

Determinants of College ROI

Matt Martinez

September 10, 2024

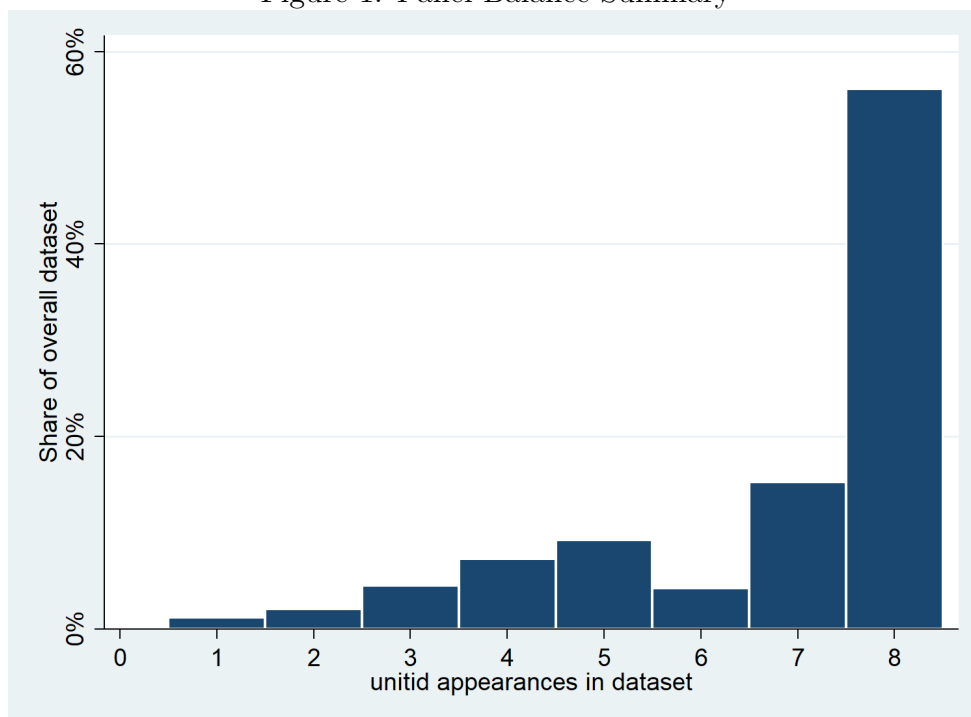
This memo summarizes the analysis performed on the College ROI dataset created from the College Scorecard data. The memo has three main sections:

1. **Pre-Analysis:** This section explores the raw data and summarizes the plan for assessing the determinants of high post-enrollment earnings/NPVs.
2. **Analysis:** This section is primarily concerned with correlation and regression analyses, with additional postestimation diagnostic testing for the regressions.
3. **Appendix:** This section includes discussions that are potentially useful for future analysis.

1 Pre-Analysis

The dataset consists of 8 unique files spanning the 2009/10 to the 2020/21 calendar years, plus a composite “Most Recent” file that fills in gaps in the most recent annual data file with the most recent prior year’s information. The panel is not perfectly balanced: just under 60% of `unitid`’s appear all 8 times in the data (excluding the “Most Recent” file):

Figure 1: Panel Balance Summary



1.1 Institution Types of Interest

Institutions can be classified in many different ways, but we will primarily focus on three categorical variables:

- **control**: control of institution (i.e., public, private non-profit, or private for-profit). 34% of the observations are for private for-profit institutions, 27% are for private nonprofit institutions, and 39% are for public institutions.
- **preddeg**: predominant undergraduate degree awarded (Associate’s, Bachelor’s, or Certificate). In the most recent data, 21% of the observations are for Associate’s-level institutions, 38% are for Bachelor’s-level institutions, and the remaining 41% are Certificate-level institutions.
- **iclevel**: level of institution (i.e., less than two-year, two-year, or four-year). In the most recent data, 30% of institutions are primarily two-year institutions, 46% are primarily four-year, and the remaining 24% are primarily less-than-two-year institutions.

These variables are *mostly* time-invariant. The variables **control**, **preddeg**, and **iclevel** are all fully populated, but have between 1% and 11% of observations that change (usually only once) over the span of the data.

The **preddeg** and **iclevel** variables are closely related. Only institutions classified as “Certificate” by the **preddeg** variable have an **iclevel** value of “less than 2-year,” across all years of data. Looking at the most recent data,¹ primarily Bachelor’s degree-granting institutions are exclusively four-year institutions, and they make up about 82% of the institutions classified as four-year. The only **iclevel** category that cuts across multiple **preddeg** values is the “2-year”, which is split roughly evenly between the primarily Associate’s-granting and Certificate-granting institutions. Around 70% of Associate’s-granting institutions are 2-year. The following analysis will focus on the **preddeg** variable instead of **iclevel**.

1.2 Dependent Variables of Interest

The dataset includes multiple post-enrollment earnings as well as calculated NPV measures by time since enrollment. However, due to the nature of the NPV calculations,² the 10-year earnings (variable name **p10**) correlates strongly with the NPV measures (variable name format **npv##**).

Table 1: ROI Variables Correlations

	p10	npv10	npv15	npv20	npv30	npv40
p10	1.00					
npv10	0.32	1.00				
npv15	0.81	0.80	1.00			
npv20	0.92	0.63	0.97	1.00		
npv30	0.96	0.49	0.92	0.98	1.00	
npv40	0.97	0.43	0.89	0.97	1.00	1.00

