Causal Models for Predicting Time-to-degree

Prof. Gregory L. Heileman, The University of Arizona

Gregory (Greg) L. Heileman currently serves as the Vice Provost for Academic Affairs and Professor of Electrical and Computer Engineering at the University of Arizona.

Melika Akbarsharifi, The University of Arizona

Melika Akbarsharifi is a PhD student in Software Engineering at the University of Arizona, studying under Professor Gregory L. Heileman. Her research at the Curricular Analytics Lab focuses on applying machine learning and data analysis to improve educational outcomes. Key contributions include developing a cohort-tracking analytics platform that supports efforts to increase graduation rates by identifying and addressing curricular barriers.

Melika has co-authored papers presented at conferences such as the ASEE Annual Conference and Exposition, exploring the intersection of curriculum complexity and student performance. Her technical expertise spans multiple programming languages and cloud computing, supporting her work on innovative educational technologies.

Kristina A Manasil, The University of Arizona

Kristi Manasil is a third-year PhD student in the College of Information Science at the University of Arizona. She received her bachelor's degree in Computer Science from the University of Arizona. Her areas of focus are data visualization, machine learning, learning analytics, and educational data mining.

Causal Models for Predicting Time-to-degree

Abstract

This paper considers the use of causal models in determining how the complexity of the curricula students confront impacts the amount of time it takes them to complete their degrees. There are many factors, in addition to curricular complexity, that influence a student's time-to-degree (TTD), and these factors confound the relationship between curricular complexity and TTD. For instance, students with higher high school GPAs tend to self-select into higher complexity programs, e.g., engineering. The fact that these high-achieving students are more successful in college serves to confound the true impact of curricular complexity on TTD. In order to better understand how various factors contribute to a student's TTD, we constructed a causal model that allows us to separately consider the influence of each factor. The development of this causal model was informed and validated using the data collected as a part of a large-scale observational study involving thirty universities and more than 750,000 students. Using this model, we are able to isolate the impact of curricular complexity on TTD using techniques commonly applied in the medical profession to assess the efficacy of particular medical treatments or interventions. Specifically, in this study we consider curricular complexity to be the "treatment" a student receives, and their TTD is the "outcome" resulting from this treatment. We then apply a methodology based on generalized propensity scores to control for the influence various factors have on TTD, including gender, first generation status, Pell Grant award status, high school GPA, and ethnicity. Using this analysis we are able to more accurately determine the average treatment effect of curricular complexity on a student's TTD.

Introduction

A causal model relating curricular complexity to time-to-degree (TTD) involves disentangling the influence of curricular complexity from the myriad other factors impacting student success outcomes. The ability to isolate the impact curricular complexity has on a student's TTD has many important implications. For instance, accurate models would directly support *a priori* reasoning around curricular reform efforts, allowing for more accurate estimates of the likely outcomes of various reform alternatives. That is, causal models in this area allow the effects of curricular "interventions" to be studied without having to actually implement them, and then wait many years to assess the outcomes. Furthermore, these models support counterfactual reasoning scenarios, including exploration of hypothetical alternatives that are not a part of the data associated with any observational study, but which can be answered by utilizing the structure of the causal model.

In this paper we present a formal model for characterizing the causal impact numerous factors have on college students' TTD. This model was constructed using the data associated with a large-scale

observational study involving thirty different institutions of higher education in the United States, including more than 750,000 students. The aim of this study was to better characterize the impact of curricular complexity on various student success outcomes, including TTD. Curricular complexity in this case refers to the structural complexity of an academic curriculum.[3] The importance of TTD is reflected in the fact that any unnecessary delays in graduation may lead to numerous negative outcomes, including an increase in student debt, a decrease in lifelong earnings, a decrease in institutional ranking, etc. This issue is particularly relevant in engineering programs, which are notoriously difficult to complete in a timely manner.

One experimental approach for determining the causal impact of curricular complexity on TTD involves constructing a randomized control trail (RCT).[1] This approach involves randomly assigning students with various backgrounds characteristics to programs with different curricular complexities at different institutions. The approach is commonly used in medical studies to determine the efficacy of particular treatments. In our case, we can consider curricular complexity as the treatment a student receives, and TTD as the outcome. If we have a sufficiently large sample of students, this RCT will isolate the impact of curricular complexity on TTD relative to any other student background factors. Specifically, the randomization ensures that at each complexity level the biases associated with confounding variables are "averaged out." Because RCTs eliminate the biases introduced by confounders, they are considered the "gold standard" for estimating the impact of particular treatments on specific outcomes. Borrowing from the terminology of the medical profession, we will use the term average treatment effect (ATE) to designate the average causal effect of a treatment or intervention on an outcome across an entire population of study participants. In the case of a RCT, the ATE is easily estimated by comparing the average outcome of the treated and untreated subpopulations in the study, or by comparing the differences between two treatment levels if there are multiple possible treatments. From this, it is easy to see the ATE represents the average change in outcome an individual in a population would expect to experience if they were exposed to one level of treatment, versus a different level of treatment. Note that the intervention in our study is not binary, i.e., a person receives the intervention or not; rather, the students in our study received many different levels (i.e., doses) of the curricular complexity intervention, and the ATE in this case must be computed at each intervention level. A dose-response function can be constructed from these ATEs to describe how the average outcome changes as the intensity (i.e., dose) of the intervention varies.

