Buy Online Pick-up in Store

Bryan Tamsir, Olivia Natasha, Hariz Hisham, and Kevin Chang

Background

We are given six questions from our boss to assess whether or not to implement BOPS.

Given 4 datasets to compare BOPS and BO-Delivery from August 1, 2010 - July 31, 2013:

1. Transaction Level Data
2. Consumer Level Data
3. Online Daily Data - Sales and Return
4. Online Daily Data - Product Category Sales and Return

Background

There are three online channel stores:

- Store #2
- Store #6
- Store #5998

Implementation of BOPS

- Stores 2 and 6 implemented BOPS on August 1st, 2011
- Store 5998 implemented BOPS on September 27th, 2012
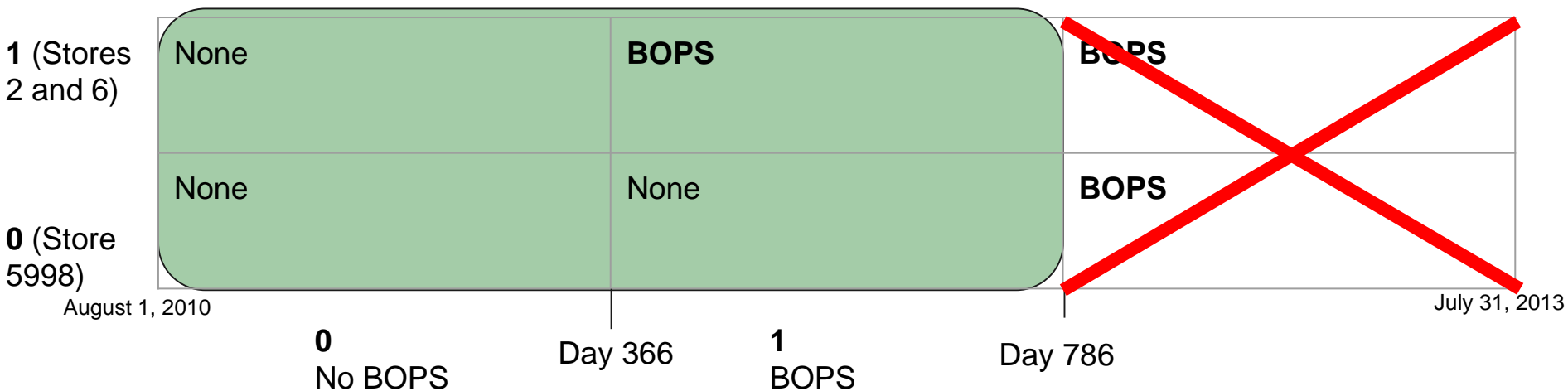 - The difference is 1 year, 1 month and 27 days.

Purpose of Analysis

In particular, we are interested in:

- **Impact of BOPS on online sales and returns (Q1 and Q2)**
- **Impact of BOPS on online customer behavior (Q3 and Q4)**
- **Product-level impact of BOPS implementation (Q5 and Q6)**

Standardization of Time

Dataset - Online Daily Sales Return

	None	**BOPS**	**BOPS**
1 (Stores 2 and 6)			
0 (Store 5998)	None	None	**BOPS**

August 1, 2010

0
No BOPS

Day 366

1
BOPS

Day 786

July 31, 2013

Question 1 -What is the impact of implementing BOPs strategy on online sales?

Dataset - Online Daily Sales Return
Model: OLS

$$log(Y_{Sales\ Value)=}\beta_0 + \beta_1\ Time\_dummy*Storegroup + \beta_2\ Avg\_female + \beta_3\ Avg\_age + \beta_4\ Avg\_income + \beta_5\ Avg\_homeowner + \beta_6\ Avg\_residency + \beta_7 Avg\_childowner$$

Model: Negative Binomial

$$log(Y_{Sales\ Quantity)=}\beta_0 + \beta_1\ Time\_dummy*Storegroup + \beta_2\ Avg\_female + \beta_3\ Avg\_age + \beta_4\ Avg\_income + \beta_5\ Avg\_homeowner + \beta_6\ Avg\_residency + \beta_7 Avg\_childowner$$

Question 2 -What is the impact of implementing BOPs strategy on online return?

Model: OLS

$$log(Y_{Return\ Value)=}\beta_0 + \beta_1\ Time\_dummy*Storegroup + \beta_2 Avg\_female + \beta_3\ Avg\_age + \beta_4\ Avg\_income + \beta_5\ Avg\_homeowner + \beta_6\ Avg\_residency + \beta_7$$
$$Avg\_childowner + \beta_8 SalesValue$$

Model: Negative Binomial

$$log(Y_{Return\ Quantity)=}\beta_0 + \beta_1\ Time\_dummy*Storegroup + \beta_2 Avg\_female + \beta_3\ Avg\_age + \beta_4\ Avg\_income + \beta_5\ Avg\_homeowner + \beta_6\ Avg\_residency + \beta_7$$
$$Avg\_childowner + \beta_8 SalesQuantity$$

Q1 Sales Value

```
Regression Results
============================================
                       Dependent variable:
                       ---------------------
                       LogSalesValue
                       HW-Robust SE
--------------------------------------------
final_day                    0.26*
                            (0.13)

storegroup                   1.90***
                            (0.14)

avg_female                  -1.52**
                            (0.47)

avg_age                     -0.16**
                            (0.05)

avg_income                   0.32**
                            (0.10)

avg_homeowner               -0.26
                            (0.48)

avg_residency               -0.07
                            (0.04)

avg_childowner               0.62
                            (0.62)

final_day:storegroup        -0.41*
                            (0.17)

Constant                     8.80***
                            (0.65)

--------------------------------------------
Observations                 2,005
R2                           0.14
Adjusted R2                  0.13
Residual Std. Error          2.01
F Statistic                  35.63***
============================================
Note:              *p<0.05; **p<0.01; ***p<0.001
```

For stores that have BOPs implemented, there is a decrease in 41% in sales value.

Q1 Sales Quantity

```
Sales Quantity Negative Binomial
============================================
                       Dependent variable:
                       ---------------------
                       salesquantity
                       HW-Robust SE
--------------------------------------------
final_day                    1.21
                            (0.13)

storegroup                  15.07***
                            (0.13)

avg_female                   0.11***
                            (0.23)

avg_age                      0.85***
                            (0.02)

avg_income                   1.37***
                            (0.04)

avg_homeowner                0.85
                            (0.20)

avg_residency                0.98
                            (0.02)

avg_childowner               0.61*
                            (0.24)

final_day:storegroup         0.78
                            (0.15)

Constant                    89.66***
                            (0.32)

--------------------------------------------
Observations                 2,005
Log Likelihood           -13,395.60
theta                    0.58*** (0.02)
Akaike Inf. Crit.        26,811.19
============================================
Note:              *p<0.05; **p<0.01; ***p<0.001
```

(IRR) Interpretation: For stores that have BOPs implemented, there is a no significant change in sales quantity.

Q2 Return Value

```
Return Value Results
=========================================
                    Dependent variable:
                    ---------------------
                    LogReturnValue
                    HW-Robust SE
-----------------------------------------
final_day           0.94***
                    (0.25)

storegroup          2.07***
                    (0.25)

salesvalue          0.0000***
                    (0.0000)

avg_female          0.09
                    (0.55)

avg_age             -0.14*
                    (0.06)

avg_income          0.29*
                    (0.11)

avg_homeowner       -0.10
                    (0.57)

avg_residency       -0.11*
                    (0.05)

avg_childowner      1.08
                    (0.68)

final_day:storegroup -1.20***
                    (0.28)

Constant            4.57***
                    (0.74)

-----------------------------------------
Observations        2,005
R2                  0.37
Adjusted R2         0.36
Residual Std. Error 2.47
F Statistic         115.74***
-----------------------------------------
Note:        *p<0.05; **p<0.01; ***p<0.001
```

Q2 Return Quantity

```
Return Quantity Exponentiated Negative Binomial
=========================================
                    Dependent variable:
                    ---------------------
                    returnquantity
                    HW-Robust SE
-----------------------------------------
final_day           1.49***
                    (0.13)

storegroup          4.55***
                    (0.13)

salesquantity       1.00***

avg_female          0.75*
                    (0.23)

avg_age             0.90***
                    (0.02)

avg_income          1.16***
                    (0.04)

avg_homeowner       0.69*
                    (0.20)

avg_residency       0.98
                    (0.02)

avg_childowner      1.45*
                    (0.24)

final_day:storegroup 0.60***
                    (0.15)

Constant            3.21***
                    (0.32)

-----------------------------------------
Observations        2,005
Log Likelihood      -8,127.20
theta               1.06*** (0.04)
Akaike Inf. Crit.   16,276.41
=========================================
Note:        *p<0.05; **p<0.01; ***p<0.001
```

For stores that have BOPs implemented, there is a 120% decrease in return value.

(IRR): For stores that have BOPs implemented, there is a 40% decrease in return quantity.

Question 3 - What is the impact of using the BOPS service on online customer purchase behavior?