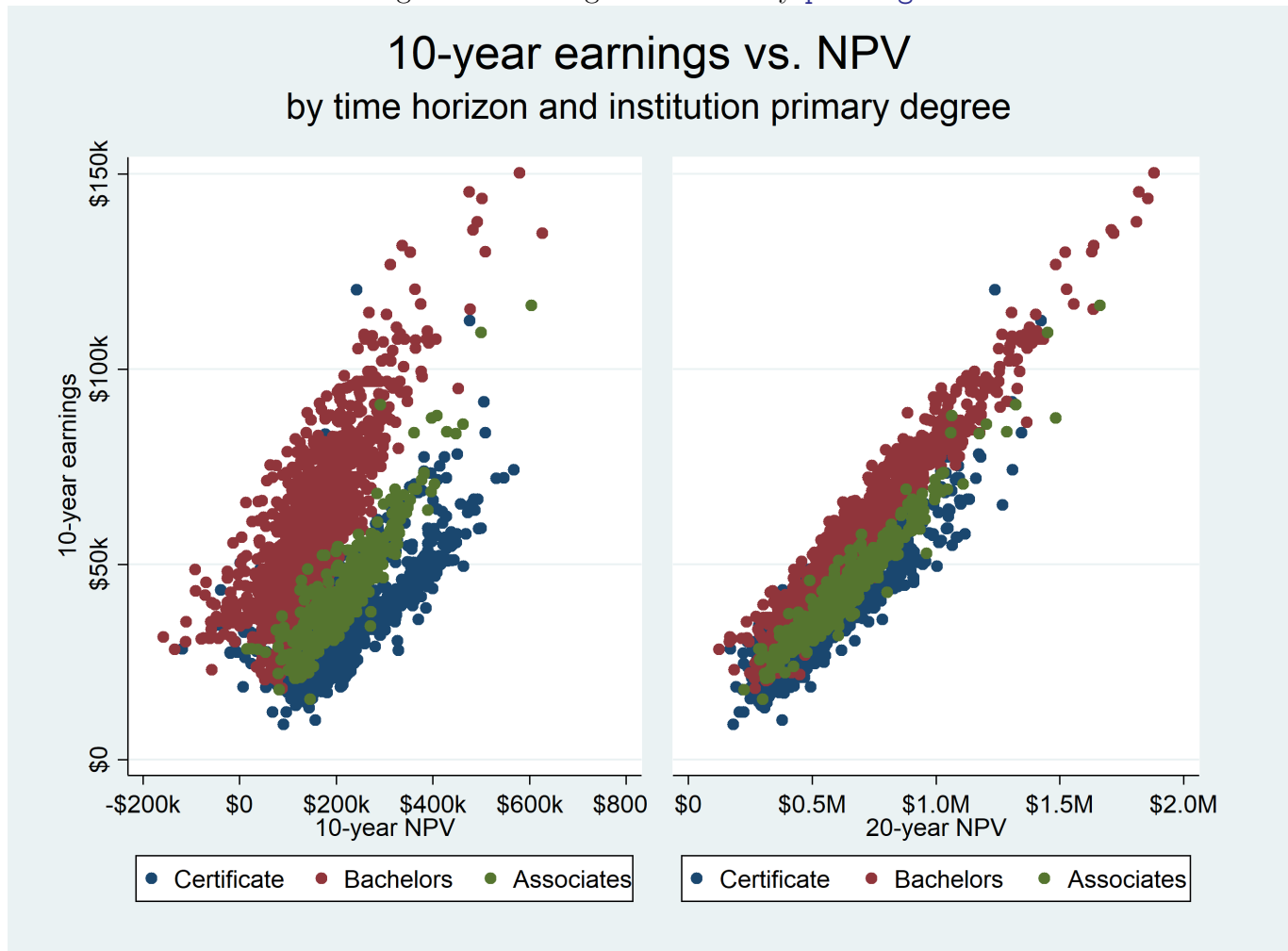
Most recent file only.

¹The following trends broadly hold across all years of data.

²We extrapolated 10- or 11-year post-enrollment earnings for all future cash flows.

Note that the later NPV calculations (`npv15` and higher) all have high degrees of correlation with each other and with 10-year earnings, but `npv10` has particularly low correlations with both 10-year earnings and the other NPV variables. This is due to underlying heterogeneity in institution program lengths: enrollees in Bachelor's programs will not have positive earnings recorded as part of their NPV calculations until six years after enrollment,³ whereas enrollees in Associate's programs will have positive earnings recorded by the fourth year after enrollment, and most⁴ enrollees in Certificate programs will have positive earnings recorded by the second year after enrollment. This heterogeneity is what drives the low correlations between `npv10` and all other dependent variables. Consider Figure 2 below, which plots `p10` against `npv10` and `npv20`, split by `preddeg`:

Figure 2: Earnings and NPVs by `preddeg`



From the left panel of Figure 2, it is clear that the correlation between `p10` and `npv10` by *institution type* are much higher than the overall correlation between `p10` and `npv10`.

NPV measures past 10 years are effectively collinear with `p10` even without controlling for `preddeg`.

³We assume students take five years to complete the degree. These students will also have *negative* cash flows in terms of `netprice` for the first five years after enrollment.

⁴We use the `iclevel` variable to categorize Certificate programs by length of program, but the majority of programs are less than two-year programs.

Even with the high degree of correlation between [p10](#) and the NPV variables, our analysis will include three outcome measures:

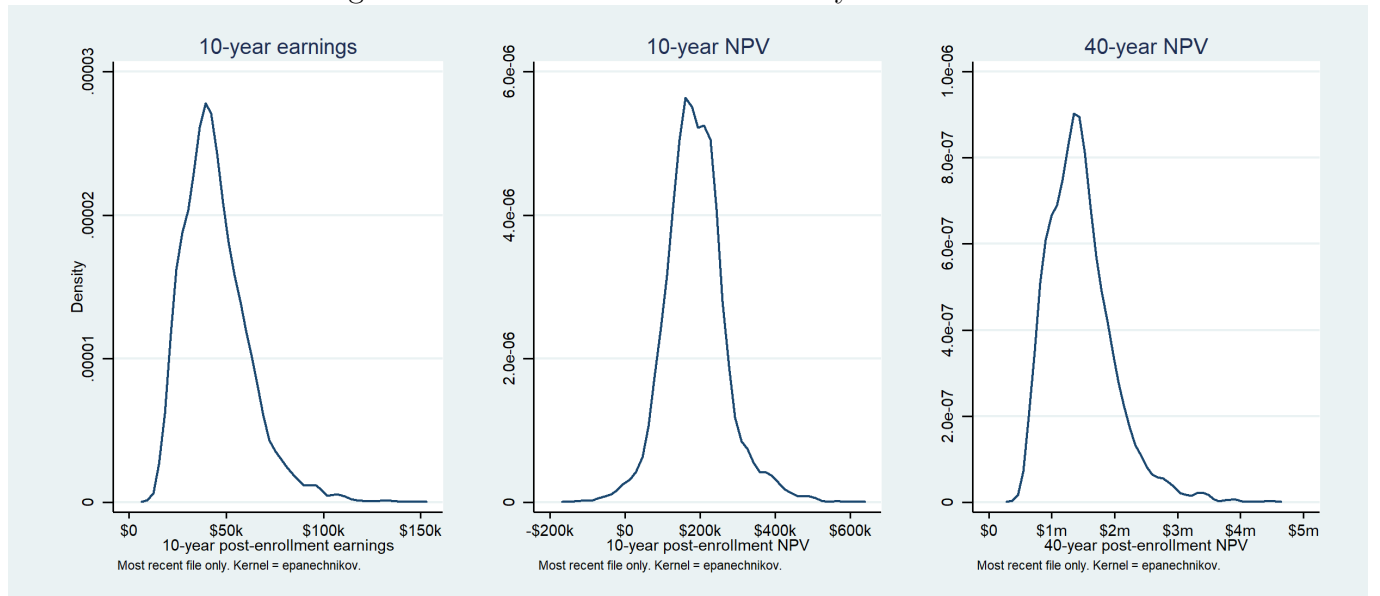
- 10-year earnings ([p10](#)): Since this measure is independent of cost, it allows us to check the impact of costs on earnings directly.
- 10-year NPV ([npv10](#)): This measure incorporates cost and is the least correlated with [p10](#), even when looking within [preddeg](#) values.
- 40-year NPV ([npv40](#)): This measure is often used in reporting. Although it is effectively collinear with 10-year earnings, it may be worth having these results handy. If necessary, it will be easy to change this to any of the other later NPV variables (e.g., [npv20](#)).

