**National Research University Higher School of Economics**

**Coursework in**

**Quantitative Methods in Economics**

**“Factors influencing wage advantage at the beginning of career”**

**Accomplished by**

Nikita Tatarinov BSE196

Daria Tarasova BSE198

Arkady Andreev BSE194

**Professors**

Elena Vakulenko

Irina Chernysheva

**1. Introduction**

The work is based on the article: Bleemer, Zachary and Aashish Mehta. 2022. "Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major." American Economic Journal: Applied Economics, 14(2):1-22.

The article examines the return of wages to studying economics at one of the American universities (University of California, Santa Cruz (UCSC)), which has a policy of preventing students with low GPA for the first year of study from choosing an economics major.

The statement of the problem in the article is due to the positive dependence of the level of wages immediately after graduation on the received specialty in one of the economic disciplines.

**2. Statement of the research problem. Urgency of the topic. Formulation of interesting substantive hypotheses**

In our project, we investigated a similar problem, namely: does studying economics provide a wage advantage at the beginning of a career path?

However, unlike the article taken as a basis, we will study a more common system of choosing specialties – when a student chooses the main one immediately upon admission. As a result, we plan to compare the inferences with the findings of Zachary Bleemer and Aashish Mehta. In addition to solving the main research problem of our project, we are investigating which system is more effective in terms of providing graduates with specialties.

Questions about choosing the type of educational program that provides a decent income in the future are significant, particularly for high school graduates. So the theme of our project is really urgent.

In addition, it is relevant for us as for students studying at the Applied Economics Minor. And although the study will be conducted specifically for the study of the main specialty, perhaps some of them will be useful to us.

The purpose of this study is to accept or reject the hypothesis that there is one fixed relationship between the specialties of economic disciplines (economics / finance) and wage increases immediately after graduation.

In addition, in the coursework, we propose to check that wage after the end of study has a positive correlation with:

* GPA;
* sex;
* level of education of parents.

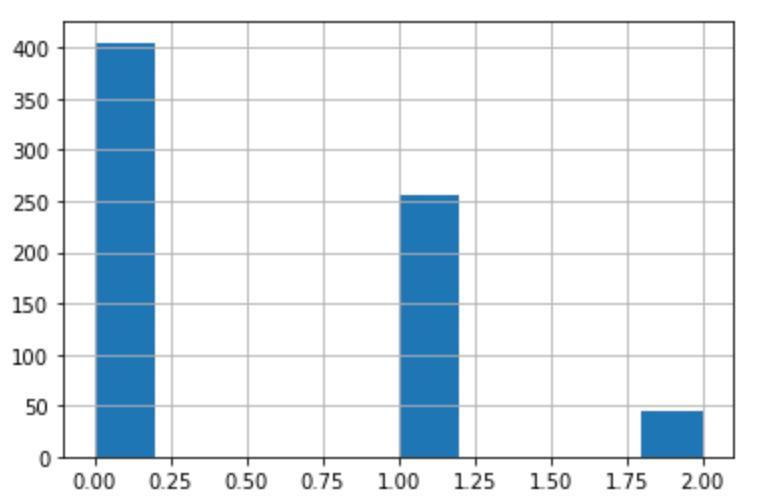
**3. Data source, description of indicators. Analysis of descriptive statistics and graphical analysis of variables**

The data source for our study was a dataset consisting of data from graduates of a university in China. The original data set consisted of two parts: the first one was obtained by an alumni survey (for example, wage immediately after graduation) and the second one was provided by the university administration and included all basic demographic indicators, academic achievements of the surveyed graduates. In total, data of 706 graduates were collected in the dataset.

The following indicators were used during the study:

* *gender* – describes the gender of the graduate (takes the values 0 if the graduate is female and 1 if the graduate is male);
* *gpa* – denotes a graduate GPA ranging from 0 to 5;
* *wagea* – determines the wage of the graduate immediately after graduation (conventional units are measured in yuan);
* *lnawage* – logarithmic values of wages (used to compare the results of models built on wagea with lnawage);
* *gradchoice* – denotes the working status of a graduate after graduation (0 - immediately got a job after graduation, 1 - continued his studies somewhere else, 2 - in anticipation of employment);
* *major* – type of graduate educational program (0 - Economics, 1 - Finance, 2 - Management, 3 - Engineering and Mathematics, 4 - Social Sciences)
* *hheconstat* – type of residential area (poor or rich);
* *Pedu* – parental education (1 - secondary school, 2 - high school, 3 - college and above);
* *Poccu* – occupation of parents (0 - unemployed, 1 - professional occupation, 2 - management, 3 - agriculture, 4 - city service).

General information and correlation matrix can be seen in Table 1-2. What stands out is the strong inverse relationship between *gpa* and *gender*, and the strong relationship between *hheconstant* and *Pedu*.



On this histogram, you can see the distribution of *gradchoice*. It can be seen that a very small part of graduates remained in anticipation of employment, which makes sense, given the high average *gpa* in the sample.

**4. Building models**

Initial data were collected from students from 5 types of educational programs (major): economics, finance, management, science and social science. We decided to divide students into 3 groups: students from economics or finance, students from management and students from social sciences. This was caused by general considerations about the structure of disciplines. For each group, 3 models of the dependence of the initial (the current one was not studied, since there is no data on work experience) wages on factors were built (tables with models and related tests are located in the appendix in the chapters with the corresponding names):

1. linear – model of wage dependence on all parameters of interest;
2. semi-logarithmic – model of the dependence of the logarithm of wages on all parameters of interest;
3. logarithmic – model of the dependence of the logarithm of wage on all parameters of interest with the logarithm of continuous parameters (in this case, *gpa*).

Each model was compiled as follows (all numbers are placed in the tables of the corresponding sections of the supplement):

1) factors were added to the model one at a time, for each new factor, the quality characteristics of the models known to us were considered - и , F-statistics and p-value, information criteria (AIC, BIC, HQ);

2) the hypothesis of the insignificance of a group of factors was tested (the choice of factors was carried out both in terms of their level of significance and in terms of their contribution to the overall level of the model) - insignificant ones were eliminated;

3) a model was built with the selected factors, the characteristics mentioned in step 1 were considered for it;

4) a Ramsey test was performed to check the need to include squares and/or cubes of factors in the models;

5) a test for multicollinearity of factors was carried out;

6) all tests known to us for heteroscedasticity were performed (that is, for violation of the 2nd assumption for a random term in the Gauss-Markov theorem);

7) a test for the normality of the residuals was performed.

Multicollinearity was not found in any of the models, while the need to add squares and/or cubes of factors was present in several models, but not in the semi-logarithmic ones. Heteroskedasticity was present in each of the models. The remains were almost always abnormal.

After building 3 models for each of the groups, the models were compared. Logarithmic and semi-logarithmic were compared, first of all, by and F-statistics, and semi-logarithmic and linear were compared using the Box-Cox test. For each of the groups, the semi-logarithmic one turned out to be the best.

Further, it was necessary to solve the issue of heteroscedasticity and abnormality of the residuals for the 3 (for each of the groups) remaining best models. The White test did not reveal heteroscedasticity in any of the cases, while the Glaser and Broesch-Pagan tests revealed it in each of the cases. In this regard, it was decided not to try to apply White's corrections, but to divide the model equation into estimates of standard deviations, expressed through estimates of the coefficients of the dependence of the residuals on one of the factors (the dependence was taken from the Glaser test). It turned out that such weighting of the model not only eliminated heteroscedasticity (for all 3 tests), but also significantly improved the quality of the models: increased from ~0.1 to ~0.7, F-statistics increased from ~10 to several hundred (p-value orders decreased significantly). As for the elimination of the abnormality of the residuals, it was not possible to cope with it for the following reasons:

1. changing the functional form would be meaningless, since our models were the best we could build;
2. it would be impossible to change the specification because no other sets with suitable data were found, in which there would be other factors;
3. adding new observations was impossible due to the same lack of other suitable data sets;
4. there was not enough time for bootstrap.

After receiving the model for each of the 3 groups, we saw an incredible similarity between the models for management and sciences: not only by the coincidence of significant factors, but also, for example, by the significant dependence of the residuals on the same factor. It was decided to conduct a Chow test for these 2 groups, as a result of which it turned out that there was no reason to reject the null hypothesis that the models were statistically different. In this regard, it was decided to combine the data of these 2 groups and build a single model.

As a result, we got 2 different models: for strictly economic disciplines (economics and finance) and for near-economic disciplines (management, science and social science). Then, we came up with the idea of ​​building a model with a new dummy variable equal to 1 for students from economics and 0 for all others. We applied the above steps and got the following final model:

| model | values |
| --- | --- |
| const | 21.2691\*\*\*  (0.546586) |
| w\_gender | 0.357434\*\*\*  (0.0437788) |
| w\_gpa | 0.336002\*\*\*  (0.0562017) |
| w\_major | 0.0723000\*  (0.0418832) |
| w\_hheconstat | −1.35428\*\*\*  (0.0578743) |
| w\_Pedu | 0.0570148\*\*  (0.0256814) |
| R2 | 0.689223 |
| R2adj | 0.687003 |
| F | 310.4836 |
| p-value | 6.5e-175 |
| AIC | 2362.919 |
| BIC | 2390.277 |
| HQ | 2373.490 |

Regarding the premises of the Gauss-Markov theorem:

1. it was possible to get rid of heteroscedasticity (according to the results of tests known to us);
2. the covariance of the residuals was not tested;
3. endogeneity has not been studied, since there are no tests for it and we are limited in data, but theoretically it would be possible to take into account the experience of internships, abilities;
4. theoretically, some variables could be omitted (such as in point 4, for example), but we did not have another set and time to build other models and compare. Extra variables were removed from the model using tests;
5. the functional form is chosen correctly according to the Ramsey test;
6. the values ​​of the factors are determined.

As for the distribution of residuals, according to the Harke-Beer test, it is not normal, but the plot of the distribution of residuals and the quantile-quantile plot visually indicate the closeness of the residuals to the normal distribution.

Thus, it is impossible to be fully confident in the unbiasedness and efficiency of the coefficients of our model (primarily due to the lack of verification of the endogeneity of variables), but the model indicators (, F-statistics, and others) are quite high. If there were more data, which also included several other variables (for example, experience), the model could be considered effective.

**5. Interpretation of findings**

As a result, we see that the *major* turned out to be slightly significant, but nevertheless, the starting wage is statistically higher for students in economics (economics/finance). That is, our hypothesis was confirmed, but we expected to see a greater influence of this feature.

*gender* and *gpa* turned out to be statistically strongly significant with a positive dependence. That is, men and students who had a high GPA receive higher salaries at the start. Such results correspond to our assumptions from paragraph 2 of this document and reality (at the same time, the dependence of wage on gender does not really correspond to common sense) - for many years there have been many studies on gender inequality in the corporate environment (for example: [1]), including in salaries. A lot of studies have also been conducted about the impact of GPA on the level of salaries at the start (for example, [2]), in general, confirming a strong positive dependence (so parents are not in vain worried about academic performance).

Unexpectedly, the wealth level of a person's area of residence (rich/poor) had a statistically strong inverse relationship. That is, people from poor areas at the start received large wages. At the same time, this parameter turned out to be highly significant for all directions of the majors, except for the economic one. This can be explained by the fact that people from poor areas seek to find a higher paying job, and those who are better off seek one that will be more interesting for them and give them more experience.

The level of education of parents turned out to be medium significant with a direct relationship, especially for graduates of the economic specialty. The assumption regarding this feature was also confirmed, although we expected less significance from it. Studies have also been conducted on this topic (for example, [3]), which came to a similar conclusion

Sources used in paragraph 5:  
1. “Гендерные различия в заработной плате выпускников вузов и учреждений СПО на начальном этапе карьеры” М.А.Кирюшина, В.Н.Рудаков - [https://www.hse.ru/data/2021/07/07/1433751016/Кирюшина,%20Рудаков.pdf](https://www.hse.ru/data/2021/07/07/1433751016/%D0%9A%D0%B8%D1%80%D1%8E%D1%88%D0%B8%D0%BD%D0%B0,%20%D0%A0%D1%83%D0%B4%D0%B0%D0%BA%D0%BE%D0%B2.pdf)

2. “Your high school GPA could affect your income” Michael T. French - <https://www.sciencedaily.com/releases/2014/05/140519092835.htm>

3. “The impact of parental income and education on the schooling of their children” Arnaud Chevalier, Colm Harmon, Vincent O’ Sullivan & Ian Walker - <https://izajole.springeropen.com/articles/10.1186/2193-8997-2-8>

**6. Comparison of the results of your research with the results of the article**

The dependence of wages on the *major* in the economy turned out to be not as strong as that of the authors of the article.

