Got Milk? An Empirical Study of Milk Consumption Patterns Based on Movers[*,†]

Nicole Golden[‡]

March 17, 2022

Abstract

This paper exploits a natural experiment in the U.S. among households who have moved to another state to study how migration affects dairy and plant-based milk consumption. Using Nielsen Consumer Data to measure outcomes, I find that after moving to a new state, on average, (i) the movers have increased their dairy milk expenditures by 1.2%, whereas the plant-based milk expenditures have decreased by 1.5%; and (ii) the new destination explains about 53% of the differences in dairy milk expenditures and only about 17% for plant-based milk expenditures. These results imply a more considerable convergence for dairy milk expenditures toward the destination state's average level. In contrast, the plant-based milk expenditures have only slightly converged toward the average level in the households' new states.

Keywords: dairy milk, plant-based milk, migration, geographic variation, demand, preferences, Nielsen Homescan data, food sustainability, food's environmental impact

JEL Codes: D12, I12, L66

[*]I am very grateful to my advisor Matthew Harding for his guidance and support as always. I also want to thank Damon Clark, Jan Brueckner, Jiawei Chen, and Yingying Lee for their helpful comments.

[†]Researcher(s) own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

[‡]University of California, Irvine, `qinqiy1@uci.edu`

1 Introduction

Herrnstadt et al. (2020) estimated that climate change would reduce the U.S. average annual real GDP growth by about 0.03% from 2020 to 2050. Burke et al. (2015)[1]forecasted that by 2100, there would be -36% change in U.S. GDP per capita due to the climate change. The diversification of food production has brought many negative ecological impacts, especially among animal-based products. Poore and Nemecek (2018) found that even for just the lowest-impact animal-based products, their ecological impact typically exceeded those of plant-based products. McClements et al. (2019) reviewed the environmental impacts, especially for dairy and plant-based milk ("plant milk")[2]. When considering some important environmental indicators compared with plant milk, dairy milk requires at least nine times as much land, causes more than ten times as much eutrophication (about twice as much compared to rice milk), and uses at least three times as much land. Compared with the most water-hungry plant-based almond milk, dairy milk still needs more than 1.5 times as much water (McClements et al., 2019). These economic damages call for dire changes in our economic policies, and there are still hopes to reduce such adverse impacts. The Nobel Prize laureate Nordhaus (2019) proposed that we should implement policies that could slow emissions as soon as possible and the most effective ones that "equalize the incremental or marginal costs of reducing emissions."

Understanding how consumers allocate their animal-based and plant-based food expenditures has critical economic implications on food sustainability-related policies. Much research has been done on incentivizing sustainable production (or consumption) to mitigate adverse environmental impacts from animal-based foods. However, little is known about how exactly households allocate their expenditures among animal-based foods and plant-based foods and how their expenditures among these two types of foods are affected by their local environments.

This paper uses dairy and plant milk as an example to answer these types of questions. Specifically, the main objectives of this study are: (1) to uncover expenditure patterns for both dairy and plant milk among households who have moved to another state and those who have not; (2) to learn how the new local environment affects household's expenditures for dairy and plant milk. I use Nielsen Consumer data for this study since it contains consumer characteristics and weekly milk purchases.

[1]Burke et al. (2015) created an interactive webpage that shows estimates of climate change for every county. See https://web.stanford.edu/~mburke/climate/map.php

[2]The plant-based milk here includes soy milk, almond milk, oat milk, and rice milk.

Nielsen updates households' Federal Information Processing Standard (FIPS) state and county code annually, which allows us to identify movers. The outcome variable is the logarithm of quarterly expenditures per person for dairy and plant milk.

The first econometric strategy exploits the fact that households move randomly[3]. This randomness allows the use of the difference-in-differences (DiD) model to identify milk expenditure patterns among movers (treated group) and non-movers (control group).

The second econometric strategy also uses the difference-in-differences model. It creates a new predictor, "size of move"[4] to measure the percentage change of milk expenditures due to moving to another state. "Size of move" measures the difference between the milk expenditures in the movers' new state and the original state. This empirical strategy difference-in-differences depends on how the environment, including milk supply, milk (or farm) regulation, taxes, and movers' peers, changes discretely when households move. If the new environment is the main drive to the milk purchases, say a household moves from state 1 to state 2, we should see a jump in the household's milk purchases to a level to that of the consumers in state 2. However, if milk purchases are only driven by household characteristics, such as their milk preferences and personal experiences, we should not expect a jump in the movers' milk purchases.

The results from the first difference-in-differences model show that, after households move to a new state, there is a 1.2% increase in dairy milk expenditures and a 1.5% decrease for plant milk expenditures. These results are surprising. It is known that dairy milk consumption has decreased at an average rate of 1% per year during the 2000s. Moreover, during the 2010s, it is decreasing at a 2.6% rate per year (USDA, 2021). In the meantime, Plant Based Foods Association (2020) found that compared with a small 0.1% sales growth in dairy milk, plant milk has experienced a 5% growth in 2019.

The fact that people are increasingly choosing plant milk motivates the second econometric strategy. The results from the second difference-in-differences model show that the new destination explains about 53% of the differences in dairy milk expenditures but only about 17% for plant-based milk expenditures. These results imply a more considerable convergence for dairy milk expenditures toward the destination state. However, its expenditures have only slightly converged toward the average level

[3]In section 3.1, I use a logit model to show that households indeed move at random.

[4]I will give a detailed explanation about its construction in Section 4

in the households' new states for plant milk.

This paper relates to several works on plant milk related topics in economics. First, there is vast literature comparing the nutritional values between dairy and plant milk. Vanga and Raghavan (2018) compared the nutritional values among various dairy and plant milk, and they concluded that soy milk might be the best dairy milk alternative in the human diet. Astolfi et al. (2020) did a similar analysis by comparing forty-one elements from multiple dairy and plant milk samples, and the authors also found that soy milk was the best alternative for human health. These works show that plant milk is gaining more popularity and thus motivate the need to study plant milk related literature.

Second, this paper relates to willingness to pay (WTP) in plant milk. Falkdalen (2017) used household scanner data in Sweden in 2011. It was estimated that the WTP for plant milk is 32% higher than the average milk price. A recent study by Yang and Dharmasena (2021) focused on soy, almond, and rice milk. Among the three, soy milk had the highest own-price elasticity. Consumers were not very sensitive to price changes due to inelastic demand. It also found that soy milk was a substitute for all four types of dairy milk[5]. The three types of plant milk were also complements among each other. This study adds to the WTP literature showing that if we use a natural experiment by separating the movers and non-movers, dairy milk is still the dominant type consumed in the U.S.

Third, some papers discuss food's environmental impacts. One of the factors that consumers choose plant milk is environmental sustainability. Poore and Nemecek (2018) compared the environmental impacts of various food products by compiling data covering 38,700 farms and 1600 processors. The finding showed that even for just the lowest-impact animal-based products, their environmental impacts typically exceeded those of plant-based products. McClements et al. (2019) looked at just dairy and plant milk products. They revealed that dairy milk had much higher GHG emissions, land use, and water use than all types of plant milk on a per-liter basis. Among plant milk, almond and rice milk required much higher water use than soy milk and oat milk, and the land use was slightly higher for the latter two. These findings suggest great environmental benefits of consuming more plant milk, but the benefits largely depend on the type of plant milk chosen. I intends to use my findings to potentially shed light on environment-related federal and state food policies since these findings reveal geographic variations.

Finally, this paper relates to the literature on determinants of a household's choice of dairy and

[5]The types of dairy milk include whole milk, 2% milk, 1% milk, and fat-free milk.

plant milk. Boaitey and Minegishi (2020) show that parental dairy consumption patterns can impact their children's dairy milk choices. An increasing number of children grow up in a household with more plant milk consumption. This phenomenon may be due to parents' perception of animal welfare and food sustainability. This study may be extended in the future to study how parents influence children's milk choices if they move to a new destination.

This paper's findings are of relevance to food and environment policymakers. A report by Grey, Clark, Shih and Associates, Limited (2018) discovered that 73% of producer returns in 2015 came from various U.S. dairy subsidies. The USDA administers a dozen of programs to regulate and assist the dairy industry[6]. The top program Milk Income Loss Contract Payment compensates dairy producers when domestic milk prices fall below a certain level. It is estimated that it had made total payments of more than $3.5 billion from 1995 to 2020 (Environment Working Group, 2021). Dairy Market Loss Assistance Program provides similar financial assistance and had made about $1 billion payments from 1995 to 2020 (Environment Working Group, 2021). The US dairy programs subsidies have amassed 6.4 billion dollars for the past 15 years (Environment Working Group, 2021). The economic consequences are that dairy farms oversupply milk and charge consumers at higher prices (Paulson et al., 2021). Although the total dairy milk sales are decreasing, this study finds that households consume more dairy milk (using a natural experiment with movers and non-movers) and less plant milk. These results may be due to dairy milk prices are not increasing (adjusted to inflation) for the past decade[7]. Due to changes in consumers' views on animal welfare and food sustainability, policymakers may need to be altered accordingly to the modern economy so that dairy farmers may benefit better in the long term instead of simply subsidizing them. From the environmental point of view, this study gives policymakers hints that dairy milk is still the dominant type of milk today. Various studies have proven that dairy milk has a much more significant impact on the environment than its plant-based opponent. Policymakers may need to implement more environmentally sustainable policies since the global warming problem has been more evident recently.

[6]For a complete list of U.S. dairy subsidy programs, see https://farm.ewg.org/progdetail.php?fips=00000&progcode=dairy.

[7]See Appendix for price index of dairy and plant milk constructed with Nielsen data.

2 Background on Dairy and Plant Milk in the U.S.

In the 1990s, One of the most acclaimed advertising campaigns, "Got Milk?" won the hearts of Americans of all ages and encouraged the consumption of milk. Or, more specifically, dairy milk. This event makes dairy milk one of the most commonly consumed beverages in the U.S. However, U.S. per capita consumption of dairy milk has been trending downward steadily since the 2000s (USDA, 2021). A recent report by (Cessna et al., 2021) states that dairy milk consumption has been decreasing at an average rate of 1% annually during the 2000s. It decreases at an even faster rate (2.6%) per year during the 2010s. In the meantime, the plant milk market has emerged and expanded quickly. Statista (2022) estimated that the global plant milk has about $21 billion market value in 2021, with a $3.1 billion market value in the U.S. Using 2019 IRI retailer data, Plant Based Foods Association (2020) found that compared with a small 0.1% sales growth in dairy milk, plant milk has experienced a 5% growth in 2019. Of course, plant-based foods are not just limited to the milk industry. There are also plant-based alternatives for meat, cheese, and yogurt. All types of plant-based foods combined totaled 11.4% in 2019, compared with a small 2.2% growth in all foods (Plant Based Foods Association, 2020).

There are several driving forces of the increasing demand for plant milk. First, people may choose plant milk over dairy milk for lactose intolerance or protein allergy reasons (Gerliani et al., 2019). Second, from a nutritional perspective, plant milk is an excellent source of macronutrients and micronutrients vital to human health[8], and it has lower calories and fat in general. Plant milk can be classified into five general categories: cereal-based, legume-based, nut-based, seed-based, and pseudo-cereal-based (Sethi et al., 2016). Take soy milk as an example; it contains a good amount of proteins, vitamins, minerals (Vanga and Raghavan, 2018). Third, plant milk is designed as functional drinks[9] that not only satisfy our desire for good taste but also nutritional needs. The emerging flexitarian lifestyle also influences plant-based foods consumption. Other factors that make people choose plant milk are environmental sustainability and farm animal welfare (Boaitey and Minegishi, 2020). This trend motivates me to study the expenditures of the two types of milk.

[8]For dairy and plant milk nutrient comparison, see Appendix A.

[9]By (USDA, 2010) definition, "(f)unctional foods are designed to have physiological benefits and/or reduce the risk of chronic disease beyond basic nutritional functions, and may be similar in appearance to conventional food and consumed as part of a regular diet."

3 Data

3.1 Nielsen Consumer Panel

This paper uses Nielsen Homescan Consumer Panel data from 2004 to 2019 to measure household-level milk purchases measured in expenditures. The panel is representative of the whole U.S. population[10]. Each household scans all their grocery purchases daily. The purchased quantities and prices for each household are recorded at Universal Product Code (UPC) level. This research cannot analyze milk purchases at the individual level since each purchase is recorded at the household level.

Each year, Nielsen surveys the household's demographic characteristics, such as age, income, education, marital status, household composition, and geographic location. By taking advantage of variations in the panel over time, this study can uncover movers and non-movers at state level[11].

3.2 Sample Construction

Four types of data files are used to construct sample data. The "products" file helps identify each type of dairy and plant milk[12]. Figure 1 shows a complete list of types of milk. Each year has a "purchases" file that records each household's weekly milk purchases. Using the total prices paid and total units, one can construct the outcome variables, which will be explained later in more detail. Each year's "trips" files record exact purchase dates, which uses each Saturday as the end of each week. The dates make it possible to aggregate the data at a quarterly level for the difference-in-differences analysis. They can also be aggregated at weekly, monthly, or annual levels, and they will be used for some other type of analysis later.

