ECON 426 - 01 **Shukhrat Khuseynov 21500372** **10.10.2018**

Homework 1

Analyze the total health care expenditure behavior of Medicare population (age 65 and above) using the mus03data_estimation.dta (any version). There is a hypothesis that the health care expenditure changes after retirement and becomes independent of income, hence evaluate the statement:

"Older generation determines their health care spending according to their quality of health rather than their income".

In order to illustrate the effects of both quality of health and income on the health care spending, relevant variables should be gathered. As representatives of the quality of life, the dummy variables, such as functional limitation (phylim) and activity limitation (actlim), along with the number of chronic problems (totchr) were taken. On the other hand, the most appropriate variables to represent income were annual household income (income) and supplement private insurance (suppins), a dummy variable. Moreover, to consider the demographic factors, variables such as age of the agent (age) and gender (female) were also added. Hence, the equation $\mathbf{totexp_i = \beta_1 + \beta_2 * phylim_i + \beta_3 * actlim_i + \beta_4 * totchr_i + \beta_5 * income_i + \beta_6 * suppins_i + \beta_7 * age_i + \beta_8 * female_i + U_i}$ is being studied.

The data was collected from Medicare population aged 65 and older to track their total care expenditure behavior. There are 3064 observations, where the arithmetic means of totexp, totchr, income and age are $7030.89, 1.75, $22.47 and 74.17, respectively. Other used variables are dummy variables. Among all variables, phylim and actlim are highly correlated (r=0.59), as well as phylim and totchr (r=0.35). Many of them are correlated with totexp to some extent. However the correlation between income and suppins is only 0.20.

Two types of regression analyses were conducted. In the first regression, OLS method was used (Shown in the table below as Reg1). In the second regression, OLS with the White's Heteroskedasticity Robust Standard Errors were deployed (Shown in table below as Reg2), which are more precise in the case of heteroskedasticity.

Models Reg1 and Reg2

	(1) Reg1	(2) Reg2
phylim	2426.5716***	2426.5716***
	(521.3988)	(536.9197)
actlim	3695.0459***	3695.0459***
	(566.0882)	(687.5856)
totchr	1939.7297***	1939.7297***
	(166.2386)	(180.9143)
income	7.2231	7.2231
	(9.2846)	(8.3685)
suppins	829.3004*	829.3004*
	(418.5569)	(411.5432)
age	-77.3720*	-77.3720*
	(33.2085)	(36.5747)
female	-1.26e+03**	-1.26e+03**
	(412.6097)	(416.6390)
_cons	7370.5665**	7370.5665**
	(2501.8671)	(2734.0622)

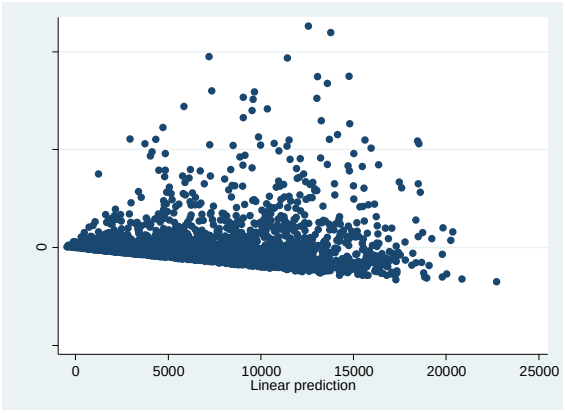
```
--------------------------------------------
N                        3064            3064
r2                     0.1234          0.1234
F                      61.4301         44.6571
ll                  -3.29e+04       -3.29e+04
--------------------------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

As seen in the table above, in both regressions the large positive effect of phylim, actlim and totchr on totexp is highly significant. The income's positive effect on totexp is not significant at α=0.05, although suppins affects positively and is significant. Lastly, both being a female and getting older negatively impacts on totexp with sufficient significance level.

In order to choose the best model among two regressions, the heteroskedasticity test of the initial one must be conducted. As a result of Breusch-Pagan test (shown below), homoskedasticity is rejected, so original model has heteroskedasticity issue. Therefore, the second regression model is appropriate in this case.

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of totexp

         chi2(1)      =    75.94
         Prob > chi2  =    0.0000
```



Furthermore, to observe how good the prediction of chosen model is, fitted values were plotted against the residuals, as seen in the graph above. It can be seen that most points are concentrated around y = 0 line, which is a case of predicted and observed values being the same, although there are still some outliers.

Finally, using the best model, several hypotheses testing can be conducted. As seen below, phylim, actlim and totchr are proved to be significant, as well as suppins. However when income is added, the joint test fails to reject the insignificance of both variables at 5% significance level.

```
. test phylim=actlim=totchr=0

( 1)  phylim - actlim = 0
( 2)  phylim - totchr = 0
( 3)  phylim = 0

       F(  3,  3056) =   96.17
            Prob > F =    0.0000
```

```
. test suppins=0

( 1)  suppins = 0

       F(  1,  3056) =    4.06
            Prob > F =    0.0440
```

```
. test income=suppins=0

( 1)  income - suppins = 0
( 2)  income = 0

       F(  2,  3056) =    2.81
            Prob > F =    0.0603
```

In a nutshell, after completing the heteroskedasticity test, the better regression among two of them was chosen. As a result of the regression output and several hypotheses tests, it can be concluded that the quality of health variables affect agents' health care expenditures more clearly and significantly than income group variables. Nevertheless, income group variables may also be accepted as explanatory ones, if to ignore annual household income (income), which is not significant, and to concentrate only on the supplement private insurance (suppins).