

**School of Business and Management**

BUS005 & BUS135

Project Handbook

Dr. Tom Kemeny

March 16, 2020

Table of contents

1. Instructions…………………………………………………………….2
2. Introduction to the Understanding Society Dataset……………………6
3. Measuring Social Class………………………………………………..6
4. Key Variables Analyzed………………………………………………7
5. Raw Stata Output……………………………………………………...8
6. Diagnostics……………………………………………………............30

Instructions

This assignment remains in many ways as it was before. It is to be completed individually. It is worth 40% of your final mark. There remains a 2,000 word limit (I still do not believe in +10%; I really mean I stop reading at 2,000 words).

One difference is that this project is now due April 23, 2020 at 23:55pm. This is two extra weeks.

The main difference from the original assignment is that, rather than undertaking a statistical analysis with Stata on the source of pay gaps of your choice, I have done the analysis. It is on the relationship between social class and pay. Your research question is:

Is there a relationship between one’s social class and pay?

To answer this question, I have conducted an analysis with descriptive and inferential components. I am providing you with all the puzzle pieces of this analysis, along with Stata code to help you understand exactly what I am doing. Your task is to complete the puzzle, by assembling whatever pieces are useful for you to write a careful and clear writeup of these results. The weekly lab outline documents will be very helpful to support you in this effort.

At a minimum, your report should present and interpret the following:

* a basic table of descriptive statistics for all included variables
* descriptive and inferential bivariate analysis
* descriptive and inferential multivariate analysis
* diagnostics

Please note that, unlike a jigsaw puzzle, you need not use ALL the pieces of evidence included at the end of this document. Many will be helpful to think through what I am doing – but that does not mean all are important to write about. As before, if you choose to include results in a table or figure, you must write about them.

To help you do all this, you must first read three peer-reviewed pieces of scientific research whose aim is to understand the links between social class and pay. These should be quantitative studies (not ones mainly based on interviews etc). I have suggested one of the three articles already:

Laurison, D., & Friedman, S. (2016). The class pay gap in higher professional and managerial occupations. *American Sociological Review*, *81*(4), 668-695.

You may find it easier to find the two additional articles by mining the bibliography of this article. I find Google Scholar helpful as well (see in particular their ‘Cited by’ link).

The structure of this writeup remains unchanged. To remind you: it must be structured like a mini-scientific article with the following major sections:

1. Introduction
2. Prior knowledge
3. Data and Methods
4. Results
5. Conclusion
6. References
7. Tables and Figures

Reproduced below from the original project instructions are section-specific notes on the project structure. Only minimal changes have been made here, as required by our change in circumstance.

**1) Introduction**

The introduction section introduces the problem or question to be studied. It usually briefly tells us the state of knowledge on the issue, and why and for whom the topic is an important one.

**2) Prior knowledge**

In this section you will succinctly synthesize what the studies you examined tell us about your particular question or hypotheses.

*Components:*

1. Theory: What does the work you read say in terms of a theory or model telling you why/how your independent variable of interest shapes your dependent variable?
2. Empirics: Write carefully about what the studies you read (plus other work they cite in their own literature review sections) show in terms evidence regarding your question or questions related to it.

**3) Data and Methods**

The purpose of this section is to primarily address two questions: 1) How was the data collected or generated? 2) How was it analysed? Since there are many different ways that research can be done, you must provide some justification for the choices that you make.

*Components:*

1. Briefly describe the data source
2. Identify, describe, and justify the main study variables (dependent, independent, and controls) to be analyzed. Justification for these typically comes from your wider reading, should be cited, and should include a logic. One additional reason to conduct your literature review is to identify possible control variables, based on what other researchers have used in their work. In our case the controls have be pre-selected, but you can briefly articulate why they might be useful to include.
3. Describe the type of statistical analysis undertaken, and why it is appropriate.

**4) Results**

The purpose of the results section is to present your key results in a logical sequence. There should be only a little interpretation here – just a statement of what the results are, as opposed to deep thinking about what they mean.

At a minimum, you should present and interpret the following:

* a basic table of descriptive statistics for all included variables;
* descriptive and inferential bivariate analysis;
* descriptive and inferential multivariate analysis; and,
* diagnostics.

Make sure you interpret key coefficients of the important regression models.

*Tips for a good Results section*

* Be selective. The point of this section is to provide statistical results that address your research question – not to present as many figures and tables as you can.
  + As a rule, do not include any figures or tables that you do not describe in words in the body of your report.
  + This is especially important since the word limit is tight. You need to be judicious about what matters and what does not.
* Make sure the numbers in your tables/figures and your text are the same.
* Make sure that the results you present directly relate to your research question.
* It is very useful to carefully read (and to *loosely* mimic) the way the studies you read phrase their results sections. I DO NOT MEAN PLAGIARIZE.
* Also remember that there are examples in the lab outlines and in my slides on how to talk about your findings.

**5) Conclusion**

The purpose of the conclusion is to interpret and describe the significance of your findings in light of what was already known about the research problem being investigated from your literature review, and to explain any new understanding or insights about the problem generated from your findings. Thus, the conclusion should connect to the introduction by way of the research questions or hypotheses you posed and the literature you reviewed. The conclusion is also your chance to acknowledge the limitations of your study.

*Components*

1. *Briefly* summarise the main findings of the study.
2. Discuss the relationship between your results and those found in prior research, including potential explanations of any discrepancies and unexpected findings.
3. Analyse the strengths and limitations of your study/analysis.
4. *Briefly* discuss wider implications of the results.

*Tips for the Conclusion:*

* Do not introduce new results in this section.
* When discussing study limitations, go beyond simply listing it to think about what it means in terms of the knowledge you have generated.

**6) References (not counted as part of word limit)**

Include bibliographic references for the articles that you summarise, as well as any resources related to the methods that you use.

**7) Tables and Figures (not counted as part of word limit)**

Put all your tables and figures (each numbered consecutively) in this final section.

*Tips for the tables and figures:*

* Tables should not be pasted directly from Stata.
* Tables and figures should be well-formatted
* Tables and figures should include clear titles, as well relevant notes.
* Tables and figures should be numbered consecutively (ie Table 1, 2, 3…; Figure 1, 2, 3…)
* Does not present estimates with a million decimal places if they are not useful. Better to round to a few decimal places - usually not more than three unless more is important. This is a substantive issue, so think of it in those terms. In other words, we don’t care about one-tenth of a pence in real life, so don’t show me that in your report.

# Introduction to the Understanding Society Dataset

In this project you will analyse results from the *Understanding Society study*, or the United Kingdom Household Longitudinal Study (UKHLS). *Understanding Society* is a panel study started in 2009 with the aims of understanding social and economic change in Britain both at the household and the individual levels. The study started with a large nationally representative sample of over 28,000 households throughout the United Kingdom. Since 2009, respondents and their household members have been interviewed annually. *Understanding Society* includes a range of measures on respondents, including measures of socio-demographics and economic status.