The critical data in identifying migrants are the "panelist" files. Each state has a unique two digit FIPS state code. By comparing the FIPS state code of the households from year to year in the data, one can identify movers and non-movers. This study defines "mover" as those households who have moved only once during the whole 14 years[13]. Since the movers' geographic location is recorded only

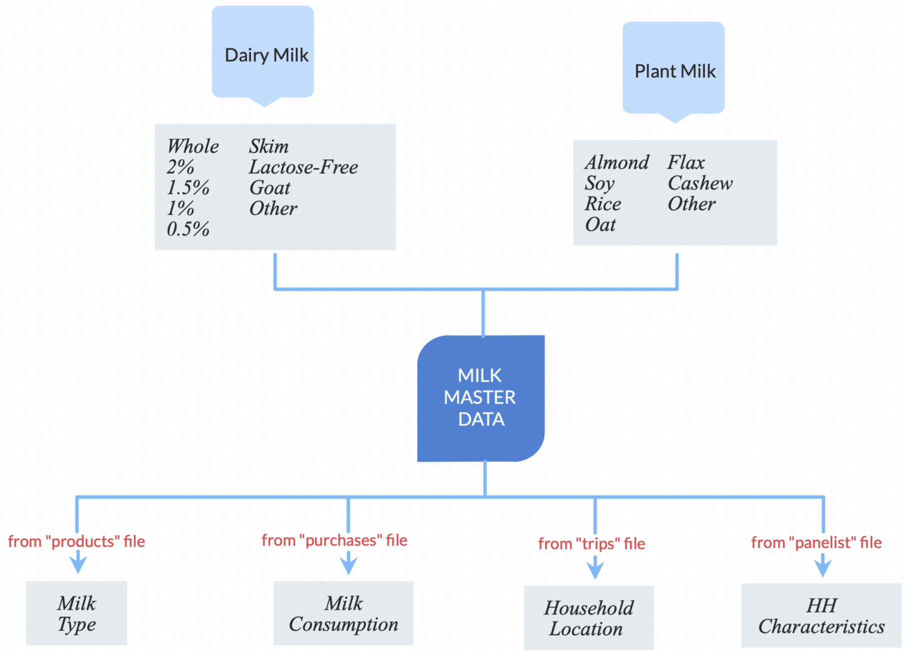
[10] Nielsen samples all 51 states and major markets, except for Alaska and Hawaii. The panelists are demographically balanced and geographically dispersed. Each panelist is assigned a unique projection factor such that each purchase is projectable to the entire U.S.

[11] The data also allows me to find movers and non-movers at county level and zip code level. The concern is that, at such a micro level, the data could be very noisy, and there is minor variation in milk purchases.

[12] Between 2006 and 2007, only soy, rice, and oat milk are identified in Nielsen data. In 2008, almond milk data was added to the data. Flax milk is not identified until the year 2011. Cashew milk has been found in the data since the year 2014.

[13] Some households who have moved more than once move between two or more states. This study excludes these

Figure 1: Data Manipulation to Obtain Master Milk Data



Notes: This data manipulation process is plotted based on her analysis.

once a year, their exact migration time cannot be identified. To avoid mismeasurement issues, I have to exclude the year of the move from the sample data[14]. These files also contain household characteristics such as household income, household size, presence of children, and race. The variables age, education, and employment status are recorded separately for male and female household heads. For this analysis, I use the household head's average age, highest education level, and highest employment status among both genders. Figure 1 illustrates a general workflow of the data manipulation process.

3.3 Outcome Variables

The main outcome variables are household-level quarterly milk expenditures per person for dairy and plant milk. First, we adjust the milk expenditures to 2019 dollars using the consumer price index for urban consumers. Then we calculate the milk expenditures per person by dividing household expenditures by the total number of persons in a household. In the main specification, the outcome

types of movers to ensure the analysis is more stable.

[14]To clarify: The main analysis is aggregated at a quarter level. Data for households who moved in 2006 is excluded since it is the base year; data for households who moved in 2019 is also excluded since we cannot observe purchases in 2020. Households who moved at any period in 2007 are defined as movers in 2008 at quarter 1; same for years between 2008 and 2018. That means between 2006 quarter 1 and 2007 quarter 4, we only observe non-movers; between 2008 quarter 1 and 2019 quarter 4, we can observe both movers and non-movers.

variable is taken logarithm of quarterly milk expenditures. We take logarithm form because the dairy and plant milk expenditures are skewed to the right. After the log transformation, the expenditure distributions become more normally distributed.

3.4 Summary Statistics

For each household, I report summary statistics for their characteristics in Table 1. The table reports the mean and standard error for each household characteristic for the mover and non-mover groups. In general, there are some differences between the two groups, but they are not too far from each other. For example, the movers are more likely to be white (non-Hispanic), have higher income, be older, and have higher education; meanwhile, movers have a smaller family size and a slightly lower employment rate. The two groups' mean marriage rate is identical in the sample.

In addition, Table 1 shows ratios of households who reside in a region or move to another region in the U.S. Again, there are some differences between the two groups, but the ranks of ratios for the four regions are the same. For example, most people move to the South region, and most non-movers reside in the South. The second most popular region people move to is the Midwest, followed by the West and Northeast regions. Same for the non-movers: the second most populous region is in the Midwest region, followed by the West and Northeast regions. For the milk purchases, we can see from Table 1 that dairy and plant milk are also similar among movers and non-movers. Based on these summary statistics, we can conclude that the movers and non-movers are generally comparable, and thus we can proceed using difference-in-differences models.

4 Empirical Strategy: Milk Consumption Patterns

This section examines changes in milk purchases after households move to another state. First, I will conduct an event study to identify the milk purchase before and after moving. Next, we will plot parallel trends with raw data. These trends help visualize the average milk expenditures among movers and non-movers over time. We will then estimate a difference-in-differences model, which shows the effect in the long run. The first differences model will then be used to study the short-term effects.

Table 1: HOUSEHOLD SUMMARY STATISTICS

	Movers (1)	Non-Movers (2)
	Demographic Characteristics	
Household Income	$72086.92	$70812.92
	($222.62)	($36.67)
Household Size	2.47	2.72
	(0.019)	(0.003)
Age	49.98	49.10
	(0.304)	(0.049)
College	0.40	0.34
	(0.023)	(0.030)
Employed	0.56	0.58
	(0.008)	(0.025)
Race: White Non-Hispanic	0.81	0.78
	(0.011)	(0.002)
Married	0.65	0.65
	(0.017)	(0.003)
	First Observed Residence	
Northeast	18.1	16.8
Midwest	22.7	25.5
South	36.7	38.4
West	22.5	19.3
	Milk Purchases	
Log(Expenditure) of Dairy Milk	0.40	0.36
Log(Expenditure) of Plant Milk	0.54	0.50
Quantity of Dairy Milk	3.84	3.82
Quantity of Plant Milk	3.48	3.44
Number of Households	4777	185230

Notes: All household characteristics are measured during the first year in the sample. Household income and milk expenditures are adjusted to 2019 dollars using the consumer price index for urban consumers. The movers are defined as those who have moved only once from 2004 to 2019. Non-movers include those households who have never moved during the sample period. The difference is statistically different from zero at a 1% significance level for all demographic characteristics.

4.1 Migration Patterns: How and Why Migrants Move

To visualize migration patterns for each state, Appendix B shows the migration maps for the top 10 most populous states in the U.S. We can see that households do move across the country, though most often they tend to move to nearby states or region[15]. Figure 2 plots separate time series plots for movers who either consume dairy or plant milk. We can observe that each year, there is a steady migration across the U.S., and the migration trend is very similar among dairy and plant milk consumers.

[15] Florida appears to be a trendy state for migrants. This phenomenon may be due to its appealing weather, scenery, and fair cost of living. Website such as My Life Elsewhere states that the U.S. is 5.8% more expensive than Florida. In terms of just milk price, Florida is more expensive. I have calculated Florida's milk price index and found that its overall milk price is higher than the U.S. average price (See Appendix F).

Figure 2: Movers Time Series by Year: Year 2004-2019

(a) Dairy Milk (b) Plant Milk

Notes: The y-axis is the accumulative fraction of movers between years 2004 and 2019.

Before conducting an event study, one concern is that households may not move randomly. If the household's decision to move is significantly influenced by, for example, their income or education level, the difference-in-differences results may not be valid since there is an endogeneity issue. I estimate a logit model to confirm that the move over time is close to random to validate the model. The logit model specification is as follows:

$$P = Pr[Ever\_Move = 1|X] = F(X'\beta) \tag{1}$$

where $F(\cdot)$ is the cdf of logistic distribution, $F(X'\beta) = \Lambda(X'\beta) = \frac{e^{X'\beta}}{1+e^{X'\beta}}$, $Ever\_Move_{it}$ is a binary outcome variable equal to one if a household has moved, and equal to zero otherwise, X is a set of household characteristics including age level, education level, income level, having children or not, race, marital status, and employment status.

Table 2 presents both coefficients and average marginal effects[16] from the logit model. A few coefficients are significant, but most of the results are not. This at least provides some evidence that households do move randomly.

[16] Marginal effects at mean (MEMs) give very similar results as average marginal effects (AMEs), hence the MEMs are not presented in the table. The marginal effects are defined as $\frac{\partial P}{\partial X_j} = F'(X'\beta)\beta_j = \Lambda(X')\beta[1 - \Lambda(X'\beta)]\beta_j = \frac{e^{X'\beta}}{1+e^{X'\beta}}\beta_j$. The MEMs are defined as $\frac{\partial P}{\partial X_j} = F'(\bar{X}\beta)\beta_j$, and the AMEs are defined as $\frac{\partial P}{\partial X_j} = \frac{\sum F'(\bar{X}\beta)}{n}\beta_j$

Table 2: Logit Model to Predict Why Households Move

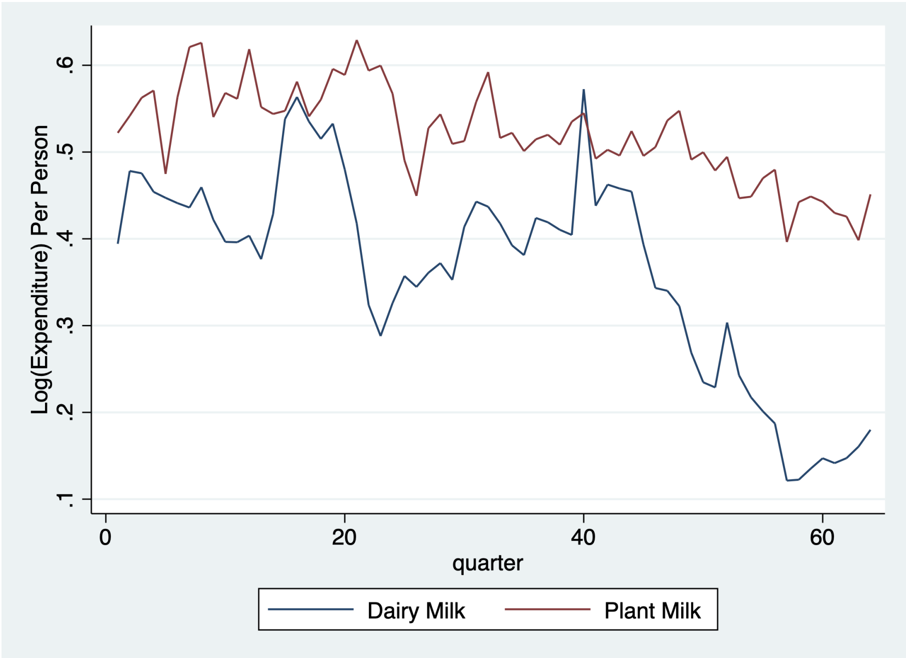
	Dairy Milk		Plant Milk	
	(1)	(2)	(3)	(4)
	Coeffs	AMEs	Coeffs	AMEs
age(young)	0.010	0	0.258	0.013
	(0.059)	(0.003)	(0.304)	(0.017)
age(midL)	-0.508	-0.015	-0.169*	-0.008
	(0.515)	(0.012)	(0.099)	(0.005)
age(midH)	0.183	0.008	-0.181	-0.008
	(0.184)	(0.008)	(0.166)	(0.007)
age(old)	-0.140	-0.005	-0.017	-0.001
	(0.308)	(0.010)	(0.135)	(0.006)
children	-0.080	-0.003	-0.143	-0.007
	(0.053)	(0.002)	(0.178)	(0.009)
no children	0.058	0.002	0.272	0.014
	(0.098)	(0.004)	(0.167)	(0.009)
educ(no college)	0.034	0.001	-0.027	-0.001
	(0.218)	(0.008)	(1.065)	(0.047)
educ(college+)	0.031	0.001	0.163	0.008
	(0.037)	(0.001)	(0.163)	(0.008)
income(low)	-0.004	0	-0.278**	-0.012
	(0.049)	(0.002)	(0.101)	(0.004)
income(med)	-0.372	-0.011	-0.046	-0.002
	(0.234)	(0.006)	(0.080)	(0.004)
income(high)	0.065	0.003	0.057	0.003
	(0.147)	(0.006)	(0.131)	(0.006)
race(white)	-0.104	-0.004*	0.02	0.001
	(0.066)	(0.002)	(0.095)	(0.004)
race(black)	-0.018	-0.001	-0.171	-0.007
	(0.039)	(0.001)	(0.152)	(0.006)
race(asian)	0.009	0	0.245*	0.011
	(0.068)	(0.002)	(0.128)	(0.005)
single	0.107***	0.004	0.175	0.009
	(0.037)	(0.001)	(0.117)	(0.006)
married	-0.065	-0.002	-0.039	-0.002
	(0.054)	(0.002)	(0.125)	(0.006)
employed	0.078	0.003	-0.231	-0.010
	(0.049)	(0.002)	(0.229)	(0.009)
unemployed	-0.073	-0.003	-0.166	-0.008
	(0.048)	(0.002)	(0.141)	(0.007)
non-movers	183266	183266	70804	70804
movers	4762	4762	2732	2732
observations	3323626	3323626	434264	434264

Notes: This logit model uses all 16 years of data. For movers, all household characteristics are from the first year they moved; For non-movers, all HH chars are from the first year of the sample. The household characteristics are groups into smaller bins for easier interpretation. The age(young) group include people who are under 34 years old, age(midL) refer to people who are between 35 to 49 years old, age(midH) include people between 50 to 64 years old, and age(old) are people who are beyond 65 years old. The children group includes only those who are under 18 years old. The no college group refers to those who hold a high school degree or lower, and the college+ group includes people who completed some college and beyond. The low-income group refers to those who earn lower than \$39,999 annually, the median income group earns between \$40,000 and \$69,999 per year, and the high-income group earns \$70,000 and more each year. *** Significant at 1 percent level.** Significant at 5 percent level.* Significant at 10 percent level.

