The Lifetime Financial Returns to a College Degree (and Many Other Uses for IPEDS Data)

Doug Webber
Temple University
This workforce solution was funded by a grant awarded to the U.S. Department of Labor's Employment and Training Administration. The solution was created by the Center for Human Resource Research on behalf of the Ohio Department of Job and Family Services and does not necessarily reflect the official position of the U.S. Department of Labor. The Department of Labor makes no guarantees, warranties, or assurances of any kind, express or implied, with respect to such information, including any information on linked sites and including, but not limited to, accuracy of the information or its completeness, timeliness, usefulness, adequacy, continued availability, or ownership. This solution is copyrighted by the institution that created it. Internal use, by an organization and/or personal use by an individual for non-commercial purposes, is permissible. All other uses require the prior authorization of the copyright owner.
Before we get started

• Thank you for having me!
• I am legally blind
  – So feel free to shout out questions in case I can’t see your hand raised
My Documented Love of IPEDS

- Papers of mine which use IPEDS data which I won’t be discussing today
- Impact of spending on student outcomes
- State Divestment and tuition
- Federal accountability reform
Overview of Today’s Talk

• Rather than focus on a single paper of mine, I am going to show a variety of results from several different papers/projects
• More of a focus on interesting (to me at least) patterns and relationships
• Less of a focus on technical minutiae (happy to go into any details you are interest in)
Motivating Questions

• How much debt should a student take on to get a degree?
• What is the likelihood that a college degree is a winning financial proposition?
• How do various factors (major, heterogeneous returns, graduation uncertainty) impact the risk associated with getting a college degree
• Do low-achieving students benefit from college?
  – Assuming that there are heterogeneous returns to college across students, the investment in college for these students could be:
    • A key to upward economic mobility, validating the massive investment in improving college access at the federal and state levels
    • A waste of both time and money if indeed these students are not “college material”
Student Debt

• Growing student debt problem
  – $30,000 average debt after leaving school
• More than $1.3 trillion in total student debt
• Forbes: One third of recent college students regret going to college
Calculating ROI is Complicated

- Usual endogeneity concerns
- Implicit and explicit costs
- Time discounting
- Graduation uncertainty
- Heterogeneous returns
  - Major, demographics, luck, etc.
  - Knowing the average return is useful, but hardly tells the whole story

The Basic Idea

• Simulate earnings for each working year, 18-65 for every educational outcome
  – Estimate raw earnings premia for each age-education group using large, nationally representative datasets
  – Estimate degree of selection bias for each age-education group using detailed panel data
  – Subtract off selection bias from raw premia to get selection-corrected estimates

• I effectively generate a simulated population which matches several hundred conditional population moments (mean and variance)
Data

- National Longitudinal Survey of Youth (NLSY): 1979 and 1997 cohorts
- American Community Survey (ACS): 2014 wave
- Education categories: High school graduate, some college, STEM, Business, Social Sciences, Arts/Humanities, Other
- College Scorecard/IPEDS
- Why use so many datasets:
  - No single perfect dataset for this question
Addressing Selection

- Cognitive ability
  - Armed Forces Qualification Test (AFQT)
- Noncognitive ability
  - Rotter score (locus of control)
  - Rosenberg scale (self esteem)
  - Big Five personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism)
Median Earnings Paths

Yearly Earnings

Age

HS College
Median Earnings Paths
Lifetime Earnings Distributions

The graph shows the lifetime earnings distributions for two different groups, represented by the lines 'kdensity total_64' and 'kdensity total_64'. The x-axis represents earnings in dollars, ranging from 0 to 10,000,000, while the y-axis represents the density of earnings, ranging from 0 to 1.5e-06.
Risk/Return and College Costs

- Likelihood of “success” (two measures) and annual costs
Applying Corrections

• Up to this point all figures have just shown raw earnings (~ $900,000 premium to college degree)

• I will now show the results after applying various permutations of the following corrections
  – Selection
  – Discounting
  – Graduation Uncertainty
<table>
<thead>
<tr>
<th></th>
<th>STEM</th>
<th>Business</th>
<th>Social Sciences</th>
<th>Arts/Hum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>532,190</td>
<td>571,480</td>
<td>391,023</td>
<td>204,166</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>361,973</td>
<td>394,506</td>
<td>315,003</td>
<td>194,190</td>
</tr>
<tr>
<td>Uncertainty + Selection</td>
<td>276,127</td>
<td>296,237</td>
<td>217,069</td>
<td>84,658</td>
</tr>
</tbody>
</table>
Graduation Uncertainty

