Online Appendix

Labor Market Power, Self-Employment, and Development

Francesco Amodio, Pamela Medina, Monica Morlacco*

This Appendix is organized as follows. Section A contains the additional tables and figures discussed in the text. Section B shows that the empirical facts shown in Section 2 are robust to the use of alternative concentration measures and variable definitions. Section C integrates Section 3 by showing additional theoretical results. Section D complements Section 4 and provides further details on model estimation.

*Amodio: francesco.amodio@mcgill.ca, McGill University and IZA, Department of Economics and ISID, 855 Sherbrooke St. West, Montreal, QC, H3A 2T7; Medina: pamela.medinaquispe@rotman.utoronto.ca, University of Toronto, UTSC and Rotman School of Management, 105 St George St, Toronto, ON M5S 3E6; Morlacco: morlacco@usc.edu, University of Southern California, Department of Economics, 3620 S. Vermont Avenue, Los Angeles, CA 90089-1422.

A Additional Tables and Figures

Table A.1: Employment Distribution Across Sectors and Transitions Across Wage Work and Self-Employment

	All Workers	Wage Workers			Self-Employed		
		All	Self-Employed at $t-1$	Self-Employed Manuf. at $t-1$	All	Wage Workers at $t-1$	Wage Workers Manuf. at $t-1$
Agriculture	0.36	0.15	0.25	0.12	0.42	0.37	0.12
Mining	0.01	0.03	0.03	0.02	0.01	0.01	0.01
Manufacturing	0.08	0.10	0.08	0.31	0.08	0.07	0.28
Utilities	0.00	0.01	0.01	0.02	0.00	0.00	0.00
Construction	0.05	0.09	0.14	0.12	0.03	0.08	0.03
Retails	0.14	0.05	0.04	0.05	0.22	0.14	0.20
Transportation	0.05	0.03	0.04	0.02	0.07	0.10	0.11
Other Services	0.30	0.56	0.42	0.35	0.17	0.24	0.25
Observations	29970	11008	1736	113	15709	1746	162

Notes. This table shows the distribution of employment across sectors amongst all workers, wage workers, and self-employed workers in the 2007-2011 panel version of ENAHO. It reports these distributions at the end of each two-year period. It also shows the distribution of wage work (and self-employment) among workers who were self-employed (wage workers) in the previous year, across all sectors and in manufacturing only.

Table A.2: Concentration, Self-Employment, and Earnings

	Self-Employed {0,1}			Log of Earnings					
					Wage Workers			Self-employed	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Employer Concentration	0.050***	0.049***	0.062***	-0.085***	-0.100***	-0.052**	-0.124**	-0.158***	-0.051
(Log of Wage-bill HHI)	(0.012)	(0.014)	(0.015)	(0.020)	(0.021)	(0.021)	(0.057)	(0.048)	(0.052)
Female	0.122***	0.121***	0.111***	-0.426***	-0.381***	-0.382***	-1.313***	-1.228***	-1.211***
	(0.023)	(0.023)	(0.023)	(0.041)	(0.036)	(0.035)	(0.072)	(0.075)	(0.078)
Age	0.018***	0.017***	0.016***	0.023**	0.027***	0.029***	0.116***	0.115***	0.110***
	(0.005)	(0.005)	(0.005)	(0.009)	(0.009)	(0.009)	(0.018)	(0.017)	(0.018)
Age sq.	-0.000	-0.000	-0.000	-0.000	-0.000**	-0.000**	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Schooling	-0.006	-0.001	0.000	0.178***	0.161***	0.156***	0.111***	0.109***	0.101***
	(0.004)	(0.004)	(0.004)	(0.008)	(0.008)	(0.007)	(0.018)	(0.017)	(0.019)
Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	7637	7637	7634	4707	4706	4698	2054	2054	2047
R^2	0.102	0.132	0.156	0.308	0.363	0.395	0.327	0.383	0.399

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Unit of observation is a working-age individual surveyed in ENAHO. A local labor market k is defined by a 2-digit industry j within a province or commuting zone g. This table reports the coefficient estimates and their standard errors obtained when estimating the regression specification: $y_{i(j,g)t} = \beta \ln HHI^{wn}_{(j,g)t} + \mathbf{X}'_{i(j,g)t}\theta + \gamma_j + \lambda_g + \delta_t + u_{i(j,g)t}$, where $y_{i(j,g)t}$ is the labor market outcome of worker i in local labor market k as defined by a manufacturing industry j within a province or commuting zone g in year t. The first regressor $\ln HHI^{wn}_{(j,g)t}$ is the log of wage-bill HHI in the market in the same year. $\mathbf{X}_{i(j,g)t}$ is a vector of individual characteristics, while γ_j, λ_g and δ_t stand for industry, location, and year fixed effects respectively. Standard errors are clustered at the local labor market level.

Table A.3: Estimates of Labor Market Power – First Stage Regression Results

$HHI^{wn} \in$ Self-employment Rate	$(0,1]$ All (1)	$(0,0.18]$ All (2)	$(0.18,0.25]$ All (3)	$(0.25,1]$ All (4)	$(0,0.25]$ Low (5)	$(0,0.25]$ High (6)	$(0.25,1]$ Low (7)	$(0.25,1]$ High (8)
				Log of Employment				
$PER_{gt} \times EC_{i(j,g)}$	0.005***							
	(0.000)							
$\times \mathbb{I}\{HHI^{wn} \in (0,0.18]\}$		-0.001	0.002***	0.004***				
		(0.002)	(0.001)	(0.001)				
$\times \mathbb{I}\{HHI^{wn} \in (0.18,0.25]\}$		0.019***	-0.026***	0.011***				
		(0.001)	(0.001)	(0.002)				
$\times \mathbb{I}\{HHI^{wn} \in (0.25,1]\}$		0.012***	-0.003**	-0.006***				
		(0.002)	(0.001)	(0.002)				
$\times \mathbb{I}\{HHI^{wn} \in (0,0.25]\} \times \mathbb{I}\{\text{Low SE}\}$					-0.010***	0.009***	0.003***	0.002**
					(0.001)	(0.000)	(0.000)	(0.001)
$\times \mathbb{I}\{HHI^{wn} \in (0,0.25]\} \times \mathbb{I}\{\text{High SE}\}$					0.012***	-0.010***	0.003***	0.002**
					(0.001)	(0.001)	(0.000)	(0.001)
$\times \mathbb{I}\{HHI^{wn} \in (0.25,1]\} \times \mathbb{I}\{\text{Low SE}\}$					0.001	0.003	-0.009***	0.010***
					(0.001)	(0.003)	(0.002)	(0.003)
$\times \mathbb{I}\{HHI^{wn} \in (0.25,1]\} \times \mathbb{I}\{\text{High SE}\}$					0.010***	0.011***	0.006***	-0.026***
					(0.002)	(0.002)	(0.002)	(0.003)
Observations	6191	6191	6191	6191	6191	6191	6191	6191
R^2	0.952	0.974	0.959	0.979	0.966	0.957	0.973	0.971

Notes. * p-value< 0.1; ** p-value<0.05; *** p-value<0.01. Unit of observation is a medium to large firm in EEA. The table reports the first stage estimates corresponding to the second stage estimates reported in Table 2. The dependent variable is the log of firm-level employment $\ln l_{i(j,g)t}$. The instrumental variable is the interaction of the cumulative number of PER projects completed in each location g up to year t (PER_{gt}) and a dummy equal to one for firms with higher than median constraints to access electricity at baseline ($EC_{i(j,g)}$). Column 1 reports the first-stage estimates associated with column 1 of Table 2. Columns 2 to 4 report the first-stage results from the three first-stage regressions associated with column 2 of Table 2. Columns 5 to 8 report those from the four first-stage regressions associated with column 3 and 4 of Table 2. Following equation 3, firm fixed effects and local labor market $\times$ year fixed effects are included in all specifications. Standard errors are clustered at the level of location g, i.e. province or commuting zone.

Table A.4: Estimates of Labor Market Power
Robustness to District $\times$ Industry $\times$ Year Fixed Effects

			Self-Employment Rate	
			Low	High
	(1)	(2)	(3)	(4)
All Markets	0.565***			
	(0.105)			
$HHI^{wn} \in (0, 0.18]$		-0.062		
		(0.064)		
$HHI^{wn} \in (0.18, 0.25]$		0.450***		
		(0.142)		
$HHI^{wn} \in (0, 0.25]$			-0.127	-0.041
			(0.309)	(0.279)
$HHI^{wn} \in (0.25, 1]$		0.861***	1.725***	-0.502
		(0.050)	(0.365)	(0.643)
SW F-statistics	376.08	848.96	282.72	1065.85
		1545.52	446.12	1211.87
		14295.91		
Observations	4954	4954	3257	1697

Notes. * p-value< 0.1; ** p-value<0.05; *** p-value<0.01. The unit of observation is a medium to a large firm in EEA. The table reports 2SLS estimates of the firm-level inverse elasticity of supply of wage labor as captured by β in equation (1). The instrumental variable is the interaction of the cumulative number of PER projects completed in each location g up to year t (PER_{gt}) and a dummy equal to one for firms with higher than median constraints to accessing electricity at baseline ($EC_{i(j,g)}$). Estimates in Columns 2 to 4 are obtained by interacting both the log of firm-level employment $\ln l_{i(j,g)t}$ and the instrument $PER_{gt} \times EC_{i(j,g)}$ with dummy variables that identify the different subsamples as discussed in the text. Low and high self-employment rates are defined as below and above the average self-employment rate across local labor markets, respectively. We report the F-statistic associated with the Sanderson-Windmeijer multivariate test of excluded instruments for each estimate. Firm fixed effects and district (instead of province or commuting zone as in baseline) $\times$ industry $\times$ year fixed effects are included in all specifications. Standard errors are clustered at the level of location g, i.e., province or commuting zone.

Table A.5: Estimates of Labor Market Power – First Stage Regression Results
Robustness to District $\times$ Industry $\times$ Year Fixed Effects

$HHI^{wn} \in$	$(0,1]$	$(0,0.18]$	$(0.18,0.25]$	$(0.25,1]$	$(0,0.25]$	$(0,0.25]$	$(0.25,1]$	$(0.25,1]$
Self-employment Rate	All	All	All	All	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log of Employment							
$PER_{gt} \times EC_{i(j,g)}$	0.006***							
	(0.000)							
$\times \mathbb{I}\{HHI^{wn} \in (0,0.18]\}$		-0.000	0.002***	0.004***				
		(0.001)	(0.000)	(0.000)				
$\times \mathbb{I}\{HHI^{wn} \in (0.18,0.25]\}$		0.024***	-0.032***	0.014***				
		(0.003)	(0.004)	(0.001)				
$\times \mathbb{I}\{HHI^{wn} \in (0.25,1]\}$		0.014***	0.001	-0.005***				
		(0.001)	(0.002)	(0.001)				
$\times \mathbb{I}\{HHI^{wn} \in (0,0.25]\} \times \mathbb{I}\{\text{Low SE}\}$					-0.008***	0.009***	0.003***	0.002***
					(0.000)	(0.000)	(0.000)	(0.000)
$\times \mathbb{I}\{HHI^{wn} \in (0,0.25]\} \times \mathbb{I}\{\text{High SE}\}$					0.013***	-0.012***	0.003***	0.003***
					(0.001)	(0.000)	(0.000)	(0.000)
$\times \mathbb{I}\{HHI^{wn} \in (0.25,1]\} \times \mathbb{I}\{\text{Low SE}\}$					0.003*	0.008***	-0.002	0.003
					(0.001)	(0.001)	(0.002)	(0.003)
$\times \mathbb{I}\{HHI^{wn} \in (0.25,1]\} \times \mathbb{I}\{\text{High SE}\}$					0.013***	0.016***	-0.002	-0.019***
					(0.001)	(0.001)	(0.002)	(0.003)
Observations	4954	4954	4954	4954	4954	4954	4954	4954
R^2	0.954	0.980	0.970	0.984	0.973	0.967	0.980	0.979

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Unit of observation is a medium to large firm in EEA. The table reports the first stage estimates corresponding to the second stage estimates reported in Table 2. The dependent variable is the log of firm-level employment $\ln l_{i(j,g)t}$. The instrumental variable is the interaction of the cumulative number of PER projects completed in each location g up to year t (PER_{gt}) and a dummy equal to one for firms with higher than median constraints to access electricity at baseline ($EC_{i(j,g)}$). Column 1 reports the first-stage estimates associated with column 1 of Online Appendix Table A.4. Columns 2 to 4 report the first-stage results from the three first-stage regressions associated with column 2 of Online Appendix Table A.4. Columns 5 to 8 report those from the four first-stage regressions associated with column 3 and 4 of Online Appendix Table A.4. Firm fixed effects and district (instead of province or commuting zone as in baseline) $\times$ industry $\times$ year fixed effects are included in all specifications. Standard errors are clustered at the level of location g, i.e. province or commuting zone.

Table A.6: Employer Concentration Across Local Labor Markets
Alternative Samples and Definitions

	Full Sample		*Merged Sample*	
	Mean	St. Dev.	Mean	St. Dev.
Number of Firms	6.39	10.37	7.25	11.19
Wage-bill HHI	0.65	0.33	0.61	0.34
Wage-bill HHI (Payroll Weighted)	0.37	0.03	0.37	0.03
Wage-bill HHI (Employment Weighted)	0.34	0.03	0.34	0.03
Employment HHI	0.63	0.35	0.59	0.35
Employment HHI (Payroll Weighted)	0.33	0.03	0.33	0.03
Employment HHI (Employment Weighted)	0.31	0.02	0.31	0.02
Percent of LLMs with 1 firm	38.78	2.27	38.78	2.29
Payroll Share of LLMs with 1 firm	7.94	1.79	7.96	1.81
Employment Share of LLMs with 1 firm	7.80	1.23	7.81	1.25
Number of Local Labor Markets	280		228	
Number of Locations	61		48	
Industries	23		22	

Notes. This table presents summary statistics and employer concentration measures derived from EEA firm-level data across Peruvian local labor markets, averaged over the years 2004 to 2011. The data are shown separately for the entire sample and the subset merged with worker-level data from ENAHO. In Local labor markets are defined by 2-digit industries within locations, with locations corresponding to Peruvian provinces or commuting zones.

