Asymmetric Cross-Price Elasticities of Private Label and National Label Organic Products

Mitchell James Wong

Advisor: Professor Benjamin Faber[1]

University of California Berkeley
Department of Economics
May 2019

Abstract

Organic products sold in retail chains across America have been booming with growth. Retailers are rushing to invest more into their organic product lines. As retailers grow their organic product lines as and demand for organics grows, organic product lines may have been able to affect the traditional cross-price elasticity asymmetry that exists between national goods and private label goods. This paper intends to find the cross-price elasticity relationship between private label organic products and national label organic products compared to their non-organic counterparts. Using the Nielsen Retail Scanner Dataset from 2014 - 2016 we find that organic private label goods are relatively less substitutable with organic national label goods than non-organic private label goods are with non-organic national label goods.

Contents:

1. Introduction:

In the American supermarket, premium national label brands have held and continue to hold a privileged status over their private label retail counterparts. Strong public perception of premium national label brands' quality compared to their private retail counterparts' perceived quality results in an asymmetric price elasticity relationship, where premium national label brand prices can affect their retail counterparts' sales but not vice versa.

However, private label retail brands have been on the up and up. O Organics, a private retail brand of Albertsons, has grown to over $1 billion in sales since its inception in 2005 (Siegner, 2018). Other private brands have followed suit: Kroger's Simple Truth product line eclipsed $2 billion in sales in 2018 (ProgressiveGrocer, 2018) and Target added a new discount brand named Smartly to its stores in October, 2018 (Stump, 2018). Private retail brands are especially desired by retailers over national label brands due to their higher profit margins (Pauwels et al. 2004) but retailers would not create and grow these brands if they didn't have a reasonable expectation of making a profit. Whether or not retailers are expanding their private label brand revenue through increased competition with their national label counterparts or by expanding the market as a whole is a question of interest in this paper. In particular, this paper is interested in seeing if organic private label brands are found by consumers to be more substitutable with national label organic brands than their non-organic retail counterparts with non-organic national label brands.

The inspiration behind studying organic private label retail brands stems from the relative proliferation of organic goods in respect to their non-organic counterparts. In the early 1990s, organic products used to be only sold in natural product stores. Since 1997, organic food sales have increased from $3.6 billion to $21.1 billion in 2008 and organic products are now sold in mainstream retailers like Wal-mart, Target, and Costco. Even though organic sales accounted for only 3% of sales in the United States in 2008, 69% of Americans purchased organic products in 2008 (Green et al. 2008). If private label

3

organic brands have proliferated as organic products have in the tastes and preferences of American consumers, then private label organic brands should be relatively more substitutable with their national label counterparts than non-organic private label brands and their national label counterparts. Thus the cross-price elasticities of private label organic products to their national label organic counterparts should be higher than the cross-price elasticities of private label non-organic products to their national label counterparts. If substitutability and competition is not actually higher among organic product lines, then the cross-price elasticities of organics should be lower than their non-organic counterparts.

The remainder of this paper is organized as follows. In section 2 I will be discussing relevant literature that have studied cross-price elasticities between private label and national label goods in American supermarkets and commenting on how this paper adds to their findings. In section 3 I discuss the nuances of the dataset I studied to reach my results and discuss the availability of the data and code. In section 4 I present my model, its objectives, and describe itsvariables. In section 5 I discuss the findings and interpretation of my model, as well as the limitations of my study. In section 6 I state my conclusions from my analysis. The sections afterwards are dedicated to the appendix and the code used for this paper.

2. Literature Review:

Several studies have observed the cross-price elasticity asymmetrical relationship between private label goods and national label goods, using different data sources.

Blattberg et al. (1989): Using 49 weeks of transactional and ad data from a Chicago retailer, Blattberg found that promotional price competition could induce asymmetrical sales effects within a retail product category. Specifically higher-price/higher-quality brands steal sales from lower-price/lower-quality brands when the higher tiered brand lowers their price. Yet the lower-price/lower-quality brands could barely draw sales from the higher tiered brands when they reduced their prices.

Allenby et al. (1991): Using a scanner panel data set of purchases of margarine, Allenby found that a higher quality brand will have higher price elasticity than a lower quality brand with the same market share due to the interaction between income and substitution effects. However, for lower quality brands, the income and substitution effects cancel each other out, which leads to lower elasticities.

Pauwels et al. (2004): Using scanner data from a Chicago supermarket chain, Pauwels found that a price cut from store brands hardly affects the sales of the most premium national brand in that product category. He also found that second-tier national brand sales are more affected by the prices of store brands.

The above literature agrees on the existence of the cross-price elasticity asymmetrical relationship between private label goods and national label goods. However, they all use proprietary supermarket data localized to one area of the United States and they did not study how this relationship would change when taking organics into account. My paper seeks to expand the existing literature by focusing on organics' effect on the cross-price elasticity asymmetry using national-level transactional panel data.

3. Data Source and Description:

In this paper I used the Nielsen Retail Scanner Dataset from 2014 - 2016. The dataset contains information from approximately 35,000 individual stores from approximately 90 retail chains in the United States of America. Nielsen estimates it covers approximately 53% of the sales of food products with these individual stores. Nielsen collects weekly transactional data by individual store (store_code_uc), by week ending date (week_end), and by individual unique product code (upc) to result in units sold (units) and weighted average price (price) by each of the preceding variables. Figure 1 summarizes the variables Nielsen provides with the scanner data. I ignored the variables for display and feature as I did not study the effects of promotions on products.

To determine which products belonged to private label retail brands and which products belonged to national label brands, I referred to the product file that Nielsen provided with their data. In that file, Nielsen uses a single masked brand code for all private label brands to protect the anonymity of the individual retailers. National brands do not get this code and have their own unique brand code that identifies each individual national brand. Thus, I indicated all of the upcs that shared the same brand code "536746" as being private label retail products. All of the upcs that had a different brand code were labeled as national label products, resulting in a retail dummy variable.

