Alice, Julia, Greg, Bochen

Project 3

Return Policy Change

Return policy change

How does a Return Period Policy Change Affect Multichannel Retailer's Sales and Return?

Data Set:

- Two years transactions data from both online and physical stores.

(1st Apr. 2013 ~ 31st Mar. 2014)

Study group:

- Stores that changed Return Period:

90days $\xrightarrow{\text{1st Oct. 2013}}$ 45 days

- VS
- Stores that remained same Return Period: 60 days

Questions

Breakdown Our analysis into 6 Questions:

What is the impact of the policy change :

1. On online channel sales?
2. On physical store sales?
3. On online channel returns?
4. On physical store returns?
5. On customers purchase behavior?
6. On customers return behavior?

Check Two Raw Data Sets

Transactions data set:

includes net purchase amount, day, return info and consumers info.

customer_id	purchase_date	transaction_id	store_number	net_purchase_amount	sku	sales_assoc_1	return	return_date	return_store	time_to_return	gender	age_band	est_income_code	ethnic_code	homeowner_code
34971695	01APR2013:00:00:00	251438	2722	351.00	19382928	1031651	0		NA	NA	M	5	6	U	O
37099670	01APR2013:00:00:00	1141263	77	1.00	15510290	1041692	0		NA	NA		NA	NA		
37102884	01APR2013:00:00:00	1099496	243	49.98	18110403	33015	0		NA	NA		NA	NA		
37106545	01APR2013:00:00:00	1054556	393	18.98	16306805	1041197	0		NA	NA		NA	NA		
37127654	01APR2013:00:00:00	335359	1581	1399.99	18243402	772414	0		NA	NA		0	6	N	R
37105756	01APR2013:00:00:00	1101497	354	1.00	15510290	1040102	0		NA	NA		NA	NA		
37155922	01APR2013:00:00:00	4445622	2	219.99	19126473	0	0		NA	NA	M	0	3	H	R
36834000	01APR2013:00:00:00	531949	1422	135.00	18277277	103426	1	01APR2013:00:00:00	1422	0	F	0	4	U	O

Store attributes data set:

includes store average price, sales info

store_number	month_index	month	year	store_average_price	store_number_of_skus	sa_gender	sa_full_time	sa_avg_years_of_exp	sa_married	sa_avg_rate_of_pay	sa_dependent	Brand_number	sales_volume_group
2	45	APR	2013	266.02	11134	0.78	0.70	4.71	0.32	10.710000	0.36	1	6
2	46	MAY	2013	261.47	12449	0.81	0.72	4.86	0.33	10.950000	0.36	1	6
2	47	JUN	2013	282.43	9414	0.77	0.72	4.61	0.36	11.050000	0.34	1	6
2	48	JUL	2013	267.57	10051	0.78	0.67	4.24	0.32	10.750000	0.31	1	6
2	49	AUG	2013	276.38	10352	0.79	0.75	4.10	0.34	10.610000	0.33	1	6
2	50	SEP	2013	264.17	9777	0.81	0.72	4.51	0.38	10.350000	0.29	1	6

Create Two variables :

1) Study group **VS** Reference group
2) Before policy change **VS** Post policy change

Study group:

Dummy Variable,

0 -> company that haven't changed their policy

1 -> company that changed their policy

Policy change:

Dummy Variable,

0 -> Before Policy changed,

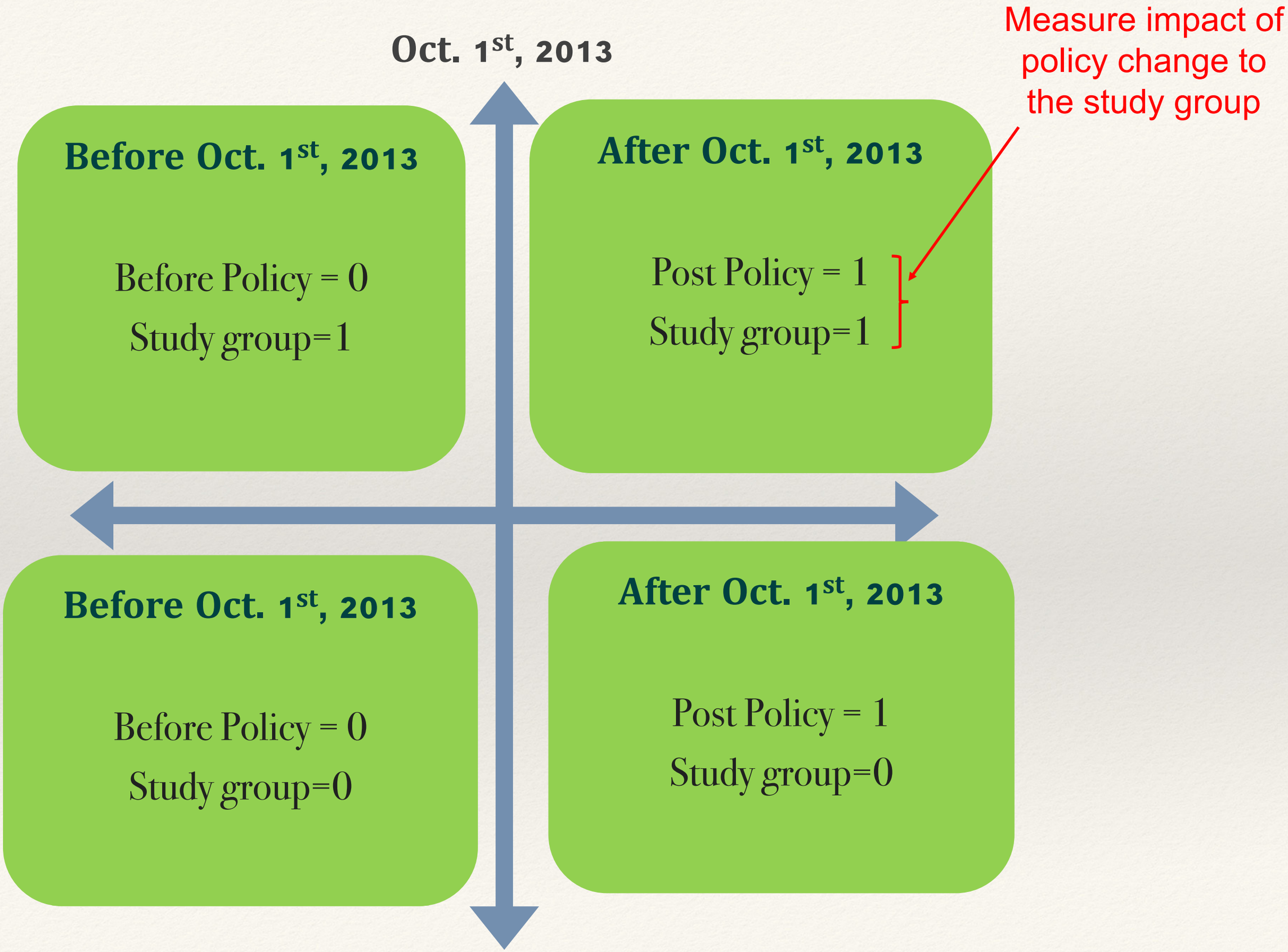
1 -> Post policy

post_policy	study_group	customer_id1	purchase_date1
1	1	100014	17DEC2013:00:00:00
1	1	100014	17DEC2013:00:00:00
1	1	100014	19DEC2013:00:00:00
0	1	100058	06JUL2013:00:00:00
0	1	100058	06JUL2013:00:00:00
0	1	100058	18JUL2013:00:00:00
1	1	100146	06MAR2014:00:00:00
1	1	100149	04MAR2014:00:00:00

After 1st Oct, 2013

This company changed their return policy

Key independent Variable



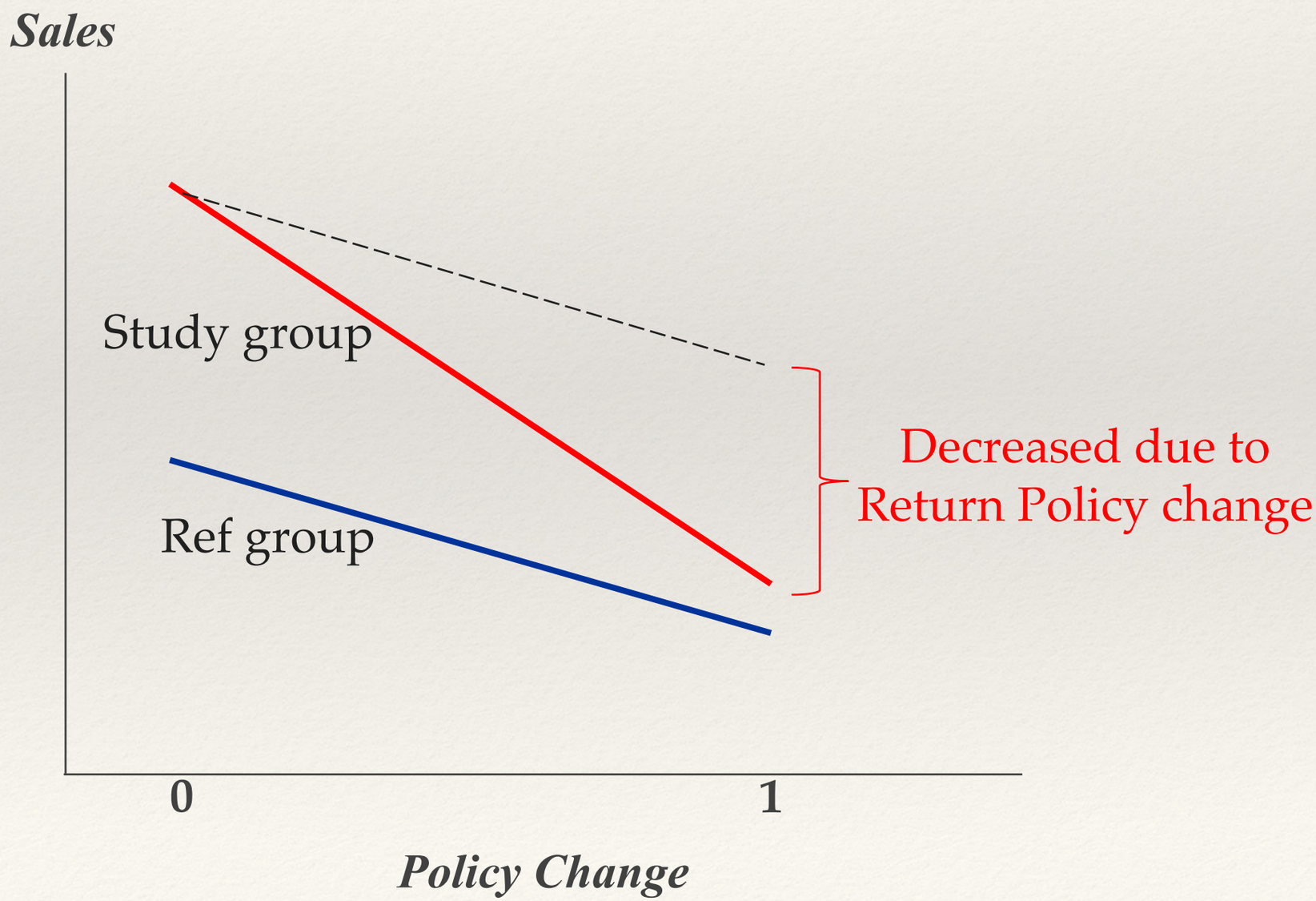
Business Question1 and Key Variables

1. Business question1: Impact of return policy change on online channel sales.
2. Key dependent variable: Online Daily Sales. (Because we only have 2years transaction data, it's better to use daily sales so that we have enough observations.)
3. Key independent variable: Policy Change:
(Dummy: before=0, post=1)

Initial hypothesis

Return policy became more strict,

Sales will decrease.



Question 1: What is the impact of the policy change **on online channel sales?**

Aggregate Data base on day, year, store number.

day	month	year	store_number	log_dailysales	ave_price	post_policy	study_group
1	4	2013	10	7.399686	30.83906	0	0
1	4	2013	2	12.155090	184.34331	0	1
1	4	2013	6	8.923589	159.70170	0	1
10	4	2013	10	6.376880	53.37182	0	0
10	4	2013	2	12.202454	158.53352	0	1
10	4	2013	6	9.314760	123.32955	0	1
100	7	2013	10	6.517745	21.80806	0	0
100	7	2013	2	12.646558	228.78582	0	1
100	7	2013	6	9.859451	225.14565	0	1
101	7	2013	2	12.288196	277.99857	0	1
101	7	2013	6	9.247975	225.70696	0	1

Question 1: What is the impact of the policy change **on online channel sales?**

Multicollinearity Check

```
> cor(df1)
                    onlineQ1.X1_time onlineQ1.X2_pc onlineQ1.store_number onlineQ1.day onlineQ1.month
onlineQ1.X1_time         1.0000000000   0.0009452633         -0.005683185  0.863346173     0.06282671
onlineQ1.X2_pc           0.0009452633   1.0000000000         -0.851567781  0.004498664    -0.00480285
onlineQ1.store_number   -0.0056831851  -0.8515677814          1.000000000 -0.011583316     0.01171046
onlineQ1.day             0.8633461733   0.0044986641         -0.011583316  1.000000000    -0.06286475
onlineQ1.month           0.0628267136  -0.0048028500          0.011710461 -0.062864749     1.00000000
onlineQ1.year            0.5503903224   0.0063387475         -0.015941469  0.726084726    -0.72951259
onlineQ1.ave_price      -0.0439175611   0.7345149901         -0.689516130  0.010685353    -0.02754466
                    onlineQ1.year onlineQ1.ave_price
onlineQ1.X1_time        0.550390322        -0.04391756
onlineQ1.X2_pc          0.006338747         0.73451499
onlineQ1.store_number  -0.015941469        -0.68951613
onlineQ1.day            0.726084726         0.01068535
onlineQ1.month         -0.729512591        -0.02754466
onlineQ1.year           1.000000000         0.02592035
onlineQ1.ave_price      0.025920351         1.00000000
> vif(df1) # delete day and then year
              Variables        VIF
1      onlineQ1.X1_time   4.328309
2        onlineQ1.X2_pc   4.300975
3 onlineQ1.store_number   3.763307
4          onlineQ1.day 143.621085
5        onlineQ1.month 148.915537
6         onlineQ1.year 311.908516
7    onlineQ1.ave_price   2.296655
```

after delete day, year, store number one by one

```
> vif(df1) # delete day,then year then store_number
                        Variables      VIF
1 onlineQ1.post_policy_numeric 1.008053
2 onlineQ1.study_group_numeric 2.177685
3                onlineQ1.month 1.004977
4           onlineQ1.ave_price 2.183188
```

Question 1: What is the impact of the policy change **on online channel sales?**

Final Model:

$$\text{log_dailysales} = \beta_0 + \beta_1\,\text{post_policy} + \beta_2\,\text{study_group} + \beta_3\,\text{post_policy} * \text{study_group} + \beta_4\,\text{ave_price} + \beta_5\,\text{month}$$

Insignificant P value
for interaction terms !!!