Table A.7: Structural Inverse Supply Elasticity – Model Estimates

			Self-Empl. Rate	
			Low	High
	(1)	(2)	(3)	(4)
All Markets	0.298			
$HHI^{wn} \in (0, 0.18]$		0.125		
$HHI^{wn} \in (0.18, 0.25]$		0.191		
$HHI^{wn} \in (0, 0.25]$			0.182	0.129
$HHI^{wn} \in (0.25, 1]$		0.373	0.461	0.267

Notes. This table presents the estimates of the average structural inverse labor supply elasticity of treated firms, obtained as $\epsilon_{iF,k} \equiv \psi_{iF,k} - 1$, across all markets (Column 1) and within different market subsets (Columns 2 to 4) in the estimated model. Low and high self-employment rates are defined as being below or above the average self-employment rate across local labor markets, respectively.

Table A.8: Median Estimates of Roy Parameters

	σ_F	σ_S	ϱ	$\hat{\mu}$ $\ln \frac{\text{Earnings}_F}{\text{Earnings}_S}$	$\ln \frac{\text{Revenues}_F}{\text{Earnings}_S}$	$\beta = 1$
Panel I. Baseline Estimates						
	0.81	0.91	0.89	-0.16	-0.1	-0.09
Panel II. Group Heterogeneity						
Group 1	0.91	0.93	0.87	-0.14	-0.02	-0.03
Group 2	0.83	0.91	0.96	-0.12	-0.11	-0.10
Group 3	0.69	0.91	0.88	-0.27	-0.16	-0.19

Notes. This table presents the median estimates of the Roy model parameters, including the variance-covariance matrix parameters (σ_F, σ_S, ϱ) and the relative mean ability ($\hat{\mu}$) across different market groups. Panel I displays the baseline estimates, where these parameters are held constant across markets to simplify the estimation process and reduce the potential impact of measurement errors. Panel II shows estimates allowing for heterogeneity across three market groups, clustered based on population terciles. The three columns under relative mean ability refer to robustness tests, as described in Appendix D.1.2, with $\ln \frac{\text{Earnings}_F}{\text{Earnings}_S}$ as our baseline.

Table A.9: Labor Market Power – Robustness

	(1) Baseline	(2) Roy Estimates	(3) Fixed Costs	(4) Entry	(5) Census Data
$\bar{\psi}_k$	1.45	1.43	1.49	1.44	1.31
— High HHI	1.71	1.71	1.78	1.67	1.56
— High self-employment	1.40	1.45	1.35	1.40	1.28

Notes. This table presents the estimates of the average markdown across different model calibrations and in different subgroups of markets. $\bar{\psi}_k$ represents the average labor market power across all markets. The second and third rows show the average labor market power in markets with high concentration and high self-employment rates, respectively.

Table A.10: Targeted Moments and Model Fit – Robustness

Moment	Baseline	Fixed Costs	Entry
Panel I. Distribution Moments			
Log Number of Firms			
Mean	0.97	2.29	1.00
Standard Deviation	0.95	1.71	0.85
Log of Sales			
Ratio p75/p25	2.94	2.92	2.93
Ratio p90/p10	5.29	5.16	5.30
CR_1, Mean	0.66	0.70	0.67
CR_1, Standard Deviation	0.29	0.25	0.26
CR_4, Mean	0.94	0.96	0.95
CR_4, Standard Deviation	0.11	0.08	0.10
Employment HHI			
Mean, Unweighted	0.57	0.55	0.56
Standard Deviation	0.34	0.30	0.31
Mean, Weighted	0.30	0.45	0.29
Percent of Markets with 1 firm	0.36	0.20	0.29
Wage-bill Share of			
Markets with 1 firm	0.07	0.08	0.01
Markets with $<$10 firms	0.89	0.49	0.91
Markets with $<$50 firms	1.00	0.72	1.00
Share of Wage Employment			
Mean	0.66	0.68	0.67
Standard Deviation	0.12	0.10	0.12
Log of Earnings$_F$/Earnings$_S$			
Mean	0.41	0.52	0.43
Standard Deviation	0.58	0.53	0.59
Log of Schooling$_F$/Schooling$_S$ (Ability$_F$/Ability$_S$)			
Ratio p75/p25	1.42	1.63	1.34
Ratio p90/p10	1.18	1.28	1.15
Panel II. Regression Coefficients			
% Wage Employment on (Log) HHI^n			
Point Estimate	-0.04	-0.01	-0.06
Standard Error	0.01	0.01	0.01
(Log) Earnings$_S$ on (Log) HHI^n			
Point Estimate	-1.17	-0.27	-1.61
Standard Error	0.12	0.16	0.12
(Log) Earnings$_F$ on (Log) HHI^n			
Point Estimate	-1.36	-0.31	-1.90
Standard Error	0.13	0.15	0.13

Notes. This table reports the moments calculated from the baseline estimated model and compares them those estimated when relaxing some parametric assumptions, as described in Section 4.5.

Table A.11: Effect of Productivity Shock in Partial and General Equilibrium

	$\Delta\bar{Y}$ (%)	
	PE	GE
Log Avg. Wage $\bar{a}_F W_F$	-0.46	2.35
Log Avg. Ability $\bar{a}_F$	0.00	0.00
Log Unit Wage W_F	-0.46	2.35
Log Markdown $\bar{\psi}_{F,k}$	0.44	0.35
Log $\overline{MRPL}_{F,k}$	-0.01	2.70
Log Price Index $P_{F,k}$	-3.30	-0.61
Log Productivity Index $Z_{F,k}$	3.28	3.24
Log $\overline{\mu}_{F,k}$	-0.01	-0.06
Log Avg. Self-Empl. earnings $\bar{a}_S W_S$	-0.70	2.09
Log Avg. Ability $\bar{a}_S$	0.50	0.54
Log Unit Earnings W_S	-1.20	1.55

Notes. This table reports the percentage change in the average of selected outcomes across markets following the same productivity shock in partial equilibrium (PE) and general equilibrium (GE). In the PE exercise, we keep aggregate income Y (the only GE variable in our model) constant at its baseline level, focusing solely on the market responses to the productivity shock. In the GE exercise, we allow aggregate income to adjust by solving for the full model.

Table A.12: Impact of Labor Market Power Across Markets

	$\bar{Y}_{\iota=1}$	$\bar{Y}_{\iota=0}$	$\bar{Y}_{\iota=1} - \bar{Y}_{\iota=0}$
Wage Employment Share	0.66	0.77	0.11
Wage-bill Concentration HHI_k^{wb}	0.57	0.63	0.06
Log Avg. Wage $\bar{a}_F W_F$	-3.76	-3.45	0.31
Log Avg. Ability $\bar{a}_F$	1.89	1.89	0.00
Log Unit Wage W_F	-5.64	-5.34	0.31
Log Markdown $\bar{\psi}_{F,k}$	0.35	0	-0.35
Log $\overline{MRPL}_{F,k}$	-5.29	-5.34	-0.04
Log Price Index $P_{F,k}$	-5.59	-5.64	-0.05
Log Productivity Index $Z_{F,k}$	0.64	0.66	0.02
Log $\overline{\mu}_{F,k}$	0.34	0.35	0.01
Log Avg. Self-Empl. earnings $\bar{a}_S W_S$	-3.46	-3.19	0.27
Log Avg. Ability $\bar{a}_S$	2.49	2.6	0.11
Log Unit Earnings W_S	-5.95	-5.79	0.15
Log Labor Income	-3.72	-3.46	0.26
Wage Labor Supply Elasticity $\epsilon(\hat{W}_k)$	0.24	-0.18	-0.42

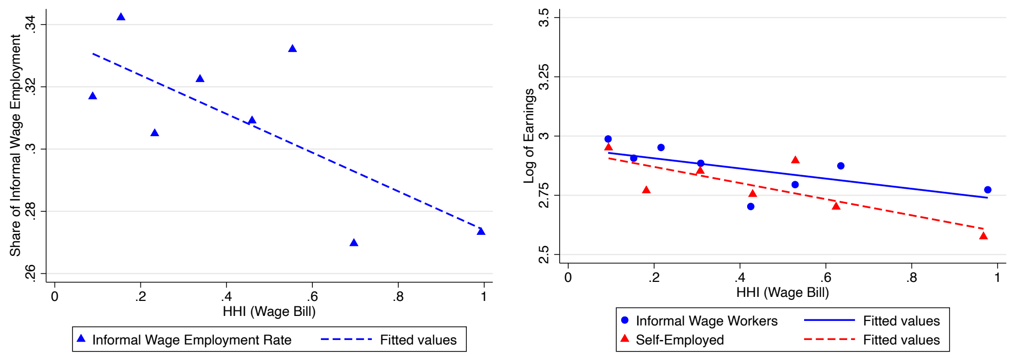
Notes. This table reports the average of selected outcomes across markets in the baseline economy ($\bar{Y}_{\iota=1}$) and in the counterfactual economy with no labor market power ($\bar{Y}_{\iota=0}$) together with the difference between the two.

Table A.13: Average Policy Impact Across Markets

		$\Delta\bar{Y}$	
	Firm Productivity ΔT_k	Fixed Cost Δf_k^e	Worker Skills $\Delta\hat{\mu}_k$
Wage Employment Share	0.67	1.37	3.57
Wage-bill Concentration HHI_k^{wb}	-0.30	-5.54	-2.49
Log Avg. Wage $\bar{a}_F W_F$	2.35	2.14	12.32
Log Avg. Ability $\bar{a}_F$	0	0	16.5
Log Unit Wage W_F	2.35	2.14	-4.18
Log Markdown $\bar{\psi}_{F,k}$	0.35	-1.93	1.49
Log $\overline{MRPL}_{F,k}$	2.70	0.21	-2.69
Log Price Index $\overline{P}_{F,k}$	-0.61	-1.43	-3.28
Log Productivity Index $\overline{Z}_{F,k}$	3.24	0.20	-0.05
Log Markup $\overline{\mu}_{F,k}$	-0.06	-1.44	-0.64
Log Avg. Self-Empl. earnings $\bar{a}_S W_S$	2.09	1.61	10.95
Log Avg. Ability $\bar{a}_S$	0.54	1.06	2.94
Log Unit Earnings W_S	1.55	0.55	8.01
Log Labor Income	2.03	1.47	10.62
Wage Labor Supply Elasticity $\epsilon(\hat{W}_k)$	-1.91	-3.69	-10.42

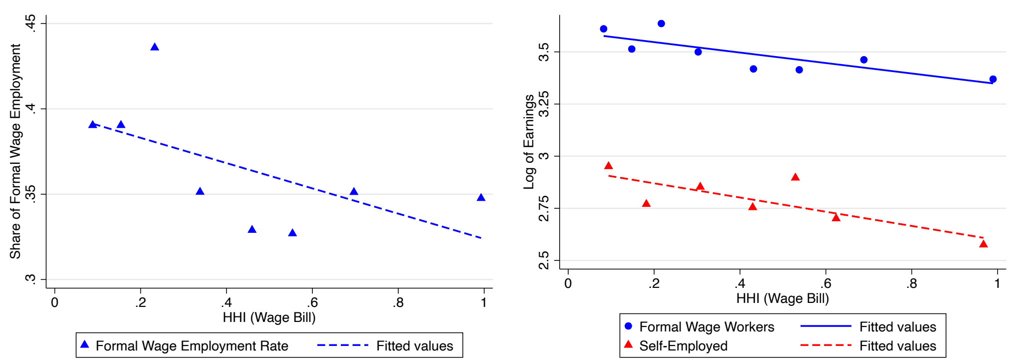
Notes. This table reports the percentage change in the average of selected outcomes across markets following the policy shocks discussed in Section 5.2.

Figure A.1: Concentration, Informal Wage Employment Rate, and Earnings



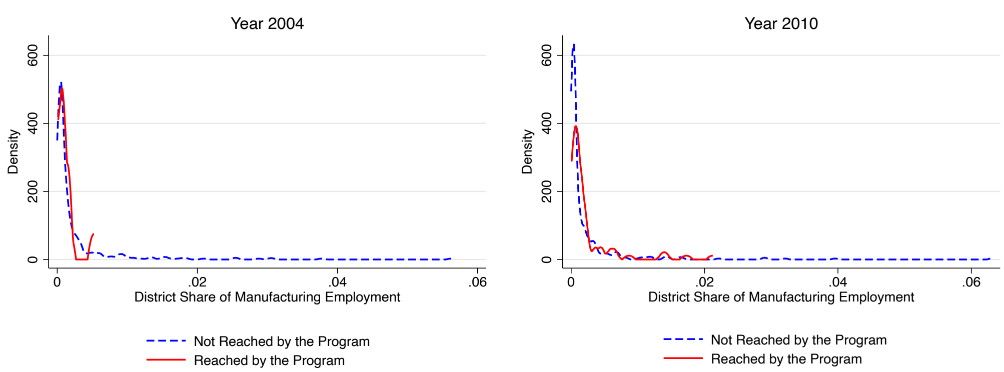
Notes. The figures illustrate the relationship between employer concentration, rate of informal wage employment (left), and earnings from both informal wage work and informal self-employment (right) across local labor markets. The left panel plots the share of informal wage workers in each decile of the wage-bill HHI distribution across local labor markets. The right panel plots the average log of daily earnings in each decile and separately for informal wage and self-employed workers. The straight lines show the linear fit based on the underlying data.

Figure A.2: Concentration, Formal Wage Employment Rate, and Earnings



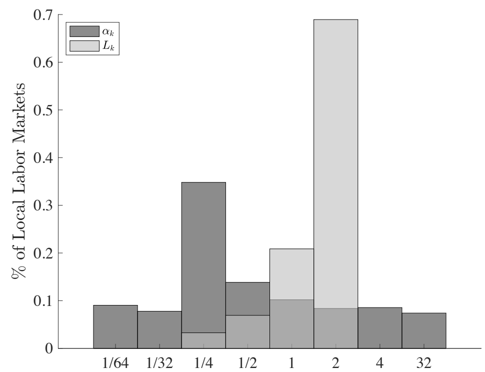
Notes. The figures illustrate the relationship between employer concentration, rate of formal wage employment (left), and earnings from both formal wage work and informal self-employment (right) across local labor markets. The left panel plots the share of formal wage workers in each decile of the wage-bill HHI distribution across local labor markets. The right panel plots the average log of daily earnings in each decile and separately for formal wage and self-employed workers. The straight lines show the linear fit based on the underlying data.

Figure A.3: Electrification Program Implementation Across Districts



Notes. The figures shows the distribution of manufacturing employment share across districts reached vs. not reached by the electrification program in the first and last year of the IV estimation sample, i.e. 2004 and 2010.