Since our data is a cross-sectional time series, we may want to consider how ROI changes over time as well as across institutions. However, Table 2 below shows that within-institution variation is significantly smaller than between-institution variation:

		Mean	Std. Dev.	N/n/T-bar
p10	overall	42,684	15,405	38,494
	between	.	15,088	6,668
	within	.	3,086	6
npv10	overall	170,686	76,270	38,494
	between	.	73,851	6,668
	within	.	23,828	6
npv40	overall	1,462,040	495,345	38,494
	between	.	483,092	6,668
	within	.	113,803	6

Finally, note that our ROI variables have strong right tails (which is to be expected):

Figure 3: ROI Variable Kernel Density Distributions



1.3 Explanatory Variables of Interest

As a starting point, there are 7 key variables we will investigate:

- `adm_rate`: admission rate.⁵
- `stem_pct`: percent of degrees awarded in STEM fields.
- `ugds`: undergraduate student enrollment.
- `grad_rate`: completion rate for first-time, full-time students (150% of expected time).
- `pctpell`: proportion of undergraduates receiving Pell grants.
- `pctfloan`: proportion of undergraduates receiving federal loans.
- `netprice`: average net price for institution's largest program.⁶

The `adm_rate` variable, which has a $[0, 1]$ range, has a significant portion of observations with a corner value; around 40% of institutions covered by the data are classified as “open admission,” (i.e., have a value of 1). These institutions are primarily Certificate- and Associate's degree-granting institutions.⁷ In contrast, 87% of Bachelor's degree-granting institutions are not open admission. Additionally, 84% of private for-profit institutions are open admissions, while 66% of public institutions are open admission. Conversely, only 17% of private nonprofit institutions are open admission.

Similarly, the `stem_pct` variable, which also has a $[0, 1]$ range, has around 40% of observations with no STEM degrees awarded. 76% of these institutions are Certificate-granting institutions. Furthermore, 77% of the institutions with no STEM degrees awarded are private, for-profit institutions.

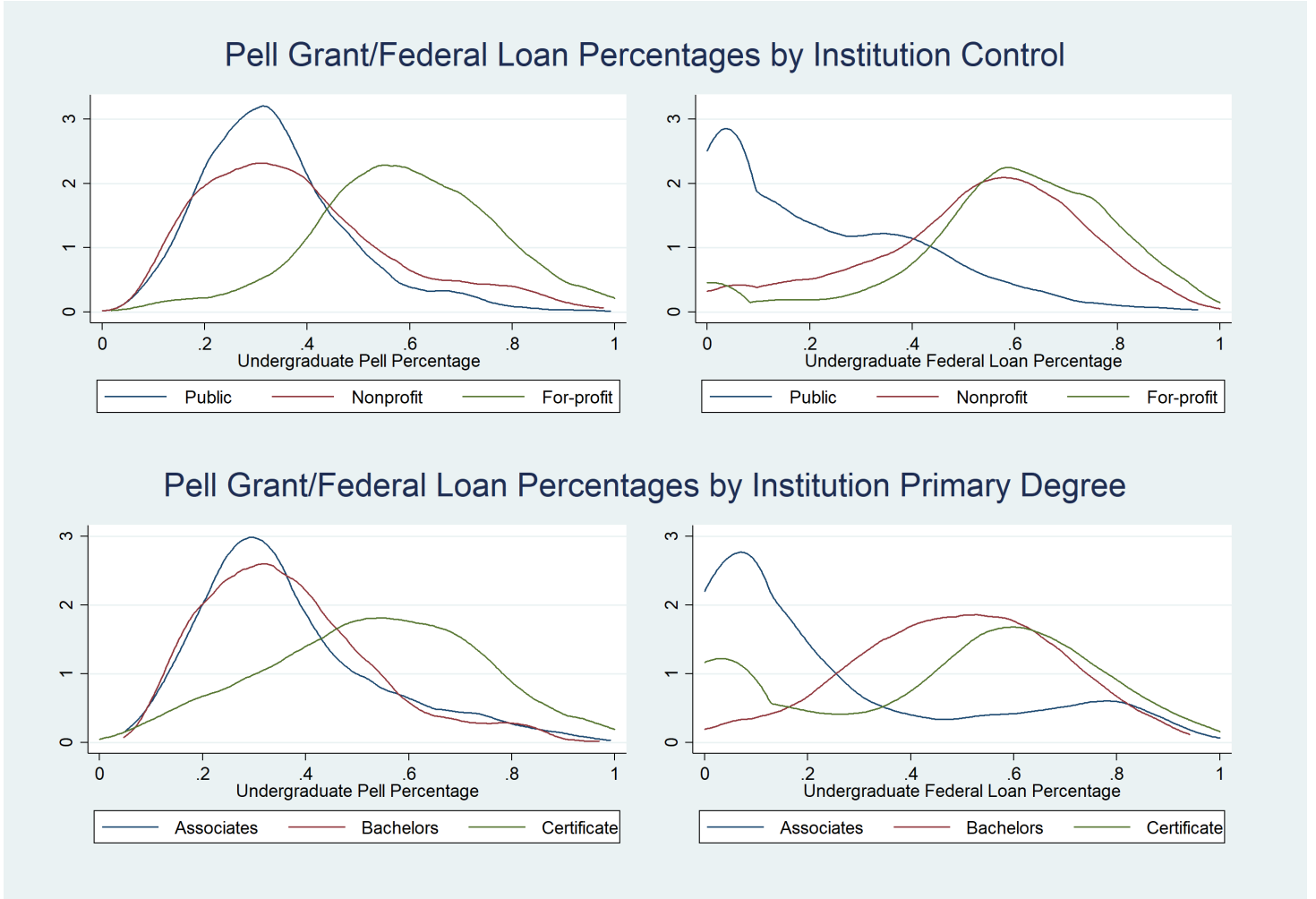
The `pctpell` and `pctfloan` variables differ significantly by institution type. Figure 4 below shows the kernel density estimates for these variables by `control` and `preddeg`. For-profit institutions and primarily Certificate-granting institutions have much higher percentages of Pell Grant recipients than nonprofit/public institutions and primarily Associate's- or Bachelor's-granting institutions. Public institutions and primarily Associate's-granting institutions have noticeably lower federal loan percentages than other institution types.

