In [24]:
```
setwd("C:/Users/Asher/OneDrive/Roma Paper/Data/World Bank Long 2013")
# Download packages to import stata file
install.packages('readstata13', dependencies=TRUE, repos='http://cran.rstudio.
com/')
library(readstata13)
install.packages('survey', dependencies=TRUE,
repos='http://cran.rstudio.com/')
library(survey)
```

```
Installing package into 'C:/Users/Asher/Documents/R/win-library/3.1'
(as 'lib' is unspecified)

  There is a binary version available (and will be installed) but the
  source version is later:
           binary source
readstata13  0.8.2  0.8.5


Warning message:
: package 'readstata13' is in use and will not be installedInstalling package
 into 'C:/Users/Asher/Documents/R/win-library/3.1'
(as 'lib' is unspecified)

  There is a binary version available (and will be installed) but the
  source version is later:
      binary source
survey 3.30-3 3.31-2


Warning message:
: package 'survey' is in use and will not be installed
```

In [70]:
```
#Import stata file
IMPORTED_DATA <- read.dta13("Data_For_R.dta")
labname <- get.label.name(IMPORTED_DATA,"type")
labtab <- get.label(IMPORTED_DATA, labname)
get.origin.codes(IMPORTED_DATA$type, labtab)
as.integer(IMPORTED_DATA$type)
#Examine data
head(IMPORTED_DATA)
dim(IMPORTED_DATA)

#RESTRICT BETWEEN 18 AND 65
dat <- subset(IMPORTED_DATA, IMPORTED_DATA$age>17 & IMPORTED_DATA$age<65)
head(dat)
dim(dat)
```

```
Warning message:
In read.dta13("Data_For_R.dta"):
  highest_education:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  reservation_wage:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  dist_willing_commute:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  willing_move:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  m2_q9_month:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  m2_q9_year:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  total_net_income:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  Mean_Net_Wages:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  m2_q16b:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  m2_q17a:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  m2_q17b:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  m2_q17c:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  hours_worked_last_week:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.Warning message:
In read.dta13("Data_For_R.dta"):
  m2_q25d:
  Missing factor labels - no labels assigned.
  Set option generate.factors=T to generate labels.x is no factor.
```

Out[70]:

Out[70]:

	HH_ID	HHM_ID	HH_type	sample_type	Region	Cluster	m1_q1_a1	m1_q1	m1_
1	10103	1	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Èâàí	HEA THE HOU
2	10103	2	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Ìàðèÿ	SPO PAR
3	10104	1	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Ïàâëèíà	HEA THE HOU

	HH_ID	HHM_ID	HH_type	sample_type	Region	Cluster	m1_q1_a1	m1_q1	m1_
4	10104	2	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Åëåíà	SOI DAU
5	10104	3	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Ìèðåëà	grai
6	10104	4	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Òàíÿ	grai

Out[70]: 7492 365

Out[70]:

	HH_ID	HHM_ID	HH_type	sample_type	Region	Cluster	m1_q1_a1	m1_q1	
4	10104	2	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Åëåíà	
7	10106	1	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Òàíÿ	
9	10107	1	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Àòàíàñ	

	HH_ID	HHM_ID	HH_type	sample_type	Region	Cluster	m1_q1_a1	m1_q1	
12	10110	2	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Íàäÿ	
15	10113	3	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Êðàñèìèð	
16	10113	4	HH from previous wave	Main sample	Blagoevgrad	101	STILL A MEMBER	Ãåðãàíà	

