Econometrica Supplementary Material

SUPPLEMENT TO "UNWILLING TO TRAIN?—FIRM RESPONSES TO THE COLOMBIAN APPRENTICESHIP REGULATION"

(*Econometrica*, Vol. 90, No. 2, March 2022, 507–550)

SANTIAGO CAICEDO
John E. Walker Department of Economics, Clemson University

MIGUEL ESPINOSA
Department of Economics, Universitat Pompeu Fabra and Management and Technology Department, Bocconi University

ARTHUR SEIBOLD
Department of Economics, University of Mannheim

THIS SUPPLEMENT CONTAINS the Online Appendices of "Unwilling to Train?—Firm Responses to the Colombian Apprenticeship Regulation" by Caicedo, Espinosa, and Seibold (2022). Appendix A shows additional tables and figures. Appendix B provides additional information on training courses and apprentices. Appendix C presents reduced-form results using an alternative sector classification. Appendix D contains model proofs and extensions. Finally, Appendix E provides details of the quantitative exercises.

Santiago Caicedo: scaiced@clemson.edu
Miguel Espinosa: miguel.espinosa@unibocconi.it
Arthur Seibold: seibold@uni-mannheim.de

 https://doi.org/10.3982/ECTA18911

APPENDIX A: ADDITIONAL FIGURES AND TABLES

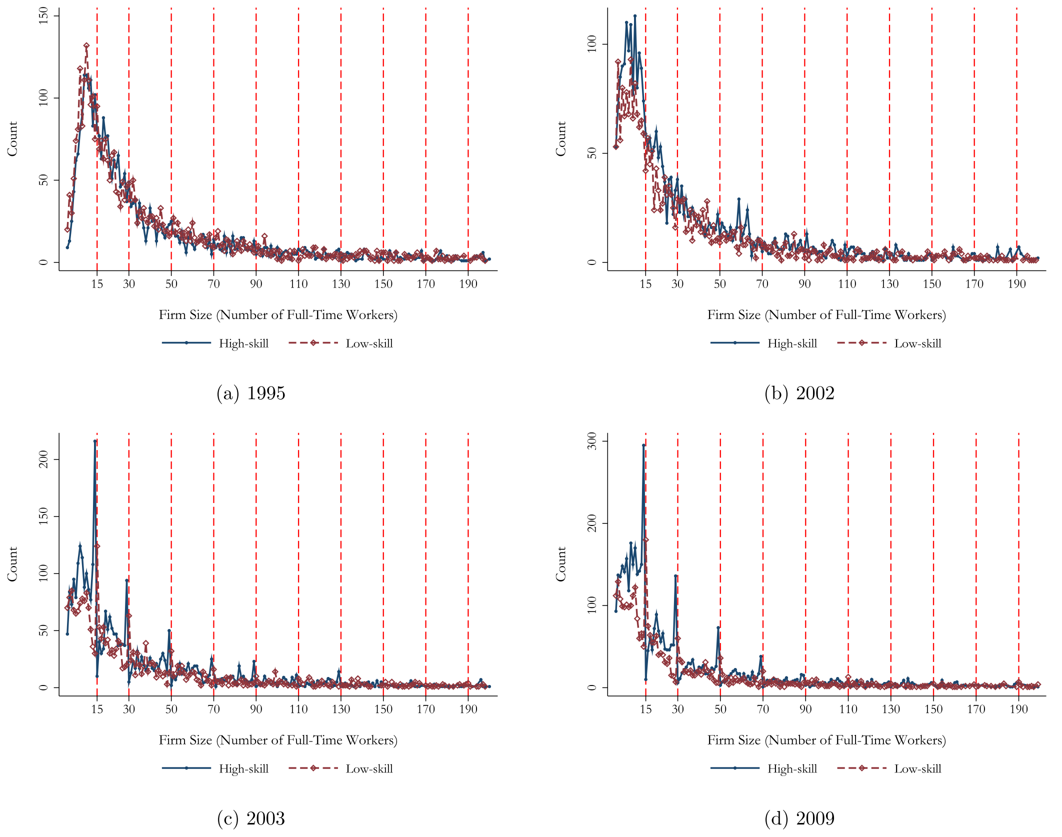
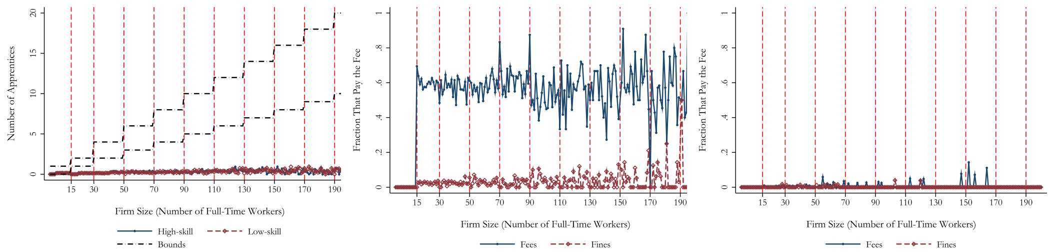
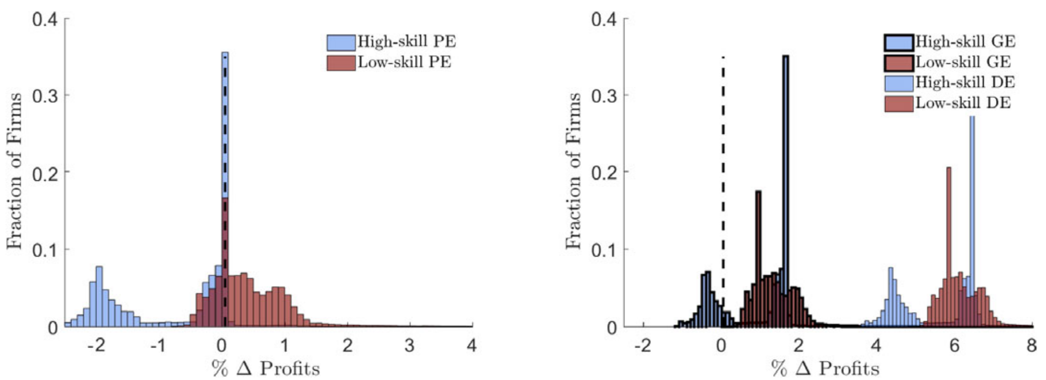


FIGURE A1.—Firm Size Distribution Pre and Post-Reform (Selected Years). *Notes*: The figure shows the distribution of full-time workers in 1995 (first available year), 2002 (last year before the reform), 2003 (first year after the reform), and 2009 (last available year), using a bin size of one. Vertical dashed lines denote the regulation thresholds. Graphs for all other years are similar and available upon request.



(a) Apprentices Prereform (b) Fees and Fines: High-skill sectors (c) Fees and Fines: Low-skill sectors

FIGURE A2.—Apprentices Prereform, Fees, and Fines. *Notes*: Panel (a) of the figure shows the average number of apprentices by firm size bin in high-skill and low-skill sector firms, prereform (1995–2002). The black dashed lines show the minimum and maximum apprentice quotas. Panels (b) and (c) show the fraction of firms paying fees and fines by sector in the post-reform years 2003 to 2009. In all panels, vertical dashed lines denote the regulation thresholds.



(a) Partial Equilibrium (b) General Equilibrium and Dynamic Effects

FIGURE A3.—Distribution of Profit Changes. *Notes*: The figure shows the distribution of percentage changes in firm profits by sector in partial equilibrium (panel a), general equilibrium and the dynamic scenario (both panel b).

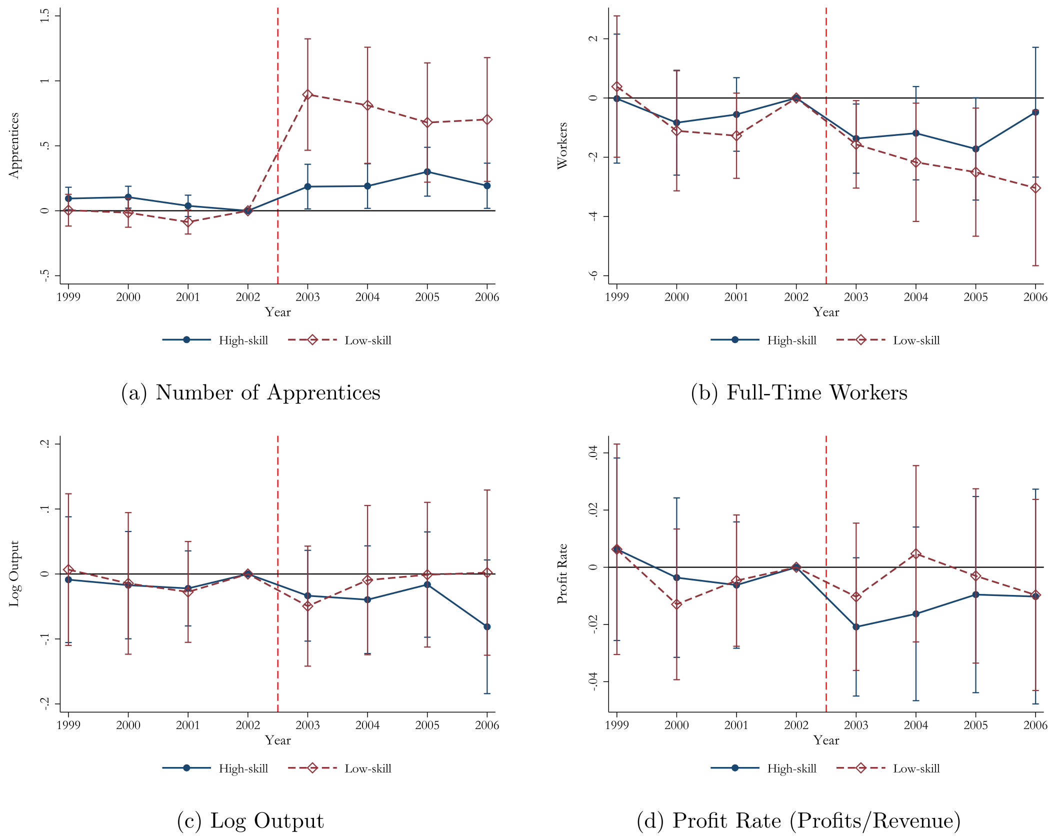


FIGURE A4.—Reduced-Form Effects of the Regulation on Firm Outcomes. *Notes*: The figure shows yearly difference-in-difference coefficients for the period 1999 to 2006, using 2002 as the base year. The vertical bars denote 95% confidence intervals based on standard errors clustered at the firm level.

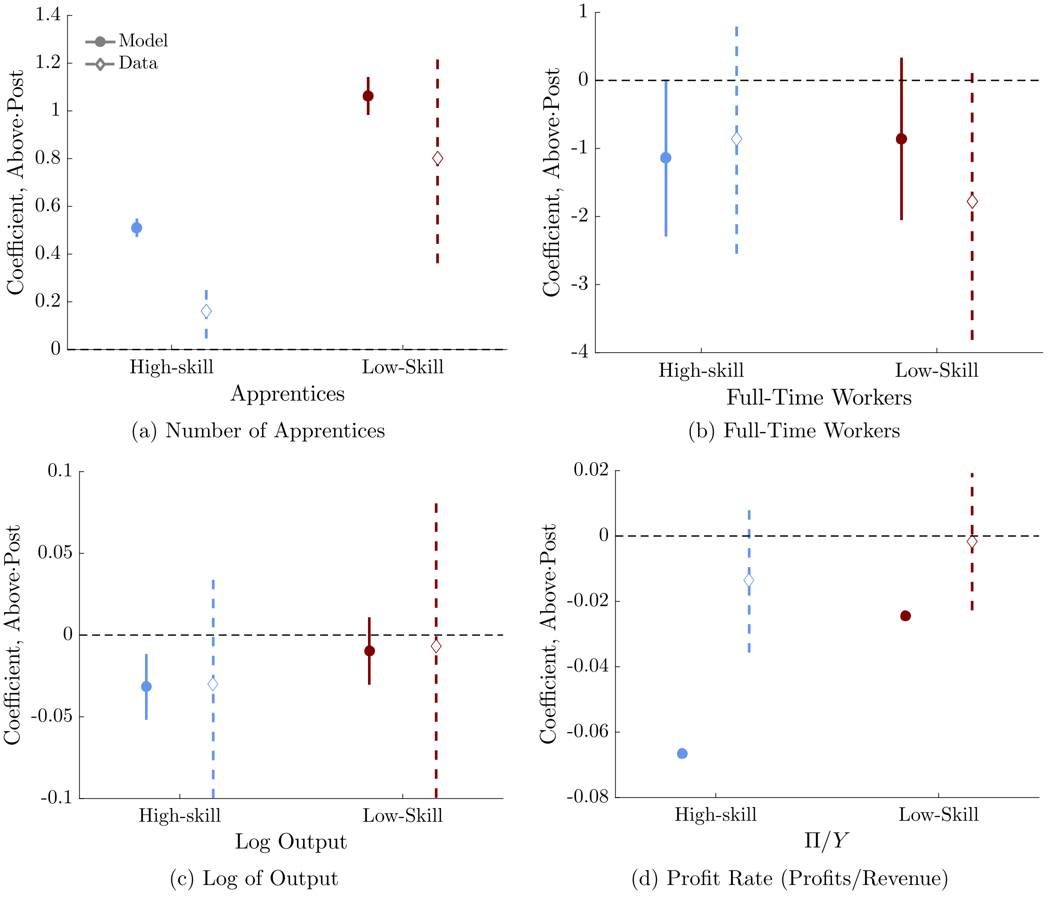


FIGURE A5.—Reduced-Form Coefficients Versus Model Prediction. *Notes*: The figure shows a comparison of reduced-form effects versus model-based effects as described in Section 5.5. The hollow diamonds depict the difference-in-difference coefficients from Table VII. The circles show analogously simulated differential effects on firms above vs. below regulation thresholds from the partial-equilibrium model. Vertical bars denote 95% confidence intervals for both sets of coefficients.

TABLE AI

CLASSIFYING SECTORS BY SKILL CONTENT.

		(1) Wood Products	(2) Textile	(3) Food/ Beverage	(4) Mineral Non-Metallic	(5) Other Manu-facturing	(6) Paper/ Editorial	(7) Metallic Products	(8) Machinery/ Equipment	(9) Chemical Products	Rank Correlation With Baseline Proxy
						Panel A: Baseline Skills Proxy					
Fraction of	Value	0.035	0.049	0.053	0.070	0.073	0.074	0.074	0.084	0.086	
Professionals	Rank	9	8	7	6	5	4	3	2	1	1.00
	High/Low	L	L	L	L	**(H)**	**H**	**H**	**H**	**H**	
						Panel B: Other Measures from Main Data					
Wage p.w.	Value	12,047	13,747	17,783	20,143	18,476	17,945	21,292	16,132	21,716	
(all workers)	Rank	9	8	6	3	4	5	2	7	1	0.68
	High/Low	L	L	L	**H**	**H**	**(H)**	**H**	L	**H**	
Wage p.w.	Value	11,791	13,051	15,235	17,040	16,628	16,799	19,981	15,508	19,636	
(prod./admin)	Rank	9	8	7	3	5	4	1	6	2	0.75
	High/Low	L	L	L	**H**	**(H)**	**H**	**H**	L	**H**	
Observations		2543	12,354	11,159	2689	5118	4585	806	10,458	7982	
						Panel C: Other Measures from Household Survey (ECH)					
Years of	Value	8.19	8.55	8.43	8.18	8.30	11.41	8.90	10.79	11.13	
Education	Rank	8	5	6	9	7	1	4	3	2	0.70
(all workers)	High/Low	L	**(H)**	L	L	L	**H**	**H**	**H**	**H**	
Fraction with	Value	0.125	0.100	0.140	0.180	0.110	0.296	0.127	0.263	0.298	
Tertiary Educ.	Rank	7	9	5	4	8	2	6	3	1	0.70
(all workers)	High/Low	L	L	**(H)**	**H**	L	**H**	L	**H**	**H**	
Fraction above	Value	0.41	0.46	0.51	0.42	0.33	0.69	0.56	0.73	0.72	
Median Wage	Rank	8	6	5	7	9	3	4	1	2	0.75
(all workers)	High/Low	L	L	**(H)**	L	L	**H**	**H**	**H**	**H**	
Observations (all workers)		920	17,142	12,343	5500	5338	2521	3657	2140	4077	

(*Continues*)

TABLE AI

Continued.

		(1) Wood Products	(2) Textile	(3) Food/ Beverage	(4) Mineral Non-Metallic	(5) Other Manu-facturing	(6) Paper/ Editorial	(7) Metallic Products	(8) Machinery/ Equipment	(9) Chemical Products	Rank Correlation With Baseline Proxy
Years of	Value	7.89	8.28	8.04	7.63	8.08	10.56	8.48	9.97	10.32	
Education	Rank	8	5	7	9	6	1	4	3	2	0.73
(prod./admin)	High/Low	L	**(H)**	L	L	L	**H**	**H**	**H**	**H**	
Fraction with	Value	0.105	0.074	0.107	0.139	0.091	0.207	0.090	0.183	0.220	
Tertiary Educ.	Rank	6	9	5	4	7	2	8	3	1	0.57
(prod./admin)	High/Low	L	L	**(H)**	**H**	L	**H**	L	**H**	**H**	
Fraction above	Value	0.39	0.45	0.49	0.39	0.32	0.65	0.54	0.70	0.69	
Median Wage	Rank	8	6	5	7	9	3	4	1	2	0.75
(prod./admin)	High/Low	L	L	**(H)**	L	L	**H**	**H**	**H**	**H**	
Observations (prod./admin)		876	16,212	11,608	5136	5103	2071	3415	1842	3544	

Note: The table shows different proxies for skill requirements by 2-digit industries. For each characteristic, the first row displays the average value among firms/workers in an industry, the second row shows the implied rank among the nine 2-digit industries, and the third row indicates whether the industry is classified as "high-skill" or "low-skill" according to this characteristic, where "L" denotes low-skill, "H" denotes high-skill, and "(H)" denotes the marginal high-skill rank 5 out of 9. Panel A shows our baseline skills proxy, the fraction of professionals, and Panel B shows rankings based on wages observed in the main firm-level data. Panel C displays rankings based on various measures from the Colombian Household Survey (ECH), corresponding to characteristics of workers by their reported industry of employment. All values shown in the table are calculated in the available pre-reform years only, namely 1995 to 2002 in Panels A and B, and 2001 to 2002 in Panel C. The last column shows the Spearman rank correlation of each characteristic with the baseline skills proxy from Panel A.