To determine which products were organic and which products were non-organic, I employed two different methodologies. The first methodology was to reference the "product extras" file for each year of data. These files described which upcs contained a USDA organic label. I took the following labels and their corresponding description codes as evidence of being an organic product: "USDA ORGANIC SEAL", "MULTIPLE USDA ORGANIC SEAL", "MULTIPLE", "ASSORTED", "ORGANIC USDA ORGANIC SEAL", "Regular", "SOLID WHITE", "USDA ORGANIC SEA", "USDAO ORGANIC SEAL" and "VARIETAL". I accepted these descriptions as being organic due to their resemblance to actual descriptions of USDA organic labels. The second methodology was to reference the upc descriptions of the product file that Nielsen provided with the data. I searched for the string "ORG" within the descriptions as the descriptions were filled with different acronyms. While Nielsen does not provide guidance to the meaning of these acronyms, I am confident that the acronym "ORG" is referencing organic products due to the fact that the acronym is used consistently across product categories and upcs. Even the upcs confirmed as organic by methodology 1 contained the acronym "ORG". I merged the findings from both methods to create an organic dummy variable to distinguish between organic and non-organic upcs.

I selected the studied product categories based on the availability of data filtered through the following requirement: whether or not there were retail and national label products sold for both organic

and non-organic products in a given store at a given time. Of the 46 different product categories that

Nielsen initially provided, 4 product categories met my research requirements. These categories are milk,

eggs, dressings, and fresh produce. It should be noted that fresh produce is packaged produce as loose

produce lack individual upcs. Nielsen also groups all types of milk together, be it whole, 2%, skim, or

soy. I treat all types of milk as equal for the purposes of this study.

 I also subset the data by semester and I define the semesters using the variable "week_end" by

allocating 26 consecutive weeks into each semester. With three years of data I am left with six semesters

of equal length, save for semester 6 where Nielsen counts an extra week by virtue of 2016 being a leap

year.

 My filtering and subsetting of the data as mentioned above reduced my original data set from

699,114,777 weekly transactional data to 59,711 semesterly data. The summary of the data is shown in

Figure 2. "Lnqdiff" is my dependent variable and my regressors are "ratiodiff" and "orginteraction". I

discuss these variables in more detail below. The variables of my final dataset are shown in Figure 3.

Figure 1.

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
8	display	Num	8	BEST12.	BEST32.
7	feature	Num	8	BEST12.	BEST32.
6	price	Num	8	BEST12.	BEST32.
5	prmult	Num	8	BEST12.	BEST32.
1	store_code_uc	Num	8	BEST12.	BEST32.
4	units	Num	8	BEST12.	BEST32.
2	upc	Num	8	BEST12.	BEST32.
3	week_end	Num	8	BEST12.	BEST32.

Figure 2.

Variable	N	Mean	Std Dev	Minimum	Maximum
lnqdiff	59711	-0.1279131	1.5856286	-12.7448660	11.2369611
ratiodif	59711	0.0364502	0.8620712	-6.6145328	8.1360616
orgindicator	59711	-0.000839290	0.2482239	-6.6145328	3.4095797

Figure 3.

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
3	Organic	Num	8		
4	RETAIL	Num	8		
7	lnqdiff	Num	8		
2	product_group_code	Num	8	BEST12.	BEST32.
6	ratiodif	Num	8		
5	semester	Num	8		
1	store_code_uc	Num	8	BEST12.	BEST32.

3.1 Data and Code Availability Statement

As per UC Berkeley's license agreement with the Kilts Center for Marketing at Chicago Booth,
no part of the data may be posted online or provided without an official license from the Kilts Center.
Data used in this study may be re-requested from the Kilts Center by someone in possession with a
license agreement with the Kilts Center for Marketing at Chicago Booth.

SAS code and STATA code used in the assembling of data and in regression analysis can be
found at the very end of this paper.

4. Model Specification:

In this paper, I look at the change in relative sales of private label products to national label products regressed by the change in relative prices of private label products to national label products and an interaction term of the organic dummy variable and this price ratio. To calculate elasticities, I take the log of each regressor and the dependent variable. The model is represented below (Figure 4):

Figure 4.

$$\Delta log(\frac{Sales_{\ private\ label}}{Sales_{\ national\ label}})_{stobp} = \beta_0 + \beta_1 \Delta log(\frac{Price_{\ private\ label}}{Price_{\ national\ label}})_{stobp} + \beta_2 D^{\ Organic} \times \Delta log(\frac{Price_{\ private\ label}}{Price_{\ national\ label}})_{stobp} + u_{stobp}$$

Where s = individual store (store_code_uc)

    t = semester

    o = organic indicator

    b = retail indicator

    p = product group (product_group_code)

To calculate the regressand, I summed the sales of all retail products and all national products by store, by product category, and by semester. I divided retail sales by national sales to get a sales ratio for each semester, store, and product category. I then took the natural log of the ratio. Afterwards, I took the difference of each logged ratio from the previous semester by store and by product category. The result is the variable "lnqdiff".

To calculate the first regressor, I took the median of all retail prices and all national prices by store, by product category, and by semester. I then took the natural log of the ratio of retail prices in the next period divided by the retail prices of the current period and the natural log of the ratio of national

prices in the next period divided by the national prices of the current period. The price ratio for both retail

and national price is the median of prices by store, by product category, and by time period. I then take

the difference of the logged retail price ratio and the logged national price ratio (illustrated in Figure 5.

below). The result is the variable "ratiodif".

Figure 5.

$$log(\frac{Price_{private\ label}t_{\ n+1}}{Price_{private\ label\ t\ n}}) - log(\frac{Price_{national\ label}t_{\ n+1}}{Price_{national\ label\ t\ n}}) = \Delta log(\frac{Price_{private\ label}}{Price_{national\ label}})_{stobp}$$

To calculate the second regressor, I multiplied the regressor "ratiodif" by the dummy variable

"Organic". The result is the variable "Orginteraction".

5. Analysis:

Figure 6 displays the regression results from the model mentioned above. Because the model is a

log-log model, β_1 tells me the cross-price elasticity of a change in the relative price ratio between median

retail non-organic prices and median national non-organic prices on the change in the relative sales ratio

of retail and national non-organic goods. β_2 is the coefficient of an interaction term between the change

in the relative logged price ratio and an organic dummy variable. β_2 tells me the cross price elasticity of

a change in the relative price ratio between median retail organic prices and median national organic

prices on the change in the relative sales ratio of retail and national organic goods.