Because it is not possible, indeed ethical, to require students to enroll in academic programs that are not of their choosing, an RCT is not practical for studying the impact of curricular complexity on a student's TTD. Rather, we have a retrospective dataset that includes the academic programs students graduated from over the 2010-2020 time period at the institutions involved in this study, along with various factors related to these students. Because this dataset lacks a controlled intervention, i.e., assignment of students to academic programs, the best we can do is design an observational study that closely approximates an RCT, and in so doing provide insights into the relationships between curricular complexity and student success outcomes without actually manipulating them. The main challenge associated with observational studies involves accounting for the various biases that are introduced due to confounders. In the following sections we first describe the data collected for this study, and then we demonstrate how various factors contained in the data set influence students' TTD. These confounding factors make it difficult to ascertain the impact of curricular complexity on TTD. Next, we describe a succession of causal models meant to describe

Complexity Range	TTD (years)	Male (percent)	HS GPA (z-score)	First Gen. (percent)	Pell Award (percent)	Hispanic (percent)
$[31.1, 54.6)$	4.074	0.33	-0.1543	0.38	0.50	0.33
$[54.6, 90.0)$	4.117	0.33	-0.0589	0.31	0.42	0.26
$[90.0, 148.0)$	4.170	0.35	-0.0008	0.30	0.39	0.23
$[148.0, 245.0)$	4.238	0.48	0.1377	0.26	0.33	0.20
$[245.0, 403.0)$	4.307	0.53	0.3024	0.27	0.40	0.22
$[403.0, 735.0]$	4.493	0.65	0.3935	0.23	0.32	0.21

Table 1: The relationship between curricular complexity and various other factors related to student academic success.

how these factors impact TDD. Finally, We describe a generalized propensity score analysis that controls for these factors, allowing us better understand the impact of curricular complexity on TTD.

The Dataset

In this observational study, we started from the set of all students in our dataset who graduated over the past ten years, and we modeled the impact curricular complexity had on the number of years it took these students to graduate. Because this backwards-looking approach starts from those who graduated, and then traces their trajectories backwards in time in order to uncover how they achieved their degrees, it serves to isolate factors such as the complexity of the curriculum a student confronted. Specifically, we know each student in this dataset confronted, at the very least, the complexity of the curriculum associated with the degree program they graduated from, i.e., the curriculum they completed, even if they had a change of major along the way. The same can be said for all other students who graduated from the same degree program.

As mentioned above, our goal is to better understand the impact of curricular complexity on TTD, accounting for various confounding factors. For instance, it is well known that students with superior high school preparation, as indicated by high school GPA, are more likely to graduate on time. However, our data shows these same students are more likely to choose majors with higher curricular complexity, thereby negatively impacting their TTD. Thus, high school GPA is a confounding variable, and there are numerous other such confounding variables related to TTD, including first generation status, ethnicity, gender, Pell Grant award status, etc. In Table 1 we breakdown the variation of these factors according to the complexity of the programs the students in our study graduated from. This table shows the percentage of males, high school GPA, and TTD all tend to increase as the complexity of a degree programs increases. On the other hand, the percentage of first generation. Pell grant award, and Hispanic students tend to decrease as curricular complexity increases. That is, it appears male students and students with higher high school GPAs tend to self select into higher complexity programs, and first generation and Pell award students tend to self select toward lower complexity programs. Because these factors also impact TTD, this selection biases needs to be accounted for when assessing the impact of curricular complexity on TTD. In Table 1, the average value of each quantity was computed within six curricular complexity intervals. Because the distribution of curricular complexity is highly skewed, a logarithmic trans-