4.2 Parallel Trend

Various reports and many news media have reported that dairy milk sales have been trending downwards since the 2000s (USDA, 2021). I use Nielsen data and plots Figure 3 to demonstrate this overall trend. We can see that the dairy milk expenditures have decreased drastically over time, whereas the plant milk expenditures have been (at least) steady over time. This trend is consistent with the current findings.

Figure 3: General Trend for Dairy and Plant Milk: Year 2004-2019 (at quarterly level)

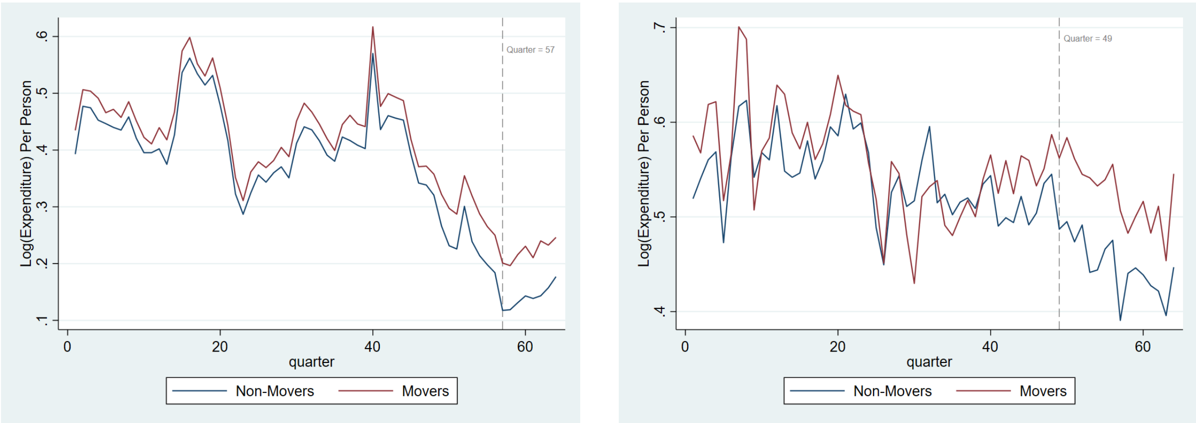


Notes: The x-axis is the accumulative quarters over years 2004 and 2019 (64 quarters in total), and the y-axis is the outcome variable: logarithm of quarterly expenditure per person for each type of milk.

However, in a difference-in-differences setting, a fundamental assumption requires no time-variant household-specific unobservables among the movers and non-movers. This assumption is the parallel trend assumption. Figure 4 compares the parallel trend for both dairy and plant milk. The trend for dairy milk somewhat exists (though some points do intersect), then there is a divergence starting from quarter 57 (or year 2018 Q1). The trend for plant milk is not very obvious. There are multiple periods that the lines intersect for movers and non-movers. There is a sharp divergence between movers and non-movers from quarter 49 (2016 Q1), and this divergence is much more significant than dairy milk.

Note that this parallel trend is plotted with raw data. In this case, the parallel trend assumes that the only control group is non-mover. This assumption does not hold here since people have moved at different periods in the sample. The difference-in-differences are comparisons between movers and non-movers, early movers and late movers, and late movers and early movers. Thus showing the comparison with only non-movers is not enough.

Figure 4: Parallel Trend: Year 2004-2019 (at quarterly level)



(a) Dairy Milk (b) Plant Milk

Notes: The x-axis is the accumulative quarters over years 2004 and 2019 (64 quarters in total), and the y-axis is the outcome variable: logarithm of quarterly expenditure per person for each type of milk. The gray dashed line marks some unusual trend. For dairy milk, before quarter 57 (2018 Q1), the general trend is downward, then there is an uptick after that. For plant milk, before quarter 49 (2016 Q1), the general trend is almost monotonic, then it goes downward.

4.3 Event Study

In the earlier section, we have seen evidence that the move occurs randomly across time. Moreover, the parallel trend gives us a visualization of the general trend for milk expenditures. Next, we can use event study to estimate and plot the post-move effects.

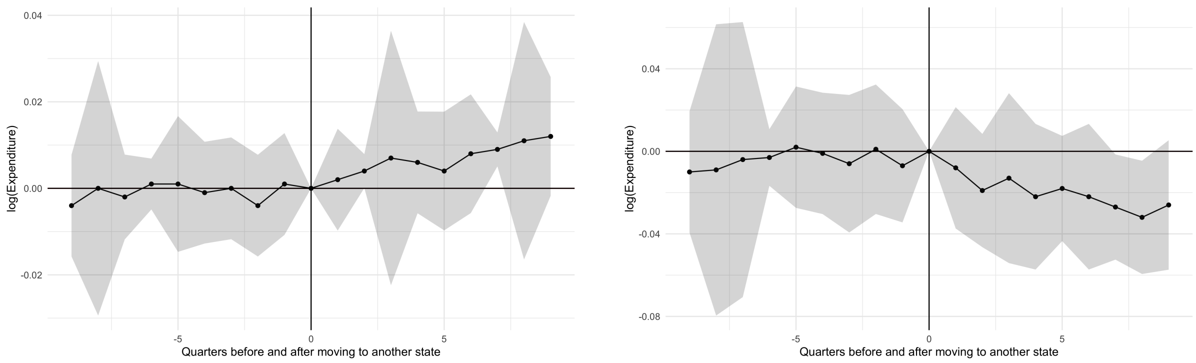
I estimate the following event study model for household i for period t:

$$ln(y)_{it} = \alpha_i + \tau_t + \sum_{r=\underline{m}, r\neq -1}^{\overline{m}} \beta_r \cdot I_{it} + X_{it}'\Gamma + \epsilon_{it} \tag{2}$$

where $ln(y)_{it}$ is the log quarterly milk expenditures per person for household i in quarter t, α_i and τ_t are household and time fixed effect respectively, I_{it} is an indicator variable set equal to one if household i in quarter t is r months after it has moved to another state, X_{it} is a set of controls for household characteristics. The household characteristics specifications are the same as in Table 2. The error term ϵ_{it} captures unobservables.

Additional indicators corresponding to "outside the event window"[17] are also added, which helps fully capture the dynamic effects of the treatment. By conditional on household fixed effects and quarter-specific shocks, we assume that all households that are r quarters away from moving to another state are identical. Here we allow $|\underline{m}| = \overline{m} = 8$, which includes 2 years pre-move and post-move sample data. The specification also includes an indicator variable +9 that captures all periods more than 8 quarters from moving to another state; for similar reasons, indicator -9 captures all periods more than 8 months before the move. Event study helps visualize the effect of moving to another state on milk expenditures and checking that all pre-trend event time, $\sum_{r=-9}^{r=-2}$, are equal to zero. The coefficients of post-trend event time, $\sum_{r=0}^{r=9}$ visualized the evolution of the treatment effect.

Figure 5: Event Study for Milk Expenditures (at quarter level)



(a) Dairy Milk (b) Plant Milk

Notes: The x-axis is the move's relative time (at a quarterly level). For example, 0 indicates the quarter of the move, -1 the quarter before the move, and 1 the quarter after the move. The data of the moving year is excluded to avoid mismeasurement issues. For example, if a household moved in 2006, "the quarter of the move" is defined as 2007 Q1.

Figure 5 are the plots for the event study with 95% confidence intervals. The pre-move coefficients are close to zero, though some standard errors appear to be significant. After the move, there was a 0.8% increase in expenditures for dairy milk and a 0.1% decrease for plant milk. The p-value from the joint significance test of the pre-move event time estimates is 0.0073 for dairy milk and 0.12 for

[17] This design draws inspiration from Harding and Rapson (2019).

plant milk. The p-value for the plant milk is only weakly significant; this still lends some evidence that there are only minor differences between the movers and non-movers before the move. The plots suggest a clear pattern: Prior to the move, there is little evidence of divergence in expenditures for both types of milk. However, after households move to a new state, they increase their dairy milk expenditures and decrease plant milk expenditures. We need to turn to the difference-in-differences model to quantify the exact expenditure changes.

4.4 Difference-in-Differences

The difference-in-differences model compares milk expenditures across households and time relative to when they move. We use the difference-in-differences model because we have data for the control group (non-movers) and the treated group (movers). In addition, we can observe data at least two years prior to the treatment for the movers. The model specification is as follows:

$$ln(y)_{it} = \alpha_i + \tau_t + \beta D_{it} + X_{it}'\Gamma + \epsilon_{it} \tag{3}$$

where D_{it} is an indicator variable which equals to one if household i has moved to another state in quarter t. The difference-in-differences specification is the same as Equation 2, except for all the coefficients before the move (r < 0) are normalized to zero, and the coefficients after the move (r $\geq$ 0) are normalized to one. β is our coefficient of interest. It uncovers the average treatment effect on the treated households (ATT), where $ATT = E[\beta|D = 1]$.

Table 3 column (1) $-$ (3) are the results for dairy milk, and column (4) $-$ (6) are for plant milk. The effect for dairy milk is only 5% (not statistically significant) if we do not control for the number of children and race in the model. When we do control for these additional variables, however, the effect more than doubles: 1.2% (and are statistically significant at 5% level) of increase in dairy milk expenditures after the move. This result makes sense since there will be more milk consumption if a family has many children. Besides, dairy milk is still the more popular type consumed among households with children, and the price for dairy milk, in general, is lower than plant milk[18]. These facts all make dairy milk a naturally popular choice. In addition, the household characteristics in this model include all levels for each variable. This additional information should give us more precise

[18]For price comparisons between dairy and plant milk, see Appendix F.

Table 3: Milk Expenditures After Move. Difference-in-Differences Estimates.

	Dairy Milk			Plant Milk		
	(1)	(2)	(3)	(4)	(5)	(6)
Post move	0.005	0.012**	0.012**	-0.015*	-0.015*	-0.015*
	(0.007)	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)
Age	Yes	Yes	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status	Yes	Yes	Yes	Yes	Yes	Yes
Employment	Yes	Yes	Yes	Yes	Yes	Yes
Children		Yes	Yes		Yes	Yes
Race			Yes			Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Household FEs	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	HH	HH	HH	HH	HH	HH
Households	188028	188028	188028	73536	73536	73536
Observations	3323626	3323626	3323626	434264	434264	434264

Notes: The household characteristics here use full levels from sample data, instead of just putting them into sub-groups. For example, there are 9 levels of "age": 1 signifies households under 25 years old, 2 signifies households between 25-29 years old, and so on. Including all levels for each variable is to ensure the precision of the difference-in-differences estimates. The model includes time fixed effects and household fixed effects. The standard errors are clustered at the household level. *** Significant at 1 percent level.** Significant at 5 percent level.* Significant at 10 percent level.

estimates[19]. There may also be cultural differences at play. For example, different races may have different milk consumption habits. If we control for the "race" factor too, it should give us a more convincing estimate.

On the other hand, we see a 1.5% decrease in plant milk expenditures (statistically significant at 10% level) on average after the move. The estimates are the same even when not controlling for the number of children and race. These results are surprising since dairy milk sales are trending downward, but dairy milk is consumed more here based on our estimates. One may argue that this decrease in plant milk expenditures could be because the households have lower income or education, but it has been proven not the case (See Table 1). The movers' characteristics (i.e., higher income and education) suggest that movers should consume more plant milk since it has higher prices (see Appendix F) and fits a healthier and more sustainable lifestyle.