From the regression, we find that β_1 is equal to 0.1411883 and is highly statistically significant

at the 99% confidence level with a z-score of 21.28. A 1% increase in the change of the price ratio leads

to a 14.1% increase in the change of the sales ratio of private label non-organic goods to national label

non-organic goods. A positive cross-price elasticity suggests that non-organic private and national label

goods are at least somewhat substitutable. We also find that β_2 is equal to -0.1497538 and is highly

statistically significant at the 99% confidence level with a z-score of -5.89. A 1% increase in the change

of the price ratio leads to a 14.97% decrease in the change of the sales ratio of private label organic goods

to national label organic goods. The fact that $\beta_1 > \beta_2$ suggests that organic private label products are

less substitutable with organic national label products when compared to non-organic private label

products with non-organic national label products. A very small R-squared of 0.0107 suggests that this is

a small effect size.

Figure 6.

```
Random-effects GLS regression                    Number of obs     =     59,711
Group variable: id                               Number of groups  =     20,801

R-sq:                                            Obs per group:
    within  = 0.0053                                         min =          1
    between = 0.0468                                         avg =        2.9
    overall = 0.0107                                         max =          5

                                                 Wald chi2(2)      =     453.13
corr(u_i, X)    = 0 (assumed)                    Prob > chi2       =     0.0000
```

lnqdiff	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
ratiodif	.1411883	.0066335	21.28	0.000	.1281869	.1541898
orginteract~n	-.1497538	.0254081	-5.89	0.000	-.1995528	-.0999549
_cons	-.2535672	.0119933	-21.14	0.000	-.2770736	-.2300608
sigma_u	1.5433252					
sigma_e	1.2824113					
rho	.59155424	(fraction of variance due to u_i)				

The limitations in my research are attributable to the data that is available to my research and the

approach as outlined in my model. My data from Nielsen only covers three years of data and only around

53% of the food product channel in American retailers. I also rely on Nielsen's definition of product

categories in subsetting my data, which leads to some overgeneralizations. An example of a

generalization is Nielsen's treatment of soy milk and regular milk as the same, where in reality the consumers of these products are very different. In my model, I take the median of prices and sales by product categories, by stores, by semesters, by retail or national, and by organic or non-organic to calculate the price and sales ratios. By taking medians, I am simplifying my data analysis at the cost of not accounting for all observations. I also divide the data into semesters to simplify my model, but by doing so I do not account for any demand shocks that may come from recurring events such as holidays.

My model also does not account for the inherent simultaneity bias that occurs between sales and prices. Since there is no way to determine from my data if a change in the price of a good is due to an endogenous or exogenous supply shock, I must accept that my coefficients contain some simultaneity bias.

6. Conclusion:

With the cross-price elasticity of organics being smaller than the cross-price elasticity of non-organics, I come to the conclusion that organic private label products are relatively less substitutable with organic national label products than non-organic private label products are with non-organic national label products. Thus the prices of private label organic products and national label organic products follow each other more closely than how their non-organic counterparts do with each other. The relatively complementary nature of organic private label brands and organic national label brands may result from retail and national brands choosing to keep their prices very similar to each other. As consumers are buying private label and national label organic products together, it would appear that a greater price dispersion between private label and national label organic products drives down demand and favors neither of the two competing parties. The pricing strategy for organic private label goods should thus be different than the pricing strategy non-organic private label goods since organic private label goods are relatively less substitutable.

7. Appendix:

Allenby, Greg M., and Peter E. Rossi. "Quality Perceptions and Asymmetric Switching Between Brands."
Marketing Science, vol. 10, no. 3, 1991, pp. 185–204., doi:10.1287/mksc.10.3.185.

Blattberg, Robert C., and Kenneth J. Wisniewski. "Price-Induced Patterns of Competition." Marketing
Science, vol. 8, no. 4, 1989, pp. 291–309., doi:10.1287/mksc.8.4.291.

Greene, Catherine R, Carolyn Dimitri, Biing-Hwan Lin, William D McBride, Lydia Oberholtzer, and
Travis A Smith. 2009. Emerging Issues in the US Organic Industry. Economic Research Service
of U.S. Department of Agriculture.

Pauwels, Koen, and Shuba Srinivasan. "Who Benefits from Store Brand Entry?" Marketing Science, vol.
23, no. 3, 2004, pp. 364–390., doi:10.1287/mksc.1030.0036.

Siegner, Cathy. "Albertsons to Grow O Organics by 50% as Brand Reaches $1B in Sales." Grocery Dive,
11 Jan. 2018,
www.grocerydive.com/news/grocery--albertsons-to-grow-oorganics-by-50-as-brand-reaches-1b-i
n-sales/534345/.

Stump, Scott. "Target Launching New Discount Brand for Household Items." TODAY.com, 8 Oct. 2018,
www.today.com/news/target-launching-new-discount-brand-household-itemst139026.

"THE RISE AND RISE AGAIN OF PRIVATE LABEL." *Nielsen*, Feb. 2018,
www.nielsen.com/content/dam/nielsenglobal/ru/docs/2018.02%20-%20Global%20-%20The%20
Rise%20and%20Rise%20Again%20of%20Private%20Label.pdf.

