Assignment2

483670 & 630516

September 11, 2022

Q1

The mean probability of having benefits in week 10 of those without search period is 73.6. For those with the search period treatment this is 57.2. The probability of still having benefits in week 30 is 54 for those without and 41.4 for those with search treatment. In the most naive estimation it seems that the search treatment decreases the probability of being on benefits both 10 and 30 weeks after the initial claim.

```
prob10nosearch <- mean(data$benefits_week10[data$searchperiod == 0])
prob10nosearch
```

```
## [1] 0.7359116
```

```
prob10search <- mean(data$benefits_week10[data$searchperiod == 1])
prob10search
```

```
## [1] 0.5723684
```

```
prob30nosearch <- mean(data$benefits_week30[data$searchperiod == 0])
prob30nosearch
```

```
## [1] 0.5403315
```

```
prob30search <- mean(data$benefits_week30[data$searchperiod == 1])
prob30search
```

```
## [1] 0.4144737
```

Q2

The balance table is provided is provided in table **??**. The potential points of concern for the balance between control and treatment groups are the differences in age, slight unbalance in the locations and the difference in shares of unknown education. Particularly those with unknown education could be problematic. As there is a significant difference in shares we drop all observations for which education is unknown. This also allows for more valid conclusions when we use education in the estimation. To do this however, we must assume that education levels are missing at random. As we have no other information on this we will make that assumption for now.

Table 1: Balance Table

variables1	Media_control1	Media_trat1	p_value1
age	39.9258850	37.2592105	0.0000000
benefits_week10	0.7359116	0.5723684	0.0000000
benefits_week30	0.5403315	0.4144737	0.0000003
children	0.1635359	0.1144737	0.0036995
educ_bachelormaster	0.2640884	0.2671053	0.8896809
educ_prepvocational	0.2176796	0.2000000	0.3763510
educ_primaryorless	0.1303867	0.1486842	0.2845937
educ_unknown	0.0143646	0.0500000	0.0000599
educ_vocational	0.3734807	0.3342105	0.0948282
female	0.3971239	0.3723684	0.3012313
location1	0.1767956	0.1131579	0.0002088
location2	0.1823204	0.2315789	0.0138131
location3	0.3734807	0.3000000	0.0015247
location4	0.1005525	0.2223684	0.0000000
location5	0.1668508	0.1328947	0.0522560
partner	0.1259669	0.1065789	0.2177481
period1	0.2640884	0.2223684	0.0475789
period2	0.2563536	0.2328947	0.2669553
period3	0.2651934	0.2855263	0.3556753
period4	0.2143646	0.2592105	0.0325346
sumincome_12monthsbefore	1.2961221	1.2590452	0.4845445
sumincome_24monthsbefore	2.7849836	2.6891123	0.3519164

```
balance <- balance_table(data, "searchperiod")
knitr::kable(balance, caption = "Balance Table")
```

```
data <- data[!(data$educ_unknown == 1), ]
```

Q3

Table 2 below provides the regression output for the effect of the search period on the probability of being on benefits after 10 and 30 weeks both without and with controls. This is the output after we have removed those observations with unknown education levels. The effect of the treatment on the 10 week mark decreases slightly in magnitude from the naive estimate. It remains significant though at 14.5 percentage points. The treatment effect increases in magnitude for being on benefits at the 30 week mark from a 10.3 percentage point to 11.8 percentage point decrease in the probability. In both cases including covariates maintains an economic and statistically significant treatment effect.

```
prob10nosearch <- mean(data$benefits_week10[data$searchperiod == 0])
prob10nosearch
```

[1] 0.7410314

```
prob10search <- mean(data$benefits_week10[data$searchperiod == 1])
prob10search
```

[1] 0.5900277

```
prob30nosearch <- mean(data$benefits_week30[data$searchperiod == 0])
prob30nosearch
```

[1] 0.544843

```
prob30search <- mean(data$benefits_week30[data$searchperiod == 1])
prob30search
```

