

Honors General Exam

Solutions

Harvard University
April 2014

PART 3: ECONOMETRICS

Immigration and Wages

Do immigrants to the United States earn less than workers born in the United States? If so, what are the reasons behind these differences? In this section of the exam you will discuss empirical strategies that could be used to investigate these questions.

Question 1: The Determinants of Wages

To investigate the extent to which immigrants earn less than US-born workers, a researcher takes a random sample of people living in the United States in 2014 and estimates a regression of the form:

$$\ln(\text{earnings}_i) = \beta_0 + \beta_1 * \text{immigrant}_i + u_i \quad (1)$$

where earnings_i are the earnings of worker i in 2013 measured in thousands of dollars and immigrant_i is an indicator variable equal to one if the individual was born outside the United States.

- (a) An observer notes that the coefficient β_1 may be biased because immigrants may have less education on average than US-born workers. Provide the sign of the bias associated with the omission of education in equation (1) assuming that the observer's assumption is correct. In predicting the sign of the bias, explain the role of (co)variances of the relevant variables.

Solution: We know that the estimated coefficient will be biased in the population by $\beta_2 \frac{\text{Cov}(\text{immigrant}, \text{education})}{\text{Var}(\text{immigrant})}$. We expect β_2 (the coefficient on education) to be positive and $\text{Cov}(\text{immigrant}, \text{education})$ is assumed to be negative. Thus, we'd expect the estimate of β_1 in equation (1) to suffer from a downward (or negative) bias.

- (b) Another observer worries that the relationship between earnings and immigrant is non-linear and suggests including immigrant_i^2 in equation (1). Do you think this is a sensible strategy? Be specific.

Solution: Since immigrant is an indicator variable, this would not be a sensible strategy. In other words, since the square of 1 is 1 and the square of 0 is 0 we gain nothing from including immigrant squared in equation

- (1). In fact, this squared variable would be perfectly collinear with the immigrant variable.
- (c) One researcher is discouraged by the potential for omitted variables since she is interested in the causal effect of being an immigrant on earnings. Thus, this researcher suggests using distance from the Mexico-US border as an instrument for $immigrant_i$ in equation (1). Assuming homogeneity across individuals, what are the two conditions that must hold for this distance to be a valid instrument? Do you think that they will hold in this context? Be specific.

Solution: The instrument has to be relevant (i.e. correlated with immigrant) and exogenous (i.e. not correlated with the error term). It seems likely that the proposed instrument would be relevant i.e. that the probability of being an immigrant increases as one gets closer to the US-Mexico border. The exogeneity of this instrument, however, is more suspect since many other variables (such as the educational level of workers, the industries that cluster in each region) may change as we get further away from the border. Thus, this instrument is probably not going to solve the endogeneity problem.

- (d) This researcher also considers gathering data on both sides of the Mexico-US border and running a specification of the form:

$$\ln(earnings_i) = \beta_0 + \beta_1 * distanceborder_i + \beta_2 * US_i + u_i \quad (2)$$

where the relevant sample would be drawn from both the United States and Mexico, $earnings_i$ measures the 2013 earnings of worker i in thousands of US dollars, $distanceborder_i$ is the distance of worker i 's residence to the Mexico-US border (where these distances are defined as negative for locations in Mexico, i.e. the distances are multiplied by -1, and positive for those in the US) and US_i is an indicator equal to one if the individual's residence is in the United States. Do you think that an estimate of β_2 in equation (2) would help uncover the causal effect of being a US resident on earnings? Provide one threat to the internal validity of such a strategy. Be specific.

Solution: In equation (2), β_2 gives the regression discontinuity design (RDD) estimate of the effect of being a resident of the United States. In other words, this strategy compares individuals living on each side of the

border (e.g. Tijuana and San Diego). Recall that for the RDD strategy to work, it must be the case that as we cross the border the only thing that changes systematically is that people become US residents. For many parts of the Mexico-US border this might be a plausible assumption since most residents on both sides of the border may be roughly similar (i.e. have similar cultural backgrounds). However, one threat to the internal validity of such a strategy is that those who actually came across the border are more driven than those that did not and thus as we cross the border we are not only changing being residents in the US, but we are also changing the composition of the population (i.e. more driven people on the US side of the border).

