers Event Study

Figure A10: **Event Study of Moves Across Zip Codes**

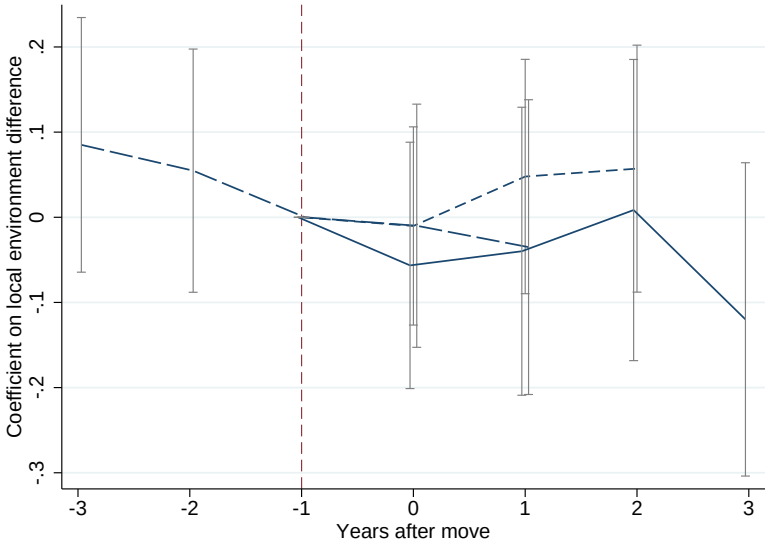


Notes: Using 2004-2015 Homescan data, these figures present results for the event study of moves across zip codes. The top left panel presents the share of shopping trips that are in the new versus old county. The top right panel presents the distribution across balanced panel households of the difference in Health Index between the new and old zip code. The bottom panels present the τ_y parameters and 95 percent confidence intervals from estimates of Equation (5): associations between household-level Health Index and the difference in average local Health Index between post-move and pre-move locations. The bottom right panel includes controls for household demographics (natural log of income, natural log of years of education, age indicators, an indicator for whether the household includes children, race indicators, employment status, weekly work hours, and total calorie need). Health Index is our overall measure of the healthfulness of grocery purchases, normalized to mean zero, standard deviation one across households. Observations are not weighted for national representativeness.

