What Impacts Tuition?
An Examination of Institutional Control and Tuition Pricing Difference

Abstract:
In this paper, we examine the tuition pricing difference between public and private not-for-profit institutions in the United States. We include three types of influential factors, educational quality, operations, and student characteristics, and conduct a best subset analysis to determine the model that has the highest adjusted R-squared value. Next, we use the extra sum of squares F-test to examine whether the inclusion of institutional control provides extra explanatory power to the model. Our result suggests that there is a significant tuition difference between public and private institutions.

Background and Significance

The surging tuition cost for higher education has attracted massive public attention. According to the 2013 College Board survey, the average tuition cost for a private, non-profit, four-year university was $31,231, 17 times higher than the cost in 1971-1972 ($1832) (Schoen, 2013). The annual rate of tuition increase was around 6 percent above the nation's inflation rate (Schoen, 2013). Tuition difference between private and public institutions are also significant. In 2015, average tuition for private, non-profit, four-year institutions exceeds out-of-state tuition for public, four-year institutions by $8512 (Bell, Baum, Ma & Pender, 2015). This gap is also expanding because of different tuition increase rates. From 2014 to 2015, annual average tuition increased by $1,122, or 3.6% for private, non-profit, four-year institutions, and $786, or 3.4% for public, four-year institutions (Bell, Baum, Ma & Pender, 2015).

Past researchers have provided a sound basis for the examination of tuition pricing. However, they mainly focused on the high increase rate of tuition, but provided few explanations for varying tuition costs among different institutions (Ehrenberg, 2000; Li, 2013). Researchers typically analyzed the composition of tuition using the supply-demand analysis, which included factors both from the supply side, e.g. operating cost of the school, and the demand side, e.g. student preference and disposable income (Yanikoski & Wilson, 1984). St. John (1992) presented a five-factor analysis of tuition pricing, which included prestige, operating budget, pricing of competitive institutions, disposable income and student aid. Throsby (1992) addressed the significant relationship between tuition and foreign student enrollment, that institutions may adopt price discrimination to secure additional sources of revenue.

To address pricing differences among U.S institutions of higher education, our study attempts to study the tuition gap between private and public institutions. In this study, we build a regression model for tuition using multiple variables that have potential impact on tuition. We summarize these tuition pricing factors into three categories: educational quality, operations and student characteristics. Next, we include the institutional control in the model. Our null hypothesis is that institutional control does not provide extra explanatory power to the model. Our alternative hypothesis is that institutional control provides significant extra explanatory power to the model.

Methodology and Analysis

Dataset Description:

The original dataset is from the Institute of Education Science (IES) 2013 Census. The dataset contains observations of 7518 institutions. We choose the 2013 Census because it provides better data coverage than the 2014 Census.

Variables Selection:

The IES dataset contains 26 quantitative variables. *Tuition* is the response variable. *Private*, a dummy variable in which 1 is assigned to private institutions, is the main variable of interest. Not all remaining variables are included in the model. Due to limited data, we exclude endowment assets and admission. For the student ethnicity data, we remove variables for Native American and Pacific Islander because these ethnic groups account for less than 3 percent of population in most institutions. The remaining ethnic variables are *Asian*, *Black*, *Hispanic* and *White*. After conducting a variance inflation indicator (VIF) check, we observe high VIFs among ethnic variables. Thus, we exclude *Hispanic* to control for multicollinearity. The final list of explanatory variables is included below. A detailed description of variables is included in Appendix 1.

Table 1. List of Explanatory Variables

Educational Quality	Operations	Student Characteristics
Graduation Rate *(Graduate.Rt)* Student-Faculty Ratio *(STuFau.Ratio)*	Average Professor Salary *(Average.Salary)* % of Students with Financial Aid *(Percent.FA)* Average Amount of Financial Aid *(Amount.FA)* # of Students Enrolled *(Enrollment)*	% of White Students *(White)* % of Black Students *(Black)* % of Asian Students *(Asian)* % of Non-resident Students *(Alien)*

Data Cleaning:

In this study, we focus on not-for-profit higher education institutions that have tuition costs. First, we remove all institutions that do not have tuition costs (either N/A or 0), which include U.S. service schools for the U.S. Armed Forces. After the adjustment, the dataset is reduced to 4298 observations. We also eliminate institutions that are private-for-profit or less-than-2-year. These institutions typically include career-oriented programs and technical schools, which do not fit in our study of higher education institutions. They also have lower costs and produce extreme values in our model. After this adjustment, the dataset contains 3096 observations, 2859 of which are complete cases.