For information on the study, refer to [www.understandingsociety.ac.uk](http://www.understandingsociety.ac.uk).

Given the complexity of *Understanding Society*, I have used just a limited subset of the data for BUS005 and 135.

Most of the data come from wave 8 of the *Understanding Society,* gatheredbetween January 2016 and June 2018.

Further information on the variables included in the dataset, including the respondents who were asked each question, as well as the wording for each question, can be found on the *Understanding Society* website here: <https://www.understandingsociety.ac.uk/documentation/mainstage>

Measuring social class

Using the underlying data on Understanding Society, I have built several measures of social class, following the general approach taken by Laurison and Friedman (2016) and others. Specifically, I use information about parental occupations to capture social class. If an individual has parents who worked in higher-status occupations, I consider them to be higher-class. I build separate measures using information on father’s and mothers’ occupation (pa\_occ and ma\_occ, respectively). To build a household measure of class capturing both parents’ occupations, I also create a measure of the average occupational score across both parents, called *pclass*. As an additional possible measure of class, I built a measure of the number of parents (max 2 ☺) that have a university degree.

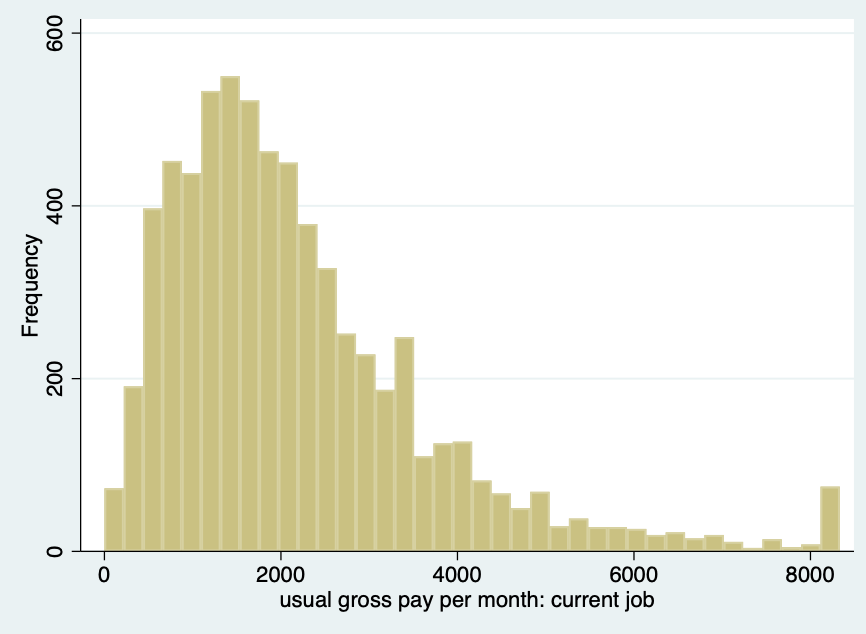
Key Variables Analyzed

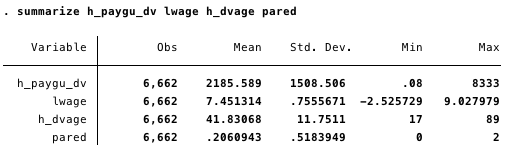
|  |  |
| --- | --- |
| **Variable name** | **Description** |
| h\_paygu\_dv | Monthly wages in £ from current job |
| lwage | Log of monthly wages from current job |
| pa\_occ | Father’s occupation |
| ma\_occ | Mother’s occupation |
| pclass | Mean of father and mother’s occupation |
| pared | Number of parents with a university degree |
| female | Sex (Male=0; Female=1) |
| h\_dvage | Age |
| eth | Ethnic group |
| h\_hiqual\_dv | Highest educational qualification |
| native | Non-UK born=0; UK born=1 |
| occ | Occupation |

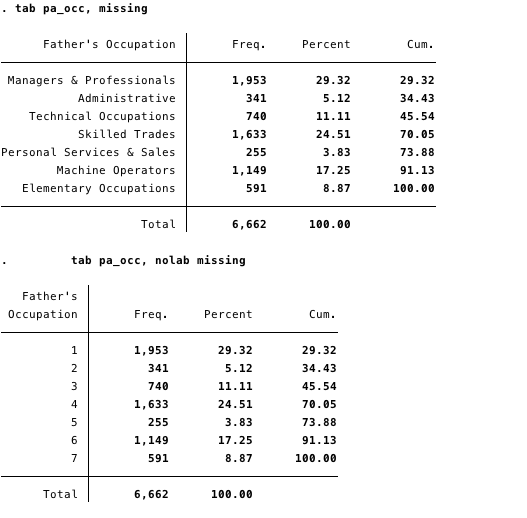
Raw Stata Output

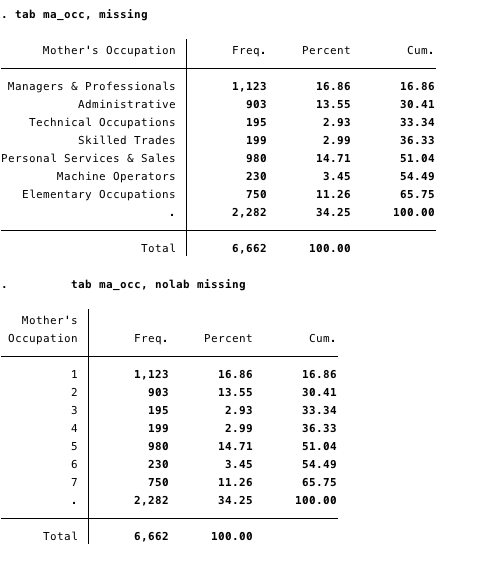
Feel free to copy and paste *figures* into your project documents, but for any tabular data, you must create well-formatted tables in MS Word by hand – see directions on QMPlus site in Module Tools. Any figures should also be accompanied by proper titles and notes.