[1] 0.4265928

```
m10b <- lm(benefits_week10 ~ searchperiod + sumincome_12monthsbefore +
    sumincome_24monthsbefore + age + female + children + partner +
    period2 + period3 + period4 + location2 + location3 + location4 +
    location5 + educ_bachelormaster + educ_prepvocational + educ_vocational,
    data = data)
m10a <- lm(benefits_week10 ~ searchperiod, data = data)
m30b <- lm(benefits_week30 ~ searchperiod, data = data)
m30a <- lm(benefits_week30 ~ searchperiod + sumincome_12monthsbefore +
    sumincome_24monthsbefore + age + female + children + partner +
    period2 + period3 + period4 + location2 + location3 + location4 +
    location5 + educ_bachelormaster + educ_prepvocational + educ_vocational,
    data = data)

stargazer(m10a, m10b, m30a, m30b, column.labels = c("", "Covariates",
    "", "Covariates"), type = "latex", title = "LPM for Benefits",
    header = FALSE, label = "tab:reg", keep = c("searchperiod"))
```

Table 2: LPM for Benefits

	Dependent variable:			
	benefits_week10		benefits_week30	
		Covariates		Covariate
	(1)	(2)	(3)	(4)
searchperiod	-0.151^{***}	-0.145^{***}	-0.103^{***}	-0.118^{**}
	(0.023)	(0.024)	(0.025)	(0.025)
Observations	1,614	1,612	1,612	1,614
R^2	0.026	0.046	0.055	0.014
Adjusted R^2	0.025	0.036	0.045	0.013
Residual Std. Error	0.463 (df = 1612)	0.460 (df = 1594)	0.489 (df = 1594)	0.497 (df = 1
F Statistic	42.411^{***} (df = 1; 1612)	4.544^{***} (df = 17; 1594)	5.447^{***} (df = 17; 1594)	22.609^{***} (df =

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

Q4

All the bounds are included in the overview table in tables **??** and **??** and they are graphically represented in the figures at the end of the document. Before starting it is important to note the direction of the treatment effect. We are looking for a negative treatment effect, meaning that when we refer to the lower bound (mathematically) we are referring to the most optimistic bound on the effect size. We will make this clear in each case. The effect sizes are also measured in decimal points and should be interpreted as percentage point changes to the probability of being on benefits at the 10 and 30 week mark respectively. The no assumption bounds for the effect sizes are $[-0.593, 0.407]$ at the 10 week mark and $[-0.558, 0.442]$ at the 30 week mark. This means that the most optimistic estimate for week 10 is a decrease in the probability of being on benefits of 59.3 percentage points and the least optimistic is an increase by 40.7 percentage points.

```
probSearch = mean(data$searchperiod)
ymin = 0
ymax = 1
```

```
# Benefits after 10 weeks - No assumption
Noass10_LB = prob10search * probSearch - prob10nosearch * (1 - probSearch) +
    (ymin + ymax) * (1 - probSearch) - ymax
Noass10_UB = prob10search * probSearch - prob10nosearch * (1 - probSearch) +
    (ymin + ymax) * (1 - probSearch) - ymin

# Benefits after 30 weeks - No assumption
Noass30_LB = prob30search * probSearch - prob30nosearch * (1 - probSearch) +
    (ymin + ymax) * (1 - probSearch) - ymax
Noass30_UB = prob30search * probSearch - prob30nosearch * (1 - probSearch) +
    (ymin + ymax) * (1 - probSearch) - ymin
```

Q5

Here we impose the restrictions of the Roy model. We assume that only those who got the treatment benefit from it. Alternatively, this means that those who didn't get the treatment would not benefit from it. For the 10 week mark the bounds become $[-0.41, 0.264]$. For the 30 week mark the bounds are $[-0.301, 0.191]$.

```
# Benefits after 10 weeks - Roy Model
Roy10_LB = -(prob10nosearch - ymin) * (1 - probSearch)
Roy10_UB = (prob10search - ymin) * probSearch

# Benefits after 30 weeks - Roy Model
Roy30_LB = -(prob30nosearch - ymin) * (1 - probSearch)
Roy30_UB = (prob30search - ymin) * probSearch
```

Q6

In MTR, treatment is assumed to be always improve the outcome. Therefore, we introduce the assumption that if an individual had a search period then this decreases the probability that they are going to be on benefits after 10 or 30 weeks after the treatment. This means that the probability that anyone is still on benefits after 10 or 30 weeks after treatment is always lower than if they had not received treatment. The MTR generally only affects the lower bound, as it pushes up the treatment effect (the worst estimate becomes better). However, in our case a better treatment effect is a more negative number. This means it works on our upper bound instead. This is why the upper bound becomes zero under MTR and the lower bound remains unchanged. The bounds under MRT for the 10 week mark the bounds become $[-0.593, 0.0]$. For the 30 week mark the bounds are $[-0.558, 0.0]$.