TABLE AII

SUMMARY STATISTICS PREREFORM.

	(1)	(2)	(3)	(4)
	Full Sample, Prereform Years			
	All Sectors	Low-Skill	High-Skill	t-test High vs. Low-Skill
Apprentices	0.18 (0.60)	0.17 (0.59)	0.19 (0.60)	0.014
Workers	55.14 (105.26)	55.50 (112.80)	54.79 (97.21)	0.419
Workers (Survey)	61.03 (142.59)	62.33 (153.32)	59.74 (131.07)	0.029
Fraction professionals	0.07 (0.12)	0.05 (0.11)	0.08 (0.13)	0.000
Fraction admin workers	0.36 (0.25)	0.37 (0.27)	0.35 (0.23)	0.000
Fraction production workers	0.57 (0.27)	0.58 (0.28)	0.57 (0.25)	0.521
Output	9,850,148 (26,403,640)	9,325,289 (24,970,662)	10,371,309 (27,744,018)	0.000
Value added	4,383,855 (12,330,128)	3,910,267 (11,378,001)	4,850,617 (13,185,101)	0.000
Profits	2,809,289 (8,899,159)	2,505,919 (8,176,656)	3,108,285 (9,548,693)	0.000
Wage bill (permanent workers)	1,283,054 (3,147,877)	1,134,166 (2,897,829)	1,430,893 (3,371,448)	0.000
Total wage bill	1,537,701 (3,684,091)	1,352,669 (3,315,540)	1,721,430 (4,008,340)	0.000
Wage per worker (permanent workers)	17,204 (15,239)	15,831 (14,841)	18,524 (15,497)	0.000
Capital/Output	0.69 (0.91)	0.66 (0.94)	0.72 (0.88)	0.000
Intermediates/Output	0.54 (0.18)	0.55 (0.19)	0.52 (0.17)	0.000
Observations	57,694	28,745	28,949	
Firms	10,244	5465	5388	

Note: The table shows summary statistics for the full sample in the pre-reform years 1995 to 2002. All monetary variables are in 2009 thousands of pesos. In columns (1) to (3), standard deviations are in parentheses. Column (4) shows the p-value of a t-test of equality of means in high- versus low-skill sectors.

TABLE AIII

CORRELATES OF RESPONSES TO REGULATION.

	(1) Choose Maximum Quota	(2) Choose Mininimum Quota	(3) Between Minimum and Maximum	(4) Pay Fee to Avoid Apprentices
High-skill Sector	−0.63 (0.0099)	0.047 (0.011)	−0.044 (0.0034)	0.57 (0.0049)
Number of workers	0.00021 (0.000060)	−0.00017 (0.000056)	−0.000019 (0.000030)	−0.0000072 (0.000025)
Wage per worker	0.0016 (0.00045)	−0.0015 (0.00050)	−0.000055 (0.00014)	0.00012 (0.00021)
Log output	−0.016 (0.0063)	−0.00056 (0.0065)	0.016 (0.0026)	−0.0011 (0.0034)
Output per worker	−0.00012 (0.000030)	0.00014 (0.000032)	−0.000028 (0.0000087)	−0.000020 (0.000015)
Profit rate	−0.17 (0.047)	0.24 (0.052)	−0.043 (0.016)	−0.037 (0.027)
Capital/Output	−0.056 (0.0079)	0.051 (0.0083)	0.0037 (0.0028)	−0.0062 (0.0038)
Intermediates/Output	−0.21 (0.044)	0.26 (0.048)	−0.055 (0.017)	0.0055 (0.023)
Mean dep. var.	0.29	0.30	0.053	0.32
Observations	21265	21265	21265	21265
R-squared	0.48	0.018	0.016	0.38

Note: The table shows regressions of indicators for different types of responses to the apprenticeship regulation on pre-reform firm characteristics, where the characteristics are averaged across prereform years by firm. All monetary variables in 2009 pesos, in units of thousands. All regressions include year fixed effects. Standard errors clustered by firm in parentheses.

TABLE AIV

CORRELATES OF BUNCHING BEHAVIOR.

	Bunchers Above	Bunchers Below	All Firms
Fraction of all firms	0.04	0.08	1.00
Share in high-skill sector	0.07	0.88	0.56
Mean number of workers	61.04	51.14	42.19
Choose maximum quota	0.72	0.18	0.21
Choose minimum quota	0.17	0.49	0.56
Between min./max.	0.06	0.02	0.03
Pay fee to avoid apprentices	0.05	0.27	0.17
Observations	2167	4154	50,691
Firms	1468	2624	10,740

Note: The table shows characteristics and indicators for the type of response to the apprenticeship regulation for three groups of firms: firms that bunch at (above) regulation thresholds (column 1), firms that bunch below regulation thresholds (column 2), and all firms (column 3). Only observations in the post-reform years 2003 to 2009 are included.

TABLE AV

REDUCED-FORM EFFECTS OF THE POLICY ON ADDITIONAL FIRM OUTCOMES.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Workers by Type							
	Workers		Professionals		Admin Workers		Production Workers	
	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill
Above*Post	−0.855	−1.781	−0.0556	−0.0649	−0.199	−0.956	−0.813	−0.898
	(0.863)	(1.036)	(0.224)	(0.185)	(0.359)	(0.445)	(0.652)	(0.781)
Mean (Prereform)	30.46	30.30	2.581	1.500	9.393	10.08	19.10	19.08
Observations	8491	6357	7471	5587	8491	6357	8491	6357
R-squared	0.904	0.894	0.767	0.867	0.887	0.859	0.874	0.883

	Panel B: Other Measures of Labor Input							
	Workers		Workers (Survey)		Temp. Workers		Outsourced Workers	
	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill
Above*Post	−0.855	−1.781	−0.757	−2.372	0.604	0.392	−0.131	2.134
	(0.863)	(1.036)	(0.891)	(1.304)	(0.643)	(1.805)	(1.137)	(2.247)
Mean (Prereform)	30.46	30.30	31.46	32.33	3.329	4.831	4.719	7.299
Observations	8491	6357	8491	6357	8491	6357	8491	6357
R-squared	0.904	0.894	0.907	0.899	0.687	0.597	0.752	0.880

	Panel C: Other Inputs							
	Log Capital		Log Intermediates		Log Energy		Log Water	
	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill
Above*Post	0.00349	0.0272	−0.0268	−0.00554	−0.0228	−0.0112	−0.0304	0.0298
	(0.0483)	(0.0689)	(0.0382)	(0.0502)	(0.0433)	(0.0497)	(0.0562)	(0.0760)
Mean (Prereform)	13.44	13.34	13.58	13.82	11.67	11.98	9.978	9.837
Observations	8491	6357	8491	6357	8491	6357	8407	6288
R-squared	0.834	0.850	0.861	0.871	0.876	0.882	0.787	0.753

	Panel D: Output, Value Added, Productivity							
	Log Output		Log Value Added		Log Output per Worker		TFP	
	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill
Above*Post	−0.0302	−0.00697	−0.0429	0.0190	0.0139	0.0122	−0.0116	−0.0128
	(0.0355)	(0.0472)	(0.0423)	(0.0553)	(0.0331)	(0.0465)	(0.0284)	(0.0361)
Mean (Prereform)	14.31	14.46	13.52	13.49	11.21	11.35	5.509	5.581
Observations	8491	6357	8491	6357	8376	6211	8058	5909
R-squared	0.866	0.862	0.814	0.786	0.772	0.813	0.794	0.822

Note: The table shows results from difference-in-difference regressions as described by equation (4). Regressions are run on the threshold sample, using years 1999 to 2006, and include year and firm fixed effects. Standard errors clustered at the firm level in parentheses.

APPENDIX B: ADDITIONAL INFORMATION ON TRAINING COURSES AND APPRENTICES

The results in this Appendix are based on several additional data sources. Two data sets allow us to gather additional information on apprentices' characteristics. The first data set is a survey of school-to-work transitions (ETET) focused on young individuals aged between 14 and 29 years. The survey was conducted by the Colombian government in 2013 and 2015. The second data set are monthly administrative records from the social security system (PILA) gathered by the Colombian Ministry of Health and Social Protection and processed by the Colombian National Statistical Agency (DANE). This data is available for 2015 and 2016. In both data sets, we pool the available years. Moreover, we use the Colombian household survey (ECH), which is available for the years 2001 to 2006. This data set is administered by DANE and constitutes the first statistical tool used by the Colombian government to study labor markets. Unfortunately, apprentices cannot be identified in the ECH data, but we can observe other workers' characteristics.

This Appendix is organized as follows. Tables BI and BII show information on specific apprenticeship training courses. Table BIII shows characteristics of apprentices and other workers based on various data sources. Figures B1 and B2 as well as Table BIV focus on apprentices in the PILA data who complete their apprenticeship during the sample period. Specifically, Figures B1 and B2 show wages of this group around the month of graduation, compared to a control group of apprentices that do not graduate throughout this period. Table BIV shows transition probabilities between firms and sectors and wages of apprentices after they complete their apprenticeship. Finally, Table BV shows the occupational distribution by sector as observed in the ECH data.

TABLE BI

LARGEST TRAINING COURSES BY ENROLLMENT (ALL SECTORS, 2015).

Sector	Training Course	Trainees Enrolled
General	Systems	2870
Services	Administrative Assistance	2433
Services	Accounting of Commercial and Financial Operations	2121
Health	Environment Management	1187
Agricultural	Agricultural Production	1057
Health	Occupational Safety	988
Services	Accounting and Finances	902
General	Analysis and Development of Information Systems	854
Commercial	Sales of Products and Services	796
Services	Administrative Management	746
Commercial	Business Management	737
Services	Kitchen	643
Electricity	Electrical Residential Installations	597
General	Software Programming	555
Electricity	Maintenance of Computer Equipment	514
Agricultural	Ecological Agricultural Systems	512
General	Human Resources	508
Construction	Building Construction	487
Services	Assistance for Organization of Archives	448
Commercial	Market Management	445
Services	Food Agroindustry	441

Note: The table shows the 20 largest training courses by enrollment in SENA in 2015. "General" refers to the training course that span several sectors (denoted as "transversal" in original data). Apprentices cannot be enrolled in more than one training course simultaneously. Source: SENA (2016).

TABLE BII

EXAMPLES OF SPECIFIC TRAINING COURSES (MANUFACTURING, 2020).

Sector	Training Course	Required Education (Years)	Length of Training (Months)
Manufacturing	Setting up Analysis and Testing Laboratories for Industry	9	15
Manufacturing	Inspection and Testing with Nondestructive Processes	9	15
Manufacturing	Industrial Manufacture of Outerwear	5	6
Manufacturing	Maintenance of Industrial Automation	9	15
Manufacturing	Paint Coating for Transportation Equipment	9	15
Manufacturing	Chemical Industry Processes	11	27
Manufacturing	Metal Joinery	9	15
Manufacturing	Light Vehicle Maintenance	9	15
Manufacturing	Maintenance of Diesel Engines	9	15
Manufacturing	Industrial Production Management	11	24
Manufacturing	Mechatronic Automotive Maintenance	11	24
Manufacturing	Cabinet Making	5	15
Manufacturing	Manufacture of Products with Fiberglass and Composite Materials	9	15
Manufacturing	Cutting of Footwear and Leather Goods	5	9
Manufacturing	Polymer Transformation by Injection	9	15
Manufacturing	Maintenance of Motorcycles and Motorcars	9	15
Manufacturing	Carpentry	5	15
Manufacturing	Food Processing	11	24
Manufacturing	Aluminum Joinery	9	12
Manufacturing	Machining on a Conventional Lathe and Milling Machine	9	15
Manufacturing	Industrial Machinery Mechanics	9	15
Manufacturing	Armed Jewelry	9	15
Manufacturing	Electrical Maintenance and Electronic Control of Automobiles	9	15

Note: The table shows examples of training courses in the manufacturing sector offered by SENA in 2020, with the corresponding education requirements and the length of the training course. Source: SENA website (http://oferta.senasofiaplus.edu.co/sofia-oferta/inicio-sofia-plus.html).

TABLE BIII

CHARACTERISTICS OF APPRENTICES.

	(1) Apprentices (ETET)	(2) Other Young Workers (ETET)	(3) Apprentices (PILA)	(4) Other Workers (PILA)	(5) Apprentices (EAM)
Fraction Female	68.4%	44.7%	57.9%	39.2%	54.7%
Age	20.94	23.35	23.51	35.18	
Fraction 14 to 19	35.7%	18.4%	19.5%	2.5%	
Fraction 20 to 24	51.0%	39.1%	50.8%	17.3%	
Fraction 25 to 29	13.3%	42.5%	15.3%	20.5%	
Education					
Primary or less	0.0%	5.4%			
Some Secondary	2.0%	14.0%			
High School	18.4%	32.7%			
Technical/vocational	49.0%	24.6%			
College	30.6%	23.3%			
Last monthly wage	513,367	808,493			
Wage relative to minimum wage			1.04	2.43	
Socioeconomic status					
Low	44.8%	66.2%			
Medium	51.0%	31.2%			
High	4.2%	2.6%			
Occupational group					
Professional					22.2%
Production Worker					39.5%
Administrative worker					38.2%
Field of training/occupation					
Humanities and arts	4.1%	5.8%			
Social sciences and business	35.7%	24.3%			
Sciences	3.1%	3.6%			
Engineering, industry and construction	23.5%	27.5%			
Agriculture	0.0%	1.3%			
Health and social services	13.3%	17.0%			
Education	1.0%	6.6%			
General services	9.2%	5.3%			
Other or not specified	10.2%	8.6%			
Sector of training/employment					
Agriculture	0.0%	0.7%			
Mining	0.0%	0.3%			
Manufacturing	24.5%	12.9%			
Electricity	3.1%	0.5%			
Construction	2.0%	6.3%			
Trade and consumer services	22.4%	35.5%			
Transport, storage, and communication	2.0%	9.7%			
Finance	6.1%	2.5%			
Real State	9.2%	9.5%			
Public administration	28.6%	13.3%			
Other services	2.0%	8.8%			

(*Continues*)

TABLE BIII

Continued.

	(1) Apprentices (ETET)	(2) Other Young Workers (ETET)	(3) Apprentices (PILA)	(4) Other Workers (PILA)	(5) Apprentices (EAM)
Satisfied with job/apprenticeship	91.8%	81.2%			
Want to move job/apprenticeship	29.6%	42.5%			
Number of observations	98	6467	99,055	1,242,423	20,546

Note: The table summarizes characteristics of apprentices and workers based on various data sources. Columns (1) and (2) shows individual characteristics from the ETET data. "Other young workers" are 14 to 29 years old. Last monthly wage in Colombian Pesos. Columns (3) and (4) show individual characteristics from the PILA data. Column (5) shows additional information from the firm-level EAM data, where the shares of apprentices (including interns) by gender and by broad occupational group are observed for a subset of firms and years. Number of observations corresponds to number of individuals in Columns (1) to (4) and to number of firms in Column (5).

TABLE BIV

FIRM AND SECTOR TRANSITIONS AFTER COMPLETING THE APPRENTICESHIP.

	(1)	(2)	(3)
		Trained in	
	All	High-Skill Sector	Low-Skill Sector
Panel A: Probability to Remain in PILA Data after Training			
Remain in Data			
Probability	23.94%	23.23%	24.48%
Observations	89,302	38,814	50,488
Panel B: Firm and Sector Transitions			
Stay in Same Firm			
Probability conditional on remaining in data	74.80%	74.39%	75.10%
Unconditional Probability	17.90%	17.28%	18.38%
Wages	1.40	1.47	1.35
	(0.80)	(0.97)	(0.64)
Move Firms within 2-Digit Industry			
Probability conditional on remaining in data	10.54%	7.45%	12.79%
Unconditional Probability	2.52%	1.73%	3.13%
Wages	1.41	1.58	1.33
	(0.72)	(0.90)	(0.62)
Move to Different Industry (High-Skill Sector)			
Probability conditional on remaining in data	7.53%	7.82%	7.32%
Unconditional Probability	1.80%	1.82%	1.79%
Wages	1.50	1.51	1.44
	(1.01)	(1.42)	(0.63)
Move to Different Industry (Low-Skill Sector)			
Probability conditional on remaining in data	7.13%	10.34%	4.79%
Unconditional Probability	1.71%	2.40%	1.17%
Wages	1.37	1.45	1.33
	(0.67)	(0.75)	(0.68)
Observations	21,376	9017	12,359

Note: The table shows information on apprentices' firm and sector transitions based on the PILA data. Panel A shows the probability of an apprentice remaining observed in the data. Panel B reports the transition probabilities and wages of apprentices moving across firms and industries during the sample period. Wages are shown in units relative to the minimum wage with standard deviations in parentheses.

TABLE BV

OCCUPATIONAL DISTRIBUTION BY SECTOR.

	(1)	(2)
	Fraction of Workers in	
	Low-Skill Sectors	High-Skill Sectors
Professionals (0,1)	0.023	0.050
Legislative Bodies/Directors (2)	0.041	0.056
Office Workers (3)	0.057	0.087
Sales/Trade Workers (4)	0.102	0.087
Hospitality/Service Workers, etc. (5)	0.065	0.025
Agricultural Workers (6)	0.026	0.005
Mining/Industrial Workers (7)	0.458	0.062
Electricians/Mechanics, etc. (8)	0.128	0.365
Craftsmen/Construction Workers, etc. (9)	0.101	0.262
Observations	109,284	58,382

Note: The table shows the fraction of workers in different occupations by sector, based on data from the Colombian Household Survey (ECH). Occupations are categorized by 1-digit occupation codes (shown in parantheses).

