Regression Report – David

Initial check

What number of requests don't have the same product_cd for all their inquiries? (Out of 319,733):
13,649

What number and percentage of *List requests* have this property?
Number: ***13,088***
Percentage: ***4.116 %***

What number and percentage of *individual List inquiries* have this property?
Number: ***217,503***
Percentage***: 5.078369 %***

Regressions run

```
subset <- df.inquiry %>% filter(pt == 0,
                        product_cd == "USHY",
                        p_type != "Broker-Dealer",
                        list_length >= 20,
                        numsublists < list_length/2,
                        5 < sublist_length)

regr1 <- felm(filled ~ trans_cost + mediancost_insublist + mediancost_outsidesublist |
                            req_id | 0 | req_id + date, data = subset)

regr2 <- felm(filled ~ trans_cost + mincost_insublist + mincost_outsidesublist | req_id | 0 |
                            req_id + date, data = subset)
```

Regression 1

	Estimate	Cluster.s.e.	t.value	pval
trans_cost	−1.6429	0.8295	−1.9806	0.0477
mediancost_insublist	2.2095	0.4294	5.1451	2.75e−07
mediancost_outsidesublist	1.4890	0.7794	1.9104	0.0561

	Estimate	Cluster.s.e.	t.value	pval
trans_cost	−1.6430	0.8295	−1.9806	0.0477
mediancost_insublist	2.1987	0.4296	5.1178	3.17e−07
mediancost_outsidesublist	1.0492	0.8593	1.2210	0.222

	Estimate	Cluster.s.e.	t.value	pval
trans_cost	−1.6429	0.8295	−1.9806	0.0477
mediancost_insublist	2.1985	0.4289	5.1260	3.04e−07
mediancost_outsidesublist	0.9374	0.7550	1.2416	0.214

	Estimate	Cluster.s.e.	t.value	pval
trans_cost	−1.6429	0.8295	−1.9806	0.0477
mediancost_insublist	2.2078	0.4297	5.1379	2.85e−07
mediancost_outsidesublist	1.3058	0.7716	1.6923	0.0906

	Estimate	Cluster.s.e.	t.value	pval
trans_cost	−1.6430	0.8296	−1.9806	0.0477
mediancost_insublist	2.1941	0.4293	5.1114	3.28e−07
mediancost_outsidesublist	0.7020	0.8398	0.8359	0.403