We assume that this is the result of the dropouts that were made at the UCSC university - the major in economics initially got more successful students (with a higher GPA). At the same time, the positive impact of *gpa* on the level of wages turned out to be high both for us and for the authors of the article taken as the basis. That is, at UCSC, the major in economics initially got students with greater potential in the future to have a higher wage, because of this, the dependence on the major in economics became much stronger.

**SUPPLEMENTS**

**ECONOMICS AND FINANCE**

General information

|  | Mean | Median | S.D. | Min | Max |
| --- | --- | --- | --- | --- | --- |
| *gender* | 0.5 | 0.5 | 0.5010 | 0 | 1 |
| *gpa* | 3.336 | 3.388 | 0.4169 | 1.996 | 4.218 |
| *wageb* | 9197 | 8046 | 6550 | 1609 | 64371 |
| *wagea* | 13396 | 10000 | 9122 | 1000 | 80000 |
| *lnbwage* | 8.967 | 8.993 | 0.5375 | 7.384 | 11.07 |
| *lnawage* | 9.329 | 9.210 | 0.5841 | 6.908 | 11.29 |
| *gradchoice* | 0.6189 | 1 | 0.5722 | 0 | 2 |
| *major* | 1.951 | 2 | 0.2167 | 1 | 2 |
| *hheconstat* | 0.5943 | 1 | 0.4920 | 0 | 1 |
| *Pedu* | 2.377 | 3 | 0.7572 | 1 | 3 |
| *Poccu* | 2.004 | 2 | 1.242 | 0 | 4 |

Correlation matrix

|  | *gender* | *gpa* | *gradchoice* | *hheconstat* | *Pedu* | *Poccu* |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 1 | -0.4176 | 0.0072 | -0.1252 | -0.0325 | 0.0562 |
| *gpa* |  | 1 | 0.2229 | 0.0754 | 0.0869 | 0.0645 |
| *gradchoice* |  |  | 1 | 0.0916 | 0.1336 | 0.0427 |
| *hheconstat* |  |  |  | 1 | 0.3240 | -0.1656 |
| *Pedu* |  |  |  |  | 1 | -0.1854 |
| *Poccu* |  |  |  |  |  | 1 |

Regression models of starting wage (linear)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 3355.21\*\*\*  (1188.23) | 5062.20\*\*\*  (1298.95) | 4973.62\*\*\*  (1305.80) | 5248.78\*\*\*  (1301.55) | 5121.99\*\*\*  (1290.74) | 5098.17\*\*\*  (1297.76) |
| *gpa* |  | 4629.68\*\*\*  (1541.32) | 4353.02\*\*\*  (1587.96) | 4318.08\*\*\*  (1575.32) | 4094.25\*\*\*  (1563.86) | 4058.18\*\*  (1575.29) |
| *gradchoice* |  |  | 778.814  (1057.76) | 520.567  (1056.03) | 237.697  (1053.56) | 226.535  (1056.94) |
| *hheconstat* |  |  |  | 2609.96\*\*  (1204.20) | 1823.27  (1241.59) | 1856.29  (1252.79) |
| *Pedu* |  |  |  |  | 1886.02\*\*  (823.738) | 1916.72\*\*  (836.612) |
| *Poccu* |  |  |  |  |  | 107.496  (476.454) |
| const | 11723.7\*\*\*  (840.206) | −4553.27  (5481.53) | −4064.01  (5527.11) | −5513.35  (5523.46) | −8597.35  (5635.89) | −8768.43  (5698.43) |
| R2 | 0.033505 | 0.070140 | 0.072346 | 0.091153 | 0.111757 | 0.111958 |
| R2adj | 0.029303 | 0.062019 | 0.060140 | 0.075138 | 0.092105 | 0.088277 |
| F | 7.973285 | 8.636825 | 5.927077 | 5.691780 | 5.686967 | 4.727720 |
| p-value | 0.005164 | 0.000242 | 0.000655 | 0.000221 | 0.000058 | 0.000148 |
| AIC | 4887.624 | 4880.659 | 4882.108 | 4879.356 | 4876.036 | 4877.983 |
| BIC | 4894.517 | 4890.999 | 4895.895 | 4896.590 | 4896.716 | 4902.111 |
| HQ | 4890.404 | 4884.829 | 4887.668 | 4886.306 | 4884.376 | 4887.714 |

Hypothesis about the significance of the coefficients *gradchoice*, *Pedu*, *Poccu*

| F | 0.833331 |
| --- | --- |
| p-value | 0.476715 |

Regression model of starting wage (linear)

| model | values |
| --- | --- |
| const | −8499.38  (5567.04) |
| *gender* | 4960.52\*\*\*  (1279.19) |
| *gpa* | 4195.45\*\*\*  (1524.76) |
| *Pedu* | 2259.22\*\*\*  (783.769) |
| R2 | 0.102835 |
| R2adj | 0.091030 |
| F | 8.711275 |
| p-value | 0.000017 |
| AIC | 4874.355 |
| BIC | 4888.142 |
| HQ | 4879.915 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 0.002119 | 0.963 |
| cubes | 0.002004 | 0.964 |
| squares and cubes | 0.001073 | 0.999 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| gender | 1.251 |
| gpa | 1.307 |
| gradchoice | 1.087 |
| hheconstat | 1.152 |
| Pedu | 1.162 |
| Poccu | 1.067 |

White test

| R2 | 0.027826 |
| --- | --- |
| TR2 | 6.455533 |
| p-value | 0.596344 |

Glejser test

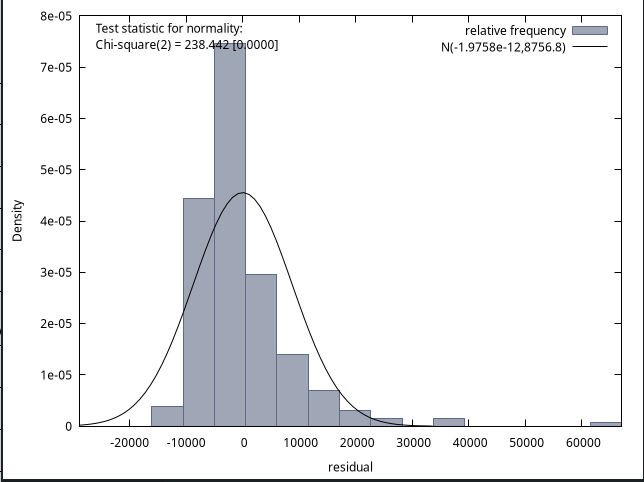
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 4408.84\*\*\*  (599.029) | 2573.29\*\*\*  (847.155) |
| *gpa* | 7567.84\*\*  (3440.32) | −562.025  1024.52 |
| *Pedu* | 2519.17\*  (1460.12) | 1318.26\*\*  (579.469) |

Breusch–Pagan test

| ESS | 71.0309 |
| --- | --- |
| LM | 35.515463 |
| p-value | 0.00000 |

Jarque-Bera test

| chi2 | 238.442 |
| --- | --- |
| p-value | 0.00000 |



Regression models of the logarithm of starting wage (semilogarithmic)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 0.223480\*\*\*  (0.0752597) | 0.375005\*\*\*  (0.0806945) | 0.369845\*\*\*  (0.0811320) | 0.386663\*\*\*  (0.0808947) | 0.376262\*\*\*  (0.0795297) | 0.376781\*\*\*  (0.0799702) |
| *gpa* |  | 0.410963\*\*\*  (0.0957514) | 0.394848\*\*\*  (0.0986629) | 0.392713\*\*\*  (0.0979104) | 0.374351\*\*\*  (0.0963581) | 0.375137\*\*\*  (0.0970716) |
| *gradchoice* |  |  | 0.0453643  (0.0657205) | 0.0295803  (0.0656349) | 0.00637548  (0.0649158) | 0.00661889  (0.0651304) |
| *hheconstat* |  |  |  | 0.159520\*\*  (0.0748443) | 0.0949852  (0.0765013) | 0.0942651  0.0771988 |
| *Pedu* |  |  |  |  | 0.154717\*\*\*  (0.0507550) | 0.154048\*\*\*  (0.0515533) |
| *Poccu* |  |  |  |  |  | −0.00234400  (0.0293598) |
| const | 9.21766\*\*\*  (0.0532167) | 7.77280\*\*\*  (0.340529) | 7.80130\*\*\*  (0.343410) | 7.71272\*\*\*  (0.343298) | 7.45973\*\*\*  (0.347258) | 7.46346\*\*\*  (0.351146) |
| R2 | 0.036922 | 0.108625 | 0.110484 | 0.127936 | 0.162376 | 0.162399 |
| R2adj | 0.032735 | 0.100840 | 0.098780 | 0.112569 | 0.143844 | 0.140063 |
| F | 8.817602 | 13.95328 | 9.439739 | 8.325487 | 8.762138 | 7.270741 |
| p-value | 0.032735 | 1.91e-06 | 6.64e-06 | 2.77e-06 | 1.30e-07 | 4.12e-07 |
| AIC | 402.1217 | 386.1720 | 387.6876 | 385.0907 | 377.7427 | 379.7361 |
| BIC | 409.0152 | 396.5122 | 401.4746 | 402.3244 | 398.4231 | 403.8633 |
| HQ | 404.9018 | 390.3421 | 393.2478 | 392.0409 | 386.0829 | 389.4663 |

Hypothesis about the significance of the coefficients *gradchoice*, *Pedu*, *Poccu*

| F | 0.525329 |
| --- | --- |
| p-value | 0.665282 |

Regression model of the logarithm of starting wage (semilogarithmic)

| model | values |
| --- | --- |
| const | 7.52109\*\*\* (0.337318) |
| *gender* | 0.370028\*\*\*  (0.0765235) |
| *gpa* | 0.377750\*\*\*  (0.0922746) |
| *Pedu* | 0.152582\*\*\*  (0.0461853) |
| R2 | 0.146259 |
| R2adj | 0.135588 |
| F | 13.70527 |
| p-value | 2.79e-08 |
| AIC | 398.4690 |
| BIC | 412.4577 |
| HQ | 404.1029 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 0.212084 | 0.646 |
| cubes | 0.215160 | 0.643 |
| squares and cubes | 0.130674 | 0.878 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| *gender* | 1.251 |
| *gpa* | 1.307 |
| *gradchoice* | 1.087 |
| *hheconstat* | 1.152 |
| *Pedu* | 1.162 |
| *Poccu* | 1.067 |

White test

| R2 | 0.029361 |
| --- | --- |
| TR2 | 7.164055 |
| p-value | 0.519040 |

Glejser test

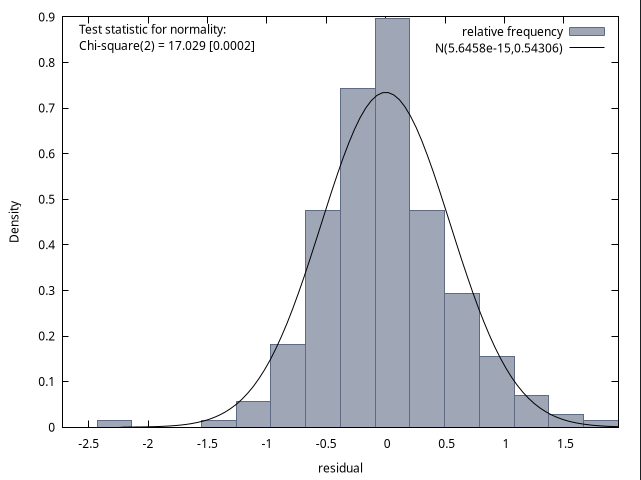
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 0.371306\*\*\*  (0.0315949) | 0.0767124\*  (0.0446819) |
| *gpa* | 0.637432\*\*\*  (0.181017) | −0.0682847  (0.0538512) |
| *Pedu* | 0.385313\*\*\*  (0.0741766) | 0.0102437  (0.0297390) |

