Brandon Ly
Collin Zoeller
Jane Shumway
Nicholas Van Slooten

12/8/2021

Business School Majors and MBB Offers

Introduction

In this paper, we explore the causal relationship between BYU business school degrees and management consultancy offers at the three most prestigious global management consulting firms — McKinsey, BCG, and Bain (MBB). Our question is, "What is the causal effect of being a major in the business school on receiving an MBB job offer?" This question is important for 1) undergraduate students interested in working in the field of management consulting who are deciding on a college major and 2) undergraduate students in non-business majors who are deciding whether or not they want to put in the effort to recruit with MBB.

Two common questions that BYU undergraduate students ask about management consulting are some version of the following:

1. "I'm a freshman/sophomore, and I want to do consulting, but I'm trying to decide what major will help me get to a top consulting firm. Should I choose a major in the business school, or something else?"
2. "I'm a mechanical engineering major, and I realized after an internship that I do not want to pursue engineering as a career path. I just learned about consulting, and I'd like to work for a top consulting firm, but am I at a disadvantage because I'm not a major in the business school?"

Our machine learning project seeks to provide a quantitative answer to those questions that can inform those students' decisions. If there is a positive relationship between majoring in the business school and receiving a MBB offer, then new students may be well-advised to choose a major in the business school, and non-business students might be dissuaded from pursuing MBB firms. If there is no relationship, then students may be more likely to consider a major outside of the business school and non-business majors could be encouraged to pursue opportunities with MBB.

Choosing a major in the business school may help students gain skills directly related to management consulting such as communication, presentation, teamwork, and leadership skills. On the other hand, popular STEM-related majors such as Economics, Mathematics, Computer Science, and Engineering, are often recommended to those interested in pursuing a career in management consulting.[1] Majoring in these fields may help students gain analytical and technical skills needed to become a management consultant. However, these STEM-related majors do not fall under BYU's Marriott School of Business.

To the degree that majoring in the business school increases (decreases) the likelihood of receiving an MBB offer, non-business (business) majors are an education-occupation mismatch for students who want to work in management consulting. Other researchers have found that education–occupation mismatches decrease earnings on average,[2] suggesting that for students considering management consulting, the choice of major is nontrivial. Our paper adds to the existing literature by using a new dataset, described below, and by identifying the type of major that best matches a career in management consulting.

Estimating the true effect of being a major in the business school on getting an offer from MBB firms will be confounded by underlying unobserved factors such as a student's intelligence or drive. We control for these factors by using proxies such as ACT and GPA scores, along with machine learning methods we describe below.

Key Terms

MBB — McKinsey, BCG, or Bain; the three most prestigious[3] global management consulting firms, also typically the most sought-after firms by BYU students who want to pursue consulting.

Non-MBB — Other top consulting firms that are not McKinsey, Bain, or BCG.

MCA — The BYU Management Consulting Association, an organization that helps BYU students to land offers to work for management consulting companies.

Data

We obtained data from the MCA archives. Since 2009, the MCA has kept a record of each undergraduate who has received and accepted consulting offers at MBB and non-MBB consulting firms along with several features about the students, including each student's major, minor, ACT score, undergraduate GPA, gender, consulting internship experience (if any), and the firm at which the student accepted a job offer. We refer to these data as "MCA Placement Data", indicating BYU undergraduate students' placement at top consulting firms. These data are summarized in Table 1 below, and we will provide more detail on each individual feature below, as well as the processes used to construct transformations of the data based on these initial features as well as mitigate missing values.

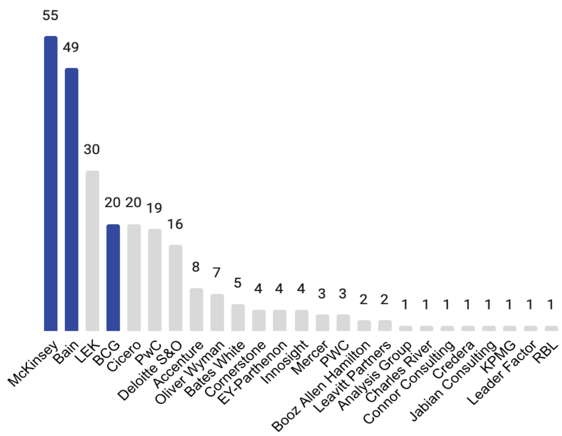
Table 1: MCA Placement Data Summary Statistics