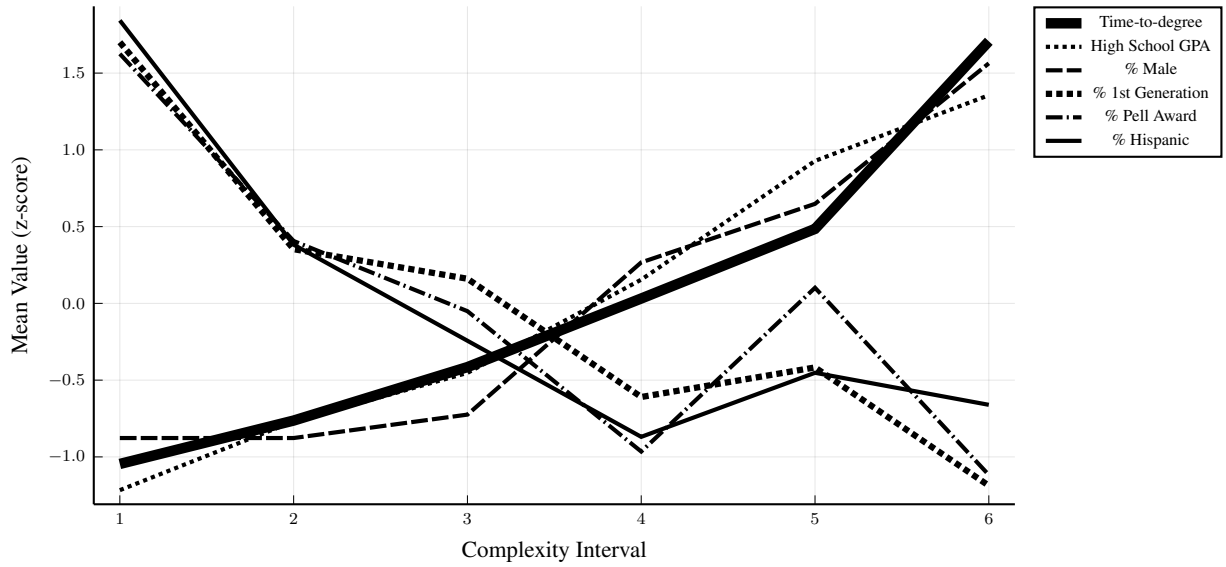

Figure 1: The average values of various factors within the six different curricular complexity intervals, where each factor was standardized using z-scores.

Complexity Range	TTD	Male	HS GPA	First Gen.	Pell Award
$[31.1, 54.6)$	34.41	40.85	52.85	39.69	47.43
$[54.6, 90.0)$	39.16	66.23	48.59	14.12	22.98
$[90.0, 148.0)$	17.21	58.55	25.70	7.86	0.50
$[148.0, 245.0)$	20.07	56.63	29.78	29.40	48.20
$[245.0, 403.0)$	39.05	67.41	71.46	14.10	4.17
$[403.0, 735.0]$	62.16	80.95	61.10	24.76	24.55

Table 2: Welch's t-tests comparing each confounding covariate across all complexity intervals.

formation was applied, and evenly spaced intervals were created in the transformed space before mapping them back to the original complexity space.[2]

In Figure 1 we show the TTD values from our dataset standardized using z-scores, along with standardized values for each of the student variables show in Table 1. This places all of the variables on the same scale, and allows us to clearly see how two of the variables (high school GPA and male), vary directly with TTD, and how the other three variables, (first generation, pell award, and Hispanic) vary inversely with TTD.

In order to validate that these differences are significant, we performed t-tests on whether the mean values of the covariates differ between the six curricular complexity groups. Specifically, Welch's t-tests were performed over each complexity interval, comparing the mean value of each of the covariates in one interval, to the mean value of the same covariate across all other intervals combined. The results are shown in Table 2 In every case, except one, these tests indicate we should reject the hypothesis that the mean values are the same for any of the covariates in one interval, as compared to all other intervals combined. Only for the Pell Award covariate in the

$[90, 148)$ complexity interval did the t-test fail to reject the hypothesis that the means are the same. Thus, these tests indicate that the covariates are significantly imbalanced across the complexity intervals.

Next, we consider how to model the relationships among the factors confounding TTD using causal models.

Causal Models

A common approach used to specify causal models involves utilizing directed graphs (digraphs). Directed edges in these digraphs are used to indicate directions of causality. For instance, Figure 2 (a) shows a two vertex digraph with curricular complexity (CC) pointing to TTD. This should be read as "curricular complexity is a cause of TTD," and the data presented in the prior section indicates that at the very least these two quantities are correlated.[1] This graph, however, is making the stronger case that the relationship is causal. Figure 2 (b) incorporates high school GPA, showing that it is a common cause of curricular complexity (i.e., students select complexity according to their high school GPA) and TTD; that is, it confounds the relationship between curricular complexity and TTD. In causal graph parlance, this is called a *fork*, which can be represented using the notation $X \leftarrow Z \rightarrow Y$, i.e., Z is a common cause of X and Y. Forks can produce spurious correlation, which can be detected in a sufficiently rich data. Specifically, if you condition on (i.e., control for) Z, and the correlation between X and Y disappears, i.e, $(X \perp Y|Z)$, then any correlation between X and Y is spurious. When we conduct this test using our data, it shows that the relationship between curricular complexity and TTD is indeed causal.

It is important to recognize that students do not actually pick a curricular complexity value, rather they pick a major, and that major at a particular institution has a particular curricular complexity, as depicted in Figure 2 (c). The path from major through curricular complexity to TTD in this figure is referred to as a *chain*, and in this chain, curricular complexity is said to *mediate* the causal impact of a major choice on TTD. As a part of this same study, we previously characterized the variability of curricular complexity according to particular disciplines, and in particular by CIP codes.[2] This work can be used to characterize the causal link between a student's major and the curricular complexity of that particular major. In this paper we focus more directly on the impact of curricular complexity on TTD. In the future we plan to more fully integrate these two as a part of a more detailed causal model. In particular, Figure 2 (d) provides a rough sketch of such a model. In this model we account for various student factors confounding the relationship between a student's choice of major and their TTD, and various institutional factors impacting both the complexity of the curricula, and TTD.

