Online Appendix for "Does Class Size Matter? How, and at What Cost?"

Desire Kedagni[1], Kala Krishna[2], Rigissa Megalokonomou[3], and Yingyan Zhao[4]

[1]Iowa State University

[2]The Pennsylvania State University, NBER and CES-IFO

[3]University of Queensland and IZA

[4]George Washington University

December 24, 2020

1 Enrollment and Number of Classes

We look at the change in number of classes with respect to the change in cohort size. The baseline specification regresses the change in class number on the change in cohort size in the 10th grade over the years. We compare this to the change in the number of classes on the change in the cohort size from 10th to 11th grade and 11th to 12th grade.

Table 1 shows that schools' decision on the class number is more sensitive to the cohort size in the 10th grade than the 11 and 12 grades. The class number remains relatively stable when the same cohort go to 11th and 12th grades. In Table 1, we look at both class size and log class size specifications. In both cases, the coefficient of dlnCohortSize is larger in the baseline, i.e., for the 10th grade.

Table 1: The Change in Class Number with Respect to the Change in Cohort Size

	(1) dClassNo	(2) dClassNo
dlnCohortSize	1.53	
	(0.06)***	
10th to 11th grade × dlnCohortSize	-0.44	
	(0.1)***	
11th to 12th grade × dlnCohortSize	-0.52	
	(0.1)***	
dCohortSize		0.031
		(0.0009)***
10th to 11th grade × dCohortSize		-0.0100
		(0.002)***
11th to 12th grade × dCohortSize		-0.012
		(0.002)***
10th to 11th grade	0.040	0.053
	(0.02)*	(0.02)***
11th to 12th grade	-0.057	-0.058
	(0.02)***	(0.02)***
R-sq	0.255	0.350
N	3247	3247

(1) *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

2 Full First Stage Estimates

Table 2 reports the full first stage estimates in Column (3) and (4) in Table 2 in the paper.

Table 2: Full First Stage Estimates for Column (3) and (4) in Table 2 in the paper.

	(1)	(2)	(3)
		First Stage	
Enrollment	0.070***	0.16***	5.56***
	(0.004)	(0.01)	(0.7)
Sq of *Enrollment*		-0.00047***	-0.013***
		(0.00007)	(0.003)
Female	0.048**	0.048**	2.17**
	(0.02)	(0.02)	(0.9)
Age	-0.049	-0.041	-2.88
	(0.09)	(0.08)	(3.4)
AgeSQ	0.00040	0.00031	0.051
	(0.002)	(0.002)	(0.08)
School FE	YES	YES	YES
School-Specific Linear Time Trend	YES	YES	YES
N	81845	81845	81845

(1) $ClassSizeSQ$ is the square of ClassSize. Female = 1 if a student is female. Age and AgeSQ control for students' age and its square.

(2) Standard errors are clustered at the class level. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

3 Estimates of the Class Size Effect using the Hoxby IV

In this section, we follow the approach of Hoxby (2000) to estimate the class size effect. It is natural to think of overall enrollment as an exogenous shock to class size. The approach taken in Hoxby (2000) goes a step further: she fits a quartic to the enrollment data and uses deviations from the quartic as the exogenous variation. In this way, she controls for trends in enrollment. We run the analogue of Hoxby's approach running the following school by school.

$$\ln e_{jt} = \alpha_{j0} + \alpha_{j1}t + \alpha_{j2}t^2 + \alpha_{j3}t^3 + \alpha_{j4}t^4 + \varepsilon_{jt}$$

For each school j, we estimate $\alpha_{j0}, ..., \alpha_{j4}$ and take the time trend out. The error term $\hat{\varepsilon}_{jt}$ serves as our IV for the class size. In Table 3, $\ln U$ denotes the deviations of the data from the predicted values, while $\ln USQ$ denotes the square of $\ln U$.

