Homework 3

Bilal Shakir

2018-01-24

I start off by loading in the required packages in the following code chunk.

```
## clearing out the global environment is a good idea before starting out

remove(list=ls())

## loading in the required datasets

pacman::p_load(tidyr, tidyverse,
               ggplot2, ggthemes,
               dplyr, RColorBrewer,
               gridExtra,
               texreg, readr,
               foreign, readstata13,
               interplot, stargazer,
               Zelig, broom,
               car, purrr,
               MASS, arm,
               pander, knitr,
               psych, cowplot,
               margins, scales,
               pastecs, plm, survival,
               ggplot2, ggrepel, pscl,
               GGally, foreign,  RTextTools,
               feather,
               lfe, dplyr, corrplot, haven)
```

Note that this assignment is based on the following paper:

Baccini, Li, and Mirkina. (2014). "Corporate Tax Cuts and Foreign Direct Investment." *Journal of Policy Analysis and Management*, vol. 32(4): 977-1006.

In the following code chunk, I load the dataset.

perform basic descriptives

Now I show some basic descriptive statistics. A glimpse of the dataset reveals that the number of variables is too large for me to describe each variable individually. As such, I only show the descriptive statistics of the variables that I use in my baseline model.

```
## glimpse the data set
glimpse(russia_data)
```

```
## Observations: 1,148
## Variables: 27
## $ id            <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, ...
## $ region        <chr> "adygea_rep", "adygea_rep", "adygea_rep", "adyge...
## $ year          <dbl> 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, ...
## $ name          <chr> "Republic of Adygea", "Republic of Adygea", "Rep...
```

```
## $ city          <dbl> 53.3, 53.8, 53.9, 53.9, 53.9, 52.8, 52.7, 52.5, ...
## $ empl_high     <dbl> 12.9, 13.3, 13.9, 19.4, 18.4, 19.0, 23.3, 20.1, ...
## $ grp           <dbl> 430.947, 508.501, 522.303, 365.162, 214.285, 196...
## $ realgrpgrowth <dbl> NA, NA, -9.1, -3.5, 5.4, 5.6, 0.7, 1.2, 3.9, 11....
## $ invrisk       <dbl> NA, 45, 50, 74, 51, 52, 40, 34, 47, 59, 45, 50, ...
## $ fdiln         <dbl> NA, 3.983413, 3.325036, 6.530149, 6.887349, 6.57...
## $ grppc_ln      <dbl> 6.863515, 7.029562, 7.056791, 6.698316, 6.167045...
## $ trade_ln      <dbl> NA, NA, 3.005683, 2.740840, 2.617396, 2.433613, ...
## $ universal     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ targeted      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ universal_cut <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ targeted_cut  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ deficit       <dbl> 0.00000, 0.00000, 0.00000, 0.00000, 0.00000, -27...
## $ crimeshareln  <dbl> 7.203405, 7.319865, 7.226209, 7.421776, 7.677864...
## $ roaddensityln <dbl> 5.267858, 5.278115, 5.298317, 5.298317, 5.313206...
## $ corruption    <dbl> 0.2376847, 0.2376847, 0.2376847, 0.2376847, 0.23...
## $ period        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, ...
## $ grp_ln        <dbl> 6.065985, 6.231467, 6.258248, 5.900341, 5.367307...
## $ fdi           <dbl> NA, 52.7000, 26.8000, 684.5000, 978.8002, 719.00...
## $ grppc         <dbl> 956.7239, 1129.5353, 1160.7142, 811.0389, 476.77...
## $ trade         <dbl> NA, NA, 19.200003, 14.499999, 12.700001, 10.4000...
## $ crimeshare    <dbl> 1343.9998, 1509.9998, 1375.0002, 1672.0000, 2160...
## $ roaddensity   <dbl> 194.0000, 196.0000, 200.0000, 200.0000, 203.0000...
```

```
## as the dataset has 27 different variables. It is too large to show the descriptives statistics for th

## descriptive for real GRP growth
realgrpgrowth_d <- stat.desc(russia_data$realgrpgrowth)

## descriptive for Gross regional Product (GRP)
grp_ln_d <- stat.desc(russia_data$grp_ln)

## descriptive for trade (log)
trade_ln_d <- stat.desc(russia_data$trade_ln)

## descriptives for roaddensity ln
roaddensityln_d <- stat.desc(russia_data$roaddensityln)

## corruption
corruption_d <- stat.desc(russia_data$corruption)

## invrisk
invrisk_d <- stat.desc(russia_data$invrisk)

descriptives <- data.frame(realgrpgrowth_d,grp_ln_d,
                    trade_ln_d, roaddensityln_d,
                    corruption_d, invrisk_d)

## Setnames
descriptives_names <- setNames(descriptives,
                             c("Real GRP growth",
                               "Gross Regional Product (GRP)",
                               "Trade_ln", "Road density",
                               "Corruption", "Investment Risk")); print(descriptives_names)
```

```
##                Real GRP growth Gross Regional Product (GRP)     Trade_ln
## nbr.val            967.0000000                 1.128000e+03 9.540000e+02
## nbr.null             7.0000000                 0.000000e+00 0.000000e+00
## nbr.na             181.0000000                 2.000000e+01 1.940000e+02
## min                -22.9000000                 4.444449e+00 9.531018e-02
## max                 78.7000000                 1.248834e+01 1.236825e+01
## range              101.6000000                 8.043894e+00 1.227294e+01
## sum               4732.2000000                 8.842756e+03 6.132707e+03
## median               5.3000000                 7.831814e+00 6.507648e+00
## mean                 4.8936918                 7.839323e+00 6.428414e+00
## SE.mean              0.2260439                 3.616141e-02 5.683694e-02
## CI.mean.0.95         0.4435938                 7.095126e-02 1.115400e-01
## var                 49.4097014                 1.475027e+00 3.081838e+00
## std.dev              7.0292035                 1.214507e+00 1.755516e+00
## coef.var             1.4363805                 1.549249e-01 2.730870e-01
##                Road density   Corruption Investment Risk
## nbr.val        1.148000e+03 1.140000e+03    1.066000e+03
## nbr.null       1.400000e+01 0.000000e+00    0.000000e+00
## nbr.na         0.000000e+00 8.000000e+00    8.200000e+01
## min            0.000000e+00 1.673717e-01    1.000000e+00
## max            6.500000e+00 5.768780e-01    8.200000e+01
## range          6.500000e+00 4.095063e-01    8.100000e+01
## sum            4.926118e+03 4.523587e+02    4.423900e+04
## median         4.815835e+00 4.000545e-01    4.150000e+01
## mean           4.291043e+00 3.968058e-01    4.150000e+01
## SE.mean        4.063563e-02 2.300597e-03    7.252974e-01
## CI.mean.0.95   7.972851e-02 4.513885e-03    1.423174e+00
## var            1.895640e+00 6.033733e-03    5.607761e+02
## std.dev        1.376823e+00 7.767711e-02    2.368071e+01
## coef.var       3.208596e-01 1.957560e-01    5.706195e-01
```

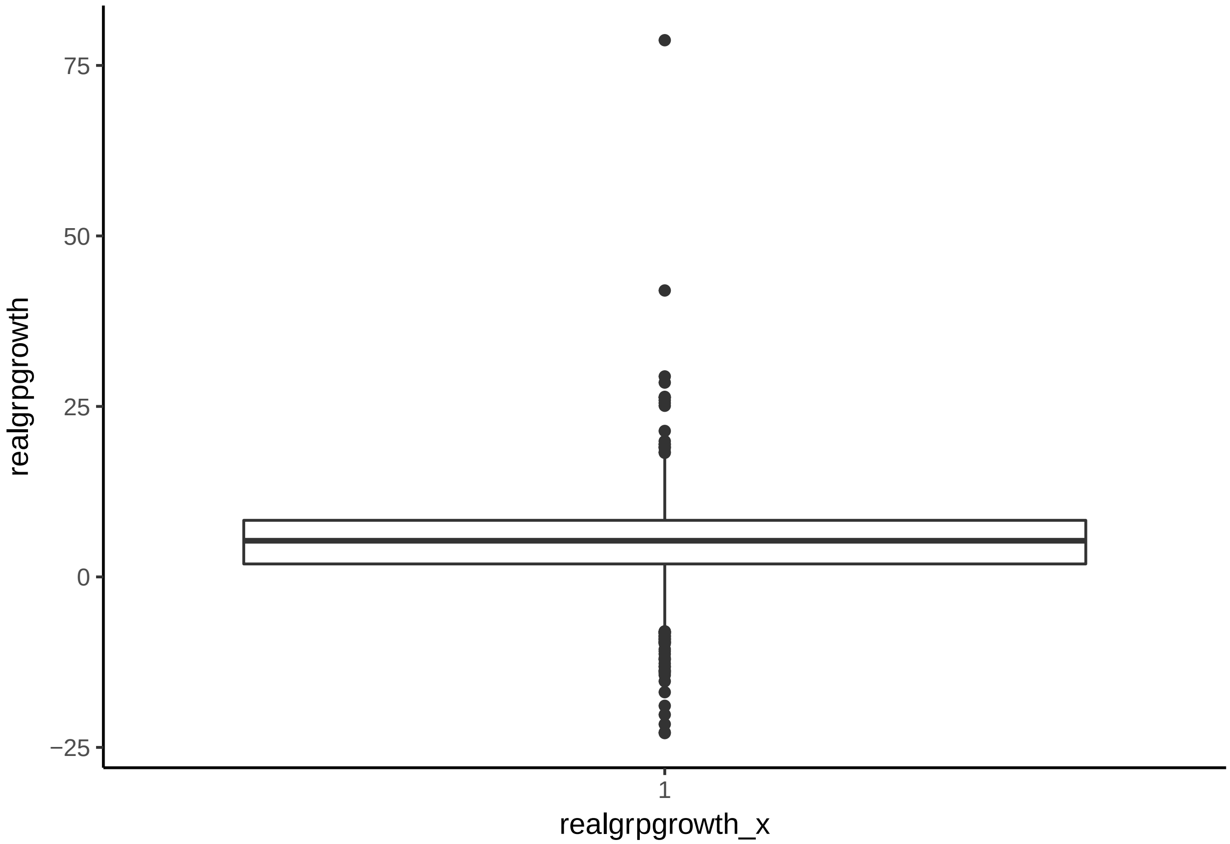
By running the basic descriptive statistics above I find that there are quite a few missing values in the data. This is especially true for variables like our DV *realgrpgrowth* and other variables like *invrisk* and *trade_ln* . Since there are missing values in the data I run a pairwise correlation as well to check for a more precise correlation between paired variables.

let's make those boxplots!

