Applied Econometrics II

Project 1 Part 2

1. The coefficient of *treatment* in Model 4 after the Omitted Variable Bias being fixed is 5097.44 and its standard error is 96.35 which makes the model statistically significant at 1% level. The interpretation for *treatment*’s coefficient in Model 4 would be on average, getting an advanced IT training increases an individual’s income by $5097.44 holding everything else constant. I do believe the coefficient is economically significant as it is depicting the true impact of training on an individual’s income. Also, the sign of this coefficient is positive unlike the sign of the coefficient of *treatment* in Model 2 when the Omitted Variable Bias is not fixed for. This depicts that individuals who are getting an advanced IT training earn more, regardless of considering ability attribute in the model. The coefficient of *treatment* in Model 4 is correct and in Model 2 is incorrect.
2. The first difference strategy is used in this model to resolve the omitted variable bias observed in Model 2. In this model we are comparing 2 time periods (2010 and 2009). We could observe the changes in dependent variables in both the years, and control for all the variables that are constant over the same time period (1 year). This way we can hold constant the unobserved factors that differ from one state to other but don’t change within. As seen in table 1, we have eliminated all variables except ‘weeks worked’ as that is the only variable changing from 2009 to 2010. Once this is done, when we run the regression, we can see the coefficient for treatment being positive which resolves the issue of omitted variable bias in model 2.

Below are the equations for solving omitted variable bias. Here, change in treatment would be the treatment given in the year 2010 as there was no treatment given in 2009 so that variable would be 0.

This would conclude that individuals who are given an IT training earn higher than those are not given the training. In model 4, resolving the omitted variable bias by using first-difference strategy proves to be a correct way of showing the true impact of giving IT training on individuals’ incomes.

|  |  |  |
| --- | --- | --- |
| **Table 1: Determinants of Income (Categorical & Interaction Variables)** | | |
|  |  |  |
| **Regressors** | **Model2** | **Model4** |
| Intercept | 117,818.05\*\*\* | -45.95 |
|  | (507.98) | (63.69) |
| Treatment | -5,884.56\*\*\* | 5,097.44\*\*\* |
|  | (81.16) | (96.35) |
| Change in weeks worked |  | 8.81 |
|  |  | (38.61) |
| Raceethnic, Gender, State, Occupation, and Industry Controlled? | YES | YES |
| Number of Weeks Worked Last Year | 937.52\*\*\* |  |
|  | (1.87) |  |
| Number of Obs | 1,788,022 | 1,788,022 |
| Adjusted R-sq | 0.5286 | 0.001563 |
| Overall Significance | 20,269.07\*\*\* | 1,399.64\*\*\* |
|  |  |  |
| Note: robust standard errors are in parentheses. \*, \*\*, and \*\*\* indicate 10%, 5%, and 1% significance levels, respectively. | | |

Appendix

SAS Code

/\*Calling the library\*/

libname AEData "~/my\_shared\_file\_links/u47408605/Data"

access=readonly;

run;

/\*Merged data\*/

data Merged\_data;

merge AEData.db1 AEData.db2 ;

by id;

run;

data Merged\_data1;

set AEData.db4 (rename=(pre\_treatment\_income=income))Merged\_data ;

run;

/\*Sorting data\*/

proc sort data=merged\_data1;

b/y id year;

run;

/\*Retrieving 20 observations from the merged data\*/

Proc Print data=Merged\_data1(obs=20);

run;

/\* Model 2 from Project 1\_Part 1\*/

ods output ParameterEstimates=PEforModel2 DataSummary=ObsModel2 FitStatistics=AdjRsqModel2 Effects=OverallSigModel2;

Proc SurveyReg data=Merged\_data plots=none;

Class raceethnic ed occ ind statefips / Ref=first;

Model2: Model income=male raceethnic ed wkswork occ ind statefips male\*raceethnic male\*ed treatment/Solution Adjrsq;

run;

/\*Resolving OVB\*/

data Stacked\_data;

set Merged\_data1 (where=(year=2009 or year=2010)); /\*having where statement here - if where is put as independent line, SAS will go through all the steps and in the end won't read\*/

