Economics C142 Problem Set 3

Vinay Maruri

February 14, 2019

1 Question 1

1(a): Show that if x_i contains a constant, then $\bar{y} = \bar{x}'\hat{\beta}$, where $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ and $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$.

Using first order conditions, $E[x_i\hat{u}_i] = 0$ must hold.
$E[y_i] = E[x_i'\hat{\beta} + \hat{u}_i]$ (we take the expectation of the sample regression)
$E[y_i] = E[x_i'\hat{\beta}] + E[\hat{u}_i]$ (by linearity)
$E[y_i] = E[x_i'\hat{\beta}]$ (using the first order condition that must hold for sample regressions)
$\frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{N}\sum_{i=1}^{N} x_i'\hat{\beta}$ (take the expectation on both sides.)
This is equivalent to $\bar{y} = \bar{x}_i'\hat{\beta}$ since we have $\bar{y}$ and $\bar{x}'$ on the right hand and left hand sides. The first order condition and statement thus hold.

1(b): Show that if x_i contains a dummy variable for membership in group g (which has N_g observations in the sample) then $\bar{y}_g = \bar{x}_g\hat{\beta}$, where $\bar{y}_g = \frac{1}{N_g}\sum_{i\in g} y_i$ and $\bar{x}_g = \frac{1}{N_g}\sum_{i\in g} x_i$.

Using first order conditions, $\sum_{i=1}^{N} D_i(y_i - x_i'\hat{\beta}) = 0$ must hold.
When the dummy variable is turned off, $D_i = 0$ and this is trivially true. We must show that when the dummy variable is turned on, $\sum_{i=1}^{N} y_i - x_i'\hat{\beta} = 0$.
E$[y_i - x_i'\hat{\beta}] = \frac{1}{N_g}\sum_{i=1}^{N} y_i - x_i'\hat{\beta}$
$\frac{1}{N_g}\sum_{i=1}^{N} y_i - x_i'\hat{\beta} = 0$ (we want the expected value to be 0)
$\frac{1}{N_g}\sum_{i=1}^{N} y_i = \frac{1}{N_g}\sum_{i=1}^{N} x_i'\hat{\beta}$
This is equivalent to $\bar{y}_g = \bar{x}_g\hat{\beta}$, as we have $\bar{y}_g$ and $\bar{x}_g$ on the right and left hand sides. This ensures the first order condition is met and thus the statement holds.

1(c): Complete the proof of the Frisch-Waugh theorem for the sample OLS regression coefficients.
For sample regressions, this first order condition must hold: $\frac{1}{N}\sum_{i=1}^{N} x_i\hat{u}_i = 0$.
Let $\hat{u}_i = y_i - x_i'\hat{\beta}$. (manipulation of sample regression)
Additionally, from first order conditions, $\frac{1}{N}\sum_{i=1}^{N} x_{(\sim)ji}\hat{\xi}_i = 0$

Let $\hat{\xi_i} = x_{ji} - x_{(\sim)ji}\hat{\pi}$ (from auxiliary regression for frisch-waugh theorem in lecture 5)
Then: $y_i = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta_K} x_{Ki} + \hat{u}_i$
(this is just sample OLS)
Then we can write: $\frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i y_i = \frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i(\hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta_K} x_{Ki} + \hat{u}_i)$
Using the first order condition for ξ, $\hat{\xi_i} \perp x_{(\sim)ji}$.
Using the auxiliary regression, $\hat{\xi_i} = x_{ji} - x_{(\sim)ji}\hat{\pi}$, $\hat{\xi_i} \perp u_i$. This is because using the first order condition for x_i and $\hat{u}_i$, each $x_i \perp \hat{u}_i$. Thus, by extension, $\hat{\xi_i} \perp u_i$.
Thus, in $\frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i y_i$, every term is 0 except $\hat{\beta}_j x_{ji}$.
$\frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i y_i = \hat{\beta}_j \frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i x_{ji}$
Substitute x_{ji} for $(\hat{\xi_i} + x_{(\sim)ji}\hat{\pi})$:
$\frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i y_i = \hat{\beta}_j \frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i(\hat{\xi_i} + x_{(\sim)ji}\hat{\pi})$
Since $\hat{\xi_i} \perp x_{(\sim)ji}$:
$\frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i y_i = \hat{\beta}_j \frac{1}{N}\sum_{i=1}^{N}\hat{\xi_i}^2$
Rearranging this equation gives us the final result: $\hat{\beta}_j = (\frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i y_i)\,(\frac{1}{N}\sum_{i=1}^{N}\hat{\xi_i}^2)^{-1}$