Feature	n	Mean	St. Dev.	Explanation
Firm Category*	**259**	**0.48**	**0.50**	**1 = "MBB"**
Business?**	**259**	**0.58**	**0.50**	**1 = Business Major**
Gender	259	0.14	0.35	1 = Female
Double Major?	259	0.05	0.23	1 = Double Major
Minor?	259	0.36	0.48	1 = Had Minor
Internship?	162	0.38	0.49	1 = Had Consulting Internship
Firm Category	259	0.48	0.50	1 = "MBB"
Firm Location	252			
ACT	221	31.37	2.88	
GPA	220	3.86	0.15	
Year	259			

*Outcome variable
**Variable of interest

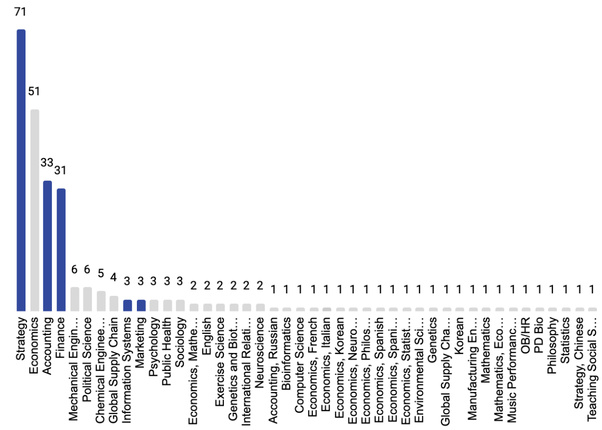
The distributions of the outcome variable (left) and variable of interest (right) in the MCA placement data are shown below in Figure 1 and Figure 2 respectively. It is important to note that more business students (58% of consulting offers) have received consulting offers than non-business students (42% of consulting offers) in the 2009-2021 time period.

*Figure 1: Consulting Offers by Firm**



*Blue firms are MBB, "Firm category" = 1

*Figure 2: BYU Consulting Offerees by Major**



*Blue majors are business, "Business?" = 1

Another item of note regarding the outcome variable and variable of interest is data that is *not* present in the dataset — namely, data for students who interviewed with consulting firms but did not receive consulting offers or applied to consulting firms but did not receive an interview. Because we do not have these data, the causal inference estimate of major choice (between business or non-business major) based on the dataset is to be interpreted as the impact of major choice on the likelihood of receiving an MBB offer relative to a non-MBB offer, not the impact of major choice on the likelihood of receiving any consulting offer at all.

The remainder of the data are control variables, including binary control variables (gender status, whether or not a person has a double major, whether or not a person has a minor), the non-binary categorical control variable (offer location), and quantitative control variables (GPA and ACT). The distribution of each of these three types of control variable is visualized below (refer to the summary table above for mean values of binary control variables).

Figure 3: Binary Variables

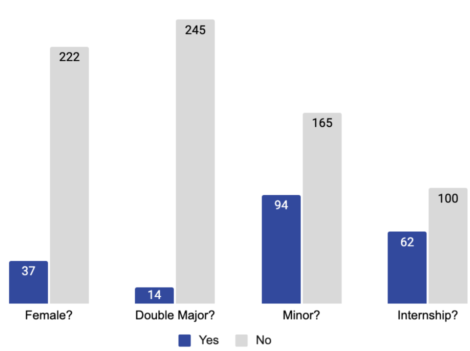


Figure 4: Non-binary Categorical (Location)

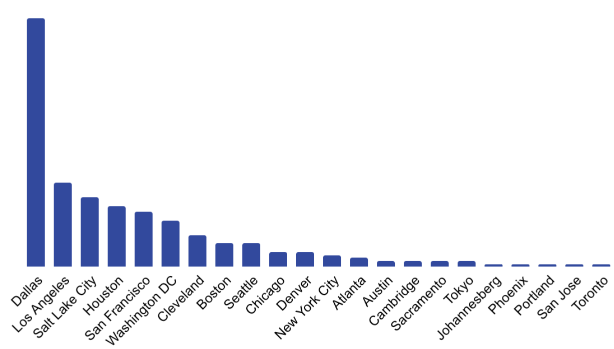
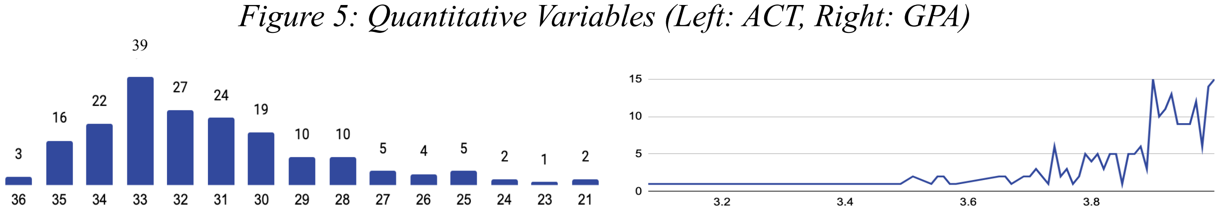