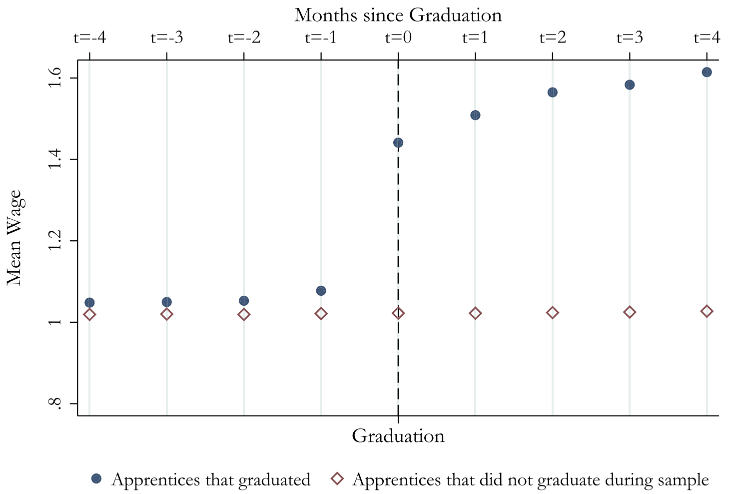
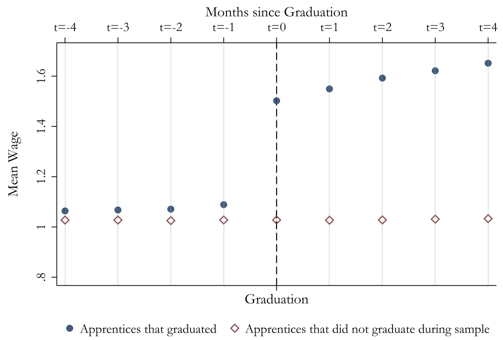


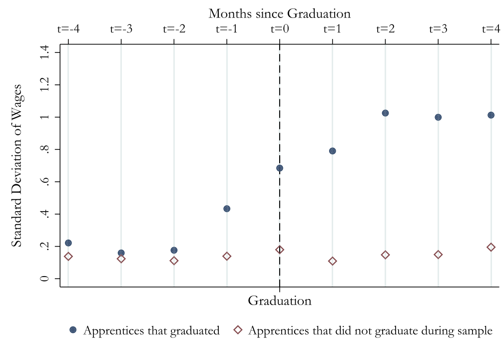
FIGURE B1.—Wages Before and After Completing the Apprenticeship. *Notes*: The figure shows wages of apprentices who graduate during the sample period versus a control group of apprentices who do not graduate during this time, based on the PILA data. Time is measured in months relative to time of completing the apprenticeship. Wages are shown in units relative to minimum wage.



(a) Mean (High-Skill Sectors)

(b) Mean (Low-Skill Sectors)

(c) Standard Deviation (High-Skill Sectors)

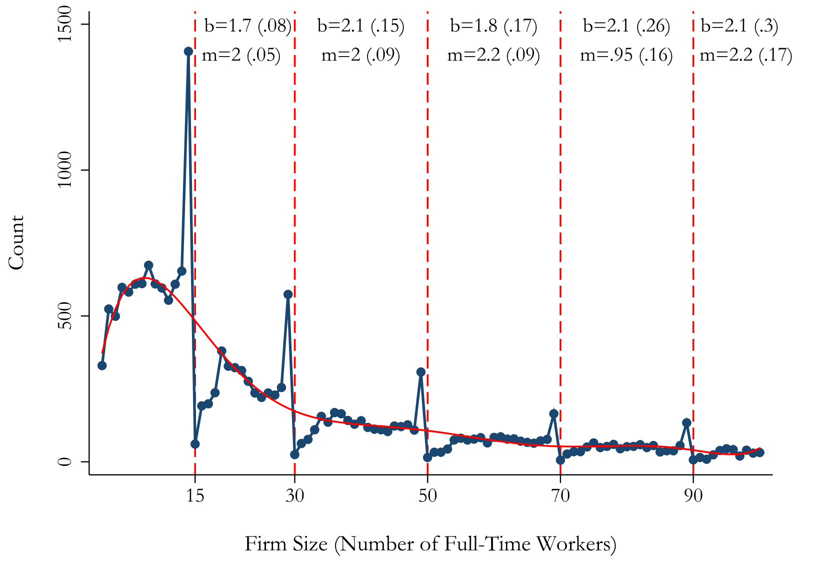


(d) Standard Deviation (Low-Skill Sectors)

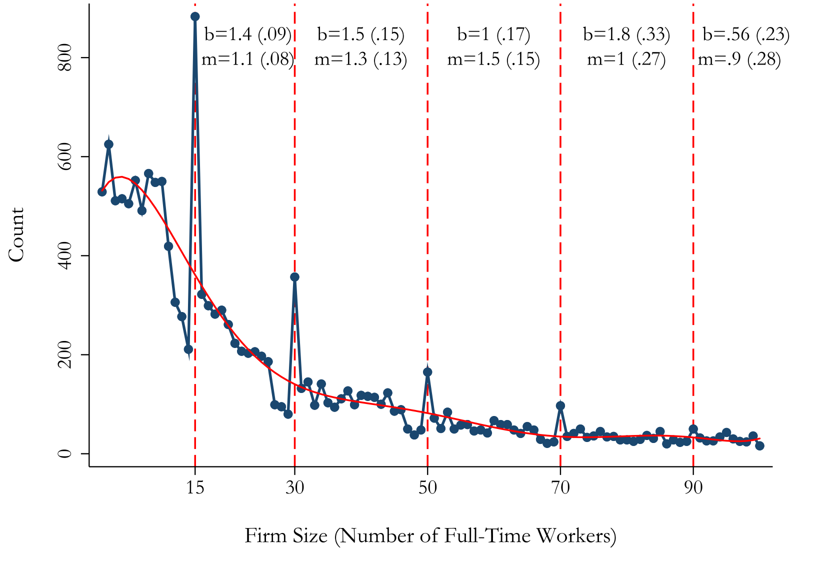
FIGURE B2.—Wages Before and After Completing the Apprenticeship by Sector. *Notes*: The figure shows the mean and standard deviation of the wages of apprentices who graduate during the sample period versus a control group of apprentices who do not graduate during this time, based on the PILA data. Time is measured in months relative to time of completing the apprenticeship. Wages are shown in units relative to minimum wage.

APPENDIX C: REDUCED-FORM RESULTS WITH ALTERNATIVE SECTOR CLASSIFICATION

In this Appendix, we replicate the main reduced-form results from Section 3 with an alternative sector classification. We focus only on those sectors which can be unambiguously classified as high- or low-skill according to the various skill proxies shown in Appendix Table AI. The set of "clearly high-skill" sectors includes paper/editorial, metallic products, machinery, and chemical products, which are classified as high-skill according to all or virtually all proxies. The set of "clearly low-skill" sectors includes wood products, textile, and food/beverage, all of which are classified as low-skill according to all or virtually all proxies. Thus, compared to the main reduced-form results, two more ambiguous sectors are excluded, namely other manufacturing and mineral nonmetallic products.

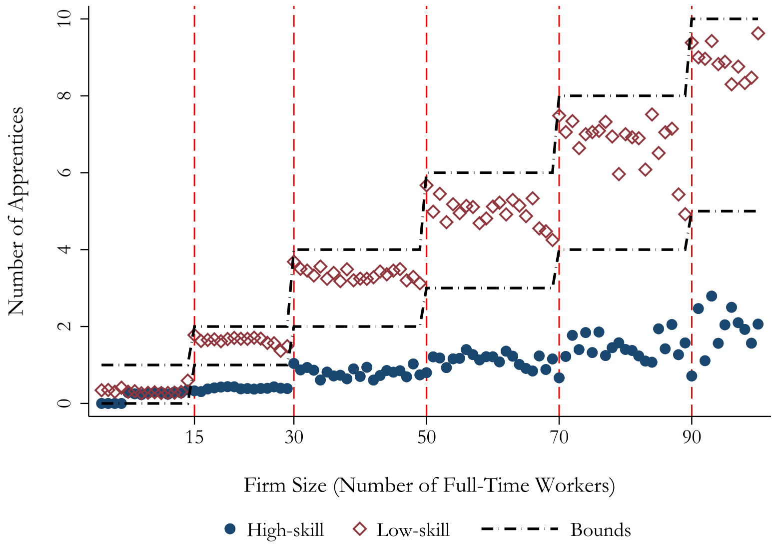


(a) High-Skill Sectors

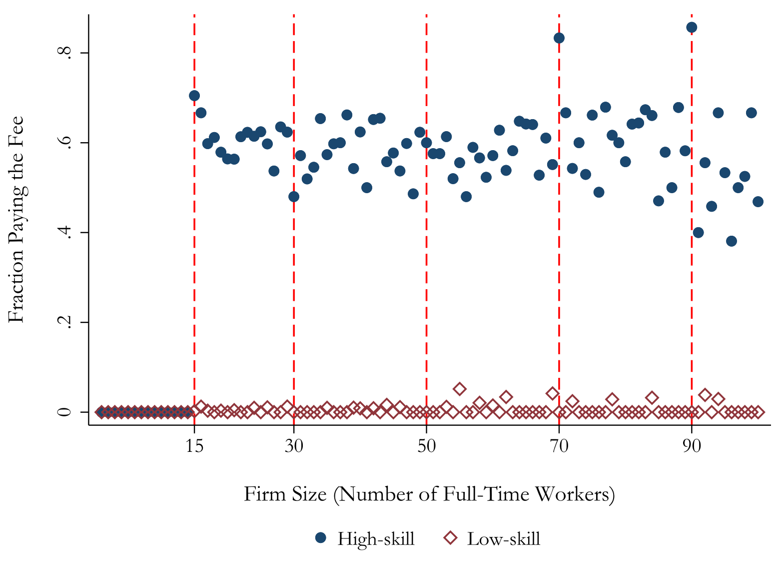


(b) Low-Skill Sectors

FIGURE C1.—Bunching Responses with Alternative Sector Definition. *Notes*: The figure replicates main text Figure 3, excluding firms in the other manufacturing and mineral non-metallic sectors, for which the assignment to high- and low-skill is less clear. The figure shows the distribution of the number of full-time workers in clearly high-skill and clearly low-skill sectors post-reform (2003–2009), using a bin size of one. The dashed vertical lines denote the regulation thresholds. The solid thin line shows the fitted counterfactual density. Excess mass b and missing mass m are reported at each threshold, with bootstrapped standard errors in parentheses.



(a) Fact 2—Number of Apprentices by Firm Size



(b) Fact 3—Share of Firms Paying Fees by Firm Size

FIGURE C2.—Number of Apprentices and Share of Firms Paying Fees with Alternative Sector Definitions. *Notes*: The figure replicates main text Figure 4, excluding firms in the other manufacturing and mineral non-metallic sectors, for which the assignment to high- and low-skill is less clear. Panel (a) shows the average number of apprentices by firm size in clearly high-skill and clearly low-skill sector firms. The horizontal dashed lines show the minimum and maximum apprentice quotas, and the vertical dashed lines denote the regulation thresholds. Panel (b) shows the fraction of firms paying fees by firm size in clearly high-skill and clearly low-skill sectors. Both panels pool the post-reform years 2003 to 2009.

APPENDIX D: MODEL PROOFS AND EXTENSIONS

D.1. *Equilibrium and Proofs*

Here, we characterize the equilibrium with and without regulation, show some additional theoretical results and provide a proof of Proposition 1.

Assumption 1 on f allows us to further characterize the optimal number of workers and apprentices, guaranteeing the existence and uniqueness of the solution. Conditions (i) and (ii) ensure the existence of a unique solution with $n > 0$. Condition (iii) implies that the optimal number of workers $n^*(z, t_a)$ and apprentices $n_a^*(z, t_a)$ are nondecreasing in z. In other words, firms with higher managerial ability z are larger. We formalize these claims in Lemma 1.

LEMMA 1: *Assumption* 1 *implies unique labor demands* $n^*(z, t_a) > 0$ *and* $n_a^*(z, t_a) \geq 0$ *solving firm z's optimization problem* (1). *Moreover, these labor demands are nondecreasing in the managerial ability,* $\frac{\partial n^*}{\partial z} \geq 0$ *and* $\frac{\partial n_a^*}{\partial z} \geq 0$.

PROOF: Similar to standard production theory, homogeneity of degree $\gamma \in (0, 1)$ in (l, l_a) implies concavity (and hence quasi-concavity) of the production function and the existence of the solution $l^*(z)$, $l_a^*(z)$. Additionally, the Inada condition on n guarantees that the solution is unique. From these labor demands, we compute the optimal number of workers n^* and apprentices n_a^*, $n^* = l^* + t_a n_a^*$, and $n_a^* = \frac{l_a^*}{\zeta_a}$.

Now, since the cross-derivatives are nonnegative, monotone comparative statics imply $\frac{\partial l^*(z)}{\partial z} \geq 0$ and $\frac{\partial l_a^*(z)}{\partial z} \geq 0$. Thus, $\frac{\partial n^*}{\partial z}$, $\frac{\partial n_a^*}{\partial z} \geq 0$. *Q.E.D.*

At an interior solution, the marginal rate of substitution between the two types of labor is equal to the ratio of marginal labor costs

$$-\frac{\frac{\partial f}{\partial l}}{\frac{\partial f}{\partial l_a}} = -\frac{w\zeta_a}{w_a + t_a w}.$$

We can use the FOCs to analyze how wages or required training time affect the optimal labor allocation decision. As usual, an increase in the relative wage of apprentices lowers demand for apprentices. Similarly, an increase in training costs decreases the demand for apprentices.

LEMMA 2: *Suppose that Assumption* 1 *holds. Then* $\frac{n_a}{n}$ *is weakly decreasing in* w_a *and* t_a, *and weakly increasing in* w.

PROOF: At an interior solution, from the firm's optimization problem,

$$\frac{dl_a}{dl} = -\frac{\frac{\partial f}{\partial l}}{\frac{\partial f}{\partial l_a}} = -\frac{w\zeta_a}{w_a + t_a w}.$$

Let $W = \frac{w\zeta_a}{w_a + t_a w}$ denote the ratio of the price of workers' and apprentices' labor. In equilibrium, if W increases, then $\frac{dl_a}{dl}$ decreases. Since f is homogeneous of degree $\gamma \in (0, 1)$ in (l, l_a), this means l_a^*/l^* increases.

Now, $\frac{n_a^*}{n^*} = \frac{1}{\zeta_a \frac{l^*}{l_a^*} + t_a}$. All comparative static results follow from this equation and the previous observations.

If w_a increases, then W decreases, so l_a^*/l decreases, implying n_a^*/n^* also decreases. Similarly, an increase in t_a implies W and l_a^*/l^* decrease. Now, this increase in t_a decreases n_a^*/n^* directly and indirectly through l_a^*/l^*, so n_a^*/n^* decreases. At the boundary, $n^* = t_a n_a^*$, and $\frac{n_a^*}{n^*} = \frac{1}{t_a}$ is also decreasing in t_a. The analogous logic applies to w. If w increases, W and l_a^*/l^* also rise, implying n_a^*/n^* increases. *Q.E.D.*

Equilibrium

All individuals in the economy are infinitely lived, have a common utility function, and are endowed with a unit of labor which they supply inelastically. Each individual i maximizes their lifetime utility

$$\max_{(c_t)_t} \sum_{t=1}^{\infty} \beta^t u(c_t) \quad \text{s.t.} \sum_{k=1}^{K} p_t^k c_t^k = I_t^i \; \forall t, \tag{6}$$

where $I_{i,t}$ denotes income in period t. Note that as for firms', individuals' decisions are static. Solving this problem implies the usual optimality conditions, $\frac{\partial u_t/\partial c_t^j}{\partial u_t/\partial c_t^k} = \frac{p_t^k}{p_t^j}$ and $\sum_{k=1}^{K} p_t^k c_t^k = I_t^i \; \forall k, t$. Assuming $u(\cdot)$ is quasiconcave, let $c_t^*(I_t^i; p_t) = (c_t^{1*}(I_t^i; p_t), \dots, c_t^{K*}(I_t^i; p_t))$ be the solution to individual i's optimization problem in period t. We use the goods market clearing conditions for each sector k to determine prices $p_t = (p_t^1, \dots, p_t^K)$,

$$C_t^k(p_t) + C_{a,t}^k(p_t) + C_{f,t}^k(p_t) = Y_t^k(p_t) \quad \forall k,$$

where $C_t^k(p_t)$ is workers' aggregate demand for good k, $C_{a,t}^k$ is apprentices' aggregate demand, $C_{f,t}^k$ is firm owners' aggregate demand and $Y_t^k(p_t)$ is aggregate production of good k:

$$C_t^k(p_t) := \sum_{j=1}^{K} L_t^j c_t^{k*}(w_t^j; p_t), \qquad C_{a,t}^k := \sum_{j=1}^{K} L_{a,t}^j c_t^{k*}(w_{a,t}; p_t),$$

$$C_{f,t}^k := \sum_{j=1}^{K} F_t^j \int\int c_t^{k*}(\pi_t^j(z, t_a); p_t)\, d\mathcal{Z}(z)\, d\mathcal{T}(t_a),$$

$$Y_t^k(p_t) := F_t^k \int\int y_t^{k*}(p_t; z, t_a)\, d\mathcal{Z}(z)\, d\mathcal{T}(t_a),$$

where F_t^j denotes the number of firms in sector j at time t.

