Spring 2014

Online or Traditional University: A Comparison of the Labor Market Returns to College Type

Leslie Schotz
Trinity College, leslie.schotz@trincoll.edu

Online or Traditional University: A Comparison of the Labor Market Returns to College Type

By

Leslie Schotz

A Thesis Submitted to the Department of Economics
of Trinity College in Partial Fulfillment of the
Requirements for the Bachelor of Science Degree

Economics 498-99

April 10, 2014

Abstract:

In a time when the rising costs of education have deterred students from seeking a college degree, finding a cost-effective alternative to a traditional university has become an increasingly important issue. This study seeks to evaluate the labor market returns to earning a degree online versus a traditional university through an econometric regression based on survey data of recent graduates from online and traditional universities. Regression analysis compares average income for graduates of online and traditional institutions while controlling for measures of school type, characteristics, selectivity, and region at the institutional level. The effect of college type on average income was statistically indistinguishable from zero. However, there were statistically significant and positive returns to college selectivity and quality. A calculation of the net present value of attending an online university versus a traditional university demonstrates that there are positive returns to attending a traditional university.

Acknowledgements

I would like to thank everyone for their continuous support throughout this year. I greatly appreciate all of your time, help, and effort; I would not have been able to complete this without you.

I would like to extend a special thank you to my thesis advisor, Chris Hoag for his unwavering support. I very much appreciate his active guidance and patience, and for agreeing to take me on as an advisee from the beginning. This thesis would not have been made possible without you.

Thank you to the other Economics Professors for offering your intelligent advice to help me put together the best possible project.

I would like to thank the librarians for helping me with the research process.

Thank you to my friends and family for acting as a constant source of support and encouragement

Contents

LIST OF TABLES 5
LIST OF FIGURES 19
INTRODUCTION 20
LITERATURE REVIEW 23
DATA 40
METHODOLOGY 42
The Regression 42
Expected Signs 45
Regression Controls 50
Initial Hypothesis 52
RESULTS 53
Analysis 53
F-test for Overall Statistical Significance 54
T-tests for Statistical Significance: 54
Interpretations on Coefficients: 55
Regression Testing: 59
Heteroskedasticity: 59
The Park Test 60
The White Test 60
Multicollinearity: 61
Simple Correlation 61
Variance Inflation Factor 62
Interactive Dummy Variable and F-test on a Subset of Coefficients 62
Final Regression 63
Net Present Value of Attending College 64
CONCLUSION 67
References 68
Appendix 1: Sample of the Data 70

LIST OF TABLES

Table 1: List of variables

Variable	Category	Definition	Unit
AVGINC2	Dependent Variable	Average income of graduates from each school	Percent
STUFACR	College Quality	Student-to-faculty ratio	Number
ENROLLMENT	College Quality	Total enrollment	Number
GRADRATE1	Student Ability	Graduation rate	Percent
NETTUITION	College Quality	Net tuition each student pays, on average	Dollars
PROFIT	College Type	Whether the school is a for-profit institution	1 = yes; 0 = no
PUBLIC	College Type	Whether the school is a public institution	1 = yes; 0 = no
TYPE	College Type	Whether the school is an online-only university	1 = yes; 0 = no
MOSTCOMP	Selectivity	Whether the school is ranked as a 'most competitive' school in the *Barron's* selectivity rankings	1 = yes; 0 = no
HIGHLYCOMP	Selectivity	Whether the school is ranked as a 'highly competitive' school in the *Barron's* selectivity rankings	1 = yes; 0 = no
VERYCOMP	Selectivity	Whether the school is ranked as a 'very competitive' school in the *Barron's* selectivity rankings	1 = yes; 0 = no
COMP	Selectivity	Whether the school is ranked as a 'competitive' school in the *Barron's* selectivity rankings	1 = yes; 0 = no
LESS COMP	Selectivity	Whether the school	1 = yes; 0 = no

		is ranked as a ‘less competitive’ school in the *Barron’s* selectivity rankings	
SPECIAL	Selectivity	Whether the school is ranked as a ‘special’ school in the *Barron’s* selectivity rankings	1 = yes; 0 = no
WEST	Region	Whether the school is located in the western region	1 = yes; 0 = no
MIDWEST	Region	Whether the school is located in the mid-western region	1 = yes; 0 = no
SOUTH	Region	Whether the school is located in the southern region	1 = yes; 0 = no

Table 2: Regression Output with the Dependent variable as the average of the log of the minimum and maximum income

Dependent Variable: AVGINC2
Method: Least Squares
Date: 04/01/14 Time: 15:25
Sample: 1 848
Included observations: 814
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.749606	0.009562	496.7238	0.0000
STUFACR	0.000738	0.000383	1.926747	0.0544
EFYTOTLT_EF2012A_	5.94E-08	9.51E-08	0.624695	0.5323
NETTUITION	6.75E-07	2.40E-07	2.806092	0.0051
GRADRATE1	-0.048978	0.032045	-1.528433	0.1268
PROFIT	0.042510	0.034204	1.242828	0.2143
PUBLIC	0.000895	0.005615	0.159324	0.8735
TYPE	-0.048145	0.032567	-1.478334	0.1397
MOSTCOMP	0.077881	0.008655	8.998841	0.0000
HIGHLYCOMP	0.048714	0.007093	6.867931	0.0000
VERYCOMP	0.024356	0.006544	3.721679	0.0002
COMP	-0.002344	0.005860	-0.399920	0.6893
SPECIAL	-0.002033	0.011726	-0.173368	0.8624
WEST	0.009874	0.007970	1.238951	0.2157
MIDWEST	-0.019701	0.005877	-3.352486	0.0008
SOUTH	-0.027613	0.004601	-6.002163	0.0000

R-squared	0.221337	Mean dependent var	4.765603
Adjusted R-squared	0.206700	S.D. dependent var	0.065475
S.E. of regression	0.058317	Akaike info criterion	-2.826399
Sum squared resid	2.713861	Schwarz criterion	-2.733977
Log likelihood	1166.344	Hannan-Quinn criter.	-2.790924
F-statistic	15.12222	Durbin-Watson stat	1.723564
Prob(F-statistic)	0.000000	Wald F-statistic	20.48921
Prob(Wald F-statistic)	0.000000		

Table 3: Regression output with the average of the minimum and maximum incomes as dependent variable

Dependent Variable: AVGINC1
Method: Least Squares
Date: 03/24/14 Time: 12:54
Sample: 1 848
Included observations: 813
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	65417.37	2366.627	27.64160	0.0000
STUFACR	-37.70554	81.23517	-0.464153	0.6427
EFYTOTLT_EF2012A_	0.010654	0.015794	0.674562	0.5001
NETTUITION	0.057921	0.021737	2.664622	0.0079
GRADRATE1	-843.1427	4636.768	-0.181838	0.8558
PROFIT	-660.8280	5396.837	-0.122447	0.9026
PUBLIC	194.9662	871.2725	0.223772	0.8230
TYPE	-2565.374	5107.033	-0.502322	0.6156
MOSTCOMP	14766.15	2220.958	6.648549	0.0000
HIGHLYCOMP	9301.127	2045.947	4.546123	0.0000
VERYCOMP	5340.205	2127.072	2.510589	0.0123
COMP	1553.836	1955.885	0.794442	0.4272
LESSCOMP	1206.147	2182.072	0.552753	0.5806
NONCOMP	2419.818	2359.344	1.025632	0.3054
WEST	2885.546	1475.129	1.956132	0.0508
MIDWEST	-3183.606	731.9400	-4.349545	0.0000
SOUTH	-4393.630	728.7645	-6.028875	0.0000

R-squared	0.266838	Mean dependent var	68173.36
Adjusted R-squared	0.252101	S.D. dependent var	10301.66
S.E. of regression	8908.998	Akaike info criterion	21.04820
Sum squared resid	6.32E+10	Schwarz criterion	21.14649
Log likelihood	-8539.093	Hannan-Quinn criter.	21.08593
F-statistic	18.10676	Durbin-Watson stat	1.737143
Prob(F-statistic)	0.000000	Wald F-statistic	21.46703
Prob(Wald F-statistic)	0.000000		

Table 4: Regression with interaction between WEST and NETTUITION

Dependent Variable: AVGINC2
Method: Least Squares
Date: 04/02/14 Time: 19:30
Sample: 1 848
Included observations: 814
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.755303	0.011720	405.7457	0.0000
STUFACR	0.000793	0.000402	1.973803	0.0487
EFYTOTLT_EF2012A_	9.62E-08	1.02E-07	0.944568	0.3452
NETTUITION	5.60E-07	1.37E-07	4.076432	0.0001
GRADRATE1	-0.046300	0.031723	-1.459520	0.1448
PROFIT	0.010782	0.046714	0.230800	0.8175
PUBLIC	0.001099	0.005634	0.195041	0.8454
TYPE	-0.016628	0.050379	-0.330056	0.7414
MOSTCOMP	0.072083	0.011173	6.451627	0.0000
HIGHLYCOMP	0.042011	0.009733	4.316205	0.0000
VERYCOMP	0.017990	0.009256	1.943710	0.0523
COMP	-0.007644	0.008615	-0.887299	0.3752
LESSCOMP	-0.007553	0.010011	-0.754471	0.4508
SPECIAL	-0.009620	0.012102	-0.794897	0.4269
WEST	-0.024944	0.041439	-0.601962	0.5474
MIDWEST	-0.019976	0.005907	-3.381804	0.0008
SOUTH	-0.028519	0.004640	-6.145925	0.0000
WEST*NETTUITION	2.88E-06	3.01E-06	0.955455	0.3396

R-squared	0.230177	Mean dependent var	4.765603
Adjusted R-squared	0.213737	S.D. dependent var	0.065475
S.E. of regression	0.058057	Akaike info criterion	-2.832903
Sum squared resid	2.683049	Schwarz criterion	-2.728929
Log likelihood	1170.992	Hannan-Quinn criter.	-2.792994
F-statistic	14.00027	Durbin-Watson stat	1.717855
Prob(F-statistic)	0.000000	Wald F-statistic	18.85550
Prob(Wald F-statistic)	0.000000		

Table 5: Regression with interaction between MIDWEST and NETTUITION

Dependent Variable: AVGINC2
Method: Least Squares
Date: 04/02/14 Time: 19:32
Sample: 1 848
Included observations: 814
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.755693	0.011783	403.6095	0.0000
STUFACR	0.000746	0.000381	1.955835	0.0508
EFYTOTLT_EF2012A_	5.76E-08	9.53E-08	0.603936	0.5461
NETTUITION	6.77E-07	2.50E-07	2.710172	0.0069
GRADRATE1	-0.047545	0.032247	-1.474393	0.1408
PROFIT	0.042536	0.034232	1.242590	0.2144
PUBLIC	0.000788	0.005621	0.140166	0.8886
TYPE	-0.054498	0.033177	-1.642626	0.1009
MOSTCOMP	0.071386	0.011083	6.441151	0.0000
HIGHLYCOMP	0.042271	0.009781	4.321767	0.0000
VERYCOMP	0.017921	0.009284	1.930363	0.0539
COMP	-0.008791	0.008774	-1.002034	0.3166
LESSCOMP	-0.008632	0.009950	-0.867586	0.3859
SPECIAL	-0.008402	0.013662	-0.614996	0.5387
WEST	0.009954	0.007974	1.248271	0.2123
MIDWEST	-0.019125	0.012592	-1.518780	0.1292
SOUTH	-0.027643	0.004648	-5.947036	0.0000
MIDWEST*NETTUITION	-4.50E-08	8.67E-07	-0.051917	0.9586

R-squared	0.221701	Mean dependent var	4.765603
Adjusted R-squared	0.205079	S.D. dependent var	0.065475
S.E. of regression	0.058376	Akaike info criterion	-2.821952
Sum squared resid	2.712593	Schwarz criterion	-2.717978
Log likelihood	1166.535	Hannan-Quinn criter.	-2.782043
F-statistic	13.33781	Durbin-Watson stat	1.725473
Prob(F-statistic)	0.000000	Wald F-statistic	18.15247
Prob(Wald F-statistic)	0.000000		

Table 6: Regression with interaction between SOUTH and NETTUITION

Dependent Variable: AVGINC2
Method: Least Squares
Date: 04/02/14 Time: 12:18
Sample: 1 848
Included observations: 814
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.755831	0.011750	404.7459	0.0000
STUFACR	0.000757	0.000379	1.996662	0.0462
EFYTOTLT_EF2012A_	5.47E-08	9.47E-08	0.578017	0.5634
NETTUITION	6.62E-07	2.51E-07	2.636709	0.0085
GRADRATE1	-0.047116	0.032574	-1.446400	0.1485
PROFIT	0.042866	0.034179	1.254138	0.2102
PUBLIC	0.000735	0.005616	0.130798	0.8960
TYPE	-0.054889	0.033125	-1.657012	0.0979
MOSTCOMP	0.071147	0.011162	6.374116	0.0000
HIGHLYCOMP	0.041904	0.009822	4.266393	0.0000
VERYCOMP	0.017739	0.009289	1.909677	0.0565
COMP	-0.008897	0.008710	-1.021504	0.3073
LESSCOMP	-0.008625	0.009956	-0.866400	0.3865
SPECIAL	-0.008515	0.013534	-0.629175	0.5294
WEST	0.009935	0.007975	1.245807	0.2132
MIDWEST	-0.019685	0.005897	-3.337871	0.0009
SOUTH	-0.029270	0.007474	-3.916182	0.0001
SOUTH*NETTUITION	1.62E-07	5.58E-07	0.290376	0.7716

