ONLINE APPENDIX: GOVERNMENT EXPENDITURE ON THE PUBLIC EDUCATION SYSTEM

CHAO FU[A], SHOYA ISHIMARU[B] AND JOHN KENNAN[A]

1. Data Details

All dollar values are measured in 2002 dollars.

1.1. **Household Data.**

1.1.1. *ELS 2002.* We determine the household type x based on the status in the 2002 interview. Household income x_1 is categorized into 5 groups: $20,000 or less, $20,001-$35,000, $35,001-$50,000, $50,001-$75,000, and $75,001 or more. In the model, each income category is assigned its median income level in our sample from the 2000 Census, as summarized in Table B1. The household education level x_2 is an indicator for the presence of at least one adult with some college education. Minority status x_3 is an indicator signifying that a student is not White or Asian.

Table B1: Discrete Income Levels

x_1	Income Range ($)	Median Income ($)	% in ELS sample
1	0-20,000	10,970	15.0
2	20,001-35,000	27,580	18.9
3	35,001-50,000	41,789	19.6
4	50,001-75,000	60,917	20.7
5	75,000-	103,636	25.8

We measure K-12 achievement k_1 based on the standardized math test test score and high school dropout status. If a person has not completed high school or obtained a GED certificate, we assign $k_1 = 1$. Otherwise we assign $k_1 = 2, \ldots, 5$ based on the 12th grade test score reported in the 2004 interview, with each quartile as a cutoff. The 12th grade test score is missing for 1,769 high school graduates, so we impute their test score using the 10th grade math and reading score, demographic characteristics, and the state of residence. The linear prediction model for the imputation is estimated using the 12,374 high school graduates whose 12th grade scores are available. The estimated linear model has an adjusted R^2 of 0.803.

Date: April, 2022.

[a:]University of Wisconsin and NBER.[b]Hitotsubashi University.

Table B2: College Choices

Enrollment pattern	Observations		Assigned o_2
Never enrolled in college	1,820	13.3%	0
Enrolled in college, no further info	182	1.3%	treated as missing
2-year college	3,483	25.5%	1
4-year public in-state	3,639	26.7%	2
2-year to 4-year public	883	6.5%	2
4-year public out-of-state	666	4.9%	3
4-year private	2,454	18.0%	3
2-year to 4-year private	516	3.8%	3
Total	13,643		

We determine college choice o_2 based on the college enrollment history reported in the 2006 and 2012 interviews. The type of the first college enrollment is reported in both 2006 and 2012 interviews, while type of the last college enrollment is reported in the 2012 interview. We assign $o_2 = 0$ if a person never enrolled in any college. For the other cases, we use the first college choice to determine $o_2 \in \{1, 2, 3\}$, except that we use the last college if a student first enrolls in a 2-year college and then goes to a 4-year college. We assign $o_2 = 1$ if a person enrolled in a 2-year college, $o_2 = 2$ if a person enrolled in a 4-year in-state public college, and $o_2 = 3$ if a person enrolled in a 4-year out-of-state public college or 4-year private college. Table B2 summarizes the number of observations for each actual enrollment pattern.

We use the degree completion status reported in the 2012 interview to determine college outcome k_2. For a 2-year college student, we assign $k_2 = 1$ if the student gets any degree or certificate and $k_2 = 0$ otherwise. For a 4-year college student, we assign $k_2 = 2$ if the student graduates with a Bachelor's degree and $k_2 = 0$ otherwise. We treat the college outcome as missing for a 2-year college student whose first enrollment is in 2010 or later and for a 4-year college student whose first enrollment is in 2007 or later, since it is likely that their final college history is incomplete. We use the outstanding college loan level in the 2006 interview to measure the annual borrowing level d during college. To compute the annual borrowing, we divide the outstanding debt by 2 for those who enrolled in college in 2004 or earlier. We use the outstanding debt as the annual borrowing for those who first enrolled in 2005.

To estimate the aid function, we use both self-reported aid information as well as the Pell grant received in the first year of college. For each college student, we first impute the aid amount $\tilde{A}$ as

$$\tilde{A} = \max\left\{(\text{Pell Grant}), \alpha \times (\text{Tuition})\right\},$$

where the fraction α is determined based on self-reported information. We assign $\alpha = 1$ if the report indicates that all college costs are covered by the aid, $\alpha = 2/3$ if at least half but not all of the cost is covered, $\alpha = 1/3$ if less than half of the cost is covered, and $\alpha = 0$ if none of the cost is covered. We then perform a Tobit regression of $\tilde{A}$ on (x, k_1) and tuition to estimate the aid function $A(\cdot)$, as described in detail in Section 2.1 below.

