Supplementary Appendix: Does plant-based milk reduce sales of dairy milk? Evidence from the almond milk craze

SA.1 Introduction

This appendix contains supplementary information for the paper *Does plant-based milk reduce sales of dairy milk? Evidence from the almond milk craze*. Section SA.2 provides additional details about the data construction and further descriptive statistics. Section SA.3 contains additional results from alternative models.

SA.2 Data

This section contains three subsections. The first subsection outlines the identification of dairy and non-dairy milks in the data. The second subsection documents the construction of price indices. The third subsection provides additional descriptive statistics.

SA.2.1 Identifying dairy and non-dairy milk products

I define a product as a dairy milk if it is in the product category "Dairy milk refrigerated" defined in the Nielsen data (note that flavored milks are in a separate product category). I define a product as a *possible* non-dairy milk if it is in the product group "Remaining drinks & shakes-refrigerated" or "Remaining drinks & shakes-non refrigerated" or if the universal produce code (UPC) is categorized as a "plant based milk" in the USDA's Branded Food Product Database (United States Department of Agriculture, 2020). I distinguish thirteen different types of non-dairy milk by the primary ingredient (almond, cashew, coconut, flax, hemp, macadamia, oat, pea, pistachio, quinoa, rice, soy, and walnut). A non-dairy milk is defined as being of a certain type (e.g., soy or almond) if it contains the name of that

primary ingredient in its ingredient list or UPC description (ingredient lists are only available for products that are included in the Branded Food Database). For products with multiple non-dairy ingredients (e.g., almond and cashew), I define the product based on the first ingredient listed. I delete any product that is defined as a shake, chocolate flavored, or horchata. This procedure does pick up a number of canned coconut milks that might be functionally different than other milks (i.e., more likely to be used in cooking) – most, if not all, of these products are eliminated when I drop products that have fewer than 28 ounces.

SA.2.2 Price index

To calculate the price index, I classify each milk product by brand and size (milk products are typically sold in 32, 64, or 128-ounce containers, with larger containers selling at a lower per-ounce price). I then impute an annual household price for each brand-size that accounts for more than 0.5% of dairy or non-dairy milk sales. This procedure results in eight different brand-sizes for both dairy and non-dairy milk. The small number of brand-sizes is due, in part, to all store brands being aggregated together in the Nielsen data.

To impute the prices of each brand-size, I regress observed prices on (a) brand-size dummies, (b) the household's state, (c) yearly dummy variables, and (d) sociodemographic variables. Sociodemographic variables include the age of the household heads, education of household heads, a second-degree polynomial of household income, household size, the presence of children in the household, dummy variables for race/ethnicity (Asian, black and Hispanic), and dummy variables for the gender/number of household heads (female head only and male head only). I then use the estimates to predict a brand-level price for each household.

After defining a time- and household-specific price for each brand-size, I aggregate dairy milk prices using a Fisher ideal price index. The dairy milk price for the ith household in the tth time period is,

$$p_{i,t} = \sqrt{\frac{\sum \tilde{p}_{i,k,t}\tilde{q}_{k,0}}{\sum \tilde{p}_{k,0}\tilde{q}_{k,0}} \frac{\sum \tilde{p}_{i,k,t}\tilde{q}_{i,k,t}}{\sum \tilde{p}_{k,0}\tilde{q}_{i,k,t}}}, \tag{1}$$

where k indexes the brand-sizes, and the tilde is used to define brand-size prices and quantities. The terms $\tilde{q}_{k0}$ and $\tilde{p}_{k0}$ denote national averages in 2013. For households who do not consume any milk in a given year, I replace $\tilde{q}_{i,k,t}$ with the national average of the kth brand-size purchased in the tth period.

SA.2.3 Descriptive statistics

This subsection has five tables that explore how the descriptive statistics in the paper vary across subsets of the data. Tables SA.1 and SA.2 are analogous to Tables 1 and 2 in the paper, but include only single person households. Table SA.3 is the same as Table 2 in the paper, except that it only includes households who remain in the dataset in all the years between 2004 and 2018. Finally, Table SA.4 is similar to Table 2 in the paper, except it uses data from 2009-2018 and categorizes households based on their dairy milk consumption in 2008.