DEFINITION 1: A competitive equilibrium is given by wages $((w_t^{k*})_k, w_{a,t}^{k*})_t$ and prices p_t^k for each sector k and each period t; and quantities of unemployed workers and untrained apprentices $((U_t^{k*})_k, U_{a,t}^*)_t$, labor demands $(n_t^{k*}(z, t_a), n_{a,t}^{k*}(z, t_a))$ for each firm (z, t_a) and consumption c_t^{k*} such that

(i) firms solve the optimization problem (1),

(ii) wage restrictions are satisfied, $w_t^{k*} \geq w_{a,t}^{k*} \geq w_{\min}$, and labor markets clear with $U_t^{k*} \geq 0$ and $U_{a,t}^* \geq 0 \; \forall t, k$.

(iii) apprentices increase total labor in each period and all sectors, as in (2),
(iv) individuals maximize utility (6) and,
(iv) the goods market clear for each sector.

Note that unemployment and untrained apprentices exist whenever the wage restrictions are binding.

Equilibrium With Regulation

Lemma 3 characterizes the solution to the optimization problem with regulation relative to the problem without regulation. We use this characterization for the proof of Proposition 1.

Let $n^*(z, t_a)$, $n_a^*(z, t_a)$ denote the optimal number of workers and apprentices a firm with managerial ability z and training costs t_a hires when solving the maximization problem (1) (without regulation); and $n^r(z, t_a)$ and $n_a^r(z, t_a)$ denote the optimal number of workers and apprentices the firm hires when solving (3) (with regulation).

LEMMA 3: *Let $j \geq 1$ be such that $n^*(z, t_a) \in [N_{j-1}, N_j)$ and $(\underline{n}_a^j, \overline{n}_a^j)$ be the corresponding minimum and maximum quotas.*

i. *If $n_a^*(z, t_a) > \overline{n}_a^j$, then*
   - *$\exists k \geq j$ such that $n^r(z, t_a) = N_k$ and $n_a^r(z, t_a) > \overline{n}_a^j$ (increase size to get more apprentices)*, or
   - *$n_a^r(z, t_a) = \overline{n}_a^j$ and $n^r(z, t_a) < n^*(z, t_a)$ (bounded by maximum quota).*

ii. *If $n_a^*(z, t_a) \in [\underline{n}_a^j, \overline{n}_a^j]$, then $n^r(z, t_a) = n^*(z, t_a)$ and $n_a^r(z, t_a) = n_a^*(z, t_a)$.*

iii. *If $n_a^*(z, t_a) < \underline{n}_a^j$, then*
   - *$\exists k < j$ such that $n^r(z, t_a) = N_k - \varepsilon$ (with $\varepsilon \to 0$) and $n_a^r(z, t_a) < \overline{n}_a^j$ (reduce size to avoid apprentices)*,
   - *$n^r(z, t_a) \geq n^*(z, t_a)$ and $n_a^r(z, t_a) < \underline{n}_a^j$ and $d_{f_a} = 1$ (pay the fee to avoid apprentices)* or
   - *$n_a^r(z, t_a) = \underline{n}_a^j$ (bounded by the minimum quota).*

PROOF: Pick any firm $z > 0$ and $t_a \geq 0$. Denote by $\pi(N)$ the maximum profit function when the number of workers is fixed to N and $\pi(N_a)$ when the number of apprentices is fixed to N_a:

$$\pi(N) = \max_{l_a \geq 0} pf\left(N - \frac{t_a}{\zeta_a} l_a, l_a\right) - wN - \frac{w_a + t_a w}{\zeta_a} l_a,$$
$$\pi(N_a) = \max_{l \geq 0} pf(l, N_a) - wl - (w_a + t_a w)N_a.$$

To simplify notation, define $\tilde{w}_a = \frac{w_a + t_a w}{\zeta_a}$ and $\tilde{t}_a = t_a / \zeta_a$. We also use the subindex notation of partial derivatives to economize on writing, $f_x := \frac{\partial f}{\partial x}$.

First, we show that $\pi(N)$ and $\pi(N_a)$ are concave. Let us start with $\pi(N_a)$. Using the envelope theorem, $\frac{\partial \pi(N_a)}{\partial N_a} = pf_{l_a} \zeta_a - (w_a + t_a w)$. We can differentiate this expression again with respect to N_a to obtain

$$\frac{\partial^2 \pi(N_a)}{\partial N_a^2} = p\left(f_{l_a l} \frac{dl^r}{dN_a} + f_{l_a l_a} \zeta_a\right), \tag{7}$$

where l^r solves the FOC of the optimization problem with fixed N_a, $pf_l(l^r, N_a) = w$. Assumption 1 implies the existence and uniqueness of this solution l^r. Totally differentiating the FOC with respect to l^r and N_a implies $\frac{dl^r}{dN_a} = -\frac{f_{ll_a}}{f_{ll}}\zeta_a \geq 0$, as $f_{ll_a} \geq 0$ and $f_{ll} \leq 0$. Replacing this derivative in (7) yields

$$\frac{\partial^2 \pi(N_a)}{\partial N_a^2} = p\left(f_{l_a l}\left(-\frac{f_{ll_a}}{f_{ll}}\zeta_a\right) + f_{l_a l_a}\zeta_a\right) = \frac{p\zeta_a}{f_{ll}}\left(f_{ll}f_{l_a l_a} - f_{ll_a}^2\right) \leq 0,$$

since concavity of f in l, l_a implies $f_{l_a l_a} \leq 0$ and $(f_{ll}f_{l_a l_a} - f_{ll_a}^2) \geq 0$. Thus, $\pi(N_a)$ is concave in N_a.

Importantly, this function is maximized at (n^*, n_a^*). So if we choose n_a away from n_a^*, profits will decrease. In Case (i), if $n_a^* > \overline{n}_a^j$, whenever the firm stays in the same jth regulation bracket, it chooses the feasible number of apprentices that is closest to n_a^*. This means the upper bound is binding, $n_a^r = \overline{n}_a^j$. Moreover, since $\frac{dl^r}{dN_a} \geq 0$ we know $l^r < l^*$ (given $n_a^* > \overline{n}_a^j$). This implies $n^r = l^r + t_a n^r = l^r + t_a \overline{n}_a^j < n^*$.

Similarly, we can show that $\pi(N)$ is concave. In this case,

$$\frac{\partial^2 \pi(N)}{\partial N^2} = pf_{ll}\left(1 - \tilde{t}_a \frac{dl_a^r}{dN}\right) + pf_{ll_a}\frac{dl_a^r}{dN}.$$

Considering the FOC and totally differentiating,

$$\frac{dl_a}{dN} = -\frac{f_{l_a l} - f_{ll}\tilde{t}_a}{f_{ll}\tilde{t}_a^2 - 2f_{ll_a}\tilde{t}_a + f_{l_a l_a}} \geq 0.$$

Substituting in the previous equation,

$$\frac{\partial^2 \pi(N)}{\partial N^2} = p\frac{\left(f_{ll}f_{l_a l_a} - f_{ll_a}^2\right)}{f_{ll}\tilde{t}_a^2 - 2f_{ll_a}\tilde{t}_a + f_{l_a l_a}} \leq 0$$

The last inequality stems again from the concavity of f, as $f_{ll}f_{l_a l_a} - f_{ll_a}^2 > 0$, $f_{ll} < 0$, and $f_{l_a l_a} < 0$.

This proves $\pi(N)$ is concave. Using a similar argument, the firm wants to get as close as possible to the optimal labor demands (n^*, n_a^*). However, now we also have to compare subsequent thresholds N_k for $k \geq j+1$, as $n_a^r(z, t_a)$ might still be larger than $\overline{n}_a^{j+1}$, so a firm might want to jump multiple thresholds to get a higher number of apprentices. In all of these cases, the firm chooses the number of workers at the threshold N_k as it is the closest to the optimal number of workers that allows the firm to get $n_a^r \in [\overline{n}_a^{k-1}, \overline{n}_a^k]$. The optimal number of apprentices in this case is $n_a^r(z, t_a) > \overline{n}_a^j$.

Case (ii) is immediate as the unconstrained optimum is within the regulation bounds, so the firm does not change its optimal decision.

For Case (iii), the proof is analogous to Case (i) whenever firms bunch just below a threshold or choose apprentices at the minimum quota. It remains to show that for relatively low $\phi_a > 0$, some firms prefer to pay the fee instead of hiring the minimum number of required apprentices.

To see this, suppose $n_a^* < \underline{n}_a^j$ and define $\pi^*(\phi_a) := pf(l^*, l_a^*) - wl^* - \tilde{w}_a l_a^* - \phi_a(\underline{n}_a^j - n_a^*)$, the profits when hiring labor as optimal without regulation, where $\tilde{w}_a := \frac{w_a + t_a w}{\zeta_a}$. Note that the optimal choice of workers and apprentices when paying the fee has to yield larger or equal profits,

$$\pi^j(\phi_a) := \max_{l_a, l \geq 0} pf(l, l_a) - wl - \tilde{w}_a l_a - \phi_a \left(\underline{n}_a^j - \frac{l_a}{\zeta_a} \right) \geq \pi^*(\phi_a).$$

Now, we know $\pi^* := pf(l^*, l_a^*) - wl^* - \tilde{w}_a l_a^* \geq \pi(N)$ and $\pi^* \geq \pi(N_a)$, for any $N, N_a \geq 0$. Also, $\pi^j(\phi_a)$ is continuous in ϕ_a, and $\lim_{\phi_a \to 0} \pi^*(\phi_a) = \pi^*$. Hence, there exists a small enough $\tilde{\phi}_a > 0$, such that $\pi^j(\tilde{\phi}_a) \geq \pi(N)$ and $\pi^j(\tilde{\phi}_a) \geq \pi(N_a)$. *Q.E.D.*

LEMMA 4: *Suppose Assumption* 1 *holds, except f does not necessarily have constant returns to scale (CRS) in (l, l_a, z). For each firm z, there exists $A(z) > 0$ such that $l_a^* = A(z) l^*$:*

i. *If $A'(z) > 0$, the parametric mapping $(n^*(z), n_a^*(z))$ is strictly convex.*
ii. *If $A'(z) = 0$, the parametric mapping $(n^*(z), n_a^*(z))$ is linear.*
iii. *If $A'(z) < 0$, the parametric mapping $(n^*(z), n_a^*(z))$ is strictly concave.*

PROOF: Take any firm $z > 0$. First, let us show that $l_a^* = A(z) l^*$. Since f is homogenous of degree γ, then $\frac{\partial f}{\partial l}$ and $\frac{\partial f}{\partial l_a}$ are homogenous of degree $\gamma - 1$. Thus, for any constant $k > 0$,

$$\frac{\dfrac{\partial f}{\partial l}(kl, kl_a; z)}{\dfrac{\partial f}{\partial l_a}(kl, kl_a; z)} = \frac{k^{\gamma-1} \dfrac{\partial f}{\partial l}(l, l_a; z)}{k^{\gamma-1} \dfrac{\partial f}{\partial l_a}(l, l_a; z)} = \frac{\dfrac{\partial f}{\partial l}(l, l_a; z)}{\dfrac{\partial f}{\partial l_a}(l, l_a; z)}.$$

So the derivatives of the isoquants are constant along any ray starting from the origin. Since $\gamma \in (0, 1)$ implies the production function is quasiconcave and the Inada condition holds for workers, there is only one point (l^*, l_a^*) such that $-\frac{\frac{\partial f}{\partial l}(l^*, l_a^*)}{\frac{\partial f}{\partial l_a}(l^*, l_a^*)} = -\frac{w\zeta_a}{w_a + t_a w}$. Together this implies l_a / l is constant whenever the derivative of the isoquant is constant. Hence, $\frac{l_a^*}{l^*} = A(z)$ for some $A(z) > 0$.

Now note that since $l = n - t_a n_a$ and $l_a = \zeta_a n_a$, $\frac{n_a^*}{n^*} = \frac{1}{\zeta_a A(z)^{-1} + t_a}$. Call this last term $B(z)$. Hence, $A'(z) > 0 \iff B'(z) > 0$.

From the equation above, $\frac{dn_a^*}{dn^*} = B(z)$, $\forall z$. Thus, if $B(z)$ is increasing in z, then $\frac{dn_a^*}{dn^*}$ is increasing in z. From Lemma (1), $\frac{dn^*}{dz} > 0$, which implies that the parametric mapping $n_a^*(n^*)$ is convex. Similarly, if $A'(z) = 0 \Rightarrow B'(z) = 0$ and so the derivative is constant for any z, $\frac{dn_a^*}{dn^*} = B \in \mathbb{R}_+$, $\forall z$. This means the parametric mapping is linear. Finally, if $A'(z) < 0 \Rightarrow B'(z) < 0$, then $\frac{dn_a^*}{dn^*}$ is decreasing in z and hence $n_a^*(n^*)$ is concave. *Q.E.D.*

COROLLARY 1: *Under Assumption* 1, *$n_a^* = Bn^*$, where B does not depend on z.*

PROOF: Assuming CRS of f on (l, l_a, z) implies $\frac{(\partial f / \partial l)}{(\partial f / \partial l_a)}$ is homogenous of degree 0 in z. So Case (ii) of Lemma 4 applies $l_a^* = Al^*$ for some $A \geq 0$ independent of z. Using the same argument in Lemma 4, $n_a^* = Bn^*$ for $B = \frac{1}{\zeta_a A^{-1} + t_a}$. *Q.E.D.*

PROPOSITION 1: *Suppose Assumption* 1 *holds and firms solve the maximization problem with regulation* (3). *Then* $\forall z \geq 0$,

Case 1: *there exist $(\overline{\frac{w_a}{w}}, \overline{t}_a)$ such that for $\frac{w_a}{w} \leq \overline{\frac{w_a}{w}}$ and $t_a \leq \overline{t}_a$,*

i. *the number of apprentices without regulation is* $n_a^* = B_u n^*$ *and is above the maximum quota,* $n_a^*(z, t_a) > \overline{n}_a^j$.
ii. *there exist cutoffs* $\{z_b^j, z_r^j\}$ *such that firms* $z \in [z_b^j, z_r^j]$ *increase their size to threshold* N_k *with* $k \geq j$.
iii. *firms choose maximum number of apprentices* $n_a^r = \overline{n}_a^j$.
iv. *firms never pay the fee.*

Case 2: there exist $(\underline{\frac{w_a}{w}}, \underline{t}_a)$ *such that for* $\frac{w_a}{w} \geq \underline{\frac{w_a}{w}}$ *or* $t_a \geq \underline{t}_a$,

i. *the number of apprentices without regulation is* $n_a^* = B_s n^*$ *and is below the minimum quota,* $n_a^*(z, t_a) < \underline{n}_a^j$.
ii. *there exist cutoffs* $\{z_b^j, z_r^j\}$ *such that firms* $z \in [z_b^j, z_r^j]$ *reduce their size* ϵ *below threshold* N_k *with* $k < j$.
iii. *firms that increase the number of apprentices choose the minimum number* $\underline{n}_a^j$.
iv. *there exists* $\overline{\phi}_a > 0$ *such that for* $\phi_a \leq \overline{\phi}_a$, *there is an additional cutoffs* z_f^j *where firms* $z \in (z_r^j, z_f^j]$ *choose to pay the fee.*

PROOF: Let $z \geq 0$. Assumption 1 and Corollary 1 imply that the solution without regulation is $n_a^* = Bn^*$ for some $B \geq 0$ that does not depend on z. Lemma 2 implies that B is a continuous nonincreasing function of $\frac{w_a}{w}$ and t_a.

Let us start with Case 1. When $\frac{w_a}{w}$ and t_a approach 0, the optimal relative number of apprentices n_a^*/n^* is unbounded. This means below some threshold $(\overline{\frac{w_a}{w}}, \overline{t}_a)$, $n_a^*(z, t_a) > \overline{n}_a^j$ with $n^*(z, t_a) \in [N_{j-1}, N_j)$ (part i). Now, for any $t_a < \overline{t}_a$, $\frac{w_a}{w} \leq \overline{\frac{w_a}{w}}$, Lemma 1 implies n_a^* and n^* are increasing and continuous in z. After some threshold z_b^j, by Lemma 3, $(N_k, n_a(N_k))$ for some $k \geq j$ is closer to the optimal labor input (n_a^*, n^*), and hence firm (z, t_a) with $z \geq z_b^j$ will bunch at this threshold N_k. Let z_r^j be productivity such that N_k is the optimal number of workers without regulation for firm (z_r^j, t_a). Then firms beyond $z > z_r^j$ do not increase their size relative to the equilibrium with no regulation, completing the proof of part (ii). Since $n_a^* > \overline{n}_a^j$ by Lemma 3, firms choose $n_a^r = \overline{n}_a^j$ (part iii). And finally, firms never pay the fee as the number of desired apprentices is above the maximum quota (part iv).

The proof of Case 2 is analogous. When either $\frac{w_a}{w}$ or t_a tend to ∞, the optimal relative number of apprentices n_a^*/n^* converges to zero. Note that for $\phi \to 0$, we can use the same argument as in the proof of Lemma 3 in order to prove that firms prefer to pay the fee instead of hiring the minimum quota of apprentices. *Q.E.D.*

Negative Marginal Productivity of Apprentices

In the following, we show that the production function has to allow for apprentices having negative marginal productivity in firms that choose to pay the fee. Consider a standard production function $\tilde{f}(n, n_a; z)$ combining managerial ability z with labor input from workers and apprentices. Let us compare two scenarios based on the components of the regulation. First, suppose that firms are required to train at least $\underline{n}_a$ apprentices paying a wage of $w_a^{\min}$. Alternatively, firms can pay a fee ϕ_a per required apprentice.

When a firm z chooses to train the apprentices, it solves

$$\pi_a(z) := \max_{n, n_a} p\tilde{f}(n, n_a; z) - wn - w_a^{\min} n_a, \quad n_a \geq \underline{n}_a.$$

Instead, the firm could pay the fee and solve

$$\pi_f(z) := \max_{n} p\tilde{f}(n, 0; z) - wn - \phi_a \underline{n}_a.$$

In Proposition 2, we show that if $\phi_a > w_a^{\min}$, then firms choose to pay the fee only if the marginal productivity of apprentices is negative.