⁵Note that the raw data has missing values for `adm_rate` about 60% of the time. However, the `openadmp` variable shows that nearly all of these institutions are open admission.

⁶Note that this variable can only be incorporated into analysis of the earnings variables (not the NPV variables) since it is a component of NPV.

⁷81% of Associate's programs and 92% of Certificate programs in the data are open admission.

Figure 4: Grant/Loan Kernel Density Distributions by Institution Types



For all of the rate- or percent-based variables, we considered changes *across* institutions as well as changes *within* institutions over time. That is, we looked at the spread of values across institutions over the entire dataset as well as the spread of each institution’s values over time. We compared the resulting “between” standard deviations to the “within” standard deviations to see if differences were pronounced across institutions or within given institutions over time.

Table 3 below shows the “within” standard deviation is an order of magnitude smaller than the “between” variation. This means that over the roughly 12-year period covered by our data, institutions (not unexpectedly) maintain many of the same characteristics over time.

Table 3: Rate Variable Summary Statistics

		Mean	Std. Dev.	N/n/T-bar
adm_rate	overall	1.402	0.490	38,493.000
	between	.	0.456	6,668.000
	within	.	0.140	5.773
grad_rate	overall	0.497	0.230	37,102.000
	between	.	0.219	6,440.000
	within	.	0.082	5.761
stem_pct	overall	0.130	0.173	38,494.000
	between	.	0.170	6,668.000
	within	.	0.050	5.773
pctpell	overall	0.494	0.218	38,494.000
	between	.	0.206	6,668.000
	within	.	0.080	5.773
pctfloan	overall	0.512	0.281	38,493.000
	between	.	0.269	6,668.000
	within	.	0.086	5.773

1.4 Econometric Approach

The first task of the analysis will be to establish correlations between the independent variables and ROI generally. One of the key questions in this analysis is what observable institution characteristics correlate with ROI. Simple correlation matrices offer a good starting point. In addition to the simple correlation analysis, we would like to take a more detailed approach. A regression framework offers a channel for us to explore trends more rigorously. Ideally, we would hope to establish a *causal* relationship between institution characteristics and ROI, but even if we are not able to do so, a regression analysis will always yield at least *linear projection parameters*, which effectively represent a zero-mean error correlation between ROI and the independent variables of interest.

Given the low within-institution variation of both the dependent and independent variables, we will begin analysis by ignoring the panel setup of our data and focus only on the “Most Recent” data produced by the College Scorecard. Although much less rich of a dataset, filtering to the “Most Recent” data may allow us to use OLS to estimate more straightforward results. We can perform postestimation testing to ensure that basic OLS assumptions hold and adjust accordingly.

Finally, in addition to analysis across all institutions, we will conduct separate analyses by [preddeg](#) (predominant undergraduate degree awarded), since Associate’s-granting institutions are fundamentally different than Bachelor’s-granting institutions, and both are fundamentally different than Certificate-granting institutions. From the variable summaries above, we have strong reason to suspect that correlations/effects may be mediated by [preddeg](#).

2 Analysis

2.1 Correlations

First, we consider the basic relationship between ROI measures and the independent variables of interest at the most aggregate level (across institution types):

Table 4: Aggregate-level ROI Correlations

	p10	npv10	npv40	adm_rate	stem_pct	ugds	grad_rate	pctpell	pctfloan	netprice
p10	1.00									
npv10	0.32	1.00								
npv40	0.97	0.43	1.00							
adm_rate	-0.57	0.07	-0.53	1.00						
stem_pct	0.53	-0.01	0.51	-0.43	1.00					
ugds	0.29	0.07	0.30	-0.17	0.27	1.00				
grad_rate	0.19	0.10	0.18	-0.24	0.06	-0.03	1.00			
pctpell	-0.51	-0.17	-0.51	0.26	-0.32	-0.25	0.01	1.00		
pctfloan	0.03	-0.21	-0.00	-0.07	-0.08	-0.21	0.24	0.46	1.00	
netprice	0.33	-0.28	0.28	-0.32	0.07	-0.11	0.34	0.02	0.59	1.00

Note: Table only uses Most Recent file.

Across institutions, there seems to be very little relationship between undergraduate student size ([ugds](#)), completion rates ([grad_rate](#)), and net price ([netprice](#)) and any of the three ROI variables. Admissions rate ([adm_rate](#)) and percent of degrees awarded in STEM fields ([stem_pct](#)) seem to correlate modestly with [p10](#), with the expected signs.⁸ Interestingly, [pctpell](#) has a moderate *negative* correlation with [p10](#) while [pctfloan](#) is almost entirely uncorrelated.

Generally speaking, all independent variables have extremely low correlations with [npv10](#). [pctfloan](#) is the only variable that has a *higher* correlation to [npv10](#) than the other outcome variables, but that correlation is only -0.21.