8. SAS Code:

```
/* ECON H195B Final Thesis Code for "Title"
   By Mitchell Wong 5/5/19*/

/* Introduce Library Name*/

LIBNAME research 'D:\SAS';

/* Import all Individual Files */

proc import out=work.data1
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\2501_2014\3608_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data2
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\2501_2014\3611_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data3
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\2505_2014\4100_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data4
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\2506_2014\3592_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data5
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\2506_2014\3625_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;
```

```sas
proc import out=work.data6
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\2506_2014\3626_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data7
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\2506_2014\3627_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data8
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\2506_2014\3628_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data9
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\2506_2014\3650_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data10
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\2506_2014\6041_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data11
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4010_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data12
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4015_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;
```

```
proc import out=work.data13
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4020_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data14
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4023_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data15
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4050_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data16
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4055_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data17
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4060_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data18
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4085_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data19
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4140_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
```

```
run;

proc import out=work.data20
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4180_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data21
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4225_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data22
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4230_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data23
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4275_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data24
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4280_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data25
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4350_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data26
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4355_20
14.tsv"
```

```
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data27
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4400_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data28
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4415_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data29
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4460_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data30
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4470_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data31
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\4475_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data32
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\6049_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data33
```

```
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\6050_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data34
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\6064_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data35
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Movement_Files\4001_2014\6070_20
14.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data36
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\2501_2015\3608_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data37
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\2501_2015\3611_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data38
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\2505_2015\4100_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data39
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\2506_2015\3592_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;
```

```
proc import out=work.data40
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\2506_2015\3625_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data41
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\2506_2015\3626_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data42
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\2506_2015\3627_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data43
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\2506_2015\3628_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data44
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\2506_2015\3650_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data45
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\2506_2015\6041_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data46
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4010_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;
```

```
proc import out=work.data47
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4015_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data48
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4020_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data49
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4023_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data50
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4050_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data51
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4055_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data52
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4060_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data53
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4085_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
```

```
run;

proc import out=work.data54
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4140_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data55
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4180_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data56
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4225_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data57
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4230_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data58
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4275_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data59
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4280_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data60
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4350_20
15.tsv"
```

```
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data61
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4355_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data62
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4400_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data63
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4415_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data64
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4460_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data65
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4470_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data66
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\4475_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data67
```

```
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\6049_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data68
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\6050_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data69
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\6064_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data70
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Movement_Files\4001_2015\6070_20
15.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data71
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\2501_2016\3608_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data72
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\2501_2016\3611_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data73
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\2505_2016\4100_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;
```

```
proc import out=work.data74
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\2506_2016\3592_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data75
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\2506_2016\3625_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data76
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\2506_2016\3626_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data77