Table SA.1: Descriptive statistics of sociodemographic and expense variables (single-person households)

	All households	Non-dairy milk purchases: Positive	Non-dairy milk purchases: Greater than dairy milk
Quantity (gallons)			
Dairy milk	11.9 (25.9)	7.45 (23.2)***	0.851 (4.16)***
Almond milk	0.330 (2.45)	2.16 (5.21)***	3.65 (8.33)***
All non-dairy milk	0.629 (3.54)	4.11 (7.09)***	6.85 (10.6)***
Price (per gallon)			
Dairy milk	4.04 (1.50)	4.36 (1.74)***	4.91 (2.04)***
Almond milk	5.75 (1.85)	5.75 (1.85)	5.76 (1.88)
All non-dairy milk	6.23 (2.12)	6.23 (2.12)	6.19 (2.00)***
Sociodemographics			
Female head only	0.509	0.592***	0.573***
Male head only	0.491	0.408***	0.427***
Average age of household heads[a]	7.13 (2.18)	6.70 (2.25)***	6.62 (2.30)***
Highest education of household heads[b]	2.14 (1.00)	2.32 (1.00)***	2.41 (1.01)***
Black	0.131	0.180***	0.210***
Hispanic	0.054	0.075***	0.072***
Asian	0.020	0.032***	0.037***
Household income[c]	7.25 (4.02)	7.70 (3.89)***	7.81 (4.09)***
Total household expenses	1,507 (1,393)	1,654 (1,446)***	1,546 (1,286)***
Data size			
Number of households	46,909	14,717	6,663
Number of observations	215,824	35,918	17,040

Weighted means with weighted standard deviations in parentheses. *, **, *** denote statistically significant differences between the households listed in the column and all other households at the 10%, 5% and 1% levels.

[a]1=Less than 25 (years); 2=25-29; 3=30-34; 4=35-39; 5=40-44, 6=45-49; 7=50-54; 8=55-64; 9=Over 65.

[b]1=No post-secondary; 2=Some college; 3=College degree; 4=Graduate degree.

[c]1=Less than 5; 2=5-10; 3=10-15; 4=15-20; 5=20-25; 6=25-30; 7=30-35; 8=35-40; 9=40-50; 10=50-60; 11=60-70; 12=70-80; 13=80-100; 14=Over 100 (in $000s).

Table SA.2: Almond milk and non-dairy milk expenses (2005-2018, single-person households)

Dairy milk expense share (2004)	2005-2009	2010-2012	2013-2015	2016-2018
Average expense share: almond milk				
Less than the 1st quartile	0.006	0.460	1.03	1.13
Between the 1st and 2nd quartiles	0.001	0.221	0.581	0.593
Between the 2nd and 3rd quartiles	0.001	0.333	1.03	1.13
Greater than the 3rd quartile	0.000	0.147	0.689	1.08
Average expense share: non-dairy milk				
Less than the 1st quartile	1.16	1.39	1.52	1.49
Between the 1st and 2nd quartiles	0.358	0.583	0.850	0.753
Between the 2nd and 3rd quartiles	0.351	0.646	1.25	1.28
Greater than the 3rd quartile	0.348	0.360	0.928	1.38
Observations				
Number of households	11,565	7,604	6,123	4,682

Table SA.3: Almond milk and non-dairy milk expenses (2005-2018, balanced panel)

Dairy milk expense share (2004)	2005-2009	2010-2012	2013-2015	2016-2018
	Average expense share: almond milk (%)			
Less than the 1st quartile	0.009	0.273	0.851	1.16
Between the 1st and 2nd quartiles	0.002	0.278	0.717	0.640
Between the 2nd and 3rd quartiles	0.004	0.277	0.866	0.813
Greater than the 3rd quartile	0.000	0.146	0.738	0.865
	Average expense share: non-dairy milk (%)			
Less than the 1st quartile	1.05	0.930	1.27	1.53
Between the 1st and 2nd quartiles	0.541	0.883	0.985	0.825
Between the 2nd and 3rd quartiles	0.337	0.642	1.11	1.02
Greater than the 3rd quartile	0.359	0.463	1.05	1.13
Observations				
Number of households	6,260	6,260	6,260	6,260

