

Analysis Plan of Gender Peer Effects

Xiaoyue Shan

1 August 2019

This document describes the main analysis strategies for gender peer effects.

1. Clean data and prepare variables
 - Data sources: baseline survey, endline survey, administrative data (registration information, scores and grades).
 - Select students in study groups for the analysis of causal peer effects.
2. Descriptive analysis
 - Show descriptive statistics of students signed up for study groups, as well as the whole sample (including those not signed up).
 - Summarize baseline and endline educational beliefs, as well as final scores, by students' own gender and the gender composition of study groups.
3. The main regression model to estimate gender peer effects

$$Y_{ig} = \alpha + \beta \text{FemalePeers}_{ig} + \gamma X_{ig} + \varepsilon_{ig}$$

- In the function, Y_{ig} and X_{ig} are the outcome and baseline characteristics of individual i in group g .
 - The outcome variable of interest includes education beliefs in endline (controlling for baseline beliefs, or the first difference of beliefs), final scores, potentially also self-efficacy in endline, evaluation of peers, major choice and dropout.
 - The baseline characteristics in X_{ig} include, for instance, gender, age, major, nationality, high school background, personalities and socioeconomic preferences.
 - FemalePeers_{ig} is the number of female peers that student i has in group g . Therefore, β measures the impact of one extra female peer.
 - Alternatively, we can replace FemalePeers_{ig} with $\text{Female}_{ig}^{LOM} = \frac{1}{3} \sum \text{Female}_{-i,g}$, which is the leave-own-out mean of peers' gender. For instance, if all peers of i are female, then $\text{Female}_{ig}^{LOM} = 1$; if only one peer of i is female, then $\text{Female}_{ig}^{LOM} = 1/3$. In this case, β measures the impact of pure-female peers versus pure-male peers.
4. Except for the impact of peer's gender, I am also interested in the impact of peer's gender attitudes and gender-related characteristics. To estimate that, simply replace Female_{ig}^{LOM} with W_{ig}^{LOM} , where W_{ig}^{LOM} measures the leave-own-out mean of peers' gender attitudes or gender-related traits (e.g., patience, agreeableness, competitiveness).
 5. It is of interest to also run a long regression with both Female_{ig}^{LOM} and W_{ig}^{LOM}
$$Y_{ig} = \alpha + \beta \text{Female}_{ig}^{LOM} + \delta W_{ig}^{LOM} + \gamma X_{ig} + \varepsilon_{ig}$$
 - In this way, we can isolate the effect of peers' gender (controlling for other gender-related traits), or the impact of peers' gender-related traits (controlling for gender itself).