Oligopsony and Collective Bargaining *

Tirza J. Angerhofer † Allan Collard-Wexler ‡
Matthew C. Weinberg §

January 22, 2026

Abstract

Employers facing limited labor market competition may suppress wages below socially optimal levels. Unions can counteract wage suppression through collective bargaining, though they may also push wages above the socially optimal level and lead to job rationing. To assess these forces, we estimate a structural model of labor supply, labor demand, and Nash-in-Nash bargaining over wages between teacher unions and school districts in Pennsylvania's K-12 public school system from 2013 to 2020. We compare negotiated equilibrium wages and employment to pure oligopsony and social planner scenarios. On average, pure oligopsony reduces wages 16 percent below the social optimum, while collective bargaining raises wages 23 percent above the optimum. Averages masks substantial district-level heterogeneity driven by bargaining power variation. Twenty-seven percent of schools have negotiated salaries below the social optimum due to cross-district externalities, where high salaries at one school cause hiring reductions that increase labor supply elsewhere.

*We thank participants at Chicago IO+, Yale, Penn, Princeton, Toronto, UT-Austin, KU Leuven, UNC, Wisconsin, Stanford, Northwestern, UC Berkeley, UCLA, UCSC, IIOC, and the Federal Trade Commission Microeconomics Conference for comments. This paper would not be possible without Hugh McCartney's contribution in sourcing the data and understanding the institutional background of teachers in Pennsylvania. We thank Kenny Su, Sunmin Kim, Seohee Kim, and Kosuke Shimamoto for research assistance. We thank Michael Dinerstein, Steven Berry, David Berger, Gregor Jarosch, Pat Kline, Paolo Somaini, and Ali Yurukoglu for comments. We have no financial interests to disclose.

†Duke University
‡Duke University and NBER
§Ohio State University

Keywords: Oligopsony, Monopsony, Collective Bargaining, Labor Markets, Teacher Unions, Nash Bargaining
JEL Codes: J31, J42, J45, J51, L13

1 Introduction

A growing body of work measures employer monopsony power and considers its welfare implications on the labor market. When facing an upward sloping labor supply curve, a firm with monopsony power suppresses wages and employment below the socially optimal level. In this paper, we consider the most traditional policy intervention for a monopsonized labor market, unionization, where workers form collective bargaining units to negotiate wages with employers. As discussed by Galbraith (1954), and earlier by Robinson (1933), collective bargaining gives workers "countervailing power" when negotiating with large employers; unions act as labor supply monopolists to counter labor demand monopsony power.[1] Ideally, introducing unions would raise wages and employment to the socially optimal level, but unions also have the potential to raise wages beyond that resulting in job rationing.

In this paper, we use a structural model of labor demand with hiring, labor supply, and Nash bargaining to quantify the equilibrium effects of collective bargaining relative to the oligopsony setting without unions (where employers post wages) and evaluate how the equilibria compare to the socially efficient outcome. We also consider the welfare effects of two alternative wage policies: minimum wages and state-level uniform wages.

We focus on the market for K-12 teachers in Pennsylvania, which is particularly well suited to understanding collective bargaining and monopsony power. First, public schools are the predominant employers of school teachers, and within fixed geographic areas, hold significant labor market power. Second, teacher wages follow a uniform schedule, which greatly simplifies the monopsony distortion. In contrast, if wages are teacher-specific, monopsony power may cause no distortion in hiring, since the employer need not raise the wages of other workers in order to make a hire. Third, in most Pennsylvania school districts, with the exception of most charter schools, wages are negotiated through a collective agreement between each school district and a union–the district's local chapter of the Pennsylvania State Education Association.

[1] See Chapter 25 of Robinson (1933).

Given that many school districts compete for the same pool of teachers, we use an oligopsony model with hiring.

Our collective bargaining model consists of three components: labor demand with an offer game, labor supply, and Nash-in-Nash bargaining. The relative elasticity of labor demand and labor supply shape the size of the union and oligopsony distortions.

Labor demand is determined by the school district's input allocation decision, in which it decides how to allocate a fixed budget to hiring teachers versus purchasing other inputs such as buildings, computers, other staff, etc. We estimate the parameters of a constant elasticity of substitution (CES) objective function that determines how the school district makes this tradeoff. We estimate an elasticity of labor demand of $\epsilon_D = -0.72$. Based on their demand and the negotiated wage, school districts send out offers to a subset of teachers. When a district's negotiated wage is set above the level that equates labor demand and supply, not all workers will receive a job offer from that district, i.e., the equilibrium point will be on the labor demand curve. Otherwise, the equilibrium point will be on the labor supply curve.

Labor supply is determined by a mixed-logit employment choice model, where each potential teacher picks their favorite employer based on wages, commuting distance to each school district, the share of students who qualify for free-lunch, charter status, and a random coefficient on the outside employment option that allows for positive correlation in teachers unobserved preferences for jobs in public education. Since our model incorporates latent constrained choice sets due to labor market rationing, a traditional estimation strategy that assumes workers can obtain employment at all possible districts would bias our estimates (Crawford et al., 2021). Instead, we estimate our labor supply model using the quit decision, since conditional on tenure, all teachers can either choose to keep their current job or "quit" and leave the public education sector. Our estimation procedure adjusts the density of worker specific unobserved heterogeneity to account for the fact that it differs for those that either quit or stay versus those that switch employers.

We find that the estimated marginal rate of substitution between commuting time and wages implies that a teacher values each hour of commuting at $ 65. The average elasticity of labor supply at the school district level is 2.8.[2] However, since

[2]Using pre-determined salary schedules as an instrument, Ransom and Sims (2010) estimate a labor supply elasticity of 3.7 for teachers in Missouri. In a meta-study not specific to teacher labor markets, Sokolova and Sorensen (2021) report average firm-level elasticities between 4-6.

teachers may only receive offers from a subset of schools in equilibrium, the average elasticity of labor supply conditional on these offers is lower at 1.1. We also estimate cross-elasticities of labor supply and find that teachers are more likely to substitute to nearby school districts than to the outside option. Thus, when a school district lowers wages, teachers switch school districts rather than leaving the teaching profession.[3]

Salary negotiations are modeled with the Nash-in-Nash framework commonly employed in the bilateral oligopoly literature in empirical industrial organization. The payoff to the union is the wage bill, while the payoff to the employer is determined by the value of their utility function defined over teachers and other inputs into education. We estimate a bargaining power parameter for each school district-collective bargaining unit pair and find substantial dispersion in bargaining power, which echoes previous work such as Grennan (2013). This is because districts with similar budgets and labor supply conditions pay very different salaries.

After estimating the model, we generate equilibrium outcomes of wages and teachers for the social planner, posted wages, and collective bargaining environments. The model predicts average wages of $50,459 for the social planner, while predicting average wages of $42,526 for the posted wage setting. Each school district always pays lower wages to teachers in the posted wage setting compared to the social planner setting. Moreover, overall employment only drops from 121,875 to 118,590 teachers when moving from the social planner to the posted wage environment. Because demand and supply for labor are relatively inelastic, the monopsony distortion has larger effects on wages than on employment. In the collective bargaining setting, average wages are $54,190 which is 27% higher than the oligopsonist solution and is also 7 percent above the social planner solution. However, the number of teachers falls from 118,590 to 110,656 as job rationing occurs under collective bargaining.

Collective bargaining, however, does not lead to uniformly higher wages for teachers. Thirty percent of school districts have *lower* wages under collective bargaining relative to the oligopsony setting due to bargaining externalities between school districts. If employment at a district is rationed because wages exceed the market-clearing level, this expands the pool of teachers that other school districts can hire.[4] The supply of labor to neighboring school districts expands, which allows them to

[3]If a school is far away from other schools, teachers may prefer the outside option when wages fall at that school.

[4]We assume that these schools are in the labor demand binds case.

lower wages. This possibility has been discussed since at least the pioneering work of Lewis (1963). The complexity of the effects of bargaining power on outcomes in markets with externalities also recalls the work of Ho and Lee (2017) who find effects of market power in the market for hospital services that can raise or lower prices to consumers due to complex externalities.

We consider two alternative wage policies: minimum wages and uniform state wages. When the equilibrium point is on the labor supply curve, a minimum wage policy can be an effective tool to reduce the oligopsony distortion. In our setting, we find that a minimum wage performs poorly due to heterogeneity in the efficient wage across employers; under minimum wages, desirable schools are on their labor demand curves and restrict hiring. We also consider state-level wage setting, which exists, e.g., in North Carolina. This policy combines a) replacing an oligopsony of school districts with a state monopsonist and b) an identical wage in each school district. The increase in labor market power should allow the state to significantly depress wages; but less desirable schools will find it difficult to hire teachers. Thus, the state would trade off equity across school districts with profits from exercising monopsony power.

Related Literature This paper relates to several strands of literature at the intersection of labor, monopsony, and industrial organization.

We contribute to the long-standing literature on the impact of unionization on wages. Notable studies include Lewis (1963), Freeman and Medoff (1984), Card (1996), Lemieux (1998), DiNardo et al. (1996), DiNardo and Lee (2004), and Farber et al. (2021). These studies estimate the extent to which unions raise wages, either by using cross-sectional comparisons, longitudinal data containing individuals that switch to or from union jobs, or quasi-experiments like close union elections. Recent estimates based on regression comparisons by Blanchflower and Bryson (2024) indicate that unionized workers earn around 17.5 percent more than observationally equivalent non-unionized workers.

Despite the fact that U.S. public sector workers are about five times more likely to be unionized than private sector workers (Card et al., 2020), few studies consider union impact on public sector wages. Notable exceptions include Blanchflower and Bryson (2010), Card et al. (2020), and Baker et al. (2024). More narrowly, teaching is the most heavily unionized occupation in the United States, yet we know of

only two studies that consider union effects on the labor market for teachers and they reach mixed conclusions. Hoxby (1996) finds that district spending per student increases after teachers unionize. She argues that the spending increase is due to increased salaries, as class sizes do not fall after unionization. Lovenheim (2009) uses a difference-in-differences design with data from teacher union election results in Iowa, Indiana, and Minnesota to estimate the impact of teacher unions on school budget outcomes. He finds a small increase in employment, but no impact on teacher pay.

Our structural model with labor demand, labor supply, and wage setting allows us to quantify the effect of unions on the monopsony distortion and welfare, as well as allowing us to evaluate counterfactual policies. Our framework relates to an older strand of labor and industrial relations theory surveyed in Farber (1986). As that literature preceded the development of many current structural methods in applied microeconomics, there are scant attempts to use this work for quantitative prediction with Azkarate-Askasua and Zerecero (2024) being a notable exception, as well as the work of Dodini et al. (2021) and Green et al. (2022). There is a larger theoretical literature on patterns of unionization, such as Horn and Wolinsky (1988b).

From industrial organization, we draw on the recent literature that looks at bilateral oligopoly, i.e., situations where both the buyers and sellers of an input have market power, and where prices are determined by Nash-in-Nash Bargaining (see Horn and Wolinsky (1988a), and Collard-Wexler et al. (2019)). This literature starts with Crawford and Yurukoglu (2012) who study negotiations in cable television, or Grennan (2013) for stents, but has reached its most fully developed form in the work on negotiations between hospitals and health insurance companies, see Gowrisankaran et al. (2015) and Ho and Lee (2017), among others. The Nash-in-Nash concept allows us to deal with both bargaining and the oligopsonic interactions in the labor market that introduce externalities between negotiations.

This work in industrial organization is complemented by production side approaches to monopsony, such as Delabastita and Rubens (2025) and Rubens (2023). Our input allocation problem is an adaptation of this type of approach to the context of the public sector where the objective of profit maximization is nonsensical, but cost minimization does hold.

Third, there are a large number of recent papers looking into monopsony power, coming from a broad cross-section of economics. Some prominent papers are Berger et al. (2022), Gottfries and Jarosch (2023), Setzler (2025), and Kroft et al. (2025). A

useful survey is provided by Kline (2025).

Finally, there is a small literature that uses equilibrium models to understand the labor market for teachers. Some recent work is Bates et al. (2025) and Biasi et al. (2021). However, while we broadly relate to this work in education, our focus is on understanding equilibrium effects of bargaining between school districts and teachers on wages and employment rather than its effect on student outcomes. Indeed, baked into the model of collective bargaining over wages is a fundamental opposition between the interests of teachers and school districts.

The paper proceeds as follows. Section 2 presents a stylized version of our model to build intuition. Section 3 discusses the institutional features of the market for teachers in Pennsylvania and the data that we will use. Section 4 presents some preliminary empirical evidence on wage setting. Section 5 estimates the structural model and Section 6 discusses counterfactuals. We conclude in Section 7.

2 Stylized Model

We begin with a stylized model of collective bargaining and monopsony power from Chapter 25 of Robinson (1933) ("Monopolistic Exploitation of Labour"). We extend the model in Section 5 to account for the institutional features of the market and to allow for estimation.

Consider the case of a monopsonist employer that has a labor demand curve given by $L^D(w)$, where w is a uniform wage paid to all employees (see Figure 1, Panel a). Workers have a labor supply curve given by $L^S(w)$. The intersection of the labor supply and labor demand curve at point A yields the social planner solution, with wages, w^*, and L^* workers hired.

The marginal hiring cost for a monopsonist is given by $w + L^S \frac{\partial w}{\partial L^S} = w\left(1 + (\epsilon^S)^{-1}\right)$, where ϵ^S is the elasticity of the labor supply curve. The marginal cost of hiring is always greater than the labor supply curve, since to hire additional workers the monopsonist needs to pay inframarginal workers more.

As such, the monopsonist will choose to hire at point B, the intersection of labor demand and the marginal hiring cost, yielding L^0 workers and a lower wage w^0. The depression of wages and employment below the competitive level implies a deadweight loss from monopsony power, which is represented by the shaded triangle.

Notice that the size of the deadweight loss, and the degree of wage depression

depend on the interaction between the elasticity of labor supply and the elasticity of labor demand. Indeed, if labor demand is completely inelastic, then there is neither deadweight loss, nor depression of wages below efficient levels, no matter the elasticity of labor supply.

Now suppose that workers can collectively bargain with the employer, which is captured in Panel b) of Figure 1. In this case, if the union's well-being increases more rapidly in wages than the employer, wages will be higher than w_0. In addition, if the union is willing to trade off less employment to obtain higher wages, wages could be bargained to w^b above w^*. In this case, labor supply at point D is greater than labor demand at point E; there will be excess supply of workers. Thus, we need to make an assumption on whether labor is chosen by the union or the employer. We follow the classic Dunlop (1944) approach used in Farber (1978), and assume that employers choose employment on the labor demand curve.[5] Then, labor will be chosen on the labor demand curve at point E, and thus $L^b = L^S(w^b)$ workers will be hired. Again there is a deadweight loss, but with different incidence on workers and employers. The relative magnitudes of the deadweight losses in the case of monopsony in Panel a) of Figure 1 versus the deadweight loss in panel b) of Figure 1 cannot be signed purely from theory. As such, we will use empirical methods to quantify the relative sizes of these deadweight losses.

3 Institutional Background and Data

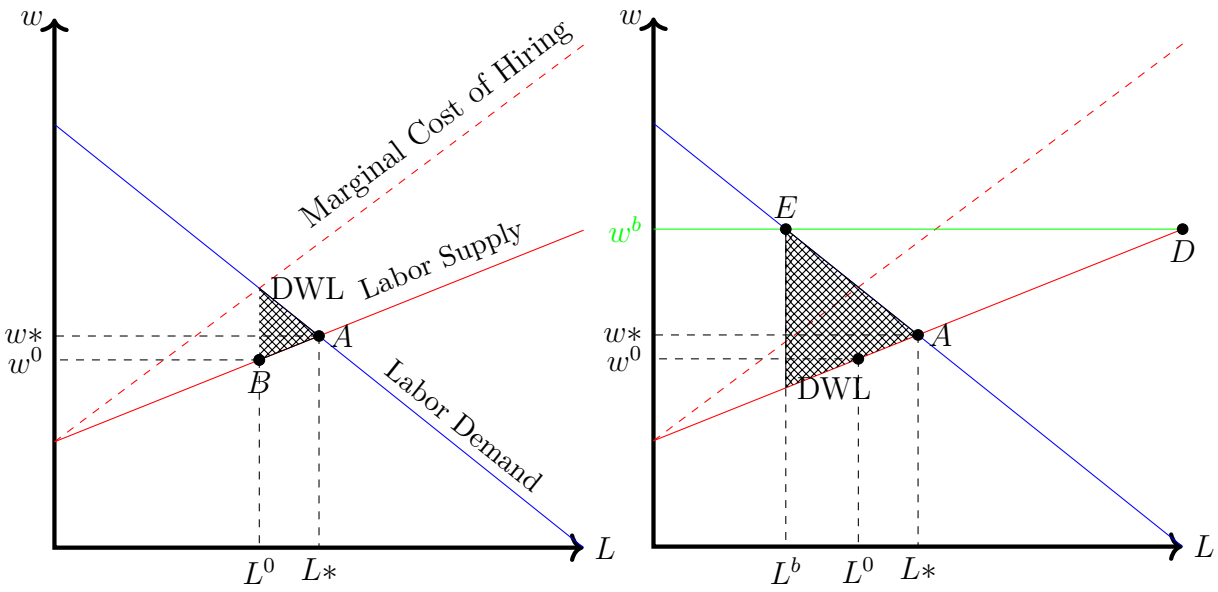
We study K-12 public school teachers, who are employed in regular public schools or charter schools in the state of Pennsylvania from 2009 to 2023.[6]

3.1 Teachers

Teachers are licensed professionals hired by schools to teach children. After three years, teachers become tenured, which is portable across all public school districts.

[5]The collective bargaining agreements for school districts in Pennsylvania do not specify employment levels in the contract, only wage schedules. The U.S. Port workers, where 50 percent of the union membership is paid, but not assigned to a particular job, provides an example for the case where employment levels are agreed upon in the union contract (see WSJ Article accessed January 25, 2025). See section 3 of Farber (1986) for more discussion on this point.

[6]For most of our analysis, we restrict our analysis to 2013-2019. First, teacher panel ids only begin in 2013. Second, we avoid using data after 2020 to avoid complications related to Covid-19.



(a) Posted Wage Monopsonist (b) Collective Bargaining

Figure 1: Welfare Analysis of Collective Bargaining and Monopsony Power

As a practical matter, once teachers obtain tenure, they are rarely fired, and also, are infrequently fired before obtaining tenure.

Teachers in Pennsylvania are paid through a school district-specific uniform wage schedule that depends only on a) years of experience, typically until 10 or 15 year in yearly steps, which transfers across school districts and b) education level (whether they have a masters or Ph.D. degree). An example schedule is shown in Figure 2 for Abington. Uniform wage schedules are important for two reasons. First, uniform wages are needed for monopsony power to distort hiring choices. If a school district can hire a marginal teacher by offering a higher wage only to that hire, the monopsony distortion on hiring would not exist.[7] Second, uniform wage schedules reveal the wage a teacher would earn in an alternative job, which is crucial for modeling labor supply.

Teachers are also compensated with benefits, such as retirement and health insurance plans, which make up about 30 percent of the cost of hiring a teacher in 2023. Benefits are standardized across school districts, so that job lock due to limited portability of benefits and employer benefit differentiation is irrelevant. For example, pension benefits are administered by the Pennsylvania Public School Employees'

[7]We assume complete information between workers and employers. Any information asymmetry would reintroduce such distortions.

APPENDIX A SALARY SCHEDULE

SCHOOL YEAR 2018-2019 (FY19)- 2.5%

Step	B	B+15	M	M+15	M+30	M+45	DOC
1	$41,998	$43,373	$45,322	$46,704	$48,286	$49,918	$51,549
2	$46,590	$48,123	$50,280	$51,740	$53,403	$55,034	$56,666
3	$49,059	$50,603	$53,328	$54,749	$56,705	$58,337	$59,969
4	$51,826	$53,532	$56,414	$57,874	$59,579	$61,211	$62,843
5	$54,422	$56,615	$59,579	$61,045	$62,711	$64,342	$65,974
6	$57,308	$59,870	$62,913	$64,332	$66,040	$67,672	$69,303
7	$60,468	$63,154	$66,156	$67,590	$69,287	$70,919	$72,551
8	$63,603	$66,484	$69,736	$71,235	$72,977	$74,609	$76,241
9	$66,728	$69,736	$73,024	$74,450	$76,109	$77,741	$79,373
10	$70,188	$73,228	$76,473	$77,899	$79,569	$81,201	$82,831
11	$74,467	$77,836	$81,655	$83,078	$84,780	$86,412	$88,043
12	$75,764	$79,167	$83,021	$84,460	$86,180	$87,812	$89,444
13	$76,655	$80,091	$83,986	$85,437	$87,172	$88,804	$90,436

Notes: B, M and DOC refer to bachelor, masters, and doctoral degree, plus additional credit hours. Steps are years of experience. Source: Page 47 of collective bargaining agreement: "Agreement Between Abington School Committee and Abington Education Association Effective September 1, 2018 to August 31, 2021" Accessed Jan 28, 2025.