Table 3: IV Estimates of Class Size Effects using the Hoxby IV

	(1)	(2)		(3)	(4)	
	Dependent Variable: GPA					
	Second Stage					
$ClassSize$	-0.23***	0.56***				
	(0.03)	(0.2)				
$ClassSizeSQ$		-0.018***				
		(0.005)				
log($ClassSize$)				-4.69***	33.4***	
				(0.6)	(7.3)	
Sq of log($ClassSize$)					-6.50***	
					(1.3)	
Female	0.90***	0.90***		0.90***	0.90***	
	(0.03)	(0.03)		(0.03)	(0.03)	
Age	-1.79***	-1.81***		-1.78***	-1.83***	
	(0.1)	(0.10)		(0.1)	(0.10)	
AgeSQ	0.028***	0.029***		0.028***	0.029***	
	(0.002)	(0.002)		(0.002)	(0.002)	
Kleibergen-Paap Statistic	152.1	32.0		158.0	55.7	
p-value	0.000	0.000		0.000	0.000	
School FE	YES	YES		YES	YES	
School-Specific Linear Time Trend	YES	YES		YES	YES	
R-sq	0.046	0.030		0.046	0.023	
N	81845	81845		81845	81845	
	First Stage					
	$ClassSize$	$ClassSize$	$ClassSizeSQ$	log($ClassSize$)	log($ClassSize$)	Sq of log($ClassSize$)
lnU	4.05***	5.79***	144.4***	0.20***	0.46***	2.28***
	(0.3)	(1.1)	(50.7)	(0.01)	(0.07)	(0.4)
$lnUSQ$		-0.54	10.6		-0.081***	-0.34***
		(0.3)	(16.7)		(0.02)	(0.1)

(1) $ClassSizeSQ$ is the square of $ClassSize$.
(2) Female = 1 if a student is female. Age and AgeSQ control for students' age and its square.
(2) Standard errors are clustered at the class level. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

The idea is that by removing the predictable time trend, we take the advantage of the unexpected change of the enrollment, which is an exogenous shock to the class size.

Columns (1) and (2) in Table 3 give the IV estimates for grade 10 for the linear and quadratic models. Columns (3) and (4) in Table 3 give the IV estimates for grade 10 for the linear-in-log and quadratic-in-log models. The standard errors are clustered at the class level. The lower panel of the table gives the relevant estimates for the first stage for convenience. The upper panel gives the estimates for the second stage in all these tables. The results show that the class size effect has the hump shape expected with a peak at around 13 in Column (2) and 16 in Column (4).

We use Hoxby's instrument so that we would expect the shock in enrollment to be positively correlated with class size as we find.[1] In Greece, school enrollment is not a choice as students attend the local school unless they choose to go to private school (which is uncommon) so that we are not worried about parents basing their enrollment on class size. Nor do schools have a say on enrollment as they have to admit all eligible students. Note that the first stage looks fine except for the level specification where the square of the residual is not significant: the coefficients on the instrument are significant at the 1% level and the instruments are not weak across all specifications in Table 3. It is interesting, and in line with the literature that women have a higher GPA.

4 Teacher Quality Data

Table 4 presents the summary statistics for the sample with teacher data available, as well as the difference in estimates for the baseline model between the restricted sample and the full sample. These 9 schools are clearly different from the full sample. They have a smaller (by 24 students) cohort size. Their class size is smaller (by 2.87 students) as well and number of classes is also smaller by (0.73). Students in these schools are also slightly older.

[1]It is worth pointing out that we have at most 12 years of data for a school while Hoxby had 24.

Table 4: Summary Statistics of the Sample with Teacher Data

	(1)	(2)	(3)
	Full	Sample with Teacher Data Available	(1) - (2)
		Individual Level Data	
GPA	11.79	11.88	-0.0916
	(3.79)	(4.11)	(-1.07)
female	0.54	0.55	-0.00600
	(0.50)	(0.50)	(-0.54)
Age	15.97	16.03	-0.0579***
	(0.60)	(0.51)	(-4.11)
N	81845	2031	
		Class Level Data	
ClassSize	22.62	19.75	2.872***
	(4.15)	(3.39)	(6.96)
N	3641	103	3744
		School Level Data	
CohortSize	76.17	52.15	24.01***
	(33.90)	(33.26)	(4.35)
ClassNo	3.37	2.64	0.728***
	(1.24)	(1.46)	(3.57)
N	1082	39	

Table 5 presents the regression of the baseline specification for the restricted sample. The same hump shape can be observed, though the coefficients are not significant.

One of the concerns is that good teachers may be assigned to large classes. Lavy and Megalokonomou (2019) who use the same data show that teachers' gender and the number of classes they teach are not correlated with class size. We also take the estimates of the teacher fixed effects in Table A.2 as a measure of teacher quality, and find that the correlation between the estimated teacher fixed effects and (log) class class is insignificantly different from zero.