Model fits the data

```
                        Estimate Std. Error t value  Pr(>|t|)
(Intercept)             5.666515   0.230132 24.6228 < 2.2e-16 ***
post_policy1            0.569407   0.334798  1.7007  0.089334 .
study_group1            1.874391   0.343236  5.4609 6.121e-08 ***
month_factor2          -0.774575   0.382468 -2.0252  0.043140 *
month_factor3          -0.650438   0.349745 -1.8597  0.063245 .
month_factor4           0.919518   0.303968  3.0251  0.002556 **
month_factor5           0.852900   0.296627  2.8753  0.004131 **
month_factor6          -0.127369   0.324153 -0.3929  0.694465
month_factor7          -0.026401   0.298309 -0.0885  0.929497
month_factor8           0.344059   0.287638  1.1962  0.231950
month_factor10         -0.411913   0.311468 -1.3225  0.186340
month_factor11          0.028352   0.305026  0.0929  0.925965
month_factor12          1.602174   0.301563  5.3129 1.359e-07 ***
ave_price               0.011331   0.001992  5.6883 1.732e-08 ***
post_policy1:study_group1 -0.041782   0.228799 -0.1826  0.855143
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.909 on 905 degrees of freedom
Multiple R-squared:  0.4834,    Adjusted R-squared:  0.4754
F-statistic:  60.5 on 14 and 905 DF,  p-value: < 2.2e-16
```

Question 1: What is the impact of the policy change **on online channel sales?**

Heteroscedasticity

```
> gqtest(onlineQ1_model1) # heteroskedasticity.

        Goldfeld-Quandt test

data:  onlineQ1_model1
GQ = 1.2185, df1 = 444, df2 = 444, p-value = 0.01879
alternative hypothesis: variance increases from segment 1 to 2

> bptest(onlineQ1_model1) # heteroskedasticity.

        studentized Breusch-Pagan test

data:  onlineQ1_model1
BP = 38.242, df = 14, p-value = 0.0004772
```

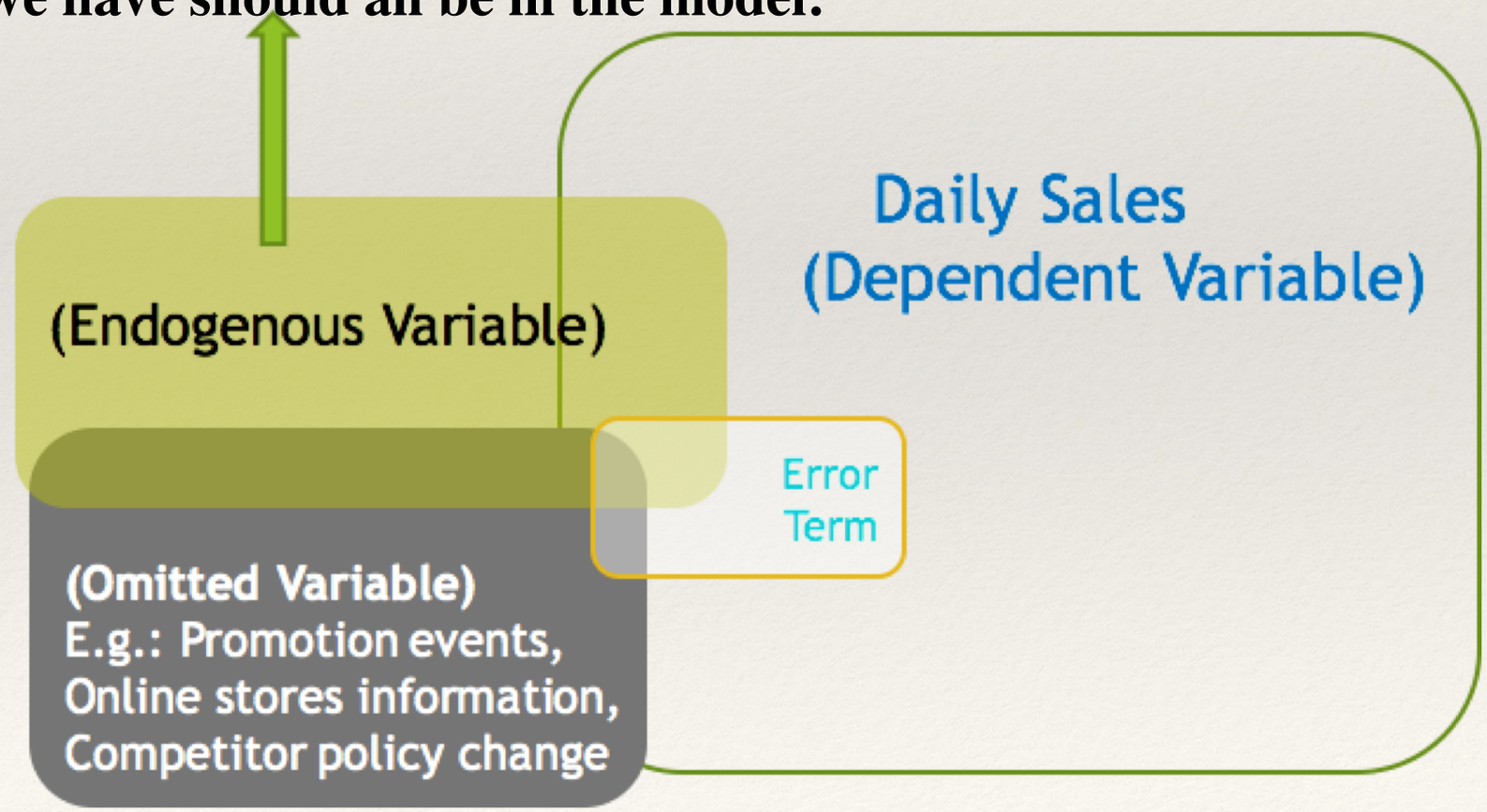
Generate Robust Standard Errors.

Question 1: What is the impact of the policy change **on online channel sales?**

Endogeneity Thinking

post_policy, study_group, store_number, day, month, year, ave_price

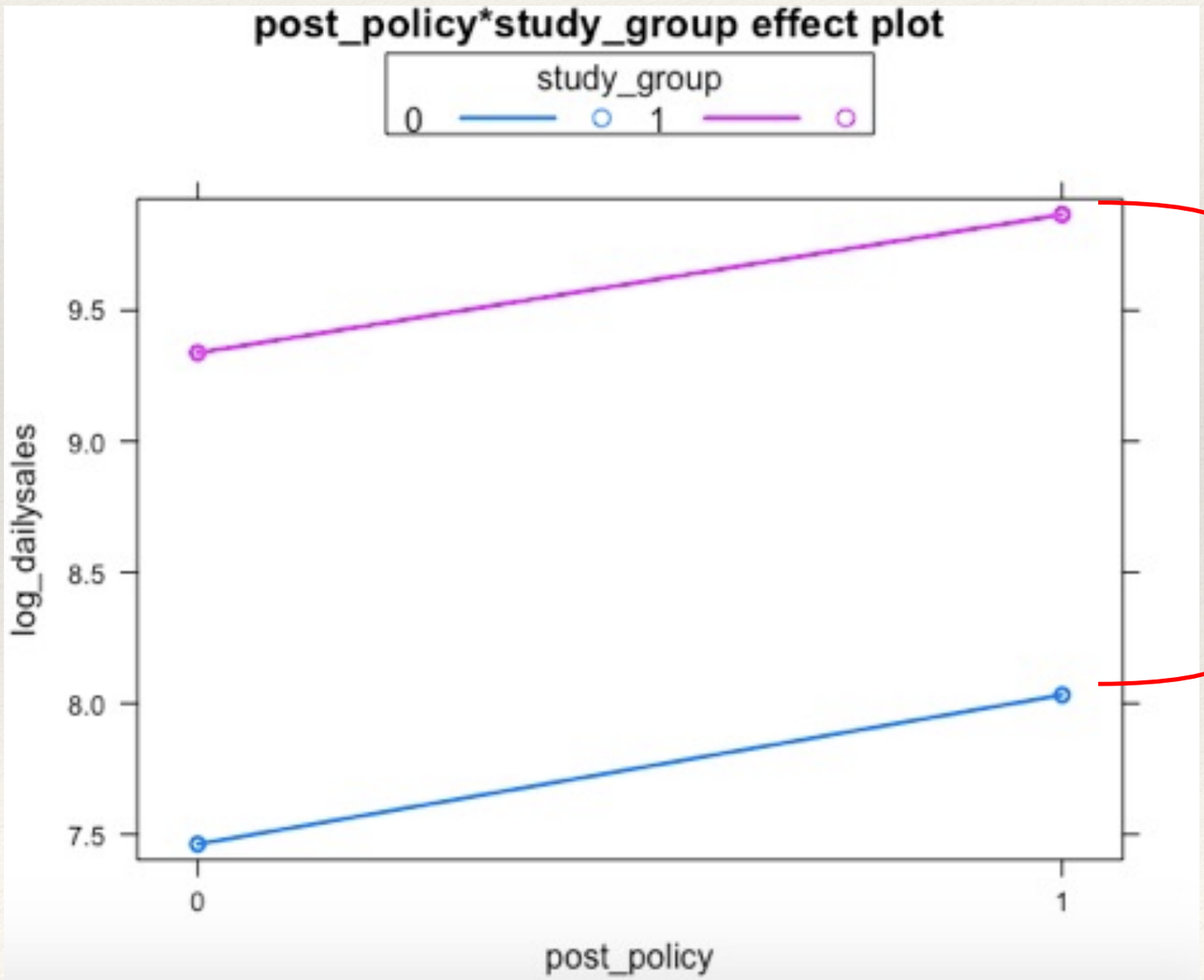
All variables we have should all be in the model.



Question 1: What is the impact of the policy change **on online channel sales?**

Final Result

There is no measurable Impact of Policy Change on online sales!



Slop has no measureable difference

Conclusion:

Return policy change does not affect online daily sales

Possible reason:

- (1) lower proportion of online customers affected by the policy change
- (2) the return displacement effect - customers accelerate their return decision in response to the policy change.
- (3)data for online stores is limited, result unreliable.

Business Question2 and Key Variables

1. Business question2: Impact of return policy change on physical channel sales.

2. Key dependent variable: Physical Sales. (we use monthly sales, because there is enough observation and we can use store attributes data to work as control variables which are monthly based.)

3. Key independent variable: Policy Change (same as Q1) (Dummy: before=0, post=1)

Question 2: What is the impact of the policy change **on physical store sales?**

Aggregate data set

based on month and store number

store_number	month_index	year	month	store_average_price	store_number_of_skus	sa_gender	sa_full_time	sa_avg_years_of_exp	sa_married	sa_avg_rate_of_pay	sa_depende
1003	45	2013	APR	383.48	116	0.71	0.57	2.00	0.43	10.42	0.
1003	46	2013	MAY	431.19	218	0.75	0.62	2.00	0.50	10.36	0.
1003	47	2013	JUN	356.14	127	0.67	0.67	2.00	0.33	10.60	0.
1003	48	2013	JUL	338.53	137	0.71	0.71	2.00	0.43	10.50	0.
1003	49	2013	AUG	326.35	146	0.75	0.62	1.87	0.37	10.57	0.
1003	50	2013	SEP	319.49	107	0.83	0.50	2.00	0.33	10.67	0.
1003	51	2013	OCT	350.69	94	0.78	0.56	1.89	0.44	10.44	0.
1003	52	2013	NOV	320.57	134	1.00	0.43	2.00	0.57	10.21	0.

sales_volume_group	month_dummy	brand_number	monthly_sales	return	post_policy	study_group
2	4	2	42895.54	22	0	1
2	5	2	96719.70	28	0	1
2	6	2	51751.14	19	0	1
2	7	2	47526.60	23	0	1
2	8	2	48177.35	15	0	1
2	9	2	28375.00	9	0	1
2	10	2	34538.25	9	1	1
2	11	2	42026.35	19	1	1
2	12	2	171784.24	87	1	1