Figure 2: Salary Schedule for Abington School District

Retirement System for all teachers.

3.1.1 Teacher Data

Our main data source is annual payroll data from Pennsylvania.[8] It contains the name, salary, experience, tenure, degree status, job description, demographic characteristics, and school district employer identification number for each teacher and non-instructional staff member from the academic years 2009-2010 through 2022-2023 inclusive. Beginning in the 2013-2014 school year, the data includes time-invariant employee identifiers, allowing us to track employees over time.

Second, we use data from InfoUSA that links names to home addresses for most Pennsylvania residents from 2010 to 2022.[9] In order to identify the commuting distance between teachers and school districts, we match the InfoUSA data to the payroll data on exact name and year. Since many teachers have common names (e.g Elizabeth Smith), we restrict attention to teachers who are uniquely matched in InfoUSA. We obtain a unique match for about 36.4% of teachers in the payroll data. We as-

[8] Penn Staff Data, accessed January 25, 2025.

[9] Note that InfoUSA rebranded as Data Axle in 2020. Data Axle's address data is generated using public sources, voter registration data, utility company data, and real estate data.

sume that teachers with unique names do not have significantly different residential patterns than other teachers. See Appendix A.2 for more details.

We use the HERE georouting application[10] to compute commuting time between a teacher's home address and the address of their current workplace. For the collective bargaining model, we approximate the full school-teacher commute time matrix by regressing our commute time data on as-the-crow-flies distance.

Table 1 presents summary statistics of the public school teachers in the data in 2017. We restrict our analysis to teachers that work at either school districts or charter schools, and avoid the 5% of public school teachers who work in juvenile penitentiaries, technical high schools, or are hired at the intermediate unit level. The majority of the remaining teachers, 94%, work in unionized public school districts with the remaining 6% working in charter schools. The median teacher is female, white, and has a master's degree. Most teachers are career teachers; teachers have 14 years of experience on average. The rate at which teachers leave the profession is U-shaped in experience: 6.5% of teachers exit after their first year on the job and the exit rate does not reach that level again until teachers have 27 years of experience.In addition, switching schools is relatively rare (only 2% every year). We find that teachers commute 20 minutes one-way on average and earn about $69,340.

In our model of labor supply, we allow teachers to choose the outside option of not teaching in the Pennsylvania public school system so that we can consider the extensive margin. These potential teachers work outside the Pennsylvania public school system as private school teachers, unemployed teachers, or in different occupations. We estimate the number of potential teachers from two sources: (1) newly licensed teachers who do not teach in Pennsylvania and (2) Pennsylvania teachers who quit in the previous four years with less than 20 years of experience. We find that about 28% of the qualified teacher pool does not work in the public school system. See Appendix A.1 for details on the construction of potential teachers.

3.2 Public Schools in Pennsylvania

In 2017, Pennsylvania had 499 regular school districts and 179 charter schools, with about 94% of public school teachers working for (and students attending) a regular

[10]We used the stata georoute package which matches to the HERE API.

Table 1: Summary Statistics of Teacher Data in 2017

	Mean	Standard Deviation	Observations
Charter Schools	0.06	0.23	109,638
Unionized	0.94	0.23	109,638
Percent Female	0.73	0.44	109,638
Percent White	0.95	0.23	109,638
Percent w/ Masters	0.60	0.49	109,638
Avg Years in Education	14.00	8.46	109,638
Avg. Commute Time	20.27	16.92	39,354
Average Wage	69,340.10	17,462.71	109,638
Exit Rate	0.04	0.20	109,638
Switching Rate	0.02	0.14	109,638

Notes: The summary statistics were computed on the 2017 population of public school teachers in Pennsylvania who work for either a public school district or a charter school that we use in our sample. We exclude teachers working at juvenile penitentiaries or technical high schools. The data was retrieved from the personnel files from the State of Pennsylvania. See Footnote 7 for details. Average commute time has missing data, because not all teachers were matched to unique names in the InfoUSA address data. In addition, the top 2.5% of commute time values were not included in the average computation.

school district.[11] Each school district belongs to one of 29 intermediate unit which manage curriculum, education support, and other administrative tasks. Intermediate units do not affect wages, and thus we treat each school district as a separate employer.[12] Figure A1 shows the geography of school districts and intermediate units in Pennsylvania.

We use data from the National Center for Education Statistics (henceforth NCES) from the common core of data,[13] which includes school characteristics like number of students, number of students who qualify for free lunch, and number of FTE teachers. We use financial data from the Annual Survey of School System Finances from the US Census Bureau,[14] which contains budget data by school district and year.

Out of 678 school districts and charter schools in 2017, we restrict our model to 486/499 school districts and 121/179 charter schools, which account for 97.5% of

[11] There are also penitentiary schools, technical career schools, and teachers hired at the intermediate unit level, but we do not consider them in our analysis. They make up 5% of public school teachers in our data.

[12] Intermediate Units hire about 3.2% of public school teachers that tend to be music teachers, PE teachers, and special needs teachers that likely travel between smaller school districts. We do not consider them in our analysis.

[13] Accessed at https://nces.ed.gov/ccd/files.asp.

[14] https://web.archive.org/web/20241203193341/https://www.census.gov/programs-surveys/school-finances.html accessed January 25, 2025.

teachers working in the 678 schools.[15] Table 2 presents summary statistics of our in sample schools. The table also includes 543 private schools, which hire about 11% of Pennsylvania's teachers.[16] Due to lack of payroll data, they are excluded from our model. Teachers who work for a private school have chosen the outside option.

Regular school districts are generally much larger than charter schools with over 3.5 times the students– 3,182 vs 866 students–and over 4 times the teachers–218.5 vs 53.5 teachers on average. School district size has large variation, with the Philadelphia School District being an order of magnitude larger than almost all other school districts. School districts tend to have lower student-teacher ratios of 14.02 students for each teacher on average compared to the charter's 15.92. Charter schools are concentrated in urban areas (72%), such as Philadelphia, and serve poorer students as measured by the percentage of students that are eligible for the government's free lunch program.[17] Private schools are small and put little competitive pressure on public school districts.

The lower panel of Table 2 contains summary statistics on school finances. School districts receive funding from federal, state, and local sources, but the majority is sourced locally (60%), especially for charter schools (80%). Local funding across districts varies substantially because it is usually sourced from local property taxes. Since school districts are much larger, they spend much more–$60.01 million compared to $11.46 million for charters. They also spend more per student–$18,230 vs $14,030 per student. Overall, schools tend to spend about the same amount on teachers as a share of their budget, i.e., about 30% of expenditures go towards teachers, while the remainder goes toward other inputs. Teachers cost about $114,435 on average, which includes salary and benefits. About 2/3 of the cost of a teacher is wages and salary.

[15] We drop 9 cyber charter schools, 38 schools with less than 20 teachers, and 24 schools due to missing data. Twenty-two are dropped due to missing wage data. In order to avoid compositional variation in teachers, we compute school wages by taking the median salary of all full-time teachers at the school with a bachelor degree, 5-7 years of experience, and salaries above $18,500. The 22 schools that were dropped because they lack a teacher that matches these constraints hired only 1.1% of all teachers and are small with a median value of 33 teachers.

[16] Private school data comes from the National Center for Education Statistics accessible at https://nces.ed.gov/surveys/pss/pssdata.asp.

[17] 11.7% of schools have missing free lunch share data in 2017. This accounts for 6.2% of schools when weighted by students. For the model, we impute this share from 2018.

Table 2: 2017 Summary Statistics of School Data

	Public			Private
	Full	School Dist	Charter	
Basic Facts				
Charter Schools	0.20	0.00	1.00	0.00
	(0.40)	(0.00)	(0.00)	(0.00)
Urban	0.27	0.16	0.72	0.42
	(0.45)	(0.37)	(0.45)	(0.49)
Num of Schools	4.62	5.52	1.00	
	(9.49)	(10.41)	(0.00)	
Num of Students	2,720.43	3,182.04	866.34	258.99
	(5,686.97)	(6,255.92)	(904.99)	(221.33)
Num of Teachers	185.57	218.44	53.54	25.93
	(335.26)	(366.77)	(44.52)	(23.00)
Num of Other Staff	39.86	46.38	13.69	
	(102.95)	(114.00)	(11.33)	
Student-Teacher Ratio	14.40	14.02	15.92	10.37
	(2.38)	(1.59)	(3.94)	(4.39)
Percent Students	0.51	0.45	0.81	
on Free Lunch	(0.26)	(0.22)	(0.25)	
Unionized	0.81	1.00	0.03	0.00
	(0.39)	(0.00)	(0.18)	(0.00)
Expenditures				
Total Expenditures (mil)	51.27	61.09	11.84	
	(156.70)	(173.65)	(12.67)	
Expenditure per Student (thou)	17.38	18.23	13.96	
	(4.76)	(4.37)	(4.77)	
Percent Funding from	0.60	0.52	0.89	
Local Sources	(0.22)	(0.18)	(0.06)	
Expenditure on Teachers (mil)	22.75	27.21	4.85	
	(40.61)	(44.25)	(3.65)	
Teacher Share of Exp	0.30	0.30	0.30	
	(0.04)	(0.04)	(0.06)	
Avg Teacher Cost	114,434.94	119,663.10	93,435.90	
	(22,013.68)	(18,002.14)	(24,126.58)	
Avg. Salary Share of	0.63	0.62	0.68	
Teacher Expenditure	(0.04)	(0.02)	(0.05)	
Obs	607	486	121	543

Notes: The table shows summary statistics for school districts, charter schools, and private schools in Pennsylvania in 2017 with 10 or more teachers employed. We only include public schools that are in our sample. The data was retrieved from NCES. The public school data comes from the School District Finance Survey (F33) and the School District Universe Survey Data. Private school data comes from the Private School Universe Survey Data that is collected every two years by the NCES.

3.3 Collective Bargaining

Teachers at every regular school district are represented by a local chapter of a union, the Pennsylvania State Education Association.[18] The union and schools negotiate a collective agreement every 3-5 years that incorporates teacher salary, retirement, health care, working hours, vacation, and termination rules, among other topics. Based on cross-referencing collective agreements with our data, we find that schools pay the wages stipulated in the contracts. Charter schools rarely negotiate with teacher unions; in 2016, only 14 out of 179 charter schools were unionized.[19] In our model, we assume that the local union chapter negotiates individually with each school district and does not consider how negotiations at one school will affect any other school district.Each school pays a single wage to all teachers that is determined as the median wage of all full-time teachers with a bachelor degree and 5-7 years of experience working at that school making at least $18,500.

4 Stylized Facts and Suggestive Evidence

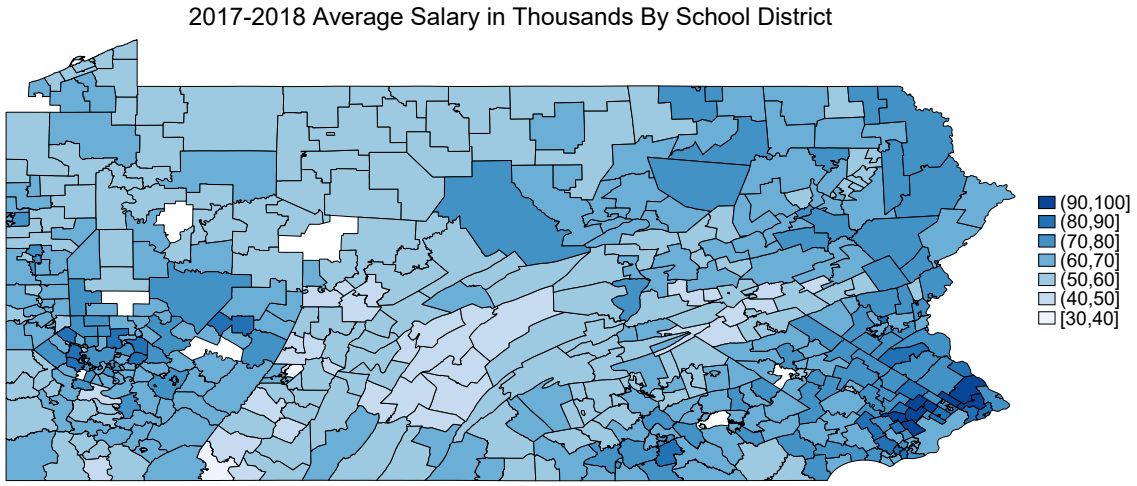
In this section we discuss the stylized facts that motivate our model of oligopsony power and collective bargaining.

4.1 Wage Variation

First, we want to understand the correlates of wages at different school districts. Salaries vary greatly across the state, from average salaries of $37,444 in Turkeyfoot Valley to $97,791 for Lower Merion (see Figure 3). Some of this salary dispersion can be explained by compositional differences in the workforce, such as having teachers with more experience and more education. However, these differences account for only one-third of the overall variance in salary dispersion, indicating large differences in salaries across employers unrelated to worker attributes. Interestingly, salary dispersion in Pennsylvania cannot be exclusively explained by local labor markets; locations such as Lower Merion and Philadelphia border each other and yet have average salaries that differ by over $20,000 per year. We believe that differences in wages can be explained at least in part by differences in bargaining power.

[18] For more information, see https://www.psea.org/.

[19] The National Alliance for Public Charter Schools collects this data.



Notes: The map displays the average real salary of teachers (average salary divided by regional price index) during 2017-2018 school year for each school district in Pennsylvania.

Figure 3: Average Salaries Vary Across Districts

4.2 Concentration

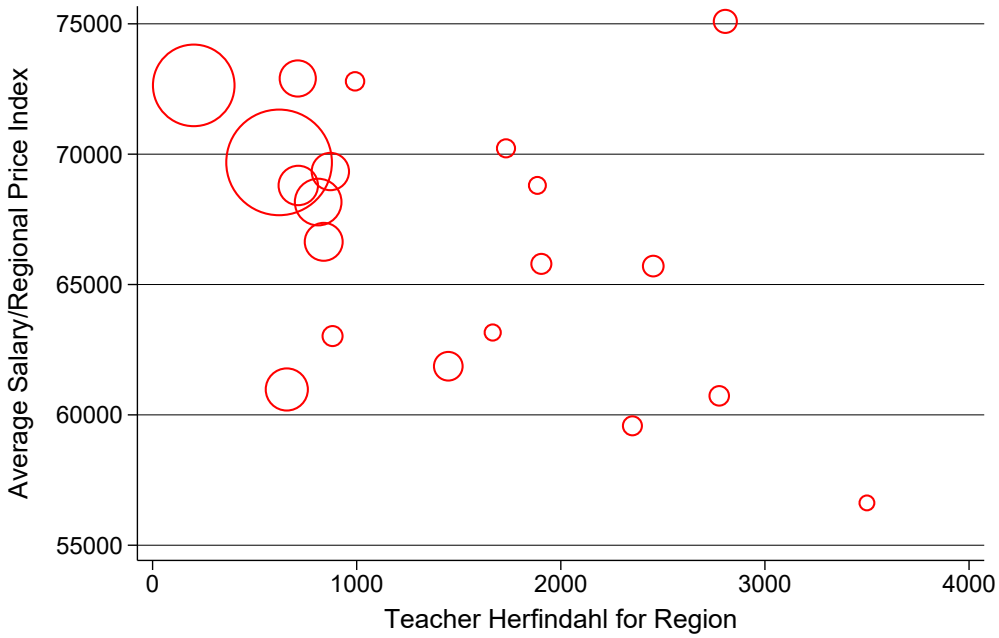
Next, we consider school district concentration in the market for teachers. We measure concentration in two ways: number of schools within a geographic region and the share of teachers within 10 miles of the school district that are hired by that school district.

First, we define local labor markets as Metropolitan Statistical Areas (MSA) from the Bureau of Labor Statistics.[20] Figure 4 shows a scatterplot of average wages at the regional level adjusted for cost of living against regional concentration measured by the HHI. The size of each circle is proportional to the number of teachers in that district. Figure 4 shows a negative relationship between wages and concentration. The OLS regression coefficient shows that an increase in HHI of 1,000 is associated with a $ 3,081 drop in wages (with a t-stat of -2.5, but with only 20 regions).[21] HHI varies from a low of 202 in Pittsburg to 3,500 for New York-Newark-Jersey City.

Second, we construct a measure of concentration that varies at the school district level. We compute the share of teachers that live within 10 miles of the school district

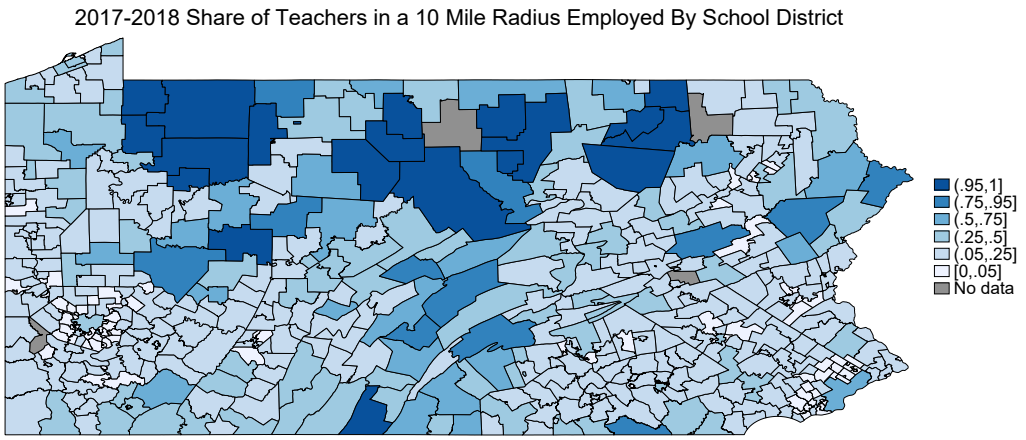
[20] See BLS Definition of Metropolitan Statistical Area, accessed Jan 28, 2025. We drop rural school districts that are not part of an MSA.

[21] We want to emphasize that these regressions show correlation, not causation. These regressions have a perilous history in empirical industrial organization as summarized by Miller et al. (2022), but have been treated with caution in the industrial organization literature since at least Demsetz (1973)'s work. One can get positive relationships between concentration and pricing without market power. Interpretation of these regressions should be done with care.



Notes: Each circle in the figure represents one of the 20 primary statistical areas in Pennsylvania. The Herfindahl is the sum of squared share of each employers teachers in the overall market, scaled by 10,000.

Figure 4: Regional HHI and Wages



Notes: The map displays the share of teachers that live within 10 miles of each school district that are employed by that district.

Figure 5: Concentration Varies Across Districts

that are employed by that school district. The higher the share of teachers a school hires, the more market power it should have. Figure 5 depicts these concentration measures at the school district level. Notice the large variation in the share of teachers within a ten mile radius that work for this school district — essentially zero in the suburban districts in Pittsburg, and over 95 percent in certain rural school districts.

Table 3 runs the regression between log salaries at a school district and either a) the log of the number of school districts within 10 miles, or b) the number of districts within 10 miles being less than 5, 5 to 15, or more than 15. A one percent increase in the number of school districts within 10 miles is associated with an increase in salary of 0.06 percent. This effect operates within a region. However, if we include other controls for density, such as the number of students living nearby, we cannot rule out zero or positive effects of concentration on wages.

4.3 Wages and Unionization: Charter Schools

Only 3% of charter schools are unionized,[22] and they pay wages that are much lower than regular school districts. In particular, they pay $50,087 on average to teachers compared to an average of $70,534 for regular school districts. Table 4 shows that the charter school salary penalty cannot be explained by teacher's experience, degree

[22]The National Alliance for Public Charter Schools collects this data: https://publiccharters.org/wp-content/uploads/2023/01/Unionized-Charter-Schools-2016-17_0.pdf.