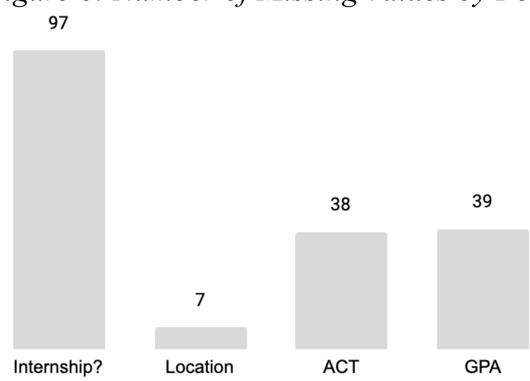
Figure 5: Quantitative Variables (Left: ACT, Right: GPA)



These visualizations paint a picture of the average consulting offeree from 2009-2021 — mostly male with one major, perhaps a minor and a consulting internship, and a relatively high ACT and GPA.

Some of these features had missing values; internship data were not recorded in 2009-2012 or 2014-2015, a few people did not list their offer location, and ACT and GPA data were sparse for 2009-2012 (see Figure 6 below). To mitigate this, for each of the four features with missing values, an indicator variable was added for "X Variable Present / Missing" (for example, if a person did not have ACT data, the "ACT Present/Missing indicator would take on a value of 0, or a person did have ACT data, the "ACT Present/Missing indicator would take on a value of 1). Missing values were then filled with 0 as a placeholder. By creating these "missing" indicators, we reduce the negative impact of a missing value on estimation power, enabling the model to recognize the value as missing rather than using the 0 placeholder as its value.

Figure 6: Number of Missing Values by Feature



Finally, we transformed the data to improve the model's estimation power. To leverage location as an input, we created indicator variables for each offer location in the dataset. After discovering heterogeneous effects of the control variables on the outcome variable by year, we also added indicator variables for each year in the dataset. Finally, to control for potential non-linear relationships between the control variables and the outcome variable, we generated polynomial features up to degree 2 (e.g. GPA^2) and interaction terms (e.g. ACT * Gender) for every possible interaction between features. This grew the data from 10 to 1128 control features.

Table 2: Data Transformation and Control Feature Expansion

Data	# control features
Original dataset	10
Adding location, year indicators	43
Adding missing value indicators	47
Adding polynomial, interaction terms	1128

Key Assumptions

In using this data and structuring our model, we make the following key assumptions:

1. **Conditional on the additional covariates in our dataset (GPA, ACT, etc.), being a business major is unrelated to other determinants of getting an offer such as intelligence, ability, ambition, parents' income, one's network with consultants, or problem solving intuition**.

    a. ***Why is this important?*** If GPA and ACT do not capture unobserved determinants of getting an offer (for example, ambition) that are correlated with being a business major or a non-business major, the impact of business major status will be overstated or understated as it will represent both its own impact and the unobserved, correlated factor(s).

    b. ***How likely is it that this assumption is correct?*** Not likely; there are many determinants that play into receiving a consulting offer (see list above), and some of them are likely correlated with one's business major status. For example, a business major is likely to have a better network with consultants than a non-business major (given that 58% of consulting offerees from BYU in the dataset were business majors).

    c. ***How did we verify this assumption?*** This assumption cannot be completely verified, though GPA and ACT are a common proxy for intelligence used in other academic studies. This is a limiting factor that may reduce the power of our model to estimate true causal impact (see "Limitations" below).