2 Question 2

2(a): Write an expression for the OLS estimate of the coefficient on immigrant statues from logwage = constant, immigrant statues, if the true model is logwage = constant, education, immigrant status.

True Model: $\log(wage) = \epsilon_i + \beta_1 Educ_i + \beta_2 Imm_i$

ϵ_i is the constant term, $Educ_i$ is the education variable, and Imm_i is the immigration variable.

Model (1): $\log(wage) = \epsilon_i + \beta_1 Imm_i$

ϵ_i is the constant term and Imm_i is the immigration variable.

An auxiliary regression we can run to OLS estimate the true coefficient of immigrant status would be:

$Educ_i = \epsilon_i + \pi_1 Imm_i$

where $Educ_i$ is the missing education variable, ϵ_i is the error term, Imm_i is the immigration variable, and π_1 is the effect of immigration on education.

We can then use Frisch-Waugh Theorem with the true model as follows:

$$\log(wage) = \epsilon_i + \beta_1 Educ_i + \beta_2 Imm_i$$

$$\log(wage) = \epsilon_i + \beta_1(\epsilon_i + \pi_1 Imm_i) + \beta_2 Imm_i$$

$$\log(wage) = \epsilon_i + \beta_1\epsilon_i + \beta_1\pi_1 Imm_i + \beta_2 Imm_i$$

$$\log(wage) = (1 + \beta_1)\epsilon_i + (\beta_1\pi_1 + \beta_2)Imm_i$$

Thus, the OLS estimate of the coefficient on immigrant status from model (1) if the true model is (5) is $(\beta_1\pi_1 + \beta_2)$.

2(b): Using a regression package, estimate the 5 models, and show the values of the terms for part (a), first for females, then for males. Your answers for females should be the same as the ones reported in the table in Lecture 5.
Code is listed here. Regression table of coefficients follows this.

```
```{r}
#starting libraries.
library(dplyr)
library(ggplot2)
library(magrittr)
library(reshape2)
library(stargazer)
library(lubridate)
library(lmtest)
library(ivpack)
library(kableExtra)
```
```{r}
#importing data
data_raw <- read.csv("/Users/EndlessWormhole/Desktop/Spring_2019/Econ_C142/
problem_set_3/ovb.csv")
```

```{r}
#split the data into male and female groups.
female <- data_raw %>% dplyr::filter(female == 1)
male <- data_raw %>% dplyr::filter(female == 0)
```

```{r}
#running model 1: constant, immigrant status on logwage.
#first females
model1female <- lm(logwage ~ imm, data = female)
```
```

```
```{r}
#next model 1 for males
model1male <- lm(logwage ~ imm, data = male)
```

```{r}
#the next model we will run is model 2: constant and education on logwage
#first females again
model2female <- lm(logwage ~ educ, data = female)
```
```{r}
#then males again
model2male <- lm(logwage ~ educ, data = male)
```

```{r}
#now we run model 3: constant and education on immigrant status.
#first females
model3female <- lm(imm ~ educ, data = female)
```

```{r}
#then males
model3male <- lm(imm ~ educ, data = male)
```

```{r}
#next we will run model 4: constant and immmigrant status on education
#first females
model4female <- lm(educ ~ imm, data = female)
```

```{r}
#then males
model4male <- lm(educ ~ imm, data = male)
```

```{r}
#finally we will run model 5: constant, immigrant status, and education
#on logwage
#first females
model5female <- lm(logwage ~ educ + imm, data = female)
```