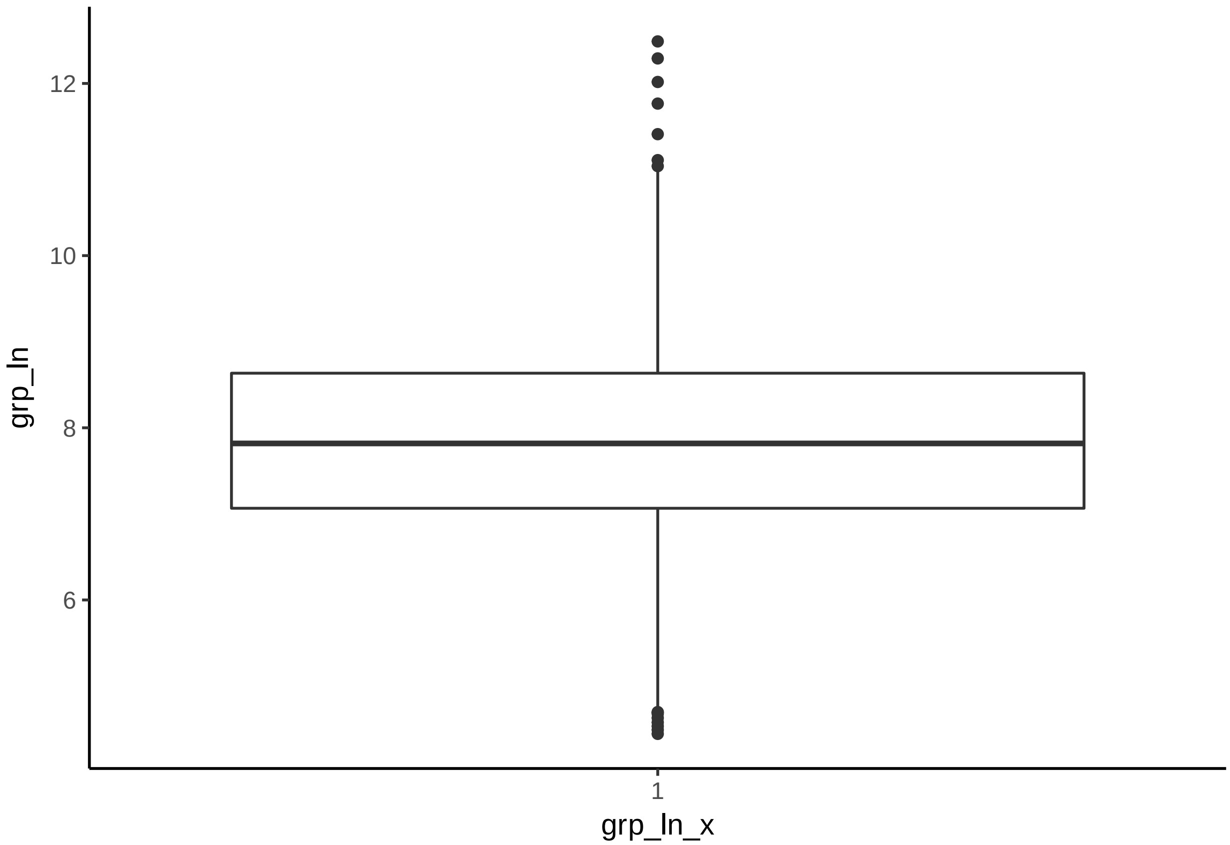
In the following code chunk I prepare to show boxplots for my variables.

```r
## Creating new categorical variables for the x axis

russia_data$grp_ln_x <- rep(1, each = 1148) %>% as.factor()
russia_data$realgrpgrowth_x <-  rep(1, each = 1148) %>% as.factor()
russia_data$trade_ln_x <- rep(2, each = 1148) %>% as.factor()
russia_data$roaddensityln_x <- rep(3, each = 1148)  %>% as.factor()
russia_data$corruption_x <- rep(4, each = 1148) %>% as.factor()
russia_data$invrisk_x <- rep(5, each = 1148)  %>% as.factor()
russia_data$realgrpgrowth_x <-  rep(1, each = 1148) %>% as.factor()

## Boxplot of realgrpgrowh

realgrpgrowth_boxplot <- ggplot(subset(russia_data,
                                !is.na(realgrpgrowth)),
                          aes(x = realgrpgrowth_x,
```

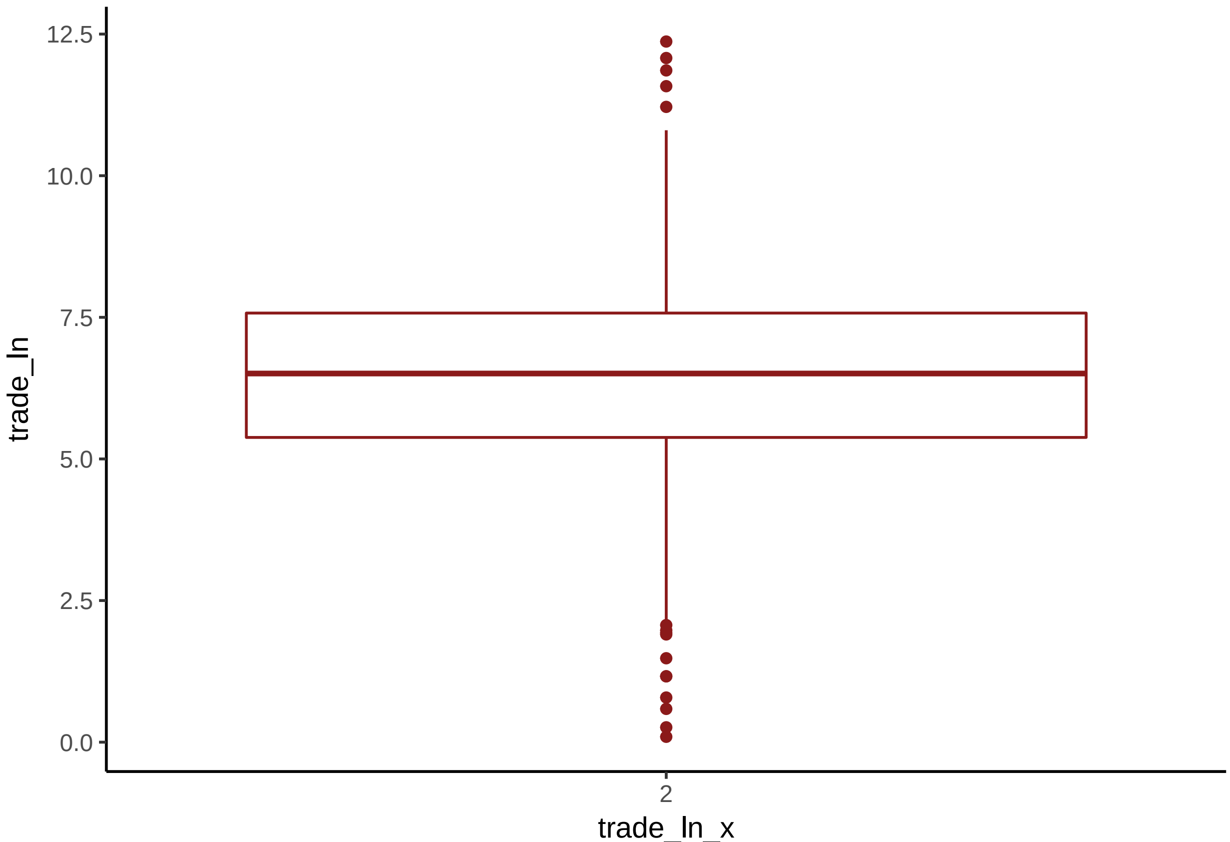
```
                             y = realgrpgrowth)) +
  geom_boxplot() +
  theme_classic(); realgrpgrowth_boxplot
```



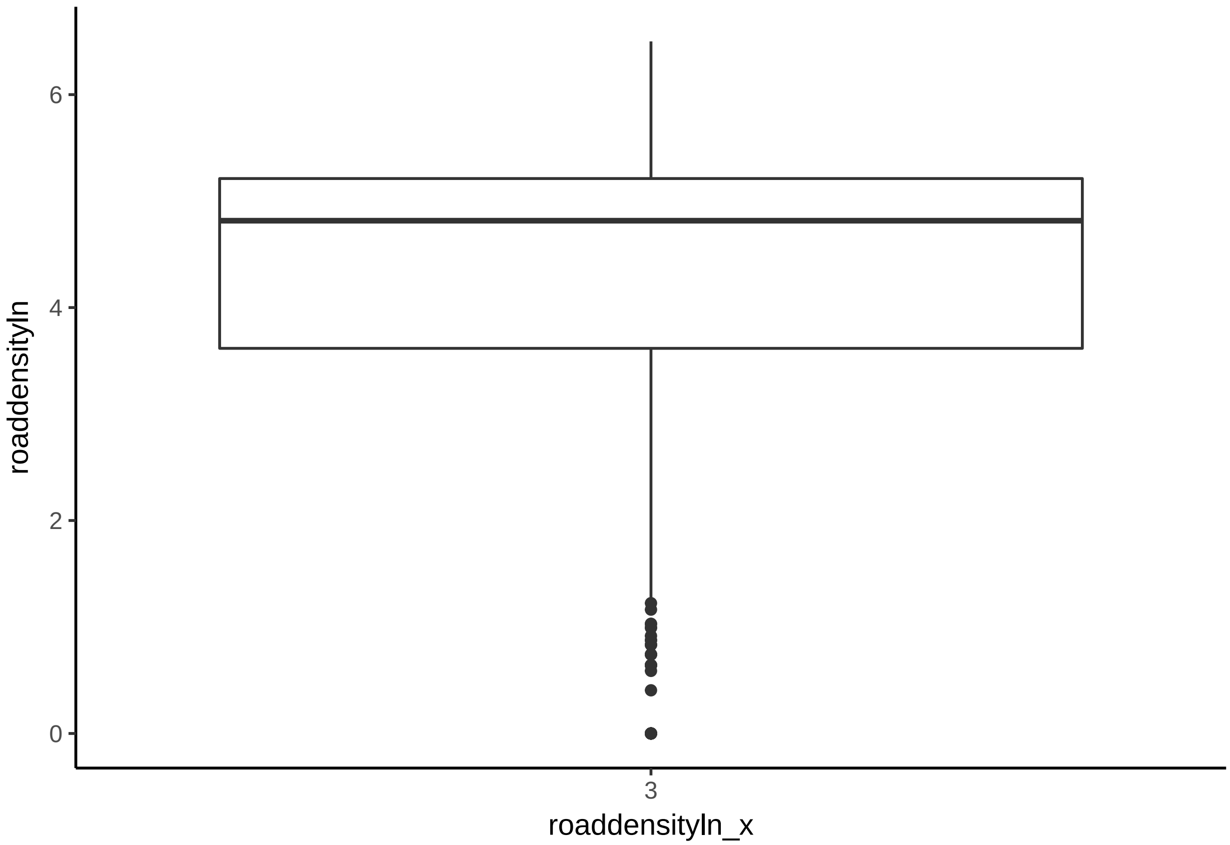

```
## Boxplot of grp_ln

grp_ln_boxplot <- ggplot(subset(russia_data,
                                 !is.na(realgrpgrowth)),
                        aes(x = grp_ln_x,
                            y = grp_ln)) +
  geom_boxplot() +
  theme_classic(); grp_ln_boxplot
```



