

University of St. Gallen

School of Economics, Law, Social Sciences, International Relations and Computer Science

TBD: Modelling this and that

By

Tobias Jucker, Jingxuan Song, Mathias Steilen, Dan Wang

Matriculation N°: / / 19-608-512 /

Course: Concepts, Applications, and Implications in Machine Learning and Artificial Intelligence

Lecturer: Prof. Dr. Clemens Stachl

29.09.2022

Contents

1	CAIML Group Project: Group 2	1
1.1	Admin	1
1.2	Data Initialisation and Inspection	1
1.3	Question 1: “Randomization”	4
1.3.1	a)	4
1.3.2	b)	5
1.3.3	c)	8
1.4	Question 2: “Descriptive Evidence”	9
1.4.1	a)	9
1.4.2	b)	10
1.4.3	c)	11
1.5	Question 3: “Treatment Effect Estimation”	12
1.5.1	a)	12
1.5.2	b)	28
1.5.3	c)	28
1.6	Question 4: “Discussion of the Results”	30
1.6.1	a)	30
1.6.2	b)	30
2	References	32

List of Figures

1	NA values present in the data	4
2	Distribution of data on participating messenger pre-experiment period	6
3	Correlation matrix for regressors employed in OLS (1, 2)	13
4	Correlation matrix for regressors employed in OLS (3)	21

List of Tables

1	Descriptive Summary for Participating Messengers: Pre-Experiment Period	5
2	Two Sample t-test: <code>rev_sample_A</code> and <code>rev_sample_B</code>	7
3	Two Sample t-test: <code>shifts_sample_A</code> and <code>shifts_sample_B</code>	8
4	Difference in mean deviation from mean for revenue between group A and group B .	10
5	Difference in mean deviation from mean for shifts between group A and group B . .	11
6	Regression Results (OLS) for Revenue	26
7	Regression Results (OLS) for Shifts	27
8	Original R squared from regular lm call	28

1 CAIML Group Project: Group 2

1.1 Admin

```
# Set working directory to source file location
setwd(dirname(rstudioapi::getActiveDocumentContext())$path))

# Load packages
library(magrittr)
library(dplyr)
library(ggplot2)
library(GGally)
library(data.table)
library(stargazer)
library(gridExtra)
library(pander)
library(plm)
library(janitor)
library(car)
```

1.2 Data Initialisation and Inspection

We start by importing the RData and converting it to the class *data.frame*.

```
# Import RData provided
load("Data/fehr.RData")
fehr <- fehr %>% as.data.frame()
```

After that, we proceed to look at the data to get a picture of the content.

```
# Inspect
glimpse(fehr)
```

```
## Rows: 386
## Columns: 11
## $ vebli      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ block      <dbl> 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3, 2, 3, 1, 2, 3, 1, 2, 3,~
## $ totrev      <dbl> 3090.9, 1210.0, 782.0, 2308.5, 2148.0, 1936.5, 3552.5, 1889~
## $ shifts      <dbl> 13, 5, 3, 8, 6, 6, 9, 4, 5, 4, 0, 4, 4, 0, 5, 11, 12, 18, 0~
## $ treatment  <dbl> 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1,~
## $ maxhigh     <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1,~
## $ shifts_fe   <dbl> 6.0000000, -2.0000000, -4.0000000, 1.3333335, -0.6666665, --
## $ totrev_fe   <dbl> 1396.6000, -484.3000, -912.3001, 177.5000, 17.0000, -194.50~
## $ experiment <dbl> 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1,~
## $ odd         <dbl> 1, 1, 1, 1, 1, 1, 0, 0, 0, NA, NA, NA, 1, 1, 1, 1, 1, 0, 0,~
## $ fahrer      <dbl> 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 6, 6, 6, 7, 7, 7,~
```

```
summary(fehr)
```

```
##      vebli      block      totrev      shifts
## Min.   :0.0000  Min.   :1.000  Min.    :  0.0  Min.    : 0.000
## 1st Qu.:0.0000  1st Qu.:1.000  1st Qu.: 302.8  1st Qu.: 2.000
## Median :0.0000  Median :2.000  Median : 1228.0  Median : 5.000
## Mean   :0.4922  Mean   :2.047  Mean    : 1926.0  Mean    : 7.244
## 3rd Qu.:1.0000  3rd Qu.:3.000  3rd Qu.: 3016.2  3rd Qu.:12.000
## Max.    :1.0000  Max.    :3.000  Max.     :12487.8  Max.     :30.000
##
##      treatment      maxhigh      shifts_fe      totrev_fe
## Min.   :0.0000  Min.   :0.0000  Min.    : -14.67  Min.    : -6177.33
## 1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:  -2.00  1st Qu.: -517.83
## Median :0.0000  Median :0.0000  Median :   0.00  Median :  -34.67
## Mean   :0.1088  Mean   :0.3212  Mean    :   0.00  Mean    :    0.00
## 3rd Qu.:0.0000  3rd Qu.:1.0000  3rd Qu.:   2.00  3rd Qu.:  487.00
## Max.    :1.0000  Max.    :1.0000  Max.    : 11.67  Max.    : 3731.67
##
```

```
##      experiment      odd      fahrer
##  Min.      :0.0000  Min.      :0.0000  Min.      : 1.00
##  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.: 33.00
##  Median :0.0000  Median :1.0000  Median : 67.50
##  Mean   :0.3342  Mean   :0.5242  Mean   : 66.66
##  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:100.00
##  Max.    :1.0000  Max.    :1.0000  Max.    :138.00
##                                     NA's    :262
```