One of the most surprising results from Table 4 is the low correlation between graduation rates and all three ROI measures. We would expect to see a high degree of correlation between completion of postsecondary education and any kind of earnings measure. This result could signal that this correlation matrix is too aggregate to be helpful, and that further breakdown is necessary.

Table 5 below shows the same correlations but broken down by [preddeg](#) (and omitting the correlations between the independent variables for simplicity):

⁸That is, lower admissions rates are correlated with higher [p10](#) and higher percentages of STEM degrees awarded are correlated with higher [p10](#).

Table 5: Correlations by primary degree granted (**preddeg**)

Associate's				Bachelor's				Certificate			
	p10	npv10	npv40		p10	npv10	npv40		p10	npv10	npv40
p10	1.00			p10	1.00			p10	1.00		
npv10	0.88	1.00		npv10	0.80	1.00		npv10	0.66	1.00	
npv40	0.96	0.88	1.00	npv40	0.98	0.83	1.00	npv40	0.94	0.75	1.00
adm_rate	-0.27	-0.16	-0.23	adm_rate	-0.40	-0.25	-0.38	adm_rate	-0.29	-0.17	-0.27
stem_pct	0.05	0.11	0.06	stem_pct	0.41	0.40	0.43	stem_pct	0.17	0.06	0.16
ugds	0.07	0.12	0.09	ugds	0.19	0.27	0.21	ugds	0.19	0.01	0.17
grad_rate	0.20	0.06	0.16	grad_rate	0.60	0.37	0.59	grad_rate	-0.05	0.13	-0.02
pctpell	-0.29	-0.42	-0.33	pctpell	-0.57	-0.33	-0.56	pctpell	-0.32	-0.23	-0.32
pctfloan	0.19	-0.06	0.14	pctfloan	-0.11	-0.28	-0.13	pctfloan	-0.05	-0.06	-0.05
netprice	0.28	-0.11	0.22	netprice	0.35	-0.26	0.29	netprice	0.12	-0.09	0.09

Note: Table only uses Most Recent file.

Table 5 shows low correlations for all explanatory variables except when **preddeg** is “Bachelor’s.” Primarily Bachelor’s-granting institutions show moderately strong positive correlations between the graduation rate and 10-year earnings ($\rho = 0.60$) and **pctpell** and 10-year earnings ($\rho = -0.57$). Looking across institution types, Table 5 shows that **adm_rate** and **stem_pct** have the strongest correlation with 10-year earnings for Bachelor’s degree institutions. However, both **adm_rate** and **stem_pct** have *lower* correlation coefficients than in Table 4. We can further disaggregate institutions by the **control** variable. Tables 6, 7, and 8 below replicate Table 5, but filtered to public, nonprofit, and for-profit institutions, respectively:

Table 6: Correlations by primary degree granted (**preddeg**) in public institutions

Associate's				Bachelor's				Certificate			
	p10	npv10	npv40		p10	npv10	npv40		p10	npv10	npv40
p10	1.00			p10	1.00			p10	1.00		
npv10	0.92	1.00		npv10	0.90	1.00		npv10	0.69	1.00	
npv40	0.95	0.90	1.00	npv40	0.98	0.91	1.00	npv40	0.91	0.77	1.00
adm_rate	-0.15	-0.10	-0.16	adm_rate	-0.31	-0.30	-0.30	adm_rate	-0.20	-0.28	-0.21
stem_pct	0.10	0.09	0.12	stem_pct	0.51	0.45	0.51	stem_pct	-0.21	-0.34	-0.23
ugds	0.17	0.10	0.17	ugds	0.41	0.35	0.39	ugds	0.07	-0.24	0.02
grad_rate	0.17	0.21	0.17	grad_rate	0.61	0.49	0.60	grad_rate	0.26	0.56	0.29
pctpell	-0.46	-0.50	-0.46	pctpell	-0.47	-0.33	-0.46	pctpell	0.05	0.27	0.11
pctfloan	0.29	0.24	0.31	pctfloan	-0.10	-0.27	-0.11	pctfloan	0.47	0.59	0.53
netprice	0.34	0.10	0.32	netprice	0.38	-0.03	0.35	netprice	0.50	0.50	0.54

Note: Table only uses Most Recent file.

For associate’s-granting public institutions, Table 6 shows **pctpell** is moderately negatively correlated with all earnings measures. For bachelor’s-granting institutions, **grad_rate** has a nearly identical correlation, but **adm_rate** and **pctpell** are smaller by 0.1 while **stem_pct** is up by 0.1 (around $\rho = 0.5$). **pct_stem** ($\rho = 0.51$). For certificate-granting institutions, **pctfloan** and **netprice** both have a correlation of around $\rho = 0.5$ across all earnings measures.

Table 7: Correlations by primary degree granted (**preddeg**) in nonprofit institutions

Associate's				Bachelor's				Certificate			
	p10	npv10	npv40		p10	npv10	npv40		p10	npv10	npv40
p10	1.00			p10	1.00			p10	1.00		
npv10	0.94	1.00		npv10	0.81	1.00		npv10	0.70	1.00	
npv40	0.97	0.92	1.00	npv40	0.98	0.83	1.00	npv40	0.96	0.78	1.00
adm_rate	-0.51	-0.52	-0.50	adm_rate	-0.44	-0.31	-0.43	adm_rate	-0.39	-0.18	-0.38
stem_pct	0.08	0.03	0.09	stem_pct	0.44	0.38	0.45	stem_pct	-0.09	-0.17	-0.12
ugds	-0.05	-0.07	-0.05	ugds	0.18	0.14	0.19	ugds	0.04	-0.18	0.01
grad_rate	0.47	0.46	0.46	grad_rate	0.65	0.42	0.65	grad_rate	-0.08	-0.02	-0.04
pctpell	-0.50	-0.39	-0.50	pctpell	-0.65	-0.35	-0.64	pctpell	-0.49	-0.34	-0.52
pctfloan	0.29	0.24	0.28	pctfloan	-0.17	-0.23	-0.17	pctfloan	0.30	0.12	0.26
netprice	0.44	0.19	0.44	netprice	0.47	-0.11	0.42	netprice	0.20	-0.15	0.19

Note: Table only uses Most Recent file.