Next, we construct scatterplots of *Tuition* against each of the explanatory variables. Out of 22 observed potential outliers, we remove 7 outliers that are resulted from exogenous factors, such as religion afiliation and special needs schools. The process is explained with more details in Appendix 2. The final dataset contains 2852 observations with complete cases.

Analysis:

First, we conduct a best subset analysis using all explanatory variables in Table 1. The model that includes all variables has the highest adjusted R-squared value of 82.09%. We then examine plots of residuals versus fitted value and each explanatory variable. To correct for observable trends in residual plots, we apply the log transformation to *Enrollment*, *Asian*, *Black* and *Alien.* Because ethnic variables contain zero values, we use the log(x+1) transformation for *Asian, Black* and *Alien*. The comparison of residual plots is provided in Appendix 3. The final model, denoted BLM.1, is:

$$\begin{aligned} Tuition = {} & -10970 + 798.0\ Average.Salary^{**} + 80.13\ Amount.FA^{**} + 77.78\ Percent.FA^{**} \\ & + 111.1\ Graduate.Rt^{**} - 213.9\ STuFau.Ratio^{**} + 449.4\ log(Enrollment)^{**} \\ & + 1379\ log(Asian + 1)^{**} - 362.5\ log(Black + 1)^{**} + 37.31\ White^{**} \\ & + 915.2\ log(Alien + 1)^{**} \end{aligned}$$

$$Adjusted\ R^2 = 83.11\% \qquad ^{**}\ indicates\ significance\ at\ 0.01\ and\ ^{*}\ at\ 0.05.$$

The transformation improves the model's explanatory power by 1.02%, and all variables in the model are significant at 0.01. We also test a variety of interaction terms but none of them improves the model or provide extra explanatory power (see Appendix 1 for a detailed list of interaction terms). Thus, BLM.1 is a better fitting model.

Next, we use BLM.1 as the base model and include *Private*, our variable of interest, in the base model. The new model, denoted BLM.2, is:

$$Tuition = -15320 + 983.8\ Average.Salary^{**} + 66.57\ Amount.FA^{**} + 56.82\ Percent.FA^{**} + 91.86\ Graduate.Rt^{**} - 205.5\ STuFau.Ratio^{**} + 1106\ log(Enrollment)^{**} + 1124\ log(Asian + 1)^{**} - 202.8 log(Black + 1)^{*} + 37.68\ White^{**} + 755.2\ log(Alien + 1)^{**} + 4828\ Private^{**}$$

$Adjusted\ R^2 = 84.59\%$ $^{**}\ indicates\ significance\ at\ 0.01\ and\ ^{*}\ at\ 0.05.$

By setting BLM.1 as the reduced model and BLM.2 as the full model, we use the extra sum of squares F-test to verify whether *Private* provides significant extra explanatory power, that is:

$$H_0: \beta_{private} = 0$$
$$H_a: \beta_{private} \neq 0$$

The test statistics is F = 273.28 and the p-value is 2.2e – 16, which is significant at 0.01. Thus, we conclude that *Private* provides significant explanatory power for predicting tuition. In other words, there is a significant difference in tuition between private and public institutions.

Discussion

Based on the results above, we find that institutional control is a significant factor in determining tuition, and BLM.2, with an adjusted R-squared value of 84.59%, has the highest explanatory power for *Tuition*. We also observe consistencies between existing research and the signs of the coefficients. The 2015 College Board survey shows positive tuition gap between private and public institutions, which supports the positive sign of *Private* (Bell, Baum, Ma & Pender, 2015). At the same time, higher educational quality – higher graduation rate and lower student-to-faculty ratio – is positively associated with *Tuition*. These evidence further support the reliability of our model.

We have to be cautious when interpreting the size of BLM.2's coefficients, since our explanatory variables are not completely independent from each other. One important assumption for the model is that institutions set their educational quality and student preference independent from their financial capabilities. In reality, however, operating budget is an important factor that affects educational quality and enrollment decisions.