Table 5: IV Estimates for the Sample with Teacher Data Available

	(1)	(2)	
	Dependent Variable: GPA		
	Second Stage		
ClassSize	0.32***	2.07	
	(0.1)	(2.0)	
ClassSizeSQ		-0.043	
		(0.05)	
Female	1.00**	1.00**	
	(0.4)	(0.4)	
Age	3.21	3.77*	
	(2.2)	(2.2)	
AgeSQ	-0.13**	-0.14**	
	(0.06)	(0.06)	
Kleibergen-Paap Statistic	18.1	4.4	
p-value	0.000	0.036	
School FE	YES	YES	
R-sq	0.051	0.045	
N	885	885	
	First Stage		
	ClassSize	*ClassSize*	*ClassSizeSQ*
Enrollment	0.17***	0.11	2.75
	(0.02)	(0.1)	(5.6)
Sq of *Enrollment*		0.00035	0.025
		(0.0008)	(0.03)

(1) *ClassSizeSQ* is the square of *ClassSize*. Female = 1 if a student is female. Age and AgeSQ control for students' age and its square.

(2) Standard errors are clustered at the class level. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

5 Nonparametric Estimates of Class Size Effects

Recent work by Chernozhukov et al. (2013) provides a framework applicable to our setting and which does not assume a particular functional form. In essence, it allows for school fixed effects in a non linear manner. The approach needs panel data, which we have, and the endogenous regressor (class size) has to be discrete.

Following their approach we specify the following model:

$$GPA_{jt} = g(CS_{jt,}\alpha_j, \varepsilon_{jt})$$

which has achievement as measured by the *average* GPA of school j in period t being a function of the *average* class size in school j in period t, a school fixed effect, α_j, and a shock, ε_{jt}, that is school and time specific. Though class size is discrete, average class size is a continuous variable. For this reason we discretize the class size into bins below.

We specify this relationship to be at the school level, because nonparametric estimation of this kind needs a long panel for each j.[2]

The assumption needed for this approach is the following:

Assumption 1 (time-homogeneity)

$$\varepsilon_{jt} \mid \boldsymbol{CS}_j,\ \alpha_j \sim F(. \mid \boldsymbol{CS}_j, \alpha_j).$$

In other words, the distribution of the shock ε_{jt}, conditional on the *vector* of average class sizes for the school at all periods (denoted by $\boldsymbol{CS}_j$) and the school itself, is time independent as the function F has no time subscript. Stated slightly differently, whatever the distribution of the shock is, its conditional distribution given the vector of average class sizes for school j does not depend on t. Chernozhukov et al. (2013) interprets this as time being randomly assigned or time being an instrument along with the distribution of factors other than class size not changing over time.

One might be concerned that Assumption 1 does not hold in the data and as a result, the approach of Chernozhukov et al. (2013) cannot be used. Fortunately, we need not take the assumption on faith. A recent paper, see Ghanem (2017) derives a statistical test to check the validity of Assumption 1. In Section 5.2, we show that using this methodology, we cannot reject the hypothesis that Assumption 1 holds in the data.[3]

5.1 Estimates

We choose to discretize class size into three bins. We do so as we will need to estimate the effect going from each bin to the other so that the number of coefficients rises rapidly with the number of bins. The first is class size below a cutoff s_0. The second bin is from s_0 to s_1, and the third is more than s_1. Since these switches are identifying the effects of interest, we need to choose s_0 and s_1 to ensure that the bins are such that these switches occur.

Let δ_{lk} be the average effect on mean GPA in a school of switching from bin l to k and $\hat{\delta}_{lk}$ be

[2] If we had specified the model to hold at the individual level, we would have a panel of length three, though we would have a lot of students. Similarly, we could have specified the model to be at the class level if we had information on which teacher was assigned to which class. In this case, we would have a panel of the same length as that for the model we use, assuming the teacher was there throughout. We do not have data on teachers and their assignment to classes we cannot use this approach.

[3] We thank Dalia Ghanem for sharing programming code with us.

its consistent estimator. In effect, what is done is the following. In each year, each school has a mean GPA and a mean class size and so falls into one of the three bins. Over the entire sample period, each school indexed by j can be in either bin 1 only, bin 1 and 2 only, bin 1 and 3 only, bin 2 and 3 only or in all three bins. We calculate the mean GPA *over time* for school j when in bin $l = 1, 2, 3$. For example, if there were 5 periods and the school was in bins $(1, 2, 1, 3, 2)$ over these time periods with GPA $(g_1, g_2, g_3, g_4, g_5)$, then the mean GPA in bin 2 would be $(g_2 + g_5)/2$ while the mean GPA in bin 1 would be $(g_1 + g_3)/2$. Their difference would capture Δ_{12}^{j} for school j. The estimated $\hat{\delta}_{12}$ would then be

$$\hat{\delta}_{12} = \frac{\sum_{j=1}^{N} d_j \Delta_{12}^{j}}{\sum_{j=1}^{N} d_j}$$

where d_j is 1 if the school was ever in both bin 1 and 2 over the entire sample period.