Simple Causal Model Analysis

One way to control for the impact of high school GPA when studying the relationship between curricular complexity and TTD is to include all of these variables in the same regression model. Figure 3 shows the results obtained using logistic regression models aimed at better understanding

[1] A corollary to the well known maxim that "correlation does not imply causation" is that "causation always produces correlation."

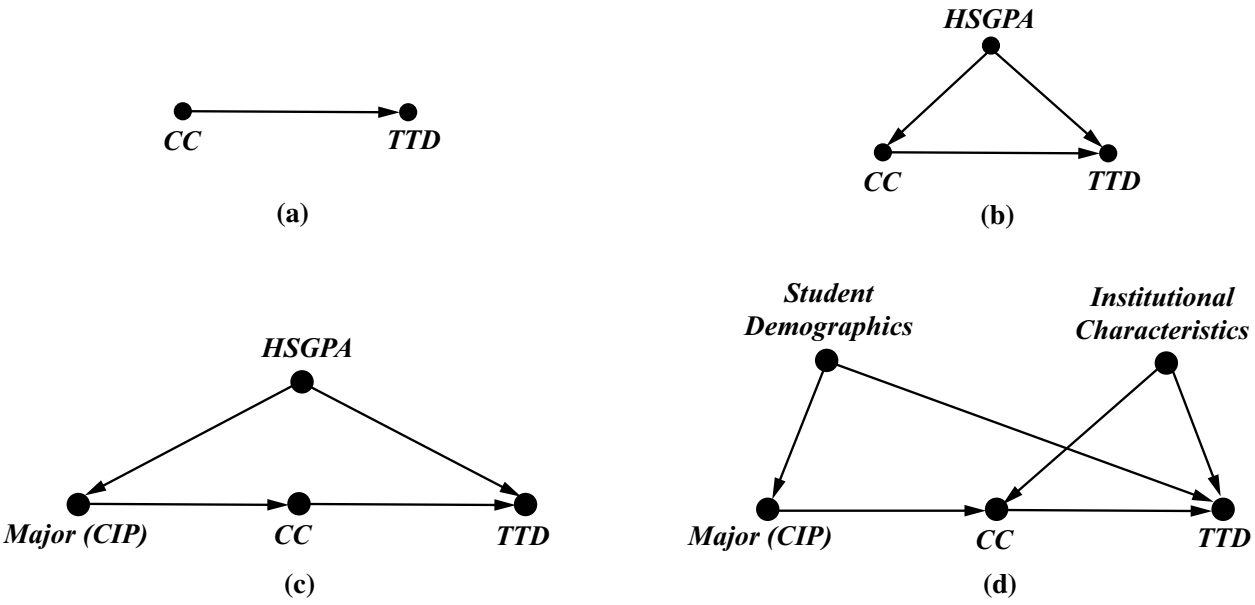

Figure 2: Causal models relating (a) curricular complexity and TTD, (b) high school GPA as a confounder between curricular complexity and TTD, (c) curricular complexity as a mediator between the choice of a major and TTD, and (d) the set of student and institutional factors influencing the relationship between curricular complexity and TTD.

the impact curricular complexity has on TTD, when controlling for high school GPA; that is, the causal model in Figure 2 (b). In this figure, the unadjusted model (without the high school GPA variable) corresponds to the thickest curve, and the other curves relate to the model that includes high school GPA, with various set values for high school GPA. Notice that for all of the high school GPA z-scores, curricular complexity has a larger impact (i.e., more negative slope), relative to the unadjusted model, as program curricular complexity grows. This approach to controlling for various confounders becomes unmanageable as more confounders are added. In the next section we will address this issue.

Generalized Propensity Score Analysis

In order to control for additional student demographic factors, we utilized a generalized propensity score (GPS) methodology that involves estimating the probability of specific treatment levels conditioned on these confounding variables.[4] These GPS values were used to construct a causal model that allowed us to control for the effects of the confounding variables. From this we are able to construct a dose-response function that characterizes the causal impact of curricular complexity on TTD. By controlling for factors such as a student's high school GPA, we found the impact of curricular complexity on TTD is more profoundly negative. Specifically, the effect of curricular complexity on TTD is more pronounced (and statistically significant) when accounting for high school GPA, first generation status, Pell award status, and ethnicity. The methodology we used is described next.