Figure A.4: Cobb-Douglas and Population Shares in the Data

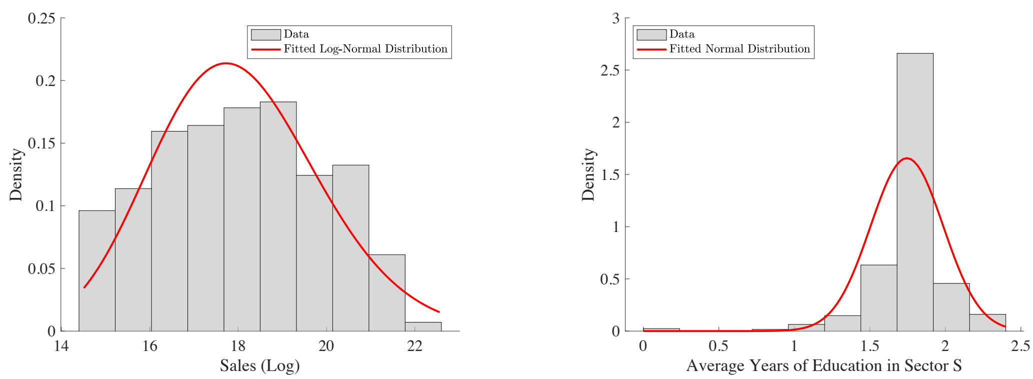


Notes. The figure displays histograms of aggregate sales and population shares across local labor markets, with shares normalized to average 1. The summary statistics are as follows: For the Cobb-Douglas shares $\tilde{\alpha}$, the mean is 0.09%. The largest Cobb-Douglas share is 2.1%, the 90th percentile is 0.33%, the median is 0.02%, and the 10th percentile is 0.002%. For the population shares $\tilde{L}_k$, the mean is 0.09%. The largest population share is 14.6%, the 90th percentile is 12.3%, the median is 0.10%, and the 10th percentile is 0.04%. The correlation coefficient from a regression of expenditure shares on population shares is 0.79, with a standard error of 0.20.

Figure A.5: Sales and Education in the Data – Fitted Distributions

(a) Variance

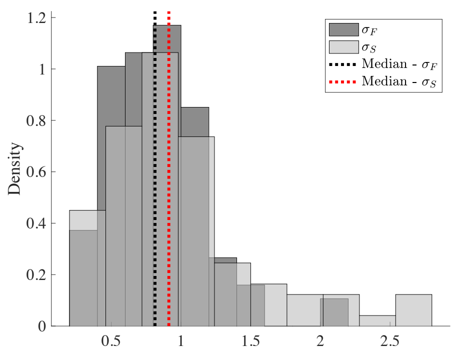
(b) Correlation



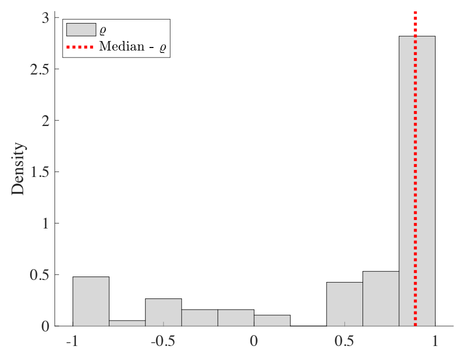
Notes. The figure displays the distributions of log sales and average years of education among self-employed workers across local labor markets, along with the fitted log normal and normal distribution, respectively.

Figure A.6: Parameter Estimates for Joint Ability Distribution

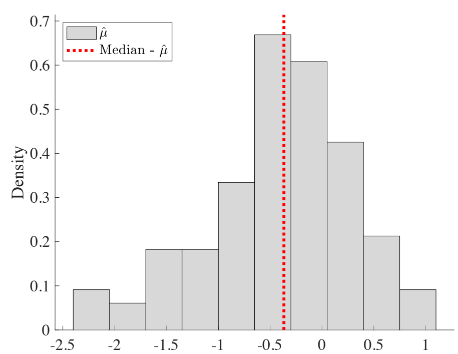
(a) Variance



(b) Correlation

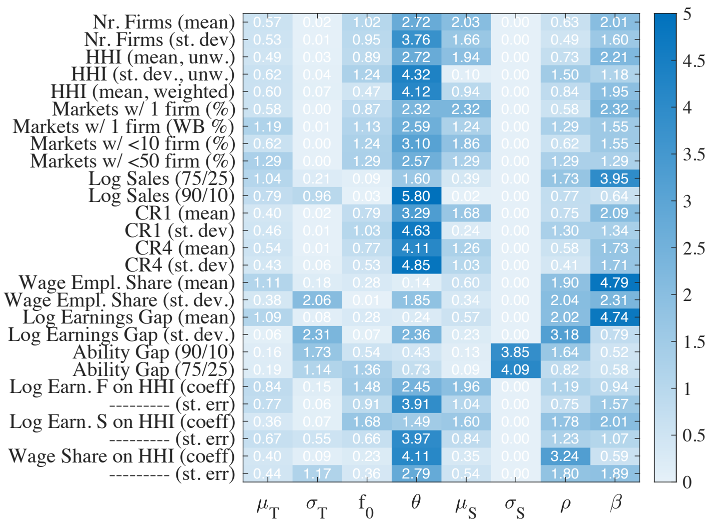


(c) Mean Comparative Advantage



Notes. The figure shows histograms of the parameter estimates for the ability distribution, as discussed in Section 4.3. Panel (a) presents histograms for the estimates of $\sigma_{F,k}$ and $\sigma_{S,k}$ across local labor markets. Panel (b) displays the histograms for the parameter ϱ_k, while panel (c) presents the estimates for $\hat{\mu}_k$. Each panel also includes the median estimate, which is used in the calibration of our baseline model.

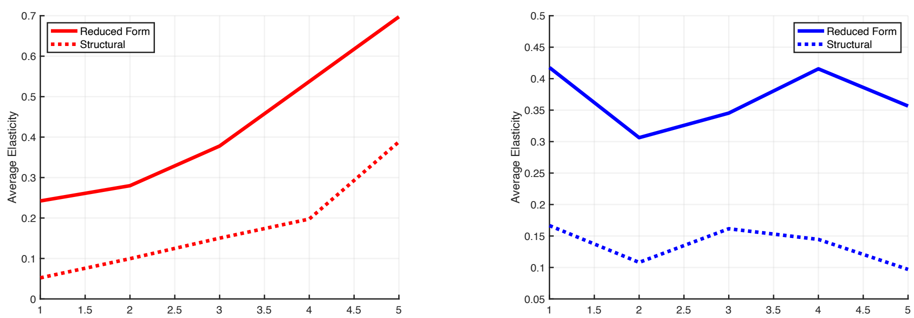
Figure A.7: Normalized Partial Derivatives of Moments with Respect to Parameters



Notes. The Jacobian matrix includes the normalized values of the elasticity of each moment i with respect to a 10% change in parameter j around its estimated value while keeping all the other parameters constant. Each row is a moment and each column is a parameter.

Figure A.8: Inverse Elasticity, Concentration, and Self-Employment Shares

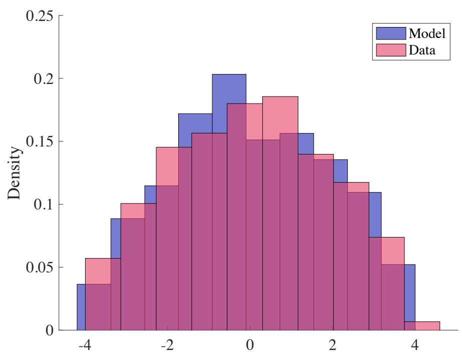
(a) HHI Quantiles (b) Self-Employment Quantiles



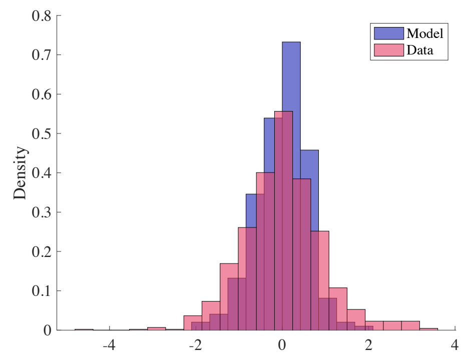
Notes. The figure presents the average reduced-form and structural inverse elasticity of wage labor across different quantiles of the wage-bill HHI and self-employment shares. Subfigure (a) shows the elasticity variation across HHI quantiles, while subfigure (b) shows the elasticity variation across self-employment quantiles. The solid red lines indicate reduced form elasticities, and the dotted red lines represent structural elasticities.

Figure A.9: Model Fit – Sales, Earnings Gap, and Number of Firms

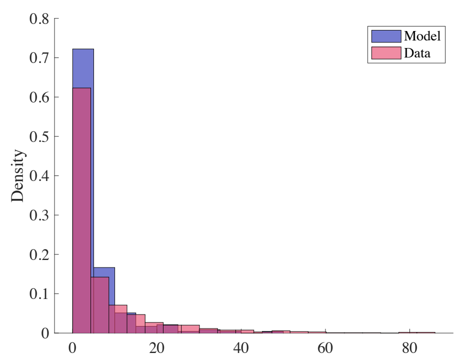
(a) Sales (Log)



(b) Earnings Gap (Log)



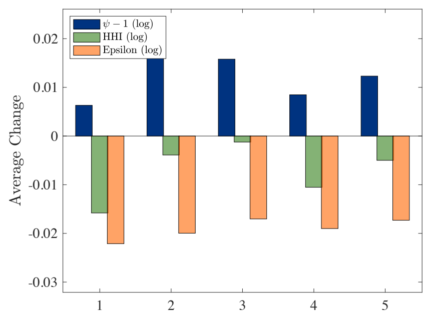
(c) Number of Firms



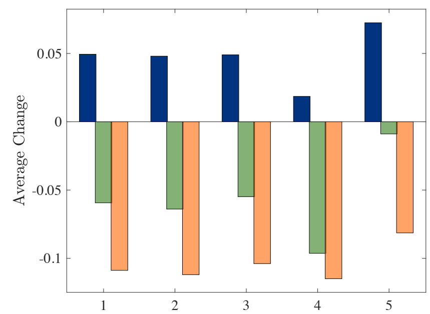
Notes. The figure shows histograms of log sales, log earnings gap, and number of firms, in the model and in the data.

Figure A.10: Effect of Policy Shocks on Wage Markdown and Its Determinants

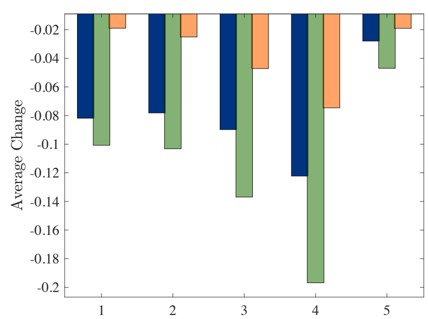
(a) Firm Productivity



(b) Worker Skills



(c) Fixed Cost



Notes. The three panels complement those in Figure 4 by illustrating the estimated change in wage markdown, concentration, and wage labor supply elasticity across local labor markets resulting from the three policy experiments. It does so for separate bins determined by the size of the wage markdown at baseline.

B Robustness of Stylized Facts

In this section, we present and discuss a series of checks to demonstrate that the empirical facts shown in Section 2 remain robust when using alternative measures of concentration, redefining self-employment, and redrawing local labor market boundaries.

Concentration. A primary concern is whether the *Encuesta Económica Anual* (EEA) firm-level dataset accurately reflects the population of firms that employ wage workers. The EEA is mandatory for—and is therefore a census of—firms with net sales above a certain threshold. To ensure consistency across years and account for changes in such threshold over time, we focus on manufacturing firms with yearly net sales exceeding 2 million Peruvian Soles—approximately 700,000 USD in 2010. Importantly, the EEA is a longitudinal survey, and the panel dimension is crucial for implementing the identification strategy used to estimate labor market power. This feature is lacking in alternative sources of firm-level data like the Economic Census.

We regard the use of EEA data and the decision to adopt a single sales threshold as inconsequential for the analysis for several reasons. First, according to 2007 Economic Census data, medium and large firms constitute the overwhelming majority of wage employment and sales in their respective local labor markets. Within manufacturing, firms with sales above the threshold account for 95% of total sales and 76% of total employment. Second, these choices do not introduce a systematic bias in the extent to which concentration varies across markets. Figure B.1 shows the distribution of wage-bill HHIs in the 2007 Economic Census and in our EEA data. The average unweighted HHI is lower in the Census data, due to the exclusion of smaller firms in the EEA. Yet, the average weighted HHI is very similar across the two. Most importantly, Figure B.2 and Table B.1 demonstrate that the two measures are strongly correlated, even after accounting for industry and location fixed effects.

A separate concern is about the use of wage-bill HHI as concentration measure. We also consider the employment HHI and the number of firms alternative measures, all defined in Section 2. Figure B.3 shows that the three measures are strongly correlated. Importantly, Tables B.2 and B.3 report the coefficient estimates that we obtain when using these measures instead of wage-bill HHI to investigate the correlation between concentration, self-employment rates and earnings. The results are very similar to the baseline ones in Table A.2.

Self-Employment. A possible concern regarding our definition of self-employment is that it includes both own-account workers and employers. First, note that employers are a small fraction of workers in the ENAHO data: out of the 40% of workers that we classify as self-employed in manufacturing, 31% are *own-account* workers while the remaining 9% are *employers*. Among these employers, only 14%—equivalent to about 1.2% of all workers in our data—have their businesses registered as legal entities and thus operate formally. Note that registration is a necessary condition for them to be included in the EEA, which is our source of data on medium to large firms.

Our decision to pool own-account workers and employers into a single self-employment category stems from the dual nature of manufacturing labor markets in poor countries, with wage employment at medium and large manufacturing firms on the one hand and self-employment on the other hand. The businesses operated by employers in this second category differ significantly from the larger firms that employ wage workers. Both workers and owners of these micro-enterprises earn low incomes, with minimal specialization between them, making these businesses resemble a group of self-employed individuals sharing a workspace rather than functioning as modern firms (Bassi et al., 2023; Atencio-De-Leon, Lee and Macaluso, 2023). These enterprises, captured in the ENAHO survey, represent the readily accessible, labor- and product-market price-taking self-employment opportunities that our model conceptualizes. For this reason, we chose to pool them with own-account self-employed workers.

We nonetheless investigate whether excluding employers from the sample affects our results. Table B.5 shows the summary statistics for manufacturing workers that we obtain when we exclude employers. Both average self-employment earnings and their dispersion are lower compared to Table 1, which is to be expected. Despite these differences, all the stylized facts hold when focusing solely on *own-account self-employment* only as the relevant alternative to wage work. This is illustrated in Figures B.6 and B.7, which closely resemble Figures 1 and 2. The same is true for the results in Table B.6, which mirror those in Table A.2.