We can observe that all variables in the data set are of type *numeric* and no variables need type conversion. However, there seem to be *NA* values present. A short overview in Figure 1 shows, that these are only present in one variable, namely *odd*. This variable indicates, whether the respective messenger has been attributed to group A (1) or group B (0). Only participating messengers were assigned, so we can conclude that the messengers who have *NA* values in column *odd* did not participate in the experiment. As *maxhigh* is an indicator variable of the participation in the experiment, we can check if this assumption holds.

```
odd_NA <- fehr %>% is.na() %>% sum()
maxhigh_zero <- sum(fehr$maxhigh == 0)

odd_NA == maxhigh_zero
```

```
## [1] TRUE
```

As shown by the count, these two variables reveal the same number of observations of messengers not participating, hence we could use either of them to proceed with filtering for the exercises.

```
colSums(is.na(fehr)) %>% barplot(las = 2)
```

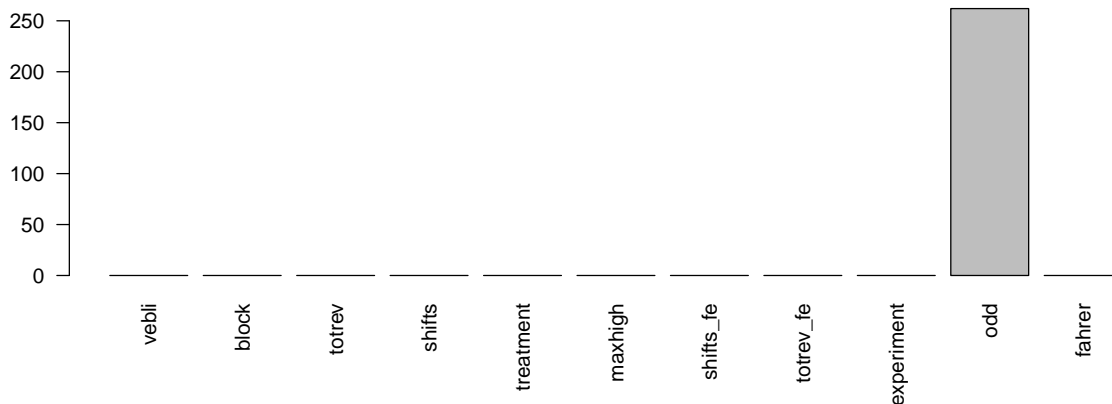


Figure 1: NA values present in the data

1.3 Question 1: “Randomization”

For now, focus on the messengers participating in the experiment (i.e., group A and group B). Analyze the differences in mean revenues and mean shifts during the four-week period prior to the experiment by answering the following questions:

1.3.1 a)

Produce a descriptive table for the pre-experiment period showing mean revenues and mean shifts by group.

For the descriptive table for the pre-experiment period by group, we need to exclude observations from block 2 and 3, as well as observations on non-participating messengers, as stated by the exercise. We then group by using the column *odd*, as we know the attribution system, but cannot rely on the ID anymore due to the pseudo issue.

```
fahr %>%
  filter(!is.na(odd)) %>% # filter out non-participating messengers
  filter(block == 1) %>% # filter out observations from later blocks
  mutate("Group" = ifelse(odd == 1, "A", "B")) %>%
  group_by(Group) %>%
  summarise("Average Revenue" = mean(totrev),
            "SD Revenue" = sd(totrev),
```

```

    "Average Shifts" = mean(shifts),
    "SD Shifts" = sd(shifts),
    N = n()) %>%
as.data.frame() %>%
stargazer(header = FALSE, type = 'latex', summary = FALSE, digits = 2,
          title = "Descriptive Summary for
          Participating Messengers: Pre-Experiment Period")

```

Table 1: Descriptive Summary for Participating Messengers: Pre-Experiment Period

	Group	Average Revenue	SD Revenue	Average Shifts	SD Shifts	N
1	A	3,500.67	2,703.25	12.14	8.06	21
2	B	3,258.99	2,319.46	10.95	7.58	19

This results in similar figures, except for revenues in Group B, which we assume has something to do with the data and the remark in the exercises, that the figures will not be perfectly equal.

1.3.2 b)

Test whether the observed differences in the pre-experiment period are statistically significant.

In order to test the statistical significance of the difference, we apply a two-sided t-test on the data. We first take a look at the distributions in our data set in Figure 2. The significant overlap of the boxplots gives us a first impression of the two means likely being equal.

```

par(mfrow = c(2,2))

gg1 <- fehr %>%
  filter(!is.na(odd)) %>%
  filter(block == 1) %>%
  mutate("Group" = ifelse(odd == 1, "A", "B")) %>%
  ggplot(aes(x = Group, y = totrrev, colour = Group)) +
  geom_boxplot(outlier.color = NA) +
  geom_jitter(alpha = 0.2, width = 0.2) +

```



```

labs(y = "Total Revenue",
     title = "Revenue Pre-Experiment") +
theme_minimal() +
theme(legend.position = "none")

gg2 <- fehr %>%
  filter(!is.na(odd)) %>%
  filter(block == 1) %>%
  mutate("Group" = ifelse(odd == 1, "A", "B")) %>%
  ggplot(aes(x = Group, y = shifts, colour = Group)) +
  geom_boxplot(outlier.color = NA) +
  geom_jitter(alpha = 0.2, width = 0.2) +
  labs(y = "Total Shifts",
       title = "Shifts Pre-Experiment") +
  theme_minimal() +
  theme(legend.position = "none")

grid.arrange(gg1, gg2, ncol = 2)