R-squared	0.221743	Mean dependent var	4.765603
Adjusted R-squared	0.205122	S.D. dependent var	0.065475
S.E. of regression	0.058375	Akaike info criterion	-2.822007
Sum squared resid	2.712444	Schwarz criterion	-2.718033
Log likelihood	1166.557	Hannan-Quinn criter.	-2.782098
F-statistic	13.34111	Durbin-Watson stat	1.726214
Prob(F-statistic)	0.000000	Wald F-statistic	18.82114
Prob(Wald F-statistic)	0.000000		

Table 7: Regression run for the joint F-test

Dependent Variable: AVGINC2
Method: Least Squares
Date: 04/02/14 Time: 14:10
Sample: 1 848
Included observations: 814

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.758652	0.015698	303.1314	0.0000
STUFACR	0.000699	0.000526	1.329161	0.1842
GRADRATE1	-0.046405	0.034908	-1.329364	0.1841
PROFIT	-0.004871	0.021941	-0.222001	0.8244
PUBLIC	0.000484	0.004977	0.097248	0.9226
MOSTCOMP	0.077077	0.014561	5.293434	0.0000
HIGHLYCOMP	0.051335	0.013852	3.706068	0.0002
VERYCOMP	0.025597	0.012985	1.971326	0.0490
COMP	-0.002930	0.012265	-0.238911	0.8112
LESSCOMP	-0.004077	0.013852	-0.294293	0.7686
SPECIAL	0.003189	0.020992	0.151937	0.8793
WEST	0.009345	0.006831	1.367948	0.1717
MIDWEST	-0.020097	0.005697	-3.527373	0.0004
SOUTH	-0.028869	0.005436	-5.311015	0.0000

R-squared	0.208968	Mean dependent var	4.765603
Adjusted R-squared	0.196113	S.D. dependent var	0.065475
S.E. of regression	0.058704	Akaike info criterion	-2.815553
Sum squared resid	2.756971	Schwarz criterion	-2.734684
Log likelihood	1159.930	Hannan-Quinn criter.	-2.784512
F-statistic	16.25667	Durbin-Watson stat	1.705989
Prob(F-statistic)	0.000000		

Table 8: Descriptive Statistics

Date: 04/01/14 Time: 15:48
Sample: 1 848

	AVGINC2	STUFACR	EFYTOTLT E...	GRADRATE1	PROFIT	PUBLIC	TYPE	MOSTCOMP	HIGHLYCOM...	VERYCOMP	COMP	LESSCOMP	SPECIAL	WEST	MIDWEST	SOUTH
Mean	4.765603	15.39926	8669.767	0.177484	0.015971	0.438575	0.018428	0.089681	0.103194	0.191646	0.472973	0.081081	0.014742	0.138821	0.222359	0.276413
Median	4.765156	15.00000	4764.500	0.179232	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Maximum	5.213017	82.00000	359464.0	0.932203	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
Minimum	4.162707	6.000000	95.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Std. Dev.	0.065475	5.391584	15908.76	0.069085	0.125438	0.496518	0.134574	0.285899	0.304399	0.393838	0.499576	0.273127	0.120593	0.345972	0.416087	0.447498
Skewness	-2.062374	3.234052	14.52113	1.352009	7.722151	0.247576	7.161385	2.872142	2.608744	1.566854	0.108266	3.069457	8.052844	2.089196	1.335357	0.999891
Kurtosis	26.65727	33.51470	300.9028	19.74133	60.63161	1.061294	52.28544	9.249200	7.805545	3.455031	1.011722	10.42157	65.84830	5.364738	2.783178	1.999781
Jarque-Bera	19559.06	33000.35	3038579.	9753.884	120740.9	135.7941	89343.16	2443.672	1706.532	340.0884	135.6713	3146.312	142765.5	781.8110	243.5122	169.5685
Probability	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Sum	3879.201	12535.00	7057190.	144.4720	13.00000	357.0000	15.00000	73.00000	84.00000	156.0000	385.0000	66.00000	12.00000	113.0000	181.0000	225.0000
Sum Sq. Dev.	3.485283	23633.24	2.06E+11	3.880280	12.79238	200.4287	14.72359	66.45332	75.33170	126.1032	202.9054	60.64865	11.82310	97.31327	140.7531	162.8071
Observations	814	814	814	814	814	814	814	814	814	814	814	814	814	814	814	814

Date: 04/01/14 Time: 16:13
Sample: 1 848

	AVGINC1	STUFACR	EFYTOTLT_E...	GRADRATE1	PROFIT	PUBLIC	TYPE	MOSTCOMP	HIGHLYCOM...	VERYCOMP	COMP	LESSCOMP	SPECIAL	WEST	MIDWEST	SOUTH
Mean	68189.93	15.38235	8650.862	0.177351	0.015931	0.437500	0.019608	0.089461	0.102941	0.191176	0.473039	0.080882	0.014706	0.138480	0.223039	0.275735
Median	67143.00	15.00000	4746.000	0.179232	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Maximum	196246.0	82.00000	359464.0	0.932203	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
Minimum	30201.00	6.000000	95.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Std. Dev.	10286.71	5.397194	15893.86	0.069322	0.125287	0.496383	0.138733	0.285583	0.304068	0.393469	0.499579	0.272822	0.120447	0.345615	0.416540	0.447158
Skewness	2.735121	3.218317	14.52952	1.321281	7.732103	0.251976	6.929646	2.876862	2.613243	1.570711	0.108000	3.074351	8.063183	2.093318	1.330632	1.003682
Kurtosis	32.56971	33.34571	301.3596	19.48929	60.78542	1.063492	49.02000	9.276333	7.829040	3.467133	1.011664	10.45164	66.01493	5.381979	2.770583	2.007377
Jarque-Bera	30745.91	32717.94	3055339.	9481.916	121662.1	136.1371	78537.29	2464.921	1721.617	342.9493	136.0046	3173.337	143852.0	788.8590	242.5887	170.5035
Probability	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Sum	55642980	12552.00	7059103.	144.7185	13.00000	357.0000	16.00000	73.00000	84.00000	156.0000	386.0000	66.00000	12.00000	113.0000	182.0000	225.0000
Sum Sq. Dev.	8.62E+10	23740.71	2.06E+11	3.916531	12.79289	200.8125	15.68627	66.46936	75.35294	126.1765	203.4069	60.66176	11.82353	97.35172	141.4069	162.9596
Observations	816	816	816	816	816	816	816	816	816	816	816	816	816	816	816	816

Table 9: Park test for student-to-faculty ratio

Dependent Variable: LOG(RESID^2)
Method: Least Squares
Date: 04/02/14 Time: 14:26
Sample: 1 848
Included observations: 814

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.836844	0.247308	-27.64510	0.0000
STUFACR	-0.054883	0.015159	-3.620571	0.0003

R-squared	0.015887	Mean dependent var	-7.681995
Adjusted R-squared	0.014675	S.D. dependent var	2.347628
S.E. of regression	2.330338	Akaike info criterion	4.532358
Sum squared resid	4409.546	Schwarz criterion	4.543911
Log likelihood	-1842.670	Hannan-Quinn criter.	4.536792
F-statistic	13.10854	Durbin-Watson stat	1.834428
Prob(F-statistic)	0.000312		

Table 10: Park test output for enrollment

Dependent Variable: LOG(RESID^2)
Method: Least Squares
Date: 04/02/14 Time: 14:27
Sample: 1 848
Included observations: 814

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.242712	0.093787	2.587896	0.0098
EFYTOTLT_EF2012A_	3.66E-06	5.18E-06	0.705815	0.4805

R-squared	0.000613	Mean dependent var	0.274404
Adjusted R-squared	-0.000618	S.D. dependent var	2.348514
S.E. of regression	2.349239	Akaike info criterion	4.548514
Sum squared resid	4481.367	Schwarz criterion	4.560067
Log likelihood	-1849.245	Hannan-Quinn criter.	4.552949
F-statistic	0.498175	Durbin-Watson stat	1.931414
Prob(F-statistic)	0.480505		

Table 11: The White Test

Heteroskedasticity Test: White

F-statistic	0.370848	Prob. F(67,746)	1.0000
Obs*R-squared	**26.23775**	Prob. Chi-Square(67)	1.0000
Scaled explained SS	448.7337	Prob. Chi-Square(67)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/02/14 Time: 14:37
Sample: 1 848
Included observations: 814
Collinear test regressors dropped from specification

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004247	0.025599	0.165907	0.8683
STUFACR^2	-1.98E-05	2.25E-05	-0.881984	0.3781
STUFACR*GRADRATE1	-0.002348	0.004415	-0.531928	0.5949
STUFACR*PROFIT	0.000916	0.001475	0.620698	0.5350
STUFACR*PUBLIC	-0.000626	0.000563	-1.111467	0.2667
STUFACR*MOSTCOMP	-2.37E-05	0.001623	-0.014584	0.9884
STUFACR*HIGHLYCOMP	0.000590	0.001513	0.390265	0.6965
STUFACR*VERYCOMP	0.000846	0.001157	0.731630	0.4646
STUFACR*COMP	0.000247	0.000986	0.250969	0.8019
STUFACR*LESSCOMP	0.000683	0.001148	0.595081	0.5520
STUFACR*SPECIAL	-0.000392	0.002109	-0.185755	0.8527
STUFACR*WEST	0.000491	0.000760	0.646077	0.5184
STUFACR*MIDWEST	4.48E-05	0.000721	0.062063	0.9505
STUFACR*SOUTH	0.000566	0.000751	0.753440	0.4514
STUFACR	0.000485	0.001465	0.331236	0.7406
GRADRATE1^2	-0.049427	0.064315	-0.768521	0.4424
GRADRATE1*PROFIT	-0.101041	0.255788	-0.395019	0.6929
GRADRATE1*PUBLIC	0.058953	0.037293	1.580789	0.1144
GRADRATE1*MOSTCOMP	0.045246	0.118017	0.383382	0.7015
GRADRATE1*HIGHLYCOMP	0.022686	0.104318	0.217469	0.8279
GRADRATE1*VERYCOMP	0.020846	0.087404	0.238500	0.8116
GRADRATE1*COMP	0.054492	0.082127	0.663510	0.5072
GRADRATE1*LESSCOMP	0.032500	0.091538	0.355041	0.7227
GRADRATE1*SPECIAL	0.006981	0.172773	0.040405	0.9678
GRADRATE1*WEST	0.059111	0.051883	1.139302	0.2549
GRADRATE1*MIDWEST	-0.039906	0.040197	-0.992775	0.3211
GRADRATE1*SOUTH	-0.010786	0.036486	-0.295627	0.7676
GRADRATE1	0.004984	0.098907	0.050390	0.9598
PROFIT^2	-0.015524	0.032902	-0.471841	0.6372
PROFIT*WEST	-0.002035	0.024488	-0.083100	0.9338
PROFIT*MIDWEST	-0.013393	0.035185	-0.380655	0.7036
PROFIT*SOUTH	-0.010639	0.024778	-0.429379	0.6678
PUBLIC^2	0.003151	0.015919	0.197942	0.8431
PUBLIC*MOSTCOMP	-0.000550	0.015549	-0.035385	0.9718
PUBLIC*HIGHLYCOMP	-0.007958	0.013788	-0.577223	0.5640
PUBLIC*VERYCOMP	-0.009912	0.012207	-0.811971	0.4171
PUBLIC*COMP	-0.003564	0.011299	-0.315375	0.7526

PUBLIC*LESSCOMP	-0.003895	0.012521	-0.311087	0.7558
PUBLIC*SPECIAL	-0.010964	0.021945	-0.499594	0.6175
PUBLIC*WEST	0.002979	0.006500	0.458331	0.6468
PUBLIC*MIDWEST	0.003521	0.005061	0.695715	0.4868
PUBLIC*SOUTH	0.000752	0.005106	0.147194	0.8830
MOSTCOMP^2	-0.008353	0.033099	-0.252358	0.8008
MOSTCOMP*WEST	-0.008465	0.017438	-0.485438	0.6275
MOSTCOMP*MIDWEST	0.006908	0.018988	0.363803	0.7161
MOSTCOMP*SOUTH	0.002702	0.014063	0.192117	0.8477
HIGHLYCOMP^2	-0.008828	0.031838	-0.277268	0.7817
HIGHLYCOMP*WEST	-0.009549	0.017004	-0.561589	0.5746
HIGHLYCOMP*MIDWEST	0.007541	0.016760	0.449967	0.6529
HIGHLYCOMP*SOUTH	0.001417	0.012403	0.114208	0.9091
VERYCOMP^2	-0.011191	0.024916	-0.449162	0.6534
VERYCOMP*WEST	-0.001397	0.015597	-0.089546	0.9287
VERYCOMP*MIDWEST	0.005364	0.015911	0.337120	0.7361
VERYCOMP*SOUTH	0.000701	0.011808	0.059394	0.9527
COMP^2	-0.007781	0.022662	-0.343356	0.7314
COMP*WEST	-0.000274	0.014812	-0.018518	0.9852
COMP*MIDWEST	0.006524	0.015357	0.424797	0.6711
COMP*SOUTH	-0.002735	0.011106	-0.246291	0.8055
LESSCOMP^2	-0.011961	0.026058	-0.459017	0.6464
LESSCOMP*WEST	-0.006007	0.016742	-0.358782	0.7199
LESSCOMP*MIDWEST	0.000303	0.017132	0.017668	0.9859
LESSCOMP*SOUTH	-0.002036	0.012500	-0.162915	0.8706
SPECIAL^2	0.003581	0.047130	0.075991	0.9394
SPECIAL*WEST	0.003063	0.034039	0.089974	0.9283
SPECIAL*MIDWEST	0.001755	0.024935	0.070362	0.9439
WEST^2	-0.011556	0.019948	-0.579315	0.5626
MIDWEST^2	0.000163	0.019890	0.008193	0.9935
SOUTH^2	-0.006941	0.017130	-0.405198	0.6854