Another limitation of our model is the prediction of highly ranked institutions and special institutions. Appendix 4 provides the plot of residuals versus fitted values and examples of observations with absolute values of residuals greater than 10,000. We find that these large residuals are mainly caused by exogenous factors such as religious affiliation and reputation. On the one hand, our model typically overestimates tuition costs of elite institutions (expected value greater than observed value), such as California Institute of Technology, Yale University and Harvard University, and religiously affiliated institutions, such as Gallaudet University and Apex School of Theology. One explanation is that even though top rank institutions provide exceptional educational environment based on the indicators, they are unlikely to charge high tuitions since this action may hinder their enrollment of prospective students. As for religiously affiliated institutions, these institutions typically offer discounted tuitions as they enroll members of churches, thus resulting in overestimation of the model. On the other hand, the model frequently underestimates famous public universities, such as Purdue University and University of Michigan – Ann Arbor, and technical schools, such as San Francisco Art Institute. Because these institutions are well recognized in their fields, our model does not explain additional tuition costs that are resulted from reputational factors.

Future work can focus on endogenizing religious affiliation and institutional rank as explanatory variables. The consistent deviation of our model provides support that these factors are influential to tuition costs.

Bibliography

Bell, D., & Baum, S., & Ma, J., & Pender, M., &(2015). *Trends in College Pricing.* College Board. Retrieved from http://trends.collegeboard.org/sites/default/files/trends-college-pricing-web-final-508-2.pdf.

Ehrenberg, R. (2000). *Tuition Rising: Why Colleges Cost So Much.* Cornell University.

Hess, F. M., & Hochleitner, T. (May, 2012). *College rankings inflation: Are you overpaying for prestige?* American Enterprise Institute. Retrieved from https://www.aei.org/publication/college-rankings-inflation-are-you-overpaying-for-prestige/.

Li, H. (May, 2013). *The Rising Cost of Higher Education: A Supply & Demand Analysis.* Leonard N. Stern School of Business.

Schoen, J. W. (2013). *Why does a college degree cost so much?* CNBC. Retrieved from http://www.cnbc.com/2015/06/16/why-college-costs-are-so-high-and-rising.html.

St. John, E. P.. (1992). *Changes in Pricing Behavior during the 1980s: An Analysis of Selected Case Studies.* The Journal of Higher Education, 63(2), 165–187. http://doi.org/10.2307/1982158.

Throsby, C. D.. (1991). *The Financial Impact of Foreign Student Enrollments.* Higher Education, 21(3), 351–358. Retrieved from http://www.jstor.org/stable/3447139.

Yanikoski, R. A., & Wilson, R. F.. (1984). *Differential Pricing of Undergraduate Education*. The Journal of Higher Education, 55(6), 735–750. http://doi.org/10.2307/1981511.