1.1.2. *ACS 2002.* We create two subsamples from ACS 2002. The first contains all primary school students, which we use to get information on the private primary school attendance rate. The second contains pairs of siblings, where a younger sibling goes to primary school and an older sibling goes to high school. We use the to get information on the joint distribution of private attendance at primary school level and at high school level. In these subsamples, we define household type x in the same way as in the ELS 2002 sample.

1.1.3. *Census of Population 2000.* We use the sample of women aged 35-40 and single men aged 37-42 to estimate the state-specific demographic distribution $F_s(x)$, the fertility rate $q_s(x)$, the federal income tax schedule $\tau_0(x_1)$, and the local tax schedule parameters $\left(\tau_s^a, \tau_s^b\right)$. As we have 20 discrete household types in our model, we estimate the demographic distribution $F_s(x)$ nonparametrically. We use the fraction of households with any child as the fertility rate $q_s(x)$ for a state-type cell with 400 or more observations. We run a regression of child dummy on state dummies and household type dummies in an additively separable specification to estimate the fertility rate for the other cells with fewer observations.

1.1.4. *Tax Schedules.* To estimate the tax schedules, we impute the federal and state income tax liabilities of each household using the Internet TAXSIM (v27) program provided by NBER. We use 2002 as the tax year, converting reported income from 2000 dollars into 2002 dollars. For households in the lowest income group, we substract the average SNAP benefit ($639; from 2005 ACS) from their federal tax liabilities. We obtain the federal income tax rate $\tau_0(x_1)$ of each income group as the ratio of the total federal tax liabilities to the total income of the group.[1] Table B3 presents the estimated federal tax schedule.

Similarly, we obtain state income tax rate τ_{js}^I of income group j in state s as the ratio of the total income tax liabilities to the total income of the group. We also take into account state and local taxes other than personal income taxes, assuming the proportional tax schedule. In particular, we

[1] Federal income tax refers to individual income tax and does not include social security payroll tax.

Table B3: Federal Tax Schedule

Income group	$\tau_0\,(x_1)$
\$0-20,000	-3.36%
\$20,001-35,000	5.04%
\$35,001-50,000	7.20%
\$50,001-75,000	9.23%
\$75,000-	17.51%

compute the ratio τ_s^N of the total revenue from taxes other than personal income and severance taxes (from Census of Governments 2002) to the total personal income of the state (from the U.S. Bureau of Economic Analysis) in state s.[2] We define $\tau_{js} = \tau_{js}^I + \tau_s^N$ as the total state tax rate of income group j in state s.

To fit into the tax formula, we perform a linear regression

$$\ln(1-\tau_{js}) = \ln(\tau_s^a) - \tau_s^b \ln\left(x_1^{(j)}\right) + \varepsilon_{js},$$

in each state s using observations $j = 1, \ldots, 5$, where $x_1^{(j)}$ is the group-specific income level defined in Table B1. The state tax shedule fitted by the regression is

$$\hat{\tau}_{js} = 1 - \tau_s^a \ \left(x_1^{(j)}\right)^{-\tau_s^b}.$$

We include the tax rate $\hat{\tau}_{3s}$ for the middle income group $j = 3$ and the progressivity parameter τ_s^b in government policy choices ψ_s. $(\hat{\tau}_{3s}, \tau_s^b)$ and (τ_s^a, τ_s^b) have a one-to-one mapping relationship.

1.1.5. *Other Data Sources.* We use IPEDS 2002 to compute college attendance costs. For the 2-year cost, we take a weighted average of in-state tuition and fees among 2-year public colleges in each state, where weights are in-state freshmen cohort size. For the public 4-year cost, we take a weighted average of in-state tuition and fees among 4-year public colleges in each state, where weights are in-state freshmen cohort size. For the private 4-year cost, we take a weighted average of out-of-state tuition and fees among 4-year public colleges and in-state tuition and fees among 4-year private colleges in the entire United States, where weights are out-of-state freshmen cohort size for public colleges and freshmen cohort size for private colleges.

We use the regional enrollment-weighted average private school tuition in SASS 2010-2011 for K-12 cost of attendance. Due to a rapid growth of tuition levels, we convert 2010 tuition levels into

[2] Severance tax will be included in endowment z_1; see Section 1.2.2 for details.