Dataset - Consumer Level Data
Model: <u>Negative Binomial</u>, Dependent Variable: <u>Sales Quantity</u>

$Y_{Sales\ Quantity} = \beta_0 + \beta_1 Bops\_in\_effect*Bops\_user + \beta_2 Avg\_female + \beta_3 Avg\_age + \beta_4 Avg\_income + \beta_5 Avg\_homeowner + \beta_6 Avg\_residency + \beta_7 Avg\_childowner$

Model: <u>OLS</u>, Dependent Variable: <u>Log(Sales Value + 1)</u>

$log(Y_{Sales\ Value}) = \beta_0 + \beta_1 Bops\_in\_effect*Bops\_user + \beta_2 Avg\_female + \beta_3 Avg\_age + \beta_4 Avg\_income + \beta_5 Avg\_homeowner + \beta_6 Avg\_residency + \beta_7 Avg\_childowner$

Q3 Sales Value

```
Sales Value Regression Results
========================================
                        Dependent variable:
                     ---------------------------
                         log(1 + salesvalue)
                     Normal SE       HW-Robust SE
                        (1)              (2)
----------------------------------------------------
bops_in_effect          0.033***         0.033***
                       (0.008)          (0.008)

bops_user              -0.208***        -0.208***
                       (0.013)          (0.013)

dummy_homeowner_code    0.029**          0.029**
                       (0.009)          (0.009)

dummy_child            -0.037***        -0.037***
                       (0.008)          (0.008)

age_band                0.005***         0.005***
                       (0.001)          (0.001)

est_income_code         0.009***         0.009***
                       (0.002)          (0.002)

length_of_residence    -0.001           -0.001
                       (0.001)          (0.001)

female                 -0.111***        -0.111***
                       (0.008)          (0.008)

bops_in_effect:bops_user 0.084***        0.084***
                       (0.019)          (0.019)

Constant                5.318***         5.318***
                       (0.013)          (0.012)
----------------------------------------------------
Observations            84,420           84,420
R2                      0.008            0.008
Adjusted R2             0.008            0.008
Residual Std. Error     1.067            1.067
F Statistic             79.147***        79.147***
----------------------------------------------------
Note:              *p<0.05; **p<0.01; ***p<0.001
```

Q3 Sales Quantity

```
Sales Quantity Regression Results
========================================
                        Dependent variable:
                     ---------------------------
                            salesquantity
                            Model-5 NB
----------------------------------------------------
bops_in_effect               1.10***
                            (0.01)

bops_user                    1.01
                            (0.01)

dummy_homeowner_code         1.02**
                            (0.01)

dummy_child                  0.99
                            (0.01)

age_band                     0.99***
                            (0.001)

est_income_code              1.02***
                            (0.002)

length_of_residence          1.00***
                            (0.001)

female                       1.46***
                            (0.01)

bops_in_effect:bops_user     1.18***
                            (0.02)

Constant                     1.77***
                            (0.01)
----------------------------------------------------
Observations                 84,420
Log Likelihood              -171,096.20
theta                        2.00*** (0.01)
Akaike Inf. Crit.            342,212.30
----------------------------------------------------
Note:              *p<0.05; **p<0.01; ***p<0.001
```

Interpretation: For a one unit increase in bops user when bops is in effect, expected sales value increases by 8.4%

(IRR): For a one unit increase in bops user when bops is in effect, the expected count increases by 18%

Question 4 - What is the impact of using the BOPS service on online customer return behavior?

We expect customers <u>using BOPS</u> to have an <u>increased likelihood of returning their purchase</u>.

*To assess this hypothesis, we use **transaction level data** to analyze, since it holds information on customers, purchase delivery method (BOPS vs home delivery), and purchase return information and also address endogeniety.*

Conceptual model:

$$Y_{Return} = \beta_0 + \beta_1 BOPS + \beta_2 log(Price) + \beta_3 Store\ Number + \beta_4 Age\ Band \\ + \beta_5 Month + \beta_6 Year + \beta_6 Product\ Category + \beta_6 Age\ Band + \\ \beta_7 Female + \beta_8 Has\ Child$$

Estimation results

$$Y_{Return} = \beta_0 + \beta_1 BOPS + \sum \beta_i Controls$$

$$X_1 = \gamma_0 + \gamma_1 Length\ of\ Residence + \sum \gamma_j Controls$$

Note: 2SLS model used in estimation

A transaction using BOPS service is associated with **an increase of 18 percentage points** in likelihood of product returns

Note: Control variables omitted for brevity
Link to full model output

Question 5 - What is the impact of implementing BOPS strategy on product-level sales and return?

Dataset - Online Daily Product Category Sales/Return

Model: <u>OLS</u>

$log(Y_{Sales\ Value)=}\beta_0 + \beta_1\ Time\_dummy*Storegroup + \beta_2\ Avg\_female + \beta_3\ Avg\_age + \beta_4\ Avg\_income + \beta_5\ Avg\_homeowner + \beta_6\ Avg\_residency + \beta_7\ Avg\_childowner + \beta_8\ Product\ Category$

Model: <u>Negative Binomial</u>

$log(Y_{Sales\ Quantity)=}\beta_0 + \beta_1\ Time\_dummy*Storegroup + \beta_2\ Avg\_female + \beta_3\ Avg\_age + \beta_4\ Avg\_income + \beta_5\ Avg\_homeowner + \beta_6\ Avg\_residency + \beta_7\ Avg\_childowner + \beta_8\ Product\ Category$

Model: <u>OLS</u>

$log(Y_{Return\ Value)=}\beta_0 + \beta_1\ Time\_dummy*Storegroup + \beta_2\ Avg\_female + \beta_3\ Avg\_age + \beta_4\ Avg\_income + \beta_5\ Avg\_homeowner + \beta_6\ Avg\_residency + \beta_7\ Avg\_childowner + \beta_8\ Product\ Category + \beta_9\ SalesValue$

Model: <u>Negative Binomial</u>

$log(Y_{Return\ Quantity)=}\beta_0 + \beta_1\ Time\_dummy*Storegroup + \beta_2\ Avg\_female + \beta_3\ Avg\_age + \beta_4\ Avg\_income + \beta_5\ Avg\_homeowner + \beta_6\ Avg\_residency + \beta_7\ Avg\_childowner + \beta_8\ Product\ Category + \beta_9\ SalesQuantity$

Q5 Sales Value

Dependent variable:	
	Logsalesvalue
	HW-Robust SE
final_day	0.18***
	(0.04)
storegroup	1.47***
	(0.04)
product_category	-0.01***
	(0.002)
avg_female	-1.06***
	(0.07)
avg_age	-0.06***
	(0.01)
avg_income	0.10***
	(0.02)
avg_homeowner	-0.13
	(0.09)
avg_residency	-0.01*
	(0.01)
avg_childowner	0.01
	(0.08)
final_day:storegroup	-0.20***
	(0.05)
Constant	7.00***
	(0.11)
Observations	21,003
R2	0.11
Adjusted R2	0.11
Residual Std. Error	1.84
F Statistic	256.84***
Note:	*p<0.05; **p<0.01; ***p<0.001

For stores that have implemented BOPS, there is a 20% decrease in sales value.

Q5 Sales Quantity

Sales Quantity Negative Binomial

Dependent variable:	
	salesquantity
	HW-Robust
final_day	1.30***
	(0.06)
storegroup	10.46***
	(0.05)
product_category	1.02***
	(0.002)
avg_female	0.63***
	(0.05)
avg_age	0.87***
	(0.01)
avg_income	1.47***
	(0.02)
avg_homeowner	1.23***
	(0.06)
avg_residency	1.05***
	(0.004)
avg_childowner	0.97
	(0.05)
final_day:storegroup	0.78***
	(0.07)
Constant	0.97
	(0.14)
Observations	21,003
Log Likelihood	-92,659.94
theta	0.49*** (0.004)
Akaike Inf. Crit.	185,341.90
Note:	*p<0.05; **p<0.01; ***p<0.001

(IRR)Interpretation: For stores that have implemented BOPS, there is a 22% decrease in sales quantity.

Q5 Return Value

```
Return Value
========================================
                     Dependent variable:
                     -------------------
                     Logreturnvalue
                     HW-Robust
----------------------------------------
final_day            0.40***
                     (0.08)

storegroup           2.43***
                     (0.08)

product_category     -0.01
                     (0.003)

avg_female           0.22*
                     (0.09)

avg_age              -0.10***
                     (0.01)

avg_income           0.12***
                     (0.02)

avg_homeowner        -0.19
                     (0.12)

avg_residency        -0.03**
                     (0.01)

avg_childowner       0.01
                     (0.11)

salesvalue           0.0001***
                     (0.0000)

final_day:storegroup -0.60***
                     (0.09)

Constant             1.59***
                     (0.15)

----------------------------------------
Observations         21,003
R2                   0.31
Adjusted R2          0.31
Residual Std. Error  2.92
F Statistic          863.34***
========================================
Note:         *p<0.05; **p<0.01; ***p<0.001
```

Interpretation: For stores that have implemented BOPS, there is a 60% decrease in return value.

Q5 Return Quantity

```
Return Quantity Negative Binomial
========================================
                     Dependent variable:
                     -------------------
                     returnquantity
                     HW-Robust
----------------------------------------
final_day            1.46***
                     (0.06)

storegroup           5.34***
                     (0.05)

product_category     0.99***
                     (0.002)

avg_female           1.15***
                     (0.03)

avg_age              0.92***
                     (0.004)

avg_income           1.07***
                     (0.01)

avg_homeowner        0.80***
                     (0.04)

avg_residency        0.98***
                     (0.003)

avg_childowner       0.98
                     (0.03)

salesquantity        1.01***
                     (0.0000)

final_day:storegroup 0.66***
                     (0.06)

Constant             0.54***
                     (0.06)

----------------------------------------
Observations         21,003
Log Likelihood       -44,007.13
theta                0.78*** (0.01)
Akaike Inf. Crit.    88,038.26
========================================
Note:         *p<0.05; **p<0.01; ***p<0.001
```

(IRR): For stores that have implemented BOPS, there is a 34% decrease in return quantity.

