Problem Set #2
Due: Wednesday, 22 November 2023

The 2008 global financial crisis and (un)employment

In April 2008, the unemployment rate in the United States stood at 5.0%. By April 2009, it had increased to 9.0%, and it had increased further, to 10.0%, by October 2009. Were some groups of workers more likely to lose their jobs than others during the 2008 global financial crisis? For example, were young workers more likely to lose their jobs than middle-aged workers? What about workers with a college degree versus those without a degree or women versus men?

The data file Employment_08_09 contains a random sample of 5440 workers who were surveyed in April 2008 and reported that they were employed full-time. A detailed description of the data is given in Employment_08_09_Description. These workers were surveyed one year later, in April 2009, and asked about their employment status (employed, unemployed, or out of the labor force). The data set also includes demographic measures for each individual. Use these data to answer the following questions.

(a) What fraction of workers in the sample were employed in April 2009? Use your answer to compute a 95% confidence interval for the probability that a worker was employed in April 2009, conditional on being employed in April 2008.

(b) Regress Employed on Age and Age^2 , using a linear probability model.

      (i)   Based on this regression, was the age a statistically significant determinant of employment in April 2009.

      (ii)   Is there evidence of a nonlinear effect of age on probability of being employed?

      (iii)   Compute the predicted probability of employment for a 20-year-old worker, a 40year-old worker, and a 60-year-old worker.

(c) Repeat (b) using a probit regression.

(d) Repeat (b) using a logit regression.

(e) Are there important differences in your answers to (b)-(d)? Explain.

(f) The data set includes variables measuring the workers' educational attainment, sex, race, marital status, region of the country, and weekly earnings in April 2008.

(i)   By adding those covariates to the linear probability model regression of point (b), investigate whether the conclusions on the effect of Age on employment from (b) are affected by omitted variable bias.

(ii)   Use the regression results to discuss the characteristics of workers who were hurt the most by the 2008 financial crisis.

(g)   **Optional**: Use the models in (b)-(d) to assess the in-sample accuracy of the classification. What is the proportion of correctly assigned classes?

(h)   **Optional**: Repeat point (g) using one or more (at your discretion) of the following classification algorithms: Naïve Bayes Classifier, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Decision trees, Random forests, K-Nearest Neighbours.