Appendix C gives difference-in-differences results using movers data only. The estimate for dairy milk is the same as using a full sample. This result may be because dairy milk data is a popular type of milk (i.e., its consumption is more stable at the national level), and it has a large number of

[19]For the results from the earlier logit model (see Table 2), each variable is put into smaller groups. For example, the variable "income" has 20 levels in the data. I have to group them so that I can fit the logit model results in one table.

households and observations, so even using just movers data, the estimate is very close (identical in this case) to the estimate using the whole sample. However, if we use only movers data for plant milk, the estimate becomes very small: a 6% decrease only in its expenditures (and not significant). The reason may be due to a considerable reduction in households and observations. Such a decrease in sample size may well affect our estimate.

To conclude, these results reveal one truth: even though dairy milk sales have been decreasing over the years, it is still the dominating type of milk consumed in the U.S. Another way to interpret it is that people's preferences for dairy milk are persistent at the U.S. national level, and it is not easy to change from dairy to plant milk. In Section 5, we will address regional milk preferences, and we will see that people's milk preferences may adjust to a new region after the move. However, their preferences do not change at the national level here because they become negligent or nonexistent when we average the effects.

4.5 First Differences

It is also interesting to see the first differences model's short-term treatment effects. As mentioned earlier, the first quarter of the move is not the quarter right after the move[20] but the first quarter of the following year after a household moves. To estimate the short-run effects, I define the following first differences model:

$$dln(y)_{it} = \tau_t + \beta D_{it} + X'_{it}\Gamma + \epsilon_{it} \tag{4}$$

where $dln(y)_{it} = ln(y)_{i,t} - ln(y)_{i,t-1}$. The term α_i from Equation 3 is dropped from the differencing since these household characteristics are time invariant.

Table 4 present the first-differences results. Same as before, column (1) – (3) are the results for dairy milk, and column (4) – (6) are for plant milk. The estimates for dairy milk vary between 6% and 9%. All of them are at least significant 5% level. The overall effect for dairy milk expenditures is reduced to half compared to the difference-in-differences estimates: just a 0.6% increase if we add all control variables. This result means that the immediate increase after the move in dairy milk expenditures is only half the size compared to the long-term change. There may be several reasons.

[20]At the beginning of the move, the households may need time to explore new surroundings to find their regular grocery stores for shopping, and their shopping record may not be as accurate.

Table 4: Milk Purchases After Move. First Differences Estimates.

	Dairy Milk			Plant Milk		
	(1)	(2)	(3)	(4)	(5)	(6)
Post move	0.009***	0.005**	0.006**	-0.020**	-0.019***	-0.020**
	(0.003)	(0.002)	(0.002)	(0.007)	(0.005)	(0.005)
Age	Yes	Yes	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status	Yes	Yes	Yes	Yes	Yes	Yes
Employment	Yes	Yes	Yes	Yes	Yes	Yes
Children		Yes	Yes		Yes	Yes
Race			Yes			Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Household FEs	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	HH	HH	HH	HH	HH	HH
Households	188028	188028	188028	73536	73536	73536
Observations	3323626	3323626	3323626	434264	434264	434264

Notes: The household characteristics here use full levels from sample data, instead of just putting them into sub-groups. For example, there are 9 levels of "age": 1 signifies households under 25 years old, 2 signifies households between 25-29 years old, and so on. Including all levels for each variable is to ensure the precision of the difference-in-differences estimates. The model includes time fixed effects and household fixed effects. The standard errors are clustered at the household level. *** Significant at 1 percent level.** Significant at 5 percent level.* Significant at 10 percent level.

First, when households just moved, they may need time to adjust to the local environment (such as finding the right grocery stores). This move may disrupt their shopping record. Second, the household in the sample may have more children (or have more people in the household in general) in the long run, thus dairy milk consumption should be higher in the longer term.

After the move, there is a more significant reduction in plant milk expenditures: a 2% decrease (statistically significant at a 5% level) compared to just a 1.5% decrease in the difference-in-differences model. The magnitude of the change is much smaller than dairy milk (25% change for plant milk, compared to 50% change for dairy milk). These results imply that changes in plant milk expenditures among movers are gradual but not drastic, whether in the short term or long term. We can see that pattern in Figure 4b. I suspect that the decrease in plant milk expenditures may be related to the availability of different types of plant milk or just the general availability of plant milk. For example, between 2006 and 2007, only soy, rice, and oat milk were available in Nielsen data. Almond milk data was available only until 2008. Flax milk was identified in 2011, and Cashew milk was found in 2014. Smaller availability could limit our plant milk sample size in terms of the total number of movers (or households) and observations[21].

[21] The evidence of the limited plant milk availability is evident when we compare the number of plant milk consumers

Same as in the difference-in-differences model, we provide first-differences estimates but use movers data only. The results are presented in Appendix D. Dairy milk has a larger effect: a 1.4% increase (significant at 1% level). However, this effect is very close to the difference-in-differences estimate using the whole sample. Again, I would argue that this may be because dairy milk is a well-accepted type of milk, so the estimate is more stable using the full sample or just movers data. Again, the estimate for plant milk is small (just 3%) and insignificant. The reason might be the smaller sample size and lesser plant milk availability.

To conclude, the general trend for dairy and plant milk still conforms to results from the difference-in-differences model. These estimates again imply that dairy milk is still the dominant type consumed in the U.S., and people are not changing their milk preferences as easily.

4.6 Heterogeneous Effects

Differences-in-differences and first differences give an overall effect. To take a closer look at the effects among households of different backgrounds (e.g., different income levels, education, and age), we need to get heterogeneous effects. I estimate Equation 2 by interacting I_{it} with same household characteristics as in Table 2. Appendix E shows that the increase in dairy milk expenditures is mostly driven by the people who are between 50 and 64 years old (the "midH" group), whose families have no children living in the households, or people who earn a lower income. People under 49 years old group (the "young" and "midL" groups), on the other hand, are lowering their dairy milk consumption. The decrease in plant milk expenditures is mostly driven by the "college" group (those who have had some college education or completed college education). They consume less maybe because this group needs time to earn a higher income to pick up a new lifestyle since plant milk is more expensive than dairy milk. The "college" group recorded in the Nielsen data also includes those exposed to but have not completed the college education. The inclusion of such a group of people may also have influenced the plant milk estimates.

I then estimate Equation 3 and 4 by interacting D_{it} with the same set of household characteristic variables. Table 5 and Table 6 present the heterogeneous treatment effects from the difference-in-differences model. Table 7 and Table 8 are heterogeneous first differences estimates. Note that in these four

in Table 2 and Table 4. There are 4762 movers for dairy milk but only 2732 for plant milk. The dairy milk has 55% more households and more than 6 times observations than plant milk.

tables, the base specifications are the same specifications as in Table 3 and Table 4. That is why their estimates are identical[22].

We can see that some results are consistent with the event study results in Appendix E. For example, in Table 5, the difference-in-differences estimates show that younger households (who are under 49 years old) are consuming less dairy milk, while people who are between 50 and 64 years old, have no child in the household, and have lower income, increase their dairy milk consumption. These results conform to the results presented in Appendix C, where the estimates are obtained using movers data only. We can also see that households with children under 12 years old also consume more dairy milk, and people with children above 12 years old or those who have a graduate education also tend to decrease their dairy milk consumption. These differences arise maybe because the difference-in-differences model uses all periods, whereas the event study only includes a year prior to and after the treatment. The short-term heterogeneous effects presented in Table 7 are very similar to Table 5 in terms of signs (increasing or decreasing) for each sub-group.

The estimates for plant milk are more different than the standard difference-in-differences results. In Table 6, it appears households who have children above 12 years old and are black consume less after the move. Households consume more plant milk if they have no children, have a high school degree, or have medium to high-level income. Right after the move, the immediate effect is that people under 34 years old, between 50-64 years old, black and Asian, decrease plant milk consumption. On the other hand, households between 35-49 years old, who have no children, have a high school degree or are white, increase their plant milk consumption (See Table 8). These results are more different compared with Appendix C. These estimates vary more than dairy milk using different models because plant milk sample size is small, and some varieties of plant milk appear in the market (thus in the data) later. That is why the estimates in plant milk vary more often if we change the models.

Overall, these results lend some evidence that a new lifestyle (i.e., healthier and more sustainable) is emerging among younger and more educated households, as reported in the news. However, some estimates are not very consistent, especially among plant milk. This motivates me to study milk consumption among households in different regions since regional consumption may be more consistent and stable. In Section 5, we are going to estimate geographic variations in milk consumption, instead of just getting average effects at national level.

[22] I listed the base specifications here again just for a reference.

Table 5: Dairy Milk Expenditure Patterns After Move. DiD Heterogeneous Effects.

	(1)	(2)	(3)	(4)	(5)	(6)
Base specification	0.012***					
	(0.006)					
Age: young		-0.120***				
		(0.029)				
Age: midL		-0.066***				
		(0.015)				
Age: midH		0.024**				
		(0.010)				
Age: old		-0.020				
		(0.013)				
Children: <12			0.050**			
			(0.023)			
Children: ≥ 12			-0.114***			
			(0.021)			
Children: mixed			0.037			
			(0.025)			
Children: none			0.139***			
			(0.022)			
Educ: HS				0.023		
				(0.019)		
Educ: college				0.004		
				(0.009)		
Educ: grad				-0.046***		
				(0.016)		
Income: low					0.053***	
					(0.014)	
Income: med					0.020	
					(0.013)	
Income: high					-0.031***	
					(0.010)	
Race: white						-0.003
						(0.008)
Race: black						-0.005
						(0.030)
Race: asian						-0.013
						(0.042)
Households	188028	188028	188028	188028	188028	188028
Observations	3323626	3323626	3323626	3323626	3323626	3323626

Notes: This analysis uses the whole sample. The dependent variable is log(expenditure) for dairy milk. It includes both time fixed effects and household fixed effects. The standard errors are clustered at a household level. For the household characteristics, the age and income group definitions are the same as in Table 2. The household characteristics are groups into smaller bins for easier interpretation. Children (< 12) refer to children who are under 12 years old, and Children (≥ 12) are for those who are 12 years old or older. If a family has children below and above 12 years old, they are placed in the mixed-age group. If a family has no children under 18 years old, they are in the no children group. The high school group includes those who hold high school degrees and lower. The college group includes those who have completed some college education or hold a college degree. The grad group only includes those who have a post-college education. To see results using data only on movers, see Appendix C. The model includes time fixed effects and household fixed effects. The standard errors are clustered at the household level. *** Significant at 1 percent level.** Significant at 5 percent level.* Significant at 10 percent level.

Table 6: Plant Milk Expenditure Patterns After Move. DiD Heterogeneous Effects.

	(1)	(2)	(3)	(4)	(5)	(6)
Base specification	-0.015*					
	(0.008)					
Age: young		-0.072				
		(0.060)				
Age: midL		-0.037				
		(0.031)				
Age: midH		-0.006				
		(0.024)				
Age: old		0.019				
		(0.028)				
Children: <12			0.008			
			(0.045)			
Children: ≥ 12			-0.100**			
			(0.039)			
Children: mixed			0.047			
			(0.049)			
Children: none			0.117***			
			(0.041)			
Educ: HS				0.081*		
				(0.046)		
Educ: college				-0.015		
				(0.018)		
Educ: grad				-0.022		
				(0.030)		
Income: low					0.031	
					(0.030)	
Income: med					0.050*	
					(0.026)	
Income: high					0.045**	
					(0.019)	
Race: white						0.003
						(0.017)
Race: black						-0.095**
						(0.044)
Race: asian						-0.023
						(0.066)
Households	73536	73536	73536	73536	73536	73536
Observations	434264	434264	434264	434264	434264	434264

Notes: The model includes time fixed effects and household fixed effects. The standard errors are clustered at the household level. *** Significant at 1 percent level.** Significant at 5 percent level.* Significant at 10 percent level.

4.7 Robustness Checks

We also perform a placebo test[23] to support the claim that the unobservables do not affect the results. For the difference-in-differences placebo test, I follow the following procedure:

1. Use only the data that came before the treatment went into effect (i.e., before households moved

[23] This method is widely used. For example, Cheng and Hoekstra (2013) adopts this same method: They performed inference using placebo estimates from pre-castle doctrine (pre-treatment only) data.

Table 7: Dairy Milk Expenditure Patterns After Move. First Differences Heterogeneous Effects.

	(1)	(2)	(3)	(4)	(5)	(6)
Base specification	0.006**					
	(0.002)					
Age: young		-0.103***				
		(0.014)				
Age: midL		-0.085***				
		(0.007)				
Age: midH		0.049***				
		(0.004)				
Age: old		-0.043***				
		(0.006)				
Children: <12			-0.015			
			(0.013)			
Children: $\geq$12			-0.005			
			(0.010)			
Children: mixed			-0.006			
			(0.014)			
Children: none			0.031***			
			(0.010)			
Educ: HS				0.063***		
				(0.008)		
Educ: college				0.003		
				(0.004)		
Educ: grad				-0.030***		
				(0.006)		
Income: low					0.013**	
					(0.006)	
Income: med					-0.013**	
					(0.006)	
Income: high					0.003	
					(0.004)	
Race: white						0.007**
						(0.003)
Race: black						-0.056***
						(0.010)
Race: asian						0.183***
						(0.016)
Households	188028	188028	188028	188028	188028	188028
Observations	3323626	3323626	3323626	3323626	3323626	3323626

Notes: This analysis uses the whole sample. The dependent variable is log(expenditure) for dairy milk. It includes both time fixed effects and household fixed effects. The standard errors are clustered at the household level. To see results using data only on movers, see Appendix D. *** Significant at 1 percent level.** Significant at 5 percent level.* Significant at 10 percent level.

to another state).