We set s_0 at 15 and s_1 at 22. $\hat{\delta}_{12} = 1.55$ and $\hat{\delta}_{23} = -.18$. Both are significantly different from zero at the 1% level. In Table 6, we vary s_1 from 21 to 24 along the rows and s_0 from 12 to 17 along the columns. For each value of s_0 and s_1 we give the estimate of $\hat{\delta}_{12}$ and $\hat{\delta}_{23}$. Note that no matter what s_0 to s_1 are set at, $\hat{\delta}_{12} > 0, \hat{\delta}_{23} < 0$ and significant. This is consistent with a nonmonotonic relationship between GPA and class size.

Table 6: Nonparametric results

s_1 \ s_0	12		13		14		15		16		17	
	δ_{12}	δ_{23}	δ_{12}	δ_{23}	δ_{12}	δ_{23}	δ_{12}	δ_{23}	δ_{12}	δ_{23}	δ_{12}	δ_{23}
21	4.99*	-0.24*	3.66*	-0.25*	2.35*	-0.24*	1.58*	-0.24*	1.04*	-0.21*	0.69†	-0.23*
	(1.05)	(0.07)	(0.86)	(0.07)	(0.68)	(0.07)	(0.53)	(0.07)	(0.44)	(0.07)	(0.32)	(0.07)
22	4.94*	-0.18*	3.61*	-0.19*	2.31*	-0.19*	1.55*	-0.18*	1.02*	-0.16†	0.67†	-0.17*
	(1.06)	(0.07)	(0.87)	(0.07)	(0.69)	(0.07)	(0.53)	(0.07)	(0.44)	(0.07)	(0.32)	(0.07)
23	4.93*	-0.19*	3.59*	-0.19*	2.30*	-0.19*	1.54*	-0.18*	1.01†	-0.17*	0.65†	-0.18*
	(1.05)	(0.07)	(0.86)	(0.07)	(0.68)	(0.07)	(0.53)	(0.07)	(0.44)	(0.07)	(0.32)	(0.07)
24	4.89*	-0.19*	3.55*	-0.20*	2.27*	-0.20*	1.52*	-0.19*	0.99†	-0.18*	0.64†	-0.18†
	(1.05)	(0.08)	(0.86)	(0.08)	(0.68)	(0.08)	(0.53)	(0.07)	(0.44)	(0.07)	(0.31)	(0.08)

(1) †, * indicate significance at the 5% and 1% levels, respectively.

5.2 Testing the time-homogeneity assumption

Assumption 1 (time-homogeneity)**.**

$$\varepsilon_{jt} \mid \boldsymbol{CS}_j,\ \alpha_j \sim F(. \mid \boldsymbol{CS}_j,\alpha_j).$$

Ghanem (2017) derived testable equality restrictions for the time-homogeneity Assumption 1. She therefore proposed a statistical test based on Kolmogorov-Smirnov and Cramer-von-Mises statistics. Below, we explain the intuition of the test. For the sake of simplicity, suppose we only have two periods. As mentioned in Chernozhukov et al. (2013), the time-homogeneity assumption is equivalent to $(\epsilon_{jt}, \alpha_j)|CS_{j.} =^d (\epsilon_{j1}, \alpha_j)|CS_{j.}$, for all t. Then this assumption implies that the conditional distribution of the second period average GPA for a school j is the same as its conditional distribution of the first period average GPA given its history of class size choices $CS_{j.} = (x, x')$. Indeed, we have:

$$\begin{aligned}
&(\epsilon_{j2}, \alpha_j)|CS_{j.} = (x, x'^d(\epsilon_{j1}, \alpha_j)|CS_{j.} = (x, x') \\
&\Rightarrow g(x', \alpha_j, \epsilon_{j2})|CS_{j.} = (x, x'^d g(x, \alpha_j, \epsilon_{j1})|CS_{j.} = (x, x') \\
&\Rightarrow g(CS_{j2}, \alpha_j, \epsilon_{j2})|CS_{j.} = (x, x'^d g(CS_{j1}, \alpha_j, \epsilon_{j1})|CS_{j.} = (x, x') \\
&\Rightarrow GPA_{j2}|CS_{j.} = (x, x'^d GPA_{j1}|CS_{j.} = (x, x'),
\end{aligned}$$

where $=^d$ means equal in distribution.

Testing procedure: bootstrap

Let T_N be a test statistic. The following summarizes the steps of the test.

1. Compute the statistic T_N for the original data $\{(GPA_{1.}, CS_{1.}), \ldots, (GPA_{N.}, CS_{N.})\}$.

2. Resample N observations $\{(GPA^*_{1.}, CS^*_{1.}), \ldots, (GPA^*_{N.}, CS^*_{N.})\}$ with replacement from the original data. Compute T^b_N, the centered statistic for the bth bootstrap sample.