R-squared	0.032233	Mean dependent var	0.003387
Adjusted R-squared	-0.054684	S.D. dependent var	0.020168
S.E. of regression	0.020712	Akaike info criterion	-4.836391
Sum squared resid	0.320016	Schwarz criterion	-4.443599
Log likelihood	2036.411	Hannan-Quinn criter.	-4.685622
F-statistic	0.370848	Durbin-Watson stat	2.078434
Prob(F-statistic)	0.999999		

Table 12: Correlation Matrix

	AVGINC2	STUFACR	EFYTOTLT	GRADRATE	NETTUITI(	PROFIT	PUBLIC	TYPE	MOSTCOM	HIGHLYCO	VERYCOMI	COMP	LESSCOMI	SPECIAL	WEST	MIDWEST	SOUTH
AVGINC2	1	-0.10558	0.042724	0.083557	0.16277	-0.00039	-0.08404	-0.00995	0.29822	0.168028	0.075932	-0.27088	-0.08702	-0.01055	0.145967	-0.10391	-0.1993
STUFACR	-0.10558	1	0.41611	-0.38065	-0.12474	0.350664	0.450495	0.366191	-0.34643	-0.10683	-0.07026	0.157677	0.114139	-0.09609	0.183238	-0.00837	0.158633
EFYTOTLT_EF2(	0.042724	0.41611	1	-0.21765	-0.01734	0.393729	0.219134	0.380105	-0.0564	0.025242	-0.00519	-0.06741	-0.02172	-0.04248	0.19134	-0.05413	0.033047
GRADRATE1	0.083557	-0.38065	-0.21765	1	0.073962	-0.28315	-0.22067	-0.31048	0.239537	0.177521	0.101119	-0.15468	-0.08243	-0.01006	-0.17052	-0.06117	-0.04986
NETTUITION	0.16277	-0.12474	-0.01734	0.073962	1	-0.05716	-0.14322	-0.07649	0.003089	0.132332	0.094122	-0.09304	-0.09264	0.107524	0.004093	0.014785	-0.1136
PROFIT	-0.00039	0.350664	0.393729	-0.28315	-0.05716	1	-0.1126	0.929786	-0.03999	-0.04322	-0.06203	-0.12069	-0.03784	-0.01558	0.090564	-0.02099	0.00891
PUBLIC	-0.08404	0.450495	0.219134	-0.22067	-0.14322	-0.1126	1	-0.1211	-0.21676	-0.07195	-0.04037	0.144541	0.118401	-0.06703	0.067601	-0.03204	0.156778
TYPE	-0.00995	0.366191	0.380105	-0.31048	-0.07649	0.929786	-0.1211	1	-0.04301	-0.04648	-0.06672	-0.1298	-0.0407	-0.01676	0.129918	-0.02933	-0.00299
MOSTCOMP	0.29822	-0.34643	-0.0564	0.239537	0.003089	-0.03999	-0.21676	-0.04301	1	-0.10647	-0.15283	-0.29734	-0.09323	-0.03839	0.01077	-0.11614	-0.09785
HIGHLYCOMP	0.168028	-0.10683	0.025242	0.177521	0.132332	-0.04322	-0.07195	-0.04648	-0.10647	1	-0.16517	-0.32135	-0.10076	-0.04149	-0.03108	0.003126	0.007055
VERYCOMP	0.075932	-0.07026	-0.00519	0.101119	0.094122	-0.06203	-0.04037	-0.06672	-0.15283	-0.16517	1	-0.46127	-0.14463	-0.05956	0.030187	0.039872	-0.06365
COMP	-0.27088	0.157677	-0.06741	-0.15468	-0.09304	-0.12069	0.144541	-0.1298	-0.29734	-0.32135	-0.46127	1	-0.2814	-0.11588	-0.05299	0.079244	0.052715
LESSCOMP	-0.08702	0.114139	-0.02172	-0.08243	-0.09264	-0.03784	0.118401	-0.0407	-0.09323	-0.10076	-0.14463	-0.2814	1	-0.03634	0.010906	-0.03978	0.078061
SPECIAL	-0.01055	-0.09609	-0.04248	-0.01006	0.107524	-0.01558	-0.06703	-0.01676	-0.03839	-0.04149	-0.05956	-0.11588	-0.03634	1	0.009851	-0.01638	-0.0756
WEST	0.145967	0.183238	0.19134	-0.17052	0.004093	0.090564	0.067601	0.129918	0.01077	-0.03108	0.030187	-0.05299	0.010906	0.009851	1	-0.21469	-0.24815
MIDWEST	-0.10391	-0.00837	-0.05413	-0.06117	0.014785	-0.02099	-0.03204	-0.02933	-0.11614	0.003126	0.039872	0.079244	-0.03978	-0.01638	-0.21469	1	-0.3305
SOUTH	-0.1993	0.158633	0.033047	-0.04986	-0.1136	0.00891	0.156778	-0.00299	-0.09785	0.007055	-0.06365	0.052715	0.078061	-0.0756	-0.24815	-0.3305	1

Table 13: VIF

Variable	VIF
STUFACR	1.010509
EFYTOTLT_EF2012A_	1.001865
GRADRATE1	1.006157
NETTUITION	1.028789
PROFIT	1.000001
PUBLIC	1.007846
TYPE	1.00011
MOSTCOMP	1.103429
HIGHLYCOMP	1.028483
VERYCOMP	1.005454
COMP	1.080504
LESSCOMP	1.00764
SPECIAL	1.000117
WEST	1.02086
MIDWEST	1.011756
SOUTH	1.040336

Table 14: The final regression with all statistically insignificant variables, excluding TYPE, removed

Dependent Variable: AVGINC2
Method: Least Squares
Date: 04/07/14   Time: 10:45
Sample: 1 848
Included observations: 814
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.746707	0.009376	506.2520	0.0000
STUFACR	0.000871	0.000329	2.644685	0.0083
NETTUITION	6.78E-07	2.40E-07	2.822419	0.0049
GRADRATE1	-0.050726	0.030475	-1.664489	0.0964
TYPE	-0.009129	0.014664	-0.622528	0.5338
MOSTCOMP	0.080340	0.007687	10.45077	0.0000
HIGHLYCOMP	0.051002	0.005686	8.968968	0.0000
VERYCOMP	0.026497	0.005211	5.085257	0.0000
WEST	0.009630	0.007853	1.226227	0.2205
MIDWEST	-0.019844	0.005859	-3.387079	0.0007
SOUTH	-0.027430	0.004602	-5.960978	0.0000

R-squared	0.220038	Mean dependent var	4.765603
Adjusted R-squared	0.210325	S.D. dependent var	0.065475
S.E. of regression	0.058183	Akaike info criterion	-2.837017
Sum squared resid	2.718389	Schwarz criterion	-2.773477
Log likelihood	1165.666	Hannan-Quinn criter.	-2.812628
F-statistic	22.65370	Durbin-Watson stat	1.716265
Prob(F-statistic)	0.000000	Wald F-statistic	29.92661
Prob(Wald F-statistic)	0.000000		

Table 15: Net Present Value Calculation

Present value of benefits of attending a traditional university versus an online university						
C = β	2565.37					
i1	8.50%	PV(i1)	$ 30,180.12			
i2	7.58%	PV(i2)	$ 33,845.52			
Present Value of the costs of the benefits of switching						
	Time	0	1	2	3	4
	Cash Flow		6256.19472	6256.195	6256.195	6256.195
	PV		5766.078083	5314.358	4898.026	4514.309
	PV(Costs, i1)		$ 20,492.77			
	Time	0	1	2	3	4
	Cash Flow		6256.19472	6256.195	6256.195	6256.195
	PV		5815.415945	5405.692	5024.835	4670.812
	PV(Costs, i2)		$ 20,916.76			
NPV(Attending traditional vs. online university)						
w/ i1	$ 9,687.35					
w/ i2	$12,928.77					

LIST OF FIGURES

Figure 1: Scatter plot of the errors versus STUFACR

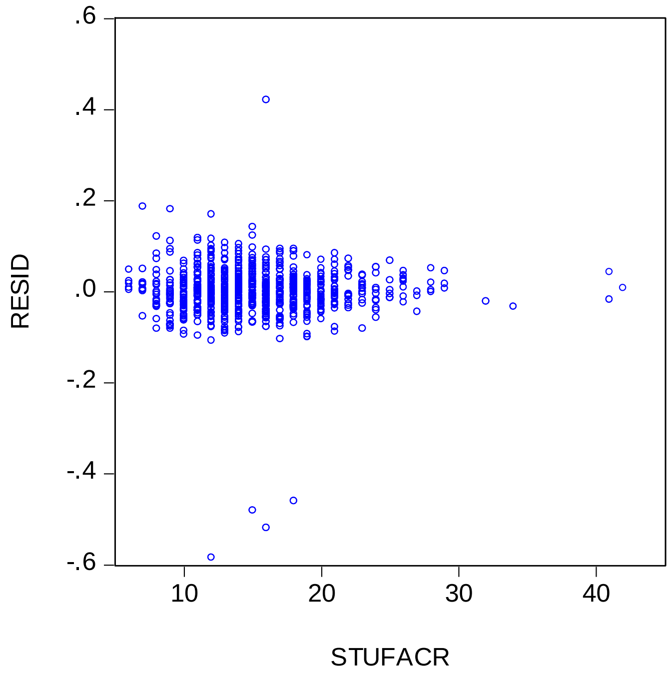
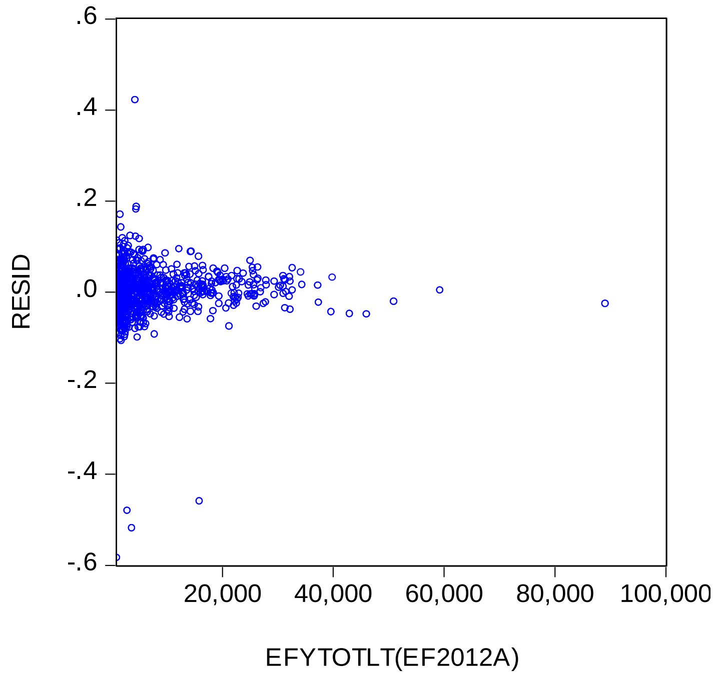


Figure 2: Scatter plot of the errors versus ENROLLMENT



INTRODUCTION

In recent years, the cost of attending college has increased dramatically, while obtaining a post-secondary degree has simultaneously become more of a necessity for the majority of jobs available in the economy. As the rising costs of education have deterred massive numbers of students from seeking a college degree, finding a cost-effective alternative to a traditional university has become an increasingly important issue in order to continue to promote and to facilitate economic growth.

In the coming years, the White House predicts that employment in jobs requiring education beyond a high school diploma will grow more rapidly than employment in jobs that do not.[1] In response to the shortage of available affordable colleges, the Obama Administration proposed the “Race to the Top” initiative to make college more affordable for students nationwide so that America will be the leader in the proportion of college

[1] "Higher Education." *The White House*. The White House, n.d. Web. 03 Apr. 2014.

graduates by 2020. According to the basic theory of human capital model, increasing the level of investment in higher education has the potential to facilitate economic growth, as increases in human capital can greatly affect economic growth. As the cost of education has grown to a level that discourages Americans from attending college, finding a viable way to increase the number of college graduates and participants has the potential to help promote economic growth.

Online universities offer a wide variety of programs and majors at a much lower cost than that of a traditional university. These schools can therefore provide an opportunity for someone who cannot afford a traditional university in this harsh economic climate to receive a college education. However, many employers are hesitant to accept online universities as a credible and equivalent alternative to a traditional university system, possibly due a general mistrust of the program as a result of a difference in learning experiences of online versus an in-person class, or an overall lack of knowledge about the various programs offered.

Today, many measure the value their college education in dollars, by comparing their salary with other recent, employed college graduates. As income has recently become one of the most popular methods for measuring the value of a college degree, this study seeks to determine whether or not an online university student can achieve the same employment outcomes as a traditional university student. This paper focuses on the effect college type has on average income by comparing the average income of online university students and traditional university students. An econometric model will be used in order to analyze the effect college type, defined as an online university or traditional university, has on average salary of recent graduates. The regression compares the average of the logarithm of the maximum and minimum salaries for each institution included in the sample. While most previous literature uses data by student, this model provides a comparison the average income of graduates from of each institution rather than a comparison of the incomes of individual students.

The independent variables included in the regression describe college quality, selectivity, type, student ability, and region. The regression incorporates multiple proxies specifically for college quality, such as student to faculty ratio and net tuition, which are similar to the proxies employed by Black and Smith (2006). This model uses the variable describing graduation rate as a proxy for both college quality and student ability, unlike previous literature, which most often used an average SAT variable to capture student ability.

Understanding the differences and similarities in labor market returns to an education earned online versus an education earned in a traditional university system would help to establish online universities as a substitute for traditional universities. However, the results from the model do not suggest that the effect college type has on average income is indistinguishable from zero, meaning there is no statistical difference between the average incomes of students who attend online universities versus traditional universities were found in the model. Additionally, all variables describing college type, such as whether or not the school is a public versus private institution and for-profit versus not-for-profit, were found to be statistically insignificant. Select variables describing the level of selectivity the school is categorized as, for example most competitive, highly competitive and very competitive schools, on the other hand, have a strong, positive effect on average income, indicating that students who attend more selective schools have higher incomes than students who attend less selective schools.