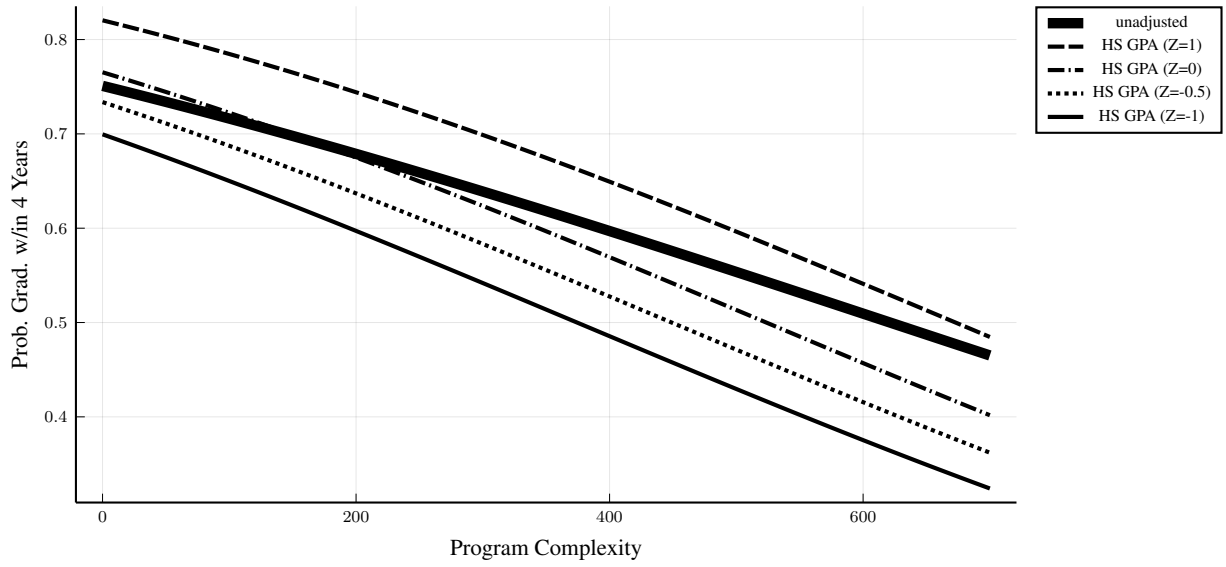

Figure 3: Logistic regression model showing the relationship between curricular complexity and graduating within four years, when controlling for high school GPA, where the high school GPA data was normalized using z-scores.

Let $i = 1, \ldots n$ denote the set of all students in the backwards dataset. The TTD outcome for student i is denoted $Y_i(T_i)$, where $T_i \in \mathcal{T}$, is the curricular complexity of the program student i completed. This is the student-level dose-response function, and we are interested in determining the average dose-response function, $\mu(t) = E[Y_i(t)],\ t \in \mathcal{T}$. For each student, there is a vector of covariates, X_i, that potentially impact both the treatment, curricular complexity, and the outcome, TTD. For ease of exposition, we will drop the i subscript when the meaning is clear.

A key assumption we will make regarding the covariates is that they can be used to unconfound the relationship between the curricular complexity and TTD. Specifically, the *weak unconfoundedness assumption* is that the outcome becomes independent of the treatment, when the treatment is conditioned on the pretreatment variables (the covariates); that is,

$$Y(t) \perp T|X,\ \forall t \in \mathcal{T}.$$

The term "weak" is used in this definition because joint independence between all potential outcomes is not required.[4] The *generalized propensity score (GPS)*, denoted $r(t, x)$, is then defined as the conditional density of the treatment given the covariates,[5]

$$r(t, x) = f_{T|X}(t|x).$$

The GPS serves to balance the treatment levels across various pretreatment variables. Specifically, within strata of the data having the same GPS values, the probability of a particular treatment level is independent of the covariates; that is,

$$X \perp 1\{T = t\}|r(t, X), \tag{1}$$

where $1\{\cdot\}$ is the indicator variable. The net result is the GPS can be used to eliminate any biases caused by the covariates when constructing the dose-response function. This is accomplished

in two steps. First, we use the student data to construct a model that estimates the conditional expectation of the outcome as a function of the treatment and the GPS,

$$\beta(t, r) = E[Y_i|T_i, r(T_i, X_i)].$$

Then, we use this model to estimate the dose-response function at a particular treatment level t, by averaging the conditional expectation over the GPS at that treatment level to obtain,

$$\mu(t) = E[\beta(t, r(t, X)]$$

Performing the last step over all $t \in \mathcal{T}$ yields a dose-response curve.