2. **MCA data correctly identifies the job offers a person accepted.**

    a. ***Why is this important?*** If many people actually accepted a different job offer than the one listed in the MCA database, the model would produce an estimator trained on incorrect data for the outcome variable, which would likely perform poorly when making out-of-sample predictions for new observations.

    b. ***How likely is it that this assumption is correct?*** Very likely; the MCA often works with BYU's strategy department to get students' job information.

    c. ***How did we verify this assumption?*** We spot-checked a small random sample using LinkedIn, and we found no incorrectly recorded full-time job offers.

3. **MCA data is not missing any observations that are correlated with control variables.**

    a. ***Why is this important?*** If many people who received consulting offers are not listed in the MCA database, and the missing data people are correlated with the control variable (e.g. people not in the database tend to be non-business majors), the model would be biased toward the structure of the training data, which could perform poorly when making out-of-sample predictions for new observations.

    b. ***How likely is it that this assumption is correct?*** Somewhat likely; however, it could be possible that people who received consulting offers from non-business majors are more likely not a part of the MCA and thus the MCA has no record of their offers.

c. ***How did we verify this assumption?*** We picked 3 years and did LinkedIn searches to find any individuals not recorded in the list who started working at consulting firms right out of undergraduate study at BYU, and the few we did find were relatively evenly distributed across majors, firm types, etc.

4. **Individuals who accepted offers from non-MBB consulting firms wanted an MBB offer but did not receive one.**

   a. ***Why is this important?*** The students with offers at non-MBB firms are the counterfactual that will enable us to find the causal relationship of major (business/non-business) and receiving an MBB offer. If students at non-MBB firms could have received MBB offers but chose not to, the data will not truly reflect the impact of "major" on receiving an MBB offer.

   b. ***How likely is it that this assumption is correct?*** Based on our experience with students interested in consulting, we anticipate this assumption is correct in the majority of cases — that is, most BYU alumni consultants with non-MBB offers originally wanted to work at an MBB firm unless they had a confounding constraint (for example, they needed to live in Utah with a spouse in school).

   c. ***How did we verify this assumption?*** After looking at the entire dataset, we discovered only one person who switched from an MBB internship to a non-MBB full-time offer, while many people switched from a non-MBB internship to an MBB full-time offer.

Methods

Machine learning methods allow us to control flexibly and robustly for covariates. We primarily use the double debiased machine learning (DDML) method outlined by Chernozhukov et al.[4] This method has three steps:

1. Predict y (whether a student got a job offer at an MBB firm) from all features other than the explanatory variable of interest d (whether a student was a business major) and recover the residuals.
2. Predict d (whether a student was a business major) from all of the other features and recover the residuals.
3. Regress the residuals of (1) on (2) to find the causal effect of being a business major on getting a job offer from an MBB firm.

Various ML methods may be used for steps (1) and (2). To select the best ML methods for steps (1) and (2), we do the following:

4. Divide the data set at random into training and test sets (.8 and .2).
5. For each ML method for both y and d,
   a. Fit the model on the training set using sample splitting.
   b. Estimate the model's out-of-sample performance via the average MSE on the test set.
6. Choose the best ML method based on the lowest test MSE for both y and d.

On step 5(a), we chose the tuning parameters for each DDML prediction step via k-fold cross-validation:

7. Used grid search over hyperparameters (e.g. C/α, γ, kernel, entropy, max depth of tree, min leaves)
8. 5-fold CV:
    a. Split the training data into 5 folds
    b. Fit the model on 4 folds while leaving out 1 fold each time to calculate MSE
    c. Select hyperparameters with the lowest average MSE

The DDML methods for predicting *y* and *d* we use include:

- Lasso Regression
- Random Forest Regression
- Support Vector Machines (SVM)
- Logistic Regression (Logit)

For our final DDML model, we choose the ML method(s) that have the lowest test MSE. Note that the optimal ML method to predict *y* may differ from the optimal method to predict *d*.

Additionally, we perform Post-Double Selection (PDS) - which assumes sparsity of covariates - to obtain an estimate of the causal effect of being a business major for comparison to the DDML model. We also perform causal forests to analyze heterogeneous treatment effects that might give us insight onto covariates of interest to include in the above models.

Note that for choosing PDS's hyperparameters, we manually select larger α's such that the remaining number of covariates is less than the natural log of our sample size: # remaining covariates < 5.

Prior to making interpretable decisions from these models, we perform Causal Forest Classification to test for any significant level of heterogeneity not accounted for in any of the selected methods. This is to check the assumption of asymptotic normality requisite for proper inference of causality.