Question 2: What is the impact of the policy change **on physical store sales?**

Multicollinearity

```
> cor(df_physical_merge)
                     post_policy_numeric study_group_numeric  month_dummy         year store_average_price    sa_gender sa_full_time
post_policy_numeric        1.0000000000        -0.001163355  0.001248065  0.576205652          0.00229400  0.012650023   0.08174840
study_group_numeric       -0.0011633551         1.000000000 -0.004548466  0.002352457          0.81218665 -0.383829491   0.34324803
month_dummy                0.0012480646        -0.004548466  1.000000000 -0.752635328         -0.01205756  0.002244872  -0.04348217
year                       0.5762056524         0.002352457 -0.752635328  1.000000000          0.01669115  0.005729375   0.08848822
store_average_price        0.0022940003         0.812186654 -0.012057557  0.016691150          1.00000000 -0.361771470   0.34296419
sa_gender                  0.0126500229        -0.383829491  0.002244872  0.005729375         -0.36177147  1.000000000  -0.21350055
sa_full_time               0.0817484030         0.343248032 -0.043482171  0.088488218          0.34296419 -0.213500547   1.00000000
sa_avg_years_of_exp        0.0293127103         0.170636271 -0.024235605  0.033060629          0.20036542  0.005406659   0.23215412
sa_married                 0.0294015203         0.415303093 -0.015640372  0.032775267          0.35034302 -0.084932707   0.22851617
sa_dependent              -0.0882533814         0.273051109 -0.033853861 -0.036893962          0.32008083 -0.066283794   0.27685430
sa_avg_rate_of_pay         0.0207234100         0.135804675 -0.003821816  0.018088109          0.28458482 -0.177153071   0.21017098
sales_volume_group        -0.0007603672         0.653598538 -0.002972871  0.001537562          0.72637944 -0.268818975   0.34928851
                     sa_avg_years_of_exp  sa_married sa_dependent sa_avg_rate_of_pay sales_volume_group
post_policy_numeric          0.029312710  0.02940152  -0.08825338        0.020723410      -0.0007603672
study_group_numeric          0.170636271  0.41530309   0.27305111        0.135804675       0.6535985376
month_dummy                 -0.024235605 -0.01564037  -0.03385386       -0.003821816      -0.0029728708
year                         0.033060629  0.03277527  -0.03689396        0.018088109       0.0015375623
store_average_price          0.200365415  0.35034302   0.32008083        0.284584816       0.7263794362
sa_gender                    0.005406659 -0.08493271  -0.06628379       -0.177153071      -0.2688189746
sa_full_time                 0.232154124  0.22851617   0.27685430        0.210170976       0.3492885082
sa_avg_years_of_exp          1.000000000  0.19412516   0.61511279       -0.076100437       0.3398760895
sa_married                   0.194125165  1.00000000   0.32067033        0.181855243       0.3180808476
sa_dependent                 0.615112791  0.32067033   1.00000000        0.076684233       0.4177849437
sa_avg_rate_of_pay          -0.076100437  0.18185524   0.07668423        1.000000000       0.2141572481
sales_volume_group           0.339876090  0.31808085   0.41778494        0.214157248       1.0000000000
```

	Variables	VIF
1	post_policy_numeric	4.308575
2	study_group_numeric	8.685361
3	brand_number	6.479653
4	month_dummy	6.627876
5	year	9.890201
6	store_average_price	3.973921
7	sa_gender	1.235937
8	sa_full_time	1.259602
9	sa_avg_years_of_exp	1.755136
10	sa_married	1.330917
11	sa_dependent	1.901914
12	sa_avg_rate_of_pay	1.264737
13	sales_volume_group	2.442652

After delete “year”, “store_average_price”,

“ brand_number” one by one.

	Variables	VIF
1	post_policy_numeric	1.050534
2	study_group_numeric	2.149176
3	month_dummy	1.211134
4	sa_gender	1.216936
5	sa_full_time	1.262959
6	sa_avg_years_of_exp	1.751522
7	sa_married	1.318470
8	sa_dependent	1.893604
9	sa_avg_rate_of_pay	1.154139
10	sales_volume_group	2.356044
11	store_number_of_skus	1.422242

Question 2: What is the impact of the policy change **on physical store sales?**

Final Model

log_monthly_sales= β_0 + β_1

X_1_time+$\beta_2 X_2$_PC +

β_3 post_policy *study_group + β_4

month + $\beta_5$sa_gender +

$\beta_6$sa_full_time +

$\beta_8$sa_avg_rate_of_exp +

$\beta_8$sa_avg_rate_of_pay +

$\beta_9$sales_volume_group + β_{10}

sa_married +

β_{11}store_number_of_skus

Results: (Robust standard error)

```
                          Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)             4.9825e+00  7.8363e-02  63.5822 < 2.2e-16 ***
post_policy1            2.5397e-01  5.4816e-02   4.6332 3.669e-06 ***
study_group1            4.6684e+00  4.1796e-02 111.6938 < 2.2e-16 ***
month_factor2          -5.0684e-01  3.0995e-02 -16.3526 < 2.2e-16 ***
month_factor3           1.8818e-01  2.7474e-02   6.8493 8.081e-12 ***
month_factor4          -2.6245e-02  3.2509e-02  -0.8073  0.419513
month_factor5          -1.0633e-02  3.4989e-02  -0.3039  0.761215
month_factor6          -4.4617e-02  3.3270e-02  -1.3411  0.179948
month_factor7          -9.8557e-02  3.2243e-02  -3.0567  0.002247 **
month_factor8           1.7887e-01  3.0782e-02   5.8109 6.499e-09 ***
month_factor10          2.0349e-01  2.8356e-02   7.1764 7.927e-13 ***
month_factor11          1.5985e-01  3.0397e-02   5.2588 1.495e-07 ***
month_factor12          3.1227e-01  5.3125e-02   5.8780 4.351e-09 ***
sa_gender               9.0728e-02  3.4434e-02   2.6349  0.008437 **
sa_full_time           -1.1820e-01  4.9239e-02  -2.4005  0.016401 *
sa_avg_years_of_exp     1.6744e-02  2.3797e-03   7.0362 2.171e-12 ***
sa_avg_rate_of_pay      5.4377e-02  4.8167e-03  11.2891 < 2.2e-16 ***
sales_volume_group      2.2267e-01  6.3977e-03  34.8051 < 2.2e-16 ***
sa_married             -1.3992e-01  3.1863e-02  -4.3914 1.144e-05 ***
store_number_of_skus    2.0552e-03  9.0454e-05  22.7210 < 2.2e-16 ***
post_policy1:study_group1 -5.6306e-01  5.0993e-02 -11.0420 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

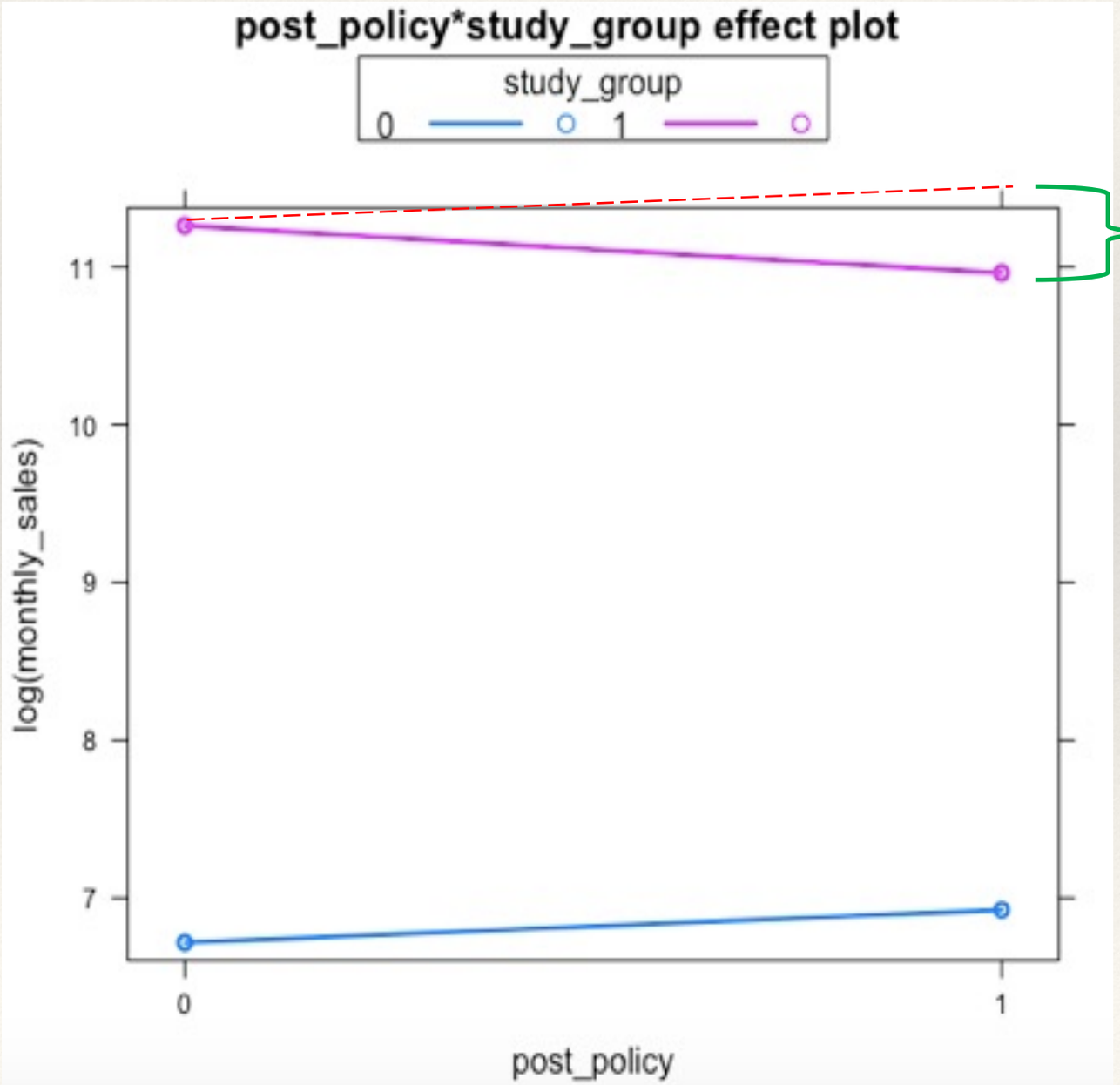
significant

Model Fit is good