While the statistical analysis of the regression suggests college type has no effect on average income, a calculation of the net present value of attending an online university versus a traditional university demonstrates that there are returns to attending a traditional university. Despite higher tuition costs, the net present value of attending a traditional university is much greater than the net present value of attending an online university. Such a large discrepancy between the two values suggests that while online universities may be more cost-effective, in the sense that online universities have lower tuition costs, and college type has not statistical impact on income, in the long run, paying higher tuition costs pays off over time in terms of increased labor market returns to attending a traditional university.

LITERATURE REVIEW

This econometric model expands upon the findings and methods of previous studies. However, every study that is considered in this section neglects to take into account online universities as a college type. Rather than investigating only college selectivity or quality, this model specifically looks at college type in two categories: online universities and traditional universities.

The study conducted by Dale and Krueger (2011) sought to examine the relationship between the college that students attended and the earnings students reported in a follow up survey in order to determine if the "selectivity" of a college affected the amount of money the student will make after graduation.[2] This study used the College and Beyond Survey linked to Detailed Earnings Records from the Social Security Administration as data on each of the students within each cohort in order to build an econometric model to estimate the return of several college characteristics. Dale and Krueger separated students into two different cohorts – one cohort consisted of students who entered college in 1989, and the other followed a cohort of students who entered college in 1976.[3]

Dale and Kreuger's first model attempted to estimate the relationship between earnings, school quality, and observable student characteristics, such as high school GPA and individual student's SAT scores. However, estimating a regression of this kind creates a difficulty with the selection, as not all characteristics that lead students to apply to selective colleges are available to researchers.[4] Neglecting to account for ability is a major problem because ability is often rewarded in the labor market, so a measure of ability should be included within the regression.[5] The equation used to estimate student's attributes should read as follows:

$$\ln(W_i) = \beta_0 + \beta_1 Q_1 + \beta_2 X_1 + \beta_3 X_2 + \varepsilon_i$$

where Q measures college selectivity, X_1 measures observable student characteristics, and X_2 measures unobservable student characteristics. In this regression, β_1 represents the monetary payoff to attending a more selective college. In previous literature, because researchers did not have access to data or did not consider the unobservable student characteristics, estimated regressions took the form:

---

[2] Dale, Stacy, and Alan B. Krueger. Estimating the return to college selectivity over the career using administrative earnings data. No. w17159. National Bureau of Economic Research, 2011.
[3] Dale and Krueger
[4] Dale and Krueger
[5] Dale and Krueger

$$\ln(W_i) = \beta_0' + \beta_1' Q_1 + \beta_2' X_i + u$$

This estimation, even if students randomly selected the college they went to, will yield biased coefficients. The coefficients describing the impact of college selectivity and observable student characteristics will not serve as accurate representations of the true population parameters. The bias in the coefficients occurs as a result of the excluded variable X_2, the variable that describes the unobservable student characteristics. Thus, excluding X_2, it is then put into the random error term U such that

$$u = \gamma x_2 + v$$

where v is uncorrelated with Q_1, X_1 and X_2, and also has a zero mean such that E(u)=0. However, U is only uncorrelated with Q_1 and X_1 if and only if X_2 is uncorrelated with the observable regressors (Q and X_1). But in this case, X_2 is correlated with the observable regressors because the unobservable student characteristics are related to the student's observable characteristics. As such, u is now correlated with the observable regressors as X_2 is now incorporated into u. Thus, this omitted variable violated the assumption that the random error term is uncorrelated with an estimate of any coefficient.

The omission of student ability from the regression causes the remaining estimated coefficients to be biased, as the student ability is relevant in explaining the dependent variable.[6] The linear projection of X_2 onto the other explanatory variables, in other words the bias created in the other independent variables, as

$$X_2 = \delta_0 + \delta_1 Q_i + \delta_2 X_1 + r$$

[6] Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. MIT press, 2010.

By definition, E(r) = 0, and the covariance between (X_j, r) = 0. Now we can infer the probability limit of the OLS estimators by regressing a variable, y, onto Q_1 and X_1, by finding an equation that does satisfy the OLS assumptions.

$$y = (\beta_0 + \gamma\delta_0) + (\beta_1 + \gamma\delta_1)Q_1 + (\beta_2 + \gamma\delta_2)v_1 + (v + \gamma\delta_j)$$

In this case, the error term has zero mean and is uncorrelated with each regressor. Now the probability limit of the OLS estimators from the regression of y can be read as

$$plim\, \widehat{\beta_j} = \beta_j + \gamma\delta_j$$

When the correlation between X_2 and a particular variable, say Q_1 is the focus, a common assumption is that all of the δ_j – the affect of the bias – in the equation except the intercept are coefficient on X_2 are zero. By writing out the effect of the bias, the correct sign and magnitude of the bias on the coefficient can be determined. According to Wooldridge, If y > 0, then X_2 and Q_1 are positively correlated. In this case, if students are not randomly choosing which school they attend, the payoff of attending a selective school is biased upwards, because the selectivity of the school attended is correlated positively with the student's unobserved abilities.[7] If the coefficients are biased and inconsistent, then we cannot make any conclusions about the population, as the estimated coefficients no longer provide an accurate portrayal of the population.

In order to account for the bias in the regression caused by the omitted variable X_2, Dale and Kruger included a proxy for the unobserved student characteristics. The proxy must be redundant, or ignorable in the structural equation, and uncorrelated with the omitted variable and each of the other regressors.[8] It is always assumed that a proxy satisfies the redundancy condition, however, it does not always satisfy the second

[7] Dale and Krueger
[8] Wooldridge

property. If the proxy turns out to be correlated with one or more of the independent variables, then the proxy is imperfect.[9] An OLS with an imperfect proxy still yields inconsistent estimated coefficients. The hope is that the bias on each of the regressors is smaller in magnitude than if the proxy was omitted from the linear regression[10]. In this study, Dale and Krueger included the average SAT score of the schools the student applied to (AVG) as a proxy for the unobservable student attributes. AVG is irrelevant in explaining wage once the unobserved student characteristics have been controlled for, yet AVG must be included in the regression because Dale and Krueger cannot obtain data on X_2. By definition, ability and ambition affect wage, thus the average SAT score of the schools the student applied to would not matter if true ability or motivation were known.

However, the concern that AVG does not control for all of the bias remains. For example, using AVG as a selection correction can still yield a biased estimate of β_1 if students' school enrollment decisions are a function of X_1 or any other variable not in the model. Previous studies found that students are more prone to matriculate to schools that provide more financial aid than others.[11] In this case, if more selective colleges provide more financial aid, the estimated coefficients will be biased upwards, because students with higher levels of X_2 will attend more selective schools, regardless of the outcomes of applications to other colleges. On the other hand, if less selective schools provide more generous financial aid, which would create incentive for students with higher levels of X_2 to attend less selective schools, the bias in the estimated coefficients will be downwards.

[9] Wooldridge
[10] Wooldridge
[11] Dale and Krueger

Dale and Krueger ultimately concluded that attending the most selective college to which a student was admitted does not necessarily benefit the student.[12] Instead, the unobserved student characteristics, such as ambition or passion, greatly affect how each student performs in the labor market. Simply relying on the prestige of the college name does not automatically generate a higher income for students.

Their results led to two different conclusions in each model. In the model that did not account for unobserved student characteristics, they found a significant and positive effect of the return to college selectivity.[13] Dale and Krueger analyzed both earnings data sets coming from the C&B earnings and the SSA earnings. With respect to C&B earnings, they found that the estimated coefficient on college SAT score/100 in the basic model is 0.068 (Table 3)[14], meaning that attending a school with a 100-point higher SAT score is correlated with about 7% higher earnings (Table 3).[15] Yet when looking at the C&B earnings within the self-revelation model, Dale and Krueger found that the estimated coefficient on school SAT score was indistinguishable from zero, suggesting school selectivity has no significant impact on student future earnings (Table 3).[16] With respect to earnings data collected from the SSA, Dale and Krueger found the estimated coefficients on school SAT score ranged from 0.048 to 0.061 – similar to the results with the C&B data (Table 3).[17] However, in the self-revelation model, the estimated coefficient on school SAT score was negative at -0.021 to -0.023 and statistically

[12] Dale and Krueger
[13] Dale and Krueger
[14] Dale and Krueger
[15] Dale and Krueger
[16] Dale and Krueger
[17] Dale and Krueger

insignificant (Table 3).[18] For both cohorts and both sources of earnings data, the return to college selectivity has a large and positive effect on earnings when unobserved student characteristics are not taken into consideration. However, when unobserved student characteristics are taken into account in the self-revelation model, the returns to college selectivity fall significantly and are statistically insignificant.[19] These results imply that students do not automatically benefit from attending the most selective college. Instead, students can maximize their benefits by instead choosing a school based on how well the school fits with the student's interests and abilities.[20]

The study conducted by Dale and Kruger provides an excellent method for accounting for the selection bias due to an omitted variable within the model. Using the average SAT score of the schools the student applied to seems to be the most efficient solution to account for bias in the estimated coefficients. Unfortunately, due to data constraints, this exact proxy variable for student ability is not included in this model. Online universities do not have an SAT requirement for students to apply and be accepted into the school. As such, online universities do keep a record of any data describing the average SAT score of the first-year class. While Dale and Kreuger's exact methodology cannot be employed in this case, their careful awareness and attention to selection bias created by omitted variables is directly applicable to this model.

In Zhang's study "Do Measures of College Quality Matter? The Effect of College Quality on Graduates' Earnings," Zhang examines the relationship between college quality and graduates' earnings. However, he specifically is interested in the different methods of measuring college quality and how changing the method affects the

[18] Dale and Krueger
[19] Dale and Krueger
[20] Dale and Krueger

relationship. Zhang concluded that simply having a single measure of college quality leads to misleading results, as higher education institutions are complex, and require more than one measure of college quality.[21]

In order to analyze the differences in the effect of measures of college quality on earnings, Zhang utilized an array of quality measures on the same data set, and then compared his results to those from previous studies that only used one measure of college quality. Furthermore, his model excludes the characteristics of high-quality institutions, such as high-level peers, superior resources, and higher levels of academic and social engagement.[22] For his data, Zhang used the second follow-up survey of the Baccalaureate and Beyond study and used the following equation for the OLS regression

$$\ln(W_i) = \beta_0 + \beta_1 Q_{ij} + \beta_2 D_i + \beta_3 F_i + \beta_4 J_i + \beta_5 A_i + \varepsilon_i$$

where Q_{ij} measures the quality of institution j he or she attended, Di measures demographic characteristics such as race and ethnicity, F_i measures family background (family income or first generation college graduate), A_i measures students' academic background, and J_i captures labor market factors (age, job tenures, and their square terms).[23] In this study, Zhang altered the definition of Q_{ij}, but used this same baseline model and the same data set to examine the possible differences among the estimates for different measures of college quality.

Zhang considered three additional measures of college quality: mean SAT scores of the entering freshman class, tuition and fees for each institution, and Carnegie institutional classifications. For the baseline model, he used *Barron's* ratings as the

[21] Zhang, Liang. "Do measures of college quality matter? The effect of college quality on graduates' earnings." The Review of Higher Education 28.4 (2005): 571-596.
[22] Zhang
[23] Zhang

single measure of college quality. The results from the OLS regression suggest that college quality, when considering only this measure of college quality, have a positive and significant impact on graduates' earnings. The estimated coefficient of Q_{ij} for most selective, private institutions was 0.1754 and a significant t-statistic (Table 3).[24] Meaning holding all other independent variables constant, students who attend high-quality, private institutions earn about 17% more than students who do not. This study found results that were consistent with Zhang's findings, as students who attend most competitive schools have incomes of about 7% more than students who do not (Table 1).

The additional measures of college quality were included in the subsequent regressions of the model. By changing and adding measures of college quality in the model, Zhang then manipulates and changes the definition of a high-quality college.[25] As a result of including additional measures, the number of students who attended a high-quality college increases (Table 5),[26] so the number of schools that are considered to be high-quality with the additional measures increased. In return, the magnitude of the effect school quality has on earnings changes and is now smaller and less significant than in the baseline model. For example, when mean SAT scores of the entering freshmen class is included as a measure of college quality, the estimated coefficient for a most selective, private institution is now 0.1005 (Table 8).[27] In other words, a student who attends a most selective, private institution earns about 10% more than a student who does not.

[24] Zhang
[25] Zhang
[26] Zhang
[27] Zhang

However, it seems as though using tuition and fees as a measure of college quality might not be the best proxy for college quality, as the costs of attending a school may fluctuate greatly as a result of variations in costs from external factors, such as government block grants and states lending different amounts to schools, and cost of living. Yet in this model, rather than simply using the sticker price of each school, the net tuition is taken into account as a way to control for the cost-effectiveness of each school and also serve as a potential proxy for school quality. Higher quality schools may be able to afford to lend more aid to students, and therefore may have a lower net tuition than others. Zhang's choice of using tuition as a measure of college quality could potentially be a source of bias in Zhang's model, as tuition and fees could potentially be considered an irrelevant proxy. Tuition and fees do not seem to be highly correlated enough with the unobservable variable, college quality.

Yet Zhang makes a solid argument for why a single measure of college quality can provide misleading results. One variable does not serve as an adequate proxy variable due to the complexity of the concept of college quality. Moreover, Zhang stresses the importance of clearly defining which variables describe college quality. As seen from this study, depending on how one defines college quality can yield different results. Zhang defined college quality only by four variables. This study will adopt parts of Zhang's method for choosing more than one measure of college quality. However, the data included in this model is not aggregated at the individual level, and instead is aggregated across individual institutions. Furthermore, due to lack of data specifically on the online schools, the variables used measures of college quality differ than those in Zhang's study.

