ECON 21110
Applied Microeconometrics
Winter 2022
Lecture 2
Linear Regression Analysis
Application: Dale & Krueger (2002) College Quality

Eyðfríð Juanna Schrøter Joensen

University of Chicago

January 20, 2022

Application: Dale & Krueger (2002) College Quality

- Is it worth it to go to a higher quality college?
- If so, for whom is it worth it?
- Two measures of college quality:
  1. Private school, P_i
     $\rightarrow$ Is there a payoff to attending a private school?
  2. Average SAT score of admitted students, SAT_{j*}
     $\rightarrow$ Is there a payoff to attending a more selective college?

Application: Dale & Krueger (2002) College Quality

- Why is this a challenging empirical question?
- Students select college – at least in part – based on variables that may not be observed to us (the researchers)
- Illustrated with a simple model with three sequential choices:
  1. A student chooses which set of colleges to apply to
  2. Colleges independently decide whether to admit or reject the student
  3. Student (and parents) choose which college the student will attend from the set of colleges he/she was admitted to
- Key idea of Dale & Krueger (2002) relates to stage 2., taking stage 1. as given and assuming that 3. is essentially random

Application: Dale & Krueger (2002) College Quality

- Assume colleges determine admissions decisions by weighting various student attributes
- **Admission** of applicant i to college j follows the decision rule:

$$\begin{aligned} Z_{ij} &= 1\left[Z_{ij}^* > C_j\right] \\ &= 1\left[\gamma_1 X_{1i} + \gamma_2 X_{2i} + e_{ij} > C_j\right] \end{aligned} \tag{1}$$

- i is admitted to j if latent quality Z_{ij}^* high enough; i.e. if and only if $\gamma_1 X_{1i} + \gamma_2 X_{2i} + e_{ij} > C_j$
- **Observed**, X_{1i}: SAT scores, high school grade point average (GPA),...
- **Unobserved**, X_{2i}: assessment of student motivation, ambition, and maturity as reflected in essay, college interview, community service, extracurricular activity, and letters of recommendation
- e_{ij} luck and other idiosyncratic i.i.d. factors
- C_j cut-off quality level college uses for admission
  $\rightarrow$ more selective colleges have higher C_j

Application: Dale & Krueger (2002) College Quality

- The population model linking college selectivity to monetary labor market returns:

$$log(Y_i) = \beta_0 + \beta_1 SAT_{j*} + \beta_2 X_{1i} + \beta_3 X_{2i} + V_{ij} \tag{2}$$

- We want to estimate the causal effect on earnings of attending a more selective college, β_1
- OLS unbiased under MLR.1-MLR.4; i.e. particularly if $\mathbb{E}[V_{ij}|SAT_{j*}, X_{1i}, X_{2i}] = 0$
- OLS estimator $\widehat{\beta}_1$ is consistent if $\epsilon_{SAT_{j*}}$ is uncorrelated with V_{ij} and unbiased under mean independence, MLR.4': $\mathbb{E}\left[V_{ij}\middle|\epsilon_{SAT_{j*}}\right] = 0$, where $\epsilon_{SAT_{j*}}$ is the error term in the regression of SAT_{j*} on all the other explanatory variables included in X_{1i} *and* X_{2i}

Application: Dale & Krueger (2002) College Quality

- Previous studies only observed X_{1i} and estimated:

$$log(Y_i) = \beta_0 + \beta_1 SAT_{j*} + \beta_2 X_{1i} + U_{ij} \tag{3}$$

- Omitting X_{2i} implies that the OLS estimator $\widetilde{\beta_1}$ will be upwards biased

$$\begin{aligned}\mathbb{E}[log(Y_i)|SAT_{j*}, X_{1i}] =& \beta_0 + \beta_1 SAT_{j*} + \beta_2 X_{1i} \\ & + \mathbb{E}[U_{ij}|\gamma_1 X_{1i} + \gamma_2 X_{2i} + e^*_{ij} > C^*_j, X_{1i}]\end{aligned}$$

because $\mathbb{E}[U_{ij}|\gamma_1 X_{1i} + \gamma_2 X_{2i} + e^*_{ij} > C^*_j, X_{1i}] > 0$ as students who were admitted to, and are more likely to attend, more selective schools must have a higher value of unobservables according to (1)