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\2506_2016\3627_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data78
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\2506_2016\3628_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data79
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\2506_2016\3650_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data80
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\2506_2016\6041_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;
```

```
proc import out=work.data81
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4010_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data82
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4015_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data83
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4020_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data84
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4023_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data85
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4050_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data86
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4055_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data87
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4060_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
```

```
run;

proc import out=work.data88
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4085_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data89
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4140_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data90
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4180_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data91
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4225_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data92
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4230_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data93
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4275_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

proc import out=work.data94
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4280_20
16.tsv"
```

```
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data95
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4350_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data96
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4355_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data97
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4400_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data98
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4415_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data99
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4460_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data100
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4470_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data101
```

```
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\4475_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data102
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\6049_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data103
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\6050_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data104
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\6064_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


proc import out=work.data105
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Movement_Files\4001_2016\6070_20
16.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;


/* Reformatting some columns from string to numeric variables*/

data data12a (DROP = feature display); set data12;
featurea = input(feature, best5.);
displaya = input(display, best5.);
run;


data data12b (DROP = featurea displaya); set data12a;
feature = input(featurea, best5.);
display = input(displaya, best5.);
run;


data data13a (DROP = feature display); set data13;
featurea = input(feature, best5.);
```

```
displaya = input(display, best5.);
run;

data data13b (DROP = featurea displaya); set data13a;
feature = input(featurea, best5.);
display = input(displaya, best5.);
run;

data data14a (DROP = feature display); set data14;
featurea = input(feature, best5.);
displaya = input(display, best5.);
run;

data data14b (DROP = featurea displaya); set data14a;
feature = input(featurea, best5.);
display = input(displaya, best5.);
run;

data data23a (DROP = feature display); set data23;
featurea = input(feature, best5.);
displaya = input(display, best5.);
run;

data data23b (DROP = featurea displaya); set data23a;
feature = input(featurea, best5.);
display = input(displaya, best5.);
run;

data data25a (DROP = feature display); set data25;
featurea = input(feature, best5.);
displaya = input(display, best5.);
run;

data data25b (DROP = featurea displaya); set data25a;
feature = input(featurea, best5.);
display = input(displaya, best5.);
run;

data data36a (DROP = feature display); set data36;
featurea = input(feature, best5.);
displaya = input(display, best5.);
run;

data data36b (DROP = featurea displaya); set data36a;
feature = input(featurea, best5.);
display = input(displaya, best5.);
run;
```

```sas
data data37a (DROP = feature display); set data37;
featurea = input(feature, best5.);
displaya = input(display, best5.);
run;

data data37b (DROP = featurea displaya); set data37a;
feature = input(featurea, best5.);
display = input(displaya, best5.);
run;

data data39a (DROP = feature display); set data39;
featurea = input(feature, best5.);
displaya = input(display, best5.);
run;

data data39b (DROP = featurea displaya); set data39a;
feature = input(featurea, best5.);
display = input(displaya, best5.);
run;

data data62a (DROP = feature display); set data62;
featurea = input(feature, best5.);
displaya = input(display, best5.);
run;

data data62b (DROP = featurea displaya); set data62a;
feature = input(featurea, best5.);
display = input(displaya, best5.);
run;

data data77a (DROP = feature display); set data77;
featurea = input(feature, best5.);
displaya = input(display, best5.);
run;

data data77b (DROP = featurea displaya); set data77a;
feature = input(featurea, best5.);
display = input(displaya, best5.);
run;

data data87a (DROP = feature display); set data87;
featurea = input(feature, best5.);
displaya = input(display, best5.);
run;

data data87b (DROP = featurea displaya); set data87a;
feature = input(featurea, best5.);
display = input(displaya, best5.);
```

```
run;

/* Merge all datasets together*/

data research.alldata;
      set data1 data2 data3 data4 data5 data6 data7 data8 data9 data10 data11
data12b data13b data14b data15 data16 data17 data18 data19 data20 data21 data22
data23b data24 data25b data26 data27 data28 data29 data30 data31 data32 data33
data34 data35 data36b data37b data38 data39b data40 data41 data42 data43 data44
data45 data46 data47 data48 data49 data50 data51 data52 data53 data54 data55
data56 data57 data58 data59 data60 data61 data62b data63 data64 data65 data66
data67 data68 data69 data70 data71 data72 data73 data74 data75 data76 data77b
data78 data79 data80 data81 data82 data83 data84 data85 data86 data87b data88
data89 data90 data91 data92 data93 data94 data95 data96 data97 data98 data99
data100 data101 data102 data103 data104 data105;
      run;

/* Import in product data to find organics and retail products*/

proc import out = work.product
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\Master_Files\Latest\products.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

/* Method 1: Acronym ORG in upc_descr defining organic*/

data tess;
   set product;
   if find(upc_descr,'ORG','i') ge 1;
   ORG = 1;
   /* equivalent WHERE clause */
   *where upcase(name) contains 'IAN';

run;

proc sort data=work.product
      out=work.onee;
      by upc;
      run;

proc sort data=work.tess
      out=work.three;
      by upc;
      run;

data work.mergedd;
```

```
      merge work.onee(keep = upc upc_descr product_group_code department_descr
brand_code_uc)
                work.three;
      by upc;
      run;


data merged2;
      set mergedd;
      IF ORG not = 1 and ORG not = 0
      THEN ORG = 0;
      RUN;


proc sort data=research.alldata
      out=research.alldata;
      by upc;
      run;


proc sort data=work.merged2
      out=work.toe;
      by upc;
      run;


data research.merged4;
      merge research.alldata(keep = store_code_uc upc week_end units price)
                work.toe;
      by upc;
      run;


data research.merged4;
      set research.merged4;
      where store_code_uc > 0;
      run;


data research.merged4;
      set research.merged4 (keep = store_code_uc upc week_end units price
upc_descr product_group_code department_descr brand_code_uc org);
      where store_code_uc > 0;
      IF brand_code_uc = 536746 THEN RETAIL = 1;
      ELSE RETAIL = 0;
      run;

/* Defining Semesters*/

data research.merged4;
      set research.merged4;
      IF week_end >= 20140104 and week_end <= 20140628 THEN semester = 1;
      IF week_end >= 20140705 and week_end <= 20141227 THEN semester = 2;
      IF week_end >= 20150103 and week_end <= 20150627 THEN semester = 3;
```

```
        IF week_end >= 20150704 and week_end <= 20151226 THEN semester = 4;
        IF week_end >= 20160102 and week_end <= 20160625 THEN semester = 5;
        IF week_end >= 20160702 and week_end <= 20161231 THEN semester = 6;
        run;


/* Method 2: USDA Organic Seal Indicator*/

proc import out=work.productextra2014
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2014\Annual_Files\products_extra_2014
.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

data productextra2014;
      set productextra2014;
      keep upc product_descr organic_claim_code organic_claim_descr
usda_organic_seal_code usda_organic_seal_descr;
      run;

proc import out=work.productextra2015