Table 3: Correlation between Concentration and Salaries

	Salary Over Cost of Living Index		
	(1)	(2)	(3)
log(Districts Within 10 Miles)	0.06	0.05	
	(0.01)	(0.01)	
≤ 5 Districts within 10 Miles			-0.10
			(0.02)
> 5 and < 15 Districts within 10 Miles			-0.06
			(0.02)
Region FE		X	
Observations	1077398	1077398	1077398
R^2	0.52	0.54	0.59

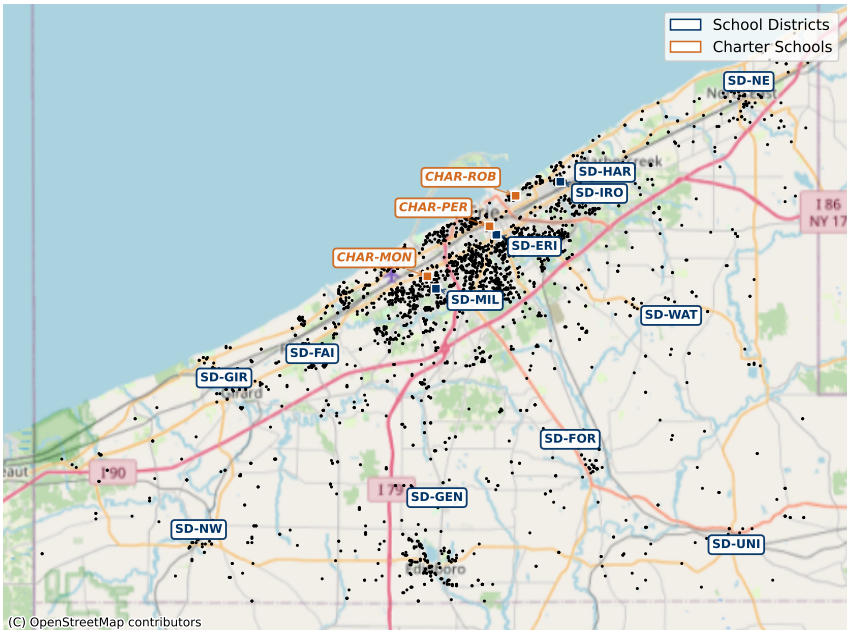
Notes: All regressions include controls for teacher experience, education, and whether the school is a charter school. Standard errors are clustered at the school district level.

status, and regional fixed effects. We believe the charter salary penalty arises due to the lack of a union rather than compensating differentials. The latter is unlikely, since teachers quit at much higher rates from charter schools, conditional on salary. Even so, we cannot interpret the difference in salary as the causal effect of unions on salaries. But, this fact motivates a modeling assumption where we allow non-unionized charter schools to post wages.

Table 4: Union Wage Gap

	Dependent Var: Log Real Wage		
Variable	(i)	(ii)	(iii)
1(Charter)	-0.340	-0.165	-0.326
	(0.016)	(0.018)	(0.004)
Teacher Covariates	No	Yes	Yes
Intermediate Unit F.E.'s	No	No	Yes
N	556394	556394	556394

Notes: The table summarizes point estimates and standard errors of a regression of the log of real salary on a charter school indicator. Column 2 adds a post-bachelor degree indicator and indicators for years of experience. Standard errors are clustered by Intermediate Unit.



Notes: The map shows the location of school districts and charter schools in Erie County, Pennsylvania in 2017 (except for Corry Area SD). Black dots represent teacher addresses, which were determined by merging teacher data with InfoUSA address data on exact names. Teacher addresses depicted here are from the years 2010-2022.

Figure 6: Schools and Teacher Location For Erie County, Pennsylvania

5 Structural Model

5.1 Overview of Model Components

We present a model of bargaining between school districts and teacher unions that incorporates a labor supply decision and an offer game. Figure 6 visually presents the data that we use to estimate the model. The figure shows the location of teachers[23] (black dots) and school districts in Erie County in northwest Pennsylvania.

Our model proceeds in three stages. First, wages are negotiated through a Nash bargaining process between school districts and teacher unions. Since negotiated wages at one school affect the labor supply of neighboring schools, we find the wage solution that leads to a Nash equilibrium, i.e., we model a Nash-in-Nash bargaining game. Our Nash bargaining model nests the posted price solution; when the bargaining weight equals one, teacher unions have zero bargaining power. Second, given wages, school districts choose which teachers to send offers to based on their labor

[23]The location also includes potential teachers, i.e., teachers who are qualified, but not teaching in 2017. See Appendix A.1 for details.

demand. Third, workers, with offers in hand, choose which school to work at.

The model is static in two important ways. First, the model assumes that hiring and job choice happen at one point in time. A model that accounts for the entire life-cycle of a teacher's career would be challenging to incorporate into the labor supply. Second, negotiations happen simultaneously.[24]

We will proceed through the three stages of the model in reverse order, that is labor supply, labor demand, and wage determination.

5.2 Labor Supply

We model the teacher labor supply decision using a mixed logit framework with constrained choice sets. Teacher $i \in \{1, \cdots, N\}$ either receives utility u_{ij} from working in a school district indexed by $j = 1, \cdots, J$:

$$\begin{aligned} u_{ij} &= \delta_{ij}(\nu_i) + \epsilon_{ij}; \\ \delta_{ij}(\nu_i) &= \psi\omega_j + x_j'\beta + \tau d_{ij} + \sigma\nu_i, \end{aligned}$$

or from leaving public education, denoted by $j = 0$:

$$u_{i0} = r + \epsilon_{i0}.$$

The teacher's utility from working at school j depends on the salary offered ω_j and a vector of non-salary attributes x_j which includes the fraction of students eligible for free lunch and an indicator for whether the employer is a charter school. The term r determines the mean value of the outside option.

The choice model allows for both observable and unobservable heterogeneity in preferences across teachers. Observable heterogeneity is captured through d_{ij}, the commuting distance between teacher i's home address and the centroid of the school district j.[25] Unobservable heterogeneity depends on two terms: 1) ϵ_{ij}, which follows a Type 1 extreme value distribution that is independent across both i and j, and

[24] Collective agreements are renegotiated every 4-5 years and are not negotiated simultaneously. Accommodating this feature would introduce dynamics into the bargaining problem which are very complex. Lee and Fong (2013) discuss the issues associated with it. Collard-Wexler et al. (2019) shows a model where forcing agreements to occur simultaneously, or allowing them to happen at different times, does not change the equilibrium outcome.

[25] The weighted average distance of the centroid to any school within a district is 1.58 miles. No school is more than 20 miles away from the centroid.

2) ν_i, which is independently distributed $\mathcal{N}(0,1)$ across teachers. The ν_i term is common across all public teaching jobs, such that, given a salary reduction, teachers will switch more often to other public education jobs rather than leaving teaching.[26]

Furthermore, this choice model allows teachers to have different choice sets. Each teacher i has a list of job offers summarized by the vector $\boldsymbol{\mathcal{O}}_i$, with element $o_{ij} = 1$ if teacher i receives an offer from district j and zero otherwise. Teachers can always choose the outside option. Each teacher chooses the employment option within their choice set that yields the highest utility. Since our ϵ errors are distributed Type 1 Extreme Value, given ν_i, we can represent the probability that teacher i chooses to work at district j as:

$$s_{ij}(\nu_i) = \frac{\exp(\delta_{ij}(\nu_i))o_{ij}}{\exp(r) + \sum_k \exp(\delta_{ik}(\nu_i))o_{ik}}. \tag{1}$$

We integrate over ν_i to get:

$$s_{ij} = \int_{\nu} s_{ij}(\nu)\phi(\nu)d\nu. \tag{2}$$

Given a vector of wages and the estimated parameters, we can aggregate the individual choice probabilities to create a firm-level teacher supply curve for each school district:

$$T_j = \sum_{i \in \boldsymbol{\mathcal{O}}_j} s_{ij}, \tag{3}$$

where $\boldsymbol{\mathcal{O}}_j$ is the set of teachers that receive an offer from school j, i.e., $\boldsymbol{\mathcal{O}}_j = \{i \text{ s.t. } o_{ij} = 1\}$.

5.3 Estimation with Unobserved Offers

The choice probability given by Equation (2) is the common mixed logit. Unfortunately, we do not observe job offer sets, which in our model are incomplete due to rationing. Erroneously assuming that teachers receive all offers would bias estimates

[26] As is common in the IO literature on discrete choice in oligopoly, it is important to understand whether teachers are more likely to switch between schools or between the teaching profession and the outside option. The logit model by itself would assume that switching schools and leaving education have the same horizontal differentiation.

of the marginal utility of job attributes towards zero.[27] Instead, we leverage the fact that conditional on ν, Independence of Irrelevant Alternatives (IIA) holds. We assume that teachers always have the option of either keeping their current job (stay) or leaving the public education sector (quit).[28] Conditional on $\nu_i = \nu$, the likelihood of quitting or staying in one's current job is given by:

$$\begin{aligned}\Pr(\text{quit}|\text{quit or stay}) &= \frac{\Pr(\text{quit})}{\Pr(\text{quit or stay})} \\ &= \frac{\exp(r)}{\exp(r) + \sum_k \exp(\delta_{ik}(\nu_i))o_{ik}} \times \frac{\exp(r) + \sum_k \exp(\delta_{ik}(\nu_i))o_{ik}}{\exp(r) + \exp(\delta_{ij}(\nu_i))} \\ &= \frac{\exp(r)}{\exp(r) + \exp(\delta_{ij}(\nu_i))}.\end{aligned} \tag{4}$$

This result follows immediately from logit-driven IIA as discussed by McFadden (1978) for estimation with a subset of choices. We will call $s_{iq}(\nu, \psi, \beta, \tau, \sigma) \equiv \Pr(\text{quit}|\text{quit or stay})$.

Thus, we can form a likelihood on quit decisions given by:

$$\begin{aligned}\mathscr{L}(\nu, \psi, \beta, \tau, \sigma) = \int_\nu \Bigg(\sum_i 1(q_i = 1) \log[s_{iq}(\nu, \psi, \beta, \tau, \sigma)] \\ + (1(q_i = 0) \log[(1 - s_{iq}(\nu, \psi, \beta, \tau, \sigma))]\Bigg) \phi(\nu) d\nu,\end{aligned} \tag{5}$$

If there were no teachers that switch districts, the density of ν conditional on quit or stay would be standard normal (e.g. $\mathcal{N}(0, 1)$) and Equation 5 would be a mixed binary logit. When switching rates are low, as in our application, misspecification bias from ignoring the selection problem is small. We propose an algorithm to correct for the selection bias and use monte carlo experiments to validate our procedure and show that the selection bias is small when switching is low.[29]

5.3.1 Selection Correction Algorithm due to Job Switchers

When some teachers switch jobs, $\phi(\nu)$ no longer equals $\phi(\nu|\text{quit or stay})$ due to a selection issue: teachers who quit teaching tend to have a lower value of working in

[27] Usually, researchers cannot observe the entire set of offers. Job search is often sequential and terminates when one has a good enough offer. For example, Bates et al. (2025) shows that principals send out very few offers to teachers to fill a position.

[28] Teachers are rarely fired, even before they receive tenure after three years.

[29] More details on the monte carlo experiments are in appendix F.

public education, i.e., ν.[30] The sample of teachers who quit or stay has a greater proportion of quitters than the full sample of teachers with movers, i.e., job switchers (see Figure A5 in the appendix). Therefore, the conditional distribution puts more weight on lower values of ν. We correct for this selection by directly estimating the conditional distribution of ν:

$$f(\nu|\text{stay or quit}, \boldsymbol{\mathcal{O}}_i) = f(\nu) \underbrace{\frac{\Big(\Pr(quit|\nu_i, \boldsymbol{\mathcal{O}}_i) + \Pr(stay|\nu_i, \boldsymbol{\mathcal{O}}_i)\Big)}{\Big(\Pr(quit|\boldsymbol{\mathcal{O}}_i) + \Pr(stay|\boldsymbol{\mathcal{O}}_i)\Big)}}_{\text{weight}}. \tag{6}$$

Notice that when the number of job movers is small, the weight in Equation 6 will be close to 1, so that estimation bias from movers is small.[31] This intuition is confirmed by the Monte-Carlos (see section F in the appendix). Table A9 shows that accounting for selection has a negligible impact on the estimated coefficients given a job move rate of 3% which is close to the 2% share in our own data (see Table 1 for the move rate). If a larger share of workers moved, then our coefficients would be biased. For these cases, which lie outside the empirical context of this paper, we can use the structure of our model to estimate each piece of the weight in Equation 6.

First, we estimate the propensity score for quitting and staying denoted as $\Pr_i(quit|\boldsymbol{\mathcal{O}}_i)$ and $\Pr(stay|\boldsymbol{\mathcal{O}}_i)$ using a non-parametric logit regression. Notice that we do not observe the offer set $\boldsymbol{\mathcal{O}}_i$. However, the model of offers and labor demand we present in section 5.5 is a deterministic function of the characteristics of teachers and of schools, which we denote as Z_i.[32] Thus, we can express the offer set as a function of Z, $\boldsymbol{\mathcal{O}}_i(Z_i)$, and $\Pr_i(quit) = \Pr(quit|\mathcal{O}_i(Z_i)) = \Pr(quit|Z_i)$. Notice that the probability of quitting combines variables that affect δ_{ij} (e.g., wages, school characteristics and commuting distances), and variables Z that affect the offer set $\boldsymbol{\mathcal{O}}_i$ and are excluded from the payoff function (e.g., number and budgets of nearby school districts).The probability of staying, $\Pr(stay)$, is estimated similarly. In the appendix, Table A10

[30] We thank Steve Berry for alerting us to this issue, and to Paulo Somaini (and Steve Berry) for useful conversation on this issue.

[31] To see this more clearly, notice that the term *weight* is also given by:

$$\frac{1 - \Pr(move|\nu_i, \boldsymbol{\mathcal{O}}_i)}{1 - \Pr(move|\boldsymbol{\mathcal{O}}_i)},$$

which converges to 1 as $\Pr(move) \to 0$.

[32] The only exception is that offers depend on the realization of logit shocks ϵ_{ij}.

shows our estimated quitting and staying propensity scores and Figure A6 shows the histogram of predicted stay, move, and quit probabilities.

Second, we tackle the probabilities conditional on ν_i and $\boldsymbol{\mathcal{O}}_i$ and focus on the probability of quitting:

$$\Pr(quit|\nu, \boldsymbol{\mathcal{O}}_i) = \frac{1}{1 + \exp(\sigma\nu)\gamma_i}, \tag{7}$$

where $\gamma_i \equiv \sum_k \exp(\delta_{ik}) o_{ik}$. Given σ, we can recover γ_i from $\Pr(quit|\boldsymbol{\mathcal{O}}_i)$ since $\Pr(quit|\boldsymbol{\mathcal{O}}_i)$ is strictly decreasing in γ_i, $\Pr(quit|\boldsymbol{\mathcal{O}}_i)$ is known (estimated), and $f(\nu)$ is a normal density.

$$\Pr(quit|\boldsymbol{\mathcal{O}}_i) = \int_\nu \Pr(quit|\nu, \boldsymbol{\mathcal{O}}_i) f(\nu) d\nu = \int_\nu \frac{1}{1 + \exp(\sigma\nu)\gamma_i} f(\nu) d\nu \tag{8}$$

Third, we estimate $\Pr(stay|\nu, \boldsymbol{\mathcal{O}}_i)$ and exploit the fact that movers and stayers have the same distribution of ν, which follows directly from IIA when teachers do not choose the outside option. The relative odds of taking any two teaching jobs j and k are independent of ν:

$$\frac{\Pr(k|\nu_i, \boldsymbol{\mathcal{O}}_i)}{\Pr(j|\nu_i, \boldsymbol{\mathcal{O}}_i)} = \frac{\exp(\delta_{ik} + \sigma\nu_i)}{\exp(\delta_{ij} + \sigma\nu_i)} = \frac{\exp(\delta_{ik})}{\exp(\delta_{ij})}.$$

Each teacher either stays, quits, or moves jobs, so we can express $\Pr(stay|\nu_i, \boldsymbol{\mathcal{O}}_i)$ in terms of $\Pr(quit|\nu_i, \boldsymbol{\mathcal{O}}_i)$, $\Pr(move|\boldsymbol{\mathcal{O}}_i)$, and $\Pr(stay|\boldsymbol{\mathcal{O}}_i)$ all of which are known:[33]

$$\Pr(stay|\nu_i, \boldsymbol{\mathcal{O}}_i) = \Big(1 - \Pr(quit|\nu_i, \mathbf{O_i})\Big)\left(1 + \frac{\Pr(move|\mathbf{O_i})}{\Pr(stay|\mathbf{O_i})}\right)^{-1} \tag{9}$$

Thus, we can construct the term labeled *weight* in Equation 6 and obtain the density of unobservable heterogeneity conditional on not switching jobs.

Since we never use choices between school districts, it is difficult to separately identify σ, at least without relying on the functional form of the normal distribution over ν. As such, we estimate r and σ using an indirect inference procedure that we discuss in Section 5.7.

[33] We use $1 = \Pr(quit|\nu_i, \boldsymbol{\mathcal{O}}_i) + \Pr(stay|\nu_i, \boldsymbol{\mathcal{O}}_i) + \Pr(move|\nu_i, \boldsymbol{\mathcal{O}}_i)$, Bayes Theorem $\Pr(move|\nu_i, \boldsymbol{\mathcal{O}}_i) = \frac{Pr(move|\boldsymbol{\mathcal{O}}_i)\Pr(\nu|move, \boldsymbol{\mathcal{O}}_i)}{\Pr(\nu)}$, and that $\Pr(\nu|move, \boldsymbol{\mathcal{O}}_i) = \Pr(\nu|stay, \boldsymbol{\mathcal{O}}_i)$.

5.4 Labor Supply Quitting Regressions

Table 5 presents estimates of the quitting versus staying decision using the likelihood given by Equation 5 conditional on different values of σ. The first three columns present maximum likelihood estimates on the sample of teachers with fewer than 15 years of experience. Column 1 corresponds to the case where $\sigma = 0$ and the model collapses to a binary logit. Each coefficient estimate is significant at the .05 level and has an intuitive sign.

Column 3 corresponds to the value of σ from our indirect inference procedure. The marginal effects (numbers in brackets under standard deviations) are essentially the same as those from the binary logit model. Assuming 180 school days per year and two trips to school per day, the coefficient on commute time and real salary implies a marginal rate of substitution (MRS) between salary and commute time of \$65.50 per hour.[34] The second column corresponds to an intermediate value of σ. Columns 1-3 show that the implied MRS between commute time and salary is almost the same for different values of σ. Column 4 shows that the parameter estimates and MRS are robust to dropping teachers without tenure, which is obtained after three years of experience. Column 5 shows that the parameters become blunted as we expand the sample to include teachers with more than 15 but less than 25 years of experience, likely because retirement becomes more salient in determining the decision to leave education. Despite this, the marginal effects are roughly the same as in column 3. Column 6 adds indicators for each of the MSA's in Pennsylvania along with an indicator for schools not in an MSA.

5.4.1 Labor Supply Elasticities

The estimated labor supply model can be used to construct a school-level labor supply elasticity matrix. The elasticity of supply to school district j with respect to salary

[34] There is an extensive literature on assessing the Value of Travel Time (VTT). The U.S. Department of Transportation, in say, Table 4 of U.S. Department of Transportation (2016) uses a value of \$25.40 for business travel, and discusses ranges found in the literature. More recent work in industrial organization also emphasizes the variation in VTT between individuals, such as Buchholz et al. (2025).

Table 5: MLE Estimates of the Quit Decision

Variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Real Salary	-0.229	-0.394	-0.696	-0.562	-0.270	-0.587	-0.531
(10k)	(0.019)	(0.033)	(0.059)	(0.072)	(0.044)	(0.069)	(0.109)
	[-0.0093]	[-0.0090]	[-0.0077]	[-0.0058]	[-0.0028]	[-0.0062]	[-0.0007]
Commute Time	0.009	0.015	0.028	0.026	0.018	0.020	0.025
(Minutes)	(0.001)	(0.003)	(0.005)	(0.005)	(0.004)	(0.006)	(0.007)
	[0.0004]	[0.0004]	[0.0003]	[0.0003]	[0.0002]	[0.0002]	[0.0000]
Free Lunch Share	0.583	1.038	1.815	1.284	1.882	1.431	1.399
	(0.162)	(0.294)	(0.544)	(0.524)	(0.443)	(0.374)	(0.797)
	[0.0238]	[0.0236]	[0.0200]	[0.0133]	[0.0197]	[0.0152]	[0.0019]
Charter	0.735	1.404	2.550	2.905	2.646	2.480	4.458
	(0.111)	(0.199)	(0.359)	(0.346)	(0.342)	(0.393)	(0.573)
	[0.0300]	[0.0319]	[0.0282]	[0.0301]	[0.0277]	[0.0263]	[0.0062]
σ	0	4.425	8.85	8.85	8.85	8.85	8.85
$Pr(quit)$	0.050	0.050	0.050	0.044	0.047	0.050	0.044
Experience	< 15 *years*	< 15 *years*	< 15 *years*	< 15 *years* > 3 *years*	< 25 *years*	< 15 *years* > 3 *years*	< 15 *years* > 3 *years*
Observations	66,745	66,745	66,745	54,962	110,299	54,962	46,467

Notes: The table reports maximum likelihood estimates of mixed binary logit models for different values of the random coefficient σ. The likelihood was computed using Gauss-Hermite quadrature with 30 nodes and estimated on full-time teachers from the 2013-14 through 2018-19 academic years. Columns 1-3 restrict the sample to teachers with less than 15 years experience. Column 5 conditions on teachers with less than 25 years of experience. Columns 4, 6 and 7 restrict the sample to teachers with more than 3 but less than 15 years of experience. All specifications include year indicators. Column 6 adjusts the quadrature weights to account for selection. Column 7 contains indicators for MSA's. Standard errors clustered by school district are in parentheses. Average marginal effects are in brackets.

at school k is given by:

$$\begin{aligned}\eta_{jk} &= \frac{\partial log(T_j)}{\partial log(\omega_k)} = \frac{\omega_k}{T_j}\sum_{i\in N}\int_\nu \frac{\partial s_{ij}(\nu)}{\partial \omega_k}\phi(\nu)d\nu \\ &= \begin{cases} \frac{\omega_k}{T_j}\sum_{i\in N}\int_\nu \psi s_{ij}(\nu)(1-s_{ij}(\nu))\phi(\nu)d\nu & \text{if } k=j, \\ \frac{\omega_k}{T_j}\sum_{i\in N}\int_\nu -\psi s_{ij}(\nu)s_{ik}(\nu)\phi(\nu)d\nu & \text{if } k\neq j \end{cases}\end{aligned} \tag{10}$$

Using 2017 wages and teachers from our data, we find that the labor supply elasticities given that teachers receive offers from all schools range between 1-2. [35] Table 6 illustrates the elasticity matrix for some of the schools in Erie County, including charter schools. The entry in column j and row k is equal to η_{jk}. Notice that the cross-wage elasticities between school districts depend on distance; schools that are farther away tend to have smaller cross-wage elasticities in magnitude. For example, compare the cross-elasticities between SD-ERI and SD-MIL to those between SD-ERI and SD-NW which are further apart.[36] Teachers can also choose to switch to the outside option. The mixed logit structure is important: when wages rise, most of the new teachers come from nearby schools, not from other occupations. Note that elasticities depend crucially on the set of job offers each teacher has. When jobs are rationed in some districts, these relationships can break down as highlighted in Section 5.8.