Breusch–Pagan test

| ESS | 13.4911 |
| --- | --- |
| LM | 6.745531 |
| p-value | 0.080466 |

Jarque-Bera test

| chi2 | 17.029 |
| --- | --- |
| p-value | 0.00020 |



Regression models of the logarithm of starting wage (logarithmic)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 0.223480\*\*\*  (0.0752597) | 0.371377\*\*\*  (0.0806472) | 0.366135\*\*\*  (0.0810389) | 0.383665\*\*\*  (0.0808084) | 0.372645\*\*\*  (0.0794846) | 0.373394\*\*\*  (0.0799510) |
| *ln(gpa)* |  | 1.27115\*\*\*  (0.300977) | 1.21891\*\*\*  (0.309471) | 1.21644\*\*\*  (0.307026) | 1.15274\*\*\*  (0.302415) | 1.15646\*\*\*  (0.304966) |
| *gradchoice* |  |  | 0.0484816  (0.0656586) | 0.0322826  (0.0655714) | 0.00948238  (0.0648701) | 0.00980885  (0.0650804) |
| *hheconstat* |  |  |  | 0.161483\*\*  (0.0749010) | 0.0973602  (0.0765982) | 0.0963767  (0.0772877) |
| *Pedu* |  |  |  |  | 0.153527\*\*\*  (0.0508388) | 0.152597\*\*\*  (0.0516504) |
| *Poccu* |  |  |  |  |  | −0.00322840  (0.0294243) |
| const | 9.21766\*\*\*  (0.0532167) | 7.62496\*\*\*  (0.380595) | 7.66032\*\*\*  (0.383971) | 7.56628\*\*\*  (0.383425) | 7.33090\*\*\*  (0.384724) | 7.33521\*\*\*  (0.387560) |
| R2 | 0.036922 | 0.106517 | 0.108648 | 0.126534 | 0.160413 | 0.160458 |
| R2adj | 0.032735 | 0.098713 | 0.096920 | 0.111142 | 0.141838 | 0.138070 |
| F | 8.817602 | 13.65014 | 9.263761 | 8.221026 | 8.636003 | 7.167216 |
| p-value | 0.032735 | 2.51e-06 | 8.33e-06 | 3.30e-06 | 1.66e-07 | 5.23e-07 |
| AIC | 402.1217 | 386.7201 | 388.1660 | 385.4634 | 378.2856 | 380.2732 |
| BIC | 409.0152 | 397.0603 | 401.9529 | 402.6971 | 398.9660 | 404.4004 |
| HQ | 404.9018 | 390.8902 | 393.7261 | 392.4136 | 386.6258 | 390.0034 |

Hypothesis about the significance of the coefficients *gradchoice*, *Pedu*, *Poccu*

| F | 0.558973 |
| --- | --- |
| p-value | 0.642611 |

Regression model of the logarithm of starting wage (logarithmic)

| model | values |
| --- | --- |
| const | 7.34141\*\*\*  (0.379449) |
| *gender* | 0.363234\*\*\*  (0.0786709) |
| *ln(gpa)* | 1.16267\*\*\*  (0.295035) |
| *Pedu* | 0.173160\*\*\*  (0.0483044) |
| R2 | 0.154189 |
| R2adj | 0.143059 |
| F | 13.85454 |
| p-value | 2.49e-08 |
| AIC | 375.9993 |
| BIC | 389.7862 |
| HQ | 381.5594 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 0.280158 | 0.597 |
| cubes | 0.286835 | 0.593 |
| squares and cubes | 0.218360 | 0.804 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| *gender* | 1.249 |
| *gpa* | 1.302 |
| *gradchoice* | 1.082 |
| *hheconstat* | 1.152 |
| *Pedu* | 1.163 |
| *Poccu* | 1.069 |

White test

| R2 | 0.036359 |
| --- | --- |
| TR2 | 8.435346 |
| p-value | 0.392139 |

Glejser test

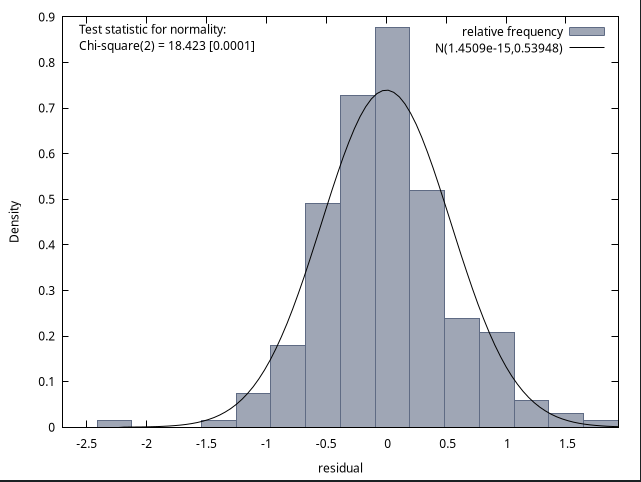
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 0.361796\*\*\*  (0.0321824) | 0.0886948\*  (0.0455128) |
| *ln(gpa)* | 0.647012\*\*\*  (0.205270) | −0.201599  (0.170735) |
| *Pedu* | 8.87168\*\*\*  (0.127196) | 0.189967\*\*\*  (0.0504795) |

Breusch–Pagan test

| ESS | 13.9431 |
| --- | --- |
| LM | 6.971535 |
| p-value | 0.072811 |

Jarque-Bera test

| chi2 | 18.423 |
| --- | --- |
| p-value | 0.00010 |



Semi-logarithmic model is better than the logarithmic one in , F. To compare linear and semi-logarithmic models, we apply the Box-Cox test. The average of the logarithm of wage *lnawage* is 9.3288.

wagea\_star = wage/exp(9.3288)

lnawage\_star = ln(wagea\_star)

RSS(wagea\_star)=142.6537

RSS(lnawage\_star)=79.31227

chi2 = 244 / 2 \* ln( 142.6537 / 79.31227 ) = 71.6173147 > x2\_1

H\_0 is rejected, i.e. the semi-logarithmic model is better (it has less RSS). It remains to eliminate heteroscedasticity. The white test does not reveal heteroscedasticity, so let's try to use the estimates of the variance of random components from the Glaser test to weight the model.

In our case, the random component will significantly depend only on *gender*:

Divide each i-th observation for the dependent variable *lnawage* and all dependent factors – *gender*, *gpa* and *Pedu* – by .

Result model

| model | values |
| --- | --- |
| const | 20.0586\*\*\*  (0.914656) |
| *w\_gender* | −1.16511\*\*\*  (0.126559) |
| *w\_gpa* | 0.385409\*\*\*  (0.0938720) |
| *w\_Pedu* | 0.171869\*\*\*  (0.0455834) |
| R2 | 0.684188 |
| R2adj | 0.680240 |
| F | 173.3150 |
| p-value | 8.78e-60 |
| AIC | 833.0127 |
| BIC | 847.0013 |
| HQ | 838.6465 |

White test

| R2 | 0.025199 |
| --- | --- |
| TR2 | 6.148440 |
| p-value | 0.630609 |

Glejser test

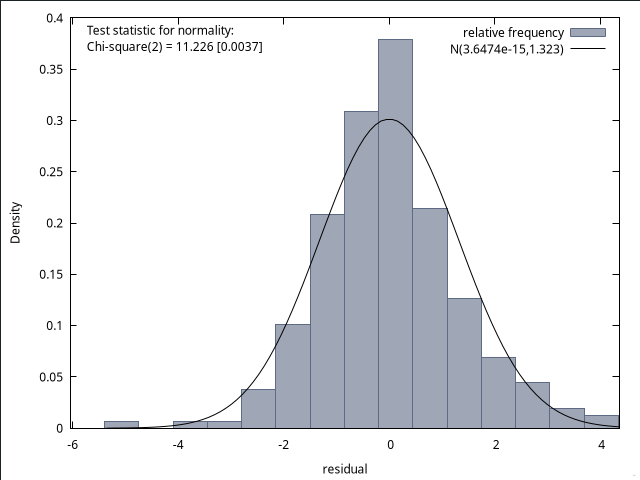
|  | const | coefficient |
| --- | --- | --- |
| *w\_gender* | 0.997813\*\*\*  (0.0771154) | 0.00292063  (0.0488598) |
| *w\_gpa* | 1.15681\*\*\*  (0.304505) | −0.0188668  (0.0362946) |
| *w\_Pedu* | 0.929245\*\*\*  (0.171513) | 0.0122583  (0.0277536) |

Breusch–Pagan test

| ESS | 6.63978 |
| --- | --- |
| LM | 3.319892 |
| p-value | 0.344884 |

Jarque-Bera test

| chi2 | 11.226 |
| --- | --- |
| p-value | 0.00365 |



**MANAGEMENT**

General information

|  | Mean | Median | S.D. | Min | Max |
| --- | --- | --- | --- | --- | --- |
| *gender* | 0.3538 | 0 | 0.4789 | 0 | 1 |
| *gpa* | 3.276 | 3.308 | 0.3494 | 2.157 | 4.092 |
| *wageb* | 7104 | 6437 | 3436 | 1824 | 21886 |
| *wagea* | 11161 | 940 | 8295 | 2500 | 10000 |
| *lnbwage* | 8.762 | 8.770 | 0.4623 | 7.509 | 9.994 |
| *lnawage* | 9.161 | 9.148 | 0.5332 | 7.824 | 11.51 |
| *gradchoice* | 0.3938 | 0 | 0.6075 | 0 | 2 |
| *major* | 3 | 3 | 0 | 3 | 3 |
| *hheconstat* | 0.5908 | 1 | 0.4925 | 0 | 1 |
| *Pedu* | 2.240 | 2 | 0.8113 | 1 | 3 |
| *Poccu* | 1.988 | 2 | 1.245 | 0 | 4 |

Correlation matrix

|  | *gender* | *gpa* | *gradchoice* | *hheconstat* | *Pedu* | *Poccu* |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 1 | -0.4650 | -0.0880 | -0.0908 | -0.0842 | 0.0695 |
| *gpa* |  | 1 | 0.1593 | 0.0000 | 0.0404 | -0.0211 |
| *gradchoice* |  |  | 1 | 0.0555 | 0.1332 | -0.0793 |
| *hheconstat* |  |  |  | 1 | 0.3625 | -0.1593 |
| *Pedu* |  |  |  |  | 1 | -0.3577 |
| *Poccu* |  |  |  |  |  | 1 |

Regression models of starting wage (linear)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 1340.72  (960.888) | 2551.20\*\*  (1077.33) | 2550.40\*\*  (1079.14) | 2850.51\*\*\*  (1073.91) | 2858.33\*\*\*  (1076.34) | 2840.43\*\*\*  (1078.67) |
| *gpa* |  | 3567.59\*\*  (1476.45) | 3576.92\*\*  (1492.24) | 3799.63\*\*  (1479.51) | 3798.56\*\*  (1481.76) | 3788.49\*\*  (1483.92) |
| *gradchoice* |  |  | −35.5426  (762.824) | −149.478  (756.321) | −166.062  (762.607) | −155.931  (764.025) |
| *hheconstat* |  |  |  | 2540.98\*\*\*  (925.929) | 2475.05\*\*  (991.713) | 2487.22\*\*  (993.478) |
| *Pedu* |  |  |  |  | 113.463  (604.941) | 195.582  (639.140) |
| *Poccu* |  |  |  |  |  | 157.669  (391.526) |
| const | 10686.7\*\*\*  (571.584) | −1427.87  (5045.61) | −1444.15  (5065.51) | −3736.12  (5083.83) | −3944.09  (5210.85) | −4413.28  (5346.20) |
| R2 | 0.005991 | 0.023694 | 0.023701 | 0.046149 | 0.046254 | 0.046740 |
| R2adj | 0.002914 | 0.017630 | 0.014576 | 0.034226 | 0.031305 | 0.028754 |
| F | 1.946856 | 3.907335 | 2.597542 | 0.046149 | 3.094113 | 2.598684 |
| p-value | 0.163885 | 0.021054 | 0.052366 | 0.004369 | 0.009629 | 0.017929 |
| AIC | 6788.581 | 6784.741 | 6786.739 | 6781.179 | 6783.143 | 6784.977 |
| BIC | 6796.149 | 6796.092 | 6801.874 | 6800.098 | 6805.846 | 6811.464 |
| HQ | 6791.601 | 6789.271 | 6792.779 | 6788.729 | 6792.204 | 6795.548 |

Hypothesis about the significance of the coefficients *gradchoice*, *Pedu*, *Poccu*

| F | 0.0786992 |
| --- | --- |
| p-value | 0.971513 |

Regression model of starting wage (linear)

| model | values |
| --- | --- |
| const | −3658.79  (5061.16) |
| *gender* | 2852.68\*\*\*  (1072.24) |
| *gpa* | 3759.62\*\*  (1463.41) |
| *hheconstat* | 2530.93\*\*\*  (923.148) |
| R2 | 0.046032 |
| R2adj | 0.037117 |
| F | 5.163124 |
| p-value | 0.001692 |
| AIC | 6779.218 |
| BIC | 6794.354 |
| HQ | 6785.259 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 2.000508 | 0.158 |
| cubes | 2.512709 | 0.114 |
| squares and cubes | 3.059091 | 0.0483 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| *gender* | 1.294 |
| *gpa* | 1.304 |
| *gradchoice* | 1.045 |
| *hheconstat* | 1.160 |
| *Pedu* | 1.304 |
| *Poccu* | 1.151 |