Keep year income treatment LagIncome LagTreatment Lagstatefips Lagmale Lagraceethnic Laged Lagocc Lagind Lagwkswork CHIncome CHTreatment CHstatefips CHmale CHraceethnic CHed CHocc CHind CHwkswork; /\*don't want to run loops on every observation so use where dependently\*/

Retain LagIncome LagTreatment Lagstatefips Lagmale Lagraceethnic Laged Lagocc Lagind Lagwkswork LagId;

IF id=Lagid then do; /\*if the condition is met, do everything between do and end\*/

CHstatefips=Statefips-LagStatefips;

CHmale=Male-Lagmale;

CHraceethnic=raceethnic-Lagraceethnic;

CHed=ed-Laged;

CHocc=occ-Lagocc;

CHind=ind-Lagind;

CHwkswork=wkswork-Lagwkswork;

CHTreatment=Treatment-LagTreatment;

CHIncome=Income-LagIncome;

output;

end;

Lagid=Id;

LagIncome=Income;

LagTreatment=Treatment;

Lagstatefips=statefips;

Lagmale=male;

Lagraceethnic=raceethnic;

Laged=ed;

Lagocc=occ;

Lagind=ind;

Lagwkswork=wkswork;

Run;

/\*Finding the changes in variables\*/

proc means data=Stacked\_data;

run;

/\*Regression\*/

ods output ParameterEstimates=PEforModel4 DataSummary=ObsModel4 FitStatistics=AdjRsqModel4 Effects=OverallSigModel4;

Proc SurveyReg data=Stacked\_data plots=none;

Model4: Model chincome=treatment chwkswork/Solution Adjrsq;

run;

/\*Cleaning up the output of the regression analysis\*/

Data Table\_Long\_Project;

length Model $10;

length Parameter $30;

set PEforModel2 PEforModel4 indsname=M;

keep Model Parameter EditedResults;

if M="WORK.PEFORMODEL2" then Model="Model2";

else if M="WORK.PEFORMODEL4" then Model="Model4";

Where Estimate ne 0;

if Probt le 0.01 then Star="\*\*\*";

else if Probt le 0.05 then Star="\*\*";

else if Probt le 0.1 then Star="\*";

Results=Estimate;

EditedResults=Cats(put(Results,comma16.2),Star);

output;

Results=stderr;

EditedResults=Cats("(",put(Results,comma16.2),")");

output;

run;

/\*Sorting the table\*/

proc sort data=Table\_Long\_Project out=Table\_Long\_Sorted;

by Model Parameter;

run;

/\*Creating separate results columns corresponding to each model\*/

data Model2Results(rename=(EditedREsults=Model2))

Model4Results(rename=(EditedREsults=Model4));

set Table\_Long\_Sorted;

if Model="Model2" then output Model2Results;

else if Model="Model4" then output Model4Results;

drop Model;

run;

/\*Creating the final results table that would include all models side-by-side\*/

data Table\_Wide;

merge Model2Results Model4Results;

by Parameter;

if mod(\_n\_,2)=1 then Regressors=Parameter;

length Order 3;

if Parameter="Intercept" then Order=1;

else if substr(Parameter,1,10)= "treatment " then Order =2;

else if substr(Parameter,1,10)="CHwkswork " then order = 3;

else if Parameter="male" then Order=4;

else if substr(Parameter,1,11)="raceethnic " then Order=5;

else if substr(Parameter,1,3)="ed " then Order=6;

else if substr(Parameter,1,15)="male\*raceethnic" then Order=7;

else if substr(Parameter,1,7)="male\*ed" then Order=8;

else if substr(Parameter,1,13)="raceethnic\*ed" then Order=9;

else if substr(Parameter,1,4)="occ " then Order =10;

else if substr(Parameter,1,4)="ind " then Order = 11;

else if substr(Parameter,1,10)= "statefips " then Order =12;

else if Parameter="wkswork " then Order = 13;

else Order=100;

run;

/\*Ordering the variables in the results table\*/

proc sort data=Table\_Wide out=Table\_Wide\_Sorted(drop=Order Parameter);

by Order;

run;