If there is significant heterogeneity, the effect of having a business major will be different for different subsets of individuals, which would diminish the accuracy of any suggestion made to a student asking if they should pursue a business major as its effect on similar individuals may be different if not accounted for in the model.

We follow the Causal Forest framework described in Wager and Athey (2018) to form a propensity tree trained via Random Forest that places individuals in a binary final node: either they received an offer or they did not. Since the intent of including Causal Forest for our purposes is purely to check for heterogeneity, we refrain from including a full summary of the theory that can be found in the original author's work. We will suffice to say that the final node is scored based on its relative homogeneity within the node: a measure of how similar observations are within each node.

Results

The results from each model were mostly consistent. DD LASSO, DD random forest, DD logit, and PDS LASSO all estimated a negative coefficient on whether someone was a business major. DD SVM estimated a positive coefficient, but along with the DD random forest and DD LASSO, the confidence interval includes zero. These results and the performances of each method are displayed in Table 3.

These double debiased ML methods appear to show that being a business major had no effect on whether a student received an offer from an MBB firm, although the logit model did have a statistically significant negative coefficient. However, the confidence intervals for the DDML methods are biased. PDS LASSO was the only method that generated unbiased confidence intervals, and its coefficient on being a business major was statistically significant at the 5% level. If our assumptions hold, this means that being a major in the business school decreases the likelihood of getting an offer from an MBB firm on average. This seems counterintuitive to conventional expectations.

Table 3: K-Fold Cross-Validated Output for Each DD Method (and PDS)

Model	DD LASSO	DD Random Forest	DD SVM	DD Logit	PDS LASSO
Parameters on y	α = 0.0104	entropy max depth = 4 min leaf samp = 1	C = 0.1 $\gamma = 1$ kernel = linear	C = 0.3594	α = 0.1
MSE_y	0.1532	0.2154	0.1615	0.223	0.2333
Parameters on d	α = 0.0096	gini max depth = 5 min leaf samp = 2	C = 0.1 $\gamma = 1$ kernel = linear	C = 2.7826	α = 0.1
MSE_d	0.2035	0.35	0.4961	0.285	0.2378
Coefficient Effect	-0.146	-0.094	0.127	-0.105	-0.248
Standard Error	0.1433	0.1335	0.0887	0.119	0.075
95% Confidence	[-0.4269, 0.1349]	[-0.3557, 0.1677]	[-0.0469, 0.3008]	[-0.3382, 0.1282]	[-0.395, -0.101]

Since you either are a business major or you aren't, the sign of the effect is what really matters, not the magnitude. If being a business major decreases the likelihood of a student receiving a job offer even a little bit relative to the other potential majors, the average student who wants to become a management consultant should never choose to be a business major.

There are multiple reasons why a non-business major might be more likely than a business major to receive an MBB offer. There is increasing value in technical and non-business skills. Since non-business majors like mechanical engineering or economics are often objectively more rigorous, certain non-business majors may signal intelligence and hard work. Additionally, non-business majors are held to a lower interview standard. Case interviews are judged based on business acumen, and non-business majors may be expected to have less. A few particular non-business majors that are very useful for becoming a management consultant may be driving the results. For example, when we include the economics major as part of the business school, the effect of being a major in the business school becomes less negative.

Table 4: K-Fold Cross-Validated Output for Each DD Method (and PDS) Before and After Including Economics as a Major in the Business School

Model	DD LASSO	DD Random Forest	DD SVM	DD Logit	PDS LASSO
Coefficient (Before)	-0.146	-0.094	0.127	-0.105	-0.248
Coefficient (After)	-0.110	0.035	0.156	-0.002	-0.233

The Causal Forest spot-check on homogeneity also revealed, as would be expected, that there was a different effect between years of the application and offer but very little heterogeneity otherwise. There are many factors that play into a firm's decision to extend an offer exogenous to the individual in question. For example, COVID-19 restrictions in late 2019 and 2020 may have limited the number of individuals permitted to work on-site or otherwise encouraged firms to bring in less interns. To account for this, we assigned each year an indicator and reevaluated each method.

Table 5 displays the results of the best-performing LASSO model after fitting and testing it to the entire dataset.