White test

| R2 | 0.008230 |
| --- | --- |
| TR2 | 2.674849 |
| p-value | 0.913366 |

Glejser test

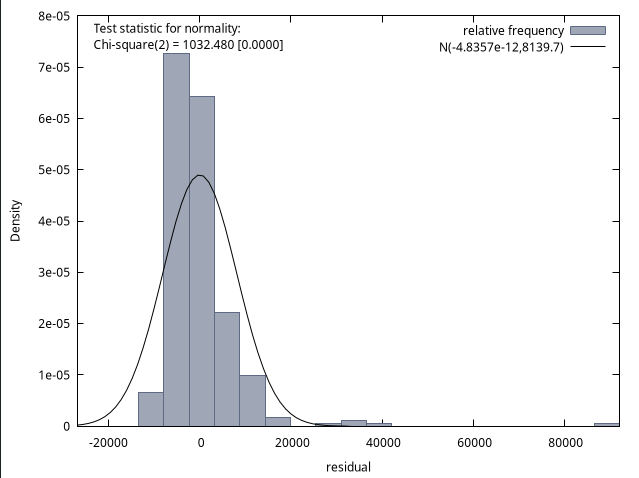
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 4717.74\*\*\*  (449.869) | 274.733  756.273 |
| *gpa* | 935.039  (3408.06) | 1184.47  (1034.57) |
| *hheconstat* | 3590.20\*\*\*  (558.409) | 2073.16\*\*\*  (726.513) |

Breusch–Pagan test

| ESS | 109.975 |
| --- | --- |
| LM | 54.987337 |
| p-value | 0.000000 |

Jarque-Bera test

| chi2 | 1032.480 |
| --- | --- |
| p-value | 0.00000 |



Regression models of the logarithm of starting wage (semilogarithmic)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 0.145194\*\*\*  (0.0614264) | 0.253787\*\*\*  (0.0682620) | 0.253755\*\*\*  (0.0683769) | 0.271282\*\*\*  (0.0681658) | 0.274627\*\*\*  (0.0681526) | 0.273335\*\*\*  (0.0682951) |
| *gpa* |  | 0.320053\*\*\* (0.0935513) | 0.320426\*\*\*  (0.0945520) | 0.333433\*\*\*  (0.0939116) | 0.332979\*\*\*  (0.0938235) | 0.332252\*\*\*  (0.0939532) |
| *gradchoice* |  |  | −0.00142239  (0.0483345) | −0.00807647  (0.0480072) | −0.0151700  (0.0482875) | −0.0144389  (0.0483736) |
| *hheconstat* |  |  |  | 0.148399\*\*\*  (0.0587730) | 0.120199\*  (0.0627942) | 0.121077\*  (0.0629012) |
| *Pedu* |  |  |  |  | 0.0485319  (0.0383042) | 0.0544581  (0.0404666) |
| *Poccu* |  |  |  |  |  | 0.0113782  (0.0247891) |
| const | 9.10971\*\*\*  (0.0365395) | 8.02290\*\*\*  (0.319702) | 8.02224\*\*\*  (0.320964) | 7.88839\*\*\*  (0.322694) | 7.79943\*\*\*  (0.329946) | 7.76557\*\*\*  (0.338490) |
| R2 | 0.017003 | 0.051481 | 0.051483 | 0.070012 | 0.074668 | 0.075281 |
| R2adj | 0.013960 | 0.045589 | 0.042619 | 0.058387 | 0.060164 | 0.057833 |
| F | 5.587101 | 8.738257 | 5.807717 | 6.022572 | 5.148236 | 4.314695 |
| p-value | 0.018685 | 0.000202 | 0.000709 | 0.000110 | 0.000148 | 0.000335 |
| AIC | 511.0284 | 501.4247 | 503.4239 | 499.0125 | 499.3811 | 501.1659 |
| BIC | 518.5961 | 512.7762 | 518.5592 | 517.9317 | 522.0841 | 527.6526 |
| HQ | 514.0487 | 505.9551 | 509.4644 | 506.5632 | 508.4419 | 511.7368 |

Hypothesis about the significance of the coefficients *gradchoice*, *Pedu*, *Poccu*

| F | 0.613439 |
| --- | --- |
| p-value | 0.606724 |

Regression model of the logarithm of starting wage (semilogarithmic)

| model | values |
| --- | --- |
| const | 7.89257\*\*\*  (0.321250) |
| *gender* | 0.271400\*\*\*  (0.0680590) |
| *gpa* | 0.331272\*\*\*  (0.0928877) |
| *hheconstat* | 0.147856\*\*\*  (0.0585954) |
| R2 | 0.069929 |
| R2adj | 0.061237 |
| F | 8.045015 |
| p-value | 0.000035 |
| AIC | 497.0413 |
| BIC | 512.1766 |
| HQ | 503.0818 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 0.430412 | 0.512 |
| cubes | 0.449568 | 0.503 |
| squares and cubes | 1.593053 | 0.205 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| *gender* | 1.294 |
| *gpa* | 1.304 |
| *gradchoice* | 1.045 |
| *hheconstat* | 1.160 |
| *Pedu* | 1.304 |
| *Poccu* | 1.151 |

White test

| R2 | 0.020092 |
| --- | --- |
| TR2 | 6.529953 |
| p-value | 0.479402 |

Glejser test

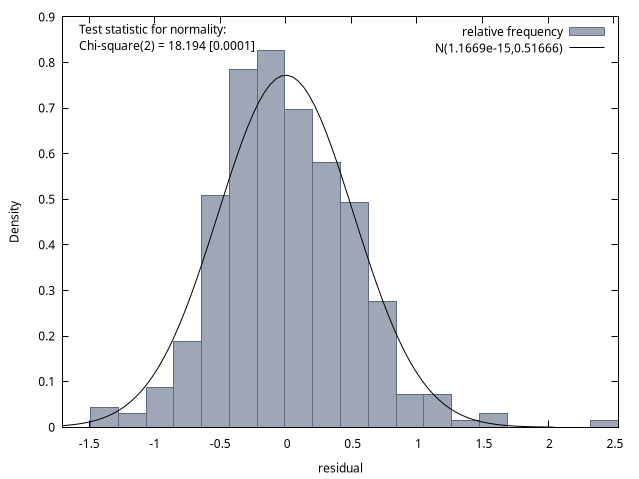
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 0.411118\*\*\*  (0.0221272) | −0.0268432  (0.0371980) |
| *gpa* | 0.175617  (0.167593) | 0.0689946  (0.0508753) |
| *hheconstat* | 0.359457\*\*\*  (0.0276587) | 0.0713700\*\*  (0.0359851) |

Breusch–Pagan test

| ESS | 16.6208 |
| --- | --- |
| LM | 8.310378 |
| p-value | 0.040014 |

Jarque-Bera test

| chi2 | 18.194 |
| --- | --- |
| p-value | 0.00011 |



Regression models of the logarithm of starting wage (logarithmic)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 0.145194\*\*\*  (0.0614264) | 0.249387\*\*\*  (0.0682469) | 0.249451\*\*\*  (0.0683722) | 0.266733\*\*\*  (0.0681639) | 0.269889\*\*\*  (0.0681501) | 0.268574\*\*\*  (0.0682996) |
| *ln(gpa)* |  | 0.958692\*\*\*  (0.290465) | 0.957339\*\*\*  (0.292975) | 0.996156\*\*\*  (0.290999) | 0.993327\*\*\*  (0.290750) | 0.990577\*\*\*  (0.291181) |
| *gradchoice* |  |  | 0.00188528  (0.0482948) | −0.00460508  (0.0479710) | −0.0116006  (0.0482530) | −0.0108844  (0.0483404) |
| *hheconstat* |  |  |  | 0.147661\*\*  (0.0588461) | 0.119737\*  (0.0628812) | 0.120585\*  (0.0629891) |
| *Pedu* |  |  |  |  | 0.0480323  (0.0383585) | 0.0537982  (0.0405262) |
| *Poccu* |  |  |  |  |  | 0.0110673  (0.0248262) |
| const | 9.10971\*\*\*  (0.0365395) | 7.94115\*\*\*  (0.355874) | 7.94199\*\*\*  (0.357064) | 7.80537\*\*\*  (0.358315) | 7.71926\*\*\*  (0.364543) | 7.68727\*\*\*  (0.371987) |
| R2 | 0.017003 | 0.049171 | 0.049175 | 0.067523 | 0.072084 | 0.072664 |
| R2adj | 0.013960 | 0.043265 | 0.040289 | 0.055867 | 0.057540 | 0.055167 |
| F | 5.587101 | 8.325901 | 5.533897 | 5.793027 | 4.956244 | 4.152950 |
| p-value | 0.018685 | 0.000298 | 0.001026 | 0.000164 | 0.000219 | 0.000492 |
| AIC | 511.0284 | 502.2152 | 504.2137 | 499.8809 | 500.2874 | 502.0843 |
| BIC | 518.5961 | 513.5667 | 519.3490 | 518.8000 | 522.9903 | 528.5711 |
| HQ | 514.0487 | 506.7456 | 510.2542 | 507.4316 | 509.3481 | 512.6552 |

Hypothesis about the significance of the coefficients gradchoice, Pedu, Poccu

| F | 0.590661 |
| --- | --- |
| p-value | 0.62152 |

Regression models of the logarithm of starting wage (logarithmic)

| model | values |
| --- | --- |
| const | 7.80768\*\*\*  (0.356954) |
| *gender* | 0.266852\*\*\*  (0.0680474) |
| *ln(gpa)* | 0.992780\*\*\*  (0.288420) |
| *hheconstat* | 0.147357\*\*  (0.0586698) |
| R2 | 0.067496 |
| R2adj | 0.058781 |
| F | 7.744869 |
| p-value | 0.000052 |
| AIC | 497.8903 |
| BIC | 513.0256 |
| HQ | 503.9308 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 0.229396 | 0.632 |
| cubes | 0.242731 | 0.623 |
| squares and cubes | 1.252919 | 0.287 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| *gender* | 1.290 |
| *gpa* | 1.295 |
| *gradchoice* | 1.040 |
| *hheconstat* | 1.160 |
| *Pedu* | 1.304 |
| *Poccu* | 1.152 |

White test

| R2 | 0.019356 |
| --- | --- |
| TR2 | 6.290648 |
| p-value | 0.506251 |

Glejser test

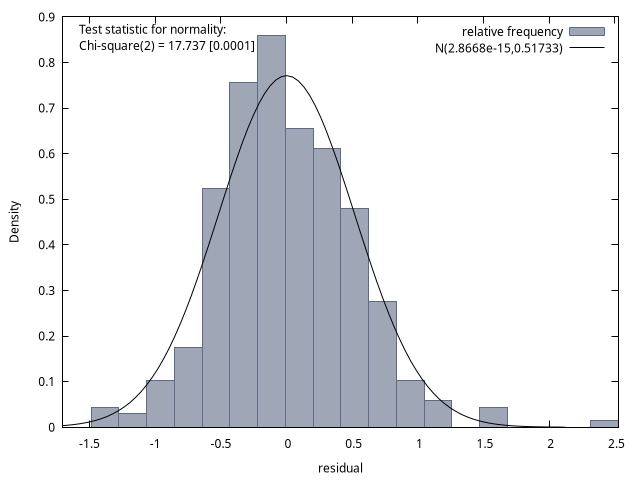
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 0.411794\*\*\*  (0.0221392) | −0.0267139  (0.0372182) |
| *ln(gpa)* | 0.161340  (0.187499) | 0.204162  (0.158123) |
| *hheconstat* | 0.360283\*\*\*  (0.0276745) | 0.0711932\*\*  (0.0360057) |

Breusch–Pagan test

| ESS | 16.4388 |
| --- | --- |
| LM | 8.219413 |
| p-value | 0.041688 |

Jarque-Bera test

| chi2 | 17.737 |
| --- | --- |
| p-value | 0.00014 |



Semi-logarithmic model is better than the logarithmic one in , F. To compare linear and semi-logarithmic models, we apply the Box-Cox test. The average of the logarithm of wage *lnawage* is 9.1611.

wagea\_star = wage/exp(9.1611)

lnawage\_star = ln(wagea\_star)

RSS(wagea\_star)=234.5160

RSS(lnawage\_star)=85.19306

chi2 = 325 / 2 \* ln( 234.6901 / 85.68609 ) = 163.731005 > x2\_1

H\_0 is rejected, i.e. the semi-logarithmic model is better (it has less RSS). It remains to eliminate heteroscedasticity. The white test does not reveal heteroscedasticity, so let's try to use the estimates of the variance of random components from the Glaser test to weight the model.