5.5 Labor Demand

In order to hire teachers, school districts send out offers $\boldsymbol{\mathcal{O}}_j = \{o_{ij}\}_{i=1}^N$ sequentially to the closest teachers, where the function $i(n)_j$ will indicate which teacher i receives the n^{th} offer from school district j.[37] This is an intuitive rationing rule that avoids introducing misallocation due to travel costs. Given this ordering of offers, the number of offers by school district j, denoted as O_j, pins down the offer set $\boldsymbol{\mathcal{O}}_j$. Thus, the labor demand decision reduces to each school choosing its number of offers, represented by the vector $\mathbf{O} \equiv \{O_1, O_2, \cdots, O_J\}$.

[35] Using a quasi-random experiment, Ransom and Sims (2010) find a labor supply elasticity of 3.7 for teachers in Missouri. Other monopsony papers (e.g., Sokolova and Sorensen (2021)) find most elasticities between 4-6 across labor markets.

[36] Figure 6 shows the locations of the schools.

[37] In practice, only teachers within 60 minutes of a school district centroid are part of the school's labor market. Given estimated travel costs, few teachers would choose these offers.

	S-CON	S-ERI	S-MIL	S-PEN	C-ROB	C-MON	Out Opt
All Offers							
S-CON	2.804	−0.115	−0.135	−0.261	−0.105	−0.133	−0.033
S-ERI	−0.03	2.411	−0.054	−0.037	−0.058	−0.055	−0.008
S-MIL	−0.232	−0.353	2.537	−0.252	−0.355	−0.35	−0.055
S-PEN	−0.321	−0.17	−0.18	2.592	−0.163	−0.178	−0.041
C-ROB	−0.0	−0.001	−0.001	−0.001	2.28	−0.001	−0.0
C-MON	−0.001	−0.001	−0.001	−0.001	−0.001	1.923	−0.0
Out Opt	−1.093	−1.054	−1.055	−1.083	−1.054	−1.054	2.305

Notes: The table shows own- and cross-elasticities of a subset of schools Erie and Crawford county when teachers get offers from every school district. S refers to a school district, while C refers to a charter school. See the map in Figure **??** for the location of the schools. Each element shows the percentage increase in teachers at the schools in the columns willing to work after a 1% increase in the wage paid by the schools in the rows.

Table 6: Elasticities: Erie and Crawford County

School districts choose O_j to maximize an objective function $W(T_j(O_j, \mathbf{O}_{-j}), X_j)$ that depends upon two inputs: the number of teachers employed, T_j, and all other educational inputs, X_j, such as non-teacher staff, equipment, buildings, and land. The number of teachers hired $T_j(O_j, \mathbf{O}_{-j})$ depends not only on the offers school j makes but also on the offers made by competing schools, denoted by the vector $\mathbf{O}_{-j} \equiv \{O_k\}_{j \neq k}$. The school's welfare function has a Constant Elasticity of Substitution (CES) form, which allows the elasticity of labor demand to be somewhat flexible. Imposing the budget constraint $B_j = X_j + w_j T_j$ yields the problem:

$$\begin{aligned} \max_{O_j} W(T_j(O_j, \mathbf{O}_{-j}), X_j) &= \Omega_j \Big(\gamma T_j(O_j, \mathbf{O}_{-j})^\rho + (1-\gamma)(B_j - w_j T_j(O_j, \mathbf{O}_{-j}))^\rho \Big)^{\frac{1}{\rho}} \\ s.t. \qquad & T_j(O_j, \mathbf{O}_{-j}) = \sum_{n=1}^{O_j} s_{i(n)jj}(\mathbf{O}_{-j}), \\ & O_j \leq N, \end{aligned} \tag{11}$$

where B_j is district j's budget, w_j is the cost of hiring a teacher, the price of other goods X_j is normalized to one, Ω_j is the district's efficiency, and N is the number of teachers in the labor market.[38]

[38]This is a model with common information, which would imply that all agents know all stochastic shocks including the logit shocks ϵ_{ij} and the normal shock ν_i. Rather than simulating the set of draws ϵ_{ij} and ν_i (and doing so for many simulation draws) and solving the model for a fixed set of draws, we sum over choice probabilities s_{ij}. As the number of teachers given offers increases to

If $O_j < N$, the district is able to hire enough teachers to meet labor demand at the prevailing wage. In this case, the optimal number of teachers hired equates the marginal rate of technical substitution between teachers and other inputs to the relative price of hiring a teacher:

$$\frac{\frac{\partial W}{\partial T}}{\frac{\partial W}{\partial x}} = \frac{\gamma}{1-\gamma} \cdot \left(\frac{T_j}{X_j}\right)^{\rho-1} = w_j. \tag{12}$$

Imposing the budget constraint and rearranging yields school j's labor demand curve:

$$T_j(w_j) = \frac{w_j^{\frac{1}{1-\rho}} B_j}{(\frac{\gamma}{1-\gamma})^{\frac{1}{\rho-1}} + w_j^{\frac{\rho}{\rho-1}}} \tag{13}$$

Since labor supply exceeds labor demand in this case, schools send out limited job offers.

When $O_j = N$ the district cannot hire as many teachers as it would like at the prevailing wage. In this case the benefit of hiring another teacher exceeds the cost:

$$\frac{\frac{\partial W}{\partial T}}{\frac{\partial W}{\partial x}} = \frac{\gamma}{1-\gamma} \cdot \left(\frac{T_j}{X_j}\right)^{\rho-1} > w_j \tag{14}$$

and the district makes offers to every teacher in the market: $T_j = \sum_{n=1}^{N} s_{i(n)jj}(\mathbf{O}_{-j})$.

An equilibrium in the offer game is defined by a vector of the number of offers sent by each district, $\mathbf{O}^*$, such that each district's number of offers, O_j^*, maximizes its payoff function given the number of offers made by other districts, $\mathbf{O}_{-j}^*$. We compute the equilibrium using best-response function iteration (see Algorithm 2 in the appendix).[39]

By log-transforming Equation 12, we obtain a reduced form equation to estimate

infinity, the sum of the choice probabilities will converge to the actual number of teachers hired. In our model, the average school district sends out over 8,000 offers, and the smallest school district makes 36 offers, and thus, the realized acceptances differ little from the expected ones. We thank Pat Kline for pointing this out to us.

[39] Although we cannot prove that there is a unique offer game equilibrium, starting with different offer guesses had no significant impact on the final offer matrix. In addition, we randomized the order of school districts and found no difference in the final offer matrix.

labor demand:[40]

$$\begin{aligned}\log(w_j T_j) - \log(X_j) &= \frac{1}{\rho - 1}\left[-log(\frac{\gamma}{1-\gamma}) + \rho \log(w_j)\right] \\ &= \gamma_1 + \rho_1 \log(w_j).\end{aligned} \tag{15}$$

Table 7 presents estimates constructed by fitting Equation 15 to the data using OLS. The independent variable, labeled teacher marginal factor cost, is total spending on teachers over the number of teachers hired, i.e., wages and benefits spent on the average teacher.[41] We exclude observations from non-unionized charter schools, since they never offer salaries high enough for labor demand to bind. Column 1 presents the results from a simple bivariate regression. Column 2 adds year fixed effects, Column 3 includes both year and region fixed effects, and Column 4 includes year and school district fixed effects. In each case, the coefficient on the teacher marginal factor cost is significantly different from zero, so that Cobb-Douglas preferences are rejected. The implied average elasticity of demand ranges from -0.68 to -0.80, and is -0.72 in our preferred specification.

The first four columns of Table 7 present results estimated on all unionized school districts, despite the fact that the first-order condition used to derive the estimating condition holds only for districts where the wage is high enough that labor demand determines employment. Column 5 presents estimates of labor demand conditional on the set of schools we classify as being on the labor demand curve given the OLS estimate of ρ in Column 4. Of the 486 school districts only 2 were not on the labor demand curve. Thus, the parameter and elasticity estimates are essentially unchanged. Figure 7 presents our estimated labor supply and demand curves for the Millcreek Township school district, which is a mid-sized school district in Erie County.

5.6 Wage Negotiation

For all unionized school districts, the wages in our model are determined via simultaneous negotiations between each school district and the union representing teachers in

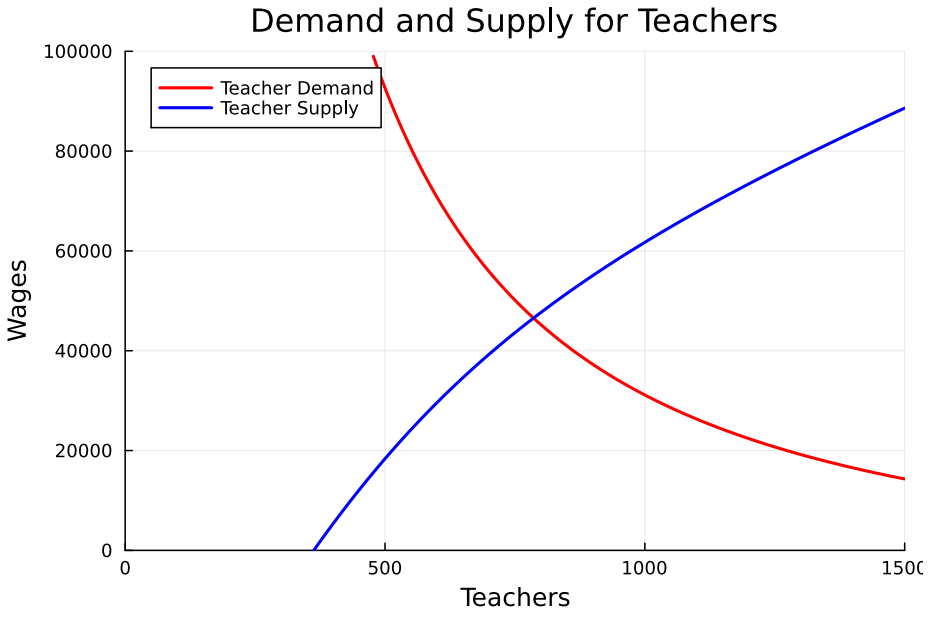
[40] Combining equations 12 and 14 would yield an inequality condition on $\log(w_j T_j) - \log(X_j)$ for the case where labor supply binds, and equality condition when labor demand is binding. In the data we find overwhelmingly that labor demand is binding, so we present the simpler OLS regression.

[41] The wage, w_j, in Equation 15 is the marginal factor cost of a teacher, which includes both wage and benefits, which better approximates the cost of hiring an additional teacher. As shown in Table 2, salaries represent 63% of the cost of hiring a teacher.

Table 7: CES School Production Function Estimates

Dependent Variable : Log of Teacher to non-Teacher Expenditure Ratio	(1)	(2)	(3)	(4)	(5)
Regression					
Teacher Marginal Factor Cost	0.3230	0.4455	0.5294	0.4127	0.4585
	(0.0529)	(0.1458)	(0.1065)	(0.0857)	(0.1050)
Parameter Estimates					
ϵ_D	-0.7991	-0.7287	-0.6832	-0.7472	-0.7215
Obs	4218	4218	4218	4218	4059
Year FE		✓	✓	✓	✓
Region FE			✓		
District FE				✓	✓

Notes: We regress the log of the teacher marginal factor cost on the log of the ratio of teacher ($mc_T \cdot T$) to non-teacher expenditure (X) for all regular school districts in our sample. Teacher Marginal Factor Cost is determined by taking total spending on teachers and dividing it by the number of teachers a school has hired. We remove non-unionized charter schools in the regression, since they hire teachers on the labor supply rather than on their demand curve. Regression (1) is a simple OLS regression. Regression (2) includes year fixed effects and regression (3) includes year and region fixed effects, where regions are defined as intermediate units. Regression (4) includes year and district fixed effects. Finally, regression (5) only includes schools that are in the labor demand binds case. There were 487 schools in the labor demand binds case (i.e., only 2 regular school districts were not hiring labor on their demand curve). All standard errors are clustered at the school district level.



Notes: The graph shows the labor supply and demand curves for Millcreek Township School District (SD-MIL) in Erie County, Pennsylvania. The demand curve comes directly from Equation 13. The labor supply curve is determined by keeping wages fixed at all other schools, while allowing SD-MIL to adjust its wages. Then we sum up how many teachers are willing to work at SD-MIL at each wage.

Figure 7: Labor Supply and Labor Demand

that district, with the outcome of each negotiation solving a bilateral Nash bargaining problem. The bilateral bargaining outcomes must constitute a Nash equilibrium so that no district or union wants to renegotiate wages given the other bargaining outcomes. Specifically, wages are set to maximize the generalized Nash Product conditional on other negotiated salaries w_{-j}:

$$\begin{aligned}\mathcal{N}(w_j|\boldsymbol{w_{-j}}) &= [W(T_j, B_j - w_j T_j))]^{\alpha_b(j)} \left[T_j(w_j - d_o)\right]^{1-\alpha_b(j)} \\ &= [W(T^*(w_j, \boldsymbol{w_{-j}}), B_j - w_j T^*(w_j, w_{-j}))]^{\alpha_b(j)} \times \\ &\quad [T^*(w_j, \boldsymbol{w_{-j}})(w_j - d_o)]^{1-\alpha_b(j)}\,,\end{aligned} \tag{16}$$

where the parameter $\alpha_b(j)$, $0 \leq \alpha_b(j) \leq 1$, represents the bargaining power of district j. The payoff to the district is given by the school objective function $W(T, X)$. In the event of disagreement, the district hires no teachers and gets a payoff of 0. The payoff to the union is the earnings of teachers in the district $T_j w_j$.[42] In the event of disagreement, the union payoff is $T_j d_0$, where we set $d_0 = 18,500$.[43]

The first-order condition of the logarithm of Equation 16 with respect to w_j is:

$$\frac{\partial \log \mathcal{N}_j}{\partial w_j} = \frac{\alpha_b(j)\frac{\partial W_j(X_j,T_j)}{\partial w_j}}{W_j(X_j, T_j)} + \frac{(1-\alpha_b(j))}{T_j}\frac{\partial T_j}{\partial w_j} + \frac{(1-\alpha_b(j))}{w_j - d_0} = 0. \tag{17}$$

The derivative, $\frac{\partial T_j}{\partial w_j}$, depends on whether labor demand or labor supply binds.[44]

Rearranging this equation allows the bargaining parameter to be expressed as a function of data and our estimated labor demand and labor supply parameters:

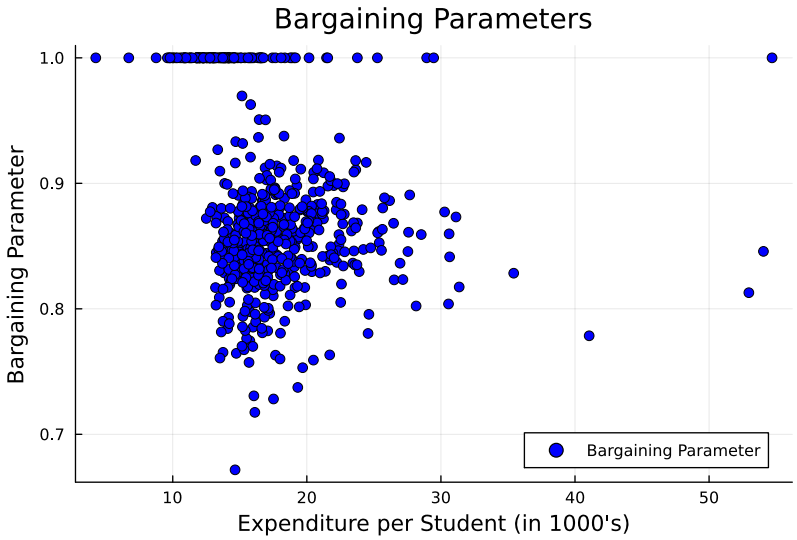
$$\alpha_b(j) = \frac{W_j(X_j, T_j)(T_j + (w_j - d_0)\frac{\partial T_j}{\partial w_j})}{(W_j(X_j, T_j))(T_j + (w_j - d_0)\frac{\partial T_j}{\partial w_j}) - (-T_j\frac{\partial W_j}{\partial X_j} + \frac{\partial T_j}{\partial \omega_j}\left[\frac{\partial W_j}{\partial T_j} - \omega_j\frac{\partial W_j}{\partial X_j}\right])T_j(\omega_j - d_o)}. \tag{18}$$

Thus, we can recover each bargaining power parameter, $\alpha_b(j)$. We assume non-

[42] In each negotiation, the union only puts weight on the teachers hired at the school district with which it is negotiating. Thus the union puts more weight on insiders than teachers that could have taught at the school.

[43] Note that this mechanically prevents the union from accepting a wage below \$18,500 a year.

[44] This model nests the monopsony model for the parameter $\alpha_j = 1$. This bargaining parameter yields the first-order condition given by $\frac{\partial W_j(B_j - w_j T_j, T_j)}{\partial w_j} = 0$, which implies that $\frac{\partial W_j}{\partial T_j} / \frac{\partial W_j}{\partial X_j} = w_j(1 + \frac{1}{\epsilon_w^T})$, the monopsony markdown Lerner formula.



Notes: The figure shows the relationship between expenditure per student and the estimated bargaining parameter for all schools in Pennsylvania. Charter schools have bargaining parameters equal to one.

Figure 8: Estimated Bargaining Power Parameters and School Budget

unionized schools post wages and set their bargaining parameter to 1.[45]

Figure 8 presents our estimates of the bargaining power parameter for each school district on the vertical axis, and school budget per student on the horizontal axis. The unweighted correlation between budget and bargaining parameters is 0.027 and the weighted correlation is 0.56 when weighting by the number of students. The lack of strong correlation indicates that wealthy schools are not significantly stronger bargainers than other schools. Our results are not being driven by budget or size of schools.

5.7 Indirect Inference Estimation

We estimate all but two parameters directly from the data using regressions and first order conditions. The remaining two parameters, σ and r, which appear in the labor supply model, are estimated via indirect inference. The parameter σ governs the random coefficient ν_i that affects the choice between teaching and not teaching. We must estimate it separately from the ϵ terms that determine how teachers choose between schools. The parameter, r, meanwhile, is the value of the outside option

[45] There are a few rare cases where bargaining power parameters are estimated to be greater than 1. In these cases, we censor the bargaining parameters at 1.

Table 8: Indirect Inference Estimates

Parameter	Estimate	Standard Error
σ	9.4	(1.75)
r	14.5	(1.4)

Notes: Standard Errors computed by bootstrapping the estimates of labor supply in Table 5 and labor demand in Table 7 and reestimating the coefficients that minimize the criterion function (equation 19) for 50 bootstrap resamples.

and rationalizes the proportion of teachers that choose the outside option. Since we do not observe the salaries of teachers working outside teaching, we estimate r through indirect inference. In our model, where the majority of teachers are employed, σ primarily plays a role in determining how difficult it becomes to hire teachers as schools go up the supply curve (i.e., the slope of the supply curve). Higher σ's make it more difficult to attract teachers from the outside option. Meanwhile, the reservation value, r which is common across all teachers, affects the level of the supply curve. The higher the r, the more desirable is the outside option.[46]

We use a GMM approach to estimate our parameters. We choose the (σ,r) pair that minimizes the usual GMM criterion function:

$$Q(\sigma, r) = \left(\mathbb{M}(\sigma, r) - \hat{\mathbb{M}}\right)' \mathbb{W} \left(\mathbb{M}(\sigma, r) - \hat{\mathbb{M}}\right), \tag{19}$$

where $\mathbb{W}$ is the identity matrix. Our moments to match are the number of teachers hired and wages, both normalized by their means, predicted by the model for each school district. For computational convenience, we restrict our procedure to Erie and Crawford Counties in Pennsylvania. Figure A4 in the appendix shows the value of the criterion function for various (σ, r) pairs.