3. Repeat points 1. and 2. B times.

4. Calculate the p-values of the tests with $p = \frac{1}{B}\sum_{b=1}^{B} 1\{T^b_N > T_N\}$. Reject if p-value is smaller than some significance level α.

For the implementation of the test, we set $B = 500$. We use the Kolmogorov-Smirnov and Cramer-von-Mises statistics (See (Ghanem, 2017) for details on the formulas). All p-values are higher than 10%, suggesting that the identifying Assumption 1 is not rejected at any 1%, 5% nor 10% significance levels. The p-values for the standard parallel trend assumption are 0.99 and 0.90 for the Kolmogorov-Smirnov and Cramer-von-Mises statistics, respectively.

6 IV Estimates of Class Size Effects using Small Schools

A concern might be that there are far fewer classes/students in the lower class size bins raising the concern that a quadratic term just adds curvature to a fitted linear regression line with the results being driven by a small number of small schools. To deal with this issue, we see if our results remain when we divide the sample into large/small schools where the cutoff of cohort size is limited to 30, 50 and 70. The number of small schools in terms of students is .8%, 5.7% and 13% respectively of the total. In the body of the paper, we present the results for large schools.

In this section, we estimate the class size effect using small schools, the results are less significant, though the same patterns emerge. Table 7 presents the estimation results. The hump shape remains and the peak occurs around 17 as well.

Table 7: IV Estimates of the Baseline Model: Small Schools

	(1)		(2)		(3)	
	Dependent Variable: GPA					
	Second Stage					
Average CohortSize	< 30		< 50		< 70	
ClassSize	0.34		0.62		-0.12	
	(0.4)		(0.5)		(0.5)	
ClassSizeSQ	-0.011		-0.018		0.0051	
	(0.01)		(0.01)		(0.01)	
Female	1.37***		1.11***		0.95***	
	(0.3)		(0.1)		(0.08)	
Age	-1.92***		-2.36***		-2.37***	
	(0.5)		(0.3)		(0.2)	
AgeSQ	0.032***		0.039***		0.039***	
	(0.010)		(0.005)		(0.004)	
Kleibergen-Paap Statistic	4.0		4.3		4.7	
p-value	0.046		0.039		0.029	
School FE	YES		YES		YES	
School-Specific Linear Time Trend	YES		YES		YES	
R-sq	0.095		0.083		0.067	
N	667		4678		10426	
	First Stage					
	ClassSize	ClassSizeSQ	ClassSize	ClassSizeSQ	ClassSize	ClassSizeSQ
Enrollment	2.99***	83.4***	0.36**	8.91	0.40***	13.3***
	(0.7)	(29.9)	(0.2)	(6.5)	(0.09)	(3.7)
Sq of Enrollment	-0.056***	-1.37	-0.0028	-0.041	-0.0031***	-0.095**
	(0.02)	(0.9)	(0.002)	(0.10)	(0.0009)	(0.04)

(1) *ClassSizeSQ* is the square of *ClassSize*. Female = 1 if a student is female. Age and AgeSQ control for students' age and its square.
(2) Standard errors are clustered at the class level. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

7 Transition Matrix of Enrollment: Simulation vs. Data

In this section, we present the transition matrix of the actual and simulated data in Tables 8 and 9, respectively. We group enrollment into ten quantiles and look at the transition probability across quantiles between $t-1$ and t. The simulated data from the estimated enrollment process with different groups mimics the transition matrix of the actual data nicely once the number of enrollment equations exceeds 6.

As one might be concerned that the way we group schools matters, we present the transition matrix of the simulated data with 1, 3, 9 and 12 enrollment equations. When we estimate the enrollment process with a single equation, the simulated transition matrix does not match the actual data as shown in Table 10. The transition probability is more dispersed compared to the data. Table 11 - Table 13 present the simulated transition matrix with 3, 9 and 12 enrollment equations. When we estimate the enrollment process with finer groups, the simulated transition matrix matches even better with the data.

We choose $K = 6$ in the paper. The more disaggregated enrollment process, the better the simulations match the data. However, a more disaggregated enrollment process also leads to more computational difficulties in the structural estimation as value functions differ by enrollment process. We choose $K = 6$ to balance these two forces.