```

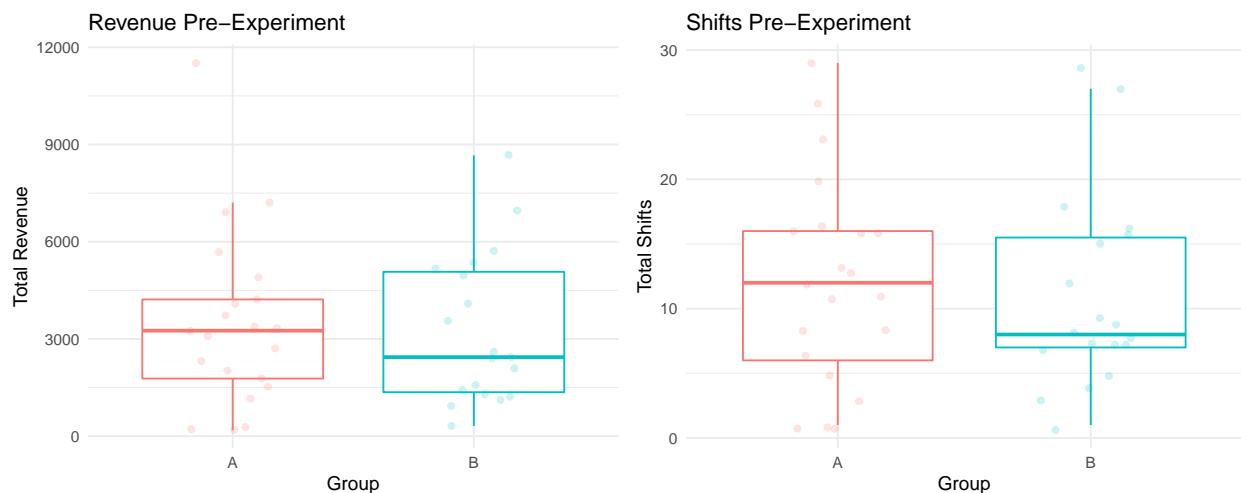


Figure 2: Distribution of data on participating messenger pre-experiment period

As we have a sufficient number of observations for the CLT to hold, the underlying distribution of

the data should not matter and we assume that the normality assumption holds for the means. Therefore, we go beyond the informal visual impression and proceed with the t-Test.

We define the null hypothesis $H_0 : \mu_A = \mu_B$ as *the mean of group A is equal to the mean of group B*. Therefore, $H_1 : \mu_A \neq \mu_B$ states that the difference in the means is significant and that they are indeed not the same. We have independent samples due to the random assignment of messengers into group A and B and we assume that the variance in both populations is roughly the same for the same reasons.

```
rev_sample_A <-
  fehr$totrev[fehr$block == 1 & fehr$maxhigh == 1 & fehr$odd == 1]

rev_sample_B <-
  fehr$totrev[fehr$block == 1 & fehr$maxhigh == 1 & fehr$odd == 0]

t_test_rev <- t.test(rev_sample_A,
                     rev_sample_B,
                     var.equal = TRUE, alternative = "two.sided")

pander(t_test_rev)
```

Table 2: Two Sample t-test: `rev_sample_A` and `rev_sample_B`

Test statistic	df	P value	Alternative hypothesis	mean of x	mean of y
0.3018	38	0.7644	two.sided	3501	3259

```
shifts_sample_A <-
  fehr$shifts[fehr$block == 1 & fehr$maxhigh == 1 & fehr$odd == 1]

shifts_sample_B <-
  fehr$shifts[fehr$block == 1 & fehr$maxhigh == 1 & fehr$odd == 0]
```

```
t_test_shifts <- t.test(shifts_sample_A,
                        shifts_sample_B,
                        var.equal = TRUE, alternative = "two.sided")

pander(t_test_shifts)
```

Table 3: Two Sample t-test: shifts_sample_A and shifts_sample_B

Test statistic	df	P value	Alternative hypothesis	mean of x	mean of y
0.4818	38	0.6327	two.sided	12.14	10.95

1.3.3 c)

What do you conclude?

In both t-tests, the p-values are very high and lie above the $\alpha = 0.1$ threshold, also indicated by the low t-statistics, implying that our means lie within the confidence intervals of our null hypothesis H_0 . Therefore, we cannot reject the null hypothesis and conclude, that both means are equal. The difference in the means for both revenues and shifts is not statistically significant.

1.4 Question 2: “Descriptive Evidence”

Let us focus again on the messengers participating in the experiment. Calculate the following quantities and test whether they are statistically different from zero:

1.4.1 a)

The difference between group A and group B in revenues for treatment period 1. This time, express the revenues as average deviations from individual means.

```
# Difference in revenues between group A and group B
fehr2 <- fehr %>%
  filter(!is.na(odd)) %>%
  group_by(fahrer) %>%
  summarise(messenger_mean = mean(totrev),
            Group = last(odd),
            rev_block1 = nth(totrev, n = 1),
            rev_block2 = nth(totrev, n = 2),
            rev_block3 = nth(totrev, n = 3)) %>%
  mutate(Group = ifelse(Group == 1, "A", "B")) %>%
  as.data.frame()

fehr3 <- fehr2 %>%
  mutate(dev_block1 = rev_block1 - messenger_mean,
         dev_block2 = rev_block2 - messenger_mean,
         dev_block3 = rev_block3 - messenger_mean) %>%
  filter(!is.na(dev_block3)) %>% # two NA values present for block 3
  group_by(Group) %>%
  summarise("Mean Deviation (Block 2)" = mean(dev_block2)) %>%
  t() %>%
  row_to_names(1) %>%
  as.data.frame() %>%
  mutate(across(c(A, B), as.numeric)) %>%
```

```
mutate("Difference" = abs(A - B))
```

Table 4: Difference in mean deviation from mean for revenue between group A and group B

	A	B	Difference
Mean Deviation (Block 2)	734.21	-220.97	955.18

1.4.2 b)