Table SA.4: Almond milk and non-dairy milk expenses (2008-2018)

Dairy milk expense share (2007)	2009-2010	2011-2013	2014-2016	2017-2018
	Average expense share: almond milk (%)			
Less than the 1st quartile	0.112	0.564	0.979	1.09
Between the 1st and 2nd quartiles	0.065	0.355	0.697	0.692
Between the 2nd and 3rd quartiles	0.052	0.348	0.834	0.771
Greater than the 3rd quartile	0.059	0.341	0.789	0.856
	Average expense share: non-dairy milk (%)			
Less than the 1st quartile	1.30	1.26	1.44	1.46
Between the 1st and 2nd quartiles	0.469	0.644	0.946	0.901
Between the 2nd and 3rd quartiles	0.405	0.681	1.09	0.989
Greater than the 3rd quartile	0.418	0.675	1.04	1.10
Observations				
Number of households	50,006	38,509	29,326	22,186

SA.3 Additional results

This section contains results of alternative specifications mentioned in section 4. These models differ form the base results in Table 3 of the paper in two ways. First, some models contain a subset of observations either from single-person households or observations after 2008. Second, some models used almond milk, as opposed to aggregate sales of non-dairy milk, as the explanatory variable. These models are summarized in Table SA.5. This table presents the coefficients on non-dairy or almond milk from each of the models. The full results of these models are contained in Tables SA.6-SA.10.

Table SA.11 contains the full results for the models listed Table 4 in the paper. Finally, Table SA.12 recreates Table 5 in the paper using almond milk as the explanatory variable, as opposed to non-dairy milk.

Table SA.5: Results from alternative models

	First differences	First differences-IV	Fixed effects
Non-dairy milk			
2009-2018	-0.452 (0.021)***	-0.410 (0.049)***	-0.591 (0.023)***
Single households	-0.401 (0.022)***	-0.510 (0.064)***	-0.492 (0.033)***
Almond milk			
All households	-0.472 (0.028)***	-0.429 (0.065)***	-0.653 (0.034)***
2009-2018	-0.465 (0.028)***	-0.423 (0.064)***	-0.612 (0.030)***
Single households	-0.381 (0.032)***	-0.384 (0.200)	-0.462 (0.041)***

Each coefficient is based on a separate regressions (full results from these regressions are in Tables SA.6-SA.10). In the regressions listed in the third, fourth and fifth rows, the quantity of non-dairy milk is replaced with the quantity of almond milk. Standard errors, clustered on the household, are in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% levels. Full results from these models are contained in the supplementary appendix (Tables SA.8-SA.12).

Table SA.6: Alternative models (2009-2018)

	First differences	First differences-IV[a]	Fixed effects
Coefficients			
Non-dairy milk (quantity)	-0.452 (0.021)***	-0.410 (0.049)***	-0.591 (0.023)***
Milk price (log)	-2.448 (0.200)***	-2.812 (0.215)***	-3.916 (0.316)***
Non-dairy milk price (log)	-1.524 (0.394)***	-1.532 (0.452)***	-3.427 (0.617)***
Income	-0.011 (0.008)	-0.014 (0.010)	-0.064 (0.014)***
Household size	0.213 (0.054)***	0.213 (0.060)***	0.422 (0.082)***
Child in the household	0.880 (0.156)***	0.949 (0.166)***	2.791 (0.231)***
Total household expenses	0.007 (0.000)***	0.007 (0.000)***	0.007 (0.000)***
Model information			
Number of households	495,322	495,322	575,702
Number of observations	108,749	108,749	111,222
R-squared	0.147	0.165	0.220

Standard errors, clustered on the household, are in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% levels. The model also includes year dummies.

[a]Robust F-statistic (strength of instruments): 53.7. Hansen J-statistic (overidentification test): 13.4 (p-value=0.337). Difference-in-Sargan test statistic (test of exogeniety of non-dairy milk): 0.251 (p-value = 0.616).