Table 8: Transition Matrix of Enrollment: Data

$e_{t=\text{First Period}}$ \ $e_{t=\text{Last Period}}$	1	2	3	4	5	6	7	8	9	10
1	0.75	0.22	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.00
2	0.18	0.50	0.26	0.05	0.01	0.00	0.00	0.01	0.00	0.00
3	0.01	0.22	0.32	0.23	0.15	0.03	0.04	0.00	0.00	0.01
4	0.00	0.06	0.22	0.28	0.26	0.07	0.09	0.00	0.01	0.00
5	0.02	0.01	0.12	0.17	0.28	0.26	0.07	0.07	0.00	0.01
6	0.01	0.00	0.01	0.10	0.24	0.24	0.21	0.13	0.03	0.02
7	0.00	0.00	0.04	0.08	0.08	0.24	0.15	0.27	0.13	0.02
8	0.00	0.00	0.01	0.05	0.07	0.10	0.24	0.23	0.25	0.05
9	0.00	0.00	0.00	0.00	0.01	0.03	0.20	0.19	0.34	0.24
10	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.11	0.20	0.65

Table 9: Transition Matrix of Enrollment: Simulation with Six Enrollment Equations

$e_{t=\text{First Period}}$ \ $e_{t=\text{Last Period}}$	1	2	3	4	5	6	7	8	9	10
1	0.76	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.22	0.52	0.25	0.01	0.00	0.01	0.00	0.00	0.00	0.00
3	0.02	0.21	0.41	0.22	0.07	0.06	0.01	0.01	0.00	0.00
4	0.00	0.03	0.22	0.33	0.32	0.07	0.02	0.00	0.01	0.00
5	0.00	0.00	0.10	0.31	0.21	0.23	0.07	0.05	0.03	0.00
6	0.00	0.01	0.00	0.11	0.23	0.26	0.25	0.10	0.04	0.00
7	0.00	0.00	0.00	0.03	0.07	0.23	0.25	0.32	0.10	0.00
8	0.00	0.00	0.00	0.00	0.06	0.15	0.22	0.36	0.16	0.04
9	0.00	0.00	0.00	0.00	0.00	0.04	0.08	0.16	0.47	0.25
10	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.06	0.22	0.70

Table 10: Transition Matrix of Enrollment: Simulation with One Enrollment Equation

e_{t-1} \ e_t	1	2	3	4	5	6	7	8	9	10
1	0.82	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.23	0.55	0.21	0.02	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.19	0.54	0.20	0.07	0.01	0.00	0.00	0.00	0.00
4	0.00	0.03	0.19	0.52	0.20	0.06	0.00	0.01	0.00	0.00
5	0.00	0.00	0.06	0.21	0.39	0.28	0.05	0.01	0.00	0.00
6	0.00	0.00	0.00	0.10	0.18	0.41	0.22	0.08	0.01	0.00
7	0.00	0.00	0.00	0.00	0.07	0.19	0.38	0.25	0.11	0.00
8	0.00	0.00	0.00	0.01	0.01	0.04	0.32	0.34	0.28	0.01
9	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.31	0.36	0.23
10	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.22	0.75

Table 11: Transition Matrix of Enrollment: Simulation with Three Enrollment Equations

e_{t-1} \ e_t	1	2	3	4	5	6	7	8	9	10
1	0.89	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.12	0.65	0.18	0.05	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.24	0.38	0.22	0.12	0.04	0.01	0.00	0.00	0.00
4	0.00	0.01	0.27	0.36	0.21	0.10	0.05	0.00	0.00	0.00
5	0.00	0.01	0.14	0.19	0.27	0.26	0.14	0.00	0.00	0.00
6	0.00	0.00	0.01	0.11	0.23	0.27	0.25	0.09	0.03	0.00
7	0.00	0.00	0.00	0.02	0.15	0.13	0.33	0.26	0.12	0.00
8	0.00	0.00	0.00	0.01	0.06	0.11	0.20	0.34	0.21	0.06
9	0.00	0.00	0.00	0.00	0.00	0.05	0.03	0.27	0.42	0.23
10	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.23	0.71

Table 12: Transition Matrix of Enrollment: Simulation with Nine Enrollment Equations

$e_{t=\text{First Period}}$ \ $e_{t=\text{Last Period}}$	1	2	3	4	5	6	7	8	9	10
1	0.84	0.15	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.20	0.55	0.23	0.01	0.01	0.00	0.00	0.00	0.00	0.00
3	0.00	0.24	0.34	0.25	0.11	0.04	0.02	0.00	0.00	0.00
4	0.00	0.06	0.25	0.29	0.20	0.12	0.04	0.03	0.01	0.00
5	0.00	0.00	0.12	0.22	0.30	0.17	0.13	0.05	0.01	0.00
6	0.00	0.00	0.04	0.10	0.19	0.28	0.22	0.14	0.03	0.00
7	0.00	0.00	0.01	0.06	0.14	0.21	0.33	0.17	0.09	0.00
8	0.00	0.00	0.00	0.02	0.06	0.17	0.18	0.29	0.23	0.04
9	0.00	0.00	0.00	0.00	0.00	0.04	0.07	0.24	0.40	0.24
10	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.08	0.18	0.73