```{r}
#then males
```
```

```
model5male <- lm(logwage ~ educ + imm, data = male)
```

```{r}
#I am choosing to report my output in a stargazer table for conciseness of code
and the submission document. Also I think it
#reports the coefficients and assorted test statistics more nicely.
#females first.
stargazer(model1female,
          model2female,
          model3female,
          model4female,
          model5female,
          title = "Estimation of the 5 models specified in
          Question 2(a) for Females",
          align = TRUE,
          dep.var.labels = c("logwage", "immigrant", "education"),
          covariate.labels = c("immigrant", "education"),
          omit.stat = c("ll"),
          column.sep.width = '0.1pt',
          no.space = TRUE,
          single.row = TRUE,
          multicolumn = F,
          header = F,
          font.size = "tiny")
```

```
the condition has length > 1 and only the first element will be usedthe conditio
\begin{table}[!htbp] \centering
  \caption{Estimation of the 5 models specified in Question 2(a) for Females}
  \label{}
\tiny
\begin{tabular}{@{\extracolsep{0.1pt}}lD{.}{.}{-3} D{.}{.}{-3} D{.}{.}{-3} D{.}{.
\\[-1.8ex]\hline
\hline \\[-1.8ex]
 & \multicolumn{5}{c}{\textit{Dependent variable:}} \\
\cline{2-6}
\\[-1.8ex] & \multicolumn{1}{c}{logwage} & \multicolumn{1}{c}{immigrant} & \mult
\\[-1.8ex] & \multicolumn{1}{c}{(1)} & \multicolumn{1}{c}{(2)} & \multicolumn{1}
\hline \\[-1.8ex]
 immigrant & -0.180^{***}$ $(0.017) & & & -1.492^{***}$ $(0.067) & -0.010$ $(0
  education & & 0.114^{***}$ $(0.002) & -0.030^{***}$ $(0.001) &
& 0.114^{***}$ $(0.002) \\
  Constant & 2.886^{***}$ $(0.007) & 1.235^{***}$ $(0.030) & 0.607^{***}$ $(0.01
 \hline \\[-1.8ex]
Observations & \multicolumn{1}{c}{10,601} & \multicolumn{1}{c}{10,601} & \multic
R$^{2}$ & \multicolumn{1}{c}{0.011} & \multicolumn{1}{c}{0.224} & \multicolumn{1
```

```
Adjusted R$^{2}$ & \multicolumn{1}{c}{0.011} & \multicolumn{1}{c}{0.224} & \mult
Residual Std. Error & \multicolumn{1}{c}{0.664 (df = 10599)} & \multicolumn{1}{c
F Statistic & \multicolumn{1}{c}{118.530$^{***}$ (df = 1; 10599)} & \multicolumn
\hline
\hline \\[-1.8ex]
\textit{Note:} & \multicolumn{5}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<
\end{tabular}
\end{table}
length of NULL cannot be changedlength of NULL cannot be changedlength of NULL c

```{r}
#males next.
stargazer(model1male,
          model2male,
          model3male,
          model4male,
          model5male,
          title = "Estimation of the 5 models specified in
          Question 2(a) for Males",
          align = TRUE,
          dep.var.labels = c("logwage", "immigrant", "education"),
          covariate.labels = c("immigrant", "education"),
          omit.stat = c("ll"),
          column.sep.width = '0.1pt',
          no.space = TRUE,
          single.row = TRUE,
          multicolumn = F,
          header = F,
          font.size = "tiny")