Implementation. In order to estimate the GPS values, $r(t, X)$, we first regress the treatment on the confounders (i.e., $T_i \sim X_i, i = 1, \ldots n$). Using the resulting model we obtain $\hat{T}_i$, the estimated treatment given the covariates for each student, along with $\hat{\sigma}$, the standard deviation of these estimates. Next, we assume

$$T = X'\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2) \tag{2}$$

That is, the treatment received by each student can be expressed as a linear model, with a normally distributed residual. In this case, the GPS values can be approximated by first estimating the residuals using $\hat{\epsilon}_i = T_i - \hat{T}_i$, and then estimating the GPS values using

$$\hat{r}(T_i, X_i) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{\hat{\epsilon}_i^2}{2\hat{\sigma}^2}\right)$$

To better align with the assumption specified in Equation (2), we first performed a logarithmic transformation of the treatment, curricular complexity, prior to fitting it with a generalized linear model (GLM) . The logarithmic transformation will yield a treatment distribution that is approximately normal.[2] We fit this data using a GLM with a Gamma distribution family, and an inverse link function. Figure 4 shows a histogram of the residual values obtained using the resulting model. A normal distribution fit to the residuals is also provided in Figure 4, demonstrating strong alignment with the assumption in Equation (2).

The GPS values generated from this approach were then applied using the inverse probability of treatment weighting (IPTW) method.[6] The basic idea in the IPTW approach involves reducing the effect of confounding through the creation of a pseudo-population obtained by weighting each student in the data set according to the inverse probability of receiving the treatment they actually received, i.e,

$$w_i = \frac{1}{\hat{r}(X_i, T_i)} \approx \frac{1}{f(T_i|X_i)}.$$

In practice, in order to avoid extreme weight values, stabilized weights are used, defined by

$$w_i^s = \frac{f(T_i)}{f(T_i|X_i)}.$$

where $f(T|X)$ is approximated as before, using the GPS values, and $f(T)$ is obtained by fitting a normal distribution to $\log T_i, i = 1, \ldots, n$.

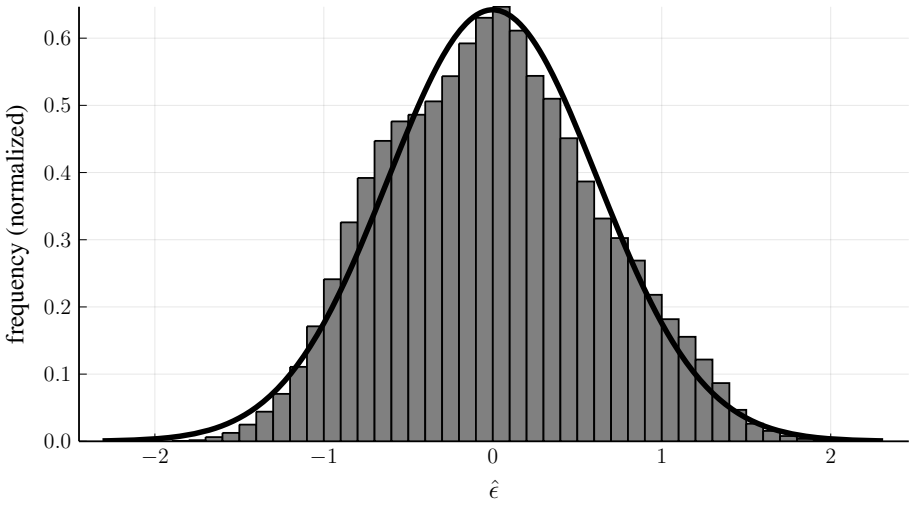

Figure 4: A histogram of the residual values $\hat{\epsilon} = \log T_i - \widehat{\log T_i}$, along with a normal distribution fit to the histogram, with $\mu = 0$ and $\sigma = 0.62$.

.

Complexity range	HS GPA	First Gen.	Pell	Male	Hispanic
[33.1 .. 54.6)	(34.59, 5.9)	(39.56, 10.81)	(47.56, 13.6)	(41.36, 5.76)	(44.8, 13.72)
[54.6 .. 90.0)	(39.22, 0.78)	(14.11, 3.0)	(22.87, 0.55)	(66.33, 1.94)	(18.83, 0.81)
[90.0 .. 148.0)	(17.21, 5.92)	(7.83, 0.68)	(0.41, 3.28)	(58.45, 13.94)	(6.77, 6.68)
[148.0 .. 245.0)	(20.07, 3.95)	(29.27, 5.23)	(48.12, 13.57)	(56.62, 7.39)	(33.32, 7.26)
[245.0 .. 403.0)	(39.11, 7.27)	(14.01, 5.08)	(4.33, 14.28)	(67.47, 4.38)	(9.16, 7.56)
[403.0 .. 735.0]	(62.2, 4.58)	(24.72, 0.7)	(24.55, 0.45)	(81.06, 6.8)	(12.51, 5.06)

Table 3: The t-statistics for assessing the mean difference between the covariate values within each complexity range, as compared to other five ranges. For each pair of numbers, the first value is for the original data set (i.e., repeated from Table 1, and the second value is for GPS-adjusted data values.