PROPOSITION 2: *If* $\phi_a > w_a^{\min}$, *then* $\pi_f(z) > \pi_a(z) \Rightarrow \frac{\partial \tilde{f}}{\partial n_a} < 0$.

PROOF: Let $z \geq 0$ and n_0^* denote the number of workers maximizing $\pi_f(z)$. By way of contradiction, suppose $\frac{\partial \tilde{f}}{\partial n_a} \geq 0$. In this case,

$$\begin{aligned}\pi_a(z) &\geq p\tilde{f}\big(n_0^*, \underline{n}_a; z\big) - wn_0^* - w_a\underline{n}_a \geq p\tilde{f}\big(n_0^*, 0; z\big) - wn_0^* - w_a\underline{n}_a \\ &> p\tilde{f}\big(n_0^*, 0; z\big) - wn_0^* - \phi_a\underline{n}_a = \pi_f(z),\end{aligned}$$

where the first two inequalities are implied by $\pi_a(z)$ being the maximum profit function and $\frac{\partial \tilde{f}}{\partial n_a} \geq 0$, and the last inequality is due to $\phi_a > w_a$. This contradicts $\pi_f(z) > \pi_a(z)$. Therefore, $\pi_f > \pi_a \Rightarrow \frac{\partial \tilde{f}}{\partial n_a} < 0$. *Q.E.D.*

In other words, if apprentices have positive productivity and it is cheaper to hire an apprentice than paying the fee, firms choose to hire the apprentice. In the model, we allow for negative marginal revenue product of apprentices simply by adding training costs.

D.2. *Additional Inputs*

In this section, we describe an extension of the model adding other inputs. We discuss a simple example to illustrate the results based on the baseline model.

Suppose there is an additional input x, with price w_x, that firms choose in each period. First, we consider the firm problem without regulation. Consider a simple Cobb–Douglas specification

$$\max_{n,n_a,x} pz^{1-\gamma}(n - t_an_a + \zeta_an_a)^{\gamma_l}x^{\gamma_x} - wn - w_an_a - w_xx \quad \text{s.t. } t_an_a \leq n$$

where γ_l is the output elasticity of labor, and γ_x the output elasticity of input x. Suppose the production function has constant returns to scale on (z, n, n_a, x), and thus $\gamma_l + \gamma_x = \gamma$. As in the baseline model, linearity in labor input implies that there are corner solutions.

A firm (z, t_a) avoids apprentices whenever $w < \frac{t_aw+w_a}{\zeta_a}$. In that case, from the FOC, $x = \frac{w}{w_x}\frac{\gamma_x}{\gamma_l}n =: \mathcal{X}n$. So the optimal input demands are $n_a^* = 0$, $n^* = (\frac{p\gamma_l\mathcal{X}^{\gamma_x}}{w})^{\frac{1}{1-\gamma}}z$, $x^* = \mathcal{X}n^*$. The corresponding output and profits are

$$y^* = \left(\frac{p\gamma_l}{w}\right)^{\gamma/(1-\gamma)}\mathcal{X}^{\frac{\gamma_x}{1-\gamma}}z, \qquad \pi^* = p^{\frac{1}{1-\gamma}}\left(\frac{\gamma_l}{w}\right)^{\gamma/(1-\gamma)}\mathcal{X}^{\frac{\gamma_x}{1-\gamma}}(1-\gamma)z.$$

On the other hand, if $w > \frac{t_aw+w_a}{\zeta_a}$ the firm seeks apprentices, so $x = \frac{w_a+t_aw}{w_x}\frac{\gamma_x}{\gamma_l}n_a =: \mathcal{X}_an_a$. The optimal input demands are $n_a^* = (\frac{p\gamma_l\zeta_a^{\gamma_l}\mathcal{X}_a^{\gamma_x}}{w_a+t_aw})^{\frac{1}{1-\gamma}}z$, $n^* = t_an_a^*$, $x^* = \mathcal{X}_an_a^*$, and output and profits are

$$y^* = \left(\frac{p\gamma_l}{w_a + t_aw}\right)^{\gamma/(1-\gamma)}\zeta_a^{\frac{\gamma_l}{1-\gamma}}\mathcal{X}_a^{\frac{\gamma_x}{1-\gamma}}z, \qquad \pi^* = p^{\frac{1}{1-\gamma}}\left(\frac{\gamma_l}{w_a + t_aw}\right)^{\gamma/(1-\gamma)}\zeta_a^{\frac{\gamma_l}{1-\gamma}}\mathcal{X}_a^{\frac{\gamma_x}{1-\gamma}}(1-\gamma)z.$$

Now, let us study the case of a particular threshold with regulation. Firms have the options of bunching at the threshold N, complying with the apprenticeship quotas by hiring the required number of apprentices $\underline{n}_a$, or paying the fee.

Suppose that a firm bunches at N to avoid training:

$$n^r = N, \qquad n_a^r = 0, \qquad x^r = \left(\frac{\gamma_x N^{\gamma_l}}{w_x}\right)^{1/(1-\gamma_x)} z^{\frac{1-\gamma}{1-\gamma_x}},$$

$$y^r = \left(\frac{p\gamma_x}{w_x}\right)^{\frac{\gamma_x}{1-\gamma_x}} N^{\frac{\gamma-\gamma_x}{1-\gamma_x}} z^{\frac{1-\gamma}{1-\gamma_x}}, \qquad \pi^r = p^{\frac{1}{1-\gamma_x}} \left(\frac{\gamma_x}{w_x}\right)^{\frac{\gamma_x}{1-\gamma_x}} (1-\gamma_x) N^{\frac{\gamma-\gamma_x}{1-\gamma_x}} z^{\frac{1-\gamma}{1-\gamma_x}},$$

If the firm has to take n_a apprentices instead, the same analysis applies:

$$n^r = \left(\frac{p\gamma_l \mathcal{X}^{\gamma_x}}{w}\right)^{\frac{1}{1-\gamma}} z - (\zeta_a - t_a) n_a, \qquad x^r = \mathcal{X} n^r.$$

Suppose the firm pays the fee:

$$n_a^r = 0, \qquad n^r = \left(\frac{p\gamma_l \mathcal{X}^{\gamma_x}}{w}\right)^{\frac{1}{1-\gamma}} z, \qquad x^r = \mathcal{X} n^r.$$

The corresponding output and profits are

$$y^r = \left(\frac{p\gamma_l}{w}\right)^{\gamma/(1-\gamma)} \mathcal{X}^{\frac{\gamma_x}{1-\gamma}} z, \qquad \pi^r = p^{\frac{1}{1-\gamma_x}} \left(\frac{\gamma_l}{w}\right)^{\gamma/(1-\gamma)} \mathcal{X}^{\frac{\gamma_x}{1-\gamma}} (1-\gamma) z - \phi_a \underline{n}_a.$$

From the equations above, we can see that the effect of adding other inputs is that firms have additional margins of substitution. Qualitatively, there are no differences to the baseline model used throughout the paper. However, this quantitatively affects the magnitude of firm responses. In terms of the estimation, the fit of the firm size distribution would be similar, but we would need information on the share of these other inputs in production to identify the parameters of the production function. Larger γ_x implies firms would respond less to the regulation as the output elasticity with respect to labor decreases.

D.3. *Multiple Types of Workers*

In this section, we describe an extension of the model including multiple types of workers. For clarity of exposition, let us suppose there are two types of workers, unskilled u and skilled s. We characterize the equilibrium in the linear labor input case, combining these types of workers in a Cobb–Douglas function.

Suppose firms are characterized by managerial ability z and training costs for each type of worker t_a^i where $i \in \{u, s\}$. First, we study the case without regulation. A firm (z, t_a^u, t_a^s) solves

$$\max_{n^i, n_a^i} p z^{1-\gamma} \left(n^u + \left(\zeta_a^u - t_a^u\right) n_a^u\right)^{\gamma_u} \left(n^s + \left(\zeta_a^s - t_a^s\right) n_a^s\right)^{\gamma_s} - \sum_{i=u}^{s} \left(w^i n^i + w_a n_a^i\right) \quad \text{s.t. } t_a^i n_a^i \le n^i, \forall i,$$

where $\sum_i \gamma_i = \gamma$, and $\zeta_a^i \in [0, 1]$ denote the apprentices' productivity in each occupation.

As before, there are corner solutions. Firms avoid apprentices of type i if $w^i < \frac{w^i t_a^i + w_a}{\zeta_a^i}$: $n_a^i = 0$ and $n^i = (\frac{p\gamma_i A^{\gamma_j}}{w^i})^{1/(1-\gamma)} z$, where $A = \frac{w^i}{w^j}\frac{\gamma_j}{\gamma_i}$ and $i \neq j$. Firms seek apprentices of type i if $w^i > \frac{w^i t_a^i + w_a}{\zeta_a^i}$: $n_a^i = (\frac{p\gamma_i (\zeta_a^i)^{\gamma_i} A^{\gamma_j}}{w^i t_a^i + w_a})^{1/(1-\gamma)} z$ and $n^i = t_a^i n_a^i$.

Now let us consider the firm decision with regulation. Suppose the firm has to train n_a apprentices due to quotas. First, we show that firms generically choose to train apprentices only in one occupation (by only one type of worker), depending on which one is relatively cheaper.

LEMMA 5: *A firm chooses to train apprentices only in occupation* $i^* = \arg\max_i w^i(\zeta_a^i - t_a^i)$.

PROOF: The net benefit of training an apprentice in occupation i is $\zeta_a^i w^i - (w^i t_a^i + w_a) = w^i(\zeta_a^i - t_a^i) - w_a$. Hence, firms choose to train apprentices only in occupation $i^* = \arg\max_i w^i(\zeta_a^i - t_a^i)$. *Q.E.D.*

Lemma 5 implies that we only have to compare the corner solutions to the choice of apprentices. Suppose the firm optimally chooses to train apprentices in occupation i. Let $x_i^* = p((\frac{\gamma_i}{w^i})^{1-\gamma_j}(\frac{\gamma_j}{w^j})^{\gamma_j})^{1/(1-\gamma)} z$, then $n_i^r = x_i^* - (\zeta_a^i - t_a^i) n_a^i$. Using Lemma 5, it is sufficient to consider the case when all apprentices are trained in occupation i, that is, $n_a^i > 0$ and $n_a^j = 0\ \forall j \neq i$.

D.4. *Dynamic Frictions*

In this section, we consider a two-period version of the model including dynamic frictions. In period t, as in the baseline model, firms hire workers to produce and to train apprentices. In period $t+1$, firms can use both workers and previously trained apprentices in production.

With competitive labor markets, trained apprentices move freely after training, yielding the same results as in the baseline model. If the cost of training apprentices in period t is larger than the cost of hiring workers ($w_t < \frac{w_t t_a + w_a}{\zeta_a}$), firms choose not to train any apprentices; while in the reverse case $w_t > \frac{w_t t_a + w_a}{\zeta_a}$, they train as many apprentices as possible, $n_{a,t}^* = (\frac{p_t \gamma \zeta_a^{\gamma}}{w_t t_a + w_a})^{1/(1-\gamma)} z$, $n_t^* = t_a n_{a,t}^*$.

Now consider labor market frictions that prevent a fraction $\rho > 0$ of apprentices from moving across firms after training. Firms can retain these apprentices at a discounted wage rate w_a^r below their marginal product. In other words, there is *wage compression*. Firms choose n_t workers and $n_{a,t}$ apprentices in period t, and n_a^r apprentices to *continue* in period $t+1$ as well as n_{t+1} workers to solve

$$\begin{aligned} \max_{n_t, n_{a,t}, n_a^r, n_{t+1}} \quad & p_t f(n_t - t_a n_{a,t}, \zeta_a n_{a,t}; z) + \beta p_{t+1} f\big(\zeta_a^r n_a^r + n_{t+1}; z\big) - w_t n_t - w_a n_{a,t} \\ & - \beta w_{t+1} n_{t+1} - \beta w_a^r n_a^r \\ \text{s.t. } & n_t \geq t_a n_{a,t},\ n_{a,t} \geq 0,\ n_a^r \geq 0,\ n_{t+1} \geq 0,\ \rho n_{a,t} \geq n_a^r, \end{aligned}$$

where ζ_a denotes the productivity of apprentices in period t, ζ_a^r the productivity of retained apprentices in $t+1$ and $\beta \in [0, 1]$ the intertemporal discount rate.

We solve the model as in the baseline case, that is, $f(n_t - t_a n_{at}, \zeta_a; z) = z^{1-\gamma}(n_t - t_a n_{at} + \zeta_a n_{at})^{\gamma}$ and $f(\zeta_a^r n_a^r + n_{t+1}; z) = (\zeta_a^r n_a^r + n_{t+1})^{\gamma}$, but consider the additional restrictions. Again, under linear labor inputs, there are corner solutions. In this case, the FOCs of the firm problem imply that with wage compression $w_a^r < \zeta_a^r w_{t+1}$, firms train apprentices if and only if

$$\underbrace{w_t}_{\text{Cost of hiring workers in } t} \geq \underbrace{\frac{w_a + t_a w_t}{\zeta_a}}_{\text{Cost of training apprentices in } t} \underbrace{- \frac{\rho\beta}{\zeta_a}\left(\zeta_a^r w_{t+1} - w_a^r\right)}_{t+1 \text{ benefit}}. \tag{8}$$

Note that equation (8) generalizes the inequality from the frictionless case by including the additional benefits from training. On the other hand, if there is no wage compression such that $w_a^r \geq \zeta_a^r w_{t+1}$, the condition for firms to train apprentices is the same as in the baseline case. Moreover, if $t_a \geq \rho^{1-\gamma}\zeta_a^\gamma(\zeta_a^r)^{1-\gamma}\frac{p_t}{p_{t+1}}\frac{w_a^r}{w_t} - \frac{w_a}{w_t}$, firms also adjust the intensive margin, training more apprentices than in the frictionless case: $n_{a,t}^f = (\frac{p_t \gamma \zeta_a^\gamma}{w_a + t_a w_t - \beta\rho(\zeta_a^r w_{t+1} - w_a^r)})^{\frac{1}{1-\gamma}} z > n_{a,t}^*$. Lemma 6 compiles these results, showing that frictions together with wage compression increase firms' willingness to train apprentices in the spirit of Acemoglu and Pischke (1998, 1999).

LEMMA 6: *With frictions and wage compression, that is, $\rho > 0$ and $\frac{w_a^r}{\zeta_a} > w_{t+1}$, the number of trained apprentices increases relative to the baseline model.*

For details of the calibration of this model extension, see Appendix E.8.

D.5. *Hiring Costs*

In this section, we describe an extension of the model that allows for fixed and variable hiring costs $\kappa_0^h, \kappa_1^h \geq 0$. We suppose that firms incur these costs if they increase their size n relative to their pre-regulation size n^*.

Formally, firm (z, t_a) solves

$$\max_{n, n_a, d_f} pf(n - t_a n_a, \zeta_a n_a; z) - wn - w_a n_a - d_f \mathcal{F}_a(n, n_a)$$

$$- \underbrace{\left(\kappa_0^h + \kappa_1^h n\right)\mathbb{1}\left(n > n^*\right)}_{\text{Hiring Costs}} \quad \text{s.t. } t_a n_a \leq n,$$

$$(n, n_a, d_f) \in \bigcup_j [N_{j-1}, N_j) \times \left[\underline{n}_a^j, \overline{n}_a^j\right] \times \{0\} \quad \text{or}$$

$$(n, n_a, d_f) \in \bigcup_j [N_{j-1}, N_j) \times \left[0, \underline{n}_a^j\right] \times \{1\}, \qquad \mathcal{F}_a(n, n_a) = \phi_a\left(\underline{n}_a^j - n_a\right)^+,$$

where $n^*(z, t_a)$ is the solution to the problem without regulation. We solve this model numerically, following an analogous procedure to in the baseline case. See Appendix E.8 for details of the calibration.

APPENDIX E: DETAILS OF THE QUANTITATIVE EXERCISES

E.1. *Parametric Fit of the Firm Size Distribution*

Figure E1 shows the fit of various parametric distributions to prereform data. The Generalized Extreme Value distribution provides the best fit out of two- and three-parameter distributions commonly used to model productivity.

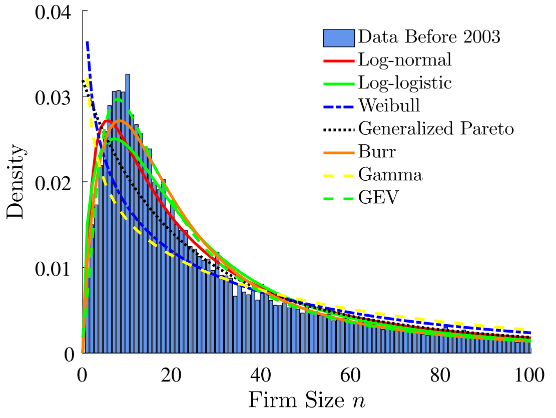


FIGURE E1.—Fitting the Prereform Firm Size Distribution. *Notes*: The figure shows the empirical firm size density pre-reform (1995 to 2002), as well as various parametric distributions fitted via maximum likelihood.

E.2. *Production Function Estimation*

Table EI shows estimated labor shares γ^k using six different methodologies: our baseline regression with time and firm fixed effects (FE), a simple OLS regression (OLS), as well as the methods proposed by Olley and Pakes (1996) (OP), Levinsohn and Petrin (2003) (LP), Wooldridge (2009) (W), and Ackerberg, Caves, and Frazer (2015) (LP-ACF). For the last four specifications, we suppose the production function depends on capital K, full-time labor l and other intermediate inputs m, which we approximate empirically using energy, water and intermediate product expenditures.