Out[70]: 4307 365

In [71]:
```
#TEST WHETHER THERE ARE VISIBLE PROBLEMS WITH VARIABLES USED, BY RUNNING A SIM
PLE LINEAR PROBABILITY MODEL
#THERE SEEM TO BE NO PROBLEMS. NEGATIVE EFFECT OF ROMA ETHNICITY ON EMPLOYMENT
 IN ALL 3. SIGNIFICANT IN ONE
Linear_Probability_Model_NO_INTERACTIONS<- lm(employed ~ Roma + child_under_6
+ child_6_to_17 + highest_education + age_factor, data=dat)
Linear_Probability_Model_INTERACTIONS_WITH_ROMA <- lm(employed ~ Roma + child_
under_6 + child_under_6_int + child_6_to_17 + child_6_to_17_int + highest_educ
ation + highest_education_int + age_factor + age_factor_int, data=dat)
Linear_Probability_Model_INTERACTIONS_PLUS_MARRIED_COGNITIVE_TEST_SCORE <-
lm(employed ~ Roma + married + IRT_Scores_Total_Cognitive_skill + child_under_
6 + child_under_6_int + child_6_to_17 + child_6_to_17_int + highest_education
+ highest_education_int + age_factor + age_factor_int, data=dat)
summary(Linear_Probability_Model_NO_INTERACTIONS)
summary(Linear_Probability_Model_INTERACTIONS_WITH_ROMA)
summary(Linear_Probability_Model_INTERACTIONS_PLUS_MARRIED_COGNITIVE_TEST_SCOR
E)
```

```
Out[71]: Call:
         lm(formula = employed ~ Roma + child_under_6 + child_6_to_17 +
             highest_education + age_factor, data = dat)

         Residuals:
             Min      1Q  Median      3Q     Max
         -0.8241 -0.5066  0.2139  0.4235  0.9462

         Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
         (Intercept)         0.137237   0.036427   3.767 0.000167 ***
         Roma               -0.177729   0.024816  -7.162 9.32e-13 ***
         child_under_6      -0.001605   0.019789  -0.081 0.935366
         child_6_to_17       0.020553   0.015976   1.287 0.198324
         highest_education   0.069878   0.005057  13.817  < 2e-16 ***
         age_factor          0.031952   0.006873   4.649 3.44e-06 ***
         ---
         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

         Residual standard error: 0.4759 on 4293 degrees of freedom
           (8 observations deleted due to missingness)
         Multiple R-squared:  0.0915,    Adjusted R-squared:  0.09044
         F-statistic: 86.47 on 5 and 4293 DF,  p-value: < 2.2e-16
```

```
Out[71]: Call:
         lm(formula = employed ~ Roma + child_under_6 + child_under_6_int +
             child_6_to_17 + child_6_to_17_int + highest_education + highest_education
         _int +
             age_factor + age_factor_int, data = dat)

         Residuals:
             Min      1Q  Median      3Q     Max
         -0.8187 -0.5099  0.2138  0.4186  1.0033

         Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
         (Intercept)            0.129468   0.038228   3.387 0.000714 ***
         Roma                  -0.108046   0.108842  -0.993 0.320918
         child_under_6          0.006693   0.022168   0.302 0.762737
         child_under_6_int     -0.048325   0.049532  -0.976 0.329305
         child_6_to_17          0.013778   0.017265   0.798 0.424898
         child_6_to_17_int      0.039451   0.045724   0.863 0.388296
         highest_education      0.068271   0.005275  12.943  < 2e-16 ***
         highest_education_int  0.016704   0.018818   0.888 0.374772
         age_factor             0.035768   0.007316   4.889 1.05e-06 ***
         age_factor_int        -0.030145   0.021481  -1.403 0.160595
         ---
         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

         Residual standard error: 0.4759 on 4289 degrees of freedom
           (8 observations deleted due to missingness)
         Multiple R-squared:  0.09243,   Adjusted R-squared:  0.09052
         F-statistic: 48.53 on 9 and 4289 DF,  p-value: < 2.2e-16
```

Out[71]:
```
Call:
lm(formula = employed ~ Roma + married + IRT_Scores_Total_Cognitive_skill +
    child_under_6 + child_under_6_int + child_6_to_17 + child_6_to_17_int +
    highest_education + highest_education_int + age_factor +
    age_factor_int, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8871 -0.4963  0.2282  0.4207  1.0215

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                        0.1579116  0.0409704   3.854 0.000118 ***
Roma                              -0.1326782  0.1250334  -1.061 0.288694
married                            0.0763573  0.0193271   3.951 7.93e-05 ***
IRT_Scores_Total_Cognitive_skill   0.0250568  0.0078418   3.195 0.001409 **
child_under_6                     -0.0249253  0.0244682  -1.019 0.308419
child_under_6_int                 -0.0563031  0.0565634  -0.995 0.319607
child_6_to_17                      0.0009976  0.0188350   0.053 0.957763
child_6_to_17_int                  0.0773576  0.0518878   1.491 0.136082
highest_education                  0.0604095  0.0058647  10.301  < 2e-16 ***
highest_education_int              0.0238661  0.0217744   1.096 0.273122
age_factor                         0.0259527  0.0086322   3.007 0.002660 **
age_factor_int                    -0.0311390  0.0243346  -1.280 0.200759
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4757 on 3739 degrees of freedom
  (556 observations deleted due to missingness)
Multiple R-squared:  0.09188,   Adjusted R-squared:  0.08921
F-statistic: 34.39 on 11 and 3739 DF,  p-value: < 2.2e-16
```