The difference between group A and group B in the number of shifts for treatment period 2. This time, express the number of shifts as average deviations from individual means.

```
# Difference in shifts between group A and group B
```

```
fehr2 <- fehr %>%
  filter(!is.na(odd)) %>%
  group_by(fahrer) %>%
  summarise(messenger_mean = mean(shifts),
            Group = last(odd),
            shifts_block1 = nth(shifts, n = 1),
            shifts_block2 = nth(shifts, n = 2),
            shifts_block3 = nth(shifts, n = 3)) %>%
  mutate(Group = ifelse(Group == 1, "A", "B")) %>%
  as.data.frame()

fehr4 <- fehr2 %>%
  mutate(dev_shifts_block1 = shifts_block1 - messenger_mean,
         dev_shifts_block2 = shifts_block2 - messenger_mean,
         dev_shifts_block3 = shifts_block3 - messenger_mean) %>%
  filter(!is.na(dev_shifts_block3)) %>%
  group_by(Group) %>%
  summarise("Mean Deviation (Block 3)" = mean(dev_shifts_block3)) %>%
```

```

t() %>%
row_to_names(1) %>%
as.data.frame() %>%
mutate(across(c(A, B), as.numeric)) %>%
mutate("Difference" = abs(A - B))

```

Table 5: Difference in mean deviation from mean for shifts between group A and group B

	A	B	Difference
Mean Deviation (Block 3)	-2.778	1.333	4.111

1.4.3 c)

What do you conclude?

Revenues: In treatment period 1, group A receives the higher wage. We can see that there is a positive deviation from the person-specific mean for group A and a negative deviation from the person-specific mean for group B. The revenues generated by group A is much larger than those of group B. The difference in revenues between group A and group B is therefore positive (955.1798). This indicates a large treatment effect.

Shifts: In treatment period 2, group B receives the treatment and works about 4 shifts more than the control group A.

In conclusion, we can determine that the group which receives the treatment generates about CHF 1000, 00 of additional revenues and also works approximately 4 shifts more compared to the control group and using the mean deviation from individuals means within both groups.

1.5 Question 3: “Treatment Effect Estimation”

1.5.1 a)

Replicate Table 3 of the paper (all 6 models).

All 6 models estimated by Fehr and Goette (2007) follow the linear form $r_{it} = \alpha_i + \delta T_{it} + d_t + e_{it}$, where:

- r_{it} is total revenue generated by messenger i during treatment period t ;
- α_i is the fixed-effect intercept for messenger i ;
- T_{it} is the dummy variable for the treatment;
- d_t is a time dummy for treatment period 1 and 2;
- e_{it} is the error term.

We create time dummies for treatment period 1 and 2 before proceeding with applying the models.

```
fehr <- fehr %>%
  mutate(treat_dummy_1 = ifelse(block == 2, 1, 0),
         treat_dummy_2 = ifelse(block == 3, 1, 0))
```

Additionally, we can quickly look at the correlation matrix in Figure 3 for our predictors and the variance inflation factors:

```
fehr %>%
  filter(!is.na(odd)) %>%
  select(fahrer, treatment, treat_dummy_1, treat_dummy_2) %>%
  ggcorr(label_round = 2, label = TRUE)
```

```
par(mfrow = c(1,1))
```

```
fehr %>%
  filter(!is.na(odd)) %>%
```

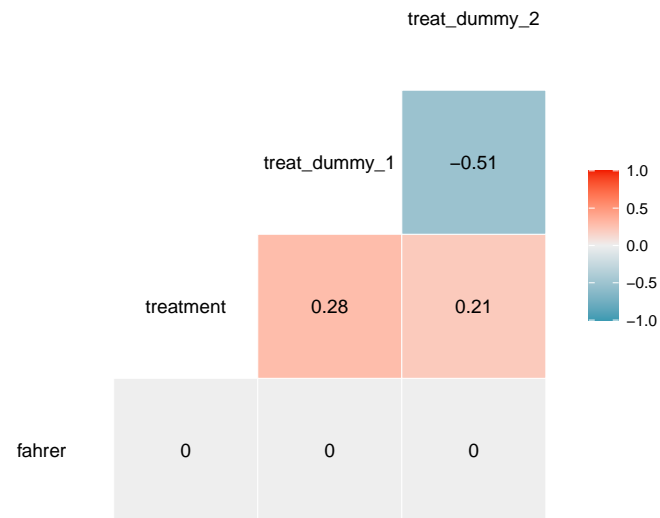


Figure 3: Correlation matrix for regressors employed in OLS (1, 2)

```
lm(formula = totrev ~ factor(fahrer) +
    treatment +
    factor(treat_dummy_1) +
    factor(treat_dummy_2)) %>%

vif() %>%
as.data.frame() %>%
select(GVIF)
```

```
##              GVIF
## factor(fahrer)  1.016129
## treatment      1.325587
## factor(treat_dummy_1) 1.735887
## factor(treat_dummy_2) 1.672764
```

Looking at the correlation matrix, we can observe that the predictors are lightly to moderately correlated. The generalised VIFs are close to 1, however. Therefore, we can assume that no issue of multicollinearity persists.