The study conducted by James Monks in 1999 sought to examine differentials in earnings across individual and institutional characteristics. The results from this study suggest that students who graduate from more selective schools earn significantly more than students who graduate from less selective schools.[28]

In this study, Monks built an econometric model to the estimate the returns of selectivity, student's ability, labor market experience, and college type on earnings.

$$\ln(W_{ijt}) = X_{0it}\beta_1 + X_{1it}\beta_2 + Q_j\beta_3 + \delta_i + \varepsilon_{ijt}$$ [29]

Where W_{ijt} is the hourly wage, X_{1it} are individual and time varying labor market experiences, X_{1it} are non-time varying individual characteristics which influence earnings, Q_j are college characteristics, δ_i is a normally distributed individual specific error component, and ε_{ijt} is a normally distributed random error.[30]

In order to create a model that fully captured the impact on earnings, Monks had to account for student ability, as ability can have a major impact on a student's future earnings. He did so by adding a group of independent variables: each student's Armed Forces Qualifications Test, race, gender, industry, and occupation. Furthermore, he used Barron's ranking and college classifications as a means to determine college selectivity, quality, and type.[31] However, even though Monks used controls for both the ability and college quality in the model, the estimated regression will yield biased estimated coefficients due to an omitted variable that would describe the institutions' and students'

---

[28] Monks, J. (2000). The returns to individual and college characteristics: Evidence from the National Longitudinal Survey of Youth. Economics of Education Review, 19(3), 279-289.
[29] Monks
[30] Monks
[31] Monks

enrollment process, as the process is not a random selection.[32] In order to account for this bias, Monks included variables that described each student's attributes, such as the student's academic ability and the student's ability to pay (family income) that may influence the enrollment process. [33] Unfortunately, this does not perfectly account for the enrollment process, so some bias in the estimated coefficients may still remain.

Due to lack of data on the enrollment process for online universities, this model does not control for the enrollment process. As a result, bias in the estimated coefficients still remains and will lead to unreliable results. Monks simply included other variables as a way to partially account for the selection bias created by the enrollment process. Monks also attempted to account for variations in earnings as a result of peer effects and classroom dynamics of race and gender by performing different and separate regressions by race and gender to allow some of the coefficients to vary.[34]

The evidence from this student suggests that students who graduate from highly or most selective schools earn significantly more, about 15% (Table 4)[35], than graduates from less selective institutions. As student's move from non-competitive, or less competitive institutions to highly or the most competitive institutions, earnings significantly increased. The estimated coefficients describing college type, quality, and students' academic ability were all large, positive and significant. More specifically, students who graduate from private colleges earned about 4.5% (Table 4)[36] more than students who graduated from a public university, while students who graduated from a

---

[32] Monks
[33] Monks
[34] Monks
[35] Monks
[36] Monks

degree-granting research institution earned about 14% (Table 4)[37] more than graduates from liberal arts colleges. Also, for every point increase in the AFQT test score, earnings increase by 11.1% (Table 4), meaning that a student with greater academic ability – measured by an increase in AFQT test score – will earn more than a student with less academic ability.[38] These results strongly suggest a positive and significant relationship between college quality and wages, even after controlling for colleges' characteristics and students' ability.

The study conducted by Brewer, Eide, and Ehrenberg (1999), investigates high school students' choice of college type based on individual and family characteristics and estimates of the net costs of attendance.[39] By analyzing the results of the regression, Brewer, Eide, and Ehrenberg were able to determine the effects of college quality on wages and earnings and how this effect varies across time. They found strong, positive economic return to attending an elite private institution, and even found some evidence to suggest that this premium has increased over time.[40]

In order to estimate the effects of college type, Brewer, Eide, and Ehrenberg regressed the logarithm of individual student's earnings or hourly wage on the student's characteristics and a set of college characteristics.[41] College quality was measured with proxied indicators of selectivity of the undergraduate body, including average SAT scores

[37] Monks
[38] Monks
[39] Brewer, Dominic J., Eric R. Eide, and Ronald G. Ehrenberg. "Does it pay to attend an elite private college?." Journal of Human Resources 34.1 (1999).
[40] Brewer, Eide, and Ehrenberg
[41] Brewer, Eide, and Ehrenberg

of the entering freshmen, and resource measures such as instructional expenditures per student, library size, and faculty per student.[42]

Furthermore, they attempted to control for the systematic selection of college type on the basis of the expected labor market payoff. Meaning that although the tuition costs of attending a highly-selective, private institution are much greater than that of a non-selective, public institutions, students will choose to attend the highly-selective school based on their expectations of labor market returns. However, the problem is that if students are investing in college quality solely on the basis of expected returns, college type cannot be treated as an exogenous determinant of earnings.[43]

Their results show clear patterns across their data cohorts, suggesting that students with higher family incomes and more highly educated parents are more likely to attend higher-quality colleges.[44] Furthermore, they found that those with greater academic talent predominate at high-quality schools, and financial aid is about twice as high for students attending private institutions than public (Table 1).[45] Interestingly, their results suggest that not only were there significant and positive returns to attending an elite college, but also these returns increased over time from 19% to 39% increases in annual earnings (Table 2).[46]

Brewer, Eide, and Ehrenberg note a potential bias in the estimated coefficient caused by the differences in financial aid given out by each school. They also attempted to control for the college selection process by estimating a reduced-form college choice

---

[42] Brewer, Eide, and Ehrenberg
[43] Brewer, Eide, and Ehrenberg
[44] Brewer, Eide and Ehrenberg
[45] Brewer, Eide and Ehrenberg
[46] Brewer, Eide and Ehrenberg

multinominal logit model.[47] In this model, they added two variables to proxy for the likelihood of being admitted to a particular institution: the availability of college openings and the students test score difference.[48] In this model, the methodology employed by Brewer, Eide, and Ehrenberg may not be applicable since online colleges technically have no limit to the number of college openings available as well as online universities do not keep SAT score data because they do not have testing requirements. The college selection process in this model is quite different than if the model was focusing at traditional institutions. Moreover, Brewer, Eide and Ehrenberg specifically added in the conclusion of the paper that although their analysis suggests the return to elite private colleges increased significantly for the 1980s cohorts as compared to the 1972 cohort, they did not in any way attempt to determine the cause of this change.[49]

Black and Smith (2006) build upon previous studies, such as Zhang (2005) and Dale and Kreuger (2002) that analyze the effects of college quality on wages.[50] Similar to the methodology employed by Zhang (2005), Black and Smith question the common practice of using a single proxy of college quality and instead examine the effect by using multiple proxies for quality. Black and Smith claim that adopting one measure of quality will underestimate the effect college type has on wages.[51] Existing literature most commonly employs a "one-factor model," in which one college quality measure is included – usually a measure of selectivity – and included in outcome equations with

[47] Brewer, Eide and Ehrenberg
[48] Brewer, Eide and Ehrenberg
[49] Brewer, Eide and Ehrenberg
[50] Black, Dan A., and Jeffrey A. Smith. "Estimating the returns to college quality with multiple proxies for quality." Journal of Labor Economics 24.3 (2006): 701-728.
[51] Black and Smith

covariates. [52] Most studies follow the methods of Heckman and Robb (1985), and assume "selection on observables" and hope that by including a sufficiently rich X with some measure of individual ability will control for the non-random matching of students and colleges.[53] However, Black and Smith, for the purposes of their study, assume selection on observables and place their attention on issues of interpreting the parameters used to capture college quality.[54]

Black and Smith built an econometric model to estimate the wage effect of college quality

$$\ln(W_{ij}) = X_i\beta + \delta S_i + \gamma Q_{ij}^{*} + \varepsilon_{ij}$$

where $\ln(W_{ij})$ is the natural logarithm of the wage rate of the *i*th person attending the *j*th college, X_i is a series of covariates, S_i is the number of years of schooling and Q_{ij} is the latent quality variables, and an error term that is assumed to be uncorrelated with the regressors.[55] For the latent college quality variables, Black and Smith included each college's faculty-student ratio, rejection rate, freshmen retention rate, mean SAT score and mean faculty salaries. This model will adopt the use of faculty-student ratio as a means to measure college quality. Yet as a result of online universities neglecting to keep SAT score information, lack of retention and rejection rates, the remaining measures of college quality used by Black and Smith (2006) will not be included in the regression.

Black and Smith found none of the estimated coefficients on the latent quality variables to be statistically different from zero, and ultimately failed to model this scale factor.[56] However, their estimators suggest a downward bias on quality of about 20%

---

[52] Black and Smith
[53] Black and Smith
[54] Black and Smith
[55] Black and Smith
[56] Black and Smith

relative to only using an SAT variable as a proxy for quality.[57] Furthermore, their results suggest that even though average SAT score is the most reliable or accurate measures of college quality, future studies need to consider whether or not this measure varies on the same scale as college quality.[58]

Most previous literature supports the claim that there are significant and positive returns to college selectivity and quality. Students who attend more selective schools earn more than students who attend less selective schools. Furthermore, most literature agrees with the methodology of including more than one measure of college quality, as a single measure of quality will inevitably yield biased coefficients. This model will incorporate multiple proxies for college quality, such as student to faculty ratio and net tuition. Unfortunately, due to data availability constraints and this study being based on institutional level data, this model will not include the other proxies for quality, similar to those used in the study done by Black and Smith (2006). Previous literature additionally stresses the importance of utilizing a proxy variable for student ability within the regression in order to attempt to control for the selection bias created by the non-random process of college enrollment and ability being rewarded in the labor market. Most studies, specifically in the research conducted by Dale and Krueger (2011), used an average SAT variable to proxy for student ability. Again, due to lack of data on the institutional level, an average SAT variable cannot be included in this regression. Instead, graduation rate will be used as a proxy for student ability.

Since online universities are fairly new institutions, previous literature does not focus on the wage effect of college type as defined as online versus traditional university.

[57] Black and Smith
[58] Black and Smith

As a result, many of the variables included in previous research to measure college quality and student ability are not available or relevant to online universities. At any rate, this paper will adopt parts of previous studies and combine each of their methods in order to estimate the labor market returns to college type.

DATA

As a result of an inability to collect individual level data in time and for online universities, this model will estimate the effect of college type on wages on an institutional level.

This study is based on data from the Payscale.com, IPEDS Data Center, the Department of Education Statistics, and *Barron's Guide to College* selectivity rankings. Payscale.com serves as an online career research center of compensation information for both employers and individuals. The website obtains its information by offering a free-of-charge salary survey that asks about specifics of individual's jobs, including compensation factors.[59] The data from each survey is subsequently run through Payscale.com's data-cleaning algorithm to validate and identify potential outliers and biases within the data. For each school, Payscale.com provides a salary range for each degree type of the school's graduates. Each year, Payscale.com releases its College Salary Report, in which it provides the ranking of the school, highest starting salary to lowest starting salary, region, school type, mid-career salary, and percent of students with high job meanings. However, because not all of the schools included in this study are on

[59] "PayScale Methodology." PayScale Methodology. N.p., n.d. Web. 08 Apr. 2014.

the College Salary Report, the starting salary given from the report could not be used in order to remain consistent with the data. To obtain AVGINC2, I took the average of the logarithms of the minimum and maximum salary for each school's bachelor's degree. AVGINC1 is just the average of the minimum and maximum salaries for each school. The magnitude of the coefficients did not significantly change for the estimated coefficients, so AVGINC2 was used as the main dependent variable in the regression (Table 2 and 3). According to labor economic theory, income is a logarithmic function rather than a linear function. As such, I converted the minimum and maximum salaries given on Payscale.com to logarithmic functions, and then took the average of the logarithms.

The collection of variables from every institution in the 2012 IPEDS universe that describe each institution's characteristics was obtained from the 2012 Institutional Characteristics survey. This analysis includes 849 colleges from the IPEDS 2012 universe, 17 of the colleges included are online-only universities. The average income of graduates is $68,189.93 (Table 8). A fully copy of the descriptive statistics can be found in the list of tables.

Only including 17 online only universities could potentially lead to inconclusive results. The extremely small sample of online universities creates a very wide confidence interval, and could have increase the magnitude of estimated coefficient on college type.

As a result of this analysis depending heavily on individuals' responses to the IPEDS and Payscale.com's survey questions, the sample for the primary analysis is restricted to survey respondents, as well as the respondents who are employed and earning an income. As the data is limited to schools and students who respond to the survey, an analysis on this data might yield inaccurate results, as the data reflects only a sample of the survey respondents rather than every school and graduate. Furthermore, AVGINC2 might be biased, as the reported salaries on Payscale.com are self-reported. The bias also may stem from the difference in ages of each college. Older colleges may have older, more established alumni reporting salaries that might be higher than the salary of a younger graduate. An abbreviated set of data can be found in the appendix.

METHODOLOGY

This project seeks to verify whether or not college type has an effect on students' salaries. By building an econometric model, the size, direction, and magnitude of the effect college type has on earnings can be estimated. In other words, this regression can determine if earning a college degree from an online university results in a higher or lower salary on average than a student would earn if the degree was obtained through a traditional university. However, instead of looking at the effect on an individual level, meaning looking at individuals' starting salaries, this study compares the effect on an institutional level, between undergraduate schools.