In [72]:
```
#CREATE VARIABLES FOR THE NET BENEFIT FROM WORKING (I.E. EARNED INCOME IF WOKR
ING - BENEFITS IF NOT WORKING)

#CREATE TWO SUBSAMPLES - ONE FOR EMPLOYED AND FOR NOT EMPLOYED
sub.sample.unemployed <- subset(dat, dat$employed==0)
sub.sample.employed <- subset(dat, dat$employed==1)

#PREDICT THE BENEFITS AN EMPLOYED PERSON WOULD RECEIVE/
#IF NOT EMPLOYED, USING THE BENEFITS OF THOSE WITH SIMILAR/
#CHARATERISTICS WHO ARE UNEMPLOYED/

BEN.UNEMPLOYED <- lm(monthly_gov_benefits ~ Roma + married + IRT_Scores_Total_
Cognitive_skill + child_under_6 + child_under_6_int + child_6_to_17 + child_6_
to_17_int + highest_education + highest_education_int + age_factor + age_facto
r_int, data=sub.sample.unemployed)

#USE BEN OF UNEMPLOYED TO PREDICT BEN OF EMPLOYED IF UNEMPLOYED
PREDICTED_MONTHLY_GOV_BEN <- predict(BEN.UNEMPLOYED, dat)

#PREDICT THE EARNINGS AN UNEMPLOYED PERSON WOULD RECEIVE/
#IF NOT UNEMPLOYED, USING THE BENEFITS OF THOSE WITH SIMILAR/
#CHARATERISTICS WHO ARE EMPLOYED/ (USE LOG B/C LEFT-SKEWED)

EARN.EMPLOYED <- lm(log(Mean_Net_Wages) ~ Roma + married + IRT_Scores_Total_Co
gnitive_skill + child_under_6 + child_under_6_int + child_6_to_17 + child_6_to
_17_int + highest_education + highest_education_int + age_factor + age_factor_
int, data=sub.sample.employed)

#USE EARNINGS OF EMPLOYED TO PREDICT EARN OF UNEMPLOYED
PREDICTED_MONTHLY_EARNINGS_lOG <- predict(EARN.EMPLOYED , dat)

#TAKE INVERSE OF LOG TO SET PREDICTED EARNINGS EQUAL TO EUROS
PREDICTED_MONTHLY_EARNINGS <- exp(PREDICTED_MONTHLY_EARNINGS_lOG)

#FIND NET BENEFIT OF WORKING
NET_BENEFIT_WORKING <- PREDICTED_MONTHLY_EARNINGS - PREDICTED_MONTHLY_GOV_BEN
summary(NET_BENEFIT_WORKING)
```

Out[72]:
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 -191.0   344.3   403.2   398.7   460.7   749.4     556
```

In [76]:
```
#TEST HOW ACCURATE MY PREDICTIONS OF EARNINGS
#BY ESTIMATING 10 PERCENT OF THE EMPLOYED SUBSAMPLE USING 90% OF THE EMPLOYED,
#AND SEEING HOW MUCH IT DIFFERS

#CREATE INDEX OF OBS
index<- 1:nrow(sub.sample.employed)

#CREATE DATAFRAME FOR 90 PERCENT OF DATA
estimated.index <- sample(index, trunc(length(index)*0.9))
estimator.set <- sub.sample.employed[estimated.index, ]

#CREATE DATAFRAME FOR 10 PERCENT OF DATA (TO BE PREDICTED)
predicted.set <- sub.sample.employed[-estimated.index, ]

#DIMENSION OF TWO SETS
dim(estimator.set)
dim(predicted.set)

# ESTIMATE LOG LINEAR MODEL FOR EARNINGS FOR ESIMATOR DATAFRAME
PERCENT.90.EARN.EMPLOYED <- lm(log(Mean_Net_Wages) ~ Roma + married + IRT_Scor
es_Total_Cognitive_skill + child_under_6 + child_under_6_int + child_6_to_17 +
child_6_to_17_int + highest_education + highest_education_int + age_factor + a
ge_factor_int, data=estimator.set)
#USING THE RESULTS ABOVE, PREDICT EARNINGS FOR OTHER 10 PERCENT OF EMPLOYED
PERCENT.10.PREDICTED.EARNINGS <- predict(PERCENT.90.EARN.EMPLOYED, predicted.s
et)
e_10.PERCENT.PREDICTED.EARNINGS <- exp(PERCENT.10.PREDICTED.EARNINGS)

#SUMMARIZE
summary(e_10.PERCENT.PREDICTED.EARNINGS)
summary(predicted.set$Mean_Net_Wages)