Table 13: Transition Matrix of Enrollment: Simulation with Twelve Enrollment Equations

$e_{t=\text{First Period}}$ \ $e_{t=\text{Last Period}}$	1	2	3	4	5	6	7	8	9	10
1	0.81	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.18	0.57	0.22	0.02	0.01	0.00	0.00	0.00	0.00	0.00
3	0.00	0.18	0.45	0.22	0.11	0.04	0.01	0.00	0.00	0.00
4	0.00	0.02	0.26	0.30	0.22	0.13	0.05	0.01	0.01	0.00
5	0.00	0.01	0.13	0.23	0.22	0.24	0.11	0.03	0.02	0.00
6	0.00	0.00	0.01	0.12	0.24	0.28	0.15	0.14	0.06	0.00
7	0.00	0.00	0.02	0.03	0.07	0.27	0.25	0.21	0.12	0.04
8	0.00	0.00	0.00	0.03	0.05	0.07	0.24	0.29	0.24	0.07
9	0.00	0.00	0.00	0.01	0.00	0.04	0.11	0.32	0.27	0.25
10	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.08	0.25	0.65

8 The Relationship between Class Number/Class Size and Enrollment

In this section, we look at the response of class number/class size to the past, current and future enrollment. We run the following regression.

$$y_{jt} = \gamma_0 + \gamma_1 e_{jt} + \gamma_2 e_{j,t-1} + \gamma_3 e_{j,t-2} + \gamma_4 e_{jt+1} + \varepsilon_{jt}$$

where, y_{jt} is the class number or log class size. We find that the class number/class size responds more to current enrollment than to the past or future enrollment as shown in Table 15.

Table 14: The Response of Class Number and Class Size to the Past, Current and Future Enrollment

	(1)	(2)	(3)	(4)
	ClassNo		lnClassSize	
$lnCohortSize_t$	1.79	1.84	0.48	0.44
	(0.04)***	(0.05)***	(0.02)***	(0.03)***
$lnCohortSize_{t-1}$	-0.069	-0.060	0.028	0.034
	(0.04)*	(0.04)	(0.02)	(0.02)
$lnCohortSize_{t-2}$		-0.042		-0.064
		(0.04)		(0.02)***
$lnCohortSize_{t+1}$	-0.064	-0.065	0.013	0.013
	(0.04)	(0.05)	(0.02)	(0.03)
R-sq	0.885	0.891	0.425	0.394
N	3033	2631	3033	2631

9 The Choice of Class Number in Response to the Change in Enrollment

To study the response of class number to the change in enrollment and show that it is more likely to increase than decrease, we run the following regression.

$$\Delta \text{Class Number}_{jt} = \gamma_0 + \gamma_1 \Delta e_{jt} \times \mathbb{1}\{\Delta e_{jt} \geq 0\} + \gamma_2 \Delta e_{jt} \times \mathbb{1}\{\Delta e_{jt} < 0\}$$
$$+\gamma_3 \text{Class Number}_{jt-1} + \gamma_4 e_{jt-1} + \varepsilon_{jt}$$

where, ΔClass Number_{jt} is the change of class number between t and $t-1$, and Δe_{jt} is the change of enrollment between t and $t-1$. γ_1 is the estimate for the response of class number to an increase in enrollment, while γ_2 is the one for the response of class number to a decrease in enrollment.

As shown in Table 15, the estimate of γ_1 is significantly larger than the estimate of γ_2.

Table 15: The Response of Class Number to the Change in Enrollment

	(1) ΔClass Number$_{jt}$
$\Delta e_{jt} \times \mathbb{1}\{\Delta e_{jt} \geq 0\}$	1.83***
	(0.1)
$\Delta e_{jt} \times \mathbb{1}\{\Delta e_{jt} < 0\}$	-1.44***
	(0.1)
Class Number$_{jt-1}$	-0.40***
	(0.02)
e_{jt-1}	0.84***
	(0.06)
p-value for $\alpha_1 \neq \alpha_2$	0.0237
R-sq	0.496
N	1044

10 Structural Estimates with Finer Groups

Table 16, Table 18, Table 20 and Table 22 present the estimates of the enrollment process when schools are divided into 1, 3, 9 or 12 categories evenly based on the average cohort size over time. Table 17, Table 19, Table 21 and Table 23 present the structural estimation of c, H, F, respectively. Overall, these estimates are similar. Firing cost is about the same as the hiring cost and about 3 times the annual wage.