length of NULL cannot be changedlength of NULL cannot be changedlength of NULL c
\begin{table}[!htbp] \centering
  \caption{Estimation of the 5 models specified in Question 2(a) for Males}
  \label{}
\tiny
\begin{tabular}{@{\extracolsep{0.1pt}}lD{.}{.}{-3} D{.}{.}{-3} D{.}{.}{-3} D{.}{.
\\[-1.8ex]\hline
\hline \\[-1.8ex]
 & \multicolumn{5}{c}{\textit{Dependent variable:}} \\
\cline{2-6}
\\[-1.8ex] & \multicolumn{1}{c}{logwage} & \multicolumn{1}{c}{immigrant} & \mult
\\[-1.8ex] & \multicolumn{1}{c}{(1)} & \multicolumn{1}{c}{(2)} & \multicolumn{1}
\hline \\[-1.8ex]
 immigrant & -0.245^{***}$ $(0.016) & & & -1.612^{***}$ $(0.067) & -0.075^{***
  education & & 0.108^{***}$ $(0.002) & -0.031^{***}$ $(0.001) &
& 0.106^{***}$ $(0.002) \\
```
```

```
  Constant & 3.156^{***}$ $(0.007) & 1.609^{***}$ $(0.027) & 0.641^{***}$ $(0.01
 \hline \\[-1.8ex]
Observations & \multicolumn{1}{c}{11,306} & \multicolumn{1}{c}{11,306} & \multico
R$^{2}$ & \multicolumn{1}{c}{0.021} & \multicolumn{1}{c}{0.219} & \multicolumn{1
Adjusted R$^{2}$ & \multicolumn{1}{c}{0.021} & \multicolumn{1}{c}{0.219} & \mult
Residual Std. Error & \multicolumn{1}{c}{0.684 (df = 11304)} & \multicolumn{1}{c
F Statistic & \multicolumn{1}{c}{246.668$^{***}$ (df = 1; 11304)} & \multicolumn
\hline
\hline \\[-1.8ex]
\textit{Note:} & \multicolumn{5}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<
\end{tabular}
\end{table}
length of NULL cannot be changedlength of NULL cannot be changedlength of NULL c

```
```
```

Table 1: Estimation of the 5 models specified in Question 2(a) for Females

	Dependent variable:				
	logwage	immigrant	education	educ	logwage
	(1)	(2)	(3)	(4)	(5)
immigrant	-0.180^{***} (0.017)			-1.492^{***} (0.067)	-0.010 (0.015)
education		0.114^{***} (0.002)	-0.030^{***} (0.001)		0.114^{***} (0.002)
Constant	2.886^{***} (0.007)	1.235^{***} (0.030)	0.607^{***} (0.019)	14.452^{***} (0.029)	1.241^{***} (0.031)
Observations	10,601	10,601	10,601	10,601	10,601
R^2	0.011	0.224	0.044	0.044	0.224
Adjusted R^2	0.011	0.224	0.044	0.044	0.224
Residual Std. Error	0.664 (df = 10599)	0.588 (df = 10599)	0.381 (df = 10599)	2.706 (df = 10599)	0.588 (df = 10598)
F Statistic	118.530^{***} (df = 1; 10599)	$3,057.517^{***}$ (df = 1; 10599)	490.608^{***} (df = 1; 10599)	490.608^{***} (df = 1; 10599)	$1,528.907^{***}$ (df = 2; 10598)
Note:					$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2: Estimation of the 5 models specified in Question 2(a) for Males

	Dependent variable:				
	logwage	immigrant	education	educ	logwage
	(1)	(2)	(3)	(4)	(5)
immigrant	-0.245^{***} (0.016)			-1.612^{***} (0.067)	-0.075^{***} (0.014)
education		0.108^{***} (0.002)	-0.031^{***} (0.001)		0.106^{***} (0.002)
Constant	3.156^{***} (0.007)	1.609^{***} (0.027)	0.641^{***} (0.018)	14.189^{***} (0.031)	1.657^{***} (0.029)
Observations	11,306	11,306	11,306	11,306	11,306
R^2	0.021	0.219	0.049	0.049	0.221
Adjusted R^2	0.021	0.219	0.049	0.049	0.221
Residual Std. Error	0.684 (df = 11304)	0.611 (df = 11304)	0.403 (df = 11304)	2.925 (df = 11304)	0.611 (df = 11303)
F Statistic	246.668^{***} (df = 1; 11304)	$3,167.320^{***}$ (df = 1; 11304)	585.605^{***} (df = 1; 11304)	585.605^{***} (df = 1; 11304)	$1,600.999^{***}$ (df = 2; 11303)
Note:					$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