/\*Step 4: Create the rows for other statistics\*/

/\* The row for Number of Obs \*/

data NumofObs(keep=Label1 Model2 Model4 );

merge ObsModel2(rename=(nvalue1=NVMoel2)) ObsModel4(rename=(nvalue1=NVMoel4));

by Label1;

where Label1="Number of Observations";

Model2=put(NVMoel2,comma16.0);

Model4=put(NVMoel4,comma16.0);

run;

/\* The row for Adj R-sq \*/

Data AdjRsq;

merge AdjRsqModel2(rename=(cvalue1=Model2)) AdjRsqModel4(rename=(cvalue1=Model4)) ;

by Label1;

Where Label1="Adjusted R-Square";

drop nvalue1;

run;

/\* The row for Overall Significance \*/

data OSM2(rename=(EditedValue=Model2)) OSM4(rename=(EditedValue=Model4)) ;

set OverallSigModel2 OverallSigModel4 indsname=M;

Where Effect="Model";

Label1="Overall Significance";

if ProbF le 0.01 then Star="\*\*\*";

else if ProbF le 0.05 then Star="\*\*";

else if ProbF le 0.1 then Star="\*";

EditedValue=Cats(Put(FValue,comma16.2),Star);

if M="WORK.OVERALLSIGMODEL2" then output OSM2;

else if M="WORK.OVERALLSIGMODEL4" then output OSM4;

keep Label1 EditedValue;

run;

Data OverallSig;

merge OSM2 OSM4 ;

by Label1;

run;

/\* Combining all rows for other statistics \*/

Data OtherStat;

set NumofObs AdjRsq OverallSig;

rename Label1=Regressors;

Run;

/\* Adding other statistics to the results table \*/

Data Table\_Wide\_Sorted\_WithStat;

set Table\_Wide\_Sorted OtherStat;

run;

/\* creating new name for variables in the regression results table through defining a new format\*/

proc format;

value $VariableName(default=50) "treatment"="Treatment"

"CHwkswork" = "Change in weeks worked"

"wkswork"="Number of Weeks Worked LAst Year"

"Number of Observations"="Number of Obs"

"Adjusted R-Square"="Adjusted R-sq"

"male"="Male"

"raceethnic 2"="Black"

"raceethnic 3"="Asian"

"raceethnic 4"="Hispanic"

"raceethnic 5"="Other "

"ed 2"="High School Graduate"

"ed 3"="Some College"

"ed 4"="College Graduate"

"ed 5"="Graduate Degree "

"occ 2"="Management Related Occupations"

"occ 3"="Professional Specialty Occupations "

"occ 4"="Technicians "

"occ 5"="Sales "

"occ 6"="Administrative Support Occupations"

"occ 7"="Service Occupations "

"occ 8"="Farming, Forestry And Fishing"

"occ 9"="Supervisors Of Production Workers"

"occ 10"="Mechanics and repairers "

"occ 11"="Construction"

"occ 12"="Other production workers "

"occ 13"="Machine operators "

"occ 14"="Motor vehicle operators "

"occ 15"="Other transportation"

"occ 16"="Military"

"ind 2"="Mining"

"ind 3"="Construction"

"ind 4"="Nondurable goods manufacturing"

"ind 5"="Durable goods manufacturing "

"ind 6"="Transportation, Communications and Utilities "

"ind 7"="Wholesale trade"

"ind 8"="Retail trade"

"ind 9"="Finance, insurance and real estate"

"ind 10"="Business and repair services"

"ind 11"="Personal services "

"ind 12"="Entertainment and recreation services "

"ind 13"="Professional and related services"

"ind 14"="Public administration 15=Military"

"male\*raceethnic 2"="Male\*Black"

"male\*raceethnic 3"="Male\*Asian"

"male\*raceethnic 4"="Male\*Hispanic"

"male\*raceethnic 5"="Male\*Other "

"male\*ed 2"="Male\*High School Graduate"

"male\*ed 3"="Male\*Some College"

"male\*ed 4"="Male\*College Graduate"

"male\*ed 5"="Male\*Graduate Degree "