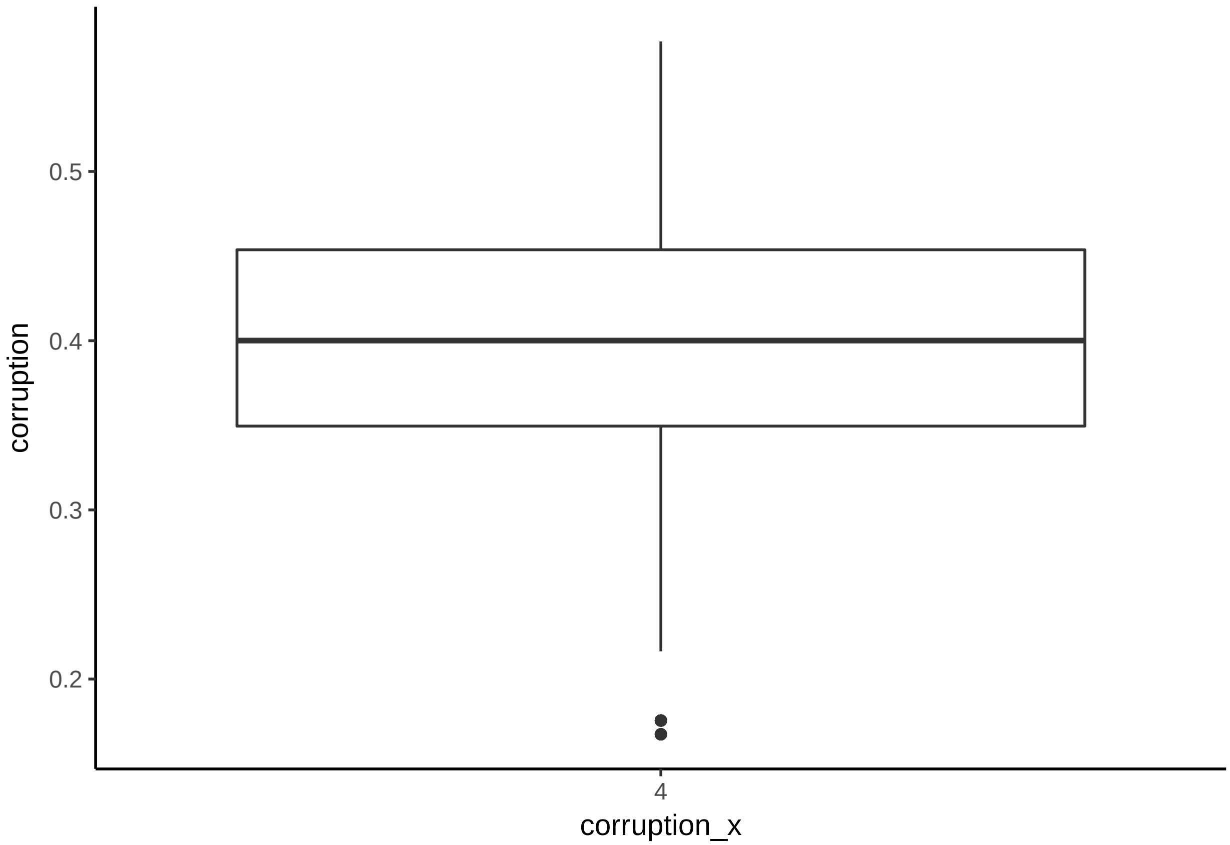

```
## Boxplot of trade_ln_x

trade_ln_boxplot <- ggplot(subset(russia_data,
                                  !is.na(trade_ln)),
                      aes(x =  trade_ln_x,
                          y =  trade_ln)) +
  geom_boxplot(color = "firebrick4") +
  theme_classic(); trade_ln_boxplot
```



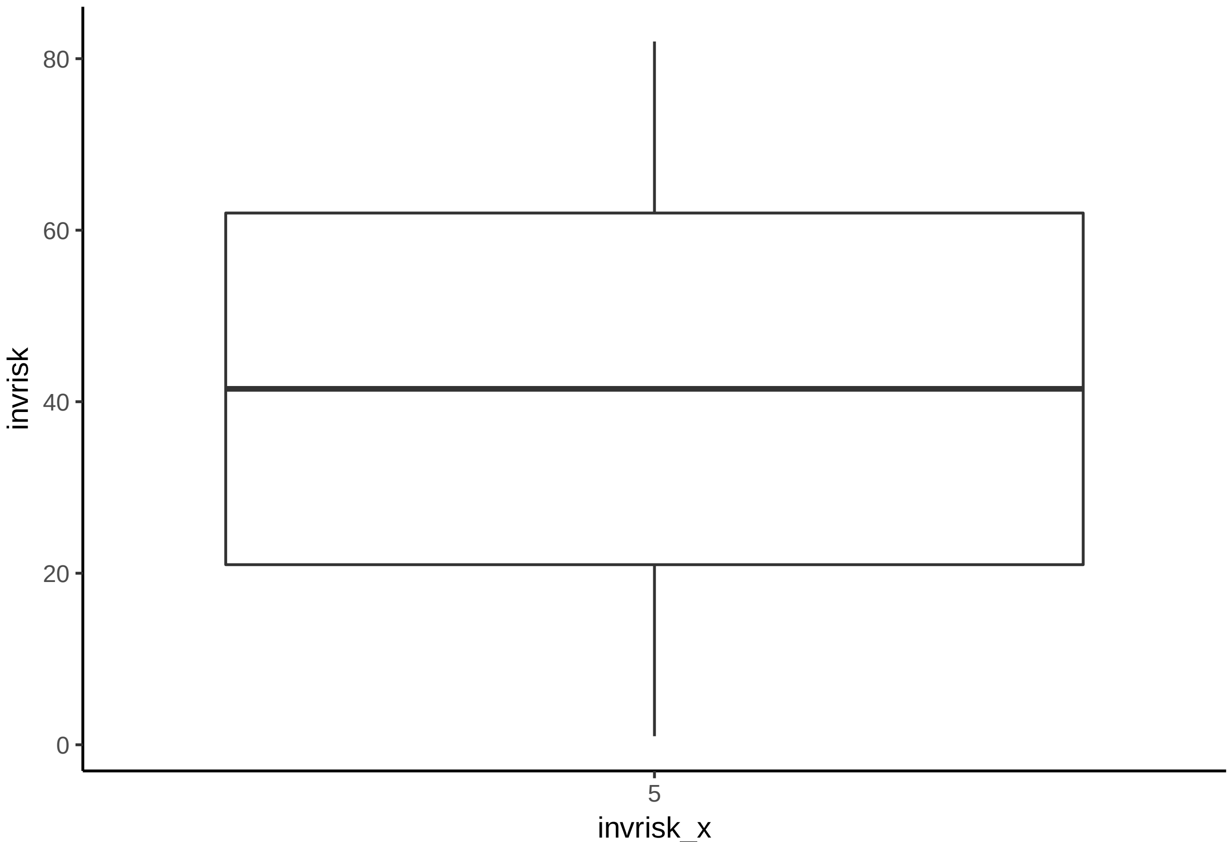

```
## Boxplot of roaddensityln_x

roadddensity_ln_boxplot <- ggplot(subset(russia_data,
                                         !is.na(roaddensityln)),
                                  aes(x = roaddensityln_x,
                         y =  roaddensityln)) +
  geom_boxplot() +
  theme_classic(); roadddensity_ln_boxplot
```



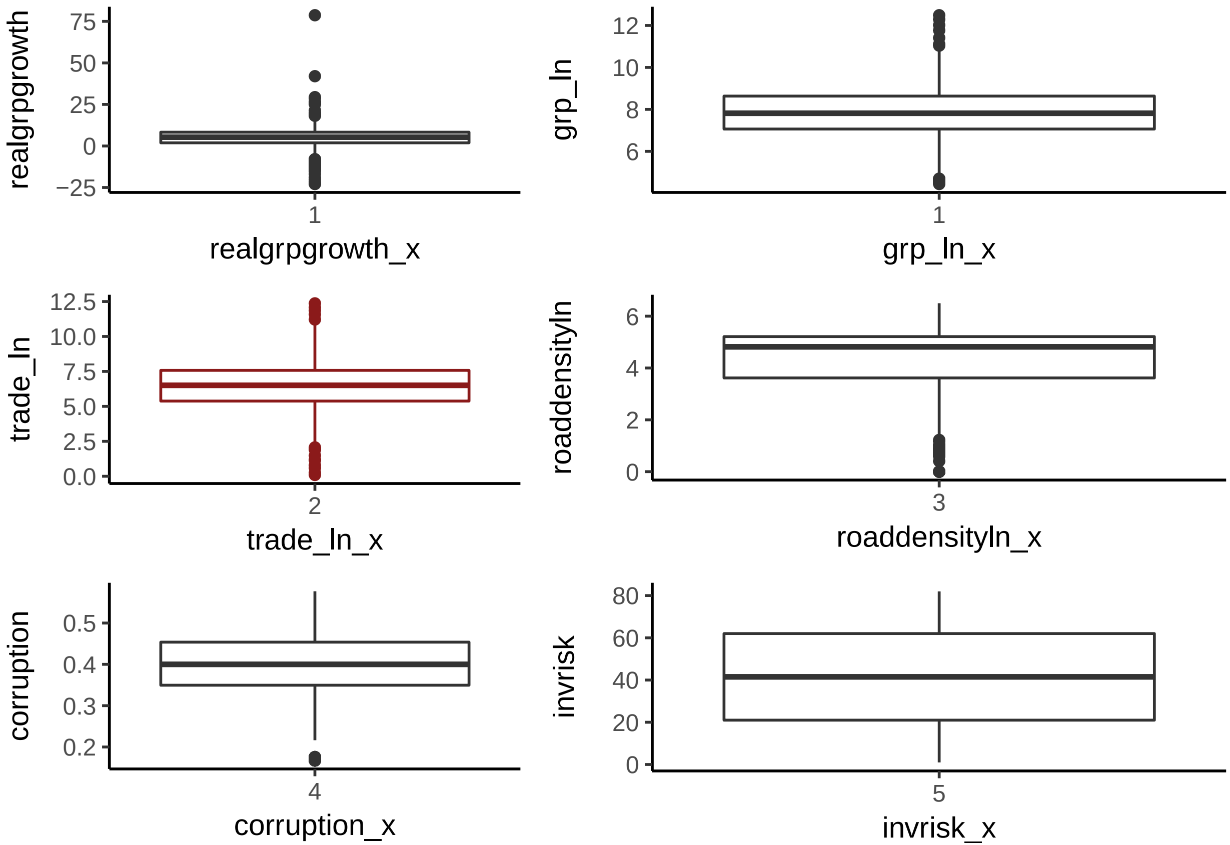

```
## Boxplot of corruption_x

corruption_boxplot <- ggplot(subset(russia_data,
                                    !is.na( corruption)),
                      aes(x = corruption_x,
                          y =  corruption)) +
  geom_boxplot() +
  theme_classic(); corruption_boxplot
```




```
## Boxplot of invrisk_x

invrisk_boxplot <- ggplot(subset(russia_data,
                                 !is.na(invrisk)),
                          aes(x = invrisk_x,
                              y = invrisk)) +
  geom_boxplot() +
  theme_classic(); invrisk_boxplot
```