1.5.1.1 Model 1: Participating messengers: In order to show individual fixed effects, we add the pseudo ID as a factor in the regression, as done by (Fehr & Goette, 2007). This removes a potential endogeneity issue caused by an omitted variable by measuring changes within individual observations across time and fixing the average effects of each individual.

```
mdl1 <- fehr %>%
  filter(!is.na(odd)) %>%
  lm(formula = totrev ~ factor(fahrer) +
      treatment +
      factor(treat_dummy_1) +
      factor(treat_dummy_2))

summary(mdl1)
```

```
##
## Call:
## lm(formula = totrev ~ factor(fahrer) + treatment + factor(treat_dummy_1) +
##     factor(treat_dummy_2), data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4376.3  -656.5    46.9   831.3  3355.6
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1611.67     911.80   1.768  0.08099 .
## factor(fahrer)2       436.70    1258.40   0.347  0.72949
## factor(fahrer)3       747.20    1258.40   0.594  0.55436
## factor(fahrer)5     -1270.41    1410.45  -0.901  0.37048
## factor(fahrer)6      1537.50    1258.40   1.222  0.22542
## factor(fahrer)7       750.17    1258.40   0.596  0.55279
## factor(fahrer)8     -106.97    1258.40  -0.085  0.93247
## factor(fahrer)9      3912.20    1258.40   3.109  0.00261 **
```

## factor(fahrer)14	884.00	1258.40	0.702	0.48445
## factor(fahrer)15	-39.97	1258.40	-0.032	0.97474
## factor(fahrer)18	1327.37	1258.40	1.055	0.29473
## factor(fahrer)19	2658.97	1258.40	2.113	0.03776 *
## factor(fahrer)21	-514.13	1258.40	-0.409	0.68397
## factor(fahrer)22	2955.60	1258.40	2.349	0.02134 *
## factor(fahrer)23	3330.23	1258.40	2.646	0.00981 **
## factor(fahrer)24	1759.17	1258.40	1.398	0.16604
## factor(fahrer)25	61.17	1258.40	0.049	0.96136
## factor(fahrer)28	-150.17	1258.40	-0.119	0.90532
## factor(fahrer)30	1557.53	1258.40	1.238	0.21949
## factor(fahrer)31	-737.97	1258.40	-0.586	0.55926
## factor(fahrer)32	-526.10	1258.40	-0.418	0.67703
## factor(fahrer)33	2173.20	1258.40	1.727	0.08808 .
## factor(fahrer)34	9107.03	1258.40	7.237	2.62e-10 ***
## factor(fahrer)35	-64.00	1258.40	-0.051	0.95957
## factor(fahrer)36	2054.90	1258.40	1.633	0.10646
## factor(fahrer)37	2193.97	1258.40	1.743	0.08514 .
## factor(fahrer)38	1836.93	1258.40	1.460	0.14833
## factor(fahrer)42	3938.03	1258.40	3.129	0.00245 **
## factor(fahrer)44	521.87	1258.40	0.415	0.67948
## factor(fahrer)45	569.77	1258.40	0.453	0.65196
## factor(fahrer)49	1044.20	1258.40	0.830	0.40916
## factor(fahrer)50	6590.67	1258.40	5.237	1.32e-06 ***
## factor(fahrer)51	5288.97	1258.40	4.203	6.88e-05 ***
## factor(fahrer)52	4278.30	1258.40	3.400	0.00106 **
## factor(fahrer)53	3433.03	1258.40	2.728	0.00785 **
## factor(fahrer)55	976.27	1258.40	0.776	0.44018
## factor(fahrer)56	1761.80	1258.40	1.400	0.16542
## factor(fahrer)57	2471.53	1258.40	1.964	0.05304 .
## factor(fahrer)58	1353.20	1258.40	1.075	0.28550
## factor(fahrer)60	931.03	1258.40	0.740	0.46158

```
## factor(fahrer)61          664.70    1258.40    0.528    0.59883
## factor(fahrer)63          -261.11    1410.45   -0.185    0.85360
## treatment                 1033.56     336.70    3.070    0.00294 **
## factor(treat_dummy_1)1    -210.97     385.31   -0.548    0.58555
## factor(treat_dummy_2)1    -574.71     378.23   -1.519    0.13264
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1541 on 79 degrees of freedom
## Multiple R-squared:  0.7404, Adjusted R-squared:  0.5958
## F-statistic: 5.121 on 44 and 79 DF,  p-value: 1.596e-10
```

However, the latter prints a lot of information, as the factor call includes the subject-specific means for every *fahrer* pseudo ID. As we want to focus on the factor loadings of the treatment and time dummy regressors, we repeat the same model using the within (fixed effects) estimator, from the *plm* package, yielding a cleaner output:

```
mdl1_plm <- fehr %>%
  filter(!is.na(odd)) %>%
  plm(formula = totrev ~ treatment +
        factor(treat_dummy_1) +
        factor(treat_dummy_2),
        index = "fahrer", model = "within")

summary(mdl1_plm)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = totrev ~ treatment + factor(treat_dummy_1) + factor(treat_dummy_2),
##      data = ., model = "within", index = "fahrer")
##
## Unbalanced Panel: n = 42, T = 2-3, N = 124
```

```
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -4376.275 -656.511    46.934    831.261   3355.558
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## treatment           1033.56     336.70  3.0696 0.002937 **
## factor(treat_dummy_1)1  -210.97     385.31 -0.5475 0.585547
## factor(treat_dummy_2)1  -574.71     378.23 -1.5195 0.132639
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    214030000
## Residual Sum of Squares: 187650000
## R-Squared:              0.12322
## Adj. R-Squared: -0.36511
## F-statistic: 3.70093 on 3 and 79 DF, p-value: 0.01508
```

This deflates the R Squared, as the information contained in the individual fixed effects dummy variables is excluded, painting a more reasonable actual accuracy of the model.