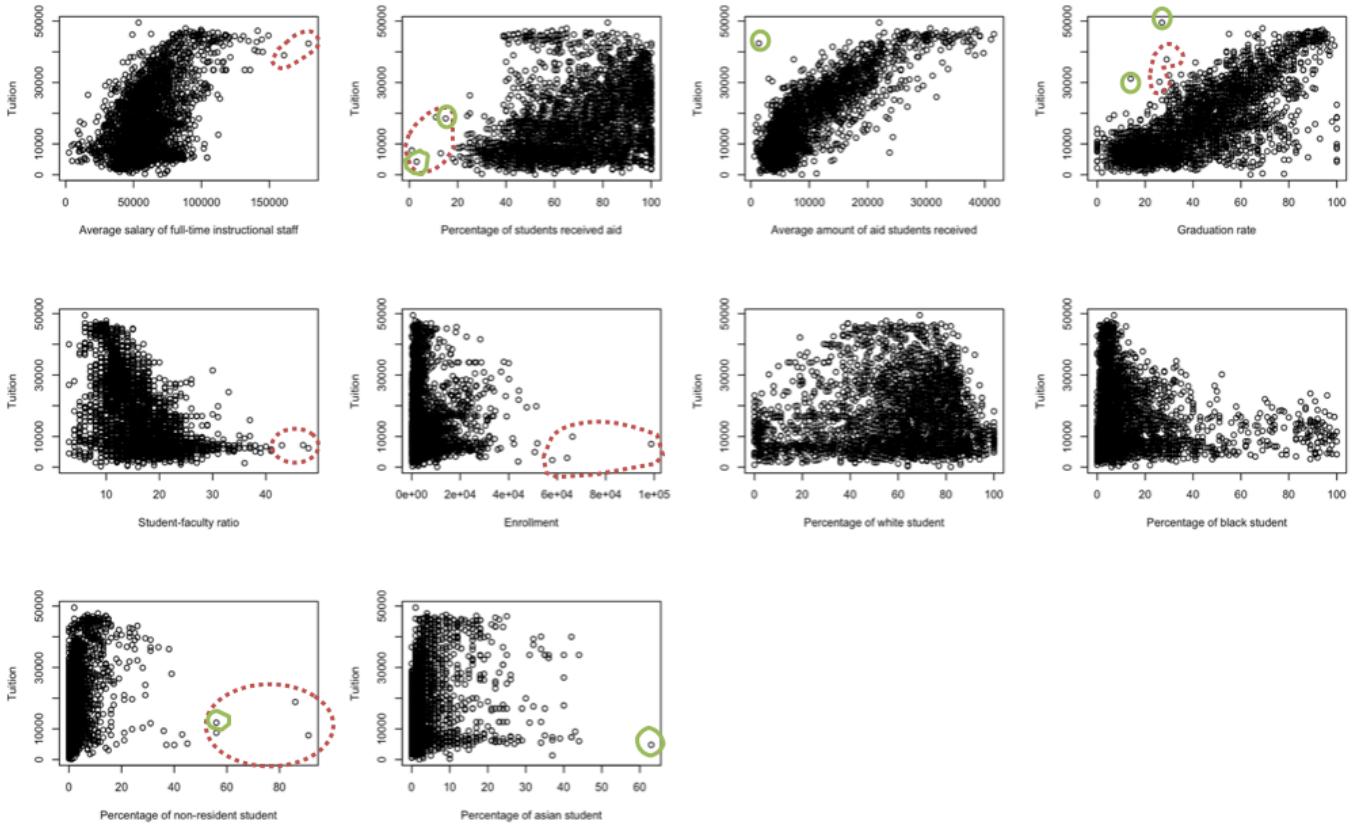
Appendix 1. Variable Description

Variable Name	Description
Tuition	Out-of-state tuition, in 2013
Average.Salary	Average salary equated to 9 months of full-time instructional staff
Percent.Aid	Percent of undergraduate students receiving federal, state, local, institutional or other sources of grant aid
Amount.Aid	Average amount of federal, state, local, institutional or other sources of grant aid dollars received by undergraduate students
Graduate.Rt	Graduation rate for the undergraduate students
StuFau.Ratio	Student-to-faculty ratio
Enrollment	Undergraduate enrollment
Asian	Percent of undergraduate enrollment that are Asian
White	Percent of undergraduate enrollment that are White
Black	Percent of undergraduate enrollment that are Black
Alien	Percent of undergraduate enrollment that are Nonresident Alien
Private	1 = private not-for-profit institutions; 0 = public institutions

**Data is retrieved from the IPEDS Database from the Institute of Education Sciences, http://nces.ed.gov/ipeds/datacenter/Default.aspx*

**Interaction terms attempted: Average.Salary*StuFau.Ratio, Average.Salary*Enrollment Average.Salary*White, Percent.Aid*Enrollment, Percent.Aid*White, Percent.Aid*Alien, Graduate.Rt*StuFau.Ratio*

Appendix 2. Elimination of Influential Outliers



In the graph above, the dashed-line circles (red) mark all 22 potential outliers, and the solid-line circles (green) mark the 7 eliminated outliers. We keep some potential outliers because they are consistent with our scope of higher education institutions. For example, we keep the two outliers in the plot of *Tuition vs. Average.Salary,* which are Harvard University and Stanford University. The 7 outliers are eliminated because their tuition costs are influenced by exogenous factors such as religion affiliation and special needs schools. These outliers include:

- Virginia Baptist College (Religion affiliation)
- Saint Charles Borromeo Seminary-Overbook (Religion affiliation)
- Webb Institute (Admitted U.S residents all receive full tuition scholarship)
- Laboure College (Religion affiliation)
- Landmark College (Special needs schools)
- World Mission University (Religion affiliation)
- Divine Word College (Religion affiliation)