All stylized facts are also robust to focusing exclusively on informal self-employment as the relevant alternative to wage work, as opposed to overall self-employment. This is not surprising considering that over 90% of self-employment is informal, also within manufacturing. The results are shown in Figure B.4, Figure B.5 and Table B.4.

Local Labor Markets. Another possible concern is with our definition of local labor markets, and specifically with whether the province or commuting zone boundaries are the appropriate geographical unit to consider. This is relevant considering that the model does not feature mobility of workers across markets. To address this concern, we consider a broader definition using 2-digit industries within departments as units of analysis, as opposed to the baseline definition of 2-digit industries within provinces or commuting zones. Departments are Level 1 administrative units, while provinces are Level 2. On average, the 24 departments (plus the Callao constitutional province) in the country contain around 9 provinces.

In Table B.7, we present summary statistics across local labor markets using the alternative definition. The number of local labor markets decreases from 280 to 179, with an average of about 10 medium to large firms in each, compared to 6 using the baseline definition. The unweighted average wage-bill HHI slightly increases to 0.68, and the percentage of markets with only one firm increases from 30% to 45%. However, the weighted concentration measures, either for payroll or employment, decrease by around one-third compared to the baseline (0.26

vs. 0.37 for wage-bill HHI and 0.21 vs. 0.31 for employment HHI). Figure B.8 displays the distributions of wage-bill HHI using both the baseline and broadened definition.

Figure B.9 mirrors Figure 1, plotting the average likelihood of transitioning into and from wage work and self-employment across the earnings distribution in the two sectors. In Figure 1, we group workers by deciles of the earning distribution within each local labor market as defined by a 2-digit industry within a province or commuting zone. In Figure B.9, we do the same but using the new broadened definition of a 2-digit industry within a department. Substantial variation exists in earnings within such large geographical areas. Despite the change in definition, the patterns in Figure B.9 are very similar to those in Figure 1.

Next, Figure B.10 mirrors Figure 2 and illustrates the relationship between employer concentration, the rate of self-employment (left), and earnings from both wage work and self-employment (right) across local labor markets, now defined as 2-digit industries within departments. Once again, the patterns are very similar to those found using the baseline definition. Higher concentration is consistently associated with higher self-employment rates and lower earnings from both wage work and self-employment. The results in Table B.8 show that these relationships still stand when controlling for individual characteristics, industry and department fixed effects—mirroring the baseline results in Table A.2.

Table B.1: Correlation Between Concentration Measures in Census and EEA Data

	EEA Wage-bill HHI			
	(1)	(2)	(3)	(4)
Census Wage-bill HHI	0.804***	0.845***	0.712***	0.715***
	(0.093)	(0.069)	(0.109)	(0.104)
	[8.60]	[12.19]	[6.56]	[6.85]
2-digit Industry FE	No	Yes	No	Yes
Location FE	No	No	Yes	Yes
Observations	194	194	169	169
R^2	0.500	0.618	0.634	0.735

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. The unit of observation is a local labor market as defined by a 2-digit industry within location, the latter corresponding to Peruvian provinces or commuting zones. The table reports the coefficient estimates and their standard errors obtained when regressing wage-bill HHI as obtained from the 2007 EEA data over the same variable obtained from the 2007 economic census, focusing on manufacturing firms. The standard errors and *t-statistics* associated with each estimate are reported in round and square brackets, respectively. Standard errors are clustered at the level of location in all specifications.

Table B.2: Concentration, Self-employment, and Earnings – Employment HHI

	Self-Employed {0,1}			Log of Earnings					
				Wage Workers			Self-employed		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Employer Concentration	0.053***	0.050***	0.063***	-0.084***	-0.096***	-0.045**	-0.122**	-0.146***	-0.032
(Log of Employment HHI)	(0.012)	(0.013)	(0.014)	(0.019)	(0.019)	(0.019)	(0.052)	(0.045)	(0.050)
Female	0.121***	0.121***	0.111***	-0.424***	-0.381***	-0.382***	-1.312***	-1.229***	-1.211***
	(0.023)	(0.023)	(0.023)	(0.041)	(0.036)	(0.035)	(0.072)	(0.075)	(0.078)
Age	0.018***	0.017***	0.016***	0.023**	0.027***	0.029***	0.116***	0.115***	0.109***
	(0.005)	(0.005)	(0.005)	(0.009)	(0.009)	(0.009)	(0.018)	(0.017)	(0.018)
Age sq.	-0.000	-0.000	-0.000	-0.000	-0.000**	-0.000**	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Schooling	-0.005	-0.001	-0.000	0.178***	0.162***	0.156***	0.111***	0.109***	0.102***
	(0.004)	(0.004)	(0.004)	(0.008)	(0.008)	(0.007)	(0.018)	(0.017)	(0.019)
Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	7637	7637	7634	4707	4706	4698	2054	2054	2047
R^2	0.104	0.133	0.156	0.308	0.363	0.395	0.327	0.382	0.399

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Unit of observation is a working-age individual surveyed in ENAHO. A local labor market k is defined by a 2-digit industry j within a province or commuting zone g. This table reports the coefficient estimates and their standard errors obtained when estimating the regression specification: $y_{i(j,g)t} = \beta \ln HHI^n_{(j,g)t} + \mathbf{X}'_{i(j,g)t}\theta + \gamma_j + \lambda_g + \delta_t + u_{i(j,g)t}$, where $y_{i(j,g)t}$ is the labor market outcome of worker i in local labor market k as defined by a manufacturing industry j within a province or commuting zone g in year t. The first regressor $\ln HHI^n_{(j,g)t}$ is the log of employment HHI in the market in the same year. $\mathbf{X}_{i(j,g)t}$ is a vector of individual characteristics, while γ_j, λ_g and δ_t stand for industry, location, and year fixed effects respectively. Standard errors are clustered at the local labor market level.

Table B.3: Concentration, Self-Employment, and Earnings – Number of Firms

	Self-Employed {0,1}			Log of Earnings					
				Wage Workers			Self-employed		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Employer Concentration	-0.035***	-0.033***	-0.041***	0.060***	0.068***	0.036***	0.083**	0.109***	0.039
(Log of Number of Firms)	(0.008)	(0.009)	(0.011)	(0.015)	(0.014)	(0.013)	(0.035)	(0.030)	(0.037)
Female	0.121***	0.120***	0.111***	-0.425***	-0.379***	-0.381***	-1.311***	-1.224***	-1.209***
	(0.023)	(0.023)	(0.023)	(0.041)	(0.036)	(0.035)	(0.072)	(0.075)	(0.079)
Age	0.018***	0.017***	0.016***	0.023**	0.028***	0.029***	0.114***	0.113***	0.109***
	(0.005)	(0.005)	(0.005)	(0.009)	(0.009)	(0.009)	(0.018)	(0.017)	(0.018)
Age sq.	-0.000	-0.000	-0.000	-0.000	-0.000**	-0.000**	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Schooling	-0.006	-0.001	-0.000	0.178***	0.162***	0.156***	0.110***	0.108***	0.102***
	(0.004)	(0.004)	(0.004)	(0.008)	(0.008)	(0.007)	(0.018)	(0.017)	(0.019)
Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	7637	7637	7634	4707	4706	4698	2054	2054	2047
R^2	0.102	0.132	0.156	0.308	0.363	0.395	0.327	0.383	0.399

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Unit of observation is a working-age individual surveyed in ENAHO. A local labor market k is defined by a 2-digit industry j within a province or commuting zone g. This table reports the coefficient estimates and their standard errors obtained when estimating the regression specification: $y_{i(j,g)t} = \beta \ln M_{(j,g)t} + \mathbf{X}'_{i(j,g)t}\theta + \gamma_j + \lambda_g + \delta_t + u_{i(j,g)t}$, where $y_{i(j,g)t}$ is the labor market outcome of worker i in local labor market k as defined by a manufacturing industry j within a province or commuting zone g in year t. The first regressor $\ln M_{(j,g)t}$ is the log of the number of firms in the market in the same year. $\mathbf{X}_{i(j,g)t}$ is a vector of individual characteristics, while γ_j, λ_g and δ_t stand for industry, location, and year fixed effects respectively. Standard errors are clustered at the local labor market level.

Table B.4: Concentration, Informal Self-Employment, and Earnings

	Informal Self-Employed {0,1}			Log of Earnings					
				Wage Workers			Self-employed		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Employer Concentration	0.049***	0.048***	0.056***	-0.085***	-0.100***	-0.052**	-0.101*	-0.143***	-0.025
(Log of Wage-bill HHI)	(0.012)	(0.014)	(0.015)	(0.020)	(0.021)	(0.021)	(0.053)	(0.045)	(0.050)
Female	0.145***	0.138***	0.129***	-0.426***	-0.381***	-0.382***	-1.233***	-1.169***	-1.161***
	(0.023)	(0.023)	(0.023)	(0.041)	(0.036)	(0.035)	(0.074)	(0.075)	(0.076)
Age	0.012***	0.011**	0.010**	0.023**	0.027***	0.029***	0.093***	0.095***	0.088***
	(0.005)	(0.004)	(0.005)	(0.009)	(0.009)	(0.009)	(0.019)	(0.019)	(0.020)
Age sq.	-0.000	-0.000	0.000	-0.000	-0.000**	-0.000**	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Schooling	-0.013***	-0.009**	-0.007**	0.178***	0.161***	0.156***	0.065***	0.063***	0.057***
	(0.004)	(0.004)	(0.003)	(0.008)	(0.008)	(0.007)	(0.018)	(0.018)	(0.019)
Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	7309	7309	7305	4707	4706	4698	1726	1725	1718
R^2	0.105	0.132	0.155	0.308	0.363	0.395	0.296	0.362	0.381

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Unit of observation is a working-age individual surveyed in ENAHO, excluding all formal self-employed workers. A local labor market k is defined by a 2-digit industry j within a province or commuting zone g. This table reports the coefficient estimates obtained when estimating the regression specification: $y_{i(j,g)t} = \beta \ln HHI^{wn}_{(j,g)t} + \mathbf{X}'_{i(j,g)t}\theta + \gamma_j + \lambda_g + \delta_t + u_{i(j,g)t}$, where $y_{i(j,g)t}$ is the labor market outcome of worker i in local labor market k as defined by a manufacturing industry j within a province or commuting zone g in year t. The first regressor $\ln HHI^{wn}_{(j,g)t}$ is the log of wage-bill HHI in the market in the same year. $\mathbf{X}_{i(j,g)t}$ is a vector of individual characteristics, while γ_j, λ_g and δ_t stand for industry, location, and year fixed effects respectively. Standard errors are clustered at the local labor market level.

Table B.5: Worker Summary Statistics – Excluding Employers

Variable	Mean	St. Dev.
Manufacturing Workers		
Wage Worker	0.61	0.49
Daily Wage	31.84	31.85
Own-Account Self-Employed	0.35	0.48
Daily Earnings from Own-Account Self-Employment	13.89	19.49
W-OAS Transition	0.05	0.21
OAS-W Transition	0.04	0.20

Notes. This table reports summary statistics from ENAHO worker-level data (Panel II), averaging across all years from 2004 to 2011, and excluding employers. Transition rates are obtained using the 2007-2011 panel version of ENAHO. Worker-level statistics are for dummy variables indicating wage work, self-employment, earnings (in PEN, 1 PEN $\approx$ 0.35 USD in 2010), and annual transitions from the wage- to self-employment sector (W-S) and vice versa (S-W).

Table B.6: Concentration, Own-Account Self-Employment, and Earnings

	Own-Account Self-Employed {0,1}			Log of Earnings					
					Wage Workers			Self-employed	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Employer Concentration	0.043***	0.045***	0.052***	-0.085***	-0.100***	-0.052**	-0.093*	-0.123***	0.036
(Log of Wage-bill HHI)	(0.012)	(0.013)	(0.014)	(0.020)	(0.021)	(0.021)	(0.051)	(0.042)	(0.052)
Female	0.167***	0.154***	0.143***	-0.426***	-0.381***	-0.382***	-1.217***	-1.178***	-1.176***
	(0.022)	(0.022)	(0.022)	(0.041)	(0.036)	(0.035)	(0.070)	(0.081)	(0.084)
Age	0.009**	0.008**	0.007	0.023**	0.027***	0.029***	0.095***	0.099***	0.094***
	(0.004)	(0.004)	(0.004)	(0.009)	(0.009)	(0.009)	(0.023)	(0.023)	(0.024)
Age sq.	0.000	0.000	0.000	-0.000	-0.000**	-0.000**	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Schooling	-0.012***	-0.008**	-0.008**	0.178***	0.161***	0.156***	0.060***	0.052***	0.039**
	(0.004)	(0.004)	(0.003)	(0.008)	(0.008)	(0.007)	(0.018)	(0.018)	(0.018)
Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	7003	7003	6999	4707	4706	4698	1470	1469	1463
R^2	0.122	0.147	0.170	0.308	0.363	0.395	0.294	0.349	0.369

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Unit of observation is a working-age individual surveyed in ENAHO, excluding all employers. A local labor market k is defined by a 2-digit industry j within a province or commuting zone g. This table reports the coefficient estimates obtained when estimating the regression specification: $y_{i(j,g)t} = \beta \ln HHI^{wn}_{(j,g)t} + \mathbf{X}'_{i(j,g)t}\theta + \gamma_j + \lambda_g + \delta_t + u_{i(j,g)t}$, where $y_{i(j,g)t}$ is the labor market outcome of worker i in local labor market k as defined by a manufacturing industry j within a province or commuting zone g in year t. The first regressor $\ln HHI^{wn}_{(j,g)t}$ is the log of wage-bill HHI in the market in the same year. $\mathbf{X}_{i(j,g)t}$ is a vector of individual characteristics, while γ_j, λ_g and δ_t stand for industry, location, and year fixed effects respectively. Standard errors are clustered at the local labor market level.

