

Discussion Referee Report

Fabian Waldinger

Referee Reports

- Average grade: 57
- Room for improvement

Paper 1: Migrants, Ancestors, and Investments

- Question: What is the effect of ancestors in a US county on FDI between that county and the home country of the ancestors
- Main specification:
$$FDI_{od} = OriginFE_o + DestinationFE_d + \beta Ancestry_{od} + Error$$
- An observation in the data set is a US county x home country pair
- One cross section of data in 2010

Paper 1: Methodology

- What is the problem with the regression equation above?
- Ancestry is endogenous (e.g. because people may move to counties with a lot of FDI)
- Solution: IV
- Use IV based on:
 - push factors: the number of people leaving a certain country in a given time period (without those that go to the particular county)
 - pull factors: the number of people from other continents settling in a particular county at a certain point in time

What to criticize in IV papers?

- In any IV paper you have to evaluate the following two conditions:
 - First stage
 - Exclusion restriction
- First stage
 - is not problematic
 - F-stats are always above 10
 - The specification with the higher order interactions of the IVs is not transparent and it is not clear what identifies the effect in those models (even if they seem to have improve the first stage a lot)

- Exclusion Restriction

- suppose people from Mexico and from Morocco migrate to 2 particular counties because they mostly work in agriculture. And there is a lot of agriculture FDI in those counties with Mexico and Morocco because that county has many agricultural firms. If migrants from Mexico and Morocco only go to these 2 counties the IV may be endogenous.
- How could be improve the IV? Use total outflows from an origin country NOT to the US as a push factor (a lot of data work but would make the IV more credible)

- OLS is most likely upward biased. Why are IV coefficients larger than OLS?
- How does ancestry affect FDI? One nice report suggested to compare listed versus other firms (on other firms insider knowledge is more important).

Paper 2: The Birth of Edge Cities in China: Measuring The Spillover Effects of Industrial Parks

- Question: How does the opening of an industrial park affect productivity in surrounding firms

- Main specifications:

$$TFP_{it} = \text{Distance to Park}_{it} + \text{Industry} \times \text{Year FE} + \text{Plant FE} + \text{Error}$$

(1)

$$TFP_{it} = \text{Distance to Park}_i + \text{Industry} \times \text{Year FE} + \text{Error}$$

(2)

- Equation (1) is essentially a DiD estimate because distance to park is reduced after the opening of a park
- Problem with both (1) and (2): Distance to Park is endogenous. So in (2) they instrument for distance to park
- Why don't they instrument for distance to park in the first specification?

IV Assumptions

- First stage:
 - they do not have a strong first stage once they control for coagglomeration
 - they do not show real first stage but probit and logit models
- Exclusion restriction:
 - IVs are problematic if firms on flat land develop faster after the opening up of China
 - they cannot control for firm fixed effects in the IV specifications because the IV is not time varying

Points that have been raised in many reports that are not problematic

- Timing of IVs relative to main data
- Dianping.com only lists some retailers: should introduce measurement error which should bias the coefficient towards 0, if they still find an effect there is no concern
- Changes in house prices (or other variables) in certain years (e.g. crisis) are not problematic because the authors include year fixed effects
- Low R-squared

Things to remember for next report

- Focus report on the main assumptions of the identification strategy
- If you have a concern think whether certain fixed effects deal with it
- Do not discuss measurement error if the authors still find an effects