2(c): Consider 3 groups of immigrants: Asian immigrants are those with (asian=1) and (hispanic=0) and (imm=1). Hispanic immigrants are those with (hispanic=1) and (imm=1). Other immigrants are those with (imm=1) who are not included in the First 2 groups. Redo the 5 models for females and for males, distinguishing the 3 groups of immigrants. So your models will have 3 separate dummies for the 3 immigrant groups, treating natives as the omitted group. Put your results in 2 new tables that are similar to the table in Lecture 5, and include these tables in your answers. **Code is listed here. Regression tables follow this.**

```
```{r}
#we generate a split of the data that creates new variables for asian, hispanic,
ovb_split_race_male <- data_raw %>% mutate(
   asiani = as.numeric(asian == 1 & imm == 1 & hispanic == 0 & black == 0),
   hispanici = as.numeric(asian==0 & hispanic == 1 & black == 0 & imm == 1),
   otheri = as.numeric(asian == 0 & hispanic == 0 & black == 1 & imm == 1))
  %>% dplyr::filter(female == 0)
```

```{r}
#we then do the same for female immigrants.
ovb_split_race_female <- data_raw %>% mutate(
   asiani = as.numeric(asian == 1 & imm == 1 & hispanic == 0 & black == 0),
   hispanici = as.numeric(asian==0 & hispanic == 1 & black == 0 & imm == 1),
   otheri = as.numeric(asian == 0 & hispanic == 0 & black == 1 & imm == 1))
  %>% dplyr::filter(female == 1)
```

```{r}
#now we re-do the models using the 3 new variables.
#first, we do model 1: constant and immigrant status on logwage.
#first females.
model1immfemale <- lm(logwage ~ asiani + hispanici + otheri,
data = ovb_split_race_female)
#next males
model1immmale <- lm(logwage ~ asiani + hispanici + otheri,
data = ovb_split_race_male)
```