E.3. *Moment Weights and Robustness*

Table EII details the weights on each group of moments in the estimation. For the bunching and missing mass moments, we use weights corresponding to the observed prereform fraction of firms at each bunching or missing mass point. For instance, for the first bunching point among high-skill sector firms we weight the bunching mass at 14 by the fraction of high-skill sector firms of size 14 in the prereform data, h^s_{14}. Additionally, we divide the missing mass moments by five (the width of the potential missing mass window we consider), in order to make the total weight on missing mass moments comparable to the weight on bunching moments. We use the same procedure for weighting the average number of apprentices by firm size and the fraction of firms paying the fee by firm size, using as weights the prereform fraction of firms h_j. Finally, we equally weight those four

TABLE EI

ESTIMATED LABOR SHARES.

	(1) FE (Baseline)	(2) OLS	(3) OP	(4) LP	(5) W	(6) LP-ACF
High-Skill	0.61	0.30	0.57	0.34	0.31	0.42
Low-Skill	0.58	0.34	0.43	0.23	0.20	0.28

Note: The table shows labor shares obtained from different estimation methods. Columns (3) to (6) are estimated using the Stata program *prodest* by Mollisi and Rovigatti (2017).

TABLE EII

BASELINE MOMENT WEIGHTS.

Weight	Moment Description
$\omega_j^k = \frac{1}{4}\frac{1}{2}h_{b(j)}$	Bunching mass points
$\omega_j^k = \frac{1}{4}\frac{1}{2}\frac{1}{5}h_{m(j)}$	Missing mass points
$\omega_j^k = \frac{1}{4}h_j$	Average number of apprentices by firm size
$\omega_j^k = \frac{1}{4}h_j$	Fraction of firms paying the fee by firm size
$\omega_j^k = \frac{1}{4}$	Fraction of firms choosing maximum number of apprentices prereform

Note: The table shows the weights on moments used in the SMM estimation.

groups of moments. Thus, the fraction of firms that choosing the maximum number of apprentices before the reform receive weight $\omega_j^k = \frac{1}{4}$.

As a robustness check, instead of equally weighting the four groups of moments, we weight them by the inverse of their variance obtained from 1000 bootstrap samples. Table EIII shows that this procedure results in somewhat extreme weights as the variance of some moments is small. For instance, virtually no high-skill sector firms choose the maximum number of apprentices before the reform, resulting in a small variance and a large weight on this moment. Similarly, very few low-skill sector firms pay the fee, attracting a large weight.

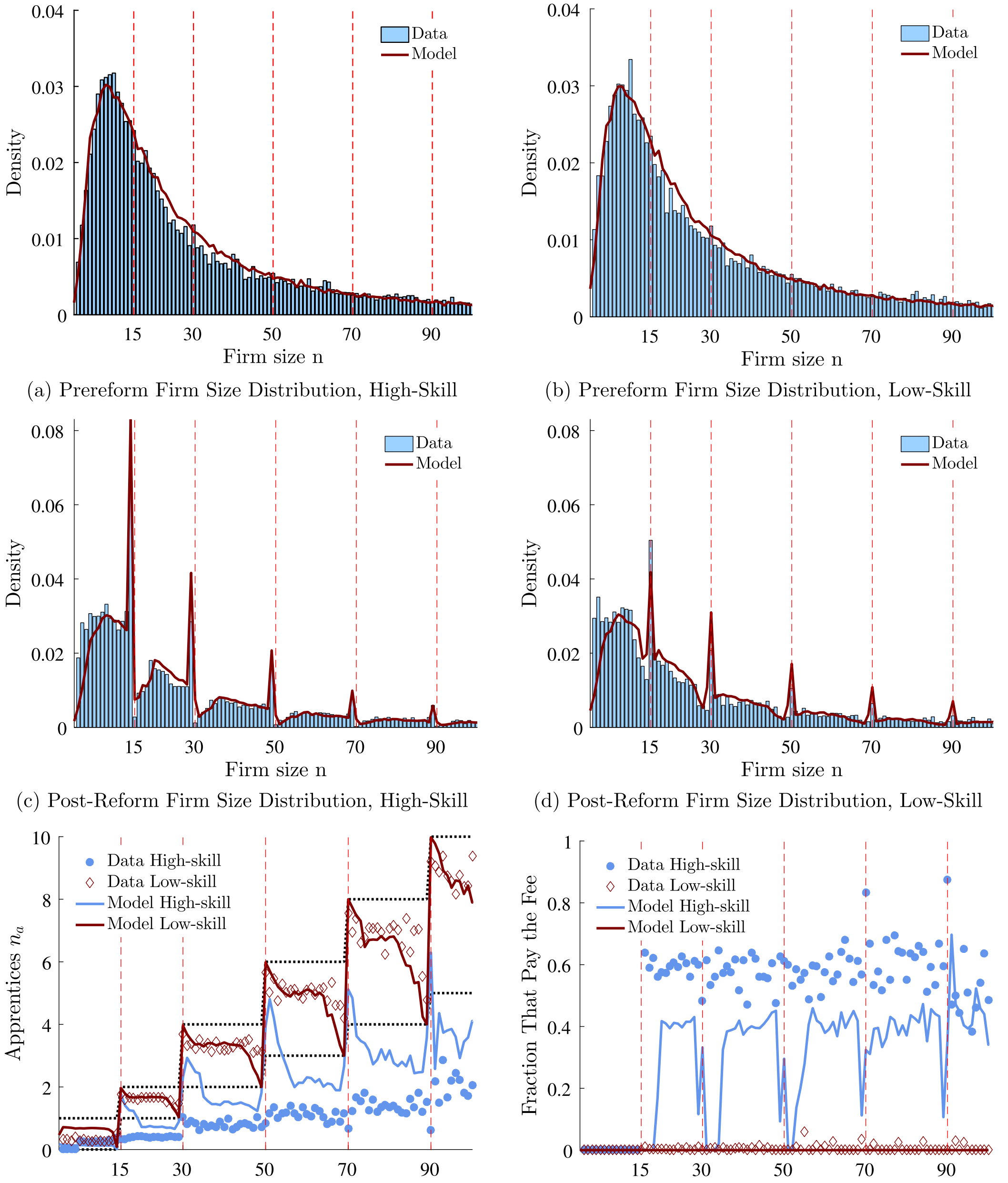
Figure E2 and Table EIV show that uniform weighting matches better the average number of apprentices (Fact 2) and the fraction paying the fee (Fact 3) for high-skill sectors. In particular, using inverse variance weights leads to somewhat overestimating the apprentice intake in high-skill sectors. Figure E3 shows that the inverse variance estimation results in a very similar estimated training cost distribution for low-skill sectors, but somewhat smaller training costs for high-skill sectors. Consistently, Table EV shows that quantitative results for low-skill sectors remain almost identical, but the inverse variance weighting over-estimates labor substitution in high-skill sectors.

TABLE EIII

INVERSE VARIANCE MOMENT WEIGHTS.

Moment Group	(1) High-skill	(2) Low-Skill
Fraction choosing maximum apprentices pre-reform	0.931	0.388
Bunching and missing mass points	0.065	0.100
Average number of apprentices by firm size	0.0001	0.0001
Fraction paying the fee by firm size	0.005	0.511

Note: The table shows weights on each group moments based on the inverse of their variance, obtained from 1000 bootstrap samples.



(a) Prereform Firm Size Distribution, High-Skill

(b) Prereform Firm Size Distribution, Low-Skill

(c) Post-Reform Firm Size Distribution, High-Skill

(d) Post-Reform Firm Size Distribution, Low-Skill

(e) Number of Apprentices

(f) Fraction Paying the Fee

FIGURE E2.—Alternative Moment Weights and Model Fit. *Notes*: The figure depicts the model fit to targeted moments using the inverse variance weights from Table EIII. Panels (a) and (b) show the distribution of firm size (number of full-time workers) for prereform (1995–2002) and panels (c) and (d) show the firm size distribution post-reform (2003–2009). Panel (e) shows the number of apprentices by firm size, and panel (f) shows the fraction of firms paying the fee by firm size, both in the post-reform period.

TABLE EIV

COMPARING FIT ACROSS ESTIMATION PROCEDURES.

	(1)	(2)	(3)	(4)	(5)
		Targeted Moments Error			*Total Score*
	Max. Apprentices Prereform	Bunching (Fact 1)	Apprentices (Fact 2)	Fee (Fact 3)	$\sum \omega \frac{\lvert\text{model}-\text{data}\rvert}{0.5\lvert\text{model}\rvert+0.5\lvert\text{data}\rvert}$
			A. Baseline		
High-Skill Sectors	0.000	0.137	0.105	0.060	0.302
Low-Skill Sectors	0.001	0.129	0.052	0.097	0.280
			B. Out-of-Sample Fit		
High-Skill Sectors	0.000	0.141	0.127	0.071	0.340
Low-Skill Sectors	0.001	0.132	0.051	0.057	0.240
			C. Inverse Variance Weights		
High-Skill Sectors	0.000	0.074	0.173	0.116	0.022
Low-Skill Sectors	0.001	0.129	0.052	0.097	0.037

Note: The table shows the model fit under the baseline estimation (panel A), the out-of-sample estimation (panel B), and the inverse variance weighting estimation (panel C). Columns (1) to (4) show the estimation error of the targeted moments $\frac{\lvert\text{model}-\text{data}\rvert}{0.5\lvert\text{model}\rvert+0.5\lvert\text{data}\rvert}$. Column (5) shows the total score function using uniform weights in panels A and B, and inverse variance weights from Table EIII in panel C.

TABLE EV

ROBUSTNESS OF QUANTITATIVE EFFECTS USING ALTERNATIVE MOMENT WEIGHTS.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Aggregate Outcomes			*Changes in Agents' Welfare $\Delta U_j/U^*$*			
	Workers	% Workers	Apprentices	% Output	Apprentices	Workers	Firms	Total
				A. Uniform Weights				
High-Skill Sectors	−1905	−0.89	4937	−0.34	0.27	−0.54	−0.46	−0.74
Low-Skill Sectors	−3519	−1.67	17,866	0.30	1.13	−0.97	0.13	0.30
Total	−5423	−1.28	22,803	−0.06	0.65	−0.73	−0.20	−0.28
				B. Inverse Variance Weights				
High-Skill Sectors	−2433	−1.14	7492	−0.37	0.41	−0.70	−0.35	−0.64
Low-Skill Sectors	−3493	−1.66	17,869	0.31	1.13	−0.96	0.13	0.31
Total	−5926	−1.40	25,361	−0.07	0.73	−0.81	−0.14	−0.22

Note: Columns (1) to (4) of the table show the effects of the apprenticeship regulation on aggregate outcomes, namely on the number of workers, the number of apprentices, and output. Columns (5) to (7) show the effect on the welfare of apprentices, workers, and firm owners. Column (8) shows the sum of welfare effects across the groups of agents from columns (5) to (7). Panel A shows results under the baseline uniform moment weights, and panel B shows results under the alternative inverse variance weights.

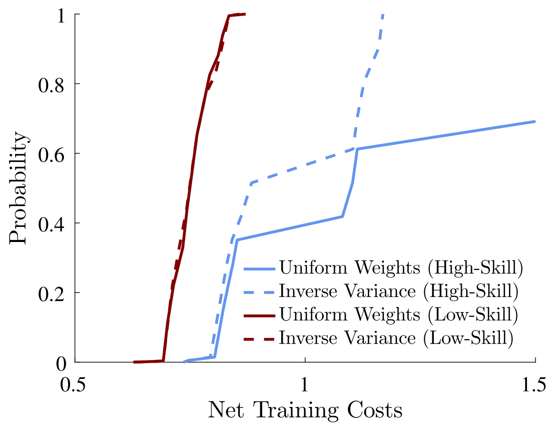


FIGURE E3.—Alternative Moment Weights and Training Costs. *Notes*: The figure shows estimated training cost distributions under the baseline weighting procedure (solid lines) and the alternative inverse variance weighting procedure (dashed lines). We normalize $\zeta_a = 1$ as in main text Figure 7.

E.4. *Out-of-Sample Fit*

In this section, we present results from an out-of-sample exercise. We randomly split the data set in half for all years and sectors, estimate the model using one-half and evaluate its fit using the other half. Table EIV shows the comparison of estimation errors of targeted moments as well as total score functions. Figure E4 depicts the out-of-sample model fit graphically.

E.5. *Truncated Normal and Uniform Training Cost Distributions*

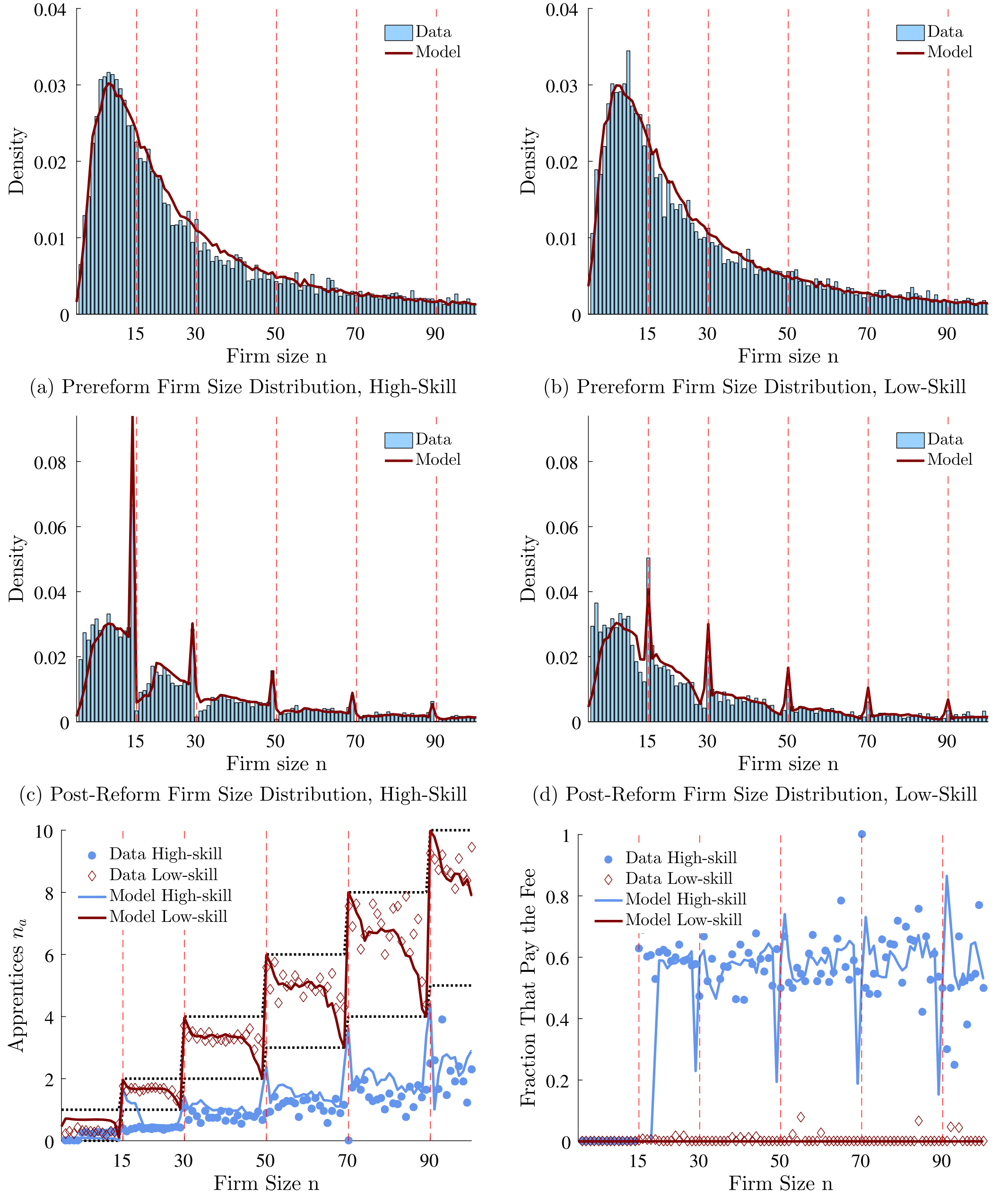
Figure E5 shows the fit of the estimated model assuming a truncated normal training cost distribution, and Figure E6 shows the fit assuming a uniform training cost distribution.

E.6. *Transferability of Training Skills*

Figure E7 shows sensitivity analysis with respect to the skills transferability parameter χ_a^k. In particular, we show how moving from $\chi_a = 0$ to $\chi_a = 1$ changes the effects on effective units of labor, output, profits, and total welfare.

E.7. *Effect of the Policy Decomposition on Aggregate Variables*

Here, we show additional details of the policy decomposition described in Section 6.1 of the paper. Table EVI shows the effects of the different components of the regulation on aggregate outcomes. We consider four scenarios, (i) only apprentice quotas, (ii) only a reduction in apprentice minimum wages, (iii) combining quotas and the lower minimum wage, and (iv) the "full" regulation adding the possibility of paying the fee. We show that with only the minimum wage reduction, firms with low training costs substitute many of their workers for apprentices. Adding the quotas attenuate the displacement of workers by establishing a maximum on the number of apprentices. The minimum quota, on the other hand, mandates firms to train, which increases training in high-skill sectors. Finally, the possibility of paying the fee dampens the negative effects for those firms with very high training costs.