2. Randomly assign a fake treatment period. Here, the fake treatment is assigned 2 quarters prior to the real treatment.

3. Estimate the same difference-in-differences model as in Section 4.4, but create a new treatment

Table 8: Plant Milk Expenditure Patterns After Move. First Differences Heterogeneous Effects.

	(1)	(2)	(3)	(4)	(5)	(6)
Base specification	-0.020***					
	(0.005)					
Age: young		-0.078**				
		(0.034)				
Age: midL		0.049***				
		(0.011)				
Age: midH		-0.084***				
		(0.018)				
Age: old		-0.001				
		(0.017)				
Children: <12			-0.003			
			(0.032)			
Children: ≥ 12			-0.012			
			(0.026)			
Children: mixed			0.052			
			(0.039)			
Children: none			0.054**			
			(0.026)			
Educ: HS				0.087***		
				(0.024)		
Educ: college				0.007		
				(0.009)		
Educ: grad				-0.014		
				(0.016)		
Income: low					0.013	
					(0.017)	
Income: med					0.031	
					(0.017)	
Income: high					0.002	
					(0.010)	
Race: white						0.034***
						(0.008)
Race: black						-0.113***
						(0.003)
Race: asian						-0.202***
						(0.005)
Households	73536	73536	73536	73536	73536	73536
Observations	434264	434264	434264	434264	434264	434264

Notes: The model includes time fixed effects and household fixed effects. The standard errors are clustered at the household level. *** Significant at 1 percent level.** Significant at 5 percent level.* Significant at 10 percent level.

variable equal to one if the households are in the treated group and zero otherwise.

4. If there is no "effect" for the fake treatment (where there should not be one), there is evidence to support the claim that there is some effect on the expenditures after the move. The parallel trend assumption would also hold.

Table 9: Placebo Tests on Movers

	Dairy Milk		Plant Milk	
	DiD	FD	DiD	FD
Placebo Treatment	0.004	0.004	-0.017	-0.012
	(0.005)	(0.006)	(0.012)	(0.013)
Time FEs	Yes	Yes	Yes	Yes
HH FEs	Yes	Yes	Yes	Yes
Clustered SEs	HH	HH	HH	HH
Observations	3263218	3263218	424055	424055

Notes: I assigned fake treatments 6 months prior to the actual treatment in this placebo test. The model includes time fixed effects and household fixed effects. The standard errors are clustered at the household level. *** Significant at 1 percent level.** Significant at 5 percent level.* Significant at 10 percent level.

Table 9 shows the placebo test results. After I drop all data after the actual treatment and then pretend that the treatment occurred two quarters earlier, I find no difference-in-differences effect. For example, both difference-in-differences and first differences are 0.4% for dairy milk. The estimate is close to zero and not statistically significant. The plant milk estimates appear to be slightly larger: between 1.2% and 1.7% of decrease, but both results are not statistically significant. One may still question the robustness of the estimates at this point. At the beginning of the study, I have used a logit model to prove that the households move across states at random. The covariates included in these models are very related to the milk consumption behavior and should explain the consumption variations well. Another concern is the confounder issue. I cannot think of other confounders that could affect milk consumption (at least given the scope of the data availability). I would conclude that these estimates are robust and can explain the average milk consumption at a national level.

5 Empirical Strategy: How Current Environment Affects Milk Consumption

Section 4 studies expenditure patterns for both types of milk. After moving to a new state, we have learned that households increase their dairy milk expenditures and decrease plant milk expenditures. How much of these variations are actually due to the current environment? This section explains these geographic variations using the same empirical strategy but with some modifications. As a convention, we will first conduct an event study to identify the milk expenditures before and after the move. We will then estimate a difference-in-differences model, which shows the effect in the long run.

5.1 Event Study

As usual, we start this section by defining an event study model. For household i for period t, the event study model is specified as follows:

$$ln(y)_{it} = \alpha_i + \tau_t + \sum_{r=\underline{m}}^{\overline{m}} \theta_r \cdot I_r \cdot \Delta_i + X_{it}'\Gamma + \epsilon_{it} \tag{5}$$

where $ln(y)_{it}$ is the log quarterly milk expenditures per person for household i in quarter t, α_i and τ_t are household and time fixed effect respectively, I_r is an indicator variable set equal to one if household i in quarter t is r months after it has moved to another state, X_{it} is a set of controls for household characteristics. The error term ϵ_{it} captures unobservables. The quarter of the move is indexed by 0. The coefficient θ_{-1} on the last quarter (before the move) is normalized to zero. The standard errors are clustered at the household level.

Term $\Delta_i = \bar{y}_{N,i} - \bar{y}_{O,i}$ is the size of the move[24]. It measures the share of the average difference between the movers' new state and the original state due to the current environment. This study follows this procedure to obtain the variable "size of the move": Use non-movers data only to compute state-level milk expenditure averages. First, average milk expenditures across households and quarters for each state (using sample weights). Then, average expenditures across all quarters. We can think of this as the expenditures of self-regressing on expenditure differences of others (who are in your destination states or origin states). As in Section 4.3, additional indicators corresponding to "outside the event window" are also added.

The coefficients θ_r measures how much milk expenditures converge towards the new states' average. If movers' milk expenditures are not changed, the coefficient θ_r is zero. If movers' milk expenditures converge towards the new states, θ_r equals one. In reality, θ_r would be between 0 and 1. It measures the percentage difference in milk consumption between the new state and the original state.

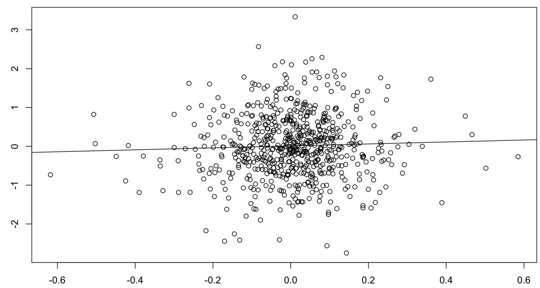
One assumption needed for the event study is that the pre-move trend in the milk expenditures is uncorrelated with the size of the move. We provide evidence by scatter plots. In Figure 6, plot (a) shows the pre-move changes for dairy milk, where the x-axis is the average size of move, the y-axis is the pre-move changes in expenditures, which is the difference of expenditures between the last quarter

[24] The empirical strategy is from Finkelstein et al. (2016). It shows how the coefficient measures the share of variation explained by the location.

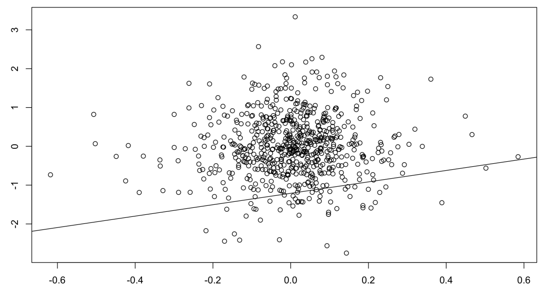
before the move and 3 quarters before the move ($y_{-1} - y_{-3}$). The slope is 0.2541 with a standard error of 0.2469, which is not significant. Plot (b) is for the post-move changes for dairy milk, where the x-axis is still the average size of move, the y-axis is the pre-post-move changes in expenditures, which is the difference of expenditures between the first quarter after the move and last quarter before the move ($y_1 - y_{-1}$). The slope is 1.4705 with a standard error of 0.2257. The result is significant at the 1% level. For plant milk, plot (c) has a slope of 0.0447 with a standard error of 0.2602, which is not significant. Plot (d) gives a slope is 0.3949 with a standard error of 0.1961. The result is significant at the 5% level. Although there may be a slight pre-trend (but not statistically significant) before the move, they are much small compared to the post-trend after the move. This lends some evidence that The event study assumption holds.

Figure 6: Changes in Milk Expenditures by the Size of Move: pre-move(a), (c) and pre-post-move(b),(d)

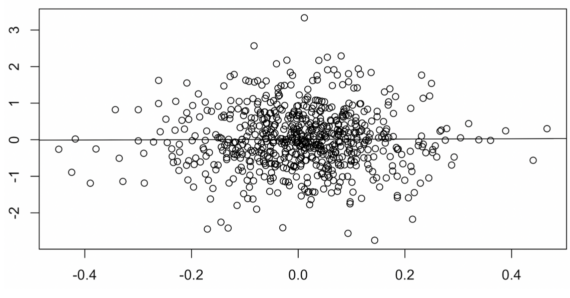
(a) Dairy Milk: Pre-Move Change



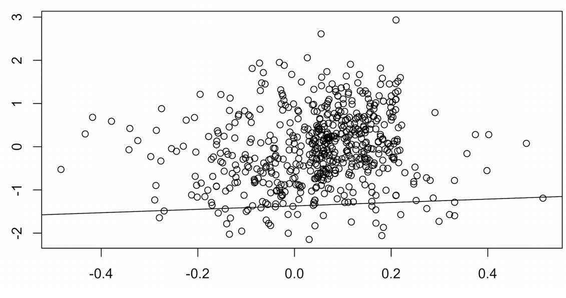
(b) Dairy Milk: Pre-Post-Move Change



(c) Plant Milk: Pre-Move Change



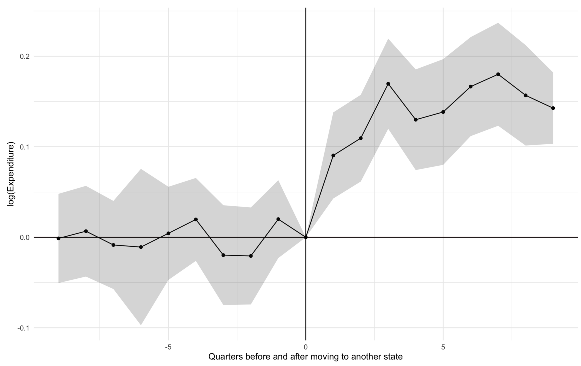
(d) Plant Milk: Pre-Post-Move Change



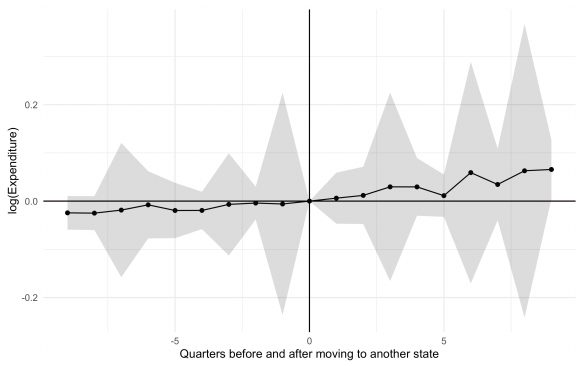
Notes: For scatter plots (a) and (c), the x-axis is the size of move, and the y-axis is the log expenditure differences between one quarter before the move and three quarters before the move (event time -1 and event time -3). For scatter plots (b) and (d), the x-axis is the size of move, and the y-axis is the log expenditure differences between one quarter after the move and one quarter before the move (event time 1 and event time -1).

Once we have proved that the trends in mover's expenditures are unlikely correlated with the "size of move," we can estimate the Equation 5 and visualize the milk expenditure changes pre-move and post-move. Figure 7 plots the results from Equation 5. The black dots connected with a black line represent all the coefficients estimated from the event study model. The shaded gray area is the estimated confidence interval. The center 0 is the time of the move. To its left, it is the pre-move trend; to its right, it is the post-move trend. For dairy and plant milk, the pre-move trends appear to be flat and close to zero (although some estimates tend to have larger confidence intervals than the others). After the move, there is a bigger upward jump for dairy milk and a smaller upward trend for plant milk. These results imply that people tend more adopt more dairy milk consumption based on the current environment than the plant milk. In Section 5.2, we are going to quantify such milk expenditure variations due to the local environments.

Figure 7: EVENT STUDY FOR MILK EXPENDITURES (AT QUARTERLY LEVEL)



(a) DAIRY MILK



(b) PLANT MILK

Notes: The x-axis is the move's relative time (at a quarterly level). For example, 0 indicates the quarter of the move, -1 the quarter before the move, and 1 the quarter after the move. The data of the moving year is excluded to avoid mismeasurement issues. For example, if a household moved in 2006, "the quarter of the move" is defined as 2007 Q1.

5.2 Difference-in-Differences

We begin this section by defining a difference-in-differences model:

$$ln(y)_{it} = \alpha_i + \tau_t + \sum_{r=\underline{m}}^{\overline{m}} \theta_r \cdot D_{it} \cdot \Delta_i + X_{it}'\Gamma + \epsilon_{it} \tag{6}$$

where D_{it} is an indicator variable which equals one if household i has moved to another state in quarter t. The difference-in-differences specification is the same as Equation 3, where all the coefficients before the move (r < 0) are normalized to zero, and the coefficients after the move (r ≥ 0) are normalized to one.