```{r}
#second, we do model 2: constant and education on logwage
#first females.
model2immfemale <- lm(logwage ~ educ + asiani + hispanici + otheri,
data = ovb_split_race_female)
```
```

```
#next males.
model2immmale <- lm(logwage ~ educ + asiani + hispanici + otheri,
data = ovb_split_race_male)
```


```{r}
#third, we do model 3: constant and education on immigrant status
#first females
model3immfemale <- lm(imm ~ educ + asiani + hispanici + otheri,
data = ovb_split_race_female)
#then males.
model3immmale <- lm(imm ~ educ + asiani + hispanici + otheri,
data = ovb_split_race_male)
```


```{r}
#fourth we do model 4: constant and immigrant status on education
#first females
model4immfemale <- lm(educ ~ asiani + hispanici + otheri,
data = ovb_split_race_female)
#next males
model4immmale <- lm(educ ~ asiani + hispanici + otheri,
data = ovb_split_race_male)
```


```{r}
#finally we do model 5: constant, education and immigrant status on logwage
#first females
model5immfemale <- lm(logwage ~ asiani + hispanici + otheri + educ,
data = ovb_split_race_female)
#next males
model5immmale <- lm(logwage ~ asiani + hispanici + otheri + educ,
data = ovb_split_race_male)
```


```{r}
#I report the results of this new class of models here.
#first males.
stargazer(model1immmale,
          model2immmale,
          model3immmale,
          model4immmale,
```

```
            model5immmale,
            title =
            "Estimation of the 5 new models specified in Question 2(c) for Males",
            align = TRUE,
            dep.var.labels = c("logwage", "immigrant", "education"),
            covariate.labels = c("immigrant", "education", "asianimmigrant",
            "hispanicimmigrant", "otherimmigrant"),
            omit.stat = c("ll"),
            column.sep.width = '0.1pt',
            no.space = TRUE,
            single.row = TRUE,
            multicolumn = F,
            header = F,
            font.size = "tiny")
```
```
the condition has length > 1 and only the first element will be usedthe conditio
\begin{table}[!htbp] \centering
  \caption{Estimation of the 5 new models specified in Question 2(c) for Males}
  \label{}
\tiny
\begin{tabular}{@{\extracolsep{0.1pt}}lD{.}{.}{-3} D{.}{.}{-3} D{.}{.}{-3} D{.}{.
\\[-1.8ex]\hline
\hline \\[-1.8ex]
 & \multicolumn{5}{c}{\textit{Dependent variable:}} \\
\cline{2-6}
\\[-1.8ex] & \multicolumn{1}{c}{logwage} & \multicolumn{1}{c}{immigrant} & \mult
\\[-1.8ex] & \multicolumn{1}{c}{(1)} & \multicolumn{1}{c}{(2)} & \multicolumn{1}
\hline \\[-1.8ex]
 immigrant & & 0.104^{***}$ $(0.002) & 0.003^{***}$ $(0.001) &
& 0.104^{***}$ $(0.002) \\
  education & 0.069^{**}$ $(0.030) & -0.059^{**}$ $(0.027) & 0.950^{***}$ $(0.00
  asianimmigrant & -0.474^{***}$ $(0.020) & -0.090^{***}$ $(0.019) & 0.965^{***}
  hispanicimmigrant & -0.253^{***}$ $(0.056) & -0.215^{***}$ $(0.051) & 0.955^{*
  otherimmigrant & 3.159^{***}$ $(0.007) & 1.673^{***}$ $(0.031) & 0.001$ $(0.01
 \hline \\[-1.8ex]
Observations & \multicolumn{1}{c}{11,306} & \multicolumn{1}{c}{11,306} & \multic
R$^{2}$ & \multicolumn{1}{c}{0.051} & \multicolumn{1}{c}{0.222} & \multicolumn{1
Adjusted R$^{2}$ & \multicolumn{1}{c}{0.051} & \multicolumn{1}{c}{0.221} & \mult
Residual Std. Error & \multicolumn{1}{c}{0.674 (df = 11302)} & \multicolumn{1}{c
F Statistic & \multicolumn{1}{c}{204.065$^{***}$ (df = 3; 11302)} & \multicolumn
\hline
\hline \\[-1.8ex]
\textit{Note:} & \multicolumn{5}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<
\end{tabular}
\end{table}
length of NULL cannot be changedlength of NULL cannot be changedlength of NULL c
```

```{r}
#next females.
#first males.
stargazer(model1immfemale,
          model2immfemale,
          model3immfemale,
          model4immfemale,
          model5immfemale,
          title = "Estimation of the 5 new models specified in Question 2(c) for
          align = TRUE,
          dep.var.labels = c("logwage", "immigrant", "education"),
          covariate.labels = c("immigrant", "education", "asianimmigrant", "hisp
          omit.stat = c("ll"),
          column.sep.width = '0.1pt',
          no.space = TRUE,
          single.row = TRUE,
          multicolumn = F,
          header = F,
          font.size = "tiny")
```

```
the condition has length > 1 and only the first element will be usedthe conditio
\begin{table}[!htbp] \centering
  \caption{Estimation of the 5 new models specified in Question 2(c) for Females
  \label{}
\tiny
\begin{tabular}{@{\extracolsep{0.1pt}}lD{.}{.}{-3} D{.}{.}{-3} D{.}{.}{-3} D{.}{.
\\[-1.8ex]\hline
\hline \\[-1.8ex]
 & \multicolumn{5}{c}{\textit{Dependent variable:}} \\
\cline{2-6}
\\[-1.8ex] & \multicolumn{1}{c}{logwage} & \multicolumn{1}{c}{immigrant} & \mult
\\[-1.8ex] & \multicolumn{1}{c}{(1)} & \multicolumn{1}{c}{(2)} & \multicolumn{1}
\hline \\[-1.8ex]
 immigrant &  & 0.112^{***}$ $(0.002) & 0.001$ $(0.001) &
& 0.112^{***}$ $(0.002) \\
  education & 0.083^{***}$ $(0.029) & 0.027$ $(0.026) & 0.959^{***}$ $(0.008) &
  asianimmigrant & -0.441^{***}$ $(0.022) & -0.058^{***}$ $(0.021) & 0.962^{***}
  hispanicimmigrant & -0.091^{*}$ $(0.054) & -0.054$ $(0.048) & 0.959^{***}$ $(0
  otherimmigrant & 2.889^{***}$ $(0.007) & 1.272^{***}$ $(0.033) & 0.031^{***}$
 \hline \\[-1.8ex]
Observations & \multicolumn{1}{c}{10,601} & \multicolumn{1}{c}{10,601} & \multic
R$^{2}$ & \multicolumn{1}{c}{0.037} & \multicolumn{1}{c}{0.225} & \multicolumn{1
Adjusted R$^{2}$ & \multicolumn{1}{c}{0.037} & \multicolumn{1}{c}{0.224} & \mult
Residual Std. Error & \multicolumn{1}{c}{0.655 (df = 10597)} & \multicolumn{1}{c
F Statistic & \multicolumn{1}{c}{135.564$^{***}$ (df = 3; 10597)} & \multicolumn
\hline
```

```
\hline \\[-1.8ex]
\textit{Note:} & \multicolumn{5}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<
\end{tabular}
\end{table}
length of NULL cannot be changedlength of NULL cannot be changedlength of NULL c
```