(a) Prereform Firm Size Distribution, High-Skill

(b) Prereform Firm Size Distribution, Low-Skill

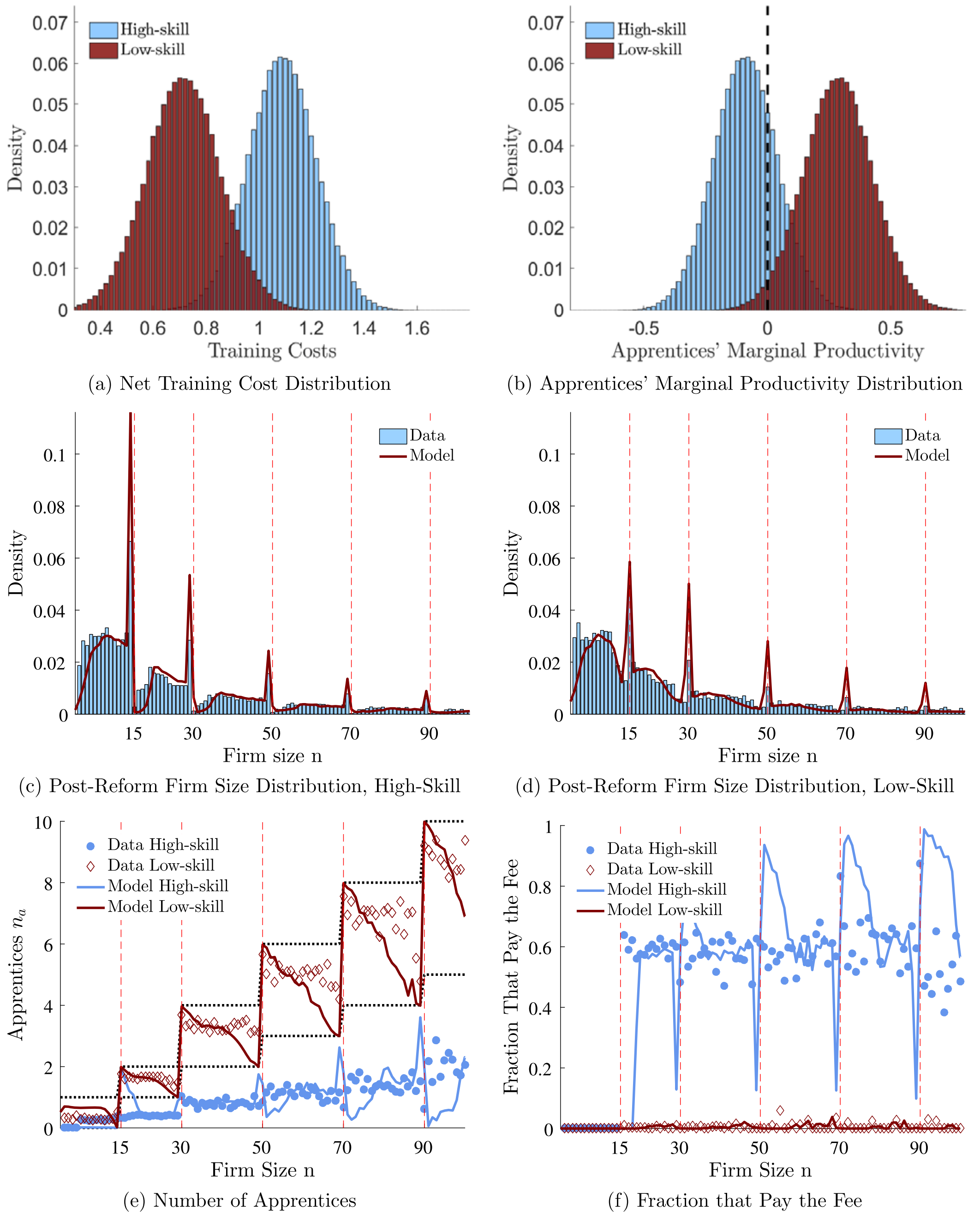
(c) Post-Reform Firm Size Distribution, High-Skill

(d) Post-Reform Firm Size Distribution, Low-Skill

(e) Number of Apprentices

(f) Fraction Paying the Fee

FIGURE E4.—Out-of-Sample Fit. *Notes*: The figure depicts the fit of the model estimated on half of the data to the other, untargeted half of the data. Panels (a) and (b) show the distribution of firm size (number of full-time workers) for pre-reform (1995–2002) and panels (c) and (d) show the firm size distribution post-reform (2003–2009). Panel (e) shows the number of apprentices by firm size, and panel (f) shows the fraction of firms paying the fee by firm size, both in the post-reform period.



(a) Net Training Cost Distribution

(b) Apprentices' Marginal Productivity Distribution

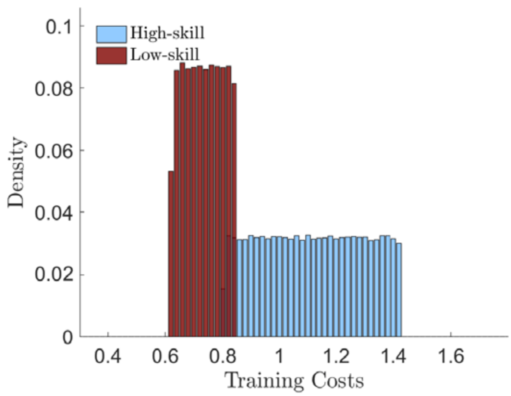
(c) Post-Reform Firm Size Distribution, High-Skill

(d) Post-Reform Firm Size Distribution, Low-Skill

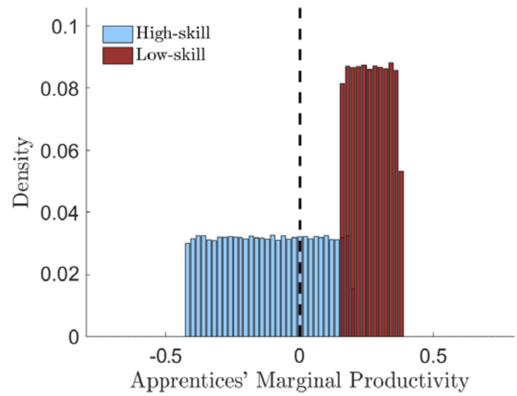
(e) Number of Apprentices

(f) Fraction that Pay the Fee

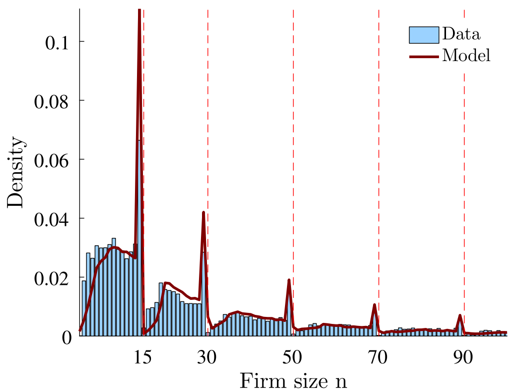
FIGURE E5.—Model Fit Under Truncated Normal Training Costs. *Notes*: The figure depicts the fit of the model under a truncated normal training cost distribution. Panels (a) and (b) show the estimated distributions of net training costs and apprentices' marginal productivity, respectively. Panels (c) and (d) show the firm size distribution post-reform (2003–2009). Panel (e) shows the number of apprentices by firm size, and panel (f) shows the fraction of firms paying the fee by firm size, both in the post-reform period.



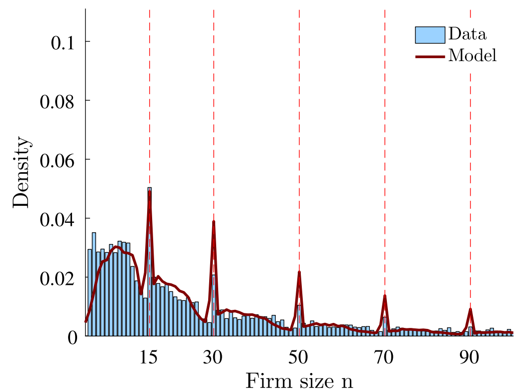
(a) Net Training Cost Distribution



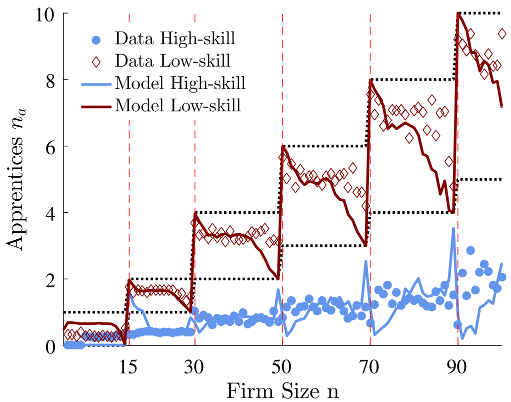
(b) Apprentices' Marginal Productivity Distribution



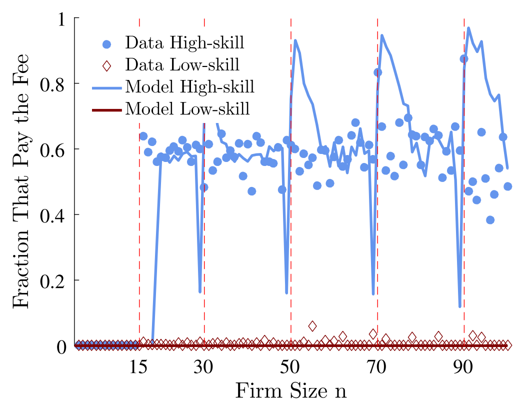
(c) Post-Reform Firm Size Distribution, High-Skill



(d) Post-Reform Firm Size Distribution, Low-Skill



(e) Number of Apprentices



(f) Fraction that Pay the Fee

FIGURE E6.—Model Fit Under Uniform Training Cost Distribution. *Notes*: The figure depicts the fit of the model under a uniform training cost distribution. Panels (a) and (b) show the estimated distributions of net training costs and apprentices' marginal productivity, respectively. Panels (c) and (d) show the firm size distribution post-reform (2003–2009). Panel (e) shows the number of apprentices by firm size, and panel (f) shows the fraction of firms paying the fee by firm size, both in the post-reform period.

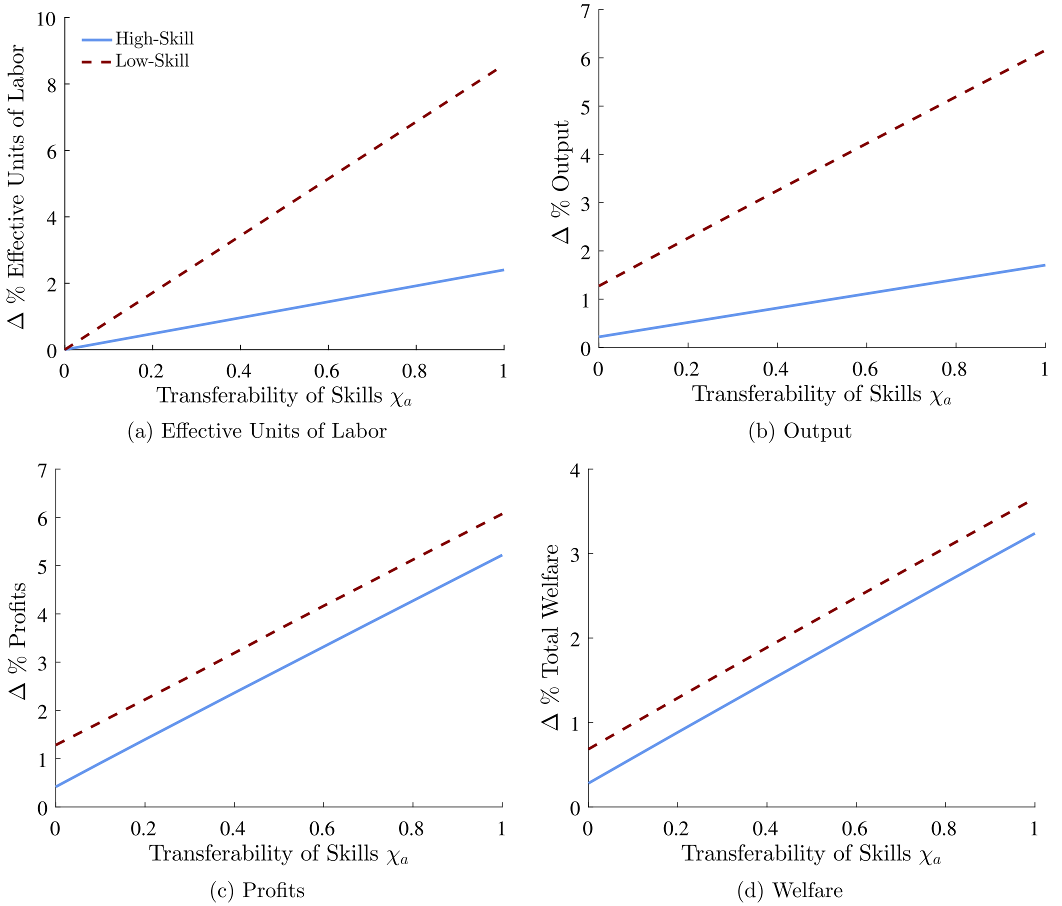


FIGURE E7.—Sensitivity Analysis: Skills Transferability. *Notes*: The figure shows dynamic effects of the regulation under different values of the skill transferability parameter χ_a. Effects are shown on effective units of labor (panel a), output (panel b), profits (panel c), and total welfare (panel d).

E.8. *Calibration of Extensions and Comparative Statics*

Dynamic Frictions. We use PILA data to calibrate variables related to the dynamic frictions model from Section D.4. Using the information from Table BIV, we calibrate the probability of apprentices staying in the same firm in the month after graduation (conditional on staying in the sample) to $\rho^s = 0.74$ and $\rho^u = 0.75$, respectively, in high-skill and low-skill sectors. These apprentices earn average wages of $w_a^{r,s} = 1.47$ and $w_a^{r,u} = 1.35$. Finally, we use the average observed wages of all other workers by sector to approximate $w_t^s = w_{t+1}^s = 3.15$ and $w_t^u = w_{t+1}^u = 2.42$. To obtain an upper bound of additional benefits from training, we set $\beta = 1$ and $\zeta_a^{r,k} = 1$, $\forall k$. Using these values, we compute the additional benefits in each sector and adjust the estimated training cost distribution (see Figure E8). Table EVII shows quantitative effects of the regulation with and without dynamic frictions. Figures E9 and E10 show comparative statics with respect to ρ and w_a^r, respectively.

TABLE EVI

POLICY DECOMPOSITION: AGGREGATE VARIABLES.

	(1) Workers	(2) % Workers	(3) Apprentices	(4) % Output
A. Only Quotas				
High-Skill	−9258	−4.34	8891	−3.17
Low-Skill	−3859	−1.83	9514	−0.42
Total	−13,118	−3.09	18,405	−1.95
B. Only $\downarrow w_a$				
High-Skill	−2345	−1.10	14,205	0.16
Low-Skill	−28,053	−13.30	156,683	3.86
Total	−30,398	−7.16	170,887	1.80
C. Quotas $+\downarrow w_a$				
High-Skill	−8024	−3.76	9786	−2.76
Low-Skill	−3519	−1.67	17,866	0.30
Total	−11,543	−2.72	27,653	−1.41
D. Full Regulation				
High-Skill	−1905	−0.89	4937	−0.34
Low-Skill	−3519	−1.67	17,866	0.30
Total	−5423	−1.28	22,803	−0.06

Note: The table shows the effects of apprenticeship regulation on aggregate outcomes, namely on the number of workers, the number of apprentices, and output. Effects are shown under four scenarios, namely only apprentice quotas (panel A), only a lower apprentice minimum wage (panel B), quotas plus the lower minimum wage (panel C), and the full regulation featuring quotas, the lower minimum wage, and the possibility of paying the fee (panel D).

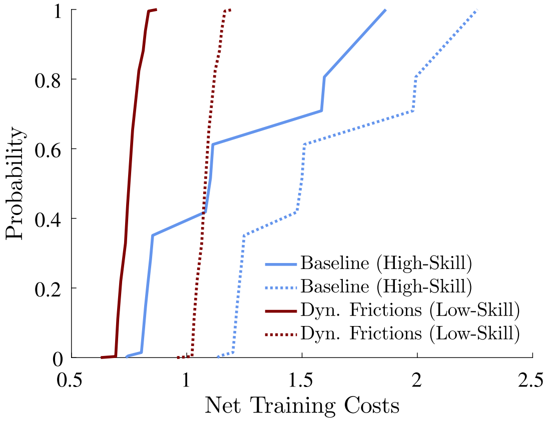


FIGURE E8.—Dynamic Frictions and Estimated Training Costs. *Notes*: The figure shows estimated training cost distributions from the baseline model (solid lines) and under dynamic frictions (dashed lines). We normalize $\zeta_a = 1$ as in main text Figure 7.

TABLE EVII

QUANTITATIVE EFFECTS WITH AND WITHOUT DYNAMIC FRICTIONS.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Aggregate Outcomes				*Changes in Agents' Welfare* $\Delta U_j/U^*$			
	Workers	Apprentices	Total Labor	% Output	Apprentices	Workers	Firms	Total
				A. Baseline $\rho = 0$				
High-Skill Sectors	−1905	4937	−1150	−0.34	0.27	−0.54	−0.46	−0.73
Low-Skill Sectors	−3519	17866	1088	0.30	1.14	−0.97	0.13	0.30
Total	−5423	22803	−62	−0.06	0.65	−0.73	−0.20	−0.28
				B. Frictions $\rho^k > 0$				
High-Skill Sectors	90	5020	−1119	−0.33	0.28	0.03	−1.03	−0.73
Low-Skill Sectors	2433	18467	1088	0.30	1.18	0.67	−1.55	0.30
Total	2524	23487	−31	−0.05	0.67	0.31	−1.26	−0.27

Note: Columns (1) to (4) of the table show the effects of the apprenticeship regulation on aggregate outcomes, namely on the number of workers, the number of apprentices, and output. Columns (5) to (7) show the effect on the welfare of apprentices, workers, and firm owners. Column (8) shows the sum of welfare effects across the groups of agents from columns (5) to (7). Panel A shows effects under the baseline model, and panel B shows effects with dynamic frictions.

Hiring Costs. We reestimate the training cost distribution under the model with hiring costs from Section D.5, setting the cost parameters to: fixed hiring costs $\kappa_0^h = 0.1$ and variable hiring costs $\kappa_1^h = 0.05$. These parameters are chosen to be large enough to have some effects on the estimated training cost distribution (see Figure E11), without radically changing the model fit (see Table EVIII). Choosing larger values of the hiring cost parameters would increasingly worsen the fit of the estimated model.