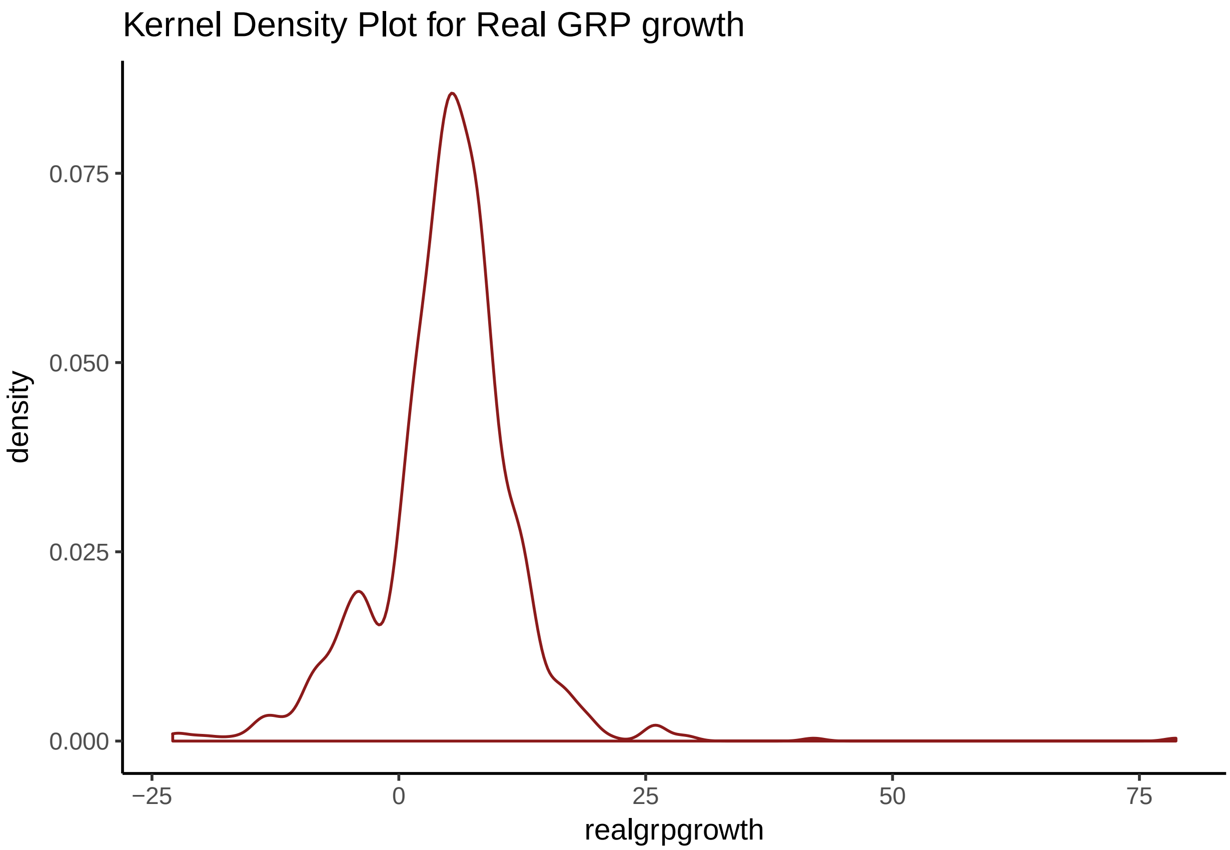

```
box_cowplot <- plot_grid(realgrpgrowth_boxplot,
          grp_ln_boxplot,
          trade_ln_boxplot,
          roadddensity_ln_boxplot,
          corruption_boxplot,
          invrisk_boxplot,
          align = "v",
          ncol = 2,
          label_fontfamily = "serif",
          label_y = 1,
          rel_widths = c(1, 1.3),
          label_size = 9); box_cowplot
```



From the figure with the boxplots of all the variables we can observe that

Density plot for realgrpgrowth

```
density_realgrpgrpwth <- ggplot(subset(russia_data,
              !is.na(realgrpgrowth)),
       aes(realgrpgrowth)) +
  geom_density(colour = "firebrick4") +
  ggtitle("Kernel Density Plot for Real GRP growth") +
  theme_classic(); density_realgrpgrpwth
```

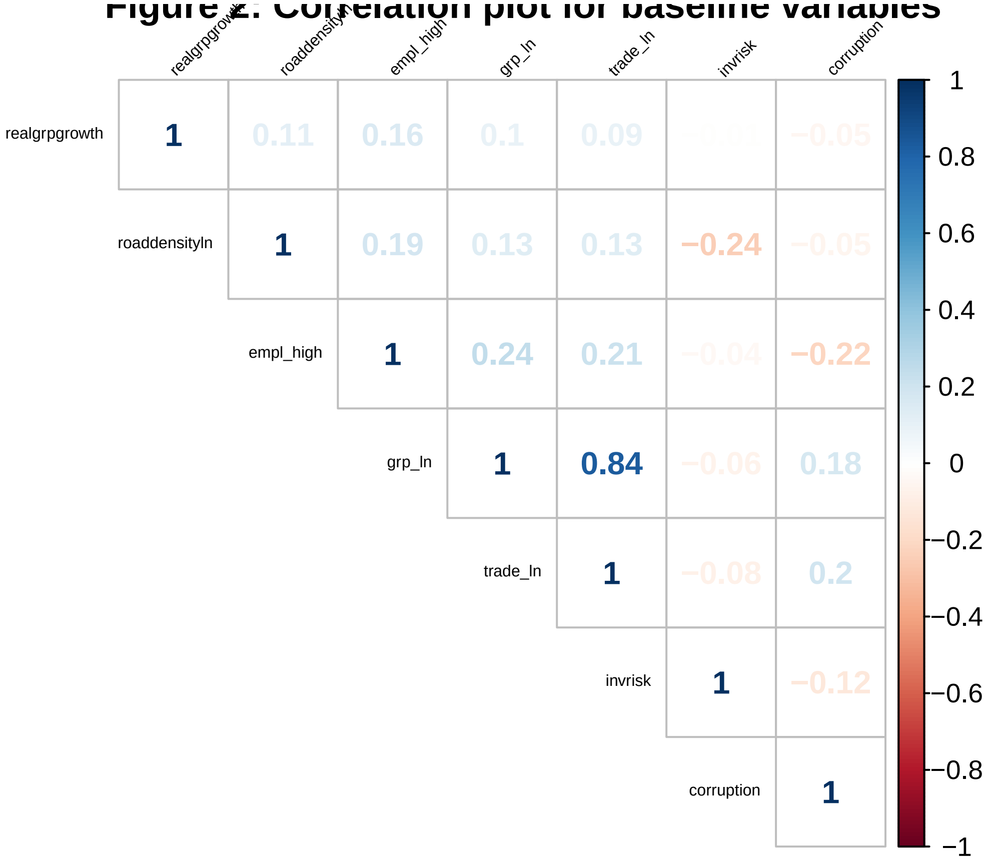


From the kernel density plot above we can see that by and large, *realgrpgrowth* is relatively evenly distributed. Therefore, there is no need for us to to log transform our dependent variable for further analysis.

correlation plot

```
russia_data %>% dplyr::select(realgrpgrowth,
                              invrisk, grp_ln, trade_ln,
                              corruption, roaddensityln, empl_high) %>%
  cor(use = "complete.obs") %>% corrplot(type = "upper",
                                         order = "hclust",
                                         tl.col = "black",
                                         tl.srt = 45,
                                         tl.cex = 0.5,
                                         method = "number",
                                         title = "Figure 2: Correlation plot for baseline variables")
```

Figure 2: Correlation plot for baseline variables



The correlation plot above showcases that from the selected variables there is only a high degree of positive correlation between *trade_ln* and *grp_ln*. The rest of the choosen explanatory variables seem to be allright.However, I include both of these variables in my models in any case as both the inclusion of both these variables is standard practice as underlined by Baccini et al (2014). Please note that I provide further description of the distribution of my data while analyzing the modelling choices for Task 1.2 below.

Task 1.2

Implement a standard D-I-D estimation, using a simple OLS and the interaction between targeted and period as well as universal and period and including the relevant controls. Discuss the results

Solution

This question requires us to run an OLS regression to capture the interaction effect between the "universal" and "targeted" treatment with "period". These interaction variables that score TRUE for both these conditions are called *universal_cut* and *targeted_cut* respectively. Recall that period is a dummy variable for the post-2003 period on the dependent variable real GRP growth rate. In other words, the period dummy signifies the year in which the treatment potentially came into effect. In addition, I add some additional control variables after getting an intuition about whould should be included in estimating realgrpgrowth after reading Baccini et al (2014) . Some of the controls that I include are investment risk, road density and corruption as they are proxies for institutions and infrastructure within different regions of Russia. Note that both trade and grp are always included in a growth model (Baccini et al 2014).

All the control variables on the RHS of my model have been lagged by 1 year. I've also included the corruption index for the different regions of Russia at the time and the share of people in employment with higher education as controls. The presence of high corruption typically impedes real grp growth and this might be true despite a tax cut. For example, a corporate tax cut in a region with a high corruption index is not likely to translate into a high real grp growth rate. One explanation for this is that despite the tax cut potential investors are not convinced to invest in a region with a high corruption index as they do not view low taxes as a credible commitment that will lead to increased profitability. We have some weak support for this hypothesis from the preliminary descriptive statistics that we have carried out such as the correlation plot which shows that there is a negative albeit very weak correlation between real grp growth and corruption.

Whereas, a greater share of people with higher education skills in a region can be indicative of a mobile and adaptable workforce. A more educated/highly skilled labor force is more likely to yield a higher real GRP growth rate than unskilled labor force.