The Regression

An OLS estimation regresses the dependent variable, the average salary of students from each institution, on the independent variables. The population relationship reads as:

$$\begin{aligned} AVGINC2_i = {} & \beta_0 + \beta_1 STUFACR_i + \beta_2 ENROLLMENT_i + \beta_3 GRADRATE1_i + \\ & \beta_4 NETTUITION_i + \beta_5 PROFIT_i + \beta_6 PUBLIC_i + \beta_7 TYPE_i + \beta_8 MOSTCOMP_i + \\ & \beta_9 HIGHLYCOMP_i + \beta_{10} VERYCOMP_i + \beta_{11} COMP_i + \beta_{12} LESSCOMP_i + \beta_{13} SPECIAL_i \\ & + \beta_{14} WEST_i + \beta_{15} MIDWEST_i + \beta_{16} SOUTH_i + \varepsilon_i \end{aligned}$$

AVGINC2 was obtained by taking the average of the logarithms of the maximum and minimum salary reported on Payscale.com.

The independent variables included in the regression describe the quality of each school, the level of selectivity, characteristics of the student body and the region of the United States in which the school is located. For this model, college type, named TYPE, is defined as online or traditional in addition to public versus private and for-profit versus not-for-profit. The estimated coefficient on TYPE, which takes on a value of 1 if the school is an online university, and 0 if the schools is a traditional university, quantifies the estimated effect college type has on average income.

Net tuition (NETTUITION) was included in the model as a measure of college quality, similarly to Zhang's method (2005). Dale and Krueger (2011) calculated net tuition by subtracting the average amount of grant aid dollars received by undergraduate students from the sticker price tuition. NETTUITION in this model was calculated in the same manner, with the sticker price of tuition coming from the Department of Education's 2012 Tuition and Enrollment Figures and the average amount of aid taken from the IPEDS 2012 Institutional Characteristic Survey.[60]For this model, the average amount of aid is assumed to be per student, rather than conditional on whether or not the student receives aid or not. However, if this variable is the average amount of aid given per the number of students who actually receive grants rather than per total student, then this variable could provide an inaccurate measure of net tuition. Furthermore, it is unclear whether or not room and board costs are included in the average aid number, as room and board chargers are excluded from the tuition sticker price.

[60] Dale and Krueger

Graduation rate (GRADRATE1) serves as a measure of student ability as well as college quality. Graduating from college requires a certain degree of student ability. In this model, due to lack of data on the online universities, a reported graduation rate in the IPEDS 2012 Institutional Characteristic Survey could not be used. As a result, GRADRATE1 was calculated by dividing the total number of graduates by total enrollment number. While this number does not yield a true graduation rate, it is the best possible estimate for graduation rate given the lack of data.

Dummy variables that describe the level of selectivity school were also included in the model. *Barron's* uses seven selectivity categories: most competitive (MOSTCOMP), highly competitive (HIGHLYCOMP), very competitive (VERYCOMP), competitive (COMP), less competitive (LESSCOMP), non-competitive and special (SPECIAL). All online universities fall under the non-competitive category, as there are minimal requirements to apply and be accepted into an online university. The special school category includes schools that teach a specialized curriculum for a specific trade or subject. Art schools and culinary schools are examples of the types of schools included within the category. The non-competitive selectivity category is omitted from the model, so all estimated coefficients are relative to non-competitive schools.

Differences in price levels as well as costs of living in the four regions of the United States – West, Midwest, South and East – may create differences in the incomes. A student who is working in New York, for example, may have a higher base salary than a student working in Alabama, simply because the cost of living in New York is much higher than Alabama. Thus, every salary in New York must be higher in order to account for higher costs of living. In order to control for these differences, the dummy variables

WEST, MIDEWEST and SOUTH were added to the regression to describe the location of the school. If a school is located in the western region of the United States, the variable WEST takes on a value of 1, and 0 for the variables MIDWEST and SOUTH. The omitted region is the eastern region, so if a school is located within the eastern region of the United States, there will be values of 0 in all of the location variables.

Expected Signs

College Quality Variables

$\beta_{STUFACR} < 0$. As the student to faculty ratio increases, AVGINC2 decreases, because lower student to faculty ratios indicate more learning in the classroom and a more interactive class environment, holding all other independent variables constant. More interactive class environments encourage students to be engaged and proactive in participating. In addition to student to faculty ratio representing a causal determinant of better learning, student to faculty ratio can also serve as a proxy for resources of the school or student ability. Schools with lower student to faculty ratios may be able to provide students with more resources, such as bigger libraries and career services centers, to help students land jobs, as the school can spend more money per student.

$\beta_{ENROLLMENT}$? 0. As the total enrollment for the undergraduate school increases, it is not clear if AVGINC2 increases or decreases. Schools with high total undergraduate enrollments have access to much larger alumni networks that provide excellent job opportunities for recent graduates. However, larger total enrollments also mean that even though there is a large alumni network available to the students, there are a much greater number of students competing with each other jobs. Moreover, schools with larger total enrollments may be experiencing significant returns to scale as schools with a high number of students are receiving tuition from a greater number of students and can thus afford to provide their students with better quality equipment, learning facilities faculty and resources. The direction of the effect total enrollment has on average income is not obvious.

$\beta_{GRADRATE1} > 0$. As the graduation rate increases, AVGINC2 increases, because schools that have high graduation rates indicate that more of their students have the motivation and resources to be able to complete a bachelor's degree program, holding all other independent variables constant. However, increases in graduation rate might occur as a result of the impact of other factors, such as family income or peer effects. Graduation rate could increase as a result of more able students attending the school, but it also can increase as a result of social aspects, such as students feeling pressure to graduate with all of their friends.

Financial Aid Variables

$\beta_{NETTUITION} > 0$. As the net tuition increases, AVGINC2 increases, because more expensive schools are often thought to be of higher quality schools, and students who attend higher-quality, more selective schools will earn more than students who do not, holding all other independent variables constant. NETTUITION can also be seen as a proxy for family income. Previous studies suggest that family income is positively correlated with better labor market outcomes, as families with higher incomes may be able to provide their children with more connections, resources and ability to obtain higher paying jobs. NETTUITION can be thought of a proxy for family income in the sense that families with higher incomes can afford schools with high tuition. Thus, as NETTUITION increase, AVGINC2 increases because family income is positively correlated with NETTUITION and AVGINC2, holding all other independent variables constant.

College Type Variables

$\beta_{PROFIT} < 0$. If the school is a for-profit school, the AVGINC2 decreases, because according to Baily, Badway and Gumport (2001), many educators believe that for-profit institutions are less committed to the humanistic educational objectives of higher education, and instead are focused on the market transactions the institution creates.[61] Therefore, students who attend a for-profit university are not receiving an education that is of a comparable quality of students who attend not-for-profit institutions. Students who receive higher-quality education will earn more than students who do not because they will be better equipped with a valuable skill set that is rewarded in the labor market, holding all other independent variables constant.

β_{PUBLIC} ? 0. If the college is a public institution, it is not obvious whether or not the AVGCINC1 increases or decreases. Some of the highest quality schools in the United States are public universities. With this in mind, it would make sense that if the school is a public university, then AVGINC2 would increase as well because higher quality and

[61] Bailey, Thomas, Norena Badway, and Patricia J. Gumport. "For-Profit Higher Education and Community Colleges." (2001).

more selective institutions are expected to have significant and positive returns, holding all other independent variables constant. Yet in the study conducted by Brewer, Eide, and Ehrenberg (1999), they found a large return to attending an elite private institution, and weak evidence to suggest a return to attending an elite public university.[62] They did not provide a specific reason for this difference in labor market returns, thus there is no basis to determine the direction of the effect that public universities have on average income. Furthermore, as public universities suffer from being under-funded by state governments, these institutions might not have as much money at their disposal to provide students with the highest quality resources as the more expensive private schools do.

β_{TYPE} ? 0. Previous research does not indicate if the type of school will increase or decrease AVGINC2, perhaps due to lack of data and research on this specific subject. One hypothesis if the school is an online-only university, AVGINC2 decreases, because online universities are non-selective institutions. More ambitious and motivated students may want to attend highly selective schools. These student characteristics are highly rewarded in the labor market and would thus create a significant and positive return to attending a highly selective school. On the other hand, if the school is an online-only university, AVGINC2 increases, because students who attend online universities can be older students who already have a job and need to complete a bachelor's degree program to advance in the labor market. Online university students, in this case, are already earning an income and are therefore more likely to have higher incomes on average than

[62] Brewer, Eide, Ehrenberg

recent graduates of traditional universities. Completing an online degree requires significant motivation and persistence, characteristics that will also be rewarded in the labor market.

Selectivity Variables

$\beta_{MOSTCOMP} > 0$. If the school is a "most competitive" school, the AVGINC1 increases, because most selective schools have a high level of the faculty, and a better output. Furthermore, if more-able students attend better and more selective schools, capturing selection bias, average income increases as ambition and motivation are student characteristics that are rewarded in the labor market. Previous literature suggests that there are significant and positive returns to college selectivity. However, the research by Dale and Kreuger (2011), Black and Smith (2006), and others stress the importance of including multiple measures of college quality. As the level of selectivity provides a single vantage point on college quality, it is unlikely that colleges only have a single quality dimension.[63]

$\beta_{HIGHLYCOMP}$, $\beta_{VERYCOMP}$, β_{COMP}, $\beta_{SPECIAL} > 0$. If the school is a "highly competitive, very competitive, competitive or special" school, the AVGINC2 increases, because at the highly competitive schools, students are still receiving a high-quality education experience, as highly competitive schools have a high level of the faculty, and a better output, holding all other independent variables constant.

[63] Black and Smith

$\beta_{LESSCOMP} < 0$. If the school is a "less competitive" school, the AVGINC2 decreases, because less competitive schools have lower levels of faculty and student output, holding all other independent variables constant.

Regression Controls

Previous research suggests that the quality of the college the student attends plays a significant role in the student's future earnings. The research conducted by Dale and Krueger (2011) supports the claim that student ability strongly influences determine successful the student will be as a participant in the labor force, rather than the characteristics of the college the student attends.[64] Whereas the research conducted by Zhang, Monks and Brewer, Eiden and Ehrenberg reinforces the idea that there are significant and positive returns to college quality and selectivity. Highly selective institutions might provide more labor market preparation given their access to more resources and overall learning experience. To control for college quality each school's student-to-faculty ratio, and college type variables will be used to proxy variables. *Barron's* index of college selectivity will serve as measures of college selectivity. The index consists of seven categories: most competitive, highly competitive, very competitive, competitive, less competitive, non-competitive and special.

However, if more motivated and ambitious students attend the more selective schools, selection bias could infiltrate the results.[65] These valuable characteristics are often rewarded in the labor market and can greatly impact a student's future earnings. Furthermore, as Dale and Krueger note, the same valuable characteristics that admissions

[64] Dale and Krueger
[65] Dale and Krueger

officers are looking for when selecting students for their college are the same to traits that employers are seeking when hiring and promoting workers.[66] Therefore, as Dale and Krueger (2011) discuss in their research, student ability should be accounted for within the regression. Dale and Krueger (2011) also state that neglecting to include a measure of student ability will yield biased estimated coefficients and render inconclusive results. Unfortunately, it is not possible to quantify such personal qualities. To correct for the bias in the estimated coefficients caused by omitting a variable that measures student ability, most previous studies, such as Dale and Kreuger (2011), used the average SAT score of the freshmen class of the colleges each student applied to as a proxy for student ability.[67] More able students may choose to apply and to attend schools with a freshmen class that have higher average SAT scores. However, using average SAT scores, the most common proxy for ability, is not available nor useful in this case, since the online universities have no SAT requirement in order for a student to be accepted. Moreover, an average SAT proxy would not be viable in this model as the data is not individual level data. In order to attempt to control for ability in this model, the graduation rate for each school will be used. According to Wooldridge, in order for a proxy variable to be viable, the proxy must be redundant, and uncorrelated with the omitted variable and each of the other regressors.[68] It is always assumed that a proxy satisfies the redundancy condition; however, it does not always satisfy the second property. [69] If the proxy turns out to be correlated with one or more of the independent variables, then the proxy is imperfect.[70] An OLS with an imperfect proxy still yields inconsistent estimated coefficients. The

---

[66] Dale and Krueger
[67] Dale and Krueger
[68] Wooldridge
[69] Wooldridge
[70] Wooldridge

hope is that the bias on each of the regressors is smaller in magnitude than if the proxy was omitted from the linear regression.[71] As it turns out, graduation rate is irrelevant in explaining wage once the unobserved student characteristics have been controlled for. By definition, ability and ambition affect wage, thus the school's graduation rate would not matter if true ability or motivation were known. However, the concern that the graduation rate does not control for all of the bias remains.

Net tuition is also included in this model in as a measure of the cost-effectiveness, in addition to college quality, of a bachelor's degree. In order to obtain net tuition, this model employs the same methodology as Dale and Krueger in their 2011 paper, and subtract the average amount of aid received by undergraduate students from the tuition sticker price. Net tuition may also be considered a proxy for family income, as this model does not employ any other controls for this variable. Family income, as demonstrated in Monks's research, serves as a proxy for a student's ability to pay for school. Net tuition may be correlated with family income, as family income increases the family's ability to pay for a high cost school increases as well. Additionally, family income might have a positive effect on income as it may be correlated with better labor market outcomes. Families with higher incomes may have access to more job opportunity connections, so it is not necessarily all of the school's resources and characteristics influencing a student's ability to get a job but now the family connections have an effect as well.

Initial Hypothesis

Results from previous research, such Black and Smith (1998), Monks (1999) and Brewer, Eiden, and Ehrenburg (1999), suggest that college type has a positive and

[71] Wooldridge

significant impact on income. In other words, it pays to attend a highly selective, private institution. In this model, college type is defined as online or traditional university. As online universities today are considered non-competitive institutions, previous research would suggest that students who attend traditional universities have higher incomes than students who attend online universities. The results do not suggest significant returns to college type. TYPE was found to be statistically insignificant, supporting the findings of Dale and Krueger (2011). The initial hypothesis predicts that TYPE would have a significant negative impact on income; however, the results do not suggest that college type has any significant impact or effect on income. On the other hand, college selectivity variables, including MOSTCOMP, HIGHLYCOMP and VERYCOMP, were found to be statistically significant, consistent with the findings of previous research.

