EPIC Summer RA Application Document

3. Econometrics Background

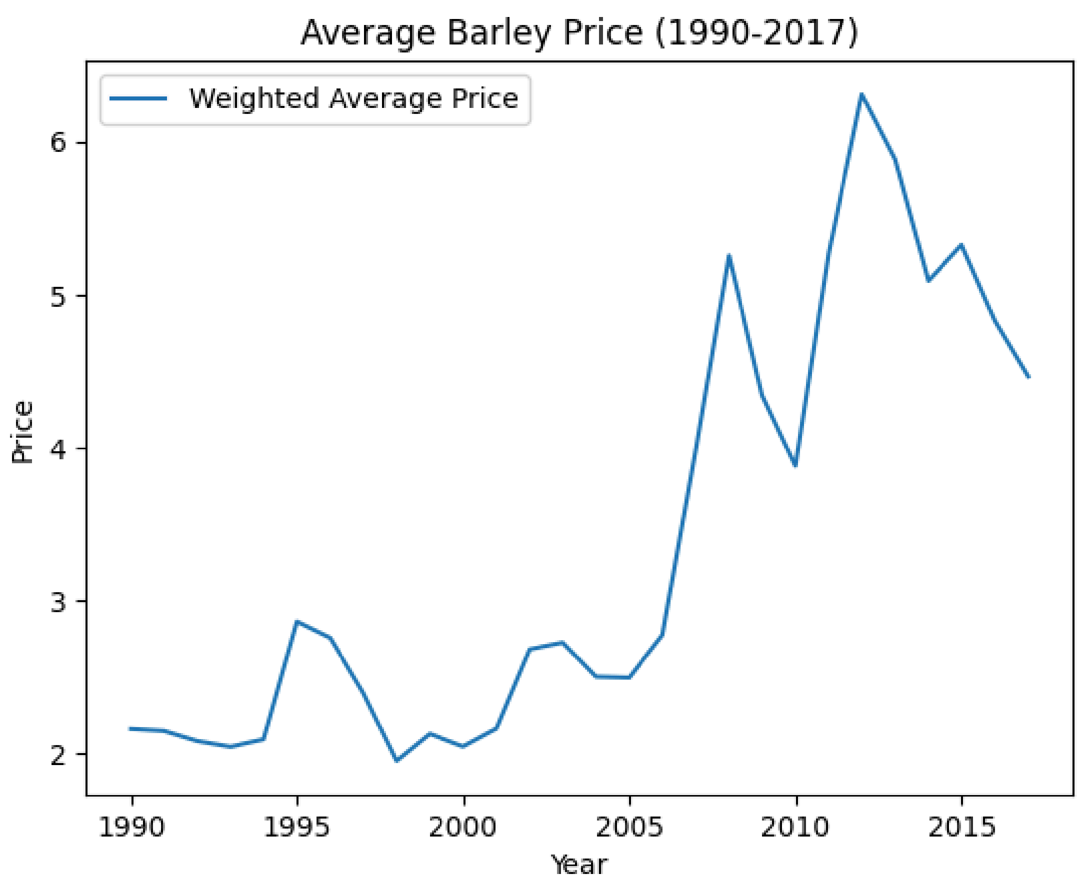
I have a solid foundation in econometrics after taking a course (ECON 140A) at my university covering OLS multivariate regression derivation and applications using STATA as well as some causal inference. I have also taken two probability and statistics theory courses (PSTAT 120A and B in the winter), and plan to take a second econometrics course (ECON 140B) in the spring.

4. Data Exploration

4.2 Dataset Cleaning

The data set provides information on barley prices and quantities produced in various U.S. states for every year from 1990 until 2017. The data has been cleaned, but it originates from data collected by the U.S. Department of Agriculture. Users should be aware that the data does not include all 50 U.S. states and some years may have more states represented than others.

4.3 Time Series Plot



After plotting the data for barley prices over time, there is a clear upwards trend. That is, the prices of barley appear to be increasing across the years. There have also been many fluctuations in the price. Notably, in 2008 there was a big drop correlating with the great recession and another drop around 2012. Even though prices seem to vary and in 2017 prices were declining, the general trend seems to be upward.