2002 levels using the following formula, instead of merely adjusting for inflation.

$$t_{2002,r} = t_{2010,r} \cdot \frac{\bar{t}_{2003}}{\bar{t}_{2010}} \cdot \left(\frac{\bar{t}_{1999}}{\bar{t}_{2003}}\right)^{1/4}$$

$t_{y,r}$ is a regional average in year y and $\bar{t}_y$ is a national average in year y. This method gives \$7,683 for NorthEast, \$4,062 for MidWest, \$6,331 for South, and \$6,844 for West.

1.2. **Government Data.**

Table B4: Policy Grid

Tax rate for middle income (%)	7.5	9.0	9.5	10.0	10.5	11.0	12.5	
Tax progressivity $\tau^b(\times 100)$	0.0	0.5	1.0	1.5	2.0	2.5		
K-12 exp (\$1,000/yr)	5.0	6.0	6.5	7.0	8.0	9.0	10.0	11.5
College exp (\$1,000/yr)	10.5	12.0	13.5	14.5	15.5	16.5	18.0	21.0
2-year tuition (\$1,000/yr)	0.0	1.4	1.9	2.3	2.7	3.7	4.7	
4-year tuition (\$1,000/yr)	0.0	2.0	2.8	3.4	4.0	4.8	6.5	8.0

1.2.1. *Policy Grid.*

1.2.2. *Census of Governments 2002.* Using Census of Governments 2002, we compute the data analogues of per-student public K-12 expenditure e_1, per-student public college expenditure e_2, and per-household endowment z_1. Per-student public K-12 expenditure e_1 is directly provided by the Census of Governments. For e_2, we divide the total higher education expenditure (excluding the component spent on auxiliary enterprises) by the total full-time-equivalent enrollment in public colleges (from IPEDS 2002). Finally, we divide the sum of the following revenue items by the total number of households (from 2002 ACS) to obtain per-household endowment z_1.

- Net liquor store revenue
- Net revenue from parking facilities
- Net interest earnings

- Net Intergovernmental revenue from federal government
- Misc general revenue: special assessments
- Misc general revenue: sale of property
- Misc general revenue: other general revenue
- Severance taxes

Although state and local governments also receive some non-tax revenue from public services such as hospitals, highways, and airports, we exclude these revenues, since spending on those services typically exceeds revenue. We include the severance tax in a state's endowment z_1, because it is a state tax imposed on the extraction of non-renewable natural resources that are intended for consumption in other states.

2. Estimation Details

2.1. **Financial Aid Function.** We estimate the financial aid function outside of the model. Using the imputed aid amount $\tilde{A}$ from the ELS 2002 data, we estimate a Tobit model

$$\tilde{A} = \max\left\{0, \mu^A_{o_2}\left(C, x, k_1\right) + \epsilon\right\}, \, \epsilon \sim N\left(0, \left(\sigma^A_{o_2}\right)^2\right),$$

for each college type $o_2 = 1, 2, 3$. We use the conditional mean of the Tobit model as the aid function $A_{o_2}(\cdot)$:

$$A_{o_2}(C, x, k_1) = \mu^A_{o_2}\left(C, x, k_1\right) \Phi\left(\frac{\mu^A_{o_2}\left(C, x, k_1\right)}{\sigma^A_{o_2}}\right) + \sigma^A_{o_2} \phi\left(\frac{\mu^A_{o_2}\left(C, x, k_1\right)}{\sigma^A_{o_2}}\right).$$

Table B5 presents the parameters of the estimated Tobit model.

2.2. **Marginal Effect.** Given the ordered logit specification, the probability distribution of K-12 outcome $\widetilde{k}_1$ is

$$Pr[\widetilde{k}_1 \leq k_1 | o_1, x, e_1, \eta, p_1] = \begin{cases} \frac{1}{1+\exp\left(\ell_1(o_1,x,e_1,\eta,p_1)-\alpha_{k_1}\right)} & k_1 = 1, \ldots, 4 \\ 1 & k_1 = 5 \end{cases},$$

where ℓ_1 is the latent outcome function defined in Appendix A.1 and $(\alpha_1, \ldots, \alpha_4)$ are cutoff parameters. When $\widetilde{k}_1$ enters the college outcome function, we transform it into continuous numerical value $c(\widetilde{k}_1)$ representing percentile ranks. In particular, we set