In our case, the random component will significantly depend only on *hheconstat*:

Divide each i-th observation for the dependent variable *lnawage* and all dependent factors − *gender*, *gpa* и *hheconstat* – by .

Result model

| model | values |
| --- | --- |
| const | 22.0297\*\*\*  (0.868425) |
| *w\_gender* | 0.259246\*\*\*  (0.0660904) |
| *w\_gpa* | 0.324788\*\*\*  (0.0904799) |
| *w\_hheconstat* | −1.42548\*\*\*  (0.0896578) |
| R2 | 0.701208 |
| R2adj | 0.698416 |
| F | 251.1090 |
| p-value | 7.52e-84 |
| AIC | 1084.868 |
| BIC | 1100.004 |
| HQ | 1090.909 |

White test

| R2 | 0.012208 |
| --- | --- |
| TR2 | 3.967541 |
| p-value | 0.783508 |

Glejser test

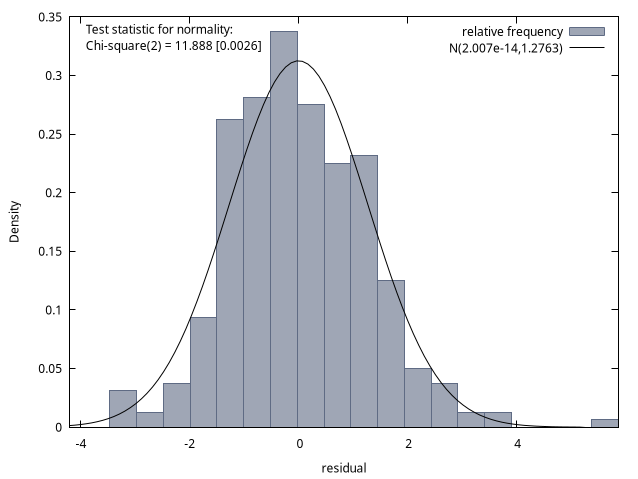
|  | const | coefficient |
| --- | --- | --- |
| *w\_gender* | 1.02282\*\*\* (0.0539045) | −0.0261077 (0.0355657) |
| *w\_gpa* | 0.686844\*\*  (0.312764) | 0.0380167  (0.0376767) |
| *w\_hheconstat* | 0.997761\*\*\*  (0.0679418) | 0.00117976  (0.0380830) |

Breusch–Pagan test

| ESS | 2.90546 |
| --- | --- |
| LM | 1.452728 |
| p-value | 0.693223 |

Jarque-Bera test

| chi2 | 11.888 |
| --- | --- |
| p-value | 0.00262 |



**SCIENCE AND ART/SOCIAL SCIENCE**

General information

|  | Mean | Median | S.D. | Min | Max |
| --- | --- | --- | --- | --- | --- |
| *gender* | 0.4088 | 0 | 0.4934 | 0 | 1 |
| *gpa* | 3.164 | 3.217 | 0.3821 | 1.931 | 3.949 |
| *wageb* | 6842 | 5472 | 3908 | 1609 | 22852 |
| *wagea* | 11269 | 10000 | 6636 | 2300 | 40000 |
| *lnbwage* | 8.696 | 8.607 | 0.5114 | 7.384 | 10.04 |
| *lnawage* | 9.184 | 9.210 | 0.5374 | 7.741 | 10.60 |
| *gradchoice* | 0.4891 | 0 | 0.6655 | 0 | 2 |
| *major* | 4.584 | 5 | 0.4947 | 4 | 5 |
| *hheconstat* | 0.5328 | 1 | 0.5008 | 0 | 1 |
| *Pedu* | 2.204 | 2 | 0.8237 | 1 | 3 |
| *Poccu* | 2.036 | 2 | 1.320 | 0 | 4 |

Correlation matrix

|  | *gender* | *gpa* | *gradchoice* | *hheconstat* | *Pedu* | *Poccu* |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 1 | -0,3614 | 0,1481 | -0,0845 | -0,0261 | 0,1237 |
| *gpa* |  | 1 | 0,0013 | -0,1350 | -0,1335 | 0,0455 |
| *gradchoice* |  |  | 1 | -0,0375 | 0,2187 | -0,1963 |
| *hheconstat* |  |  |  | 1 | 0,3758 | -0,0630 |
| *Pedu* |  |  |  |  | 1 | -0,3790 |
| *Poccu* |  |  |  |  |  | 1 |

Regression models of starting wage (linear)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 4256.64 \*\*\* (1098.03) | 5136.83 \*\*\*  (1163.14) | 5233.92 \*\*\*  (1181.43) | 5644.51 \*\*\* (1170.25) | −4318.96 \*\*\* (−4318.96) | 5750.01 \*\*\* (1197.50) |
| *gpa* |  | 3144.66 \*\*  (1501.8) | 3190.90 \*\*  (1508.69) | 3857.08 \*\*  (1503.40) | 3841.87 \*\* (1518.79) | 3902.44 \*\* (1526.45) |
| *gradchoice* |  |  | −421.291  (816.712) | −391.073  (801.093) | −371.855 (832.785) | −440.238 (843.566) |
| *hheconstat* |  |  |  | 2683.84 \*\*  (1071.98) | −371.855 \*\* (832.785) | 2785.23 \*\* (1167.46) |
| *Pedu* |  |  |  |  | −63.8753 (719.949) | −214.152 (768.705) |
| *Poccu* |  |  |  |  |  | −252.767 (444.662) |
| const | 9529.07 \*\*\*  (702.015) | −780.723  (4972.54) | −760.665  (4986.37) | −4481.22  (5111.25) | −4318.96  (−4318.96) | −3711.01 (5564.65) |
| R2 | 4256.64 | 0.128676 | 0.130416 | 0.169837 | 0.169887 | 0.171945 |
| R2adj | 0.093504 | 0.115671 | 0.110801 | 0.144681 | 0.138203 | 0.133727 |
| F | 15.02825 | 9.894479 | 6.648888 | 6.751228 | 5.361962 | 4.499070 |
| p-value | 0.000165 | 0.000098 | 0.000322 | 0.000056 | 0.000162 | 0.000357 |
| AIC | 2788.598 | 2786.187 | 2787.913 | 2783.558 | 2785.549 | 2787.209 |
| BIC | 2794.438 | 2794.947 | 2799.593 | 2798.157 | 2803.069 | 2807.649 |
| HQ | 2790.971 | 2789.747 | 2792.660 | 2789.491 | 2792.669 | 2795.515 |

Hypothesis about the significance of the coefficients *gradchoice*, *Pedu*, *Poccu*

| F | 0.188755 |
| --- | --- |
| p-value | 0.903898 |

Regression model of starting wage (linear)

| model | values |
| --- | --- |
| const | −4510.76 (5096.24) |
| *gender* | 5555.61 \*\*\* (1152.68) |
| *gpa* | 3816.13 \*\* (1496.75) |
| *hheconstat* | 2691.72 \*\* (1068.78) |
| R2 | 0.168338 |
| R2adj | 0.149579 |
| F | 8.973590 |
| p-value | 0.000019 |
| AIC | 2781.805 |
| BIC | 2793.485 |
| HQ | 2786.551 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 0.022916 | 0.88 |
| cubes | 0.035649 | 0.851 |
| squares and cubes | 0.092289 | 0.912 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| *gender* | 1.245 |
| *gpa* | 1.213 |
| *gradchoice* | 1.124 |
| *hheconstat* | 1.218 |
| *Pedu* | 1.429 |
| *Poccu* | 1.227 |

White test

| R2 | 0.072035 |
| --- | --- |
| TR2 | 9.868836 |
| p-value | 0.196129 |

Glejser test

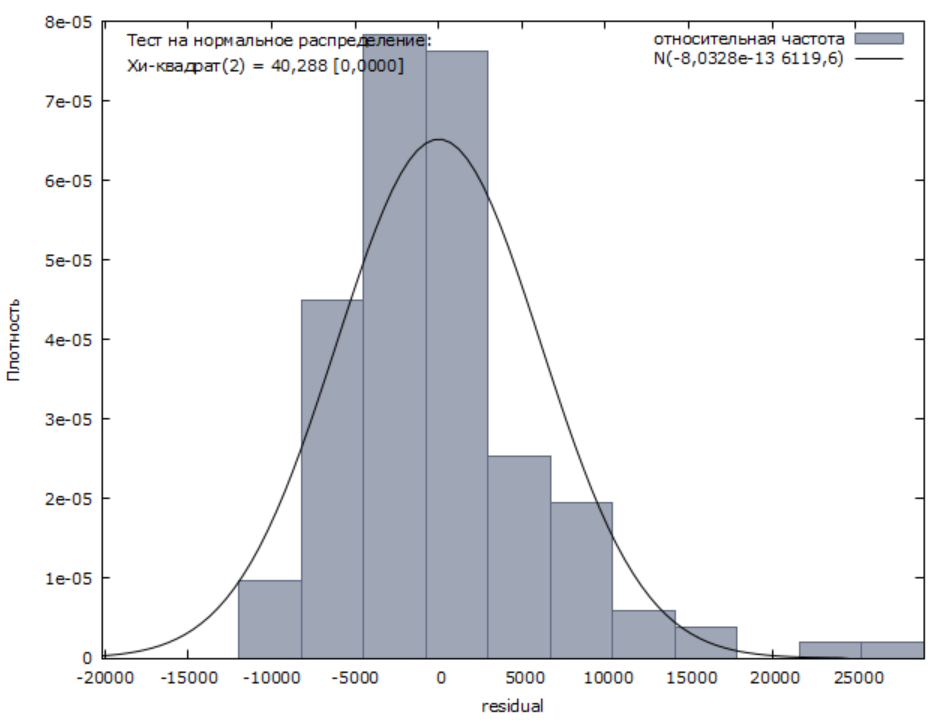
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 3301,62 \*\*\* (449,777) | 2501,22 \*\*\* (703,499) |
| *gpa* | 3879,54 (3027,21) | 140,474 (949,886) |
| *hheconstat* | 3528,14 \*\*\* (520,773) | 1493,63 \*\* (713,423) |

Breusch–Pagan test

| ESS | 0.196129 |
| --- | --- |
| LM | 26.434932 |
| p-value | 0.000008 |

Jarque-Bera test

| chi2 | 40,288 |
| --- | --- |
| p-value | 0,00000 |



Regression models of the logarithm of starting wage (semilogarithmic)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 0.350497 \*\*\* (0.0887475) | 0.420090 \*\*\* (0.0940765) | 0.431814 \*\*\* (0.0954380) | 0.464571 \*\*\* (0.0945904) | 0.461653 \*\*\* (0.0951387) | 0.468452 \*\*\* (0.0967838) |
| *gpa* |  | 0.248635 \*\* (0.121475) | 0.431814 \*\* (0.0954380) | 0.307366 \*\* (0.121519) | 0.301530 \*\* (0.122683) | 0.305171 \*\* (0.123370) |
| *gradchoice* |  |  | −0.0508748 (0.0659755) | −0.0484640 (0.0647517) | −0.0410900 (0.0672697) | −0.0452005 (0.0681784) |
| *hheconstat* |  |  |  | 0.214116 \*\* (0.0866472) | 0.228790 \*\* (0.0936331) | 0.232587 \*\* (0.0943562) |
| *Pedu* |  |  |  |  | −0.0245092 (0.0581553) | −0.0335425 (0.0621280) |
| *Poccu* |  |  |  |  |  | −0.0151942 (0.0359383) |
| const | 9.04095 \*\*\* (0.0567400) | 8.22580 \*\*\* (0.402186) | 8.22822 \*\*\* (0.402808) | 7.93140 \*\*\* (0.413138) | 7.99366 \*\*\* (0.439974) | 8.03020 \*\*\* (0.449744) |
| R2 | 0.171945 | 0.130747 | 0.134616 | 0.172880 | 0.174000 | 0.175134 |
| R2adj | 0.133727 | 0.117774 | 0.115096 | 0.147816 | 0.142473 | 0.137063 |
| F | 4.499070 | 10.07771 | 6.896358 | 6.897470 | 5.519121 | 4.600226 |
| p-value | 0.000357 | 0.000084 | 0.000237 | 0.000045 | 0.000121 | 0.000287 |
| AIC | 206.6326 | 204.4150 | 205.8038 | 201.6083 | 203.4227 | 205.2344 |
| BIC | 212.4725 | 213.1749 | 217.4838 | 216.2082 | 220.9426 | 225.6743 |
| HQ | 209.0058 | 207.9748 | 210.5503 | 207.5414 | 210.5423 | 213.5407 |