Table 8 presents the estimates of σ and r by indirect inference. We compute standard errors coming from sampling error when estimating labor supply parameters, i.e. two-step error, by bootstrap.

Table 9: Unconstrained and Constrained Elasticities: Erie County

	S-CON	S-ERI	S-MIL	S-PEN	C-ROB	C-MON	Out Opt
All Offers							
S-CON	2.804	−0.115	−0.135	−0.261	−0.105	−0.133	−0.033
S-ERI	−0.03	2.411	−0.054	−0.037	−0.058	−0.055	−0.008
S-MIL	−0.232	−0.353	2.537	−0.252	−0.355	−0.35	−0.055
S-PEN	−0.321	−0.17	−0.18	2.592	−0.163	−0.178	−0.041
C-ROB	−0.0	−0.001	−0.001	−0.001	2.28	−0.001	−0.0
C-MON	−0.001	−0.001	−0.001	−0.001	−0.001	1.923	−0.0
Out Opt	−1.093	−1.054	−1.055	−1.083	−1.054	−1.054	2.305
Constrained Offers							
S-CON	0.634	0.0	0.0	−0.011	−0.004	−0.01	−0.019
S-ERI	0.0	0.664	−0.313	−0.005	−0.973	−0.93	−0.087
S-MIL	0.0	−0.167	0.648	0.0	−0.086	−0.098	−0.047
S-PEN	−0.015	−0.001	0.0	0.85	−0.016	−0.018	−0.025
C-ROB	−0.001	−0.032	−0.005	−0.002	1.957	−0.318	−0.004
C-MON	−0.002	−0.045	−0.009	−0.003	−0.464	1.432	−0.006
Out Opt	−1.256	−1.229	−1.257	−1.25	−1.826	−1.846	1.579

Notes: The table shows own- and cross-elasticities of a subset of schools in Erie County. SD refers to a school district, while CHAR refers to a charter school. See the map in Figure 6 for school locations. Each cell shows the percentage increase in teachers at the schools in the columns willing to work after a 1% increase in the wage paid by the schools in the rows. The top table shows the elasticities when schools give out all offers, i.e., they do not constrain their hiring. In the bottom table, school districts can limit hiring by sending out fewer offers.

5.8 Labor Supply Elasticities with Constrained Offers

We recompute the labor supply elasticities while conditioning on the equilibrium set of offers made to teachers. The panel labeled *Constrained Offers* in Table 9 presents labor supply elasticities given equilibrium job offer sets.[47] The top panel presents the same elasticities from Table 6 for comparison. Three observations are worth noting. First, the elasticity of supply to each school is lower when jobs are rationed, as expected because teachers have fewer options to switch to given a salary reduction. On average, the own-salary elasticites are 73 percent of elasticities when schools make offers to all teachers. Relatedly, the elasticity of the outside option with respect to each school's salary increases. Second, a wage change can have no impact on supply to other schools when the teachers impacted by the wage change do not get offers

[46]Figure A3 shows how varying σ and r affect aggregate labor supply. r primarily has a level effect, while σ affects both the level and slope of the supply curve.

[47]Note that a change in wage may affect the offers a school decides to distribute. But, if we allow schools to change their offers, we would be estimating demand elasticities, rather than labor supply elasticities.

from potential substitute schools. Third, non-unionized charter schools become closer substitutes to regular school districts under restricted offers, since they send offers to all teachers. In equilibrium, this allows charter schools to pay lower salaries.

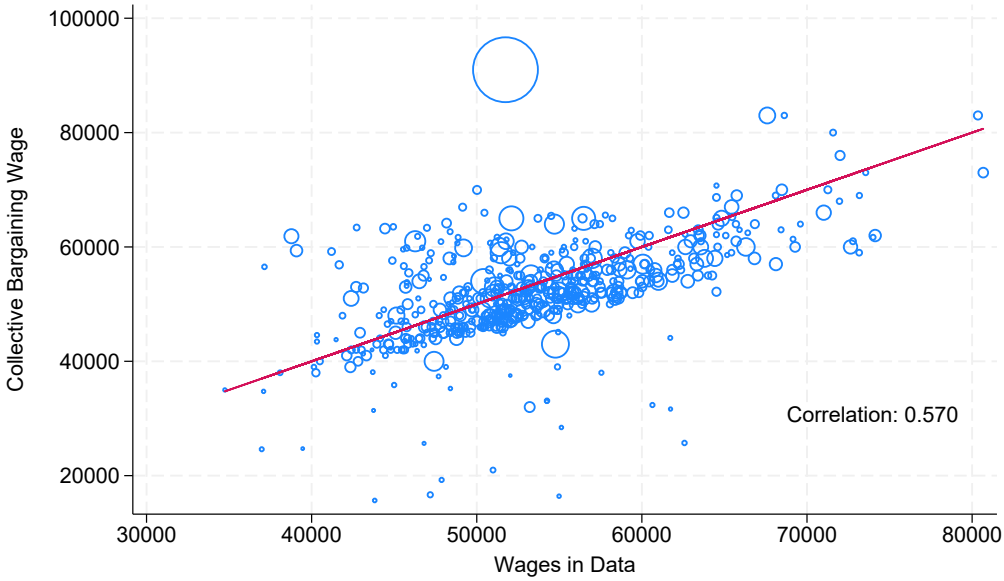
5.9 Model Validation

Before showing counterfactual exercises, we first assess how well our model fits the data. Table 11 presents the average salary and the total number of teachers hired in the data and under the model of collective bargaining. Our model predicts average wages of $ 55,100 in contrast to average wages of $ 54,190 in the data. Likewise, our model predicts that 110,656 teachers will be hired, which is close to the 112,460 that were actually hired in 2017 in Pennsylvania. Moving beyond averages, Figure 9a plots the predicted wages from collective bargaining against wages in the data. The correlation between predicted and real wages is 0.57. Figure 9b shows that fit is substantially better for districts on the labor demand curve, with a correlation coefficient of 0.86, versus schools on the labor supply curve, where this same correlation is 0.23. When labor demand binds, the Nash bargaining weights, α_j. absorb much of the wage variation; we lack this structure for schools where labor supply binds. In addition, the model struggles to explain why districts with large free lunch shares do not pay better.[48]

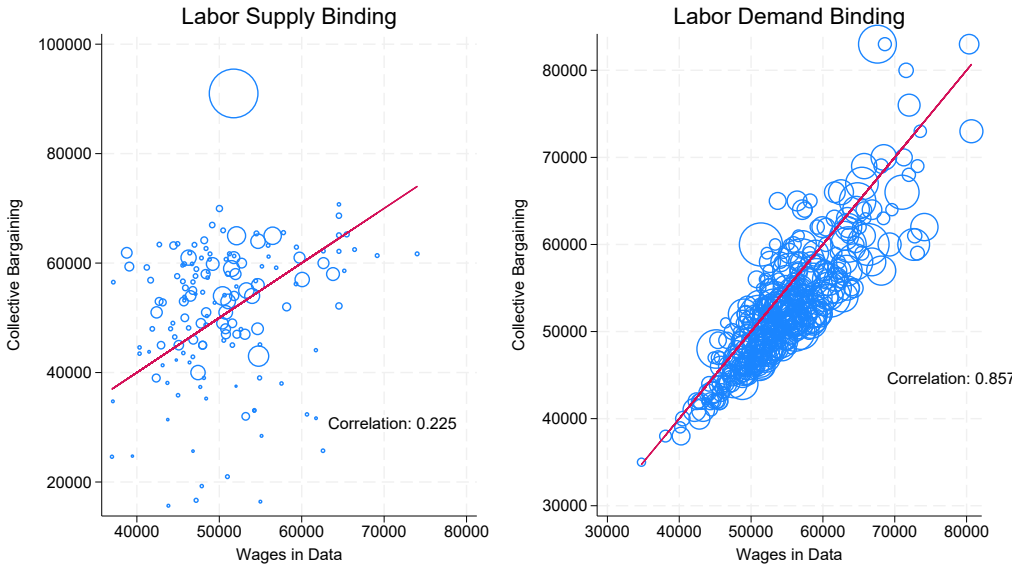
As the number of teachers hired by a district is strongly correlated with budgets, the model does a better job of explaining the number of teachers hired with a correlation that is essentially one. More interesting is that our model predicts an average student to teacher ratio of 15.8 versus 14.9 in the data.

To validate the model's prediction of whether labor supply or labor demand binds, we use data on teacher shortages. Schools on the labor supply curve face shortages; they would like to hire more teachers at the prevailing wage, but cannot do so. We use data on emergency permits for long-term positions for the year 2017, which can be used as a proxy for shortages (see Appendix D.3 for more details). Table A6 shows that schools where labor supply binds hire almost 6 times more teachers using emergency permits than schools where labor demand binds on average (weighted by

[48] An attentive reader would note that the largest district in our data, the Philadelphia school district, over-predicts wages. One of the issues with such a large school district, both in terms of students and geography is that the travel distance from teacher house to centroid of the school district may over-predict true commuting times.



(a) Predicted and Actual Wages



(b) Predicted and Actual Wages

Figure 9: Model Fit

Table 10: Percent Teachers hired with Emergency Permits

	LD Bind	LS Bind
Shortage	0.4%	4.0%
Number	419	106

Notes: The table determines the average percentage of teachers hired at a school using an emergency permit in 2017-18, which is a proxy for a teacher shortage (weighted by teachers). Our model predicted that 419 schools were bound by labor demand, whereas 105 schools had labor supply binding. The table indicates that schools where labor supply binds hired almost 6 times more teachers using emergency permits than those where are model predicted that labor demand binds.

teachers). This fact suggests that our model accurately identifies schools bound by labor supply.

6 Counterfactuals

6.1 Welfare of Collective Bargaining and Oligopsony

Our first counterfactuals address the challenge posed in Section 2: we compute the social planner solution and compare it to the pure oligopsony (posted wages) and collective bargaining settings. Our goal is to understand how oligopsony and collective bargaining individually distort the market. We will often refer to pure oligopsony as a "no-union" counterfactual. Such policies have been implemented recently in the market for school teachers in the United States; e.g., Wisconsin passed Act 10 in 2011 which effectively restricted collective bargaining for Wisconsin teachers.

The social planner and oligopsony equilibria are computed similarly. For the social planner, we find the equilibrium that sets labor supply equal to labor demand for every school district. For oligopsony, we set the marginal rate of technical substitution equal to each school's marginal factor cost. Importantanly, in both models, all teachers receive offers from every school; otherwise, schools and the social planner leave welfare on the table.[49],[50] Note that the oligopsony solution is nested by Nash-

[49]Consider the oligopsony scenario. Suppose school j does not give an offer to teacher $\tilde{i}$. If this teacher would receive an offer and accept with probability $s_{\tilde{i}j} > \epsilon$, the school district could still hire T_j teachers by making the additional offer and lowering wages to $\tilde{w}_j = w_j - \delta$ for some small δ. Thus, not making an offer to $\tilde{i}$ cannot maximize school welfare W_j, a contradiction. The proof for the planner problem proceeds in a similar way with the observation that there are values of the ϵ shock such that teacher $\tilde{i}$ has the highest value for a job at school j, and thus, not offering them a job would not maximize total surplus. This result arises because schools do not care about which teacher accepts a job, so higher utility of accepting a job is enough.

[50]We do not attempt to quantify welfare in the three settings, because workers mechanically

in-Nash, specifically for $\alpha_j^b = 1$ for all j. See Sections C.5 and C.4 in the appendix for more details on computation.

Model	Average Wage	Teachers Hired
Data	54,190	104,472
Collective Bargaining	55,100	110,656
Planner	50,459	121,875
Posted	42,526	118,590

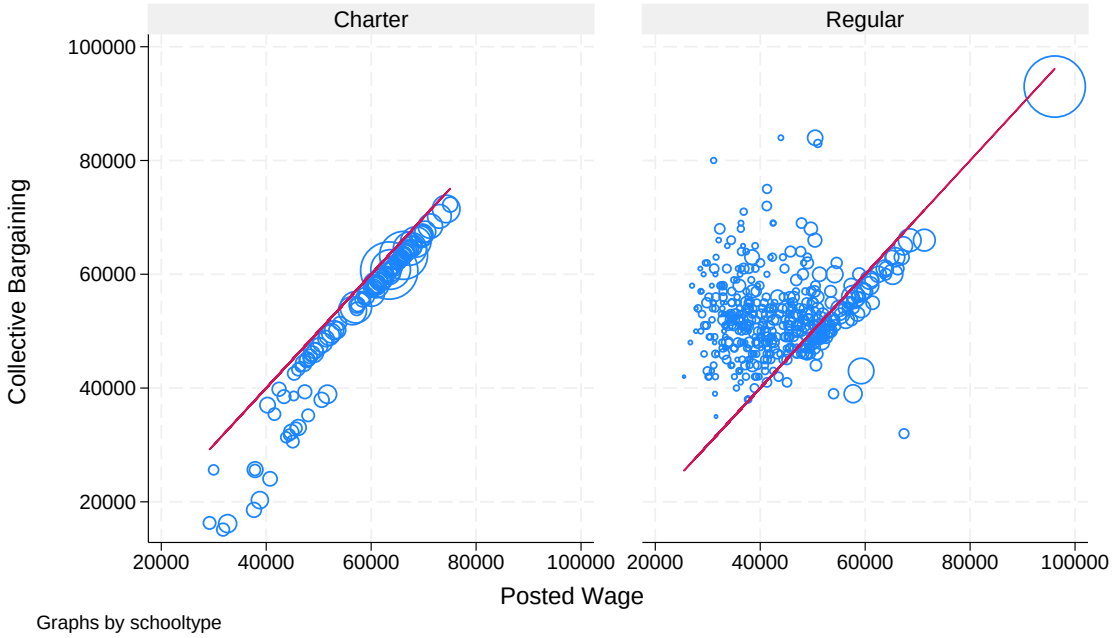
Notes: Average Wage over all teachers shown. Posted corresponds to the situation where employers set wages with no unions, planner to the social planner solution, where labor demand meets labor supply, and Collective Bargaining to the situation where school districts and teacher unions bargain collectively over wages.

Table 11: Welfare, Collective Bargaining and Oligopsony Distortion

Table 11 presents the equilibrium predictions of the model both in terms of average wages and the total number of employed teachers, for collective bargaining, the social planner, and oligopsony with posted wages. We find that the planner would set average wages of $ 50,459, which is above the oligopsony model prediction of wages of $ 42,536, or a 16% decrease in wages due to oligopsony power. Associated with this decrease in wages is a decrease in the number of teachers from 121,875 to 118,590, or a two percent fall. As teacher supply is fairly inelastic in terms of substitution between teaching and the outside job rather than between schools, a 16% drop in wages only results in a 2% change in employment. Thus, the oligopsony distortion mostly leads to wage transfers from teachers to schools, rather than large quantity distortions and accompanying deadweight loss. Second, the average elasticity of teacher supply is $\epsilon_S = 2.8$, which would imply a markdown in the monopsonist model in section 2 of 27 percent. However, wages are not depressed by nearly this amount because labor demand is also fairly inelastic, with a labor demand elasticity of $\epsilon_D = -0.72$. Thus, to understand the effect of monopsony power on wages, at least relative to the efficient level, it is important to understand wage sensitivity on both the supply and demand side. In addition, the wage distortion due to oligopsony power is fairly uniform across school districts in the data.

The collective bargaining model predicts wages of $ 55,100 which is above both the planner and the posted solutions. Collective bargaining is strong enough to eliminate the oligopsony distortion on wages, but it introduces a new distortion because wages

benefit from having large choice sets making welfare implausibly large under the social planner and pure oligopsony settings where offers are sent to all teachers. This is a long-standing issue in the literature, and reflects the product variety effect.



Notes: The figure shows the model predicted wages from collective bargaining on the y-axis and the predicted wages from the posted wage scenario on the x-axis. Circle size is proportional to the number of teachers at the school. The 45 degree line is plotted in red.

Figure 10: Collective Bargaining and Posted Wage Predictions

are too high. Given a 23% increase in wage, the collective bargaining model leads to 7% fewer teachers being employed, reducing teachers from 121,000 in the planner solution to 110,656 hired teachers. This is because under collective bargaining rationing of job offers occurs in 390 out of 525 districts.

Considering the entire distribution of salaries in the two scenarios reveals a more complete picture. Figure 10 presents the collective bargaining wages on the vertical axis, and the posted — or no-union — wages on the horizontal axis for all the 525 school districts in our data. The 45 degree line shows the case where these two wages would be the same. A striking fact is that collective bargaining leads to increased wages at 333 schools, but *decreased* wages at 194 schools. The direct effect of collective bargaining increasing wages is simple: increased bargaining power for teacher unions increases wages. Eliminating unions should thus decrease wages. But, eliminating unions also has an indirect effect: previously unionized school districts lower wages and expand hiring making the labor market for teachers more competitive, which

pushes wages up at rival school districts.[51] For example, Figure 10 shows that banning unions *raises* the wages for all previously non-unionized charter schools.

6.2 Minimum Wage

In this counterfactual, we consider how a minimum wage for teachers across the state of Pennsylvania would affect the equilibrium allocation of teachers. In particular, we ban unions but allow the state to set a minimum wage for all schools, which is the policy in 9 states.[52] Of course, the social planner solution could be implemented if the state could set a unique minimum wage *for each school district*. The state would set the school level minimum wage $\underline{w}_j = w_j^*$ to the social efficient wage. The complication with setting minimum wages has to do with having a single wage floor for the entire state. We find that minimum wages can raise wages to the social planner level; however, they cannot achieve social planner hiring due to inefficiencies at the minimum wage constraint.

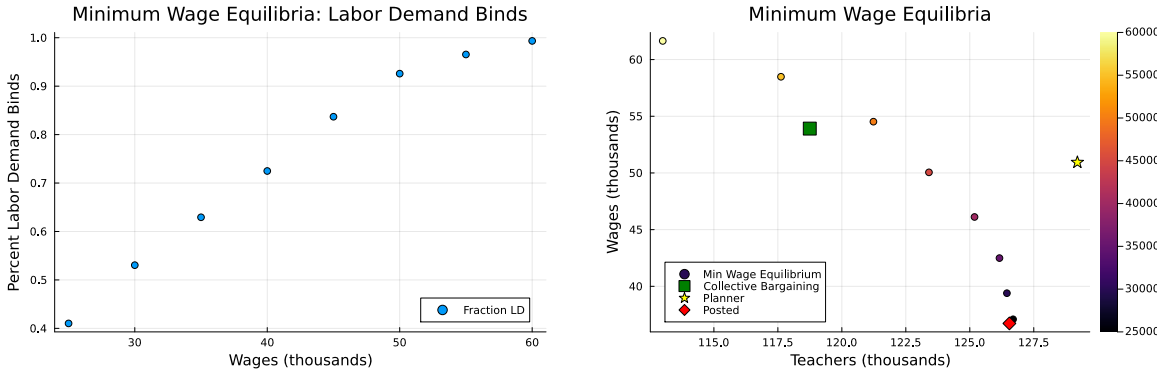
Consider a school that finds the minimum wage binding. Too many teachers want to work at the school, i.e., labor demand is below labor supply. Thus, the school will implement the rationing rule and offer fewer jobs than it would in the wage posting setting. Thus, wages increase, but rationing leads to lower equilibrium values of teachers hired.[53]

Figure 11 shows the equilibrium points of various minimum wage policies that range from \$25,000 to \$60,000 in \$5,000 increments. Panel 11a shows how changing the minimum wage on the horizontal axis increases the share of schools where labor demand binds. For instance, at a wage of \$40,000, 65 percent of schools have binding labor demand. Panel 11b presents the average wage and number of teachers hired for the state as the minimum wage increases. As minimum wages increase, so do average wages; for example, at a minimum wage of \$40,000, the average wage would

[51] In oligopsony with logit teacher preferences, wages are strategic complements, that is, raising wage at one school district raises wages at rival school districts (see Caplin and Nalebuff (1991)). Thus, it is likely (but not strictly a theorem in a BLP context), that oligopsony with a mixed logit also yields strategic complementarity. In contrast, if all schools were on the labor demand curve, then wages are likely (but again this is not a theorem given the random coefficients on distance) to be strategic substitutes. If a school raises wages, then that school will hire fewer teachers. This expands labor supply at rival schools leading to lower wages.