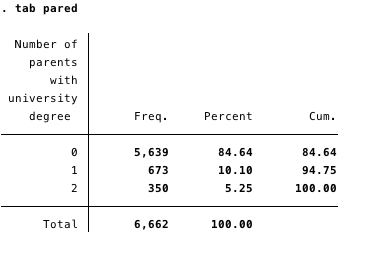
.histogram h\_paygu\_dv, freq

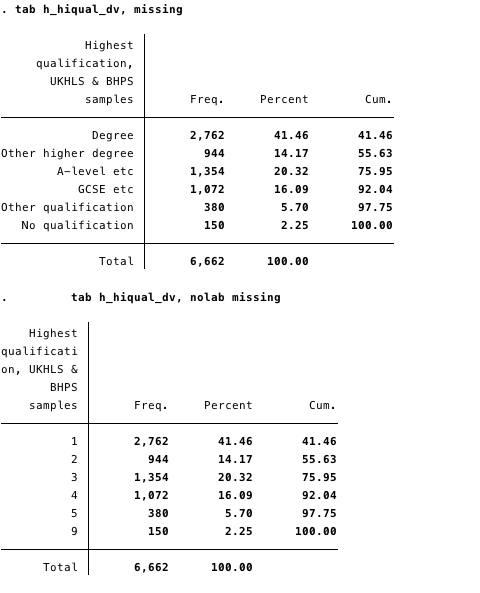


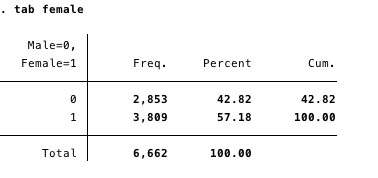


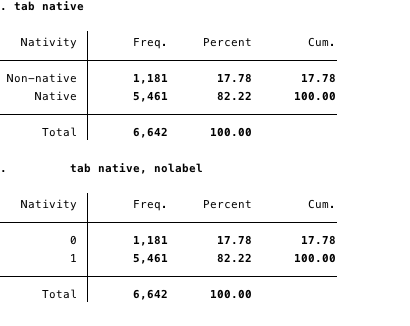