Table 10 shows the estimates for Equation 6 using all periods of data. For dairy milk, we can see about 53% (significant at 10% level) of the new state minus the original state difference in the expenditure changes after the move. For plant milk, the expenditure changes are only about 17% (significant at 5% level). These results imply a more considerable convergence for dairy milk expenditures toward the destination state's average level. In contrast, the plant milk expenditures have only slightly converged toward the average level in the households' new states. This variation may be reasonable since, in Section 4.4, we see that dairy milk is still the dominant type of milk consumed in the U.S. Another reason is that there are more movers (more observations) among dairy consumers than plant consumers since plant milk is a more recent product.

Table 10: Change in Milk Expenditures After Move.Difference-in-Differences Estimates

	Log(Expenditure) Dairy Milk (1)	Log(Expenditure) Plant Milk (2)
$\Delta \cdot$ *Post move*	0.534*	0.173**
	(0.169)	(0.076)
Time Fixed Effects	Yes	Yes
Household Fixed Effects	Yes	Yes
Clustered SE	Household	Household
Movers	4762	2732
Observations	148786	22484

Notes: Δ_i is the difference in the average logarithm of milk expenditures between the new state and the original state. Sample only includes movers who have moved only once to another state. The model includes time fixed effects and household fixed effects. The dependent variable is the logarithm of milk expenditures. Standard errors are clustered at the household level. The coefficients are significant at 1 percent level. *** Significant at 1 percent level.** Significant at 5 percent level.* Significant at 10 percent level.

The results again are estimated at a national level. We have seen the overall trend using all data from all periods in the previous and this section. We know that movers increase their dairy milk expenditures but decrease plant milk expenditures after the move. Households tend to adopt more dairy milk consumption behaviors in the new state than plant milk. At this point, one might be curious about comparing specific regions and seeing their milk consumption behavior, given that the regional economy and culture can be vastly different.

Table 11: Dairy Milk Purchases After Move (Movers Between Midwest and Every Other Region). Difference-in-Difference Estimates.

	Dairy Milk: Log(Expenditure)					
	Origin: Midwest			Destination: Midwest		
	West	Northeast	South	West	Northeast	South
	(1)	(2)	(3)	(4)	(5)	(6)
Post move	0.013	0.127**	0.080***	-0.016	-0.020	-0.034*
	(0.023)	(0.127)	(0.017)	(0.035)	(0.041)	(0.020)
Age	Yes	Yes	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status	Yes	Yes	Yes	Yes	Yes	Yes
Employment	Yes	Yes	Yes	Yes	Yes	Yes
Children	Yes	Yes	Yes	Yes	Yes	Yes
Race	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Household FEs	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	HH	HH	HH	HH	Yes	Yes
Households	195	53	502	147	65	286
Observations	5948	1503	16609	4194	2074	8547

Notes: These estimates are re-estimated for Equation 3 using data for each region. All household characteristic specifications are the same as in 3. The model includes time fixed effects and household fixed effects. The standard errors are clustered at the household level. *** Significant at 1 percent level.** Significant at 5 percent level.* Significant at 10 percent level.

Table 12: Plant Milk Purchases After Move (Movers Between Midwest and Every Other Region). Difference-in-Difference Estimates.

	Plant Milk: Log(Expenditure)					
	Origin: Midwest			Destination: Midwest		
	West	Northeast	South	West	Northeast	South
	(1)	(2)	(3)	(4)	(5)	(6)
Post move	0.064	0.644**	0.009	-0.091	-0.327	-0.024
	(0.052)	(0.182)	(0.040)	(0.071)	(0.466)	(0.036)
Age	Yes	Yes	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status	Yes	Yes	Yes	Yes	Yes	Yes
Employment	Yes	Yes	Yes	Yes	Yes	Yes
Children	Yes	Yes	Yes	Yes	Yes	Yes
Race	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Household FEs	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	HH	HH	HH	HH	Yes	Yes
Households	108	27	272	72	29	147
Observations	966	186	2129	443	134	1221

Notes: These estimates are re-estimated for Equation 3 using data for each region. All household characteristic specifications are the same as in 3. The model includes time fixed effects and household fixed effects. The standard errors are clustered at the household level. *** Significant at 1 percent level.** Significant at 5 percent level.* Significant at 10 percent level.

In Table 11 and Table 12, I present results for Equation 3 using data for each specific region. For each table, columns (1) - (3) are the estimates for households who move from Midwest to West, Northeast, or South. Columns (4) - (6) are results for those who move from the opposite direction: from West, Northeast, or South to Midwest. Here is the rationale why I use Midwest as the base region (or as a reference): Multiple sources point out that Midwest has the lowest plant milk sales volume (Good Food Institute, 2020) (The Good Food Institute, 2020) (Ipsos Retail Performance, 2022). In this case, it may be more interesting to see how movers from Midwest to the other regions change their dairy and plant milk consumption. Likewise, it is worth investigating how households who move from relatively higher plant milk consumption regions[25] to a low consumption region (Midwest) change their milk consumption.

Looking at households who move from Midwest to the other regions, they all increase dairy milk expenditures (1.3%, 12.7% [significant at 5% level], and 8% [significant at 1% level] respectively). The increases are substantial for the Northeast and South regions. Moreover, if people from other regions move to Midwest, they appear to decrease dairy milk consumption (1.6%, 2%, and 3.4% [significant at 10% level] respectively). The changes in columns (4) - (6) are much smaller than the changes in columns (1) - (3) on average. These estimates s that dairy milk is still the dominant type in the U.S., and people are not decreasing that much dairy milk expenditures even moving to a new environment. In an overall picture, one can interpret these results for dairy milk as follows: People who migrate from a lower plant milk consumption region or dairy milk dominant region (Midwest in this case) to a higher plant milk consumption region may still maintain their old consumption behavior. That is to say, if their original dairy milk consumption is high, they will continue to consume more dairy milk even after moving to a region where dairy milk is less prevalent. This also works the other way around: Households who move from a higher plant milk consumption region or less dairy milk dominant region (in this case, every region other than Midwest) to a lower plant milk consumption region may persistently consume less dairy milk. A downside of these estimates is that the number of households dropped drastically once separating the data by regions. Hence, we need to treat these estimates with precaution but use them as a reference.

The plant milk estimates are more interesting. In Table 12, households who move from the Midwest

[25] Using Nielsen data, I was able to show that the level of plant milk expenditures matches the map shown in Good Food Institute (2020). Midwest has the lowest average log(expenditure): 0.492 (sd = 0.643), followed by South 0.494 (sd = 0.631), Northeast 0.509 (sd = 0.654), and West 0.533 (sd = 0.663).

to other regions increase their plant milk expenditures (6.4%, 64.4% [significant at 5% level], and 0.9% respectively). The increase is smallest for South. Furthermore, it is the other way around if people move from other regions to Midwest: they all decrease their plant milk consumption (9.1%, 32.7%, and 2.4%, respectively). One may argue that Midwestern households who move to the other regions increase plant milk consumption because there is more variety of plant milk. Another way to see these results is that, although dairy milk is a dominant type of milk in the U.S., people could adopt a new eating style and local culture if their environment changes. In this case, households start to drink more plant milk if they move to the West, Northeast, or South. Of course, the percentage change varies from region to region. Same as in the dairy milk estimates, the limitation is that the sample size becomes too small once we limit the data to a specific region. We need to interpret these results with precaution and only use them at best as reference.

6 Conclusions and Discussions

This paper exploits a natural experiment in the U.S. among households who have moved to another state to estimate how migration affects dairy and plant milk consumption. I merged multiple data files related to milk types, milk purchases, household locations, and characteristics from Nielsen Consumer Data to get two main findings. First, after moving to a new state, households increased their dairy milk expenditures by 1.2%, whereas the plant milk expenditures decreased by 1.5%. The second finding shows that the new destination explains about 53% of the differences in dairy milk expenditures and only about 17% for plant-based milk expenditures. A trivial finding is that people have a persistent dairy milk consumption behavior: if their original dairy milk consumption is high, then their dairy consumption will remain high even moving to another region; and it is the same the other way around. On the other hand, households appear to pick up local plant milk consumption behavior. If they move from a region where plant milk is less prevalent, they could increase plant milk consumption; and if they move from a region where plant milk is more popular, they may reduce plant milk consumption. Overall, I conclude that dairy milk is still the dominant type consumed in the U.S., thus it is not easy for consumers to switch their preferences to plant milk. From the geographic variation point of view, the local environment has a more considerable influence on dairy milk consumption than plant milk.

These findings are of relevance to food and environment policymakers. First, if dairy milk is still the dominant type of milk, it will continue to bring negative environmental impact as shown in McClements et al. (2019). Second, since the local environment influences households to consume more dairy milk, this may require federal and local governments to implement policies that would nudge people to consume more environmentally friendly and sustainable foods (in this case, plant milk). We have already seen evidence that people can adopt more environmentally friendly consumption behavior in the last section (or see the trivial findings summarized in the last paragraph). Nordhaus (2019) discussed that implementing policies that help reduce climate impact may be the most realistic option. The other options such as "carbon removal" (i.e., removing excessive CO_2 from the air) or "geoengineering"[26] are not only more costly, technically difficult, but also not tested (or not testable unless it is done). This evidence suggests an urgency for the U.S. to adopt more environmentally sustainable policies as soon as possible.

There are several limitations to this study. Nielsen data only covers some major markets and a few Census Region levels in the U.S.[27]. This sampling method may limit interpretations of the study since plant milk may not be broadly available in smaller cities. Even if it is available, its higher price (than dairy milk) may limit households from choosing it. Future study could focus on just metropolitan areas defined by USDA[28] since in this case, both dairy and plant milk will be appropriately represented. Another limitation is that there may be recording errors in Nielsen data since it is self-recorded. Einav et al. (2010) demonstrated that such errors[29] could be present and have shown some correction methods. I could learn from this paper and use the correction methods to correct potential recording errors.

[26] Nordhaus (2019) explained that the "carbon removal" option requires removing "200 or 400 or 1,000 billion tons of CO2 from the atmosphere at a reasonable cost." Still, this option may not be technologically possible for now. The other option, "geoengineering," involves "making the earth 'whiter' or more reflective, so that less sunlight reaches the surface of the earth." This option artificially makes the earth cooler but not necessarily "greener," it is also not tested in reality.

[27] Nielsen data sample is stratified into 61 geographic areas in total, including 52 major markets and 9 remaining Census Divisions in the U.S.

[28] USDA provides Market Groups data which defines 26 metropolitan areas in the U.S. See https://www.ers.usda.gov/webdocs/DataFiles/52760/qfahpd1codebook.xls?v=349.5QFAHPD-1–Codebook for its definitions.

[29] Einav et al. (2010) describes the potential recording errors in Nielsen data as follows, "First, there are potential concerns about sample selection. Because of the time commitment, households who agree to participate in the sample might not represent the population of interest. Second, households who agree to participate in the sample might record their purchases incorrectly." Since Einav et al. (2010) was published in 2010, I suspect that the Nielsen data is aware of such issue and has been adjusting its data collection method over the years to minimize such errors.

References

Astolfi, M. L., Marconi, E., Protano, C., and Canepari, S. (2020). Comparative elemental analysis of dairy milk and plant-based milk alternatives. *Food Control*, 116:107327.

Boaitey, A. and Minegishi, K. (2020). Determinants of household choice of dairy and plant-based milk alternatives: evidence from a field survey. *Journal of Food Products Marketing*, 26(9):639–653.

Burke, M., Hsiang, S. M., and Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*, 527(7577):235–239.

Cessna, J., Dong, D., Kuchler, F., and Stewart, H. (2021). Examining the decline in U.S. per capita consumption of fluid cow's milk, 2003–18.

Cheng, C. and Hoekstra, M. (2013). Does strengthening self-defense law deter crime or escalate violence? evidence from expansions to castle doctrine. *Journal of Human Resources*, 48(3):821–854.

Einav, L., Leibtag, E., and Nevo, A. (2010). Recording discrepancies in nielsen homescan data: Are they present and do they matter? *QME*, 8(2):207–239.

Environment Working Group (2021). Dairy program subsidies in the United States totaled $6.4 billion from 1995-2020.

Falkdalen, M. (2017). Valuation on characteristics in the milk market.

Finkelstein, A., Gentzkow, M., and Williams, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *The quarterly journal of economics*, 131(4):1681–1726.

Gerliani, N., Hammami, R., and Aider, M. (2019). Production of functional beverage by using protein-carbohydrate extract obtained from soybean meal by electro-activation. *LWT*, 113:108259.

Good Food Institute (2020). U.S. retail market data for the plant-based industry.

Grey, Clark, Shih and Associates, Limited (2018). U.S. federal and state subsidies to agriculture.

Harding, M. and Rapson, D. (2019). Does absolution promote sin? A conservationist's dilemma. *Environmental and resource economics*, 73(3):923–955.