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2015\Annual_Files\products_extra_2015
.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

data productextra2015;
      set productextra2015;
      keep upc product_descr organic_claim_code organic_claim_descr
usda_organic_seal_code usda_organic_seal_descr;
      run;

proc import out=work.productextra2016
datafile =
"D:\Globus\Extracted\nielsen_extracts\RMS\2016\Annual_Files\products_extra_2016
.tsv"
dbms=dlm replace; getnames = yes; datarow=2;delimiter='09'x;
run;

data productextra2016;
      set productextra2016;
      keep upc product_descr organic_claim_code organic_claim_descr
usda_organic_seal_code usda_organic_seal_descr;
      run;

data productextra;
```

```sas
        set productextra2014 productextra2015 productextra2016;
        run;

data twoo;
        set productextra;
        IF    usda_organic_seal_code = 612
        or usda_organic_seal_code = 122
        or usda_organic_seal_code = 373
        or usda_organic_seal_code = 123
        or usda_organic_seal_code = 799
        or usda_organic_seal_code = 487
        or usda_organic_seal_code = 4
        or usda_organic_seal_code = 297
        or usda_organic_seal_code = 796
        or usda_organic_seal_code = 507
        THEN ORG2 = 1;
        ELSE ORG2 = 0;
RUN;

proc sort data=work.product
        out=work.one;
        by upc;
        run;

proc sort data=work.twoo
        out=work.two;
        by upc;
        run;

data work.mergedd;
        merge work.one(keep = upc upc_descr product_group_code department_descr
brand_code_uc)
                        work.two;
        by upc;
        run;

/* Finalizing Organic indicator*/

data merged2;
        set mergedd;
        IF ORG2 not = 1 and ORG2 not = 0
        THEN ORG2 = 0;
        RUN;

proc sort data=research.alldata
        out=work.onee;
        by upc;
        run;
```

```sas
proc sort data=work.merged2
      out=work.toe2;
      by upc;
      run;


/* Start of sales variable creation*/

data research.merged5;
      merge research.merged4(keep = store_code_uc upc week_end units price
upc_descr product_group_code ORG RETAIL semester)
                  work.toe2;
      by upc;
      run;


data research.merged5;
      set research.merged5 (keep = store_code_uc week_end upc units price
upc_descr product_group_code org retail semester department_descr org2);
      where store_code_uc >0;
      run;


data research.merged5;
      set research.merged5;
      IF ORG = 0 and ORG2 = 0 then Organic = 0;
      ELSE Organic = 1;
      run;


data research.allstuff;
      set research.merged5 (keep = store_code_uc upc units price
product_group_code organic retail semester week_end);
      IF retail = 0 then quantitynational = units;
      ELSE quantitynational = 0;
      IF retail = 1 then quantityretail = units;
      ELSE quantityretail = 0;
      IF retail = 0 then pricenational = price;
      ELSE pricenational = 0;
      IF retail = 1 then priceretail = price;
      ELSE priceretail = 0;
      run;


proc sort data = research.allstuff;
      by store_code_uc upc week_end semester;
      run;


data research.organicq;
      set research.allstuff;
      where organic = 1;
      run;
```

```
data research.nonorganicq;
      set research.allstuff;
      where organic = 0;
      run;
proc sql;
      create table research.neworganicq as
            select store_code_uc,
                        upc,
                        price,
                        product_group_code,
                        organic,
                        retail,
                        week_end,
                        semester,
                        sum(quantitynational) as quantitynational,
                        sum(quantityretail) as quantityretail
            from research.organicq
            group by store_code_uc,product_group_code,semester;
      quit;

proc sql;
      create table research.newnonorganicq as
            select store_code_uc,
                        upc,
                        price,
                        product_group_code,
                        organic,
                        retail,
                        week_end,
                        semester,
                        sum(quantitynational) as quantitynational,
                        sum(quantityretail) as quantityretail
            from research.nonorganicq
            group by store_code_uc,product_group_code,semester;
      quit;

data research.newnonorganicq;
      set research.newnonorganicq;
      where quantityretail > 0;
      run;

data research.neworganicq;
      set research.neworganicq;
      where quantityretail > 0;
      run;

proc sort data=research.neworganicq
```

```sas
      out=main1 nodupkey;
      by store_code_uc semester;
      run;


proc sort data=research.newnonorganicq
      out=main2 nodupkey;
      by store_code_uc semester;
      run;


data research.consolidatedquantity;
      set research.neworganicq research.newnonorganicq;
      run;


proc sort data = research.consolidatedquantity
      out = research.main
      NODUPKEY;
      BY store_code_uc upc week_end semester organic;
RUN;


proc sort data = research.main
      out = research.four
      nodupkey;
      by store_code_uc semester organic;
      run;


proc sort data = research.four;
      by store_code_uc product_group_code semester organic;
      run;


data research.four;
      set research.four;
      where quantityretail >0;
      lnqratio = log(quantityretail/quantitynational);
      run;


data research.five;
      set research.four (DROP = upc price week_end quantitynational
quantityretail);
      if store_code_uc = lag(store_code_uc)
            and product_group_code = lag(product_group_code)
            and semester NE lag(semester)
      then lagqratio = lag(lnqratio);
      else lagqratio = .;
      lnqdiff = lnqratio - lagqratio;
      run;


/*start price sorting here*/
```

```
data research.organicp;
      set research.allstuff;
      where org = 1;
      run;


proc sort data = research.organicq;
      by store_code_uc semester;
      run;


proc sort data = research.nonorganicq;
      by store_code_uc semester;
      run;


/* This code finds the median of the prices by store by product type code by
semester*/
proc sql;
      create table research.pricevariable1 as
            select store_code_uc,
                        upc,
                        price,
                        product_group_code,
                        organic,
                        retail,
                        week_end,
                        semester,
                        median(pricenational) as medianpricenational,
                        median(priceretail) as medianpriceretail
            from research.organicq
            group by store_code_uc, upc, semester;
      quit;

data research.pricevariableretail;
      set research.pricevariable1;
      where retail = 1;
      run;


data research.pricevariablenational;
      set research.pricevariable1;
      where retail = 0;
      run;


proc sql;
      create table research.pricevariable2 as
            select store_code_uc,
                        upc,
                        price,
                        product_group_code,
                        organic,
```

```sas
                              retail,
                              week_end,
                              semester,
                              median(pricenational) as medianpricenational,
                              median(priceretail) as medianpriceretail
                 from research.nonorganicq
                 group by store_code_uc, upc, semester;
          quit;


data research.pricevariableretail2;
      set research.pricevariable2;
      where retail = 1;
      run;


data research.pricevariablenational2;
      set research.pricevariable2;
      where retail = 0;
      run;
/* Retail Organic Milk Price*/
data orgmilk;
set research.pricevariableretail;
where product_group_code = 2506;
run;


proc sort data = orgmilk
      dupout = lmao
      nodupkey;
      by store_code_uc semester;
      run;


data orgmilk;
      set orgmilk;
      if store_code_uc = lag(store_code_uc)
             and product_group_code = lag(product_group_code)
             and semester NE lag(semester)
      then lagprice = lag(medianpriceretail);
      else lagprice = .;
      Pratioretail = log(medianpriceretail/lagprice);
      run;


data orgmilkretail;
      set orgmilk (DROP = upc price week_end medianpricenational);
      where pratioretail NE .;
      run;


/* National Organic Milk Price*/
data orgmilk2;
set research.pricevariablenational;
```

```
where product_group_code = 2506;
run;

proc sort data = orgmilk2
      dupout = lmao2
      nodupkey;
      by store_code_uc semester;
      run;

data orgmilk2;
      set orgmilk2;
      if store_code_uc = lag(store_code_uc)
            and product_group_code = lag(product_group_code)
            and semester NE lag(semester)
      then lagprice = lag(medianpricenational);
      else lagprice = .;
      Prationational = log(medianpricenational/lagprice);
      run;

data orgmilknational;
      set orgmilk2 (DROP = upc price week_end medianpricenational);
      where prationational NE .;
      run;

/*Retail Organic Egg Price*/

data orgegg;
set research.pricevariableretail;
where product_group_code = 2505;
run;

proc sort data = orgegg
      nodupkey;
      by store_code_uc semester;
      run;

data orgegg;
      set orgegg;
      if store_code_uc = lag(store_code_uc)
            and product_group_code = lag(product_group_code)
            and semester NE lag(semester)
      then lagprice = lag(medianpriceretail);
      else lagprice = .;
      Pratioretail = log(medianpriceretail/lagprice);
      run;

data orgeggretail;
      set orgegg (DROP = upc price week_end medianpricenational);
```