Table 7 shows moderate correlations between **adm_rate**, **grad_rate**, and **pct_pell** and the earnings measures for primarily associate's-granting nonprofit institutions (all around $\rho = 0.5$). For primarily bachelor's-granting nonprofit institutions, **grad_rate** correlates strongly with 10-year earnings ($\rho = 0.65$), while **pct_pell** has a strong negative correlation with 10-year earnings ($\rho = -0.65$). **adm_rate**, **stem_pct**, and **net_price** all have moderate correlations with 10-year earnings. For primarily certificate-granting nonprofit institutions, the only moderately strong correlation comes from **pct_pell**.

Table 8: Correlations by primary degree granted (**preddeg**) in for-profit institutions

Associate's				Bachelor's				Certificate			
	p10	npv10	npv40		p10	npv10	npv40		p10	npv10	npv40
p10	1.00			p10	1.00			p10	1.00		
npv10	0.89	1.00		npv10	0.90	1.00		npv10	0.56	1.00	
npv40	0.97	0.87	1.00	npv40	0.98	0.91	1.00	npv40	0.93	0.67	1.00
adm_rate	-0.05	-0.04	-0.00	adm_rate	-0.27	-0.21	-0.24	adm_rate	-0.28	-0.03	-0.23
stem_pct	0.03	0.11	-0.00	stem_pct	0.01	-0.01	0.06	stem_pct	0.17	0.09	0.17
ugds	0.12	0.14	0.15	ugds	-0.07	0.05	-0.03	ugds	0.15	-0.05	0.12
grad_rate	0.07	0.10	0.06	grad_rate	0.33	0.23	0.29	grad_rate	0.01	0.13	0.02
pctpell	-0.30	-0.22	-0.32	pctpell	-0.24	-0.09	-0.25	pctpell	-0.15	-0.13	-0.16
pctfloan	0.07	0.07	0.03	pctfloan	0.36	0.34	0.29	pctfloan	0.23	0.09	0.21
netprice	0.27	-0.07	0.24	netprice	0.27	-0.15	0.18	netprice	0.46	0.09	0.41

Note: Table only uses Most Recent file.

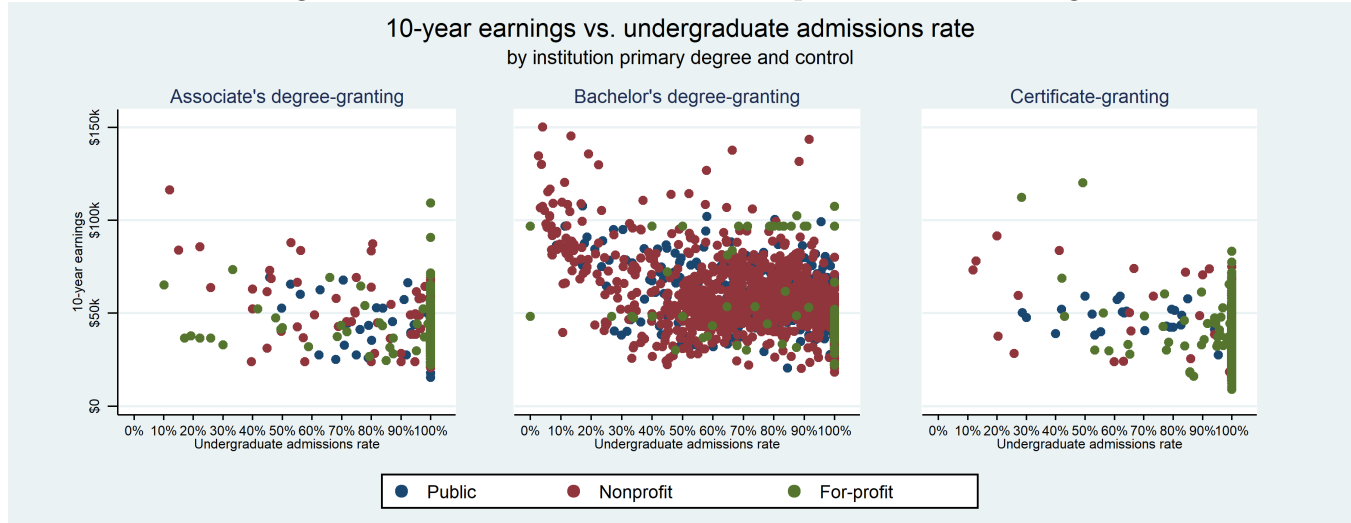
Table 8 shows no correlations that are even moderately strong, across all values of **preddeg**, except for **netprice** and 10-year earnings ($\rho = 0.46$) for primarily certificate-granting institutions.

2.2 Graphical Analysis

Although the tables above are helpful, graphical representations of key correlations are often easier to interpret quickly. We will focus on [p10](#) as the key outcome measure.

One surprising result from above is the lack of correlation between admissions rates ([adm.rate](#)) and 10-year earnings ([p10](#)):

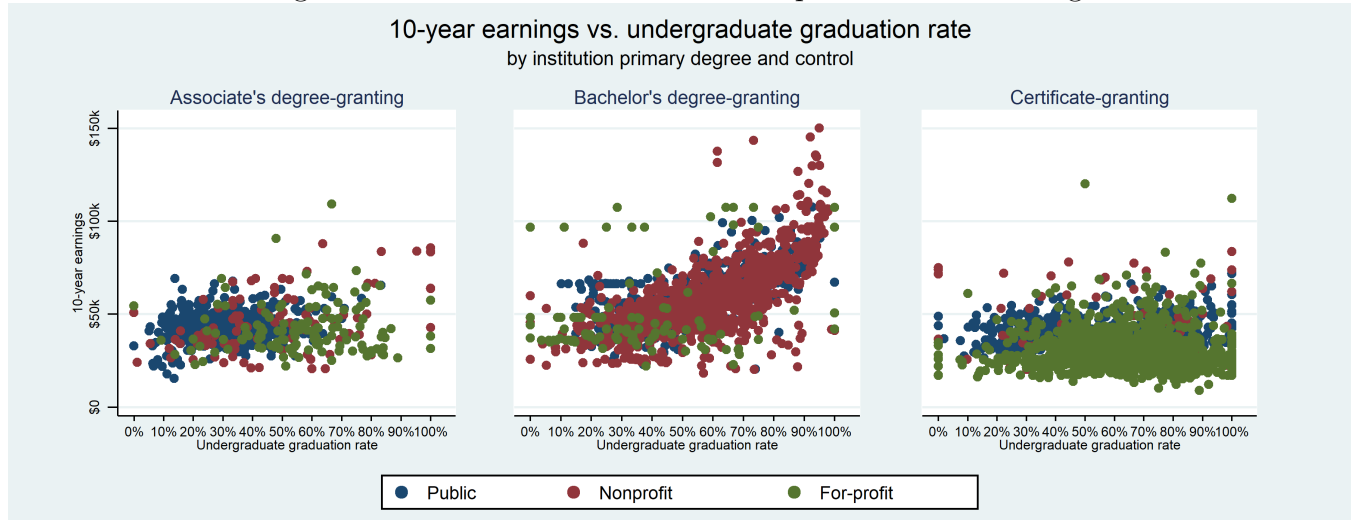
Figure 5: Admissions Rate Relationship to 10-Year Earnings