Table B.7: Employer Concentration Across Broadened Local Labor Markets

	Full Sample		*Broadened LLMs*	
	Mean	St. Dev.	Mean	St. Dev.
Number of Firms	6.39	10.37	10.24	22.83
Wage-bill HHI	0.65	0.33	0.68	0.35
Wage-bill HHI (Payroll Weighted)	0.37	0.03	0.26	0.03
Wage-bill HHI (Employment Weighted)	0.34	0.03	0.34	0.03
Employment HHI	0.63	0.35	0.66	0.36
Employment HHI (Payroll Weighted)	0.33	0.03	0.21	0.02
Employment HHI (Employment Weighted)	0.31	0.02	0.21	0.02
Percent of LLMs with 1 firm	38.78	2.27	44.65	2.58
Payroll Share of LLMs with 1 firm	7.94	1.79	6.65	1.72
Employment Share of LLMs with 1 firm	7.80	1.23	6.38	1.10
Number of Local Labor Markets	280		179	
Number of Locations	61		23	
Industries	23		23	

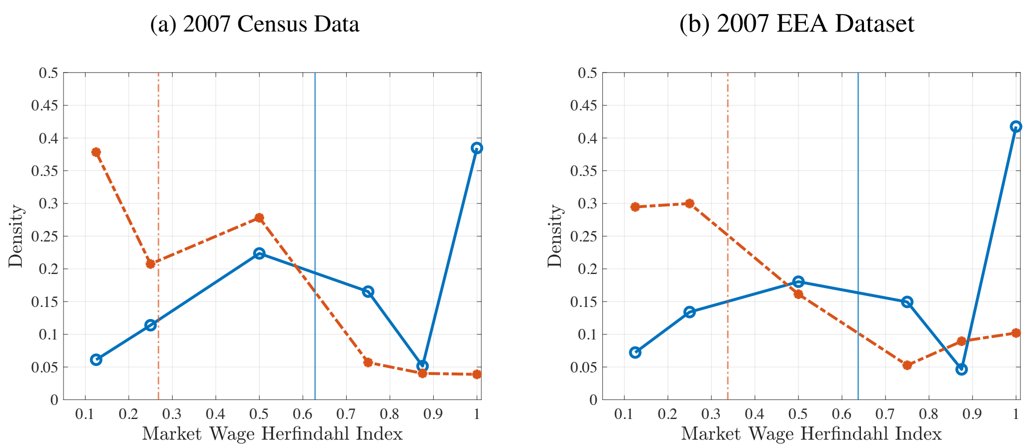
Notes. This table presents summary statistics and employer concentration measures derived from EEA firm-level data across Peruvian local labor markets, averaged over the years 2004 to 2011. The data are shown separately for the entire sample and the full sample where local labor markets are more broadly defined as 2-digit industries within Peruvian departments instead of provinces or commuting zones.

Table B.8: Concentration, Self-Employment, and Earnings – Broadened Local Labor Markets

	Self-Employed {0,1}			Log of Earnings					
				Wage Workers			Self-employed		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Employer Concentration	0.044***	0.050***	0.058***	-0.057***	-0.075***	-0.048**	-0.103**	-0.151***	-0.024
(Log of Wage-bill HHI)	(0.010)	(0.010)	(0.017)	(0.019)	(0.018)	(0.020)	(0.044)	(0.026)	(0.055)
Female	0.130***	0.130***	0.106***	-0.427***	-0.397***	-0.396***	-1.345***	-1.252***	-1.216***
	(0.026)	(0.026)	(0.027)	(0.045)	(0.039)	(0.037)	(0.074)	(0.073)	(0.074)
Age	0.023***	0.023***	0.019***	0.021**	0.025***	0.027***	0.098***	0.088***	0.094***
	(0.004)	(0.004)	(0.004)	(0.009)	(0.009)	(0.009)	(0.015)	(0.014)	(0.014)
Age sq.	-0.000***	-0.000***	-0.000*	-0.000	-0.000*	-0.000**	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Schooling	-0.015***	-0.011***	-0.002	0.180***	0.165***	0.156***	0.115***	0.119***	0.092***
	(0.004)	(0.004)	(0.003)	(0.009)	(0.007)	(0.007)	(0.019)	(0.019)	(0.017)
Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	9613	9613	9596	5573	5573	5541	2937	2937	2904
R^2	0.119	0.150	0.206	0.300	0.349	0.409	0.343	0.401	0.458

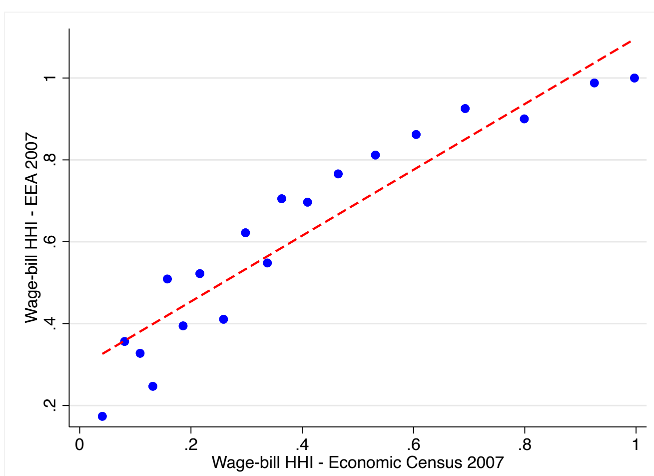
Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Unit of observation is a working-age individual surveyed in ENAHO. A local labor market k is defined by a 2-digit industry j within a department g. This table reports the coefficient estimates and their standard errors obtained when estimating the regression specification: $y_{i(j,g)t} = \beta \ln HHI^{wn}_{(j,g)t} + \mathbf{X}'_{i(j,g)t}\theta + \gamma_j + \lambda_g + \delta_t + u_{i(j,g)t}$, where $y_{i(j,g)t}$ is the labor market outcome of worker i in local labor market k as defined by a manufacturing industry j within a province or commuting zone g in year t. The first regressor $\ln HHI^{wn}_{(j,g)t}$ is the log of wage-bill HHI in the market in the same year. $\mathbf{X}_{i(j,g)t}$ is a vector of individual characteristics, while γ_j, λ_g and δ_t stand for industry, location, and year fixed effects respectively. Standard errors are clustered at the local labor market level.

Figure B.1: Employer Concentration Across Local Labor Markets in Census and EEA Data



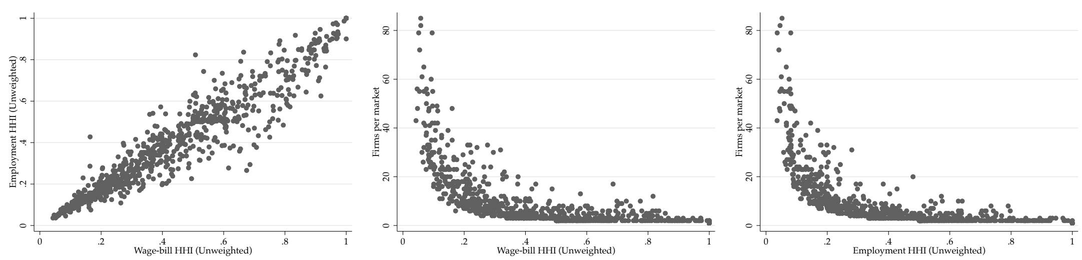
Notes. The figure plots the distribution of wage-bill HHI computed from the 2007 Peruvian Economic Census (left panel) and the same distribution computed from the 2007 EEA dataset (right panel) across local labor markets in the manufacturing sector. The blue solid line in both panels corresponds to the unweighted average, while the dashed line corresponds to the weighted average, where weights are given by the local labor market's share of nation-wide payroll.

Figure B.2: Correlation Between Concentration Measures in Census and EEA Data



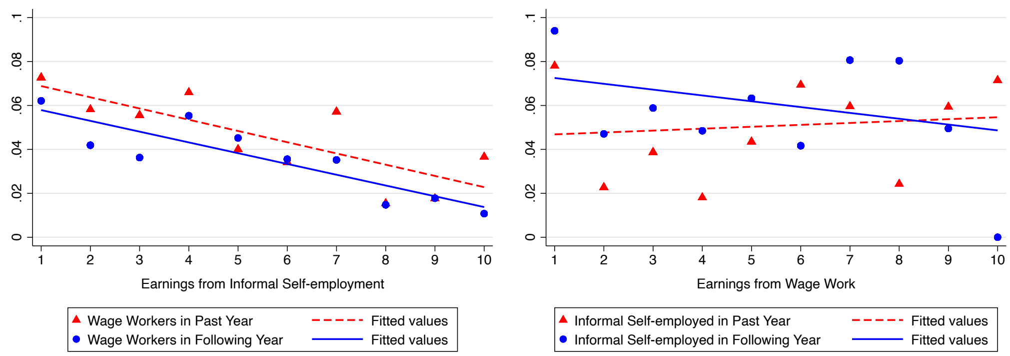
Notes. The figure illustrates the correlation between wage-bill HHI across local labor markets computed from the 2007 Peruvian Economic Census and the 2007 EEA dataset. Both variables are grouped into equal-sized bins, each point showing the average within bins. The dashed line shows the linear fit based on the underlying data, its slope equal to the corresponding parameter in the first column of Online Appendix Table B.1.

Figure B.3: Correlation Between Employer Concentration Measures



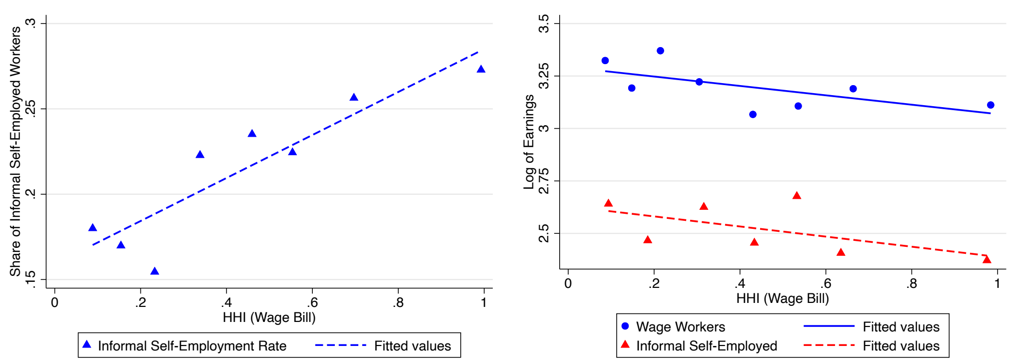
Notes. The figure plots the raw correlation of the three employer concentration measures – wage-bill HHI, employment HHI, and number of firms (bottom center panel) – one against the other across all local labor market-level observations. Wage-bill and employment HHI are strongly positively correlated and they are both strongly negatively correlated to the number of firms.

Figure B.4: Transitions Probabilities and Earnings – Informal Self-Employment



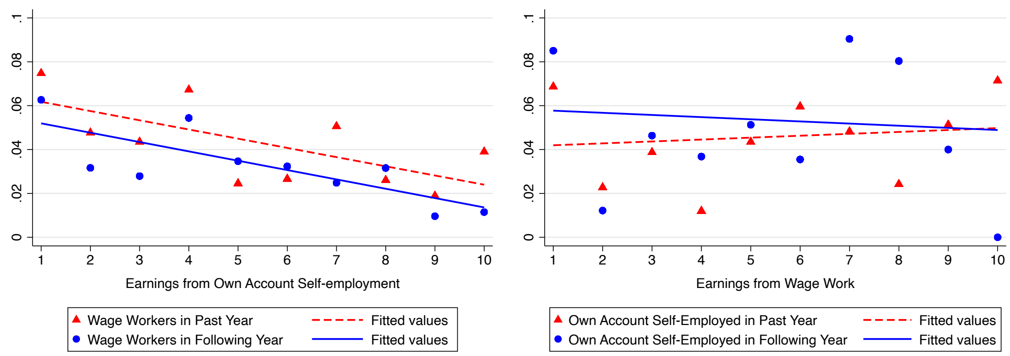
Notes. The figures illustrate the relationship between the likelihood of transitioning from and into wage work and informal self-employment, and earnings. The left panel plots average yearly transition probabilities into and from wage work across deciles of the informal self-employment earnings distribution. Similarly, the right panel plots average yearly transition probabilities into and from informal self-employment across the wage work earnings distribution deciles. The straight lines show the linear fit based on the underlying data.

Figure B.5: Concentration, Informal Self-Employment Rate, and Earnings



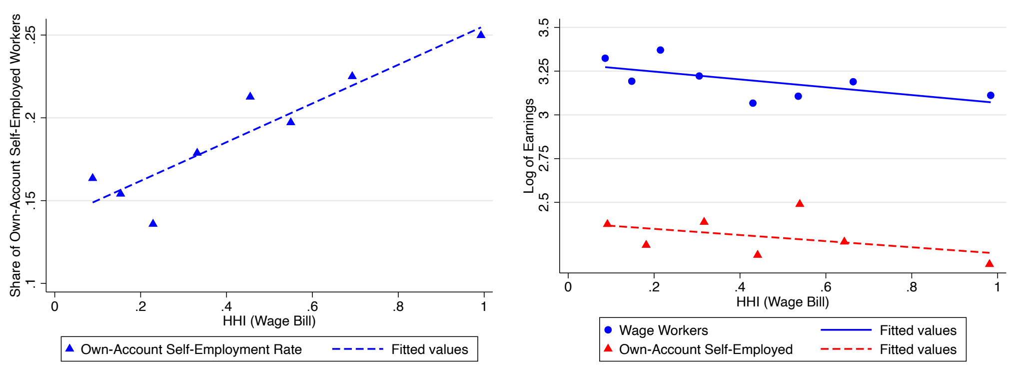
Notes. The figures illustrate the relationship between employer concentration, rate of informal self-employment (left), and earnings from both wage work and informal self-employment (right) across local labor markets. The left panel plots the share of informal self-employed workers in each decile of the wage-bill HHI distribution across local labor markets. The right panel plots the average log of daily earnings in each decile and separately for wage and informal self-employed workers. The straight lines show the linear fit based on the underlying data.

Figure B.6: Transitions Probabilities and Earnings – Own-Account Self-Employment



Notes. The figures illustrate the relationship between the likelihood of transitioning from and into wage work and own-account self-employment (thus excluding employers), and earnings. The left panel plots average yearly transition probabilities into and from wage work across deciles of the own-account self-employment earnings distribution. Similarly, the right panel plots average yearly transition probabilities into and from own-account self-employment across the wage work earnings distribution deciles. The straight lines show the linear fit based on the underlying data.