Application: Dale & Krueger (2002) College Quality

- Similarly, the population model linking college quality to monetary labor market returns:

$$log(Y_i) = \beta_0 + \delta_1 P_i + \beta_2 X_{1i} + \beta_3 X_{2i} + V_{ij} \tag{4}$$

- We want to estimate the causal effect of attending a private school, δ_1
- Recall δ_1 gives the difference in conditional means by private school status:

$$\begin{aligned}\delta_1 &= \mathbb{E}[log(Y_i)|P_i = 1, X_{1i}, X_{2i}] - \mathbb{E}[log(Y_i)|P_i = 0, X_{1i}, X_{2i}] \\ &= \mathbb{E}[log(Y_{1i})|P_i = 1, X_{1i}, X_{2i}] - \mathbb{E}[log(Y_{0i})|P_i = 0, X_{1i}, X_{2i}]\end{aligned}$$

- OLS unbiased under MLR.1-MLR.4; i.e. particularly if $\mathbb{E}[V_{ij}|P_i, X_{1i}, X_{2i}] = 0$
- OLS estimator $\widehat{\delta_1}$ is consistent if ϵ_{P_i} is uncorrelated with V_{ij} and unbiased under mean independence, MLR.4': $\mathbb{E}\left[V_{ij}|\epsilon_{P_i}\right] = 0$, where ϵ_{P_i} is the error term in the regression of P_i on all the other explanatory variables included in X_{1i} *and* X_{2i}

Application: Dale & Krueger (2002) College Quality

- To illustrate: What are we estimating if we compare the average earnings between the two groups of students who attend private and public schools?

$$\begin{aligned}
&\underbrace{\mathbb{E}\left[log(Y_{1i}) \mid P_i = 1\right] - \mathbb{E}\left[log(Y_{0i}) \mid P_i = 0\right]}_{\text{observed difference in average earnings}} \\
= \;& \mathbb{E}\left[log(Y_{1i}) \mid P_i = 1\right] - \mathbb{E}\left[log(Y_{0i}) \mid P_i = 1\right] \\
& + \mathbb{E}\left[log(Y_{0i}) \mid P_i = 1\right] - \mathbb{E}\left[log(Y_{0i}) \mid P_i = 0\right] \\
= \;& \underbrace{\mathbb{E}\left[log(Y_{1i}) - log(Y_{0i}) \mid P_i = 1\right]}_{\text{average treatment on the treated (TT)}} \\
& + \underbrace{\mathbb{E}\left[log(Y_{0i}) \mid P_i = 1\right] - \mathbb{E}\left[log(Y_{0i}) \mid P_i = 0\right]}_{\text{selection bias}}
\end{aligned}$$

- Need to make sure that Y_{0i} observed for those who attend public $P_i = 0$ (or less selective) schools is a good counterfactual for Y_{0i} for those who attend private $P_i = 1$ (or more selective) schools

Application: Dale & Krueger (2002) College Quality

- **Key idea**: If, conditional on admission, students choose to attend schools for reasons that are independent of X_{2i} and V_{ij}, then students who were accepted and rejected by the same set of schools have the same value of U_{ij}
- Even if the researcher does not observe X_{2i}, then college admission administrators have evaluated students and independently reached the same conclusions regarding their abilities, ambitions, and motivation
- **Solution 1.**: include an unrestricted set of dummy variables indicating students receiving the same admissions decisions
- This is essentially **matching**

Application: Dale & Krueger (2002) College Quality

TABLE 2.1
The college matching matrix

		Private			Public			
Applicant group	Student	Ivy	Leafy	Smart	All State	Tall State	Altered State	1996 earnings
A	1		Reject	Admit		Admit		110,000
	2		Reject	Admit		Admit		100,000
	3		Reject	Admit		Admit		110,000
B	4	Admit			Admit		Admit	60,000
	5	Admit			Admit		Admit	30,000
C	6		Admit					115,000
	7		Admit					75,000
D	8	Reject			Admit	Admit		90,000
	9	Reject			Admit	Admit		60,000

Note: Enrollment decisions are highlighted in gray.