E.9. *Additional Results on Counterfactual Exercises*

Subsidizing Apprenticeship Training

The linearity of the model implies corner solutions. Lemma 7 characterizes this solution.

LEMMA 7: *Suppose a firm* (t_a, z) *solves* (5):

i. *If* $(1-\varsigma)w_a + w(1+\tau)t_a > w(1+\tau)$, *then the firm avoids apprentices and chooses*

$$n^* = \left(\frac{p\gamma}{w(1+\tau)}\right)^{1/(1-\gamma)} z, \qquad n_a^* = 0.$$

TABLE EVIII

HIRING COSTS, TRAINING COSTS, AND MODEL FIT.

	(1) Baseline	(2) Fixed Hiring Cost $\kappa_0^h = 0.1$	(3) Variable Hiring Cost $\kappa_1^h = 0.05$
Average t_a in High-Skill	1.188	1.210	1.045
Average t_a in Low-Skill	0.753	0.750	0.740
Total Score SMM	0.58	0.63	0.58

Note: The table shows average training costs by sector and the model fit (measured by the score function of the SMM estimation) under the baseline model (column 1), the model with fixed hiring costs (column 2) and the model with variable hiring costs (column 3).

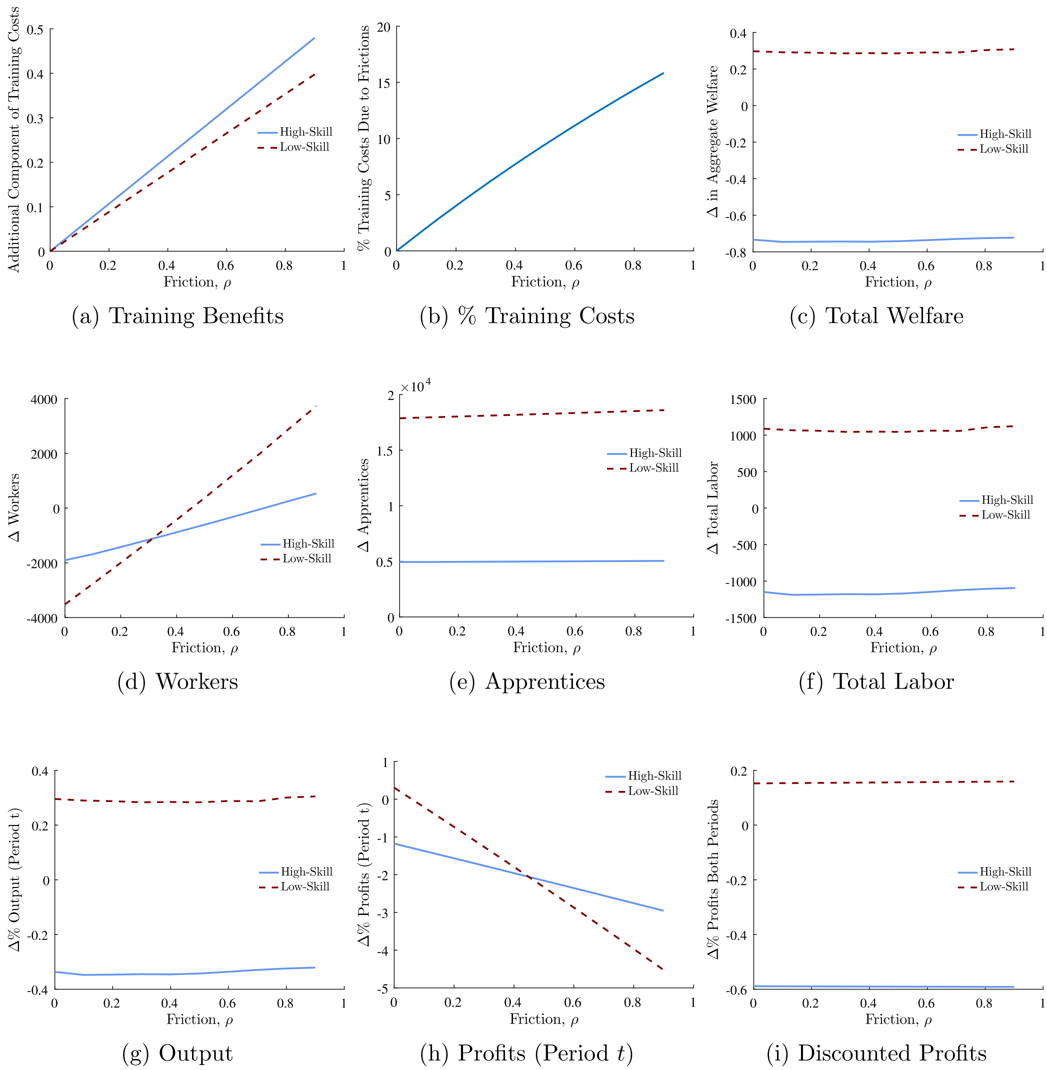


FIGURE E9.—Comparative Statics: Frictions. *Notes*: The figure shows comparative statics with respect to the probability of staying in the firm (ρ), based on estimating the dynamic model described in Section D.4.

ii. *If* $(1-\varsigma)w_a + w(1+\tau)t_a < w(1+\tau)$, *then the firm trains apprentices and chooses*

$$n^* = t_a n_a^*, \qquad n_a^* = \left(\frac{p\gamma\zeta_a^\gamma}{\tilde{w}}\right)^{1/(1-\gamma)} z, \quad \textit{where } \tilde{w} := w_a(1-\varsigma) + w(1+\tau)t_a.$$

Case (i) describes the situation where the cost of training apprentices is higher than the total cost of hiring workers, once the tax and the subsidy are taken into account. The tax scheme harms firms with high t_a, since these training costs are not directly covered by the subsidy. On the other hand, in Case (ii), when the tax and the subsidy are high enough, firms are incentivized to train. Thus, the demand for apprentices depends on whether the subsidy covers all monetary training expenses.

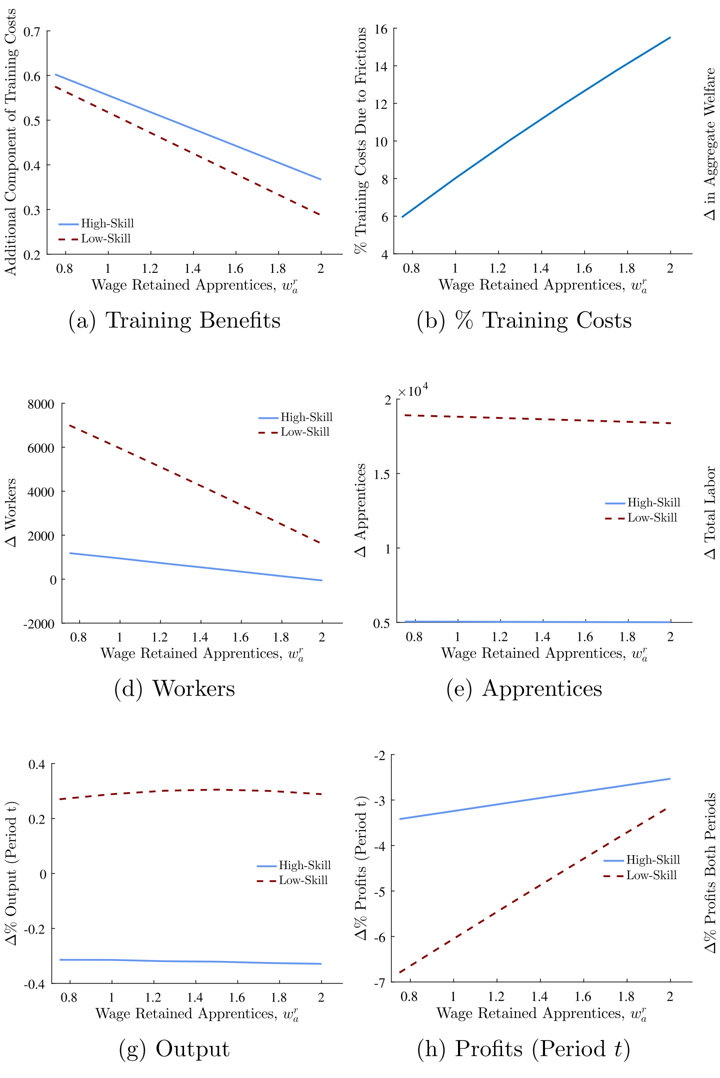


FIGURE E10.—Comparative Statics: Wage Compression. *Note*: The figure shows the comparative statics with respect to the wages of retained apprentices (w_a^r), based on estimating the dynamic model described in Section D.4.

Additionally, we ensure the policy is budget-balanced. Total revenue from payroll taxes is $\text{Rev}(\tau, \varsigma) := \sum_k \tau w^k N^k(\tau, \varsigma)$. Total subsidy payments to firms equal

$$\text{Sub}(\tau, \varsigma) := \sum_k F^k \int \int \mathcal{S}^*(t_a, z; \tau, \sigma)\, d\mathcal{Z}^k(z)\, d\mathcal{T}^k(t_a),$$

where F^k is the number of firms in sector k and $\mathcal{S}^*(t_a, z; \tau, \sigma)$ denotes the amount of subsidy optimally taken by firm (t_a, z) when facing policy (τ, σ). In Section 6 of the paper, we consider the set of budget-balanced policies (τ, ς) such that $\text{Rev}(\tau, \varsigma) - \text{Sub}(\tau, \varsigma) = 0$.

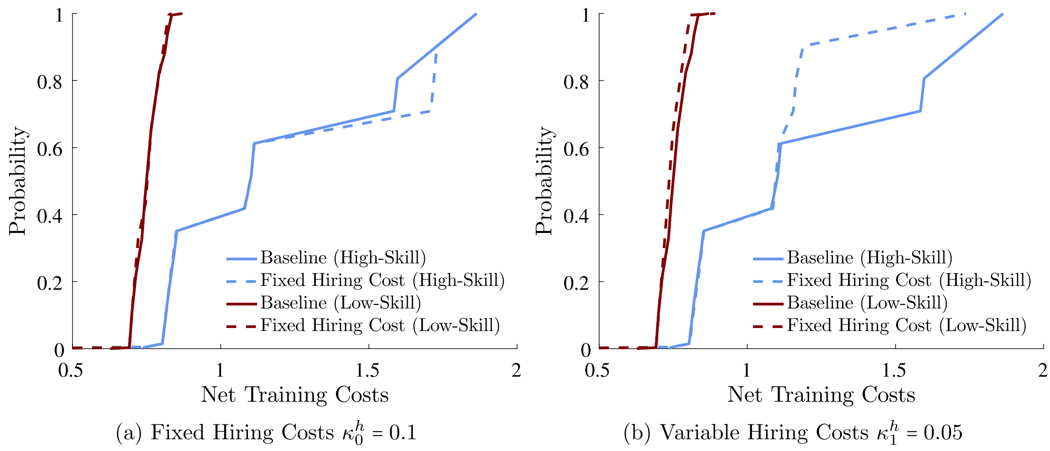


FIGURE E11.—Hiring Costs and Estimated Training Costs. *Notes*: The figure shows estimated training cost distributions from the baseline model (solid lines in both panels), from the model with with fixed hiring costs (dashed lines in panel a), and from the model with variable hiring costs (dashed lines in panel b). We normalize $\zeta_a = 1$ as in main text Figure 7.

Sector-Specific Apprentice Minimum Wage

In the following, we describe the details of computing the sector-specific minimum wage counterfactual policy. Let $n_a^{k,\min}(z, t_a; w_a^k, w^k)$ denote the solution to the firm maximization problem when the wage of workers is w^k and the apprentice wage is w_a^k. For each sector k, we compute the minimum wage for apprentices w_a^{*k} that solves

$$N_a^{*k} = F^k \int \int n_a^{k,\min}\big(z, t_a; w_a^{*k}, w^k\big)\, d\mathcal{Z}^k(z)\, d\mathcal{T}^k(t_a), \tag{9}$$

where N_a^{*k} denotes the total number of apprentices trained in sector k under the original regulation. We take worker wages and structural parameter estimates from the baseline model.

Under linear labor inputs, a firm (t_a, z) trains apprentices if the total cost of training is smaller than that of hiring workers, and chooses $n_a^{k,\min} = (\frac{p^k \gamma^k (\zeta_a^k)^{\gamma^k}}{w_a^k + w^k t_a})^{1/(1-\gamma^k)} z$ and $n^{k,\min} = t_a n_a^{k,\min}$. Conversely, the firm does not train apprentices if $\frac{w_a^k + w^k t_a}{\zeta_a^k} > w^k$, such that $n_a^{k,\min} = 0$ and $n^{k,\min} = (\frac{p^k \gamma^k}{w^k})^{1/(1-\gamma^k)} z$. We solve equation (9) numerically and obtain apprentice minimum wages of $w_a^{*s} = 0.77$ in high-skill sectors and of $w_a^{*u} = 0.98$ in low-skill sectors.

In Figure E12, we target the same total number of apprentices as under the baseline regulation, but change the allocation of apprentices across sectors. We show that the effect of changing the fraction of apprentices trained in high-skill sectors on the number of apprentices per worker, number of workers, total output, and total welfare.

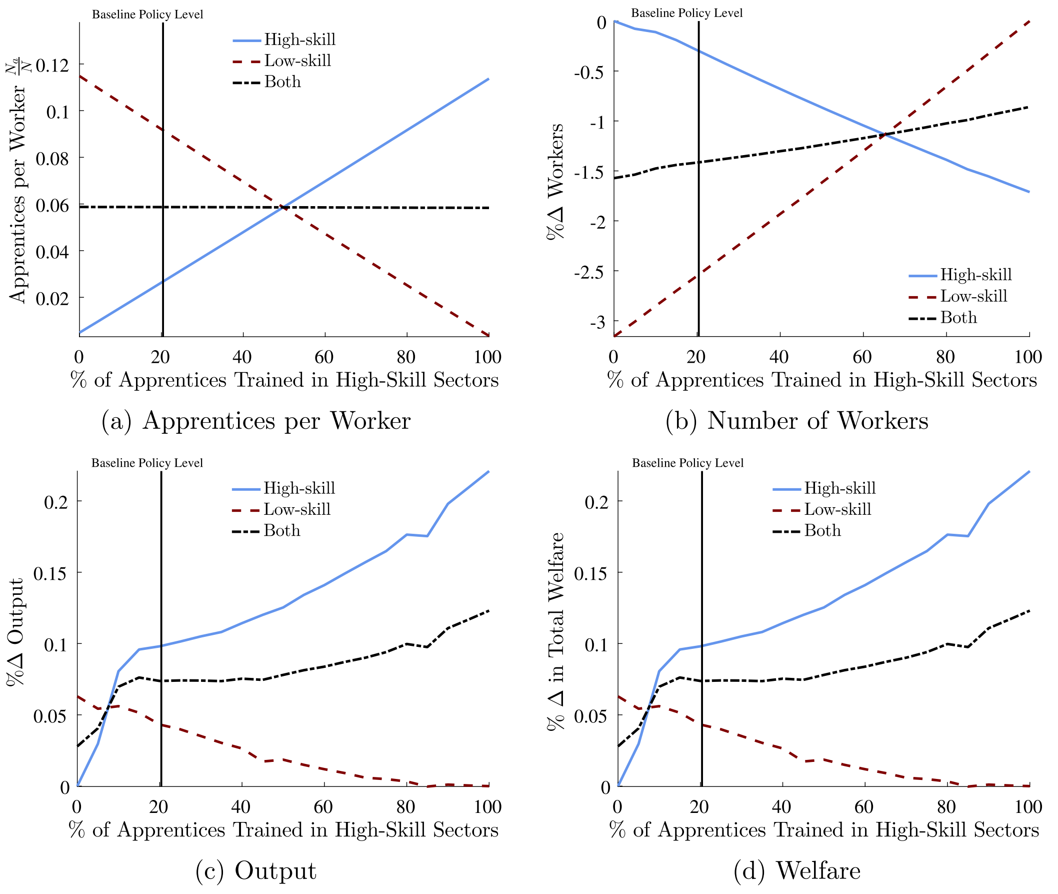


FIGURE E12.—Varying the Share of Training in High-Skill Sectors. *Notes*: The figure shows the effects of varying the share of training in high-skill sectors by varying sector-specific minimum wages, holding the total number of apprentices fixed. Panel (a) shows the number of apprentices per worker, panel (b) shows the percentage change in the number of workers, panel (c) shows the percentage change in aggregate output, and panel (d) the change in total welfare measured as the change in total aggregate utility relative to prereform utility, $\Delta\mathcal{U}/\mathcal{U}^*$.

REFERENCES

ACEMOGLU, D., AND J.-S. PISCHKE (1998): "Why Do Firms Train? Theory and Evidence," *The Quarterly Journal of Economics*, 113 (1), 79–119. [29]

——— (1999): "The Structure of Wages and Investment in General Training," *Journal of Political Economy*, 107 (3), 539–572. [29]

ACKERBERG, D. A., K. CAVES, AND G. FRAZER (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83 (6), 2411–2451. [30]

CAICEDO, S., M. ESPINOSA, AND A. SEIBOLD (2022): "Unwilling to Train?—Firm Responses to the Colombian Apprenticeship Regulation," *Econometrica*, 90 (2), 507–550. [1]

LEVINSOHN, J., AND A. PETRIN (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, 70 (2), 317–341. [30]

MOLLISI, V., AND G. ROVIGATTI (2017): "Theory and Practice of TFP Estimation: The Control Function Approach Using Stata," Working Paper. [30]

OLLEY, G. S., AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64 (6), 1263–1297. [30]

SENA (2016): "Data Report on Vocational Education and Training—Information and Analysis That Contribute to the Development of Colombian Vocational Education and Training." [11]
WOOLDRIDGE, J. M. (2009): "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters*, 104 (3), 112–114. [30]

Co-editor Oriana Bandiera handled this manuscript.

Manuscript received 1 September, 2020; final version accepted 17 July, 2021; available online 4 August, 2021.