"statefips 2"="Alaska"

"statefips 4"="Arizona"

"statefips 5"="Arkansas"

"statefips 6"="California"

"statefips 8"="Colorado"

"statefips 9"="Connecticut"

"statefips 10"="Delaware"

"statefips 11"="District of Columbia"

"statefips 12"="Florida"

"statefips 13"="Georgia"

"statefips 15"="Hawaii"

"statefips 16"="Idaho"

"statefips 17"="Illinois"

"statefips 18"="Indiana"

"statefips 19"="Iowa"

"statefips 20"="Kansas"

"statefips 21"="Kentucky"

"statefips 22"="Louisiana"

"statefips 23"="Maine"

"statefips 24"="Maryland"

"statefips 25"="Massachusetts"

"statefips 26"="Michigan"

"statefips 27"="Minnesota"

"statefips 28"="Mississippi"

"statefips 29"="Missouri"

"statefips 30"="Montana"

"statefips 31"="Nebraska"

"statefips 32"="Nevada"

"statefips 33"="New Hampshire"

"statefips 34"="New Jersey"

"statefips 35"="New Mexico"

"statefips 36"="New York"

"statefips 37"="North Carolina"

"statefips 38"="North Dakota"

"statefips 39"="Ohio"

"statefips 40"="Oklahoma"

"statefips 41"="Oregon"

"statefips 42"="Pennsylvania"

"statefips 44"="Rhode Island"

"statefips 45"="South Carolina"

"statefips 46"="South Dakota"

"statefips 47"="Tennessee"

"statefips 48"="Texas"

"statefips 49"="Utah"

"statefips 50"="Vermont"

"statefips 51"="Virginia"

"statefips 53"="Washington"

"statefips 54"="West Virginia"

"statefips 55"="Wisconsin"

"statefips 56"="Wyoming"

"statefips 60"="American Samoa"

"statefips 66"="Guam"

"statefips 69"="Northern Mariana Islands"

"statefips 72"="Puerto Rico"

"statefips 74"="U.S. Minor Outlying Islands"

"statefips 78"="U.S. Virgin Islands";

Run;

/\* Printing the clean results table \*/

ods excel file="/home/u60659161/MySAS/Project1\_Part2.xlsx" options(Embedded\_Titles="ON" Embedded\_Footnotes="ON"); /\*Use the path to your MySAS folder \*/

Title "Table 1: Determinants of Income (Categorical & Interaction Variables)";

footnote justify=left "Note: robust standard errors are in parentheses. \*, \*\*, and \*\*\* indicate

10%, 5%, and 1% significance levels, respectively.";

proc print data=Table\_Wide\_Sorted\_withstat noobs;

var Regressors;

var Model2 Model4 /style(header)={just=center} style(data)={just=center tagattr="type:String"};

format Regressors $VariableName.;

run;

ods excel close;