Table 3: Estimation of the 5 new models specified in Question 2(c) for Males

	Dependent variable:				
	logwage	immigrant	education	educ	logwage
	(1)	(2)	(3)	(4)	(5)
immigrant		0.104^{***} (0.002)	0.003^{***} (0.001)		0.104^{***} (0.002)
education	0.069^{**} (0.030)	-0.059^{**} (0.027)	0.950^{***} (0.008)	1.226^{***} (0.120)	-0.059^{**} (0.027)
asianimmigrant	-0.474^{***} (0.020)	-0.090^{***} (0.019)	0.965^{***} (0.006)	-3.676^{***} (0.080)	-0.090^{***} (0.019)
hispanicimmigrant	-0.253^{***} (0.056)	-0.215^{***} (0.051)	0.955^{***} (0.016)	-0.368 (0.226)	-0.215^{***} (0.051)
otherimmigrant	3.159^{***} (0.007)	1.673^{***} (0.031)	0.001 (0.010)	14.220^{***} (0.028)	1.673^{***} (0.031)
Observations	11,306	11,306	11,306	11,306	11,306
R^2	0.051	0.222	0.787	0.170	0.222
Adjusted R^2	0.051	0.221	0.787	0.170	0.221
Residual Std. Error	0.674 (df = 11302)	0.610 (df = 11301)	0.191 (df = 11301)	2.733 (df = 11302)	0.610 (df = 11301)
F Statistic	204.065^{***} (df = 3; 11302)	804.838^{***} (df = 4; 11301)	$10{,}432.160^{***}$ (df = 4; 11301)	773.137^{***} (df = 3; 11302)	804.838^{***} (df = 4; 11301)
Note:					$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 4: Estimation of the 5 new models specified in Question 2(c) for Females

	Dependent variable:				
	logwage	immigrant	education	educ	logwage
	(1)	(2)	(3)	(4)	(5)
immigrant		0.112^{***} (0.002)	0.001 (0.001)		0.112^{***} (0.002)
education	0.083^{***} (0.029)	0.027 (0.026)	0.959^{***} (0.008)	0.504^{***} (0.116)	0.027 (0.026)
asianimmigrant	-0.441^{***} (0.022)	-0.058^{***} (0.021)	0.962^{***} (0.007)	-3.421^{***} (0.088)	-0.058^{***} (0.021)
hispanicimmigrant	-0.091^{*} (0.054)	-0.054 (0.048)	0.959^{***} (0.015)	-0.338 (0.213)	-0.054 (0.048)
otherimmigrant	2.889^{***} (0.007)	1.272^{***} (0.033)	0.031^{***} (0.010)	14.458^{***} (0.027)	1.272^{***} (0.033)
Observations	10,601	10,601	10,601	10,601	10,601
R^2	0.037	0.225	0.782	0.128	0.225
Adjusted R^2	0.037	0.224	0.782	0.127	0.224
Residual Std. Error	0.655 (df = 10597)	0.588 (df = 10596)	0.182 (df = 10596)	2.585 (df = 10597)	0.588 (df = 10596)
F Statistic	135.564^{***} (df = 3; 10597)	767.302^{***} (df = 4; 10596)	$9{,}524.920^{***}$ (df = 4; 10596)	516.448^{***} (df = 3; 10597)	767.302^{***} (df = 4; 10596)
Note:					$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$