Hypothesis about the significance of the coefficients *gradchoice*, *Pedu*, *Poccu*

| F | 0.30282 |
| --- | --- |
| p-value | 0.823311 |

Regression model of the logarithm of starting wage (semilogarithmic)

| model | values |
| --- | --- |
| const | 7.92774 \*\*\* (0.412426) |
| *gender* | 0.453554 \*\*\* (0.0932836) |
| *gpa* | 0.302291 \*\* (0.121129) |
| *hheconstat* | 0.215093 \*\* (0.0864940) |
| R2 | 0.169370 |
| R2adj | 0.150634 |
| F | 9.039789 |
| p-value | 0.000017 |
| AIC | 200.1885 |
| BIC | 211.8684 |
| HQ | 204.9349 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 0.259290 | 0.611 |
| cubes | 0.260255 | 0.611 |
| squares and cubes | 0.133868 | 0.875 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| *gender* | 1.245 |
| *gpa* | 1.213 |
| *gradchoice* | 1.124 |
| *hheconstat* | 1.218 |
| *Pedu* | 1.229 |
| *Poccu* | 1.227 |

White test

| R2 | 0.050928 |
| --- | --- |
| TR2 | 6.977084 |
| p-value | 0.431270 |

Glejser test

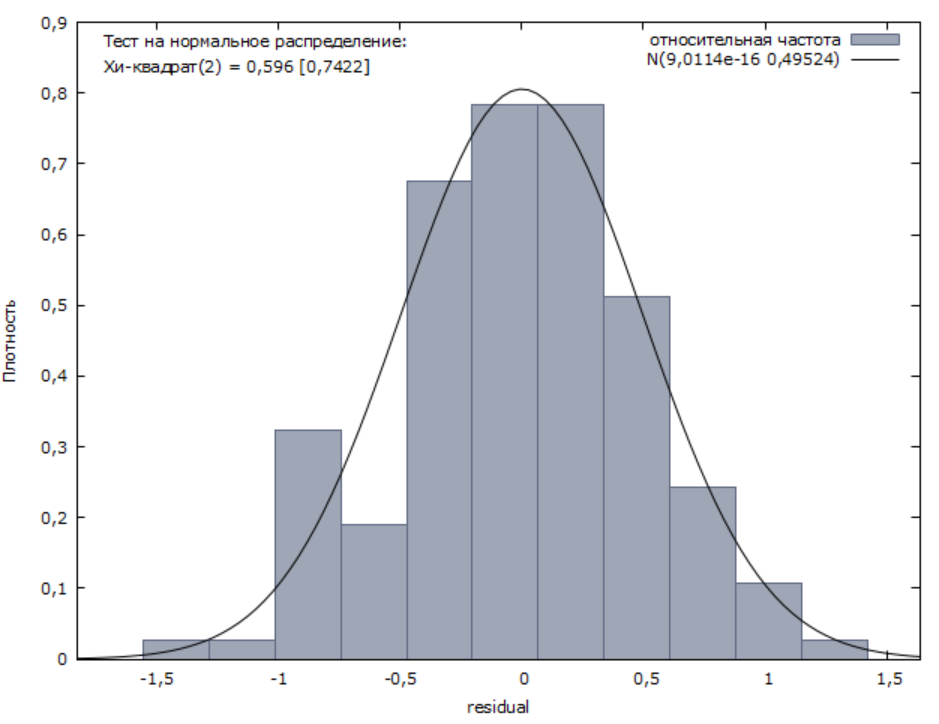
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 0,360351 \*\*\* (0,0332563) | 0,0624052 (0,0520165) |
| *gpa* | 0,432962 \*\* (0,215151) | −0,0148865 (0,0675106) |
| *hheconstat* | 0,329342 \*\*\* (0,0370173) | 0,106066 \*\* (0,0507112) |

Breusch–Pagan test

| ESS | 8.90594 |
| --- | --- |
| LM | 4.452972 |
| p-value | 0.216524 |

Jarque-Bera test

| chi2 | 0,596 |
| --- | --- |
| p-value | 0,74222 |



Regression models of the logarithm of starting wage (logarithmic)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
| --- | --- | --- | --- | --- | --- | --- |
| *gender* | 0.350497 \*\*\* (0.0887475) | 0.414818 \*\*\* (0.0940459) | 0.425884 \*\*\* (0.0953705) | 0.459120 \*\*\* (0.0945688) | 0.456143 \*\*\* (0.0950944) | 0.462582 \*\*\* (0.0967183) |
| *ln(gpa)* |  | 0.700538 \* (0.364851) | 0.714145 \* (0.365918) | 0.881625 \*\* (0.365507) | 0.863885 \*\* (0.368794) | 0.873737 \*\* (0.370778) |
| *gradchoice* |  |  | −0.0491457 (0.0660720) | −0.0465333 (0.0648555) | −0.0387900 (0.0673503) | −0.0427011 (0.0682547) |
| *hheconstat* |  |  |  | 0.214334 \*\* (0.0869751) | 0.229811 \*\* (0.0939456) | 0.233457 \*\* (0.0946769) |
| *Pedu* |  |  |  |  | −0.0258629 (0.0582434) | −0.0345389 (0.0622430) |
| *Poccu* |  |  |  |  |  | −0.0145626 (0.0360032) |
| const | 9.04095 \*\*\* (0.0567400) | 8.21315 \*\*\* (0.434776) | 8.21709 \*\*\* (0.435535) | 7.89640 \*\*\* (0.446829) | 7.96290 \*\*\* (0.472546) | 7.99774 \*\*\* (0.481827) |
| R2 | 0.171945 | 0.127574 | 0.131188 | 0.169401 | 0.170649 | 0.171692 |
| R2adj | 0.133727 | 0.114552 | 0.111591 | 0.144231 | 0.138995 | 0.133462 |
| F | 4.499070 | 9.797319 | 6.694194 | 6.730356 | 5.390973 | 4.491062 |
| p-value | 0.000357 | 0.000107 | 0.000305 | 0.000058 | 0.000153 | 0.000363 |
| AIC | 206.6326 | 204.9143 | 206.3456 | 202.1834 | 203.9773 | 205.8050 |
| BIC | 212.4725 | 213.6742 | 218.0255 | 216.7833 | 221.4972 | 226.2449 |
| HQ | 209.0058 | 208.4741 | 211.0920 | 208.1164 | 211.0970 | 214.1113 |

Hypothesis about the significance of the coefficients *gradchoice*, *Pedu*, *Poccu*

| F | 0.289307 |
| --- | --- |
| p-value | 0.833058 |

Regression model of the logarithm of starting wage (logarithmic)

| model | values |
| --- | --- |
| const | 7.89115 \*\*\* (0.445953) |
| *gender* | 0.448804 \*\*\* (0.0932987) |
| *ln(gpa)* | 0.869542 \*\* (0.364452) |
| *hheconstat* | 0.869542 \*\* (0.364452) |
| R2 | 0.166162 |
| R2adj | 0.147353 |
| F | 8.834439 |
| p-value | 0.000022 |
| AIC | 200.7166 |
| BIC | 212.3965 |
| HQ | 212.3965 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 0.385312 | 0.536 |
| cubes | 0.386351 | 0.535 |
| squares and cubes | 0.195710 | 0.195710 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| *gender* | 1.238 |
| *l\_gpa* | 1.209 |
| *gradchoice* | 1.121 |
| *hheconstat* | 1.222 |
| *Pedu* | 1.429 |
| *Poccu* | 1.227 |

White test

| R2 | 0.051689 |
| --- | --- |
| TR2 | 7.081397 |
| p-value | 0.420454 |

Glejser test

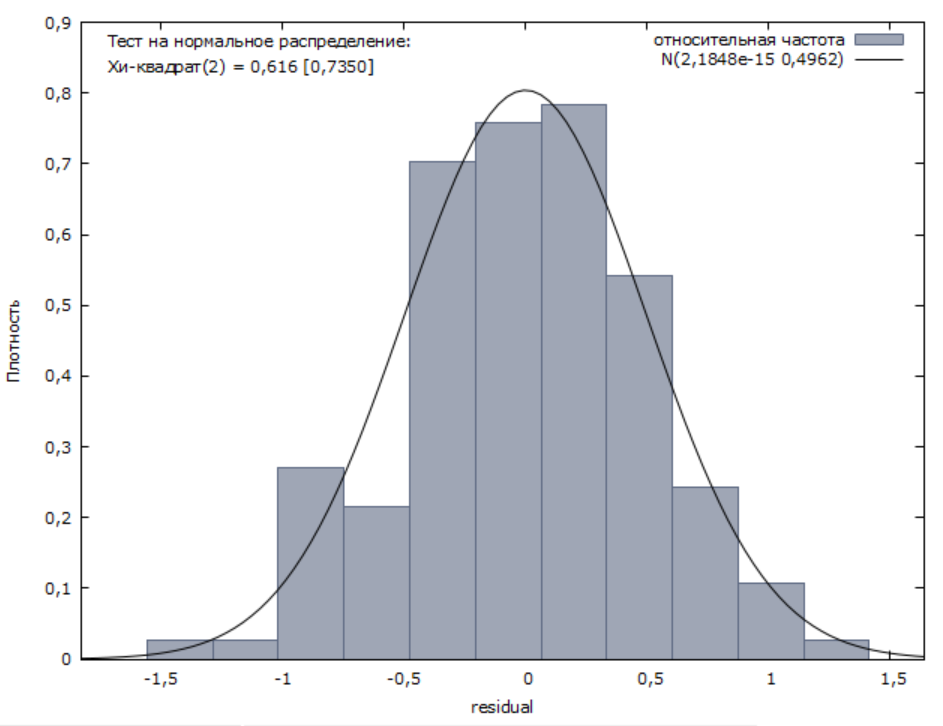
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 0,361675 \*\*\* (0,0332395) | 0,0623629 (0,0519901) |
| *ln(gpa)* | 0,420659 \* (0,233394) | −0,0292735 (0,202752) |
| *hheconstat* | 0,329199 \*\*\* (0,0369667) | 0,108787 \*\* (0,0506419) |

Breusch–Pagan test

| ESS | 9.55226 |
| --- | --- |
| LM | 4.776130 |
| p-value | 0.188943 |

Jarque-Bera test

| chi2 | 0,616 |
| --- | --- |
| p-value | 0,73505 |



Semi-logarithmic model is better than logarithmic one in , F. To compare linear and semi-logarithmic model, we apply the Box-Cox test. The average of the logarithm of wage *lnawage* is 9.1842.

wagea\_star = wage/exp(9.1842)

lnawage\_star = ln(wagea\_star)

RSS(wagea\_star)=52.48111

RSS(lnawage\_star)=32.62045

chi2 = 137 / 2 \* ln( 52.48111 / 32.62045 ) = 32.5727024 > x2\_1

H\_0 is rejected, i.e. the semi-logarithmic model is better (it has less RSS). It remains to eliminate heteroscedasticity. The white test does not reveal heteroscedasticity, so let's try to use the estimates of the variance of random components from the Glaser test to weight the model.

In our case, the random component will significantly depend only on *hheconstat*:

Divide each i-th observation for the dependent variable *lnawage* and all dependent factors − *gender*, *gpa* и *hheconstat* – by .