#DIFFERENCE BETWEEN PREDICTED AND SAMPLED
resid_wage <- (predicted.set$Mean_Net_Wages - e_10.PERCENT.PREDICTED.EARNINGS)
plot(resid_wage, ylab="Difference in Surveyed and Predicted Wages", xlab="Surv
eyed Monthly Wages")
plot(predicted.set$Mean_Net_Wages, resid_wage [resid_wage<1000])
plot(density(resid_wage, na.rm = TRUE))
plot(predicted.set$Mean_Net_Wages, resid_wage/predicted.set$Mean_Net_Wages)
```

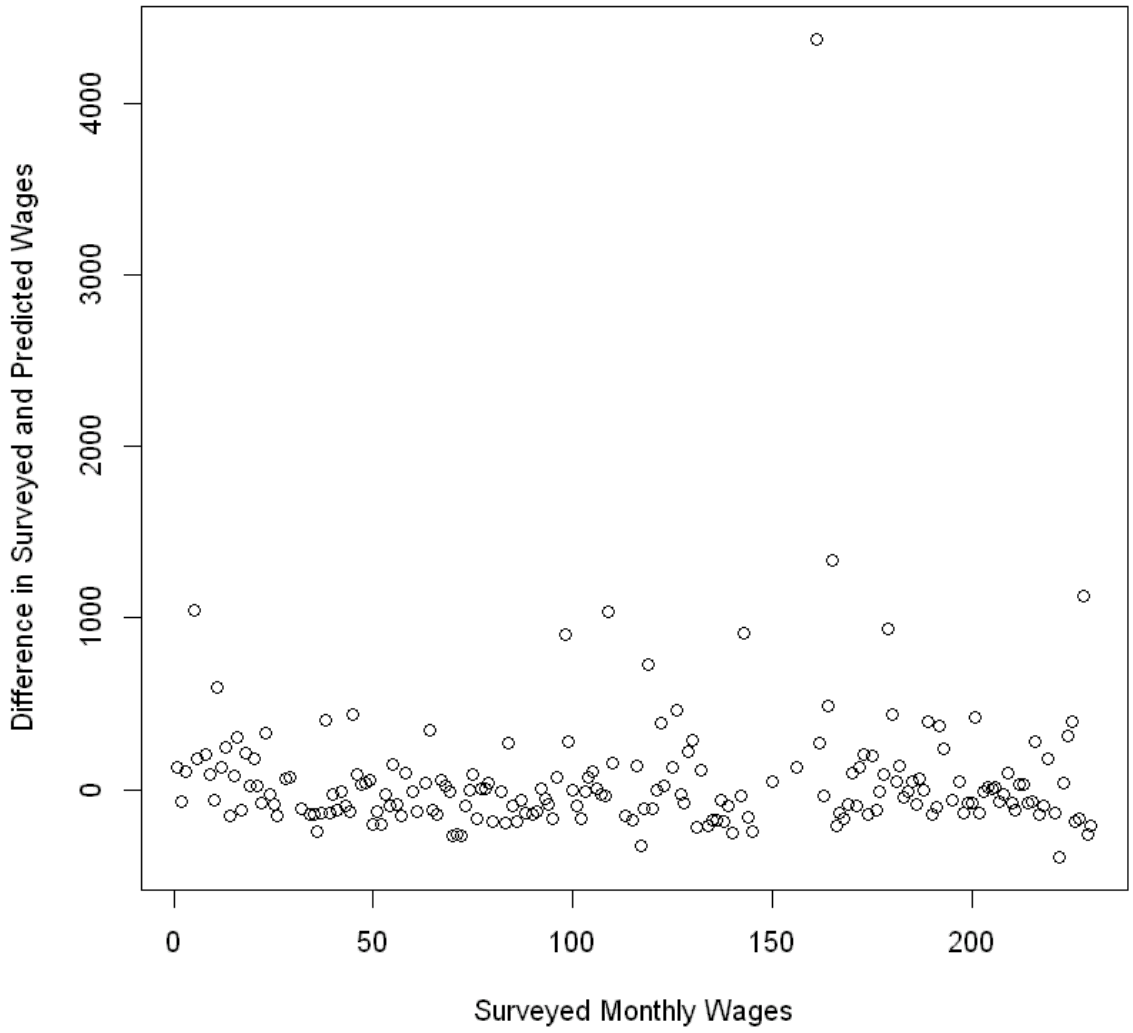
```
Out[76]:      2064  365
```

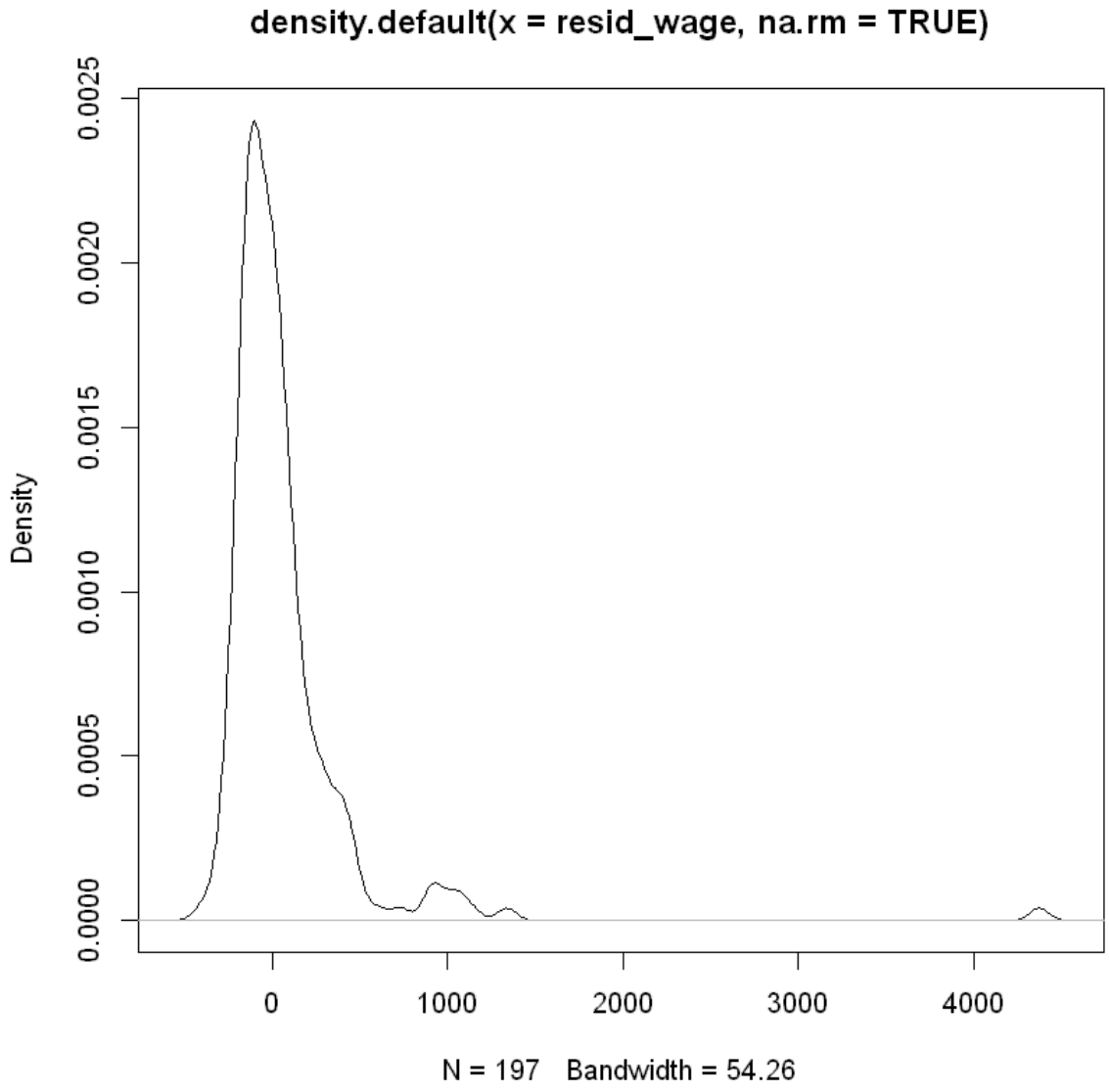
```
Out[76]:      230  365
```

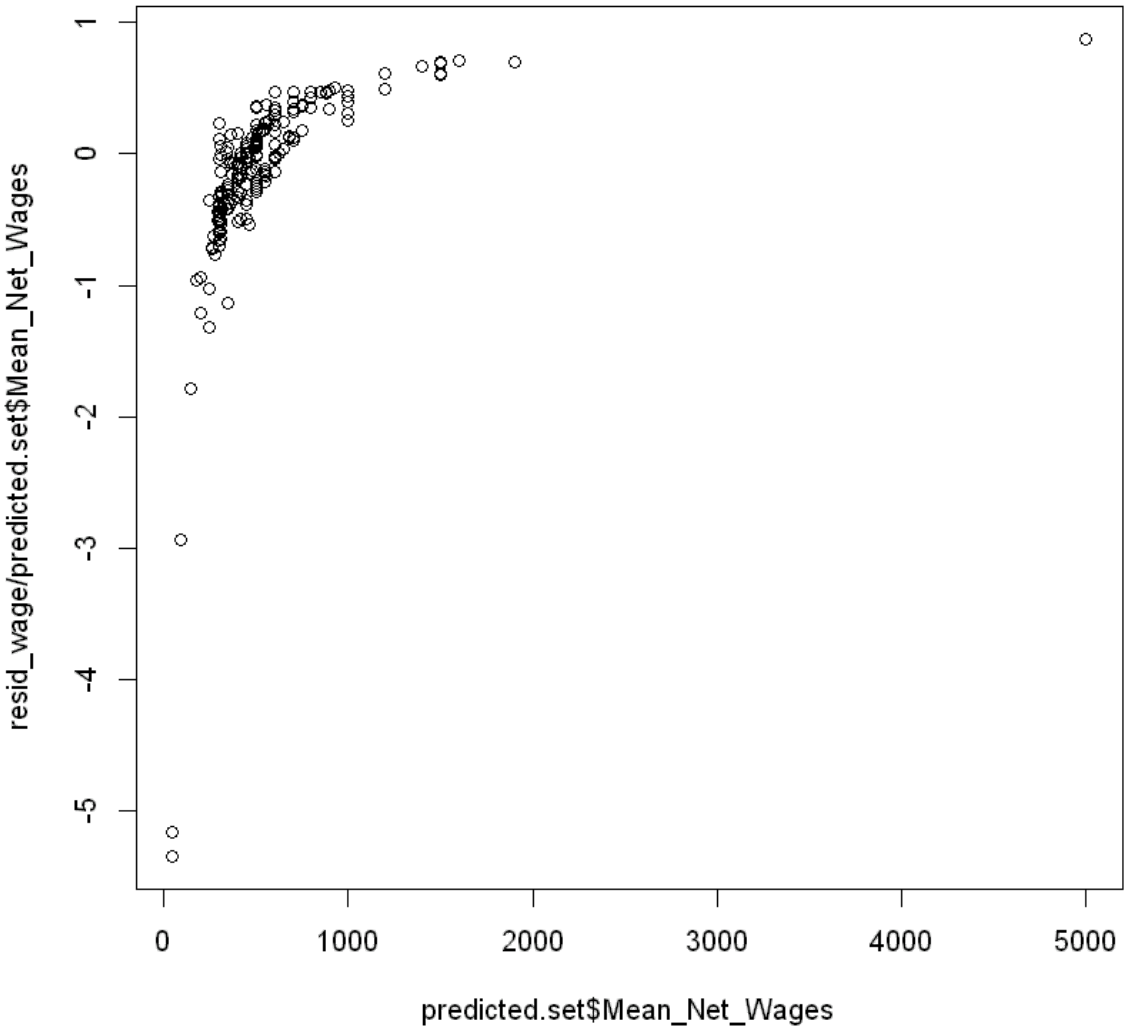
```
Out[76]:     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
            231.3   427.5   462.7   480.3   531.1   749.8      33
```