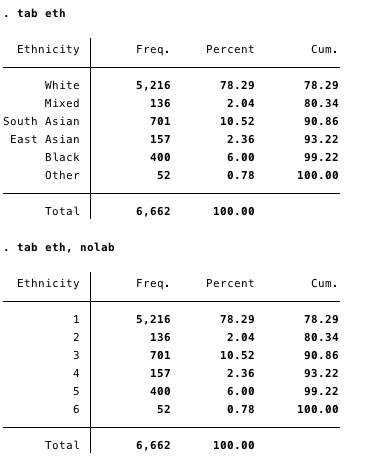


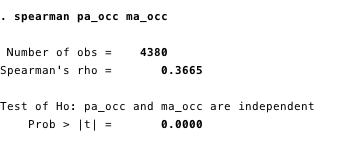


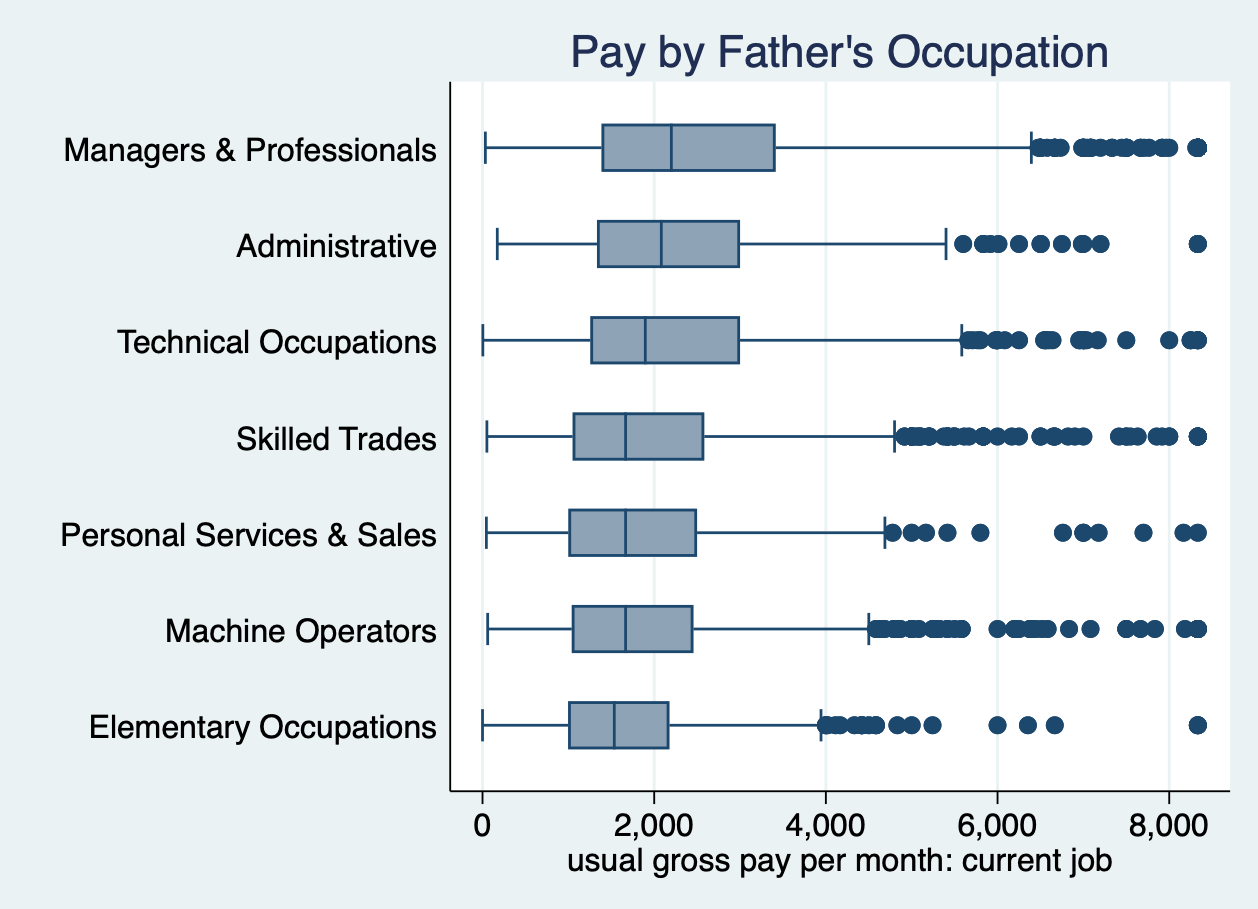


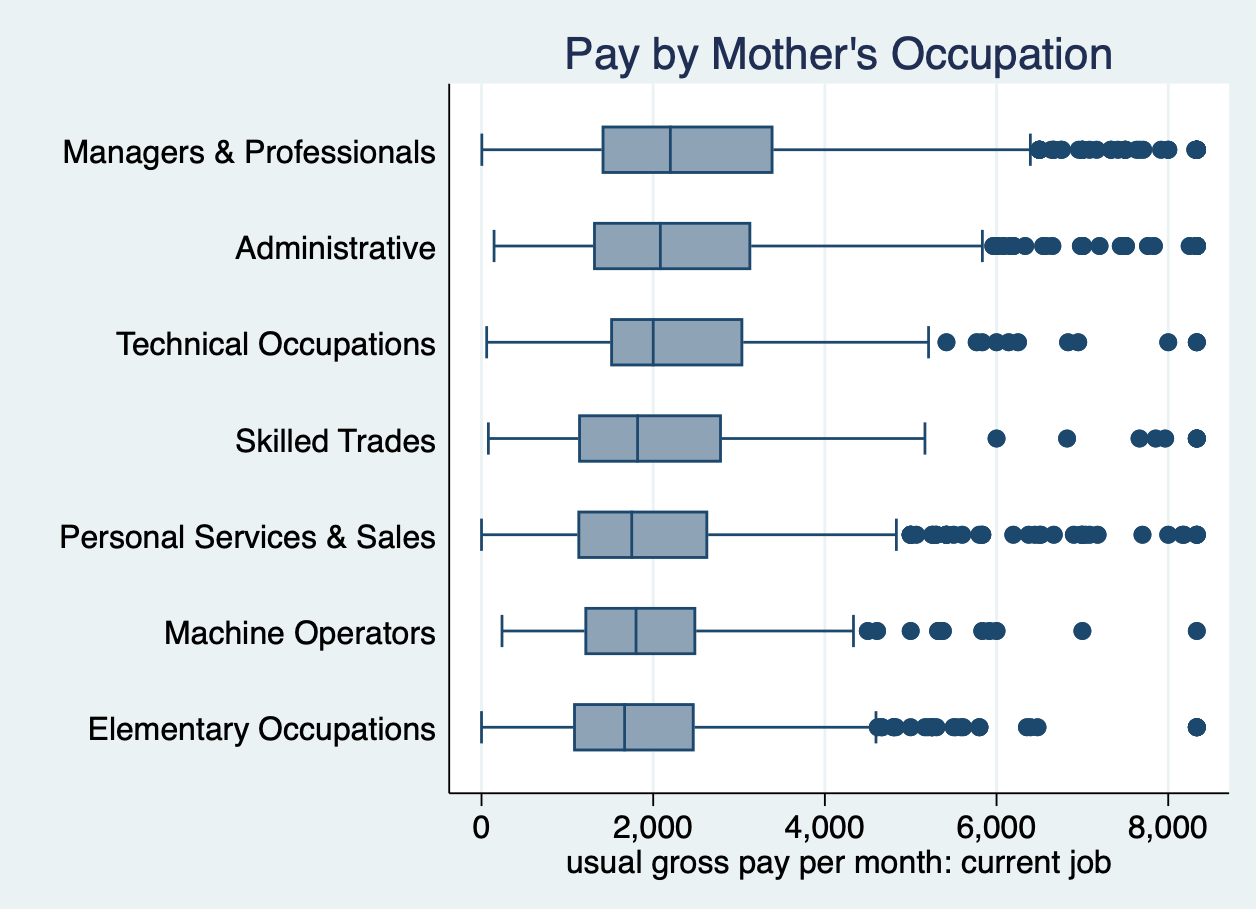


