

# Applied Econometrics Using Stata - Assignment 2

Instructor: Garima Agarwal

February 2024

## Instructions

- Please submit a pdf of your answers along with a stata do file and log file. Your answers should be typed and converted to a pdf.
- Any figures or results from stata can be added to your assignment. Please do not use screenshots for results. You can create tables if relevant.
- Make sure you add comments in the code to make it easy to follow your steps.
- Deadline: 19-Feb 11:59pm through a moodle submission link.
- Please name your files “metrics\_assign2\_firstname\_lastname.pdf”, “metrics\_assign2\_firstname\_lastname.do” and “metrics\_assign2\_firstname\_lastname.smcl”.

This assignment is based on a 2008 paper titled “Predicting the Quality and Prices of Bordeaux Wine” by Orley Ashenfelter. This work created a lot of controversy! See these [old](#) and [current](#) videos for some context.

The author estimates a model to estimate the price of wine based on the growing conditions the crop experiences. Use the dataset wine.csv to answer the questions that follow.

The variables used are as below:

- year: year in which grapes were harvested to make wine.
- price: Average market price for Bordeaux vintages according to 1990–1991 auctions.
- lnprice: natural log of price
- winterrain: winter rainfall (in mm).
- agst: Average Growing Season Temperature (in Celsius degrees).
- harvestrain: harvest rainfall (in mm).
- age: age of the wine measured as the number of years stored in a cask.
- francepop: population of France at Year (in thousands).

1. Suppose we wish to predict the natural log of the price of wine. Which regressors predict this variable best?
  - (a) Explore graphically. [10]
  - (b) Explore correlations. [10]
  - (c) Explore simple regressions. [10]
  - (d) Ashenfelter claims that weather is the key determinant of wine prices. Replicate Figures 1 and 2 from the paper and explain his arguments. [10]
2. Let us consider a model to predict log prices based on the paper.
  - (a) Replicate columns (1) and (2) of Table 2. Run a stata command to create the table output. [10]
  - (b) Comment on the goodness of fit of models (1) and (2). Do results change if we consider log prices rather than log prices relative to 1961 as our dependent variable. [10]
  - (c) Are the regression coefficients in model (2) as per expectations? Interpret the point estimates and comment on their individual and joint statistical significance. [10]
  - (d) Notice that the dataset has another variable that denotes the population of France. Estimate a model with this additional regressor and compare the results to those from (2). Should this variable be included in the model? Conduct a formal test. [10]
  - (e) Suppose we estimate the same specification as model (2) but use prices instead of log prices as the dependent variable. Show the results of both models. Comment on the goodness of fit of this model compared to (2) using the  $R^2$  criterion. [10]
3. Do you have thoughts on the limitations of this econometric exercise? [10]