**Applications of Statistics and Econometrics in Finance and Accounting**

**Spring 2023**

**Dr. Kinga Tchórzewska**

**Empirical Group Project Instructions**

**“Wealth and Health: exploring the relationship between earnings, obesity, and other health outcomes and behaviours.”**

There are two parts to this project: the first part is an empirical analysis of the relationship between earnings and weight, and the second one is a policy brief.

For part A, you will analyse data from the British Household Panel Survey (BHPS). BHPS is a panel dataset of individuals from a random cross-section of British Households. The first wave was collected in 1991, and after wave 18 (in 2008) the individuals forming the cross-sectional sample were incorporated in a new panel dataset called Understanding Society (USoc).

You are given three waves of data from the BHPS: wave 14 (2004), wave 16 (2006) and wave 19 (2010-2011), which was collected as part of wave 2 of Understanding Society. This is an unbalanced panel for many reasons: first, some survey questions changed from year to year. For example, height and weight questions are only included in waves 14 and 16, whereas wave 19 (which is part of USoc wave 2) has many more questions about sports activities. Also, some subjects dropped out or were added to the sample.

For more information on the BHPS, see: <https://www.iser.essex.ac.uk/bhps>

For more information on USoc, see: <https://www.understandingsociety.ac.uk/>

For Part B, you will need to write a review of potential government interventions aiming to curb obesity and its costs to the taxpayer. You are given some readings and references, but you are expected to find and use additional sources of information, statistics, etc, to back up your claims.

**Part A: Health and Wealth: exploring the relationship between labour market outcomes and health condition and behaviour (60%)**

Your main goal is to establish the causal link (if there is one) of body weight-for-height on labour market outcomes. The project guides you through a few different approaches that have been used in the literature. You will need to present and interpret the empirical evidence, as well as discuss the methodology and its potential caveats in each case.

Possible labour market outcomes include labour earnings, current employment status (employed/unemployed), socioeconomic position, etc. One common measure of weight for height is the Body Mass Index (BMI), which is the ratio of one’s weight in kilograms to one’s height in metres squared (). One’s BMI defines whether they are classified an underweight, normal weight, overweight or obese[[1]](#footnote-1).

There are two datasets for you to use: bhps\_aofe.dta and bhps\_usco2\_aofe.dta. The former combines wave 14 and wave 16 of BHPS, and the latter appends these waves to the sample of BHPS respondents in the second wave of Understanding Society. Full documentation, codebooks and questionnaires are provided for you on Learn.

You are provided with a shortened version of these datasets, but there are still many variables that pertain to labour market outcomes, demographics, and general health. You don’t have to use all the variables that you are provided with, and you will have to generate some new variables (e.g. you will need the variables on height and weight to create a BMI index). Use your judgment and be selective. For example, you can create indices out of some existing variables, and there might be different indices that one can come up with. Also, some variables have non-missing values for very few observations, so it will be hard to use those.

Generally, questions are intentionally open-ended, and there will very often be more than one ways to answer each question.

Before you start your project, familiarise yourself with the datasets, and especially how the variables might be different from wave to wave. You can do this, for example, by looking at summary statistics and tabulations. It is also important to understand what questions in the surveys the variables correspond to. The best way to check this is to see which part of the survey questionnaire each variable comes from. In some cases, only individuals who provided a certain answer in a question will have data for subsequent questions which ask them to elaborate on their answers. For example, detailed questions about health problems or disabilities are only answered by those who said they were suffering from health issues. This means that some variables will have a lot of missing values. Also, the questions in each wave might differ from questions in other waves, even if they try to elicit the same information from respondents. As you proceed in the project, use your judgment to generate new variables, indices, etc, which capture this information across waves.

1. [20%] Using any wave of the BHPS, explore the cross-sectional relationship between labour market outcomes and weight for height, gradually adding relevant controls. Is this relationship similar across genders?
2. [10%] Explore whether one’s weight for height is correlated with other health characteristics, such as physical health and disability status, or mental health. Are these correlations similar across genders?
3. [10%] Explore whether one’s weight for height is correlated with health-related behaviours such as smoking or being physically active (e.g. by practising sports).
4. [10%] Is weight for height correlated with other demographic variables, such as education, socioeconomic status, or marital status? Are there differences across genders?
5. [10%] Outline the potential endogeneity problem in the relationship between labour market outcomes and weight for height, which makes it hard to establish a causal link just by looking at a cross-sectional relationship.
6. [10%] Use the combined waves 14 and 16 of the BHPS (bhps\_aofe.dta) to generate first-differenced estimates of the relationship between labour market outcomes and weight for height, gradually adding relevant controls. Comment on the advantages and caveats of this method compared to the estimation in part (1).
7. [10%] In waves 14 and 16 of the BHPS, some individuals were matched to their parents. Using this feature of the data, use parental weight for height as an instrument for own weight for height, in order to get 2SLS estimates of the effect of weight for height on labour market outcomes. Check if the effects are constant across genders.
8. [10%] Comment on the reasoning behind choosing the instrumental variables mentioned in part (7). Clearly state the assumptions required for the results of part (7) to be consistent estimates of the causal effect of weight for height on labour market outcomes. Are these conditions satisfied here? What would cast doubt on the approach taken in part (8)?
9. [10%] Summarise and interpret your findings in order to answer the question: does obesity affect labour market outcomes, and by how much?

**Part B: Policy brief (40%)**

You work for the Department of Health, and you are preparing a policy brief for the Health Secretary. The Department is looking into different policies to combat obesity and its consequences. There are many ideas on the table: for example, taxation of fatty or sugary foods, subsidies for gym subscriptions, full NHS coverage of bariatric surgery, anti-obesity drugs, and Weight Watchers membership, as well as changes in the guidelines for school meals. You are required to present and discuss a number of policies, and come up with policy recommendations. It is up to you to decide how many and which policies to focus on. As part of this report, you need to 1) provide some background and lay out the problem and the scope for government intervention, 2) cite and comment on existing evidence of policy evaluations (e.g. from other countries), exposing the economic arguments for and against each policy, and 3) come up with a recommended course of action for the Department of Health.

1. BMI has been criticised as not being an accurate measure of one’s weight for height, because it does not take muscle mass into account. If you want, you can look for other measures and use them instead, or contrast them with BMI. Also, you can use BMI directly or generate categorical variables for whether one is overweight, obese, etc. [↑](#footnote-ref-1)