$$c(k_1) = \frac{Pr[\widetilde{k}_1 < k_1] + Pr[\widetilde{k}_1 \leq k_1]}{2}.$$

Table B5: Aid Function

Parameter	2-year	4-year public	4-year private
Intercept	-0.62	1.37	5.10
Inc=2	0.14	0.01	-1.77
Inc=3	-0.27	-1.11	-1.38
Inc=4	-0.34	-1.10	-2.91
Inc=5	-0.75	-1.38	-4.52
College	-0.02	-0.27	0.31
Minority	0.31	0.63	1.33
$k_1 = 2$	-0.03	0.13	0.62
$k_1 = 3$	-0.01	0.31	0.92
$k_1 = 4$	0.06	0.86	2.94
Tuition*(Inc=1)	1.11	0.17	
Tuition*(Inc=2)	0.92	-0.01	
Tuition*(Inc=3)	0.81	0.06	
Tuition*(Inc=4)	0.47	-0.01	
Tuition*(Inc=5)	0.62	-0.08	
Tuition^2	-0.11	0.02	
σ_{o_2}	2.31	2.10	6.19

Tuition and aid are in \$1,000 units.

For example, the share of high school droupouts ($k_1 = 1$) is 0.052, so we set $c(1) = 0.026$. The marginal effect of the log K-12 expenditure $\ln e_1$ on $c(\widetilde{k}_1)$ is therefore defined as

$$\frac{\partial}{\partial \ln e_1} E\left[c(\widetilde{k}_1)|o_1, x, e_1, \eta, p_1\right]$$

$$= \sum_{k_1=1}^{5} c(k_1) \cdot \frac{\partial}{\partial \ln e_1}\left\{Pr[\widetilde{k}_1 \leq k_1|o_1, x, e_1, \eta, p_1] - Pr[\widetilde{k}_1 \leq k_1 - 1|o_1, x, e_1, \eta, p_1]\right\}$$

$$= \sum_{k_1=1}^{4} (c(k_1+1) - c(k_1)) \cdot \frac{\partial}{\partial \ln e_1}\ell_1(o_1, x, e_1, \eta, p_1) \cdot Pr[\widetilde{k}_1 > k_1|o_1, x, e_1, \eta, p_1] \cdot Pr[\widetilde{k}_1 \leq k_1|o_1, x, e_1, \eta, p_1]$$

We focus on the marginal effect for those who attend public schools, so we set $\frac{\partial}{\partial \ln e_1}\ell_1(o_1, x, e_1, \eta, p_1) = \mu_1^6(x_1)$, which varies across low income ($x_1 = 1, \ldots, 3$) and high income ($x_1 = 4, 5$) groups. While the marginal effect also depends on the outcome distribution $Pr[\widetilde{k}_1 \leq k_1|o_1, x, e_1, \eta, p_1]$, we evaluate it at the population average $Pr[\widetilde{k}_1 \leq k_1|o_1 = 1]$.

3. Examining Potential Selection into Private Schools

Table 3 presents OLS and IV estimates of the private high school effect using our ELS2002 sample. The first three columns present the estimated coefficient b_0 (of private high school attendance) from three sets of OLS regressions of the following form

$$y_i = o_{1i} b_0 + x_i b_1 + z_{c_i} b_2 + d_{s_i} + \varepsilon_i, \tag{3.1}$$

where y_i is an education outcome as listed in Table 3, including the math and reading test scores in the 10th and 12th grades, a dummy for high school completion, a dummy for college enrollment and a dummy for 4-year college graduation. On the right side of (3.1), o_{1i} is a binary variable that equals 1 if i's high school is private, and 0 otherwise, x_i is a vector household characteristics, z_{c_i} is a vector of county-level controls and d_{s_i} is a state dummy (s_i is the state that i belongs to). The first two specifications do not control for county-level characteristics, i.e., b_2 is restricted to be zero. In the first specification, x_i is the vector of household characteristics as specified in our model, i.e., education, minority and the 5 income quintile dummies. In all the other specifications, income groups are finer, with 13 income groups defined by the ELS survey.[3] In the third specification, we control for two county-level characteristics: local-level per-student K-12 expenditure and an urban status dummy indicating a county that belongs to an urban area with a population above 750,000. The b_0 estimates are very similar across the three specifications, suggesting that there is little selection into private schools based on additional observable variables that are present in the second and third regression, but are omitted from the model.