Q6 - Sales Value

*The marginal effect plots for all product categories **Sales Value***

*We found that **only 2 out of the 21** categories have a significant difference.*

1. *Product #4 (Diamond Fashion)*
2. *Product #14 (Pre-Owned)*
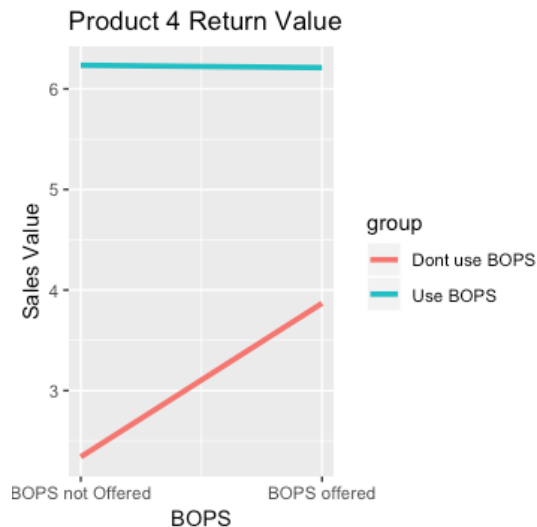

Sales Value Product Category Graph

Product 4 Sales Value

Sales Value vs BOPS
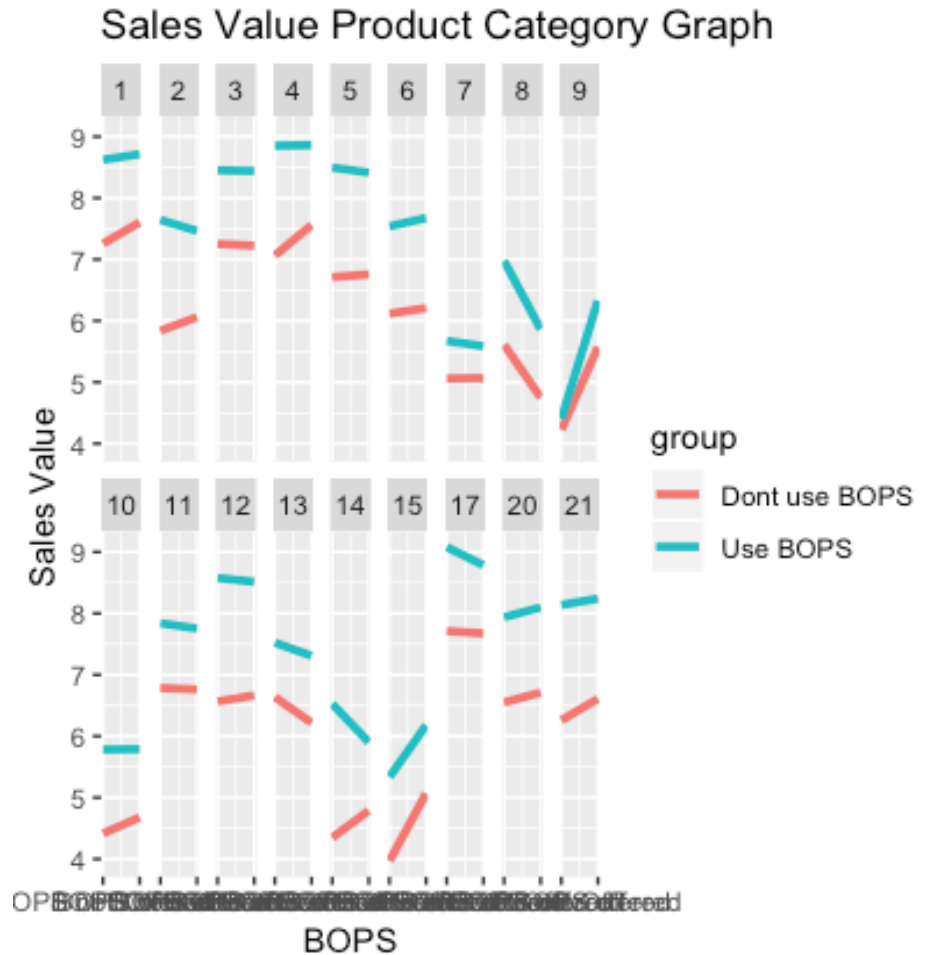
group
- Dont use BOPS
- Use BOPS
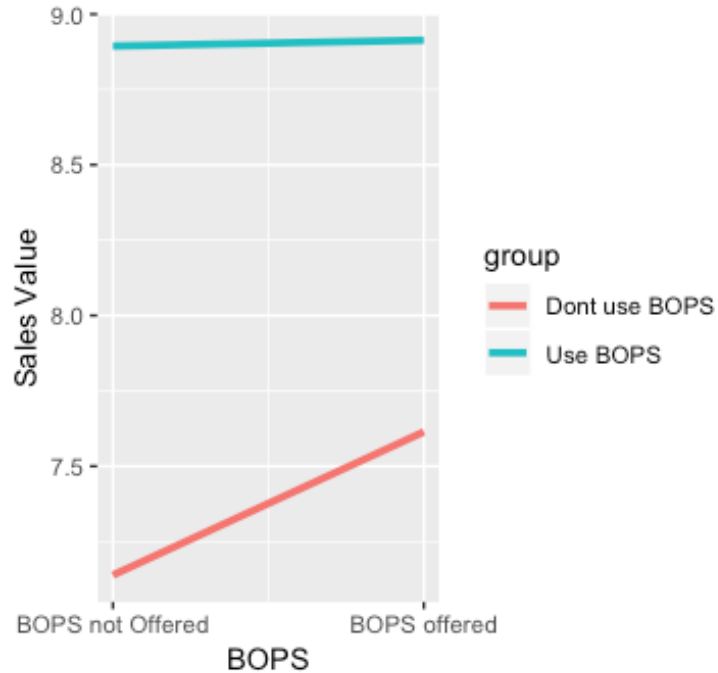
```
final_day:storegroup          -0.46**
                               (0.15)
```

Interpretation: When BOPS is available and in use, it led to a decrease of 46% in sales value for Diamond Fashion products

Product 14 Sales Value

Sales Value vs BOPS

group
- Dont use BOPS
- Use BOPS
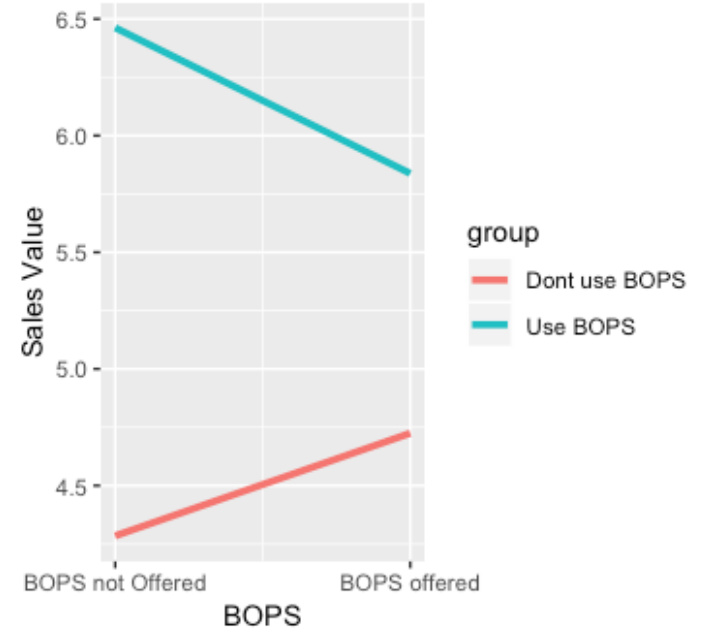
```
final_day:storegroup          -1.06***
                               (0.19)
```

Interpretation: When BOPS is available and in use, it led to a decrease of 106% in sales value for Pre-Owned products

Q6 - Return Value

The Marginal plot for all categories
Return Value

*We found that **3 out of the 21** categories have a significant difference*

1. *Product #4 (Diamond Fashion)*
2. *Product #14 (Pre-Owned)*
3. *Product #21 (Sterling Silver)*

Marginal Plot for Return Value
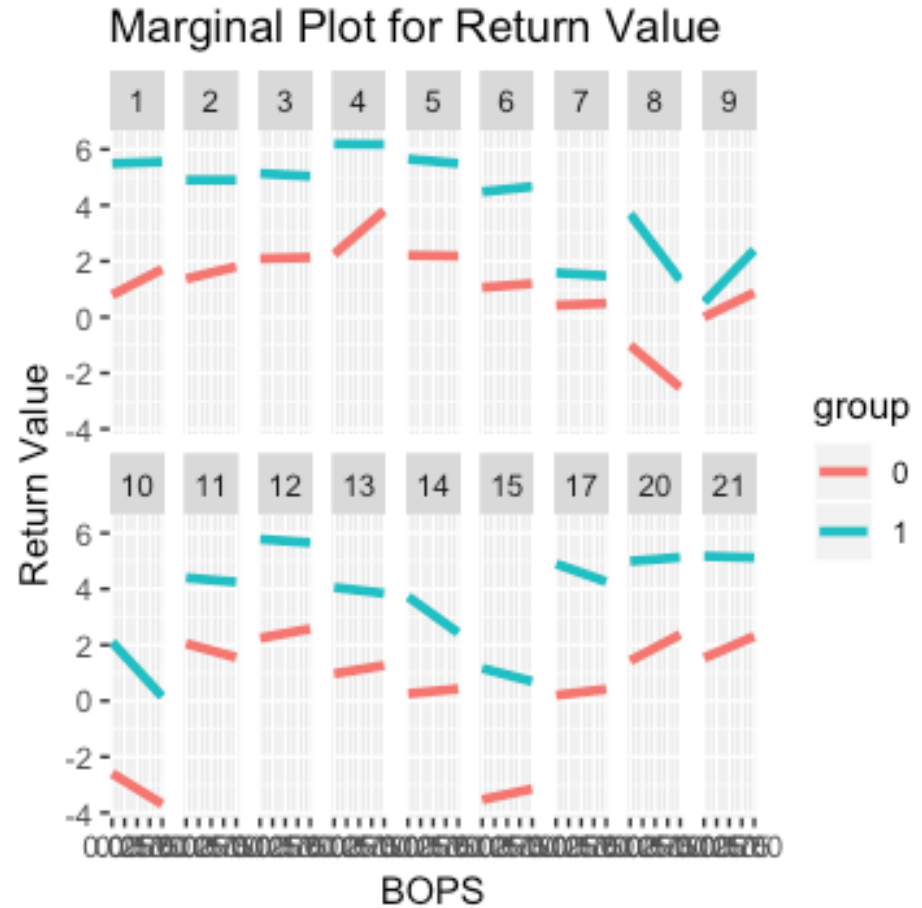
Product Categories with Significant Changes
With BOPS:
In these product categories, the use of BOPS led to a decrease of 155%, 148%, and 90% in sales value.