Result model

| model | values |
| --- | --- |
| const | 23.9369\*\*\*  (1.24638) |
| *w\_gender* | 0.481743\*\*\*  (0.0900628) |
| *w\_gpa* | 0.312087\*\*  (0.121141) |
| *hheconstat* | −2.32044\*\*\*  (0.175397) |
| R2 | 0.868627 |
| R2adj | 0.865664 |
| F | 293.1289 |
| p-value | 2.07e-58 |
| AIC | 460.7068 |
| BIC | 472.3867 |
| HQ | 465.4532 |

White test

| R2 | 0.032282 |
| --- | --- |
| TR2 | 4.422626 |
| p-value | 0.730013 |

Glejser test

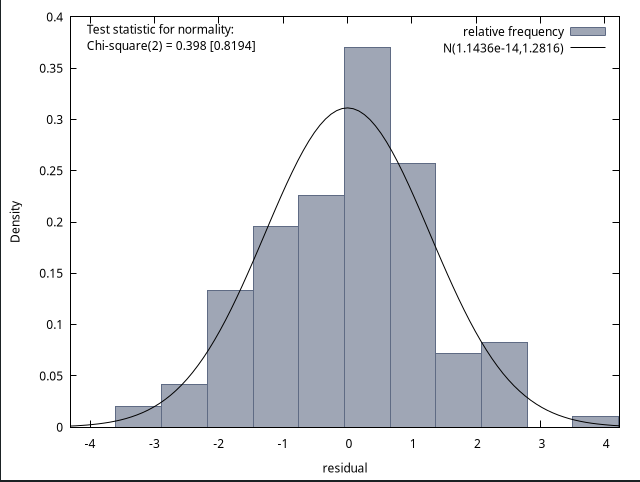
|  | const | coefficient |
| --- | --- | --- |
| *w\_gender* | 0.919590\*\*\*  (0.0849970) | 0.0737151  (0.0491457) |
| *w\_gpa* | 1.02052\*\*\*  (0.347296) | −0.00240899  (0.0406845) |
| *w\_hheconstat* | 0.999030\*\*\*  (0.0970351) | 0.00106630  (0.0578794) |

Breusch–Pagan test

| ESS | 4.78509 |
| --- | --- |
| LM | 2.392545 |
| p-value | 0.495024 |

Jarque-Bera test

| chi2 | 0.398 |
| --- | --- |
| p-value | 0.81939 |



**Сhou test**

| F | 1.1448 |
| --- | --- |
| p-value | 0.3348 |

**RESULT MODEL**

General information

|  | Mean | Median | S.D. | Min | Max |
| --- | --- | --- | --- | --- | --- |
| *gender* | 0.4150 | 0 | 0.4931 | 0 | 1 |
| *gpa* | 3.275 | 3.317 | 0.3845 | 1.931 | 4.218 |
| *wageb* | 7777 | 6566 | 4923 | 1609 | 64371 |
| *wagea* | 11954 | 10000 | 8360 | 1000 | 100000 |
| *lnbwage* | 8.820 | 8.790 | 0.5101 | 7.384 | 11.07 |
| *lnawage* | 9.224 | 9.210 | 0.5567 | 6.908 | 11.51 |
| *gradchoice* | 0.4901 | 0 | 0.6147 | 0 | 2 |
| *major* | 0.3456 | 0 | 0.4759 | 1 | 2 |
| *hheconstat* | 0.5807 | 1 | 0.4938 | 0 | 1 |
| *Pedu* | 2.280 | 2 | 0.7976 | 1 | 3 |
| *Poccu* | 2.004 | 2 | 1.257 | 0 | 4 |

Correlation matrix

|  | *gender* | *gpa* | *gradchoice* | *major* | *hheconstat* | *Pedu* | *Poccu* |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *gender* | 1 | -0.4065 | 0.0159 | 0.1253 | -0.0999 | -0.0439 | 0.0759 |
| *gpa* |  | 1 | 0.1549 | 0.1151 | 0.0088 | 0.0335 | 0.0228 |
| *gradchoice* |  |  | 1 | 0.1523 | 0.0470 | 0.1619 | -0.0642 |
| *major* |  |  |  | 1 | 0.0199 | 0.0881 | 0.0007 |
| *hheconstat* |  |  |  |  | 1 | 0.3530 | -0.1421 |
| *Pedu* |  |  |  |  |  | 1 | -0.3050 |
| *Poccu* |  |  |  |  |  |  | 1 |

Regression models of starting wage (linear)

|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *gender* | 2874.33\*\*\*  (629.799) | 4199.36\*\*\*  (678.774) | 0.345209\*\*\*  (0.0445589) | 0.329136\*\*\*  (0.0452062) | 4252.64\*\*\*  (687.935) | 4273.66\*\*\*  (687.653) | 4246.96 690.40342\*\*\*  (46.96 690.403) |
| *gpa* |  | 4179.87\*\*\*  (870.437) | 0.354462\*\*\*  (0.0578319) | 0.335897\*\*\*  (0.0584737) | 3992.36\*\*\*  (885.082) | 4001.25\*\*\*  (884.528) | 3973.36\*\*\*  (887.028) |
| *gradchoice* |  |  | 0.0149415  (0.0330523) | 0.00685853\*\*\*  (0.0332374) | 51.4871  (503.152) | −49.7418  (508.117) | −42.3753  (508.645) |
| *major* |  |  |  | 0.0855397  (0.0433228) | 1215.20\*  (655.322) | 1154.33\*  (656.370) | 1150.56\*  (656.787) |
| *hheconstat* |  |  |  |  | 2605.24\*\*\*  (615.727) | 2290.93\*\*\*  (655.913) | 2300.19\*\*\*  (656.578) |
| *Pedu* |  |  |  |  |  | 567.259  (409.951) | 621.192  (426.081) |
| *Poccu* |  |  |  |  |  |  | 118.491  (253.346) |
| const | 10761.5\*\*\*  (405.727) | −3476.33  (2991.75) | 7.91219\*\*\*  (0.196554) | 7.95405\*\*\*  (0.197292) | −4842.56  (3026.40) | −5920.83\*  (3123.19) | −6186.42\*  (3176.12) |
| R2 | 0.028737 | 0.059584 | 0.094249 | 0.099258 | 0.088512 | 0.091001 | 0.091286 |
| R2adj | 0.027357 | 0.056908 | 0.090378 | 0.094118 | 0.082001 | 0.083199 | 0.082173 |
| F | 20.82907 | 22.27066 | 24.34902 | 19.31180 | 13.59493 | 11.66303 | 10.01696 |
| p-value | 5.92e-06 | 4.19e-10 | 5.35e-15 | 4.38e-15 | 1.11e-12 | 1.78e-12 | 5.81e-12 |
| AIC | 14738.11 | 14717.33 | 1113.605 | 1111.689 | 14701.27 | 14701.34 | 14703.12 |
| BIC | 14747.23 | 14731.00 | 1131.843 | 1134.487 | 14728.63 | 14733.25 | 14739.59 |
| HQ | 14741.64 | 14722.61 | 1120.652 | 1120.498 | 14711.84 | 14713.67 | 14717.21 |

Hypothesis about the significance of the coefficients *gradchoice*, *Poccu*

| F | 0.114161 |
| --- | --- |
| p-value | 0.892131 |

Regression models of starting wage (linear)

| model | values |
| --- | --- |
| const | −5883.78\*  (3097.98) |
| *gender* | 4268.88\*\*\*  (685.432) |
| *gpa* | 3987.84\*\*\*  (873.244) |
| *major* | 1147.26\*  (651.923) |
| *hheconsta* | 2291.06\*\*\*  (655.447) |
| *Pedu* | 561.481  (405.392) |
| R2 | 0.090989 |
| R2adj | 0.084496 |
| F | 14.01354 |
| p-value | 4.48e-13 |
| AIC | 14699.35 |
| BIC | 14726.70 |
| HQ | 14709.92 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 1.414702 | 0.235 |
| cubes | 1.493923 | 0.222 |
| squares and cubes | 0.774619 | 0.461 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| *gender* | 1.273 |
| *gpa* | 1.278 |
| *gradchoice* | 1.074 |
| *major* | 1.074 |
| *hheconstat* | 1.155 |
| *Pedu* | 1.269 |
| *Poccu* | 1.115 |

White test

| R2 | 0.015887 |
| --- | --- |
| TR2 | 11.216511 |
| p-value | 0.845090 |

Glejser test

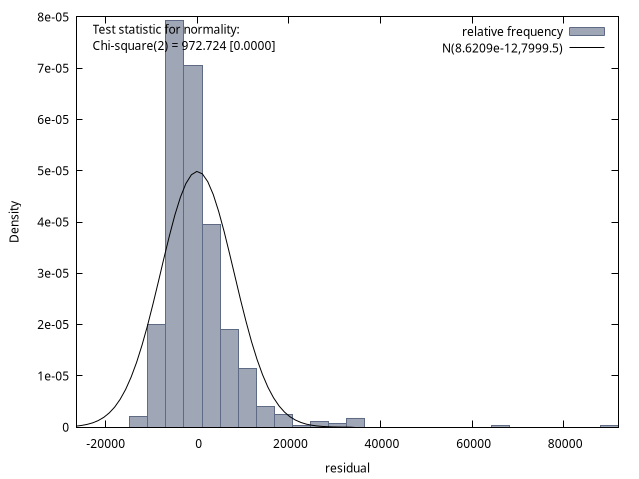
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 4417.53\*\*\*  (299.548) | 1610.25\*\*\*  (464.980) |
| *gpa* | 3590.34\*\*  (1981.89) | 456.667  (601.086) |
| *major* | 4722.18\*\*\*  (284.667) | 1052.12\*\*  (484.221) |
| *hheconstat* | 4010.70\*\*\*  (352.848) | 1851.27\*\*\*  (463.018) |
| *Pedu* | 3864.01\*\*\*  (698.606) | 535.768\*  (289.192) |

Breusch–Pagan test

| ESS | 170.557 |
| --- | --- |
| LM | 85.278572 |
| p-value | 0.00000 |

Jarque-Bera test

| chi2 | 972.724 |
| --- | --- |
| p-value | 0.00000 |



Regression models of the logarithm of starting wage (semilogarithmic)

|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *gender* | 0,233141 \*\*\* (0,0416352) | 0,346971 \*\*\* (0,0443631) | 0,345209 \*\*\* (0,0445589) | 0,329136 \*\*\* (0,0452062) | 0,350285 \*\*\* (0,0449961) | 0,352569 \*\*\* (0,0448690) | 0,351015 \*\*\* (0,0450499) |
| *gpa* |  | 0,359081 \*\*\* (0,0568898) | 0,354462 \*\*\* (0,0578319) | 0,335897 \*\*\* (0,0584737) | 0,347618 \*\*\* (0,0578910) | 0,348585 \*\*\* (0,0577150) | 0,346961 \*\*\* (0,0578800) |
| *gradchoice* |  |  | 0,0149415 (0,0330523) | 0,00685853 (0,0332374) | −0,000111581 (0,0329099) | −0,0111141 0,0331544 | −0,0106852 (0,0331899) |
| *major* |  |  |  | 0,0855397 \*\* (0,0433228) | 0,0796502 \* (0,0428629) | 0,0730338 \* (0,0428278) | 0,0728143 \* (0,0428564) |
| *hheconstat* |  |  |  |  | 0,165665 \*\*\* (0,0402732) | 0,131503 \*\*\* (0,0427980) | 0,132042 \*\*\* (0,0428428) |
| *Pedu* |  |  |  |  |  | 0,0616550 \*\* (0,0267491) | 0,0647951 \*\* (0,0278024) |
| *Poccu* |  |  |  |  |  |  | 0,00689879 (0,0165312) |
| const | 9,12679 \*\*\* (0,0268221) | 7,90366 \*\*\* (0,195534) | 7,91219 \*\*\* (0,196554) | 7,95405 \*\*\* (0,197292) | 7,81614 \*\*\* (0,197949) | 7,69894 \*\*\* (0,203787) | 7,68348 \*\*\* (0,207247) |
| R2 | 0,042640 | 0,093985 | 0,094249 | 0,099258 | 0,120518 | 0,127152 | 0,127370 |
| R2adj | 0,041280 | 0,091407 | 0,090378 | 0,094118 | 0,114236 | 0,119660 | 0,118618 |
| F | 31,35569 | 36,46262 | 24,34902 | 19,31180 | 19,18458 | 16,97110 | 14,55435 |
| p-value | 3,08e-08 | 8,57e-16 | 5,35e-15 | 4,38e-15 | 6,45e-18 | 2,36e-18 | 9,43e-18 |
| AIC | 1148,727 | 1111,810 | 1113,605 | 1111,689 | 1096,826 | 1093,480 | 1095,304 |
| BIC | 1157,846 | 1125,489 | 1131,843 | 1134,487 | 1124,184 | 1125,397 | 1131,781 |
| HQ | 1152,251 | 1117,096 | 1120,652 | 1120,498 | 1107,397 | 1105,813 | 1109,399 |

Hypothesis about the significance of the coefficients *gradchoice*, *Poccu*

| F | 0,143198 |
| --- | --- |
| p-value | 0,866608 |

Regression model of the logarithm of starting wage (semilogarithmic)

| model | values |
| --- | --- |
| const | 7,70722 \*\*\* (0,202157) |
| *gender* | 0,351501 \*\*\* (0,0447274) |
| *gpa* | 0,345589 \*\*\* (0,0569829) |
| *major* | 0,0714540 \* (0,0425408) |
| *hheconstat* | 0,131533 \*\*\* (0,0427707) |
| *Pedu* | 0,0603640 \*\* (0,0264536) |
| R2 | 0,127012 |
| R2adj | 0,120776 |
| F | 20,36868 |
| p-value | 5,20e-19 |
| AIC | 1091,594 |
| BIC | 1118,951 |
| HQ | 1102,165 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 0,022198 | 0,882 |
| cubes | 0,023181 | 0,879 |
| squares and cubes | 0,050230 | 0,951 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| *gender* | 1,273 |
| *gpa* | 1,278 |
| *gradchoice* | 1,074 |
| *major* | 1,074 |
| *hheconstat* | 1,155 |
| *Pedu* | 1,269 |
| *Poccu* | 1,115 |