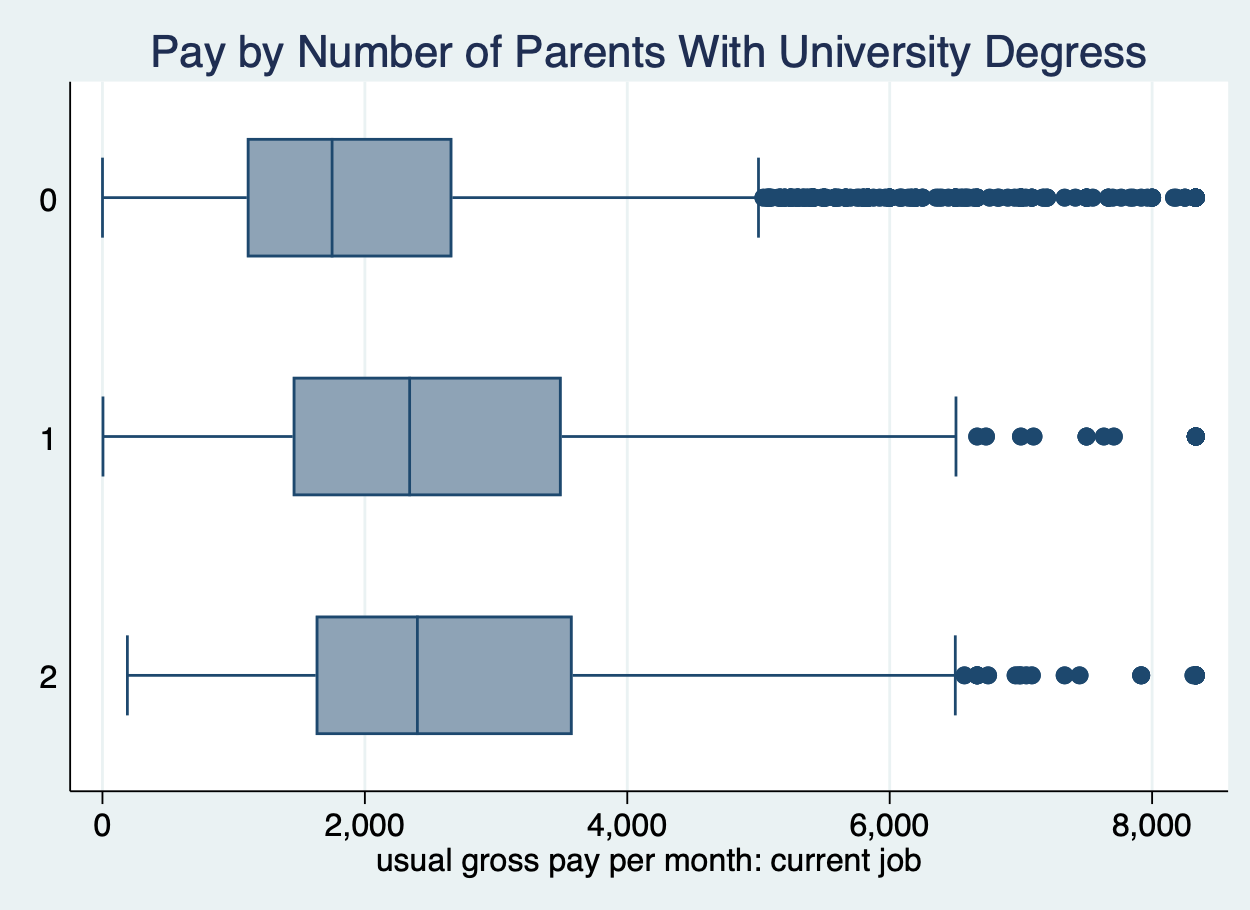


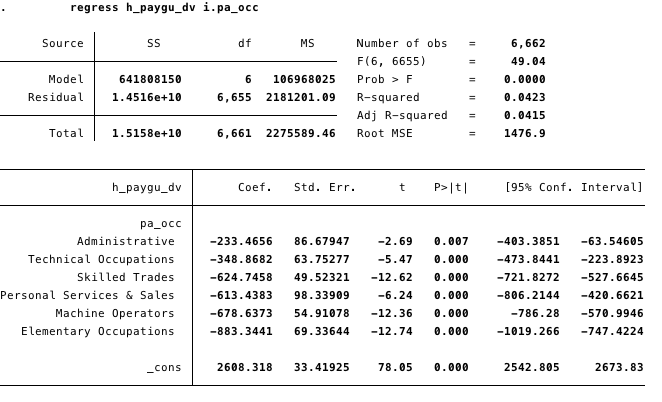


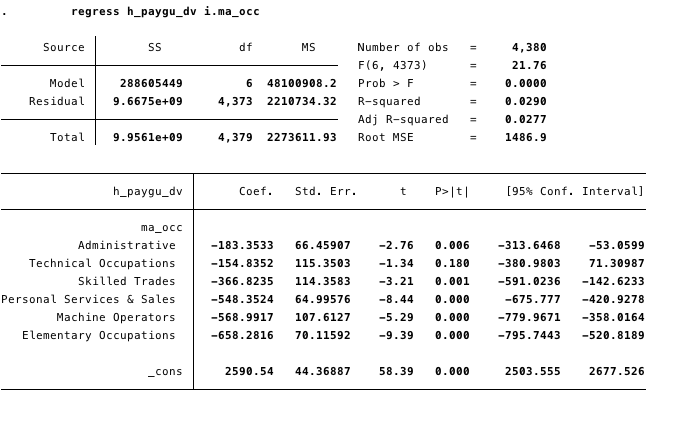


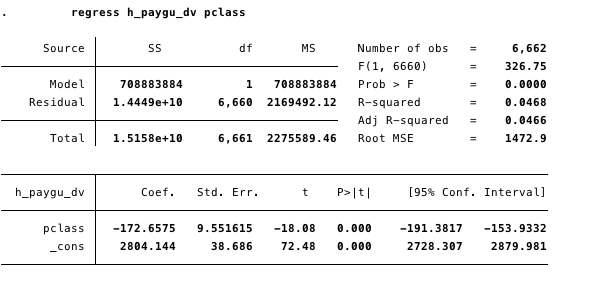


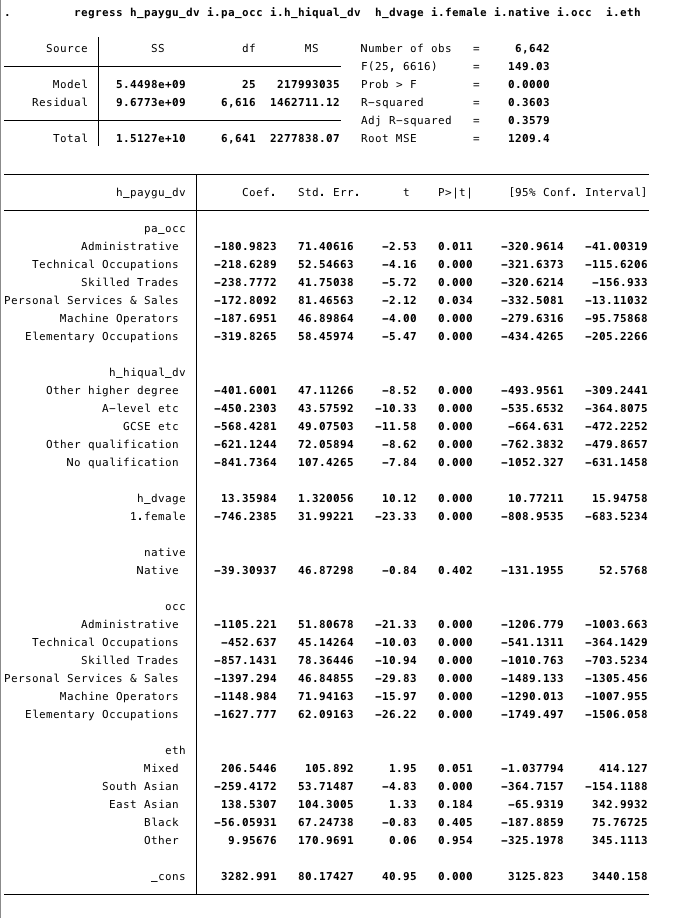