• Comparing the results with and without correcting for graduation uncertainty illustrates how important obtaining the college diploma is
  – This is not a new finding, rates of student loan default are much higher for non-completers

• Need both to get an accurate picture of college ROI
Putting It All Together
More Cool Data

• Although I can’t credibly address selection at individual major level, it is still interesting to look at the full CDF’s
  – [Link](#) on my website
Heterogeneous Returns

Distribution of Lifetime Earnings by Major

Projected Lifetime Earnings

Measure (Weights):
- Finance
- Fine Arts
- Foreign Languages
- Forestry
- General Business
- Geography
- Graphic Design
- Health Services (G.
- High School
- History
- Hospitality
- Human Resources
- Human/Community...
- Humanities:
- Industrial Engineer...
- Information Sci...
- Interdisciplinary S...
- International Bus...
- Journalism
- Liberal Arts
- Linguistics/Compu...
ROI By College

![Graph showing ROI by college type]

- **Density**
- **x**

Legend:
- Blue: Public
- Red: Private Non-Profit
- Green: Private For-Profit

- TEMPLE UNIVERSITY
  College of Liberal Arts
College Return/Marginal Student

• Ost, Ben, Weixiang Pan, and Douglas Webber. "The returns to college persistence for marginal students: regression discontinuity evidence from university dismissal policies." Journal of Labor Economics (Forthcoming)
Is everyone “College Material”?

“Kid, I’m Sorry, but You’re Just Not College Material”

Is exactly what we should be telling a lot of high school students.

Americans who say “college isn’t for everyone” never mean their own kids.

Why College Isn’t (And Shouldn’t Have to Be) For Everyone

03/22/2015 08:53 pm ET | Updated May 22, 2015
Failure To Get Into Private College To Be Most Financially Responsible Act Of 17-Year-Old’s Life
Identification strategy

• Students on probation must meet GPA standards in following terms, otherwise they are dismissed.
• We use RD design based on this policy.
• So we are effectively comparing the future labor market outcomes for students just above and just below the dismissal threshold.
Identification strategy intuition

Student 1
- GPA<2.0
- Academic Probation
- GPA=1.99
- Academic Dismissal
- Earnings

Student 2
- GPA<2.0
- Academic Probation
- GPA=2.01
- No Academic Dismissal
- Earnings
Main contributions

• Novel identification strategy for estimating returns to college.
• Estimate returns to college for particularly interesting group.
  – The RD estimates are a LATE for those just around the GPA standard.
• Big returns for these students, implicit and explicit costs of college are recouped by (on average) age ~30
  – Caveat: Applies only to students in Ohio public schools; returns on the intensive margin (continuing college once already started).
Data

• Administrative data on all 13 four-year colleges in Ohio. (2000-2011)
  – Courses attempted
  – Term and cumulative GPA
  – Basic demographics
  – Also include data on 2-year college students.
  – We focus analysis on those starting at 4-year schools.

• Administrative UI data on weekly earnings. (2003-2012)
  – Quarterly earnings by employer
  – Weeks worked per employer
Data advantages

• Administrative data on both earnings and college enrollment.

• Earnings data covers vast majority (90+%) of labor force in Ohio.
  – Excludes self-employed, farmers and federal employees.

• Higher education data is universe of public 2- and 4-year college students in Ohio.
Data disadvantages

• No measure of hours worked.
  – Unfortunately common
• Cannot observe earnings for those that leave state.
  – We provide suggestive evidence that this is not a concern for our results, as well as bounding exercises
Missing earnings

• If someone isn’t observed in data
  – 68% chance they left state, work for federal government or are self employed.
  – 32% chance they are out of LF or unemployed.