In order to test how well these GPS values balance the data set, we test the extent to which the covariates become independent of the treatment, when conditioned on the GPS values, i.e., the condition in Equation (1). We estimate this using the methodology described by Hirano and Imbens.[4] Specifically, for each covariate, we test whether the GPS-adjusted mean value in one complexity range is different from the mean values in the other ranges combined. This is accomplished by discretizing the level of the treatment, and the GPS value. Specifically, we compute $r(m_j, X_i), j = 1, \ldots, 6$ and $i = 1, \ldots, n$, where m_j is the median complexity value in the j-th complexity range. We test the condition in Equation (1) by first blocking on the values of $r(m_1, X_i), i = 1, \ldots, n$, with blocks defined by the quintiles of these $r(m_1, X_i)$ values. For each quintile, we compute the mean value of each covariate within each complexity range. This was repeated for $r(m_2, X_i), \ldots, r(m_6, X_i)$ for each covariate. Next, each of the six different mean values are combined, weighted by the number of observations in each GPS range. Then, a t-statistic is computed under the assumption the GPS-adjusted means are the same within each complexity range. The results of these tests are provided in Table 3. Note the t-statistics in Table 3 are smaller

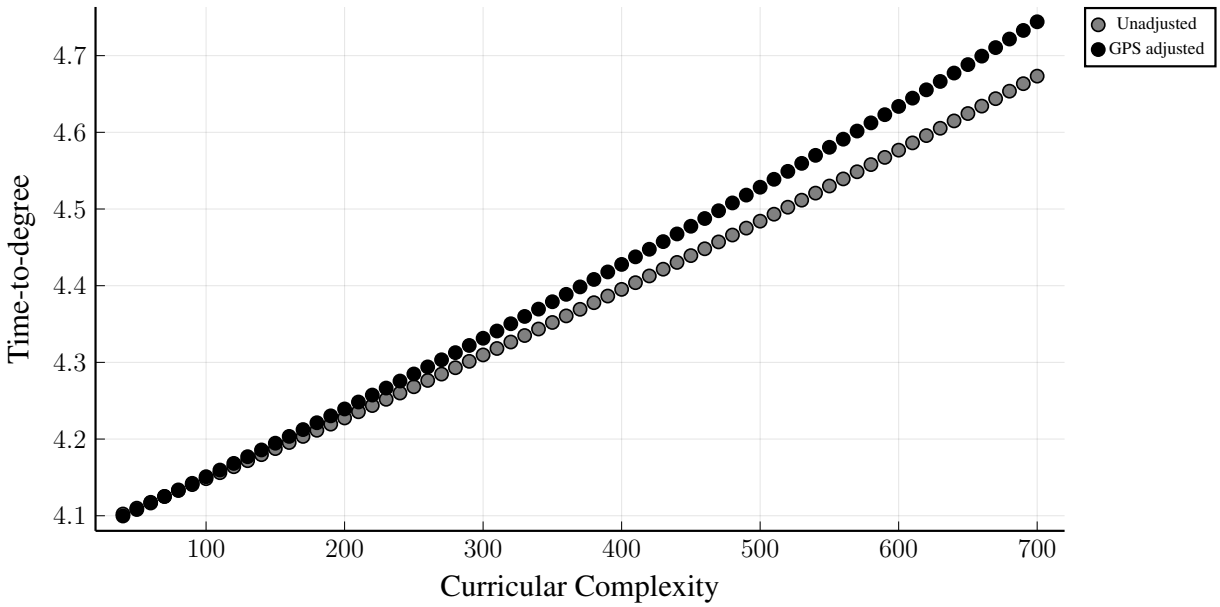

Figure 5: Unadjusted and GPS-adjusted TTD as a function of curricular complexity. The GPA-adjusted curve corresponds to the dose-response function for curricular complexity relative to the TTD outcome.

in each case except two, and in most cases significantly smaller, than the t-statistics provided in Table 2, indicating the GPS adjustments improve the balance of the data.

Finally, we regress the outcome on the weighted treatments values, i.e., $Y_i \sim w_i^s(T_i)$, to obtain an estimate of the impact of curricular complexity on TTD, controlling for the various confounding factors. Here, once again, we use a GLM with a Gamma distribution family, and an inverse link function. The results of this analysis are shown in Figure 5. The dose-response function, corresponding to the GPS-adjusted curve, shows the expected outcome if every student received the same dose of complexity. This figure also shows the impact of curricular complexity on TTD is not constant, rather it grows in proportion to curricular complexity. For the highest complexity programs, when accounting for various confounders, we see the increase in the average TTD is nearly 0.1 years as compared to the unadjusted model.

The GPS model we have constructed allows us to ask interesting counterfactual questions. For instance, what would happen to the 4-year graduation rates of the engineering programs in our study if they reduced their curricular complexity. Figure 6 shows that for every 50 point reduction in the curricular complexity of engineering programs, the 4-year graduation rate for each of the engineering programs in this study would increase by slightly more then two percentage points. Similarly, if these engineering students were to instead enroll in less complex non-engineering programs, the institutions in this study would realize 4-year graduation rate improvements in proportion to the number of engineering students at each institution..