Question 2: The Effects of Immigration on US Employment

Does immigration increase unemployment among the existing residents of the host country? A variety of studies have investigated this question using population shocks. Imagine for the sake of this part of the exam that 55 cities on the east coast of the United States accepted 1.3 million refugees from the ongoing Syrian Civil War in 2014. Thus, these cities would have a surge in their labor forces in 2014. For the purposes of this part of the exam, let i denote city and t years. Furthermore, $refugee_i$ is an indicator variable equal to one if city i received at least one Syrian refugee in 2014.

- (a) The researcher begins by estimating the following regression using cross-sectional data on 1324 US cities in 2015 and obtains:

$$unemployment_i = \widehat{0.079} - 0.03 * refugee_i \quad (3)$$

(0.001) (0.001)

where $unemployment_i$ is the unemployment rate in city i in 2015 and $refugee_i$ is as explained above. Interpret both the intercept and the coefficient on $refugee_i$ in equation (3). Do these results suggest that this population shock affected unemployment?

Solution: Cities that did not receive Syrian immigrants had an unemployment rate of 7.9 percent in 2015. This was lower by .3 percentage points in cities that received at least one Syrian refugee. This regression does not provide support for the claim that this population shocks affected unemployment since the coefficient on refugee is small in absolute value and not statistically significant. Of course, one should always be worried about omitted variables in such a context.

- (b) An observer is frustrated by the zeros in front of the coefficients in equation (3). Suggest how you would modify regression (3) so that each of the estimated coefficients is multiplied by one hundred.

Solution: Multiplying the unemployment variable (the dependent variable) by one hundred and re-estimating the regression will give the desired results.

- (c) The researcher decides to ignore the observer's frustrations, gathers data for all 1324 cities for 2013 and estimates a regression of the form (using data from 2015 and 2013):

$$\widehat{unemployment}_{it} = 0.080 - 0.019 * refugee_i + 2 \cdot 10^{-4} * year2015_t + 0.015 interaction_{it} \quad (4)$$

(0.001) (0.003) (8 · 10⁻⁴) (0.004)

where the variables are as explained above and

$$interaction_{it} = refugee_i * year2015_t$$

Interpret all of the coefficients in equation (4). Do these results provide support for the claim that the Syrian population shock increased unemployment?

Solution: The intercept's coefficient indicates that in 2013, in areas that were not going to receive refugees, the mean unemployment rate was 8 percent. The coefficient on *refugee* suggests the unemployment rate was 1.9 percentage points lower in areas that were to receive refugees. The coefficient on *year2015* indicates that between 2013 and 2015, the unemployment rate fell by 0.02 percentage points in areas that did not receive refugees. Finally, the coefficient on the interaction term is the dif-in-dif coefficients and shows that areas that received refugees saw their unemployment rate increase by 1.5 percentage points between 2013 and 2015 when compared to those evolution of the non-treated areas over the same period. This result is consistent with the claim that the refugee shock increased unemployment.

- (d) Explain how you would use the coefficients in equation (4), to provide a point estimate of:

$$E[unemployment|refugee = 1, year = 2015] - E[unemployment|refugee = 1, year = 2013]$$

Be specific.

Solution:

- $2 * 10^{-4} = E[\text{unemployment}|\text{refugee} = 0, \text{year} = \widehat{2015}] - E[\text{unemployment}|\text{refugee} = 0, \text{year} = 2013]$
- $0.015 = E[\text{unemployment}|\text{refugee} = 1, \text{year} = \widehat{2015}] - E[\text{unemployment}|\text{refugee} = 1, \text{year} = 2013]$
 $-E[\text{unemployment}|\text{refugee} = 0, \text{year} = \widehat{2015}] + E[\text{unemployment}|\text{refugee} = 0, \text{year} = 2013]$

So,

$$E[\text{unemployment}|\text{refugee} = 1, \text{year} = \widehat{2015}] - E[\text{unemployment}|\text{refugee} = 1, \text{year} = 2013]$$
$$= 0.015 + 2 * 10^{-4}$$