```
Residual standard error: 0.5235 on 6701 degrees of freedom
Multiple R-squared:  0.9417,    Adjusted R-squared:  0.9415
F-statistic:  5407 on 20 and 6701 DF,  p-value: < 2.2e-16
```

Question 2: What is the impact of the policy change **on physical store sales?**

Finial Result: As expected, when return policy become more strict, sales of physical stores will decrease.



Sales decreased due to return policy change

Compare to sales before policy change, sales of each physical store after policy change were decreased by 56.31%.

Business Question3 and Key Variables

1. Business question3: Impact of return policy change on online channel returns.
2. Key dependent variable: Online Daily Return Amount. (Because we only have 3 online stores with 12 months of data, it's better aggregate at the daily level for more observations)
3. Key independent variable: Policy Change:
(Dummy: before=0, post=1)

Q5:Consumer Purchase behavior

Unit of analysis : consumer

Purchase behavior (dependent variable)

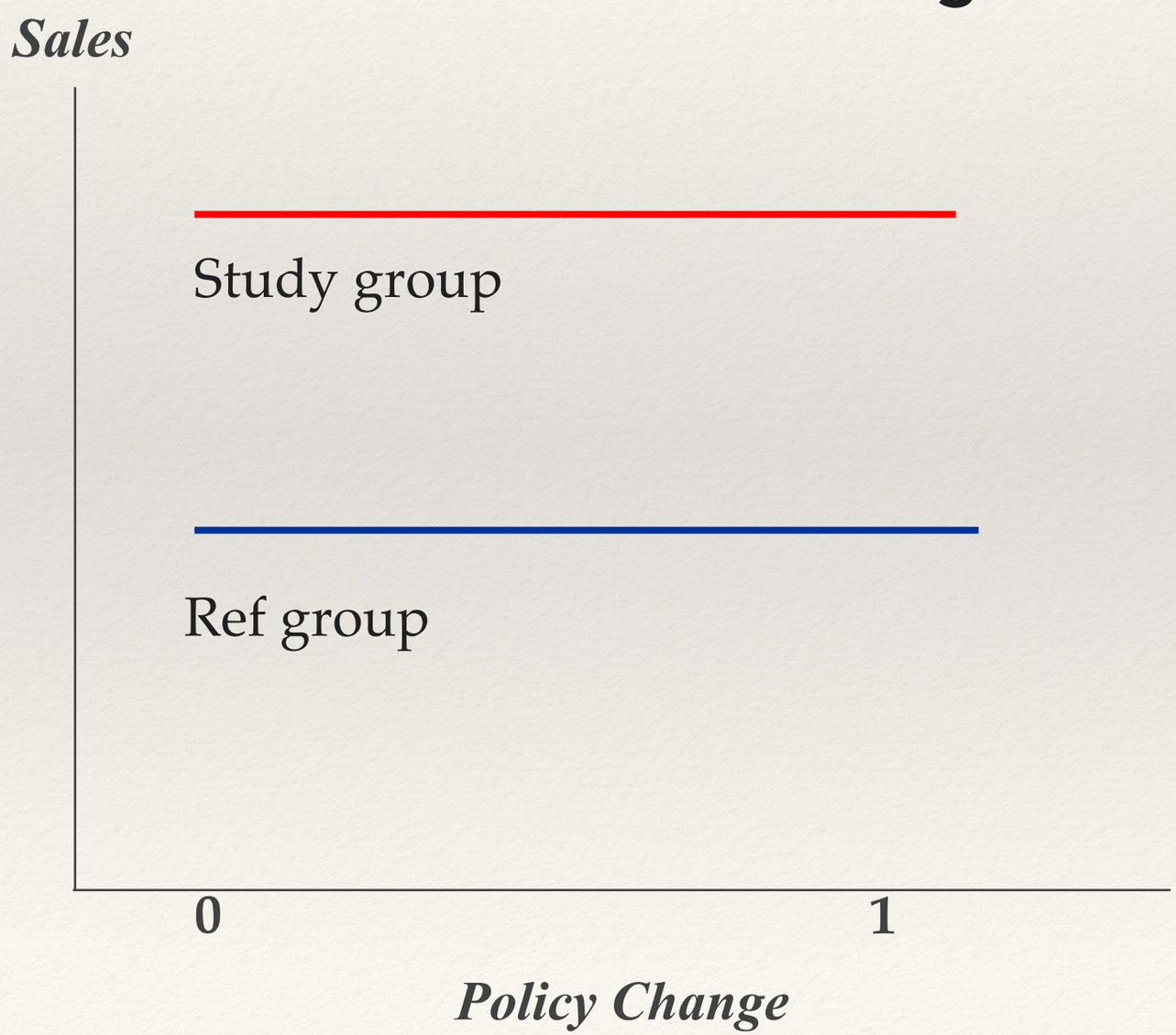
net purchase amount > OLS model

unit purchase amount > Poisson, negative binomial

Initial hypothesis

Return policy became more strict,

Returns will not change.



Question 1: What is the impact of the policy change **on online channel returns?**

Aggregate Data base on day, year, store number.

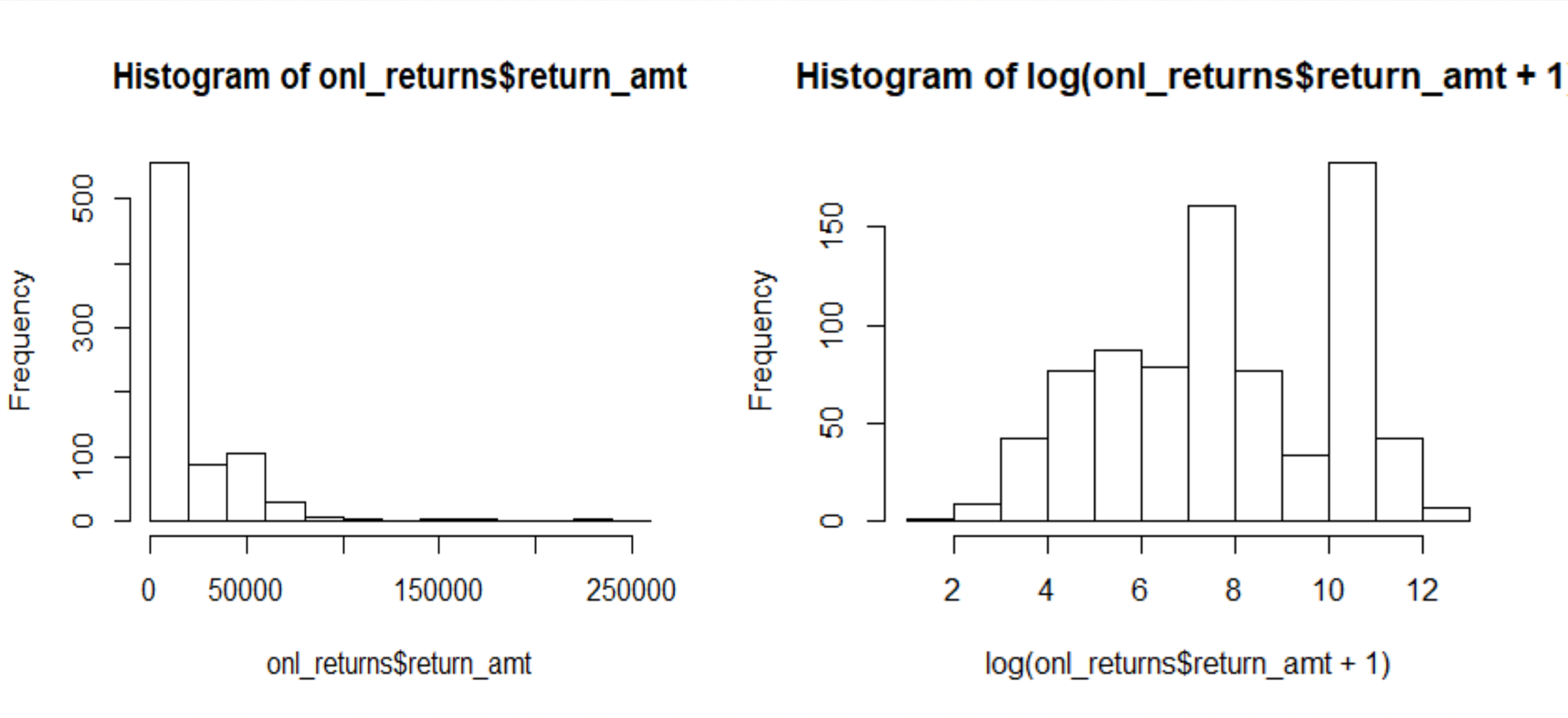
day	month	year	store_number	log_dailysales	ave_price	post_policy	study_group
1	4	2013	10	7.399686	30.83906	0	0
1	4	2013	2	12.155090	184.34331	0	1
1	4	2013	6	8.923589	159.70170	0	1
10	4	2013	10	6.376880	53.37182	0	0
10	4	2013	2	12.202454	158.53352	0	1
10	4	2013	6	9.314760	123.32955	0	1
100	7	2013	10	6.517745	21.80806	0	0
100	7	2013	2	12.646558	228.78582	0	1
100	7	2013	6	9.859451	225.14565	0	1
101	7	2013	2	12.288196	277.99857	0	1
101	7	2013	6	9.247975	225.70696	0	1

Question 1: What is the impact of the policy change **on online channel sales?**

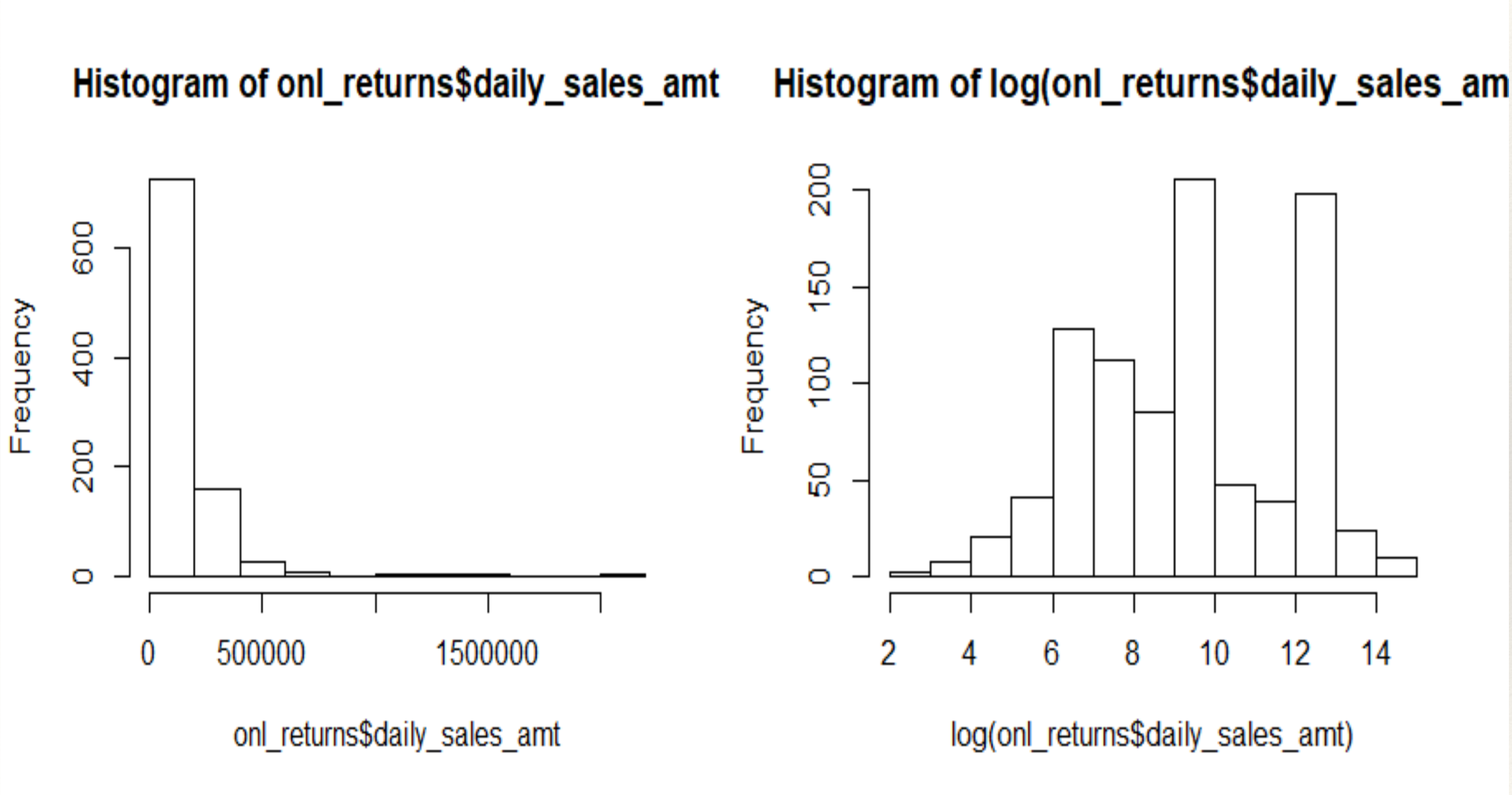
Aggregate Data base on day, year, store number.

day	month	year	store_number	log_dailysales	ave_price	post_policy	study_group
1	4	2013	10	7.399686	30.83906	0	0
1	4	2013	2	12.155090	184.34331	0	1
1	4	2013	6	8.923589	159.70170	0	1
10	4	2013	10	6.376880	53.37182	0	0
10	4	2013	2	12.202454	158.53352	0	1
10	4	2013	6	9.314760	123.32955	0	1
100	7	2013	10	6.517745	21.80806	0	0
100	7	2013	2	12.646558	228.78582	0	1
100	7	2013	6	9.859451	225.14565	0	1
101	7	2013	2	12.288196	277.99857	0	1
101	7	2013	6	9.247975	225.70696	0	1

Data summary
-----Use log transformed dependent variables

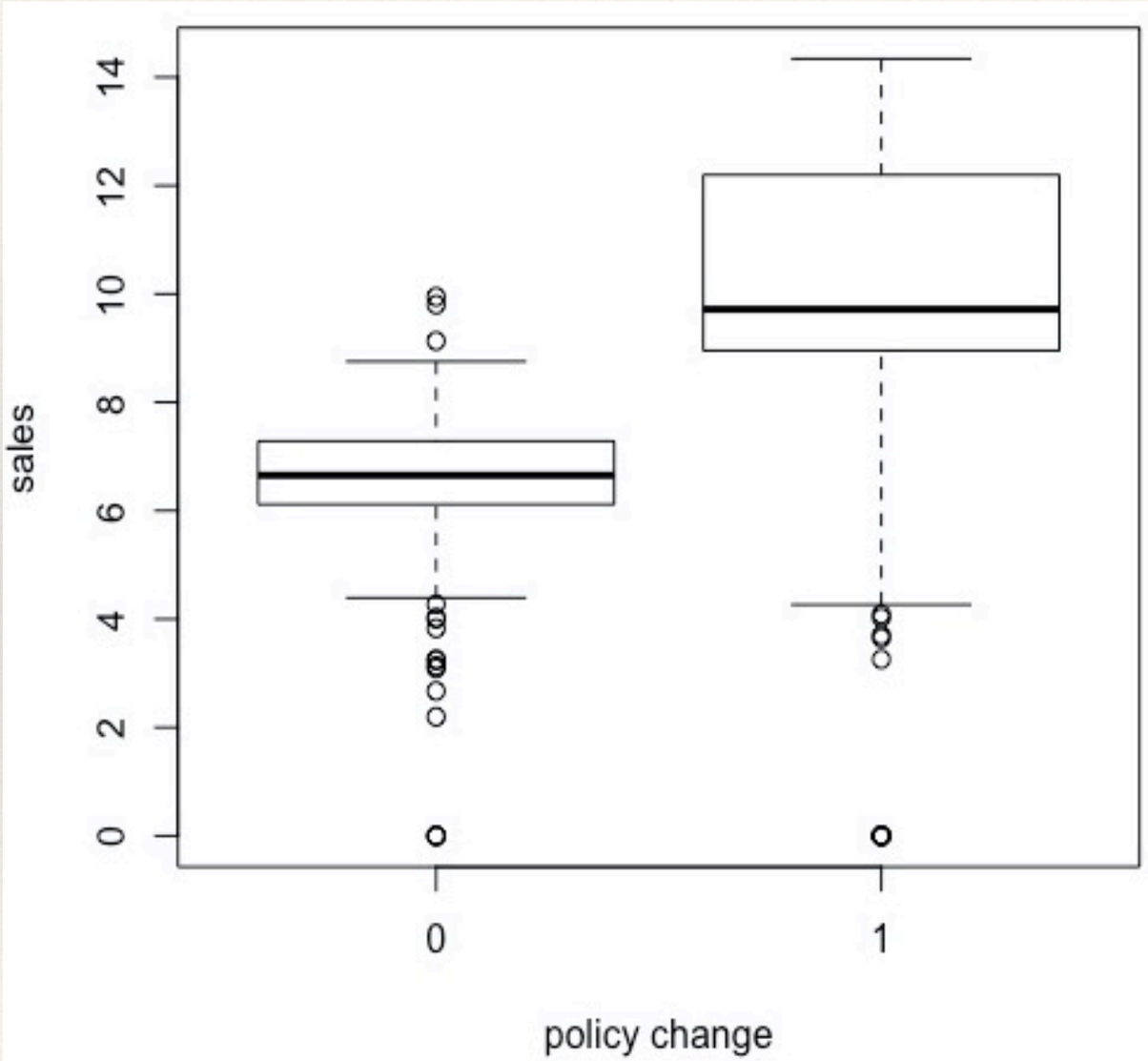


Use log transform data for daily_sales_amt

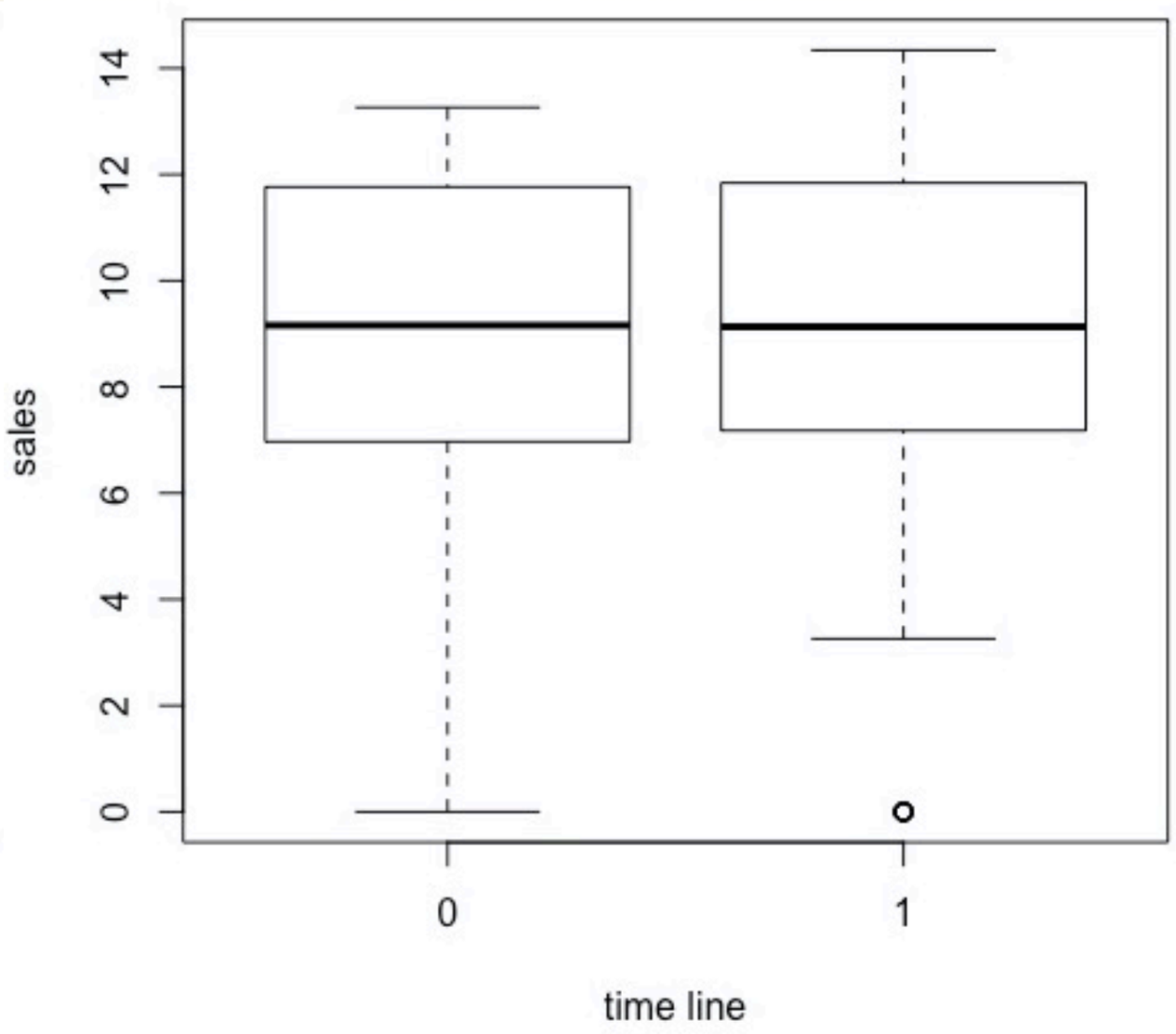


Data Visualization

Study group VS reference



Policy Change
Before Vs Post



Initial-Final OLS Model

- **Initial-Final Model:**

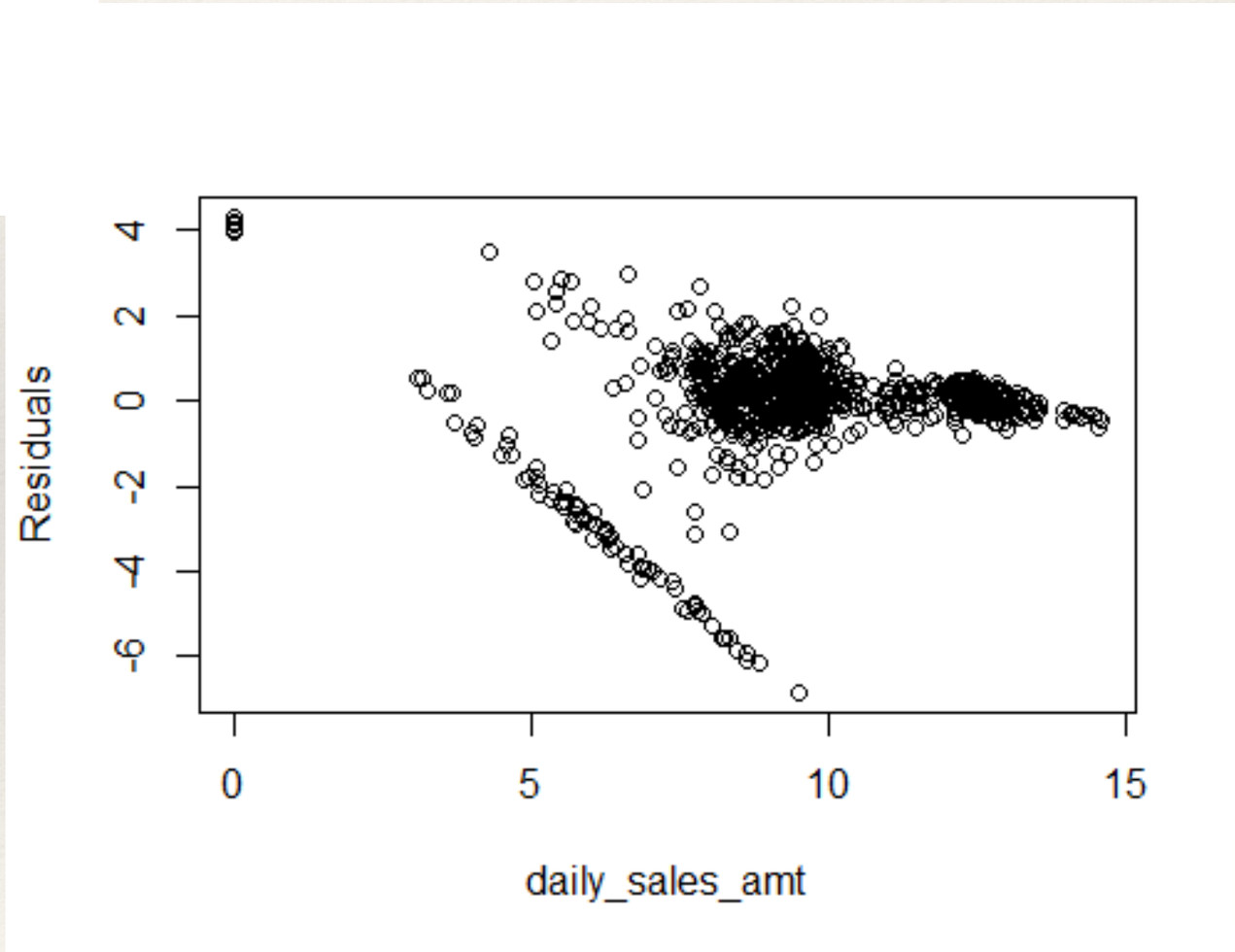
Ln(return_amt+1)= β_0 + β_1 post_policy+$\beta_2$study_group +β_3 post_policy *study_group +β_4 ln(daily_sales_amt+1) +$\beta_5$avg_age +β_6 avg_income

Final Model Results:

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              -3.651187   0.370759  -9.848   <2e-16 ***
post_policy               0.026310   0.158791   0.166    0.868
study_group               0.035993   0.137256   0.262    0.793
log(daily_sales_amt + 1)  1.176611   0.019421  60.585   <2e-16 ***
avg_income_code          -0.082809   0.056909  -1.455    0.146
avg_age_band             -0.007134   0.032676  -0.218    0.827
post_policy:study_group  -0.210016   0.187564  -1.120    0.263
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.282 on 924 degrees of freedom
Multiple R-squared:  0.822,     Adjusted R-squared:  0.8209
F-statistic: 711.3 on 6 and 924 DF,  p-value: < 2.2e-16
```