Figure A11: **Event Study of Movers with Different Balanced Sample Windows**



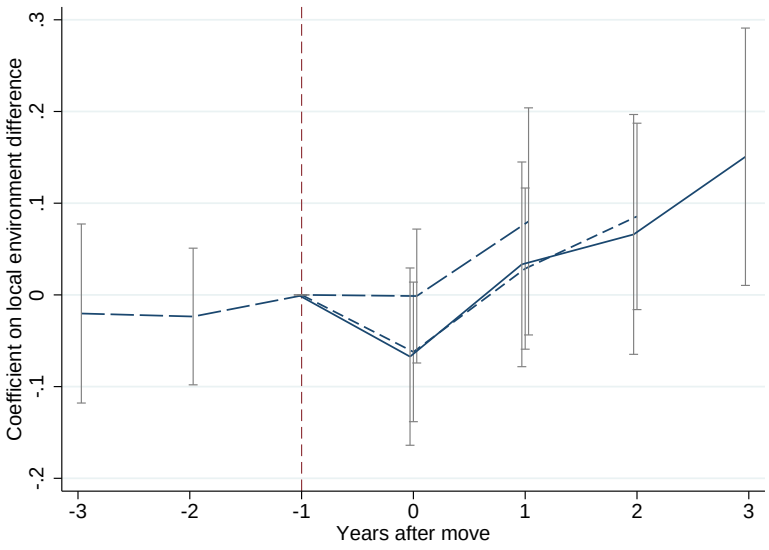
(a) **Without Controls**



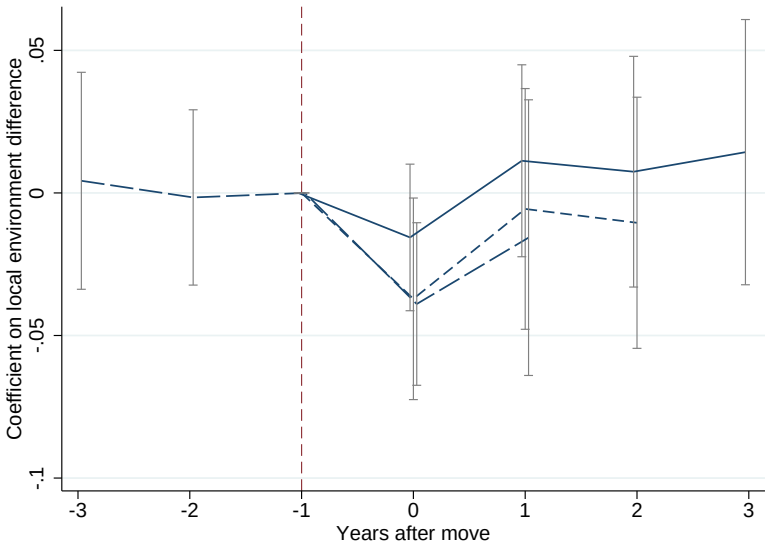
(b) **With Controls**

Notes: Using 2004-2015 Homescan data, these figures present the τ_y parameters and 95 percent confidence intervals from estimates of Equation (5): associations between household-level Health Index and the difference in average local Health Index between post-move and pre-move locations. Each figure superimposes three different estimates identified off of balanced panels for different windows around the move. Panel (b) includes controls for household demographics (natural log of income, natural log of years of education, age indicators, an indicator for whether the household includes children, race indicators, employment status, weekly work hours, and total calorie need). Health Index is our overall measure of the healthfulness of grocery purchases, normalized to mean zero, standard deviation one across households. Observations are not weighted for national representativeness.

Figure A12: **Event Study: Income Changes in Mover Households**



(a) **Moves Across Counties**



(b) **Moves Across Zip Codes**

Notes: Using 2004-2015 Homescan data, these figures present the τ_y parameters and 95 percent confidence intervals from estimates of Equation (5): associations between natural log of household income and the difference in average local Health Index between post-move and pre-move locations. All regressions control for year indicators and household fixed effects. Each figure superimposes three different estimates identified off of balanced panels for different windows around the move. Health Index is our overall measure of the healthfulness of grocery purchases, normalized to mean zero, standard deviation one across households. Observations are not weighted for national representativeness. The regressions are the same as in Figure 9, except with natural log of household income as the dependent variable and no controls for household demographics.

Table A8: **Association of Income with Local Area Health Index Using Movers**

	(1)	(3)
Zip code average Health Index	-0.00658	
	(0.0100)	
County average Health Index		0.0971
		(0.0335)***
Observations	560,492	565,914

Notes: This table uses 2004-2015 Nielsen Homescan data at the household-by-year level. The sample excludes observations where less than 50 percent of trips to RMS stores are not in the household's end-of-year county of residence. The dependent variable is the natural log of household income. Health Index is our overall measure of the healthfulness of grocery purchases, normalized to mean zero, standard deviation one across households. All regressions control for year indicators and household fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

Table A9: **Association of Coke Market Share with Local Area Coke Market Share Using Movers**

	(1)	(2)
County average Coke market share	0.138	0.136
	(0.0547)**	(0.0547)**
Household demographics	No	Yes
Observations	323,710	323,710

Notes: This table uses 2004-2015 Nielsen Homescan data at the household-by-year level. The sample excludes observations where less than 50 percent of trips to RMS stores are not in the household's end-of-year county of residence. Coke market share equals Coke calories purchased / (Coke + Pepsi calories purchased). Household demographics are natural log of income, natural log of years of education, age indicators, an indicator for whether the household includes children, race indicators, employment status, weekly work hours, and total calorie need. All regressions also control for year indicators and household fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

D Appendix to Demand Model Estimation

D.A GMM Estimation

Our GMM estimator is defined as follows:

$$\left(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\phi}}, \hat{\tilde{\boldsymbol{\beta}}}, \hat{\xi}\right) = \underset{(\boldsymbol{\delta},\boldsymbol{\phi},\tilde{\boldsymbol{\beta}},\xi)}{\arg\min} \left(\frac{1}{IJT}\sum_i\sum_j\sum_t \boldsymbol{g}_{ijt}\right)' \boldsymbol{W} \left(\frac{1}{IJT}\sum_i\sum_j\sum_t \boldsymbol{g}_{ijt}\right).$$