• We focus on earnings, conditional on employment in Ohio and test for differential attrition.
Data disadvantages

- Cannot observe college enrollment for those that leave state or go to private institution.
  - Our data cover approximately 75% of 4-year schools and nearly all non-selective 4-year schools in Ohio.
- Causes us to potentially overstate enrollment effects and understate impact on earnings.
  - We may incorrectly classify a dismissed student as not ever getting a college degree, but then see them making a lot of money in future employment.
• Called each school to collect
  – Information on academic probation and dismissal policies.
  – Our policies are likely measured with some error.
• We identify University wide policies, but many schools have different standards for different areas of study.
• Where possible, we coded policy changes over time but we did not find many.
Probation and Dismissal Policies

• Dismissed students can apply for readmission.

• Typically requires 1-year waiting period.

• Some schools have 1-year suspension followed by screening process.
Dismissal data

- Our data do not include actual probation or dismissal status.
- We infer these based on school policies combined with observed GPA.
- Many students successfully appeal dismissal decisions, especially in the first year of college.
# Descriptive statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Full Student Sample</th>
<th>At-Risk Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ever on probation</td>
<td>0.24</td>
<td>1.00</td>
</tr>
<tr>
<td>Ever at risk of dismissal</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>Ever dismissed</td>
<td>0.06</td>
<td>0.53</td>
</tr>
<tr>
<td>Female</td>
<td>0.55</td>
<td>0.45</td>
</tr>
<tr>
<td>Black</td>
<td>0.08</td>
<td>0.17</td>
</tr>
<tr>
<td>Age at college entry</td>
<td>19.14</td>
<td>19.02</td>
</tr>
<tr>
<td>BA from starting institution</td>
<td>0.57</td>
<td>0.27</td>
</tr>
<tr>
<td>BA from any Ohio public institution</td>
<td>0.60</td>
<td>0.29</td>
</tr>
<tr>
<td>AA from any Ohio public institution</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>Total credits earned from starting institution</td>
<td>105.08</td>
<td>76.12</td>
</tr>
<tr>
<td>Total credits earned from any institution</td>
<td>109.32</td>
<td>79.97</td>
</tr>
<tr>
<td>Years to degree</td>
<td>4.17</td>
<td>4.62</td>
</tr>
<tr>
<td>Missing weekly earnings</td>
<td>0.38</td>
<td>0.32</td>
</tr>
<tr>
<td>Weekly earnings</td>
<td>805.85</td>
<td>647.09</td>
</tr>
<tr>
<td>Age when weekly earnings measured</td>
<td>28.80</td>
<td>28.70</td>
</tr>
<tr>
<td>Observations</td>
<td>218,030</td>
<td>21,664</td>
</tr>
</tbody>
</table>
Earnings trajectories

Quarterly Earnings, excluding zeroes

Quarters, starting with summer preceding first enrollment

- Gets BA
- Does not get BA
Forcing variable

• Forcing variable is either term GPA or cumulative GPA depending on school’s policies.

• Sample is those we identify to be at risk of dismissal.
  – If school gives 3 terms to rehabilitate GPA, we are focused on the GPA in the 3\textsuperscript{rd} consecutive term below standard.
Empirical approach

• Stack all the schools and all the terms so some students appear multiple times.
• Include term fixed effects so that we only use cross-sectional variation.
• Cluster standard errors by student to account for serial correlation within a person.
  – Also cluster standard errors on running variable as suggested by Lee and Card (2008).
Empirical approach

• Local linear regression (Gelman and Imbens, 2014)

• Normalize GPA so that performance threshold is at zero.
First stage: Enrollment in t+1

Discontinuity: -0.128*** (0.0147)
1\textsuperscript{st} year students

Discontinuity: -0.044 (0.0290)
2nd year or higher standing

Discontinuity: -0.152*** (0.0156)
Sample restriction

• Policies only (normally) bind for upper year students.
  – We see this in the data, and anecdotally from talks with administrators.

• We focus remainder of analysis on upper year students.
Main outcome

• Log weekly earnings measured 7-12 years after college entry.

• For 2000 cohort we see them for up to 12 years.

