Employing Machine-Learning Approaches to Predict Incomes of Recent College Graduates

Benjamin T. Skinner
University of Florida
btskinner

Olivia Morales
University of Florida
olivianmo

William R. Doyle
Vanderbilt University
wdoyle42

UF | College of Education
UNIVERSITY *of* FLORIDA

16 November 2022
Association for the Study of Higher Education Annual Conference | Las Vegas, NV

The Prediction Business for Returns to Higher Education

What we know

- Wealth of scholarship to support economic returns from higher education
- Prior analyses limited by access to institution-level data not widely available
  - The publication of the College Scorecard (2015) changed the higher education research landscape, with national institutional-level data now publicly accessible

Examining existing research

- Standard econometric approaches to estimating earnings are vulnerable to researcher bias and produce less accurate predictions
- Machine learning methods provide a more useful alternative (predictions made algorithmically, based on features of the data)

Our project

- Marrying the institutional/program-level data available in the Scorecard with machine learning methods to predict incomes for recent graduates

Data

College Scorecard

- 2010-2020
- Includes over 2,000 variables for 6,700 accredited higher education institutions (HEIs) in the U.S.
- Significant portions of the data are privacy suppressed

American Community Survey (ACS)

- 2015-2019
- Geography of interest: County level (using FIPS)
- Matching HEIs with unique county FIPS codes
- Helps recover information lost from missing data in the Scorecard

Methodology and Model Specification

$$y_i = \alpha + c_i\gamma + \epsilon_i$$

where

y_i is the outcome variable (median earnings of graduates one year after graduation)
α is the intercept
c_i is a vector of covariates
γ is a vector of coefficients for those covariates
ϵ_i is an error term

Thank You!

Olivia N. Morales
oliviamorales@ufl.edu
olivianmo
onmorales23

UF | College of Education
UNIVERSITY *of* FLORIDA

References I

Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2017). *Predicting student dropout in higher education*. doi: 10.48550/arXiv.1606.06364

Beaulac, C., & Rosenthal, J. S. (2019). Predicting university students' academic success and major using random forests. *Research in Higher Education*, *60*(7), 1048-1064. doi: 10.48550/arXiv.1802.03418

Boland, W. C., Gasman, M., Castro Samayoa, A., & Bennett, D. (2021). The effect of enrolling in Minority Serving Institutions on earnings compared to non-minority serving institutions: A College Scorecard analysis. *Research in Higher Education*, *62*, 121–150.

Bombardieri, M. (2020, October 28). *Tapping local support to strengthen community colleges*. Retrieved from `https://www.americanprogress.org/article/tapping-local-support-strengthen-community-colleges/`

Card, D. (1993, October). *Using Geographic Variation in College Proximity to Estimate the Return to Schooling* (Working Paper No. 4483). National Bureau of Economic Research. doi: 10.3386/w4483

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In L. N. Christofides, E. K. Grant, & R. Swidinsky (Eds.), *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*. Toronto: University of Toronto Press.

Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics*, *3*, *Part A*, 1801–1863.

Card, D. (2001, September). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, *69*(5), 1127–1160. doi: 10.1111/1468-0262.00237

Carnevale, A. P., Rose, S. J., & Cheah, B. (2011). *The college payoff: Education, occupation, lifetime earnings*. Georgetown University Center on Education and the Workforce. Retrieved from `http://hdl.handle.net/10822/559300`

The college scorecard: Institutional characteristics and graduate outcomes. [Article]. (2016, 02). *NACE Journal*, *76*(3), 16–22. Retrieved from `https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=eue&AN=122470898&site=ehost-live`

Dowd, A. C., & Grant, J. L. (2006). Equity and efficiency of community college appropriations: The role of local financing. *The Review of Higher Education*, *29*(2), 167-194. doi: 10.1353/rhe.2005.0081

Doyle, W. R., & Skinner, B. T. (2016). Estimating the education-earnings equation using geographic variation. *Economics of Education Review*, *53*, 254–267. doi: 10.1016/j.econedurev.2016.03.010

Elu, J. U., Ireland, J., Jeffries, D., Johnson, I., Jones, E., Long, D., … Trotman, J. (2019, September). The earnings and income mobility consequences of attending a Historically Black College/University: Matching estimates from 2015 U.S. Department of Education College Scorecard Data. *The Review of Black Political Economy*, *46*(3), 171–192. Retrieved from `https://doi.org/10.1177/0034644619866201`

Hastie, T., Tibshirani, R., & Friedman, J. (2016). *The elements of statistical learning: Data mining, inference and prediction, second edition*. Springer.