[52] Source.

[53] In our counterfactual, we allow schools where minimum wage binds to play the same offer game that they do with collective bargaining.



(a) Minimum Wages and Binding Labor Demand

(b) Minimum Wages Equilibria: Employment and Wages

Notes: The figure maps minimum wage policies to an equilibrium outcome. In particular, we show the equilibria for minimum wages from \$30,000 - \$60,000 in \$5,000 increments. The red diamond shows the posted equilibrium, the green square the collective bargaining equilibrium and the yellow star the planner outcome.

Figure 11: Minimum Wage Counterfactual

be \$47,000. Equilibrium employment, meanwhile, first increases and then decreases in minimum wage as schools shift from hiring on the labor supply curve to the labor demand curve.

Importantly, minimum wages can increase wages to the social planner level, but employment is always substantially below the social planner outcome. This occurs because attractive schools, that could hire enough teachers at lower wages, respond to higher wages by rationing jobs.

6.3 State Level Wage Setting

Many states determine teacher compensation by setting a flat wage schedule for the entire state, rather than having local school districts decide on teacher compensation. For example, North Carolina paid a base salary of \$41,870 in 2023-2024 for all teachers in the public school system with a bachelor's degree and three years of experience, including charter schools, along with some adjustments for the local cost of living.[54] Statewide wage posting clearly enhances the monopsony power of schools because it internalizes spillovers between employers. We find, however, that state-level wage schedules generate inequality between schools, since many schools struggle to hire enough teachers when they cannot differentiate themselves with higher wages.

[54] https://www.dpi.nc.gov/documents/fbs/resources/salaryschedule2324hb259pdf/download?attachment Accessed October 1, 2025, Table 1.

To assess the effect of statewide wage posting, we endow the state agency with an objective function that aggregates the welfare of individual school districts.[55] We consider a state regulator that sets a state-level wage $\bar{w}$, and allocates budgets B_j to each school district to minimize total expenditures ($\sum_j B_j$), subject to the constraint that school districts are no worse off than under an initial wage w_j^0 and budget B_j^0. That is, budgets are adjusted to keep each district's welfare constant at its initial level, consistent with the concept of compensating variation. This can be thought of as a Pareto constraint at the school district level when setting state wages $\bar{w}$. Section E in the appendix provides further details on our method.

We initially examine a simplified market that clearly shows the mechanisms at play when setting a state-level wage. We then show state-level wage results for the State of Pennsylvania. In particular, a state wage: a) allows the state to internalize wage externalities between school districts and thereby depresses wages, but b) when significant school heterogeneity exists, an equalized state wage cannot provide compensating wage variation, which leads to inefficiency. We begin by analyzing a market with identical schools.

6.3.1 Identical School Districts

We consider a simplified market with 10 identical schools and 500 identical teachers and present the entire calibration in Table A7 in Appendix E. Both schools and teachers are in the same location. All schools have a budget of \$750,000 and the same school-level characteristics, so that each school yields identical mean utility to teachers δ_{ij}. Table 12a shows that the posted wage would be \$ 53,247 for each school. The state-level wage that minimizes the total spending of all school districts is substantially lower at \$ 22,000. This would reduce the spending on schools by 7.2 percent, which is substantial given that Table 2 showed that schools spend a third of their budget on teachers. The state has a lot of monopsony power because most of the diversion of teachers when wages fall is to other school districts, rather than to the outside option, as was shown in the elasticity matrix in Table 6. Indeed, the number of teachers only falls by 1.7% between the posted and state wage setting outcomes even though wages have dropped by almost 50%.

[55] In the case of a private firm, it would be straightforward to simply consider joint profit maximization. Since school districts do not maximize profit, we must take a different approach.

6.3.2 Nine Identical School Districts, One Charter School

Additional complications arise when we move to a setting with differentiated teachers and schools so that teachers have varying willingness to work at different schools. When a school is particularly unattractive, perhaps due to its remote location, charter status, or high free lunch share, a flat state wage precludes the school from offering higher wages to compensate for the lack of amenities leading to a precipitous drop in teachers hired. Returning to our simple example, we assign charter status to the last school, while the remaining nine schools are unchanged. The average posted wages would then be $ 53,549 (versus $53,247 before), but the wages of the charter school would be $ 68,320. If the state imposed a flat wage across all school districts, the wage that would minimize expenditures, again subject to the constraint that no school district is worse off, would be a state wage of $ 70,000 which is far above the level of posted wages.[56] In addition, the state would need to spend 6.3 percent *more* to implement such a policy relative to posted wages. Heterogeneity in school amenity values makes uniform wage policies difficult to implement, unless the state is willing to accept welfare losses at some schools to lower wages.

6.3.3 Pennsylvania School Districts

Schools in Pennsylvania have significant heterogeneity in amenity values, making it difficult for the state to exploit its monopsony power under the constraint of keeping all schools just as well off as before. We focus our analysis on the greater State College Region in Pennsylvania and present results in Table 12b, but heterogeneity in non-wage amenities is a greater issue for a statewide wage. We choose as a starting point the wages and hiring outcomes from the posted wages.The constraint on school welfare is somewhat arbitrary. The optimal state wage would be set to $ 57,000 and would increase spending by 14.8%. Thus, while state wage setting could lower wages due to the ability to coordinate wages between districts, it is not effective in areas with significant heterogeneity across school districts. As such, there is little benefit, at least with the type of Pareto constraint we impose, in having uniform wage setting at the state level. Indeed, states that impose uniform wage schedules tend to allow schools with hiring difficulties to set wages higher, which could be one way to balance

[56] A high wage makes non-charter schools hire fewer teachers, increasing the residual labor supply from which the charter school can hire.

(a) State Wage Setting for Case Study Market

	Symmetric Schools	One Charter
Average Posted Wage	53,247	53,549
Charter Posted Wage	.	68,320
State Wage	22,000	70,000
Change in State Budget	-7.2%	6.3 %
Change in Number of Teachers	-2%	-17%

Notes: The table shows a simulated region with 10 school districts and 1000 teachers, all of which are in the same location. In column 1 all schools are regular school districts, while in column 2, one school is a charter school. State Wages denote the budget minimizing uniform state wage where the starting point is the posted wage equilibrium.

(b) State Wage Setting for State College

Average Posted Wage	34,441
State Wage	57,000
Change in State Budget	14.8%

Notes: The table shows the state college region. State Wages denote the budget minimizing uniform state wage where the starting point is the posted wage equilibrium.

these incentives.[57]

7 Conclusion

Many labor markets are characterized by a few dominant firms, that can be represented with a model of oligopsony power. The public sector labor market, where the government is the sole employer is particularly susceptible to labor market power and the distortions it imposes. In response, the public sector has the highest percentage of workers represented by unions in the American economy.

We find the potential for substantial monopsony power in the market for public school teachers in Pennsylvania, with wages under a pure wage posting arrangement 16 percent below the social planner solution on average. However, collective bargaining allows public teachers to have countervailing power that pushes their wages 23 percent higher on average than in the oligopsony scenario, which is even higher than wages under the social planner. Thus, we find that collective bargaining not only corrects the distortions from oligopsony, but introduces an upward distortion in wages. The effect of collective bargaining, however, depends crucially on bargaining power.

[57] North Carolina, for example, allows schools to provide a hiring bonus if they are having trouble hiring.

Where bargaining power is too strong, collective bargaining adds an additional distortion that pushes wages too high. In particular, there is a vector of bargaining parameters between 0 and 1 that would lead to the socially optimal equilibrium. In the schoolteacher environment, the union's bargaining power seems to be too strong given that collective bargaining wages are somewhat above the wages the planner would set.

Much of the wage differences between the collective bargaining and oligopsony settings are not about greater or lower monopsony power, but instead about the ability to negotiate higher or lower wages. Moreover, raising wages, say due to a higher bargaining power parameter, has ambiguous effects on workers, since either labor supply or labor demand might be binding. When labor demand binds, newly unemployed workers will push supply curves at rival schools out, allowing them to reduce wages. More vexing is the implication that extrapolation from existing empirical studies on the impact of unions on wages might be quite difficult since it would need to be conditioned on which side of the market is binding.

In addition, we use our structural model to evaluate two counterfactual policies: a minimum wage and a statewide uniform wage schedule. We find that heterogeneity in amenity values across schools mitigates the effectiveness of minimum wages in reducing monopsony power and also mitigates the ability of the state to exercise increased monopsony power when setting uniform wages. Overall, this paper highlights the challenges of predicting the effect of policies in the labor market when oligopsony power and unions are present, especially when complex externalities between employers are taken into account.

References

Azkarate-Askasua, Miren, and Miguel Zerecero. 2024. "Union and firm labor market power."Technical report, Mannheim.

Baker, Michael, Yosh Halberstam, Kory Kroft, Alexandre Mas, and Derek Messacar. 2024. "The Impact of Unions on Wages in the Public Sector: Evidence from Higher Education." Working Paper 32277, National Bureau of Economic Research.

Bates, Michael, Michael Dinerstein, Andrew C Johnston, and Isaac Sorkin.

2025. "Teacher labor market policy and the theory of the second best." *Quarterly Journal of Economics*.

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2022. "Labor market power." *American Economic Review* 112 (4): 1147–1193.

Biasi, Barbara, Chao Fu, and John Stromme. 2021. "Equilibrium in the market for public school teachers: District wage strategies and teacher comparative advantage."Technical report, National Bureau of Economic Research.

Blanchflower, David G., and Alex Bryson. 2010. "The Wage Impact of Trade Unions in the UK Public and Private Sectors." *Economica* 77 (305): 92–109.

Blanchflower, David G, and Alex Bryson. 2024. "Unions, wages and hours." Working Paper 32471, National Bureau of Economic Research.

Buchholz, Nicholas, Laura Doval, Jakub Kastl, Filip Matejka, and Tobias Salz. 2025. "Personalized Pricing and the Value of Time: Evidence From Auctioned Cab Rides." *Econometrica* 93 (3): 929–958.

Caplin, Andrew, and Barry Nalebuff. 1991. "Aggregation and imperfect competition: On the existence of equilibrium." *Econometrica* 1 25–59.

Card, David. 1996. "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis." *Econometrica* 64 (4): 957–979.

Card, David, Thomas Lemieux, and W. Craig Riddell. 2020. "Unions and wage inequality: The roles of gender, skill and public sector employment." *Canadian Journal of Economics/Revue canadienne d'économique* 53 (1): 140–173.

Collard-Wexler, Allan, Gautam Gowrisankaran, and Robin S Lee. 2019. ""Nash-in-Nash" bargaining: a microfoundation for applied work." *Journal of Political Economy* 127 (1): 163–195.

Crawford, Gregory S., Rachel Griffith, and Alessandro Iaria. 2021. "A survey of preference estimation with unobserved choice set heterogeneity." *Journal of Econometrics* 222 (1): 4–43.

Crawford, Gregory S, and Ali Yurukoglu. 2012. "The welfare effects of bundling in multichannel television markets." *American Economic Review* 102 (2): 643–685.

Delabastita, Vincent, and Michael Rubens. 2025. "Colluding against workers." *Journal of Political Economy*.

Demsetz, Harold. 1973. "Industry structure, market rivalry, and public policy." *The Journal of Law and Economics* 16 (1): 1–9.

DiNardo, John, Nicole Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica* 64 (5): 1001–44.

DiNardo, John, and David S. Lee. 2004. "Economic Impacts of New Unionization on Private Sector Employers: 1984–2001." *The Quarterly Journal of Economics* 119 (4): 1383–1441.

Dodini, Samuel, Kjell Salvanes, and Alexander Willen. 2021. "The Dynamics of Power in Labor Markets: Monopolistic Unions versus Monopsonistic Employers." Working Paper 9495, CESifo.

Dunlop, John T. 1944. *Wage Determination Under Trade Unions.*. New York, NY: Macmillan.

Farber, Henry S. 1978. "Individual preferences and union wage determination: the case of the united mine workers." *Journal of Political Economy* 86 (5): 923–942.

Farber, Henry S. 1986. "The Analysis of Union Behavior." In *Handbook of Labor Economics*, edited by Ashenfelter, Orley, and Layard Volume 2. Chap. 18 1039–1089.

Farber, Henry S, Daniel Herbst, Ilyana Kuziemko, and Suresh Naidu. 2021. "Unions and Inequality over the Twentieth Century: New Evidence from Survey Data*." *The Quarterly Journal of Economics* 136 (3): 1325–1385.

Freeman, Richard B., and James L. Medoff. 1984. *What Do Unions Do*. New York, NY: Basic Books.

Galbraith, John Kenneth. 1954. "Countervailing Power." *The American Economic Review* 44 (2): 1–6.

Gottfries, Axel, and Gregor Jarosch. 2023. "Dynamic Monopsony with Large Firms and Noncompetes."Technical report, National Bureau of Economic Research.

Gowrisankaran, Gautam, Aviv Nevo, and Robert Town. 2015. "Mergers when prices are negotiated: Evidence from the hospital industry." *American Economic Review* 105 (1): 172–203.

Green, David A, Benjamin Sand, and Iain G Snoddy. 2022. "The impact of unions on nonunion wage setting: Threats and bargaining."Technical report, Working Paper UBC.

Grennan, Matthew. 2013. "Price discrimination and bargaining: Empirical evidence from medical devices." *American Economic Review* 103 (1): 145–177.

Ho, Kate, and Robin S Lee. 2017. "Insurer competition in health care markets." *Econometrica* 85 (2): 379–417.

Horn, Henrick, and Asher Wolinsky. 1988a. "Bilateral monopolies and incentives for merger." *The RAND Journal of Economics* 408–419.

Horn, Henrik, and Asher Wolinsky. 1988b. "Worker substitutability and patterns of unionisation." *The Economic Journal* 98 (391): 484–497.

Hoxby, Caroline. 1996. "How Teachers' Unions Affect Education Production." *The Quarterly Journal of Economics* 111 (3): 671–718.

Kline, Patrick M. 2025. "Labor Market Monopsony: Fundamentals and Frontiers."Technical Report 33467, National Bureau of Economic Research, https://www.nber.org/papers/w33467.

Kroft, Kory, Yao Luo, Magne Mogstad, and Bradley Setzler. 2025. "Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry." *American Economic Review* 115 (9): 2926–2969.

Lee, Robin S, and Kyna Fong. 2013. "Markov perfect network formation: An applied framework for bilateral oligopoly and bargaining in buyer-seller networks."

Lemieux, Thomas. 1998. "Estimating the Effects of Unions on Wage Inequality in a Panel Data Model with Comparative Advantage and Nonrandom Selection." *Journal of Labor Economics* 16 (2): 261–91.

Lewis, H. Gregg. 1963. *Unionism and Relative Wages in the United States.*. Chicago, IL: University of Chicago Press.

Lovenheim, Michael F. 2009. "The Effect of Teachers' Unions on Education Production: Evidence from Union Election Certifications in Three Midwestern States." *Journal of Labor Economics* 27 (4): 525–587.

McFadden, D. 1978. "Modeling the choice of residential location." *Transportation Research Record* 672 72–77.

Miller, Nathan et al. 2022. "On the misuse of regressions of price on the HHI in merger review." *Journal of Antitrust Enforcement* 10 (2): 248–259.

Ransom, Michael, and David P. Sims. 2010. "Estimating the Firm's Labor Supply Curve in a "New Monopsony" Framework: Schoolteachers in Missouri." *Journal of Labor Economics* 28 (2): 331–355.

Robinson, Joan. 1933. *The economics of imperfect competition*. London: Macmillan and Co., Ltd..

Rubens, Michael. 2023. "Management, productivity, and technology choices: evidence from US mining schools." *The RAND Journal of Economics* 54 (1): 165–186.

Setzler, Bradley. 2025. "Labor and Product Market Power, Endogenous Quality, and the Consolidation of the US Hospital Industry." Working Paper 34180, National Bureau of Economic Research.

Sokolova, Anna, and Todd Sorensen. 2021. "Monopsony in Labor Markets: A Meta-Analysis." *ILR Review* 74 (1): 27–55.

U.S. Department of Transportation. 2016. "The Value of Travel Time Savings: Departmental Guidance for Conducting Economic Evaluations." Departmental Guidance Revision 2, Office of the Secretary of Transportation, 2016 Update.

A Data Construction

A.1 Potential Teachers

Unfortunately, we do not have data on the number of qualified teachers who are not teaching, which we refer to as potential teachers. Instead, we approximate the size of this group from data on certifications and exits. We do our analysis by region since the exit rate varies across regions. We find that the potential teachers make up about 28.5% of total teachers. After calculating the number of total teachers in each region, we assign addresses to these teachers by sampling with replacement from our address dataset.

A.1.1 Calculating the Number of Potential Teachers

There are two types of non-teachers: teachers who get certified and never teach and teachers who exit teaching prematurely.

Certifications/New Teachers: The state of Pennsylvania publishes data on the number of people who have received teacher certificates from 2010 to the present.[58]

Teachers need an Instructional I certificate to practice as teachers.[59] Once received, it lasts for 6 years, but is usually replaced with an Instructional II certificate before its expiration. The number of Instructional I certificates issued is a useful proxy for how many new teachers are qualified to work in Pennsylvania and have put in effort, such as taking exams, to pursue a teaching career. Even though certificates are valid for 6 years, we assume that after 3 years a teacher who does not teach will never teach and has exited the potential pool.

Next, we find the number of teachers who entered the teacher dataset for the first time and have less than two years of experience. By subtracting the number of starting teachers from the number of newly licensed teachers, we can get a value of the number of teachers in each year who are licensed, but not teaching. In 2017, this adds 8,494 potential teachers. We allocate these teachers proportionally across regions since our certification data is not region specific.

Exiting Teachers: Next, we consider teachers who have quit teaching or taken a gap year. To avoid counting retirees as part of the pool, we restrict our sample of exiting teachers to those with less than 15 years of experience in education. We assume that after 3 years out of the labor market, a teacher who permanently exits will not return to teaching.

For each year beginning in 2014, we calculate the number of teachers who permanently exit and have less than 15 years of experience. We also add teachers who have

[58] Data were downloaded from https://www.pa.gov/agencies/education/data-and-reporting/school-staff/educator-preparation-and-certification/act-82-of-2018-report-on-educator-preparation-and-certification.

[59] Some teachers can get emergency permits to teach due to shortages. We ignore these permits here.

Table A1: Potential Teachers by Region

Region	Actual Teachers	Certificate Teachers	Exiters	Total
Philadelphia	34,408	2,599	11,037	48,044
Pittsburgh	18,710	1,413	2,858	22,981
Harrisburg	17,348	1,310	3,422	22,080
NW	10,525	795	1,476	12,796
NE	10,438	788	1,231	12,457
State College	6,869	519	1,085	8,473
Allentown	14,162	1,070	2,453	17,685
Total	112,460	8,494	23,562	144,516

Notes: The table shows the number of teachers and the potential teachers by region. Actual represents the number of teachers in each region in our teacher dataset. Certificates represents the number of teachers who have been certified but are not teaching. Exiters represents the number of teachers who have exited and have less than 15 years of experience in education. Total sums actual certificate and exiters to get the total number of qualified teachers in each region.

taken a gap year. Table A1 shows the number of teachers in our teacher dataset as well as the number of potential teachers that are added.

We believe our potential teacher numbers match well to other statistics. The number of Pennsylvania private school teachers in 2017 was 23,630, which is about two-thirds of our non-teachers.[60]. In addition, the Bureau of Labor Statistics says that the number of teachers in Pennsylvania is 152,980, which also includes teachers at technical career and penitentiary schools (about 6,000 additional teachers that were not included in actual teachers).[61]

A.1.2 Potential Teachers in the Model

For the structural model, we need to add potential teachers to our teacher data. In the model, teachers enter as addresses. We have 150,189 unique teacher addresses that were collected from InfoUSA. These addresses include any unique address where a teacher has lived between 2010-2022.

We sample with replacement from our teacher address database by region until we reach the total number of teachers desired, i.e., column "Total" in A1.

[60] https://nces.ed.gov/surveys/pss/tables1314.asp

[61] https://web.archive.org/web/20250129185820/https://www.bls.gov/oes/tables.htm. This number includes elementary, middle, and secondary school teachers, along with special education teachers and career technical education teachers. We do not include preschool teachers.

Table A2: 2017 Statistics Of Teachers Matched to InfoUSA

	Unique	None	Multiple
Percent Female	0.79	0.79	0.65
	(0.41)	(0.41)	(0.48)
Percent White	0.95	0.93	0.95
	(0.22)	(0.26)	(0.22)
Percent w/ Masters	0.62	0.56	0.61
	(0.48)	(0.50)	(0.49)
Avg Years in Education	14.35	12.13	14.56
	(8.28)	(8.50)	(8.43)
Charter Schools	0.06	0.10	0.07
	(0.24)	(0.30)	(0.25)
Average Wage	70,011.82	65,436.94	70,159.40
	(17,531.14)	(17,548.66)	(17,350.65)
Obs	41,359	26,371	44,730

Notes: The table shows summary statistics for teachers who were matched once, not at all, or more than once to the InfoUSA data in 2017.