From the figure above, we can see that a near-majority of associate's-level and certificate-level institutions are open admissions, which explains the low correlations. For bachelor's-level institutions, there seems to be a stronger relationship between earnings and admissions rates for institutions with admissions rates below 50%.

Another interesting result from the tables above is that graduation rates only seem to correlate with 10-year earnings for bachelor's-level institutions and nonprofit associate's-level institutions. Again, a scatter plot allows further examination of the relationship.

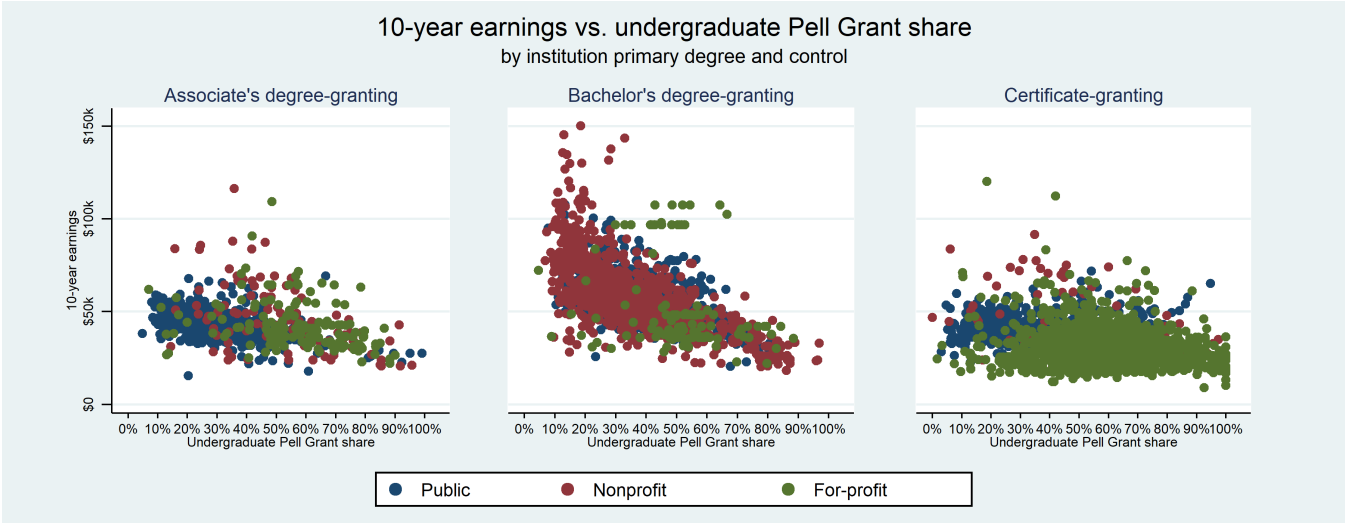
Figure 6: Graduation Rate Relationship to 10-Year Earnings



From Figure 6, we can see that many associate’s-level programs have a lower general range of graduation rates and the largest benefits of bachelor’s-level institution graduation rates come for institutions at the higher end of the range. Certificate-granting institutions do not appear to show any relationship between graduation rates and 10-year earnings, nor do for-profit institutions across all program types.

A third striking result from the tables above was the persistent, negative relationship between the percentage of Pell Grant recipients and 10-year earnings:

Figure 7: Pell Grant Share Relationship to 10-Year Earnings



It appears that the relationship between Pell Grant share and 10-year earnings is present in associate’s-level institutions but not nearly as strong as in bachelor’s-level institutions.

2.3 Regressions

Although we have some sense of correlations, we do not yet have a sense of the magnitude of the relationships between the ROI variables and our independent variables of interest. Table 9 below shows a simple OLS regression of the three key logged ROI variables (`ln_p10`, `ln_npv10`, `ln_npv40`) on the numeric independent variables, controlling for key institution characteristics. Note that the excluded category is public, primarily Bachelor's degree-granting institutions:

Table 9: Aggregate Regressions

	(1) ln_p10	(2) ln_npv10	(3) ln_npv40
adm_rate	-0.23*** (0.00)	-0.22*** (0.00)	-0.27*** (0.00)
stem_pct	0.30*** (0.00)	0.35*** (0.00)	0.31*** (0.00)
ugds	0.00*** (0.00)	0.00*** (0.01)	0.00*** (0.00)
grad_rate	0.15*** (0.00)	0.44*** (0.00)	0.20*** (0.00)
pctpell	-0.51*** (0.00)	-0.44*** (0.00)	-0.64*** (0.00)
pctfloan	0.30*** (0.00)	0.27*** (0.00)	0.42*** (0.00)
netprice	0.00*** (0.00)		
Associate's	-0.03** (0.01)	0.59*** (0.00)	0.01 (0.34)
Certificate	-0.14*** (0.00)	0.59*** (0.00)	-0.08*** (0.00)
Private for-profit	-0.35*** (0.00)	-0.38*** (0.00)	-0.26*** (0.00)
Private nonprofit	-0.14*** (0.00)	-0.35*** (0.00)	-0.10*** (0.00)
Constant	10.85*** (0.00)	11.90*** (0.00)	14.46*** (0.00)
Observations	4490	4441	4490
R^2	0.676	0.341	0.596