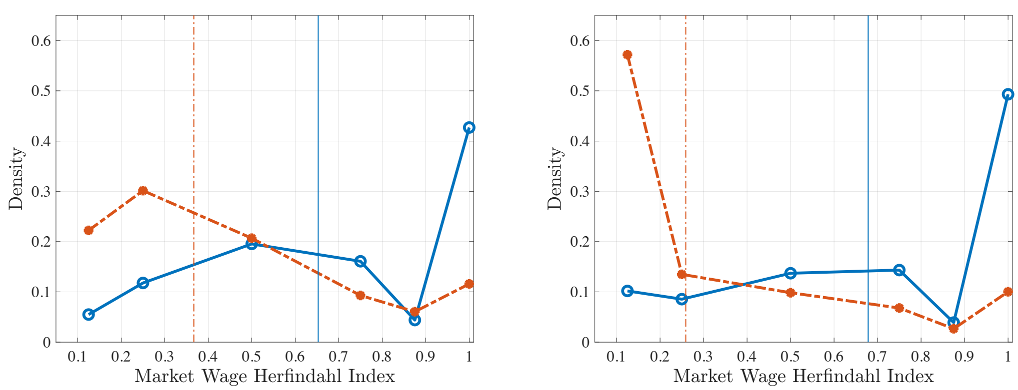
Figure B.7: Concentration, Own-Account Self-Employment Rate, and Earnings



Notes. The figures illustrate the relationship between employer concentration, rate of informal self-employment (left), and earnings from both wage work and informal self-employment (right) across local labor markets. The left panel plots the share of informal self-employed workers in each decile of the wage-bill HHI distribution across local labor markets. The right panel plots the average log of daily earnings in each decile and separately for wage and informal self-employed workers. The straight lines show the linear fit based on the underlying data.

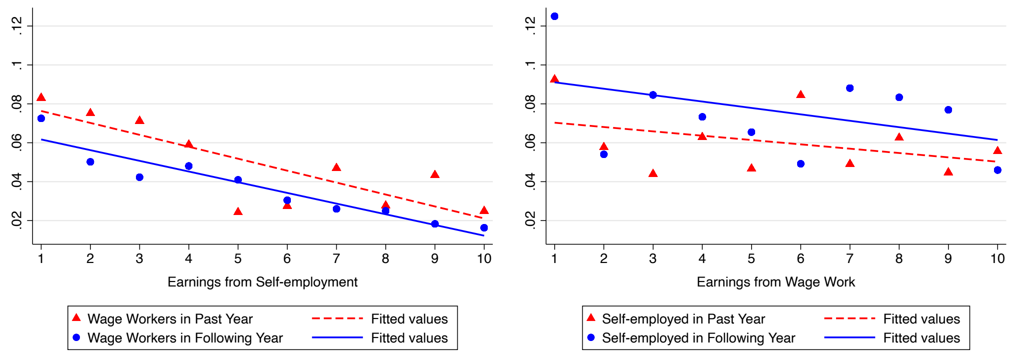
Figure B.8: Employer Concentration Across Local Labor Markets – Baseline vs. Broadened

(a) Baseline (b) Broadened



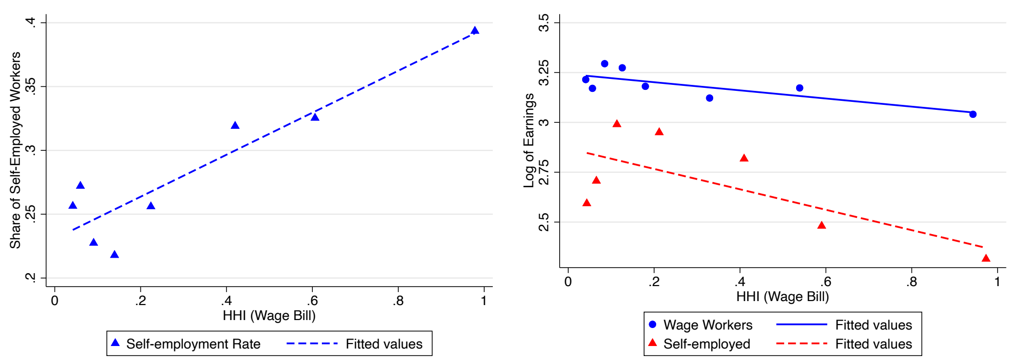
Notes. The figure plots the distribution of wage-bill HHI computed from the EEA dataset using the baseline (left panel) and broadened (right panel) definition of local labor markets. At baseline, local labor markets are defined by a 2-digit industry within a province or commuting zone. In the broadened definition, they are 2-digit industries within a department. The blue solid line in both panels corresponds to the unweighted average, while the dashed line corresponds to the weighted average, where weights are given by the local labor market's share of nation-wide payroll.

Figure B.9: Transition Probabilities and Earnings – Broadened Local Labor Markets



Notes. The figures illustrate the relationship between the likelihood of transitioning from and into wage work and self-employment, and earnings, where the latter are residuals from a regression of daily earnings over the full set of province or commuting zone fixed effects. The left panel plots average yearly transition probabilities into and from wage work across deciles of the self-employment earnings distribution within local labor markets as defined by 2-digit industries within departments. Similarly, the right panel plots average yearly transition probabilities into and from self-employment across the wage work earnings distribution deciles derived at the same level. The straight lines show the linear fit based on the underlying data.

Figure B.10: Concentration, S.-E. Rate, and Earnings – Broadened Local Labor Markets



Notes. The figures illustrate the relationship between employer concentration, rate of self-employment (left), and earnings from both wage work and self-employment (right) across local labor markets as defined by 2-digit industries within departments. The left panel plots the share of self-employed workers in each decile of the wage-bill HHI distribution across local labor markets. The right panel plots the average log of daily earnings in each decile and separately for wage and self-employed workers. The straight lines show the linear fit based on the underlying data.

C Theory Appendix

This section provides further details on the theory. Section C.1 describes the approach to solving the model's general equilibrium (GE). Section C.2 explores the implications of assuming a log-normal distribution for workers' ability, a restriction applied in our empirical analysis.

C.1 Model Solution

With segmented labor markets, interactions across markets occur solely through changes in expenditures $Y_k = \alpha_k Y$, where $\{\alpha_k\}_{k\in(0,1)}$ are the constant expenditure shares. Consequently, given Y, the equilibrium in each market can be determined independently of the others.

This feature of the model allows decomposing its solution into a market equilibrium component and a general equilibrium component. The market equilibrium refers to the process of solving for equilibrium in each local labor market given Y. Each market equilibrium, in turn, provides a value for Y based on market-clearing conditions. The final step involves determining the general equilibrium Y by solving the corresponding fixed-point algorithm.

C.1.1 Market Equilibrium

Let $\mathbf{\Lambda}_k \equiv \{s_{iF,k}, s^N_{iF,k}, \mu_{iF,k}, \psi_{iF,k}\}_{i=1}^{M_k}$ represent the vector of output and employment shares, markups, and markdowns for each active firm in market k. Let $\mathbf{K}$ denote the vector of number of entrants, relative wages, and $\mathbf{\Lambda}_k = \left\{M_k, \hat{W}_k, \mathbf{\Lambda}_k\right\}_{k\in(0,1)}$ in each k. A market equilibrium consists of the vector $\hat{\mathbf{K}}$ such that, given a value for Y, expenditure shares $\{\alpha_k\}_{k\in(0,1)}$, and model primitives $\left\{\{z_{iF,k}\}_{i=1}^{\bar{M}^*_k}, f^e_k, G_k\right\}$, all model equations (4)-(15) are satisfied.

Equilibrium given M_k Let's first assume that the number of entrants M_k is known in each k. From equations (10) and (11), we have:

$$\frac{Y_{F,k}}{Y_{S,k}} = Z_{F,k}\hat{N}_k(\hat{W}_k), \tag{C.1}$$

where $\hat{N}_k(\hat{W}_k) \equiv \frac{N_{F,k}}{N_{S,k}}$ is the relative labor supply, which is a known function of $\hat{W}_k$, given $G_k(\cdot)$, and $Z_{F,k} \equiv \left(\sum_{i=1}^{M_k} s_{iF,k}^{\frac{\eta}{\eta-1}} z_{iF,k}^{-1}\right)^{-1}$ is the productivity index for sector F in market k. With CES preferences, the left-hand side of (C.1) is equal to:

$$\frac{Y_{F,k}}{Y_{S,k}} = \zeta^\rho \left(\frac{P_{F,k}}{P_{S,k}}\right)^{-\rho}. \tag{C.2}$$

Using the pricing rule from (14) and aggregating across firms, we find:

$$P_{F,k} = \left(\sum_{i=1}^{M_k} (p_{iF,k})^{1-\eta} \right)^{\frac{1}{1-\eta}} = \Phi_{F,k} W_{F,k},$$

where $\Phi_{F,k} \equiv \left(\sum_{i=1}^{M_k} \left(\frac{\mu_{iF,k}\psi_{iF,k}}{z_{iF,k}} \right)^{1-\eta} \right)^{\frac{1}{1-\eta}}$ is a market-level index reflecting the aggregate effects of productivity, markups, and markdowns. The self-employment sector is competitive, so aggregate prices reflect marginal cost: $P_{S,k} = W_{S,k}$. Combining these, we obtain:

$$\hat{N}_k(\hat{W}_k) \left(\hat{W}_k \right)^{\rho} = \zeta^{\rho} \left(\Phi_{F,k} \right)^{-\rho} Z_{F,k}^{-1}. \tag{C.3}$$

Equation (C.3) represents the first equilibrium block. The unknowns in this equation are the relative wage $\hat{W}_k$ and the market-level indices $Z_{F,k}$ and $\Phi_{F,k}$, which depend on the vector $\mathbf{\Lambda}_k \equiv \{s_{iF,k}, s^N_{iF,k}, \mu_{iF,k}, \psi_{iF,k}\}_{i=1}^{M_k}$ of shares, markups and markdowns.

The second equilibrium block is defined by the following expressions for firms' market shares, markups, and markdowns, which together form a fixed-point problem given $\hat{W}_k$:

$$s_{iF,k} = \left(\frac{p_{iF,k}}{P_{F,k}} \right)^{1-\eta} = \frac{\left(\frac{\mu_{iF,k}\psi_{iF,k}}{z_{iF,k}} \right)^{1-\eta}}{\sum_{i=1}^{M_k} \left(\frac{\mu_{iF,k}\psi_{iF,k}}{z_{iF,k}} \right)^{1-\eta}}, \tag{C.4}$$

$$\mu_{iF,k} = \frac{\varepsilon_{iF,k}}{\varepsilon_{iF,k} - 1}, \quad \text{where} \quad \varepsilon_{iF,k} = \left[\frac{1}{\eta}(1 - s_{iF,k}) + \frac{1}{\rho} s_{iF,k} \right]^{-1}, \tag{C.5}$$

$$\psi_{iF,k} = \left(\frac{s^N_{iF,k}}{\epsilon_{F,k}(\hat{W}_k)} + 1 \right), \quad \text{with} \quad s^N_{iF,k} = \frac{s_{iF,k}^{\frac{\eta}{\eta-1}} (z_{iF,k})^{-1}}{\sum_{i=1}^{M_k} s_{iF,k}^{\frac{\eta}{\eta-1}} (z_{iF,k})^{-1}}. \tag{C.6}$$

Given these expressions, we can now outline an algebraic algorithm to solve for the market equilibrium. Specifically, given M_k and market-level draws $\{z_{iF,k}\}_{i \in [1, M_k]}$, the equilibrium in market k consists of a relative wage $\hat{W}_k$ and a vector $\mathbf{\Lambda}_k$ such that:

1. Given $\hat{W}_k$, $\mathbf{\Lambda}_k$ solves the fixed-point problem defined by equations (C.4)-(C.6).
2. Given $\mathbf{\Lambda}_k$, the wage $\hat{W}_k$ solves equation (C.3).

The market equilibrium can be found by applying the implied iterative fixed-point procedure.

Solving for M_k As is standard in the literature, we solve the entry problem by considering a sequential entry game where shadow firms with higher productivity draws move first. The equilibrium number of entrants in each market can be determined using the following iterative procedure. For each candidate number of entrants, $M_k = 1, .., \bar{M}_k^*$, we find the equilibrium $(\hat{W}_k, \mathbf{\Lambda}_k)$ using the procedure outlined above. We then compute the profits of the marginal entrant $\underline{i}$ in market k as:

$$\pi_{\underline{i}F,k} = s_{\underline{i}F,k}\gamma_{F,k}Y_k\left(1 - \frac{1}{\mu_{\underline{i}F,k}\psi_{\underline{i}F,k}}\right) - f^e_{\underline{i},k},$$

where $f^e_{\underline{i},k}$ is the entry cost for firm $\underline{i}$, which depends on the firm's ranking of entry, and $\gamma_{F,k}$ is the expenditure share on sector F goods, which solves:

$$\left(\frac{\gamma_{F,k}}{1-\gamma_{F,k}}\right) = \zeta^{\rho}\left(\hat{W}_k\right)^{1-\rho}\Phi^{\rho-1}_{F,k}.$$

An equilibrium of the entry game is achieved when the equilibrium profits for the marginal entrant $\underline{i}$ are positive, while those for any additional entrant would be negative. With sequential entry, this entry game has a unique cutoff equilibrium, meaning that only firms with productivity above a certain cutoff enter the market.

Simplified Entry Game Solving for the exact equilibrium values of M_k can be computationally intensive, as it requires solving for $(\hat{W}_k, \mathbf{\Lambda}_k)$ for each candidate value, which involves a fixed-point algorithm. To mitigate this complexity, we adopt a simplified entry game approach, inspired by Gaubert and Itskhoki (2021), where firms are treated as 'naive' at the entry stage. Specifically, we assume that upon entry, firms expect to charge the minimum markup and markdown as if they were infinitesimal. For markups, this means setting $\mu_i = \frac{\eta}{\eta-1}$; for markdowns, it implies $\psi_{iF,k} = 1$ for all i. Moreover, under the assumption that all firms behave atomistically, the market shares simplify to:

$$s_{iF,k} = \frac{(z_{iF,k})^{\eta-1}}{\sum_i (z_{iF,k})^{\eta-1}},$$

and the market index $\Phi_{F,k}$ becomes:

$$\Phi_{F,k} = \left(\frac{\eta-1}{\eta}\right)Z_{F,k}, \quad \text{where} \quad Z_{F,k} = \left[\sum_i (z_{iF,k})^{\eta-1}\right]^{\frac{1}{\eta-1}}.$$

As a result, profits simplify to:

$$\pi_{\underline{i}F,k} = s_{\underline{i}F,k}\gamma_{F,k}\frac{\alpha_k Y}{\eta} - f^e_{\underline{i}}, \tag{C.7}$$

where $\gamma_{F,k}$ is given by:

$$\frac{\gamma_{F,k}}{1-\gamma_{F,k}} = \zeta^{\rho}\left(\frac{\eta}{\eta-1}\right)^{1-\rho} Z^{\rho-1}_{F,k}\left(\hat{W}_k\right)^{1-\rho}, \tag{C.8}$$

and where $\hat{W}_k$ can be found by simple inversion, solving the 'simplified' equilibrium condition:

$$\hat{W}_k^{\rho} \hat{N}_k = \left(\frac{\eta}{\eta - 1} \right)^{-\rho} \zeta^{\rho} Z_{F,k}^{\rho - 1}. \tag{C.9}$$

The number of entrants M_k can be determined using the iterative procedure described above with these simplified expressions.