Higher Education Appropriations Subcommittee. (2022). *House of representatives staff analysis* (Tech. Rep.). Florida House of Representatives. Retrieved from `https://www.myfloridahouse.gov/Sections/Documents/loaddoc.aspx?PublicationType=Committees&CommitteeId=3089&Session=2022&DocumentType=proposed%20committee%20bill%20analyses&FileName=pcb01a.HEA.pdf`

Hout, M. (2012). Social and economic returns to college education in the united states. *Annual Review of Sociology*, *38*(1), 379-400. doi: 10.1146/annurev.soc.012809.102503

Hout, M., Levanon, A., & Cumberworth, E. (2011). Job loss and unemployment. In *The great recession* (p. 82-126). New York, NY: Russell Sage Foundation.

Hoynes, H., Miller, D. L., & Schaller, J. (2012). Who suffers during recessions? *American Economic Association*, *26*(3), 27-48. doi: 10.1257/jep.26.3.27

Huntington-Klein, N. (2017). *The search: the effect of the college scorecard on interest in colleges*. California State University - Fullerton. Retrieved from `https://www.nickchk.com/Huntington-Klein_2017_The_Search.pdf`

References II

Hurwitz, M., & Smith, J. (2018). Student responsiveness to earnings data in the College Scorecard. *Economic Inquiry*, *56*(2), 1220–1243.

Iacobucci, W., Dall, T., Chakrabarti, R., Reynolds, R., & Jones, K. (2021). *Florida nurse workforce projections: 2019 to 2035* (Tech. Rep.). IHS Markit. Retrieved from `https://www.fha.org/uploads/1/3/4/0/134061722/ihs_florida_nurse_workforce_report.pdf`

Iatrellis, O., Savvas, I. K., Fitsilis, P., & Gerogiannis, V. C. (2021). A two-phase machine learning approach for predicting student outcomes. *Education and Information Technologies*, *26*, 69-88. doi: 10.1007/s10639-020-10260-x

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, *86*, 4-29. doi: 10.1162/003465304323023651

Kim, C., & Tamborini, C. R. (2019). Are they still worth it? the long-run earnings benefits of an associate degree, vocational diploma or certificate, and some college. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *5*(3), 64-85. doi: 10.7758/RSF.2019.5.3.04

Kuhn, M., & Silge, J. (2022). *Tidy modeling with r: A framework for modeling in the tidyverse.* O'Reilly Media. Retrieved from `https://www.tmwr.org`

Li, Q., & Lin, N. (2010). The bayesian elastic net. *Bayesian Analysis*, *5*(1), 151-170. doi: 10.1214/10-BA506

Mabel, Z., Libassi, C., & Hurwitz, M. (2020, April). The value of using early-career earnings data in the College Scorecard to guide college choices. *Economics of Education Review*, *75*, 101958. doi: 10.1016/j.econedurev.2020.101958

Office of the Press Secretary. (2013, February 12). *Remarks by the president in the state of the union address.* Retrieved from `https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/remarks-president-state-union-address`

Oreopoulous, P., & Petronijevic, U. (2013). *Making college worth it: A review of research on the returns to higher education.* National Bureau of Economic Research. Retrieved from `https://www.nber.org/system/files/working_papers/w19053/w19053.pdf`

Posillico, J. J. (2017). *Impacting student access through federal policy changes: How college presidents interpret the College Scorecard* (Unpublished doctoral dissertation). Northeastern University.

Seaman, J., Bell, B. J., & Trauntvein, N. (2017). Assessing the value of a college degree in outdoor education or recreation: Institutional comparisons using the College Scorecard and surveys of faculty and employers. *Journal of Outdoor Recreation, Education & Leadership*, *9*(1), 26–41. Retrieved from `https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=s3h&AN=121227337&site=ehost-live`

Skinner, B. T., & Doyle, W. R. (2021). Do civic returns to higher education differ across subpopulations? an analysis using propensity forests. *Journal of Education Finance*, *46*(4), 519–562.

State Higher Education Executive Officers Association. (2021). *State higher education finance (SHEF) report.* State Higher Education Executive Officers Association. Retrieved from `https://shef.sheeo.org/wp-content/uploads/2022/06/SHEEO_SHEF_FY21_Remark.pdf`

Terenzini, P. T. (1996). Rediscovering roots: Public policy and higher education research. *The Review of Higher Education*, *20*(1), 5-13. doi: 10.1353/rhe.1996.0006

Zeineddine, H., Braendle, U., & Farah, A. (2021). Enhancing prediction of student success: Automated machine learning approach. *Computers and Electrical Engineering*, *89*, 106903. doi: 10.1016/j.compeleceng.2020.106903

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*, *67*(2), 301-320. doi: 10.1111/j.1467-9868.2005.00503.x

Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics*, *37*(4), 1733-1751. doi: 10.1214/08-AOS625