Table 16: Estimates of Enrollment Process with One Category of School Size

	Full Sample
γ_1	0.91
sd	(0.00)
γ_0	6.80
sd	(0.28)
N	1062

(1) The standard errors are presented in parentheses.

Table 17: Estimates of the Structural Model with One Category of School Size

	c	H	F	σ
All	83.34	258.46	247.02	200.15
sd	(53.61)	(66.87)	(64.64)	(5.74)
Euro	€20,572	€63,801	€60,978	

(1) The standard errors are presented in parentheses.

Table 18: Estimates of Enrollment Process with Three (Equal sized) Categories

Enrollment Quantile	0-33	33-67	68-100
γ_1	0.79	0.48	0.66
sd	(0.01)	(0.02)	(0.01)
γ_0	8.70	38.76	37.88
sd	(0.44)	(1.34)	(1.36)
N	344	355	363

(1) The standard errors are presented in parentheses.

Table 19: Estimates of the Structural Model with Three (Equal Sized) Categories

	c	H	F	σ
All	90.46	246.26	237.81	208.60
sd	(57.49)	(68.16)	(64.76)	(5.48)
Euro	€20,572	€56,001	€54,081	

(1) The standard errors are presented in parentheses.

Table 20: Estimates of Enrollment Process with Nine Categories of School Size

Enrollment Quantile	0-11	11-22	23-33	34-44	45-55
γ_1	0.50	0.48	0.44	0.32	0.49
sd	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
γ_0	12.72	20.63	31.59	44.50	37.71
sd	(1.06)	(1.20)	(2.34)	(1.91)	(2.65)
N	109	113	122	117	116
Enrollment Quantile	56-67	68-78	79-89	90-100	
γ_1	0.14	0.35	0.21	0.54	
sd	(0.05)	(0.02)	(0.05)	(0.04)	
γ_0	72.24	61.08	82.83	60.75	
sd	(3.86)	(2.21)	(5.08)	(5.13)	
N	122	126	120	117	

(1) The standard errors are presented in parentheses.

Table 21: Estimates of the Structural Model with Nine Categories of School Size

	c	H	F	σ
All	89.54	252.28	235.87	216.46
sd	(54.43)	(64.68)	(61.67)	(5.34)
Euro	€20,572	€57,961	€54,192	

(1) The standard errors are presented in parentheses.

Table 22: Estimates of Enrollment Process with Twelve Categories of School Size

Enrollment Quantile	0-8	9-17	18-25	26-33	34-42	43-50
γ_1	0.40	0.40	0.38	0.52	0.35	0.28
sd	(0.05)	(0.05)	(0.03)	(0.07)	(0.03)	(0.04)
γ_0	13.92	20.63	29.35	28.39	42.17	51.64
sd	(1.09)	(1.45)	(1.68)	(3.88)	(2.05)	(3.00)
N	83	84	97	80	92	81
Enrollment Quantile	51-58	59-67	68-75	76-83	84-92	93-100
γ_1	0.53	0.11	0.39	-0.01	0.37	0.50
sd	(0.05)	(0.05)	(0.03)	(0.05)	(0.10)	(0.04)
γ_0	35.35	75.99	56.47	101.21	69.69	67.66
sd	(3.97)	(4.18)	(2.29)	(5.31)	(10.55)	(5.70)
N	84	98	92	92	86	93

(1) The standard errors are presented in parentheses.

Table 23: Estimates of the Structural Model with Twelve Categories of School Size

	c	H	F	σ
All	88.83	255.91	236.14	219.25
sd	(54.41)	(64.87)	(61.46)	(5.47)
Euro	€20,572	€59,270	€54,691	

(1) The standard errors are presented in parentheses.

References

Chernozhukov, Victor, Ivan Fernandez-Val, Jinyong Hahn, and Whitney Newey, "Average and quantile effects in nonseparable panel models," *Econometrica*, 2013, *81* (2), 535–580.

Ghanem, Dalia, "Testing identifying assumptions in nonseparable panel data models," *Journal of Econometrics*, 2017, *197* (2), 202–217.

Hoxby, Caroline M., "The effects of class size on student achievement: New evidence from population variation," *The Quarterly Journal of Economics*, 2000, *115* (4), 1239–1285.

Lavy, Victor and Rigissa Megalokonomou, "Persistency in Teachers' Grading Bias and Effects on Longer-Term Outcomes: University Admissions Exams and Choice of Field of Study," Technical Report, National Bureau of Economic Research 2019.