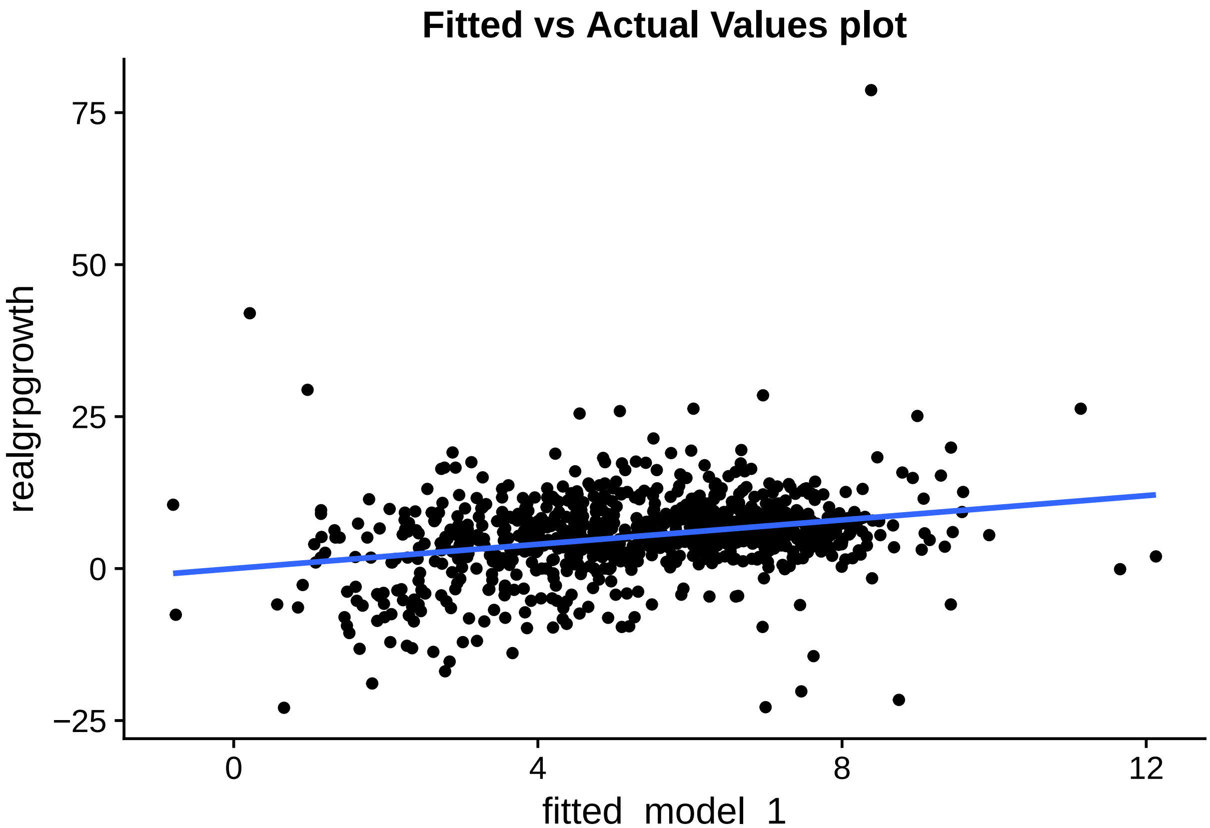
```
## Generating lags

  russia_data <- read_dta("Dataset_Russia_Stata.dta") %>%
    arrange(region) %>%
    group_by(region) %>%
    mutate(
      grp_ln_lag = lag(grp_ln),
      grpgrowth_lag = lag(realgrpgrowth),
      trade_ln_lag = lag(trade_ln),
      roaddensityln_lag = lag(roaddensityln),
      corruption_lag = lag(corruption),
      invrisk_lag = lag(invrisk),
      empl_high_lag = lag(empl_high))

model_1 <- lm(realgrpgrowth ~ invrisk_lag +
                grp_ln_lag + trade_ln_lag +
                roaddensityln_lag + corruption_lag +
                universal + period + universal_cut +
                targeted + targeted_cut + empl_high_lag,
              na.action = na.exclude,
              data = russia_data)

screenreg(list(model_1))
```

```
##
## =============================
##                    Model 1
## -----------------------------
## (Intercept)          6.26 **
##                     (2.40)
## invrisk_lag          0.01
##                     (0.01)
## grp_ln_lag          -1.55 ***
##                     (0.36)
## trade_ln_lag         0.88 ***
##                     (0.25)
## roaddensityln_lag    0.70 ***
##                     (0.20)
## corruption_lag      -1.33
##                     (3.15)
## universal            1.17
##                     (1.22)
## period               3.71 ***
```

```
##                       (0.62)
## universal_cut         -2.49
##                       (1.65)
## targeted               0.88
##                       (0.75)
## targeted_cut          -1.14
##                       (1.01)
## empl_high_lag          0.02
##                       (0.05)
## -----------------------------
## R^2                    0.08
## Adj. R^2               0.06
## Num. obs.            870
## RMSE                   6.71
## =============================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

```
# ORANGE - UNIVERSAL
# YELLOW - TARGETED

russia_data$fitted_model_1 <- predict(model_1)
russia_data$resid_model_1 <- resid(model_1)

## plotted

ggplot(subset(russia_data,
              !is.na(realgrpgrowth)),
       aes(fitted_model_1,
           realgrpgrowth)) +
  geom_point() +
  geom_smooth(method = "lm",
              se = FALSE) +
  ggtitle("Fitted vs Actual Values plot")
```

```
## Warning: Removed 97 rows containing non-finite values (stat_smooth).

## Warning: Removed 97 rows containing missing values (geom_point).
```



The regression results above show a negative relationship between the variable "universal_cut" and the dependent variable, real GRP growth. Whereas, "targeted_cut" is a variable that captures the interaction between the "targeted" treatment and the dummy variable period signifying cases of observations in the post-2003 period. Consequently, cases coded as "1" in this variable are regions which implemented discriminatory tax-cuts that came into force in/after 2003. Regions of Russia that implement discriminatory tax cuts are coded as "0" if the discriminatory tax intervention occurred before 2003. Moreover, "0" is coded along the cases in the dataset which had not implemented discriminatory tax cuts at all.

The key point that we take away is a lack of the statistical significance of both universal_cut and targeted_cut on our DV. This is applicable for our other explanatory variables (universal and targeted as well).

One result that is particular important to highlight is the positive relationship between "Period" and real GRP growth. The significance of the post-2003 dummy shows a positive relationship between the year 2003 as a threshold with respect to real GRP growth among regions of Russia in the data. This is indicative of the significance of allowing autonomous taxation to regional government and an increase in positive real GRP growth. Moreover, other results that I believe are pertinent to discuss are the statistically significant positive relationships between the real GRP growth and "trade_ln_lag" and "roaddensityln_lag." Whereas, "grp_ln_lag" is negatively correlated with the dependent variable. The fitted vs actual values plot in the code chunk above shows that while our model isn't perfect, however, at the same time it is not completely useless either.

Task 1.3

3. include region and year fixed effects and re-estimate the model in point 2 using OLS regression. Discuss the results;

Solution 1.3

Now, I include year and region fixed effects to the baseline formula that I used in Task 1.2. The interaction variables have region and fixed effects added to them between the treatment and period. I retain the control variables from my previous regression (Solution 1.2).

```
## OLS with year and region fixed effects -- incorrect specification

model_2 <- felm(realgrpgrowth ~ invrisk_lag + grp_ln_lag + trade_ln_lag +
                roaddensityln_lag + corruption_lag + universal_cut +
                  targeted_cut + empl_high_lag | region + year,
                data = russia_data)

screenreg(list(model_2))
```

```
##
## =================================
##                        Model 1
## ---------------------------------
## invrisk_lag              0.01
##                         (0.01)
## grp_ln_lag              -5.85 ***
##                         (1.31)
## trade_ln_lag            -0.09
##                         (0.38)
## roaddensityln_lag        1.77
##                         (1.72)
## corruption_lag          -3.09
##                         (3.99)
## universal_cut           -2.71 *
##                         (1.34)
## targeted_cut            -1.24
##                         (0.82)
## empl_high_lag           -0.18 *
##                         (0.08)
## ---------------------------------
## Num. obs.              870
## R^2 (full model)         0.45
## R^2 (proj model)         0.04
## Adj. R^2 (full model)    0.38
## Adj. R^2 (proj model)   -0.08
## =================================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

```
model_3 <- felm(realgrpgrowth ~ invrisk_lag + grp_ln_lag + trade_ln_lag +
                roaddensityln_lag + corruption_lag + universal_cut +
                  targeted_cut + empl_high_lag + grpgrowth_lag | region + year,
                data = russia_data)

screenreg(list(model_3))
```

```
##
## =================================
##                        Model 1
## ---------------------------------
```

```
## invrisk_lag               0.01
##                          (0.01)
## grp_ln_lag               -5.85 ***
##                          (1.34)
## trade_ln_lag             -0.09
##                          (0.38)
## roaddensityln_lag         1.77
##                          (1.72)
## corruption_lag           -3.09
##                          (3.99)
## universal_cut            -2.71 *
##                          (1.35)
## targeted_cut             -1.24
##                          (0.82)
## empl_high_lag            -0.18 *
##                          (0.08)
## grpgrowth_lag            -0.00
##                          (0.04)
## ---------------------------------
## Num. obs.               870
## R^2 (full model)          0.45
## R^2 (proj model)          0.04
## Adj. R^2 (full model)     0.38
## Adj. R^2 (proj model)    -0.08
## =================================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

```r
stargazer(model_1, model_2, model_3,
          header = FALSE,
          add.lines = list(c("Fixed Effects", "No", "Yes", "Yes")),
          omit.stat = c("adj.rsq", "f"),
          model.numbers = FALSE,
          model.names = FALSE,
          dep.var.caption = "Dependent Variable: Real GRP growth",
          dep.var.labels.include = FALSE,
          column.labels = c("Model 1", "Model 2", "Model 3"),
          title = "Difference-in-Differences OLS",
          covariate.labels = c("Investment risk",
                               "GRP (log)",
                               "Trade (log)",
                               "Road density (log)",
                               "Corruption",
                               "Non-discriminatory cut",
                              "Period",
                              "Non-discriminatory cut X Period",
                              "Discriminatory cut",
                              "Discriminatory cut X Period",
                              "Share of people employed in higher education",
                              "Lagged GRP growth (log)"),
          notes = "All control variables have been lagged one year.",
          type = "text")
```

```
##
## Difference-in-Differences OLS
## ===============================================================================
```