Table SA.7: Alternative models (single households)

	First differences	First differences-IV[a]	Fixed effects
Coefficients			
Non-dairy milk (quantity)	-0.401 (0.022)***	-0.510 (0.064)***	-0.492 (0.033)***
Milk price (log)	-0.860 (0.237)***	-0.866 (0.237)***	-1.914 (0.402)***
Non-dairy milk price (log)	-0.668 (0.536)	-0.850 (0.540)	-2.742 (0.948)**
Income	-0.010 (0.010)	-0.009 (0.010)	-0.027 (0.017)
Total household expenses	0.005 (0.000)***	0.005 (0.000)***	0.005 (0.000)***
Model information			
Number of households	159,567	159,567	207,179
Number of observations	32,780	32,780	38,264
R-squared	0.110	0.110	0.136

Standard errors, clustered on the household, are in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% levels. The model also includes year dummies.

[a]Robust F-statistic (strength of instruments): 11.6. Hansen J-statistic (overidentification test): 12.1 (p-value=0.438). Difference-in-Sargan test statistic (test of exogeniety of non-dairy milk): 3.73 (p-value = 0.053).

Table SA.8: Alternative models (almond milk as the explantory variable)

	First differences	First differences-IV[a]	Fixed effects
Coefficients			
Almond milk (quantity)	-0.472 (0.028)***	-0.429 (0.065)***	-0.653 (0.034)***
Milk price (log)	-2.459 (0.191)***	-2.750 (0.211)***	-4.153 (0.334)***
Almond milk price (log)	-1.613 (0.436)***	-1.692 (0.513)***	-3.971 (0.806)***
Income	-0.009 (0.008)	-0.019 (0.009)*	-0.103 (0.015)***
Household size	0.228 (0.050)***	0.207 (0.055)***	0.745 (0.086)***
Child in the household	0.970 (0.145)***	1.062 (0.159)***	3.471 (0.228)***
Total household expenses	0.007 (0.000)***	0.007 (0.000)***	0.008 (0.000)***
Model information			
Number of households	637,763	637,763	796,133
Number of observations	126,589	126,589	129,160
R-squared	0.145	0.161	0.242

Standard errors, clustered on the household, are in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% levels. The model also includes year dummies.

[a]Robust F-statistic (strength of instruments): 60.3. Hansen J-statistic (overidentification test): 9.05 (p-value=0.338). Difference-in-Sargan test statistic (test of exogeniety of non-dairy milk): 0.201 (p-value = 0.654).

Table SA.9: Alternative models (2009-2018 with almond milk as the explantory variable)

	First differences	First differences-IV[a]	Fixed effects
Coefficients			
Almond milk (quantity)	-0.465 (0.028)***	-0.423 (0.064)***	-0.612 (0.030)***
Milk price (log)	-2.458 (0.200)***	-2.819 (0.215)***	-3.903 (0.317)***
Almond milk price (log)	-1.611 (0.435)***	-1.655 (0.511)**	-3.636 (0.660)***
Income	-0.011 (0.008)	-0.015 (0.010)	-0.065 (0.014)***
Household size	0.213 (0.054)***	0.213 (0.060)***	0.423 (0.082)***
Child in the household	0.864 (0.155)***	0.956 (0.166)***	2.789 (0.232)***
Total household expenses	0.006 (0.000)***	0.007 (0.000)***	0.007 (0.000)***
Model information			
Number of households	495,322	495,322	575,702
Number of observations	108,749	108,749	111,222
R-squared	0.145	0.162	0.216

Standard errors, clustered on the household, are in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% levels. The model also includes year dummies.

[a]Robust F-statistic (strength of instruments): 60.1. Hansen J-statistic (overidentification test): 9.29 (p-value=0.318). Difference-in-Sargan test statistic (test of exogeniety of non-dairy milk): 0.174 (p-value = 0.677).