RESULTS

Analysis

TYPE is statistically insignificant, meaning the type of college a student attends, online versus traditional, has no significant impact on income, holding all other independent variables constant. Only seven variables out of the initial regression were found to be statistically significant (Table 2). The results suggest strong, positive returns to college selectivity and quality. In addition, students who graduate from schools located in the eastern region of the United States have higher average incomes than students who graduate from schools located in the mid-western or southern region of the country. 20.7% of the variance in the dependent variable is explained by the variance in the independent variables (Table 2).

F-test for Overall Statistical Significance

The F-test for overall significance provides more statistically concrete evidence to describe the accuracy of the regression, as it is a formal hypothesis test for the overall fit of the model.[72] The F-test determines whether or not as a group the independent variables have a statistically significant effect on the dependent variable. The test is conducted by taking the explained sum of squares divided by the number of independent variables in the regression, that quantity is then divided by the residual sum of squares divided by the degrees of freedom.[73] If the resulting number is greater than the critical value, then we can conclude that there is a linear relationship between the dependent variable and the independent variables. On the other hand, if the calculated f-statistic is less than the critical value, then there is no linear relationship. The null-hypothesis, H_0 states that the estimated coefficients are not statistically different from zero; whereas the alternative hypothesis, H_1, states the estimated coefficients are statistically different from zero. The calculated f-statistic for the model is 14.189 (Table 2). As the calculated F-statistic is larger than the critical value, the null hypothesis is rejected, and therefore the group of estimated coefficients is statistically different from zero. Thus, the f-statistic concludes that there is a linear relationship between average income and the independent variables.

T-tests for Statistical Significance:

The t-test for significant provides a formal hypothesis test to determine whether or not the individual independent variable has a statistically significant impact in the value

[72] Studenmund, A. H. Using econometrics: A practical guide 5th Edition. Vol. 321311558. Addison Wesley, New York) ISBN, 2005.
[73] Studenmund

of the dependent variables. The test shows whether or not there is a large enough difference between the value of the estimated coefficient and zero, relative to the spread of the estimated coefficient, to confirm a linear relationship between the individual independent variable and the dependent variable.[74] The t-statistic is calculated by dividing the standard error into the difference of the estimated coefficient from zero.[75] If the calculated t-statistic is greater than the critical value, then we can conclude that there is a statistically significant relationship between the independent variable and AVGINC1.

STUFACR, NETTUITION, MOSTCOMP, HIGHLYCOMP, VERYCOMP, MIDWEST AND SOUTH were all found to be statistically significant (Table 2).

Interpretations on Coefficients:

Only the variables found to be statistically significant, as well as the coefficient on TYPE are further discussed. The remaining estimated coefficients were found to not be statistically significant from zero, and thus have no statistical impact on the dependent variable.

For every one student increase in the student-to-faculty ratio, average income increases by 0.0738% (Table 2), holding all other independent variables constant. Student-to-faculty ratio, similarly to Black and Smith (2006) serves as a measure of college quality. Student-to-faculty ratio can additionally serve as a proxy for school resources. STUFACR has a significant and positive relationship with average income. However, this result is unexpected student-to-faculty ratio most often has a negative relationship with quality, meaning schools with higher ratios are considered lower quality

[74] Studenmund
[75] Studenmund

schools. Yet this result is not very worrisome as a one student increase in the ratio only accounts for a small change in average income, holding all other independent variables constant.

Net tuition, the other measure of college quality, was also statistically significant. For every one dollar increase in net tuition, the average income of graduates from each school increases by 0.0000647% (Table 2), holding all other independent variables constant. Thus, if net tuition increases by $10,000, then the average income of graduates will increase by 0.647%, holding all other variables constant. More expensive schools should provide more resources to improve the quality of education than less expensive schools, as is often assumed, in accordance with basic economic principles that higher prices reflect higher quality products. Thus, that if the price of the education is higher, than the quality of the education must be high as well. As a result, schools with higher tuition are able to pay for more expensive teachers, better equipment, provide more library resources and services, and other resources so that the college can provide a higher quality college education for their students. Students who receive a higher quality education will be better prepared when entering the labor market. However the magnitude of the estimated coefficient suggests that students who attend more expensive schools will only be marginally rewarded in the labor market. For significant increase in tuition, $10,000 for example, average income only increases slightly, suggesting that while this variable has a statistical impact on average income, attending more expensive schools does not have a large effect on average income. As NETTUITION served as a proxy for college quality and family income in this model, this result does not fully imply large, positive returns to college quality and family income.

This model only included two measures of college quality and neglected to include an average SAT variable as a proxy for quality due to lack of data on the online universities as well as the data being institutional level rather than individual level data. While the model includes more than one proxy for quality, as Black and Smith (2006) and Zhang (2005) note failing to adopt multiple proxies for quality may result in an underestimation of the wage effect of college quality and type.[76] On the other hand, as both of these independent variables serve as measures of college quality, the results imply that there are in fact significant and positive returns to attending higher quality institutions.

If the school is an online-only university, the average income of graduates from each school decreases by 5.45% (Table 2), holding all other independent variables constant. Although the t-test for significance indicates that college type has no effect on average income. The data may lead to inconclusive results because of having an extremely limited sample of online-only institutions. A limited sample of online universities causes the confidence intervals to be very wide. In order for there to be an effect of college type, there would have to be a difference in incomes of students attending online versus traditional universities of almost $10,000 (Table 3), suggesting college type has no effect on average income, as such a large difference in incomes between online and traditional universities is unlikely.

Furthermore, the results suggest significant returns to attending more selective schools, as the estimated coefficients on the three most-selective ranking categories collectively were statistically significant and positive. For example, if the school is a most competitive school, the average income of graduates increases by about 7.1%

[76] Black and Smith

(Table 2), holding all other independent variables constant; if the school is a highly competitive school, the average income of graduates increases by about 4.2% (Table 2), holding all other independent variables constant; finally, if the school is a very competitive school, the average income of graduates increases by about 1.8% (Table 2), holding all other independent variables constant. As the level of selectivity increases, the wage effect of selectivity increases as well. However, less competitive schools have no effect on increasing or decreasing the average income of graduates. These results suggest that students who attend more selective schools, ranging from very competitive to most competitive, will have on the average a higher income than students who attend less competitive schools.

However, selectivity bias within the model created by an omitted variable bias could, in this case, be artificially inflating the overall effect of college type has on average income. More students might attend more selective schools. The characteristics of higher quality, such as ambition, motivation, and overall intelligence, are the same characteristics that are greatly valued by employers and are thus rewarded in the labor market. In other words, highly motivated and ambitious students are more likely to become highly motivated and productive workers. Yet this model does not account for these student characteristics within the model. Thus, this unobservable variable that describes student ability is excluded from this model. Unfortunately, the omission of this relevant independent variable results in biased estimated coefficients on the other independent variables. If the estimated coefficients are biased, the results are misleading. In this case, we expect the direction of the bias on the other estimated coefficients to be positive, as the direction of the expected value of the relationship between the included independent variables and student ability (positive) is multiplied by the direction of the coefficient on student ability (positive).[77] Thus, omitting student ability would then increase the overall effect college selectivity has on average income.

Regression Testing:

Heteroskedasticity:

Heteroskedasticity is the violation of the assumption of homoscedasticity, meaning that the observations of the error term are drawn from a distribution that has a constant variance.[78] If the assumption of homoscedasticity is met, then it can be assumed that all of the observations of the error term are being drawn from the same distribution. With pure heteroskedasticity, the variance of the error term is not constant; meaning the variance of the distribution of the error term depends on a specific observation.[79] Although heteroskedasticity does not create bias in the estimated coefficients, OLS is now not the minimum-variance estimator as well as it creates bias within the standard errors. Bias in the standard errors results in unreliable hypothesis testing. [80]

In order to detect heteroskedasticity within the model, scatter plots of each non-dummy independent variable are plotted against the dependent variable. This provides an informal method to determine if one or more independent variables is heteroskedastic, as we can see in the graphs whether or not the spread of the error term or the dependent variable changes as the values of the independent variables increase.[81] However there are two formal statistical tests to detect heteroskedasticity: the Park Test and the White Test.

---

[77] Studenmund
[78] Studenmund
[79] Studenmund
[80] Studenmund
[81] Studenmund

The Park Test looks to determine if there is heteroskedasticity in the model with respect to specific independent variables.[82]

The scatter plot graphs potentially indicate heteroskedasticity within the model due to STUFACR and ENROLLMENT (Graph 1 and 2). For both variables, as the independent variable increases, the spread of the error term decreases. In a homoscedastic model, the spread of the error term would remain constant as the independent variable increases.

The Park Test

The Park Test is conducted by regression the log of the sum of the squared residuals and the log of the independent variable. Two-tailed t-tests are then conducted on each of the coefficients. Heteroskedasticity exists within the model if and only if the calculated t-statistic is greater than the critical value.[83]

The estimated coefficients for the log of student-to-faculty ratio was found to be statistically significant (Table 9), while the estimated coefficient of the log of total enrollment was statistically insignificant (Table 10). Furthermore, the scatter plot graph (Graph 1) suggests that as STUFACR increases, the spread of the error term decreases. Thus, there is heteroskedasticity within the model with respect to STUFACR.

The White Test

The White Test determines whether or not there is heteroskedasticity with respect to one or more variables in the model. This test is conducted by conducting a one-sided

[82] Studenmund
[83] Studenmund

Chi-squared test of the sum of the squared residuals. Heteroskedasticity exists within the model if the calculated chi-squared value is greater than the critical value.[84]

However, because 26.23775 (Table 11) is less than 26.3, the chi-sqaured critical value, we can conclude that there is not heteroskedasticity with respect to one or more independent variables in the model. The model will still be corrected for heteroskedasticity as the park test revealed heteroskedasticity in the model with respect to STUFACR.

Multicollinearity:

Mulicollinearity is a violation of one of the classical assumptions, which states that the explanatory variables are not perfect linear functions of each other.[85] In other words, if perfect multicollinearity exists, then the variation in one independent variable can be completely explained by movements in a different independent variable.[86] On the other hand, imperfect multicollinearity occurs if there is a linear relationship between two or more independent variables that significantly affects the estimation of the coefficients of the variables.[87] Multicollinearity creates bias in the standard errors of the coefficients, and thus leads to unreliable hypothesis testing.

Two methods commonly used to determine if multicollinearity exists in the model is to look at the simple correlation between two independent variables, partial correlation, and the variance inflation factor test (VIF).

Simple Correlation

[84] Studenmund
[85] Studenmund
[86] Studenmund
[87] Studenmund

The correlation matrix gives the simple linear correlations between each individual explanatory variables and the dependent variable, as well as the explanatory variables with the each other explanatory variable. The correlation matrix is used to detect multicollinearity. If an independent variable is highly correlated with the dependent variable but even more highly correlated with another independent variable, then multicollinarity may exist within the model. The correlation matrix (Table 12) does not suggest any evidence of multicollinearity within the model.

Variance Inflation Factor

If multicollinarity is apparent in the model, the standard errors of the coefficients are artificially inflated. Calculating the Variance Inflation Factor will illustrate how much the standard errors are actually increased. To calculate the VIF, the difference between 1 and the R^2 from a regression run with just the independent variable against the dependent variable, that quantity divided by 1.[88] If the calculated VIF has a value of five or greater, then multicollinearity exists in the model with respect to that independent variable.

None of the Variation Inflation Factors (Table 14) have values greater than five, and thus do not suggest multicollinarity exits in the model.

Interactive Dummy Variable and F-test on a Subset of Coefficients

The independent variables included in the regression are broken up into categories of college quality, type, enrollment, selectivity and region. Even though not every, or none of the variables included in a subset of coefficients are statistically significant, the subset as a group may have a linear relationship with average income. In order to

[88] Studenmund

determine whether or not there is a relationship between the subset and average income, a regression must be run excluding the variables in a single subset, giving the $RSS_{Restricted}$. The RSS from the original regression, including the subset, is subtracted from the $RSS_{Restricted}$, divided by the number of variables in the subset, and that quantity is divided by the original RSS divided by the degrees of freedom. A F-test is then performed, if the null hypothesis is rejected, then we can conclude that the coefficients of the subset are statistically different zero, and there is a significant linear relationship between the subset and the dependent variable.

The regional variables were placed in the regression as an attempt to adjust for differences in costs of living throughout the different regions of the United States. The regional variables themselves account for differences in income, the dependent variable. However, net tuition might be dependent on region, as regions with higher costs of living may charge higher tuition per student. A series of interactive dummy variables, interacting each region with net tuition, were tested to attempt to adjust for the difference in net tuition based on region. None of the interactive terms were significantly different from zero (Table 4, 5, 6). In order to determine whether or not the interaction term should remain in the regression, even though it was found to be statistically insignificant, an F-test on a subset of coefficients including TYPE and the interaction terms, WEST*NETTUITION, MIDWEST*NETTUITION, and SOUTH*NETTUITION, was run. The results of the f-test led to a rejection of the null hypothesis, and to conclude that there is a significant linear relationship between the subset of coefficients and average income.

Final Regression

For the final regression, all insignificant variables were removed from the regression (Table 15). However TYPE was kept in the regression, as it is the variable this model is focusing on. The regression was again corrected for heteroskedasticity. Excluding all statistically insignificant variables did not yield large changes in the significant estimated coefficients. All of the estimated coefficients of the statistically significant variables but STUFACR increased very slightly, but remained essentially unchanged. The estimated coefficient on TYPE, on the other hand, decreased to -0.009122 (Table 15), suggesting that the magnitude of the effect TYPE has on AVGINC2 is almost zero. Even though WEST was found to be statistically insignificant, it remains in the final regression as the group of regional variables has a statistically significant impact on AVGINC2.