1.5.1.2 Model 2: All messengers as Veloblitz: Model two looks at all messengers at Veloblitz. As we have a binary variable *vebli*, we can easily filter for these messengers' observations using *dplyr* again. Henceforth, we will not include the whole summary for the base R fixed effects regression call, as it unnecessarily makes this replication paper longer. Additionally, the summarised results for *plm* OLS calls can be found in Table 6 and Table 7.

```
mdl2 <- fehr %>%
  filter(vebli == 1) %>%
  lm(formula = totrev ~ factor(fahrer) +
      treatment +
```

```
factor(treat_dummy_1) +
factor(treat_dummy_2))
```

Doing the same using the within estimator from plm package for cleaner output:

```
mdl2_plm <- fehr %>%
  filter(vebli == 1) %>%
  plm(formula = totrev ~ treatment +
      factor(treat_dummy_1) +
      factor(treat_dummy_2),
      index = "fahrer", model = "within")

summary(mdl2_plm)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = totrev ~ treatment + factor(treat_dummy_1) + factor(treat_dummy_2),
##      data = ., model = "within", index = "fahrer")
##
## Unbalanced Panel: n = 66, T = 1-3, N = 190
##
## Residuals:
```

	Min.	1st Qu.	Median	3rd Qu.	Max.
	-4436.352	-577.899	-32.302	644.690	3295.482

```
##
## Coefficients:
```

	Estimate	Std. Error	t-value	Pr(> t)
treatment	1094.50	282.87	3.8692	0.0001774 ***
factor(treat_dummy_1)1	-370.62	264.58	-1.4008	0.1638365
factor(treat_dummy_2)1	-656.23	261.41	-2.5103	0.0133814 *

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    257310000
## Residual Sum of Squares: 225240000
## R-Squared:              0.12464
## Adj. R-Squared:        -0.36729
## F-statistic: 5.7432 on 3 and 121 DF, p-value: 0.0010446
```

1.5.1.3 Model 3: All messengers as Flash and Veloblitz: For Model 3, we need to add a dummy variable for the non-treated at Veloblitz. Fehr and Goette (2007) state that this binary variable was created based on the “whole nontreated group at Veloblitz” in order to show how Veloblitz messengers differ from Flash messengers. At this point, it is highly relevant that this dummy variable is not created solely based on the treatment dummy, as this would lead to collinearity and make the new dummy variable worthless. Hence, the dummy must conform with the rule that no treatment is applied and the messenger must be at Veloblitz. Additionally, during the first block, no person was part of a control group, as no treatment had been applied at that stage in time. Therefore, the dummy variable must take the value 0 for all messengers during Block 1. Following this reasoning, which we found by trial and error, the replication takes on the right values.

```
fehr <- fehr %>%
  mutate(nt_vb_dummy = ifelse(treatment == 0 & vebli == 1, 1, 0),
         nt_vb_dummy = ifelse(nt_vb_dummy == 1 & block == 1, 0, nt_vb_dummy))
```

```
mdl3 <- fehr %>%
  lm(formula = totrev ~ factor(fahrer) +
      treatment +
      factor(treat_dummy_1) +
      factor(treat_dummy_2) +
      factor(nt_vb_dummy))
```

Doing the same using the within estimator from plm package for cleaner output:

```

mdl3_plm <- fehr %>%
  plm(formula = totrev ~ treatment +
        factor(treat_dummy_1) +
        factor(treat_dummy_2) +
        factor(nt_vb_dummy),
        index = "fahrer", model = "within")

summary(mdl3_plm)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = totrev ~ treatment + factor(treat_dummy_1) + factor(treat_dummy_2) +
##      factor(nt_vb_dummy), data = ., model = "within", index = "fahrer")
##
## Unbalanced Panel: n = 138, T = 1-3, N = 386
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -5831.9332 -474.0468   -5.4398   491.5033  3426.5876
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## treatment           1035.801    344.051   3.0106 0.002881 **
## factor(treat_dummy_1)1 -264.758    223.643  -1.1838 0.237628
## factor(treat_dummy_2)1 -650.479    223.643  -2.9086 0.003966 **
## factor(nt_vb_dummy)1   -54.398    305.213  -0.1782 0.858690
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    463310000
## Residual Sum of Squares: 416060000

```

```
## R-Squared:      0.102
## Adj. R-Squared: -0.41692
## F-statistic: 6.92883 on 4 and 244 DF, p-value: 2.6802e-05
```

```
fehr %>%
  select(fahrer, treatment, treat_dummy_1, treat_dummy_2, nt_vb_dummy) %>%
  ggcorr(label_round = 2, label = TRUE)
```

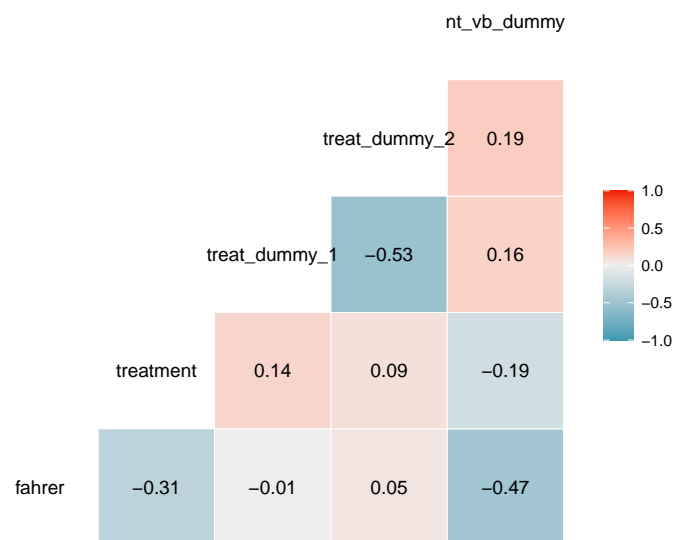


Figure 4: Correlation matrix for regressors employed in OLS (3)

Again, the correlation between the regressors is moderate and we assume that no multicollinearity issue persists and that we can trust the p-values.

1.5.1.4 Model 4: Participating messengers: Model 4 to 6 now take the same logic, but regress *shifts* on all previous regressors. We won’t include the same explanations again, but proceed with the code.