Define $\boldsymbol{Y}$ as the vector of product group calorie consumption Y_{ijt}, $\boldsymbol{F}\left(\tilde{\boldsymbol{\beta}}, \xi\right)$ as the vector of implicit prices $F_{ijt} = \left(\tilde{p}_{ijt} - \sum_{c=2}^{C} \tilde{\beta}_c \tilde{a}_{ijct} - \xi\right)$, $\boldsymbol{D}$ as a stacked matrix of the two dummy variable matrices ($\boldsymbol{D}_j$ and $\boldsymbol{D}_m$), $\boldsymbol{Z}$ as a matrix with all of our vectors of instruments ($\boldsymbol{D}$, the nutrient content $\tilde{\boldsymbol{a}}$, and the price instruments $\boldsymbol{P}$), and $\boldsymbol{Pr_D} = \left(\boldsymbol{D'ZWZ'D}\right)^{-1}\boldsymbol{D'ZWZ'}$ as a projection matrix. We can simplify the estimation problem by solving for our vectors of linear coefficients, $\boldsymbol{\delta}$ and $\boldsymbol{\phi}$, as analytic functions of $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\xi}$:

$$(\boldsymbol{\delta}, \boldsymbol{\phi}) = \boldsymbol{Pr}_D\left(\ln(\boldsymbol{Y}) - \boldsymbol{F}\left(\tilde{\boldsymbol{\beta}}, \xi\right)\right). \tag{35}$$

Substituting Equation (35) back into Equation (22), we can re-write the GMM estimator in terms of $\tilde{\boldsymbol{\beta}}$ and $\hat{\xi}$:

$$\left(\hat{\tilde{\boldsymbol{\beta}}}, \hat{\xi}\right) = \underset{(\tilde{\boldsymbol{\beta}},\xi)}{\arg\min} \left(\frac{1}{IJT}\sum_i\sum_j\sum_t g_{ijt}\left(\tilde{\boldsymbol{\beta}}, \xi\right)\right)' \boldsymbol{W} \left(\frac{1}{IJT}\sum_i\sum_j\sum_t g_{ijt}\left(\tilde{\boldsymbol{\beta}}, \xi\right)\right).$$

At the true value, the gradient for this problem is:

$$-2\boldsymbol{G}\left(\tilde{\boldsymbol{\beta}}, \xi\right)' \boldsymbol{W}\boldsymbol{G}\left(\tilde{\boldsymbol{\beta}}, \xi\right) = 0$$

where the Jacobian of the moments, $\boldsymbol{G}\left(\tilde{\boldsymbol{\beta}}, \xi\right)$, is

$$\boldsymbol{G}\left(\tilde{\boldsymbol{\beta}}, \xi\right) = \frac{1}{IJT}\begin{bmatrix} \tilde{\boldsymbol{a}}'\left(\boldsymbol{I} - \boldsymbol{DPr_D}\right) \\ \boldsymbol{P}'\left(\boldsymbol{I} - \boldsymbol{D}_m\boldsymbol{Pr}_{\boldsymbol{D}_m}\right) \\ \boldsymbol{D}'\left(\boldsymbol{I} - \boldsymbol{DPr_D}\right) \end{bmatrix} \nabla_\beta \boldsymbol{F}\left(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{a}}; \tilde{\boldsymbol{\beta}}\right). \tag{36}$$

In the above equation, $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{Pr}_{\boldsymbol{D}_m}$ is a projection matrix using $\boldsymbol{D}_m$.