Regression 2

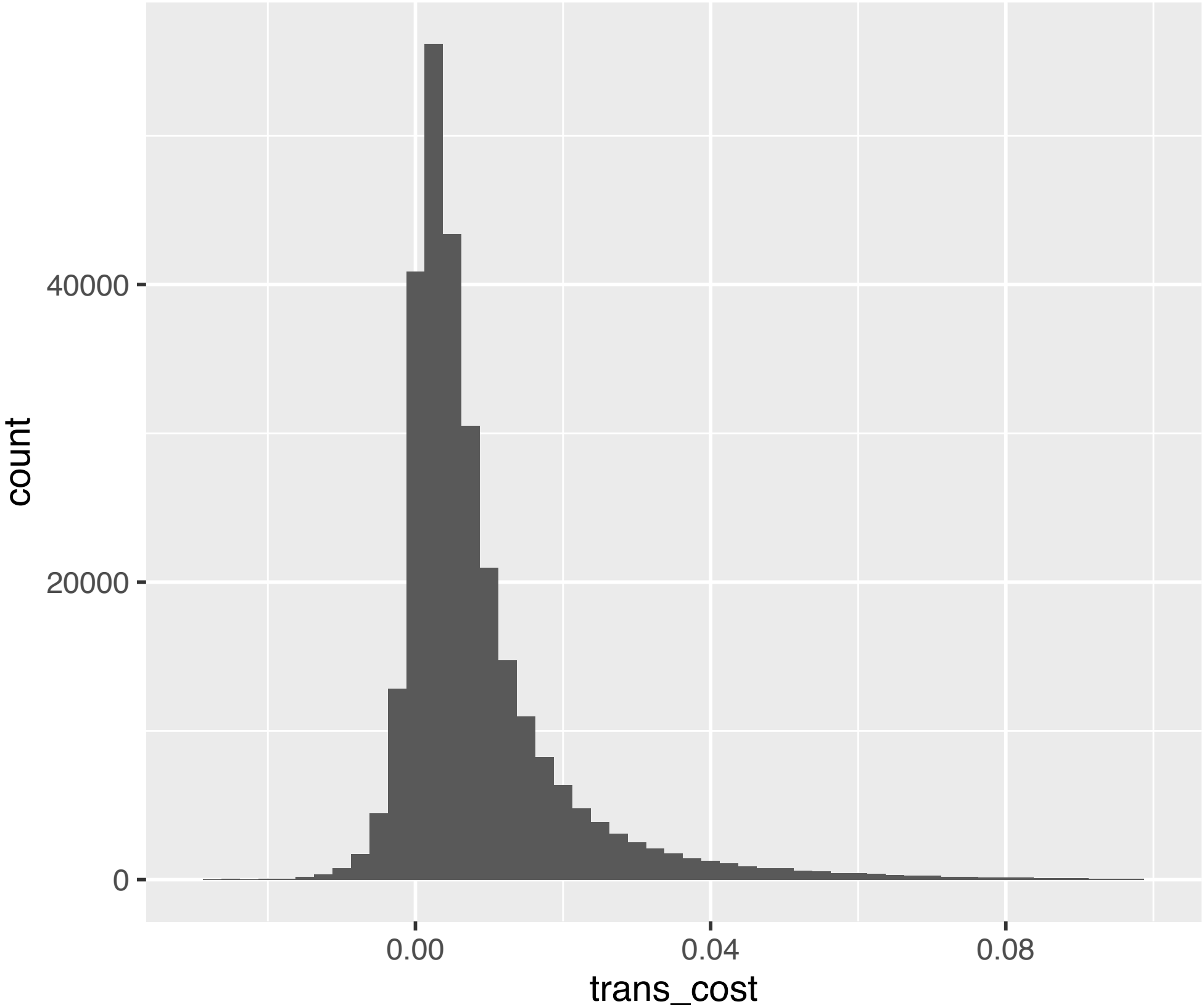
	Estimate	Cluster.s.e.	t.value	pval
trans_cost	–1.7335	0.8546	–2.0284	0.0426
mincost_insublist	–0.2064	0.0860	–2.3989	0.0165
mincost_outsidesublist	–0.1712	0.1033	–1.6576	0.0974

	Estimate	Cluster.s.e.	t.value	pval
trans_cost	–1.7336	0.8546	–2.0284	0.0426
mincost_insublist	–0.2065	0.0861	–2.3984	0.0165
mincost_outsidesublist	–0.1741	0.1032	–1.6866	0.0917

	Estimate	Cluster.s.e.	t.value	pval
trans_cost	–1.7334	0.8546	–2.0284	0.0426
mincost_insublist	–0.2064	0.0860	–2.3991	0.0165
mincost_outsidesublist	–0.1699	0.1034	–1.6440	0.1002

	Estimate	Cluster.s.e.	t.value	pval
trans_cost	–1.7282	0.8490	–2.0357	0.0418
mincost_insublist	–0.1999	0.0789	–2.5349	0.0113
mincost_outsidesublist	0.1542	0.1187	1.2992	0.1939

	Estimate	Cluster.s.e.	t.value	pval
trans_cost	–1.7282	0.8490	–2.0357	0.0418
mincost_insublist	–0.1999	0.0789	–2.5348	0.0113
mincost_outsidesublist	0.1528	0.1190	1.2843	0.1991