C.2 Sorting and Log Normality

This section examines the properties of the model when the joint ability distribution $G_k(a_F, a_S)$ is log-normally distributed, as specified in equation (24). For simplicity, we focus on a single market k and omit the market-level subscript unless needed.

C.2.1 Aggregate Wage Labor Supply Elasticity

We begin by providing sufficient conditions for the wage work supply elasticity to be a decreasing function of the relative wage $\hat{W}$. Without loss of generality, let $L = 1$. The aggregate supply of wage work $N_F(\hat{W})$ and its log can be expressed as:

$$\begin{aligned} N_F(\hat{W}) &= \Pr(h \in \text{wage sector}) \times \mathbb{E}\left(a_F \mid a_F \hat{W} \geqslant a_S \right), \\ \ln N_F(\hat{W}) &= \ln \Pr(h \in \text{wage sector}) + \ln \mathbb{E}\left(a_F \mid a_F \hat{W} \geqslant a_S \right). \end{aligned} \tag{C.10}$$

We want to find conditions for $\epsilon_F(\hat{W}) \equiv \frac{\partial \ln N_F}{\partial \ln \hat{W}}$ to decrease with $\hat{W}$, which is equivalent to show that $\ln N_F$ is a concave function of $\ln \hat{W}$, i.e. $\frac{\partial^2 \ln N_F}{\partial \ln \hat{W}^2} < 0$.

Under log-normality we have

$$\Pr(h \in \text{wage sector}) = \Phi\left(\frac{\ln \hat{W} + \hat{\mu}}{\sigma^*} \right) = \Phi(c_F), \tag{C.11}$$

with $\sigma^* \equiv \sqrt{\sigma_F^2 + \sigma_S^2 - 2\varrho\sigma_F\sigma_S}$ and $\hat{\mu} \equiv \mu_F - \mu_S$.

Following Heckman and Sedlacek (1985) and knowing that $\ln \mathbb{E}(x) \approx \mathbb{E}(\ln x) + \frac{1}{2}\text{Var}(\ln x)$, we also get:

$$\begin{aligned} \ln \mathbb{E}\left(a_F \mid a_F \hat{W} \geqslant a_S \right) &\approx \mathbb{E}\left(\ln a_F | a_F \hat{W} \geqslant a_S \right) + \frac{1}{2}\text{Var}\left(\ln a_F | a_F \hat{W} \geqslant a_S \right) \\ \ln \mathbb{E}\left(a_F \mid a_F \hat{W} \geqslant a_S \right) &\approx \mu_F + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*} \right) \lambda(c_F) + \frac{1}{2}\left\{ \sigma_F^2 + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*} \right)^2 \lambda'(c_F) \right\} \end{aligned} \tag{C.12}$$

where $\lambda(x) \equiv \phi(x)/\Phi(x)$ is a decreasing and convex function of x.

Let $\alpha = \frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}$. Substituting (C.11) and (C.12) into equation (C.10) and taking the derivative with respect to $\ln \hat{W}$ we get the wage labor supply elasticity:

$$\epsilon_F(\hat{W}) \equiv \frac{\partial \ln N_F}{\partial \ln \hat{W}} = \frac{1}{\sigma^*}\left[\lambda(c_F) + \alpha\lambda'(c_F) + \frac{1}{2}\alpha^2\lambda''(c_F)\right] > 0, \tag{C.13}$$

This is positive for any given α. When $\alpha \leq 0$, since $\lambda(\cdot), \lambda''(\cdot) > 0$, and $\lambda'(\cdot) \in (-1, 0)$, the expression is always positive. When $\alpha > 0$, for any given c_F the expression has a minimum at $\alpha = -\frac{\lambda'(c_F)}{\lambda''(c_F)}$, and is always positive when evaluated at that minimum.

Finally, taking the second derivative we get:

$$\frac{\partial^2 \ln N_F(\hat{W})}{\partial \ln \hat{W}^2} = \frac{1}{(\sigma^*)^2}\left[\lambda'(c_F) + \alpha\lambda''(c_F) + \frac{1}{2}\alpha^2\lambda'''(c_F)\right]. \tag{C.14}$$

which can be shown to be negative if $\alpha \in (-1, 1)$, thus ruling out cases of particularly extreme negative or positive selection (see Section C.2.2). This is because it is negative for $\alpha = -1$ and $\alpha = 1$ at any given value of c_F. It is also monotone increasing in α for any $c_F \leq 1$ since $-\frac{\lambda''(\cdot)}{\lambda'''(\cdot)} \leq -1 < \alpha$. For all $c_F > 1$, the term in squared brackets is bounded between $\lambda'(1) + \alpha\lambda''(1)$ and zero and always negative for $\alpha < -\frac{\lambda'(1)}{\lambda''(1)} \approx 1$. Hence the following proposition.

Proposition C.1. *When the joint ability distribution is log-normal and $\alpha \in (-1, 1)$, the aggregate supply elasticity of wage labor $\epsilon_F(\hat{W})$ is positive and decreases with the relative efficiency unit wage $\hat{W}$.*

Note that our estimates of $(\sigma_F, \sigma_S, \varrho, \hat{\mu}) = (0.81, 0.91, 0.89, -0.12)$ imply $\alpha \approx 0$. Equations (C.10)-(C.13) also imply a one-to-one negative relationship between $\epsilon_F(\hat{W})$ and the equilibrium self-employment share. Denoting the self-employment share as χ_S, we have:

$$1 - \chi_S = \Phi(c_F) \Leftrightarrow c_F = \Phi^{-1}(1 - \chi_S).$$

Since the function $\Phi(x)$ is monotone increasing, its inverse function $\Phi^{-1}(x)$ is also monotone increasing. As the self-employment share χ_S increases, $\Phi^{-1}(1 - \chi_S)$ decreases, implying that $\epsilon_F(\hat{W})$ increases. Hence the following proposition.

Proposition C.2. *When the joint ability distribution is log-normal and $\alpha \in (-1, 1)$, the aggregate supply elasticity of wage labor $\epsilon_F(\hat{W})$ is a one-to-one increasing function of the self-employment share.*

C.2.2 Ability Distribution and Sectoral Earnings

In Section 3.3, we argued that the scope and sign of the selection channel depend on the parameters of the workers' ability distribution, particularly those governing absolute and comparative advantage. Here, we illustrate this point using the log-normal case.

Following Heckman and Sedlacek (1985), the mean average ability in each sector, defined as the log endowment of efficiency units of labor, can be written as:

$$
\begin{aligned}
A_F &\equiv \mathbb{E}\left(\ln a_F \mid a_F \hat{W} \geqslant a_S\right) = \mu_F + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_F), \\
A_S &\equiv \mathbb{E}\left(\ln a_S \mid a_F \hat{W} < a_S\right) = \mu_S + \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_S),
\end{aligned}
\tag{C.15}
$$

where $c_S = -c_F$ and all other parameters have been defined in Section C.2.1.

Now, consider a shock ϑ to the economic environment that lowers the relative wage per efficiency unit $\hat{W}$, thereby shrinking the wage employment sector. Given the system in (C.15), we can express the response of average ability in the two sectors as:

$$
\begin{aligned}
\frac{dA_F}{d\vartheta} &= \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right) \cdot \frac{d\lambda(c_F)}{dc_F} \cdot \frac{dc_F}{d\vartheta}, \\
\frac{dA_S}{d\vartheta} &= \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right) \cdot \frac{d\lambda(c_S)}{dc_S} \cdot \frac{dc_S}{d\vartheta}.
\end{aligned}
\tag{C.16}
$$

By construction, $\frac{dc_F}{d\vartheta} < 0$ and $\frac{dc_S}{d\vartheta} > 0$, implying that $\frac{d\lambda(c_F)}{dc_F} < 0$ and $\frac{d\lambda(c_S)}{dc_S} < 0$. This indicates that the signs of $\frac{dA_F}{d\vartheta}$ and $\frac{dA_S}{d\vartheta}$ depend solely on $\left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)$ and $\left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)$, respectively.

If the two abilities are uncorrelated ($\varrho = 0$) or negatively correlated ($\varrho < 0$), it follows that $\frac{dA_F}{d\vartheta} > 0$ and $\frac{dA_S}{d\vartheta} < 0$. In this case, average ability will decrease in the wage employment sector and increase in self-employment as the relative wage $\hat{W}$ increases.

Conversely, if $\varrho > 0$ and $\sigma_F^2 < \varrho\sigma_F\sigma_S < \sigma_S^2$, then $\left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right) < 0$ and $\left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right) > 0$, implying $\frac{dA_F}{d\vartheta} < 0$ and $\frac{dA_S}{d\vartheta} < 0$. This means that if abilities are more dispersed in self-employment compared to the wage employment sector, and provided that ϱ is positive and sufficiently high, average ability will increase in both the wage employment and self-employment sectors as the relative wage $\hat{W}$ increases.

D Estimation Appendix

This section outlines the model's estimation strategy. Section D.1 covers the identification and direct estimation of the Roy parameters, focusing on the parameters of the variance-covariance matrix Σ_k and the relative mean comparative advantage $\hat{\mu}_k$. Section D.2 provides details on the identifying variation of remaining parameters using the Method of Simulated Moments. Section D.3 describes the computational algorithm used to solve the model given estimated parameters. Section D.4 provides details on non-targeted model fit.

D.1 Identification of Ability Distribution Parameters

D.1.1 Variance-Covariance Matrix

We focus on a single market k and omit the market-level subscript unless needed. We denote the share of workers in the wage sector as c_1, and write it as:

$$c_1 \equiv \Pr(h \in \text{wage sector}) = \Phi\left(\frac{\ln \hat{W} + \hat{\mu}}{\sigma^*}\right). \tag{D.1}$$

with $\sigma^* \equiv \sqrt{\sigma_F^2 + \sigma_S^2 - 2\varrho\sigma_F\sigma_S}$. Following Heckman and Sedlacek (1985), the mean log earnings in the wage employment and self-employment sectors, which we denote as c_3 and c_4, can be expressed as:

$$\begin{aligned} c_3 &\equiv \mathbb{E}\left(\ln a_F W_F \mid a_F\hat{W} \geq a_S\right) = \ln W_F + \mu_F + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_F), \\ c_4 &\equiv \mathbb{E}\left(\ln a_S W_S \mid a_F\hat{W} \leq a_S\right) = \ln W_S + \mu_S + \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_S), \end{aligned} \tag{D.2}$$

and the corresponding variances, denoted as c_5 and c_6, as:

$$\begin{aligned} c_5 &\equiv \text{Var}\left(\ln a_F W_F \mid a_F\hat{W} \geq a_S\right) = \sigma_F^2 + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)^2\left[-\lambda(c_F)c_F - \lambda^2(c_F)\right], \\ c_6 &\equiv \text{Var}\left(\ln a_S W_S \mid a_F\hat{W} \leq a_S\right) = \sigma_S^2 + \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)^2\left[-\lambda(c_S)c_S - \lambda^2(c_S)\right], \end{aligned} \tag{D.3}$$

where c_F is as defined in equation (C.11), and $c_S = -c_F$.

The variables $c_1, .., c_6$ can be observed for each local labor markets where a cross section of workers' earnings across the two sectors is available. Similarly, equation (C.11) shows that we can easily recover the terms c_F (and c_S) from simple inversion of the observed employment shares in the two sectors, from which we can also get $\lambda(c_F), \lambda(c_S)$.

From simple algebra, one can derive the following system of equations holding for each market:

$$\begin{cases} c_1 & = \Phi(c_F), \\ c_3 - c_4 & = \sigma^* c_F + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_F) - \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(-c_F), \\ c_5 & = \sigma_F^2 + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)^2 \left[\lambda(c_F)c_F - \lambda^2(c_F)\right], \\ c_6 & = \sigma_S^2 + \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)^2 \left[-\lambda(-c_F)c_F - \lambda^2(-c_F)\right]. \end{cases} \tag{D.4}$$

The one above is a system of $4\times\bar{K}$ equations in $4\times\bar{K}$ unknowns $(c_{F,k}, \sigma_{F,k}, \sigma_{S,k}, \varrho_k)$, where $\bar{K}$ is the number of local labor markets in our data where earnings data are available for both sectors. It follows that observing multiple individuals in each sector in a given market k will suffice for the identification of the parameter vector $\mathbf{\Theta}_k = (\sigma_{F,k}, \sigma_{S,k}, \varrho_k)$ in that market. We recover the vector $\mathbf{\Theta}_k = (\sigma_{F,k}, \sigma_{S,k}, \varrho_k)$ from a constrained Minimum Distance Estimation (MDE) procedure, where we restrict the variance coefficients to be non-negative and the correlation parameter to be $\varrho_k \in [-1, 1]$.

D.1.2 Mean Comparative Advantage

We now address the identification of the mean comparative advantage $\hat{\mu}$. From Equations (C.15), we derive the following expression for the relative mean log abilities across sectors:

$$\begin{aligned} \mathbb{E}\left(\ln\frac{a_F}{a_S} \mid a_F\hat{W} \geq a_S\right) &\equiv \mathbb{E}\left(\ln a_F \mid a_F\hat{W} \geq a_S\right) - \mathbb{E}\left(\ln a_S \mid a_F\hat{W} \leq a_S\right) \\ &= \hat{\mu} + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_F) - \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_S). \end{aligned} \tag{D.5}$$

While the terms involving the variance-covariance parameters are known, the left-hand side of (D.5) is unobserved, hindering the identification of $\hat{\mu}$. To overcome this, we use years of education as a proxy for abilities and assume the average log ability in sector I relates to average log education as $\mathbb{E}(\ln a_I \mid \cdot) = \delta + \beta\mathbb{E}(\ln \text{edu}_I) + \varepsilon_I$, where δ is market-specific. This gives:

$$\mathbb{E}\left(\ln\frac{a_F}{a_S} \mid a_F\hat{W} \geq a_S\right) = \beta\mathbb{E}\left(\ln\frac{\text{edu}_F}{\text{edu}_S}\right) + \varepsilon. \tag{D.6}$$

Here, β represents the elasticity of ability with respect to education, assumed constant across markets and sectors, while ε is a zero-mean i.i.d. error term. By examining sectoral differences, we control for market-level factors affecting the education-ability mapping.