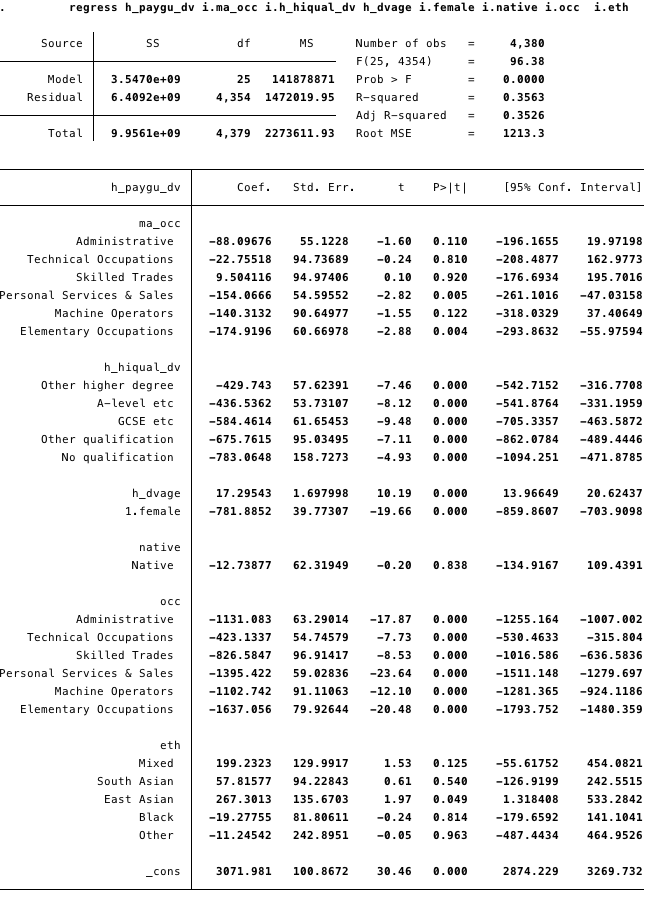


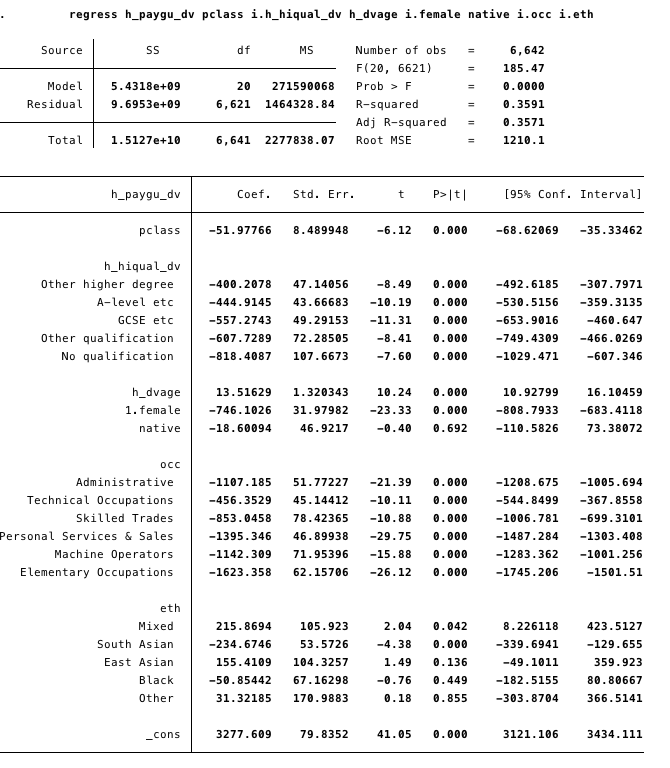


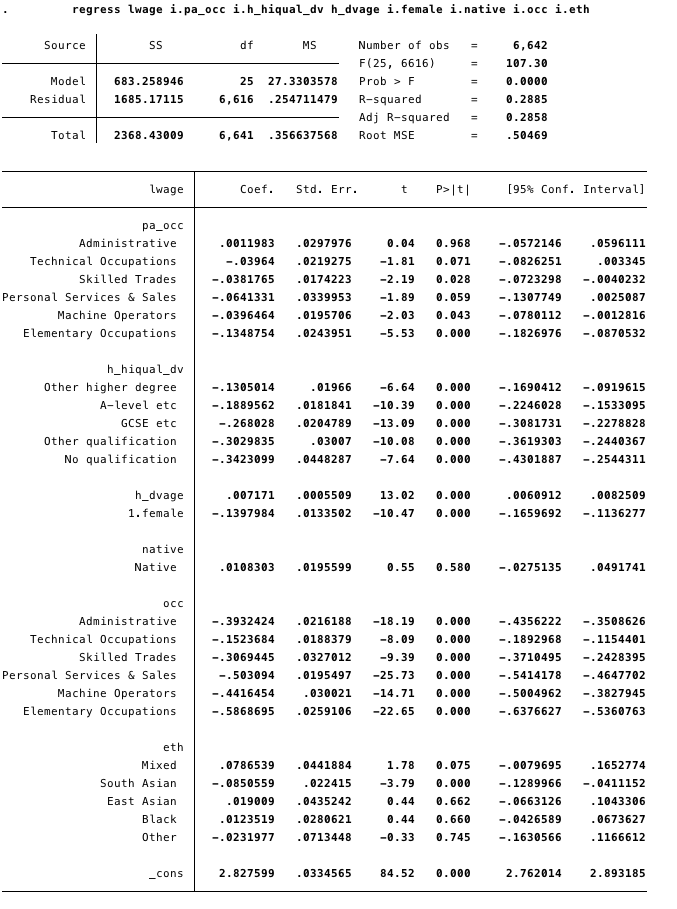


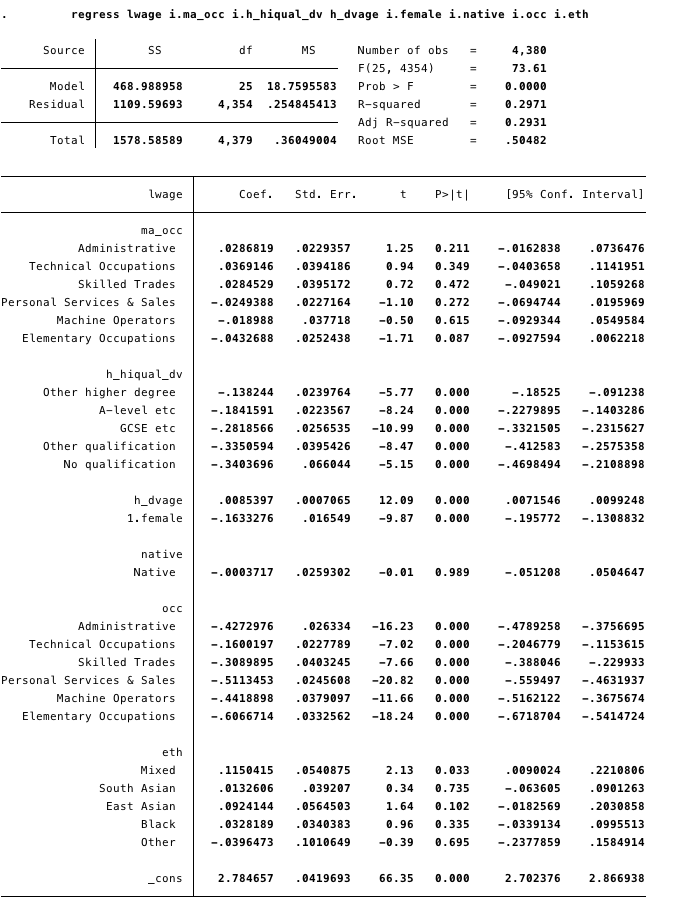