• For 2005 cohort we see them for up to 7 years.
Mid-run earnings

Log weekly earnings

GPA in critical term

Discontinuity: -0.056*** (0.0203)
Identification threats

• Endogenous sorting
  – Higher ability students are more easily able to manipulate the threshold

• Attrition
Identification threat #1

• Students may sort around threshold to avoid probation/dismissal.
  – Policies are publicly known
  – Students with stronger dedication to school may strive to get GPA just above threshold.
Empirical tests

• Histogram test
• Covariate smoothness test
Heaping at whole numbers
Histogram

Fraction

GPA in critical term

0
-1

0.02
0.04
0.06
0.08

-0.5

0

0.5

1
Covariate smoothness
Covariate smoothness estimates

<table>
<thead>
<tr>
<th>Sample</th>
<th>All at-risk students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credits earned at t-1</td>
<td>-1.420</td>
</tr>
<tr>
<td></td>
<td>(1.532)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.0259</td>
</tr>
<tr>
<td></td>
<td>(0.0189)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.00121</td>
</tr>
<tr>
<td></td>
<td>(0.0130)</td>
</tr>
<tr>
<td>Age at college entry</td>
<td>-0.0514</td>
</tr>
<tr>
<td></td>
<td>(0.0628)</td>
</tr>
<tr>
<td>Term GPA in term 1</td>
<td>0.0247</td>
</tr>
<tr>
<td></td>
<td>(0.0313)</td>
</tr>
<tr>
<td>Credits attempted in time t</td>
<td>0.0294</td>
</tr>
<tr>
<td></td>
<td>(0.133)</td>
</tr>
<tr>
<td>Ln(weekly earnings) term t-1</td>
<td>0.00541</td>
</tr>
<tr>
<td></td>
<td>(0.0328)</td>
</tr>
<tr>
<td>Employed in year t-1</td>
<td>0.0138</td>
</tr>
<tr>
<td></td>
<td>(0.0159)</td>
</tr>
<tr>
<td>Predicted mid-run ln(weekly earnings)</td>
<td>0.00713</td>
</tr>
<tr>
<td></td>
<td>(0.00693)</td>
</tr>
<tr>
<td>Predicted BA</td>
<td>0.000992</td>
</tr>
<tr>
<td></td>
<td>(0.00884)</td>
</tr>
<tr>
<td>Predicted total credits earned</td>
<td>-0.127</td>
</tr>
<tr>
<td></td>
<td>(0.683)</td>
</tr>
<tr>
<td>Predicted ln(weekly earnings) term t+1</td>
<td>0.00169</td>
</tr>
<tr>
<td></td>
<td>(0.0144)</td>
</tr>
<tr>
<td>Observations</td>
<td>14,071</td>
</tr>
</tbody>
</table>
Does sorting substantially bias our estimates?

- Histogram test suggests possibility of sorting.
- All covariates are smooth.
- If discontinuity is driven by sorting on unobservables, these are completely uncorrelated with observables (including prior wages/employment).
More evidence on sorting

• Oster (Forthcoming)
  – Estimate degree of unobservable sorting necessary to explain away our results
  – Ranges from 800% (very conservative assumption) to 23200% greater than observable selection
  – Large effects due to fact that observables are highly predictive of future wages, but there is no discontinuity in any observable factor
More evidence on sorting

- Gerard, Rokkanen and Rothe (2016)
  - Assume unobservable sorting is infinitely large and not related to observable sorting
  - Use discontinuity in CDF to derive bounds on estimates
  - Intuition similar to Manski/Lee bounding.
  - Results in paper, bounds are tight even under extreme assumption
Identification threat #2

• We don’t have earnings for people who leave Ohio or are not employed.

• If college causes a different type of person to be missing, this can bias estimates.
Potential for attrition bias

Employed in Ohio

Unemployed in OH
Left state

Employed in Ohio

Unemployed in OH
Left state

Treatment group

Control group
Suppose treatment causes leaving state

- **Treatment group**
  - Employed in Ohio
  - Unemployed in OH
  - Left state

- **Control group**
  - Employed in Ohio
  - Unemployed in OH
  - Left state
Test for differential attrition

Missing earnings year 7-12

GPA in critical term

Discontinuity: -0.017 (0.0143)
Intermediate outcomes

• We look at
  – Total credits earned.
    • 30 credits = 1 year of school
  – Credits earned at other public schools
  – BA receipt from first school
  – BA receipt from other public school
  – AA receipt from other public school
Total credits earned