Heteroscedasticity

```
        Goldfeld-Quandt test

data:  modelOLSfin_onl
GQ = 1.2474, df1 = 459, df2 = 458, p-value = 0.009068
alternative hypothesis: variance increases from segment 1 to 2

> bptest(modelOLSfin_onl) #heteroskedasticity

        studentized Breusch-Pagan test

data:  modelOLSfin_onl
BP = 197.1, df = 6, p-value < 2.2e-16
```



Heteroscedasticity

Solution:

Generate Robust Standard Errors.

Robust SEs: Insignificant

	Estimate	Robust SE	LL	UL
(Intercept)	-0.974039692	0.01555669	-0.9919792	-0.91597650
post_policy	0.026658941	0.13358302	-0.2044442	0.32489583
study_group	0.036648369	0.15402135	-0.2252514	0.38708202
log(daily_sales_amt + 1)	2.243363215	0.11809577	2.0199620	2.48329049
avg_income_code	-0.079473084	0.08017962	-0.2239425	0.09189054
avg_age_band	-0.007108849	0.05405846	-0.1076059	0.10470568
post_policy:study_group	-0.189428971	0.13494401	-0.4151018	0.12331580

Q3 Conclusion

- Policy change for online channel does NOT affect returns

Business Question4 and Key Variables

Impact of return policy change on physical channel returns.

1. Key dependent variable:

BMS Monthly Return Amount.
(aggregated at monthly level)

2. Key independent variables:

Policy Change:
(Dummy: before=0, post=1)

Study Group:
(Dummy: reference=0, study=1)

Final OLS Model (for return $ amount)

Ln(return_amt+1)= β_0 + β_1 post_policy+$\beta_2$study_group +β_3 post_policy *study_group +β_4 ln(monthly_sales_amt+1) +$\beta_5$avg_age +β_6 avg_income +β_7 ln(avg_years_of_exp+1) +$\beta_8$avg_rate_of_pay

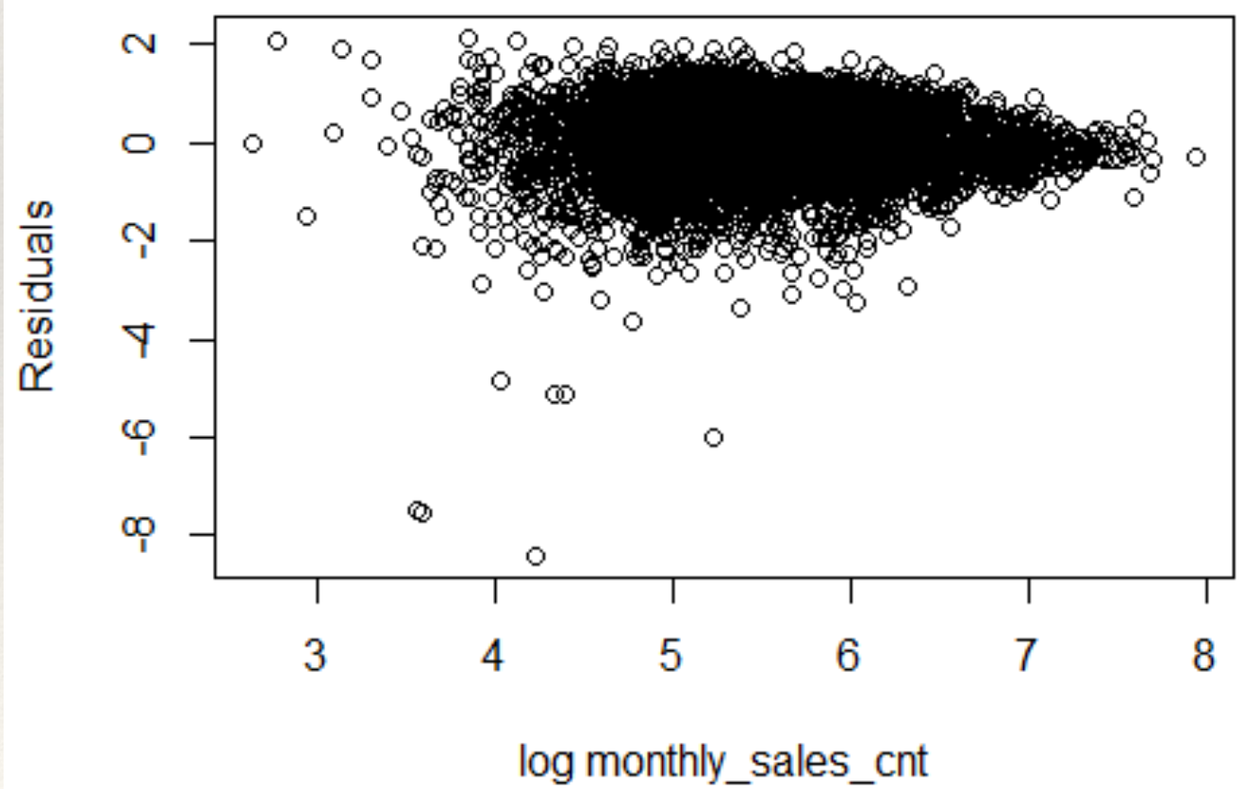
Heteroscedasticity

```
        Goldfeld-Quandt test

data:  modelOLSfin_bm
GQ = 0.84648, df1 = 3346, df2 = 3345, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(modelOLSfin_bm) #heteroskedasticity

        studentized Breusch-Pagan test

data:  modelOLSfin_bm
BP = 185.31, df = 8, p-value < 2.2e-16
```

Heteroscedasticity

Solution:

Generate Robust Standard Errors.