White test

| R2 | 0,030902 |
| --- | --- |
| TR2 | 21,816809 |
| p-value | 0,191867 |

Glejser test

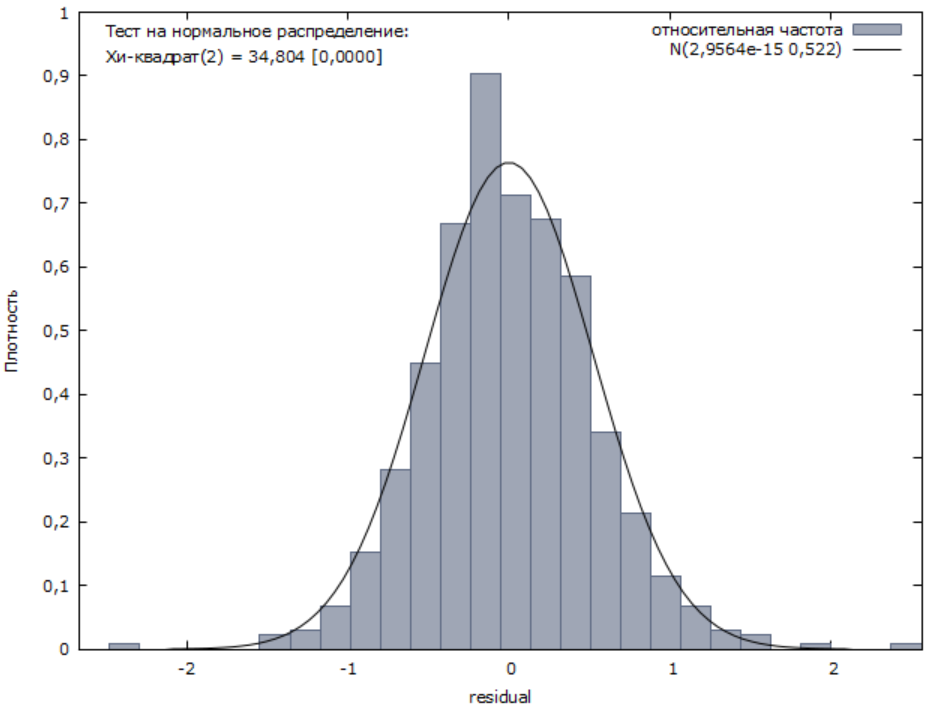
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 0,391489 \*\*\* (0,0160577) | 0,0316928 0,0249259 |
| *gpa* | 0,424876 \*\*\* (0,105510) | −0,00617874 0,0320000 |
| *major* | 0,402166 \*\*\* (0,0151989) | 0,00716473 0,0258535 |
| *hheconstat* | 0,364003 \*\*\* (0,0188827) | 0,0699779 \*\*\* (0,0247784) |
| *Pedu* | 0,369906 \*\*\* (0,0372421) | 0,0152323 0,0154166 |

Breusch–Pagan test

| ESS | 29,5806 |
| --- | --- |
| LM | 14,790320 |
| p-value | 0,011297 |

Jarque-Bera test

| chi2 | 34,804 |
| --- | --- |
| p-value | 0.00000 |



Regression models of the logarithm of starting wage (logarithmic)

|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *gender* | 0.233141\*\*\*  (0.0416352) | 0.342752\*\*\*  (0.0444290) | 0.340661\*\*\*  (0.0445976) | 0.324275\*\*\*  (0.0451848) | 0.345876\*\*\*  (0.0449824) | 0.347954\*\*\*  (0.0448580) | 0.346404\*\*\*  (0.0450438) |
| *ln(gpa)* |  | 1.07664\*\*\* (0.177011) | 1.05982\*\*\*  (0.179478) | 1.00414\*\*\*  (0.181045) | 1.04421\*\*\*  (0.179273) | 1.04551\*\*\*  (0.178741) | 1.04033\*\*\*  (0.179293) |
| *gradchoice* |  |  | 0.0190606  (0.0330303) | 0.0102432  (0.0332241) | 0.00323202  (0.0328944) | −0.00759983  (0.0331400) | −0.00718376  (0.0331753) |
| *major* |  |  |  | 0.0900383\*\*  (0.0433005) | 0.0841056\*  (0.0428378) | 0.0776298\*  (0.0428051) | 0.0774067\*  (0.0428340) |
| *hheconstat* |  |  |  |  | 0.166489\*\*\*  (0.0403419) | 0.132683\*\*\*  (0.0428761) | 0.133204\*\*\*  (0.0429205) |
| *Pedu* |  |  |  |  |  | 0.0609812\*\*  (0.0267898) | 0.0640735\*\*  (0.0278473) |
| *Poccu* |  |  |  |  |  |  | 0.00678735  (0.0165605) |
| const | 9.12679\*\*\*  (0.0268221) | 7.81208\*\*\*  (0.217730) | 7.82344\*\*\* (0.218720) | 7.86908\*\*\*  (0.219305) | 7.72167\*\*\*  (0.219762) | 7.60739\*\*\*  (0.224786) | 7.59307\*\*\*  (0.227620) |
| R2 | 0.042640 | 0.090501 | 0.090933 | 0.096506 | 0.117966 | 0.124456 | 0.124667 |
| R2adj | 0.041280 | 0.087914 | 0.087048 | 0.091350 | 0.111666 | 0.116941 | 0.115889 |
| F | 31.35569 | 34.97670 | 23.40668 | 18.71908 | 18.72410 | 16.56020 | 14.20156 |
| p-value | 3.08e-08 | 3.30e-15 | 1.90e-14 | 1.24e-14 | 1.72e-17 | 6.65e-18 | 2.63e-17 |
| AIC | 1148.727 | 1114.519 | 1116.184 | 1113.843 | 1098.871 | 1095.657 | 1097.487 |
| BIC | 1157.846 | 1128.198 | 1134.423 | 1136.641 | 1126.229 | 1127.574 | 1133.964 |
| HQ | 1152.251 | 1119.805 | 1123.232 | 1122.652 | 1109.442 | 1107.990 | 1111.582 |

Hypothesis about the significance of the coefficients *gradchoice*, *Poccu*

| F | 0.110252 |
| --- | --- |
| p-value | 0.895624 |

Regression models of the logarithm of starting wage (logarithmic)

| model | values |
| --- | --- |
| const | 7.61319\*\*\*  (0.223210) |
| *gender* | 0.347286\*\*\*  (0.0447331) |
| *ln(gpa)* | 1.03970\*\*\*  (0.176815) |
| *major* | 0.0765055\*  (0.0424946) |
| *hheconstat* | 0.132701\*\*\*  (0.0428470) |
| *Pedu* | 0.0600991\*\*  (0.0264942) |
| R2 | 0.124391 |
| R2adj | 0.118136 |
| F | 19.88865 |
| p-value | 1.44e-18 |
| AIC | 1093.710 |
| BIC | 1121.068 |
| HQ | 1104.281 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 0.023779 | 0.877 |
| cubes | 0.024857 | 0.875 |
| squares and cubes | 0.054756 | 0.947 |

Multicollinearity test

| factor | VIF |
| --- | --- |
| *gender* | 1.269 |
| *gpa* | 1.267 |
| *gradchoice* | 1.070 |
| *major* | 1.069 |
| *hheconstat* | 1.156 |
| *Pedu* | 1.269 |
| *Poccu* | 1.115 |

White test

| R2 | 0.031579 |
| --- | --- |
| TR2 | 22.294847 |
| p-value | 0.173644 |

Glejser test

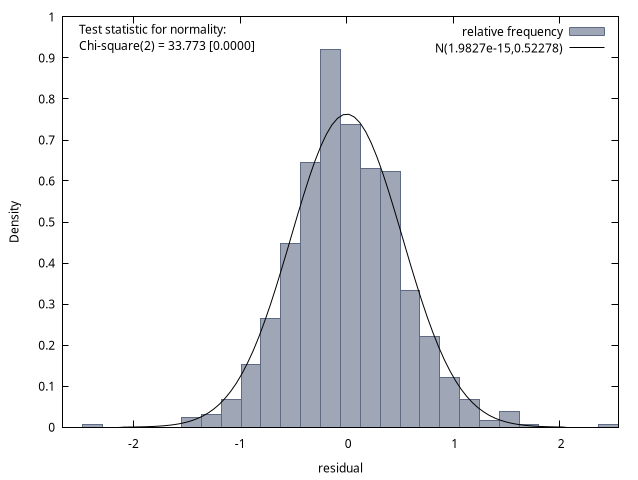
|  | const | coefficient |
| --- | --- | --- |
| *gender* | 0.392174\*\*\*  (0.0160615) | 0.0322764  (0.0249319) |
| *ln(gpa)* | 0.429580\*\*\*  (0.117879) | −0.0203672  (0.0994471) |
| *major* | 9.16794\*\*\*  (0.0256721) | 0.160876\*\*\*  (0.0436686) |
| *hheconstat* | 0.364511\*\*\*  (0.0188858) | 0.0707002\*\*\*  (0.0247826) |
| *Pedu* | 0.368831\*\*\* (0.0372495) | 0.0161104  (0.0154196) |

Breusch–Pagan test

| ESS | 29.2221 |
| --- | --- |
| LM | 14.611048 |
| p-value | 0.012160 |

Jarque-Bera test

| chi2 | 33.773 |
| --- | --- |
| p-value | 0.00000 |



Semi-logarithmic model is better than the logarithmic one in , F. To compare linear and semi-logarithmic models, we apply the Box-Cox test. The average of the logarithm of wage lnawage is 9.224.

wagea\_star = wage/exp(9.224)

lnawage\_star = ln(wagea\_star)

RSS(wagea\_star)=435,8703

RSS(lnawage\_star)=190,7393

chi2 = 706 / 2 \* ln( 435,8703 / 190,7393 ) = 291.732313 > x2\_1

H\_0 is rejected, i.e. the semi-logarithmic model is better (it has less RSS). It remains to eliminate heteroscedasticity. The white test does not reveal heteroscedasticity, so let's try to use the estimates of the variance of random components from the Glaser test to weight the model.

In our case, the random component significantly depends only on *hheconstat*:

Divide each i-th observation for the dependent variable *lnawage* and all dependent factors – *gender*, *gpa*, *major*, *hheconstat* и *Pedu* – by .

Result model

| model | values |
| --- | --- |
| const | 21.2691\*\*\*  (0.546586) |
| *w\_gender* | 0.357434\*\*\*  (0.0437788) |
| *w\_gpa* | 0.336002\*\*\*  (0.0562017) |
| *w\_major* | 0.0723000\*  (0.0418832) |
| *w\_hheconstat* | −1.35428\*\*\*  (0.0578743) |
| *w\_Pedu* | 0.0570148\*\*  (0.0256814) |
| R2 | 0.689223 |
| R2adj | 0.687003 |
| F | 310.4836 |
| p-value | 6.5e-175 |
| AIC | 2362.919 |
| BIC | 2390.277 |
| HQ | 2373.490 |

White test

| R2 | 0.021919 |
| --- | --- |
| TR2 | 15.474658 |
| p-value | 0.561333 |

Glejser test

|  | const | coefficient |
| --- | --- | --- |
| *w\_gender* | 0.965765\*\*\*  (0.0391147) | 0.0329955  (0.0240403) |
| *w\_gpa* | 1.04421  (0.208363) | −0.00539599  (0.0252889) |
| *w\_major* | 0.992372\*\*\*  (0.0370546) | 0.00913753  (0.0252780) |
| *w\_hheconstat* | 0.999952  (0.0463905) | 0.000197524  (0.0264186) |
| *w\_Pedu* | 1.01127\*\*\*  (0.0935796) | −0.00196773  (0.0157791) |

Breusch–Pagan test

| ESS | 9.30357 |
| --- | --- |
| LM | 4.651783 |
| p-value | 0.459836 |

Ramsey RESET test

|  | F | p-value |
| --- | --- | --- |
| squares | 0,004502 | 0,947 |
| cubes | 0,003718 | 0,951 |
| squares and cubes | 0,099065 | 0,906 |

Jarque-Bera test

| chi2 | 23.695 |
| --- | --- |
| p-value | 0.00001 |