Table SA.10: Alternative models (single households with almond milk as the explantory variable)

	First differences	First differences-IV[a]	Fixed effects
Coefficients			
Almond milk (quantity)	-0.381 (0.032)***	-0.384 (0.200)	-0.462 (0.041)***
Milk price (log)	-1.004 (0.242)***	-0.882 (0.239)***	-1.890 (0.405)***
Almond milk price (log)	-0.112 (0.551)	-0.176 (0.616)	-2.551 (1.161)*
Income	-0.005 (0.009)	-0.011 (0.011)	-0.031 (0.017)
Total household expenses	0.005 (0.000)***	0.005 (0.000)***	0.005 (0.000)***
Model information			
Number of households	159,567	159,567	207,179
Number of observations	32,780	32,780	38,264
R-squared	0.097	0.104	0.128

Standard errors, clustered on the household, are in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% levels. The model also includes year dummies.

[a]Robust F-statistic (strength of instruments): 4.17. Hansen J-statistic (overidentification test): 14.4 (p-value=0.001). Difference-in-Sargan test statistic (test of exogeniety of non-dairy milk): 0.876 (p-value = 0.349).

Table SA.11: Estimation results (2009-2018 differences)

	First differences	First differences-IV[a]	First differences
Dependent variable			
Dependent variable	ΔDairy milk (2009-2018)	ΔDairy milk (2009-2018)	ΔDairy milk (2004-2009)
ΔNon-dairy milk (2009-2018)	-0.649 (0.059)***	-0.517 (0.181)**	-0.011 (0.077)
Milk price (log)	-4.563 (1.019)***	-4.418 (1.012)***	-4.879 (1.360)***
Coefficients			
Non-dairy milk price (log)	-8.719 (3.563)*	-8.173 (3.639)*	0.586 (4.113)
Income	-0.129 (0.042)**	-0.131 (0.042)**	-0.226 (0.054)***
Household size	0.457 (0.306)	0.435 (0.306)	1.296 (0.331)***
Model information			
Child in the household	6.351 (0.764)***	6.398 (0.762)***	4.396 (0.829)***
Total household expenses	0.007 (0.000)***	0.007 (0.000)***	0.008 (0.000)***
Number of observations	22,174	22,174	9,117

All models include the same explanatory variables as Table **??**. Standard errors are in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% levels. Full results are contained in the supplementary appendix (Table SA.13)

[a]The instrument set includes all levels of non-dairy milk prior to 2009 (missing values are replaced with zero). Robust F-statistic (strength of instruments): 27.4. Hansen J-statistic (overidentification test): 6.56 (p-value=0.161). Difference-in-Sargan test statistic (test of exogeniety of non-dairy milk): 0.542 (p-value = 0.462).

Table SA.12: Interaction between almond milk and previous milk consumption

	First differences	First differences (control function)	Fixed effects	First differences (2009-2018)
Non-dairy milk coefficient by milk consumption in 2004 or 2008[a]				
Less than the 1st quartile	-0.242 (0.038)***	-0.227 (0.101)*	-0.075 (0.042)	-0.205 (0.028)***
Between the 1st and 2nd quartiles	-0.422 (0.090)***	-0.407 (0.127)**	-0.469 (0.068)***	-0.412 (0.054)***
Between the 2nd and 3rd quartiles	-0.586 (0.054)***	-0.571 (0.109)***	-0.809 (0.116)***	-0.467 (0.046)***
Greater than the 3rd quartile	-0.707 (0.085)***	-0.692 (0.123)***	-1.431 (0.140)***	-0.785 (0.127)***
Model information				
Number of households	33,403	33,403	34,046	50,169
Number of observations	245,385	245,385	257,559	297,331
R-squared	0.137	0.137	0.249	0.153

Standard errors, clustered on the household, are in parentheses. In the second model, standard errors are bootstrapped using 500 iterations . *, **, *** represent statistical significance at the 10%, 5%, and 1% levels. The model also includes year dummies.

[a]Refers to milk consumption in 2004 (models 1-3) or milk consumption in 2008 (model 4).

References

United States Department of Agriculture. 2020. Branded Food Products Database. Available at: https://data.nal.usda.gov/dataset/usda-branded-food-products-database, accessed on May 2, 2020.