The Returns to College Education: A Reassessment of Heterogeneous Treatment Effects

Stefanie Lightner
University of Minnesota

Eric Grodsky
University of Wisconsin-Madison

Andrew Halpern-Manners
Indiana University

Background

Comparative advantage posits that students make educational choices that maximize the present discounted value of their lifetime earnings subject to constraints imposed by their resources. Students are assumed to behave rationally based on both observed and unobserved endowments; those who go to college when we expect otherwise have an unobserved advantage that leads to success in both college and the labor market.

Brand and Xie (2010) present an alternative theory, ***negative selection***, which suggests that disadvantaged and less academically prepared students—those *least* likely to complete college—experience a greater college return than those with a high propensity for college.

In their analysis, Brand and Xie estimate the propensity to complete college and stratify on this predicted probability. Assuming independence between unobserved attributes and college completion for each stratum, they estimate the returns to completing college by stratum and plot the linear difference in the college premium across propensity bands.

We take another look at the data to address the following:

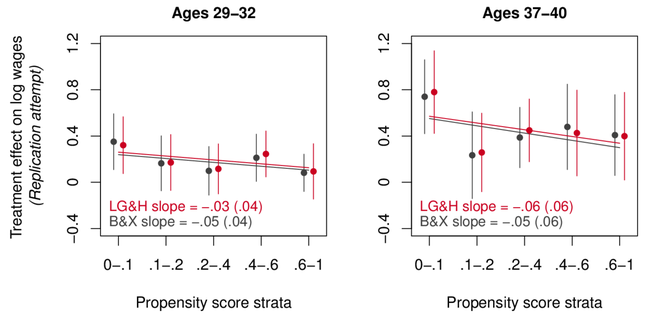
- **Potentially important omitted variables in the college completion model**
- **Robustness to alternative construction of key measures**
- ***Attendance* as the proper treatment for reasons of behavioral theory and policy**

We use data from the National Longitudinal Survey of Youth, 1979.

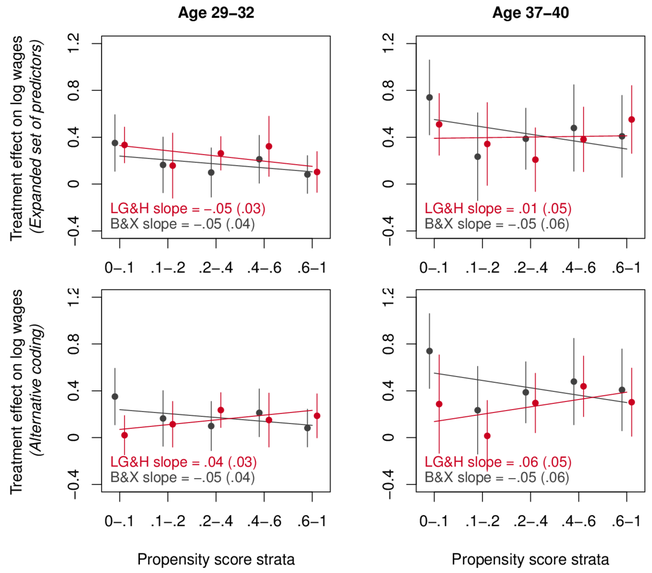
Analysis plan

1. To develop a richer model of college completion, we added the following variables to Brand and Xie's original specification:
   - **Self-esteem**
   - **Locus of control**
   - **Educational aspirations and expectations**
2. We also examined models that included alternate versions of key predictors:
   - **Parents years of schooling (0-20) → *Categorical measure***
   - **Family income (measured in 1979) → *Mean family income (1979-1983)***
   - **College completion (16+ years of education by 1990) → *BA+ by 1990***
3. In the third part of our analysis, we consider a different (and possibly more relevant) treatment: **college attendance**.
4. Finally, we conducted a Monte Carlo simulation study to show the role of unobserved characteristics on estimation bias.
   - The only parameter that we varied across simulations was the correlation (r) between the error in the college completion equation (e_c) and the error in the earnings equation (e_e). Correlations were allowed to run from -.4 to .4.
   - When generating the earnings data, we assumed a 15% return to a college degree ***and homogeneity across propensity score strata***.

Replication*

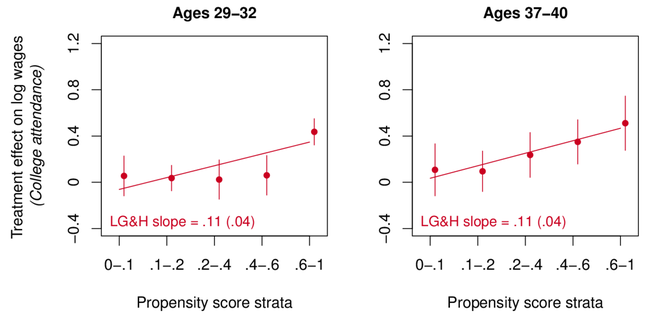


Influence of model specification



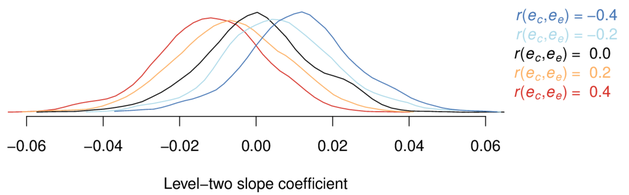
*B&X's slopes are from their 2010 paper and are static for comparison. All results are for men only.

Returns to college *attendance*



When we consider college attendance, a more appropriate treatment for behavior modeling and policy considerations, we see the opposite trend. ***The students who benefit the most from attending college tend to be those who are most likely to enroll.***

Monte Carlo simulation study



Heterogeneous treatment effects are observed when the errors for the college completion and earnings equations are correlated. A negative correlation yields a positive trend across strata (as indicated by the blue lines) and a positive correlation yields a negative trend (as indicated by the red and orange lines). ***The estimates are only unbiased (i.e., properly centered over 0) when the error terms are unrelated to one another***. The role of unobserved characteristics in this simulation is consistent with comparative advantage.

Conclusions

Estimated returns to college completion are sensitive to propensity specification making them ill-suited to the task of estimating heterogeneous treatment effects. Results depend heavily on how the model performs at the tails of the propensity distribution, where these models are the weakest, and on uncorrelated errors in the college completion and earnings equations.