```
##                                                   Dependent Variable: Real GRP growth
##                                                -------------------------------------------------
##                                                     Model 1         Model 2         Model 3
## -----------------------------------------------------------------------------------------------
## Investment risk                                      0.010           0.012           0.012
##                                                     (0.010)         (0.013)         (0.013)
##
## GRP (log)                                          -1.549***       -5.854***       -5.846***
##                                                     (0.362)         (1.308)         (1.343)
##
## Trade (log)                                         0.880***        -0.095          -0.094
##                                                     (0.250)         (0.381)         (0.382)
##
## Road density (log)                                  0.700***         1.767           1.768
##                                                     (0.198)         (1.721)         (1.722)
##
## Corruption                                          -1.331          -3.088          -3.091
##                                                     (3.153)         (3.988)         (3.991)
##
## Non-discriminatory cut                               1.168
##                                                     (1.220)
##
## Period                                              3.706***
##                                                     (0.620)
##
## Non-discriminatory cut X Period                     -2.490         -2.706**        -2.710**
##                                                     (1.647)         (1.340)         (1.348)
##
## Discriminatory cut                                   0.876
##                                                     (0.752)
##
## Discriminatory cut X Period                         -1.140          -1.237          -1.237
##                                                     (1.006)         (0.820)         (0.821)
##
## Share of people employed in higher education         0.017         -0.184**        -0.184**
##                                                     (0.049)         (0.084)         (0.084)
##
## Lagged GRP growth (log)                                                             -0.001
##                                                                                     (0.036)
##
## Constant                                            6.264***
##                                                     (2.402)
##
## -----------------------------------------------------------------------------------------------
## Fixed Effects                                         No              Yes             Yes
## Observations                                          870             870             870
## R2                                                   0.076           0.451           0.451
## Residual Std. Error                            6.711 (df = 858) 5.453 (df = 772) 5.456 (df = 771)
## ===============================================================================================
## Note:                                                          *p<0.1; **p<0.05; ***p<0.01
##                                                  All control variables have been lagged one year.
# ORANGE - UNIVERSAL
# YELLOW - TARGETED
```

The distinguishing feature of the inclusion of year and region fixed effects can be seen is that now the *universal_cut* variable from the regression table above is statistically significant. In other words, we now have a statistically significant effect for the difference in differences. Moreover, there is potential that there was possibility of decreased omitted variable bias of our estimates in our model.

Moreover, we do not include "period" in this model specification as it is replaced instead with pre-coded treatment variables. These pre-coded treatment variables have already accounted for the time-period making its inclusion in our model unnecessary. On the other hand, the inclusion of interaction variables and fixed effects assists in a robust controlling of time in our model. The use of cut variables reduces the effect of the inclusion period might have on the regression as time has already been accounted for in the regression.

Now, coming back to the tax cuts themselves we can see that there is a negative correlation between both 'universal_cut' and 'targeted_cut' with respect to the dependent variable real GRP growth. In other words, both non-discriminatory and discriminatory tax cuts are negatively correlated with real GRP growth. Yet, the relationship is only statistically significant for 'universal_cut' which is the non discriminatory tax cut post-2003 period. The implication here seems to be that Russia should increase non-discriminatory taxation to stimulate growth.

Lastly, its interesting to see that a lot of the controls except the lagged variable for GRP lose and share of people in higher education lose their statistical significance in this model. This increases the validity of the argument for the inclusion of region fixed effects in the model as we may control for only time-variant variation in this quasi-experiment.

Task 1.4

4. include leads for both targeted and universal regions. Discuss the results;

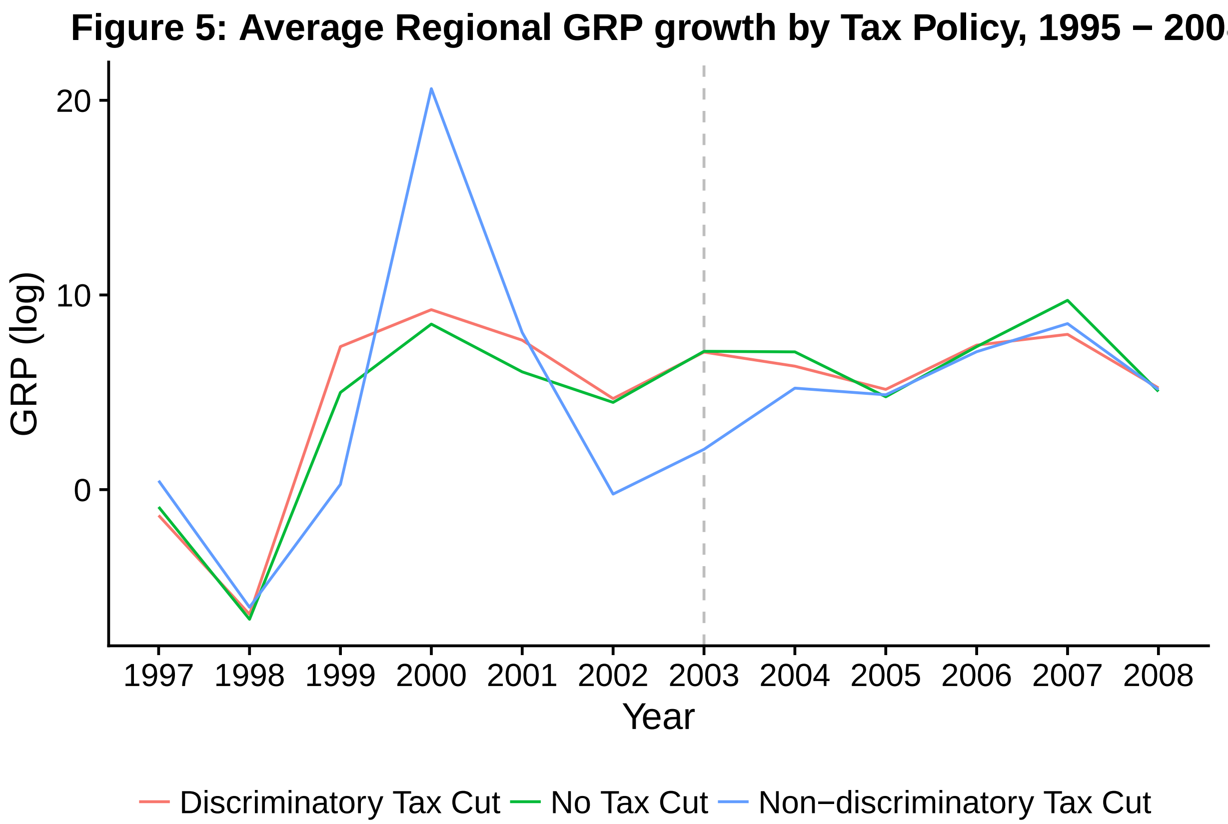
Solution 1.4

One of the basic assumption of the Difference in Difference (DID) model is that the treatment and control group will follow the same trend without the treatment or intervention. This can also be termed as the parallel trends assumption. In the code chunk below I generate these leads and test their significance. If the leads are significant then that is evidence of anticipatory effects, which is problematic for the parallel trends assumption.

```r
# ORANGE - UNIVERSAL
# YELLOW - TARGETED

## generating new variables
russia_data$treatment <- ifelse(russia_data$universal == 1,
                            "Non-discriminatory Tax Cut",
                            ifelse(russia_data$targeted == 1,
                                   "Discriminatory Tax Cut",
                                   "No Tax Cut"))

## now making our plot to visualize the trends

figure_5 <- ggplot(russia_data,
       aes(year, realgrpgrowth, col = treatment)) +
  geom_vline(xintercept = 2003,
             linetype = "dashed",
             col = "grey") +
#  geom_jitter(alpha = 0.2) +
  stat_summary(fun.y = mean, geom = "line") +
  scale_x_continuous(breaks = seq(1995, 2008, by = 1)) +
```

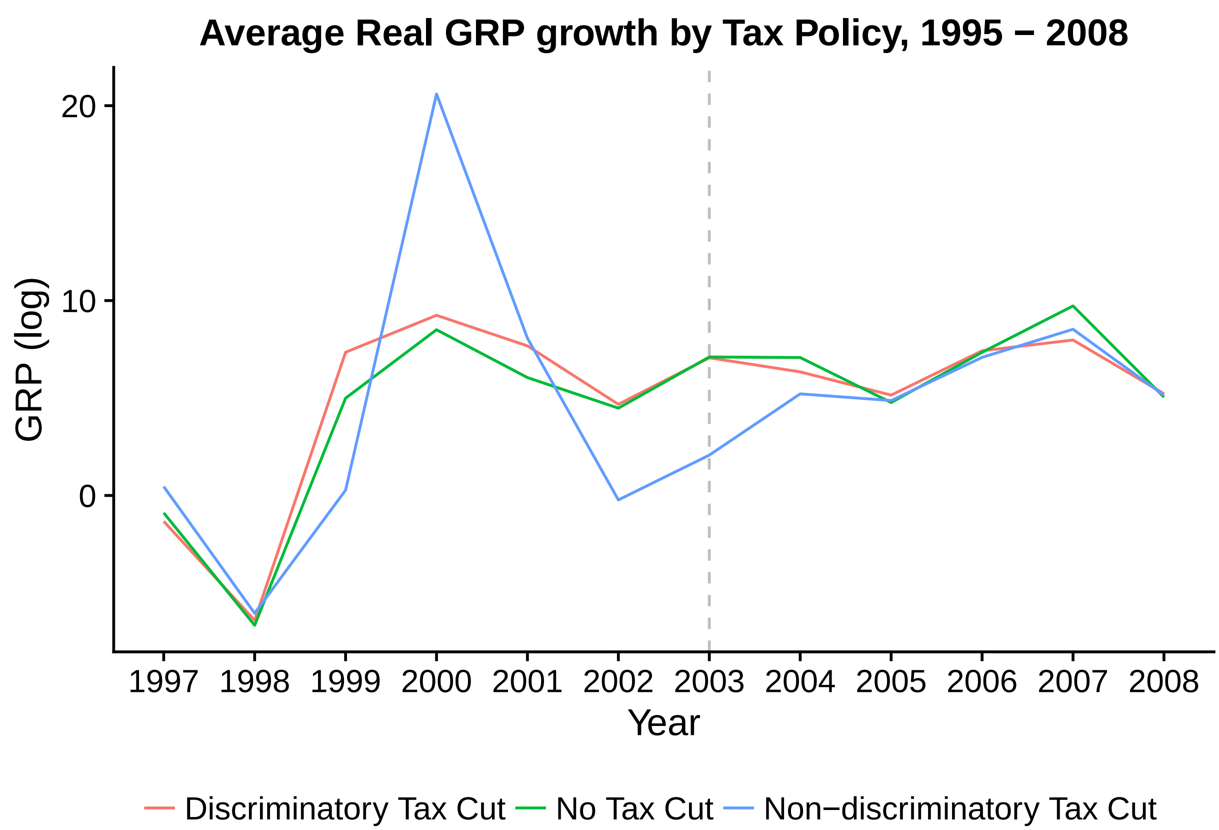
```
  xlab("Year") +
  ylab("GRP (log)") +
  ggtitle("Figure 5: Average Regional GRP growth by Tax Policy, 1995 - 2008") +
  guides(col = guide_legend(title = NULL)) +
  theme(legend.position = "bottom"); figure_5
```