Appendix 3. Comparison of Residual Plots before and after Log-transformation

Figure 1. Residual Plots before Log-transformation

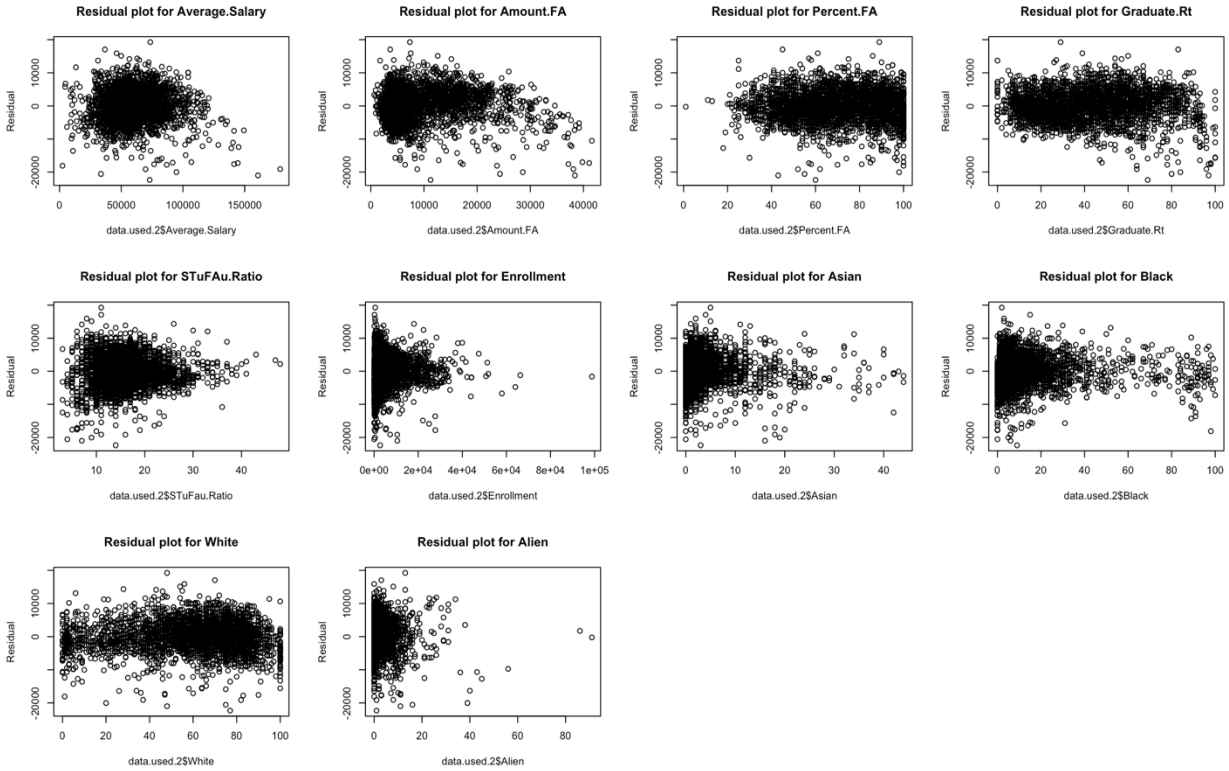
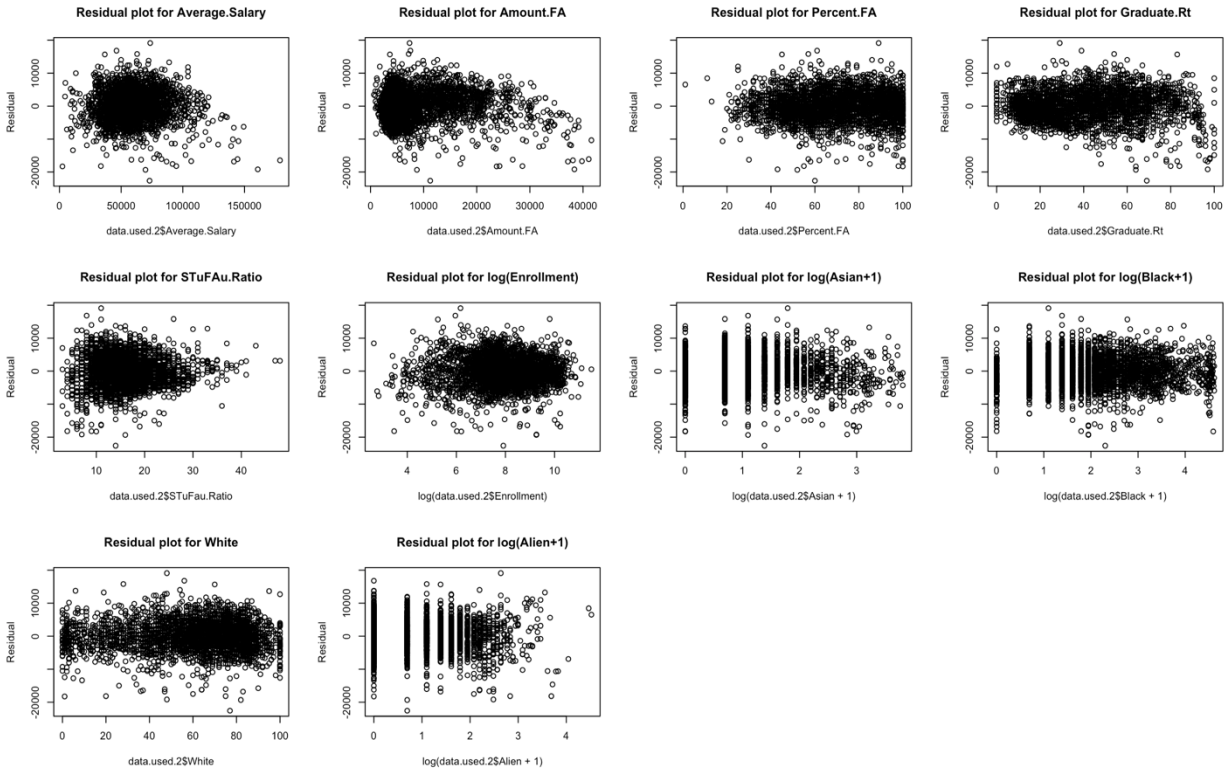
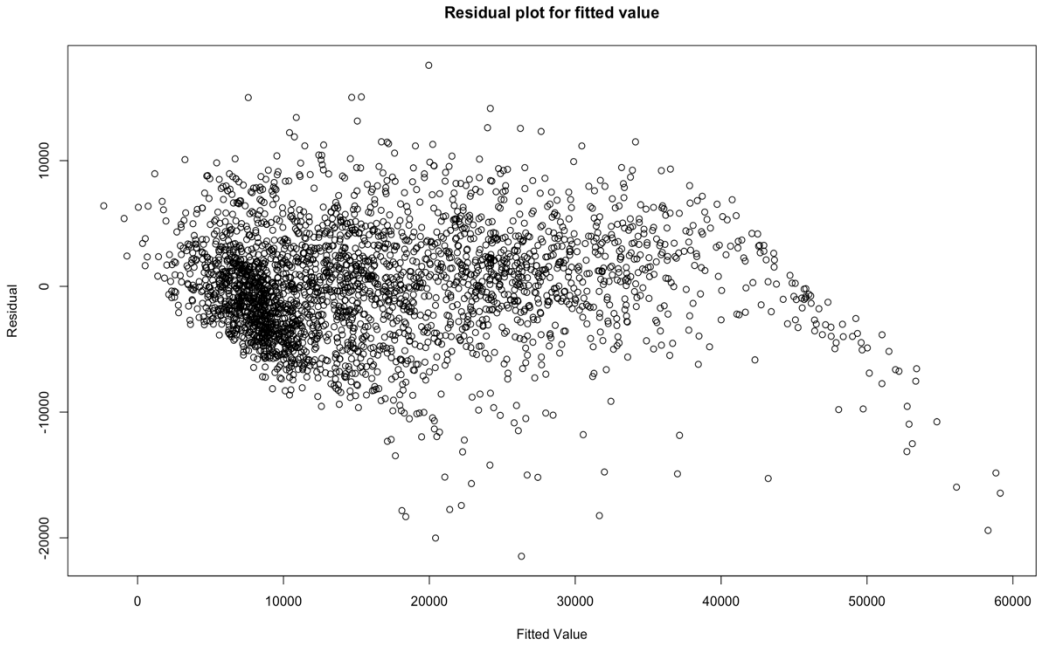


Figure 2. Residual Plots after Log-transformation



Appendix 4. Full Model Discussion



Sample List of Overestimated Institutions

Institution Name	Tuition	Residuals	Fitted Value
Yale University	44000	-14837	58837
Stanford University	42690	-16445	59135
Princeton University	40170	-15965	56135
Harvard University	38891	-19407	58298
Gallaudet University	13424	-18232	31656
Talmudic College of Florida	12250	-15188	27438
Moody Bible Institute	9120	-13160	22280
Apex School of Theology	4800	-12321	17121
Baptist Missionary Association Theological Seminary	4200	-13470	17670

Sample List of Underestimated Institutions

Institution Name	Tuition	Residuals	Fitted Value
California Institute of the Arts	39976	12316	27660
San Francisco Art Institute	37536	17576	19960
Indiana University-Purdue University-Indianapolis	28565	11460	17105
The Boston Conservatory	38800	12559	26241
School of the Museum of Fine Arts-Boston	38324	14144	24180
University of Michigan-Ann Arbor	41617	11166	30451
American Musical and Dramatic Academy	31520	11287	20233
Ohio State University-Mansfield Campus	22632	12220	10412
Cornish College of the Arts	33550	10115	23435