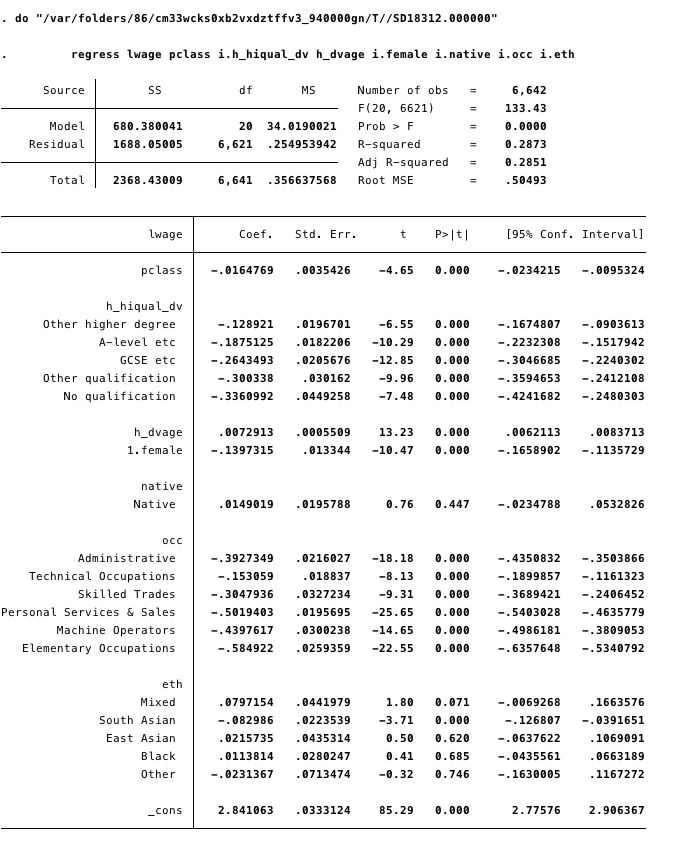


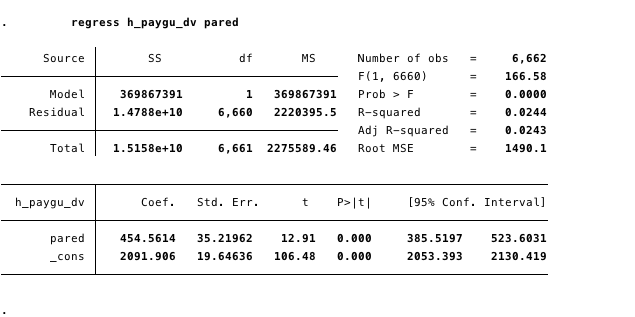


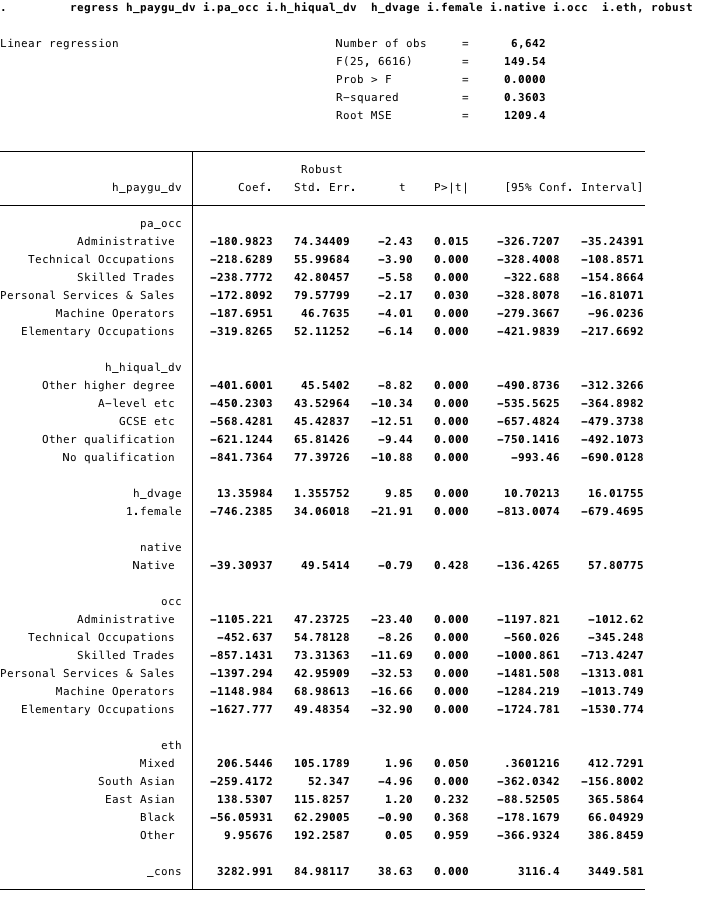
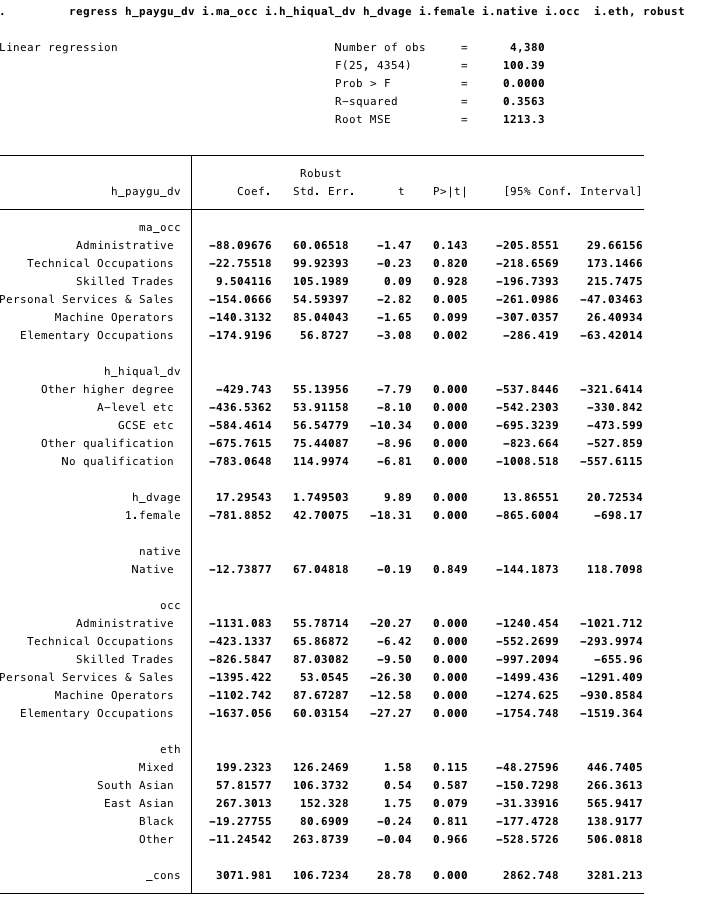
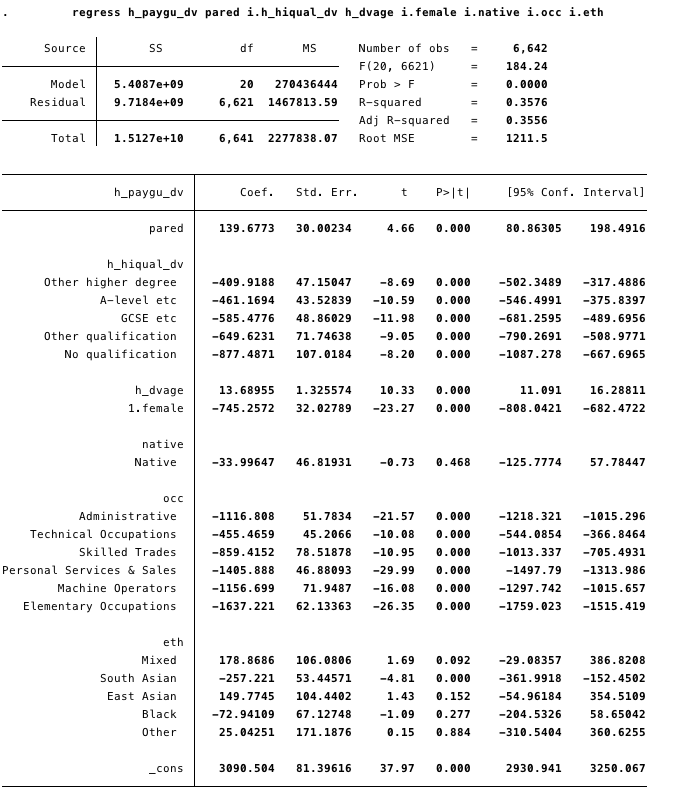