OLS Model Results

Robust SEs

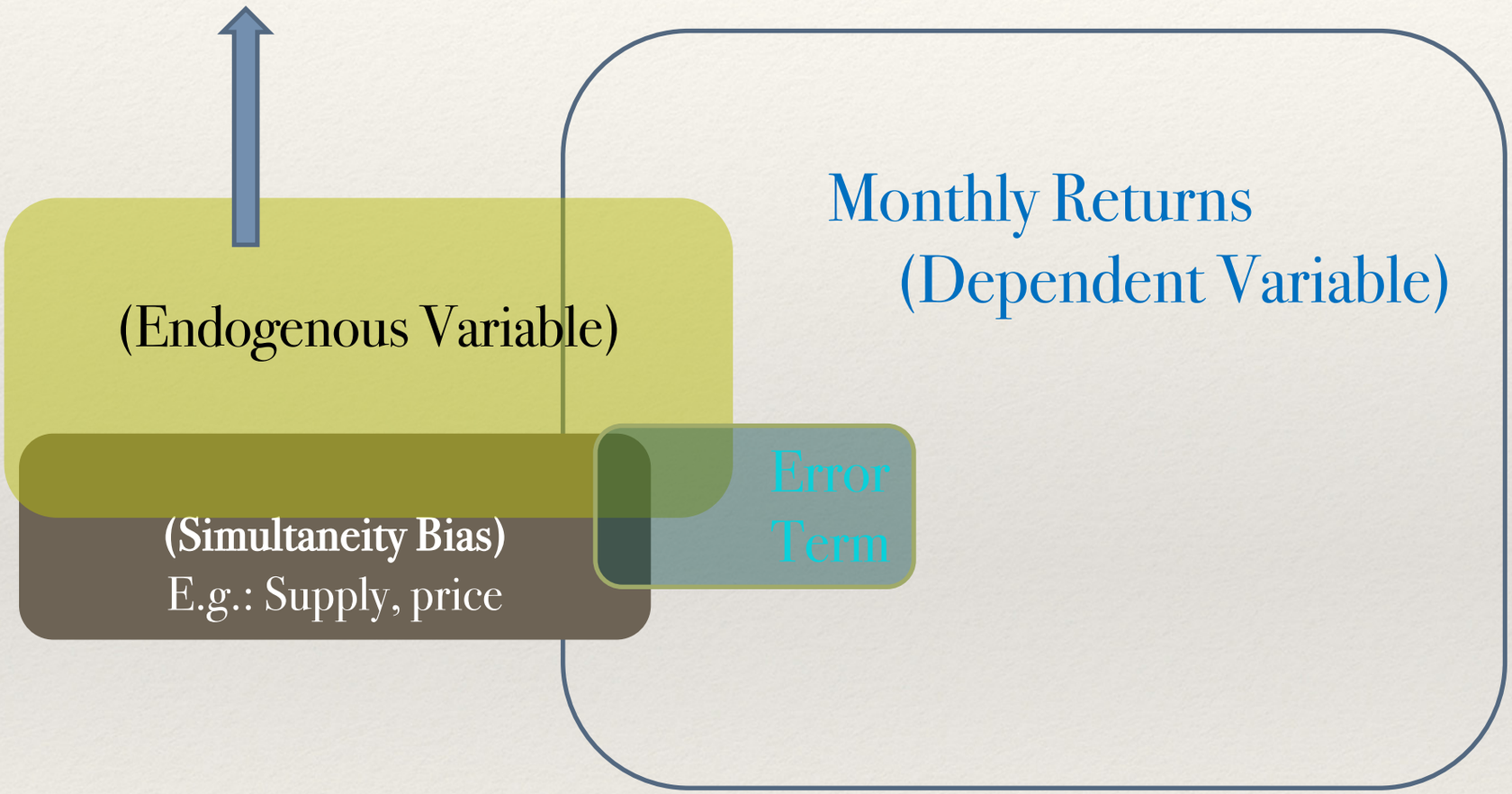
(return $ amount)

```
                                Estimate Std. Error t value  Pr(>|t|)
(Intercept)                    -0.2247852  0.1313039 -1.7119   0.08695 .
post_policy                     0.2086464  0.0445185  4.6867 2.831e-06 ***
study_group                     3.3339990  0.0371356 89.7790 < 2.2e-16 ***
log(monthly_sales_cnt + 1)      0.9879844  0.0169754 58.2009 < 2.2e-16 ***
log(sa_avg_years_of_exp + 1)    0.0931758  0.0194652  4.7868 1.731e-06 ***
sa_avg_rate_of_pay              0.1044172  0.0070701 14.7688 < 2.2e-16 ***
avg_income_code                 0.0162909  0.0113458  1.4359   0.15109
avg_age_band                   -0.0412729  0.0087301 -4.7277 2.318e-06 ***
post_policy:study_group        -0.4560070  0.0482647 -9.4480 < 2.2e-16 ***
---
```

Endogeneity Possibilities

Returns-MonthlySales simultaneity bias?

All variables we have should all be in the model instead of working as Ivs.



IV Assessment

- No possible IV candidates

Initial-Final Negative Binomial Model

- **Using number of returns as dependent var**
- **Initial-Final Model:**

Ln(return_amt+1)= β_0 + β_1 post_policy+$\beta_2$study_group +β_3 post_policy *study_group +β_4 ln(monthly_sales_amt+1) +$\beta_5$avg_age +β_6 avg_income+β_7 ln(avg_years_of_exp+1)

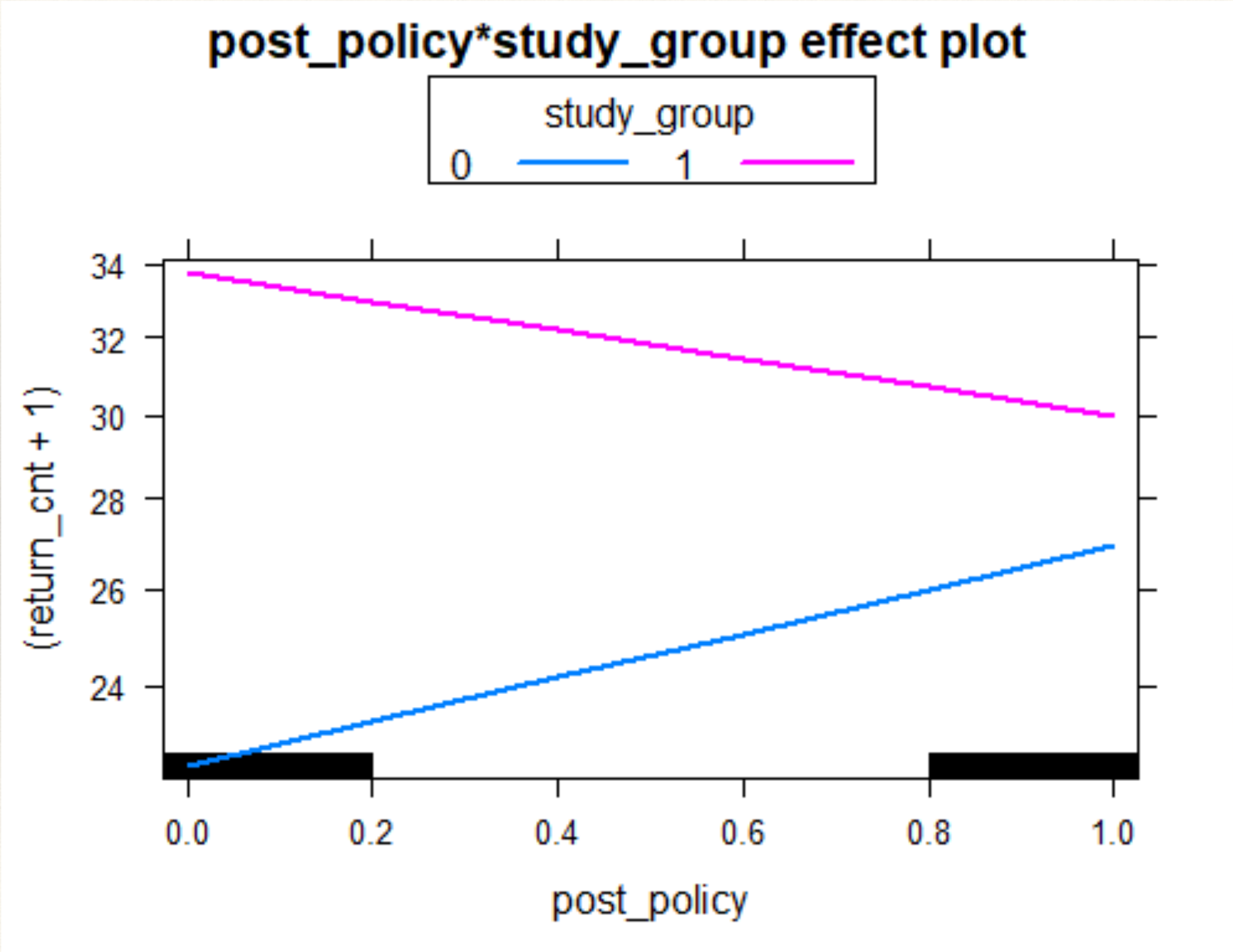
+$\beta_8$avg_rate_of_pay+β_9 θ

NegBin Results

	Estimate	Robust SE	Pr(>|z|)	LL
(Intercept)	-0.906069170	0.005837187	2.912574e-58	-0.91684075
post_policy	0.198605048	0.038023563	4.246142e-218	0.12634850
study_group	0.503156136	0.036181742	0.000000e+00	0.43388677
log(monthly_sales_cnt + 1)	1.467933886	0.016715753	0.000000e+00	1.43538752
sa_avg_rate_of_pay	0.039965430	0.003712898	0.000000e+00	0.03271355
avg_age_band	-0.010980893	0.005454749	0.000000e+00	-0.02161462
avg_income_code	-0.000328355	0.006328452	0.000000e+00	-0.01265549
log(sa_avg_years_of_exp + 1)	-0.002933052	0.009431013	0.000000e+00	-0.02124755
post_policy:study_group	-0.258599074	0.024328536	5.668017e-204	-0.30478194

UL

Control vs Study Marg Effects

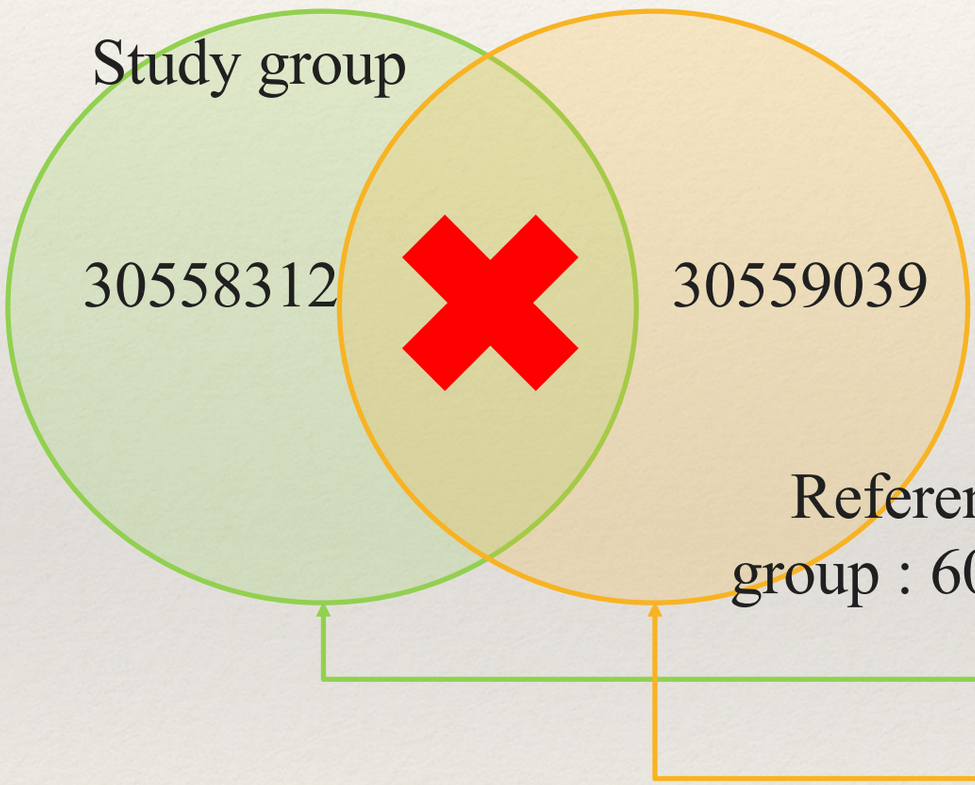


Q4 Conclusion

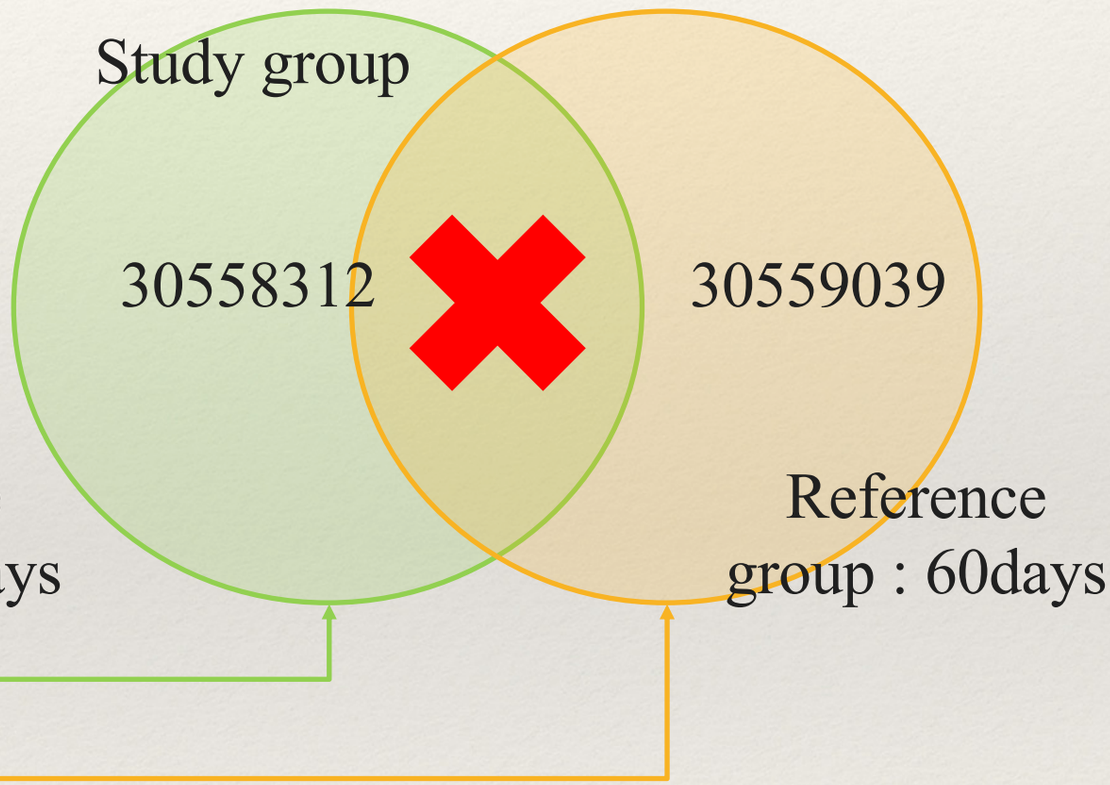
- Policy change leads to 25.86% decrease in the number of returns and a 46.6% decrease in the return dollar amount for bms channel

To set groups for policy change..