Removing the statistically insignificant variables created almost no changes in the regression. Thus we can conclude that their place in the regression is irrelevant. Furthermore, the estimated coefficient on TYPE getting close to zero in the final regression confirms our hypothesis that TYPE has no impact on average income.

Net Present Value of Attending College

While the results from this model suggest that college type has no statistical wage effect, college type may still have an impact on income. Online universities cost, on average, $6256.20 less per year in net tuition than traditional universities (Table 15). As the model predicts that college type has no statistical effect on average income, it is important to determine whether or not students earn extra income by attending a traditional university, despite the higher tuition costs.

In order to determine whether or not traditional university students earn additional income even with higher tuition costs, the net present value of the benefits of attending a traditional university as opposed to an online university is thus calculated. If the net present value of the benefits of attending a traditional university versus an online university is positive, then the additional income students earn outweighs the additional tuition costs they must pay in order to attend a traditional university. Net present value is calculated by

$$NPV = PV(Benefits) - PV(Costs)$$

where the PV(Benefits) represents the present value of an annuity with the estimated coefficient on TYPE (Table 3), and PV(Costs) represents the present value of the costs of attending a traditional university versus an online university, the difference in net tuition for traditional and online universities.

To obtain the present value of the costs of attending a traditional university as opposed to an online university was calculated by

$$PV(Costs) = \sum_{t=1}^{4} \frac{(difference\ in\ net\ tuition)}{(1+i)^t}$$

The present value of the benefits of attending a traditional university versus an online university, meaning the difference in average income between each type, was calculated by treating the difference in average income as an annuity.

$$PV(Benefits) = \frac{\widehat{\beta_{TYPE}}}{i} * \left(1 - \frac{1}{(1+i)^{45}}\right) * \frac{1}{(1+i)^4}$$

The estimated coefficient on TYPE, for the purpose of this exercise, was taken from the regression output using the average value of each independent variable for each type of school (Table 3) as a measure of the difference in average income earned by students who attend traditional universities versus online universities. Students who attend traditional universities earn $2565.37 (Table 15) more than students who attend online universities, holding all other independent variables constant. For the purpose of this calculation, it is assumed that students graduate and earn their degree in 4 years, and are all 22 years old when they graduate. Moreover, it is assumed that students work from ages 22 until the standard retirement age, 67 years old. Average income is assumed to remain constant for the students working life.

For both calculations, *i* represents the discount rate. Two different discount rates were used: a discount rate that takes into account the rate of return of schooling, i_1, and the return on a financial portfolio of a blend of stocks and bonds, i_2. The rate of return to education for an average worker typically ranges from 5-12%.[89] Thus, the value of i_1 was calculated by taking the average of 5% and 12%. The rate of return on stocks and bonds was calculated using a blend of 60% stocks and 40% bonds, with the rate of returns on stocks and bonds coming from historical prices and returns on the S&P 500 and 10-year US Treasury Bonds from 1970 to 2013 (Table 15) respectively.

Net present value was then calculated by subtracting the present value of the cost of attending college from the present value of the benefits of attending college. The resulting net present value of the benefits of attending a traditional university, when using a discount rate of the returns to college tuition, was $9,687.35 (Table 15). If the rate of return on a financial portfolio of stocks and bonds is used, the net present value of the benefits of attending a traditional university is $12,928.74 (Table 15), suggesting that

[89] Ehrenberg, Ronald G., and Robert S. Smith. "Modern labor economics." (2010).

students who attend traditional universities earn more additional income than online university students, despite higher tuition costs.

Thus, students who attend traditional universities will experience greater labor market returns, even after paying higher tuition. Therefore, online universities cannot serve as a cost-effective substitute for traditional university. Instead, the results imply that even though traditional universities have, on average, higher costs, in the long run it is worth it to pay almost double the amount per year in tuition costs to attend a traditional university.

CONCLUSION

While college type, defined as online-only versus traditional university, even after the final run, has no statistical impact on average income, the results suggest that college selectivity has a strong, significant and positive impact on average income. Furthermore, the model finds significant and positive returns to college quality. However, while the proxies for college quality, student to faculty ratio and net tuition, were found to be statistically significant, the magnitude of the effect these variables have on average income is minimal. Large changes in both of the college quality variables only yield marginal changes in average income. Perhaps this effect would have been stronger if more proxies of college quality were included within the regression. Moreover, a better measure of student ability could have changed the results as well. Graduation rate, which was the variable initially expected to proxy for student ability and college quality, was also found to be statistically insignificant. As almost all of the previous literature finds strong, positive returns to student ability, finding and incorporating a statistically significant proxy for student ability would relieve the model of selection bias and provide accurate results. Unfortunately, due to lack of data for online universities as well as using institutional level data, finding a viable proxy for student ability proves to be difficult so the bias remains within the model.

The results ultimately do not support the claim that online universities cannot serve as a perfect substitute for a traditional university. While the results do not provide evidence of attending an online university having a positive and significant impact on students' average income, they neither suggest the presence of a negative impact on average income if the student attended an online university. However, the net present value calculation illustrates that despite higher tuition costs of traditional universities, attending a traditional university yields a higher net present value than an online university. Even though with respect to labor market outcomes, online universities may not be able to serve as a perfect substitute for a college education experience, it has the potential to with serious improvements in technology to create a better learning environment, trust in the working world, and more growth in the industry.

References

1. "Higher Education." *The White House*. The White House, n.d. Web. 03 Apr. 2014.
2. Dale, Stacy, and Alan B. Krueger. *Estimating the return to college selectivity over the career using administrative earnings data*. No. w17159. National Bureau of Economic Research, 2011.
3. Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. MIT press, 2010.

4. Zhang, Liang. "Do measures of college quality matter? The effect of college quality on graduates' earnings." *The Review of Higher Education* 28.4 (2005): 571-596.
5. Monks, James. "The returns to individual and college characteristics: Evidence from the National Longitudinal Survey of Youth." *Economics of Education Review* 19.3 (2000): 279-289.
6. Brewer, Dominic J., Eric R. Eide, and Ronald G. Ehrenberg. "Does it pay to attend an elite private college?." *Journal of Human Resources* 34.1 (1999).
7. Black, Dan A., and Jeffrey A. Smith. "Estimating the returns to college quality with multiple proxies for quality." *Journal of Labor Economics* 24.3 (2006): 701-728.
8. "PayScale Methodology." *PayScale Methodology*. N.p., n.d. Web. 08 Apr. 2014.
9. Bailey, Thomas, Norena Badway, and Patricia J. Gumport. "For-Profit Higher Education and Community Colleges." (2001).
10. Studenmund, A. H. *Using econometrics: A practical guide 5th Edition*. Vol. 321311558. Addison Wesley, New York) ISBN, 2005.
11. Ehrenberg, Ronald G., and Robert S. Smith. "Modern labor economics." (2010).

Appendix 1: Sample of the Data

Name	UNITID		Location	STUFA(	GRTYPI	GRTOTLT	EFFYLE\	EFYTOTLT(EF2012A)	Min	MAX	AVGINC1	AVGINC	GRADRATE1	GRADRATE	TUITION	UAGRNTA	NETTUITION
Colorado Technical University - Online	444158		CO	27	2	1390	2	39772	32013	92178	$62,095.00	$62,095.00	0.03494921	0.13979684	12368	4738	7630
American InterContinental University Online	445027		GA	42	2	2020	2	27100	33387	98949	66168	66168	0.07453875	0.29815498	14043	4919	9124
Johnson & Wales University Online	460349		RI	6			2	120		34746	73543	73543	0			3499	-3499
University of Phoenix Online	372213		AZ	41	2	53137	2	359464	36011	102813	69412	69412	0.14782287	0.59129148	9216	3751	5465
Strayer University	131803			15	2	18	2	2184	39533	114767	77150	77150	0.00824176	0.03296703	14850	1656	13194
Walden University	125231		CA	20			2	13823	30263	88562	59412	59412	0		10725	3045	7680
Capella University	413413		MN	28			2	13933	35287	98592	66939	66939	0		11952	3208	8744
American Sentinel University	460738		CO	13			2	1960	47182	86162	66672	66672	0		9120	11298	-2178
Ashford University	154022		CA	21	2	286	2	153446	29786	81731	55758	55758	0.00186385	0.00745539	9648	4970	4678
Brandman University	262086		CA	13			2	5878	25447	88084	56765	56765	0			4909	-4909
Columbia Southern University	450933		AL	82	2	125	2	25169	39791	89046	46800	46800	0.00496643	0.01986571	4800	2275	2525
American Public University System	449339		WV	23	2	16	2	89175	29590	90433	48100	48100	0.00017942	0.00071769	6000	3296	2704
Grantham University	442569		MO	17			2	14397	41788	119075	80431	80431	0		6360	1056	5304
Jones International University	444723		IL	46	2	4	2	4620	34817	100000	67408.5	67408.5	0.0008658	0.0034632	12720	3942	8778
Sullivan University	157793		KY	18	2	251	2	6449	36494	87793	62143	62143	0.03892076	0.15568305	17520	2516	15004
Western Governors University	433387		UT	41	2	108	2	34271	37204	100888	69046	69046	0.00315135	0.01260541	5780	4219	1561
Northcentral University	444130		AZ	2			2	680					0			4976	-4976
Harvey Mudd College	115409		CA	8	2	180	2	784	59643	134002	96822.5	73300	0.22959184	0.91836735	44159	24421	19738
United States Naval Academy	164155		MD	9		1190	2	4536	62798	161486	112142	77100	0.26234568	1.04938272	0		0
Stevens Institute of Technology	186867		NJ	9		483	2	2575	51800	134140	92970	64900	0.18757282	0.75029126	41670	21133	20537
Babson College	164580		MA	13		443	2	2015	42055	134876	88465.5	59700	0.21985112	0.87940447	41888	28569	13319
Princeton University	186131		NJ	6		1228	2	5327	34862	137506	86184	56100	0.23052375	0.92209499	38650	35654	2996
United States Military Academy	197036		NY	7		1256	2	4592	57348	179956	118652	74000	0.27351916	1.09407666	0		0
Stanford University	243744		CA	12		1646	2	7063	41596	133781	87688.5	61300	0.23304545	0.93218179	41250	36893	4357
Harvard University	166027		MA	7		1679	2	10564	36216	135408	85812	55300	0.15893601	0.63574404	41250	37239	4011
Brown University	217156		RI	8		1464	2	6435	36309	139011	87660	52300	0.22750583	0.91002331	42808	31476	11332
Massachusetts Institute of Technology	166683		MA	8		1001	2	4503	51634	153967	102800.5	68600	0.22229625	0.88918499	41770	32572	9198
Colgate University	190099		NY	9		743	2	2871	37622	127809	82715.5	51800	0.25879485	1.03517938	44330	34425	9905
Yale University	130794		CT	6		1313	2	5405	38491	142790	90640.5	50000	0.24292322	0.97169288	42300	39324	2976

Name	PROFIT	TYPE	PUBLIC	MOSTCOMP	HIGHLYCOMP	VERYCOMP	COMP	LESSCOMP	NONCOMP		SPECIAL	West	MIDWEST	SOUTH
Colorado Technical University - Online	1	1	0	0	0	0	0	0	1	1	0	1	0	0
American InterContinental University Online	1	1	0	0	0	0	0	0	1	1	0	0	0	1
Johnson & Wales University Online	0	1	0	0	0	0	0	0	1	1	0	0	0	0
University of Phoenix Online	1	1	0	0	0	0	0	0	1	1	0	1	0	0
Strayer University	1	1	0	0	0	0	0	0	1	1	0	0	0	0
Walden University	1	1	0	0	0	0	0	0	1	1	0	1	0	0
Capella University	1	1	0	0	0	0	0	0	1	1	0	0	1	0
American Sentinel University	1	1	0	0	0	0	0	0	1	1	0	1	0	0
Ashford University	1	1	0	0	0	0	0	0	1	1	0	1	0	0
Brandman University	0	1	0	0	0	0	0	0	1	1	0	1	0	0
Columbia Southern University	1	1	0	0	0	0	0	0	1	1	0	0	0	1
American Public University System	1	1	0	0	0	0	0	0	1	1	0	0	0	1
Grantham University	1	1	0	0	0	0	0	0	1	1	0	0	0	0
Jones International University	1	1	0	0	0	0	0	0	1	1	0	0	1	0
Sullivan University	1	1	0	0	0	0	0	0	1	1	0	0	0	1
Western Governors University	0	1	0	0	0	0	0	0	1	1	0	1	0	0
Northcentral University	1	1	0	0	0	0	0	0	1	1	0	1	0	0
Harvey Mudd College	0	0	0	1	0	0	0	0	0	1	0	0	0	0
United States Naval Academy	0	0	1	1	0	0	0	0	0	1	0	0	0	0
Stevens Institute of Technology	0	0	0	0	1	0	0	0	0	1	0	0	0	0
Babson College	0	0	0	0	1	0	0	0	0	1	0	0	0	0
Princeton University	0	0	0	1	0	0	0	0	0	1	0	0	0	0
United States Military Academy	0	0	1	1	0	0	0	0	0	1	0	0	0	0
Stanford University	0	0	0	1	0	0	0	0	0	1	0	1	0	0
Harvard University	0	0	0	1	0	0	0	0	0	1	0	0	0	0
Brown University	0	0	0	1	0	0	0	0	0	1	0	0	0	0
Massachusetts Institute of Technology	0	0	0	1	0	0	0	0	0	1	0	0	0	0
Colgate University	0	0	0	1	0	0	0	0	0	1	0	0	0	0
Yale University	0	0	0	1	0	0	0	0	0	1	0	0	0	0