The covariance matrix of our full GMM estimator, $\boldsymbol{\Theta}^{GMM} \equiv \left(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\phi}}, \hat{\tilde{\boldsymbol{\beta}}}\hat{\xi}\right)$, is $cov\left(\boldsymbol{\Theta}^{GMM}\right) =$

$(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W}\boldsymbol{\Omega}\boldsymbol{W}\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}$, with Jacobian matrix

$$G = \frac{1}{IJT}\sum_i\sum_j\sum_t \begin{bmatrix} \overrightarrow{0}'_J & -\tilde{a}_{ijt}D'_m & -\tilde{a}_{ijt}\nabla_\beta F'_{ijt} \\ -P_{jmt}D'_j & -P_{jmt}D'_m & -P_{jmt}\nabla_\beta F'_{ijt} \\ -D_jD'_j & -D_jD'_m & -D_j\nabla_\beta F'_{ijt} \\ -D_mD'_{ijt} & -D_mD'_m & -D_m\nabla_\beta F'_{ijt} \end{bmatrix}$$

and covariance matrix

$$\boldsymbol{\Omega} = E\left(\boldsymbol{g}_{ijt}\left(\boldsymbol{\Theta}^{GMM}\right)\boldsymbol{g}_{ijt}\left(\boldsymbol{\Theta}^{GMM}\right)'\right).$$

When computing our standard errors, we cluster by household as follows:

$$\hat{\Omega} = \frac{1}{IJT}\sum_i\sum_{j,j'}\sum_{t,t'}\boldsymbol{g}_{ijt}\left(\hat{\tilde{\boldsymbol{\beta}}},\hat{\xi}\right)\boldsymbol{g}_{ij't'}\left(\hat{\tilde{\boldsymbol{\beta}}},\hat{\xi}\right)'.$$