40000
20000
0
count
0.00
0.04
0.08
trans_cost

RegressionSummary

David

2024-03-18

```
subset <- read_csv("~/Desktop/Portfolio_Trades_my_computer/data_minimizing/working_files/subset.csv")
```

```
## Rows: 297381 Columns: 16
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (3): product_cd, p_type, request_type
## dbl (13): req_id, req_quantity, trade_quantity, spread, filled, trans_cost, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
regr1 <- felm(filled ~ trans_cost + mediancost_insublist +
                mediancost_outsidesublist | req_id | 0 | req_id, data = subset)
summary1 <- summary(regr1, cluster = c("req_id"))
print(summary1)
```

```
##
## Call:
##    felm(formula = filled ~ trans_cost + mediancost_insublist + mediancost_outsidesublist |      req_
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.2535  -0.2189   0.0078   0.1995   2.1468
##
## Coefficients:
##                            Estimate Cluster s.e. t value Pr(>|t|)
## trans_cost                  -1.6429       0.8295  -1.981   0.0477 *
## mediancost_insublist         2.2095       0.4294   5.145 2.75e-07 ***
## mediancost_outsidesublist    1.4890       0.7794   1.910   0.0561 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3766 on 188840 degrees of freedom
##   (101568 observations deleted due to missingness)
## Multiple R-squared(full model): 0.447   Adjusted R-squared: 0.4265
## Multiple R-squared(proj model): 0.01872   Adjusted R-squared: -0.01751
## F-statistic(full model, *iid*):21.89 on 6972 and 188840 DF, p-value: < 2.2e-16
## F-statistic(proj model): 11.47 on 3 and 6969 DF, p-value: 1.68e-07
```

```
regr2 <- felm(filled ~ trans_cost + mincost_insublist +
                mincost_outsidesublist | req_id | 0 | req_id, data = subset)
summary2 <- summary(regr2, cluster = c("req_id"))
print(summary2)
```

```
##
## Call:
##    felm(formula = filled ~ trans_cost + mincost_insublist + mincost_outsidesublist |      req_id | 0
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0910  -0.2188   0.0074   0.1997   2.5319
##
## Coefficients:
##                          Estimate Cluster s.e. t value Pr(>|t|)
## trans_cost              -1.73346      0.85460  -2.028   0.0426 *
## mincost_insublist       -0.20641      0.08604  -2.399   0.0165 *
## mincost_outsidesublist -0.17123      0.10330  -1.658   0.0974 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3765 on 188840 degrees of freedom
##   (101568 observations deleted due to missingness)
## Multiple R-squared(full model): 0.4474   Adjusted R-squared: 0.427
## Multiple R-squared(proj model): 0.01944   Adjusted R-squared: -0.01676
## F-statistic(full model, *iid*):21.93 on 6972 and 188840 DF, p-value: < 2.2e-16
## F-statistic(proj model):   2.9 on 3 and 6969 DF, p-value: 0.03364
```

Simple regression model of filled on trans_cost for non-dealer HY requests

without list fixed effects

```
model1_fe <- lm(filled ~ trans_cost, data = subset)
summary(model1_fe)
```

```
##
## Call:
## lm(formula = filled ~ trans_cost, data = subset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.4846  -0.5083   0.4638   0.4801  15.6350
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.5246294  0.0009569  548.24   <2e-16 ***
## trans_cost  -1.7719032  0.0247742  -71.52   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4955 on 282674 degrees of freedom
```

```
##   (14705 observations deleted due to missingness)
## Multiple R-squared:  0.01777,    Adjusted R-squared:  0.01777
## F-statistic:  5115 on 1 and 282674 DF,  p-value: < 2.2e-16
```

Simple regression model of filled on trans_cost for non-dealer HY requests

with list fixed effects

This is run with felm, because lm with req_id fixed effect is computationally very intensive (vector memory is exceeded)

```
model2_nofe <- felm(filled ~ trans_cost | req_id, data = subset)
summary(model2_nofe)
```

```
##
## Call:
##    felm(formula = filled ~ trans_cost | req_id, data = subset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6516  -0.2248  -0.0003   0.1987   9.0649
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## trans_cost -1.21635    0.01997   -60.9   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.379 on 271211 degrees of freedom
##   (14705 observations deleted due to missingness)
## Multiple R-squared(full model): 0.4487   Adjusted R-squared: 0.4254
## Multiple R-squared(proj model): 0.01349   Adjusted R-squared: -0.02821
## F-statistic(full model):19.25 on 11464 and 271211 DF, p-value: < 2.2e-16
## F-statistic(proj model):  3708 on 1 and 271211 DF, p-value: < 2.2e-16
```