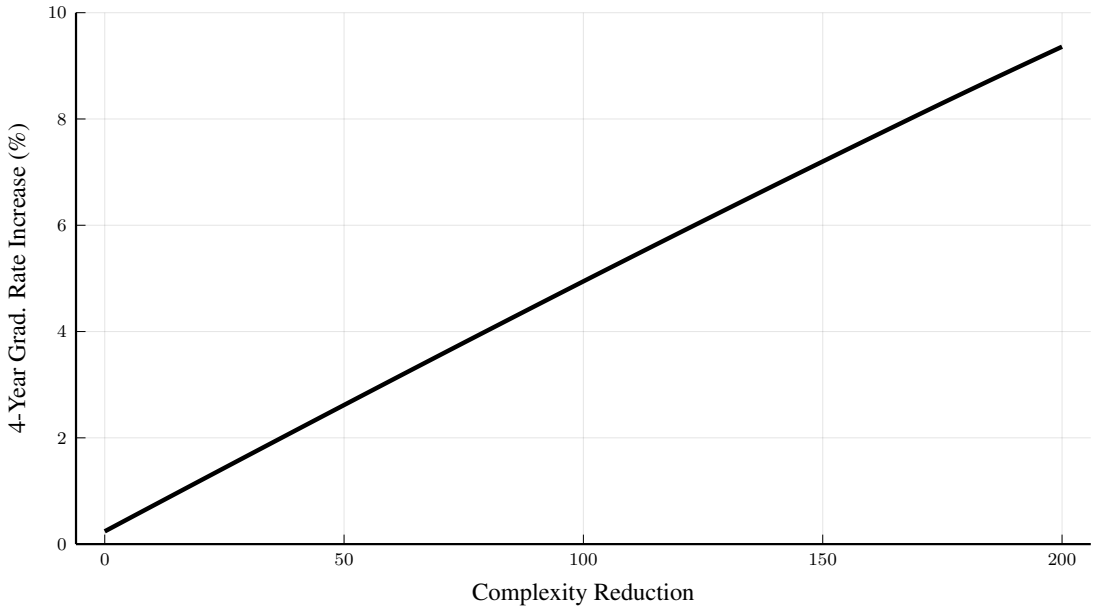

Figure 6: Percentage point improvement in 4-year graduation rates for the engineering programs our study, under various curricular complexity reduction scenarios.

Discussion

In this paper we investigated the impact of curricular complexity on TTD, in the presence of covariates that confound this relationship. We began by studying perhaps the most important covariate influencing the impact of curricular complexity on TTD, namely, high school GPA. High school GPA in this case is actually a proxy for assessing a student's prior preparation. We then utilized a methodology based on generalized propensity scores to account for other covariates. These GPS values were then used to construct a model that allowed us to control for the effects of these confounding variables. From this we are able to construct a dose-response function that characterizes the causal impact of curricular complexity on TDD. By controlling for factors such as a student's high school GPA, we found the impact of curricular complexity on TTD is more profoundly negative. Specifically, the effect of curricular complexity on TTD is more pronounced (and statistically significant) when accounting for high school GPA, first generation status, Pell award status, and ethnicity. Because students with higher high school GPAs tend to self select into higher complexity programs, while first generation and Pell award students tend to select away from higher complexity programs, the ability to accurately account for these in a causal model allows us to more accurately assess the true impact of curricular complexity on TTD. Another important aspect of this causal model is that it supports causal inferencing, allowing us to reason about counterfactual events; that is, events that did not actually occur. Our next step in this research will involve utilizing causal models for this purpose. For instance, the model described here allowed us to answer, "what would happen if we were able to reduce the complexity of all undergraduate degree programs in the College of Engineering by 10%?" Additional questions might include, "what would happen if we increased the number of women in our engineering program, the number of first generation students, etc." The usefulness of such a model in guiding student success efforts is

easy to see and will be the subject of subsequent research.

References

[1] A. Bhide, P. S. Shah, and G. Acharya. A simplified guide to randomized controlled trials. *Acta Obstet Gynecol Scand.*, 97(4):380–387, 2018.

[2] G. L. Heileman, C. T. Abdallah, K. A. Manasil, M. Akbarsharifi, R. Akbarsharifi, and A. Slim. Are engineering degrees really more complex? characterizing the complexities of academic programs by discipline. In *2025 ASEE Annual Conference & Exposition*, Montreal, Quebec, Canada, June 2025. ASEE Conferences. peer.asee.org/55453.

[3] G. L. Heileman, C. T. Abdallah, A. Slim, and M. Hickman. Curricular analytics: A framework for quantifying the impact of curricular reforms and pedagogical innovations. *www.arXiv.org*, arXiv:1811.09676 [cs.CY], June 2018. arxiv.org/abs/1811.09676.

[4] K. Hirano and G. W. Imbens. The propensity score with continuous treatments. In *Applied Bayesian Modeling and Causal Inference from Imcomplete-Data Perspectives*. John Wiley & Sons, 2004.

[5] G. W. Imbens. The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710, 2000.

[6] P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.