Category 4

Category 14

Category 21

final_day:storegroup   -1.55***
                               (0.34)

Product 4 Return Value

final_day:storegroup   -1.48***
                               (0.29)

Product 14 Return Value

final_day:storegroup   -0.90**
                               (0.30)

Product 21 Return Value
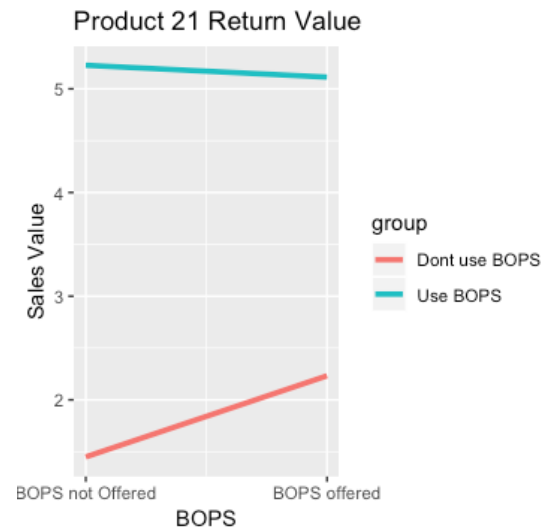
To recap

Question	Model	Outcome	Dataset
Q1 - Impact of BOPs on online channel sales.	OLS	**Decrease in sales value by 40%.** **No change in sales quantity.**	**Online Daily Data - Sales and Return**
Q2 - Impact of BOPs on online channel returns.	Negative Binomial	**Decrease in return value by 120%** **Decrease in return quantity by 40%.**	**Online Daily Data - Sales and Return**
Q3 - Impact of using BOPS service on online customer purchase behavior.	OLS and Neg Bin	**Increase in sales value by 8.4%** **Increase in sales quantity by 18%**	**Consumer Level Data**
Q4 - Impact of using BOPs service on online customer return behavior.	OLS, Probit, and IV/2SLS	**Increase in return likelihood by 18 p.p.**	**Transaction Level Data**
Q5 - Impact of implementing BOPS strategy on product level sales and returns.	OLS and Neg Bin	**Decrease in sales value by 20% and 22% in sales quantity. Decrease in return value by 60% and 34% decrease in return quantity.**	**Online Daily Data - Product Category**
Q6 - Impact of implementing BOPS strategy vary across product categories.	OLS	**The effect of implementing BOPS strategy depends on product category number.**	**Online Daily Data - Product Category**

Managerial Insight

- We **do not recommend** a universal BOPS rollout because the effects differ at different levels of analysis:
  - Limit BOPS to loyal customers since repeat customers bought more when using BOPS.
  - Target specific product categories such as product #21, sterling silver
  - We gather more data on online user transaction
    - ROPO - Research Online, Purchase Offline
  - We gather more data on more channels that have implemented BOPS

Limitations

- We don't know how much the implementation of BOPS costs and if it is scalable.
  - If implementation of BOPS results in insignificant change in net profits, implementation results in a loss.
- We indicate that a consumer uses BOPS due to convenience (near the store), however we do not have any data of how far consumers are from the store
  - Person using BOPS likely lives near store, making it convenient for them to return in store (vs by mail). Dataset lacks distance to store information.
- We don't know if customer returns are converted to in-store credit or exchange.
  - Without this information, we could be overestimating effect of BOPS on returns.
- We were not given full details on the return policy of the stores and online channels
- Cost for customers to use BOPS versus being delivered

Appendix

Summary Statistics - Online Daily Returns/Sales before replacing NA with mean

```
Online Sales and Return Level Descriptive Statistics
=================================================================================
Statistic          N      Mean      St. Dev.   Min   Pctl(25)  Median   Pctl(75)      Max
---------------------------------------------------------------------------------
store_number     2,005  1,579.44   2,639.11    2       2        6        5,998       5,998
year             2,005  2,011.18     0.70    2,010   2,011    2,011      2,012       2,012
month_index      2,005    25.78      7.26     13      20       26         32          38
month_dummy      2,005     6.62      3.41      1       4        7          9          12
bops_in_effect   2,005     0.39      0.49      0       0        0          1           1
day              2,005   404.81    221.04      1      215      411        593         785
salesvalue       2,005  91,493.44 164,109.90 0.00  7,248.05 16,249.74 148,777.00 1,413,919.00
returnvalue      2,005  14,406.62  24,454.22 0.00    681.43  2,387.69  24,014.74   192,876.80
salesquantity    2,005   523.04    989.23      1      38       98         760        8,933
returnquantity   2,005    55.16     97.77      0       3        10         88          876
avg_female       1,583     0.52      0.18     0.00    0.47     0.53       0.60        1.00
avg_age          1,634     5.07      1.67     0.00    4.64     5.05       5.64       13.00
avg_income       1,633     5.38      0.89     1.00    5.17     5.40       5.57        9.00
avg_homeowner    1,633     0.66      0.19     0.00    0.62     0.67       0.72        1.00
avg_residency    1,633     7.06      2.13     0.00    6.66     7.04       7.48       15.00
avg_childowner   1,633     0.35      0.17     0.00    0.33     0.36       0.40        1.00
LogSalesValue    2,005     9.85      2.16     0.00    8.89     9.70      11.91       14.16
LogReturnValue   2,005     7.42      3.10     0.00    6.53     7.78      10.09       12.17
LogSalesQuantity 2,005     4.73      1.99     0.00    3.64     4.58       6.63        9.10
LogReturnQuantity2,005     2.67      1.74     0.00    1.39     2.40       4.49        6.78
final_day        2,005     0.56      0.50      0       0        1          1           1
storegroup       2,005     0.74      0.44      0       0        1          1           1
```

24

Summary Statistics - Replace NA with Average

```
Online Sales and Return Level Descriptive Statistics
=================================================================================
Statistic          N     Mean       St. Dev.     Min   Pctl(25)  Median    Pctl(75)      Max
---------------------------------------------------------------------------------
store_number       2,005 1,579.44   2,639.11     2     2         6         5,998         5,998
year               2,005 2,011.18   0.70         2,010 2,011     2,011     2,012         2,012
month_index        2,005 25.78      7.26         13    20        26        32            38
month_dummy        2,005 6.62       3.41         1     4         7         9             12
bops_in_effect     2,005 0.39       0.49         0     0         0         1             1
day                2,005 404.81     221.04       1     215       411       593           785
salesvalue         2,005 91,493.44  164,109.90   0.00  7,248.05  16,249.74 148,777.00    1,413,919.00
returnvalue        2,005 14,406.62  24,454.22    0.00  681.43    2,387.69  24,014.74     192,876.80
salesquantity      2,005 523.04     989.23       1     38        98        760           8,933
returnquantity     2,005 55.16      97.77        0     3         10        88            876
avg_female         2,005 0.52       0.16         0.00  0.49      0.53      0.57          1.00
avg_age            2,005 5.03       1.51         0.00  4.73      4.87      5.49          13.00
avg_income         2,005 5.37       0.81         1.00  5.24      5.34      5.52          9.00
avg_homeowner      2,005 0.66       0.17         0.00  0.64      0.66      0.70          1.00
avg_residency      2,005 7.04       1.92         0.00  6.75      7.00      7.33          15.00
avg_childowner     2,005 0.36       0.15         0.00  0.34      0.37      0.39          1.00
LogSalesValue      2,005 9.85       2.16         0.00  8.89      9.70      11.91         14.16
LogReturnValue     2,005 7.42       3.10         0.00  6.53      7.78      10.09         12.17
LogSalesQuantity   2,005 4.73       1.99         0.00  3.64      4.58      6.63          9.10
LogReturnQuantity  2,005 2.67       1.74         0.00  1.39      2.40      4.49          6.78
final_day          2,005 0.56       0.50         0     0         1         1             1
storegroup         2,005 0.74       0.44         0     0         1         1             1
---------------------------------------------------------------------------------
```

Q1 Analysis of Control Variables

1. **Avg_Female** - the ratio of all female customers to all customers
2. **Avg_Age** - average age_band of all customers at the given aggregation level and note that this is not the actual age. It is the average of age groups. The higher it is, the older the customer profile is
3. **Avg_Income** - the average income_band of all customers at the given aggregation level. It is the average of income groups. The higher it is, the richer the customer profile is
4. **Avg_Homeowner** - the ratio of customers who own their house to all customers (homeowners + renters)
5. **Avg_Residency** - the average number of years spent in the current address for customers
   Avg_Childowner - the ratio of customers who have at least one child to all customers

Exclude: Store Number, Year, Product Category, Month, Month_index, Month_dummy, and Bops_In_Effect

Q1 Sales Value/Quantity - Heteroskedasticity

Sales Value

Sales Quantity

```
> gqtest(sales1)

        Goldfeld-Quandt test

data:  sales1
GQ = 0.61199, df1 = 993, df2 = 992, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(sales1) #Heteroskedastic

        studentized Breusch-Pagan test

data:  sales1
BP = 59.584, df = 9, p-value = 0.000000001612
```

```
> gqtest(salesquantnb)

        Goldfeld-Quandt test

data:  salesquantnb
GQ = 0.0044378, df1 = 993, df2 = 992, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(salesquantnb) #Heteroskedastic

        studentized Breusch-Pagan test

data:  salesquantnb
BP = 63.574, df = 9, p-value = 0.000000000273
```

Fixed with HW Robust Standard errors!