MTS normally affects the upper bound as it restricts the best estimate of the treatment, assuming that individuals assigned to the treatment have more favorable outcomes than individuals who were not assigned to the treatment. In our case though this is the lower bound. We introduce the assumption that individuals who got a search period have generally higher chances finding a job within 10 (30) weeks than individuals who were not assigned to treatment. The bounds under MTS for the 10 week mark the bounds become $[-0.151, 0.407]$. For the 30 week mark the bounds are $[-0.118, 0.442]$.

Combining both assumptions yields the MTS/MTR bounds of $[-0.151, 0.00]$ and $[-0.118, 0.0]$ for the 10 and 30 year mark respectively.

```
# Benefits after 10 weeks - MTS/MTR
MTS10_LB = prob10search - prob10nosearch
MTS10_UB = prob10search * probSearch - prob10nosearch * (1 - probSearch) +
    (ymin + ymax) * (1 - probSearch) - ymin

MTR10_LB = prob10search * probSearch - prob10nosearch * (1 - probSearch) +
    (ymin + ymax) * (1 - probSearch) - ymax
MTR10_UB = 0

MTSMTR10_LB = MTS10_LB
MTSMTR10_UB = MTR10_UB

# Benefits after 30 weeks - MTS/MTR
MTS30_LB = prob30search - prob30nosearch
MTS30_UB = prob30search * probSearch - prob30nosearch * (1 - probSearch) +
    (ymin + ymax) * (1 - probSearch) - ymin

MTR30_LB = prob30search * probSearch - prob30nosearch * (1 - probSearch) +
    (ymin + ymax) * (1 - probSearch) - ymax
MTR30_UB = 0

MTSMTR30_LB = MTS30_LB
MTSMTR30_UB = MTR30_UB
```

Q7

```
prob10search_bachelormaster = mean(data$benefits_week10[data$searchperiod ==
    1 & data$educ_bachelormaster == 1])
prob10nosearch_bachelormaster = mean(data$benefits_week10[data$searchperiod ==
    0 & data$educ_bachelormaster == 1])

prob10search_vocational = mean(data$benefits_week10[data$searchperiod ==
    1 & data$educ_vocational == 1])
prob10nosearch_vocational = mean(data$benefits_week10[data$searchperiod ==
    0 & data$educ_vocational == 1])

prob10search_prepvocational = mean(data$benefits_week10[data$searchperiod ==
    1 & data$educ_prepvocational == 1])
prob10nosearch_prepvocational = mean(data$benefits_week10[data$searchperiod ==
    0 & data$educ_prepvocational == 1])

# Goes against the MIV assumption!
```

```
prob10search_primaryorless = mean(data$benefits_week10[data$searchperiod ==
    1 & data$educ_primaryorless == 1])
prob10nosearch_primaryorless = mean(data$benefits_week10[data$searchperiod ==
    0 & data$educ_primaryorless == 1])

# Does not tell us anything, so ignore prob10search_unknown =
# mean(data$benefits_week10[data$searchperiod==1 &
# data$educ_unknown==1]) prob10nosearch_unknown =
# mean(data$benefits_week10[data$searchperiod==0 &
# data$educ_unknown==1])
```

Checking the data for the assumptions of the MIV reveals that Vocational, Prepvocational and Primaryorless go against the MIV assumption. The assumption is that those with more education have a lower probability of being unemployed. This seems to be violated slightly for the ordering of education below bachelor/master education. We thus reduce the MIV to whether a person has a bachelor or not, and check again whether the MIV assumption holds:

```
prob10search_NObachelormaster = mean(data$benefits_week10[data$searchperiod ==
    1 & data$educ_bachelormaster == 0])
prob10nosearch_NObachelormaster = mean(data$benefits_week10[data$searchperiod ==
    0 & data$educ_bachelormaster == 0])
```

Now, the MIV assumption holds as for both cases $d = \{0, 1\}$, the probability of persons with a bachelor or master degree being still unemployed after 10 weeks is lower than the probability of persons without a bachelor or master degree. Despite the fact that it is a theoretical assumption that doesn't necessarily have to hold in the data to make the estimate there is also a logical explanation for the trend in education observed in the data. Logically it makes sense that one would expect it is easier to find a job with a higher education level. Why the other eductation levels may not have this effect is because the differences between the education levels may not be high enough. The biggest difference in educational qualification probably exists for those with a bachelor or master degree. In contrast, the others may give people similarly marketable skills. There is thus a logical explanation for what me observe in the data.