Herrnstadt, E., Dinan, T., et al. (2020). Cbo's projection of the effect of climate change on US economic output: Working paper 2020-06. Technical report.

Ipsos Retail Performance (2022). Vegan trends in the U.S.

McClements, D. J., Newman, E., and McClements, I. F. (2019). Plant-based milks: A review of the science underpinning their design, fabrication, and performance. *Comprehensive reviews in food science and food safety*, 18(6):2047–2067.

Nordhaus, W. (2019). Climate change: The ultimate challenge for economics. *American Economic Review*, 109(6):1991–2014.

Paulson, N., Schnitkey, G., Swanson, K., and Zulauf, C. (2021). US dairy market and policy overview.

Plant Based Foods Association (2020). U.S. plant-based retail market worth $5 billion, growing at 5x total food sales.

Poore, J. and Nemecek, T. (2018). Reducing food's environmental impacts through producers and consumers. *Science*, 360(6392):987–992.

Sethi, S., Tyagi, S. K., and Anurag, R. K. (2016). Plant-based milk alternatives an emerging segment of functional beverages: a review. *Journal of food science and technology*, 53(9):3408–3423.

Statista (2022). U.S. plant-based milks - statistics facts.

The Good Food Institute (2020). Plant-based: The business case.

USDA (2010). ARS research related to functional foods.

USDA (2021). Food availability (per capita) data system.

Vanga, S. K. and Raghavan, V. (2018). How well do plant based alternatives fare nutritionally compared to cow's milk? *Journal of food science and technology*, 55(1):10–20.

Yang, T. and Dharmasena, S. (2021). US consumer demand for plant-based milk alternative beverages: Hedonic metric augmented Barten's synthetic model. *Foods*, 10(2):265.

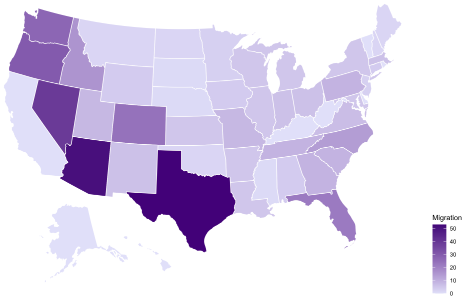
A Appendix: Plant-Based Milk and Dairy Milk Nutrient Comparison

Table A.1: Nutritional Profile of Different Types of Milk

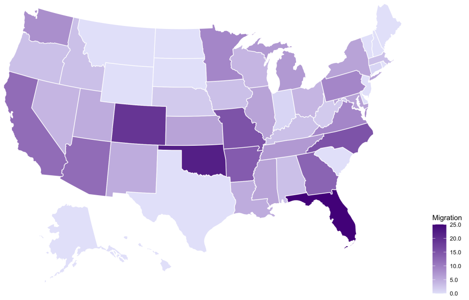
Nutrients	Whole	Skim	Soy	Almond	Oat	Coconut	Rice
Calories	150 kcal	80 kcal	80 kcal	30 kcal	90 kcal	40 kcal	120 kcal
Total Fat	8 g	0 g	4 g	2.5 g	1.5 g	4 g	2.5 g
Saturated Fat	5 g	0 g	0.5 g	0 g	0 g	3 g	0 g
Sodium	120 mg	100 mg	75 mg	125 mg	120 mg	45 mg	100 mg
Carbohydrates	12 g	12 g	3 g	1 g	19 g	1 g	23 g
Dietary Fiber	0 g	0 g	2 g	$<$1 g	2 g	0 g	0 g
Total Sugars	12 g	12 g	1 g	0 g	4 g	0 g	10 g
Added Sugar	0 g	0 g	0 g	0 g	4 g	0 g	0 g
Protein	8 g	8 g	7 g	1 g	2 g	0 g	1 g
Calcium	28%	30%	20%	30%	25%	35%	30%
Folate	3%	3%	10%	-	-	-	-
Iron	1%	1%	6%	2%	2%	2%	4%
Magnesium	6%	7%	8%	2%	-	-	-
Phosphorus	21%	25%	6%	-	-	-	15%
Potassium	9%	10%	7%	2%	8%	6%	1%
Riboflavin	24%	26%	30%	-	10%	-	-
Vitamin A	11%	10%	15%	15%	20%	20%	10%
Vitamin B12	18%	20%	120%	-	10%	35%	25%
Vitamin D	31%	29%	15%	10%	20%	10%	25%
Vitamin E	-	-	-	25%	-	20%	-

Notes: Nutrient data comes from https://totaste.com/got-milk-which-one/. These measures are similar to the results published in Vanga and Raghavan (2018).

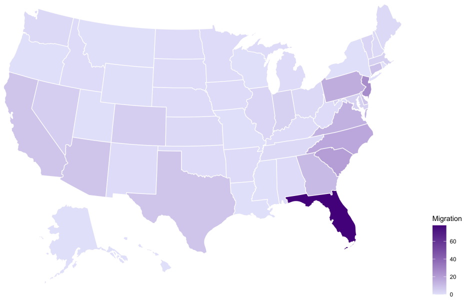
B Appendix: Migration Map for Top 10 Most Populous States in the U.S.



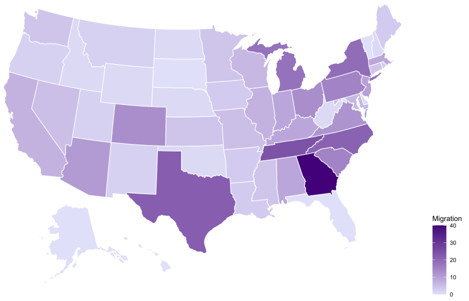
(a) California



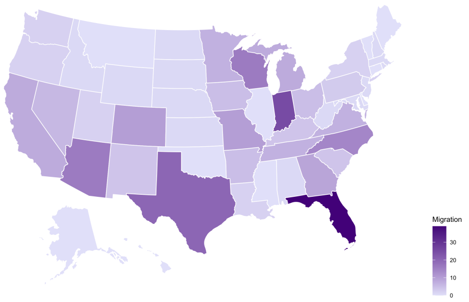
(b) Texas



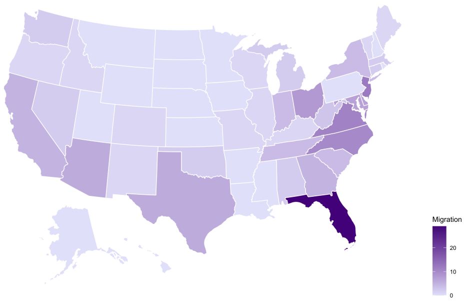
(c) New York



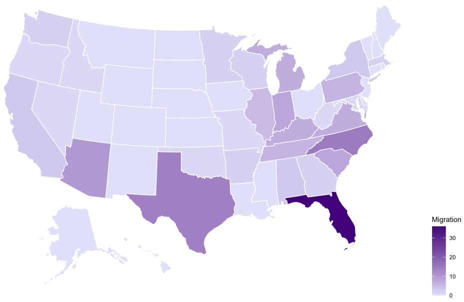
(d) Florida



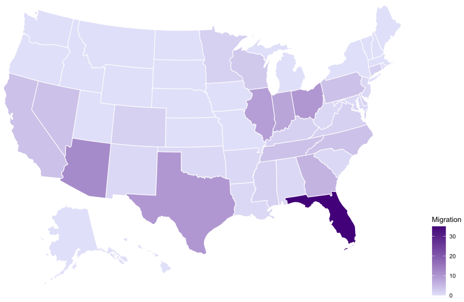
(e) Illinois



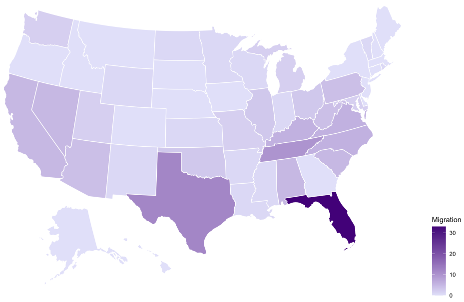
(f) Pennsylvania



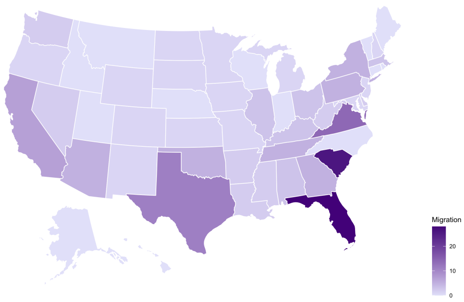
(g) OHIO



(h) MICHIGAN



(i) GEORGIA



(j) CAROLINA

C Appendix: DiD Results (Movers Data Only)

Table C.1: Dairy Milk Expenditure Patterns After Move. DiD Estimates.

	(1)	(2)	(3)	(4)	(5)	(6)
Base specification	0.012**					
	(0.006)					
Age: young		-0.141***				
		(0.028)				
Age: midL		-0.077***				
		(0.016)				
Age: midH		0.050***				
		(0.011)				
Age: old		-0.029*				
		(0.015)				
Children: <12			0.090***			
			(0.029)			
Children: ≥ 12			-0.124***			
			(0.025)			
Children: mixed			0.045			
			(0.032)			
Children: none			0.161***			
			(0.026)			
Educ: HS				0.029		
				(0.029)		
Educ: college				0.015*		
				(0.009)		
Educ: grad				-0.032*		
				(0.015)		
Income: low					0.037**	
					(0.016)	
Income: med					0.007	
					(0.014)	
Income: high					-0.009	
					(0.011)	
Race: white						0.011
						(0.008)
Race: black						0.009
						(0.030)
Race: asian						-0.004
						(0.043)
Households	4762	4762	4762	4762	4762	4762
Observations	148786	148786	148786	148786	148786	148786

Notes: This difference-in-differences analysis uses data only on movers. The dependent variable is log(expenditure) for dairy milk. It includes both time fixed effects and household fixed effects. The standard errors are clustered at the household level.

Table C.2: Plant Milk Expenditure Patterns After Move. DiD Estimates.

	(1)	(2)	(3)	(4)	(5)	(6)
Base specification	-0.006					
	(0.013)					
Age: young		-0.118**				
		(0.055)				
Age: midL		-0.079**				
		(0.034)				
Age: midH		0.044*				
		(0.026)				
Age: old		-0.014				
		(0.035)				
Children: <12			0.024*			
			(0.053)			
Children: ≥ 12			-0.084			
			(0.046)			
Children: mixed			0.047			
			(0.067)			
Children: none			0.119**			
			(0.048)			
Educ: HS				0.078		
				(0.055)		
Educ: college				0.002		
				(0.020)		
Educ: grad				0.001		
				(0.030)		
Income: low					0.033	
					(0.034)	
Income: med					0.035	
					(0.029)	
Income: high					-0.020	
					(0.021)	
Race: white						0.023
						(0.019)
Race: black						-0.098**
						(0.044)
Race: asian						0.025
						(0.067)
Households	2732	2732	2732	2732	2732	2732
Observations	22484	22484	22484	22484	22484	22484

D Appendix: First Differences Results (Movers Data Only)

Table D.1: Dairy Milk Expenditure Patterns After Move. First Differences Estimates.

	(1)	(2)	(3)	(4)	(5)	(6)
Base specification	0.014***					
	(0.038)					
Age: young		-0.118***				
		(0.017)				
Age: midL		-0.085***				
		(0.009)				
Age: midH		0.026***				
		(0.006)				
Age: old		-0.027***				
		(0.008)				
Children: <12			0.061***			
			(0.016)			
Children: ≥ 12			-0.058***			
			(0.013)			
Children: mixed			0.020			
			(0.019)			
Children: none			0.087***			
			(0.013)			
Educ: HS				0.056***		
				(0.010)		
Educ: college				-0.023		
				(0.005)		
Educ: grad				-0.011***		
				(0.008)		
Income: low					0.001	
					(0.008)	
Income: med					-0.042***	
					(0.008)	
Income: high					-0.020***	
					(0.006)	
Race: white						-0.018***
						(0.005)
Race: black						0.030**
						(0.013)
Race: asian						-0.156***
						(0.021)
Households	188028	188028	188028	188028	188028	188028
Observations	3323626	3323626	3323626	3323626	3323626	3323626

Notes: This analysis uses data only on movers. The dependent variable is log(expenditure) for dairy milk. It includes both time fixed effects and household fixed effects. The standard errors are clustered at the household level.

Table D.2: Plant Milk Expenditure Patterns After Move. First Differences Estimates.