|  |  |  |
| --- | --- | --- |
| **Table 2: Determinants of Income (Categorical & Interaction Variables)** | | |
|  |  |  |
| **Regressors** | **Model2** | **Model4** |
| Intercept | 117,818.05\*\*\* | -45.95 |
|  | (507.98) | (63.69) |
| Treatment(OVB) | -5,884.56\*\*\* | 5,097.44\*\*\* |
|  | (81.16) | (96.35) |
| Change in weeks worked |  | 8.81 |
|  |  | (38.61) |
| Male | 5,945.98\*\*\* |  |
|  | (249.03) |  |
| Black | -444.92\*\*\* |  |
|  | (172.35) |  |
| Asian | 751.82\*\*\* |  |
|  | (240.74) |  |
| Hispanic | -4,155.41\*\*\* |  |
|  | (172.25) |  |
| Other | -618.51\* |  |
|  | (338.86) |  |
| High School Graduate | -1,050.95\*\*\* |  |
|  | (207.97) |  |
| Some College | 60,284.86\*\*\* |  |
|  | (204.72) |  |
| College Graduate | 68,687.00\*\*\* |  |
|  | (225.01) |  |
| Graduate Degree | 85,625.02\*\*\* |  |
|  | (251.88) |  |
| Male\*Black | -6,378.54\*\*\* |  |
|  | (250.39) |  |
| Male\*Asian | -7,607.83\*\*\* |  |
|  | (335.95) |  |
| Male\*Hispanic | -4,512.33\*\*\* |  |
|  | (230.21) |  |
| Male\*Other | -5,294.75\*\*\* |  |
|  | (474.58) |  |
| Male\*High School Graduate | 3,064.74\*\*\* |  |
|  | (272.42) |  |
| Male\*Some College | 5,025.64\*\*\* |  |
|  | (263.60) |  |
| Male\*College Graduate | 17,564.39\*\*\* |  |
|  | (287.49) |  |
| Male\*Graduate Degree | 31,727.09\*\*\* |  |
|  | (317.67) |  |
| Mechanics and repairers | -30,017.03\*\*\* |  |
|  | (255.97) |  |
| Construction | -30,173.87\*\*\* |  |
|  | (267.29) |  |
| Other production workers | -29,243.41\*\*\* |  |
|  | (279.12) |  |
| Machine operators | -36,891.14\*\*\* |  |
|  | (233.08) |  |
| Motor vehicle operators | -37,034.75\*\*\* |  |
|  | (246.14) |  |
| Other transportation | -34,620.91\*\*\* |  |
|  | (230.56) |  |
| Military | -22,034.45\*\*\* |  |
|  | (2,041.66) |  |
| Management Related Occupations | -15,427.30\*\*\* |  |
|  | (212.98) |  |
| Professional Specialty Occupations | -20,501.14\*\*\* |  |
|  | (155.11) |  |
| Technicians | -17,175.30\*\*\* |  |
|  | (227.74) |  |
| Sales | -24,671.22\*\*\* |  |
|  | (173.56) |  |
| Administrative Support Occupations | -31,823.61\*\*\* |  |
|  | (158.18) |  |
| Service Occupations | -32,993.72\*\*\* |  |
|  | (161.74) |  |
| Farming, Forestry And Fishing | -36,474.85\*\*\* |  |
|  | (366.88) |  |
| Supervisors Of Production Workers | -16,367.28\*\*\* |  |
|  | (805.80) |  |
| Business and repair services | 5,023.54\*\*\* |  |
|  | (372.93) |  |
| Personal services | -3,501.58\*\*\* |  |
|  | (388.66) |  |
| Entertainment and recreation services | -3,091.14\*\*\* |  |
|  | (423.65) |  |
| Professional and related services | 2,193.71\*\*\* |  |
|  | (350.63) |  |
| Public administration 15=Military | 12,917.19\*\*\* |  |
|  | (376.08) |  |
| ind 15 | 430.40 |  |
|  | (2,023.23) |  |
| Mining | 31,313.34\*\*\* |  |
|  | (616.57) |  |
| Construction | 5,840.04\*\*\* |  |
|  | (390.27) |  |
| Nondurable goods manufacturing | 16,267.92\*\*\* |  |
|  | (388.07) |  |
| Durable goods manufacturing | 16,249.60\*\*\* |  |
|  | (371.30) |  |
| Transportation, Communications and Utilities | 15,554.88\*\*\* |  |
|  | (373.63) |  |
| Wholesale trade | 14,139.61\*\*\* |  |