Application: Dale & Krueger (2002) College Quality

The six columns in Table 2.2 on the next slide represent the following specifications:

1. $log(Y_i) = \beta_0 + \delta_1 P_i + U_{ij}$
2. $log(Y_i) = \beta_0 + \delta_1 P_i + \beta_2 SAT_i + U_{ij}$
3. $log(Y_i) = \beta_0 + \delta_1 P_i + \beta_{21} SAT_i + \beta_{22} log(PI)_i + \beta_2 X_{1i} + U_{ij}$
4. $log(Y_i) = \beta_0 + \delta_1 P_i + \sum_{j=1}^{150} \beta_{3j} Group_{ij} + V_{ij}$
5. $log(Y_i) = \beta_0 + \delta_1 P_i + \beta_2 SAT_i + \sum_{j=1}^{150} \beta_{3j} Group_{ij} + V_{ij}$
6. $log(Y_i) =$
   $\beta_0 + \delta_1 P_i + \beta_{21} SAT_i + \beta_{22} log(PI)_i + \beta_2 X_{1i} + \sum_{j=1}^{150} \beta_{3j} Group_{ij} + V_{ij}$

Application: Dale & Krueger (2002) College Quality

TABLE 2.2
Private school effects: Barron's matches

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.135	.095	.086	.007	.003	.013
	(.055)	(.052)	(.034)	(.038)	(.039)	(.025)
Own SAT score ÷ 100		.048	.016		.033	.001
		(.009)	(.007)		(.007)	(.007)
Log parental income			.219			.190
			(.022)			(.023)
Female			−.403			−.395
			(.018)			(.021)
Black			.005			−.040
			(.041)			(.042)
Hispanic			.062			.032
			(.072)			(.070)
Asian			.170			.145
			(.074)			(.068)
Other/missing race			−.074			−.079
			(.157)			(.156)
High school top 10%			.095			.082
			(.027)			(.028)
High school rank missing			.019			.015
			(.033)			(.037)
Athlete			.123			.115
			(.025)			(.027)
Selectivity-group dummies	No	No	No	Yes	Yes	Yes

Notes: This table reports estimates of the effect of attending a private college or university on earnings. Each column reports coefficients from a regression of log earnings on a dummy for attending a private institution and controls. The results in columns (4)–(6) are from models that include applicant selectivity-group dummies. The sample size is 5,583. Standard errors are reported in parentheses.

Application: Dale & Krueger (2002) College Quality

TABLE 2.2
Private school effects: Barron's matches

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.135	.095	.086	.007	.003	.013
	(.055)	(.052)	(.034)	(.038)	(.039)	(.025)
Own SAT score ÷ 100		.048	.016		.033	.001
		(.009)	(.007)		(.007)	(.007)
Log parental income			.219			.190
			(.022)			(.023)
Female			−.403			−.395
			(.018)			(.021)
Black			.005			−.040
			(.041)			(.042)
Hispanic			.062			.032
			(.072)			(.070)
Asian			.170			.145
			(.074)			(.068)
Other/missing race			−.074			−.079
			(.157)			(.156)
High school top 10%			.095			.082
			(.027)			(.028)
High school rank missing			.019			.015
			(.033)			(.037)
Athlete			.123			.115
			(.025)			(.027)
Selectivity-group dummies	No	No	No	Yes	Yes	Yes

Notes: This table reports estimates of the effect of attending a private college or university on earnings. Each column reports coefficients from a regression of log earnings on a dummy for attending a private institution and controls. The results in columns (4)–(6) are from models that include applicant selectivity-group dummies. The sample size is 5,583. Standard errors are reported in parentheses.

Application: Dale & Krueger (2002) College Quality

- What if there is heterogeneity (β_{i1} and δ_{i1}) students are better informed about their own quality of the college match and respond to it?

- **Solution 2.**: include controls for how many schools student applied to and the average quality of these schools

- This means that the 150 $Group_{ij}$ dummy variables in Table 2.2 will be replaced by the four control variables listed in the bottom rows of Table 2.3; i.e. average SAT score of schools applied and dummy variables for whether the student sent two, three or four+ applications

Application: Dale & Krueger (2002) College Quality

TABLE 2.3
Private school effects: Average SAT score controls

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.212 (.060)	.152 (.057)	.139 (.043)	.034 (.062)	.031 (.062)	.037 (.039)
Own SAT score ÷ 100		.051 (.008)	.024 (.006)		.036 (.006)	.009 (.006)
Log parental income			.181 (.026)			.159 (.025)
Female			−.398 (.012)			−.396 (.014)
Black			−.003 (.031)			−.037 (.035)
Hispanic			.027 (.052)			.001 (.054)
Asian			.189 (.035)			.155 (.037)
Other/missing race			−.166 (.118)			−.189 (.117)
High school top 10%			.067 (.020)			.064 (.020)
High school rank missing			.003 (.025)			−.008 (.023)
Athlete			.107 (.027)			.092 (.024)
Average SAT score of schools applied to ÷ 100				.110 (.024)	.082 (.022)	.077 (.012)
Sent two applications				.071 (.013)	.062 (.011)	.058 (.010)
Sent three applications				.093 (.021)	.079 (.019)	.066 (.017)
Sent four or more applications				.139 (.024)	.127 (.023)	.098 (.020)