	(1)	(2)	(3)	(4)	(5)	(6)
Base specification	-0.003					
	(0.009)					
Age: young		-0.030				
		(0.039)				
Age: midL		-0.067***				
		(0.023)				
Age: midH		-0.029*				
		(0.016)				
Age: old		0.018				
		(0.023)				
Children: $<$12			0.060			
			(0.040)			
Children: $\geq$12			-0.045			
			(0.033)			
Children: mixed			0.077			
			(0.051)			
Children: none			0.043			
			(0.034)			
Educ: HS				0.126***		
				(0.032)		
Educ: college				-0.074***		
				(0.014)		
Educ: grad				-0.021		
				(0.021)		
Income: low					0.066***	
					(0.022)	
Income: med					0.062***	
					(0.021)	
Income: high					-0.123***	
					(0.015)	
Race: white						-0.050***
						(0.013)
Race: black						-0.068**
						(0.028)
Race: asian						-0.097**
						(0.049)
Households	73536	73536	73536	73536	73536	73536
Observations	434264	434264	434264	434264	434264	434264

E Appendix: Event Study: Heterogeneous Effects

Table E.1: Event Study: Heterogeneous Effects For Dairy Milk)

	age(young)	age(midL)	age(midH)	age(old)	children(< 12)	children(≥12)	children(mixed)	children(none)
Event Time								
-5	-0.011	0.008	-0.002	-0.007	0.006	-0.017	-0.04	0.017
	(0.021)	(0.016)	(0.010)	(0.016)	(0.030)	(0.027)	(0.036)	(0.028)
-4	-0.041*	-0.007	0.009	0	0.007	-0.042	0.012	0.059**
	(0.025)	(0.015)	(0.010)	(0.015)	(0.031)	(0.028)	(0.037)	(0.028)
-3	-0.051**	-0.015	0.017*	-0.011	-0.012	-0.034	0.020	0.054*
	(0.025)	(0.016)	(0.010)	(0.015)	(0.031)	(0.027)	(0.036)	(0.028)
-2	-0.057**	-0.011	0.014	-0.021	0.023	-0.056**	-0.002	0.071**
	(0.025)	(0.016)	(0.011)	(0.016)	(0.031)	(0.028)	(0.037)	(0.029)
-1	-0.045*	-0.004	0.010	0.006	0.025	-0.061**	0.004	0.084***
	(0.025)	(0.017)	(0.011)	(0.017)	(0.034)	(0.030)	(0.040)	(0.031)
0	0	0	0	0	0	0	0	0
1	-0.135***	-0.040**	0.035**	-0.041**	-0.033	-0.020	-0.019	0.038**
	(0.033)	(0.018)	(0.011)	(0.016)	(0.038)	(0.033)	(0.042)	(0.034)
2	-0.103***	-0.054**	0.028**	-0.032*	0.014	-0.049	-0.002	0.062*
	(0.032)	(0.018)	(0.012)	(0.017)	(0.036)	(0.031)	(0.041)	(0.032)
3	-0.107***	-0.045**	0.029**	-0.026	-0.036	-0.007	-0.049	0.024
	(0.035)	(0.019)	(0.012)	(0.017)	(0.038)	(0.032)	(0.041)	(0.033)
4	-0.028	-0.044**	0.021*	-0.005	0.038	-0.056*	0.004	0.077**
	(0.042)	(0.019)	(0.012)	(0.017)	(0.038)	(0.033)	(0.042)	(0.034)
5	-0.006	-0.047**	0.028**	-0.021	0.022	-0.033	-0.022	0.055
	(0.040)	(0.019)	(0.012)	(0.017)	(0.039)	(0.035)	(0.048)	(0.036)

Table E.2: (Cont.) Event Study: Heterogeneous Effects For Dairy Milk

	educ(HS)	educ(college)	educ(grad)	income(low)	income(med)	income(high)	race(white)	race(black)	race(asian)
Event Time									
-5	-0.006	-0.004	0.022	0.004	0.001	-0.001	0.000	-0.012	0.026
	(0.021)	(0.008)	(0.015)	(0.015)	(0.014)	(0.010)	(0.007)	(0.027)	(0.044)
-4	-0.030	0.012	-0.017	0.010	0.008	-0.001	-0.001	0.005	0.051
	(0.020)	(0.008)	(0.014)	(0.015)	(0.014)	(0.009)	(0.007)	(0.027)	(0.037)
-3	-0.015	0.012	-0.020	0.024	0.021	-0.007	0.001	0.023	0.029
	(0.020)	(0.008)	(0.014)	(0.015)	(0.014)	(0.010)	(0.007)	(0.030)	(0.038)
-2	0.011	0.000	-0.004	0.023	0.018	-0.011	-0.006	0.050*	-0.003
	(0.020)	(0.008)	(0.015)	(0.015)	(0.015)	(0.010)	(0.007)	(0.029)	(0.037)
-1	0.021	0.006	-0.009	0.035**	0.015	-0.008	0.004	0.032	0.013
	(0.021)	(0.008)	(0.016)	(0.016)	(0.015)	(0.010)	(0.007)	(0.031)	(0.039)
0	0	0	0	0	0	0	0	0	0
1	0.016	0.008	-0.029*	0.030*	0.035**	-0.019*	0.001	-0.011	0.053
	(0.022)	(0.009)	(0.016)	(0.017)	(0.016)	(0.011)	(0.007)	(0.033)	(0.043)
2	0.010	0.000	-0.018	0.039**	0.031*	-0.026**	-0.005	0.030	0.026
	(0.023)	(0.009)	(0.017)	(0.017)	(0.016)	(0.011)	(0.008)	(0.034)	(0.046)
3	-0.013	0.005	-0.006	0.015	0.001 -	0.005	-0.001	-0.007	0.060
	(0.023)	(0.010)	(0.016)	(0.017)	(0.017)	(0.011)	(0.008)	(0.032)	(0.045)
4	0.033	0.010	-0.036**	0.029*	-0.009 -	0.003	0.005	-0.008	0.040
	(0.024)	(0.009)	(0.018)	(0.017)	(0.018)	(0.012)	(0.008)	(0.033)	(0.045)
5	0.016	0.013	-0.024	0.035**	-0.013 -	0.001	0.005	0.023	0.027
	(0.022)	(0.009)	(0.017)	(0.017)	(0.017)	(0.011)	(0.008)	(0.036)	(0.040)

Table E.3: Event Study: Heterogeneous Effects For Plant Milk

	age(young)	age(midL)	age(midH)	age(old)	children(<12)	children(≥12)	children(mixed)	children(none)
Event Time								
-5	0.025	0.060	-0.067***	0.051	-0.017	0.002	-0.188*	-0.036
	(0.051)	(0.043)	(0.024)	(0.043)	(0.080)	(0.065)	(0.101)	(0.067)
-4	0.060	0.068	-0.072***	0.032	0.060	-0.091	-0.045	0.068
	(0.052)	(0.042)	(0.025)	(0.039)	(0.131)	(0.124)	(0.140)	(0.125)
-3	-0.001	0.039	-0.041	0.058	-0.013	-0.050	-0.012	0.050
	(0.051)	(0.043)	(0.027)	(0.041)	(0.068)	(0.053)	(0.099)	(0.056)
-2	-0.007	0.010	-0.034	0.013	-0.009	-0.025	-0.112	0.012
	(0.050)	(0.045)	(0.032)	(0.042)	(0.066)	(0.056)	(0.104)	(0.060)
-1	0.021	0.063	-0.064**	0.011	-0.066	0.020	-0.256**	-0.053
	(0.056)	(0.042)	(0.026)	(0.039)	(0.072)	(0.062)	(0.108)	(0.064)
0	0	0	0	0	0	0	0	0
1	-0.088	0.022	-0.052*	0.049	-0.106	-0.035	-0.029	0.024
	(0.060)	(0.045)	(0.028)	(0.039)	(0.087)	(0.077)	(0.112)	(0.079)
2	-0.065	-0.052	0.001	0.030	0.086	-0.148*	0.018	0.164**
	(0.058)	(0.045)	(0.030)	(0.041)	(0.086)	(0.079)	(0.101)	(0.081)
3	-0.147**	-0.020	0.038	-0.043	0.069	-0.083	0.058	0.105
	(0.063)	(0.043)	(0.029)	(0.044)	(0.072)	(0.063)	(0.125)	(0.066)
4	-0.104	-0.016	-0.013	-0.029	0.093	-0.067	0.141	0.034
	(0.083)	(0.055)	(0.029)	(0.043)	(0.077)	(0.060)	(0.107)	(0.064)
5	-0.022	0.049	-0.046	0.014	-0.034 -	0.024	0.035	0.006
	(0.094)	(0.053)	(0.034)	(0.044)	(0.108)	(0.089)	(0.119)	(0.091)

Table E.4: (Cont.) Event Study: Heterogeneous Effects For Plant Milk

	educ(HS)	educ(college)	educ(grad)	income(low)	income(med)	income(high)	race(white)	race(black)	race(asian)
Event Time									
-5	-0.059	-0.054***	0.074*	-0.068*	-0.003	-0.014	0.002	-0.038**	0.035
	(0.076)	(0.020)	(0.039)	(0.041)	(0.038)	(0.025)	(0.050)	(0.019)	(0.129)
-4	0.044	-0.031	-0.040	-0.016	-0.019	-0.026	-0.041	-0.035**	-0.077
	(0.053)	(0.020)	(0.036)	(0.040)	(0.036)	(0.024)	(0.046)	(0.018)	(0.083)
-3	0.045	-0.024	0.006	-0.037	-0.028	0.003	-0.033	-0.019	0.096
	(0.057)	(0.022)	(0.033)	(0.043)	(0.035)	(0.024)	(0.055)	(0.018)	(0.123)
-2	0.062	-0.037*	0.007	-0.045	-0.039	-0.005	-0.001	-0.033*	-0.026
	(0.086)	(0.020)	(0.038)	(0.042)	(0.040)	(0.023)	(0.067)	(0.018)	(0.077)
-1	0.056	-0.057***	0.037	0.009	0.040	-0.053**	-0.002	-0.042**	0.090
	(0.063)	(0.020)	(0.036)	(0.040)	(0.036)	(0.025)	(0.050)	(0.018)	(0.115)
0	0	0	0	0	0	0	0	0	0
1	0.119**	-0.044**	-0.028	0.019	0.023	-0.058**	0.001	-0.043**	0.117
	(0.058)	(0.022)	(0.037)	(0.041)	(0.040)	(0.023)	(0.047)	(0.019)	(0.100)
2	0.032	-0.008	-0.027	0.030	-0.019	-0.02	-0.017	-0.015	0.069
	(0.066)	(0.023)	(0.036)	(0.042)	(0.039)	(0.025)	(0.048)	(0.020)	(0.094)
3	0.065	-0.005	0.024	-0.043	0.029	0.007	0.028	0.004	0.093
	(0.067)	(0.022)	(0.037)	(0.044)	(0.038)	(0.025)	(0.046)	(0.020)	(0.110)
4	0.065	-0.061**	0.075*	-0.014	-0.008	-0.032	-0.083	-0.021	0.120
5	0.053	-0.030	-0.012	0.051	-0.009	-0.050*	-0.048	-0.024	0.039
	(0.065)	(0.026)	(0.043)	(0.042)	(0.054)	(0.028)	(0.070)	(0.021)	(0.111)

F Appendix: Price Index for Dairy and Plant Milk

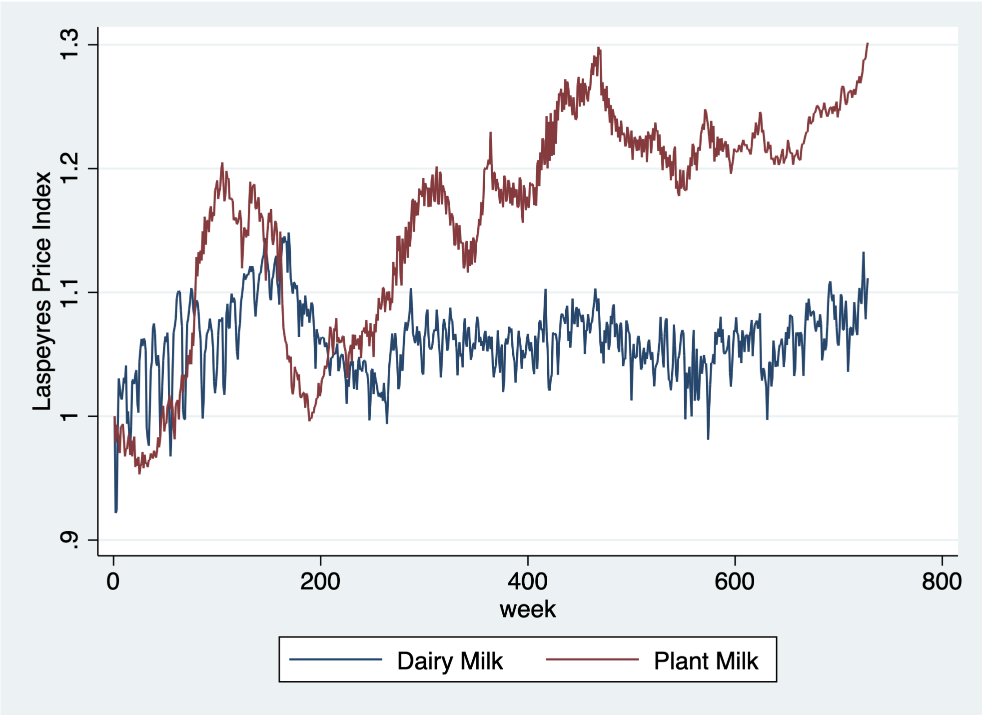


Figure F.1: Price Index for Dairy and Plant Milk