4.4 Summary Table

Average State Production for Barley per Decade (in millions)

State	1990-1999	2000-2009	2010-2017
IDAHO	56.985	50.399	52.4838
MINNESOTA	32.95	7.7295	5.1984
MONTANA	59.218	36.875	39.5806
NORTH DAKOTA	120.154	79.5335	42.1862
WYOMING	8.859	6.211	6.814

5. Short Answer

1. Y(p) = c + Bp,

   Y is production output, c is a constant regressor intercept, B (beta) is the coefficient for price, p is price

2. OLS Regression Estimate: Y(p) = 2.137p + 6.6272 + e

   e = error term representing effects from all omitted variables

   a. The result can be interpreted by saying that there is an estimated change of 2.137 million bushels of barley for every one dollar increase in price holding all else constant (e=0). On the other hand, you may use the coefficient to estimate a possible production output given the price.

   b. With this data, we cannot conclude that the price increase is causing the production increase because there may be other omitted variables correlated with price that are affecting barley production. Since R-squared (0.0135) is quite low, there are most likely other important factors to consider

3. Some other variables that we should control for could be annual average temperature, and average annual rainfall. Temperature and rainfall climate changes caused from an increase in greenhouse gas density in the atmosphere over time can heavily influence barley production (Cammarano et. al, 2019). These variables may simultaneously be associated with barley price as they could be impacting markets of close substitutes to barley. For example, if higher temperatures and lower rainfall happened to decrease the supply of another grain that is a substitute to barley, then that grain's price may increase, leading to an increase in demand for barley, thus resulting in a higher price and production of barley.

4.

a. Y(p) = (c + a) + Bp + e,
   Y = production, c = constant regressor intercept, a (alpha entity) = state fixed effect, B (beta) = coefficient, p = price, e = error term representing effects from all omitted variables

   OLS Estimate: Y(p) = 21.509 - 3.1895p + e

b. This value can be interpreted by saying there is an estimated decrease in about 3.19 million bushels of barley for a one dollar increase in price when the time-invariant effects impacting individual state production over the years are averaged out.

c. It is important to account for state-specific characteristics that do not change over time when trying to isolate the effect of price on production in the U.S.. There are various factors that may be impacting production in certain states that do not change over time. For instance, some states may have a climate better suited for growing barley, or one state may specialize in the production of a good that requires barley so there is a higher demand there than in other states. I prefer this specification because the R-squared value is higher than no fixed effects (0.1134 > 0.0135) meaning the regression is a better fit.

5. Upon running the regression using year and state fixed effects, the estimated coefficient on price is -5.5551 and the standard error is 3.9206. The regression ran standard errors clustered by firm and year because by fixing year and state effects we are taking on the assumption that errors in OLS estimator are within each state as well as across states in each year.

   a. Y(p) = (a + t + c) + Bp + e

      Y = production, a (alpha entity) = state fixed effect, t (alpha time) = year fixed effects, c = constant regressor intercept, B (beta) = coefficient on price, p = price, e = error term representing effects from all omitted variables

      OLS Estimator: Y(p) = 28.118 - 5.5551p + e

   b. The coefficient on price can be interpreted as a decrease in 5.555 million bushels of barley for a one dollar increase in price if fixed effects in each state as well as fixed effects across states that are influencing barley production each year are averaged out.

   c. Some time specific characteristics that are the same across states may be national trends that affect all states for a given year. For example if there was a recession or if oil prices were particularly high one year due to a conflict, then it may cost more to produce barley that year and production could be lower. I do not prefer this specification over the state fixed effect because R-squared (0.0586 < 0.1134) is lower meaning the model is a looser fit. Moreover, unlike the previous regression, the coefficient estimate is not significant at the 5% confidence level, meaning we fail to reject the null hypothesis that price has any effect on production.

6. If 10% of observations were dropped purely randomly then there should be little to no effect on the estimated coefficient. This is because the coefficient estimator's formula is Cov(price, production) / Var(price) and both the covariance and variance are functions divided by sample size, therefore, the change would cancel out. However, the standard error would increase with a smaller sample size because the standard error is calculated by dividing the standard deviation by the sample size. If the deletions were not random, this would affect both the coefficient estimator as well as the standard error. Since the coefficient estimator is constructed on averages, disproportionately removing some data will change the average values, thus also changing the coefficient estimator.

7. Economic theory says that as price increases, quantity supplied should also increase. However, when we regress production on price (best regression fit of the three) with state fixed effect, the coefficient returns negative with 95% certainty. This goes against the basic economic theory of an upward sloping supply curve, but it does not mean the coefficient is incorrect. Price may just be negatively associated with production, but most likely is not causing the decrease. Instead, there may be other variables that are causing the reduction in quantity that are correlated with increases in price. Therefore, the model does not uncover the true relationship between price and supply, but it raises interesting questions as to what else may be contributing to production decreases that is stronger than the effect of increasing prices.