A.2 InfoUSA Teacher Matching

We use InfoUSA data from 2010-2022 that includes names and addresses of people living in Pennsylvania, which were compiled from public sources, voter registration data, utility company data, and real estate data. We match our public school data to the InfoUSA data using exact name and year. About 36.4% of the sample had a unique match, 24.4% of the sample had no match, and 39.2% had multiple matches. We only include unique matches and make the assumption that teachers with unique names do not have residential patterns that are significantly different from their peers.

Table A2 shows summary statistics across the three match possibilities. Teachers that were matched once or multiple times have very similar statistics. Teachers who were not matched tend to have less experience, lower pay, and are more likely to work for charter schools.

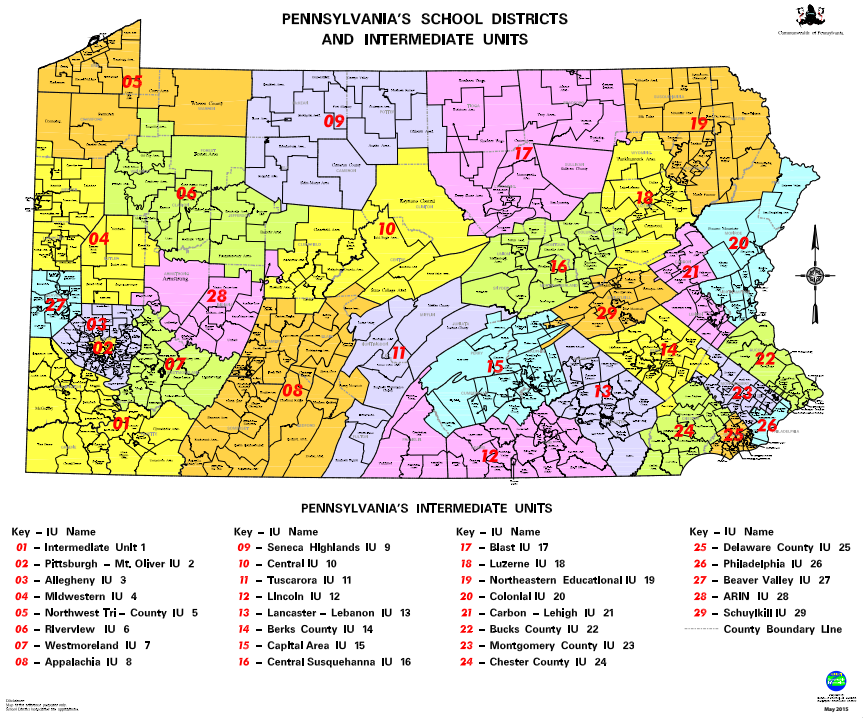


Figure A1: Map of School Districts in Pennsylvania

A.3 Additional Figures

B Proofs and Derivations

B.1 Nash Bargaining

Let $\mathbf{w} = \{w_j\}_{j=1}^{J}$, i.e., the vector of wages at all schools $j \in \mathcal{J}$. The problem is:

$$\max_{w_j | w_{-j}} \quad \Big(W_j(X_j, T_j(\mathbf{w}))\Big)^{\alpha_j} \Big(T_j(\mathbf{w})(w_j - d_o)\Big)^{1-\alpha_j}$$
$$s.t. \;\; B_j = w_j T_j(\mathbf{w}) + X_j. \tag{20}$$

Let's starting by transforming the problem into logs:

$$\alpha_j log\Big(W_j(X_j, T_j(\mathbf{w}))\Big) + (1-\alpha_j) log\Big(T_j(\mathbf{w})\Big) + (1-\alpha_j) log\Big(w_j - d_o\Big)$$

FOC:

$$\frac{\alpha_j \frac{\partial W_j(X_j, T_j(\mathbf{w}))}{\partial w_j}}{W_j(X_j, T_j(\mathbf{w}))} + \frac{(1-\alpha_j)}{T_j(\mathbf{w})} \frac{\partial T_j(\mathbf{w})}{\partial w_j} + \frac{(1-\alpha_j)}{w_j - d_o} = 0$$

For simplicity, we define:

$$m = \frac{\partial W_j(X_j, T_j(\mathbf{w}))}{\partial w_j} = (-T_j(\mathbf{w})\frac{\partial W_j}{\partial X_j} + \frac{\partial T_j(\mathbf{w})}{\partial \omega_j}\Big[\frac{\partial W_j}{\partial T_j(\mathbf{w})} - w_j\frac{\partial W_j}{\partial X_j}\Big]).$$

Then we have:

$$\frac{\alpha_j m}{W_j(X_j, T_j(\mathbf{w}))} + (1-\alpha_j)[\frac{1}{w_j - d_o} + \frac{1}{T_j}\frac{\partial T_j(\mathbf{w})}{\partial w_j}] = 0$$

Solving step by step to get α's:

$$\begin{aligned}
\frac{\alpha_j m}{W_j(X_j, T_j(\mathbf{w}))} &= -(1-\alpha_j)[\frac{1}{w_j - d_o} + \frac{1}{T_j(\mathbf{w})}\frac{\partial T_j(\mathbf{w})}{\partial w_j}], \\
\frac{\alpha_j m}{W_j(X_j, T_j(\mathbf{w}))} &= -(1-\alpha_j)\frac{T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j}}{T_j(\mathbf{w})(w_j - d_o)}, \\
\alpha_j m T_j(\mathbf{w})(w_j - d_o) &= -(1-\alpha_j)W_j(X_j, T_j(\mathbf{w}))(T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j}), \\
\alpha_j m T_j(\mathbf{w})(w_j - d_o) - \alpha_j W_j(X_j, T_j(\mathbf{w}))(T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j}) & \\
= -W_j(X_j, T_j(\mathbf{w}))(T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j}), & \\
\alpha_j(m T_j(\mathbf{w})(w_j - d_o) - W_j(X_j, T_j(\mathbf{w}))(T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j})) & \\
= -W_j(X_j, T_j(\mathbf{w}))(T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j}), &
\end{aligned}$$

$$\begin{aligned}
\alpha_j &= \frac{-W_j(X_j, T_j(\mathbf{w}))(T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j})}{m T_j(\mathbf{w})(w_j - d_o) - W_j(X_j, T_j(\mathbf{w}))(T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j})}, \\
\alpha_j &= \frac{W_j(X_j, T_j(\mathbf{w}))(T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j})}{(W_j(X_j, T_j(\mathbf{w})))(T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j}) - m T_j(\mathbf{w})(w_j - d_o)},
\end{aligned}$$

$$\alpha_j = \frac{W_j(X_j, T_j(\mathbf{w}))(T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j})}{(W_j(X_j, T_j(\mathbf{w})))(T_j(\mathbf{w}) + (w_j - d_o)\frac{\partial T_j(\mathbf{w})}{\partial w_j}) - (-T_j(\mathbf{w})\frac{\partial W_j}{\partial X_j} + \frac{\partial T_j(\mathbf{w})}{\partial w_j}\Big[\frac{\partial W_j}{\partial T_j(\mathbf{w})} - w_j\frac{\partial W_j}{\partial X_j}\Big])T_j(\mathbf{w})(w_j - d_o)}. \tag{21}$$

C Code Appendix

C.1 Labor Supply

Let $\mathbf{w} = \{w_j\}_{j=1}^J$ and $\mathcal{O}_i$ is a J-length vector with element $o_{ij} = 1$ if teacher i receives an offer from district j and zero otherwise. The conditional choice probability that teacher i chooses school j, $s_{ij}(\mathbf{w}, \mathcal{O}_i)$ is:

$$\begin{aligned} s_{ij} &= \int_\nu \frac{o_{ij}\exp(\psi\omega_j + x_j'\beta + \tau d_{ij} + \sigma\nu)}{\exp(r) + \sum_{k=1}^J o_{ik}\exp(\psi w_k + x_k'\beta + \tau d_{ik} + \sigma\nu)}\phi(\nu)d\nu \\ &\approx \sum_{n=1}^N w_n \frac{o_{ij}\exp(\psi\omega_j + x_j'\beta + \tau d_{ij} + \sigma\nu_n)}{\exp(r) + \sum_{k=1}^J o_{ik}\exp(\psi w_k + x_k'\beta + \tau d_{ik} + \sigma\nu_n)} \end{aligned} \tag{22}$$

where w_n are Quadrature weights, and ν_n are Quadrature nodes. In practice, we use 10 halton draws to approximate the ν parameter.

It will also be convenient to define the unconstrained offer supply function given by $s_{ij}(\mathbf{w})$, which is just the supply function if all workers get offers, that is, for all i, $\mathcal{O}_i$ such that $o_{ij} = 1$ for all j.[62]

Let N be the total number of teachers. The firm makes offers to the closest workers first. Thus, there is an ordering of workers represented by the function $i(n)_j = \{i_1, i_2, \cdots, i_N\}$ which indicates that teacher i receives the n^{th} offer from school district j.

Labor Supply is thus $L_j^S(\mathbf{w}, \mathbf{O}_{-j}) = \sum_{n=1}^N s_{i(n)_j j}(\mathbf{w}, (o_{i(n)_j j} = 1, \mathcal{O}_{i(n)_j, -j}))$, where $\mathbf{O}_{-j} = \{O_1, O_2, ..., O_J\}/\{O_j\}$. Remember that due to our strict ordering, O_j, the number of offers sent out, is sufficient to pin down $\mathcal{O}_j = \{o_{ij}\}_{i=1}^N$. Labor supply if all firms post wages, i.e., everyone gets an offer, is just $L_j^S(\mathbf{w}) = \sum_{i=1}^N s_{ij}(\mathbf{w})$.

C.2 Labor Demand

Labor Demand $L_j^D(w_j)$ is given by:

$$L_j^D(w_j) = \frac{w_j^{\frac{1}{\rho-1}} B_j}{\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{\rho-1}} + w_j^{\frac{\rho}{\rho-1}}}. \tag{23}$$

The elasticity of labor demand is given by:

$$\frac{\partial log(L_j^D)}{\partial log(w_j)} = \frac{1}{\rho - 1} - \frac{\frac{\rho}{\rho-1} w^{\frac{\rho}{\rho-1}}}{(\frac{\gamma}{1-\gamma})^{\frac{1}{\rho-1}} + w^{\frac{\rho}{\rho-1}}}. \tag{24}$$

[62] We restrict offers to teachers within a 60-minute commute.

C.3 Offer Game

To hire $L_j^D(w_j)$ workers, the firm needs to make offers to more than L^D workers, since some of these workers will choose to work at another school district or the outside option. The school will offer jobs in sequence, and $i(\bar{n})_j$ is the last worker that is offered a job.

Define the *best-response offer* from school j as $O_j(\mathbf{w}, \mathbf{O}_{-j})$. Pick $\bar{n} = O_j(\mathbf{w}, \mathbf{O}_{-j})$ such that

$$L_j^D(w_j) = \sum_{n=1}^{\bar{n}} s_{i(n)_j j}(\mathbf{w}, (o_{i(n)_j j} = 1, \mathcal{O}_{i(n)_j, -j})). \tag{25}$$

In practice, we use Algorithm 1.

Algorithm 1 Offer Best-Response

procedure BEST-RESPONSE OFFER($O_j(\mathbf{w}, \mathbf{O}_{-j})$)
  $\bar{n} = 0$
  $L_j^* = L^D(w_j)$.
  $T_j = 0$
  while $T_j \leq L_j^*$ or $\bar{n} \leq N$ **do**
    $\bar{n} = \bar{n} + 1$
    Update offer matrix $\mathcal{O}_{i(\bar{n})_j, j} = 1$.
    $T_j = T_j + s_{i(\bar{n})_j j}(\mathbf{w}, \mathcal{O}_{i(\bar{n})_j})$
  end while
  return $O_j = \bar{n}$
end procedure

An equilibrium of this offer game is the vector $\mathbf{O}^* = \{O_1^*, O_2^*, ..., O_N^*\}$ such that offers are a best response, given the offers sent by other schools. That is $O_j^* = O_j(\mathbf{w}, \mathbf{O}_{-j}^*)$ for all j. See Algorithm 2. Note that all non-unionized school districts automatically send out offers to every teacher, i.e., $\bar{n} = N$.

C.4 Planner Solution

The planner solution is simply the wage vector $\mathbf{w}$ such that labor supply and labor demand intersect for each school district. Define excess supply as $\xi_j(\mathbf{w}) = L_j^S(\mathbf{w}) - L_j^D(\mathbf{w})$. The planner solution $\mathbf{w}^*$ is such that $\xi_j(\mathbf{w}^*) = 0$ for all j.

In practice, we define the vector function $\xi(\mathbf{w}) \equiv \begin{bmatrix} \xi_1(\mathbf{w}) \\ \xi_2(\mathbf{w}) \\ \cdots \\ \xi_J(\mathbf{w}) \end{bmatrix}$, and look for the zero of this vector function using a root finding algorithm.

Algorithm 2 Offer Game

procedure OFFER GAME($\mathbf{O}(\mathbf{w})$)
  Start with $\bar{n}_j = L^D(w_j)$ initial offers sent so that $o_{ij} = 1$ for all j and for $i \leq i(\bar{n})_j$.
  $k = 0$ (iteration counter).
  while $|\mathbf{O}^k - \mathbf{O}^{k-1}| > 0$ **do**
    Update iteration counter $k = k + 1$
    for $j = 1, \cdots, J$ **do**
      Compute Best-Response for j: $O'_j = O_j(\mathbf{w}, \mathbf{O}^k_{-j})$
      Update offer matrix for that school: $O^k_j = O'_j$
    end for
  end while
  return O
end procedure

C.5 Posted Prices Solution

For the posted price solution, each school district sets its marginal rate of technical substitution for teachers equal to its marginal factor cost:

$$\frac{\frac{\partial W}{\partial T}}{\frac{\partial W}{\partial x}} = \frac{\gamma}{1-\gamma} \cdot \left(\frac{T_j}{X_j}\right)^{\rho-1} = w_j(1 + \frac{1}{\epsilon_j}). \tag{26}$$

We can define $\xi_j^{PO}(\mathbf{w}) = MRS_j - w_j(1+\frac{1}{\epsilon_j})$. Thus, the posted solution $\mathbf{w}^*$ is such that $\xi_j^{PO}(\mathbf{w}^*) = 0$ for all j.

In practice, we use a best-response algorithm. Schools make wage decisions holding other school wages fixed. We iterate until we reach convergence where $\xi_j^{PO}(\mathbf{w}^*) = 0$ for all j.

C.6 Nash-in-Nash Bargaining

For the Nash-in-Nash Bargaining solution, we first define how to find the wage that maximizes the Nash-Product. Find $w_j = w_j^{NB}(\mathbf{w}_{-j})$ which maximizes $\mathscr{N}_j(w_j^{NB}|w_{-j})$ (see Equation 20).

Take the first-order condition of the Nash Product given by $\frac{\partial \mathscr{N}_j(w'_j|w_{-j})}{\partial w_j} = 0$. We find $w_j^{NB}(\mathbf{w}_{-j})$ such that $\xi^{NB}(w_j^{NB}|\mathbf{w}_{-j}) = \frac{\partial \mathscr{N}}{\partial w_j} = 0$.

A Nash-in-Nash equilibrium is defined as a vector of wages $\mathbf{w}^{NN}$ such that all schools $j \in \mathscr{J}$ are choosing wages that maximize their Nash Product given the wages of other schools, that is $w_j^{NN} = w_j^{NB}(\mathbf{w}_{-j}^{NN})$ for all $j \in \mathscr{J}$.

We follow Algorithm 3. Note that non-unionized charter schools have $\alpha_j = 1$. In practice, the problem for these school collapses to Equation 26.

Algorithm 3 Nash-in-Nash Bargaining

procedure NASH-IN-NASH BARGAINING($\mathbf{w}$)
  Start with inital wage guess $\mathbf{w}^0 = \{w_j^0\}_{j=1}^J$.
  Iteration Counter $k = 0$.
  while $|\mathbf{w}^k - \mathbf{w}^{k-1}| > \epsilon$ **do**
    for $j = 1, \cdots, J$ **do**
      Find w_j' such that $\frac{\partial \mathcal{N}_j(w_j'|w_{-j}^k)}{\partial w_j} = 0$. ▷ This is a root finding exercise in one dimension.
      Update $w_j^k = w_j'$
    end for
  end while
  return w
end procedure

C.7 Parallel Implementation

Given there are 607 schools and 144,516 actual and potential teachers, we need to find a way to speed up the computation of the Nash-in-Nash equilibrium, which otherwise can take over a month to compute. To do this we use the julia distributed.jl package along with SlurmClusterManagers.jl to run the for loops in parallel across 512 nodes on the Duke DCC compute cluster. This is done by making school level choices run in parallel for both planner, poster, and Nash Bargaining solutions, holding the choices of other schools implicitely fixed.

C.8 Partition Analysis

Even with parallel implementation, the Nash Bargaining computation would take a month to run. Instead, we run our analysis using a partition of Pennsylvania into separate labor markets. We choose seven regions composed of groups of counties. We pick county borders to minimize school district interactions across borders, i.e., borders where there are few school districts directly next to each other across the border. Figure A2 shows how we partition Pennsylvania. Using this method, computing the Nash Bargaining solution goes from a month if we pick the market to be the State of Pennsylvania to about 5 days with a seven market partition.

As our counterfactuals for the Planner and Posted outcomes for the entire state only take half a day to run, versus a month for the Nash Bargaining solution, we can compare how the partition performs relative to the whole state. We find that equilibrium wages are similar across the two methods (see Table A6). We find that wages are somewhat different around border regions, as would be expected. However, these differences average out when we consider the whole state.

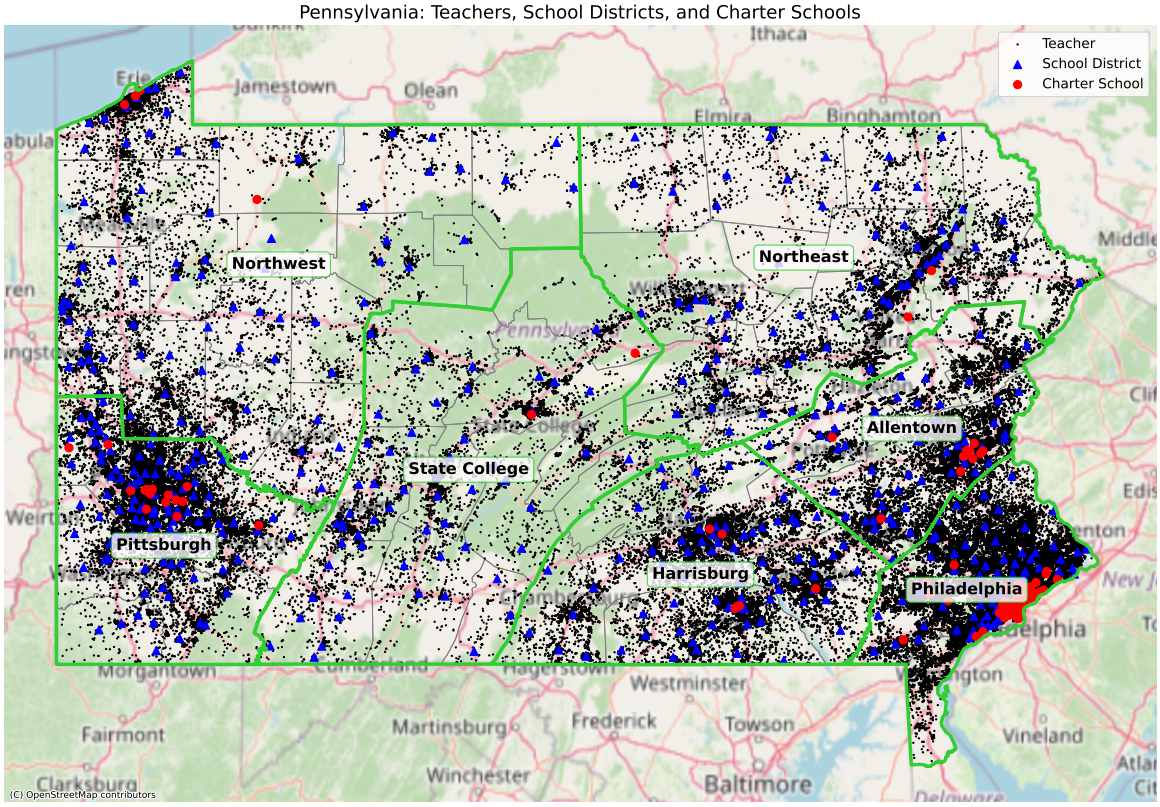


Figure A2: Pennsylvania Partition

Notes: The map depicts teacher locations and school locations in Pennsylvania. It also shows the seven regions that we used to partition Pennsylvania when we compute Nash Bargaining for the whole state.