41

```sas
        where pratioretail NE .;
        run;

/* National Organic Egg Price*/
data orgegg2;
set research.pricevariablenational;
where product_group_code = 2505;
run;

proc sort data = orgegg2
        nodupkey;
        by store_code_uc semester;
        run;

data orgegg2;
        set orgegg2;
        if store_code_uc = lag(store_code_uc)
                and product_group_code = lag(product_group_code)
                and semester NE lag(semester)
        then lagprice = lag(medianpricenational);
        else lagprice = .;
        Prationational = log(medianpricenational/lagprice);
        run;

data orgeggnational;
        set orgegg2 (DROP = upc price week_end medianpricenational);
        where prationational NE .;
        run;

/*Retail Organic Produce Price*/

data orgproduce;
set research.pricevariableretail;
where product_group_code = 4001;
run;

proc sort data = orgproduce
        nodupkey;
        by store_code_uc semester;
        run;

data orgproduce;
        set orgproduce;
        if store_code_uc = lag(store_code_uc)
                and product_group_code = lag(product_group_code)
                and semester NE lag(semester)
        then lagprice = lag(medianpriceretail);
        else lagprice = .;
```

```
        Pratioretail = log(medianpriceretail/lagprice);
        run;


data orgproduceretail;
        set orgproduce (DROP = upc price week_end medianpricenational);
        where pratioretail NE .;
        run;

/* National Organic Produce Price*/
data orgproduce2;
set research.pricevariablenational;
where product_group_code = 4001;
run;

proc sort data = orgproduce2
        nodupkey;
        by store_code_uc semester;
        run;


data orgproduce2;
        set orgproduce2;
        if store_code_uc = lag(store_code_uc)
                and product_group_code = lag(product_group_code)
                and semester NE lag(semester)
        then lagprice = lag(medianpricenational);
        else lagprice = .;
        Prationational = log(medianpricenational/lagprice);
        run;

data orgproducenational;
        set orgproduce2 (DROP = upc price week_end medianpricenational);
        where prationational NE .;
        run;

/*Retail Organic Dressings Price*/

data orgdressings;
set research.pricevariableretail;
where product_group_code = 3001;
run;

proc sort data = orgdressings
        nodupkey;
        by store_code_uc semester;
        run;


data orgdressings;
        set orgdressings;
```

```
        if store_code_uc = lag(store_code_uc)
                and product_group_code = lag(product_group_code)
                and semester NE lag(semester)
        then lagprice = lag(medianpriceretail);
        else lagprice = .;
        Pratioretail = log(medianpriceretail/lagprice);
        run;


data orgdressingsretail;
        set orgdressings (DROP = upc price week_end medianpricenational);
        where pratioretail NE .;
        run;


/* National Organic Dressings Price*/
data orgdressings2;
set research.pricevariablenational;
where product_group_code = 3001;
run;


proc sort data = orgdressings2
        nodupkey;
        by store_code_uc semester;
        run;


data orgdressings2;
        set orgdressings2;
        if store_code_uc = lag(store_code_uc)
                and product_group_code = lag(product_group_code)
                and semester NE lag(semester)
        then lagprice = lag(medianpricenational);
        else lagprice = .;
        Prationational = log(medianpricenational/lagprice);
        run;


data orgdressingsnational;
        set orgdressings2 (DROP = upc price week_end medianpricenational);
        where prationational NE .;
        run;


/* Merge all organic together*/

/* Milk*/
data organicmilkprice;
        merge orgmilkretail(keep = store_code_uc product_group_code organic
RETAIL semester pratioretail)
                        orgmilknational;
        by store_code_uc product_group_code semester;
        run;
```

```sas
data organicmilkprice;
      set organicmilkprice (DROP = medianpriceretail lagprice);
      where pratioretail NE . and prationational NE .;
      run;


/* Egg*/
data organiceggprice;
      merge orgeggretail(keep = store_code_uc product_group_code organic RETAIL
semester pratioretail)
                 orgeggnational;
      by store_code_uc product_group_code semester;
      run;


data organiceggprice;
      set organiceggprice (DROP = medianpriceretail lagprice);
      where pratioretail NE . and prationational NE .;
      run;


/* Produce*/
data organicproduceprice;
      merge orgproduceretail(keep = store_code_uc product_group_code organic
RETAIL semester pratioretail)
                 orgproducenational;
      by store_code_uc product_group_code semester;
      run;


data organicproduceprice;
      set organicproduceprice (DROP = medianpriceretail lagprice);
      where pratioretail NE . and prationational NE .;
      run;


/* Dressings*/
data organicdressingsprice;
      merge orgdressingsretail(keep = store_code_uc product_group_code organic
RETAIL semester pratioretail)
                 orgdressingsnational;
      by store_code_uc product_group_code semester;
      run;


data organicdressingsprice;
      set organicdressingsprice (DROP = medianpriceretail lagprice);
      where pratioretail NE . and prationational NE .;
      run;


/* All Organic Prices*/

data allorganicprices;
```

```
        set organicmilkprice organiceggprice organicproduceprice
organicdressingsprice;
        ratiodif = pratioretail - prationational;
        run;


/* Retail Non-organic Milk Price*/
data milk;
set research.pricevariableretail2;
where product_group_code = 2506;
run;


proc sort data = milk
        nodupkey;
        by store_code_uc semester;
        run;


data milk;
        set milk;
        if store_code_uc = lag(store_code_uc)
                and product_group_code = lag(product_group_code)
                and semester NE lag(semester)
        then lagprice = lag(medianpriceretail);
        else lagprice = .;
        Pratioretail = log(medianpriceretail/lagprice);
        run;


data milkretail;
        set milk (DROP = upc price week_end medianpricenational);
        where pratioretail NE .;
        run;


/* National Non-organic Milk Price*/
data milk2;
set research.pricevariablenational2;
where product_group_code = 2506;
run;


proc sort data = milk2
        nodupkey;
        by store_code_uc semester;
        run;


data milk2;
        set milk2;
        if store_code_uc = lag(store_code_uc)
                and product_group_code = lag(product_group_code)
                and semester NE lag(semester)
        then lagprice = lag(medianpricenational);
```

```
        else lagprice = .;
        Prationational = log(medianpricenational/lagprice);
        run;


data milknational;
        set milk2 (DROP = upc price week_end medianpricenational);
        where prationational NE .;
        run;


/*Retail Non-organic Egg Price*/

data egg;
set research.pricevariableretail2;
where product_group_code = 2505;
run;


proc sort data = egg
        nodupkey;
        by store_code_uc semester;
        run;


data egg;
        set egg;
        if store_code_uc = lag(store_code_uc)