```
## Warning: Removed 181 rows containing non-finite values (stat_summary).
```




```
ggplot(russia_data,
       aes(year, realgrpgrowth, col = treatment)) +
  geom_vline(xintercept = 2003, linetype = "dashed", col = "grey") +
#  geom_jitter(alpha = 0.2) +
  stat_summary(fun.y = mean, geom = "line") +
  scale_x_continuous(breaks = seq(1995, 2008, by = 1)) +
  xlab("Year") +
  ylab("GRP (log)") +
  ggtitle("Average Real GRP growth by Tax Policy, 1995 - 2008") +
  guides(col = guide_legend(title = NULL)) +
  theme(legend.position = "bottom")
```

```
## Warning: Removed 181 rows containing non-finite values (stat_summary).
```




```
# ORANGE - UNIVERSAL
# YELLOW - TARGETED

# Generate leads
russia_data <- russia_data %>%
  mutate(
    universal_lead = ifelse(universal == 1 & year %in% c(2001, 2002), 1, 0),
    targeted_lead = ifelse(targeted == 1 & year %in% c(2001, 2002), 1, 0)
  )

# ORANGE - UNIVERSAL
# YELLOW - TARGETED


model_4 <- felm(realgrpgrowth ~ invrisk_lag + grp_ln_lag +
                  trade_ln_lag + roaddensityln_lag + corruption_lag
                + universal_cut + targeted_cut + grpgrowth_lag +
                  universal_lead + targeted_lead + empl_high_lag | region + year,
                data = russia_data)

screenreg(list(model_3, model_4))
```

```
##
## =============================================
##                          Model 1    Model 2
## ---------------------------------------------
## invrisk_lag                0.01       0.01
##                           (0.01)     (0.01)
```

```
## grp_ln_lag              -5.85 ***   -5.78 ***
##                         (1.34)      (1.35)
## trade_ln_lag            -0.09       -0.08
##                         (0.38)      (0.38)
## roaddensityln_lag        1.77        1.82
##                         (1.72)      (1.72)
## corruption_lag          -3.09       -3.07
##                         (3.99)      (3.99)
## universal_cut           -2.71 *     -4.02 *
##                         (1.35)      (1.56)
## targeted_cut            -1.24       -1.45
##                         (0.82)      (0.96)
## empl_high_lag           -0.18 *     -0.18 *
##                         (0.08)      (0.08)
## grpgrowth_lag           -0.00        0.01
##                         (0.04)      (0.04)
## universal_lead                      -3.37
##                                     (2.04)
## targeted_lead                       -0.56
##                                     (1.24)
## ---------------------------------------------
## Num. obs.              870         870
## R^2 (full model)         0.45        0.45
## R^2 (proj model)         0.04        0.04
## Adj. R^2 (full model)    0.38        0.38
## Adj. R^2 (proj model)   -0.08       -0.08
## =============================================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

We can see that there is a negligible difference in the R squared after including 1 year leads for both universal and discriminatory tax cuts for the different regions of Russia from the previous question. Overall, the results seem largely consistent with the results achieved in the earlier specification of the model. The inclusion of one year leads for both discriminatory and non-discriminatory tax cuts are not statistically significant. What is notable, however, is the fact that the inclusion of one year leads results in a decrease in the coefficient of universal_cut on realgrpgrowth. Though the statistical significance remains the same and the direction of the relationship remains negative.

The lack of potential anticipatory effects increases our confidence in the exogeneity of the treatment (or the tax cuts). This implies that actors did not anticipate the annoucement of tax cuts in the fiscal regions of Russia. Moreover, we can observe from Figure 5 above that with the exception of one outlier region which is pulling up the parallel trends, by and large the different regions of Russia seem to be following the same trends. In other words, regions that implemented no tax cut as well as those that did (whether it be discriminatory or non-discriminatory) are all following the same trend and there is little evidence of any anticipatory effects.

Furthermore, we also find evidence that as there is no anticipation of tax cuts, an indiscriminate tax cuts has an improved effect on real grp growth.This is in contrast with a discriminatory tax cut that is statistically insignificant. One explanation for this can be perhaps that since institutions in Russia are weak there is a great deal of an inter relationship between the business sector and the government. In other words, the discriminate tax cuts are ineffective in stimulating growth as actors can potentially reorient agreed upon and negotiated in advance. However, at the same time the leads are insigifnificant so such an explanation of prior negotiations is question. That said, there is evidence, however, that non-discriminatory tax cutes are better for generating stimulus for the Russian economy if they were implemented in 2002.

Task 1.5

5. include region-specific trends to test the parallel trend assumption. Discuss the results