Table A3: Partition Comparison

Penn Pl Wage	Partition Pl Wage	Penn Po Wage	Partition Po Wage
48,427.72	48,907.63	40,339.61	40,785.36
(20,013.99)	(20,028.02)	(20,203.36)	(20,242.74)

Notes: We ran the planner and posted counterfactuals for the whole state of Pennsylvania and also using the partition method. The results here show that the partition methodology produces similar results as the entire state of Pennsylvania.

Table A4: Fringe Benefits

District Cost Per Teacher	(1)	(2)
Average Wage	1.498	1.421
	(0.027)	(0.036)
Constant	7959.845	5810.115
	(2329.756)	(2693.499)
Observations	578	5652

Notes: The table shows the regression of average wage on the marginal cost of a teacher. This regression allows us to estimate the fringe benefits from the wage. Column (1) shows the regression for just the year 2017. Column (2) shows the regression for all years with school and year fixed effects. Standard errors are clustered at the school district level.

D Estimation Details

D.1 Fringe Benefits

A school's cost of hiring a teacher is not just a wage, but also includes benefits. In our model, we set the cost of a teacher equal to wages plus a fringe benefit. In practice, we ran a regression of the average wage on the marginal cost of a teacher (expenditure on teachers divided by the number of teachers hired). Table A4 shows this relationship.

D.2 Indirect Inference on σ and r

We estimate σ and r using an indirect inference procedure as described in Section 5.7. σ affects the distribution of the random coefficient ν_i that affects the choice between teaching and not teaching, and affects the slope of the labor supply curve. The parameter, r, represents the value of the outside option and rationalizes the proportion of teachers that choose the outside option. It mostly affects the level of the supply curve. Figure A3 illustrates the difference in how these parameters affect the aggregate supply curve. We derive aggregate labor supply curves at the market level for Erie and Crawford county in Pennsylvania. We change the wages in the data by a certain percentage (e.g., 50% of the wages), and then sum the conditional choice probabilities to determine how many teachers would be employed. Since we focus on labor supply here, we force schools to send out offers to everyone. The blue curve in Figure A3 represents the aggregate supply curve for our estimated parameters. If we double r (moving from the blue to the red curve), the outside option becomes more attractive; fewer teachers choose to work in public education jobs. This change shifts the labor supply curve inwards, but mostly acts as a level shift. Next, we can see that changing σ has both a level and a slope effect (moving from the red curve to the green curve). First, notice that since σ is normally distributed with a mean of zero, about half of teachers should have a positive ν_i coefficient. Since, in our model, most

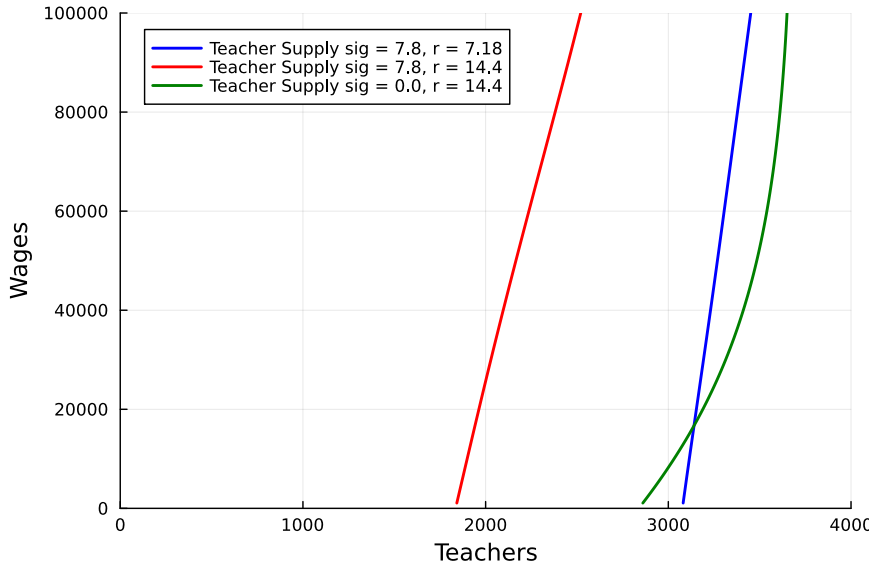


Figure A3: Aggregate Supply Curves in Erie County, Pennsylvania

Notes: The figure depicts aggregate labor supply curves for Erie County, Pennsylvania under different parameter values of σ and r. Aggregate supply curves are created by changing wages at each school district by the same percent, and then summing teachers willing to work at any school district. r has a level effect, while σ has both a level and a slope effect.

(more than half of) teachers teach, increasing *sigma* makes it more difficult to hire, because it makes the marginal teacher dislike teaching more. This leads to the level shift that pushes the labor supply curve out when we reduce sigma to zero (green curve). But σ also governs how difficult it becomes to hire teachers as schools go up the supply curve (i.e., the slope of the supply curve). Lower σ's mean that the labor supply curve is more elastic, because it is easier to hire teachers from the outside option. Thus, we also notice a slope effect.

Figure A4 depicts a heatmap of the value of the criterion function for different values of σ and r. The criterion function is minimized when σ is between 7 and 8, and r is between 6 and 8.

D.3 Teacher Shortages

When the labor supply curve binds, school districts would like to hire more teachers at the negotiated wage, but not enough teachers want to work for them. Thus, school districts where labor supply binds should report shortages, which gives us a natural test to evaluate the ability of the model to correctly identify school districts where labor supply binds. Using data from the 2017-2018 school year (which we used to calibrate the model), we show that the school districts where labor supply binds

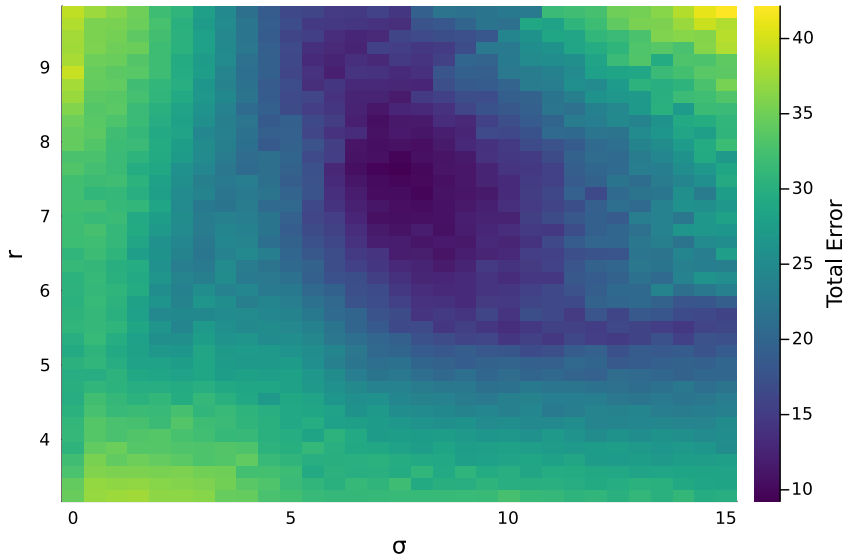


Figure A4: HeatMap of Indirect Inference Criterion Function

Notes: The figure is a heatmap of the values of the criterion function for different pairs, (σ, r). The pairs with the least errors occur in the range between 6-8.

according to our model, are significantly more likely to have a shortage. [63]

Our preferred measure of shortages comes from emergency permit data, which can be used as a proxy for shortages.[64] Emergency permits allow teachers without the proper certifications to teach at schools. "An emergency permit is issued by the Department upon the request of an employing public school entity (LEA) when a position has been advertised and no fully qualified and properly certificated applicant is available".[65] They are used primarily at schools that struggle to hire teachers. The number of emergency permits, therefore, tells us how many under-qualified teachers school districts have hired. In particular, we focus on teachers hired for long-term positions either with or without educational obligations. The 2017-2018 data demonstrates that schools where labor supply binds are more likely to have shortages. They hire almost 12 times more teachers via emergency permits than schools where labor demand binds (see Table A5.

Next, we looked at data on teacher shortage areas from the US Department of

[63] Vacancy data from 2024-2025 exists, but it was not collected in earlier years. If we assume that shortages are stable over time, then the vacancy data also supports our model. Schools with labor supply binding are more likely to have vacancies than those without. Data available at: https://ceepablog.wordpress.com/2025/03/28/teacher-shortages-examining-pennsylvania-teacher-vacancies-in2023-24-and-2024-25/

[64] Accessed September 12, 2025 https://www.pa.gov/agencies/education/data-and-reporting/school-staff/educator-preparation-and-certification.html#accordion-43b97ca7af-item-61eeb6dfb7

[65] See https://www.pa.gov/agencies/education/programs-and-services/educators/certification/current-pa-educators/emergency-permits, Accessed September 12, 2025.

Table A5: Percent Teachers hired with Emergency Permits

	LD Bind	LS Bind
Shortage	1.0%	11.6%
Number	487	120

Notes: The table determines the average percentage of teachers hired at a school using an emergency permit in 2017-18, which is a proxy for a teacher shortage (weighted by teachers). Our model predicted that 487 schools were bound by labor demand, whereas 120 schools had labor supply binding. The table indicates that schools where labor supply binds hired almost 12 times more teachers using emergency permits than those where are model predicted that labor demand binds.

Table A6: Teacher Shortage Areas

	LD Bind	LS Bind
Shortage	6.2%	100%
Number	484	2

Notes: Using 2017 data from the US Department of Education, we characterize school districts into whether there is a teacher shortage area or not. The table shows that schools where labor supply binds are much more likely to be in a teacher shortage area, than those where labor demand binds. We did not include charter schools for this result.

Education.[66] This dataset designates shortage areas at the school district level, intermediate unit level, and state level. We only use the school district level data, since the data does not have good coverage of charter schools. As shown in Table A6, both of our school districts where labor supply binds are considered to be in a shortage area in 2017, while only 6.2% of schools where labor demand binds have shortages.

E State-Level Wage Setting

To assess the effect of such a statewide wage posting, we need to endow the state agency with an objective function. In general, when looking at the effects of consolidation we would simply add up firm profits, but this is not possible in our setting since the school district objective functions are not in comparable units. Instead, we rely on compensating variation. Consider the indirect utility function $V_j(w_j, B_j) \equiv \max_{T_j \leq \sum_i s_{ij}(w_j)} W(w_j T_j, B_j - w_j T_j)$, or the maximum utility that a school district achieves given a wage w_j, budget B_j, and the labor supply constraint. Given a wage w_j^o and budget B_j^o, the compensating variation for movement to a wage w' would be the budget change $B' - B_j^o$ that yields the same indirect utility. That is, we want $B^{CV}(w')$ such that $V_j(w_j^o, B_j^o) = V_j(w', B^{CV})$. Note that because of externalities between school districts operating on competing offers, we are actually looking for a market equilibrium in terms of this compensating variation. Consider the vector of wages $\vec{w} = \{w_1, w_2, \cdots, w_J\}$ and budgets $\vec{B} = \{B_1, B_2, \cdots, B_J\}$. We find a vector of $\vec{B}^{CV} = \{B_1^{CV}, B_2^{CV}, \cdots, B_J^{CV}\}$ such that given the wage change to $\vec{w}'$ we have:

[66] Available at: https://tsa.ed.gov/#/reports.

Table A7: Parameters for Identical Schools

Parameter	Value
Number of Schools	10
Number of Teachers	500
Budget per school	750,000
Mean δ_j for Regular School	10.0
Mean δ_j for Charter	9.0

Note: All other parameters are those used in the body of the paper.

1. $V_j(\vec{w}, \vec{B}) \equiv \max_{T_j \leq \sum_i s_{ij}(w_j, w_{-j}, \boldsymbol{\mathcal{O}}_{-j})} W(w_j T_j, B_j - w_j T_j)$, where the choice of T_j also defines an offer set $\boldsymbol{\mathcal{O}}_j$. This is the equilibrium of the labor demand system in section 5.5 with a new set of wages and budgets.
2. $\vec{B}^{CV}$ such that $V_j(\vec{w}^o, \vec{B}^o) = V_j(\vec{w}', \vec{B}^{CV})$ for $j = 1, \cdots, J$.

The state budget is given by $\mathbf{B} = \sum_j B_j$. Suppose then that the state agency chooses a state wage $\bar{w}$ to minimize total expenditures subject to the constraint that no school is worse off under this allocation.[67] That is:

$$\min_{\bar{w}} \sum_j B_j^{CV}(\bar{w}). \tag{27}$$

E.1 Identical School District Simulation

For the simulation of identical school districts we use the following calibration in Table A7.

F Monte-Carlo

We perform a Monte-Carlo to assess the quitting regression used to estimate labor supply. We generate data on schools (5) and teachers (1000). To generate teacher choices, we begin by generating offers $\mathcal{O}_i = \{o_{ij}\}_{j=1}^J$ such that:

$$o_{ij} = \mathbf{1}(\gamma_0 + \gamma_1 w_j + \gamma_2 d_{ij} + \gamma_3 c_j + \gamma_4 f_j + \epsilon_{ij}^o > 0) \tag{28}$$

where c designates charter schools, f the free lunch share, and $\epsilon_{ij}^o \sim \mathcal{N}(0,1)$. That is, we estimate a probit model. For one of the options, called the *stay* option, teachers receive an offer with certainty, thus, $o_{ij} = 1$ if j is teacher i's stay school. Teachers

[67] Clearly this exercise depends on the exact starting point to compute compensating variation. We show the results with starting points at the real, planner, posted, and bargaining solutions

Table A8: Parameters for Monte-Carlo

Offer Parameters	Low Switching	High Switching
γ_0	-1	20
γ_1	-0.00001	-0.00001
γ_2	-0.04	-0.04
Implied Move Rate	4%	80%
Utility Parameters		
β_0	3	3
β_1	0.00005	0.00005
β_2	-0.01	-0.01
β_3	-0.5	-0.5
β_4	-1.2	-1.2
σ	13	13

can always choose the outside option. Thus, the stay and quit decisions are always in the offer set.

Second, we generate utilities for teachers given by:

$$u_{ij} = \beta_0 + \beta_1 w_j + \beta_2 d_{ij} + \beta_3 c_j + \beta_4 f_j + \sigma \nu_i + \epsilon_{ij}, \tag{29}$$

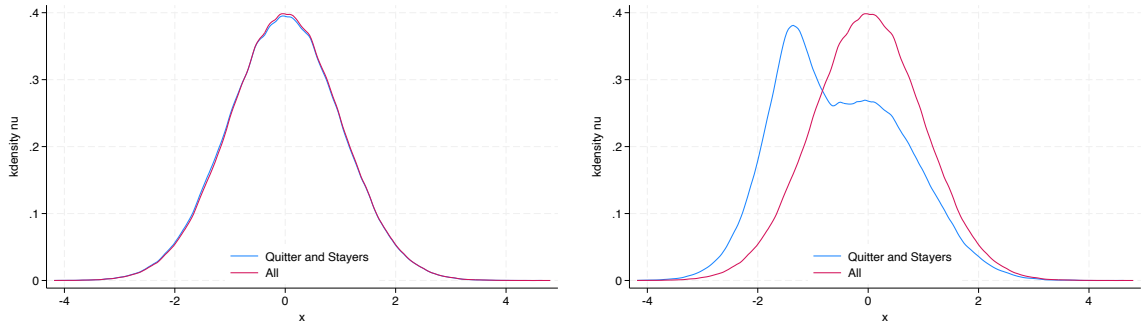
where ϵ_{ij} are logit draws and $\nu_i \sim \mathcal{N}(0,1)$, c designates charter schools, f the free lunch share.

Third, teachers choose j with the highest u_{ij} among offers $\mathcal{O}_i$: $j_i^* \equiv j \; s.t. u_{ij} \geq u_{ik} \forall k \; s.t. o_{ik} == 1$.

We use the parameters in Table A8; the utility parameters, i.e. the β's, mimic our empirical results, while the offer parameters are used to target a specific move rate, 4% or 80%.

Figure A5 shows the distribution of ν, conditional on teachers choosing to stay in their current job or quit the teaching profession, plotted in blue. The normal distribution, i.e. the unselected distribution with movers, is plotted in red. Panel a) shows the case with a 5 percent move rate, which is greater than the 2% move rate in our data, and Panel b) shows an 80 percent move rate. The distributions in Panel a) are very hard to distinguish, and indeed, accounting for selection makes little difference in our estimates. In Panel b), the selected distribution of ν is substantially different from the unselected one, since we remove a substantial fraction of movers who tend to have higher tastes of working in public education.

Table A9 presents results from our Monte-Carlo. Only Column 1 with a high move rate of 80 percent and $\sigma = 13$ has significantly different results from the truth. Column 2, where $\sigma = 0$, shows that the selection problem disappears when σ is not relevant. Similarly, Column 3 shows that there is no selection issue with a high σ,



(a) ν distribution: 5% move rate (b) ν distribution: 80% move rate

Figure A5: Selection on ν

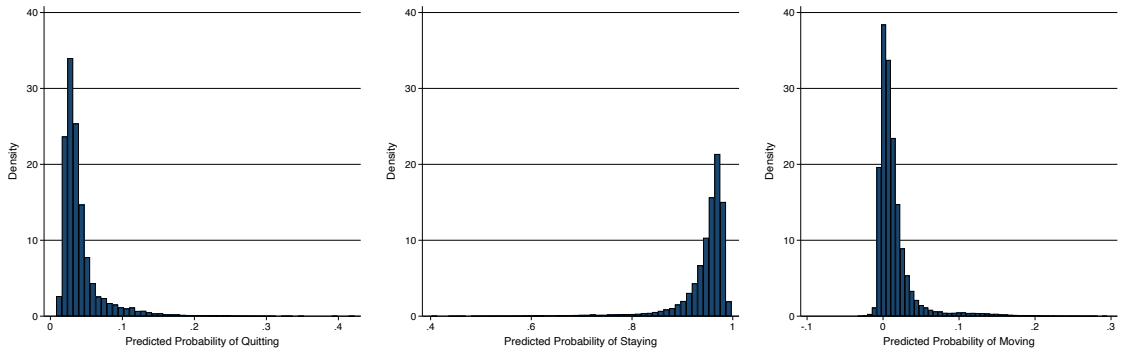


Figure A6: Predicted Probability of Quitting, Staying and Moving

ere $\sigma = 13$, but with no moves. Column 4 and 5 present results for high values of σ and a low move rate (4%) that is above what we see in the data. Finally, Column 6 presents our selection correction for the same specification as column 5.

Table A9: Monte Carlo Results

	True	(1)	(2)	(3)	(4)	(5)	(6)
Parameters							
σ		13	0	13	8	13	13
Move Rate		80%	80%	0%	4%	4%	4%
Selection Correction							X
Estimate							
Wage	0.50	0.26	0.50	0.52	0.49	0.51	0.52
		(0.60)	(0.10)	(0.17)	(0.11)	(0.17)	(0.07)
Distance	-1.00	-0.64	-0.97	-0.98	-0.95	-0.99	-1.03
		(2.25)	(0.34)	(0.72)	(0.45)	(0.69)	(0.29)
Charter	-0.50	-0.33	-0.50	-0.51	-0.48	-0.51	-0.49
		(0.84)	(0.13)	(0.19)	(0.13)	(0.19)	(0.09)
Free Lunch	-1.20	-0.57	-1.21	-1.26	-1.20	-1.25	-1.16
		(0.82)	(0.12)	(0.23)	(0.13)	(0.21)	(0.09)
Constant	3.00	23.47	2.98	3.61	3.45	3.44	4.09
		(2.51)	(0.44)	(0.73)	(0.49)	(0.72)	(0.29)
Replications		100	100	100	100	100	50

Notes: Standard deviation in parenthesis.

Table A10: Logit estimates of Quitting and Staying

	(1) quitter	(2) stayer
Real Salary (10k)	-0.110***	0.194***
	(0.013)	(0.012)
Free Lunch Share	0.560***	-0.749***
	(0.070)	(0.062)
Travel Time	0.001**	-0.001***
	(0.000)	(0.000)
Charter	0.191***	-0.591***
	(0.058)	(0.050)
Num of School Districts	0.117	-0.119
	(0.089)	(0.078)
Num of Charter	0.016	-0.002
	(0.031)	(0.028)
Num of Teachers	0.000***	-0.000**
	(0.000)	(0.000)
Avg. Free Lunch Share	-2.086***	2.553***
	(0.654)	(0.618)
Observations	75,324	75,324
Pseudo R2	0.057	0.084

Notes: Standard errors in parentheses. All regressions include fixed effects for years, intermediate units and job descriptions. Avg. Free Lunch Share refers to the average free lunch share weighted by the number of teachers.