                and product_group_code = lag(product_group_code)
                and semester NE lag(semester)
        then lagprice = lag(medianpriceretail);
        else lagprice = .;
        Pratioretail = log(medianpriceretail/lagprice);
        run;


data eggretail;
        set egg (DROP = upc price week_end medianpricenational);
        where pratioretail NE .;
        run;


/* National Non-organic Egg Price*/
data egg2;
set research.pricevariablenational2;
where product_group_code = 2505;
run;


proc sort data = egg2
        nodupkey;
        by store_code_uc semester;
        run;


data egg2;
```

```
        set egg2;
        if store_code_uc = lag(store_code_uc)
              and product_group_code = lag(product_group_code)
              and semester NE lag(semester)
        then lagprice = lag(medianpricenational);
        else lagprice = .;
        Prationational = log(medianpricenational/lagprice);
        run;


data eggnational;
        set egg2 (DROP = upc price week_end medianpricenational);
        where prationational NE .;
        run;


/*Retail Non-organic Produce Price*/

data produce;
set research.pricevariableretail2;
where product_group_code = 4001;
run;


proc sort data = produce
        nodupkey;
        by store_code_uc semester;
        run;


data produce;
        set produce;
        if store_code_uc = lag(store_code_uc)
              and product_group_code = lag(product_group_code)
              and semester NE lag(semester)
        then lagprice = lag(medianpriceretail);
        else lagprice = .;
        Pratioretail = log(medianpriceretail/lagprice);
        run;


data produceretail;
        set produce (DROP = upc price week_end medianpricenational);
        where pratioretail NE .;
        run;


/* National Non-organic Produce Price*/
data produce2;
set research.pricevariablenational2;
where product_group_code = 4001;
run;


proc sort data = produce2
```

```
        nodupkey;
        by store_code_uc semester;
        run;


data produce2;
        set produce2;
        if store_code_uc = lag(store_code_uc)
               and product_group_code = lag(product_group_code)
               and semester NE lag(semester)
        then lagprice = lag(medianpricenational);
        else lagprice = .;
        Prationational = log(medianpricenational/lagprice);
        run;


data producenational;
        set produce2 (DROP = upc price week_end medianpricenational);
        where prationational NE .;
        run;

/*Retail Non-organic Dressings Price*/

data dressings;
set research.pricevariableretail2;
where product_group_code = 3001;
run;

proc sort data = dressings
        nodupkey;
        by store_code_uc semester;
        run;

data dressings;
        set dressings;
        if store_code_uc = lag(store_code_uc)
               and product_group_code = lag(product_group_code)
               and semester NE lag(semester)
        then lagprice = lag(medianpriceretail);
        else lagprice = .;
        Pratioretail = log(medianpriceretail/lagprice);
        run;

data dressingsretail;
        set dressings (DROP = upc price week_end medianpricenational);
        where pratioretail NE .;
        run;

/* National Non-organic Dressings Price*/
data dressings2;
```

```
set research.pricevariablenational2;
where product_group_code = 3001;
run;


proc sort data = dressings2
      nodupkey;
      by store_code_uc semester;
      run;


data dressings2;
      set dressings2;
      if store_code_uc = lag(store_code_uc)
            and product_group_code = lag(product_group_code)
            and semester NE lag(semester)
      then lagprice = lag(medianpricenational);
      else lagprice = .;
      Prationational = log(medianpricenational/lagprice);
      run;


data dressingsnational;
      set dressings2 (DROP = upc price week_end medianpricenational);
      where prationational NE .;
      run;


/* Merge all non-organic together*/

/* Milk*/
data nonorganicmilkprice;
      merge milkretail(keep = store_code_uc product_group_code organic RETAIL
semester pratioretail)
                  milknational;
      by store_code_uc product_group_code semester;
      run;


data nonorganicmilkprice;
      set nonorganicmilkprice (DROP = medianpriceretail lagprice);
      where pratioretail NE . and prationational NE .;
      run;


/* Egg*/
data nonorganiceggprice;
      merge eggretail(keep = store_code_uc product_group_code organic RETAIL
semester pratioretail)
                  eggnational;
      by store_code_uc product_group_code semester;
      run;


data nonorganiceggprice;
```

```
      set nonorganiceggprice (DROP = medianpriceretail lagprice);
      where pratioretail NE . and prationational NE .;
      run;


/* Produce*/
data nonorganicproduceprice;
      merge produceretail(keep = store_code_uc product_group_code organic
RETAIL semester pratioretail)
                 producenational;
      by store_code_uc product_group_code semester;
      run;


data nonorganicproduceprice;
      set nonorganicproduceprice (DROP = medianpriceretail lagprice);
      where pratioretail NE . and prationational NE .;
      run;


/* Dressings*/
data nonorganicdressingsprice;
      merge dressingsretail(keep = store_code_uc product_group_code organic
RETAIL semester pratioretail)
                 dressingsnational;
      by store_code_uc product_group_code semester;
      run;


data nonorganicdressingsprice;
      set nonorganicdressingsprice (DROP = medianpriceretail lagprice);
      where pratioretail NE . and prationational NE .;
      run;


/* All Non-Organic Prices*/

data research.allnonorganicprices;
      set nonorganicmilkprice nonorganiceggprice nonorganicproduceprice
nonorganicdressingsprice;
      ratiodif = pratioretail - prationational;
      run;


/*All Prices*/

data research.pricevariablefinal;
      set research.allorganicprices research.allnonorganicprices;
      run;


proc sort data = research.pricevariablefinal;
      by store_code_uc product_group_code organic semester;
      run;
```

```
/*Fixing Sales Data */

data research.salesvariablefinal;
      set research.five (DROP = lnqratio lagqratio);
      where lnqdiff NE .;
      run;


proc sort data = research.salesvariablefinal;
      by store_code_uc product_group_code organic semester;
      run;


/* Merge for final dataset*/
data research.finaldataset;
      merge research.pricevariablefinal
                  research.salesvariablefinal;
      by store_code_uc product_group_code organic semester;
      run;


data research.finaldataset;
      set research.finaldataset (drop = pratioretail prationational);
      where ratiodif NE . and lnqdiff NE .;
      run;


proc sort data = research.finaldataset
      out = test3;
      by store_code_uc product_group_code semester;
      run;


proc export
      data = research.finaldataset
      outfile = "d:\sas\finaldataset.csv"
      replace;
      run;


data research.finaldataset;
      set research.finaldataset;
      orgindicator = organic*ratiodif;
      run;
```

9. STATA Code:

import delimited \\Client\D$\SAS\finaldataset.csv

egen id = group(store_code_uc product_group_code organic)

g orginteraction = ratiodif*organic

xtset id semester

xtreg lnqdiff ratiodif orginteraction