The left-hand-side of equation D.6 can be decomposed as:

$$\mathbb{E}\left(\ln\frac{a_F}{a_S} \mid a_F\hat{W} \geq a_S\right) = \mathbb{E}\left(\ln\frac{N_F}{N_S} \mid a_F\hat{W} \geq a_S\right) - \mathbb{E}\left(\ln\frac{Em_F}{Em_S} \mid a_F\hat{W} \geq a_S\right), \tag{D.7}$$

which expresses the mean (log) ability gap as the difference between the mean (log) effective labor supply gap and relative employment in the two sectors. For log-normal distributions, the mean (log) effective labor supply gap can be expressed as:

$$\mathbb{E}\left(\ln\frac{N_F}{N_S} \mid a_F\hat{W} \geq a_S\right) = \ln\frac{N_F}{N_S} - \frac{1}{2}\left(c_5 - c_6\right), \tag{D.8}$$

where c_5 and c_6 are the observed variances of log earnings in sectors F and S.

Combining equations (D.6)-(D.8), we obtain:

$$\ln\frac{N_F}{N_S} = \alpha + \beta^*\mathbb{E}\left(\ln\frac{\bar{\text{edu}}_F}{\bar{\text{edu}}_S}\right) + \gamma_1\left(\ln\frac{c_1}{1-c_1}\right) + \gamma_2\left(c_5 - c_6\right) + u, \tag{D.9}$$

where $\beta^* = \beta\frac{\rho-1}{\rho}$ and u represents measurement error.

While the left-hand-side is unobserved, we can use our model's equation to write it as:[1]

$$\ln\frac{N_F}{N_S} \propto \frac{\rho}{\rho-1}\ln\left(\frac{\text{Revenues}_F}{\text{Earnings}_S}\right) - \ln Z_F. \tag{D.10}$$

We also consider an alternative specification where we measure the left hand side as:[2]

$$\ln\frac{N_F}{N_S} \propto \frac{\rho}{\rho-1}\ln\left(\frac{\text{Earnings}_F}{\text{Earnings}_S} + \ln\Phi_F\right) - \ln Z_F. \tag{D.11}$$

Assuming u is uncorrelated with education, this provides a consistent estimate of $\hat{\beta} = \hat{\beta}^*\frac{\rho}{\rho-1}$ for a given choice of ρ.[3] With $\hat{\beta}$ estimated, we impute the mean log ability gap from equation (D.6) and finally derive $\hat{\mu}$ using equation (D.5) market by market.

Robustness A potential concern with our approach is that the estimate of β may be biased if (i) there is measurement error in the earnings data or (ii) the orthogonality assumption is violated. While considering two alternative methods to measure $\ln\frac{N_F}{N_S}$ partially addresses the

[1]Our model implies:

$$\begin{aligned}\frac{N_F}{N_S} &= \frac{Y_F}{Y_S}\left(Z_F\right)^{-1} \\ &= \zeta^{-\frac{\rho}{\rho-1}}\left(\frac{R_F}{R_S}\right)^{\frac{\rho}{\rho-1}}\left(Z_F\right)^{-1},\end{aligned}$$

where $\frac{Y_F}{Y_S}$ is relative output across sectors, which under the CES structure of demand, can be expressed as a function of revenues. In the self-employment sector, total revenues coincide with self-employment earnings ($R_S = \text{Earnings}_S$). Taking logs yields equation (D.10).

[2]The rationale for this approach is that since firm revenues are measured from the firm survey and self-employment earnings from the worker survey, the variable $\ln\frac{\text{Revenues}_F}{\text{Earnings}_S}$ could be measured with error. We therefore express $\text{Revenues}_F = \text{Earnings}_F \cdot \Phi_F$ using the model's structure, leading to equation (D.11). This approach reduces measurement error by measuring relative earnings only with worker-level data. The disadvantage is that the wedge Φ_F is not observed, so that we need to rely on additional proxy controls to mitigate its potential confounding effects.

[3]For implementation, we use a $\rho = 2.7$, equal to the estimated value using the MSM method.

first concern, we also consider the simpler case of setting $\beta = 1$, imposing

$$\mathbb{E}\left(\ln \frac{a_F}{a_S} \mid a_F \hat{W} \geq a_S\right) = \mathbb{E}\left(\ln \frac{\text{edu}_F}{\text{edu}_S}\right).$$

While more restrictive, this approach is less susceptible to measurement error and omitted variable bias.

D.1.3 Results

Figure A.6 presents histograms of the estimated variance-covariance parameters and mean comparative advantage. Panel (a) shows the distributions of $\hat{\sigma}_F$ and $\hat{\sigma}_S$ across local labor markets, panel (b) depicts the histogram of the correlation parameter $\hat{\varrho}$, and panel (c) illustrates the distribution of the estimated mean comparative advantage $\hat{\mu}$. While there is some heterogeneity across markets, these estimates are generally well-behaved and centered around their means.

As explained in the main text, we set the $\boldsymbol{\Sigma_k}$ parameters and $\hat{\mu}$ constant across markets. Panel I of Table A.8 reports their median values. When computing the median of the $\boldsymbol{\Sigma_k}$ parameters, we exclude markets where the constrained minimum-distance procedure failed to find an interior solution to limit the influence of outliers. We verify that this exclusion does not significantly affect the correlation patterns between absolute and comparative advantage.

For $\hat{\mu}$, the table provides estimates from the three alternative cases discussed earlier. We calibrate the baseline model using the average of their median values.

D.2 Sensitivity of Model Moments to Parameters

To examine the relationship between model parameters and the generated moments, we follow a method similar to that of Kaboski and Townsend (2011), computing the sensitivity of each moment to each model parameter. The process involves the following steps:

1. We begin with the estimated vector of parameters, denoted by Φ^*, and create 18 alternative parameter vectors. For each parameter j, we generate two variations: one where Φ_j is reduced by 5%, $\Phi^- = \{\Phi^*_{-j}, 0.95 \cdot \Phi^*_j\}$, and one where it is increased by 5%, $\Phi^+ = \{\Phi^*_{-j}, 1.05 \cdot \Phi^*_j\}$. In both cases, all other parameters remain unchanged.
2. Using these adjusted parameter vectors, we then simulate the model to obtain the corresponding moment vectors. For each parameter change, we calculate the difference in each moment r, denoted as $\Delta_{jr} = m_r(\Phi^+) - m_r(\Phi^-)$. This difference quantifies how moment r changes when parameter j is altered by 10%, while keeping the other parameters fixed.

To facilitate comparison across moments, we normalize Δ_{jr} for each parameter such that, after rounding, the sum of the values across all parameters equals 9. This normalization creates an interpretable scale: if all parameters have an equal influence on a particular moment, the corresponding row in the matrix will show a value of 1 for each parameter. Alternatively, if only three parameters significantly affect a moment and their impacts are equal, each will show a value of 3, with the rest being 0, and so on.

The resulting Jacobian matrix, displayed in Figure A.7, clearly highlights the parameters that most strongly affect each moment. This intuitive mapping links specific parameters to the moments we would expect, as detailed in Section 4.2.2.

D.3 Full Estimation Algorithm

This section outlines the algorithm used to solve the model given the estimated parameters.

1. Using the estimated Roy parameters, Σ and $\hat{\mu}$, derive the functional expressions for labor supply and labor supply elasticity for a benchmark '0'-market, where we set $\mu_S = 0$.
2. For given parameter values $(\mu_T, \sigma_T, \theta)$, (f_0, f_1), (μ_μ, σ_μ), and (ρ, η, ζ), draw K local labor market productivities T_k from a log-normal distribution with parameters (μ_T, σ_T). Similarly, draw K mean absolute advantage parameters μ_k from a log-normal distribution with parameters (μ_μ, σ_μ). The seed for all random draws remains constant during estimation.[4]
3. For given values of parameter θ and realization of T_k in each market $k = 1, .., K$, we draw productivities of potential entrants $\{z_{iF,k}\}_{i=1}^{\bar{M}}$ as follows. We follow Eaton, Kortum and Sotelo (2012) and draw the productivity of the most productive firm and each firm thereafter, with spacings following an exponential distribution. Specifically, denote $U_k^{(n)} \equiv T_k z_{F,k}^{(n)-Z_\theta}$, where n is the rank of the firm in market k. Then $U_k^{(1)}$, $(U_k^{(2)} - U_k^{(1)})$, $(U_k^{(3)} - U_k^{(2)})$, etc., are i.i.d. exponential with cdf $G_U(u) = 1 - e^{-u}$ (Eaton, Kortum and Sotelo, 2012). We use the transformation to convert the exponential draws into productivity draws $\{z_{iF,k}\}_{i=1}^{\bar{M}}$. We cap the number of shadow firms $\bar{M}$ at 85, which is the maximum number of firms observed in the data.
4. With the calibrated value of local labor market shares and populations $\{\alpha_k, L_k\}_{k=1}^{K}$, the normalization $P = 1$, and given the functional forms for $\hat{N}(\cdot)$ and $\epsilon_F(\cdot)$, draws of $\{T_k, \mu_{\mu_k}, \{z_{iF,k}\}_{i=1}^{\bar{M}}\}_{k=1}^{K}$, and the remaining model parameters, we implement the following fixed point procedure:
   (a) Take an initial guess for aggregate income Y_0, which completes the general equilibrium vector $\mathbf{X} = (Y, 1)$.
   (b) Given $\mathbf{X}$, we solve for the market equilibrium $\mathbf{K} = \{\mathbf{M}, \hat{\mathbf{W}}, \mathbf{\Lambda}\}$, as detailed in Appendix C.1.1.
   (c) Given $\mathbf{K}$, use the parameters $\mu_{F,k}$ and $\mu_{S,k}$ to obtain the corresponding labor supply functions for each market by affinity with the functions of the benchmark market. Note that the market equilibrium is not affected by the specific value of these parameters, as it only depends on $\hat{N}$ and ϵ_F, which only depend on Σ and $\hat{\mu}$.

[4] To avoid mechanical correlations between the different distributions, we use separate random seed values for each distribution and verify that the correlations are close to zero.

(d) Given $\mathbf{K}$, use the general equilibrium conditions to solve for the new values of Y.

(e) Update the initial values of Y_0 taking the midpoint between the initial vector from step (a) and the new vector from step (c), and loop over until convergence.

(f) Upon convergence of the equilibrium vector $(\mathbf{X}, \mathbf{K})$, simulate the model and calculate the moment vector $\{m_k(\Phi)\}_{k=1}^{K}$ for all markets $k = 1, .., K$, corresponding to parameter vector Φ.

5. On a grid for parameters Φ with 100,000 points, evaluate the moment function $m_k(\Phi)$, with moments described in Table 4, and the associated MSM loss function:

$$\mathcal{L}(\Phi) = \hat{\mathbf{f}}(\Phi)' \mathbb{W} \hat{\mathbf{f}}(\Phi),$$
$$\text{where} \quad \hat{\mathbf{f}}(\Phi) \equiv f(m_k(\Phi)) - f(\tilde{m}_k)$$

and where $\tilde{m}_k$ are the values of the moments in our empirical dataset, the function $f(\cdot)$ is the simple average: $f(x_k) = K^{-1} \sum_k x_k$, and $\mathbb{W}$ is the weighting matrix, which we chose to be diagonal and inversely proportional to $\tilde{\mathbf{m}}$.[5] We use a Halton sequence to define the grid points, so that it covers the whole parameter space more efficiently than if points were regularly spaced.

6. With the results from the first Halton grid, we recompute a second finer Halton grid of 50,000 points. We restrict this grid to be wide enough to encompass the 50 best fitting parameter values of the previous grid, but exclude the regions with the highest loss function. We iterate this procedure several times, until convergence to a narrow region of the parameter space.

7. We take as our estimate (the global minimizer) the point of local convergence with the lowest loss function, $\hat{\Phi} = \arg\min_{\Phi} \ \mathcal{L}(\Phi)$.

D.4 Model Fit

To assess the model's fit, we replicate the reduced-form elasticities from Table 2 by simulating the shock to firm productivity and labor demand used in Section 2.5.

We proceed in three steps. First, we identify the counterpart of treated firms in the model. We extend the model by assigning an electricity wedge (τ) to all firms, as estimated from the data. Specifically, we restrict the sample to the baseline year and use OLS to estimate the parameters from the following specification:

$$\tau_{i(j,g)t_0} = \beta_\tau z_{i(j,g)t_0} + e_{i(j,g)t_0},$$

where $\tau_{i(j,g)t_0}$ represents the electricity wedge estimated at the firm level, as detailed in Section 2.5, $z_{i(j,g)t_0}$ denotes value added per worker, and $e_{i(j,g)t_0}$ is a normally distributed i.i.d. shock.

[5]Because of the poor fit in matching the correlation coefficient mgintudes between concentration and log earnings, we downweight these moment sby a factor of 30 to improve overall precision.

We estimate $\hat{\beta}_\tau = 0.08$, significant at the 1% level.

In the model, we then calculate the associated $\hat{\tau}_{ik}$ for each firm as

$$\hat{\tau}_{ik} = \hat{\beta}_\tau z_{ik} + e_{ik},$$

where z_{ik} is the firm-level productivity draw in local labor market k and $e_{ik} \sim \mathcal{N}(0,1)$. Once each firm is assigned a wedge, we consider a firm treated if their $\hat{\tau}_{ik}$ is greater than the economywide median, mirroring the strategy described in Section 2.5.

Second, from the data, we back up the size of the electrification shock in terms of productivity improvements—specifically, the average productivity increase at the firm level induced by the treatment. We do this by estimating using OLS the parameters from the following specification:

$$\ln z_{i(j,g)t} = \gamma PER_{gt} \times EC_{i(j,g)} + \phi_i + \delta_{(j,g)t} + v_{i(j,g)t}, \tag{D.12}$$

which is akin to the specification in equation (3) but has the log of value added per worker as dependent variable. We estimate $\gamma = 0.0019$, significant at the 5% level. Given that the mean of the treatment ($PER_{gt} \times EC_{i(j,g)}$) is approximately 12, the average treatment effect is an increase of 2.3% in firm-level productivity.

Third, starting from the baseline model equilibrium, we simulate a 2.3% productivity shock for the treated firms. We then obtain the firm-level inverse elasticity by taking the ratio between the (log) wage and the (log) employment responses $\hat{\epsilon}_{iF,k} \equiv \Delta \ln W_{F,k}/\Delta \ln n_{iF,k}$. We do this for treated firms only and in markets with more than one firm, consistent with the within-market identification strategy and the LATE nature of the estimates in Table 2.