|  | (404.53) |  |
| Retail trade | -2,090.53\*\*\* |  |
|  | (356.49) |  |
| Finance, insurance and real estate | 14,504.02\*\*\* |  |
|  | (373.57) |  |
| Delaware | 3,664.20\*\*\* |  |
|  | (717.80) |  |
| District of Columbia | 16,231.84\*\*\* |  |
|  | (824.00) |  |
| Florida | 1,295.63\*\*\* |  |
|  | (328.89) |  |
| Georgia | 1,211.40\*\*\* |  |
|  | (356.62) |  |
| Hawaii | 2,028.68\*\*\* |  |
|  | (606.30) |  |
| Idaho | -3,564.48\*\*\* |  |
|  | (581.17) |  |
| Illinois | 3,161.69\*\*\* |  |
|  | (341.48) |  |
| Indiana | -819.20\*\* |  |
|  | (381.60) |  |
| Iowa | -4,045.42\*\*\* |  |
|  | (452.66) |  |
| Alaska | 6,362.57\*\*\* |  |
|  | (777.97) |  |
| Kansas | -2,713.63\*\*\* |  |
|  | (466.07) |  |
| Kentucky | -2,075.93\*\*\* |  |
|  | (423.50) |  |
| Louisiana | 1,311.52\*\*\* |  |
|  | (423.88) |  |
| Maine | -5,418.49\*\*\* |  |
|  | (611.01) |  |
| Maryland | 9,656.03\*\*\* |  |
|  | (386.79) |  |
| Massachusetts | 6,297.75\*\*\* |  |
|  | (376.59) |  |
| Michigan | -269.78 |  |
|  | (356.08) |  |
| Minnesota | -144.85 |  |
|  | (393.42) |  |
| Mississippi | -1,759.90\*\*\* |  |
|  | (481.28) |  |
| Missouri | -2,660.48\*\*\* |  |
|  | (386.42) |  |
| Montana | -6,338.92\*\*\* |  |
|  | (674.12) |  |
| Nebraska | -5,192.02\*\*\* |  |
|  | (536.34) |  |
| Nevada | 5,759.28\*\*\* |  |
|  | (479.10) |  |
| New Hampshire | 842.41 |  |
|  | (590.91) |  |
| New Jersey | 12,482.92\*\*\* |  |
|  | (361.46) |  |
| New Mexico | -1,529.49\*\*\* |  |
|  | (550.22) |  |
| New York | 7,241.72\*\*\* |  |
|  | (327.95) |  |
| North Carolina | -497.22 |  |
|  | (356.48) |  |
| North Dakota | -5,417.39\*\*\* |  |
|  | (761.72) |  |
| Ohio | -1,074.47\*\*\* |  |
|  | (346.00) |  |
| Arizona | 1,989.11\*\*\* |  |
|  | (389.57) |  |
| Oklahoma | -2,604.77\*\*\* |  |
|  | (440.50) |  |
| Oregon | -1,235.46\*\*\* |  |
|  | (431.29) |  |
| Pennsylvania | 426.77 |  |
|  | (342.99) |  |
| Rhode Island | 3,217.70\*\*\* |  |
|  | (653.84) |  |
| South Carolina | -1,041.83\*\* |  |
|  | (412.80) |  |
| South Dakota | -5,775.72\*\*\* |  |
|  | (734.16) |  |
| Tennessee | -1,488.31\*\*\* |  |
|  | (385.33) |  |
| Texas | 3,231.08\*\*\* |  |
|  | (322.53) |  |
| Utah | -2,406.89\*\*\* |  |
|  | (477.93) |  |
| Arkansas | -2,691.53\*\*\* |  |
|  | (475.93) |  |
| Vermont | -5,242.04\*\*\* |  |
|  | (781.70) |  |
| Virginia | 6,206.60\*\*\* |  |
|  | (365.49) |  |
| Washington | 3,543.43\*\*\* |  |
|  | (378.09) |  |
| West Virginia | -2,080.06\*\*\* |  |
|  | (568.44) |  |
| Wisconsin | -1,901.63\*\*\* |  |
|  | (386.63) |  |
| Wyoming | -1,747.37\*\* |  |
|  | (849.84) |  |
| California | 6,760.92\*\*\* |  |
|  | (315.65) |  |
| Colorado | 1,081.48\*\*\* |  |
|  | (397.68) |  |
| Connecticut | 11,968.19\*\*\* |  |
|  | (434.86) |  |
| Number of Weeks Worked LAst Year | 937.52\*\*\* |  |
|  | (1.87) |  |
| Number of Obs | 1,788,022 | 1,788,022 |
| Adjusted R-sq | 0.5286 | 0.001563 |
| Overall Significance | 20,269.07\*\*\* | 1,399.64\*\*\* |
|  |  |  |
| Note: robust standard errors are in parentheses. \*, \*\*, and \*\*\* indicate 10%, 5%, and 1% significance levels, respectively. | | |