We can apply the formulas to work out the LB and UB of those with and without the treatment. However, there is a step that needs to be taken as we are looking not for the UB and LB of the expected probabilities, but of the treatment effect. We thus need to use $\Delta Y = E[Y_1] - E[Y_0]$. To work out the UB (most conservative estimate) of ΔY we must subtract the upper bound of $E[Y_0]$ from the LB of $E[Y_1]$. For the LB (most optimistic estimate) of ΔY we must subtract the LB of $E[Y_0]$ from the UB of $E[Y_1]$. This gives us the MIV bounds for the treatment effect of $[-0.25, 0.062]$ for the 10 week and $[-0.284, 0.005]$ for the 30 week mark.

```
prob_BachelorMaster = mean(data$educ_bachelormaster)
```

```
# MIV for Benefits after 10 weeks

prob10nosearch_LB = (1 - prob_BachelorMaster) * mean(data$benefits_week10[data$searchperiod ==
    0 & data$educ_bachelormaster == 0]) + prob_BachelorMaster * max(mean(data$benefits_week10[data$searc
    0 & data$educ_bachelormaster == 0]), mean(data$benefits_week10[data$searchperiod ==
    0 & data$educ_bachelormaster == 1]))
prob10nosearch_UB = (1 - prob_BachelorMaster) * min(mean(data$benefits_week10[data$searchperiod ==
    0 & data$educ_bachelormaster == 0]), mean(data$benefits_week10[data$searchperiod ==
    0 & data$educ_bachelormaster == 1])) + prob_BachelorMaster * mean(data$benefits_week10[data$searchpe
    0 & data$educ_bachelormaster == 1])
```

Table 3: 10 Week Bounds

Model	Lowerbound	Upperbound
No Assumptions	-0.5929368	0.4070632
Roy Model	-0.4095415	0.2639405
MTS	-0.1510037	0.4070632
MTR	-0.5929368	0.0000000
MTS+MTR	-0.1510037	0.0000000
MIV	-0.2503187	-0.0615845