```
mdl4 <- fehr %>%
  filter(!is.na(odd)) %>%
  lm(formula = shifts ~ factor(fahrer) +
      treatment +
```



```
factor(treat_dummy_1) +
factor(treat_dummy_2))
```

Doing the same using the within estimator from plm package for cleaner output:

```
mdl4_plm <- fehr %>%
  filter(!is.na(odd)) %>%
  plm(formula = shifts ~ treatment +
      factor(treat_dummy_1) +
      factor(treat_dummy_2),
      index = "fahrer", model = "within")

summary(mdl4_plm)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = shifts ~ treatment + factor(treat_dummy_1) + factor(treat_dummy_2),
##      data = ., model = "within", index = "fahrer")
##
## Unbalanced Panel: n = 42, T = 2-3, N = 124
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -13.950000  -2.283333   0.020833   2.410227  11.383333
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## treatment              3.9864    1.1356   3.5104 0.0007423 ***
## factor(treat_dummy_1)1  -1.2750    1.2995  -0.9812 0.3295098
## factor(treat_dummy_2)1  -2.5614    1.2756  -2.0079 0.0480716 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    2513.3
## Residual Sum of Squares: 2134.5
## R-Squared:              0.15074
## Adj. R-Squared: -0.32226
## F-statistic: 4.67417 on 3 and 79 DF, p-value: 0.004662
```

```
mdl5 <- fehr %>%
  filter(vebli == 1) %>%
  lm(formula = shifts ~ factor(fahrer) +
      treatment +
      factor(treat_dummy_1) +
      factor(treat_dummy_2))
```

1.5.1.5 Model 5: All messengers as Veloblitz: Doing the same using the within estimator from plm package for cleaner output:

```
mdl5_plm <- fehr %>%
  filter(vebli == 1) %>%
  plm(formula = shifts ~ treatment +
      factor(treat_dummy_1) +
      factor(treat_dummy_2),
      index = "fahrer", model = "within")

summary(mdl5_plm)
```

```
## Oneway (individual) effect Within Model
```

```
##
```

```
## Call:
```

```
## plm(formula = shifts ~ treatment + factor(treat_dummy_1) + factor(treat_dummy_2),
```

```
##      data = ., model = "within", index = "fahrer")
##
## Unbalanced Panel: n = 66, T = 1-3, N = 190
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -14.03971  -1.83016  -0.17725   2.12184  11.29362
##
## Coefficients:
##
##              Estimate Std. Error t-value Pr(>|t|)
## treatment           4.08254    0.99486  4.1036 7.414e-05 ***
## factor(treat_dummy_1)1 -1.57040    0.93052 -1.6877  0.094053 .
## factor(treat_dummy_2)1 -2.63127    0.91938 -2.8620  0.004962 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    3242
## Residual Sum of Squares: 2786
## R-Squared:              0.14066
## Adj. R-Squared: -0.34228
## F-statistic: 6.60169 on 3 and 121 DF, p-value: 0.00036088
```

1.5.1.6 Model 6: All messengers as Flash and Veloblitz: At this stage, we require the same dummy variable as in Model 3.

```
fehr <- fehr %>%
  mutate(nt_vb_dummy = ifelse(treatment == 0 & vebli == 1, 1, 0),
         nt_vb_dummy = ifelse(nt_vb_dummy == 1 & block == 1, 0, nt_vb_dummy))
```

```
mdl6 <- fehr %>%
  lm(formula = shifts ~ factor(fahrer) +
      treatment +
```

```

factor(nt_vb_dummy) +
factor(treat_dummy_1) +
factor(treat_dummy_2))

```

Doing the same using the within estimator from plm package for cleaner output:

```

mdl6_plm <- fehr %>%
  plm(formula = shifts ~ treatment +
        factor(nt_vb_dummy) +
        factor(treat_dummy_1) +
        factor(treat_dummy_2),
        index = "fahrer", model = "within")

summary(mdl6_plm)

```

```
## Oneway (individual) effect Within Model
```

```
##
```

```
## Call:
```

```
## plm(formula = shifts ~ treatment + factor(nt_vb_dummy) + factor(treat_dummy_1) +
```

```
##     factor(treat_dummy_2), data = ., model = "within", index = "fahrer")
```

```
##
```

```
## Unbalanced Panel: n = 138, T = 1-3, N = 386
```

```
##
```

```
## Residuals:
```

```
##      Min.      1st Qu.      Median      3rd Qu.      Max.
```

```
## -14.041867 -1.708534 -0.052489  1.904869  11.291466
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t-value Pr(>|t|)
```

```
## treatment              3.43517    1.24945   2.7493 0.006418 **
```

```
## factor(nt_vb_dummy)1  -0.62856    1.10841  -0.5671 0.571177
```

```
## factor(treat_dummy_1)1 -0.71660    0.81218  -0.8823 0.378475
```

```
## factor(treat_dummy_2)1 -2.21560    0.81218 -2.7280 0.006836 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    6136
## Residual Sum of Squares: 5487.1
## R-Squared:              0.10575
## Adj. R-Squared: -0.41101
## F-statistic: 7.21333 on 4 and 244 DF, p-value: 1.6643e-05
```

This model does not reflect the exact same factor loadings, but the same R squared. It is entirely possible that an error occurred on our part, but it might also be due to the data.

In a last step, we put everything together in two separate tables. The R^2 in the table are from the *plm* regressions, hence not the same as in the paper. However, we will show the latter below as well for completeness.

The original R^2 (Table 8 below):