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9 shows that primarily Certificate-granting institutions, less-than 2-year institutions, and private for-profit institutions are negatively correlated with the logged ROI variables. For a more granular view, Table 10 below shows the same regressions but performed individually across the three values of `preddeg`:

Table 10: Aggregate Regressions

	Associate's			Bachelor's			Certificate		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ln_p10	ln_npv10	ln_npv40	ln_p10	ln_npv10	ln_npv40	ln_p10	ln_npv10	ln_npv40
adm_rate	-0.18** (0.08)	-0.20* (0.11)	-0.17** (0.09)	-0.12*** (0.03)	-0.04 (0.07)	-0.11*** (0.03)	-0.35*** (0.13)	-0.30 (0.24)	-0.42*** (0.13)
stem_pct	0.06 (0.06)	0.13* (0.07)	0.04 (0.06)	0.38*** (0.04)	0.58*** (0.10)	0.41*** (0.04)	0.30*** (0.05)	-0.01 (0.10)	0.27*** (0.06)
ugds	0.00*** (0.00)	0.00*** (0.00)	0.00*** (0.00)	0.00*** (0.00)	0.00 (0.00)	0.00*** (0.00)	0.00*** (0.00)	-0.00*** (0.00)	0.00*** (0.00)
grad_rate	0.11** (0.05)	0.17** (0.07)	0.10* (0.05)	0.37*** (0.04)	0.43*** (0.09)	0.38*** (0.04)	0.07** (0.03)	0.33*** (0.05)	0.08*** (0.03)
pctpell	-0.67*** (0.05)	-0.62*** (0.07)	-0.75*** (0.05)	-0.75*** (0.05)	-0.57*** (0.09)	-0.78*** (0.05)	-0.35*** (0.03)	-0.39*** (0.06)	-0.46*** (0.04)
pctfloan	0.40*** (0.05)	0.38*** (0.05)	0.45*** (0.04)	0.22*** (0.03)	-0.13* (0.07)	0.22*** (0.03)	0.32*** (0.03)	0.40*** (0.04)	0.50*** (0.03)
netprice	0.00** (0.00)			0.00 (0.00)			0.00*** (0.00)		
Private for-profit	-0.16*** (0.04)	-0.33*** (0.05)	-0.14*** (0.04)	-0.00 (0.04)	-0.43*** (0.10)	0.02 (0.04)	-0.46*** (0.02)	-0.46*** (0.03)	-0.38*** (0.02)
Private nonprofit	-0.08** (0.04)	-0.23*** (0.06)	-0.08** (0.04)	-0.05*** (0.01)	-0.30*** (0.03)	-0.05*** (0.01)	-0.22*** (0.04)	-0.54*** (0.10)	-0.17*** (0.04)
Public	0.00 (.)	0.00 (.)	0.00 (.)	0.00 (.)	0.00 (.)	0.00 (.)	0.00 (.)	0.00 (.)	0.00 (.)
Constant	10.88*** (0.09)	12.60*** (0.12)	14.45*** (0.09)	10.86*** (0.05)	11.91*** (0.10)	14.33*** (0.05)	10.82*** (0.13)	12.63*** (0.24)	14.55*** (0.13)
Observations	928	928	928	1699	1659	1699	1863	1854	1863
R^2	0.388	0.276	0.378	0.564	0.283	0.545	0.462	0.202	0.384

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3 Appendix

3.1 `preddeg` vs. `iclevel`

In order to compute the ROI measures, we assume that students are enrolled toward obtaining the institution's predominant degree (using the `preddeg` variable). If the institution is predominantly an associate's degree institution, then we assume that the student is enrolled for three years while if the predominant degree is a bachelor's degree then we assume that the student is enrolled for five years.

About 40 percent of institutions in the College Scorecard data are predominantly certificate institutions. These institutions can be further classified by the length of time required to earn the relevant certificate using the `iclevel` variable: less than two years, two years, or (rarely) four years. For these institutions, we assume that students are enrolled for only one year if the data suggests that the institution requires less than two years of study. If a predominantly certificate institution is classified as requiring two years of study, we assume the student follows the same path as a student enrolled in a primarily associate's degree institution (three years of enrollment). In the rare cases where a primarily certificate institution is classified as requiring four years of study, we assume the student follows the same path as a student enrolled in a primarily bachelor's degree institution (five years of enrollment).

Although the College Scorecard does classify predominantly associate's degree institutions as either two-year or four-year institutions, many associate's degree institutions are community colleges that also grant bachelor's degrees. This makes easy delineations based on the two-year or four-year categories difficult to interpret cleanly, so for simplicity we assume that all primarily associate's degree institutions enroll students for three years.

3.2 Panel Data Analysis

If we want to use the entire dataset, we will need to account for the fact that our data is a cross-sectional time series. We will ideally want to control for unobserved individual (i.e., institution-level) effects. However, we need to determine whether to use a **random effects** (RE) model or a **fixed effects** (FE) model. According to Oscar Torres-Reyna of Princeton:⁹

If you have reason to believe that differences across entities have some influence on your dependent variable but are not correlated with the predictors, then you should use random effects. An advantage of random effects is that you can include time invariant variables (i.e. gender). In the fixed effects model these variables are absorbed by the intercept. Random effects assume that the entity's error term is not correlated with the predictors which allows for time-invariant variables to play a role as explanatory variables... [However, w]henver there is a clear idea that individual characteristics of each entity or group affect the regressors, use fixed effects.

Before performing any analysis, it is difficult to predict with certainty whether RE or FE models are more appropriate. On one hand, we would like to check whether observable, time-invariant institution characteristics such as `control` and `preddeg` are important determinants of ROI. In this

⁹See pp. 40-41, 43 of "[Panel Data Analysis: Fixed and Random Effects using Stata.](#)"

case, we would like to use an RE model. However, we might think that there are unobservable institution-level characteristics, such as reputation, that might be correlated with several of our independent variables, such as admission rate, number of undergraduates, and completion rates.¹⁰ Thus, we could begin with RE models, but we would perform Hausman (DWH) specification tests on our models to see if we need to switch to FE models.

¹⁰For example, I heard someone offhandedly say at Princeton that the university invests heavily in advising/mentoring programs for students at risk of failing partially because keeping graduation rates high helps maintain the university's prestige as a premier institution.