Q1/2 - Online Daily Sales/Returns Multicollinearity

```
                        finalosdr.avg_age finalosdr.avg_childowner finalosdr.avg_homeowner finalosdr.avg_residency finalosdr.avg_female
finalosdr.avg_age             1.0000000               0.13977814              0.21368975              0.24988711              0.04267430
finalosdr.avg_childowner      0.1397781               1.00000000              0.29972108              0.12219873              0.03177948
finalosdr.avg_homeowner       0.2136898               0.29972108              1.00000000              0.33552799              0.03972333
finalosdr.avg_residency       0.2498871               0.12219873              0.33552799              1.00000000              0.09759838
finalosdr.avg_female          0.0426743               0.03177948              0.03972333              0.09759838              1.00000000
finalosdr.avg_income          0.1136697               0.07497034              0.30774555              0.15455566             -0.07291792
                        finalosdr.avg_income
finalosdr.avg_age             0.11366971
finalosdr.avg_childowner      0.07497034
finalosdr.avg_homeowner       0.30774555
finalosdr.avg_residency       0.15455566
finalosdr.avg_female         -0.07291792
finalosdr.avg_income          1.00000000
> corrplot(corr, type = "upper", tl.pos = "td", method = "circle", tl.cex = 0.5, tl.col = 'black', diag = TRUE)
> #There is no multicollinearity in in demographic variables.
> vifcor(df) # All less than 3 if time variables are not included in data frame.
No variable from the 6 input variables has collinearity problem.

The linear correlation coefficients ranges between:
min correlation ( finalosdr.avg_female ~ finalosdr.avg_childowner ):  0.03177948
max correlation ( finalosdr.avg_residency ~ finalosdr.avg_homeowner ):  0.335528

---------- VIFs of the remained variables --------
             Variables      VIF
1       finalosdr.avg_age 1.097856
2 finalosdr.avg_childowner 1.106963
3  finalosdr.avg_homeowner 1.330623
4  finalosdr.avg_residency 1.182994
5    finalosdr.avg_female 1.019647
6    finalosdr.avg_income 1.121032
```
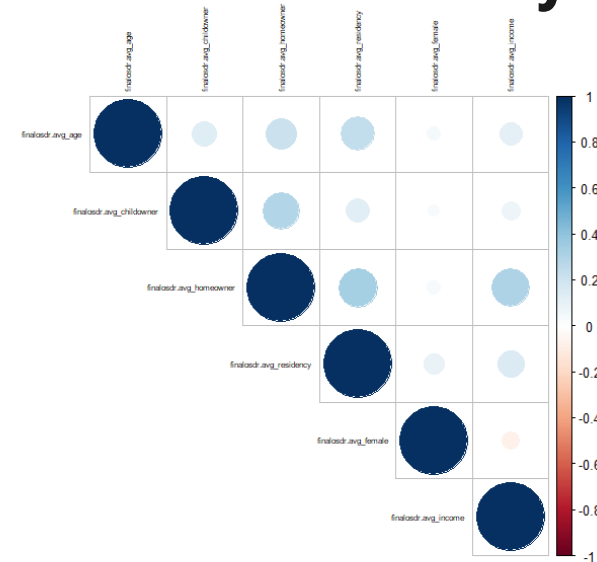
Question 3 - Multicollinearity

	consumerdata.store_number	consumerdata.age_band	consumerdata.length_of_residence	consumerdata.bops_in_effect
consumerdata.store_number	1.0000000000	0.000318228	-0.001161459	-0.0005940805
consumerdata.age_band	0.0003182280	1.000000000	0.140020955	0.0000000000
consumerdata.length_of_residence	-0.0011614587	0.140020955	1.000000000	0.0000000000
consumerdata.bops_in_effect	-0.0005940805	0.000000000	0.000000000	1.0000000000
consumerdata.dummy_homeowner_code	-0.0009660944	0.187316469	0.289011449	0.0000000000
consumerdata.dummy_child	-0.0005233652	0.006137120	-0.025386381	0.0000000000
consumerdata.female	-0.0022130349	0.072834704	0.028422111	0.0000000000

	consumerdata.dummy_homeowner_code	consumerdata.dummy_child	consumerdata.female
consumerdata.store_number	-0.0009660944	-0.0005233652	-0.002213035
consumerdata.age_band	0.1873164691	0.0061371201	0.072834704
consumerdata.length_of_residence	0.2890114486	-0.0253863805	0.028422111
consumerdata.bops_in_effect	0.0000000000	0.0000000000	0.000000000
consumerdata.dummy_homeowner_code	1.0000000000	0.2085613093	0.007922949
consumerdata.dummy_child	0.2085613093	1.0000000000	0.038860043
consumerdata.female	0.0079229486	0.0388600428	1.000000000

Q3 - Sales Quantity, Choosing a Model

```
Model 1: salesquantity ~ bops_in_effect * bops_user + dummy_homeowner_code +
    dummy_child + age_band + est_income_code + length_of_residence +
    female
Model 2: salesquantity ~ bops_in_effect * bops_user + dummy_homeowner_code +
    dummy_child + age_band + est_income_code + length_of_residence +
    female
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  11 -171095
2  10 -223989 -1 105787  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Model 1: salesquantity ~ bops_in_effect * bops_user + dummy_homeowner_
    dummy_child + age_band + est_income_code + length_of_residence +
    female
Model 2: salesquantity ~ 1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  11 -171095
2   2 -173071 -9 3952.4  < 2.2e-16 ***
```

Q3 - Sales Quantity Heteroskedasticity

Normal Q-Q Plot
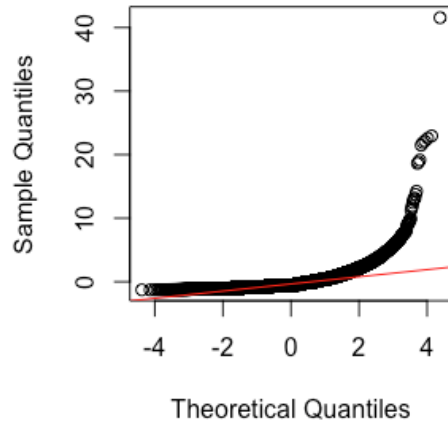

```
> gqtest(model5a)

        Goldfeld-Quandt test

data:  model5a
GQ = 0.1403, df1 = 42200, df2 = 42200, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(model5a) #both tests showed no heteroskedasticity

        studentized Breusch-Pagan test

data:  model5a
```
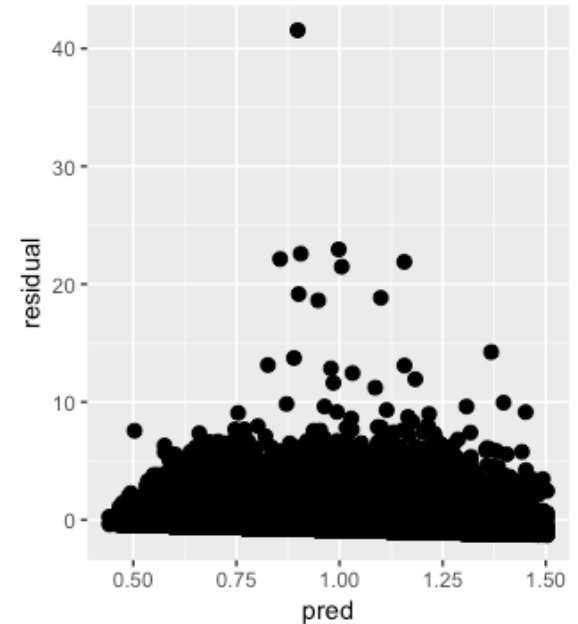
Q3 - Sales Value Heteroskedasticity
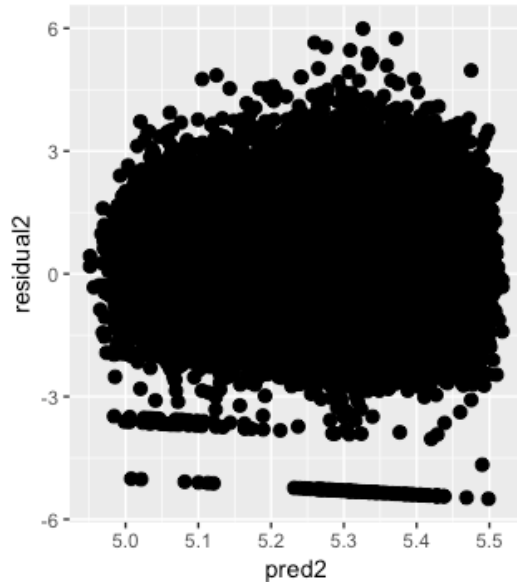

```
> gqtest(model6) #showed no heteroskedasticity

        Goldfeld-Quandt test

data:  model6
GQ = 0.87052, df1 = 42200, df2 = 42200, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(model6) #showed heteroskedasticity

        studentized Breusch-Pagan test

data:  model6
BP = 493.98, df = 9, p-value < 2.2e-16
```


Normal Q-Q Plot
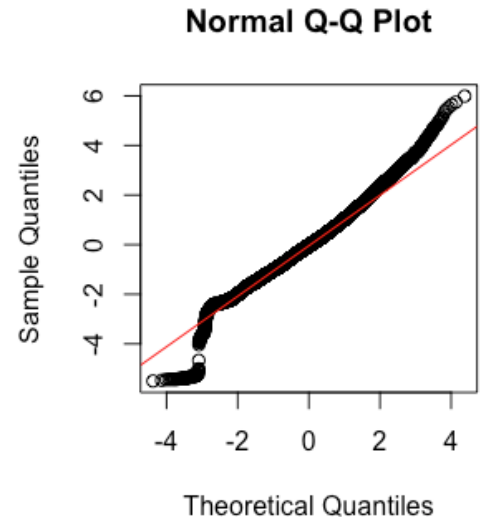
Q3 - Marginal Plot
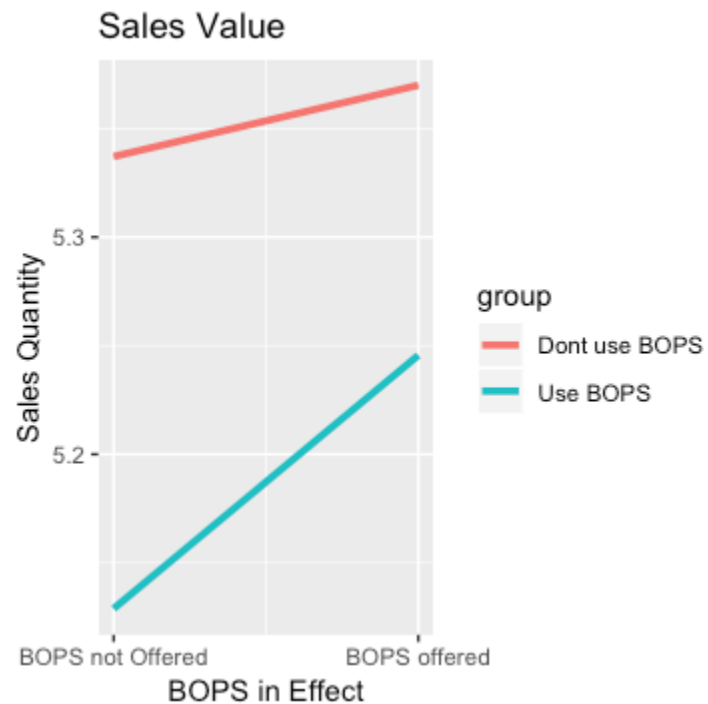
Question 4 - Probit model approach

BOPS coefficient is significant. However, interpreting the coefficient for a probit model is difficult to understand

```
Regression Results
===============================================
                    Dependent variable:
                  ---------------------------
                            return
                  Normal SE        HW-Robust SE
                     (1)                (2)
-----------------------------------------------
bops             0.105***            0.105***
                  (0.004)            (0.004)

Constant         -2.270***          -2.270***
                  (0.021)            (0.018)


-----------------------------------------------
Observations     1,170,564          1,170,564
Log Likelihood   -362,820.700       -362,820.700
Akaike Inf. Crit. 725,717.400       725,717.400
===============================================
Note:            *p<0.05; **p<0.01; ***p<0.001
```

Note: Control variables omitted for brevity
Link to full model output

Question 4 - Comparable to OLS model

To compare OLS and probit models, we first have to **generate the marginal effects** of the probit model *(left column)*.