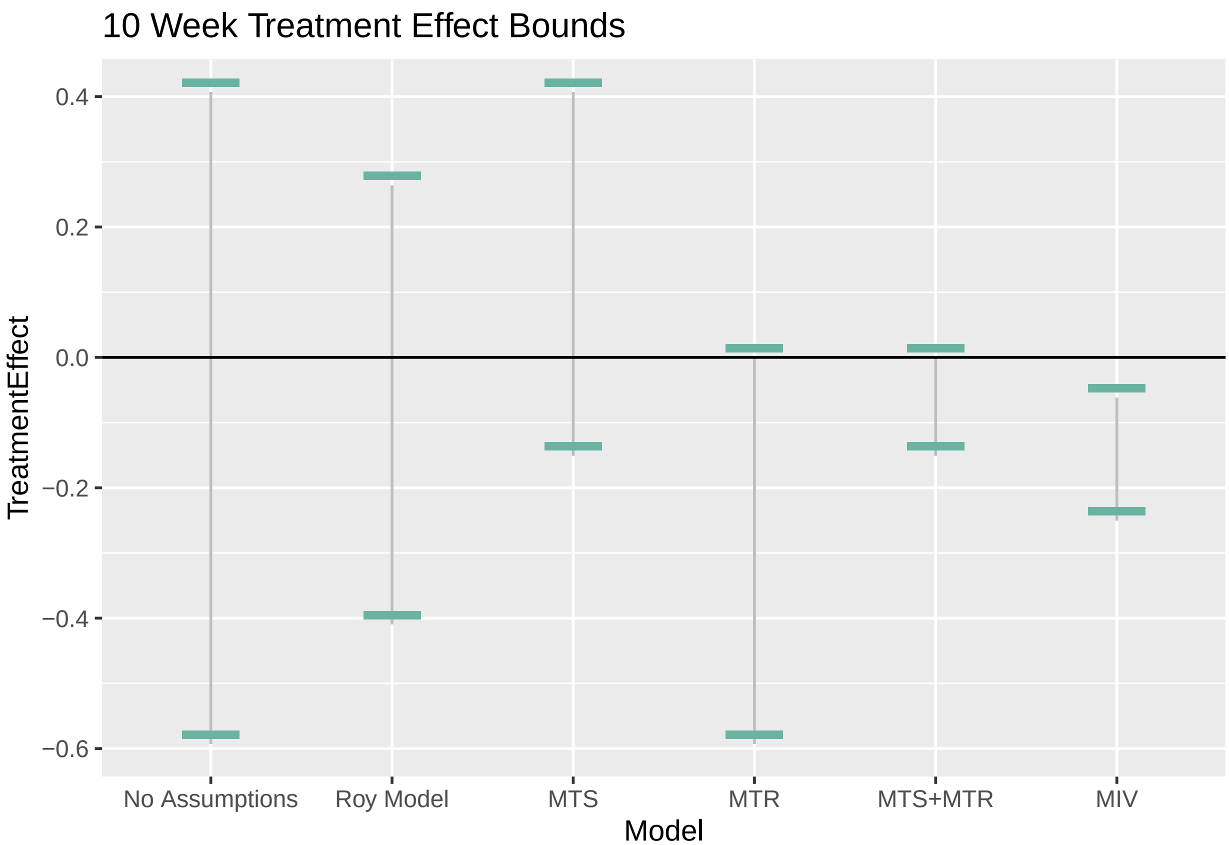
```
prob10search_LB = (1 - prob_BachelorMaster) * mean(data$benefits_week10[data$searchperiod ==
    1 & data$educ_bachelormaster == 0]) + prob_BachelorMaster * max(mean(data$benefits_week10[data$searc
    1 & data$educ_bachelormaster == 0]), mean(data$benefits_week10[data$searchperiod ==
    1 & data$educ_bachelormaster == 1]))
prob10search_UB = (1 - prob_BachelorMaster) * min(mean(data$benefits_week10[data$searchperiod ==
    1 & data$educ_bachelormaster == 0]), mean(data$benefits_week10[data$searchperiod ==
    1 & data$educ_bachelormaster == 1])) + prob_BachelorMaster * mean(data$benefits_week10[data$searchpe
    1 & data$educ_bachelormaster == 1])

MIV10_UB = prob10search_LB - prob10nosearch_UB

MIV10_LB = prob10search_UB - prob10nosearch_LB
```

```
# MIV for Benefits after 30 weeks

prob30nosearch_LB = (1 - prob_BachelorMaster) * mean(data$benefits_week30[data$searchperiod ==
    0 & data$educ_bachelormaster == 0]) + prob_BachelorMaster * max(mean(data$benefits_week30[data$searc
    0 & data$educ_bachelormaster == 0]), mean(data$benefits_week30[data$searchperiod ==
    0 & data$educ_bachelormaster == 1]))
prob30nosearch_UB = (1 - prob_BachelorMaster) * min(mean(data$benefits_week30[data$searchperiod ==
    0 & data$educ_bachelormaster == 0]), mean(data$benefits_week30[data$searchperiod ==
    0 & data$educ_bachelormaster == 1])) + prob_BachelorMaster * mean(data$benefits_week30[data$searchpe
    0 & data$educ_bachelormaster == 1])

prob30search_LB = (1 - prob_BachelorMaster) * mean(data$benefits_week30[data$searchperiod ==
    1 & data$educ_bachelormaster == 0]) + prob_BachelorMaster * max(mean(data$benefits_week30[data$searc
    1 & data$educ_bachelormaster == 0]), mean(data$benefits_week30[data$searchperiod ==
    1 & data$educ_bachelormaster == 1]))
prob30search_UB = (1 - prob_BachelorMaster) * min(mean(data$benefits_week30[data$searchperiod ==
    1 & data$educ_bachelormaster == 0]), mean(data$benefits_week30[data$searchperiod ==
    1 & data$educ_bachelormaster == 1])) + prob_BachelorMaster * mean(data$benefits_week30[data$searchpe
    1 & data$educ_bachelormaster == 1])

MIV30_UB = prob30search_LB - prob30nosearch_UB

MIV30_LB = prob30search_UB - prob30nosearch_LB
```