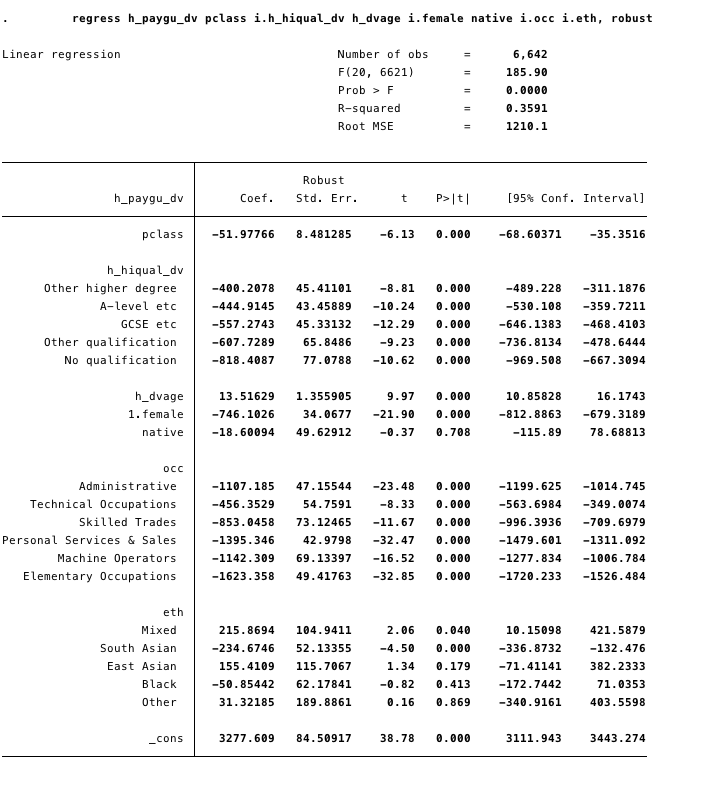


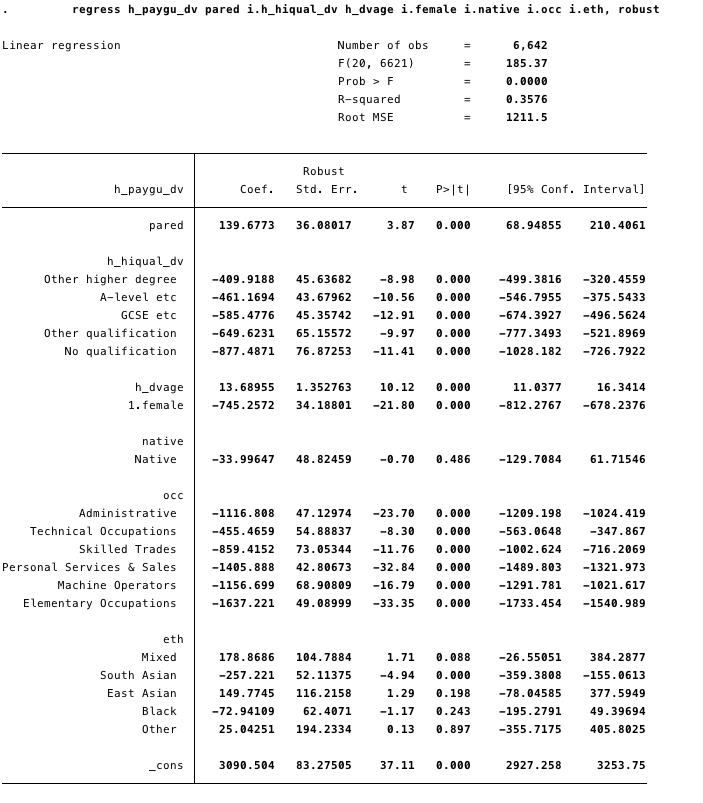












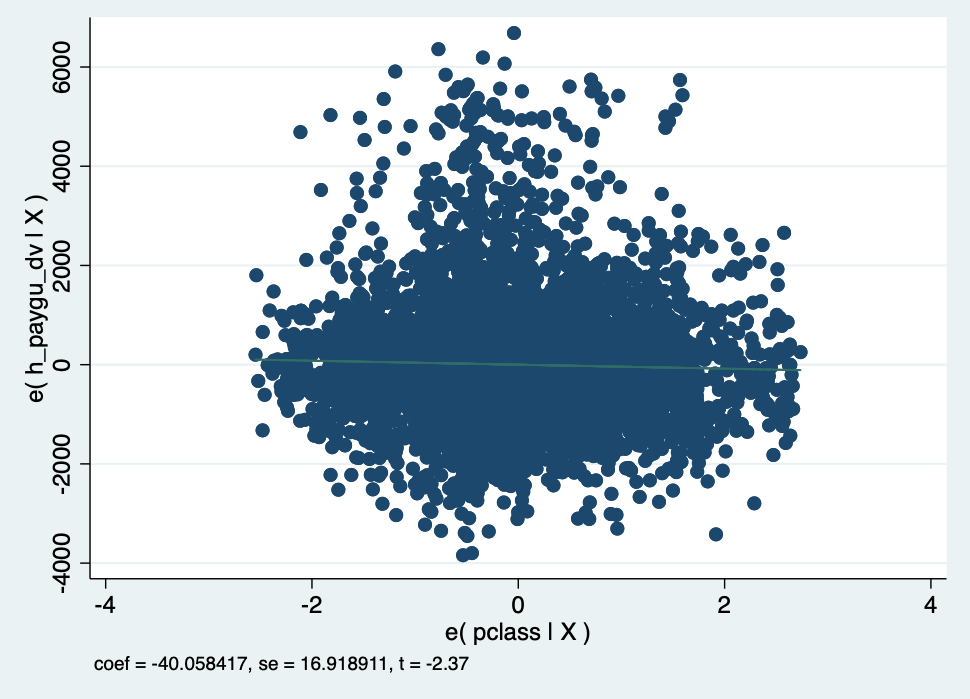
Diagnostics

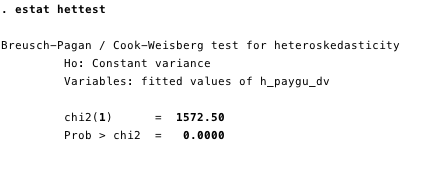
All diagnostics are based on the following regression:

regress h\_paygu\_dv i.pa\_occ i.h\_hiqual\_dv h\_dvage i.female i.native i.occ i.eth

(Diagnostic results broadly similar for other non-robust multivariate specifications, including those with a logged dependent variable)

AVPLOT





RVFplot (residuals v fitted)