<Pre policy change : 90days>



<Post policy change : 45days>



Data Preparation

customer_id	post_policy	study_group	net_purchase_amount	units	gender	age_band	est_income_code	ethnic_code	homeowner_code	length_of_residence	child
165152	0	1	1881.50	7	M	0	8	N	O	15	N
165152	1	1	644.06	7	M	0	8	N	O	15	N
166011	0	1	259.99	2	M	10	9	N	O	9	N
166011	1	1	247.40	2	M	10	9	N	O	9	N
166800	0	1	649.99	2	M	0	6	N	O	8	N
166800	1	1	251.10	2	M	0	6	N	O	8	N
172063	0	0	50.99	5	F	3	4	N	R	15	N
172063	1	0	92.48	5	F	3	4	N	R	15	N
173216	0	1	99.99	3	M	0	3	S	R	10	Y
173216	1	1	194.98	3	M	0	3	S	R	10	Y
173487	0	1	3797.99	4	M	9	5	N	O	4	Y
173487	1	1	279.00	4	M	9	5	N	O	4	Y
178190	0	1	1349.99	4	M	0	4	N	O	7	N
178190	1	1	547.68	4	M	0	4	N	O	7	N

Customer behavior on net purchase amount

OLS

log_net_purchase_amount= β0 + β1 post_policy +β2study_group+β3 post_policy *study_group + β4 gender + β5 age_band + β6 est_income_code + β7 ethnic_code + β8 homeowner_code + β9length_of_residence + β10child + β11log_units

OLS result

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3.13346445	0.05130922	61.0702	< 2.2e-16	***
post_policy	-0.05202848	0.03272731	-1.5898	0.1118924	
study_group	1.88232325	0.02269959	82.9232	< 2.2e-16	***
genderM	0.32267121	0.00868365	37.1585	< 2.2e-16	***
genderU	0.12964933	0.01237528	10.4765	< 2.2e-16	***
age_band	-0.00064057	0.00095528	-0.6706	0.5025023	
est_income_code	0.01862362	0.00175965	10.5837	< 2.2e-16	***
ethnic_codeD	0.02570657	0.07811304	0.3291	0.7420848	
ethnic_codeF	-0.11174097	0.04848487	-2.3047	0.0211876	*
ethnic_codeG	-0.12979627	0.04550984	-2.8520	0.0043446	**
ethnic_codeH	-0.05046807	0.04361801	-1.1570	0.2472558	
ethnic_codeI	-0.13276411	0.04846656	-2.7393	0.0061581	**
ethnic_codeJ	-0.07623895	0.05717367	-1.3335	0.1823827	
ethnic_codeM	-0.11271840	0.06440705	-1.7501	0.0801048	.
ethnic_codeN	-0.10642321	0.04312082	-2.4680	0.0135876	*
ethnic_codeO	0.11801538	0.05286822	2.2323	0.0256000	*
ethnic_codeP	0.01330574	0.07193560	0.1850	0.8532549	
ethnic_codeR	0.12849968	0.06009976	2.1381	0.0325103	*
ethnic_codeS	-0.11557424	0.04406428	-2.6229	0.0087208	**
ethnic_codeU	-0.10552647	0.04671034	-2.2592	0.0238749	*
ethnic_codeX	-0.19271648	0.07298003	-2.6407	0.0082753	**
ethnic_codeZ	-0.15779349	0.04418383	-3.5713	0.0003554	***
homeowner_codeR	-0.01649571	0.00948707	-1.7388	0.0820803	.
length_of_residence	-0.00443621	0.00073834	-6.0084	1.880e-09	***
childY	-0.08261089	0.00807501	-10.2304	< 2.2e-16	***
log_units	0.54078835	0.00665261	81.2897	< 2.2e-16	***
post_policy:study_group	-0.18071224	0.03364595	-5.3710	7.846e-08	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Interaction terms
p value significant

18% decrease

Robust standard error is generated

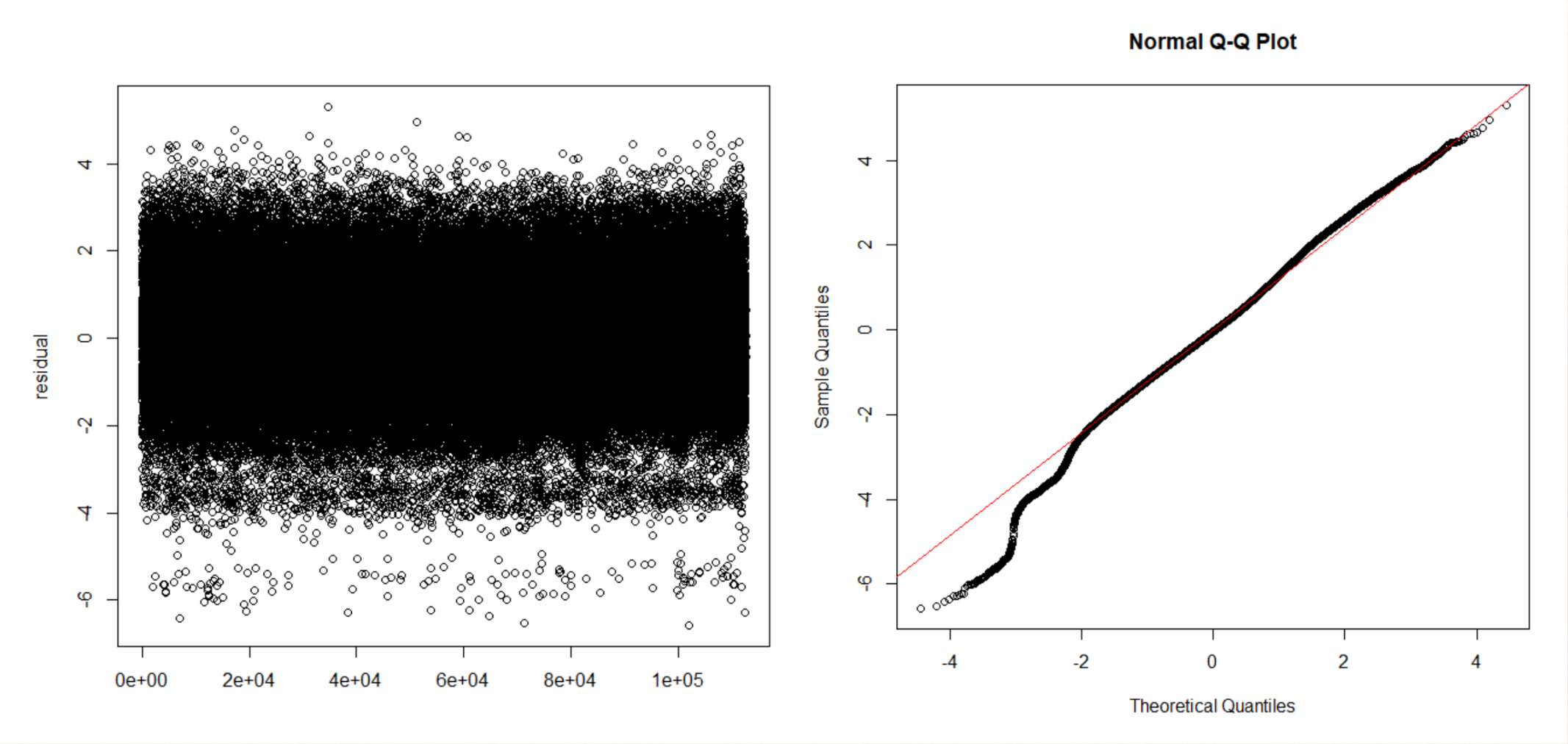
Multicollinearity - OLS

```
> cor(df)
                          finalQ5.post_policy finalQ5.study_group finalQ5.gender_num finalQ5.age_band
finalQ5.post_policy                         1         0.000000000        0.000000000       0.00000000
finalQ5.study_group                         0         1.000000000        0.074637206      -0.01709192
finalQ5.gender_num                          0         0.074637206        1.000000000      -0.22813538
finalQ5.age_band                            0        -0.017091925       -0.228135381       1.00000000
finalQ5.est_income_code                     0         0.002330681       -0.049239870       0.08447609
finalQ5.ethnic_code_num                     0        -0.010204930       -0.005830595       0.02450824
finalQ5.homeowner_code_num                  0        -0.003228570       -0.094163772       0.11786657
finalQ5.length_of_residence                 0        -0.020089444       -0.123115843       0.10973007
finalQ5.child_num                           0        -0.016603684       -0.034490552       0.01586947
finalQ5.log_units                           0        -0.097644439       -0.098229633      -0.00615571
                          finalQ5.est_income_code finalQ5.ethnic_code_num finalQ5.homeowner_code_num
finalQ5.post_policy                    0.000000000            0.0000000000                0.0000000000
finalQ5.study_group                    0.002330681           -0.0102049298               -0.0032285696
finalQ5.gender_num                    -0.049239870           -0.0058305947               -0.0941637725
finalQ5.age_band                       0.084476094            0.0245082368                0.1178665654
finalQ5.est_income_code                1.000000000           -0.0221302706                0.3331002507
finalQ5.ethnic_code_num               -0.022130271            1.0000000000                0.0000852284
finalQ5.homeowner_code_num             0.333100251            0.0000852284                1.0000000000
finalQ5.length_of_residence            0.110180868            0.0157423610                0.2938595909
finalQ5.child_num                      0.117495206           -0.0129810638                0.2556209530
finalQ5.log_units                      0.013845916            0.0079563967                0.0237078168
                          finalQ5.length_of_residence finalQ5.child_num finalQ5.log_units
finalQ5.post_policy                        0.00000000       0.000000000       0.000000000
finalQ5.study_group                       -0.02008944      -0.016603684      -0.097644439
finalQ5.gender_num                        -0.12311584      -0.034490552      -0.098229633
finalQ5.age_band                           0.10973007       0.015869473      -0.006155710
finalQ5.est_income_code                    0.11018087       0.117495206       0.013845916
finalQ5.ethnic_code_num                    0.01574236      -0.012981064       0.007956397
finalQ5.homeowner_code_num                 0.29385959       0.255620953       0.023707817
finalQ5.length_of_residence                1.00000000       0.018424543       0.023590749
finalQ5.child_num                          0.01842454       1.000000000       0.006695499
finalQ5.log_units                          0.02359075       0.006695499       1.000000000
```

```
> vif(df) #no multicolinearity
                    Variables      VIF
1          finalQ5.post_policy 1.001229
2          finalQ5.study_group 1.019718
3           finalQ5.gender_num 1.085142
4             finalQ5.age_band 1.071521
5      finalQ5.est_income_code 1.131291
6      finalQ5.ethnic_code_num 1.001995
7   finalQ5.homeowner_code_num 1.315552
8  finalQ5.length_of_residence 1.125528
9            finalQ5.child_num 1.076620
10           finalQ5.log_units 1.022461
```

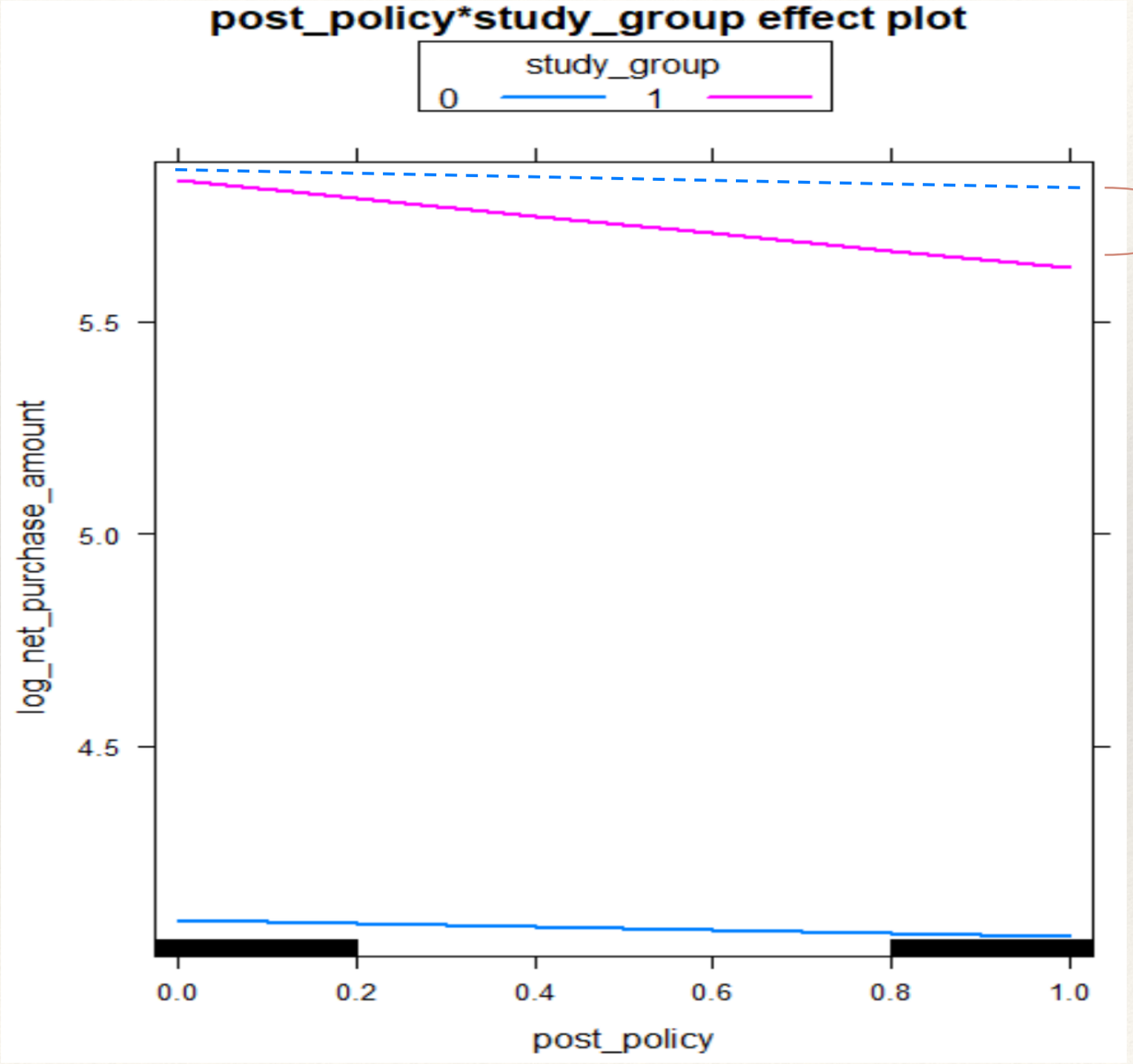
No multicolinearity

OLS – plotting residuals



It doesn't perfectly look normal

Graph



18% decrease : read interaction term coefficients

Customer behavior on item purchase amount

Poisson

units ~ β0 + β1 post_policy +β2study_group + β3 post_policy *study_group + β4 net_purchase_amount + gender + β5 age_band + β6est_income_code + β7 ethnic_code + β8homeowner_code + β9 length_of_residence + β10child,
familiy = "poisson"

```
> with(modelQ5_possion, null.deviance - deviance)
[1] 34136.58
> with(modelQ5_possion, df.null - df.residual)
[1] 26
> with(modelQ5_possion, pchisq(null.deviance - deviance, df.null - df.residual, lower.tail = FALSE))
[1] 0
```

Model fit insignificant!