Discontinuity: -4.850*** (1.1438)
Credits at other schools

Discontinuity: 0.815** (0.3771)
BA from first school

BA from starting institution

GPA in critical term

Discontinuity: -0.108*** (0.0146)
BA from other schools

BA earned at other schools

GPA in critical term

Discontinuity: -0.001 (0.0034)
Robustness

• Both earnings and educational outcome results are robust.
  – Varying bandwidth
  – Including covariates
  – Donut RD
## Robustness

<table>
<thead>
<tr>
<th>Variable</th>
<th>.25</th>
<th>.25</th>
<th>.25</th>
<th>.50</th>
<th>.50</th>
<th>.50</th>
<th>.75</th>
<th>.75</th>
<th>.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.449)</td>
<td>(1.244)</td>
<td>(1.336)</td>
<td>(1.141)</td>
<td>(1.026)</td>
<td>(1.057)</td>
<td>(0.962)</td>
<td>(0.917)</td>
<td>(0.906)</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>9163</td>
<td>9163</td>
<td>8711</td>
<td>14618</td>
<td>14618</td>
<td>14166</td>
<td>18045</td>
<td>18045</td>
<td>17593</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.282</td>
<td>0.328</td>
<td>0.331</td>
<td>0.291</td>
<td>0.328</td>
<td>0.329</td>
<td>0.295</td>
<td>0.327</td>
<td>0.328</td>
</tr>
<tr>
<td>BA from starting institution</td>
<td>-0.0989***</td>
<td>-0.107***</td>
<td>-0.103***</td>
<td>-0.108***</td>
<td>-0.117***</td>
<td>-0.117***</td>
<td>-0.134***</td>
<td>-0.142***</td>
<td>-0.144***</td>
</tr>
<tr>
<td></td>
<td>(0.0183)</td>
<td>(0.0174)</td>
<td>(0.0196)</td>
<td>(0.0147)</td>
<td>(0.0135)</td>
<td>(0.0146)</td>
<td>(0.0133)</td>
<td>(0.0121)</td>
<td>(0.0126)</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>9157</td>
<td>9157</td>
<td>8705</td>
<td>14606</td>
<td>14606</td>
<td>14154</td>
<td>18026</td>
<td>18026</td>
<td>17574</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.0926</td>
<td>0.123</td>
<td>0.124</td>
<td>0.109</td>
<td>0.136</td>
<td>0.138</td>
<td>0.115</td>
<td>0.143</td>
<td>0.144</td>
</tr>
</tbody>
</table>
Robustness

<table>
<thead>
<tr>
<th>Variable</th>
<th>Bandwidth</th>
<th>.25</th>
<th>.25</th>
<th>.25</th>
<th>.50</th>
<th>.50</th>
<th>.50</th>
<th>.75</th>
<th>.75</th>
<th>.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid-run log weekly earnings</td>
<td>-0.0660**</td>
<td>(0.0259)</td>
<td>-0.0455*</td>
<td>(0.0237)</td>
<td>-0.0497*</td>
<td>(0.0260)</td>
<td>-0.0563***</td>
<td>(0.0203)</td>
<td>-0.0453**</td>
<td>(0.0187)</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>6448</td>
<td>6448</td>
<td>6132</td>
<td>10228</td>
<td>10228</td>
<td>9912</td>
<td>12533</td>
<td>12533</td>
<td>12217</td>
<td></td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.00966</td>
<td>0.139</td>
<td>0.141</td>
<td>0.00951</td>
<td>0.133</td>
<td>0.134</td>
<td>0.0110</td>
<td>0.134</td>
<td>0.134</td>
<td></td>
</tr>
</tbody>
</table>
Instrumental Variable Estimates

- Scale earnings estimate by credits earned estimate to get IV estimate of returns to a year of schooling.

- Our estimate imply 25% return to an additional year of college

- $0.0485/(5.68/30)$
Tracing out the returns

Net Present Value Over the Lifecycle

-20000
0
20000
40000
60000

Quarters since possible dismissal

Private Return
Social Return
Social Return (including deadweight loss)
Conclusion

• Low-achieving students benefit from college

• No implication regarding dismissal policy efficacy.

• Dismissal policies might
  – Provide incentives to work hard
  – Protect signaling value of degree
  – Hurt the dismissed students.