```
r2 <- data.frame("Model" = c(1:6),
                 "R_sq" = c(summary mdl1)$r.squared,
                           summary mdl2)$r.squared,
                           summary mdl3)$r.squared,
                           summary mdl4)$r.squared,
                           summary mdl5)$r.squared,
                           summary mdl6)$r.squared))
```

1.5.2 b)

Why do the authors estimate three different models per outcome? What changes across models?

The goal of the analysis is to isolate the effect the wage increase has on the labor supply of the messengers. This can only be achieved if the treatment of one group does not affect the behavior of the untreated group. Otherwise, a new situation may have been created and the

Table 6: Regression Results (OLS) for Revenue

	<i>Dependent variable:</i>		
	totrev		
	(1)	(2)	(3)
treatment	1,033.56*** (336.70)	1,094.50*** (282.87)	1,035.80*** (344.05)
factor(treat_dummy_1)1	-210.97 (385.31)	-370.62 (264.58)	-264.76 (223.64)
factor(treat_dummy_2)1	-574.71 (378.23)	-656.23** (261.41)	-650.48*** (223.64)
factor(nt_vb_dummy)1			-54.40 (305.21)
Observations	124	190	386
R ²	0.12	0.12	0.10
Adjusted R ²	-0.37	-0.37	-0.42
F Statistic	3.70** (df = 3; 79)	5.74*** (df = 3; 121)	6.93*** (df = 4; 244)

Note:

*p<0.1; **p<0.05; ***p<0.01

R squared are deflated from plm method. See original R squared below.

Table 7: Regression Results (OLS) for Shifts

	<i>Dependent variable:</i>		
	shifts		
	(1)	(2)	(3)
treatment	3.99*** (1.14)	4.08*** (0.99)	3.44*** (1.25)
factor(nt_vb_dummy)1			-0.63 (1.11)
factor(treat_dummy_1)1	-1.28 (1.30)	-1.57* (0.93)	-0.72 (0.81)
factor(treat_dummy_2)1	-2.56** (1.28)	-2.63*** (0.92)	-2.22*** (0.81)
Observations	124	190	386
R ²	0.15	0.14	0.11
Adjusted R ²	-0.32	-0.34	-0.41
F Statistic	4.67*** (df = 3; 79)	6.60*** (df = 3; 121)	7.21*** (df = 4; 244)

Note:

*p<0.1; **p<0.05; ***p<0.01

R squared are deflated from plm method. See original R squared below.

Table 8: Original R squared from regular lm call

Model		R_sq
1	1	0.74
2	2	0.79
3	3	0.75
4	4	0.69
5	5	0.74
6	6	0.70

R squared were deflated from plm method.

treatment effect on the treated is altered by unaccounted influences, leading to a selection bias. Comparing the estimations where only the participation groups have been considered with the ones including the non-participants from Veloblitz and the messengers from Flash allows to control for such change in behavior. That is why the latter are called control groups. If their behavior is independent of the introduction of the treatment, a casual link between the wage increase and the labour supply can be assumed.

When implementing the control groups, the sample size increases leading to a increasingly randomized selection, preventing a selection bias.

1.5.3 c)

What do you conclude?

The results suggest that a wage increase causes an increase in the number of shifts taken and the total labor supply by individuals. Regardless if the non-participants of Veloblitz and/or of Flash are considered, the treatment effect stays roughly the same across the different models. It is further statistically significant on a $\alpha < 0.01$ level. Analogously, the dummy variable indicating the effect of being an employee of Flash in models 3 and 6 are small and likely statistically insignificant, as seen by the high p-value in regression 3 and 6.

As can be seen in Table 8, the R^2 is relatively high, suggesting that most of the variance in revenue can be explained by the model. It can thus be assumed that the most decisive variables affecting the outcome are considered in our model.

The consistency of the significant treatment effect while controlling for other effects in combination with the high R^2 allow us to assume a casual effect between wage and labour supply.

1.6 Question 4: “Discussion of the Results”

1.6.1 a)

Based on your preferred estimate, calculate the intertemporal elasticity of substitution with regard to labor supply.

We take the average revenue of group A and B during the two treatment periods as basis for the labor supply and the ~ CHF 1'000.00 from the treatment dummy as the treatment effect.

Comparing the resulting elasticities of the periods thereby yield the upper- and lower bound. We can calculate the intertemporal elasticity of substitution caused by a 25% wage increase as follows: Average Revenue Period 1: $\frac{1}{2}(4'131.33 + 3'005.75) = 3'568.54$, Average Revenue Period 2: $\frac{1}{2}(2'734.03 + 3'675.57) = 3'204.8$ Lower bound:

$$\frac{1}{25\%} \frac{1'000}{3'568.45} = 1.12$$

Upper bound:

$$\frac{1}{25\%} \frac{1'000}{3'204.8} = 1.25$$

1.6.2 b)

Do messengers increase their effort per hour, the number of hours worked or both? Motivate your answer.

The regressions 4-6 in Question 3 show a positive effect of the wage increase on the number of hours worked of roughly 4 shifts. The wage elasticity of shifts pictures the intensity of said effect which can be calculated analogously to the elasticity of substitution above.

We take the average number of shifts of group A and B again for both treatment periods, resulting in a upper and lower bound of the estimated elasticity.

- Average Shifts Period 1: $\frac{1}{2}(14 + 9.85) = 11.925$
- Average Shifts Period 2: $\frac{1}{2}(8.73 + 12.55) = 10.64$

The wage increase of 25% this leads to the following results:

Lower bound:

$$\frac{1}{25\%} \frac{4}{11.925} = 1.34$$

Upper bound:

$$\frac{1}{25\%} \frac{4}{10.64} = 1.50$$

The results indicate a wage elasticity of shifts between 1.34 and 1.5. It is clearly larger than the elasticity with regard to the total labor supply between 1.12 and 1.25. Since total labor supply is defined as the sum of the effect on shifts and the effect on effort per shift, the wage elasticity of the latter has to be negative in our case. This suggests that the effort per shift decreases in the wage increase. Since the positive effect on the number of shifts is larger however, a wage increase leads to a positive overall labor supply.

2 References

Fehr, E., & Goette, L. (2007). Do workers work more if wages are high? evidence from a randomized field experiment. *American Economic Review*, 97(1), 298–317.