Since the marginal effects coefficient of BOPS in the probit model **is similar to that of the OLS model**, we can assume that the OLS model is a good estimator as well.

```
Regression Results
===============================================
                      Dependent variable:
               --------------------------------
                            return
                    probit              OLS
              Marg Eff w/ RobStdErr  OLS w/ RobStdErr
                     (1)                (2)
-----------------------------------------------
bops              0.0176***          0.0158***
                  (0.0007)           (0.0007)


-----------------------------------------------
Observations      1,170,564          1,170,564
R2                                   0.0279
Adjusted R2                          0.0279
Log Likelihood   -362,820.7000
Akaike Inf. Crit. 725,717.4000
Residual Std. Error                  0.2956
F Statistic                          908.3410***
===============================================
Note:                *p<0.05; **p<0.01; ***p<0.001
```

Note: Control variables omitted for brevity
Link to full model output

Question 4 - Addressing endogeneity

Conceptually, the key ind. var. BOPS is a decision variable that is correlated with our dependent variable. Thus, endogeneity is present in our model:

- We address this endogeneity by utilizing **length of residence** as an instrumental variable for BOPS
- Why length of residence as proxy?
  - We believe **the longer a customer has lived at an address, the less likely they are to use BOPS service** (due to familiarity of safety within a neighborhood)

$$Y_{Return} = \beta_0 + \beta_1 BOPS + \sum \beta_i Controls$$

$$X_1 = \gamma_0 + \gamma_1 Length\ of\ Residence + \sum \gamma_j Controls$$

Question 4 - Summary Statistics

```
   customer_id          purchase_date        transaction_id       store_number         price              sku
Min.   :      100348   Length:1170568   Min.   :  50002    Min.   :   2.0   Min.   :    0.00   Min.   :11373024
1st Qu.:    28454419   Class :character 1st Qu.:3262163    1st Qu.:   2.0   1st Qu.:   44.24   1st Qu.:17951286
Median :    31964480   Mode  :character Median :3787321    Median :   2.0   Median :   89.99   Median :18214460
Mean   : 23247802315                    Mean   :3751126    Mean   : 180.4   Mean   :  170.16   Mean   :18424852
3rd Qu.:    35423835                    3rd Qu.:4236822    3rd Qu.:   2.0   3rd Qu.:  186.99   3rd Qu.:18697427
Max.   :919650001519                    Max.   :4702552    Max.   :5998.0   Max.   :39422.00   Max.   :80006100

     return             age_band          est_income_code   homeowner_code      length_of_residence     child
Min.   :0.00000    Min.   : 0.000    Min.   : 0.000    Length:1170568    Min.   : 0.000    Length:1170568
1st Qu.:0.00000    1st Qu.: 0.000    1st Qu.: 0.000    Class :character  1st Qu.: 2.000    Class :character
Median :0.00000    Median : 5.000    Median : 5.000    Mode  :character  Median : 6.000    Mode  :character
Mean   :0.09983    Mean   : 4.832    Mean   : 4.856                      Mean   : 7.178
3rd Qu.:0.00000    3rd Qu.: 8.000    3rd Qu.: 8.000                      3rd Qu.:13.000
Max.   :1.00000    Max.   :13.000    Max.   :13.000                      Max.   :15.000

     year              month            month_index       product_category  month_dummy        week_index          bops
Min.   :2011    Length:1170568    Min.   :25.00     Min.   : 1.0    Min.   : 1.000    Min.   : 53.0    Min.   :0.0000
1st Qu.:2012    Class :character  1st Qu.:30.00     1st Qu.: 5.0    1st Qu.: 4.000    1st Qu.: 78.0    1st Qu.:0.0000
Median :2012    Mode  :character  Median :39.00     Median : 9.0    Median : 8.000    Median :117.0    Median :0.0000
Mean   :2012                      Mean   :37.03     Mean   :10.1    Mean   : 7.387    Mean   :107.1    Mean   :0.2376
3rd Qu.:2013                      3rd Qu.:42.00     3rd Qu.:12.0    3rd Qu.:12.000    3rd Qu.:129.0    3rd Qu.:0.0000
Max.   :2013                      Max.   :48.00     Max.   :21.0    Max.   :12.000    Max.   :157.0    Max.   :1.0000
                                                    NA's   :4

    female            hasChild           ownHome           logprice
Min.   :0.000    Min.   :0.0000    Min.   :0.0000    Min.   : 0.000
1st Qu.:0.000    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.: 3.812
Median :0.490    Median :0.0000    Median :1.0000    Median : 4.511
Mean   :0.488    Mean   :0.4019    Mean   :0.6644    Mean   : 4.466
3rd Qu.:1.000    3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.: 5.236
Max.   :1.000    Max.   :1.0000    Max.   :1.0000    Max.   :10.582
```

Notes on data clean-up:

- **Data was subset** to only include data points when **bops was implemented**.
- Missing data points were replaced with sample mean (e.g. *age_band*, *length_of_residence*, *female*), except for "Product Category" variable.
- "hasChild" is dummy variable for "child".
- "ownHome" is dummy variable for "homeowner_code".
- "Logprice" created as "price" variable was right-skewed.

Question 4 - Standardization of Time

Since we are concerned with the effect of BOPS on return, we will exclude all data points when BOPS was not implemented (i.e. BOPS = n/a):

| None | BOPS | BOPS |

August 1, 2010

Day 366

Day 786

July 31, 2013

Question 4 - Full Probit Output

Regression Results	Normal SE (1)	HW-Robust SE (2)
Dependent variable: return		
bops	0.105*** (0.004)	0.105*** (0.004)
logprice	0.223*** (0.003)	0.223*** (0.002)
factor(store_number)6	-0.034*** (0.007)	-0.034*** (0.007)
factor(store_number)5998	-0.184*** (0.011)	-0.184*** (0.010)
factor(month_dummy)2	-0.130*** (0.007)	-0.130*** (0.007)
factor(month_dummy)3	-0.063*** (0.008)	-0.063*** (0.009)
factor(month_dummy)4	-0.117*** (0.009)	-0.117*** (0.009)
factor(month_dummy)5	-0.164*** (0.008)	-0.164*** (0.008)
factor(month_dummy)6	-0.100*** (0.009)	-0.100*** (0.009)
factor(month_dummy)7	-0.133*** (0.009)	-0.133*** (0.009)
factor(month_dummy)8	-0.095*** (0.010)	-0.095*** (0.010)
factor(month_dummy)9	-0.117*** (0.010)	-0.117*** (0.010)
factor(month_dummy)10	-0.094*** (0.009)	-0.094*** (0.009)
factor(month_dummy)11	-0.135*** (0.008)	-0.135*** (0.008)
factor(month_dummy)12	-0.179*** (0.007)	-0.179*** (0.007)
factor(year)2012	-0.046*** (0.005)	-0.046*** (0.005)
factor(year)2013	-0.067*** (0.007)	-0.067*** (0.007)
factor(product_category)2	0.348*** (0.011)	0.348*** (0.012)
factor(product_category)3	-0.040** (0.012)	-0.040** (0.015)
factor(product_category)4	0.032*** (0.009)	0.032*** (0.010)
factor(product_category)5	-0.025** (0.010)	-0.025* (0.011)
factor(product_category)14	0.179*** (0.020)	0.179*** (0.020)
factor(product_category)15	-2.393 (25.607)	-2.393*** (0.147)
factor(product_category)17	-0.186*** (0.027)	-0.186*** (0.032)
factor(product_category)20	0.219*** (0.011)	0.219*** (0.013)
factor(product_category)21	0.009 (0.010)	0.009 (0.011)
age_band	-0.003*** (0.0004)	-0.003*** (0.0004)
female	0.169*** (0.004)	0.169*** (0.004)
hasChild	0.014*** (0.003)	0.014*** (0.003)
Constant	-2.270*** (0.021)	-2.270*** (0.018)
Observations	1,170,564	1,170,564
Log Likelihood	-362,820.700	-362,820.700
Akaike Inf. Crit.	725,717.400	725,717.400
Note:		*p<0.05; **p<0.01; ***p<0.001

Question 4 - Likelihood Ratio Test & Predictive Power

```
Likelihood ratio test

Model 1: return ~ bops + factor(store_number) + logprice + age_band +
    female + hasChild
Model 2: return ~ 1
  #Df  LogLik Df Chisq Pr(>Chisq)
1   8 -365500
2   1 -380104 -7 29208  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To test model fit, a new probit model with fewer control variables was created (Model 1) and compared to a null model:

- The test is significant, indicating that the model with variables beats the null model.
- We can conclude that our initial model (with more variables) fits the data.