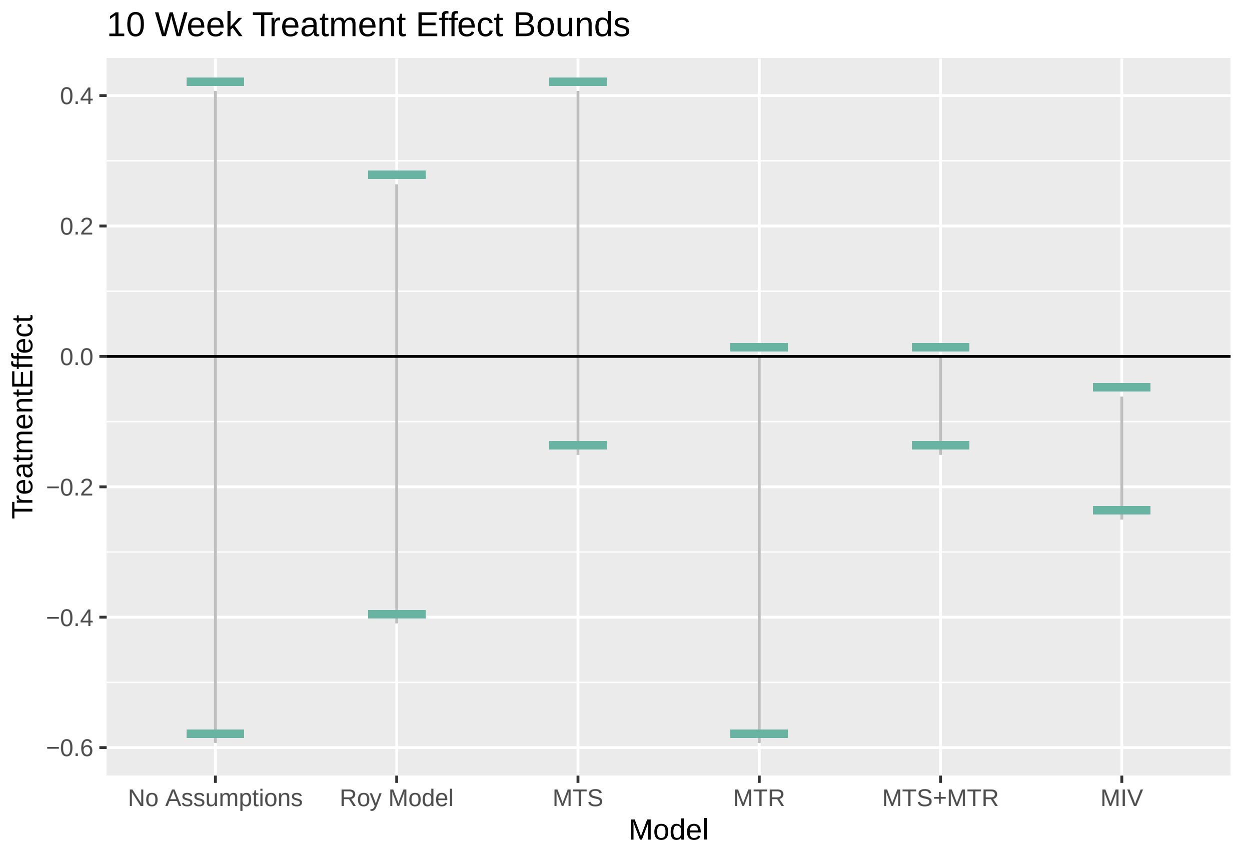
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Table 4: 30 Week Bounds

Model	Lowerbound	Upperbound
No Assumptions	-0.5576208	0.4423792
Roy Model	-0.3011152	0.1908302
MTS	-0.1182503	0.4423792
MTR	-0.5576208	0.0000000
MTS+MTR	-0.1182503	0.0000000
MIV	-0.2839641	0.0047081




```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```


10 Week Treatment Effect Bounds
TreatmentEffect
0.4
0.2
0.0
−0.2
−0.4
−0.6
No Assumptions
Roy Model
MTS
MTR
MTS+MTR
MIV
Model