```
Out[76]:     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
             50.0   350.0   470.0   548.8   600.0  5000.0
```

```
Error in xy.coords(x, y, xlabel, ylabel, log): 'x' and 'y' lengths differ
```




density.default(x = resid_wage, na.rm = TRUE)
Density
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
0
1000
2000
3000
4000
N = 197 Bandwidth = 54.26


resid_wage/predicted.set$Mean_Net_Wages
predicted.set$Mean_Net_Wages
1
0
-1
-2
-3
-4
-5
0
1000
2000
3000
4000
5000

In [87]:
```
#Introduce a probit model, that is able to estimate the marginal effect of Rom
a on the probability of being employed.
probit_roma <- glm(employed ~ Roma, family=binomial(link="probit"), data=dat)
probit_1 <- glm(employed ~ NET_BENEFIT_WORKING + Roma + IRT_Scores_Total_Cogni
tive_skill, family=binomial(link="probit"), data=dat)
probit_2 <- glm(employed ~ NET_BENEFIT_WORKING + Roma + IRT_Scores_Total_Cogni
tive_skill + married + child_under_6 + child_under_6_int + child_6_to_17 + chi
ld_6_to_17_int + highest_education + highest_education_int + age_factor + age_
factor_int, family=binomial(link="probit"), data=dat)
probit_3 <- glm(employed ~ Roma + IRT_Scores_Total_Cognitive_skill + married +
child_under_6 + child_under_6_int + child_6_to_17 + child_6_to_17_int + highes
t_education + highest_education_int + age_factor + age_factor_int, family=bino
mial(link="probit"), data=dat)
probit_4 <- glm(employed ~ Roma + married + child_under_6 + child_under_6_int
+ child_6_to_17 + child_6_to_17_int + highest_education + highest_education_in
t + age_factor + age_factor_int, family=binomial(link="probit"), data=dat)
probit_5 <- glm(employed ~ Roma + NET_BENEFIT_WORKING + IRT_Scores_Total_Cogni
tive_skill + married + child_under_6 + child_6_to_17  + highest_education + ag
e_factor, family=binomial(link="probit"), data=dat)
summary(probit_roma)
summary(probit_1)
summary(probit_2)
summary(probit_3)
summary(probit_4)
summary(probit_5)
```

```
Out[87]: Call:
         glm(formula = employed ~ Roma, family = binomial(link = "probit"),
             data = dat)

         Deviance Residuals:
            Min      1Q  Median      3Q     Max
         -1.303  -1.303   1.057   1.057   1.659

         Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
         (Intercept)  0.18107    0.02053    8.82   <2e-16 ***
         Roma        -0.84812    0.06270  -13.53   <2e-16 ***
         ---
         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

         (Dispersion parameter for binomial family taken to be 1)

             Null deviance: 5941.3  on 4298  degrees of freedom
         Residual deviance: 5746.4  on 4297  degrees of freedom
           (8 observations deleted due to missingness)
         AIC: 5750.4

         Number of Fisher Scoring iterations: 4
```

```
Out[87]: Call:
         glm(formula = employed ~ NET_BENEFIT_WORKING + Roma + IRT_Scores_Total_Cognit
         ive_skill,
             family = binomial(link = "probit"), data = dat)

         Deviance Residuals:
             Min       1Q   Median       3Q      Max
         -1.8673  -1.2184   0.7922   1.0725   2.0904

         Coefficients:
                                           Estimate Std. Error z value Pr(>|z|)
         (Intercept)                     -0.7199487  0.0964959  -7.461 8.59e-14 ***
         NET_BENEFIT_WORKING              0.0021177  0.0002234   9.478  < 2e-16 ***
         Roma                            -0.2225037  0.0881051  -2.525   0.0116 *
         IRT_Scores_Total_Cognitive_skill  0.0234436  0.0231234   1.014   0.3107
         ---
         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

         (Dispersion parameter for binomial family taken to be 1)

             Null deviance: 5175.2  on 3750  degrees of freedom
         Residual deviance: 4899.5  on 3747  degrees of freedom
           (556 observations deleted due to missingness)
         AIC: 4907.5

         Number of Fisher Scoring iterations: 4
```

Out[87]:
```
Call:
glm(formula = employed ~ NET_BENEFIT_WORKING + Roma + IRT_Scores_Total_Cognit
ive_skill +
    married + child_under_6 + child_under_6_int + child_6_to_17 +
    child_6_to_17_int + highest_education + highest_education_int +
    age_factor + age_factor_int, family = binomial(link = "probit"),
    data = dat)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.846  -1.175   0.760   1.027   2.443

Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                     -0.055513   0.528914  -0.105 0.916410
NET_BENEFIT_WORKING             -0.003365   0.002055  -1.638 0.101511
Roma                            -1.134369   0.525713  -2.158 0.030946 *
IRT_Scores_Total_Cognitive_skill  0.152654   0.055473   2.752 0.005925 **
married                          0.313765   0.084162   3.728 0.000193 ***
child_under_6                   -0.263522   0.136252  -1.934 0.053103 .
child_under_6_int               -0.741458   0.376148  -1.971 0.048702 *
child_6_to_17                   -0.149055   0.105496  -1.413 0.157688
child_6_to_17_int                0.156348   0.159102   0.983 0.325758
highest_education                0.364934   0.126098   2.894 0.003803 **
highest_education_int            0.096948   0.070091   1.383 0.166615
age_factor                      -0.006843   0.050400  -0.136 0.891997
age_factor_int                   0.033976   0.101008   0.336 0.736594
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5175.2  on 3750  degrees of freedom
Residual deviance: 4810.1  on 3738  degrees of freedom
  (556 observations deleted due to missingness)
AIC: 4836.1

Number of Fisher Scoring iterations: 4
```

Out[87]:
```
Call:
glm(formula = employed ~ Roma + IRT_Scores_Total_Cognitive_skill +
    married + child_under_6 + child_under_6_int + child_6_to_17 +
    child_6_to_17_int + highest_education + highest_education_int +
    age_factor + age_factor_int, family = binomial(link = "probit"),
    data = dat)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.946  -1.171   0.732   1.040   2.396

Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                     -0.904478   0.111688  -8.098 5.57e-16 ***
Roma                            -0.534931   0.377581  -1.417  0.15656
IRT_Scores_Total_Cognitive_skill  0.068687   0.021373   3.214  0.00131 **
married                          0.206226   0.052545   3.925 8.68e-05 ***
child_under_6                   -0.067678   0.065747  -1.029  0.30330
child_under_6_int               -0.188934   0.164979  -1.145  0.25213
child_6_to_17                    0.002807   0.050860   0.055  0.95598
child_6_to_17_int                0.234525   0.151610   1.547  0.12189
highest_education                0.160847   0.016350   9.838  < 2e-16 ***
highest_education_int            0.120396   0.068461   1.759  0.07864 .
age_factor                       0.066210   0.023172   2.857  0.00427 **
age_factor_int                  -0.083602   0.070923  -1.179  0.23849
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5175.2  on 3750  degrees of freedom
Residual deviance: 4812.9  on 3739  degrees of freedom
  (556 observations deleted due to missingness)
AIC: 4836.9

Number of Fisher Scoring iterations: 4
```

Out[87]:
```
Call:
glm(formula = employed ~ Roma + married + child_under_6 + child_under_6_int +

    child_6_to_17 + child_6_to_17_int + highest_education + highest_education
_int +
    age_factor + age_factor_int, family = binomial(link = "probit"),
    data = dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7698  -1.1615   0.7106   1.0439   2.3581

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)           -0.92550    0.10497  -8.817  < 2e-16 ***
Roma                  -0.58134    0.33096  -1.757   0.0790 .
married                0.22518    0.04896   4.600 4.23e-06 ***
child_under_6         -0.05167    0.06132  -0.843   0.3994
child_under_6_int     -0.11642    0.14387  -0.809   0.4184
child_6_to_17         -0.01710    0.04794  -0.357   0.7213
child_6_to_17_int      0.13656    0.13324   1.025   0.3054
highest_education      0.18096    0.01482  12.214  < 2e-16 ***
highest_education_int  0.10128    0.05933   1.707   0.0878 .
age_factor             0.04880    0.02177   2.242   0.0250 *
age_factor_int        -0.05504    0.06303  -0.873   0.3825
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5941.3  on 4298  degrees of freedom
Residual deviance: 5502.3  on 4288  degrees of freedom
  (8 observations deleted due to missingness)
AIC: 5524.3

Number of Fisher Scoring iterations: 4
```

Out[87]:
```
Call:
glm(formula = employed ~ Roma + NET_BENEFIT_WORKING + IRT_Scores_Total_Cognit
ive_skill +
    married + child_under_6 + child_6_to_17 + highest_education +
    age_factor, family = binomial(link = "probit"), data = dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9394  -1.1671   0.7199   1.0386   2.0865

Coefficients:
                                   Estimate Std. Error z value Pr(>|z|)
(Intercept)                      -0.7706860  0.2071604  -3.720 0.000199 ***
Roma                             -0.5276717  0.1183156  -4.460 8.20e-06 ***
NET_BENEFIT_WORKING              -0.0005448  0.0007163  -0.761 0.446876
IRT_Scores_Total_Cognitive_skill  0.0824275  0.0280456   2.939 0.003292 **
married                           0.2293771  0.0573597   3.999 6.36e-05 ***
child_under_6                    -0.1427793  0.0864317  -1.652 0.098548 .
child_6_to_17                     0.0056405  0.0594455   0.095 0.924406
highest_education                 0.2009488  0.0456392   4.403 1.07e-05 ***
age_factor                        0.0464699  0.0257168   1.807 0.070764 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5175.2  on 3750  degrees of freedom
Residual deviance: 4821.0  on 3742  degrees of freedom
  (556 observations deleted due to missingness)
AIC: 4839