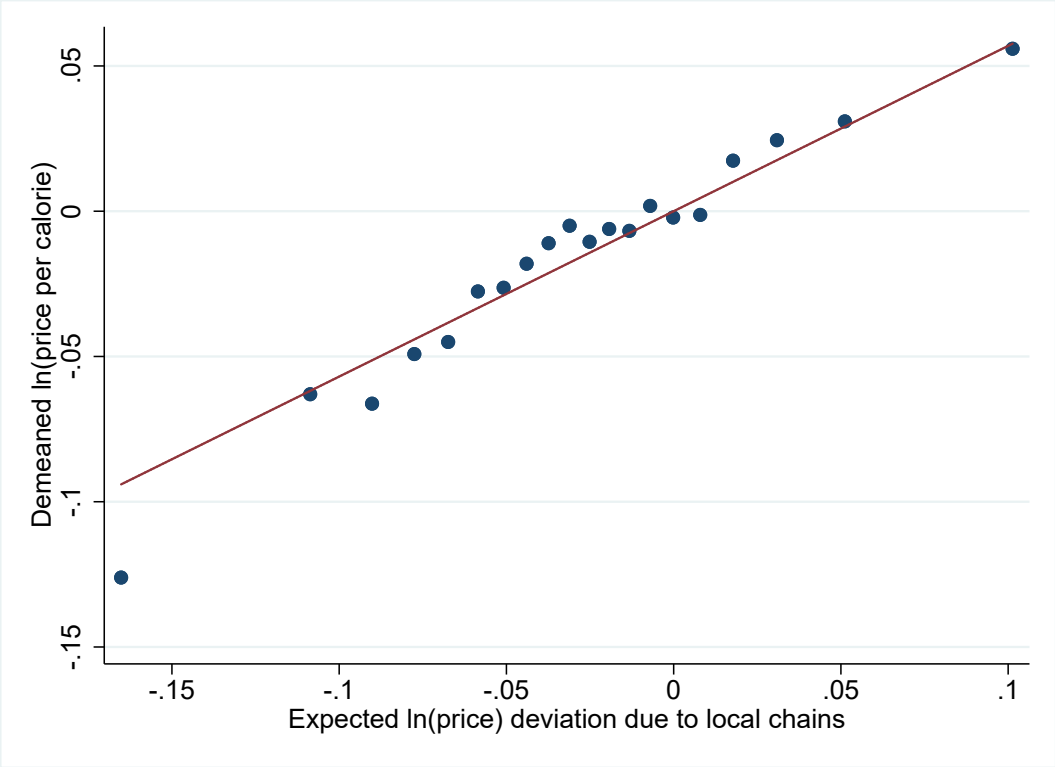
D.B Additional Tables and Figures

Table A10: **Preferences for Nutrients by Household Income**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Income group	Unsat. Fat	Sat. Fat	Fiber	Protein	Sugar	Sodium	Cholest.	Fruit	Veg	Unobs. Nutrient	Conve-nience	Shelf Life	WTP for Health Index
$Inc \leq 25k$	-0.83***	9.82***	9.98***	4.40***	1.94***	-29.76***	-13.47***	0.08***	0.42***	0.09***	0.17***	-0.90***	0.52***
	(0.11)	(0.15)	(0.17)	(0.29)	(0.04)	(1.86)	(4.04)	(0.02)	(0.02)	(0.02)	(0.002)	(0.02)	(0.012)
$25k < Inc \leq 50k$	-1.77***	13.35***	10.17***	6.61***	1.91***	-29.93***	-13.45***	0.24***	0.65***	-0.05***	0.22***	-0.80***	0.72***
	(0.07)	(0.08)	(0.06)	(0.14)	(0.02)	(0.81)	(1.84)	(0.01)	(0.004)	(0.01)	(0.001)	(0.01)	(0.01)
$50k < Inc \leq 70k$	-3.15***	11.12***	10.478***	7.350***	1.39***	-30.06***	-13.45***	0.38***	0.75***	0.09***	0.24***	-0.82***	0.91***
	(0.07)	(0.09)	(0.10)	(0.13)	(0.02)	(0.85)	(1.91)	(0.01)	(0.004)	(0.01)	(0.001)	(0.01)	(0.01)
$70k < Inc$	-3.57***	10.50***	10.59***	8.24***	1.26***	-30.14***	-13.45	0.60***	0.93***	0.11	0.28***	-0.88***	1.12***
	(0.58)	(0.55)	(0.41)	(0.96)	(0.15)	(5.41)	(11.57)	(0.04)	(0.03)	(0.07)	(0.01)	(0.05)	(0.05)

Notes: This table presents GMM estimates of the preference parameters $\tilde{\beta}_c$ from Equation (16), adding convenience and shelf life as additional product characteristics. Shelf life is measured in years and top-coded at one year. Convenience is a score ranging from 0 to 3, defined as follows. 0: basic ingredients. These are raw or minimally processed foods used in producing a meal or snack that are generally composed of a single ingredient, such as milk, dried beans, rice, grains, butter, cream, fresh meat, poultry, and seafood. 1: complex ingredients, such as bread, pasta, sour cream, sauce, canned vegetables, canned beans, pickles, cereal, frozen meat/poultry/seafood, canned meat/poultry/seafood, and lunch meat. 2: ready-to-cook meals and stacks. These are foods that require minimal preparation involving heating, cooking, or adding hot water, such as frozen entrees, frozen pizzas, dry meal mixes, pudding mixes, soup, chili, and powdered drinks. 3: ready-to-eat meals and snacks. These are foods that are intended to be consumed as is and require no preparation beyond opening a container, including refrigerated entrees and sides, canned and fresh fruit, yogurt, candy, snacks, liquid drinks, and flavored milk. Shelf life data are from Okrent and Kumcu (2016), while convenience data are from the U.S. government's FoodKeeper app (HHS 2015). Magnitudes of nutrient estimates represent willingness to pay for a kilogram of the nutrient instead of a kilogram of carbohydrates. Value of fruit and vegetables accounts for value over and beyond macronutrient characteristics of the fruit and vegetables. "WTP for Health Index" in column 13 equals $\sum_c \hat{\tilde{\beta}}_c G_c r_c$, where $G_c = 1$ for "healthy" nutrients, $G_c = -1$ for "unhealthy" nutrients, and r_c is the recommended daily intake of nutrient c detailed in Appendix Table A1. Standard errors, clustered by household, are in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

Figure A13: **Binned Scatterplot of First Stage Price Regression**



Notes: This figure presents a binned scatterplot of a regression of natural log price per calorie on our price instrument P_{jmt}, using Homescan data at the household-by-product group-by-year level. Zip-3 and product group fixed effects are residualized out before plotting. There are 20 equally sized bins, and all income groups are included.