```
> print(paste('Accuracy', 1 - misClasificError))  # Accuracy = 90.02% - Damn good.
[1] "Accuracy 0.900165560650898"
```

This model has a very good predictive power

Question 4 - Full OLS v Probit

```
Regression Results
===================================================
                        Dependent variable:
                   --------------------------------
                               return
                     probit           OLS
                Marg Eff w/ RobStdErr OLS w/ RobStdErr
                       (1)              (2)
---------------------------------------------------
bops                 0.0176***        0.0158***
                    (0.0007)         (0.0007)

logprice             0.0359***        0.0292***
                    (0.0003)         (0.0002)

factor(store_number)6    -0.0054***   -0.0052***
                    (0.0011)         (0.0011)

factor(store_number)5998 -0.0264***   -0.0290***
                    (0.0013)         (0.0015)

factor(month_dummy)2   -0.0196***     -0.0285***
                    (0.0010)         (0.0013)

factor(month_dummy)3   -0.0097***     -0.0151***
                    (0.0013)         (0.0017)

factor(month_dummy)4   -0.0176***     -0.0184***
                    (0.0012)         (0.0015)

factor(month_dummy)5   -0.0241***     -0.0327***
                    (0.0010)         (0.0014)
```

```
factor(month_dummy)7   -0.0197***     -0.0287***
                    (0.0013)         (0.0017)

factor(month_dummy)8   -0.0144***     -0.0205***
                    (0.0014)         (0.0018)

factor(month_dummy)9   -0.0176***     -0.0241***
                    (0.0014)         (0.0019)

factor(month_dummy)10  -0.0143***     -0.0178***
                    (0.0014)         (0.0017)

factor(month_dummy)11  -0.0203***     -0.0283***
                    (0.0011)         (0.0015)

factor(month_dummy)12  -0.0272***     -0.0359***
                    (0.0011)         (0.0013)

factor(year)2012       -0.0075***     -0.0078***
                    (0.0008)         (0.0008)

factor(year)2013       -0.0106***     -0.0093***
                    (0.0010)         (0.0012)

factor(product_category)2  0.0689***  0.0411***
                    (0.0028)         (0.0030)

factor(product_category)3  -0.0063**  -0.0109***
                    (0.0022)         (0.0038)

factor(product_category)4  0.0052**   -0.0193***
                    (0.0017)         (0.0025)

factor(product_category)5  -0.0041*   -0.0340***
                    (0.0017)         (0.0025)
```

```
factor(product_category)14   0.0325***   -0.0037
                    (0.0041)         (0.0041)

factor(product_category)15   -0.0891***  -0.1732***
                    (0.0003)         (0.0092)

factor(product_category)17   -0.0264***  -0.0246**
                    (0.0040)         (0.0085)

factor(product_category)20   0.0404***   0.0405***
                    (0.0027)         (0.0035)

factor(product_category)21   0.0015      -0.0353***
                    (0.0017)         (0.0025)

age_band             -0.0005***        -0.0004***
                    (0.0001)         (0.0001)

female                0.0272***         0.0274***
                    (0.0006)         (0.0006)

hasChild              0.0022***         0.0021***
                    (0.0005)         (0.0006)

--------------------------------------------------
Observations           1,170,564        1,170,564
R2                                      0.0279
Adjusted R2                             0.0279
Log Likelihood       -362,820.7000
Akaike Inf. Crit.    725,717.4000
Residual Std. Error                     0.2956
F Statistic                             908.3410***
==================================================
Note:                   *p<0.05; **p<0.01; ***p<0.001
```

Question 4 - IV Reg Diagnostics

Using length of residence as an instrument variable. Note:

- Sargan test unavailable since only one instrument used
- F-statistic of 163 > 10 - the instrument in use is relevant
- Significant Hausman test, so we will interpret the IVReg/2SLS model estimation

```
Diagnostic tests:
                    df1     df2 statistic p-value
Weak instruments      1 1170526   163.533 < 2e-16 ***
Wu-Hausman            1 1170525     9.816 0.00173 **
Sargan                0     NA        NA      NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3043 on 1170526 degrees of freedom
Multiple R-Squared: -0.03044,   Adjusted R-squared: -0.03047
Wald test: 842.4 on 37 and 1170526 DF,  p-value: < 2.2e-16
```

Question 4 - Full IV Reg Output

```
Regression Results
========================================================
                          Dependent variable:
                          --------------------------
                                  return
                          Normal SE        HW-Robust SE
                            (1)                (2)
--------------------------------------------------------
bops                     0.1892***          0.1892***
                         (0.0570)           (0.0569)

logprice                 0.0290***          0.0290***
                         (0.0003)           (0.0003)

factor(store_number)6     0.0115*            0.0115*
                         (0.0056)           (0.0056)

factor(store_number)5998 -0.0578***        -0.0578***
                         (0.0096)           (0.0096)

factor(month_dummy)2     -0.0314***        -0.0314***
                         (0.0016)           (0.0017)

factor(month_dummy)3     -0.0201***        -0.0201***
                         (0.0023)           (0.0024)

factor(month_dummy)4     -0.0205***        -0.0205***
                         (0.0016)           (0.0017)

factor(month_dummy)5     -0.0344***        -0.0344***
                         (0.0015)           (0.0015)

factor(month_dummy)6     -0.0273***        -0.0273***
                         (0.0023)           (0.0024)
```

```
factor(month_dummy)7     -0.0350***        -0.0350***
                         (0.0026)           (0.0027)

factor(month_dummy)8     -0.0162***        -0.0162***
                         (0.0023)           (0.0024)

factor(month_dummy)9     -0.0331***        -0.0331***
                         (0.0035)           (0.0035)

factor(month_dummy)10    -0.0245***        -0.0245***
                         (0.0028)           (0.0028)

factor(month_dummy)11    -0.0412***        -0.0412***
                         (0.0045)           (0.0045)

factor(month_dummy)12    -0.0479***        -0.0479***
                         (0.0041)           (0.0042)

factor(year)2012         -0.0284***        -0.0284***
                         (0.0068)           (0.0068)

factor(year)2013         -0.0326***        -0.0326***
                         (0.0078)           (0.0078)

factor(product_category)2 0.0408***         0.0408***
                         (0.0024)           (0.0031)

factor(product_category)3 -0.0067*          -0.0067
                         (0.0033)           (0.0041)

factor(product_category)4  -0.0059           -0.0059
                         (0.0048)           (0.0051)
```

```
factor(product_category)15 -0.1453          -0.1453***
                          (0.1524)          (0.0139)

factor(product_category)17 -0.0400***       -0.0400***
                          (0.0083)          (0.0101)

factor(product_category)20 0.0513***         0.0513***
                          (0.0045)          (0.0050)

factor(product_category)21 -0.0283***       -0.0283***
                          (0.0031)          (0.0034)

age_band                   0.0001            0.0001
                          (0.0002)          (0.0002)

female                     0.0230***         0.0230***
                          (0.0016)          (0.0016)

hasChild                   0.0014*           0.0014*
                          (0.0006)          (0.0006)

Constant                  -0.0094           -0.0094
                          (0.0083)          (0.0084)

--------------------------------------------------------
Observations              1,170,564         1,170,564
R2                        -0.0304           -0.0304
Adjusted R2               -0.0305           -0.0305
Residual Std. Error       0.3043            0.3043
========================================================
Note:            *p<0.05; **p<0.01; ***p<0.001
```

Question 5 - Heteroskedasticity

Sales Value

```
> gqtest(salesval) #No Heteroskedasticity

        Goldfeld-Quandt test

data:  salesval
GQ = 0.464, df1 = 10491, df2 = 10490, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(salesval) #Heteroskedasticity

        studentized Breusch-Pagan test

data:  salesval
BP = 945.55, df = 10, p-value < 2.2e-16
```

```
> gqtest(salesquantnb) #No heteroskedasticity

        Goldfeld-Quandt test

data:  salesquantnb
GQ = 0.0077548, df1 = 10491, df2 = 10490, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(salesquantnb) #Heteroskedasticity

        studentized Breusch-Pagan test

data:  salesquantnb
BP = 79.827, df = 10, p-value = 0.0000000000005428
```

Return Value

```
> gqtest(returnval)#No heteroskedasticity

        Goldfeld-Quandt test

data:  returnval
GQ = 0.81164, df1 = 10490, df2 = 10489, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(returnval) #Heteroskedastic

        studentized Breusch-Pagan test

data:  returnval
BP = 3181.5, df = 11, p-value < 2.2e-16
```

```
> gqtest(returnquantnb) #No heteroskedasticity

        Goldfeld-Quandt test

data:  returnquantnb
GQ = 0.019753, df1 = 10490, df2 = 10489, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(returnquantnb) #No heteroskedastic