To deal with the potential selection into private schools based on unobserved household characteristics, we follow the literature and run two 2SLS regressions. In the first set of 2SLS, we follow the IV strategy in Evans and Schwab (1995) and Neal (1997), who use the contemporary Catholic affiliation as the instrumental variable. Specifically, we use parental Catholic affiliation as the IV.[4] In the second 2SLS specification, we follow Cohen-Zada (2009) and Kim (2011) and use the historical share of Catholics in a county as the IV.[5] The 2SLS estimates of b_0, as shown in the last two

[3]The 13 income groups in the ELS2002 are \$0, \$1-\$1,000, \$1,001-5,000, \$5,001-10,000, \$10,001-15,000, \$15,001-20,000, \$20,001-25,000, \$25,001-35,000, \$35,001-50,000, \$50,001-75,000, \$75,001-100,000, \$100,001-250,000, and \$250,001 or more. In our model, income groups are \$0-20,000, \$20,001-35,000, \$35,001-50,000, \$50,001-75,000, and \$75,001 or more.

[4]Instrumenting private school attendance with Catholic variables has two potential problems. First, the instrument can influence outcomes through the switch from non-Catholic private schools to Catholic schools, instead of through public schools to private schools. Second, the instrument would identify the treatment effect of Catholic schools through the switch from public schools to Catholic schools, not the average treatment effect of private schools. These two problems are expected to be inconsequential for our estimates, since the OLS-estimated effects of Catholic schools and non-Catholic private schools are similar.

[5]The Census of Religious Bodies provides the number of Catholic members in each county in 1890, 1906, 1916, 1926, and 1936. We report the results using the 1936 data compiled by the IPUMS National Historical Geographic Information System, which involves the smallest inconsistency of geographic units between the historical data and the current data. We get similar 2SLS regession results using the older data, which are available upon request.

Table B6: OLS and IV regression results

	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	N
10th Grade Math	0.201	0.178	0.179	0.832	0.385	15,058
	(0.035)	(0.034)	(0.034)	(0.342)	(0.700)	
10th Grade Reading	0.282	0.264	0.266	0.663	0.291	15,058
	(0.037)	(0.037)	(0.037)	(0.301)	(0.829)	
12th Grade Math	0.285	0.255	0.253	0.921	0.723	14,450
	(0.036)	(0.035)	(0.035)	(0.354)	(0.693)	
High School Completion	0.025	0.024	0.025	0.153	0.254	14,825
	(0.003)	(0.003)	(0.003)	(0.076)	(0.221)	
Attend College	0.091	0.090	0.087	0.512	0.966	14,213
	(0.008)	(0.008)	(0.008)	(0.163)	(0.531)	
4-year College Graduate	0.168	0.154	0.150	0.657	1.177	12,370
	(0.017)	(0.016)	(0.016)	(0.172)	(0.525)	
First-stage Coefficient	-	-	-	0.068	0.156	15,058
				(0.011)	(0.083)	
Income Groups in Control	5	13	13	13	13	
County Variables in Control	no	no	yes	yes	yes	
Instrument	-	-	-	Catholic	% Catholic in County	

Note: Test scores are standardized to have a unit standard deviation. Standard errors are clustered at the county level (459 clusters). Control variables are income group dummies, indicator of any parent with some college education, minority dummy, and state dummies. County-level control variables in columns (3)-(5) are urban status (=1 if a county belongs to an urbanized area with population 750,000 or higher) and per-student public K-12 expenditure.

columns of Table 3, are uniformly larger than the OLS estimates across all outcomes, which fails to suggest positive selection into private schools based on unobservables.[6]

References

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling," *Journal of Human Resources*, 2005, *XL*, 791–821.

Cohen-Zada, Danny, "An Alternative Instrument for Private School Competition," *Economics of Education Review*, 2009, *28* (1), 29–37.

Evans, William N. and Robert M. Schwab, "Finishing High School and Starting College: Do Catholic Schools Make a Difference?," *The Quarterly Journal of Economics*, 11 1995, *110* (4), 941–974.

[6]As cautioned by Altonji et al. (2005), these 2SLS estimates are open to question, since the quality of the instruments is doubtful.

Kim, Young-Joo, “Catholic schools or school quality? The effects of Catholic schools on labor market outcomes,” *Economics of Education Review*, 2011, *30* (3), 546–558.

Neal, Derek, “The Effects of Catholic Secondary Schooling on Educational Achievement,” *Journal of Labor Economics*, 1997, *15* (1, Part 1), 98–123.