Table 5: Output and Performance of Best-Performing Method

LASSO Final Regression	
LASSO test MSE for y	0.1532
Optimal Alpha (y)	0.0104
LASSO Test MSE for d	0.2035
Optimal Alpha (d)	0.0096
Coefficient Effect:	-0.1458
95% Confidence Interval:	[-0.4269, 0.1349]

Of all the methods, DD LASSO had the best performance (the lowest MSE) for predicting both y and d. This could be because LASSO assumes sparsity. If only a few of the covariates actually mattered for predicting y and d, LASSO would be ideal for identifying the most important features and throwing out the others.

Conclusion

We find that BYU undergraduate non-business majors are more likely to receive an MBB offer (relative to a non-MBB offer) than undergraduate business majors. This can inform students' decision in choosing a major tailored toward a consulting career, as picking a major outside of the business school (e.g. economics) may increase their chances of obtaining a top consulting offer. Additionally, non-business major students may take heart in our results which show that these students might not be at a disadvantage at securing a top consulting offer. Some implications of our findings include that the business school should consider eliminating the strategy major (a business major designed to prepare students for consulting) and that non-business major colleges should advertise consulting as a potential career path.

Limitations on our results include:

1. Given grade inflation in the BYU business school (for instance, BYU's Business Strategy program curves to the average grade of incoming students, often 3.8 or higher) could mean that someone with a given GPA in the business school is performing worse academically than someone with the same GPA outside the business school, causing the results to be artificially biased toward non-business students.
2. Non-business school students are less likely to seek consulting jobs in general. Since our sample precludes that students have already received a consulting offer, the non-business school students may have confounding factors (i.e. drive, networking) that bias our results toward them.
3. Even if this effect is accurately estimated, it does not provide a complete picture of what a student needs to do to receive a consulting offer.
    a. Students may change to a non-business major and still not get a consulting offer.
    b. This is the optimal choice to reduce randomness, not eliminate it.
4. The accuracy of this estimate is based on key assumptions (shown above) that may not be completely true.
    a. For example, not all people who chose to work at a non-MBB firm would have wanted to choose MBB.
    b. Mitigant: Even if some of the assumptions do not hold perfectly, they hold in the majority of cases, so the sign of the causal effect is likely correct, which is the most important part of the estimate in terms of economic significance.
5. The conclusion is highly dependent upon the majors we include as "business majors". When the Economics major is included in "business majors", the effect of being a business major on getting an MBB offer actually becomes positive (or less negative).
6. The conclusion is not statistically significant according to naive standard error calculations. Positive values for the effect of business major on getting an MBB offer are contained within the naive confidence intervals calculated.
    a. Further analysis should be done to determine more accurate standard errors (i.e. via bootstrapping techniques).

Possible future extensions of our results could include the following:

1. Gather more student data to model the causal impact of individual major choice for students
    - Who are not at BYU
    - Who apply to consulting firms and do not receive an interview
    - Who interview and do not receive consulting offers
2. Expand the target audience to include all BYU alumni who get consulting offers to show the long-term impact of major choice
    - People who go to graduate school and then apply and receive consulting offers
    - People who work for a few years and then apply and receive consulting offers
3. Expand the question to model the causal impact of major choice on receiving any high-paying job offer
    - Are students who choose business majors more likely to receive a $100K salary?

Notes

[1] Management Consultant Careers, What Degrees Are Best for a Career in Management Consulting
[2] For example, see Nordin et al. (2010), Zhu (2012), and Rios‑Avila and Saavedra-Caballero (2019).
[3] Best Consulting Firms to Work For
[4] Victor Chernozhukov et al., "Double/Debiased Machine Learning for Treatment and Structural Parameters," *The Econometrics Journal* 21, no. 1 (February 1, 2018): pp. C1-C68.

References

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal* 21, no. 1 (February 1, 2018): C1–C68.

Nordin, Martin, Inga Persson, and Dan-Olof Rooth. "Education–Occupation Mismatch: Is There an Income Penalty?" *Economics of Education Review* 29, no. 6 (December 2010): 1047–59.

Rios‑Avila, Fernando and Fabiola Saavedra-Caballero. "It Pays to Study for the Right Job: Exploring the Causes and Consequences of Education-Occupation Job Mismatch." *Levy Economics Institute*, Working Papers Series, no. 922 (February 5, 2019).

Wager, Stefan and Susan Athey. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests" *Journal of American Statistical Association* 113:523 (June 6, 2018): 1228-1242.

Zhu, Rong. "The Impact of Major–Job Mismatch on College Graduates' Early Career Earnings: Evidence from China." *Education Economics* 22, no. 5 (February 22, 2012): 511–28.