Poisson

Interaction terms p value is insignificant

```
Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              2.207e+00  1.849e-02 119.331  < 2e-16 ***
post_policy             -1.584e-03  1.277e-02  -0.124 0.901238
study_group             -6.177e-01  9.331e-03 -66.204  < 2e-16 ***
net_purchase_amount      5.837e-05  2.890e-07 201.964  < 2e-16 ***
genderM                 -2.516e-01  3.102e-03 -81.118  < 2e-16 ***
genderU                 -8.998e-02  4.318e-03 -20.837  < 2e-16 ***
age_band                -4.547e-03  3.491e-04 -13.025  < 2e-16 ***
est_income_code          5.521e-03  6.433e-04   8.581  < 2e-16 ***
ethnic_codeD             6.004e-03  2.648e-02   0.227 0.820614
ethnic_codeF            -5.003e-02  1.739e-02  -2.877 0.004011 **
ethnic_codeG            -3.389e-02  1.630e-02  -2.079 0.037660 *
ethnic_codeH            -2.589e-02  1.560e-02  -1.660 0.096983 .
ethnic_codeI            -6.026e-02  1.753e-02  -3.437 0.000589 ***
ethnic_codeJ             3.042e-02  2.002e-02   1.520 0.128555
ethnic_codeM            -9.497e-03  2.329e-02  -0.408 0.683478
ethnic_codeN            -3.658e-02  1.541e-02  -2.374 0.017575 *
ethnic_codeO            -6.378e-02  1.846e-02  -3.455 0.000550 ***
ethnic_codeP            -1.647e-02  2.571e-02  -0.641 0.521756
ethnic_codeR            -7.578e-02  2.185e-02  -3.467 0.000526 ***
ethnic_codeS            -7.533e-03  1.574e-02  -0.479 0.632200
ethnic_codeU             1.554e-01  1.650e-02   9.419  < 2e-16 ***
ethnic_codeX            -8.908e-02  2.761e-02  -3.227 0.001253 **
ethnic_codeZ            -2.844e-02  1.579e-02  -1.801 0.071641 .
homeowner_codeR         -2.556e-02  3.461e-03  -7.386 1.52e-13 ***
length_of_residence      8.422e-05  2.679e-04   0.314 0.753229
childX                  -9.626e-04  2.936e-03  -0.328 0.743010
post_policy:study_group  1.655e-02  1.308e-02   1.265 0.205719
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 330291  on 112613  degrees of freedom
Residual deviance: 296155  on 112587  degrees of freedom
AIC: 653668

Number of Fisher Scoring iterations: 7
```

Negative Binomial

units ~ β0 + β1 post_policy +β2study_group + β3 post_policy *study_group + β4 net_purchase_amount + gender + β5 age_band + β6est_income_code + β7 ethnic_code + β8homeowner_code + β9 length_of_residence + β10child

```
> with(modelQ5_NB, null.deviance - deviance)
[1] 14025.11
> with(modelQ5_NB, df.null - df.residual)
[1] 26
> with(modelQ5_NB, pchisq(null.deviance - deviance, df.null - df.residual, lower.tail = FALSE))
[1] 0
```

Model fit significant!!

```
> X2 <- 2 * (logLik(modelQ5_NB) - logLik(modelQ5_possion))
> X2
'log Lik.' 113118.2 (df=28)
> pchisq(X2, df = 1, lower.tail=FALSE) #likelihood test for poisson and NB - NB better
'log Lik.' 0 (df=28)
```

Comparison test significant!!

- Negative Binomial is better!!

Negative Binomial

```
Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              2.148e+00  2.973e-02  72.266  < 2e-16 ***
post_policy             -4.605e-02  2.364e-02  -1.948 0.051440 .
study_group             -6.111e-01  1.705e-02 -35.844  < 2e-16 ***
net_purchase_amount      9.980e-05  1.158e-06  86.175  < 2e-16 ***
genderM                 -2.574e-01  4.672e-03 -55.107  < 2e-16 ***
genderU                 -1.182e-01  6.652e-03 -17.762  < 2e-16 ***
age_band                -3.291e-03  5.248e-04  -6.271 3.59e-10 ***
est_income_code          4.202e-03  9.684e-04   4.339 1.43e-05 ***
ethnic_codeD             3.197e-02  4.058e-02   0.788 0.430727
ethnic_codeF            -3.700e-02  2.642e-02  -1.400 0.161374
ethnic_codeG            -1.003e-02  2.485e-02  -0.403 0.686587
ethnic_codeH            -1.066e-02  2.382e-02  -0.447 0.654635
ethnic_codeI            -2.094e-02  2.670e-02  -0.784 0.433005
ethnic_codeJ             4.078e-02  3.065e-02   1.330 0.183449
ethnic_codeM             7.412e-03  3.536e-02   0.210 0.833975
ethnic_codeN            -2.075e-02  2.353e-02  -0.882 0.377779
ethnic_codeO            -1.971e-02  2.818e-02  -0.699 0.484287
ethnic_codeP             9.359e-03  3.910e-02   0.239 0.810822
ethnic_codeR            -3.946e-02  3.302e-02  -1.195 0.231991
ethnic_codeS             7.242e-03  2.403e-02   0.301 0.763108
ethnic_codeU             8.800e-02  2.548e-02   3.454 0.000553 ***
ethnic_codeX            -5.659e-02  4.118e-02  -1.374 0.169400
ethnic_codeZ            -2.374e-03  2.410e-02  -0.099 0.921524
homeowner_codeR         -2.748e-02  5.207e-03  -5.277 1.31e-07 ***
length_of_residence      1.062e-03  4.035e-04   2.632 0.008485 **
childY                   9.464e-04  4.437e-03   0.213 0.831094
post_policy:study_group  6.857e-02  2.403e-02   2.854 0.004319 **

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(3.6634) family taken to be 1)

    Null deviance: 112624  on 112613  degrees of freedom
Residual deviance:  98599  on 112587  degrees of freedom
AIC: 540551

Number of Fisher Scoring iterations: 1


              Theta:  3.6634
          Std. Err.:  0.0243

 2 x log-likelihood:  -540495.4360
```

Interaction terms p value is significant

Negative Binomial result - exponentiate

	Estimate	2.5 %	97.5 %
(Intercept)	8.5707862	8.0877898	9.0835172
post_policy	0.9549976	0.9120077	1.0000132
study_group	0.5427348	0.5249494	0.5610477
net_purchase_amount	1.0000998	1.0000971	1.0001025
genderM	0.7730239	0.7659939	0.7801175
genderU	0.8885545	0.8770868	0.9001739
age_band	0.9967145	0.9956969	0.9977332
est_income_code	1.0042104	1.0023056	1.0061188
ethnic_codeD	1.0324904	0.9534694	1.1181799
ethnic_codeF	0.9636805	0.9150902	1.0147912
ethnic_codeG	0.9900225	0.9429889	1.0393298
ethnic_codeH	0.9894011	0.9443096	1.0365661
ethnic_codeI	0.9792819	0.9293827	1.0318024
ethnic_codeJ	1.0416203	0.9809348	1.1060400
ethnic_codeM	1.0074394	0.9400233	1.0797119
ethnic_codeN	0.9794613	0.9353554	1.0255660
ethnic_codeO	0.9804812	0.9278266	1.0360798
ethnic_codeP	1.0094029	0.9349377	1.0898768
ethnic_codeR	0.9613057	0.9010704	1.0255613
ethnic_codeS	1.0072688	0.9609728	1.0557156
ethnic_codeU	1.0919924	1.0388914	1.1477339
ethnic_codeX	0.9449781	0.8717256	1.0244361
ethnic_codeZ	0.9976288	0.9516417	1.0457598
homeowner_codeR	0.9728927	0.9630391	0.9828478
length_of_residence	1.0010627	1.0002699	1.0018561
childY	0.9990540	0.9904126	1.0077711
post_policy:study_group	1.0709736	1.0219770	1.1223203

7.09% increase

Heteroskedasticity – OLS

```
> gqtest(modelQ5_OLS)

        Goldfeld-Quandt test

data:  modelQ5_OLS
GQ = 1.042, df1 = 56280, df2 = 56280, p-value = 5.303e-07
alternative hypothesis: variance increases from segment 1 to 2
```

significant!!

```
> bptest(modelQ5_OLS)

        studentized Breusch-Pagan test

data:  modelQ5_OLS
BP = 6846.6, df = 26, p-value < 2.2e-16
```

significant!!

- We have heteroskedasticity
- Fix with Robust standard error

Heteroskedasticity – Negative binomial

```
> gqtest(modelQ5_NB)

        Goldfeld-Quandt test

data:  modelQ5_NB
GQ = 0.4714, df1 = 56280, df2 = 56280, p-value = 1
alternative hypothesis: variance increases from segment 1 to 2

> bptest(modelQ5_NB)

        studentized Breusch-Pagan test

data:  modelQ5_NB
BP = 1256.6, df = 26, p-value < 2.2e-16
```

insignificant!!

significant!!

- We have heteroskedasticity
- Fix with Robust standard error

Final result for Negative binomial

	Estimate	Robust SE	Pr(>|z|)	LL	UL
(Intercept)	8.5707862	6.834366e-01	4.469281e-36	7.3306821	10.0206740
post_policy	0.9549976	9.102260e-02	9.414274e-26	0.7922663	1.1511539
study_group	0.5427348	2.883495e-06	0.000000e+00	0.4684477	0.6288025
net_purchase_amount	1.0000998	6.834366e-01	1.433745e-01	1.0000941	1.0001055
genderM	0.7730239	9.102260e-02	2.019632e-17	0.7626735	0.7835147
genderU	0.8885545	2.883495e-06	0.000000e+00	0.8684211	0.9091547
age_band	0.9967145	6.834366e-01	1.447341e-01	0.9952236	0.9982077
est_income_code	1.0042104	9.102260e-02	2.662407e-28	1.0014006	1.0070281
ethnic_codeD	1.0324904	2.883495e-06	0.000000e+00	0.9376198	1.1369603
ethnic_codeF	0.9636805	6.834366e-01	1.585247e-01	0.8999901	1.0318781
ethnic_codeG	0.9900225	9.102260e-02	1.489087e-27	0.9258111	1.0586875
ethnic_codeH	0.9894011	2.883495e-06	0.000000e+00	0.9275550	1.0553710
ethnic_codeI	0.9792819	6.834366e-01	1.518924e-01	0.9124762	1.0509788
ethnic_codeJ	1.0416203	9.102260e-02	2.533487e-30	0.9582472	1.1322474
ethnic_codeM	1.0074394	2.883495e-06	0.000000e+00	0.9250282	1.0971926
ethnic_codeN	0.9794613	6.834366e-01	1.518175e-01	0.9191383	1.0437433
ethnic_codeO	0.9804812	9.102260e-02	4.675468e-27	0.9067043	1.0602612
ethnic_codeP	1.0094029	2.883495e-06	0.000000e+00	0.9141525	1.1145780
ethnic_codeR	0.9613057	6.834366e-01	1.595531e-01	0.8857832	1.0432673
ethnic_codeS	1.0072688	9.102260e-02	1.831215e-28	0.9439086	1.0748820
ethnic_codeU	1.0919924	2.883495e-06	0.000000e+00	1.0074171	1.1836679
ethnic_codeX	0.9449781	6.834366e-01	1.667613e-01	0.8487140	1.0521607
ethnic_codeZ	0.9976288	9.102260e-02	5.934748e-28	0.9351489	1.0642831
homeowner_codeR	0.9728927	2.883495e-06	0.000000e+00	0.9586263	0.9873714
length_of_residence	1.0010627	6.834366e-01	1.429896e-01	0.9999149	1.0022117
childY	0.9990540	9.102260e-02	4.991231e-28	0.9865735	1.0116924
post_policy:study_group	1.0709736	2.883495e-06	0.000000e+00	0.8881831	1.2913831

Interaction terms p value is still significant

Consumer Purchase behavior

After return policy changed, net purchase amount decreased by 18% , unit purchase amount increased by 7.09%.

- Customer bought less expensive items more!!