Solution 1.5

Note that in the following code chunk we include region specific trends to test the parallel trends assumption. Note that if the universal_cut and targeted_cut dummies are still significant after the inclusion of region specific trends, that would be evidence that the parallel trends holds. However, if the trends kill the significance of the treatment, this implies that the parallel trends assumption does not hold.

```
# ORANGE - UNIVERSAL
# YELLOW - TARGETED

# this includes region year fixed effects

model_5 <- lm(realgrpgrowth ~ invrisk_lag + grp_ln_lag +
                trade_ln_lag + roaddensityln_lag + corruption_lag
              + universal_cut + targeted_cut  +
                factor(region)*year + empl_high_lag,
              data = russia_data)

screenreg(list(model_5))
```

```
##
## ===================================================
##                                       Model 1
## ---------------------------------------------------
## (Intercept)                          -4905.65 ***
##                                      (1155.06)
## invrisk_lag                              0.01
##                                         (0.02)
## grp_ln_lag                              -8.42 ***
##                                         (0.63)
## trade_ln_lag                            -0.93
##                                         (0.52)
## roaddensityln_lag                        4.23
##                                         (2.19)
## corruption_lag                          -6.79
##                                         (8.60)
## universal_cut                           -5.07
##                                         (2.73)
## targeted_cut                            -0.73
##                                         (1.46)
## factor(region)alania_rep              3128.84
##                                      (1613.10)
## factor(region)altai_kr                2037.74
##                                      (1627.81)
## factor(region)altai_rep               1155.73
##                                      (1659.88)
## factor(region)amur_obl                 367.18
##                                      (1782.20)
## factor(region)arkhangelsk_obl         2837.74
##                                      (1653.30)
```

```
## factor(region)astrakhan_obl              2840.88
##                                         (1669.66)
## factor(region)bashkortostan_rep          874.60
##                                         (1613.04)
## factor(region)belgorod_obl               593.47
##                                         (1618.41)
## factor(region)bryansk_obl               -148.18
##                                         (1799.07)
## factor(region)buryatia_rep              2153.23
##                                         (1636.20)
## factor(region)chechen_rep                 24.20
##                                           (13.63)
## factor(region)chelyabinsk_obl            424.01
##                                         (1658.39)
## factor(region)chukotka_ao                -34.16
##                                         (1714.58)
## factor(region)chuvash_rep               -913.70
##                                         (1773.43)
## factor(region)dagestan_rep             -1532.65
##                                         (1691.53)
## factor(region)ingushetia_rep             -64.23
##                                         (1631.75)
## factor(region)irkutsk_obl               1428.35
##                                         (1685.23)
## factor(region)ivanovo_obl               1546.12
##                                         (1618.68)
## factor(region)jewish_aobl               -702.96
##                                         (1650.37)
## factor(region)kabardin_rep              4378.74 **
##                                         (1632.81)
## factor(region)kaliningrad_obl           -112.41
##                                         (1658.48)
## factor(region)kalmykia_rep              3885.43 *
##                                         (1802.44)
## factor(region)kaluga_obl                 566.48
##                                         (1717.18)
## factor(region)kamchatka_kr              2678.40
##                                         (1615.05)
## factor(region)karachay_rep              2150.98
##                                         (1629.56)
## factor(region)karelia_rep               2890.12
##                                         (1661.44)
## factor(region)kemerovo_obl              1841.76
##                                         (1671.91)
## factor(region)khabarovsk_kr             2382.14
##                                         (1673.75)
## factor(region)khakassia_rep             2128.41
##                                         (1610.59)
## factor(region)kirov_obl                 2809.55
##                                         (1607.50)
## factor(region)komi_rep                  2542.53
##                                         (1661.03)
## factor(region)kostroma_obl              2094.97
##                                         (1624.75)
```

```
## factor(region)krasnodar_kr              1905.45
##                                        (1647.23)
## factor(region)krasnoyarsk_kr            2343.62
##                                        (1671.01)
## factor(region)kurgan_obl                 638.86
##                                        (1656.42)
## factor(region)kursk_obl                 2430.13
##                                        (1631.77)
## factor(region)leningrad_obl             1478.51
##                                        (1694.16)
## factor(region)lipetsk_obl               1017.26
##                                        (1666.31)
## factor(region)magadan_obl               1933.16
##                                        (1619.25)
## factor(region)mariel_rep                1075.45
##                                        (1632.74)
## factor(region)mordovia_rep              1332.19
##                                        (1664.73)
## factor(region)moscow                     372.14
##                                        (1630.82)
## factor(region)moscow_obl                1330.77
##                                        (1726.54)
## factor(region)murmansk_obl              2739.18
##                                        (1613.87)
## factor(region)nizhnynovgorod_obl        2256.95
##                                        (1609.98)
## factor(region)novgorod_obl              2741.63
##                                        (1612.78)
## factor(region)novosibirsk_obl           1596.45
##                                        (1683.20)
## factor(region)omsk_obl                   896.01
##                                        (1614.32)
## factor(region)orenburg_obl               518.37
##                                        (1652.12)
## factor(region)oryol_obl                 3476.97 *
##                                        (1614.08)
## factor(region)penza_obl                 1230.03
##                                        (1667.33)
## factor(region)perm_kr                   1164.70
##                                        (1860.05)
## factor(region)primorsky_kr              1546.26
##                                        (1628.16)
## factor(region)pskov_obl                 2240.43
##                                        (1666.93)
## factor(region)rostov_obl                 -99.88
##                                        (1787.36)
## factor(region)ryazan_obl                2682.65
##                                        (1619.65)
## factor(region)sakhalin_obl              -181.20
##                                        (1661.30)
## factor(region)samara_obl                2220.49
##                                        (1650.76)
## factor(region)saratov_obl               1801.77
##                                        (1658.64)
```

```
## factor(region)smolensk_obl              2757.74
##                                         (1667.79)
## factor(region)spetersburg               1099.14
##                                         (1753.51)
## factor(region)stavropol_kr              1590.74
##                                         (1677.81)
## factor(region)sverdlovsk_obl            1369.74
##                                         (1621.66)
## factor(region)tambov_obl                2379.06
##                                         (1659.21)
## factor(region)tatarstan_rep             1136.52
##                                         (1708.44)
## factor(region)tomsk_obl                 2969.33
##                                         (1622.47)
## factor(region)tula_obl                   803.24
##                                         (1619.95)
## factor(region)tver_obl                  1619.59
##                                         (1631.74)
## factor(region)tyumen_obl                4656.22 **
##                                         (1707.57)
## factor(region)tyva_rep                  2535.62
##                                         (1624.33)
## factor(region)udmurt_rep                1610.21
##                                         (1821.23)
## factor(region)ulyanovsk_obl             1905.74
##                                         (1609.02)
## factor(region)vladimir_obl              2022.27
##                                         (1682.46)
## factor(region)volgograd_obl             1712.54
##                                         (1716.51)
## factor(region)vologda_obl               2877.18
##                                         (1684.27)
## factor(region)voronezh_obl               614.31
##                                         (1652.93)
## factor(region)yakutia_rep               2552.99
##                                         (1715.87)
## factor(region)yaroslav_obl              2543.96
##                                         (1676.49)
## factor(region)zabaykalsky_kr            1941.47
##                                         (1643.29)
## year                                       2.47 ***
##                                           (0.58)
## empl_high_lag                             -0.25 **
##                                           (0.10)
## factor(region)alania_rep:year             -1.56
##                                           (0.81)
## factor(region)altai_kr:year               -1.01
##                                           (0.81)
## factor(region)altai_rep:year              -0.57
##                                           (0.83)
## factor(region)amur_obl:year               -0.17
##                                           (0.89)
## factor(region)arkhangelsk_obl:year        -1.40
##                                           (0.82)
```

```
## factor(region)astrakhan_obl:year          -1.41
##                                           (0.83)
## factor(region)bashkortostan_rep:year      -0.42
##                                           (0.81)
## factor(region)belgorod_obl:year           -0.29
##                                           (0.81)
## factor(region)bryansk_obl:year             0.08
##                                           (0.90)
## factor(region)buryatia_rep:year           -1.06
##                                           (0.82)
## factor(region)chelyabinsk_obl:year        -0.20
##                                           (0.83)
## factor(region)chukotka_ao:year             0.03
##                                           (0.86)
## factor(region)chuvash_rep:year             0.46
##                                           (0.89)
## factor(region)dagestan_rep:year            0.78
##                                           (0.84)
## factor(region)ingushetia_rep:year          0.03
##                                           (0.81)
## factor(region)irkutsk_obl:year            -0.69
##                                           (0.84)
## factor(region)ivanovo_obl:year            -0.77
##                                           (0.81)
## factor(region)jewish_aobl:year             0.35
##                                           (0.82)
## factor(region)kabardin_rep:year           -2.18 **
##                                           (0.82)
## factor(region)kaliningrad_obl:year         0.06
##                                           (0.83)
## factor(region)kalmykia_rep:year           -1.94 *
##                                           (0.90)
## factor(region)kaluga_obl:year             -0.27
##                                           (0.86)
## factor(region)kamchatka_kr:year           -1.33
##                                           (0.81)
## factor(region)karachay_rep:year           -1.07
##                                           (0.81)
## factor(region)karelia_rep:year            -1.43
##                                           (0.83)
## factor(region)kemerovo_obl:year           -0.90
##                                           (0.83)
## factor(region)khabarovsk_kr:year          -1.17
##                                           (0.84)
## factor(region)khakassia_rep:year          -1.06
##                                           (0.80)
## factor(region)kirov_obl:year              -1.39
##                                           (0.80)
## factor(region)komi_rep:year               -1.25
##                                           (0.83)
## factor(region)kostroma_obl:year           -1.04
##                                           (0.81)
## factor(region)krasnodar_kr:year           -0.93
##                                           (0.82)
```

```
## factor(region)krasnoyarsk_kr:year         -1.15
##                                            (0.83)
## factor(region)kurgan_obl:year              -0.31
##                                            (0.83)
## factor(region)kursk_obl:year               -1.20
##                                            (0.81)
## factor(region)leningrad_obl:year           -0.72
##                                            (0.85)
## factor(region)lipetsk_obl:year             -0.50
##                                            (0.83)
## factor(region)magadan_obl:year             -0.96
##                                            (0.81)
## factor(region)mariel_rep:year              -0.53
##                                            (0.82)
## factor(region)mordovia_rep:year            -0.66
##                                            (0.83)
## factor(region)moscow:year                  -0.16
##                                            (0.81)
## factor(region)moscow_obl:year              -0.65
##                                            (0.86)
## factor(region)murmansk_obl:year            -1.35
##                                            (0.81)
## factor(region)nizhnynovgorod_obl:year      -1.11
##                                            (0.80)
## factor(region)novgorod_obl:year            -1.36
##                                            (0.81)
## factor(region)novosibirsk_obl:year         -0.78
##                                            (0.84)
## factor(region)omsk_obl:year                -0.43
##                                            (0.81)
## factor(region)orenburg_obl:year            -0.25
##                                            (0.82)
## factor(region)oryol_obl:year               -1.73 *
##                                            (0.81)
## factor(region)penza_obl:year               -0.61
##                                            (0.83)
## factor(region)perm_kr:year                 -0.56
##                                            (0.93)
## factor(region)primorsky_kr:year            -0.76
##                                            (0.81)
## factor(region)pskov_obl:year               -1.11
##                                            (0.83)
## factor(region)rostov_obl:year               0.07
##                                            (0.89)
## factor(region)ryazan_obl:year              -1.33
##                                            (0.81)
## factor(region)sakhalin_obl:year             0.11
##                                            (0.83)
## factor(region)samara_obl:year              -1.09
##                                            (0.82)
## factor(region)saratov_obl:year             -0.89
##                                            (0.83)
## factor(region)smolensk_obl:year            -1.37
##                                            (0.83)
```

```
## factor(region)spetersburg:year             -0.53
##                                              (0.88)
## factor(region)stavropol_kr:year            -0.78
##                                              (0.84)
## factor(region)sverdlovsk_obl:year          -0.66
##                                              (0.81)
## factor(region)tambov_obl:year              -1.18
##                                              (0.83)
## factor(region)tatarstan_rep:year           -0.55
##                                              (0.85)
## factor(region)tomsk_obl:year               -1.47
##                                              (0.81)
## factor(region)tula_obl:year                -0.39
##                                              (0.81)
## factor(region)tver_obl:year                -0.80
##                                              (0.81)
## factor(region)tyumen_obl:year              -2.30 **
##                                              (0.85)
## factor(region)tyva_rep:year                -1.26
##                                              (0.81)
## factor(region)udmurt_rep:year              -0.79
##                                              (0.91)
## factor(region)ulyanovsk_obl:year           -0.94
##                                              (0.80)
## factor(region)vladimir_obl:year            -1.00
##                                              (0.84)
## factor(region)volgograd_obl:year           -0.84
##                                              (0.86)
## factor(region)vologda_obl:year             -1.42
##                                              (0.84)
## factor(region)voronezh_obl:year            -0.30
##                                              (0.83)
## factor(region)yakutia_rep:year             -1.26
##                                              (0.86)
## factor(region)yaroslav_obl:year            -1.26
##                                              (0.84)
## factor(region)zabaykalsky_kr:year          -0.96
##                                              (0.82)
## ---------------------------------------------------
## R^2                                           0.41
## Adj. R^2                                      0.27
## Num. obs.                                   870
## RMSE                                          5.94
## ===================================================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

```r
stargazer(model_4, model_5,
          keep = c("grp_ln_lag", "trade_ln_lag", "roaddensityln_lag",
            "corruption_lag", "invrisk_lag",
              "universal_cut", "targeted_cut", "grpgrowth_lag"),
          header = FALSE,
          add.lines = list(c("Fixed Effects", "Yes", "Yes"),
                           c("Region-Specific Trends", "No", "Yes")),
          omit.stat = c("adj.rsq", "f"),
```

```r
         model.numbers = FALSE,
         model.names = FALSE,
         dep.var.caption = "Dependent Variable: Real GRP growth",
         dep.var.labels.include = FALSE,
         column.labels = c("Model 4", "Model 5"),
         title = "Difference-in-Differences Parallel Trend Diagnostics",
         notes = "All control variables have been lagged one year.",
         type = "text")
```

```
##
## Difference-in-Differences Parallel Trend Diagnostics
## ========================================================================
##                               Dependent Variable: Real GRP growth
##                          -----------------------------------------------
##                                 Model 4                  Model 5
## ------------------------------------------------------------------------
## invrisk_lag                      0.012                    0.007
##                                 (0.013)                  (0.016)
##
## grp_ln_lag                     -5.783***                -8.420***
##                                 (1.345)                  (0.626)
##
## trade_ln_lag                    -0.083                   -0.930*
##                                 (0.382)                  (0.521)
##
## roaddensityln_lag                1.823                    4.233*
##                                 (1.722)                  (2.187)
##
## corruption_lag                  -3.072                   -6.792
##                                 (3.989)                  (8.604)
##
## universal_cut                  -4.020**                 -5.067*
##                                 (1.564)                  (2.727)
##
## targeted_cut                    -1.451                   -0.729
##                                 (0.958)                  (1.456)
##
## grpgrowth_lag                    0.007
##                                 (0.036)
##
## ------------------------------------------------------------------------
## Fixed Effects                     Yes                      Yes
## Region-Specific Trends            No                       Yes
## Observations                      870                      870
## R2                               0.453                    0.408
## Residual Std. Error        5.454 (df = 769)         5.937 (df = 703)
## ========================================================================
## Note:                                         *p<0.1; **p<0.05; ***p<0.01
##                           All control variables have been lagged one year.
```

The crucial insight from the table above is that the statistical significance of *universal_cut* is not killed after the inclusion of the region specific trends in our model. This provides support for the parallel trends assumption. In fact, the coefficient on the non-discriminatory tax cut decreases even further after the inclusion of region specific trends. With regards to overall goodness of fit, we can observe that, for example, the r-squared remains largely similar as earlier.

However, the retention of the statistical significance of “universal_cut” is desirable in this case as it indicative of the fact that there are no characteristics of the individual regions that affect the impact of tax policies on real GRP growth. In other words, this allows us to accept the parallel trend assumption that the regions are comparable and therefore our DID estimation is valid.

Task 1.6 (optional)

Optional: try a D-I-D with treatment intensity with or without region and year fixed effects. Note that the intensity variable must be chosen in a theoretically informed way. Provide the graphical interpretation of the results of the interaction term.

Solution