Number of Fisher Scoring iterations: 4
```

In [107]:
```
#install.packages('mfx', dependencies=TRUE, repos='http://cran.rstudio.com/')
#library(mfx)
#MARGINAL EFFECTS
probitmfx(formula=employed ~ NET_BENEFIT_WORKING + Roma + IRT_Scores_Total_Cog
nitive_skill + married + child_under_6 + child_under_6_int + child_6_to_17 + c
hild_6_to_17_int + highest_education + highest_education_int + age_factor + ag
e_factor_int, data=dat)
probitmfx(formula=employed ~ Roma + NET_BENEFIT_WORKING + IRT_Scores_Total_Cog
nitive_skill + married + child_under_6 + child_6_to_17  + highest_education +
age_factor, data=dat)
```

Out[107]:
```
Call:
probitmfx(formula = employed ~ NET_BENEFIT_WORKING + Roma + IRT_Scores_Total_
Cognitive_skill +
    married + child_under_6 + child_under_6_int + child_6_to_17 +
    child_6_to_17_int + highest_education + highest_education_int +
    age_factor + age_factor_int, data = dat)

Marginal Effects:
                                        dF/dx   Std. Err.       z     P>|z|

NET_BENEFIT_WORKING              -0.00133601  0.00081598 -1.6373 0.1015686

Roma                             -0.40716431  0.14550666 -2.7983 0.0051380 **

IRT_Scores_Total_Cognitive_skill  0.06060305  0.02202803  2.7512 0.0059381 **

married                           0.12456003  0.03322093  3.7494 0.0001772 **
*
child_under_6                    -0.10482089  0.05393557 -1.9434 0.0519623 .

child_under_6_int                -0.28103060  0.12504790 -2.2474 0.0246155 *

child_6_to_17                    -0.05925583  0.04194721 -1.4126 0.1577650

child_6_to_17_int                 0.06144668  0.06169367  0.9960 0.3192518

highest_education                 0.14487784  0.05007473  2.8932 0.0038130 **

highest_education_int             0.03848801  0.02783075  1.3829 0.1666859

age_factor                       -0.00271670  0.02000880 -0.1358 0.8919988

age_factor_int                    0.01348825  0.04010083  0.3364 0.7366006

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

dF/dx is for discrete change for the following variables:

[1] "Roma"              "married"           "child_under_6"
[4] "child_under_6_int" "child_6_to_17"     "child_6_to_17_int"
```

Out[107]:
```
Call:
probitmfx(formula = employed ~ Roma + NET_BENEFIT_WORKING + IRT_Scores_Total_
Cognitive_skill +
    married + child_under_6 + child_6_to_17 + highest_education +
    age_factor, data = dat)

Marginal Effects:
                                        dF/dx   Std. Err.       z     P>|z|

Roma                                -0.20701044  0.04437911 -4.6646 3.092e-06 **
*
NET_BENEFIT_WORKING                 -0.00021621  0.00028425 -0.7606  0.446886

IRT_Scores_Total_Cognitive_skill  0.03271054  0.01112996  2.9390  0.003293 **

married                            0.09114306  0.02275118  4.0061 6.173e-05 **
*
child_under_6                      -0.05682646  0.03443262 -1.6504  0.098868 .

child_6_to_17                       0.00223815  0.02358522  0.0949  0.924397

highest_education                   0.07974451  0.01811293  4.4026 1.069e-05 **
*
age_factor                          0.01844113  0.01020509  1.8071  0.070754 .

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

dF/dx is for discrete change for the following variables:

[1] "Roma"          "married"       "child_under_6" "child_6_to_17"
```

In [ ]:
```
#ABOVE SHOWS THE LOWEST AND HIGHEST ESTIMATED ELASTICITY OF EMPLOYMENT FOR ROM
A
#BOTH RESULTS ARE SIGNIFICANT.
#THE ELASTICITY RANGES FROM -0.2 TO -0.4.
#IN BOTH SPECIFICATIONS THE ESIMATED EFFECT OF NET BENEFIT FROM WORKING IS EXT
REMELY SMALL AND INSIGNIFICANT
#IF THE METHODS ARE TAKEN AS CONVINCING, THESE RESULTS SUGGEST THAT BULGARIAN
 ROMA, CONTROLLING FOR LIKELY HIGHER MARGINAL TAXES
#FROM THE WELFARE SYSTEM AND OTHER RELEVANT ARE LESS LIEKLY TO BE EMPLOYED.
#BEING ROMA LOWER THE PROBABILITY OF BEING EMPLOYED, INDEPENDENT OF EDUCATION,
 NUMBER OF CHILDREN, AGE, MARTIAL STATUS, ETC,
#BY BETWEEN 20 AND 40 PERCENT.
```