        studentized Breusch-Pagan test

data:  returnquantnb
BP = 848.96, df = 11, p-value < 2.2e-16
```

Question 5 - Choosing Negative Binomial
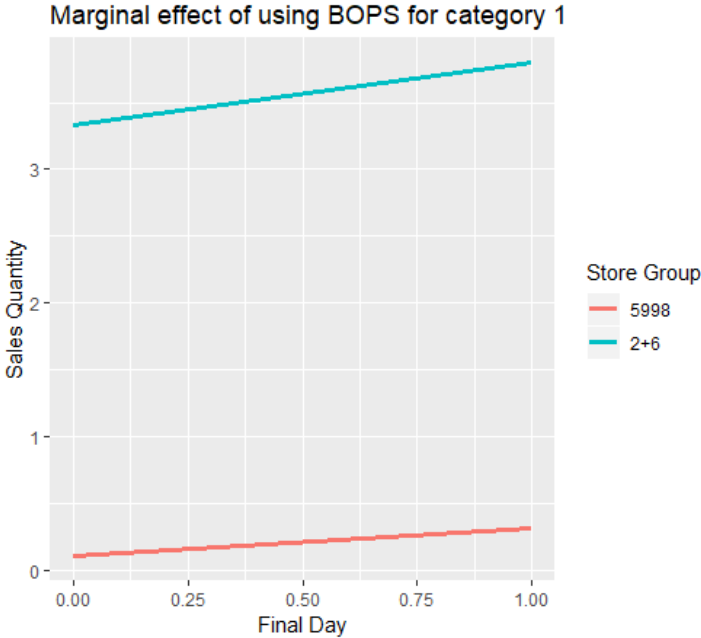
Sales Quantity

Return Quantity

Q6 - Sales Quantity

The marginal effect plots for all product categories

*We found that **4 out of the 21** categories have a significant difference*
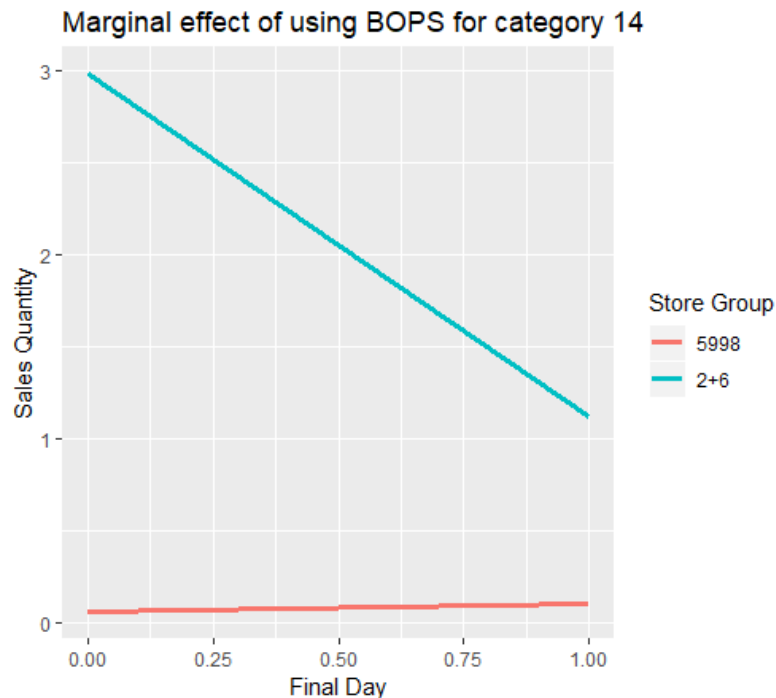
Sales Quantity- Product Category 1
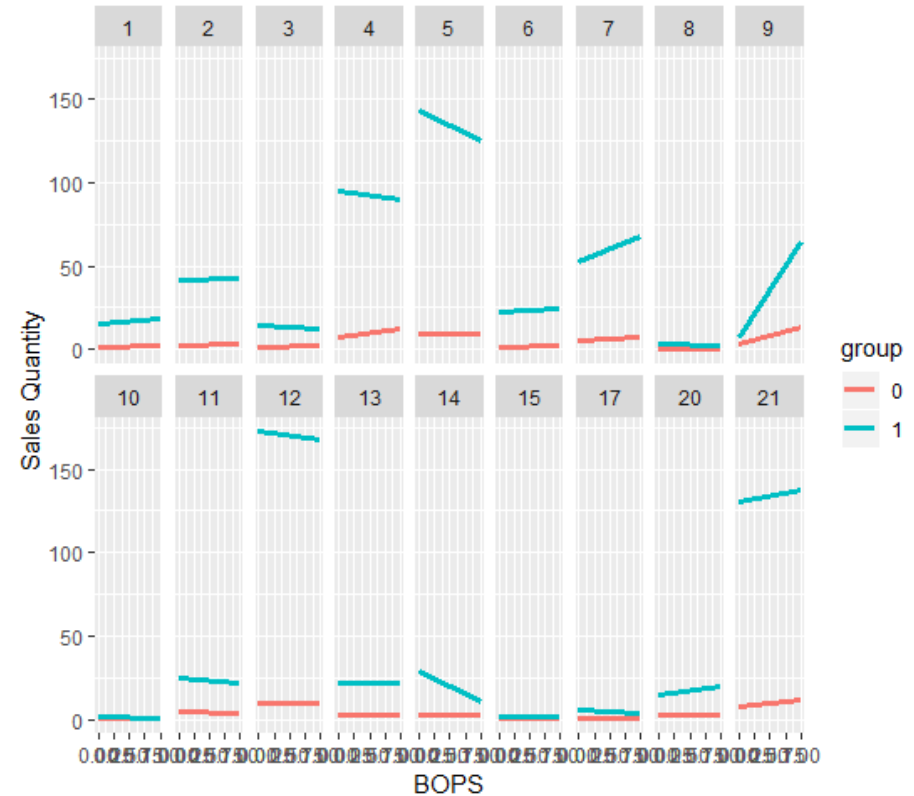
Marginal effect of using BOPS for category 1

Category 1	
	Dependent variable:
	returnquantity Sales Quantity
final_day	1.05*
	(0.42)
storegroup	3.40***
	(0.39)
avg_female	0.21
	(0.15)
avg_age	-0.11***
	(0.02)
avg_income	0.19***
	(0.03)
avg_homeowner	-0.69***
	(0.17)
avg_residency	-0.02
	(0.01)
avg_childowner	0.33*
	(0.16)
final_day:storegroup	-0.92*
	(0.43)
Constant	-2.26***
	(0.44)
Observations	1,381
Log Likelihood	-2,892.23
theta	0.64*** (0.04)
Akaike Inf. Crit.	5,804.46
Note:	*p<0.05; **p<0.01; ***p<0.001

47

Sales Quantity-Product Category 4

Marginal effect of using BOPS for category 4
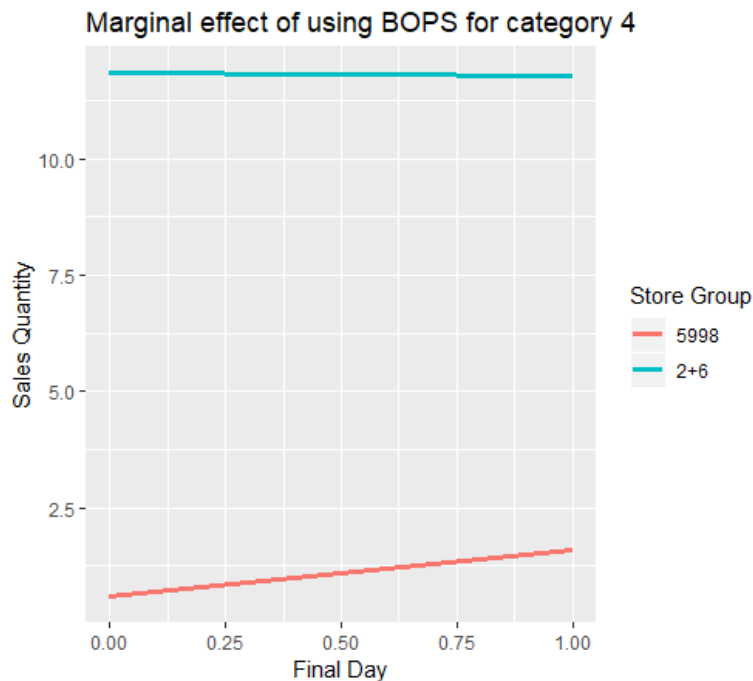
Store Group
— 5998
— 2+6

Category 4	
	Dependent variable:
	returnquantity Sales Quantity
final_day	0.98*** (0.18)
storegroup	2.98*** (0.16)
avg_female	-2.87*** (0.19)
avg_age	-0.07** (0.03)
avg_income	0.36*** (0.05)
avg_homeowner	-1.66*** (0.25)
avg_residency	-0.09*** (0.02)
avg_childowner	-0.68** (0.23)
final_day:storegroup	-0.99*** (0.19)
Constant	1.15*** (0.31)
Observations	1,838
Log Likelihood	-5,383.05
theta	0.52*** (0.02)
Akaike Inf. Crit.	10,786.10
Note:	*p<0.05; **p<0.01; ***p<0.001

Sales Quantity- Product Category 14

Marginal effect of using BOPS for category 14

Category 14	
	Dependent variable:
	returnquantity Sales Quantity
final_day	0.53 (0.68)
storegroup	3.88*** (0.54)
avg_female	−0.20 (0.18)
avg_age	−0.09*** (0.02)
avg_income	0.05 (0.04)
avg_homeowner	−0.08 (0.21)
avg_residency	−0.03 (0.02)
avg_childowner	−0.24 (0.18)
final_day:storegroup	−1.50* (0.69)
Constant	−2.09*** (0.59)
Observations	1,079
Log Likelihood	−1,763.42
theta	0.49*** (0.04)
Akaike Inf. Crit.	3,546.84
Note:	*p<0.05; **p<0.01; ***p<0.001

Sales Quantity- Product Category 21
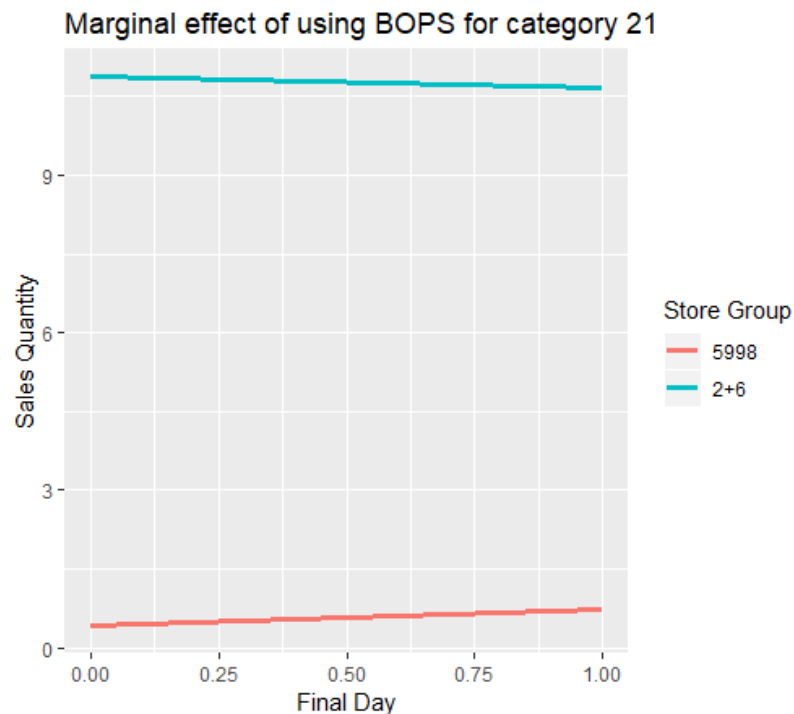
Marginal effect of using BOPS for category 21

	Dependent variable:
	returnquantity Sales Quantity
final_day	0.532*** (0.187)
storegroup	3.248*** (0.166)
avg_female	-2.855*** (0.201)
avg_age	-0.250*** (0.027)
avg_income	0.359*** (0.052)
avg_homeowner	0.287 (0.280)
avg_residency	-0.031 (0.022)
avg_childowner	-0.667*** (0.238)
final_day:storegroup	-0.552*** (0.204)
Constant	0.279 (0.346)
Observations	1,835
Log Likelihood	-5,033.139
theta	0.500*** (0.020)
Akaike Inf. Crit.	10,086.280

Category 21

Note: *p<0.1; **p<0.05; ***p<0.01