Notes: This table reports estimates of the effect of attending a private college or university on earnings. Each column shows coefficients from a regression of log earnings on a dummy for attending a private institution and controls. The sample size is 14,238. Standard errors are reported in parentheses.

Application: Dale & Krueger (2002) College Quality

- The observed earnings premium to attending a private school seems to be driven by more ambitious and motivated students (i.e. students with higher X_{2i}) attending private schools
- What if we examine college selectivity as measured by average peer ability (aka SAT score) instead?
- This means that the explanatory variable of primary interest P_i in Table 2.3 will be replaced by SAT_{j*} in Table 2.4

Application: Dale & Krueger (2002) College Quality

TABLE 2.4
School selectivity effects: Average SAT score controls

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
School average SAT score ÷ 100	.109 (.026)	.071 (.025)	.076 (.016)	−.021 (.026)	−.031 (.026)	.000 (.018)
Own SAT score ÷ 100		.049 (.007)	.018 (.006)		.037 (.006)	.009 (.006)
Log parental income			.187 (.024)			.161 (.025)
Female			−.403 (.015)			−.396 (.014)
Black			−.023 (.035)			−.034 (.035)
Hispanic			.015 (.052)			.006 (.053)
Asian			.173 (.036)			.155 (.037)
Other/missing race			−.188 (.119)			−.193 (.116)
High school top 10%			.061 (.018)			.063 (.019)
High school rank missing			.001 (.024)			−.009 (.022)
Athlete			.102 (.025)			.094 (.024)
Average SAT score of schools applied to ÷ 100				.138 (.017)	.116 (.015)	.089 (.013)
Sent two applications				.082 (.015)	.075 (.014)	.063 (.011)
Sent three applications				.107 (.026)	.096 (.024)	.074 (.022)
Sent four or more applications				.153 (.031)	.143 (.030)	.106 (.025)

Notes: This table reports estimates of the effect of alma mater selectivity on earnings. Each column shows coefficients from a regression of log earnings on the average SAT score at the institution attended and controls. The sample size is 14,238. Standard errors are reported in parentheses.

Application: Dale & Krueger (2002) College Quality

- Table 2.5 assesses omitted variables bias to corroborate causal interpretation
  - What if we omitted own SAT score?
  - What if we omitted own parental income?

Application: Dale & Krueger (2002) College Quality

TABLE 2.5
Private school effects: Omitted variables bias

	Dependent variable					
	Own SAT score ÷ 100			Log parental income		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	1.165 (.196)	1.130 (.188)	.066 (.112)	.128 (.035)	.138 (.037)	.028 (.037)
Female		−.367 (.076)			.016 (.013)	
Black		−1.947 (.079)			−.359 (.019)	
Hispanic		−1.185 (.168)			−.259 (.050)	
Asian		−.014 (.116)			−.060 (.031)	
Other/missing race		−.521 (.293)			−.082 (.061)	
High school top 10%		.948 (.107)			−.066 (.011)	
High school rank missing		.556 (.102)			−.030 (.023)	
Athlete		−.318 (.147)			.037 (.016)	
Average SAT score of schools applied to ÷ 100			.777 (.058)			.063 (.014)
Sent two applications			.252 (.077)			.020 (.010)
Sent three applications			.375 (.106)			.042 (.013)
Sent four or more applications			.330 (.093)			.079 (.014)

Notes: This table describes the relationship between private school attendance and personal characteristics. Dependent variables are the respondent's SAT score (divided by 100) in columns

Application: Dale & Krueger (2002) College Quality

- The average earnings return to attending a more selective college is not statistically significant
- BUT Dale & Krueger (2002) show evidence that going to a more selective (or private) college pays off more for those with lower family income
- Conclusion:
  College match is important, but the impact of college quality is still not well understood – despite high willingness to pay for it
- Discussion: Are the identifying assumptions credible?